J. Mol. Biol. (1975) **94**, 441–448

A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase

F. Sanger and A. R. Coulson

Medical Research Council
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

(Received 20 December 1974)

A simple and rapid method for determining nucleotide sequences in single-stranded DNA by primed synthesis with DNA polymerase is described. It depends on the use of *Escherichia coli* DNA polymerase I and DNA polymerase from bacteriophage T4 under conditions of different limiting nucleoside triphosphates and concurrent fractionation of the products according to size by ionophoresis on acrylamide gels. The method was used to determine two sequences in bacteriophage ϕX174 DNA using the synthetic decanucleotide A-G-A-A-A-T-A-A-A-A and a restriction enzyme digestion product as primers.

1. Introduction

In previous papers (Sanger *et al.*, 1973; Donelson, J. E., Barrell, B. G., Weith, H. L., Kössel, H. & Schott, H., unpublished data) we have described the determination of two nucleotide sequences in bacteriophage DNA using DNA polymerase primed by synthetic oligonucleotides. In this method the oligonucleotide was bound to a specific complementary region on the single-stranded DNA, and nucleoside triphosphates were added by the DNA polymerase to the 3′ end of the primer. By using ^{32}P-labelled triphosphates, a radioactive complementary copy of a defined part of the template DNA was synthesized and subjected to sequence determination, which was greatly facilitated by use of the ribosubstitution technique.

This paper describes a relatively rapid and simple alternative procedure for deducing sequences by primed synthesis with DNA polymerase.

2. Principle of the Method

Figure 1 illustrates the principle of the method by considering its application to a small hypothetical sequence in a DNA chain. DNA polymerase I is first used to extend the primer oligonucleotide and copy the template in the presence of the four deoxyribotriphosphates, one of which is labelled with ^{32}P. Ideally this synthesis should be non-synchronous and as random as possible, so that the maximum number of oligonucleotides of different length, all starting from the primer, is formed. This mixture is then purified on an agarose column to remove the excess triphosphates and samples are re-treated in various ways as follows:—

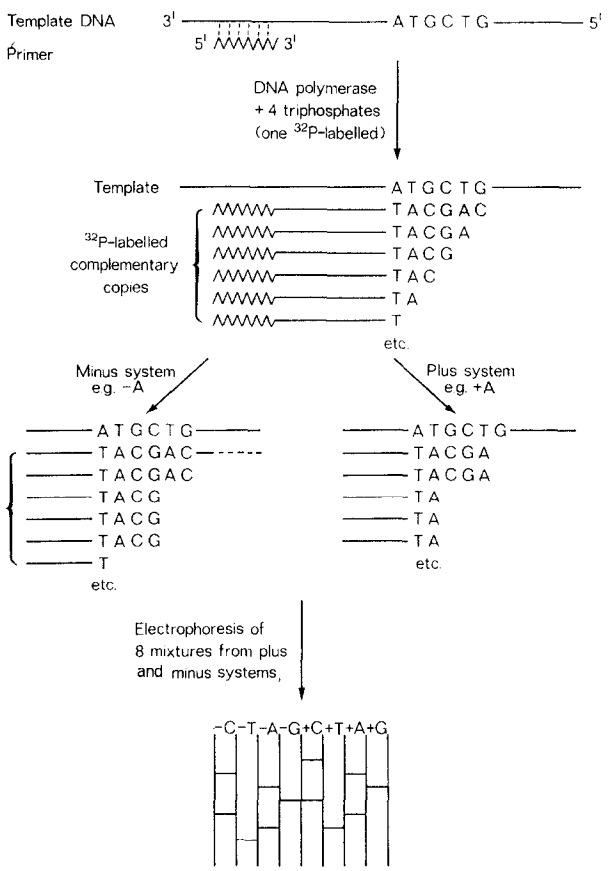

FIG. 1. The principle of the method.

(a) *The "minus" system*

In their original work on the "sticky ends" of phage λ DNA, Wu & Kaiser (1968) showed that if DNA polymerase acted in the absence of one triphosphate, synthesis would proceed accurately up to a position where the missing triphosphate should have been incorporated, and they used this principle to deduce sequences by assaying the relative amount of each nucleotide incorporated in the presence of different triphosphate mixtures (Wu & Taylor, 1971). The minus system described here uses a similar principle.

The random mixture of oligonucleotides, which is still hybridized to the template DNA, is reincubated with DNA polymerase I in the presence of three deoxyribotriphosphates. Synthesis then proceeds as far as it can on each chain: thus, if dATP is the missing triphosphate (the —A† system), each chain will terminate at its 3′ end at a position before an A residue. Separate samples are incubated, with each one of the four triphosphates missing.

† Since this paper is concerned only with DNA the symbols A, C, G and T refer to the deoxyribonucleosides.

The four incubation mixtures are then denatured to separate the newly-synthesized strands from the template, subjected to electrophoresis on acrylamide gel in the presence of 8 M-urea, and a radioautograph prepared. In this fractionation system mobility is essentially proportional to size, so that the various synthesized oligonucleotides (which have a common 5′ end) will be arranged according to size. Ideally each oligonucleotide should be separated from its neighbour, which contains one more residue. The radioautograph from the —A system will contain bands corresponding to positions before the A residues in the synthesized chain. Thus the positions of As are located. Similarly the relative positions of the other residues may be located and, ideally, the sequence of the DNA read off from the radioautograph. This system alone is usually not sufficient to establish a sequence, so a second similar system is normally used in conjunction with it.

(b) *The "plus" system*

This system makes use of the method of Englund (1971,1972) who showed that, in the presence of a single deoxyribotriphosphate, DNA polymerase from bacteriophage T4-infected *Escherichia coli* (T4 polymerase) will degrade double-stranded DNA from its 3′ end, but that this exonuclease action will stop at residues corresponding to the one triphosphate that is present. This method is applied to the random oligonucleotide mixture obtained above. Samples are incubated with T4 polymerase and a single triphosphate and then fractionated by electrophoresis on acrylamide gel. Thus in the +A system only dATP is present and all the chains will terminate with A residues. The positions of A residues will be indicated by bands on the radioautograph. Usually these will be in products one residue larger than the corresponding bands in the —A system, but if there is more than one consecutive A residue the distance between the bands in the —A and +A systems will indicate the number of such consecutive residues. In the example illustrated in Figure 1 the smallest oligonucleotide gives a band in the —T position, indicating that the next residue after its 3′ terminus will be a T. This is confirmed by the presence of a band in the +T position in the next largest oligonucleotide. The bands in the +T and —A positions in this product show that its 3′ terminus is T and the residue following is an A, thus defining the dinucleotide sequence T-A. Similarly the next largest oligonucleotide defines the dinucleotide A-C, and so establishes the sequence T-A-C.

3. Materials and Methods

(a) *Chemicals and enzymes*

^{32}P-labelled deoxyribonucleotide triphosphates were synthesized by the method of Symons (1974) or obtained from New England Nuclear at a specific activity of about 100 mCi/μmol. Phage φX DNA was a gift from H. L. Weith. Phage φX RF (replicative form) was a gift from J. W. Sedat. *E. coli* DNA polymerase (nach Klenow) was obtained from Boehringer Chemical Corporation. *Haemophilus influenzae* restriction enzyme (*Hin*d II and III) was prepared by the method of Smith & Wilcox (1970), and some was a gift from T. Maniatis. T4 polymerase was prepared by the method of Goulian *et al.* (1968), and some was a gift from K. Murray and A. G. Isaksson.

Two experiments will be described here to illustrate the use of the method.

(b) *Experiment 1 (Plate I)*

The initial reaction mixture (100 μl) contained 0·02 M-Tris·HCl (pH 7·4), 0·01 M-$MgCl_2$, 0·01 M-mercaptoethanol, 0·05 mM-dATP, -dTTP and -dCTP, 10 μCi of [^{32}P]dGTP (100

Ci/mmol), 6 μg ϕX174 DNA, 1·0 μg of the decanucleotide primer (Schott, 1974) and 16 units DNA polymerase. Incubation was at 0°C. After 2 min 50 μl were removed and added to 5 μl 0·2 M-EDTA to terminate the reaction. After 8 min 30 μl were removed and inhibited, and the remaining 20 μl incubated for a total of 30 min. The combined incubation mixtures were then applied to a column of Agarose (Bio-Gel A-0·5 m, 200 to 400 mesh, Bio-Rad Laboratories) using a 1-ml disposable plastic pipette. The column was made up and run in 2·0 mM-Tris·HCl (pH 7·4), 0·04 mM-EDTA. The fractionation was followed using a hand radiation monitor and the front band containing the synthesized DNA collected manually. It was concentrated to about 100 μl, and 5-μl samples taken for reincubation with 5 μl of the plus and minus mixtures and 1 μl of the appropriate enzyme.

The minus mixtures contained 0·04 M-Tris·HCl (pH 7·4), 0·02 M-$MgCl_2$, 0·02 M-mercaptoethanol, and the appropriate three dNTPs in 0·02 mM concentration. Incubation with DNA polymerase (0·8 unit in 1 μl) was for 30 min at 0°C.

The plus mixtures contained 0·13 M-Tris·HCl (pH 8·0), 0·013 M-$MgCl_2$, 0·02 M-mercaptoethanol, and the one triphosphate in 0·4 mM concentration. Incubation with T4 polymerase (approx. 0·02 unit in 1 μl) was for 30 min at 37°C.

The reactions were stopped by the addition of 1 μl 0·2 M-EDTA. 25 μl of freshly deionized formamide containing 0·3% xylene cyanol FF and 0·3% bromphenol blue were added and the solutions were heated at 95 to 100°C for 3 min before layering onto the acrylamide gel.

Electrophoresis was carried out on a 15% acrylamide gel (20 cm × 40 cm × 0·1 cm) at room temperature according to the method of Peacock & Dingman (1967). The buffer in the gel and in the anode compartment was Tris–glycine (3·028 g Tris-base, 14·4 g glycine/l)–8 M-urea. The cathode contained the same buffer without urea. It was run at 400 V until the bromphenol blue marker had travelled 30 cm from the origin (approx. 16 h). The gel was covered with "Saran wrap" and radioautographed.

(c) *Experiment 2 (Plate II)*

*Hin*d fragment 1 was prepared by digestion of 100 μg of phage ϕX RF with *Hin*d II + III and purified by ionophoresis on a 5% acrylamide gel (Edgell *et al.*, 1972). The amount obtained from 20 μg RF was mixed with 6 μg single-stranded ϕX DNA in 40 μl water and heated to 95°C for 3 min. 4 μl H × 10 buffer (66 mM-Tris·HCl (pH 7·4), 66 mM-$MgCl_2$, 100 mM-mercaptoethanol, 0·5 M-NaCl) was added and it was incubated at 67°C for 4 h. 20 μl of this annealed material was then incubated with DNA polymerase under the conditions described in section (b), above, in a volume of 50 μl. The labelled triphosphate was [^{32}P]dATP. Half of the solution was incubated for 2 min at 0°C and half for 8 min. These were combined and purified on an Agarose column and the eluate concentrated to 40 μl. 2-μl samples were then taken for reincubation with 2 μl of the plus and minus mixtures and 1 μl of the enzyme solution.

The minus mixtures were prepared by mixing equal volumes of the H × 10 buffer and 0·1 mM solutions of each of the three appropriate triphosphates. The plus mixtures contained 1 vol. H × 10 buffer, 1 vol. of a 0·2 mM solution of the triphosphate and 2 vol. of water. After incubation with enzyme as described above, 1 μl *Hin*d II + III enzyme was added (this was assayed as sufficient enzyme to digest 2·5 μg bacteriophage λ DNA in 1 h at 37°C) and the solutions were reincubated at 37°C for 15 min. The reactions were terminated with EDTA and fractionated on a 12% acrylamide gel. A voltage of 1000 V was applied until the bromphenol blue marker had travelled 35 cm from the origin (4 to 5 h). The gel was allowed to become relatively hot during the run and no cooling was applied.

4. Results

(a) *The sequence primed by A-G-A-A-A-T-A-A-A-A (experiment 1)*

The method was initially tested out using the primer A-G-A-A-A-T-A-A-A-A on ϕX DNA. The sequence of the first 41 residues primed by this decamer was already known (Robertson *et al.*, 1973; Donelson *et al.*, unpublished data) and is

shown in Figure 2. Plate I shows the results of one experiment. In interpreting the results it is essential to know the relative positions of the bands from the different samples on the radioautograph. This is easy where samples are run side by side but involves careful measuring for ones further apart. Thus, to show that the bands in

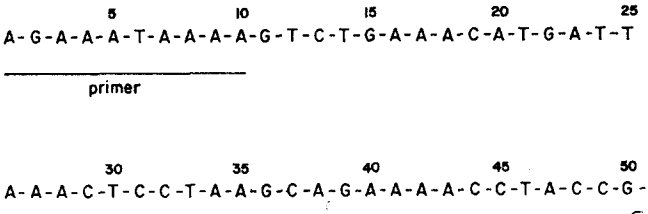

FIG. 2. The sequence of the first 41 residues primed by the decanucleotide on ϕX DNA (Donelson *et al.*, unpublished data).

the +A and −T system are both in position 20, the strong −C band in position 28 was used as a reference. The distance between this −C and the −T in position 20 was the same as the sum of the distance from the −C to the +G in position 22 and that from this +G to the +A in position 20. Plate I also shows the exact lining up of the different bands, from the radioautograph.

If we consider position 18, there are two bands present, one in the +A and one in the −C system. This indicates that the 3′ residue of the oligonucleotide in this position is an A and that the next residue in the chain will be a C. A dinucleotide, A-C, is thus defined. As would be expected, the next position contains a band in the +C column, thus confirming the identification of the C. It also has a −A, which establishes A as the next residue. This is clearly followed by a sequence T-G; however at this stage the sequence is less easy to deduce, due to the presence of artifact bands, which appear to occur in this region and which will be discussed below. From positions 23 to 51 the sequence can be read off readily, and agrees well with the known sequence. Beyond residue 51 a partial sequence can be predicted but there are some uncertain parts, especially where "runs" of the same residue are concerned. For instance, it is clear that there are runs of T and A residues in the position labelled b, suggesting a sequence A_{4-6} T_2 A_{3-4} T_{4-6}, but the exact number of residues cannot be decided. This partial sequence will be discussed in a separate paper in connection with results obtained by other techniques.

(b) *Priming with fragments from restriction enzyme digests*

Besides using synthetic oligonucleotides as primers for DNA polymerase, it is possible to use the specific fragments obtained by the action of restriction enzymes (Maniatis *et al.*, 1974). The method is essentially the same as that used for the oligonucleotide primer except that the products from the incubations with the plus and minus mixtures are digested with the restriction enzyme before applying to the acrylamide gel for electrophoresis. Thus the radioactive oligonucleotides all have the same 5′ terminus, which corresponds to the original cleavage point of the enzyme. Plate II shows the result of an experiment in which fragment 1 from a digest prepared by the action of the restriction enzyme from *H. influenzae* (*Hin*d II + III) on double-stranded ϕX DNA (Edgell *et al.*, 1972) was used as primer on the single-stranded ϕX

DNA. The results here were particularly clear-cut and it was possible to predict a sequence of 70 residues that proved to be essentially correct. This DNA sequence codes for a part of the largest coat protein (from gene *F*) of the bacteriophage and has been largely determined by Sedat J. W., Ziff, E. B. & Galibert, F. (unpublished data) by direct partial digestion methods (see Galibert *et al.*, 1974), and by Blackburn (unpublished data) using transcription methods. Data were also obtained on peptides from the protein by Air (unpublished data). Although neither method gave the complete sequence, the results from the four techniques, which depend on entirely different approaches, left no doubt as to the correct sequence. These results will be discussed in a separate communication.

5. Discussion

In order that the method should give reliable results, various criteria must be satisfied. First, the products synthesized must all have the same 5′ terminus. *E. coli* DNA polymerase I normally has 5′ exonuclease activity; however, this can be removed by digestion with subtilisin and fractionation on Sephadex (Klenow & Henningsen, 1970). Thus DNA polymerase treated in this way was used in this work.

It is essential that the oligonucleotides are fractionated according to size. In preliminary experiments in which 20% acrylamide gels were used this was found not to be the case, some oligonucleotides migrating faster than corresponding smaller ones. This happened particularly in the region marked a in Plate I. The effect has not been encountered on 12% gels, though migration rate is not always exactly proportional to size, and the effect of the addition of a single residue is not exactly constant (see below). The reason for the anomalous migratory behaviour at higher gel concentrations is not clear, though it has been noted before (see Ikemura & Dahlberg, 1973). It may be connected with the secondary structure of the oligonucleotides and is certainly much worse in non-denaturing conditions.

Ideally oligonucleotides of all possible lengths should be present in the initial product of synthesis, so that all residues are represented in the plus and minus systems. In fact it is difficult to achieve this and under all conditions studied we have found that certain products are formed in relatively high yield, whereas others are absent. This suggests that the polymerase acts at different rates at different sites. It may be that this effect is partly related to the secondary structure of the template. It can also depend on the relative concentration of the triphosphates used. Thus if a low concentration of the ^{32}P-labelled triphosphate is used, "piling-up" frequently occurs before these residues. We find that the best results are obtained if synthesis is carried out for short times with a relatively high concentration of polymerase. Nevertheless it is frequently found that some expected products are missing (e.g. the +G in position 36, Plate I), and this constitutes a limitation of the method and is one reason why it is necessary to use both the plus and minus systems.

The main difficulty with the method occurs when consecutive runs of a given mononucleotide are present. Ideally if all oligonucleotides are present in the initial product of synthesis, each component of a run should appear as a band in the minus system, though the smallest component will be the strongest since it will be formed by extension of other smaller products—whereas the larger components are merely the unchanged oligonucleotides from the initial incubation. In Plate I it is possible to see the run of three As in positions 25–27 and of four As in positions 39–42. However it

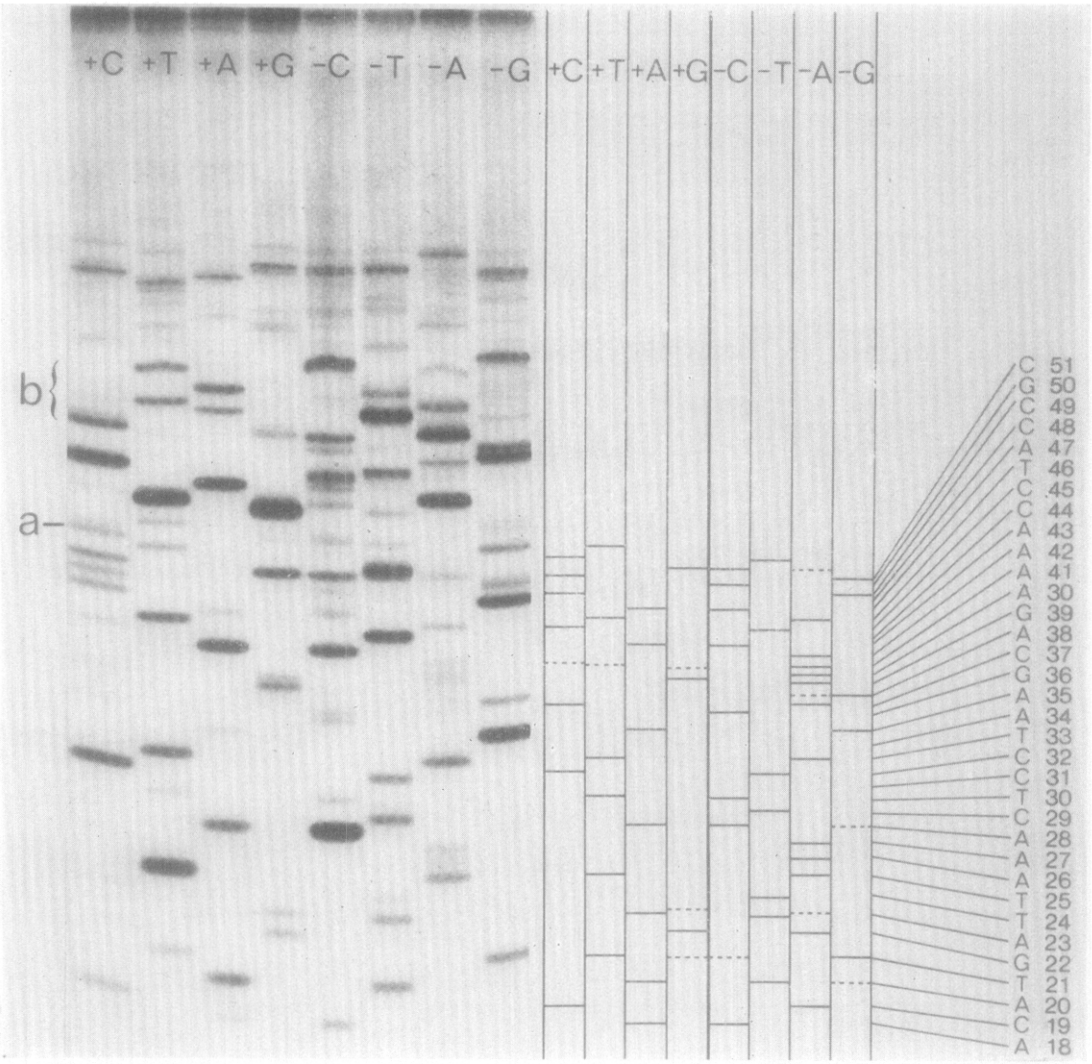

PLATE I. Radioautograph of experiment 1 (see Materials and Methods) in which the decanucleotide was used as a primer. The diagram shows the interpretation and the sequence deduced. The dashed lines represent artifact bands (see text). The bromphenol blue marker was 1 cm below the section of the gel shown in the Plate and the xylene cyanol FF marker opposite the product 50 nucleotides long.

[*facing p. 446*

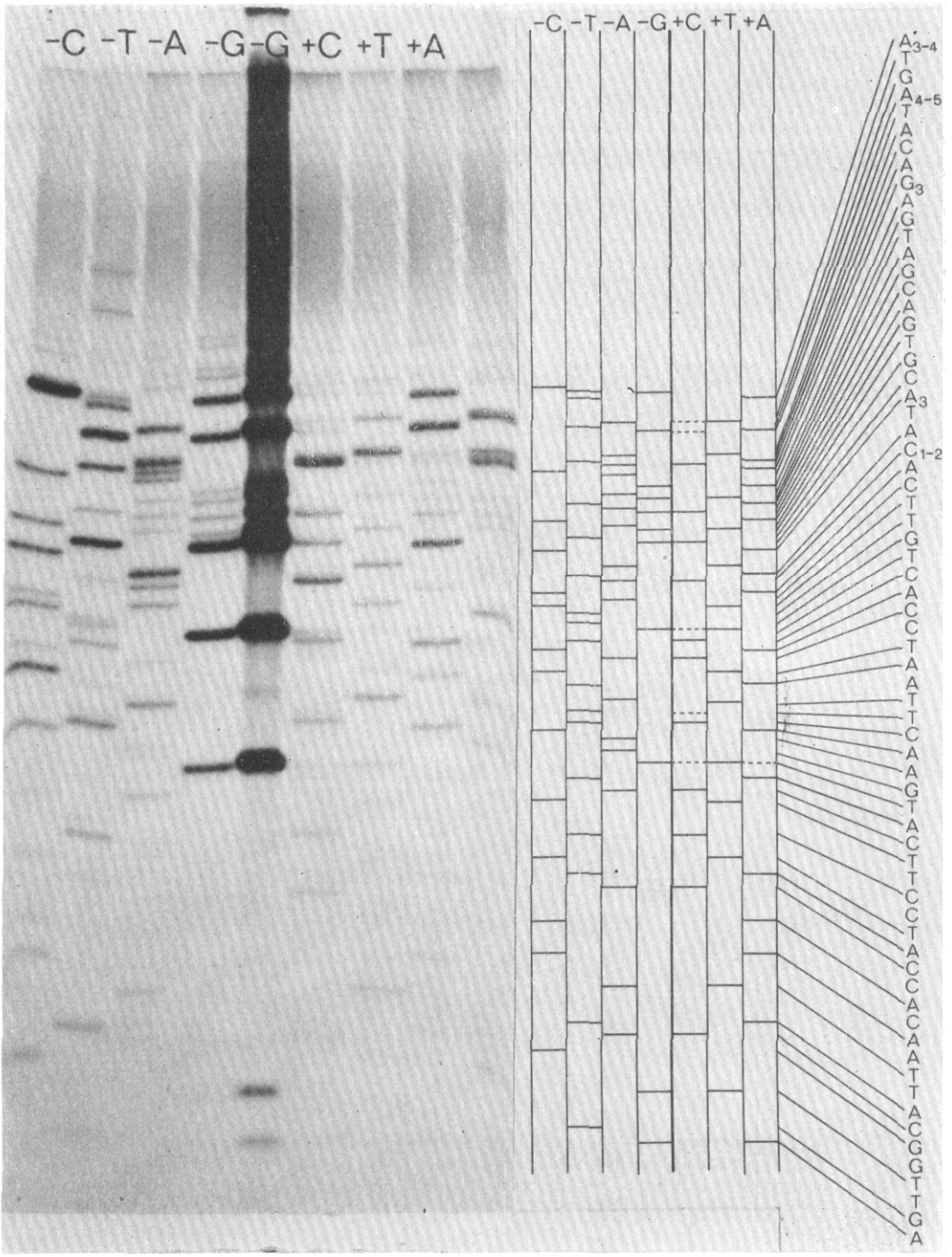

PLATE II. Radioautograph of experiment 2 (see Materials and Methods) in which *Hind* fragment 1 was used as a primer, and diagram illustrating the interpretation and sequence. The very dark centre sample labelled −G contained five times as much material as the other samples. The +G sample gave a number of artifact bands and it was not used in the interpretation. This was probably due to contamination of the +G mixture used in this experiment. A later experiment with a fresh mixture gave satisfactory results though resolution of the bands was less good.

is usually not possible to see the runs in this way and we have mainly used the distances between bands to deduce the size of a given run. In Figure 3 the change in distance travelled due to the addition of a single mononucleotide residue (the "jump") is plotted against the size of the oligonucleotides fractionated in Plates I and II. In general these values are sufficiently consistent; however they are less accurate for larger oligonucleotides and there are certain anomalies. In particular it will be seen from Plate I that position a appears to give four bands corresponding to —T,—C, +C and +G. Since the previous position contains a —G and there is a +T in a subsequent one, the most likely explanation is that the sequence is G-C-T and the products ending in G and C are moving at the same rate, although the expected

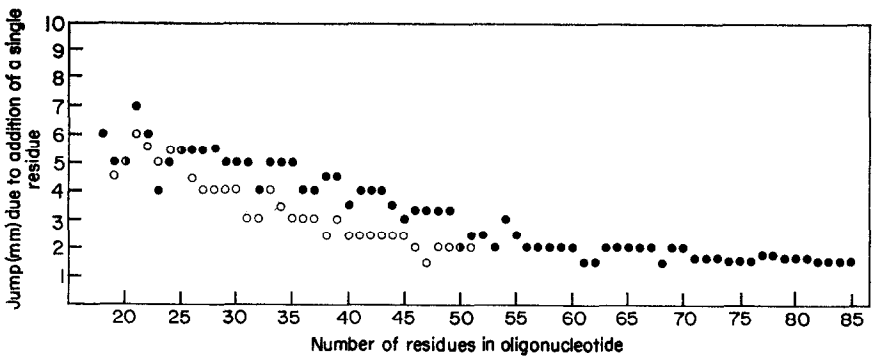

FIG. 3. The effect of the addition of a single nucleotide residue to oligonucleotides of different length on the distance travelled on ionophoresis in acrylamide gel. (○) Results from experiment 1 (Plate I); (●) results from experiment 2 (Plate II).

jump in this region should be 1·5 to 2·0 mm. Clearly in such a situation it is not possible to say how many C residues are present; however, fortunately such anomalies appear to be rare. Most of them can be avoided by using more severely denaturing conditions and low concentration gels as in the experiment shown in Plate II: thus by running the electrophoresis at 1000 V, so that the gel was at a higher temperature, some separation of the above two oligonucleotides was obtained.

One problem with methods involving primed synthesis with DNA polymerase is that one frequently finds a considerable blank incorporation of triphosphate in the absence of added primer. This is usually worse with ϕX than with phage f1 DNA and considerably worse with larger DNAs (see Maniatis *et al.*, 1974). In general it does not affect the present method very seriously since the blank seems to represent some type of random incorporation and results in a general darkening of the radioautograph at the top half of the gel, corresponding to oligonucleotides of over 100 residues long.

Another problem with the method, which is not fully understood, is the occasional appearance of "artifact" bands on the gel. These are usually faint bands and are marked as broken lines in Plates I and II. They can usually be recognized by the fact that they are not consistent in the plus and minus systems. The most serious one in Plate I is the +G band in position 23. It is clearly anomalous since there is no corresponding —G band, and two +G bands should never occur in adjacent positions. These artifact bands are usually not reproducible from one experiment to another, but their presence does emphasize the need for caution in interpreting

the results. They are usually found clustered in a certain region (e.g. positions 21 to 23, Plate I) and may represent positions where the polymerases react particularly slowly—perhaps due to some secondary structure of the DNA.

The above method represents a rapid and simple technique for determining sequences in specific regions of a DNA chain, if suitable primers are available. Thus, if successfully carried out, it is possible to deduce a sequence of 50 nucleotides in a few days. At present, due to the limitations discussed above, it cannot be regarded as a completely reliable method and it is necessary to have confirmatory data such as amino acid sequences or transcription results. It has nevertheless been found extremely useful and it is hoped that its reliability will increase as more experience is gained.

Some preliminary experiments in connection with this approach were done with J. E. Donelson. We wish to thank H. Kössel and H. Schott for providing the decanucleotide primer, K. Murray and A. G. Isaksson for a gift of T4 DNA polymerase, T. Maniatis for *Hin*d enzyme, H. L. Weith for ϕX DNA, and J. W. Sedat for ϕX RF DNA.

REFERENCES

Edgell, M. H., Hutchison, C. A. & Sclair, M. (1972). *J. Virol.* **9**, 574–582.
Englund, P. T. (1971). *J. Biol. Chem.* **246**, 3269–3276.
Englund, P. T. (1972). *J. Mol. Biol.* **66**, 209–224.
Galibert, F., Sedat, J. & Ziff, E. (1974). *J. Mol. Biol.* **87**, 377–407.
Goulian, M., Lucas, Z. J. & Kornberg, A. (1968). *J. Biol. Chem.* **243**, 627–638.
Ikemura, T. & Dahlberg, J. E. (1973). *J. Biol. Chem.* **248**, 5024–5032.
Klenow, H. & Henningsen, I. (1970). *Proc. Nat. Acad. Sci., U.S.A.* **65**, 168–175.
Maniatis, T., Ptashne, M., Barrell, B. G. & Donelson, J. E. (1974). *Nature (London)*, **250**, 394–397.
Peacock, A. C. & Dingman, C. W. (1967). *Biochemistry*, **6**, 1818–1827.
Robertson, H. D., Barrell, B. G., Weith, H. L. & Donelson, J. E. (1973). *Nature New Biol.* **241**, 38–40.
Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H. & Fischer, D. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 1209–1213.
Schott, H. (1974). *Die Makromolekulare Chemie*, **175**, 1683–1693.
Smith, H. O. & Wilcox, K. W. (1970). *J. Mol. Biol.* **51**, 379–391.
Symons, R. H. (1974). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 29, part E, pp. 102–115, Academic Press, New York and London.
Wu, R. & Kaiser, A. D. (1968). *J. Mol. Biol.* **35**, 523–527.
Wu, R. & Taylor, E. (1971). *J. Mol. Biol.* **57**, 491–511.