DNA sequence analysis: a formula to predict electrophoretic mobilities of oligonucleotides on cellulose acetate*

Robert Bambara, Ernest Jay and Ray Wu[†]

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

Received 19 August 1974

ABSTRACT

A simple method has recently become available for sequence analysis of large oligonucleotide fragments. Sequences are derived from the characteristic mobility shifts of the sequential partial degradation products of the oligonucleotide on two dimensional homochromatography. We have now developed an empirical formula for predicting the relative mobilities of each of the partial products in the first dimension (electrophoresis on cellulose acetate gel). The formula allows a more precise interpretation of the sequence of the oligonucleotide. It eliminates the ambiguities present in the method previously reported for sequence analysis by simple inspection of the mobility shifts. In order to amplify the mobility shifts so that they may be more easily and accurately measured, methods have been developed for preparing and fractionating the oligonucleotides on 40 x 40 cm DEAE-cellulose plates. Both improvements have proven valuable for direct sequence analysis by mapping.

INTRODUCTION

In recent years, considerable effort has been devoted to understanding the mechanisms of control in gene expression and repression. Full understanding of these mechanisms depends, in part, on the elucidation of the nucleotide sequence of the DNA itself. Since the discovery of site specific restriction enzymes¹⁻⁹ considerable attention has been focused on mapping various functions expressed by different DNA segments. Digestion of long DNA molecules with these restriction enzymes produces specific short fragments of DNA which can assist in more precise localization of various biological functions. As the number of restriction enzymes being isolated increases, it is likely that unique specific segments of DNA, 20-50 nucleotides in length, will become available for sequence analysis. Therefore, it is essential to have simple methods of rapid direct sequence determination of these short oligonucleotides. Our laboratory has been concerned with developing such methods or improving existing methods.

Two dimensional systems for fractionation and sequence analysis of oligonucleotides. It has been known for some time that the relative electrophoretic mobilities of oligonucleotides on cellulose acetate at pH 3.5 can be used to provide some sequence information from these compounds^{10,11}. More recently, this system has been combined with homochromatography on DEAE-cellulose to produce a two-dimensional (2-D) fractionation system¹². Ling^{13,14} has used this 2-D system to determine the sequence of several long pyrimidine tracts from bacteriophage DNA. In his mobility-shift method, a pyrimidine tract was digested with an exonuclease to produce a series of degradation products of decreasing length^{10,15}. The digestion products were fractionated on the 2-D system, and the relative mobility of each oligonucleotide was compared with an oligonucleotide one nucleotide longer (or shorter). The addition of a dpC to any oligonucleotide always causes a mobility decrease on the first dimension (electrophoresis on cellulose acetate at pH 3.5). The addition of a dpT always causes a mobility increase. Thus, there is no ambiguity in distinguishing these two pyrimidine nucleotides in sequence analysis of the original oligonucleotide.

Distinction of three nucleotides by mobility shifts is required for RNA sequence analysis when the oligoribonucleotides are produced by base-specific RNases. Sequences of two dodecaribonucleotides were determined by Rensing and Schoenmakers¹⁶ using a combination of several analyses. They have discussed the possible use of mobility in this 2-D homochromatography system to give tentative sequences of dephosphorylated 3'-G terminated oligonucleotides which contain pU, pC and pA. They also pointed out the "occasional difficulty in distinguishing removal of AMP from the one of CMP". In DNA sequence analysis by mobility shifts, distinction of all four nucleotides is required. Several investigators¹⁷⁻²¹ have attempted to use the 2-D homochromatography system to analyze partial enzymatic digests of oligodeoxynucleotides. However, since it is not always possible to use electrophoresis on cellulose acetate (the first dimension) to differentiate the mobility shift resulting from the addition (or removal) of a dpA from dpG with phosphorylated oligonucleotides or a dpA from dpC with dephosphorylated oligonucleotides, it has not been possible to determine sequences unambiguously by mobility shifts. Recently, it has been shown that addition of purines can be distinguished from addition of pyrimidines for both phosphorylated²¹ and dephosphorylated²⁰ oligonucleotides on the homochromatography dimension. Distinction between dpA and dpC can then be made on the second dimension. However, since dpG and dpA are both purines, the second dimension is not helpful. Because of such difficulties, together with the

problem in the interpretation of mobilities of short oligonucleotides, a second method has always been used to establish the sequence. In some cases, the second method, though more tedious, is complete by itself and therefore the mobility method was only used to confirm the sequence instead of to determine the sequence^{17,18}. In other cases, the second method resolves some ambiguities in the mobility method¹⁹ or complements the mobility method^{20,21}.

Other workers have used different fractionation systems for sequence analysis. Murray & Murray²² used two one-dimensional electrophoresis systems to distinguish mobility shifts caused by the addition of any one of the four deoxynucleotides to a short oligodeoxynucleotide (2 to 5 nucleotides in length). This method can usually determine the sequence of oligonucleotides up to hexanucleotides. With longer oligodeoxynucleotides, they used a different 2-D system which includes electrophoresis on cellulose acetate followed by chromatography on PEI thin layer plates. Similar problems exist with this system, as with 2-D homochromatography, i.e., although a portion of the sequence of an oligonucleotide can frequently be deduced by mobility shifts, unambiguous determination of the entire sequence cannot be made.

The purpose of this communication is to present an empirical formula which is capable of predicting the electrophoretic mobility of any oligonucleotide on cellulose acetate at pH 3.5. The formula has been used successfully to predict mobility shifts and mobilities for many oligonucleotides up to very large sizes. In addition to predicting the mobilities of partial degradation products, the formula will also be valuable for verification of the composition of short oligonucleotides (from di- to tetranucleotides) determined by some other methods. Thus, with the help of this formula, exact sequence of an oligonucleotide can usually be unambiguously determined by using mobility-shift as the primary or sole method.

Theory on electrophoretic mobility of particles. A considerable amount of theory is available describing the movement of particles in an electric field^{23,24}. The velocity u of any charged particle in an insulating medium is given by

$$u = Eq/f$$

where E is the electric field strength, q is the charge on the particle and f is the frictional coefficient describing the drag on the particle.

Electrophoretic mobility, or movement at a standard field strength, is defined as,

$$U = \frac{u}{E} = \frac{q}{f}$$

For a spherical molecule, $f = C(M)^{1/3}$ where M is the molecular weight of the particle and C is a constant.

Under experimental conditions, this theoretical mobility is modified by the effects of counter ion mobility, solvent currents, supporting material structure, particle shape and flexibility, and a number of other factors. Most of these effects are not yet rigorously described. Experimentally, the frictional drag of large flexible polymers is found to obey the relation

$$f = C(M)^a$$

where "a" is a relatively constant value for any specific polymer, usually between 0.4 and 1.0²⁵.

On both filter paper²⁶ and zone²⁷ electrophoresis, the mobilities of short oligonucleotides (less than 20) were found to increase in mobility with increase in size, suggesting that $a < 1$. Longer oligonucleotides on zone electrophoresis all seemed to have the same mobility, suggesting $a \approx 1$.

Mobilities of oligonucleotides on cellulose acetate gel at pH 3.5 also seem to display this general behavior. Our own data show that very large DNA fragments have small mobilities in this system, suggesting that $a > 1$.

An equation to describe the mobilities of a limited size range of oligonucleotides in this system should simply have the form,

$$U = \frac{q}{C(M)^a}$$

where "a" is considered constant. In fact, the form $U = q/C(\text{molecular volume})^{1/3}$ has been used successfully to describe the mobilities of di and trinucleotides on paper electrophoresis at pH 3.5²⁸.

In the case of oligonucleotides, since the molecular weight (M) is approximately proportional to the number of nucleotides (N) the mobility can be expressed as

$$U = \frac{Q}{C'(N)^a}$$

where $Q = \sum q$ represents the sum of the net charge of each of the N number of nucleotides, and C' is the new constant for use with N.

From our experiments with deoxyoligonucleotides of defined sequences and various chain lengths (N), it was observed that in order for the formula to apply to a wide range of chain length it was necessary to express "a" as a function of N. As the chain length (N) increases, "a" also has to increase. Thus, for oligonucleotides it was found that

$$U = \frac{Q}{C'(N)^{f(N)}}$$

After testing with numerous oligonucleotides it was found that the empirical expression $a = b(1 + K \ln N)$ allows excellent agreement with the mobilities of oligonucleotides of wide size range. A logarithmic term was used because, from

experimental data, the value of a should increase rapidly with size for smaller oligonucleotides, but more slowly for larger oligonucleotides. In this expression, b and K are constants. Thus the mobilities of oligonucleotides are given by

$$U = \frac{Q}{C'(N)^b(1 + K \ln N)}$$

For sequence analysis, we have found it more convenient to express the mobilities of oligonucleotides relative to dpT which has a net charge of 1.0 at pH 3.5. Then, $C' = 1$, so the expression simplifies to:

$$U_T = \frac{Q}{N^b(1 + K \ln N)}$$

where U_T is the mobility of the oligonucleotide with respect to dpT. The most suitable values for the constants b and K were experimentally determined. The usefulness of this expression for calculating the expected mobility shifts in a series of sequential degradation products with respect to sequence analysis will be critically evaluated and discussed.

MATERIALS AND METHODS

Wet cellulose acetate strips (Cellogel) were obtained from Reeve Angel Co. Dry Cellulose acetate strips were obtained from Schleicher and Schuell Inc., New Hampshire. The terms "wet" and "dry" refer to packaging and storage conditions. Both types are soaked in running buffer just prior to use. The 40 x 40 cm DEAE-cellulose thin layer plates were prepared using home-made equipment. 40 x 40 glass plates were obtained from a local supplier. A large mounting board similar to the small one supplied by Brinkmann was made from plexiglass. An adjustable applicator (Figure 1) 40 x 6 x 7 cm was constructed with 3/16 inch thick plexiglass. The thickness of the DEAE-cellulose layer can be varied. Whatman paper of known thickness can be used to assist in setting of the thickness. The DEAE-cellulose plates (300 μ thick) were made from a 7.5 : 1 mixture of cellulose MN 300 HR and DEAE-cellulose 300 purchased from Brinkmann. The DEAE-cellulose plates with a 3 MM Whatman paper wick hanging from the top to the back were run as described earlier²¹.

RESULTS AND DISCUSSION

Two dimensional fractionation systems have been extensively used for separating oligonucleotides of various sizes and compositions^{10,13,14,29,30}. The most useful system for the first dimension has been found to be electrophoresis on cellulose acetate gel at pH 3.5. Unlike most of the second dimension separation systems, which separate mainly on the basis of size, this system separates approximately on the basis of composition of the oligonucleotide. Oligonucleotides which are rich in thymidine have very high mobility. Those

FIGURE 1 -- Applicator for Preparation of 40 x 40 cm DEAE-Cellulose Plates. The applicator (A) is constructed with 3/16 inch thick plexiglass. An adjustable slit (B) made of 1/16 inch plexiglass is attached to (A) by two screws. The thickness of the DEAE-cellulose layer can be varied by loosening and adjusting (B) upwards or downwards. While pouring the cellulose mixture into the applicator (A), a 3/16 inch thick plexiglass strip (C) is placed against (B) to prevent leakage while pouring prior to spreading of the plates by pulling away from (C).

rich in guanosine, adenosine and cytidine have respectively slower mobilities. This observation can be explained as the result of a net charge decrease from dpT through dpG and dpA to dpC at pH 3.5. Mobilities of homologous series of oligoribonucleotides resulting from RNA digests have been examined on two dimensional fractionation systems employing electrophoresis on cellulose acetate at pH 3.5 as the first dimension separation¹¹. The addition of a pC to any oligoribonucleotide was found to cause a decrease in mobility in the first dimension. The addition of a pU and pG caused an increase in mobility. The addition of a pA usually caused little mobility change. However, depending on the composition of the oligonucleotide, small changes in either direction were also observed. Oligodeoxyribonucleotides have recently been demonstrated to behave similarly^{17-20,29}. This information suggested that the mobility of oligonucleotides on cellulose acetate at pH 3.5 follows the simple electrophoretic equation

$$U = \frac{q}{C(M)^a}$$

where q represents the net charge of the molecule, M represents the molecular weight and C and a are constants. Table 1 shows the pKa and q values for

TABLE 1
Values of q for Mononucleotides

Mononucleotide	pKa (heterocyclic bases)	q from pKa	effective q
dpT	-	-1.0	-1.0
dpG	2.9	-0.72	-0.85
dpA	4.4	-0.41	-0.30
dpC	4.6	-0.08	-0.15
dTp	-	-1.0	-1.0
dGp	2.9	-0.73	-0.82
dAp	-	-	-0.31
dCp	-	-	-0.16
rpU	-	-1.0	-1.0
rpG	2.4	-0.92	-
rpA	3.7	-0.39	-
rpC	4.5	-0.10	-
rUp	-	-1.0	-1.0
rGp	2.3	-0.94	-0.86
rAp	3.7	-0.39	-0.35
rCp	4.3	-0.16	-0.17

each of the four ribo- and deoxyribomononucleotides. We have found that for electrophoresis on cellulose acetate at pH 3.5 for a small range of chain length (5-12) the same equation with $C = 1$ and $a = 0.66$ is sufficient to describe the mobilities of oligonucleotides with respect to dpT. However, in order for the equation to apply to a wide range of chain length, the expression $a = b(1 + K \ln N)$ was found to be most suitable, and the mobilities of oligonucleotides can simply be described by

$$U_T = \frac{Q}{N^{b(1 + K \ln N)}}$$

where U_T is the mobility of the oligonucleotide with respect to dpT, N is the number of component nucleotides and Q is the sum of q of each component nucleotide.

Effective net charge (q). It has been mentioned that from the values of pK_a for the dissociable groups³¹ of the four mononucleotides dpT, dpG, dpA and dpC one can calculate their net charges at any pH. Table 1 shows that the calculated q at pH 3.5 for dpT = -1.0, dpG = -0.72, dpA = -0.41 and dpC = -0.08. The effective charge of each nucleotide will be slightly different from these values because the pK_a 's for the nucleotides were determined in a different buffer and also because of differential shielding of the charged species by other ionic components of the electrophoretic buffer. Table 1 shows the electrophoretic mobility of the four nucleotides with respect to dpT. Since the value of q for dpT was defined as -1.0 at pH 3.5 these relative mobilities represent the effective charge for each of the nucleotides. Thus the effective charge q (experimentally determined) for dpT = 1, dpG = 0.85, dpA = 0.30 and dpC = 0.15. The total net charge Q of an oligonucleotide is then the sum of the effective charges of all the component nucleotides in the oligonucleotide. The values for the 3'-mononucleotides are similar to those of the 5'-mononucleotides. For convenience, the mobilities of oligonucleotides on cellulose acetate have been measured relative to the orange G dye marker. This yellow dye was found to migrate with a defined mobility relative to dpT. Using this value, the mobilities of all oligonucleotides can be defined with respect to dpT. The relative mobilities of orange G and dpT were found to be a function of voltage and the distance of migration. On wet cellulose acetate, as the distance of migration is increased, the orange G marker moves farther with respect to dpT. Although this effect is small, it is necessary to correct for it to locate the position of dpT from the position of the dye marker. Because of this observation, it is advisable to add a [¹⁴C] dpT as a marker during electrophoresis. The advantages are that the mobility of dpT can be accurately

FIGURE 2 -- Panel A shows the 2-D homochromatography pattern for degradation products of the oligonucleotide $d(^{32}\text{pA-G-T-C-C-A-T-C-A-C-T-T-A-A})$. A $[^{14}\text{C}]\text{dpT}$ standard was included in the sample. Panel B shows the predicted positions of the oligonucleotides in the first dimension. The second dimension mobilities are the same as in panel A. Predictions are based on the mobility of the standard dpT .

measured from the autoradiogram so the U_T for other nucleotides can be easily calculated without having to make the correction mentioned above, and also because it can be easily distinguished from the ^{32}P -labeled oligonucleotides, since it will not expose a second x-ray film placed on top of the first.

Determination of b and k. The effective charge q previously determined for each mononucleotide and the mobilities of numerous standard oligodeoxyribonucleotides were used to determine the optimal values for b and K . The values of b and K were chosen so that the predicted oligonucleotide mobilities best matched the observed mobilities. Table 2 shows a comparison of the observed mobilities of standard oligonucleotides and DNA fragments with those predicted by the equation using the values 0.50 for b and 0.165 for K . With these constants, the equation predicts the mobilities quite accurately from mononucleotide to 5000 nucleotide-long DNA. More importantly, the direction of mobility change from the addition of any given nucleotide to an oligonucleotide is accurately predicted.

Use of predicted mobilities. Table 2 shows some of the deoxyoligonucleotides which have been examined. The sequences of these oligonucleotides have previously been determined by some other methods. In most of the cases, the predicted mobilities of the nucleotides are very close to the observed value. More importantly, one can see that the equation predicts the observed direction of mobility changes for every nucleotide. It is especially useful for differentiating the addition of a dpG from a dpA. Since both cause a large mobility change (d value) in the second dimension²¹ and the addition of dpA can cause a mobility increase, though to a smaller extent than that caused by the addition of dpG, the use of the equation becomes necessary. The last two columns in Table 2 show the calculated mobilities of the oligonucleotides if dpA had been added instead of the correct dpG and vice versa. It is obvious that the incorrect addition either causes a mobility change in the opposite direction or if in the correct direction, it is far from the observed value. For example, Table 2 shows the values for the partial products of $d(\text{pGTAA})$. The nucleotide added to oligomer 3 in the second dimension (not shown) must be a purine²¹ (either dpA or dpG). The observed mobility shift in the first dimension is +0.16 (from 0.88 to 1.04). The predicted shift for the addition of a dpG is +0.21 (from 0.84 to 1.05) and that for a dpA is -0.07 (from 0.84 to 0.77), in the opposite direction. Thus the sequence is clearly pGTAA and not pATAA.

Figure 2 and Table 2 show the 2-D homochromatography fractionation of the partial venom digest of $d(\text{pAGTCCATCACTTAA})$. In both the figure and the table both the observed and the predicted mobilities are shown. Spot 1 has a mo-

TABLE 2
Electrophoresis of Phosphorylated Oligonucleotides

N	Deoxy- oligonucleotide	U_T (pred.)	U_T (obs.)	U_T (pred.) if G added	U_T (pred.) if A added
2	pAA	0.41	0.44		
3	pTAA	0.84	0.88		
4	pGTAA	1.05	1.04		0.77
2	pTT	1.35	1.35		
3	pGTT	1.49	1.52		1.20
4	pAGTT	1.34	1.46	1.58	
5	pCAGTT	1.19	1.33		
1	pAG	0.78	0.69		
2	pCAG	0.69	0.62		
1	pGC	0.68	0.66		
2	pGCC	0.60	0.60		
3	pGCCC	0.56	0.56		
1	pA	0.30	0.39		
2	pAG	0.78	0.83		0.40
3	pAGT	1.13	1.19		
4	pAGTC	0.98	0.99		
5	pAGTCC	0.88	0.90		
6	pAGTCCA	0.86	0.90	1.04	
7	pAGTCCAT	1.04	1.05		
8	pAGTCCATC	0.96	1.02		
9	pAGTCCATCA	0.94	1.00	1.06	
10	pAGTCCATCAC	0.89	0.97		
11	pAGTCCATCACT	1.00	1.07		
12	pAGTCCATCACTT	1.10	1.14		
13	pAGTCCATCACTTA	1.07	1.11	1.21	
14	pAGTCCATCACTTAA	1.04	1.09	1.18	
1	pA	0.30	0.28		
2	pAG	0.78	0.68		0.40
3	pAGG	1.05	0.92		0.76
4	pAGGT	1.28	1.19		
5	pAGGTC	1.14	1.05		
6	pAGGTCTG	1.25	1.18		1.08
7	pAGGTCCG	1.15	1.09		

TABLE 2 (con't)

N	Deoxy- oligonucleotide	U _T (pred.)	U _T (obs.)	U _T (pred.) if G added	U _T (pred.) if A added
7	pAGGTCGC	1.15	1.09		
8	pAGGTCGCC	1.07	1.07		
9	pAGGTCGCCG	1.15	1.25		1.03
10	pAGGTCGCCGC	1.08	1.22		
11	pAGGTCGCCGCC	1.03	1.19		
12	pAGGTCGCCGCCC	0.97	1.16		
2	pCA	0.31	0.28		
3	pCAG	0.68	0.69		0.40
4	pCAGT	0.98	0.96		
5	pCAGTT	1.19	1.21		
6	pCAGTTT	1.34	1.34		
7	pCAGTTTG	1.42	1.38		1.31
8	pCAGTTTGA	1.35	1.36	1.50	
9	pCAGTTTGAG	1.40	1.39		1.29
10	pCAGTTTGAGC	1.31	1.35		
11	pCAGTTTGAGCA	1.27	1.31	1.38	
12	pCAGTTTGAGCAT	1.34	1.38		
2	pCA	0.31	0.35		
2	pTT	1.38	1.39		
2	pGT	1.28	1.29		
2	pAT	0.90	0.89		
2	pAA	0.42	0.47		
2	pCC	0.21	0.18		
2	pAG	0.79	0.72		
2	pTG	1.28	1.28		
1	pA	0.30	0.32		
2	pAA	0.41	0.51	0.78	
3	pAAT	0.84	0.76		
4	pAATT	1.10	1.03		
5	pAATTG	1.24	1.16		1.08
6	pAATTGT	1.39	1.33		
7	pAATTGTT	1.50	1.40		
8	pAATTGTTA	1.49	1.39	1.53	
9	pAATTGTTAT	1.51	1.44		

bility characteristic of dpA^* . It was isolated and found to co-chromatograph with a dpA standard. The mobility change from spots 1 to 2 gave a large d value²¹ in the second dimension which indicated the addition of a purine (dpG or dpA). It also gave a large mobility increase (+0.44 from 0.39 to 0.83) in the first dimension indicating the addition of a dpG . The addition of dpA should give very little mobility change (+0.10 from 0.30 to 0.40). The mobility changes from oligomer 2 to 3, 4 and 5 in the second dimension gave small d values indicative of the addition of three consecutive pyrimidines²¹. From both the apparent shifts, and the relative shifts determined by the formula, the addition of pT and two consecutive dpC 's were assigned. The mobility change from oligomer 5 to 6 gave a large d value in the second dimension and no mobility increase (0.90 to 0.90) in the first dimension. This could have resulted from the addition of either dpG or dpA . From calculation, if a dpG had been added the mobility in the first dimension would give a mobility increase (+0.16 from 0.88 to 1.04). If a dpA had been used it would give a small mobility decrease (-0.02 from 0.88 to 0.86). Thus, the observed mobility must have been due to the addition of a dpA and not a dpG . Another pT and then dpC were added to give oligomer 8 followed by a purine to give oligomer 9. The addition of the purine nucleotide to oligomer 8 gave a small mobility decrease (-0.02 from 1.02 to 1.00). Addition of dpA is predicted to give a small decrease (-0.02 from 0.96 to 0.94), but addition of dpG is predicted to give a large increase (+0.10 from 0.96 to 1.06). Again the addition of dpA is confirmed. It should be emphasized that, only the relative mobility value between two adjacent oligonucleotides is needed for sequence analysis. Accurate interpretation of the data can usually be made even when the observed mobility deviates appreciably from the predicted values as in this case. The predicted mobility change for the addition of a dpA to oligomer 8 is -0.02 and that for the addition of a dpG is +0.10. Since the observed value was -0.02, the addition of a dpA was indicated. Following this sequential analysis of mobility changes as the length of the oligonucleotide increases one can unambiguously arrive at the complete sequence of the oligonucleotide.

Table 3 shows the observed and calculated mobilities of very large DNA fragments. It can be observed that the calculated values correspond quite well to the observed mobilities. Figure 3 summarizes the average mobilities of various sizes of oligonucleotides. Thus, the formula does predict the mobilities of even very large polynucleotides. However, it must be pointed out that although the formula can be used to predict the mobilities of very large oligonucleotides, it can only be used to assist sequence analysis by mobility shifts

FIGURE 3 -- Effect of Size on Average Mobility of Oligonucleotides on Cellulose Acetate Gel at pH 3.5. The graph shows predicted mobilities of oligonucleotides all having an average charge of 0.57 per molecule. The average charge is the statistical average of the values of q for the four deoxyribonucleotides.

TABLE 3

Electrophoresis of SV₄₀ DNA fragments obtained from digestion with Haemophilus Influenza endonuclease (Hin)

Fragment	N	U _T (pred.) [*]	U _T (obs.) [†]
Hin A	1200	0.31	0.38
Hin B	900	0.38	0.51
Hin C and D	500	0.53	0.63
Hin E and F	400	0.57	0.69
Hin H-K	200-250	0.73	0.57-72
SV ₄₀	5500	0.10	0.17
* Values for U _T (predicted) by taking the average q value of 0.57.			
† Run in 7 M Urea to prevent secondary structure formation.			

for oligonucleotides less than 30 nucleotides long. It is interesting to observe that the predicted and observed mobility of oligonucleotides increases with chain length up to about 40 nucleotides. After that the mobility starts decreasing until 10,000 nucleotides which hardly migrates.

Table 4 lists predicted and observed mobility data for oligonucleotides derived from the left and right hand 3'-ends of ϕ 80 DNA. Again the formula can be used to confirm sequence assignments, particularly those of addition of pA and pG.

Table 5 shows the values determined for two dephosphorylated oligomers. Comparative maps of actual and predicted mobilities of one of these oligonucleotides are shown in figure 4. It is not surprising to find small deviations from the calculated values, particularly for the small oligonucleotides since the number of nucleotides (N) is not directly proportional to molecular weight. As the number of nucleotides increase to 4 or 5 the difference becomes negligible. Nevertheless, use of the mobility formula is a help in the analysis of dephosphorylated oligonucleotides. Table 5 shows that although predicted mobilities can be used to distinguish addition of dpA from dpG, they often cannot distinguish addition of dpA from dpC, a problem encountered when dephosphorylated oligonucleotides are sequenced on 2-D homochromatography. This is because of the small charge difference between dpA

FIGURE 4 -- Panel A shows the 2-D homochromatography pattern for degradation products of the dephosphorylated oligonucleotide d(T-C-G-C-C-G-C-C-C). The oligonucleotide was labeled with ^{32}P at the 5'-end of each dpC. A [^{14}C]dpT standard was run independently. Panel B shows the predicted positions of the oligonucleotides in the first dimension. The second dimension mobilities are the same as in panel A. Predictions are based on the mobility of the standard dpT. A total of 3,000 cpm of ^{32}P was used for this map. If each spot were to be eluted and analyzed further, several times more starting counts of oligonucleotides would be required.

and dpC, 0.15 units, compared to the larger charge difference between dpA and dpG, 0.55 units. Fortunately, however, distinction between addition of a purine from a pyrimidine can be made by examination of oligonucleotide mobility in the second dimension, homochromatography^{19,21}.

This method of sequence analysis should also apply to ribonucleotides. Table 1 shows the calculated net charge and the effective net charge experimentally determined for each of the ribonucleotides. This method should be most useful for sequence analysis of the short RNA fragments generated by pancreatic RNase, T₁ and other specific endonucleases.

TABLE 4
Electrophoresis of Oligonucleotides Derived from
the 3'-Termini of Bacteriophage ϕ 80 DNA.

N	Deoxy- oligonucleotide	U _T (pred.)	U _T (obs.)	U _T (pred.) if G added	U _T (pred.) if A added
2	pGA	0.78	0.68		
3	pCGA	0.68	0.60		
4	pCCGA	0.91	0.82		0.68
5	ptGCCGA	1.14	1.13		
6	pCTGCCGA	1.03	1.00		
7	ptCTGCCGA	1.19	1.21		
8	pATCTGCCGA	1.14	1.18	1.25	
9	pGATCTGCCGA	1.22	1.34		1.12
10	ptGATCTGCCGA	1.32	1.40		
2	pGG	1.15	1.04		
3	pCGG	0.97	0.90		
4	pACGG	0.92	0.89	1.16	
5	ptACGG	1.14	1.11		
6	ptTACGG	1.30	1.33		
7	pGTTACGG	1.38	1.41		1.26
8	pGGTTACGG	1.45	1.50		1.34

It should be possible to apply the formula to predict mobilities of deoxy-oligonucleotides run on wet cellulose acetate at pH 3.5 ± 0.05 in any laboratory, using the same values of q, b and K which we have used. Table 6 lists the values of the denominator of the mobility formula for values of N up to 20. For different pH, dry cellulose acetate, or oligoribonucleotide separa-

TABLE 5

Electrophoresis of Dephosphorylated Oligonucleotides

N	Deoxy- oligonucleotide	$U_T(\text{pred.})$	$U_T(\text{obs.})$	$U_T(\text{pred.})$ if A added
2	TC	0.10	0.14	
3	TCG	0.52	0.46	0.24
4	TCGC	0.49	0.45	0.56
5	TCGCC	0.47	0.44	0.53
6	TCGCCG	0.67	0.65	0.53
7	TCGCCGC	0.64	0.64	0.68
8	TCGCCGCC	0.61	0.64	0.65
9	TCGCCGCCC	0.58	0.64	0.62
4	CCCC	0.13	0.22	
5	CGCCC	0.16	0.24	0.22
6	CCGCC	0.19	0.25	0.24
7	GCCGCC	0.40	0.49	0.28
8	CGCCGCC	0.40	0.49	0.49
9	TCGCCGCC	0.58	0.63	

TABLE 6

Values of the Denominator of the Mobility Formula

N	Denominator	N	Denominator
1	1	11	5.33
2	1.45	12	5.77
3	1.91	13	6.21
4	2.34	14	6.65
5	2.77	15	7.10
6	3.19	16	7.54
7	3.62	17	7.99
8	4.04	18	8.45
9	4.47	19	8.91
10	4.90	20	9.38

tions, the values of q and probably b and K have to be slightly modified. We have found that use of dry cellulose acetate slightly changes the relative positions of the mononucleotides with respect to the yellow dye marker and with respect to each other. Use of urea in either the wet or dry cellulose acetate system seems to have little effect, at least on oligonucleotides given in Table 2.

It has previously been demonstrated that of all available fractionation systems, 2-D homochromatography seems most generally applicable and informative. From these results, it is concluded that, in most cases, DNA sequences can be unambiguously determined by this mapping system. The use of the mobility formula aids accurate objective interpretation of the nucleotide sequences, replacing the subjective method of inspection.

It must be stressed that only the predicted relative mobilities of oligonucleotides are needed for the sequence determination. In all of the data we have obtained so far, the predicted relative mobilities of oligonucleotides have described the correct sequences. We are collecting more data to further substantiate the validity of the mobility formula.

Preliminary evidence suggests that homopolymers tend to give larger systematic error in observed mobility than heteropolymers. Similarly, stretches of the same nucleotide within an oligonucleotide also may cause some deviation from predicted behavior³². We expect that in some cases, an observed mobility may fall in between the predicted values for two possibilities (e.g., dpA or dpG). Under such circumstances, simple supplementary analysis such as 3'-end analysis, base composition or pyrimidine tract analysis should be carried out.

Moreover, we have found that the use of 40 x 40 cm DEAE-Cellulose plates allows easier and more accurate mobility measurements of oligonucleotides. Such accuracy is often required for measuring small mobility differences between large oligonucleotides.

We expect that the use of the mobility formula and the 40 x 40 cm thin layer plates will significantly aid reliable interpretation of sequence from 2-D homochromatography. The results presented here have put us one step closer to perfecting a simple method for DNA sequence analysis based solely on mobility shifts. Such a method is not only rapid but it allows sequence determination with lesser amounts of radioactive oligonucleotides, since they do not have to be isolated from the 2-D fractionation system for further analyses.

*This is paper XVIII in a series on NUCLEOTIDE SEQUENCE ANALYSIS OF DNA. Paper XVII is by Ghangas and Wu, J. Biol. Chem. 1974. This work was supported by research grants CA 14989 and GM 13887 from the National Institutes of Health and grant GB 40035 from the National Science Foundation. The authors wish to thank Chen-pei D. Tu and Kenneth Mariani for some of the data which appear in table 2.

†Authors in alphabetical order

REFERENCES

1. Kelly, T.J., Jr. and Smith, H.O. (1970) J. Mol. Biol. 51, 393.
2. Hedgpeth, J., Goodman, H.M. and Boyer, H. W. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3448.
3. Mertz, J. and Davis, R.W. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3370.
4. Gromkova, R. and Goodgal, S.H. (1972) J. Bacteriol. 109, 987.
5. Sharp, P.A., Sugden, B. and Sambrook, J. (1973) Biochemistry 12, 3055.
6. Meselson, H., Yuan, R. and Heywood, J. (1972) J. Bacteriol. 109, 987.
7. Sugisaki, H. and Takanami, M. (1973) Nature New Biology 246, 138.
8. Takanami, M. (1973) FEBS Letters 34, 318.
9. Takanami, M. and Koji, H. (1973) FEBS Letters 29, 267.
10. Sanger, R., Brownlee, G.G. and Barrell, B.G. (1965) J. Mol. Biol. 13, 373.
11. Barrell, B.G.G. (1971) Procedures in Nucleic Acids, eds. Cantoni, G.L. and Davis, D.R., Harper & Rowe, New York, vol. 2, pp. 751-779.
12. Brownlee, G.G. and Sanger, F. (1969) Europ. J. Biochem. 11, 395.
13. Ling, V. (1972) J. Mol. Biol. 64, 87.
14. Ling, V. (1972) Proc. Nat. Acad. Sci. U.S. 69, 742.
15. Holley, R.W., Madison, J.T. and Zamir, A. (1964) Biochem. Biophys. Res. Comm. 17, 389.
16. Rensing, R.F.E. and Schoenmakers, J.G.G. (1973) Eur. J. Biochem. 33, 8.
17. Ziff, E.B., Sedat, J.W. and Galibert, F. (1973) Nature New Biology 241, 34.
18. Sanger, F., Donelson, J.E., Coulson, A.R., Kossel, H. and Fischer, D. (1973) Proc. Nat. Acad. Sci. U.S. 70, 1209.
19. Ghangas, G., Jay, E., Bambara, R. and Wu, R. (1973) Biochem. Biophys. Res. Comm. 54, 998.
20. Wu, R., Tu, D.C. and Padmanabhan, R. (1973) Biochem. Biophys. Res. Comm. 55, 1092.
21. Jay, E., Bambara, R., Padmanabhan, R. and Wu, R. (1974) Nucleic Acid Research 1, 331.
22. Murray, K. and Murray, N.E. (1973) Nature 243, 134.
23. Van Holde, K.E. (1971) Physical Biochemistry, Prentice Hall, N.J.
24. Tanford, C. (1961) Physical Chemistry of Macromolecules, Wiley and Sons, N.Y.
25. Tanford, C., Kawahara, K. and Lapanje, S. (1967) J. Am. Chem. Soc. 89, 729.
26. Markham, R. and Smith, J.D. (1952) Biochem. J. 52, 558.
27. Olivera, B.M., Baine, P. and Davidson, N. (1964) Biopolymers 2, 245.
28. Markham, R. and Smith, J.D. (1951) Nature Lond. 168, 406.
29. Murray, K. (1970) Biochem. J. 118, 831.
30. Southern, G.M. and Mitchell, A. (1971) Biochem. J. 123, 613.
31. Sober, H.A., ed. (1970) Handbook of Biochemistry, 2nd edition, pp. J-209-210.
32. Tu, D., Jay, E., Bahl, C. and Wu, R. unpublished observations.