cule and its biological implications were discussed in talks by F. H. C. Crick and J. D. Watson, and studies on the biosynthesis of nucleic acid were presented by Ellis Bolton. These four papers are reproduced in the following pages. The symposium was concluded by two additional informative papers: "Genetic Functions of Desoxyribonucleic Acids in Bacterial Transformations," by Rollin D. Hotchkiss, Rockefeller Institute for Medical Research, and "Nucleic Acid and Viral Growth," by A. D. Hershey, Carnegie Institution of Washington, Cold Spring Harbor.

CHEMICAL STRUCTURE OF THE NUCLEIC ACIDS

BY ALEXANDER R. TODD

CHEMICAL LABORATORY, UNIVERSITY OF CAMBRIDGE

Communicated by Linus Pauling, May 7, 1954

During the last four years a remarkable transformation has occurred in our knowledge of the nucleic acid structure, and there is reason to believe that many of the major features may now be regarded as established. It is the purpose of this brief review to outline the main developments which have taken place and to indicate the chemical evidence upon which they rest.

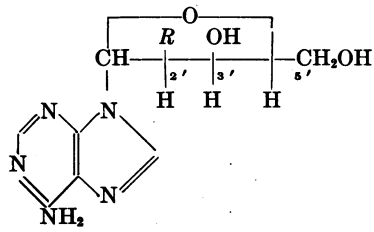
The known nucleic acids are of two types—the ribonucleic acids, yielding upon total hydrolysis the sugar D-ribose, and the deoxyribonucleic acids, yielding 2-deoxy-D-ribose. Graded hydrolysis of nucleic acids gives first the nucleotides and then the nucleosides. From ribonucleic acids, four nucleosides—adenosine, guanosine, uridine, and cytidine—are obtained, and from deoxyribonucleic acids, six—deoxyadenosine, deoxyguanosine, deoxycytidine, thymidine, 5-methyl-deoxycytidine, and 5-hydroxymethyl-deoxycytidine—of which the two last-named appear to be of limited occurrence. Schematically, the hydrolytic breakdown of the nucleic acids may be represented thus:

Nucleic acids
↓
Nucleotides
↓
Nucleosides + Phosphoric acid
↓
Purines or pyrimidines + D-ribose or 2-deoxy-D-ribose

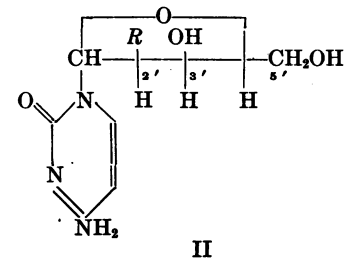
Ever since it was realized that the nucleotides are phosphates of the nucleosides and that the nucleic acids can be regarded as polynucleotides, many attempts have been made to formulate general structures for the latter. Titrimetric studies by Levene and Simms[1] led to a general concept of nucleic acid structure in which the individual nucleoside residues were joined together by phosphodiester linkages, and eliminated from consideration ether or pyrophosphate linkages. This simple formulation has become generally accepted, and, although other types of linkage have been suggested from time to time, such suggestions lack corroborative chemical evidence. Of these suggestions, the phosphotriester linkage, which might represent a branching point in a polynucleotide, has perhaps been most widely canvassed, and

it will be mentioned later in connection with the branching problem. Essentially, however, the nucleic acids may be regarded as polydiesters of phosphoric acid.

It is clear that an understanding of the details of nucleic acid structure must rest upon a knowledge of the precise structure of the individual nucleosides, and of the position, as well as the nature, of the internucleotidic linkage. These problems my colleagues and I in Cambridge have endeavored to study, using the methods of organic chemistry, and I believe that a consideration of the development of our studies and of related biochemical work affords the simplest exposition of nucleic acid structure as we know it today. The earlier work of Levene, in particular, established some of the general features of the group and simplified our studies at least in their earlier phases. As a preliminary phase in our work, we extended earlier studies by establishing in every detail the structure and stereochemical configuration of the natural ribonucleosides and confirmed our findings by total synthesis. They are in each case β-D-ribofuranosides, the sugar residue being attached at N_9 in the purine nucleosides and at N_3 in the pyrimidine nucleosides;[2] typical examples are adenosine (I in the accompanying diagram; R = OH), and cytidine (II; R = OH). The total synthesis of the deoxyribonucleosides has not yet been achieved, but their structure as β-2-deoxy-D-ribofuranosides with the sugar attached at N_9 in the purine and N_3 in the pyrimidine nucleosides has been rigidly established in the cases of deoxyadenosine (I; R = H), deoxycytidine (II; R = H), and thymidine.[3] In the case of deoxyguanosine the β-configuration, although allotted by analogy with the others, is almost certainly correct.



I

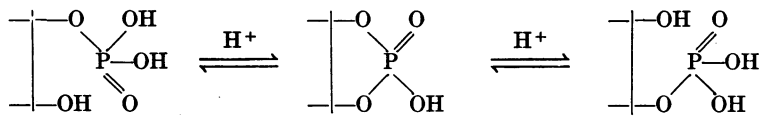


II

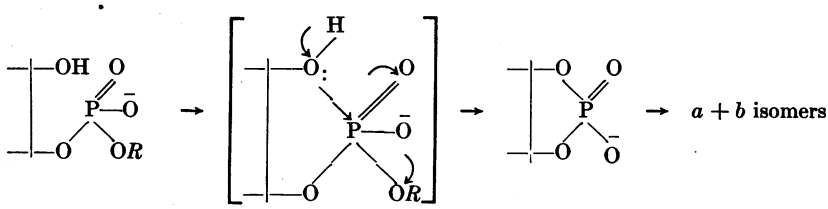
The next step in the direction of the determination of nucleic acid structure was clearly the synthesis of the various simple nucleotides derivable from the nucleosides by the introduction of phosphate groups into the glycosidic portion of the molecule. For this purpose a number of new and flexible methods of phosphorylation were devised,[2] the most successful involving the use of dibenzyl phosphorochloridate $[(C_6H_5CH_2O)_2POCl]$[4] as phosphorylating agent. Using it with appropriately protected intermediates, the unambiguous synthesis of all the ribonucleoside-5′-phosphates[5] was effected, as well as of the main deoxyribonucleoside-3′- and -5′-phosphates[6] (except for the deoxyguanosine phosphates, whose synthesis has still to be completed). The unambiguous synthesis of the 2′- and 3′-phosphates of the ribonucleosides proved a much more difficult undertaking, partly because of the difficulty of obtaining ribonucleosides blocked simultaneously in the 3′- and 5′-positions and partly because of phosphoryl migration, which is discussed below. Brown and Todd[7] were able to prepare adenosine-2′-phosphate and adenosine-3′-phosphate by phosphorylating 5′-trityladenosine, but they were unable at the time to say which was the 2′- and which the 3′-compound; the same

was, of course, true of other ribonucleoside 2′- and 3′-phosphates prepared in the same way from the appropriate 5′-trityl derivatives.

Until 1949 it had been believed that alkaline hydrolysis of ribonucleic acids yielded only four nucleotides, which were regarded on evidence, now known to be of doubtful validity, as the 3′-phosphates of the four nucleosides. In that year, however, Carter and Cohn[8] applied ion-exchange chromatography to alkaline hydrolyzates of ribonucleic acids and showed that they contained not four but eight nucleotides made up of four pairs of isomers—an *a* and a *b* nucleotide corresponding to each of the four ribonucleosides. Since it is clear, on the basis of evidence of many sorts, that the four *a* nucleotides all have the phosphate residue in the same position and the *b* nucleotides likewise, it simplifies matters to discuss the problem in one specific case—the adenylic acids *a* and *b*—bearing in mind that what is said here about this pair applies equally to each of the other pairs. The synthetic 2′- and 3′-phosphates of adenosine were shown to be identical with adenylic acids *a* and *b*, although, as already mentioned, it could not then be decided which was 2′- and which was 3′-phosphate. Brown and Todd[7] observed that, although quite stable in alkali, acid solutions of either isomer undergo phosphoryl migration, giving an equilibrium mixture of both. This interconversion of the *a* and *b* nucleosides can be written formally as follows:



Even more interesting was the behavior of simple monoesters of the *a* and *b* nucleotides. Diesters of phosphoric acid are normally very resistant toward alkaline hydrolysis. The esters of the *a* and *b* nucleotides are, however, very labile to alkali, undergoing hydrolysis with loss of the ester group and formation in each case of a mixture of the three *a* and *b* nucleotides.[7] This ease of hydrolysis is also shown by esters of the glycerophosphoric acids, and it is clear that in such cases the structural feature causing alkali-lability in a diester of phosphoric acid is the presence of a *cis*-hydroxyl group on the carbon atom adjacent to the point of attachment of the phosphate residue in one of the esterifying groups. The process of hydrolysis postulated is indicated schematically below. The formation of the cyclic phosphate as an intermediate has been demonstrated, and such cyclic phosphates have been synthesized[9] in the nucleotide series and shown to have the expected properties.



It should be noted here that the expulsion of the esterifying group *R* is an inevitable consequence of the cyclization step—one of the esterifying groups clearly must be removed if the phosphorus atom is to retain its normal valency, and mere

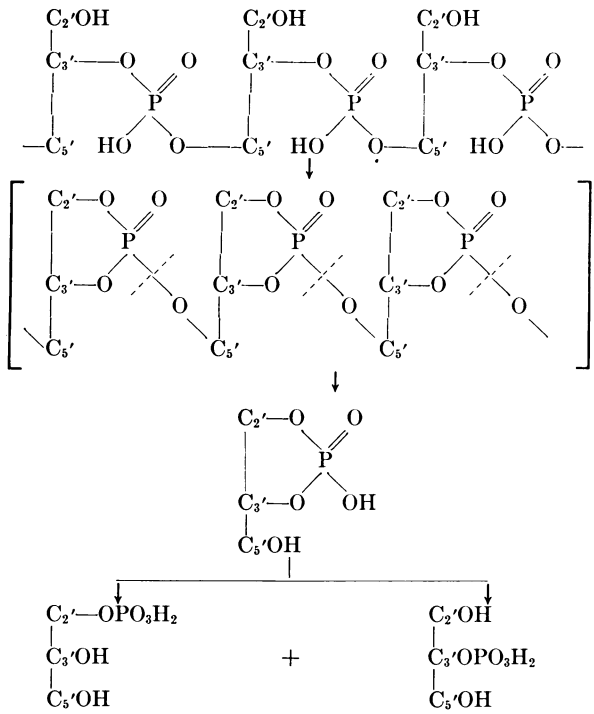
rupture of either linkage with the nucleoside residue would not give any degradation of the molecule; a neutral cyclic triester of phosphoric acid is not an intermediate in the process (although it may at times be convenient to write the intermediate as such in formal schematic representations of the process in complex molecules). The production of a mixture of *a* and *b* nucleotides upon hydrolysis of the cyclic phosphate is to be expected, since the two ester linkages in the latter are effectively equivalent.

Although the structural theory of the ribonucleic acids was developed during the period when the terms *a* and *b* had to be used for the isomeric nucleotides, pending a resolution of the 2′- and 3′-phosphate problem, this resolution has now been achieved, and it is convenient to record the result at this point. Brown, Fasman, Magrath, and Todd[10] were able to prepare a homogeneous *x*:5′-diacetyladenosine and to phosphorylate it. Subsequent removal of protecting groups yielded *exclusively* adenylic acid *a;* this demonstrated that their efforts to avoid any phosphoryl migration during the various operations had been successful (if not, a mixture of *a* and *b* would have been produced) and hence that their diacetyladenosine was, in fact, the *b*:5′-compound. The diacetyl derivative was now tosylated, and the position of the stable tosyl group was rigidly established by degradative methods as $C_{2'}$. It follows rigidly that the diacetyladenosine used was 3′:5′-diacetyladenosine and hence that adenylic acid *a* is adenosine-2′-phosphate. It was also demonstrated by Cochran and Woolfson[10] by X-ray crystallographic analysis that adenylic acid *b* is adenosine-3′-phosphate. Khym and Cohn[11] have also reached the same conclusion by an ingenious use of differential hydrolysis on ion-exchange resins. Much indirect evidence of a physical nature has also been advanced which generally confirms the conclusion that the other *a* nucleotides are also ribonucleoside-2′-phosphates and the *b* isomers ribonucleoside-3′-phosphates.[12]

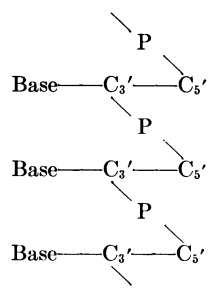
It has long been known that ribonucleic acids undergo very ready hydrolysis with alkali, yielding simple nucleotides; no larger fragments have ever been characterized. The deoxyribonucleic acids, on the other hand, are comparatively stable to alkali and are not degraded to simple nucleotides. As a result of their observations on the alkali-lability of the esters of *a* and *b* nucleotides, Brown and Todd[13] were able to advance a simple explanation of the hydrolytic behavior of the nucleic acids and thence to develop a general theory of nucleic acid structure. Structurally the polynucleotides are strictly analogous to the nucleotide esters (the esterifying group in their case is a polynucleotide chain) and must behave similarly on hydrolysis. The scheme (p. 752) represents the alkaline hydrolysis of a ribonucleic acid based on this analogy; the scheme predicts total hydrolysis exclusively to a mixture of the 2′- and 3′-phosphates of the respective nucleosides, which is, in fact, observed. In the scheme, the expression $C_{2'}$—$C_{3'}$—$C_{5'}$ is used to represent the sugar residue of a nucleoside, the hydroxyls in the 2-, 3-, and 5-positions of the chain being the only ones concerned in linkages or reactions.

The reason for the alkali-stability of the deoxyribonucleic acids is at once apparent—having no hydroxyl group at C_2, they cannot undergo the cyclization process, and so they show the normal stability of simple diesters of phosphoric acid. It is interesting to reflect that easy alkaline hydrolysis depends on the sugar in a nucleic acid being ribose—no other pentose would satisfy the structural re-

quirement for lability. This mechanism of hydrolysis does not of itself fix the internucleotidic linkage in ribonucleic acids, although it requires that one point of attachment be C_2 or C_3. The 3′:5′ linkage shown in the scheme was preferred to the 2′:5′ in the original postulate[13] mainly by analogy with the deoxyribonucleic acids, in which only the 3′:5′ linkage is possible, and it has been fully justified by later work. It is convenient at this point to discuss the structural features of the two types of acid separately.



Deoxyribonucleic Acids.—These acids are formulated as linear polynucleotides with a recurring 3′:5′ phosphodiester linkage, as shown below:[13]

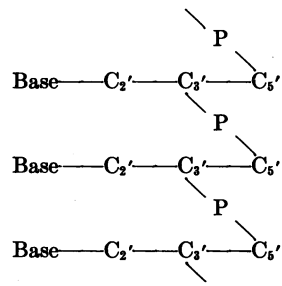


This formulation accords with their stability toward alkali and with the fact that the nucleotides formed on enzymic hydrolysis have been shown to be nucleoside-5′-phosphates by enzymic methods[14] and by synthesis.[15] The production of 3′:5′-diphosphates of the pyrimidine nucleosides on acid hydrolysis[16] is also in

accord with this structure and shows the participation of both $C_{3'}$ and $C_{5'}$ in the internucleotidic linkage. The initial action of acid on deoxyribonucleic acids is to hydrolyze the purine glycosidic linkages, yielding the apurinic acids.[17] In the second phase of its action these apurinic acids are degraded mainly by elimination reactions rather than by hydrolysis, since each purine-free sugar residue in them has a phosphate linkage at $C_{3'}$, i.e., in the β-position to a carbonyl or potential carbonyl group. In this way pyrimidine nucleoside-3′:5′-diphosphates would be expected to arise from those sites in the original molecule where purine and pyrimidine nucleoside residues were adjacent.

The alkali-stability of the deoxyribonucleic acids would seem to preclude branched-chain structures. The only type of branching which could be expected would be on phosphorus, and the resulting phosphotriester linkages would be unstable to alkali. So far no method has been devised to determine the sequence of residues in deoxyribonucleic acids.

Ribonucleic Acids.—Brown and Todd[13] postulated a recurring 3′:5′-linked polynucleotide structure for the main chain in ribonucleic acids, i.e.,

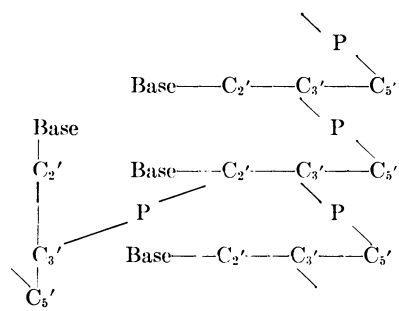


That $C_{2'}$ or $C_{3'}$ is one point of attachment of the internucleotidic linkage follows, as indicated above from the results of alkaline hydrolysis. $C_{5'}$ is also involved, since, for example, treatment of ribonucleic acids with snake-venom diesterase[18] or with ribonuclease followed by intestinal phosphatase[19] yields large amounts of the nucleoside-5′-phosphates. The postulated mechanism of alkaline hydrolysis has been further confirmed by the actual isolation of cyclic nucleoside-2′:3′-phosphates in very mild barium carbonate and ammonia hydrolyzates.[20] Convincing evidence in support of this structure has come from enzyme studies, particularly from those using ribonuclease.

Crystalline pancreatic ribonuclease acts on ribonucleic acids to yield a variety of products. Short action yields, in addition to larger fragments, substances which have been identified as the cyclic 2′:3′-phosphates of cytidine and uridine.[21] Further action yields the pyrimidine nucleoside-3′-phosphates and a variety of small polynucleotides, mainly di- and tri-, in which the terminal residue bearing the phosphorus at $C_{3'}$ (or in a 2′:3′-cyclic group) is always a pyrimidine nucleoside residue.[22] This apparent specificity for linkages attached to pyrimidine nucleosides has been confirmed and defined by studies of the enzyme's action on synthetic nucleotide esters[23] and on nucleoside 2′:3′-cyclic phosphates.[24] It was found that ribonuclease attacks the cyclic 2′:3′-phosphates of uridine and cytidine, yielding exclusively the 3′-phosphates; it has no action on the corresponding adenosine and guanosine derivatives. Moreover, although it has no action on

simple esters of pyrimidine or purine nucleoside-5′-phosphates or on esters of purine nucleoside-2′- or -3′-phosphates or of pyrimidine nucleoside-2′-phosphates, it hydrolyzes esters of pyrimidine nucleoside-3′-phosphates smoothly to the nucleoside-3′-phosphates, the cyclic 2′:3′-phosphates being detectable as intermediates in the process. From these results it follows that the pyrimidine nucleoside residues in ribonucleic acid have the phosphate group of the internucleotidic linkage at $C_{3'}$ and that ribonuclease acts in a similar way to alkali, except that it attacks only pyrimidine nucleotide sites in the nucleic acid and yields $C_{3'}$-nucleotides rather than a mixture of $C_{2'}$ and $C_{3'}$. Similar studies of the action of spleen nuclease (which attacks both purine and pyrimidine nucleotide sites in polyribonucleotides) on simple nucleotide esters show that the purine nucleoside residues are also linked at $C_{3'}$.[25] It would therefore seem that the postulated 3′:5′-linked polynucleotide structure for the main chain of the ribonucleic acids can be taken as established, and all other evidence obtained in recent years points in the same direction. There is no evidence for any other type of internucleotidic linkage.

The possibility that some ribonucleic acids have a branched-chain structure has been frequently canvassed. Bearing in mind the primary requirement of alkali-lability in any ribonucleic acid structure and the mechanism of alkaline hydrolysis, it is clear that, apart from branching on phosphorus through phosphotriester linkages which would be theoretically admissible, only one type of branching from $C_{2'}$ in nucleoside residues in the main polynucleotide chain is possible,[13] viz.,



In this type of branching, $C_{2'}$ in one residue in the main chain is joined by a phosphodiester linkage to $C_{3'}$ in the first residue of the branching chain. Linkage to $C_{5'}$ in the first residue of the branch would give a structure in which neither that residue nor the branching residue in the main chain would contain a free hydroxyl so situated as to permit the cyclization step in alkaline hydrolysis; as a result, alkaline hydrolysis of a polynucleotide containing such a structure would yield fragments larger than simple nucleotides. There has been much discussion as to whether ribonucleic acids are branched or unbranched, and no decisive evidence either way has been adduced. It may be that both branched and unbranched structures occur, but at the moment one can only say that chemical considerations, although they indicate *how* branching may occur, are silent on whether it does occur. Physical studies or the further progress of synthetic work may decide this important point.

There remains, of course, the problem of the sequence of nucleoside residues in the ribonucleic acids, for it is in the sequence of residues that the difference between the various acids doubtless lies. A method suitable for this purpose has been proposed by Brown, Fried, and Todd.[26] This method depends on the oxidation of a terminal nucleoside residue containing a free 2′:3′-α-glycol grouping with

periodate, followed by expulsion of the oxidized residue under very weakly alkaline reaction by means of an elimination reaction, leaving the rest of the molecule intact. Its validity has been tested in a number of oligonucleotides,[27] but it is clear that it cannot be applied to the nucleic acids themselves until a truly homogeneous ribonucleic acid can be made available. It is worthy of mention that, given a homogeneous ribonucleic acid, this stepwise degradation method of sequence determination should in theory serve to establish the presence or absence of branching; on degrading the main chain, the method would come to a halt when a branching point was reached, since there would be no free α-glycol system in the residue exposed at that point, and hence periodate action would not occur.

This review of present chemical evidence on nucleic acid structure is necessarily incomplete, for it is impossible in a limited space to include all the mass of work which has been done in the field. Nevertheless, I have endeavored to select the main features, in an attempt to give a rational account of our present views and the basic evidence for them. Certain aspects which require further investigation have been indicated, but it seems reasonable to conclude that the broad outline of the chemical structure of this important group of natural substances is now clear enough to open the way to new advances in our knowledge of their behavior and biological functions.

[1] P. A. Levene and H. S. Simms, *J. Biol. Chem.*, **65**, 519, 1925; **70**, 327, 1927.

[2] For a collected account of this work see, e.g., G. W. Kenner, *Fortschr. Chem. org. Naturstoffe*, **8**, 96, 1951.

[3] D. M. Brown and B. Lythgoe, *J. Chem. Soc.*, p. 1990, 1950; W. Andersen, D. H. Hayes, A. M. Michelson, and A. R. Todd, *J. Chem. Soc.*, p. 1882, 1954.

[4] F. R. Atherton, H. T. Openshaw, and A. R. Todd, *J. Chem. Soc.*, p. 382, 1945.

[5] J. Baddiley and A. R. Todd, *J. Chem. Soc.*, p. 648, 1947; A. M. Michelson and A. R. Todd, *J. Chem. Soc.*, p. 2476, 1949.

[6] A. M. Michelson and A. R. Todd, *J. Chem. Soc.*, p. 951, 1953; p. 34, 1954; D. H. Hayes, A. M. Michelson, and A. R. Todd, *J. Chem. Soc.*, 1954 (in press).

[7] D. M. Brown and A. R. Todd, *J. Chem. Soc.*, p. 44, 1952.

[8] C. E. Carter and W. E. Cohn, *Federation Proc.*, **8**, 190, 1949.

[9] D. M. Brown, D. I. Magrath, and A. R. Todd, *J. Chem. Soc.*, p. 2708, 1952.

[10] D. M. Brown, G. D. Fasman, D. I. Magrath, A. R. Todd, W. Cochran, and M. M. Woolfson, *Nature*, **172**, 1184, 1953; *J. Chem. Soc.*, p. 1448, 1954.

[11] J. X. Khym and W. E. Cohn, *J. Am. Chem. Soc.*, **76**, 1818, 1954.

[12] *Cf.*, e.g., L. F. Cavalieri, *J. Am. Chem. Soc.*, **74**, 5804, 1952; J. J. Fox, L. F. Cavalieri, and N. Chang, *J. Am. Chem. Soc.*, **75**, 4315, 1953; A. M. Michelson and A. R. Todd, *J. Chem. Soc.*, p. 34, 1954; H. S. Loring, M. L. Hammell, L. W. Levy, and H. W. Bortner, *J. Biol. Chem.*, **196**, 821, 1952.

[13] D. M. Brown and A. R. Todd, *J. Chem. Soc.*, p. 52, 1952.

[14] C. E. Carter, *J. Am. Chem. Soc.*, **73**, 1537, 1951.

[15] A. M. Michelson and A. R. Todd, *J. Chem. Soc.*, p. 951, 1953; p. 34, 1954; and as yet unpublished data.

[16] C. A. Dekker, A. M. Michelson, and A. R. Todd, *J. Chem. Soc.*, p. 947, 1953.

[17] C. Tamm, M. E. Hodes, and E. Chargaff, *J. Biol. Chem.*, **195**, 49, 1952.

[18] W. E. Cohn and E. Volkin, *Arch. Biochem. and Biophys.*, **35**, 465, 1952.

[19] W. E. Cohn and E. Volkin, *Nature*, **167**, 483, 1951.

[20] R. Markham and J. D. Smith, *Biochem. J.*, **52**, 552, 1952.

[21] R. Markham and J. D. Smith, *Research (London)*, **4**, 344, 1951; *Nature*, **168**, 406, 1951.

[22] R. Markham and J. D. Smith, *Biochem. J.*, **52**, 558, 1952.

[23] D. M. Brown and A. R. Todd, *J. Chem. Soc.*, p. 2040, 1953.

[24] D. M. Brown, C. A. Dekker, and A. R. Todd, *J. Chem. Soc.*, p. 2715, 1952.

[25] D. M. Brown, L. A. Heppel, and R. J. Hilmoe, *J. Chem. Soc.*, p. 40, 1954.

[26] D. M. Brown, M. Fried, and A. R. Todd, *Chemistry & Industry*, p. 352, 1953.

[27] P. R. Whitfeld and R. Markham, *Nature*, **171**, 346, 1953.