Volume 13 Number 7 1985 Nucleic Acids Research

A new method for sequence analysis of oligodeoxyribonucleotides

D.M.Black and P.T.Gilham

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received 21 November 1984; Revised and Accepted 20 February 1985

ABSTRACT

The structure of an oligodeoxyribonucleotide may be determined by a simple two-dimensional separation on a polyethyleneimine-cellulose thin layer sheet. Chromatography in the first dimension fractionates by chain length a nested set of fragments that are generated by subjecting the oligomer to partial spleen phosphodiesterase degradation and then labelling their non-common ends with ^{32}P using polynucleotide kinase. A subsequent in situ treatment with nuclease Bal 31 produces labelled mononucleotides, and these are identified by chromatography in the second dimension. Since the method does not identify the 3' terminal nucleotide, a convenient procedure involving 3' end labelling followed by enzymatic digestion to monomers has been developed for this purpose. This approach to sequence analysis also has the advantage of permitting assignment of the identity and location of any modified or unusual bases within the oligonucleotide.

INTRODUCTION

Oligodeoxyribonucleotides are used extensively in molecular biology as primers for DNA synthesis, as linkers for joining DNA in cloning experiments, and as probes in studies of nucleic acid function. They are also widely employed as models in physical studies of nucleic acid structure with nuclear magnetic resonance and crystallographic techniques. These applications require a simple and reliable procedure for determining both the sequence and purity of small DNA fragments.

The two approaches currently used for sequence analysis of oligodeoxyribonucleotides are mobility shift ("wandering spot") analysis (1,2) and the chemical cleavage technique (3,4). The first approach uses a two-dimensional fractionation of a nested set of fragments that is derived from the oligonucleotide by partial exonuclease action. The set members, which contain labels at their common ends, are separated by electrophoresis on cellulose acetate at low pH in the first dimension followed by transfer to a DEAE-cellulose thin layer sheet for homochromatography in the second dimension. Since the mobility of any member of the nested set depends on its size and

© IRL Press Limited, Oxford, England.

base composition, the sequence of the oligomer is derived by comparing each member's position on the sheet with that of the member containing one nucleotide less. In the chemical cleavage technique, the labelled oligomer is cut at specific sites in several different base-excising reactions, and the products are separated by length using polyacrylamide gel electrophoresis. The sequence may then be inferred from the gel pattern by observing which specific reaction is responsible for cleavage at each position of the chain.

In the mobility shift technique, the sequence cannot be determined reliably by visual inspection alone; corrections relative to standards must be made for the position of each fragment (2) and, even when the analysis is performed carefully, some deviations and ambiguities may occur (5). In the chemical cleavage method, extraneous bands may appear in the final autoradiogram (4), making interpretation difficult. A deficiency common to each of the methods, moreover, is that the procedures are not readily adaptable to the analysis of oligonucleotides containing modified bases. Considerable difficulty would be experienced in identifying any number of modified nucleotides from differences in their mobility shifts and, in order to use the chemical cleavage technique for assigning their sequence positions, new specific base-excising reactions would have to be found.

The new method described here and also in a preliminary report (6) avoids the above problems and allows direct reading of the sequence from the final autoradiogram without ambiguities. In addition, since the method involves visualizing the mononucleotide at each position of the chain rather than inferring its identity from mobility shifts or cleavage patterns, modified residues may also be identified, an advantage that arises from the fact that the members of the nested set are labelled at their non-common ends, instead of the common ends as is the case in the currently used sequencing techniques. A method of sequence analysis employing fragments labelled at the non-common 3' or 5' ends has been previously reported (7). It is a rather lengthy procedure, however, requiring separation of the oligomers by two-dimensional homochromatography, followed by elution of each fragment from the thin layer plate and enzymatic digestion in solution to monomers, in order to identify the nucleotide at each position of the chain.

MATERIALS AND METHODS

Materials.

Lyophilized spleen phosphodiesterase was obtained from P-L Biochemicals, Inc.; lyophilized calf intestinal phosphatase and polynucleotide kinase were

products of Boehringer Mannheim Biochemicals; nuclease Bal 31 and terminal deoxynucleotidyl transferase were purchased from Bethesda Research Laboratories, Inc. Oligomers were chemically synthesized by previously published techniques (8-10). Nucleotide markers and yeast tRNA (type X, catalog no. R-9001) were purchased from Sigma Chemical Co. The [γ-^{32}P]ATP (3000 Ci/mmol) was purchased from Amersham Corp.; [α-^{32}P]ATP (300 Ci/mmol) was the product of ICN Radiochemicals. N^6-methyldeoxyadenosine 5'-phosphate was prepared by a previously reported method (11).

PEI-cellulose MN 300, UV-254 pre-coated plastic sheets, 20 X 20 cm, (Brinkmann Instruments Co.) are washed prior to use. Each sheet is soaked in 1 liter of 10% NaCl for 10 min., then in 1 liter of distilled water for 10 min., and again in water, with occasional agitation and with no intermediate drying. The same NaCl solution may be used for washing 4 sheets sequentially, but fresh distilled water is used for each sheet. Finally, the sheets are dried at room temperature for several hours and stored at -20° wrapped in foil.

TLC Solvent A is 2.7 M Tris-HCl buffer-deionized formamide (1:1, v/v). The Tris-HCl buffer is prepared by mixing 206 g of Tris base and 158 g of Tris-HCl with water up to a final volume of 1 liter. The formamide is deionized by adding 2 g of mixed bed resin [AG 501-X8(D) from Bio-Rad Laboratories], to 100 ml of formamide, and allowing it to stand for 30 min. TLC Solvent B is 1.8 M Tris-HCl (pH 8.0)-deionized formamide (1:1, v/v). The Tris-HCl for this solvent is prepared by adjusting the pH of a 1.8 M solution of Tris base with HCl at 25°. When making Solvents A and B, the Tris-HCl and formamide are mixed together just before use. TLC Solvent C is 0.2 M sodium acetate, adjusted to pH 4.35 with acetic acid. TLC Solvent D is prepared by diluting 10 ml of acetic acid with water, adjusting the pH to 3.65 with pyridine, and adding water to a final volume of 100 ml (12). TLC Solvent E is i-PrOH-AcOH-1% aqueous $(NH_4)_2SO_4$ (45:35:20, v/v), (11). Solvents D and E are freshly prepared each week.

Step 1. Partial spleen phosphodiesterase digestion.

The reaction mixtures contain 10/x nmole of oligomer, where x is the chain length, in 50 µl of 10 mM sodium 2-(N-morpholino)ethanesulfonate (MES) (pH 6.5, 25°). For molecules up to undecamers in size, the mixtures are treated with 5 mU of spleen phosphodiesterase; for longer molecules, 10 mU of the phosphodiesterase is added. Incubation is at 37°, and 5 µl aliquots withdrawn at 1, 3, 6, 10, 15, 20, 30, 40, 50, and 60 min. are added to a single Eppendorf tube kept on solid carbon dioxide. The combined aliquots are

then heated at 100° for 5 min. to inactivate the enzyme.

Step 2. Removal of 3' phosphate groups.

To 5 µl of the spleen phosphodiesterase digest is added 5 µl of 250 mM Tris-HCl (pH 7.6), 10 µl of water, and 1 µl of calf intestinal phosphatase solution [0.01 U/µl in 100 mM Tris-HCl (pH 8.0)-glycerol (1:1, v/v)]. Incubation is at 37° for 15 min., and the enzyme is inactivated by heating at 100° for 5 min.

Step 3. Labelling with polynucleotide kinase.

To the phosphatase-treated reaction mixture are added 5 µl of a solution that contains 50 mM $MgCl_2$ and 25 mM dithiothreitol, and 1 µl of polynucleotide kinase (5-6 units). The mixture is combined with 33 pmol of dried [γ-^{32}P]ATP and incubated at 37° for 15 min.

Step 4. Separation by chain length by TLC in the first dimension.

One µl of the labelled oligomer mixture is applied to a point 1.5 cm from the bottom and 1.5 cm from one side of a thin layer. Chromatography tanks are filled to a depth of 0.5 cm with developing solutions. For molecules up to undecamers in length, the sheet is developed with water to the origin, and then, without intermediate drying, with Solvent A to the top (about 6 h.). For longer molecules such as the icosamer, two thin layer sheets are used. The first sheet, used to separate the shorter fragments, is developed as for short molecules described above. The second sheet, used to separate the longer fragments, is prepared by stapling a wick of Whatman 3MM paper to the top. The wick, which is the width of the sheet and extends from the covered tank, permits chromatography to proceed for an extended period of time. The sheet is developed with water to the origin, and then, without intermediate drying, with Solvent B for 38 h. After development, each sheet is dried in a stream of cool air for 5 min., soaked in 500 ml of methanol for 5 min. to remove Tris and formamide, soaked in another 500 ml of methanol for 5 min., and dried.

Step 5. In situ enzymatic digestion to mononucleotides.

A solution containing 3 U of nuclease Bal 31 and 1 A_{260} unit of tRNA in 100 µl of 20 mM Tris-HCl (pH 8.0, 25°)-12 mM $CaCl_2$-12 mM $MgCl_2$-1 mM EDTA is streaked from a 20 µl capillary at about 4 µl/cm over the line of oligomers on the thin layer. The sheet is immediately covered with Saran wrap, clamped between two Lucite sheets, and incubated at 37° for 1 h.

Step 6. Identification of 5' terminals by TLC in the second dimension.

Identification of the labelled nucleotides is effected with one of the solvent systems listed (together with relevant mobilities) in Table 1. In

Table 1. Mobilities of Nucleotides Relative to pdG.

TLC Solvent	5'-Phosphates of dG	dA	dT	dC	m^6dA	m^5dC
C	1.0	2.9	4.7	5.9	2.9	6.7
D	1.0	2.6	2.4	3.2	2.9	3.3
E	1.0	11.5	6.5	19.0	16.0	23.5

order to prevent streaking of the spots, the sheet is soaked for 5 min. in a mixture of 5 ml of the selected solvent and 500 ml of water prior to chromatography. The sheet is dried, developed to the top in the second dimension with the selected solvent, dried, and subjected to autoradiography.

Identification of 3' terminals.

The oligomer (200 pmol) is incubated with 15 pmol of [α-^{32}P]ATP and 17 U of terminal transferase in 10 µl of 10 mM potassium 4-(2-hydroxyethyl)-1-piperazineethanesulfonate (HEPES) (pH 7.2, 25°)-5 mM $MgCl_2$-0.2 mM dithiothreitol, at 37° for 2 h. The oligomers are then degraded to 3' mononucleotides by adding 10 µl of 100 mM MES, pH 6.5 and 2 µl of spleen phosphodiesterase solution (10 mU/µl in H_2O), and incubating at 37° for 2 hours. One µl of the digest is spotted on a PEI-cellulose thin layer sheet next to a spot containing a set of 3' nucleotide standards, dried, developed to the top in Solvent D, and subjected to autoradiography.

RESULTS AND DISCUSSION

The sequencing strategy depends on the conversion of the molecule to a nested set of oligomers that have 3' common ends and ^{32}P labels at their 5' non-common ends. The set is created by treating the oligomer with spleen phosphodiesterase, freezing portions of the mixture at timed intervals, and then heating the combined portions to inactivate the enzyme. At this point, the solution contains nucleoside 3' phosphates as well as the desired set and, in order to prevent labelling of these monomers in the next step, their phosphate groups are removed by treatment with calf intestinal phosphatase. Following inactivation of phosphatase, the set of fragments is labelled with polynucleotide kinase and [γ-^{32}P]ATP. For the analysis of small oligomers (with chain lengths of up to about 12), the members of the set are separated by length on a single PEI-cellulose thin layer sheet. For larger oligomers, however, it is preferable to use two sheets: one to separate the smaller members of the set and the other, developed for a longer time, to fractionate the larger species. Nuclease Bal 31 is used to digest the oligomers in situ to 5' mononucleotides, which are then separated by chromatography in the

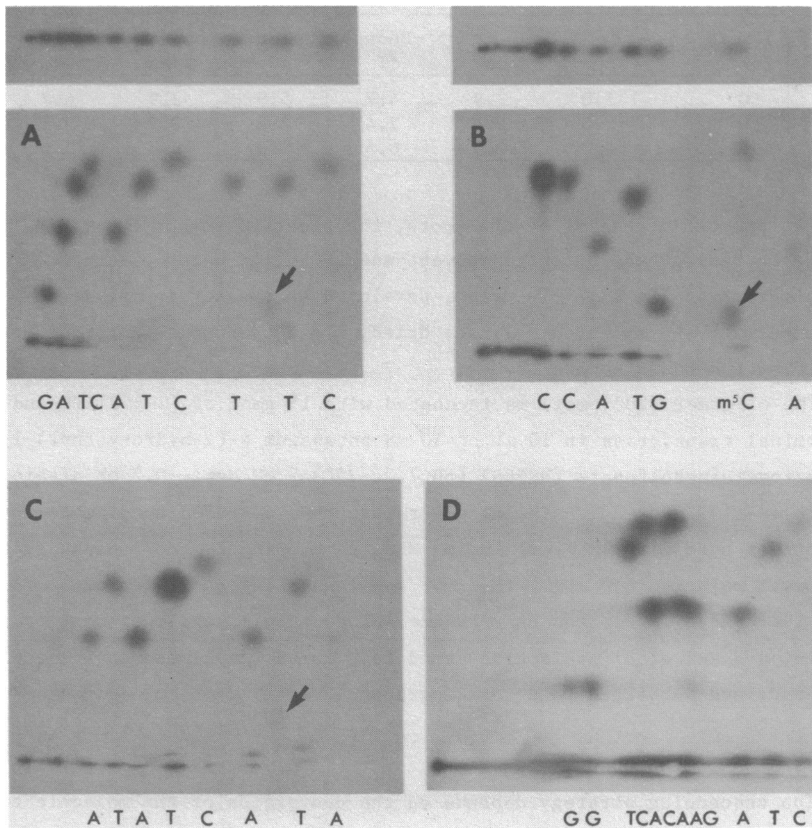

Figure 1. TLC patterns obtained from the analysis of oligomers: A, GATCATCTTCTp; B, CCATGm^{5}CAT; C, ATATCATAT; D, GGTCACAAGATCTGAAGCAG. In each case, the first dimension separation proceeded from left to right with TLC Solvent A (for A-C) or TLC Solvent B (for D) and the second dimension from bottom to top using TLC Solvent C. At the top of the figure are the intermediate patterns obtained for molecules A and B by autoradiography after the first dimension separation and prior to the second. In each of the patterns A-C the spot indicated by an arrow arises from contaminating inorganic phosphate. The pattern of the icosamer D represents the analysis of the longer members of the nested set obtained by extended chromatography in the first dimension. The sequence corresponding to the right half of this molecule was obtained by using the standard first dimension separation that is illustrated in the analyses of molecules A-C.

second dimension. Subsequent autoradiography gives a pattern of spots which allows the sequence to be read. The technique has been tested on a number of molecules including one containing a 5-methyldeoxycytidine residue: GATCATCTTCTp, CCATGm^{5}CAT, ATATCATAT, CGAGTTTGACGp, CGCTAAACTCGp, AGTCAT,

ATCTTATTTTCTGTTTGTA, TGACGTGA, and GGTCACAAGATCTGAAGCAG; the TLC patterns corresponding to the analyses of three of these are shown in Figure 1 A-C. The patterns obtained from the other test oligomers were equally unambiguous in interpretation except the one corresponding to a particular region in the sequence of the icosamer, the last member in the above list. The pattern obtained from the analysis of the left half of this molecule is reproduced in Figure 1 D in order to demonstrate a potential limitation to the size and structure of large oligomers that can be completely sequenced by the method. The arrangement of spots in this pattern indicate a less than satisfactory first dimension resolution of the nested set members that have chain lengths of 14, 13, and 12, with 5' terminals of A, A, and G, respectively. The difficulty seems to be related to the chain lengths of the set members as well as the nature of their 5' terminals because the shorter oligomers corresponding to the sequences to the right of this area are well resolved by chromatography with TLC Solvent A, even though they contain a similar arrangement (A,A,G) of 5' terminals deriving from the sequence near the 3' end of the icosamer. We have not investigated this problem further so it is possible that there are other special sequence arrangements that may be difficult to resolve when located near the left hand end of a long oligonucleotide.

The spleen phosphodiesterase used to form the nested sets seems to have little base specificity, and the digestion conditions described above can be used successfully for molecules ranging up to icosamers in size. However, as may be seen in the TLC patterns, spots derived from dinucleotides in the nested sets are not as intense as the others; whether this is due to some cleavage specificity in the phosphodiesterase or a preferred labelling of larger molecules by the kinase has not been investigated.

There are no purification steps in the sequencing method and this requires use of modified experimental protocols for the two labelling procedures in order to avoid the appearance of extraneous spots in the final patterns. In the case of the kinase reaction, the oligomer is added in molar excess over the [γ-^{32}P]ATP, so that no ATP remains at the end of the incubation. Nevertheless, a radioactive contaminant that co-migrates with inorganic phosphate always appears on the patterns; it runs near the trimer in the first dimension and just below pdG with Solvent C in the second (Fig. 1). The inorganic phosphate is present in commercial preparations of [γ-^{32}P]ATP, and its radioactive spot becomes more intense with prolonged incubation of the kinase reaction, an increase that probably results from a side reaction of the kinase in which it catalyzes the hydrolysis of the γ-phosphate group (13).

The chromatographic separation in the first dimension is a modification of a previously reported step-gradient TLC system using high concentrations of Tris-HCl and urea (14). While oligomers of a nested set are well resolved in this gradient system, we find that the stepwise elution procedure is somewhat tedious and that equivalent resolution can be achieved by using a single solvent containing a high concentration of formamide. The mobility of an oligomer in this formamide system is mainly dependent on its chain length, although there is some variation caused by differing base compositions, with dG residues causing extra retardation. It will be necessary to take this into account if mixed oligonucleotide probes are to be sequenced by this method.

Several enzymes have been tested for their ability to digest oligomers that are adsorbed to PEI-cellulose sheets. Snake venom phosphodiesterase and the nucleases S1 and P1 tend to leave some dimers and larger oligomers undigested. Dimers migrate below pdG while the larger species remain at the origin during the second chromatographic step with Solvent C. In contrast to these enzymes, nuclease Bal 31 has the capacity to completely degrade oligonucleotides along the line where the capillary containing the enzyme solution contacts the thin layer sheet; undigested oligomers on each side of the line remain at the origin during chromatography in the second dimension. However, enzymatic digestion proceeds to completion only when the enzyme solution contains tRNA, which presumably competes with the labelled oligomers for binding sites on the PEI-cellulose, thus rendering them more accessible to enzyme attack. The products of the digestion are nucleoside 5'-phosphates, and the four common nucleotides as well as the 5'-phosphates of 5-methyldeoxycytidine and N^6-methyldeoxyadenosine are well separated by one or other of the solvent systems listed in Table 1.

The 3' terminal residue of the oligomer is not detected in the two-dimensional separation technique because, at the end of the phosphodiesterase and phosphatase treatments (Steps 1 and 2), this residue exists as a nucleoside, which is not labelled by the kinase. However, we have adapted a 3' terminal labelling procedure for use in identifying the nucleotide at this position if this information is desired. It involves the use of terminal deoxynucleotidyl transferase to add a single ^{32}P labelled adenosine 5'-phosphate to the 3' end of a separate sample of the oligomer, followed by degradation of the product with spleen phosphodiesterase. The terminal deoxyribonucleoside is thus released as a labelled 3' phosphate, which is then identified by a one-dimensional separation on a PEI-cellulose sheet. The relative mobilities of the 3' nucleotides in Solvent D used for this

separation are essentially the same as those listed for the 5' nucleotides in Table 1; inorganic phosphate migrates below dGp, and other contaminants in commercial ATP preparations remain at the origin. Under the conditions normally used for terminal transferase reactions, several ribonucleotide residues are added and, in the present application, this would cause difficulty in interpreting the TLC patterns because two nucleotide spots would always be present. However, we have found that, by using a large oligomer to ATP ratio (10:1 or greater) and by substituting Mg^{++} for the commonly used Co^{++} ion, multiple additions are not observed.

The procedures described above constitute a simple and accurate method for determining the sequences of oligodeoxyribonucleotides of up to about twenty nucleotides in length. The analysis employs readily available equipment and is quite sensitive, using less than 0.1 A_{260} unit of oligomer for the spleen phosphodiesterase reaction mixture with only a small portion of this mixture required for application to the PEI-cellulose thin layer sheet. The method should be particularly useful in the analysis of chemically synthesized oligomers, because it has the capacity to determine sequence, state of purity, and the nature and location of unusual or modified bases.

ACKNOWLEDGEMENTS

We thank Dr. J.A. Shapiro of MSDRL, Merck & Company and Ms. M.J. Brunden for providing some of the oligonucleotides for analysis. This work was supported by grants from the National Institutes of Health (GM 11518, GM 19395 and Training Grant GM 07211).

REFERENCES

1. Brownlee, G.G., and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
2. Tu, C-P.D., Jay, E., Bahl, C.P. and Wu, R. (1976) Anal. Biochem. 74, 73-93.
3. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman L. and Moldave, K. Eds., Vol. 65, pp. 499-560, Academic Press, New York.
4. Banaszuk, A.M., Deugau, K.V., Sherwood, J., Michalak, M. and Glick, B.R. (1983) Anal. Biochem. 128, 281-286.
5. Tu C-P.D. and Wu, R. (1980) in Methods in Enzymology, Grossman L. and Moldave, K. Eds., Vol. 65, pp. 620-638, Academic Press, New York.
6. Black, D. M. and Gilham, P. T. (1984) Fed. Proc. 43, 2048.
7. Wu, R., Tu, C-P.D. and Padmanabhan, R. (1973) Biochem. and Biophys. Res. Comm. 55, 1092-1099.
8. Gough, G.R., Singleton, C.K., Weith, H.L. and Gilham, P.T. (1979) Nucleic Acids Res. 6, 1557-1570.
9. Gough, G.R., Collier, K.J., Weith, H.L. and Gilham, P.T. (1979) Nucleic Acids Res. 7, 1955-1964.
10. Gough, G.R., Brunden, M.J., Nadeau, J.G. and Gilham, P.T. (1982) Tetrahedron Lett. 23, 3439-3442.

11. Griffin, B. E. and Reese, C. B. (1963) Biochim. Biophys. Acta 68, 185-192.
12. Bernardi, A. (1974) Anal. Biochem. 59, 501-507.
13. van de Sande, J.H., Kleppe, K. and Khorana, H.G. (1973) Biochemistry 12, 5050-5055.
14. Randerath, K., Randerath, E., Chia L.S.Y., Gupta, R.C. and Sivirajan, M. (1974) Nucleic Acids Res. 1, 1121-1141.