Deoxynucleotide substitution: a new technique for sequence analysis of RNA

A.Van de Voorde, R.Rogiers, J.Van Herreweghe, H.Van Heuverswyn, G.Volckaert and W.Fiers

Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, 9000 Gent, Belgium

Received 12 July 1974

ABSTRACT

It is possible to replace in a normal transcription reaction catalyzed by E.coli RNA polymerase one of the four precursors by the corresponding deoxynucleoside triphosphate . These deoxynucleotide-substituted RNA's offer interesting prospects for nucleotide sequence analysis. Indeed by the use of U_2-RNase with dG-RNA, or pancreatic RNase with dC-RNA or dU-RNA, base specific cleavage can be obtained at any of the four residues. In this way overlap of at least six residues in length can be obtained for any site in the RNA. The technique offers also great benefit for solving the sequence of the more difficult T_1-oligonucleotides. Some examples in the sequence analysis of SV40 DNA-Hind fragments are reported.

INTRODUCTION

Berg et al. (1) have made the interesting observation that E.coli DNA polymerase I will insert ribonucleotide residues into a polydeoxynucleotide chain if Mn^{++} replaces Mg^{++} in the reaction mixture. This ribosubstitution technique has provided a valuable tool in nucleotide sequence analysis of DNA as it allows base specific cleavage (2, 3). Conversely, it is possible to incorporate deoxynucleotide residues into RNA using a reaction mixture which contains template DNA, E.coli RNA polymerase and Mn^{++} in addition to Mg^{++} (4, 5). In this paper we report the use of this deoxynucleotide substitution technique for nucleotide sequence analysis of RNA.

Abbreviations used: The ribonucleotide residues are indicated as U, C, A and G; the corresponding deoxyribonucleotides as dU, dC, dA and dG; P-RNase, pancreatic ribonuclease; d-oligonucleotide, deoxyoligonucleotide terminated by a ribonucleotide at the 3'-end; T_1 and P-oligonucleotide, oligonucleotide released after complete digestion with T_1-RNase or P-RNase respectively.

Indeed, as many ribonucleases are base-specific and cleave the phosphodiesterbond via a 2', 3' cyclic phosphodiester, they can no longer cleave an RNA chain adjacent to a deoxynucleotide residue. As shown in Table 1, base-specific cleavage can be realized on this basis by the digestion of the appropriately substituted RNA chain with U_2-ribonuclease (specific for purine nucleotides) or with pancreatic ribonuclease (specific for pyrimidine nucleotides). This procedure allows in principle to solve two main problems in nucleotide sequence analysis, viz. the ordering of T_1- and P-oligonucleotides and the sequence determination of the longer T_1-oligonucleotides.

MATERIALS

Calf thymus DNA was treated at room temperature with $NaIO_4$ in order to eliminate ribose-containing contaminants. Residual periodate was removed with glycol. Hind-fragment DNA was obtained by digestion of SV40 DNA with *H. influenzae Rd* restriction enzyme. *E. coli* RNA-polymerase preparations were gifts from Dr. R. Schilperoort, Dr. H. Schaller or Dr. R. Burgess. The preparations were further purified by a glycerol gradient centrifugation in the presence of 0.5 M KCl. The specific activity was in the range of 1,000 U/mg.

α-^{32}P-nucleoside triphosphate was either prepared in the laboratory or bought from NEN Chemicals. The unlabeled nucleoside triphosphates were obtained from P-L Biochemicals, Sigma or Boehringer Co., and were further purified by chromatography on a Dowex 1 column (x8, 200-400 mesh) using a gradient of 1.5 M HCOOH to 2 M HCOOH and 1.0 M NH_4OOCH. Deoxynucleoside triphosphate from Boehringer was further treated with periodate.

"Total ribonuclease", was a mixture of Aspergillus oryzae ribonucleases obtained by heat treatment of a crude extract according to Hiramaru et al. (7).

METHODS

Synthesis of labeled, substituted RNA. The reaction mixture (0.05 ml) contained 0.1 µg heat denatured DNA (calf thymus or SV40 Hind-fragment), 10 µg E.coli RNA polymerase, 0.15 mM of three nucleoside triphosphates, one of which was α-^{32}P -labeled (10-20 mC/mmole), 0.15 mM of the complementary deoxynucleoside triphosphate, in 0.04 M Tris-buffer, pH 7.8, 0.075 M KCl, 4 x 10^{-5}M phosphate, 0.1 mM EDTA and $MgCl_2$ and Mn Cl_2 as indicated below. Incubation was at 37°C until incorporation levelled off (usually 2-3 hrs). The reaction was stopped by addition of 10 µl 0.5 % SDS + 0.1 M EDTA, and the product was isolated by filtration through a 0.9 cm x 12 cm Sephadex G50-column. The peak fractions were collected, phenolized and precipitated.

Fingerprinting and nucleotide sequence analysis. In order to judge the ribose-contamination and to prepare deoxyoligonucleotides containing a ribonucleotide at the 3'end (d-oligonucleotide), the substituted RNA was hydrolyzed at all ribonucleotide linkages, either by digestion with "total" RNase (in 0.05 M Na-acetate buffer, pH 4.5) or by alkali. The digest was separated by conventional two-dimensional electrophoresis (8). Likewise the T_1-RNase or P-RNase digests were fingerprinted either by two-dimensional electrophoresis or electrophoresis-homochromatography (8, 9).

RESULTS AND DISCUSSION

Application and scope of the deoxynucleotide substitution technique

As most ribonucleases, e.g. T_1, U_2 and P, hydrolyse the internucleotide bond via a 2', 3'-cyclic phosphodiester intermediate, they are unable to cleave next to a deoxyribose-residue. On this basis, RNA in which an appropriate residue is replaced by the deoxynucleotide homologue, can be cleaved selectively at any of the four base residues (Table 1). One of the main problems in normal nucleotide sequence work of RNA's is to order the T_1-oligonucleotides. However, it is obvious that by the use of an appropriate deoxysubstitution-RNase combination a direct overlap of at least four nucleotides at any site in the RNA chain can be obtained (Fig. 1). As furthermore the nucleotide

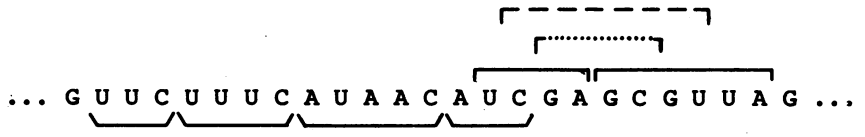

Figure 1

An example of the deoxynucleotide substitution technique.

A hypothetical sequence is shown. On top overlaps between T_1-oligonucleotides are indicated which can be obtained by dC, or dU or dG substitution (from top to bottom respectively). Note that these overlaps are in fact extended by two residues as also the preceding and the succeeding nucleotide is known. The 16 residues long T_1-oligonucleotide can readily be solved on the basis of normal double digestion and of dU-substitution; the products obtained by P - RNase digestion of the dU-oligonucleotide are indicated underneath (cfr. text for further details).

TABLE 1

BASE-SPECIFIC CLEAVAGE OF RNA

Substitution	Enzyme	Base specificity
-	T_1- RNase	G
dG	U_2- RNase	A
dU	P - RNase	C
dC	P - RNase	U

preceding the tetranucleotide overlap is known from the specificity of the enzyme used, and as the nucleotide following the tetranucleotide can be deduced from nearest neighbour information, it follows that for any site in the polynucleotide chain an effective overlap of six residues can be derived. In principle this is sufficient to order all oligonucleotides in a chain nearly 1,000 residues in length.

Also the sequencing of long T_1-oligonucleotides, especially those rich in pyrimidines, may often pose some difficulties. Of course, in theory they can be solved by partial digestion with exonucleases, but this requires rather much radioactivity and stringent purity of the exonuclease preparations. Deoxynucleotide substitution also helps considerably in solving such sequencing problems. Consider for example the first T_1-oligonucleotide (n = 16) shown in Fig. 1. P - RNase analysis leads to (5 U, 3 C, 2 A-U, A-A-C)G, U_2-RNase to (5 U, C, 2 C-A, U-A-A) (U,C)G and combined with nearest neighbour information this gives (5 U,C) C-A-U-A-A-C-A-U-C-G. dU-substitution allows to isolate dU-dU-C and dU-dU-dU-C, only the latter of which is labeled in an α- ^{32}P -ATP reaction; this information is sufficient to deduce the complete sequence.

Synthesis of deoxynucleotide substituted RNA.

The reaction rate (and final yield) depended for each substitution on the concentration of divalent ions. Optimal incorporations were obtained under the following conditions :

d GTP : 5 mM Mg^{++} and 1.5 mM Mn^{++}
d UTP : 5 mM Mg^{++} and 5 mM Mn^{++}
d CTP : 10 mM Mg^{++} and 1.5 mM Mn^{++}

In general the yields amounted to approximately 5 to 10 % relative to the reaction mixture containing all 4 ribonucleoside triphosphates. As α- ^{32}P -labeled precursors of high specific activity are commercially available, sufficient radioactivity can still be accumulated in the products for normal analyses.

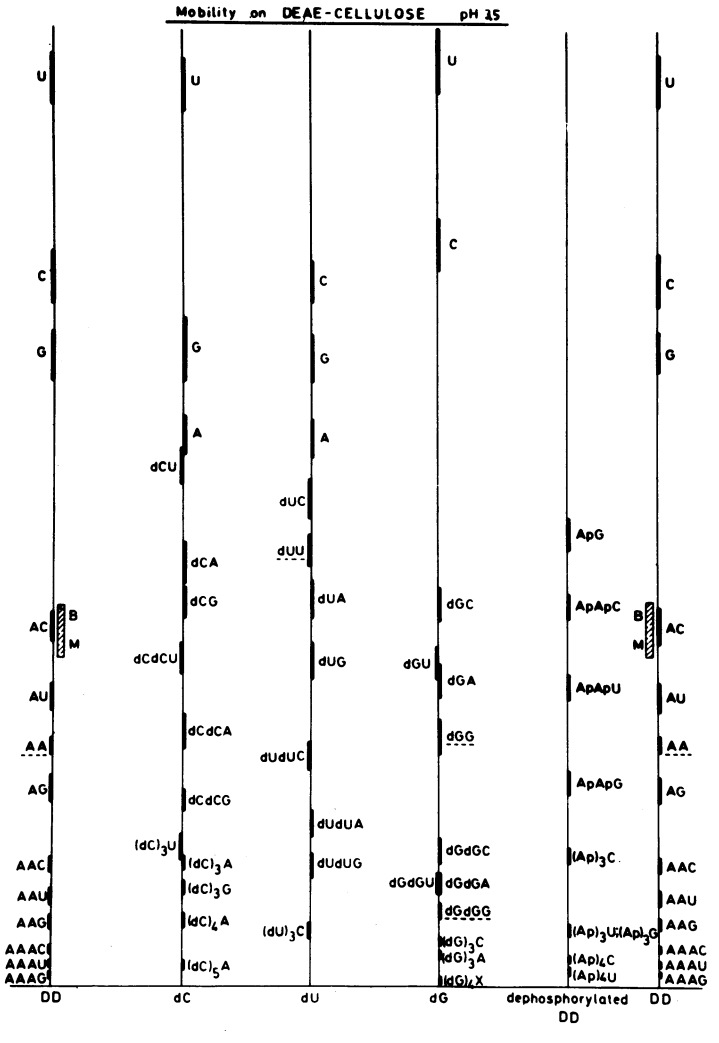

Figure 2

Relative mobility of deoxynucleotide-containing oligonucleotides.

Oligonucleotides, containing one or more deoxyresidues followed by a 3'-terminal ribonucleotide were prepared from dC - dU- respectively dG-containing transcripts and characterized as described in the text. The mobility of each d-oligonucleotide in a DEAE-paper electrophoresis system at pH 3.5 (8) was determined relative to a standard double digest shown at the right and the left hand lanes (DD; T_1-RNase + P - RNase). The mobility of a dephosphorylated double digest is also shown; these products are used as references for identification of the 3'-ends of transcripts. B M means blue dye marker (xylene cyanol FF). Odd products are underlined by dashes; e.g. dU-U is formed as a result of some ribo U contamination in the dU-RNA.

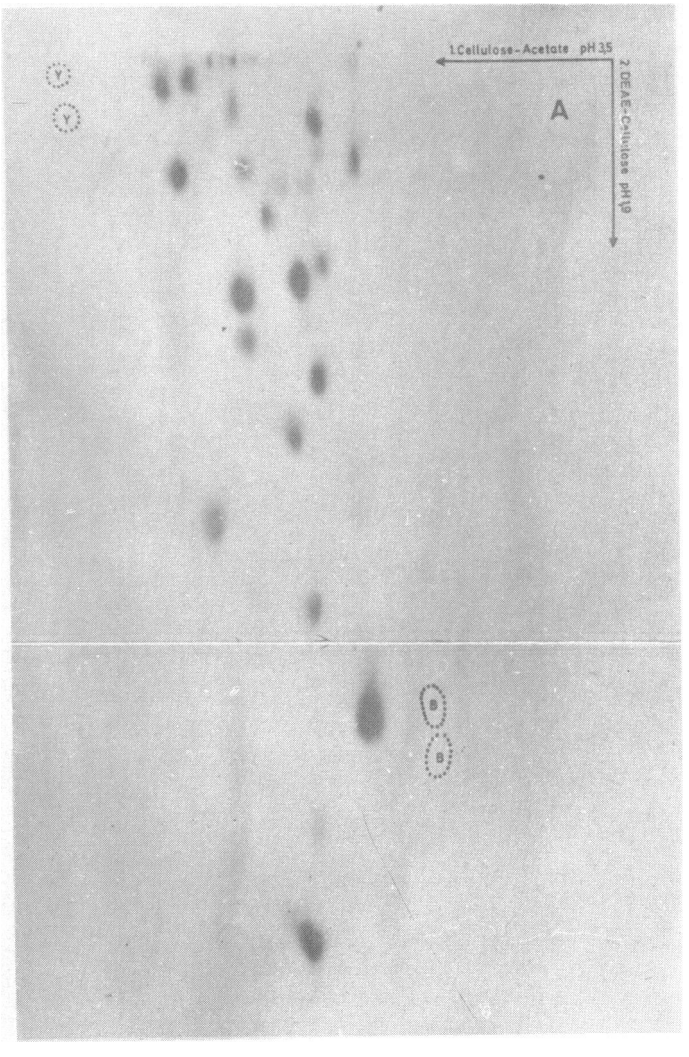

Figure 3

Fingerprints of normal and dC-substituted transcripts of SV40 Hind-fragment H.

A. SV40 Hind-fragment H was prepared and transcribed as described elsewhere (6). The labeled precursor was α-^{32}P -ATP. The transcript was digested with P. RNase and fingerprinted (8). B and Y are standard dye markers.

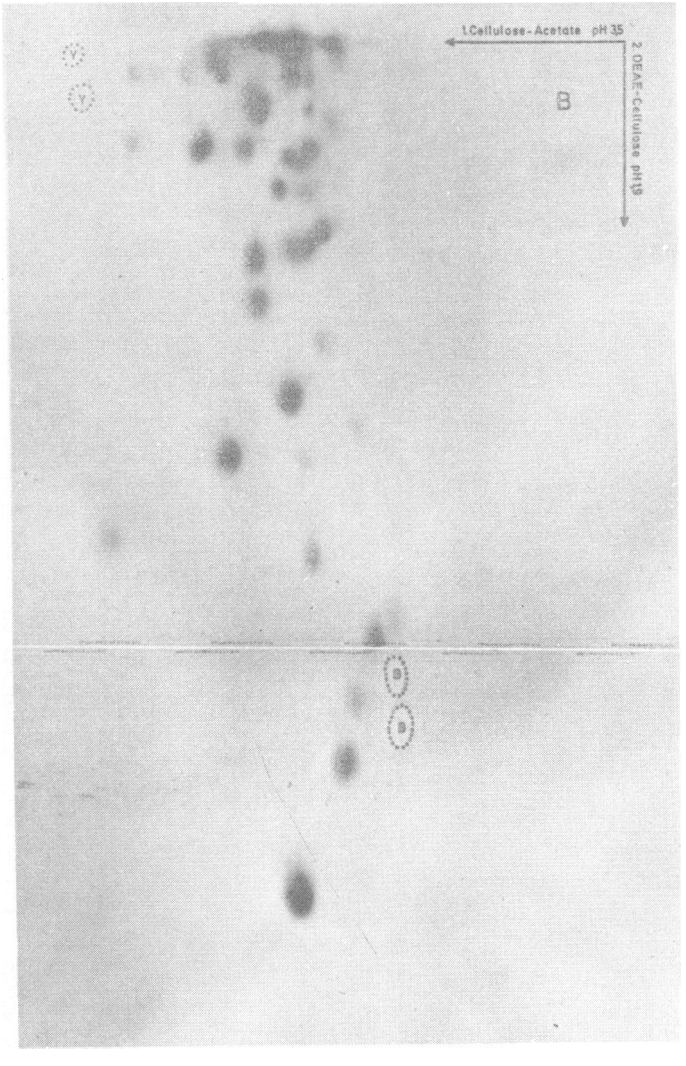

Figure 3B

B. Same as A, except that the reaction mixture contained dCTP instead of CTP, 1.5 mM $MnCl_2$ and 5 mM $MgCl_2$.

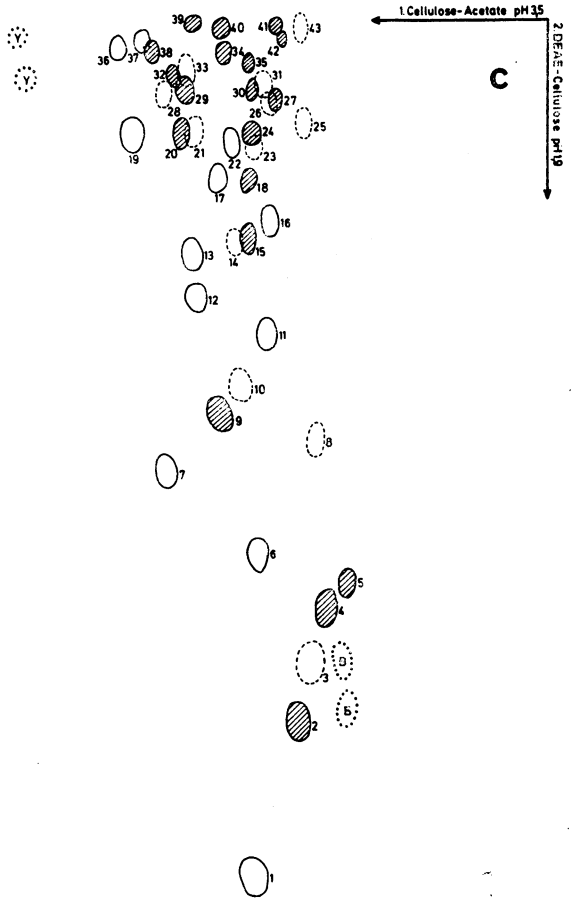

Figure 3C

C. Diagrammatic explanation of the results shown in panels A and B. Open, full line spots are present in A and B, Open, dashed line spots have disappeared in B or are considerably weaker. Shaded spots are only present in B

1. A-U	6. A-A-U	10. G-G-C	19. G-G-U
2. A-C-U + C-A-U	7. G-U	12. G-A-U	
3. A-G-C	9. G-C-U	13. A-G-U	

Characterization of the deoxyribonucleotide-containing oligonucleotides

The substituted RNA was digested with "total" RNase or with alkali and fingerprinted as described under Methods. Even under our best conditions we could still observe some ribonucleotide contamination of the type being substituted. Further work is in progress to eliminate this noise completely. In addition, three series of oligonucleotides were present on each fingerprint, e.g. in the dC-substitution these corresponded to $(dC)_n$ U, $(dC)_n$ A and $(dC)_n$ G. The number of spots in a series indicated that E.coli polymerase is able to incorporate 2, 3 and even more deoxyresidues in a row. The identity of each product could usually be deduced from its position on the map and was further confirmed by spleen exonuclease hydrolysis.

In order to allow rapid identification of all the d-oligonucleotides a reference system was developed. The mobility of all d-oligonucleotides on DEAE-paper electrophoresis at pH 3.5 was determined relative to a standard double digest (D D; T_1-RNase + P. RNase). These results are summarized in Fig. 2. On this basis, most d-oligonucleotides can be readily identified. For example if a T_1-oligonucleotide from a map of a dU-substituted RNA produces upon digestion with "total" RNase a spot which moves just ahead of A-C and another one which moves like A-A-C, this T_1-oligonucleotide must contain the sequences dU-A and dU-dU-G.

An example of the deoxynuclectide substitution technique

As discussed above, the major application of this method is in ordering of oligonucleotides on the basis of the new overlap which can be obtained. Such results are illustrated in Fig. 3 for a transcript of the SV40 Hind-fragment H. A comparison between the fingerprints of a P - RNase digest of a normal transcript and the P - RNase digest of a dC-substituted product reveals that in the latter most of the C-terminal oligonucleotides are absent or at least greatly reduced. On the other hand a series of new spots appear, and further analysis reveals that all these contain at least one dC-residue and hence are formed by fusing two or more oligonucleotides from the former digest. For example the spot 14 (A-G-G-

C) reappears as 40 (G-C-C-A-A-C-A-G-G-C-U), the spot 25 (A-A-A-A-A-G-C) reappears as 39 (A-A-A-A-A-G-C-G-G-G-U) and the spot 31 (A-G-A-G-A-A-C) reappears as 34 (A-G-A-G-A-A-C-A-A-A-U). Further work is in progress on the longer overlapping sequences which can be isolated by the use of homochromatography in the second dimension.

CONCLUSION

The deoxynucleotide substitution technique allows hydrolyses of the RNA chain specifically at any of the 4 base residues. Hence this method is very promising for nucleotide sequence studies of in vitro synthesized RNA. Indeed in principle it allows the ordering of all oligonucleotides in a chain of nearly 1,000 residues in length, thus eliminating the need for the time consuming, and inefficient partial digestions. Moreover, the sequence determination of the longer T_1-oligonucleotides can be considerably simplified.

Other applications of the technique can also be considered. For example, it seems quite possible that a deoxyribonuclease could be found which splits only between two deoxynucleotide residues (10); this would cleave the substituted RNA chain into oligonucleotides of average chain length 16. The advantages of this are obvious.

One drawback of the technique is that the transcription is only 5-10 % as efficient relative to the normal reaction. This can be remedied by using higher specific activity precursors. Also other RNA-polymerases may perhaps be found which are less stringent regarding the sugar mdety of the precursors. Sofar it does not appear that the Mn^{++}-ions induce an unacceptable level of errors in the transcription. Although it is obvious from the results presented here that the polymerase can incorporate two and more deoxyresidues in a row, we do not yet know if the transcripts are shorter than normal and especially if all regions of the template are transcribed. Further work to study these aspects is in progress.

ACKNOWLEDGEMENT

It is our pleasure to thank Dr. R. Burgess, H. Schaller and R. Schilperoort for generous gifts of RNA-polymerase.

We are grateful to José Van der Heyden and Greet Van Houwe for expert technical assistance. J.V.H. thanks the I.W.O.N.L. for a fellowship; R.G. works in the Laboratory of Histology and Genetics.
This research was supported by the "A.S.L.K.-Kankerfonds".

REFERENCES

1 Berg, P., Fancher, H. & Chamberlin, M.(1963) in Symposium on Informational Macromolecules, pp. 467, ed. Vogel, H., Bryson, V. & Lampen, J.O., Academic Press, New York
2 Sanger, F.(1973) in Virus Research, pp. 573-599, ed. Fox, C.F. & Robinson, W.S., Academic Press, New York
3 Salser, W., Fry, K., Brunk, C. & Poon, R. (1972) Proc. Nat. Acad. Sci. (U.S.A.)69, 238-242
4 Chamberlin, M.J.(1966) in Procedures in Nucleic Acid Research, pp. 513, ed. Cantoni, G.L. & Davies, D.R., Harper & Row, New York
5 Hurwitz, J., Yarbrough, L. & Wickner, S. (1972) Biochem. Biophys. Res. Comm. 48, 628-635.
6 Fiers, W., Danna, K., Rogiers, R., Vandevoorde, A., Van Herreweghe, J., Van Heuverswyn, H., Volckaert, G. & Yang, R. (1974) Cold Spring Harbor Symposium (in press)
7 Hiramaru, M., Uchida, T. & Egami, F. (1966) Anal. Biochem. 17, 135-142
8 Sanger, F. & Brownlee, G.G. (1967) in Methods in Enzymology, vol. 12A, pp. 361, ed. Grossman, L. & Moldave, K., Academic Press, New York
9 Brownlee, G.G. & Sanger, F. (1969) European J. Biochem. 11, 395
10 Bernardi, G., Ehrlich, S.D. & Thiery, J.P. (1973) Nature New Biology 246(150),36-40.