

---

DNA sequence analysis: a general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping\*

---

Ernest Jay, Robert Bambara, R. Padmanabhan and Ray Wu

---

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14850, USA

---

Received 10 December 1973

---

#### ABSTRACT

Several electrophoretic and chromatographic systems have been investigated and compared for sequence analysis of oligodeoxyribonucleotides. Three systems were found to be useful for the separation of a series of sequential degradation products resulting from a labeled oligonucleotide: (I) 2-D electrophoresis<sup>†</sup>; (II) 2-D PEI-cellulose; and (III) 2-D homochromatography. System (III) proved generally most informative regardless of base composition and sequence. Furthermore, only in this system will the omission of an oligonucleotide in a series of oligonucleotides be self-evident from the two-dimensional map. The sequence of up to fifteen nucleotides can be determined solely by the characteristic mobility shifts of its sequential degradation products distributed on the two-dimensional map. With this method, ten nucleotides from the double-stranded region adjacent to the left-hand 3'-terminus and seven from the right-hand 3'-terminus of bacteriophage  $\lambda$  DNA have been sequenced. Similarly, nine nucleotides from the double-stranded region adjacent to the left-hand 3'-terminus and five nucleotides from the right-hand terminus of bacteriophage  $\phi$ 80 DNA have also been sequenced. The advantages and disadvantages of each separation system with respect to sequence analysis are discussed.

#### INTRODUCTION

Sequence determination of DNA molecules has lagged far behind sequence determination of RNA molecules. Among the various reasons,<sup>1</sup> two major difficulties have been the lack of base specific endonucleases and the large size of DNA molecules. The recent discovery of several base-sequence specific restriction endonucleases<sup>2-8</sup> has allowed the production of a population of smaller segments of homogeneous DNA molecules. This advance, coupled with the development of DNA sequence analysis by the "primer extension" method using DNA polymerase<sup>9-13</sup> to

---

<sup>†</sup>Abbreviations used: 2-D electrophoresis (two-dimensional electrophoresis), electrophoresis on cellulose acetate strip at pH 3.5 (pyridine-acetate buffer) followed by electrophoresis on DEAE-cellulose paper (DE 81) at pH 1.9 (formic-acetic acid buffer) or pH 3.5 (pyridine acetate buffer). 2-D PEI-cellulose, electrophoresis on cellulose acetate strip at pH 3.5 (pyridine-acetate buffer) followed by chromatography on PEI-cellulose thin layer plate (cellulose impregnated with polyethyleneimine) in LiCl containing 7 M urea. 2-D homochromatography, electrophoresis on cellulose acetate strip at pH 3.5 (pyridine-acetate

allow production of short, labeled DNA fragments of defined sequence, makes it highly desirable to develop simple, rapid methods of sequencing fragments in the 10 to 20 nucleotide size range. Nucleotide sequence studies at the termini of long DNA molecules have already appeared in the literature<sup>14-19</sup> but the majority of the sequence information has so far been limited to three to nine nucleotides. Furthermore, the procedures usually required laborious analyses and combinations of multiple experiments. A major restriction has been the limitation of poor fractionation of larger oligomers by electrophoresis on DEAE-cellulose paper.

In our present work, we have investigated various well known 2-D fractionation systems for DNA sequence analysis solely on the basis of the mobilities of the oligonucleotides distributed on the 2-D map. In this investigation, we have terminally labeled synthetic oligonucleotides, of chain lengths up to fourteen, with either T4 polynucleotide kinase and  $\gamma$ -<sup>32</sup>P-ATP at the 5'-end<sup>20</sup> or with calf thymus deoxynucleotidyl terminal transferase<sup>21</sup> and  $\alpha$ -<sup>32</sup>P-rNTP at the 3'-end<sup>22</sup>. The 5'-labeled oligomers were partially digested by snake venom phosphodiesterase, and the 3'-labeled oligomers partially digested by spleen phosphodiesterase to produce labeled sequential degradation products down to mononucleotides. In the case of double-stranded DNA such as  $\lambda$  and  $\phi$ 80 DNA, terminal labeling was performed by DNA polymerase-catalyzed partial repair-synthesis<sup>23</sup> with the appropriate  $\alpha$ -<sup>32</sup>P-dNTP. In this method, only one 3'-end is selectively labeled in a given experiment. Thus, the left-hand 3'-end was labeled with one <sup>32</sup>P-dAMP and the right-hand 3'-end with one <sup>32</sup>P-rGMP separately.<sup>19</sup> After terminal labeling, the DNA was digested with pancreatic DNase to produce a series of oligonucleotides, each ending in the same 3'-labeled terminus. These sequential partial degradation products were fractionated in one or more of the 2-D systems. The sequence of the oligonucleotides was determined by the characteristic mobility shifts of the degradation products standardized by several synthetic oligomers. Specific sequence information could be obtained for up to fifteen nucleotides by fractionating the partial products on 2-D homochromatography (system III).

---

*(Continuation of footnote from preceding page.)*

buffer) followed by homochromatography on DEAE-cellulose thin layer plate in 2% partially hydrolyzed yeast RNA containing 7 M urea. Homo-mix, a solution of partially hydrolyzed RNA containing 7 M urea for use in homochromatography. d-value (differential value), the difference in mobility between two adjacent oligonucleotides in homochromatography. m-value, as defined by Sanger *et al.*<sup>27</sup> d<sub>5</sub>N, <sup>32</sup>P-labeled 5'-deoxyribonucleotide. d<sub>3</sub>N, <sup>32</sup>P-labeled 3'-deoxyribonucleotide.

## MATERIALS

*DNA.*  $\lambda$  DNA was isolated from the lambda phage obtained by heat induction of *E. coli* strain W3350 ( $\lambda$ C<sub>I</sub> 857) or in larger amounts by heat induction of *E. coli* strain M5107 ( $\lambda$ C<sub>I</sub> 857 S7). One sample of  $\lambda$  DNA was a gift of Dr. G. Changas.  $\phi$ 80 DNA was isolated from the bacteriophage  $\phi$ 80 which is available as a prophage in the *E. coli* strain Y mel ( $\phi$ 80) inducible by ultraviolet irradiation

*Enzymes.* *E. coli* DNA polymerase used was as reported previously.<sup>23,24</sup> Bacterial alkaline phosphatase (BAP-c), pancreatic DNase and snake venom phosphodiesterase were purchased from Worthington Biochemical Co. The venom phosphodiesterase was further purified<sup>25</sup> before use. Purified spleen phosphodiesterase with no detectable phosphatase activity was a gift of Dr. G. Bernardi. Calf thymus deoxynucleotidyl terminal transferase was a gift of Dr. R. Roychoudhury.

*Labeled nucleoside triphosphates and chemicals.*  $\alpha$ -<sup>32</sup>P-dATP (20-75 mCi/ $\mu$ mole),  $\alpha$ -<sup>32</sup>P-UTP and  $\alpha$ -<sup>32</sup>P-GTP (60-100 mCi/ $\mu$ mole) were purchased from New England Nuclear Co., and  $\gamma$ -<sup>32</sup>P-ATP (60-100 mCi/ $\mu$ mole) were purchased from ICN (Irvine, California). The yeast RNA for the preparation of the Homo-mix was purchased from Gallard-Schlesinger Chemical Mfg, New York, and the urea was from Allied Chemicals. The DEAE-cellulose plates were made from a 7.5:1 mixture of cellulose MN-300 HR and DEAE-cellulose 300 purchased from Brinkmann Instruments, and were developed in home-made one-inch wide plexiglass tanks. The PEI-cellulose on plastic sheets (20 x 20 cm) were purchased from Brinkmann Instruments.

## METHODS

*Preparation of Homo-mixes I to VI (Table I).* 20 gm of yeast RNA (free acid) was suspended in water (88-80 ml). While cooling in an ice-bath, the appropriate amounts of 5N KOH (12-18 ml) were added and the mixture was shaken at 37°C for 24 hours for Homo-mixes I through V and for 48 hours for Homo-mix VI. 100 ml of water was added and the mixtures were neutralized with 1N HCl. 420 gm of solid urea was added and the volume made up to 1 liter with deionized water to give a solution containing 2% partially hydrolyzed RNA in 7M urea. The mixtures were filtered before use.

*Phosphorylation of an oligonucleotide with T4 polynucleotide kinase*<sup>20</sup> and  $\gamma$ -<sup>32</sup>P-ATP. 100-200 pmoles of oligodeoxynucleotide in 5  $\mu$ l of solution containing 66 mM Tris-HCl (pH 7.8), 6.6 mM MgCl<sub>2</sub>, 15 mM dithiothreitol and 66  $\mu$ M  $\gamma$ -<sup>32</sup>P-ATP was incubated at 37°C with 2-4 units of T4 polynucleotide kinase for 2 hours. The 5'-<sup>32</sup>P labeled oligonucleotide was purified and desalted on a G-50 (fine) Sephadex column after elution with 0.05 M triethylammonium bicarbonate (pH 8.0).

TABLE I

RNA (gm)	H <sub>2</sub> O (ml)	5N KOH (ml)	KOH/RNA *	Time of reaction at 37°C (hr)	Longest oligomer expected to move
20	88	12	1.2	24	36
20	86	14	1.4	24	30
20	84	16	1.6	24	22
20	82	18	1.8	24	16
20	80	20	2.0	24	14
20	80	20	2.0	48	12

\* Molar ratio of KOH to RNA is expressed as ratio of KOH to mononucleotides.

*Labeling an oligonucleotide with a single <sup>32</sup>P-ribomonucleotide using calf thymus deoxynucleotidyl terminal transferase and α-<sup>32</sup>P-NTP.* 100-200 pmoles of oligonucleotide in 5 μl of solution containing 240 mM potassium cacodylate, 50 mM Tris (pH 7.6), 1.6 mM CoCl<sub>2</sub>, 0.16 mM dithiothreitol, 150 μM α-<sup>32</sup>P-UTP or -GTP and 1.6 μg (30 units) calf thymus deoxynucleotidyl terminal transferase, was incubated at 37°C for 4 hours. The mixture was then treated with 20 μl of 0.35 M KOH for 12 hours at 37°C, then brought to pH 8.5 by the addition of 1N HCl. Bacterial alkaline phosphatase (5 μg) was added and incubated at 45°C for 2 hours. This procedure removes any additional ribonucleotides incorporated after the first.<sup>22</sup> The labeled oligonucleotide was purified and desalted on a G-50 Sephadex column. It was eluted from the column with 0.05 M triethylammonium bicarbonate.

*Removal of the 3'-terminal ribonucleoside from an oligonucleotide-<sup>32</sup>P-rN.* The labeled oligonucleotide-<sup>32</sup>P-rN was oxidized in a solution containing 4 μl water, 1 μl 2N triethylammonium bicarbonate and 2 μl 0.1 M NaIO<sub>4</sub> at room temperature for 1 hour. 5 μl of 1 M n-propylammonium bicarbonate (pH 7.5) was added and the solution was kept at 45°C for 3 hours. 1 μl of 0.2 M glycerol was added to destroy the excess NaIO<sub>4</sub>. The triethylammonium bicarbonate was removed by evaporation in a desiccator. The product is oligonucleotide-3'-<sup>32</sup>P.

*Labeling of the r-strand of λ DNA and φ80 DNA with one <sup>32</sup>P-rGMP using DNA polymerase and α-<sup>32</sup>P-GTP.* A mixture containing 50 μg of DNA, 80 mM imidazole (pH 6.9), 70 mM KCl, 180 mM NaCl, 0.5 mM MnCl<sub>2</sub>, 10 mM dithiothreitol, 2 μM α-<sup>32</sup>P-rGTP (80-100 mCi/μmole) and 10-20 units of *E. coli* DNA polymerase<sup>1,26</sup> in a final volume of 0.6 ml was incubated at 5°C for 4 hours. The level of incorporation reached a plateau after 2-3 hours. Samples of 5 μl were removed

at 1 hour intervals to determine the level of incorporation. The labeled DNA was precipitated three times with 10% trichloroacetic acid after addition of 50-100 µg carrier salmon sperm DNA. The precipitate was taken up in 0.4 M NaOH and incubated for 12 hours at 37°C, then brought to pH 8.5 by bubbling with CO<sub>2</sub> gas. Bacterial alkaline phosphatase (5 µg) was added and incubated at 45°C for 2 hours. This procedure removes any additional ribonucleotides incorporated after the first. The DNA was then ready for pancreatic DNase digestion.

*Labeling the l-strand of λ DNA and φ80 DNA with <sup>32</sup>P-dAMP using DNA polymerase and α-<sup>32</sup>P-dATP.* A mixture containing 100 µg of DNA, 70 mM potassium phosphate (pH 6.9), 10 mM MgSO<sub>4</sub>, 10 mM dithiothreitol, 70 mM NaCl, 1 µM α-<sup>32</sup>P-dATP (20-80 mCi/µmole) and 10-20 units of *E. coli* DNA polymerase<sup>23</sup> in a final volume of 1 ml was incubated at 5°C. Aliquots of 5-20 µl were removed every 30 minutes to determine the level of incorporation. Incubation was continued until after a plateau was reached (2-3 hours). Excess EDTA was added to the mixture and it was dialyzed for 24 hours against 10 mM tris-HCl (pH 7.8) - 1 mM EDTA. After concentration to about 0.5 ml, the λ DNA was purified in an alkaline sucrose gradient of 5-20% (w/v) sucrose in 1 M NaCl, 0.1 M NaOH and 0.001 M EDTA. The centrifugation was continued for 6 hours at 36,000 rev/min at 5°C in an SW 41 rotor. Samples of 0.3 ml were collected from the bottom and the fractions containing the intact labeled DNA were pooled for precipitation in 10% trichloroacetic acid.

*Partial snake venom phosphodiesterase digestion of 5'-<sup>32</sup>P-labeled oligonucleotides.* The 5'-labeled oligonucleotide was digested at 37°C in 10 µl containing 30 µg RNA carrier (partially hydrolyzed yeast RNA), 50 mM tris-HCl (pH 8), 5 mM MgCl<sub>2</sub> and 0.2 µg of venom phosphodiesterase. Five 2 µl samples were removed at increasing time intervals and blown into 50 µl of 1 M ammonium hydroxide containing 1 mM EDTA. Samples were taken after 10, 20, 40, 60 and 90 minutes for oligonucleotides with chain lengths of 5-6 and up to 3 hours for chain lengths up to 12-14. The combined digest was dried in a desiccator before dissolving in 2-5 µl of water for electrophoresis.

*Partial spleen phosphodiesterase digestion of 3'-<sup>32</sup>P-labeled oligonucleotides.* The oligonucleotide-3'-<sup>32</sup>P was digested at 37°C in 10 µl containing 30 µg RNA carrier, 3 mM potassium phosphate (pH 6), 0.3 mM EDTA, 0.1% Tween 80 and 0.25 µg of spleen phosphodiesterase. Samples were taken at 5, 10, 20, 40 and 60 minutes for chain lengths of 12-14.

*Partial pancreatic DNase digestion of 3'-terminal labeled DNA.* The DNA after labeling reaction was mixed with 50 µg of carrier salmon sperm DNA and

precipitated three times in 15% trichloroacetic acid. After three extractions with ether, the DNA was dissolved in 200  $\mu$ l of 0.1 M ammonium hydroxide. The solution was neutralized by bubbling with CO<sub>2</sub> gas and divided into 4 tubes each containing 5 mM MgCl<sub>2</sub> and pancreatic DNase (0.4, 1.0, 2.0, 6.0  $\mu$ g). After incubation at 37°C for 2.5 hours, the mixtures were boiled for 2 minutes, dried and dissolved in 5  $\mu$ l of water for electrophoresis.

*2-D homochromatography.* 2-5  $\mu$ l of the sample was applied to the cellulose acetate strip (2.5 x 52 cm) and 0.2  $\mu$ l of the standard mixture of three dyes<sup>27</sup> was applied 0.5 cm from each side of the sample. Electrophoresis was carried out at pH 3.5 for 20-30 minutes at 3,000 V. After transfer of the oligomers from the cellulose acetate strip onto the DEAE-cellulose on glass plate,<sup>28</sup> the DEAE-cellulose plate was allowed to air-dry at room temperature for 2-4 hours. A Whatman 3 MM paper was clipped to the top of the plate and uniformly pressed against the DEAE-cellulose by a plastic strip (2 cm x width of plate). The DEAE-cellulose plate was first developed in water after 0.2  $\mu$ l of the standard dye mixture had been spotted 2 cm apart along the origin of the second dimension. Chromatography was then carried out in the appropriate Homo-mix (in plexiglass tanks (24 x 24 x 2.5 cm or 44 x 24 x 2.5 cm) until the blue dye marker was one inch from the top of the plate (3 hours for 20 x 20 and 10-12 hours for 20 x 40 plates). We found it necessary to use multiple application of the dye markers (2 cm apart) in order to specify exactly the position of the solvent front along the entire width of the plate.

### RESULTS AND DISCUSSION

(I) *2-D electrophoresis: cellulose acetate at pH 3.5 followed by DEAE-cellulose paper at pH 1.9 or 3.5.* This 2-D fractionation system was introduced by Sanger *et al.*,<sup>27</sup> for sequence analysis of oligoribonucleotides. From their studies it was concluded that the sequence of a short oligoribonucleotide can be deduced from the mobility of its sequential degradation products by a set of *m*-values for each product on a DEAE-cellulose paper separation at either pH 1.9 or 3.5. The *m*-value is defined as  $[(\text{mobility } pA\dots pX) - (\text{mobility } pA\dots pXpY)] / (\text{mobility } pA\dots pXpY)$ . The *m*-value is characteristic for each of the four mononucleotides represented by *pY*. However, for a given fractionation of a T1 RNase or pancreatic RNase digest, because of the specificity of the cleavages which produce the oligoribonucleotides, differences between only 3 ribonucleotides need to be distinguished. In cases involving a DNA digest, mobility differences between all four deoxynucleotides must be considered. With this aim in mind, both Murray<sup>29</sup> and we<sup>19,13</sup> have attempted to develop a mapping technique for DNA

sequence analysis. It was observed that the *m*-values given by Sanger *et al.*<sup>27</sup> for RNA apply also to DNA, with few exceptions. It was also evident that the direction of the shifts of mobilities of these partial degradation products in the 2-D system in the first dimension are characteristic of the four mononucleotides (Fig. 1). The addition of a dpC causes a mobility shift to the right, a dpT or dpG causes a shift to the left, and a dpA, in most cases, gives a vertical shift. In the case of the addition of a dpA to an oligonucleotide containing only dpC or a mixture of dpC and dpA, a small shift to the left is observed which can easily be distinguished from the large shift due to the addition of a dpG or dpT. The addition of a dpA to an oligomer containing mostly dpT and dpG will cause a very small shift to the right, much smaller than the shift observed for addition of a dpC.

One major difficulty in applying this technique to DNA is the fact that both pG and pT have overlapping *m*-values, and the addition of either causes a mobility shift in the first dimension to the left, which makes their unambiguous distinction impossible. After using various synthetic oligodeoxynucleotides (such as pT-T-T-T-T, pG-G-G-C-G, pC-A-G-T-T, pG-T-A-A, pA-G-T-C-C) to ascertain that the *m*-values are applicable to DNA, we have labeled the right-hand end of  $\lambda$  DNA with one <sup>32</sup>P-rGMP.<sup>19</sup> After digestion with pancreatic DNase to obtain labeled fragments of increasing chain length with the same 3'-terminus, the digest was subjected to 2-D electrophoresis fractionation (Fig. 1). Oligomer 1 was characterized to be dpGrpG by co-electrophoresis with a synthetic d(pGpG) marker. Oligomer 2 shifted to the right in the first dimension which is indicative of a gain of a dpC. The small *m*-value (0.06) in the second dimension is also in agreement with the addition of a dpC. A vertical shift and an *m*-value of 0.40 indicates a gain of dpA in oligomer 3. Oligomer 4 was shifted to the left with an *m*-value of 1.4, which can be due to a gain of either a dpG or dpT. The *m*-values in the pH 3.5 system (not shown) also agree with this sequence. To differentiate whether oligomer 4 was due to a gain of a dpG or dpT, the spots were eluted and were run in another system (one-dimensional homochromatography) which will be discussed later.

Similarly, the left-hand end of  $\lambda$  DNA was labeled with a <sup>32</sup>P-dAMP and the pancreatic DNase digest was fractionated in the 2-D electrophoresis system (Fig. 2). Accordingly, oligomer 1 was characterized by co-electrophoresis with d(pGpA) and oligomer 9 was deduced as dp(G or T)-(G or T)-A-C-C-C-(G or T)-C-G-A. Again, the ambiguity between dpG or dpT was resolved by running the eluted oligonucleotides in one-dimensional homochromatography. Each oligomer was further characterized by complete spleen phosphodiesterase digestion which gave the expected

nearest neighbor  $dGp^*$  for both the natural 3'-ends (right- and left-hand ends) of  $\lambda$  DNA.<sup>24</sup> This data again shows that 2-D electrophoresis alone cannot give the complete sequence of an oligonucleotide containing dpG or dpT. Furthermore, since the range of the m-values is rather large, omission of a single nucleotide, particularly dpC from an adjacent oligomer, will not be self-evident.

$\phi 80$  DNA<sup>30</sup> has the same cohesive end sequence as  $\lambda$  DNA,<sup>24</sup> and so it is possible to label its 3'-ends in the same way as in  $\lambda$  DNA. The right-hand 3'-end was labeled with a  $^{32}P$ -rGMP and the pancreatic DNase digest was run on 2-D electrophoresis. The dimer dpGrpG was characterized by co-electrophoresis with a standard and all of the spots showed the correct  $dGp^*$  nearest neighbor. All of the spots appeared in the same positions as the corresponding spots in the experiment with the 3'-terminally labeled fragments from the right-hand end of  $\lambda$  DNA. The spots were isolated and further characterized by one-dimensional homochromatography.

Figure 3 shows the 2-D electrophoresis separation of the pancreatic DNase digest of the left-hand terminally labeled  $\phi 80$  DNA. Just as with  $\lambda$  DNA, the dimer co-electrophoresed with d(pGpA). The trimer and tetramer spots are in the same positions as the corresponding oligomers of  $\lambda$  DNA. The spots corresponding to the pentamer and larger fragments are in different positions from the corresponding spots from  $\lambda$  DNA (compare figures 2 and 3). From this separation the sequence appeared to be (G or T)-C-(G or T)-(G or T)-C-G-A. These spots were also isolated and run on homochromatography to differentiate dpG from dpT.

(II) *Chromatography on PEI-cellulose.* Various laboratories have reported the use of PEI-cellulose thin layer chromatography<sup>31</sup> (cellulose impregnated with polyethyleneimine) for the fractionation of oligoribonucleotides.<sup>18,32,34</sup> We have also investigated the possibility of using the PEI-cellulose system for mapping the partial digestion products of short oligodeoxynucleotides. Murray<sup>18</sup> has recently used pyridine-formate (pH 3.5) buffer with PEI-cellulose plate for DNA sequence analysis of short oligonucleotides. From our preliminary studies using this system, we found that although it is capable of fractionating larger oligonucleotides than the 2-D electrophoresis systems, it has the following disadvantages: (a) In this system, although the m-values for the gain of a dpG or a dpT are much larger than those for a dpA or dpC, they vary considerably depending on the  $R_F$  values of the oligomers being compared. Because of the variance and overlap of mobility shifts, this system like the 2-D electrophoresis system cannot unequivocally distinguish a dpG from a dpT. (b) In order to move larger oligonucleotides, higher concentration of pyridine-



formate has to be used. Under such conditions, the smaller oligonucleotides either move with the solvent front or are compressed near the solvent front, particularly if the smaller oligonucleotides are rich in dpC and dpA.

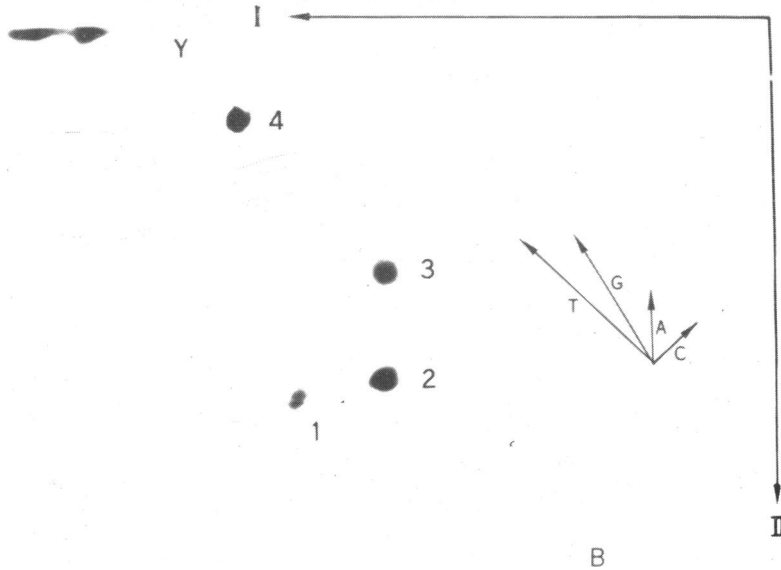
With these points in mind, it would be advantageous to have a system where the  $m$ -values remain reasonably constant regardless of the  $R_f$  values and also, where both the smaller and larger oligonucleotides will be well fractionated. Chromatography on PEI cellulose in different concentrations of LiCl containing 7 M urea seems to meet these requirements.<sup>31</sup>

Figure 4 shows the mobilities of the partial venom phosphodiesterase digestion of 5'-<sup>32</sup>P-T-T-T-T and 5'-<sup>32</sup>P-d(G-G-G-C-G) on one-dimensional PEI-cellulose thin layer chromatography in various concentrations of LiCl containing 7 M urea. We observe that the  $m$ -values get smaller with increasing LiCl concentrations, but they remain rather constant for each nucleotide tested. The  $m$ -value for the gain or loss of a dpG residue is twice that for a dpT or dpC. Thus, this system can be used in conjunction with 2-D electrophoresis for differentiating between a dpG or dpT addition.

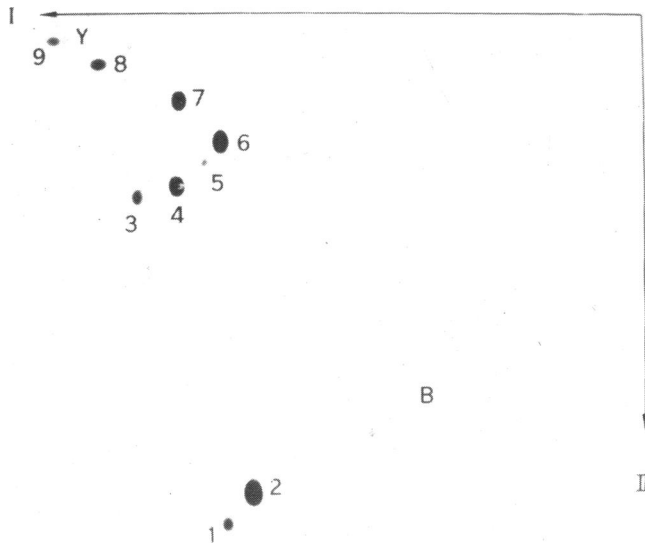
Since chromatography on PEI-cellulose plates in LiCl-7 M urea can distinguish the addition of a dpG from a dpT, and electrophoresis on cellulose acetate distinguishes dpC, dpA and dpT or dpG, a 2-D fractionation by electrophoresis on cellulose acetate, pH 3.5, followed by chromatography on PEI-cellulose should be sufficient to define the sequence of an oligonucleotide by the mobility shifts of its partial degradation products. Figure 5 shows the map of the partial venom phosphodiesterase digest of a 5'-<sup>32</sup>P-labeled synthetic tetradecamer d(pA-G-T-C-C-A-T-C-A-C-T-T-A-A).<sup>‡</sup> From the mobilities of the intermediates in this PEI-cellulose system, it could be seen that the  $m$ -values for the gain of a dpC and dpT are much smaller than the  $m$ -value for the gain of a dpG. However, the  $m$ -value for the gain of a dpA is also small. Since the gain of a dpC or a dpA has a characteristic shift in the first dimension, and the gain of a dpG from a dpT can be clearly distinguished in the second dimension, this 2-D fractionation system seems useful for sequence analysis by the characteristic mobility shift method.

(III) *2-D homochromatography system.* Homochromatography was introduced by Brownlee and Sanger<sup>33</sup> for fractionating and obtaining large partial digestion ( $T_1$  or pancreatic RNase) fragments for further RNA sequence analysis. Recently, Sanger *et al.* have used this system for fractionating DNA digests.<sup>12</sup> Ling also used homochromatography for fractionating the partial digests of

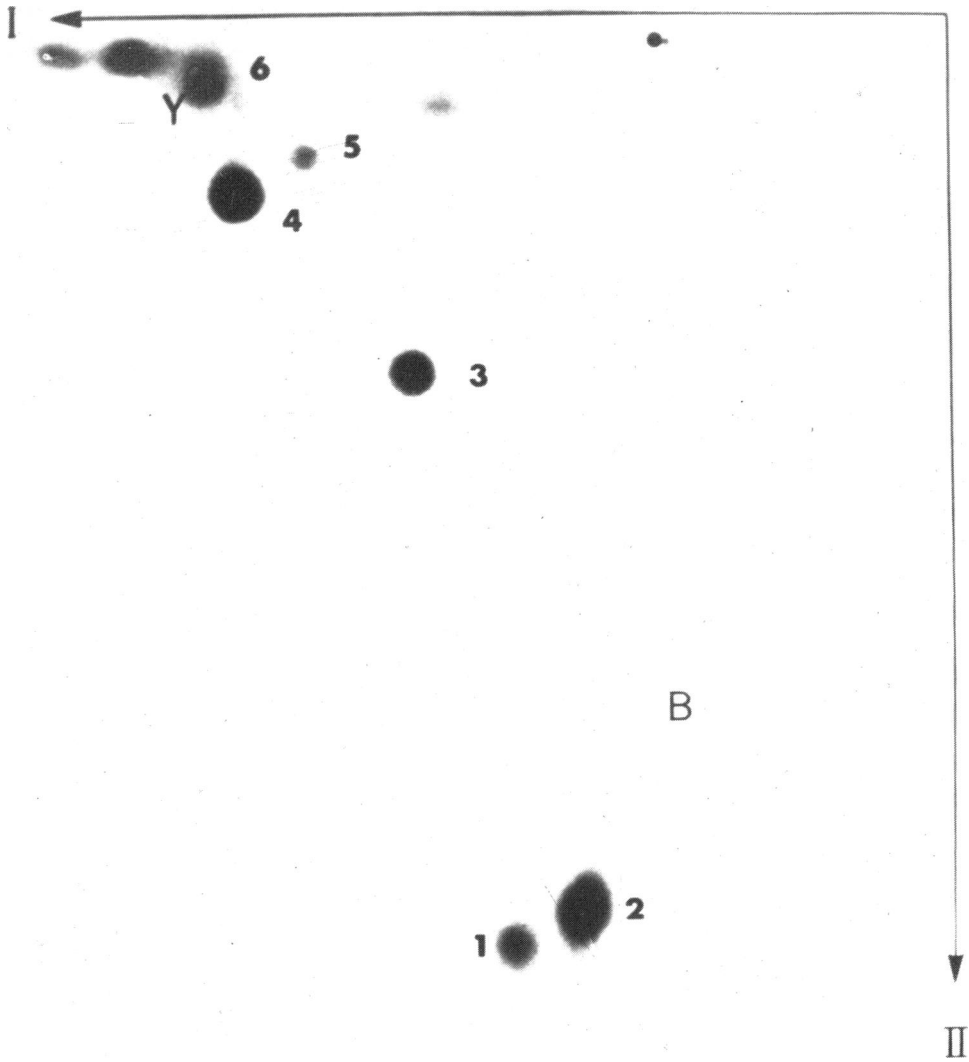
<sup>‡</sup> The synthesis of the tetradecamer will be published elsewhere.



**FIGURE 1** -- 2-D electrophoresis fingerprint of the pancreatic DNase fragments of  $\lambda$  DNA labeled at the right-hand 3'-end with  $^{32}\text{P}$ -rGMP. B and Y represent blue and yellow dyes, respectively.



**FIGURE 2** -- 2-D electrophoresis fingerprint of the pancreatic DNase fragments of  $\lambda$  DNA labeled at the left-hand 3'-end with  $^{32}\text{P}$ -dAMP.



**FIGURE 3** -- 2-D electrophoresis fingerprint of the pancreatic DNase fragment of  $\phi 80$  DNA labeled at the left-hand 3'-end with  $^{32}\text{P}$ -dAMP.

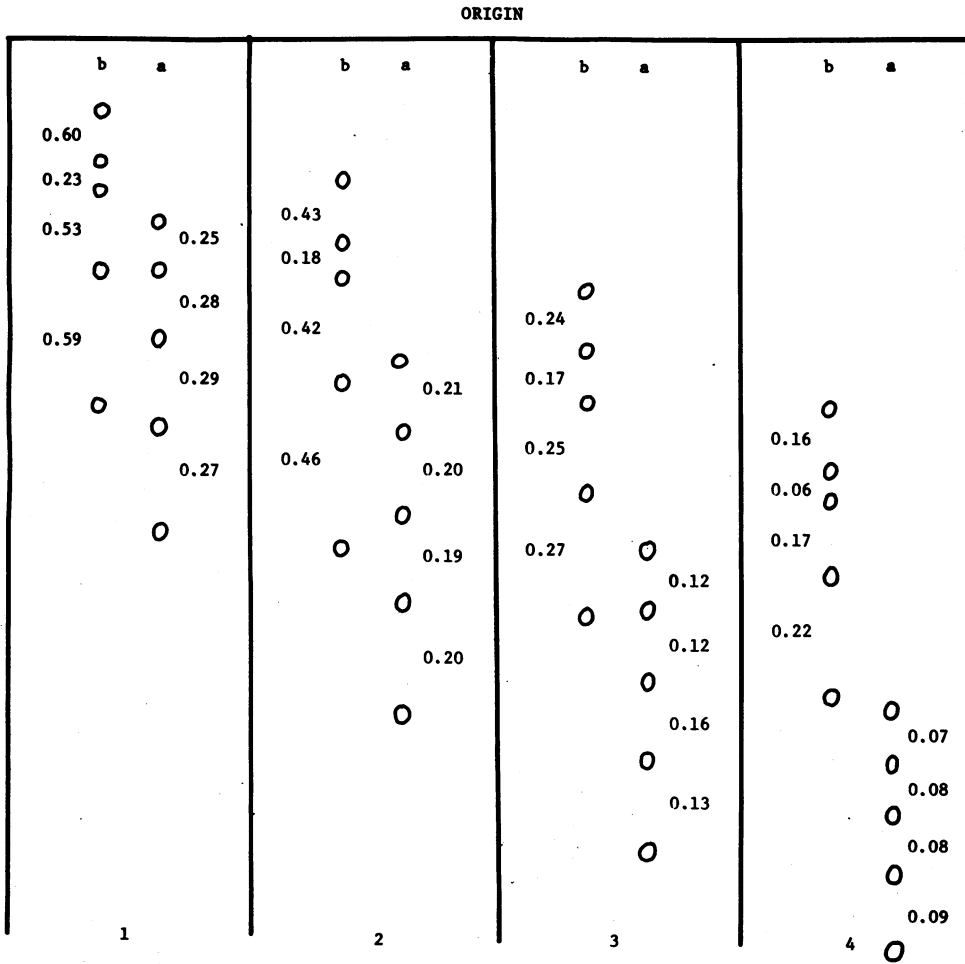


FIGURE 4 -- One-dimensional chromatography and the m-values of the partial snake venom phosphodiesterase digest of (a)  $^{32}\text{pT-T-T-T}$  and (b)  $\text{d}(^{32}\text{pG-G-G-C-G})$  on PEI-cellulose thin layer plates in 7 M urea and (1) 0.6 M LiCl, (2) 0.7 M LiCl, (3) 0.8 M LiCl, (4) 0.9 M LiCl.

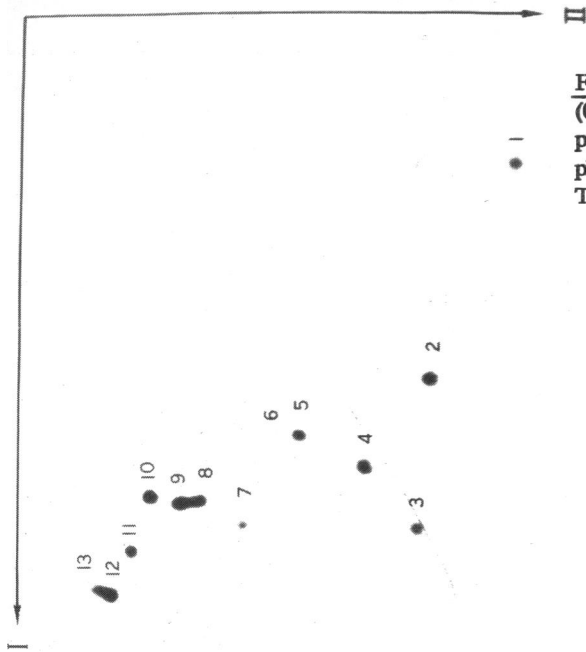
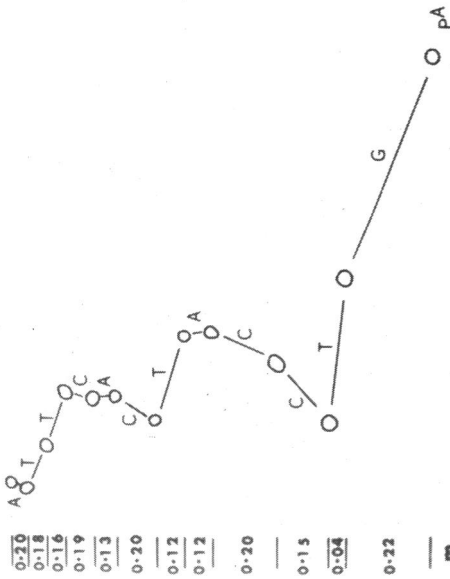


FIGURE 5 -- 2-D PEI-cellulose (0.85 M LiCl-7 M urea) fingerprint of the partial venom phosphodiesterase digest of d(<sup>32</sup>pA-G-T-C-C-A-T-C-A-C-T-T-A-A).



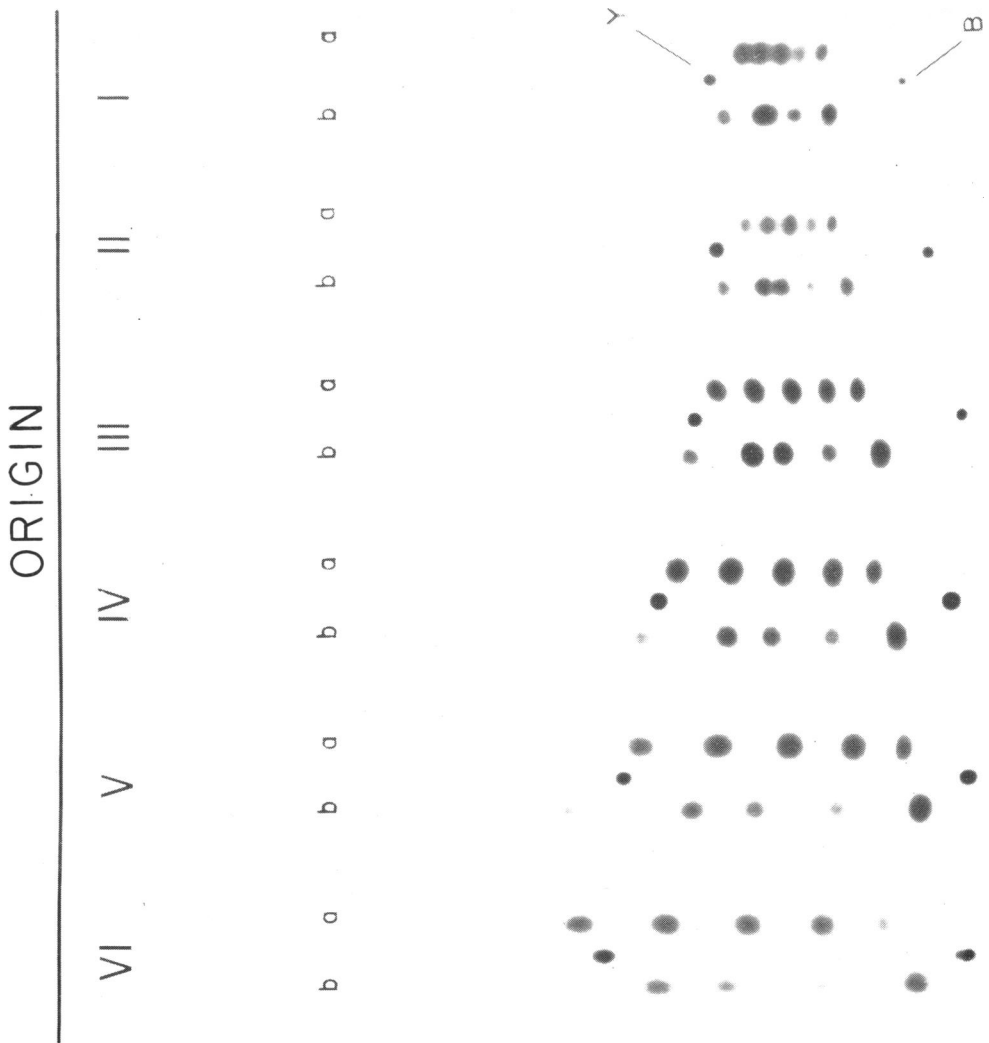
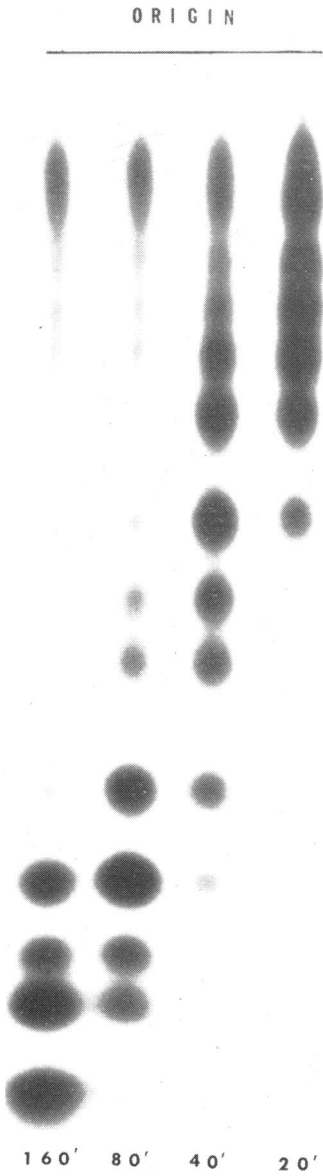


FIGURE 6 — One-dimensional homochromatography of the partial venom phosphodiesterase digest of (a)  $^{32}\text{pT-T-T-T-T}$  and (b)  $\text{d}(^{32}\text{pG-G-G-C-G})$  in Homo-mixes I through VI (Table I).



**FIGURE 7** -- One-dimensional homochromatography of the 5'-labeled tetradecamer d(<sup>32</sup>pA-G-T-C-C-AT-C-A-C-T-T-A-A) at increasing time of incubation.

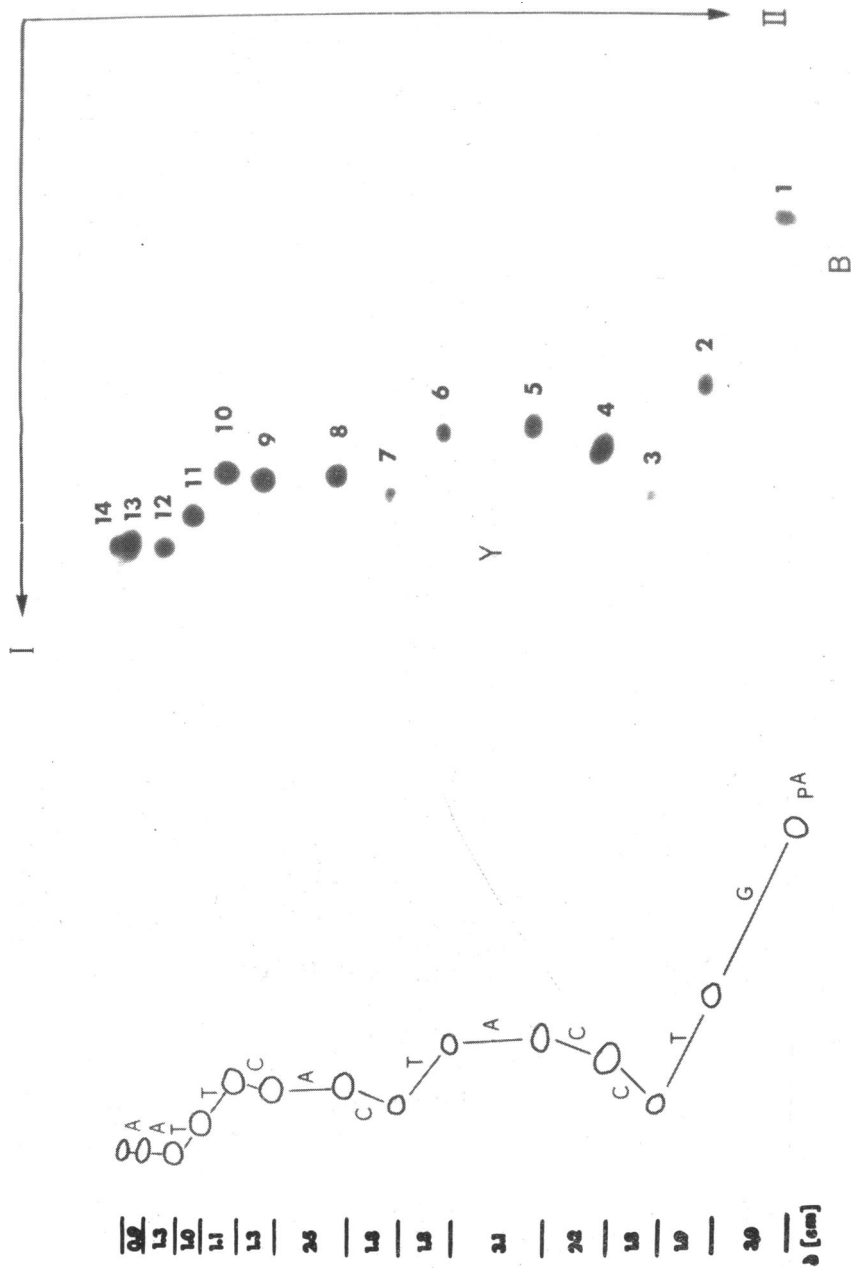


FIGURE 8 -- 2-D homochromatography (Homo-mix IV, 20 x 40 cm) fingerprint of the venom phosphodiesterase digest of d(32pA-G-T-C-C-A-T-C-A-C-T-T-A-A).



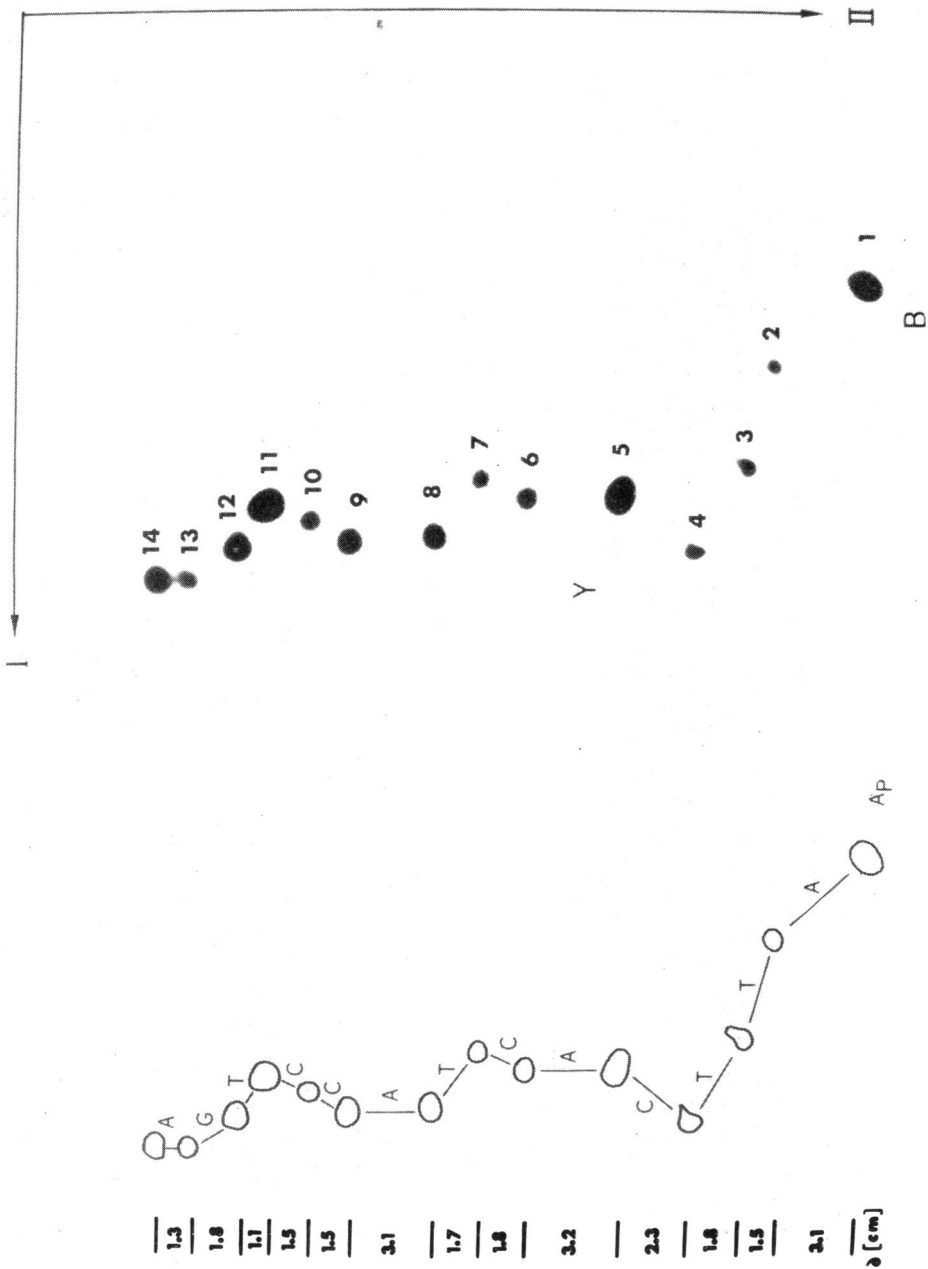
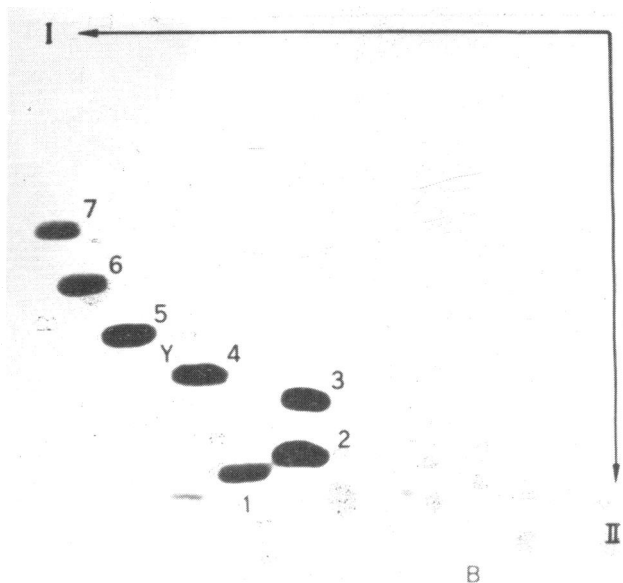
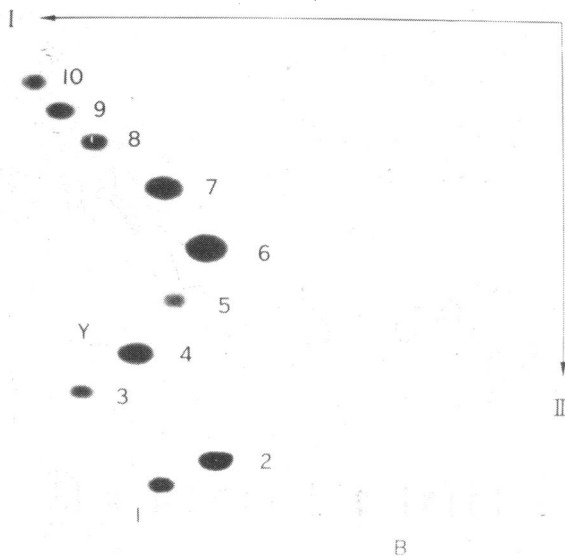


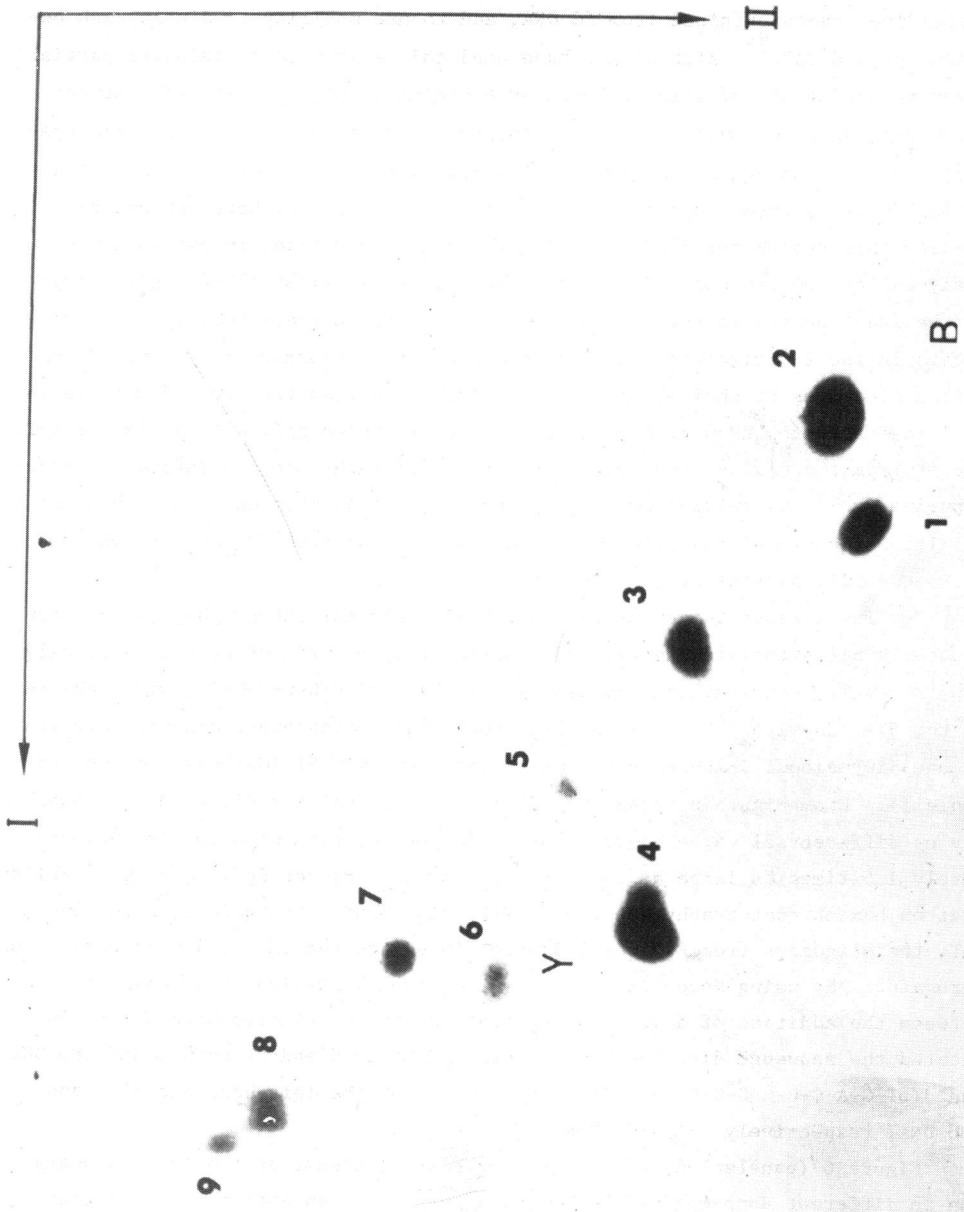
FIGURE 9 -- 2-D homochromatography (Homo-mix V, 20 x 40 cm) fingerprint of the spleen phosphodiesterase digest of the 3'-<sup>32</sup>P-labeled tetradecamer d(A-G-T-C-C-A-T-C-A-C-T-T-A-A-<sup>32</sup>p).



**FIGURE 10** -- 2-D homochromatography (Homo-mix VI, 20 x 20 cm) fingerprint of the pancreatic DNase fragments of  $\lambda$  DNA labeled at the right-hand 3'-end with one  $^{32}\text{P}$ .



**FIGURE 11** -- 2-D homochromatography (Homo-mix VI, 20 x 20 cm) fingerprint of the pancreatic DNase fragments of  $\lambda$  DNA labeled at the left-hand 3'-end with one  $d^{32}\text{P}$ .



**FIGURE 12** — 2-D homochromatography (Homo-mix VI, 20 x 20 cm) fingerprint of the pancreatic DNase fragments of  $\phi 80$  DNA labeled at the left-hand 3'-end with one  $d^{32}pA$ .

pyrimidine tracts obtained from fd DNA, and to use mobility to distinguish between dpT and dpC.<sup>35</sup> Ziff *et al.* have used this system to fractionate partial exonuclease digests of oligonucleotides containing dpT, dpC and dpA, but not for direct sequence analysis.<sup>36</sup> It appeared that in this chromatographic system, one can distinguish the change of mobility due to a loss or a gain of a purine (A or G) from a pyrimidine (C or T) nucleotide. We have extensively studied this system for distinguishing all four nucleotides in DNA sequence analysis for two reasons: (a) Very large oligonucleotides (10-60 units long) can be fractionated in this system by using the appropriate Homo-mix. In contrast, in the 2-D electrophoresis system, sequence information is limited to oligonucleotides of chain-lengths 5-10. (b) Since from the first dimension in cellulose acetate, pH 3.5, one can usually distinguish dpC, dpA and dpG or dpT, and if from the second dimension in homochromatography one can indeed distinguish a purine from a pyrimidine nucleotide (dpG from dpT), then one can deduce the complete sequence of the oligonucleotide solely from the 2-D mobility shifts of its sequential partial degradation products.

We first wished to demonstrate whether by homochromatography, one can consistently differentiate the gain of a purine from a pyrimidine, and especially dpG from dpT. The synthetic pentamers, dT(pT)<sub>4</sub> and d(G-G-G-C-G), were labeled at the 5'-ends with <sup>32</sup>P. The partial venom phosphodiesterase digests were run on one-dimensional DEAE-cellulose plate with Homo-mix VI (Methods section and Table I). From Figure 6 (panel 6), it is evident that the difference of mobility or differential value (d-value) for the loss or gain of a dpG is approximately 1.5 times as large as that of a pyrimidine, dpT or dpC. Having established that on homochromatography one can clearly distinguish the gain of a pG from a dpT, the oligomers from Figures 1-3 were eluted and run on one-dimensional homochromatography using Homo-mix VI. This (not shown) resolved the ambiguity between the addition of a pG or pT encountered in 2-D electrophoresis and indicated the sequence d(pT-T-A-C-G) for the right-hand end of both λ and φ80 DNA, and d(pT-G-A-C-C-C-G-C-G) and d(pT-C-T-G-C-G) for the left-hand end of λ and φ80 DNA, respectively.

Figure 6 (panels 1-6) also shows the partial digest of the two pentamers run in different Homo-mixes. From this figure, one can estimate the largest oligonucleotide that can be fractionated in each of the six mixtures. Thus the largest oligonucleotide that will move significantly in Homo-mixes I through VI, respectively, can be estimated to be approximately 36, 30, 22, 16, 14 and 12. Of course much bigger pieces can be fractionated by increasing the RNA concentration, but the resolution (d-value) will decrease accordingly. One advantage

in our method of preparation of the Homo-mixes is the use of limited amounts of alkali for the hydrolysis of the RNA. This procedure eliminates the dialysis step used by Sanger and co-workers.<sup>28,33</sup> During the dialysis, most of the smaller oligomers are lost. These might otherwise have been advantageous for the uniform fractionation of the small, labeled oligonucleotides under investigation.

To substantiate the validity of using the mobility shifts of the partial sequential degradation products of an oligomer on 2-D homochromatography system, several synthetic dodecamers<sup>13</sup> and tetradecamers have been used. Figure 7 shows the one-dimensional homochromatography (Homo-mix VI) of the products from the partial venom phosphodiesterase digestion of the 5'-<sup>32</sup>P-labeled tetradecamer d(pA-G-T-C-C-A-T-C-A-C-T-T-A-A) for varying lengths of time. Figure 8 shows the fingerprint (20 x 40 cm) of the combined partial venom phosphodiesterase digest of the same 5'-<sup>32</sup>P-labeled tetradecamer. Spot 1 was assigned as dpA\* from its position in both dimensions and from comparison with standard dpA. Addition of a dpG causes a shift to the left in the first dimension and a large mobility change (d-value = 3.0 cm) in the second dimension to spot 2. Addition of a dpT again causes a shift in the first dimension to the left but a small mobility change (d-value = 1.9 cm) in the second dimension to spot 3. A shift to the right to spot 4 is indicative of the addition of a dpC residue which also gives a small mobility change in the second dimension (1.8 cm). Thus, if one proceeds from the mononucleotide to the largest oligonucleotide in this 2-D system in this manner, one can arrive at the sequence d(pA-G-T-C-C-A-T-C-A-C-T-T-A-A) without ambiguity. The differential values in the second dimension decrease slightly but regularly with increasing chain length. Thus, to differentiate a purine or pyrimidine shift, one must compare its differential value with those of its neighboring spots.

Figure 9 shows the 2-D fingerprint of the partial spleen phosphodiesterase digestion of the same tetradecamer labeled at the 3'-end with <sup>32</sup>P prepared as mentioned in the Methods section to give d(A-G-T-C-C-A-T-C-A-C-T-T-A-Ap\*). This 3'-labeled tetradecamer was then partially digested with spleen phosphodiesterase and fractionated. From the pattern shown in Figure 9, one can arrive at the same tetradecamer sequence by following the mobility shifts of the partial products.

From these results, one observes that homochromatography has a further advantage over the other systems. 2-D electrophoresis (system I) alone cannot be accepted to provide the complete sequence of an oligonucleotide from the mobilities of its partial degradation products because the accepted range of m-values for each of the bases is so wide that it is not obvious if a single

oligomer in the partial digest is missing. In homochromatography (system III), since the d-value for the gain of a purine nucleotide is about 1.5 that of a pyrimidine nucleotide, it would become obvious if one of the intermediate oligomers in the digest was missing. For example, the mobility change for the consecutive gain of two pyrimidines is too large for the gain of a purine. Therefore, if one observes that the d-value between two oligomers is larger than that expected for the gain of a purine nucleotide, one can conclude that one intermediate between the two oligomers must have been missing.

Figure 10 shows the 2-D fingerprint of the pancreatic DNase digest of  $\lambda$  DNA labeled at the right-hand 3'-end with a single rGMP, as described previously, on electrophoresis on cellulose acetate followed by homochromatography in Homo-mix VI. Following the 2-D "mobility shift" rules starting from oligomer 1, which was characterized to be dpGrpG, one can arrive at the sequence d(pG-G-T-A-C-G)rpG for spot 7. Thus, in addition to distinguishing the gain of dpG from dpT which 2-D electrophoresis fails to do (Figure 1), this system also provided further complete sequence information.

Figure 11 shows the fingerprint of the pancreatic DNase digest of  $\lambda$  DNA labeled at the left-hand 3'-end with  $^{32}$ P-dAMP on 2-D homochromatography system. Following the mobility shift rules, once again, one can resolve the ambiguity encountered in the 2-D electrophoresis system mentioned earlier, and independently arrive at the sequence d(pT-T-G-A-C-C-G-C-GpA) for spot 10.

Figure 12 shows the corresponding information as obtained from  $\phi$ 80 DNA. Using the mobility shift rules one can read off the sequence d(pT-G-A-T-C-T-G-C-GpA). Thus, the left-hand sequences from the 3'-ends are similar in  $\lambda$  DNA and  $\phi$ 80 DNA but not identical.<sup>19</sup> This technique can readily pinpoint the differences.

From these results, it was concluded that the 2-D electrophoresis on cellulose acetate, pH 3.5, followed by homochromatography on DEAE-cellulose (system III) is more useful and informative than the other two systems studied for DNA sequence analysis. Oligonucleotides of up to pentadecamer, regardless of composition and base sequence, can be unambiguously sequenced solely from the mobility shifts of their partial degradation products on the 2-D map with the suitable Homo-mix. Occasionally, the relative mobility of the mono- and dinucleotides in the first dimension (see Figure 9) does not strictly follow the mobility rule. However, combining data from partial venom phosphodiesterase digest with that from partial spleen phosphodiesterase digest will confirm and double check the sequence. The validity of this method has been completely confirmed by using two other terminal labeling procedures followed by sequence analysis.<sup>13</sup>

## REFERENCES

- 1 Wu, R., Donelson, J., Padmanabhan, R. and Hamilton, R. (1972) *Bulletin de L'Institut Pasteur* 70, 203.
- 2 Kelly, T.J., Jr. and Smith, H.O. (1970) *J.Mol.Biol.* 51, 393.
- 3 Hedgpeth, J., Goodman, H.M. and Boyer, H.W. (1972) *Proc.Nat.Acad.Sci.U.S.* 69, 3448.
- 4 Mertz, J. and Davis, R.W. (1972) *Proc.Nat.Acad.Sci.U.S.* 69, 3370.
- 5 Meselson, H., Yuan, R. and Heywood, J. (1972) *Ann.Rev.Biochem.* 41, 447.
- 6 Gromkova, R. and Goodgal, S.H. (1972) *J.Bacteriol.* 109, 987.
- 7 Sack, G.H., Jr. and Nathans, D. (1973) *Virology* 51, 517.
- 8 Sharp, P.A., Sugden, B. and Sambrook, J. (1973) *Biochemistry* 12, 3055.
- 9 Wu, R. (1972) *Nature New Biology* 236, 198.
- 10 Padmanabhan, R., Padmanabhan, Raji and Wu, R. (1972) *Biochem.Biophys.Res. Comm.* 48, 1295.
- 11 Loewen, P.C. and Khorana, H.G. (1973) *J.Biol.Chem.* 248, 3489.
- 12 Sanger, F., Donelson, J.E., Coulson, A.R., Kössel, H. and Fischer, D. (1973) *Proc.Nat.Acad.Sci.U.S.* 70, 1209.
- 13 Wu, R., Tu, C.D. and Padmanabhan, R. (1973) *Biochem.Biophys.Res.Comm.*, 55, 1092.
- 14 Englund, P. (1972) *J.Mol.Biol.* 66, 209.
- 15 Murray, K. (1973) *Biochemistry* 131, 569.
- 16 Brezinski, D.P. and Wang, J.C. (1973) *Biochem.Biophys.Res.Comm.* 50, 398.
- 17 Weigel, P.H., Englund, P.T., Murray, K. and Old, R.W. (1973) *Proc.Nat. Acad.Sci.U.S.* 70, 1151.
- 18 Murray, K. and Murray, N.E. (1973) *Nature* 243, 134.
- 19 Ghangas, G., Jay, E., Bambara, R. and Wu, R. (1973) *Biochem.Biophys.Res. Comm.* 54, 998.
- 20 Richardson, C.C. and Weiss, B. (1967) *J.Mol.Biol.* 23, 405.
- 21 Kato, K., Goncalves, J.M., Houtes, G.E. and Bollum, F.J. (1967) *J.Biol. Chem.* 242, 2780.
- 22 Roychoudhury, R. and Kössel, H. (1971) *Eur.J.Biochem.* 22, 310.
- 23 Wu, R. (1970) *J.Mol.Biol.* 51, 501.
- 24 Wu, R. and Taylor, E. (1971) *J.Mol.Biol.* 57, 491.
- 25 Sulkoski, E. and Laskowski, M.Sr. (1971) *Biochim.Biophys.Acta* 240, 443.
- 26 Berg, P., Fancker, H. and Chamberlin, M. (1963) in *Informational Macromolecules*, Academic Press, New York, p. 467.
- 27 Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965) *J.Mol.Biol.* 13, 373.
- 28 Barrell, B.G. (1971) in *Procedures in Nucleic Acids*, eds. Cantoni, G.L. and Davis, D.R., Harper & Rowe, New York, Vol. 2, pp. 751-779.
- 29 Murray, K. (1970) *Biochem.J.* 118, 831.
- 30 Bambara, R., Padmanabhan, R. and Wu, R. (1973) *J.Mol.Biol.* 75, 741.
- 31 Randerath, K. and Randerath, E. (1967) in *Methods of Enzymology*, Vol. 12A, p. 323, eds. Grossman, L. and Moldave, K., Academic Press, New York and London.
- 32 Southern, E.M. and Mitchell, A. (1971) *Biochem.J.* 123, 613.
- 33 Brownlee, G.G. and Sanger, F. (1969) *Eur.J.Biochem.* 11, 395.
- 34 Griffin, B.E. (1971) *FEBS Letters* 15, 165.
- 35 Ling, V. (1972) *J.Mol.Biol.* 64, 87; *Proc.Nat.Acad.Sci.U.S.* 69, 742.
- 36 Ziff, E.B., Sedat, J.W. and Galibert, F. (1973) *Nature New Biology* 241, 3.

\* This is paper XIII in a series on NUCLEOTIDE SEQUENCE ANALYSIS OF DNA. Paper XII is by Wu, Tu and Padmanabhan (Reference 13). This work was supported by Grants GB-31485X from N.S.F. and GM-18887 from N.I.H.