Neighboring nucleotide interactions during DNA sequencing gel electrophoresis

John M. Bowling, Kaylon L. Bruner¹, Joan L. Cmarik² and Clark Tibbetts^{1*}

Department of Mechanical Engineering, School of Engineering, ¹Department of Microbiology and Immunology and ²Center in Molecular Toxicology, Department of Biochemistry, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA

Received January 9, 1991; Revised and Accepted April 26, 1991

ABSTRACT

Electrophoretic separation of oligonucleotides in denaturing polyacrylamide gels is primarily a function of length-dependent mobility. The 3' terminal nucleotide sequence of the oligonucleotide is a significant, secondary determinant of mobility and separation. Oligomers with 3'-ddT migrate more slowly than expected on the basis of length alone, and thus are better separated from the preceding, shorter oligomers in the sequencing ladder. Oligomers with 3'-ddC are relatively faster than expected, and are therefore less separated. At the 3' penultimate position, -dC- increases and -dT- reduces separation. Purines at the 3' terminal or penultimate positions of oligonucleotides affect separation less than the pyrimidines. These results suggest specific interactions among neighboring nucleotides with important effects on the conformation of oligonucleotides during electrophoresis. These interactions are compared to compression artifacts, which represent more extreme anomalies of length-dependent separation of oligonucleotides. Knowledge of base-specific effects on electrophoretic behavior of DNA oligomers supplements the usual information available for determination of sequences; additionally it provides an avenue to thermodynamic and hydrodynamic investigations of DNA structure.

INTRODUCTION

Contemporary DNA sequencing depends on incremental separation of single-stranded DNA oligomers in denaturing polyacrylamide gels. The system of gel electrophoresis supports determination of DNA sequences from a single sample over a range from one to hundreds of nucleotides (1-3). Many DNA sequences, however, present local domains of anomalous oligomer yields or separations. These regions challenge efforts to design and automate objective and uniform procedures for acquisition and interpretation of DNA sequencing data.

Compression artifacts are striking anomalies of oligonucleotide separations in DNA sequencing ladders. The mobilities of compression-prone oligomers are not uniformly related to their relative lengths, and this compromises local sequence determination. Compressions are generally attributed to idiosyncratic DNA sequences, prone to form secondary structures such as hairpins or stem-loops, even in the denaturing environment of 7 to 8 M urea (1, 3-5). Electrophoresis at elevated temperatures can often restore compression-prone sequences to more uniform order in the ladder with respect to oligomer length (6). Purine deoxynucleotide analogues have also been used, during the DNA synthesis stage of sequencing, to provide products with reduced propensity to compression artifacts (7-11). Complementary strand sequence determinations often provide unambiguous results through regions compressed in the first strand's ladder. None of these approaches to dealing with mobility artifacts is universally applicable with satisfactory results. We decided to investigate anomalies of electrophoretic separation of oligonucleotides with the objective of better understanding their sequence-specific determinants.

The manual, autoradiographic approach to DNA sequencing presents a static view of oligomer ladders after a fixed period of electrophoresis. Mobility differences of oligomers are determined from their relative distances of migration from the loading well. This method has the disadvantage that longer oligomers traverse shorter distances, and are physically closer to their immediately shorter or longer neighbors. Recently introduced, automated DNA sequencers enable real-time detection of oligomers as they pass a fixed position in the gel (12-14). Mobility differences, in this case, are evaluated from the differences in time required for electrophoretic transport over a fixed distance from the loading well. The average temporal separations of oligomers are more uniform, over the range of oligomer lengths in a typical sequencing ladder, than are the average physical separations of the oligomers determined from a static view. Otherwise, the factors affecting the mobilities of oligonucleotides during gel electrophoresis are identical for the manual, static view approach and the automated, real-time approach.

* To whom correspondence should be addressed

In this report we use an automated DNA sequencer (14) to analyze the mobilities of oligonucleotides in DNA sequencing ladders. We expected the instrument to be useful for the study of mobility artifacts such as sequence-specific compressions. The improved precision of mobility determinations with the instrument led to discovery of a more subtle, but at the same time more pervasive anomaly of the electrophoretic separations of oligonucleotides. The temporal separations of successive oligomers are observed to vary extensively from oligomer to oligomer in a typical DNA sequencing ladder. The major determinant of this variance of separation is the 3' terminal DNA sequence of two to three nucleotides.

MATERIALS AND METHODS

Fluorescent labeling of DNA sequencing ladders with 3'-labeled terminators

Samples of DNA from phage M13 mp18, or spontaneous or induced mutants, were sequenced using the dideoxynucleotide chain termination strategy (5) and protocols developed for the fluorescence-based Du Pont GenesisTM 2000 automated DNA analysis system (14).

Annealing of template, 3 μ g M13 phage DNA, and 15 ng universal primer (the 17-mer from -40) was performed in 20 μ l, 65 mM Tris-HCl, pH 7.5, 80 mM NaCl, 32 mM MgCl₂ by heating 5 min at 95°C, followed by incubation for 10 to 20 min at 37°C. DNA synthesis followed using the annealed primer-template DNA and 3 to 6 units of SequenaseTM (United States Biochemical Corp.) in a single tube as 30 μ l, with 47 mM Tris-HCl, pH 7.5, 59 mM NaCl, 23 mM MgCl₂, 2 mM MnCl₂, and 8 mM dithiothreitol; C⁷-deaza-dGTP, C⁷-deaza-dATP, dCTP and dTTP were each present at 7.5 μ M; the 3' dideoxy terminators were included as 0.075 μ M ddGTP-505, 0.15 μ M ddATP-512, 0.15 μ M ddCTP-519 and 0.30 μ M ddTTP-526. All nucleotides used were components of NEN-DuPont GenesisTM system DNA sequencing kits. After 10 min at 37°C, 30 μ l of 5 M ammonium acetate, pH 7 and 150 μ l of cold ethanol were added to precipitate the DNA. After washing the pellet once with cold 70% ethanol, the dried pellet was dissolved in 6 μ l 95% formamide, 5 mM Na₂EDTA, pH 8, then heated 5 min at 95°C. One half of the sample was then loaded into a single well of the gel for electrophoresis.

Fluorescent labeling of DNA sequencing ladders with 5'-labeled primer

Fluorophore-labeled oligomers were developed by Du Pont for lane tracking in the operation of the GenesisTM 2000 automated DNA analysis system. The SF-505 succinyl fluoroscein fluorophore (14) was attached via a phosphoramidite (15) to the 5' end of an oligonucleotide representing all but the last base of the universal sequencing primer. This labeled oligomer was provided by George Tice (Du Pont, Glasgow, DE). The same annealing conditions were employed as described above, using 3 μ l of the 5' labeled primer (diluted 1:10000, estimated 15 ng). For DNA synthesis, the conditions were also the same as described above, except the 3-labeled dideoxynucleotide terminators were replaced by the unlabeled dideoxynucleotides ddGTP, ddATP, ddCTP and ddTTP (Pharmacia), 0.075 μ M each. Identification of the oligonucleotides in these uniformly labeled ladders was straightforward, by comparisons with four parallel ladders in which samples were prepared by inclusion of

only one of the four terminators. The fluorescence ratios of these 5'-labeled oligomer ladders were uniformly representative of the SF-505 succinyl fluoroscein fluorophore (14).

DNA sequencing gels

Electrophoresis was performed through 7% polyacrylamide gels (19:1 monomer to bis; 0.4 mm $\times$ 370 mm $\times$ 125 mm), containing 8 M urea buffered with 0.100 M Tris-borate, pH 8.8, 1 mM Na₂EDTA. Gels were exhaustively prerun (2 to 3 hours at 1000 volts) to constant resistance of 100,000 ohm, prior to loading. The gel electrophoresis of DNA samples was performed at constant voltage, in the range of 500 to 1500 V, depending on the experiment. Continuous temperature measurements were performed with a 4 mm diameter, flat-faced remote thermistor probe placed on the surface of the outer gel plate, approximately 5 cm above the detector window. The thermistor was previously calibrated at 4°C and 100°C using distilled water/ice and boiling water. Typical gels run at 750 V maintained a plate surface temperature of 32°C.

Data acquisition and analysis

GenesisTM 2000 instrument control, data acquisition and analysis were performed using an Apple Macintosh II computer. The

Figure 1. (A) Electrophoresis of an M13 mp18 DNA sequencing ladder at 750 volts. The time required for oligomers to traverse the gel from well to detector is a uniform and monotonic function of oligomer length. (B) Separations of successive oligomers increase gradually from about 90 seconds to 150 seconds over the range of 30 to 400 nucleotides, subject to large variance throughout this range. A third-order polynomial fit of the separation data, with coefficients determined by method of least squares, is shown as the line through the scatter of data points. This function is used to evaluate the average temporal separation of oligomers as a function of oligomer length.

entire 7 cm detector window of the GenesisTM 2000 was repetitively scanned as 90 horizontal steps, at intervals ranging from 4 to 10 seconds, over 6 to 48 hours of electrophoresis. The time per scan was selected to provide data separation of successive oligomers by 10 to 30 scan intervals, peak to peak, throughout the run. The elapsed time (from the computer's internal clock) and the digitized fluorescence output from the two photometers (PMT-A and PMT-B) of the GenesisTM 2000 were recorded for each scan and subsequently analyzed in spreadsheet format with graphics (Microsoft Excel v. 2.2, CricketGraph v. 1.2).

Identification of terminal nucleotides of oligomers in the sequencing ladders is based on the ratio of fluorescence (as $\ln(\text{PMT-B}/\text{PMT-A})$, reference 14) associated with each of the base-specific terminators. The times required for electrophoresis from well to detector were evaluated by fluorescence intensity weighting of elapsed times corresponding to the four scans of maximum signals for each oligomer.

Primer for sequencing mutations in the lac Z gene amino terminal coding region

We have used the GenesisTM 2000 DNA analysis system to characterize mutations in the *lac Z* gene, either spontaneous mutations or those induced by an activated ethylene dibromide mutagen. A specific oligonucleotide primer was synthesized for these experiments (M13 mp18 nucleotides 5'-6423 to 3'-6406, as reverse complement), to sequence through the amino-terminal coding region of the gene. This primer anneals to complementary DNA sequence of M13 mp18 template, 100 bases upstream from the conventional universal primer. ('Upstream' and 'downstream' references are given with respect to the 5' to 3' polarity of the oligomer product DNA strands.)

RESULTS

Temporal view of oligonucleotide ladder

The Du Pont GenesisTM 2000 automated DNA sequencer permits identification of 3' dideoxynucleotide terminators on each successive oligonucleotide, based on spectral properties of four base-specific fluorophores (14). In the following experiments we use these fluorescence measurements, together with the known DNA sequence of M13 mp18 (16), to establish the length of oligomers in the DNA sequencing ladders. We program the instrument to simultaneously record the elapsed time of electrophoresis (seconds) with each laser-photometer scan of the detector window. An M13 mp18 DNA sequencing ladder

Table 1. Relative Separation of Successive Oligonucleotides

3' terminal nucleotide	Fluorophore on 3' dideoxynucleotide	Fluorophore on 5' primer
—ddN	1.000 ± 0.192 (347)	1.000 ± 0.194 (316)
—ddT	1.177 ± 0.108 (85)	1.166 ± 0.124 (79)
—ddG	1.049 ± 0.150 (95)	1.031 ± 0.173 (86)
—ddA	0.959 ± 0.126 (77)	0.966 ± 0.143 (70)
—ddC	0.815 ± 0.163 (90)	0.838 ± 0.170 (81)

Relative separations of oligonucleotides. The relative separations of oligonucleotides from an M13 mp18 DNA sequencing ladder are shown as average ± standard deviation (number of events). Parallel DNA sequencing ladders were labeled with fluorescent 3' terminators or 5' primer and these provide statistically indistinguishable results. The relatively slowest species have 3' -ddT, the fastest have 3' -ddC.

undergoing electrophoresis at 750 V is analyzed in Figure 1A, showing the uniform, nearly linear relation between the time (seconds) for electrophoresis from well to detector and the length of the oligonucleotides (bases). Linearity of this function corresponds to an inverse relationship between mobility of the oligomers and their length.

In this experiment, the oligonucleotides are separated by 1 to 4 minutes, peak to peak, over the range of almost 400 nucleotides, Figure 1B. The scatter of individual points indicates large variance of the temporal separation of successive oligomers throughout this range. We use a third order polynomial function, determined by least squares curve-fitting, to describe the average temporal separation as a function of oligomer length. The relative separation of any adjacent pair of oligomers, (N)-mer and (N-1)-mer, can then be expressed as the ratio of actual time between their respective maximum signals and the average time difference expected for an oligomer of length N. On transformation of the data in Figure 1B in this manner, the entire ladder has an average relative separation of 1.000 ± 0.192 (standard deviation), Table 1. A wide interval, from 0.4 to 1.6 times the average relative separation, is required to include 99% of the oligomers.

Figure 2. The separations of successive oligonucleotides, taken from Figure 1B, were divided by the average separation function of oligomer length to evaluate relative separations (dimensionless). (A) The portion of the data set is shown which corresponds to those oligomers with 3'-terminal dideoxypyrimidines. There is nearly total partition of the 3' -ddT-terminated oligomers (black circles, slower than average) and those oligomers which end with 3' -ddC (open circles, faster than average). (B) The relative separations are illustrated for those oligomers which end with 3' dideoxypurines. These relative separations of these species differ less from average than the pyrimidine-terminated oligomers. The 3' -ddG oligomers (black circles) as a group are faster than average, the 3' -ddA oligomers (open circles) are slower than average.

The data shown in Figure 1 were obtained using a sample prepared with fluorophore-labeled dideoxynucleotides as 3' chain terminators (14). Statistically identical results are obtained with DNA samples prepared using a 5' oligonucleotide primer, labeled with dG-505 fluorophore, together with unlabeled dideoxynucleotide terminators. These 5'-labeled oligomers have an average relative separation of 1.000 ± 0.194 , in parallel tracks of the same gel, Table 1.

Other investigators have remarked on the uniform and monotonic decrease of mobility with oligomer length, and they have suggested this may facilitate localization of low yield oligomers in the ladder (14, 17). The large variance in separations from oligomer to oligomer in DNA sequencing ladders, which we report and characterize here, may compromise the utility of such a search algorithm.

The 3' terminal dideoxynucleotides and relative separation of oligomers

Each oligomer of length N differs from the corresponding $(N-1)$ -mer by an additional nucleotide at the 3' terminus of the chain. We thus considered that the 3' terminal dideoxynucleotides of each oligomer may be specific determinants of the variance of relative separation. Figure 2A shows relative separations of those oligonucleotide chains terminating with dideoxypyrimidines. A nearly total partition of these species is apparent. The chains with 3' ddT are consistently slower than average over the range of 400 nucleotides length. Those terminating with 3' ddC are consistently faster than average. The relative temporal separations of dideoxypurine-terminated oligomers, Figure 2B, are closer to

Figure 3. Relative separations of oligonucleotide are correlated with the 3' terminal dinucleotide sequences. The relative separations of the M13 mp18 oligomers were sorted according to the 16 possible 3' terminal dinucleotide combinations. The groups are arranged from slower to faster, in the order 3' -ddT, -ddG, -ddA, -ddC. The black squares indicate the average relative separation for each group; the error bars correspond to the standard deviation of values for each group. The groups of faster and slower oligomers are symmetrically distributed about the average relative temporal separation. Statistically indistinguishable results were obtained from ladders which were generated with unlabeled dideoxynucleotides and a fluorescent dG-505 labeled 5' primer.

the average than are those with 3' dideoxypyrimidines. Statistically, chains ending with 3' ddG are slower than average and those with 3' ddA are faster than average. The average relative separations of these groups of oligomers, based on the

Figure 4. Data streams from DNA sequences which differ by a single base change. Three selected examples are illustrated here: (A) a C $\rightarrow$ T transition at nucleotide 6322, (B) a purine transition mutation at nucleotide 6366, and (C) a C $\rightarrow$ A transition at nucleotide 6350. For each comparison of wild type M13 (open circles) and mutant (black circles) DNA, sequencing ladders were recorded during electrophoresis at 750 volts in parallel lanes of the same gel. (The different relations of time axes to oligomer lengths, evident in the three panels of data, reflect different times of initiation of data recording after loading of the samples on the different gels.) For each comparison, the wild type and mutant DNA data streams (PMT-A channel only) were aligned by superposition of the ladders through the oligomers preceding the base difference. This is accomplished with a single shift of one data stream by no more than one or two scan intervals. Such alignment is maintained over thousands of scans of identical sequences. The base changes distinguish the parallel ladders of wild type and mutant oligonucleotides which, up to the points of the lesions, are physically identical and indistinguishable in terms of their relative temporal separations of successive oligomers. Beyond the locus of base differences, the longer oligomers inherit altered mobilities.

3' dideoxynucleotide, are presented in Table 1. There is a statistically significant gradient of relative separations, increasing in the following order:

(faster) ---ddC < ---ddA < ---ddG < ---ddT (slower).

This relation between the identity of the 3' dideoxynucleotide and mobility of the oligomer applies to DNA samples in which the reporting fluorophore is attached to the individual 3' dideoxynucleotide terminators (14) or to the 5' end of the DNA sequencing primer. The difference between the average relative separations of the most similar groups, those with 3' dideoxypurines, corresponds to 3 to 4 standard errors of mean (probability $p < 0.001$ that the differences are due to random variation alone). The most distinct oligomer groups, those with 3' dideoxypyrimidines, have means which differ by 14 to 17 times the standard error. The probability that this is due to chance alone is infinitesimal ($p < 10^{-40}$).

Two different groupings of the data were evaluated as negative controls. The variation of relative separation was analyzed for four groups with respect to the identity of nucleotide (N+1), the next nucleotide of the DNA sequence, which is not physically part of the labeled oligomer of length (N). The oligomers were also analyzed as four groups according to length and without regard to their 3' terminal nucleotides. None of these control groups differs significantly from the total group of oligomers in terms of average relative separations, 1.00 ± 0.19 . The variance associated with average relative separation in each of these

negative control groups was consistently greater than for any of the groups based on the identity of the 3' dideoxynucleotide.

The 3' proximal deoxynucleotides and relative separations of oligomers

We next evaluated the possible role of the 3' proximal deoxynucleotide (N-1) as a co-determinant of the variance of relative temporal separation of oligomers. An array of

Figure 5. Resolution of a compression artifact by distal single base change. Portions of aligned data streams, both photometer channels, are shown for (A) wild type M13 mp18 and a (B) mutant which has a C → T change at nucleotide 6381. The oligomer corresponding to the base change is labeled C or T in the respective data streams, appearing after 12100 seconds of electrophoresis at 750 volts. Eight bases downstream from the locus of the mutation (12850 seconds) there is a compression artifact characteristic of the wild type sequence; the sequence order, 5'-CTGGCG-3', is not reflected in the order of appearance of the successive oligonucleotides at the detector window. The same sequence is not compressed in the oligonucleotide ladder of the mutant. After seven or eight additional nucleotides are added to the ladder, beyond the compressed 3' C and 3' G species, the wild type and mutant data streams again approach coincidence. From this point the data streams are offset by a constant 29 seconds. This offset is attributed to the C-T difference inherited by all those oligomers which extend beyond the locus of the mutation (and the compression).

Figure 6. (A) Anomalous oligomer separations due to compression artifact are resolved by raising the temperature of the gel during electrophoresis. Electrophoresis was performed at 500, 750, 1000, 1250, and 1500 volts (constant), with running temperatures measured in the range of 25°C to 60°C. The separations of the compressed oligonucleotide (i) 5' ---TTACGCCAGCTGGCG 3' (closed circles) and uncompressed oligonucleotide (ii) 5' ---TTACGCCAGCTGG 3' (open circles) were evaluated relative to the longer, uncompressed oligonucleotide (iii), using a shorter oligomer as a reference (see text). The compressed structures appear to 'melt' over the range from 27°C to 35°C. At the lowest temperatures the oligomers (i) and (ii) effectively comigrate in spite of their two nucleotide length difference. As temperature is raised, species (i) and (ii) resolve, and for identification we have assumed that they appear in temporal order with respect to their lengths. (B) The average relative separations for oligomers ending with each of the four bases were evaluated from parallel data streams of the same gels used in the above analysis of the compression. The averages at each temperature were determined for the oligomers from 19 to 160 bases long, using the universal primer on M13 mp18 DNA template. The order of the 3' nucleotide-specific differences was maintained throughout the range of 27°C, 500 V to 55°C, 1500 V, although the magnitude of differences gradually decreases. We also observed about a 3-fold decrease in the standard deviations of distributions of relative separations (for each group of oligomers based on 3' dideoxynucleotide) as temperature increases in this interval.

distributions, based on the sixteen possible 3' terminal dinucleotide sequences, Figure 3, leads to the generalization that relative temporal separations increase in the order

(faster) ---dT-ddN < ---dR-ddN < ---dC-ddN (slower).

It is interesting to note that the pyrimidines are associated again with the most extreme differences in mobility, but they contribute in opposite order when located in the 3' proximal and 3' terminal positions. The same order and magnitudes of results were obtained with the reporter fluorophores placed at the 3' end or the 5' end of the oligomers (results not shown).

Relative separations of oligomers with single base changes

Figure 4 shows parallel data streams from the DNA sequencing ladders of wild type M13 mp18 and three mutants which differ by single base changes. Data from only one of the two photometers is shown for each sample. Up to the altered nucleotide at position (N), the two data streams of each example are coincident. The data streams become asynchronous at the positions corresponding to the base changes of the mutants. Thus oligomers which are identical through their first (N-1) deoxynucleotides, have different mobilities due to their different 3' dideoxynucleotide terminators at position (N). The data streams of these and other mutants we have screened are consistent with the conclusions derived above by reference to average separations of oligomers.

A second feature of the parallel data streams for wild type and mutant DNA is that they continue asynchronously beyond the position of base change. This suggests that conformational differences due to the interactions involving the 3' terminal and proximal nucleotides are inherited by longer oligomers, at least in part, as local structure at that position.

Resolution of a compression artifact by distal single base change

In the course of our survey of *lac Z* gene mutations, we observed one mutant for which oligomer mobility changes were particularly dramatic, Figure 5. The compared sequences are aligned below:

```
wt:      -T-T-A-C-G-C-C-A-G-C-T-G-G-C-G-A-A-A-G-G-G-G-
          *
mutant:  -T-T-A-C-G-T-C-A-G-C-T-G-G-C-G-A-A-A-G-G-G-G-
```

Oligomer separations differ at the point of mutation, due to the difference of 3' ddC and 3' ddT at the ends of otherwise identical oligomers. Wild type oligomers which extend 8 and 9 bases beyond the point of the mutation are subject to a severe compression artifact. The oligomers of wild type DNA ending with ddC (base 6372) and ddG (base 6373), underlined above, migrate faster than oligomers which are one or two bases shorter. The mutant sequence shows no evidence of compression at all. It is interesting to note that as the length of the oligomer increases to include the next three dA's and five dG's following the compression, the data streams of the mutant and wild type sequences reapproach synchrony, offset only by the mobility difference due to the upstream dC → dT change.

Compressions and terminal nucleotide interactions at elevated temperatures

Conventional practice prescribes running sequencing gels at high temperatures in order to abrogate the base pairing, hydrogen

Figure 7. Proposed hairpin loop and extended conformations of a DNA sequence subject to compression artifact (see Figure 5 and text). The asterisk indicates the dC residue of the sequence which is changed in a mutation whose DNA no longer exhibits the downstream compression artifact. Oligomers of the compression-prone sequence and those of the mutant comigrate up to the dG residue which is underlined in the lower structure. The compression involves the next two oligomers, indicating at least three base pairs are required to stabilize, if only transiently, such a small hairpin structure.

bonding interactions associated with compression artifacts. These conditions can be achieved by electrophoresis at elevated fields; Joule heating leads to higher running temperatures. We used this approach to transform the compression artifact, described in the previous section, to uniformity of the mobility-length relation. M13 mp18 DNA samples were analyzed by electrophoresis through 7% polyacrylamide gels, at fields which led to temperatures in the range of 25°C to 60°C, Figure 6.

The average separations of oligomers in a DNA sequencing ladder vary with the strength of the electrical field. In this experiment the average separations of successive oligomers decreased from about 200 seconds to 45 seconds as voltage of electrophoresis increased from 500 to 1500 volts. We measured separations from the shorter reference oligomer to oligomers (i) and (ii) below, relative to the separation of the longer species (iii).

	-T-T-A-C-G-C-C-A-G-C-T	reference oligomer
(i)	-T-T-A-C-G-C-C-A-G-C-T-G-G	not compressed
(ii)	-T-T-A-C-G-C-C-A-G-C-T-G-G-C-G	compressed
(iii)	-T-T-A-C-G-C-C-A-G-C-T-G-G-C-G-A	not compressed

The compression-prone oligomer (ii) migrates more slowly as temperature increases from 27°C to 35°C, Figure 6A, approaching the average position in the ladder expected for uncompressed oligomers of its length. The relative separation of the oligomer which is not subject to compression (i) does not vary significantly with temperature in the range of this experiment.

The same DNA samples were analyzed with regard to the systematic mobility differences which we attributed to specific 3' terminal dinucleotide sequences. As shown in Figure 6B, there is only modest attenuation of the differences of oligomers ending with different dideoxynucleotides. As temperature increases, the variance associated with the base-specific average relative separations decreases. This appears to be due to the decreased involvement of particular oligomers in compressions at the higher temperatures.

Table 2. Relative Separation of Successive Oligonucleotides

a) Fastest Sequences (-dN-dT-ddC, 0.735 ± 0.072)	mean ± standard deviation (events)	
	Fluorophore on 3' dideoxynucleotide	Fluorophore on 5' primer
-dG-dT-ddC	0.698 ± 0.062 (4)	0.769 ± 0.251 (4)
-dA-dT-ddC	0.751 ± 0.061 (4)	0.753 ± 0.159 (5)
-dC-dT-ddC	0.727 ± 0.083 (5)	0.702 ± 0.045 (5)
-dT-dT-ddC	0.753 ± 0.081 (7)	0.718 ± 0.085 (5)
b) Slowest Sequences (-dN-dC-ddT, 1.285 ± 0.080)		
-dG-dC-ddT	1.254 ± 0.093 (9)	1.228 ± 0.058 (7)
-dA-dC-ddT	1.240 ± 0.051 (3)	1.258 ± 0.039 (3)
-dC-dC-ddT	1.326 ± 0.044 (6)	1.316 ± 0.073 (5)
-dT-dC-ddT	1.371 ± 0.019 (2)	1.349 ± 0.091 (2)

Effects of the anteproimal (N-2) nucleotide on relative separations of oligomers which have 3' heteropyrimidine dinucleotide sequences. Separations of the fastest species are relatively unaffected by the upstream (N-2) position nucleotide. The slowest species, with the sequence 3' -dNdCddT, are significantly partitioned for N = R and N = Y. Similar results are obtained with the fluorophore labels at either end of the oligomers.

DISCUSSION

Oligonucleotide reptation and conformation at the 3' terminus

The large majority of oligonucleotides in a DNA sequencing ladder appear in regular temporal order at the detector window of an automated DNA sequencer. The data in Figure 1A follow a nearly linear progression in time as a function of oligomer length. In static view, autoradiographic sequencing experiments these oligomers appear as an ordered array of decreasing displacements with increasing oligomer length. The results obtained in either format translate into an inverse relationship of oligomer mobility and length. The prevalent model for electrophoresis of polynucleotides in gels, based upon end-first migration or reptation through the gel matrix (18-23), predicts such an inverse relationship of oligomer mobility and length.

Compression artifacts and mobility differences due to terminal nucleotide interactions are documented in this report. These are sequence-specific anomalies of oligonucleotide electrophoresis in denaturing polyacrylamide gels. These phenomena may be reconciled as variations in local conformation which affect the average mobility of specific oligomers during electrophoretic transport according to the reptation model.

More elaborate models for electrophoresis of polynucleotides have been proposed, which incorporate higher order effects. These consider transient entrapment of long molecules at high fields, independence of length and mobility for very long molecules, and reorientation kinetics during variation of field strength and direction (24-32). The relatively porous gels used for DNA sequence analysis and the flexibility of short, single-stranded oligonucleotides may limit the usefulness of such modeling to polynucleotides which are too long for current DNA sequencing methods.

The 3' terminal bases as conformational determinants of oligonucleotides

The variation of DNA oligomer separations, shown in Figure 1A, reflects differences of oligomer mobility in sequencing gels on the order of 0.1 to 3%. The identities of the 3' terminal dideoxynucleotides and 3' proximal deoxynucleotides are clearly

associated with differences in the mobilities of different oligonucleotides in a DNA sequencing ladder, Figure 2. The order of effects with respect to base identity is opposite at the proximal and terminal positions, Figure 3. Pyrimidines at either position impose larger effects than do the purines. The effects tend to follow a polarity of keto to amino nucleotides, T-G-A-C. Can simple mass or charge differences of the oligomers, with respect to their 3' terminal nucleotides, account for these results?

Average dideoxynucleotides contribute about 315 to oligomer molecular weight, and 3' ddT residues contribute 13 mass units more than corresponding 3' ddC residues. This 4% difference of the terminal nucleotide mass corresponds to less than 0.04% difference in oligomers of length greater than 100 monomers. The greatest mobility differences observed are for oligomers terminating with 3' -dCddT and 3' -dTddC (otherwise identical), yet these have negligible mass differences. Thus mobility differences **do not** reflect small differences in oligomer mass related to the terminal nucleotide (sequence).

The same polarity order of 3' nucleotides is observed for electrophoretic mobility differences in the 'wandering-spot' method of nucleotide sequence determination (33). High voltage paper electrophoresis at acidic pH, leads to 3' nucleotide-specific mobility differences. These are attributable to average electrical charge of the oligomers, determined by the different pK_a values of the 3' terminal nucleotides. Denaturing polyacrylamide gels used for DNA sequencing, however, are buffered between pH 8 and 9. Such conditions should neither protonate nor titrate protons from the four bases at either internal or terminal positions. Thus mobility differences **do not** reflect small differences in oligomer charge related to the terminal nucleotide (sequence).

Since neither mass nor charge differences account for the 3' nucleotide-specific mobility differences, then altered conformation due to nucleotide-specific interactions becomes an appealing model. We suggest that stronger interactions between 3' terminal and proximal nucleotides leads to a more compact conformation and thus more rapid migration through the gel.

Crystallographic and spectroscopic studies of oligonucleotide homopolymers reveal some structural determinants which affect conformation in single-stranded oligomers (20). These include base stacking and various opportunities for hydrogen bonding between adjacent nucleotides, bound water molecules, and the phosphodiester backbone. Integrated constraints of bond lengths and angles are further limited by the restriction of pyrimidines to *anti*- configuration about their C1'-N1 bonds. Homopyrimidine oligomers have particularly distinctive structures. Poly-cytidine chains appear capable of base stacking, with successive bases inside a tight helix. Evidence also has been presented for a base-outward configuration in solution, with interaction of cytidine's 4-amino group, through hydrogen bonding, with the position-2 carbonyl oxygen of the upstream proximal pyrimidine (35). A terminal dideoxythymidine would not have such an opportunity for the hydrogen bonding with its 3' proximal pyrimidine neighbor. Poly-thymidylate chain conformation has been suggested from crystallographic analysis of thymidylate dimer structure. In this case a wide helix is predicted with bases unstacked, facing outward (36, 37).

We propose that the amino groups present on the aromatic rings of dideoxycytidine and dideoxyadenosine favor hydrogen bonding with acceptor sites of the immediately upstream deoxynucleotide. The ring carbonyl of thymidine is in particularly suitable orientation for such an interaction at the 3' proximal nucleotide

position. When such interactions occur, conformation is more compact, and the oligomer is relatively faster than it would be with terminal dinucleotide sequences that do not support this type of interaction.

Purines, in contrast with the pyrimidines, can present in both *syn* and *anti* configurations with respect to their C1'-N1 bonds. This may present opportunities for extended hydrogen bonding interactions with adjacent nucleotides, involving their C7 ring nitrogens. In most of the experiments presented here we have utilized the C7-deazapurines to obviate this potential complication. The exception is those experiments in which the 5'-primer bears the reporting fluorophore. These ladders were generated using ddGTP and ddATP as purine-specific chain terminators. We have not observed significant mobility differences between the purine-terminated oligomers in these ladders and those generated with the fluorophores attached to C7 of the deazapurine dideoxynucleotides. It would be premature to conclude, however, that there are no aromatic stacking or hydrogen bonding interactions of the fluorophore probe with the nucleotides in the proximity of the ends of 5''- or 3' -labeled oligomers.

Our model of neighboring nucleotide interactions at the 3' terminus leads to a prediction about the role of nucleotides in the 3' anteproximal position (N-2). In the case of the 3' terminal sequence —dT-ddC, strong interaction between the terminal pyrimidines is associated with relatively rapid migration. The interaction of nucleotides at positions (N-1) and (N), together with the tendency of dT not to interact with its upstream neighbor, suggest that the (N-2) nucleotide **would not** strongly influence the mobility difference of these (N)-mers and their corresponding (N-1)-mers.

In contrast, the limited capacity for interaction between the terminal pyrimidines in the slow —dC-ddT oligomers, together with the tendency of dC to interact with its upstream neighbor, suggest that this (N-2) nucleotide **does** contribute to the mobility differences of these (N)-mers and their corresponding (N-1)-mers.

Table 2 describes the distributions of relative separations for oligomers which terminate with a 3' heteropyrimidine dimer sequence. The fastest species, ending as —dT-ddC, shows no differences with regard to the base at the (N-2) position (Table 2a). The slowest species, ending as —dC-ddT, are presented in Table 2b. These oligomer groups do show a significant ($p < .001$) dependence of their relative temporal separation on presence of a purine or pyrimidine in the (N-2) position

(faster) ---dR-dC-ddT < ---dY-dC-ddT (slower).

The same results are obtained when the fluorophore label is part of the base-specific 3' terminator dideoxynucleotides or on the 5' primer oligonucleotide.

Significant differences in the separation of successive oligomers are strongly correlated with the 3' terminal sequence of two to three nucleotides. A limited number of adjacent nucleotide-specific hydrogen bonding interactions, perhaps only one or two, may account for conformation differences and resulting differences in mobility during electrophoresis. The free energy change associated with these proposed interactions is probably very small, on the order -1 kcal/mole or less. Base-stacking interactions of adjacent nucleotides may also play a role, but the energetics are not likely to be any more favorable. The presence of 8 M urea as a competitor for hydrogen bonding accounts for the denaturing properties of the electrophoresis buffer. We reconcile the proposed weak interactions of the 3' terminal

nucleotides in this environment, because the interaction is intramolecular and occurs over very limited distances between adjacent nucleotides.

Denaturation of the compression-prone 3' terminal DNA sequence

The compression artifact presented in Figure 5 is associated with a 3' terminal DNA sequence that suggests formation of short hairpin structures of three to four base pairs, Figure 7. The mobilities of the compressed oligomers increase as temperature is raised from 28°C to 37°C, as shown in Figure 6A, consistent with a melting, helix-coil transition. At the higher temperatures, the compression-prone oligomers assume the expected positions in the ordered sequencing ladder.

Co-migrations of the shorter oligomers of the wild type and mutant sequences, up to the ddG 6374 underlined in Figure 7, suggest that at least three potential base pairs are required for a hairpin structure to appear as a significant compression effect (below 32°C). The dC → dT difference of the mutant DNA sequence (asterisk) lies in the middle of the base pairing region to which we attribute the compression. The mismatch of a base-pair in the center of this short and patently labile structure supports the absence of the compression artifact in the mutant's DNA sequencing ladder.

This particular example of compression seems to represent a lower limit benchmark for the type of interactions that can lead to compression artifacts in conventional sequencing ladders. Under the conditions of electrophoresis we have examined, above 25°C, no compression-prone oligomers have been observed with less than three conventional base pairs about a short hairpin loop. We also note that as the three successive dA and five dG residues are added beyond the sequences with base-pairing potential, the ladders of the wild type and mutant DNA reapproach coincidence. Extension of the single strand beyond the stem structure appears to mechanically distend the chain, further destabilizing the labile stem. Free energies for the formation of the proposed compressed structures in neutral buffer would be very low. A crude estimate can be made for the 3 to 4 G:C pairs and 4 base loop of the hairpins: $-\Delta G_{25^\circ\text{C}} < 6$ to 9 kcal (38). It is difficult to appreciate the significance of such a labile structure at or above room temperature with 8 M urea in the gel buffer.

The compressed oligonucleotides probably reflect an equilibrium of base-paired and extended conformations during electrophoresis, their observed mobility an average based on the fraction of time spent in each configuration. This equilibrium is determined by the relative rates of formation and extension of the short hairpin. The hairpin is intrinsically unstable, with high first order rate of extension. However, formation of the hairpin is favored as an intramolecular nucleation or annealing event (also first order) that is driven by the proximity (high local concentration) of the interacting sequences.

This line of reasoning also applies to the local interactions of neighboring nucleotides at the 3' terminus described earlier. Those interactions may be an order of magnitude weaker, in terms of numbers of hydrogen bonds or free energy of formation. But at the same time the distance over which alternative conformations may be represented is also an order of magnitude smaller. The magnitudes of separation differences based on the 3' terminal nucleotides gradually decrease as temperature is raised (Figure 6B). Over the restricted range of only two to three nucleotides, the conformational transition, which may accompany thermal disruption of neighboring nucleotide interactions, is likely to have only limited effects upon oligomer mobility.

Applications in screening DNA sequences for mutations or mutant alleles

An attractive feature of an automated DNA sequencer is its capacity to screen large numbers of samples for nucleotide changes resulting from mutation. The data streams from samples of wild type and mutant DNA should differ by the intensities of signals from the radioactive or fluorescent label(s) at the position of base change(s). Our results suggest that single base changes will also alter the time distribution of the data stream at and beyond the locus of mutation.

The polymerase chain reaction is often used to amplify regions of diploid chromosomal DNA for direct sequence analysis. The yields of oligomers corresponding to the distinguishing nucleotides of different alleles may differ greatly, because of template sequence-dependent, competitive incorporation of the deoxynucleotides and their dideoxynucleotide analogues. Strong signals from oligomers of one allele may mask the signals of oligomers from the other. In such circumstances, the less represented allele may be more easily detected by changes in the time course of the heterozygous sample's data stream, compared to that of homozygous controls.

SUMMARY

Local, nucleotide-specific interactions lead to differences in the mobilities of homologous oligonucleotides in DNA sequencing ladders. A systematic variation of conformation and mobility is apparent, determined by the sequence of two to three adjacent, 3' terminal nucleotides. Less common are compression artifacts, which appear to result from conventional base-pairing interactions, with hairpin formation, over a range of 10 to 20 bases at the 3' ends of oligomers. Further analysis of the electrophoretic transport properties determined by the sequences of oligonucleotides should be useful to complement and corroborate data from conventional approaches to the determination of DNA sequences. Beyond the application of its original design, the automated DNA sequencing instrument can be an important analytical tool for biophysical studies of polynucleotide structures and interactions.

ACKNOWLEDGMENTS

We are particularly grateful to Dr. Ed Chait, for his early appreciation of our interest in automated DNA sequencing, and for his efforts to initiate a program of collaborative research between Du Pont Co. and Vanderbilt University. George Tice, Jay Bass and Dave Porter of Du Pont Co. were of great assistance in consultations regarding the Genesis™ 2000 instrument system and the specialized reagents used in our experiments. Further important contributions to this project have been provided by Dr. Al Strauss, Department of Mechanical Engineering, Vanderbilt University. We appreciate the thoughtful comments of Drs. Gisela Mosig (Department of Molecular Biology), Stephen Lloyd (Department of Biochemistry) and Bruce Jacobson (Oak Ridge National Laboratory, Division of Biology), which they shared after reading early drafts of this manuscript. This work has been supported, in part, by grants from the DHHS, National Cancer Institute, CA-34126 (C.Tibbetts, P.I) and CA 44353 (F. Guengerich, P. I., Department of Biochemistry, Center in Molecular Toxicology, Vanderbilt University).

REFERENCES

1. Sanger, F. and Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448.
2. Maxam, A. M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
3. Maniatis, T., Jeffrey, A. and van deSande, H. (1975) *Biochemistry* **14**, 3787–3794.
4. Kramer, F. R. and Mills, D. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5334–5338.
5. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
6. Sanger, F. and Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
7. Mills, D. R. and Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2232–2235.
8. Gough, J. A. and Murray, N. E. (1983) *J. Mol. Biol.* **166**, 1–19.
9. Mizusawa, S., Nishimura, S. and Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319–1324.
10. Barr, P. J., Thayer, R. M., Laybourn, P., Najarian, R. C., Seela, F. and Tolan, D. R. (1986) *BioTechniques* **4**, 428–432.
11. Jensen, M. A., Zagursky, R. J., Trainor, G. L., Cocuzza, A. J., Lee, A. and Chen, E. Y. (1990) *DNA Sequence* (accepted for publication).
12. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connel, C. R., Heiner, C., Kent, S. B. H. and Hood, L. E. (1986) *Nature (London)* **321**, 674–679.
13. Ansorge, W., Sproat, B. S., Stegemann, J. and Schwager, C. (1987) *Nucleic Acids Res.* **15**, 4593–4602.
14. Prober, J. M., Trainor, G. L., Dam, G. L., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. and Baumeister, K. (1987) *Science* **238**, 336–341.
15. Hobbs, F. W., European Patent Application No. 359225.
16. Yanisch-Perron, C., Vierstra, J. and Messing, J. (1985) *Gene* **33**, 103–119.
17. Beck, S. and Pohl, F. M. (1984) *EMBO J.* **3**, 2905–2909.
18. Lerman, L. S. and Frisch, H. L. (1982) *Biopolymers* **21**, 995–997.
19. Lumpkin, O. J. and Zimm, B. H. (1982) *Biopolymers* **21**, 2315–2316.
20. Bean, C. P. and Hervet, H. (1983) *Biophys. J.* **41**, A289.
21. Bean, C. P. and Hervet, H. (1983) *Bull. Am. Phys. Soc.* **28**, 444.
22. Edmondson, S. P. and Gray, D. M. (1984) *Biopolymers* **23**, 2725–2742.
23. Hervet, H. and Bean, C. P. (1987) *Biopolymers* **26**, 727–742.
24. Birren, B. W., Simon, M. I. and Lai, E. (1990) *Nucleic Acids Res.* **18**, 1481–1487.
25. Deutsch, J. M. (1988) *Science* **240**, 922–924.
26. deGennes, P. G. (1971) *J. Chem. Phys.* **55**, 572–578.
27. Slater, G. W. and Noolandi, J. (1985) *Biopolymers* **24**, 2181–2184.
28. Noolandi, J., Rousseau, J., Slater, G. W., Turmel, C. and Lalande, M. (1987) *Phys. Rev. Letters* **58**, 2428–2431.
29. Holzwarth, G., McKee, C. B., Steiger, S. and Crater, G. (1987) *Nucleic Acids Res.* **15**, 10031–10044.
30. Rubenstein, M. (1987) *Phys. Rev. Letters* **59**, 1946–1949.
31. Smith, S. B., Aldridge, P. K. and Callis, J. B. (1989) *Science* **243**, 203–206.
32. de la Cruz, M. O., Deutsch, J. M. and Edwards, S. F. (1986) *Phys. Review A* **33**, 2047–2055.
33. Tu, C.-P. D. and Wu, R. (1980) *Methods in Enzymology* **65**, 620–638.
34. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Chapter 13, 'Synthetic, Homopolymer Nucleic Acids Structures', Springer-Verlag, New York.
35. Broido, M. S. and Kearns, D. R. (1982) *J. Amer. Chem. Soc.* **104**, 5207–5216.
36. Camerman, N., Fawcett, J. K. and Camerman, A. (1976) *J. Mol. Biol.* **107**, 601–621.
37. Wood, D. J., Hruska, F. E. and Ogilvie, K. K. (1974) *Can. J. Chem.* **52**, 3353–3366.
38. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. and Gralla, J. (1973) *Nature New Biology* **246**, 40–41.