Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕX174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

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

Contributed by F. Sanger, October 3, 1977

ABSTRACT **A new method for determining nucleotide sequences in DNA is described. It is similar to the "plus and minus" method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441–448] but makes use of the 2′,3′-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage ϕX174 and is more rapid and more accurate than either the plus or the minus method.**

The "plus and minus" method (1) is a relatively rapid and simple technique that has made possible the determination of the sequence of the genome of bacteriophage ϕX174 (2). It depends on the use of DNA polymerase to transcribe specific regions of the DNA under controlled conditions. Although the method is considerably more rapid and simple than other available techniques, neither the "plus" nor the "minus" method is completely accurate, and in order to establish a sequence both must be used together, and sometimes confirmatory data are necessary. W. M. Barnes (*J. Mol. Biol.*, in press) has recently developed a third method, involving ribo-substitution, which has certain advantages over the plus and minus method, but this has not yet been extensively exploited.

Another rapid and simple method that depends on specific chemical degradation of the DNA has recently been described by Maxam and Gilbert (3), and this has also been used extensively for DNA sequencing. It has the advantage over the plus and minus method that it can be applied to double-stranded DNA, but it requires a strand separation or equivalent fractionation of each restriction enzyme fragment studied, which makes it somewhat more laborious.

This paper describes a further method using DNA polymerase, which makes use of inhibitors that terminate the newly synthesized chains at specific residues.

Principle of the Method. Atkinson *et al.* (4) showed that the inhibitory activity of 2′,3′-dideoxythymidine triphosphate (ddTTP) on DNA polymerase I depends on its being incorporated into the growing oligonucleotide chain in the place of thymidylic acid (dT). Because the ddT contains no 3′-hydroxyl group, the chain cannot be extended further, so that termination occurs specifically at positions where dT should be incorporated. If a primer and template are incubated with DNA polymerase in the presence of a mixture of ddTTP and dTTP, as well as the other three deoxyribonucleoside triphosphates (one of which is labeled with ^{32}P), a mixture of fragments all having the same 5′ and with ddT residues at the 3′ ends is obtained. When this mixture is fractionated by electrophoresis on denaturing acrylamide gels the pattern of bands shows the distribution of dTs in the newly synthesized DNA. By using analogous terminators for the other nucleotides in separate incubations and running the samples in parallel on the gel, a pattern of bands is obtained from which the sequence can be read off as in the other rapid techniques mentioned above.

Two types of terminating triphosphates have been used—the dideoxy derivatives and the arabinonucleosides. Arabinose is a stereoisomer of ribose in which the 3′-hydroxyl group is oriented in *trans* position with respect to the 2′-hydroxyl group. The arabinosyl (ara) nucleotides act as chain terminating inhibitors of *Escherichia coli* DNA polymerase I in a manner comparable to ddT (4), although synthesized chains ending in 3′ araC can be further extended by some mammalian DNA polymerases (5). In order to obtain a suitable pattern of bands from which an extensive sequence can be read it is necessary to have a ratio of terminating triphosphate to normal triphosphate such that only partial incorporation of the terminator occurs. For the dideoxy derivatives this ratio is about 100, and for the arabinosyl derivatives about 5000.

METHODS

Preparation of the Triphosphate Analogues. The preparation of ddTTP has been described (6, 7), and the material is now commercially available. ddA has been prepared by McCarthy *et al.* (8). We essentially followed their procedure and used the methods of Tener (9) and of Hoard and Ott (10) to convert it to the triphosphate, which was then purified on DEAE-Sephadex, using a 0.1–1.0 M gradient of triethylamine carbonate at pH 8.4. The preparation of ddGTP and ddCTP has not been described previously; however we applied the same method as that used for ddATP and obtained solutions having the requisite terminating activities. The yields were very low and this can hardly be regarded as adequate chemical characterization. However, there can be little doubt that the activity was due to the dideoxy derivatives.

The starting material for the ddGTP was *N*-isobutyryl-5′-*O*-monomethoxytrityldeoxyguanosine prepared by F. E. Baralle (11). After tosylation of the 3′-OH group (12) the compound was converted to the 2′,3′-didehydro derivative with sodium methoxide (8). The isobutyryl group was partly removed during this treatment and removal was completed by incubation in NH_3 (specific gravity 0.88) overnight at 45°. The didehydro derivative was reduced to the dideoxy derivative (8) and converted to the triphosphate as for the ddATP. The monophosphate was purified by fractionation on a DEAE-Sephadex column using a triethylamine carbonate gradient (0.025–0.3 M) but the triphosphate was not purified.

ddCTP was prepared from *N*-anisoyl-5′-*O*-monomethoxytrityldeoxycytidine (Collaborative Research Inc., Waltham, MA) by the above method but the final purification on DEAE-Sephadex was omitted because the yield was very low and the solution contained the required activity. The solution was used directly in the experiments described in this paper.

An attempt was made to prepare the triphosphate of the intermediate didehydrodideoxycytidine because Atkinson *et*

Abbreviations: The symbols C, T, A, and G are used for the deoxyribonucleotides in DNA sequences; the prefix dd is used for the 2′,3′-dideoxy derivatives (e.g., ddATP is 2′,3′-dideoxyadenosine 5′-triphosphate); the prefix ara is used for the arabinose analogues.

QIAGEN Exhibit 1016
Page 1

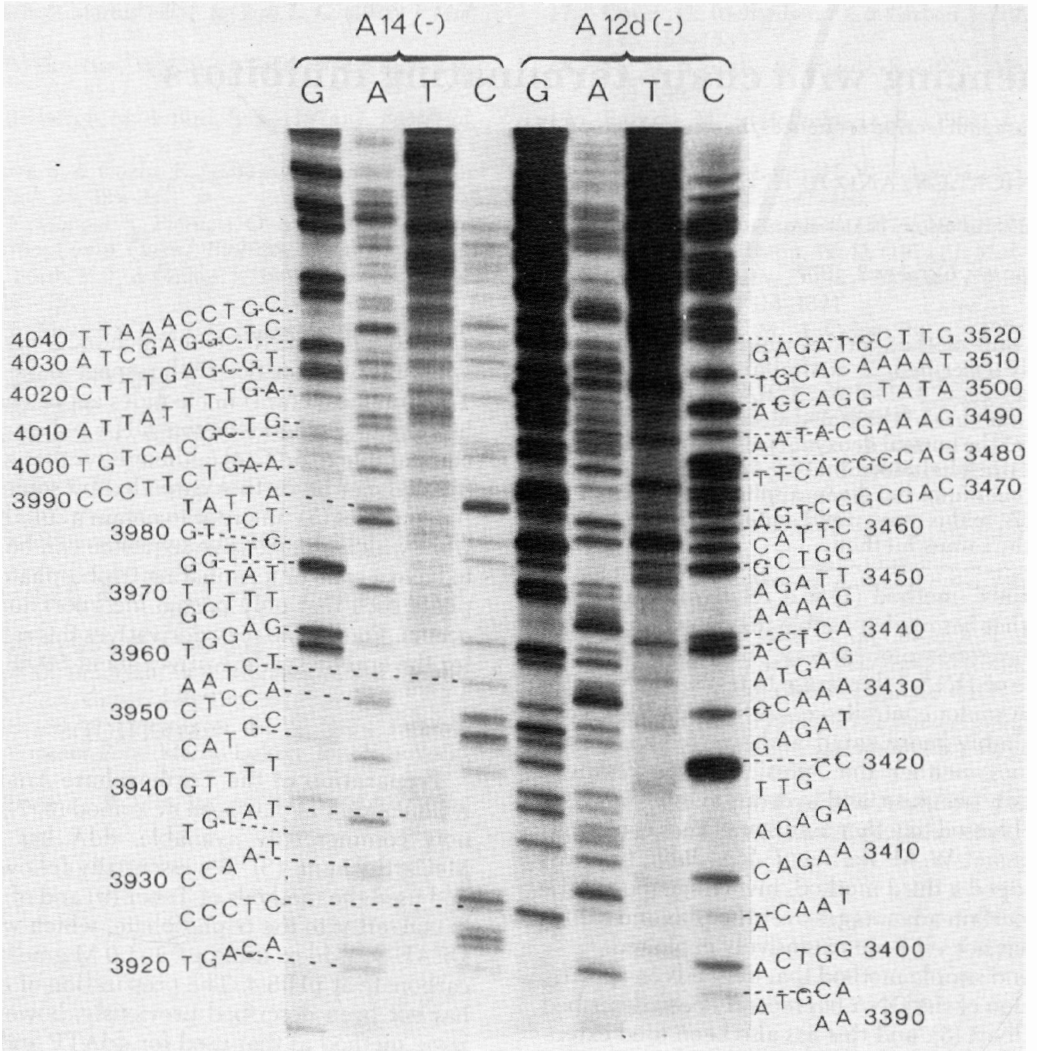

FIG. 1. Autoradiograph of the acrylamide gel from the sequence determination using restriction fragments A12d and A14 as primers on the complementary strand of ϕX174 DNA. The inhibitors used were (left to right) ddGTP, ddATP, ddTTP, and araCTP. Electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10 cm of the gel is not shown. The DNA sequence is written from left to right and upwards beside the corresponding bands on the radioautograph. The numbering is as given in ref. 2.

al. (4) have shown that the didehydrodideoxy-TTP is also active as a terminator. However, we were unsuccessful in this. These compounds seem much less stable than the dideoxy derivatives.

araATP and araCTP were obtained from P-L Biochemicals Inc., Milwaukee, WI.

Sequencing Procedure. Restriction enzyme fragments were obtained from ϕX174 replicative form and separated by electrophoresis on acrylamide gels. The material obtained from 5 μg of ϕX174 replicative form in 5 μl of H_2O was mixed with 1 μl of viral or complementary strand ϕX174 DNA (0.6 μg) and 1 μl of H $\times$ 10 buffer (13) and sealed in a capillary tube, heated to 100° for 3 min, and then incubated at 67° for 30 min. The solution was diluted to 20 μl with H buffer and 2 μl samples were taken for each incubation and mixed with 2 μl of the appropriate "mix" and 1 μl of DNA polymerase (according to Klenow, Boehringer, Mannheim) (0.2 units). Each mix contained 1.5 $\times$ H buffer, 1 μCi of [α-^{32}P]dATP (specific activity approximately 100 mCi/μmol) and the following other triphosphates.

ddT: 0.1 mM dGTP, 0.1 mM dCTP, 0.005 mM dTTP, 0.5 mM ddTTP

ddA: 0.1 mM dGTP, 0.1 mM dCTP, 0.1 mM dTTP, 0.5 mM ddATP

ddG: 0.1 mM dCTP, 0.1 mM dTTP, 0.005 mM dGTP, 0.5 mM ddGTP

ddC: 0.1 mM dGTP, 0.1 mM dTTP, 0.005 mM dCTP, approximately 0.25 mM ddCTP

(The concentration of the ddCTP was uncertain because there was insufficient yield to determine it, but the required dilution of the solution was determined experimentally.)

araC: 0.1 mM dGTP, 0.1 mM dTTP, 0.005 mM dCTP, 12.5 mM araCTP

Incubation was at room temperature for 15 min. Then 1 μl of 0.5 mM dATP was added and incubation was continued for a further 15 min. If this step (chase) was omitted some termination at A residues occurred in all samples due to the low concentration of the [α-^{32}P]dATP. With small primers, where it was unnecessary to carry out a subsequent splitting (as in the experiment shown in Fig. 1), the various reaction mixtures were denatured directly and applied to the acrylamide gel for electrophoresis (1). If further splitting was necessary (see Fig. 2), 1 μl of the appropriate restriction enzyme was added shortly after the dATP "chase," and incubation was at 37°.

The single-site ribo-substitution procedure (N. L. Brown, unpublished) was carried out as follows. The annealing of template and primer was carried out as above but in "Mn buffer" (66 mM TrisCl, pH 7.4/1.5 mM 2-mercaptoethanol/

QIAGEN Exhibit 1016
Page 2

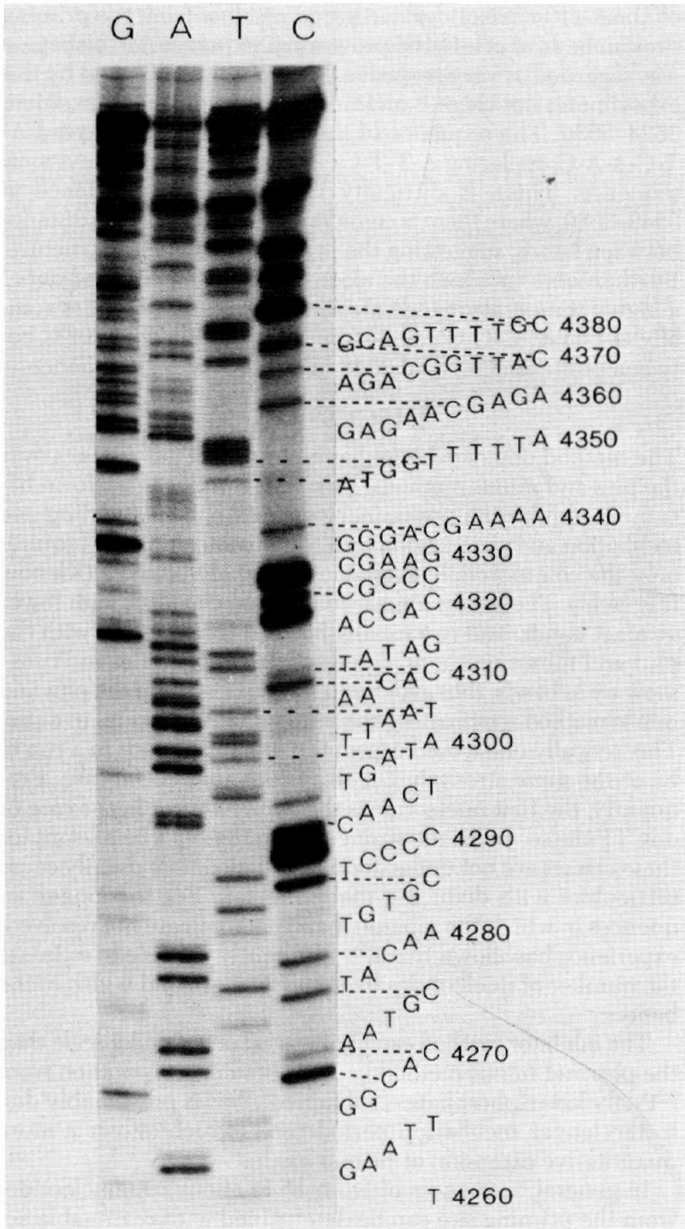

FIG. 2. Autoradiograph from an experiment using fragment R4 as primer on the complementary strand of ϕX174 DNA. Conditions were as in Fig. 1 with the following exceptions: ddCTP was used as inhibitor instead of araCTP. After incubation of the solutions at room temperature for 15 min, 1 μl of 0.5 mM dATP and 1 μl of restriction enzyme *Hae* III (4 units/μl) were added and the solutions were incubated at 37° for 10 min. The *Hae* III cuts close to the *Hin*dII site and it was used because it was more readily available. The electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10.5 cm of the gel is not shown.

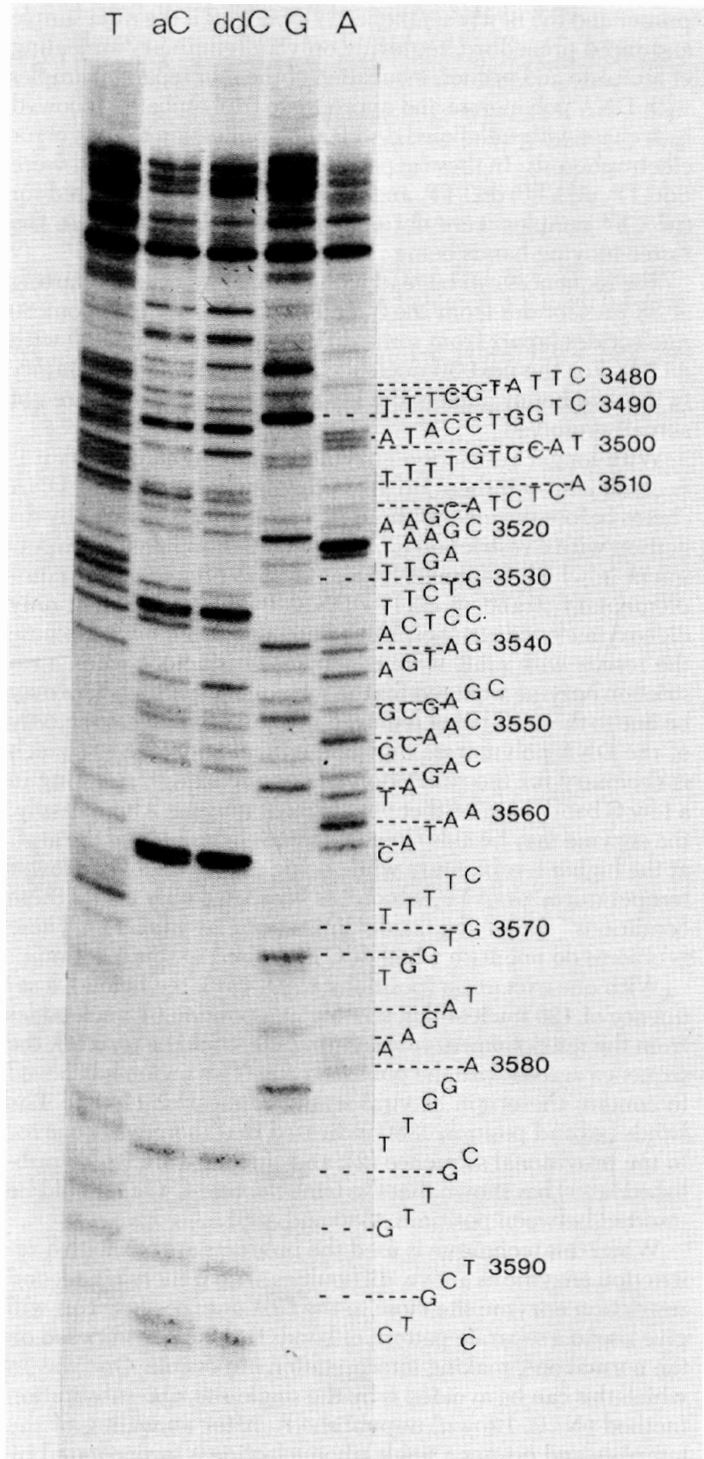

FIG. 3. Autoradiograph of an experiment with fragment A8 as primer on the viral strand of ϕX174 DNA using the single-site ribosubstitution method. Electrophoresis was on a 12% gel at 40 mA for 6 hr. The top 5.5 cm is not shown. Inhibitors used were (left to right) ddTTP, araCTP, ddCTP, ddGTP, and ddATP.

0.67 mM $MnCl_2$) rather than in H buffer. To 7 μl of annealed fragment was added 1 μl of 10 mM rCTP, 2 μl of H_2O, and 1 μl of 10 × Mn buffer. Five microcuries of dried [α-^{32}P]dTTP (specific activity approximately 1 mCi/μmol) was dissolved in this and 1 unit DNA polymerase (Klenow) was added. Incubation was for 30 min in ice. One microliter of 0.2 M EDTA was added before loading on a 1-ml Sephadex G-100 column. Column buffer was 5 mM Tris, pH 7.5/0.1 mM EDTA. The labeled fragment was followed by monitor, collected in a minimum volume (approximately 200 μl), dried down, and redissolved in 30 μl of 1 × H buffer. Samples (2 μl) of this were taken for treatment as above. Following the chase step, 1 μl of 0.1 M EDTA and 1 μl of pancreatic ribonuclease A at 10 mg/ml were added and incubated for 60 min at 37°.

RESULTS

Figs. 1–3 show examples of the use of the method for determining sequences in the DNA of ϕX174. In the experiment shown in Fig. 1 two small restriction enzyme fragments (A12d and A14, ref. 2) were used as primers on the complementary strand and there was no final digestion step to cut between the

primer and the newly synthesized DNA. This is the most simple and rapid procedure, requiring only a preliminary annealing of template and primer, incubation of the four separate samples with DNA polymerase and appropriate triphosphates, followed by a chase with unlabeled dATP and application to the gel for electrophoresis. In these experiments the inhibitors used were ddGTP, ddATP, ddTTP, and araCTP. The conditions used for the "T" samples were not entirely optimal, resulting in the faster-moving bands being relatively weak.

The sequences can be read with reasonable accuracy starting at 88 nucleotides from the 5′ end of the primer for about 80 nucleotides (apart from some difficulty at position 3459 with A12d). For the next 50 nucleotides there is some uncertainty in the number of nucleotides in "runs" because bands are not actually resolved.

With longer restriction enzyme fragments as primers it is necessary to split them off from the newly synthesized DNA chains before the electrophoresis. This is normally done by digestion with a restriction enzyme. Fig. 2 shows such an experiment in which fragment R4 was used as primer on the complementary strand of ϕX174 DNA. In this experiment only dideoxynucleoside triphosphates were used as inhibitors because the results with araC were much less satisfactory when a restriction enzyme was used for the subsequent splitting. This may be due to the araC being removed by the 3′-exonuclease activity of the DNA polymerase during the incubation at 37° (which is necessary for the restriction enzyme splitting), resulting in a few C bands being either very faint or missing. Alternatively, the enzyme may be able to extend some chains beyond the araC at the higher temperature while being unable to do so at lower temperatures. araATP, which has been used only under these conditions, shows the same limitations as araCTP. These problems do not arise when ddCTP is used in this reaction.

With one exception (positions 4330–4343, see below), a sequence of 120 nucleotides, starting at a position 61 nucleotides from the restriction enzyme splitting site, could be read off; the sequence agreed with the published one. This region is believed to contain the origin of viral strand replication (2, 14). The bands beyond position 4380 indicated that there was an error in the provisional sequence (2), and further work (to be published later) has shown that the trinucleotide C-G-C should be inserted between positions 4380 and 4381.

When this technique is used the products are cut with a restriction enzyme as above, difficulties arise if there is a second restriction enzyme site close to the first one, because this will give rise to a separate pattern of bands that is superimposed on the normal one, making interpretation impossible. One way in which this can be avoided is by the single-site ribo-substitution method (N. L. Brown, unpublished). After annealing of the template and primer a single ribonucleotide is incorporated by incubation with DNA polymerase in the presence of manganese and the appropriate ribonucleoside triphosphate. Extension of the primer is then carried out with the separate inhibitors as above and the primer is split off at the ribonucleotide by ribonuclease or alkali. The method is particularly suitable for use with fragments obtained with the restriction enzyme *Alu*, which splits at the tetranucleotide sequence A-G-C-T. This enzyme is in fact inhibited by single-stranded DNA and cannot be used for the subsequent splitting of the primer from the newly synthesized DNA chain. The initial incorporation is carried out in the presence of rCTP and [α-^{32}P]dTTP. The incorporation of the ^{32}P facilitates subsequent purification on the Sephadex column.

Fig. 3 shows an example of the use of this method with fragment A8 on the viral strand of ϕX174 DNA. A sequence of about 110 nucleotides starting 33 residues from the priming site can be read off. In the provisional sequence (2) this region was regarded as very tentative. Most of it is confirmed by this experiment, but there is a clear revision required at positions 3524–3530. The sequence of the viral strand should read A-T-C-A-A-C, replacing A-T-T-C- - -A-C given in the provisional sequence. There is difficulty in reading the sequence at 3543–3550, where there is considerable variation in the distance between bands, suggesting the presence of a looped structure. Further work in which the electrophoresis was carried out at a higher temperature indicates that the sequence here is actually G-C-T-C-G-C-G (viral strand); i.e., an insertion of C between positions 3547 and 3548 in the provisional sequence.

DISCUSSION

The method described here has a number of advantages over the plus and minus methods. First, it is simpler to perform because it requires no preliminary extension, thus avoiding one incubation and purification on a Sephadex column. It requires only the commercially available DNA polymerase I (Klenow fragment). The results appear to be more clear-cut with fewer artefact bands, and can usually be read further than with the plus and minus methods. Intermediate nucleotides in "runs" show up as bands, thus avoiding a source of error in the plus and minus method—estimating the number of nucleotides in a run. Theoretically one would expect the different bands in a run to be of the same strength, but this is not always the case. Frequently, the first nucleotide is the strongest, but in the case of ddCTP the second is the strongest (see Fig. 2). The reasons for these effects are not understood, but they do not usually cause difficulties with deducing the sequences. For the longer sequences in which the separate bands in a run are not resolved, experience has shown that it is frequently possible to estimate the number of nucleotides from the strength and width of the band.

The inhibitor method can also be used on a smaller scale than the plus and minus method because better incorporation from ^{32}P-labeled triphosphates is obtained. This is presumably due to the longer incubation period used, which allows a more quantitative extension of primer chains.

In general, sequences of from 15 to about 200 nucleotides from the priming site can be determined with reasonable accuracy using a single primer. Frequently it is possible to read the gels further and, on occasions, a sequence of about 300 nucleotides from the priming site has been determined. Occasional artefacts are observed, but these can usually be readily identified. It seems likely that these are usually due to contaminants in the fragments. The most serious difficulties are due to "pile-ups" of bands, which are usually caused by the DNA forming base-paired loops under the conditions of the acrylamide gel electrophoresis. These pile-ups are seen as a number of bands in the same position or unusually close to one another on the electrophoresis. They generally occur at different positions when the priming is carried in opposite directions along the DNA over the same sequence. An example of this effect is seen in Fig. 2 at position 4330, where there is a single strong band in the G channel that in fact represents four G residues. They are presumably forming a stable loop by pairing with the four Cs at positions 4323–4326. Another example is in Fig. 3 at positions 3545–3550. This effect is likely to be found in all the rapid techniques that use gel electrophoresis.

It is felt that for an accurate determination of sequence one should not rely completely on single results obtained by this method alone but that confirmation should be obtained by some

QIAGEN Exhibit 1016
Page 4

other technique or by priming on the opposite strand. This consideration probably applies to all other available methods also. The main disadvantage of the present method is the difficulty in obtaining all the inhibitors—particularly ddGTP, which is not commercially available.

We wish to thank Dr. K. Geider for a gift of ddTTP, Dr. F. E. Baralle for a gift of *N*-isobutyryl-5′-*O*-monomethoxytrityldeoxyguanosine, and Dr. M. J. Gait for useful advice on the synthetic work.

1. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448.
2. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M. & Smith, M. (1977) *Nature* **265**, 687–695.
3. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
4. Atkinson, M. R., Deutscher, M. P., Kornberg, A., Russell, A. F. & Moffatt, J. G. (1969) *Biochemistry* **8**, 4897–4904.
5. Hunter, T. & Francke, B. (1975) *J. Virol.* **15**, 759–775.
6. Russell, A. F. & Moffatt, J. G. (1969) *Biochemistry* **8**, 4889–4896.
7. Geider, K. (1974) *Eur. J. Biochem.* **27**, 555–563.
8. McCarthy, J. R., Robins, M. J., Townsend, L. B. & Robins, R. K. (1966) *J. Am. Chem. Soc.* **88**, 1549–1553.
9. Tener, G. M. (1961) *J. Am. Chem. Soc.* **83**, 159–168.
10. Hoard, D. E. & Ott, D. G. (1965) *J. Am. Chem. Soc.* **87**, 1785–1788.
11. Büchi, H. & Khorana, H. G. (1972) *J. Mol. Biol.* **72**, 251–288.
12. Robins, M. J., McCarthy, J. R. & Robins, R. K. (1966) *Biochemistry* **5**, 224–231.
13. Air, G. M., Sanger, F. & Coulson, A. R. (1976) *J. Mol. Biol.* **108**, 519–533.
14. Slocombe, P. M. (1976) Ph.D. Dissertation, University of Cambridge.