Proc. Nat. Acad. Sci. USA
Vol. 70, No. 4, pp. 1209–1213, April 1973

Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage f1 DNA

(octadeoxyribonucleotide/pulse-labeling/homochromatography/intercistronic region/evolution)

F. SANGER*, J. E. DONELSON*, A. R. COULSON*, H. KÖSSEL†, AND D. FISCHER†

* Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, England; and † Institut für Biologie III (Genetik und Molekularbiologie), Universität Freiburg, 78 Freiburg, W. Germany

Contributed by F. Sanger, January 16, 1973

ABSTRACT **A sequence of 50 residues in f1 DNA has been determined by the extension of a chemically synthesized octadeoxyribonucleotide by *Escherichia coli* DNA polymerase I, with radioactive nucleoside triphosphates and f1 DNA template. The polymerized product was synthesized either in the presence of manganese and a mixture of ribo- and deoxyribotriphosphates or in a magnesium-containing reaction with one or more of the four triphosphates absent. The sequence determination depended largely on fractionation of the polymerized products by two-dimensional "homochromatography." This approach and the techniques for the subsequent sequence analysis should be of general use for determining other sequences of DNA. Several features of this sequence suggest that it is located in an intercistronic region of f1 DNA.**

Techniques for determining the nucleotide sequence of various radioactive RNA species are well developed (1). Since naturally occurring DNA molecules are very large, the application of many of these techniques to DNA sequence analysis involves one of two initial approaches. Either a uniformly labeled radioactive DNA molecule must be degraded to small fragments (20–100 nucleotides long) that can be isolated for sequence studies (2–4), or the radioactivity must be introduced selectively *in vitro* into a small specific region of an unlabeled DNA molecule (5–8). This report describes a method that uses the second approach with *Escherichia coli* DNA polymerase I (9).

The aminoacid sequence Trp-Met-Val in positions 26–28 of the major coat protein (49 amino acids) of bacteriophage fd has been reported (10). Since the plus strand is the "sense" strand (11), and since tryptophan and methionine are each coded for by a unique codon, the corresponding octadeoxyribonucleotide sequence A-C-C-A-T-C-C-A in the minus strand can be predicted. We have chemically synthesized this sequence on the assumption that it would serve as a unique primer, i.e., a unique starting point, for the incorporation of nucleoside triphosphates by *E. coli* DNA polymerase I, using fd or f1 plus strand as template. The sequence of nucleotides polymerized onto this unique primer–template could then be studied.

The octanucleotide functioned efficiently as a primer on f1 and fd DNA, but we soon realized that the sequence made was not that expected from the published sequence of the coat protein. However, as the priming appeared to be specific, the reaction could be used for the development of methods for DNA sequencing, and has led to the elucidation of a sequence of 50 residues in f1 DNA that may be of considerable biological interest. This paper is a preliminary account of the work; details of the sequence analyses will be published elsewhere.

Previous successful use of the pulse-labeling technique (12, 13) has been with RNA, and has relied mainly on the analysis of ribonuclease T1 digests of the synthetic product obtained after various time intervals. We have used some different approaches that depend largely on the fractionation of the synthetic products of different lengths by a two-dimensional "homochromatography" (14). This system was used directly to determine the sequence of the first 12 residues, using only nearest-neighbor analysis and depurination (15). To determine further sequences it was necessary to have a specific method of degrading the DNA. Berg *et al.* (16) have shown that it is possible to incorporate a ribonucleotide into a deoxynucleotide sequence with DNA polymerase I if manganese is present in the incubation medium instead of magnesium. If a single ribotriphosphate is used with the other three deoxyribotriphosphates, it is incorporated to give bonds that are susceptible to alkali or an appropriate ribonuclease. In theory, this reaction offers an attractive approach to sequencing DNA as different ribonucleotides can be introduced, giving rise to different products, which should overlap one another. We have, therefore, applied this method to the octanucleotide-primed synthesis and, using largely rCTP incorporation, have isolated specific fragments by digestion with pancreatic ribonuclease A and ordered them by studying their distribution in the synthetic products of differing lengths.

MATERIALS

The octadeoxyribonucleotide A-C-C-A-T-C-C-A was synthesized chemically by procedures to be presented elsewhere. DNA was extracted with phenol from phage f1 (a gift of H. D. Robertson). [α-^{32}P]Deoxynucleoside triphosphates (10–100 Ci/mmol) were synthesized by the method of Symons (17). *E. coli* DNA polymerase I was the gift of T. Jovin.

RESULTS

Octanucleotide-Primed Incorporation in the Absence of One or More Nucleotide Triphosphates. Preliminary experiments showed that the octanucleotide could serve as a primer with either fd or f1 DNA as template. In order to determine the sequence of the first few nucleotides incorporated onto the octanucleotide, various incubations were conducted in the absence of one or more nucleoside triphosphates. After a short

Abbreviations: The symbols A, C, G, and T refer to the deoxyribonucleosides, and rC and rG refer to the ribonucleosides.

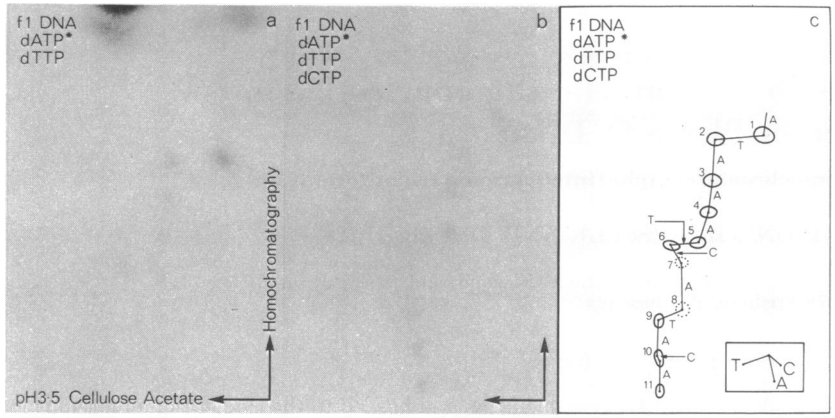

FIG. 1. Fractionation of products synthesized in the absence of one or more nucleoside triphosphates. (*a*) Radioautograph of a fractionation. The reaction mixture (10 μl) contained 0.6 μg of f1 DNA, 0.2 μg of octanucleotide in 50 mM Tris·HCl (pH 7.4), 30 mM NaCl, 8 mM $MgCl_2$, 2 mM dithiothreitol, 0.1 mM dTTP, 1–5 μM [^{32}P]dATP (50 Ci/mmol), and 1.0 unit of DNA polymerase I. After 1 min at 8°, EDTA was added to 20 mM and the reaction mixture was spotted directly on cellulose acetate for ionophoresis at pH 3.5. The second dimension was in a 3% homomixture hydrolyzed for 30 min (2). (*b*) Same as *a*, except that 0.1 mM dCTP was also present during the incubation for 2 min at 8°. (*c*) Composite diagram of *a* and *b* showing the course of synthesis of the oligonucleotide chain. The *inset* shows the predicted effect of the addition of one residue to an oligonucleotide (2, 3).

incubation, the reactions were run directly on the two-dimensional system, which resolved the [^{32}P]deoxyribotriphosphates that had not reacted from labeled oligonucleotides. In a reaction containing TTP and [^{32}P]dATP, six oligonucleotides of increasing size were observed (Fig. 1*a*) and, from their relative mobilities (2, 3) and nearest-neighbor analyses, the sequence of the first six nucleotides polymerized onto the octanucleotide could be deduced. As polymerization stopped at this point, the next position must be either a C or G residue. Fig. 1*b* shows the results of an experiment in which dCTP was also present in the reaction mixture, and Fig. 1*c* shows a stylized version of the autoradiograph. Again, from nearest-neighbor analysis and depurination of the larger oligonucleotides, a sequence of 12 nucleotides could be deduced. In many experiments spots **7** and **8** (Fig. 1*c*) were much weaker than the other spots, or were not even observed; this decrease was due to variations in the asynchronous incorporation of nucleotides by the polymerase, because the relative mobilities and nearest-neighbor analyses of the larger oligonucleotides remained the same. Spot **10** is a mixture of two unresolved oligonucleotides differing by the addition of a 3′-terminal pC residue. This combination was detected by a nearest-neighbor analysis of the larger oligonucleotide in spot **11**. From position differences of these synthesis products on the fingerprint, much of their sequence could be "read off" directly, in the same manner as that of degradation products of partial exonuclease digests (2, 3).

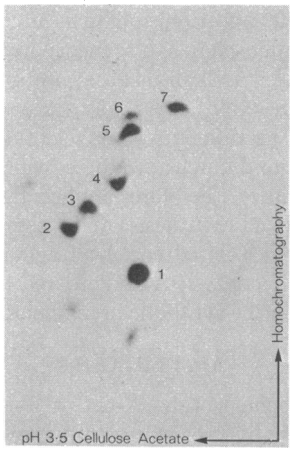

FIG. 2. Radioautograph of a fractionation of a ribonuclease A digest of the mixed product formed by octanucleotide-primed synthesis. The reaction mixture (20 μl) contained 67 mM Tris·HCl (pH 7.4), 0.67 mM $MnCl_2$, 1.5 mM 2-mercaptoethanol, 1.0 mM rCTP, 0.05 mM each of dGTP and dTTP, 1.2 μg of f1 DNA, 0.4 μg of octanucleotide, 2–4 units of DNA polymerase I, and about 0.5 μCi of [^{32}P]dATP (50 Ci/mmol). Incubation was at 25° for 10 min, and 0.13 μCi of [^{32}P] was incorporated. The products were purified by an unpublished procedure of V. Ling involving precipitation with ethanol and with HCl, and digested with ribonuclease A (enzyme to substrate ratio 1:3, 37°, 2 hr). The first dimension was separated by ionophoresis at pH 3.5 on cellulose acetate, the second in a 5% homomixture hydrolysed for 10 min (14).

Incorporation Experiments in the Presence of rCTP. When f1 DNA and the octanucleotide were incubated with DNA polymerase I in the presence of manganese, rCTP, and the other three deoxyribotriphosphates, rCMP was incorporated in place of dCMP, and subsequent treatment with ribonuclease A cleaved the polynucleotide chain at the rC residues. These products could then be fractionated by the two-dimensional system. Fig. 2 shows a fingerprint in which [^{32}P]dATP was used. There are seven strong spots and a number of fainter ones, which are more apparent if longer times of incubation are used. The strong spots are numbered **1–7**, and are considered here since they represent the first 50 residues incorporated onto the octanucleotide. Most of the other spots presumably represent sequences from further along the chain, but there are a few that are present as faint spots at early stages of incorporation and represent some type of artifact, perhaps a second initiation site with the octanucleotide or miscopying by the polymerase. The sequences of the seven components were determined, by nearest-neighbor analysis with different [^{32}P]deoxyribotriphosphates (5), depurination (3), and partial digestion with exonucleases (2, 3, 19), and are shown in Table 1. The partial digestion of spot **1** by spleen exonuclease showed that it contained the octanucleotide sequence, and the remaining part of the sequence confirmed that structure already described above.

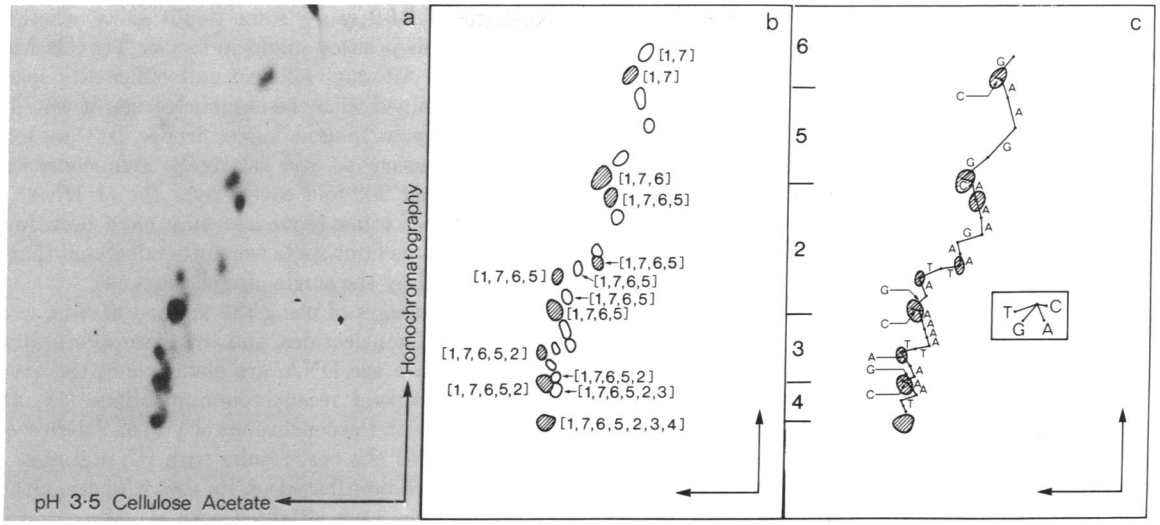

FIG. 3. Fractionation of the products synthesized by DNA polymerase I in the presence of the octanucleotide, Mn, and rCTP. (*a*) Radioautograph of a fractionation. Incubation was for 4 min at 10°, as described in Fig. 2. The product was purified by precipitation with ethanol, and applied to a cellogel strip for fractionation in the first dimension. The second dimension was in homomixture *b* (14). (*b*) Composite diagram showing spots observed in several experiments similar to that shown in *a*, but with various times of incubation and different labels (dATP and dGTP). The oligonucleotides appearing as strong spots in *a* are indicated by *shading*. The products from the spots were digested with ribonuclease A (50 μg per spot, overnight, 37°), and the oligonucleotides were identified by one-dimensional chromatography in a 3% homomixture hydrolysed for 10 min, or by ionophoresis on DEAE-paper in 7% formic acid (19), together with marker oligonucleotides 1–7 from a digest of unfractionated material. These are shown in *brackets* on the diagram. (*c*) Diagram showing the relationship between the position of the spots and the course of synthesis of the oligonucleotide chain. Each spot (Fig. 3*b*) or predicted spot is indicated by a *dot*. The dots are joined by *lines*, whose directions indicate the residue added. The *numbers* to the left of the diagram show at what level on the fingerprint the different sequences, represented by oligonucleotides 2–6 (Table 1), appear in the synthetic product.

Order of the Oligonucleotides. Various methods were investigated for determining the order of the oligonucleotides listed in Table 1. The most successful was to fractionate on the two-dimensional system the product synthesized by DNA polymerase I in the presence of Mn and rCTP. By use of a relatively strong unhydrolyzed "homomixture" (14), fragments of up to 60 residues could be fractionated and then digested with pancreatic ribonuclease A to identify the components. Fig. 3*a* shows such a fractionation.

The presence of certain relatively strong spots well separated from each other indicates that the rate of synthesis by the polymerase is not constant along the chain, but is held up at certain residues. Much of this effect is due to the rC residues, which the polymerase has difficulty in incorporating, though in some experiments there was also an apparent delay at the radioactively labeled residues—presumably because the corresponding triphosphate was present in very low concentrations. The spots from the fingerprint were eluted and digested with ribonuclease A; the digestion products were identified by one-dimensional chromatography. The combined results from several experiments are summarized in Fig. 3*b*. There were several products that did not correspond to any of the markers and appeared in only one or two fragments, but not in the larger fragments. These were intermediates from the 3′ end of the synthesized chain. After digestion with ribonuclease A, this 3′-hydroxyl terminus is liberated as an oligonucleotide that is different from the markers and, if it is labeled, will appear on the analysis.

From the results in Fig. 3*b* the oligonucleotides can be arranged in the order **1–7–6–5–2–3–4**. It was important to be certain that no other residues could be interspersed between the recognized oligonucleotides numbered **1–7**. Table 1 shows that all these oligonucleotides have Ap at their 5′ ends. Thus, if there were any product lying between them in the sequence, it would be labeled in an experiment where [^{32}P]dATP was used. As no such products were present, we concluded that the above oligonucleotides are linked directly in the sequence shown in Fig. 4*a*. Two further lines of evidence support this conclusion. Tentative sequences were ascribed to the products derived from an alkaline digest of the material from an incorporation with rGTP; these sequences precluded the presence of any intermediates.

Fig. 3*c* is a diagram showing the relationship between the course of synthesis of the DNA chain and the relative positions of the spots on the two-dimensional fractionation system. This diagram is based on the above results and on the expected effect of the addition of a residue to a nucleotide chain, as indicated in the inset and as already discussed above (Fig. 1). It is not possible to use the fingerprint to determine the sequence directly, since many intermediates are not seen and

TABLE 1. *Oligonucleotides isolated from a ribonuclease A digest of the mixed product containing rC*

Oligonucleotide (Fig. 2)	Sequence
1	A-C-C-A-T-C-C-A-A-T-A-A-A-T-rC(A)
2	A-A-A-G-A-A-T-T-A-G-rC(A)
3	A-A-A-A-T-T-A-A-G-rC(A)
4	A-A-T-A-A-A-G-rC(rC)
5	A-A-G-G-rC(A)
6	A-G-G-rC(A)
7	A-T-A-rC(A)

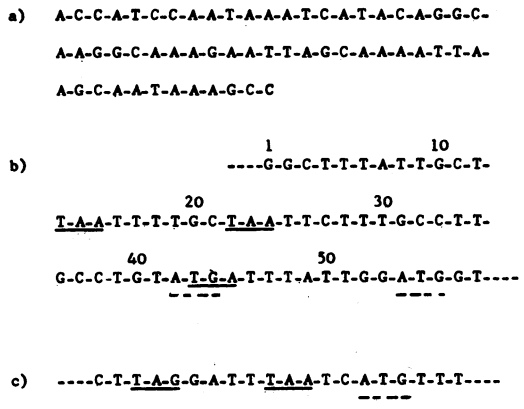

FIG. 4. DNA sequence in phage f1 DNA. (*a*) Sequence of the minus strand, as determined in this paper. The first eight residues are from the synthetic octanucleotide used for priming. (*b*) The compliment of the sequence in (*a*); this is the sequence in the plus strand and in the messenger RNA. *Underlining* indicates termination codons; *broken underlining* indicates initiation codons. (*c*) Sequence around the initiating site of a structural protein of ϕX DNA (4).

the differences in position are very small for the larger products. The fact that it is possible to draw such a diagram gives useful confirmation of the sequence and precludes the possibility of any extensive sequence being missed. Certain features, such as the two –T–T– sequences, are readily recognized on the fingerprint (Fig. 3*a*).

DISCUSSION

The methods described in this paper should be of fairly general application to the determination of sequences in DNA, though no doubt each situation will present its own problems. In theory, this procedure, requires only a synthetic oligonucleotide complimentary to a known sequence and conditions where polymerase will copy faithfully. In practice, the choice of the octanucleotide primer in this work was entirely fortuitous as it does not prime in the expected position in the coat protein cistron. The recent report by Snell and Offord (20) on the aminoacid sequence of the coat protein of the related phage ZJ-2 raises some doubt as to whether the sequence in this position might in fact be Trp-Ala-Met instead of Trp-Met. Although efficient and sufficiently specific priming was obtained with the octanucleotide, it would probably be advantageous to use a longer primer. In these experiments it was necessary to use relatively high concentrations of primer (about 200-fold excess over the f1 DNA), and it is probable that other sequences may have been formed to a lesser extent as faint spots were often detected that could not be explained by the single unique sequence.

The advantages of using the method of Berg *et al.* (16) to introduce ribonucleotides, and therefore specifically susceptible bonds into the DNA, are obvious and this has been the subject of several recent communications (18, 21, 22). In agreement with the conclusions of Van de Sande *et al.* (21), we have obtained the best results with rC, and most of the sequences have been based on its use, whereas, although good incorporations were obtained with rG, the fingerprints were rather complex and somewhat variable. This may have been partly due to the nature of the sequence involved. Others (21, 22) have concluded that DNA polymerase I has difficulty in inserting two consecutive ribonucleotides. The first –C–C– is at the 3′ end of the deduced sequence, whereas the first G residues are in a –G–G–C–A–A–G–G– sequence. It may be that in a different situation the use of rG may be more advantageous. As might be expected from the large number of A residues, almost no incorporation of rA could be obtained.

An important question to be considered is whether DNA polymerase I is copying faithfully under these special conditions and, therefore, whether the sequence determined is the exact compliment of the plus strand of f1 DNA. In particular, the unusually large number of A residues present in apparently homologous sequences raised the possibility of some artifact of incorporation. Van de Sande *et al.* (21) report that misincorporation with rCTP may occur at 37°, but not at 10°. Most of our experiments were done at 25°, but all essential results have now been confirmed at 10°. Further evidence that the copying is faithful is the agreement with the results obtained under normal conditions with three deoxyribotriphosphates. However, the main evidence is that it is possible to deduce a unique sequence with no evidence of randomiza-

FIG. 5. Hypothetical course of evolution of the sequence in positions 1–39 (Fig. 4*b*). *Underlining* indicates the position in which the next mutation event takes place.

tion. If misincorporation were occurring, it would not be specific and the products would be expected to be much more heterogeneous.

The sequence determined (Fig. 4*a*) corresponds to that of the minus strand of f1 DNA. The complimentary sequence, which corresponds to the plus strand and the messenger RNA, is shown in Fig. 4*b*. The last eight residues are based on the assumption that the synthetic octanucleotide used as a primer initiates by hybridizing with its exact complimentary sequence on the template DNA. This assumption seems likely, since such a small oligonucleotide would not be expected to prime unless it paired perfectly, but there is no direct proof.

The sequence is remarkable in that it is very rich in T residues and shows considerable repetition of homologous sequences. Thus, T–T–G–C is repeated four times, T-G-C-C-T twice, and T–A–A–T–T twice. The most likely explanation is the way in which the DNA has evolved. It is probable that the increase in size of a DNA molecule is brought about by duplication of existing sequences, which will then evolve by mutation to give functional DNA. Thus, at early stages of evolution one would expect to find repeated sequences. Such repetitions have been found in RNA sequences (23) and particularly in satellite DNA (24). Fig. 5 shows one way in which the sequence from position 1–39 could have evolved from the pentanucleotide T–G–C–C–T by a process involving six duplications, two insertions of single residues (which could be regarded as single residue duplications), and only five base changes. Clearly there are many other ways in which the evolution could have taken place, but this scheme does illustrate that a sequence of the type found here might be found shortly after an increase in DNA size has occurred by duplication of relatively short sequences. This mechanism would imply either that the DNA section has evolved relatively recently, or else that it acquired an essential function at an early stage in evolution and thus remained stable. This latter possibility is suggested by the further speculation given below.

Robertson *et al.* (4) have recently determined the sequence of a ribosomal binding site in ϕX 174 DNA that was found to contain the initiation point for one of the structural proteins of the virus (25). This sequence is compared in Fig. 4*c* with the sequence of f1 DNA described here, and there are certain interesting homologies. Both are rich in T residues and contain the sequence G–A–T–T–T–A shortly before an A–T–G triplet (positions 54–56). The sequence Pu–Pu–T–T–T–Pu is characteristically found in this position relative to initiation sites for several protein cistrons (4).

This finding suggests the possibility that the A–T–G of this f1 DNA fragment might be an initiation site for a protein. Moreover, the sequence contains three termination codons, T–A–A in positions 13–15 and 22–24, and T–G–A in position 43–45. All three are in phase with each other, so that only the first one could actually function as a protein chain terminator, but the other two might act as a defense mechanism against incomplete termination. A similar situation is found at the end of the coat protein cistron of phage R17 (26), which has three termination codons in phase with each other, the first two being adjacent. It is thus tempting to speculate that the sequence could represent the intercistronic space between two protein cistrons, one of which ends in the aminoacid sequence Gly–Phe–Ile–Ala and the next one of which starts with formylMet–Val. It is hoped that further work will reveal whether this is the case. If it were not, there are two possible aminoacid sequences for which the DNA fragment could code, since two of the three possible phases do not contain termination codons.

Some preliminary experiments in connection with this work were carried out with Dr. V. Ling. We thank Dr. T. Jovin for a gift of DNA polymerase I and Dr. H. D. Robertson for phage f1. J.E.D. is supported by the Helen Hay Whitney Foundation. H.K. wishes to acknowledge support from the Deutsche Forschungsgemeinschaft (SFB 46) and from EMBO.

1. Barrell, B. G. (1971) *Procedures in Nucleic Acids*, eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 2, pp. 751–779.
2. Ling, V. (1972) *J. Mol. Biol.* **64**, 87–102.
3. Ziff, E. B., Sedat, J. W. & Galibert, F. (1971) *Nature New Biol.* **241**, 34–37.
4. Robertson, H. D., Barrell, B. G., Weith, H. L. & Donelson, J. E. (1973) *Nature New Biol.* **241**, 38–40.
5. Wu, R. & Taylor, E. (1971) *J. Mol. Biol.* **57**, 491–511.
6. Padmanabhan, R. & Wu, R. (1972) *J. Mol. Biol.* **65**, 447–467.
7. Padmanabhan, R., Padmanabhan, R. & Wu, R. (1972) *Biochem. Biophys. Res. Commun.* **48**, 1295–1302.
8. Oertel, W. & Schaller, H. (1972) *FEBS Lett.* **27**, 316–320.
9. Kornberg, A. (1969) *Science* **163**, 1410–1418.
10. Asbeck, V. F., Beyreuther, K., Köhler, H., Wettstein, G. & Braunitzer, G. (1969) *Hoppe-Seyler's Z. Physiol. Chem.* **350**, 1047–1066.
11. Hayashi, M., Hayashi, M. N. & Spiegelman, S. (1963) *Proc. Nat. Acad. Sci. USA* **50**, 664–671.
12. Billeter, M. A., Dahlberg, J. E., Goodman, H. M., Hindley, J. & Weissmann, C. (1969) *Nature* **224**, 1083–1086.
13. Blattner, F. R. & Dahlberg, J. E. (1972) *Nature New Biol.* **237**, 227–232.
14. Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395–399.
15. Burton, K. (1967) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 12, part A, pp. 222–224.
16. Berg, P., Fancher, H. & Chamberlin, M. (1963) in *Symposium on Informational Macromolecules*, eds. Vogel, H., Bryson, V. & Lampen, J. O. (Academic Press, New York & London), pp. 467–483.
17. Symons, R. H. (1969) *Biochim. Biophys. Acta* **190**, 548–550.
18. Salser, W., Fry, K., Brunk, C. & Poon, R. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 238–242.
19. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398.
20. Snell, D. T. & Offord, R. E. (1972) *Biochem. J.* **127**, 167–178.
21. Van de Sande, J. H., Loewen, P. C. & Khorana, H. G. (1972) *J. Biol. Chem.* **247**, 6140–6148.
22. Wu, R., Donelson, J. E., Padmanabhan, R. & Hamilton, R. (1972) *Bull. Inst. Pasteur, Paris* **70**, 203–233.
23. Brownlee, G. G., Sanger, F. & Barrell, B. G. (1968) *J. Mol. Biol.* **34**, 379–412.
24. Southern, E. M. (1970) *Nature* **227**, 794–798.
25. Air, G. & Bridgen, J. (1973) *Nature New Biol.* **241**, 40–41.
26. Nichols, J. L. (1970) *Nature* **225**, 147–151.