A Proposed General Procedure for Isolating End-Groups of Nucleic Acids*

R. V. TOMLINSON AND G. M. TENER†

From the Department of Biochemistry, University of British Columbia, Vancouver

Received December 17, 1962

Model experiments have been done to illustrate the following general method for isolating the end-groups of nucleic acids. The nucleic acid is degraded with an appropriate nuclease such that the polynucleotide containing the original terminal nucleotide residue has either no, or two, terminal phosphate groups, whereas the rest of the polynucleotides, which arise from inside the nucleic acid chain, contain only one terminal phosphate residue. The mixed polymers are separated according to their charge on DEAE-cellulose at *p*H 7.5 with 7 M urea in the eluting solution and an increasing concentration of sodium chloride. Each fraction is isolated and treated with phosphomonoesterase. The polymer derived from the end now differs by two charges from the other polymers in the fraction and can be separated from them by rechromatography in the same system. The method can be used to determine both ends of DNA and RNA, whether or not they are phosphorylated. A modification of the procedure which makes use of a cyanoethyl ester of a deoxyribopolynucleotide has also been studied.

An initial step in determining the primary structure of a strand of either ribo- or deoxyribonucleic acid is the identification of its two terminal nucleoside residues and their immediate neighbors. In this paper we propose a general procedure for isolating these end-groups and have tested the procedure in model experiments. In addition, an alternate procedure of more limited application has been described.

EXPERIMENTAL[1]

A standard DEAE-cellulose (chloride) column (20 × 1.8 cm) was prepared as described in the accompanying paper (Tomlinson and Tener, 1963). The column was eluted with a linear gradient formed by running 1 liter of 0.3 M sodium chloride, 7 M in urea, into 1 liter of 7 M urea. Both solutions were buffered by the addition of 20 ml of 1 M Tris chloride *p*H 7.5 or 7.8 or, alternatively, 20 ml of 1 M sodium acetate *p*H 4.7 or 5.1. Fractions of 10 ml were collected at a flow rate of 1 ml per minute.

Mixed polynucleotides were obtained by digesting salmon testis DNA with pancreatic deoxyribonuclease and isolating the resultant polynucleotides on DEAE-cellulose columns, and the dephosphorylated polynucleotides were obtained by digesting these polynucleotide fractions with phosphomonoesterase from *E. coli* as described in the accompanying paper (Tomlinson and Tener, 1963). The phosphomonoesterase was purchased from Worthington Biochemicals Inc., Freehold, New Jersey.

Cyanoethyl Ester of the Pentanucleotide Fraction.—One hundred optical density units (271 mμ) of the pentanucleotide fraction were converted into the pyridinium salt by passage through a column of Dowex-50 (pyridinium) resin. The solution was concentrated to dryness *in vacuo*, 5 ml of anhydrous pyridine added, and the solution again concentrated to dryness. The residue was dissolved in a mixture of 1 ml of pyridine, 0.2 ml of 4-benzylpyridine, and 2 ml of hydracrylonitrile. Three hundred mg of dicyclohexylcarbodiimide was added and the mixture was allowed to stand for 24 hours at room temperature. Unreacted dicyclohexylcarbodiimide was decomposed with 50 ml of water and the solution filtered to remove dicyclohexyl urea. The filtrate was poured through a column of DEAE-cellulose (carbonate). After this column had been washed well with water, the product was eluted with 200 ml 2 M ammonium bicarbonate. The effluent was concentrated to dryness *in vacuo* at 30° and last traces of ammonium bicarbonate eliminated by adding water and reevaporating to dryness. The product was characterized by chromatographing it again on a DEAE-cellulose column with 7 M urea in the eluting solution, first at *p*H 5.0, where it was indistinguishable from the starting material, and then at *p*H 7.5, where it was eluted ahead of the starting material in the region where polymers carrying one less charge would be expected. In addition, when this new product was treated with alkali, it gave back the original pentanucleotide starting material.

Removal of the Cyanoethyl Group.—The cyanoethyl group was removed from the ester by treating it with 0.5 N sodium hydroxide at 100° for 10 minutes (Tener, 1961). The reaction mixture was neutralized to *p*H 8 by careful addition of Dowex-50 (hydrogen) prior to chromatography.

RESULTS AND DISCUSSION

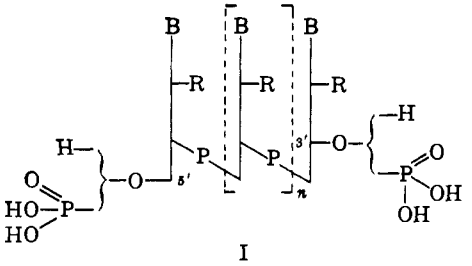
Most nucleic acids are high-molecular-weight linear polymers composed of nucleoside residues linked together by phosphodiester bonds from the 3′-position of one residue to the 5′-position of the next. The terminal residues, however, are unique in that they can have either a free hydroxyl or a phosphomonoester group at either their 3′- or 5′-positions. Structure I illustrates these features.

The unique nature of the terminal nucleoside residues has been utilized in several methods which have been proposed for their identification. Methods for the study of RNA employing enzymic and alkaline degradation were developed by Markham and Smith (1952) (see also Lane and Allen, 1961). For example, alkaline degradation of tobacco mosaic virus RNA (TMV-RNA) liberated adenosine from the 3′-hydroxyl end of the molecule (Sugiyama and Fraenkel-Conrat, 1961). The rest of the nucleosides in the RNA were liberated

* This work has been supported by the United States Public Health Service, Grant C-5342, and was first reported at the Pacific Slopes Biochemical Conference, Seattle, September 6, 1962.

† Medical Research Associate, Medical Research Council of Canada.

[1] The following abbreviations have been used: Tris, tris(hydroxymethyl)aminomethane; RNA, ribonucleic acid; DNA, deoxyribonucleic acid; the convention recommended by *J. Biol. Chem.* for abbreviating polynucleotides is used, with Y being used to represent a deoxyribonucleoside residue, e.g., pYpYpY is the mixture of trinucleotides with the terminal phosphate on the 5′-position. CEpYpYpYpYpY is the cyanoethyl ester of the pentanucleotide fraction.



I

B = purine or pyrimidine base, backbone = ribose (R = OH) or deoxyribose (R = H), with either hydroxyl or phosphate groups at the terminal 3'- and 5'-positions

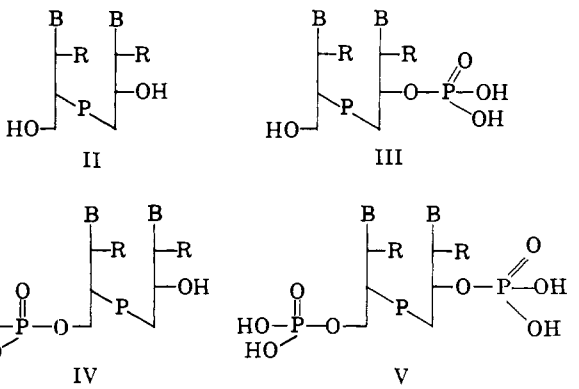
as nucleoside 2'- and 3'-phosphates and thus were readily separated from the end-group. Had the nucleoside at the other end of the chain carried a terminal phosphate group, it would have been liberated as a nucleoside-(2')3',5'-diphosphate, and it likewise could have been separated from the bulk of the nucleotides and identified. The absence of a nucleoside diphosphate in the digest indicated that in TMV-RNA this end nucleoside does not have a 5'-phosphate group. Yeast *s*-RNA does, however, bear a phosphate group at this position, since it liberates guanosine-(2')3',5'-diphosphate on alkaline hydrolysis (Singer and Cantoni, 1960; Zillig *et al.*, 1960). The use of alkaline degradation to determine the end-groups of a nucleic acid is limited, though, to RNA, and it can be used only when the respective end nucleosides have a 3'-hydroxyl or a 5'-phosphomonoester group.

Similar unique end-groups or even end-sequences can result from enzymic digestion of either RNA or DNA. For example, Whitfeld (1962) has extended the studies on TMV-RNA by degrading it with pancreatic ribonuclease as well as Taka-Diastase T_1 and *B. subtilis* ribonuclease. His work confirmed the presence of adenosine at the 3'-hydroxyl end. In addition, it showed that the penultimate base was a pyrimidine, and that probably the third position was also occupied by a pyrimidine.

The publication by Ralph *et al.* (1962) describes a method for labeling a terminal phosphate residue in RNA with C^{14}-aniline to facilitate its detection. The method was used to determine nucleotide sequences at the 5'-hydroxyl end of yeast *s*-RNA. The mixed polymers released by ribonuclease digestion were chromatographed on columns of DEAE-cellulose and the various fractions collected. Those containing the C^{14}-labeled phosphoranilidates were isolated and, in part, characterized. In this way, the authors demonstrated that although the majority of the *s*-RNA chains terminate with guanosine, some do terminate with either adenosine or uridine. In addition, it was apparent that there were few, if any, common terminal sequences. The use of this technique is limited to detecting base sequences at the 5'-hydroxyl end of phosphorylated RNA, or at the 3'- or 5'-hydroxyl ends of phosphorylated DNA. In addition, a problem still existed in separating the labeled ends from the other digestion products.

In the accompanying article (Tomlinson and Tener, 1963) we reported that chromatography on DEAE-cellulose columns in the presence of 7 M urea separated mixed polynucleotides according to their net negative charge. This separation has suggested the following general method for isolating the end-groups of a nucleic acid chain. First, the nucleic acid is degraded by a nuclease which will break up the chain in such a way that the terminal polynucleotide residue(s) carries two or no primary phosphate groups, whereas the rest of the polymers resulting from the digestion carry only one primary phosphate group. Four different types of nucleases are required: two for degrading DNA to give deoxyribopolynucleotides terminated in either 3'- or 5'-phosphate groups, and two for degrading RNA to the corresponding ribopolynucleotides. Examples of such enzymes are pancreatic deoxyribonuclease and spleen deoxyribonuclease (see Laskowski, 1961) for degrading DNA and Taka-Diastase T_1 and pancreatic ribonucleases (see Anfinson and White, 1961) and the nuclease from mung bean sprouts (Sung and Laskowski, 1962) for degrading RNA.

As an example, the four different types of compounds which might arise on nuclease digestion of Structure I are illustrated below with compounds containing two nucleoside residues (the mononucleoside and higher polynucleotide analogs could also be formed).



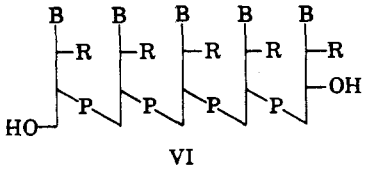
II III IV V

Compounds II and V could arise from either end of a nucleic acid chain, depending on whether it was phosphorylated and on what nuclease was used. The rest of the chain could appear as either compound III or IV. At *p*H 7.5, compounds II to V would have one, three, three, and five negative charges, respectively.

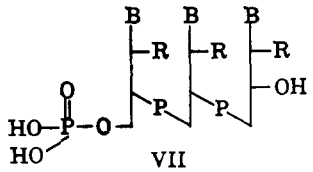
The second step is to separate the enzymic digest in the urea-containing chromatographic system, and to isolate the polymers with various charges as discrete peaks. Compound II would be eluted before the mononucleotides (two charges), whereas compound V would come off with the tetranucleotides (five charges). Each fraction is next treated with phosphomonoesterase to remove all primary phosphate groups, and then rechromatographed in the same system (except in the case of II where it would not be necessary). Since the terminal residue now differs by two charges from the remaining polymers in its peak, it readily separates, thus permitting its isolation and identification.

The method has been tested with a 1:10 mixture of the pentanucleoside tetraphosphate (VI, YpYpYpYpY) and the trinucleotide (VII, pYpYpY). These could be among the products formed by pancreatic deoxyribonuclease digestion of a DNA with no phosphate group on the 5'-hydroxyl terminus. The pentanucleoside tetraphosphate represents the terminal residue. When this mixture was chromatographed on a DEAE-cellulose column at *p*H 7.5 in 7 M urea, both components were eluted together (Fig. 1a) because both carry four negative charges.

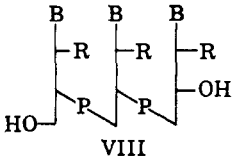
However, after the mixture was treated with phosphomonoesterase to convert the trinucleotide into a trinucleoside diphosphate (VIII, YpYpY), the two components were readily separated by rechromatographing them in the same system. The results of such rechromatography are shown in Figure 1c. Compounds VI and VIII are well separated, so that VI could be easily isolated and then identified.



VI



VII



VIII

As another approach to separating the two components, the use of a DEAE-cellulose column run at *p*H 5.1 was explored. At this *p*H the secondary phosphate dissociation is repressed so that a trinucleotide will have about three negative charges. The results from this study are shown in Figure 1b. The separation was not complete, although it is obvious that the trinucleotide was being eluted from the column earlier than the pentanucleoside tetraphosphate. However, it was not half way between the four-charged pentanucleoside tetraphosphate and the two-charged trinucleoside diphosphate which had been added as a marker. Therefore at *p*H 5.1 the trinucleotide must have a net charge slightly greater than three. But any attempt to reduce this partial charge by lowering the *p*H of the eluting solution caused the peak to split, presumably the result of protonation of cytosine residues. This approach for isolating the terminal residue, therefore, is not as satisfactory as the use of phosphomonoesterase. The latter technique should be adequate for isolating small polynucleotide end-groups at least to the decanucleotide level.

The lack of appropriate compounds has prevented us from doing model experiments to test the method under the alternative situation, where the terminal residue bears two primary phosphate groups, as in compound V. However, on chromatography of the nuclease digest such a compound would be expected to come off with the tetranucleotides. After phosphomonoesterase treatment, it would yield a dinucleoside phosphate such as compound II, and then could be readily separated from the tetranucleoside triphosphates. This expectation has been realized in studies of the end-groups of yeast *s*-RNA (Bell *et al.*, 1963).

By using an appropriate nuclease, analogous experiments can be done to isolate the polynucleotide residue at the 3′-hydroxyl end of the molecule; and in certain instances, both the 3′- and 5′-hydroxyl ends can be isolated simultaneously.

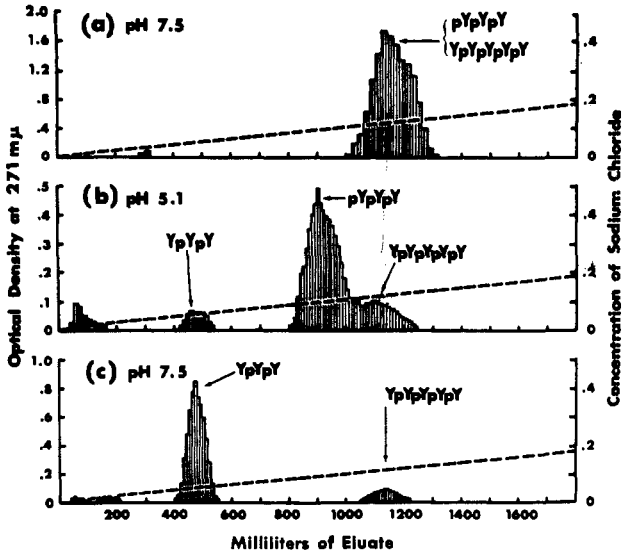
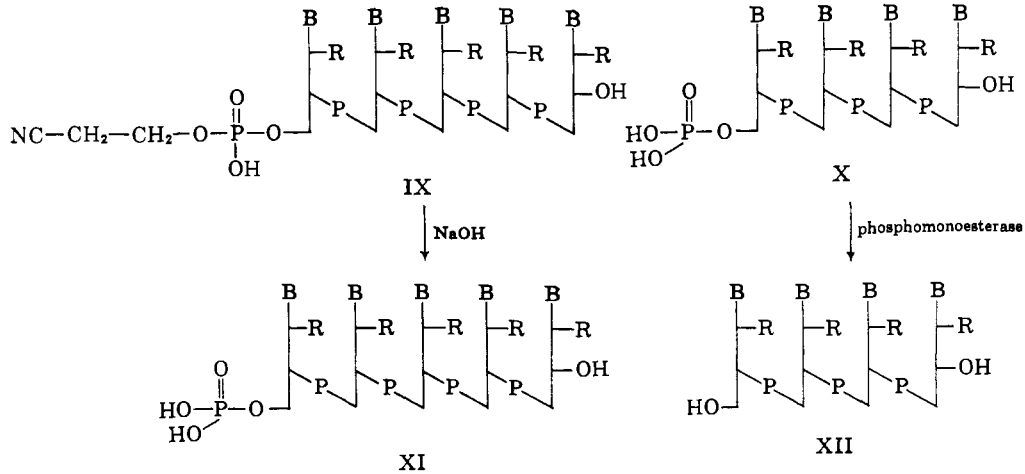


FIG. 1.—Chromatography of a mixture of trinucleotide (pYpYpY) and pentanucleoside tetraphosphate (YpYpYpYpY) fractions in a 10:1 ratio on a DEAE-cellulose (chloride) column (20 × 1.8 cm), using a gradient of sodium chloride as shown and the eluting mixture 7 M in urea: (a) at *p*H 7.5, (b) at *p*H 5.1 with added trinucleoside diphosphate (YpYpY) as marker, (c) at *p*H 7.5 after treatment with phosphomonoesterase.

The separation of long terminal sequences from equally long contaminating polymeric nucleotides may require that a difference greater than two negative charges be achieved. Theoretically, in those cases where a DNA carries a terminal phosphate at either the 3′- or 5′-hydroxyl ends, the difference can be increased to three negative charges. It is proposed that this phosphate group in the DNA be esterified with hydracrylonitrile to produce the cyanoethyl ester. The nucleic acid is then degraded by an appropriate nuclease, and the resulting polynucleotides are separated according to their negative charges. The terminal polynucleotide residue could appear as compound IX (CEpYpYpYpYpY), and this would be eluted at *p*H 7.5 along with the tetranucleotide (X, pYpYpYpY), since both carry five negative charges. Treatment of the mixture with phosphomonoesterase would remove the terminal phosphate from the tetranucleotide to produce XII, but it would not touch the cyanoethyl ester (IX). Mild alkaline treatment can then be used to remove the cyanoethyl group, converting IX to the pentanucleotide (XI). At *p*H 7.5, XI and XII differ by three negative charges. This



IX X

NaOH phosphomonoesterase

XI XII

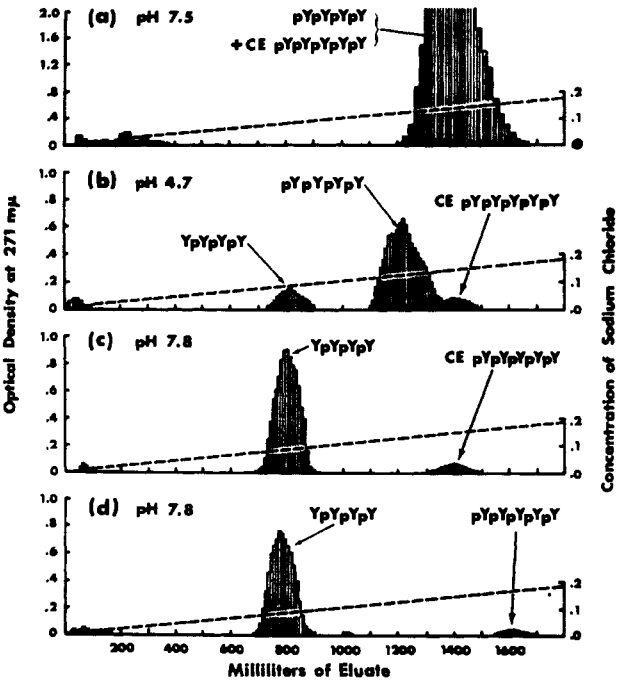


FIG. 2.—Chromatography of a mixture of tetranucleotide (pYpYpYpY) and the cyanoethyl ester of pentanucleotide (CEpYpYpYpYpY) fractions in a 10:1 ratio on a DEAE-cellulose (chloride) column (20 × 1.8 cm), using a gradient of sodium chloride as shown and the eluting mixture 7 M in urea: (a) at *p*H 7.5, (b) at *p*H 4.7 with added tetranucleoside triphosphate (YpYpYpY) as a marker, (c) at *p*H 7.8 after phosphomonoesterase treatment, and (d) at *p*H 7.8 after phosphomonoesterase and alkaline treatment.

enhanced charge difference should permit a greater resolution of the mixture than is possible with compounds of only two-charge difference. Model experiments with the mixed polymers IX and X demonstrated the feasibility of this approach.

Figure 2a shows that a 10:1 mixture of X and IX emerged as a single peak after 1400 ml of *p*H 7.5 eluant had passed through the column. After phosphomonoesterase treatment of the mixture and rechromatography, the cyanoethyl ester of the pentanucleotide again was eluted with a peak at 1400 ml whereas the dephosphorylated tetranucleotide now came off the column after 800 ml. These results are shown in Figure 2c. When the phosphomonoesterase digestion was followed by treatment with mild alkali, the results shown in Figure 2d were obtained. The pentanucleotide, having lost its cyanoethyl group and having acquired one more negative charge, now emerged from the column after 1600 ml. The elution characteristics of the tetranucleoside triphosphate were unchanged by the additional treatment. However, it should not usually be necessary to use this modification for the isolation of the end-groups, since in most cases a difference of two negative charges is sufficient to achieve the separation. In addition, there are complications associated with chemical treatment which could lead to the undesirable release of nonterminal phosphate groups. Also, large polynucleotides and nucleic acids are difficult to dissolve in pyridine, although by the use of 4-benzyl pyridine reported in this paper or the dialysis technique of Bautz and Hall (1962) this objection might be circumvented.

Figure 2b is an elution pattern obtained by chromatographing the mixture of IX and X at an acid *p*H. The tetranucleotide (X) was not eluted half way between the cyanoethyl ester of the pentanucleotide (IX) and the tetranucleoside triphosphate (XII) which had been added as a marker. Therefore, it was evident that, even at *p*H 4.7, the secondary phosphate dissociation in compound X was not completely repressed. Moreover, since some separation within the peak was noticeable, presumably due to protonation of cytosine residues, further lowering of the *p*H would not be expected to improve the separation. Again, this approach for isolating the terminal residue proved to be impractical.

This cyanoethyl ester modification of the sequence determination technique is limited to DNA which has a terminal phosphate group. On the other hand, the basic technique itself is not only completely general but also much simpler. It can be used for determining both end-sequences of DNA or of RNA, regardless of whether they have a terminal phosphate residue.

In the model experiments described here synthetic mixtures of the polynucleotide and the end-group were prepared in the ratio of 10:1 in order to illustrate the procedures readily. However, such a ratio does not represent the limits of the technique. If one employs optical density to follow the materials coming off chromatographic columns, it should be possible to detect the presence of the end-group polymer in a peak when this polymer constitutes only 1% of the total optical density. If the peak represents 20% of the optical density in the nuclease digest, it should be possible to detect, isolate, and identify the end-sequence of a nucleic acid having at least 500 nucleotide residues. The sensitivity of the method can be increased at least 100-fold by utilizing radioisotopes. The limitation of the technique, therefore, is not in the methods available for detecting the polynucleotide end-groups.

However, one requirement of this method is to have nucleases of adequate purity, which will degrade the nucleic acids in the desired manner without producing any adverse side reactions. For example, the nuclease must be completely free of phosphomonoesterase activity. Such enzymes are now available and some can be obtained commercially. A second limitation is the availability of pure nucleic acids. Some relatively homogeneous nucleic acids which should be suitable for end-group studies can be isolated from bacteriophage and viruses. However, the method can be used even on a mixture of nucleic acids such as yeast *s*-RNA if the distribution of end-sequences is sought.

REFERENCES

Anfinson, C. B., and White, F. H., Jr. (1961), *in* The Enzymes, vol. 5, Boyer, P., Lardy, H., and Myrbäck, K., eds., New York, Academic, p. 95.

Bautz, E. F. K., and Hall, B. D. (1962), *Proc. Nat. Acad. Sci. U. S. 48*, 400.

Bell, D., Tomlinson, R. V., and Tener, G. M. (1963), *Biochem. Biophys. Res. Comm. 10*, 304.

Lane, B. G., and Allen, F. W. (1961), *Biochem. Biophys. Acta 47*, 36.

Laskowski, M., Sr. (1961), *in* The Enzymes, vol. 5, Boyer, P., Lardy, H., and Myrbäck, K., eds., New York, Academic, p. 123.

Markham, R., and Smith, J. D. (1952), *Biochem. J. 52*, 565.

Ralph, R. K., Young, R. J., and Khorana, H. G. (1962), *J. Am. Chem. Soc. 84*, 1490.

Singer, M. F., and Cantoni, G. L. (1960), *Biochim. Biophys. Acta 39*, 182.

Sugiyama, T., and Fraenkel-Conrat, H. (1961), *Proc. Nat. Acad. Sci. U. S. 47*, 1393.

Sung, S., and Laskowski, M., Sr. (1962), *J. Biol. Chem. 237*, 506.

Tener, G. M. (1961), *J. Am. Chem. Soc. 83*, 159.

Tomlinson, R. V., and Tener, G. M. (1963), *Biochemistry 2*, 697 (this issue).

Whitfeld, P. R. (1962), *J. Biol. Chem. 237*, 2895.

Zillig, W., Schachtschabel, D., and Krone, W. (1960), *Z. Physiol. Chem. 318*, 100.