Journal of Chromatography A, 652 (1993) 23–30
Elsevier Science Publishers B.V., Amsterdam

CHROM. 25 138

Simultaneous determination of the migration coefficient of each base in heterogeneous oligo-DNA by gel filled capillary electrophoresis

Takashi Satow* and Tadao Akiyama

Beckman Instruments (Japan) Ltd., 6, Sanbancho, Chiyoda-ku, Tokyo 102 (Japan)

Akiko Machida

11-11, 5-Chome, Nijoh, Kitagou, Shiraishi-ku, Sapporo 003 (Japan)

Yushi Utagawa and Hidesaburo Kobayashi

Faculty of Science, Josai University, Keyakidai, Sakado-shi, Saitama-ken 350-02 (Japan)

ABSTRACT

The direct determination of migration coefficients was achieved by analysing the migration time of heterogeneous oligo-DNA with a gel filled capillary using the Gauss least-squares method for the observation functions, assuming that the migration time of oligo-DNA is dependent on its base composition and chain length. By using the coefficients obtained, the migration time of oligo-DNA of any known sequence that does not have a secondary structure can be estimated with an accuracy of less than 0.5-mer of cytidine. In addition, from the deviation of the actual migration time from the calculated migration time in certain specially designed base sequences, the existence of a secondary structure such as a hairpin structure was strongly suggested even in the presence of 7 *M* urea. From the investigation of the effects of secondary structure on migration time, it was concluded that this approach will give qualitative information on secondary structure, which may be applicable in work such as single strand conformational polymorphism (SSCP) analysis or antisense DNA analysis, in which secondary structure plays an important role in accelerating or decelerating migration times. The results of the analysis also predict the apparent chain length reversal from short to long together with a reduction in the actual chain length in DNA sequencing using capillary electrophoresis.

INTRODUCTION

Since capillary electorphoresis (CE) was first successfully achieved by Jorgenson and Lukacs [1], many different separation modes, such as micellar electrokinetic capillary chromatography (MECC) [2] have been developed not only for small compounds but also for biological macromolecules, as reviewed by Kuhr [3] and by Kuhr and Monnig [4]. One of the most useful separation modes is capillary gel electrophoresis (CGE) originally developed by Karger's group for protein [5] and DNA [6] separations. Especially for DNA separations, the performance of CGE has been continuously improved through improvements of the gel matrices and experimental conditions. Recently a few groups have achieved as high as $1 \cdot 10^7$–$3 \cdot 10^7$ theoretical plates [7–9] by using homogeneous oligo-DNA as the sample.

On the other hand, only a limited number of studies have been carried out using heterogeneous oligo-DNA samples to describe the migra-

* Corresponding author.

0021-9673/93/$06.00

tion behaviour of the sample components in the gel [10,11]. As the conclusion was that the oligo-DNA was not separated by chain length alone [10], and as the migration time slightly changed in each run, we have developed a novel means of determining the migration coefficient for each base in a single run by using several heterogeneous oligo-DNAs as standards followed by statistical analysis with the Gauss least-squares method for observation functions [12].

EXPERIMENTAL

CE instrument

For CE analysis, a Beckman (Fullerton, CA, USA) P/ACE System 2100 was used with Beckman Gold Workstation software on an IBM personal computer to operate the P/ACE instrument and to analyse the data. Some of the data were saved in the ASCII format on an IBM computer and transferred to an Apple Macintosh computer for further analysis with Microsoft Excel. For electrophoresis the polarity was reversed (cathode on the injection side and anode on the detector side).

The gel-filled capillary employed was eCAP gel U100P (Beckman) of 37 cm total length and 100 μm I.D. containing polyacrylamide gel in 7 *M* urea. Monitoring was effected with an on-column UV detector at 254 nm.

Oligo-DNA

Standard heterogeneous oligo-DNAs were a gift from Takara (Kyoto, Japan). Other oligo-DNAs including specially designed sequences for the analysis of the secondary structure effects were synthesized with a Milligen DNA synthesizer (Cyclone Plus DNA Synthesizer; Millipore, Bedford, MA, USA). The base sequences of the heterogenous oligo-DNAs are summarized in Table I and Fig. 3.

Procedure

In most instances, the sample concentration was adjusted to 0.01–0.005 absorbance units at 260 nm with water. For each injection we used 40–60 μl of sample solution in a modified polyethylene micro vial. Samples were injected electrokinetically at 5 kV for 2 S. Electrophoresis was performed at 300 V/cm, the capillary being kept at 30°C during the experiments by using the liquid cooling system of the P/ACE instrument.

Statistical analysis (Gauss least-squares method for observation functions)

Pilot experiments with heterogeneous oligo-DNA showed that the separation was not based on the chain length alone. We therefore assumed that the migration time is described as follows, assuming that the migration time of heterogeneous oligo-DNA is probably a function of the chain length and the base composition:

$$[A]a + [G]g + [T]t + [C]c + k = M \quad (1)$$

where [A], [G], [T] and [C] are the numbers of each base in the chain, a, g, t and c are the migration coefficients (in minutes per number of bases), k is a constant or the time required for oligo-DNA of zero length to move to the detector (in minutes) and M is the migration time in minutes obtained from an electropherogram.

It is possible to determine the parameters of the above function by using five heterogeneous oligo-DNAs of different base compositions by constructing simultaneous functions if there is no experimental error. However, in reality, all the experimentally obtained values include some degree of error. To minimize the effects of experimental errors, the statistical analysis of migration times was done basically following the Gauss least-squares method for observation functions [12], which is popular in experimental physics for calculating the coefficients of function(s).

We modified eqn. 1 into the following forms to remove the effects of observation errors and to obtain five different functions for the calculation of the parameters (coefficients):

$$[AA]a + [AG]g + [AT]t + [AC]c + [A]k = [AM]$$

$$[AG]a + [GG]g + [GT]t + [CG]c + [G]k = [GM]$$

$$[AT]a + [GT]g + [TT]t + [CT]c + [T]k = [MT]$$

$$[AC]a + [CG]g + [CT]t + [CC]c + [C]k = [CM]$$

$$[A]a + [G]g + [T]t + [C]c + [K]k = [M] \quad (2)$$

where K is the number of the constant (=1), $[AA] = \Sigma_i A_i A_i$, $[AT] = \Sigma_i A_i T_i$ and $[A] = \Sigma_i A_i$.

By solving the simultaneous eqns. 2, we can determine the migration coefficients (a, g, t and c) for A, G, T and C. The actual calculation was time consuming as we had to repeat a large number of simple calculations. We therefore developed a template on Microsoft Excel with a Macintosh computer to obtain the normal equations and to solve the simultaneous equations simply by inputting the base compositions and migration times of standard heterogeneous oligo-DNAs actually measured in a single experiment. This template also gives the estimated migration time of any known sequence. A limited number of templates for Microsoft Excel on Macintosh are available on request from the authors for research purposes only.

RESULTS AND DISCUSSION

Analysis of standard sample

CGE is a high-resolution technique and shows its resolving power particularly when using oligo-DNAs as the sample. For example, a few groups have achieved more than 10^7 theoretical plates/m [7–9]. In most work to improve the resolution of the gel-filled capillary, homogeneous oligo-DNAs were used as the sample, and showed a very linear relationship between chain length and migration time, which gives equal time differences between the peak of an n-mer and the $(n + 1)$-mer. Unlike homogeneous oligo-DNA, however, no linear calibration graph is obtained when using heterogeneous oligo-DNAs. A typical electorpherogram obtained with a heterogeneous oligo-DNA standard is shown in Fig. 1 and the base sequences are summarized in Table I. The peak assignment was done by varying the concentration of each oligo-DNA.

As can be seen in Fig. 1, the fact that the migration time differences between the 20–21-mers, 29–30-mers, 30–31-mers, 40–41-mers and 50–51-mers were different strongly suggested an effect of base composition on the migration time, as was reported by Guttman *et al.* [10]. Rapid heat treatment to eliminate the possibility of complex formation through hydrogen bonding had no significant effect on the separation.

Repetitive injection caused a slight change in

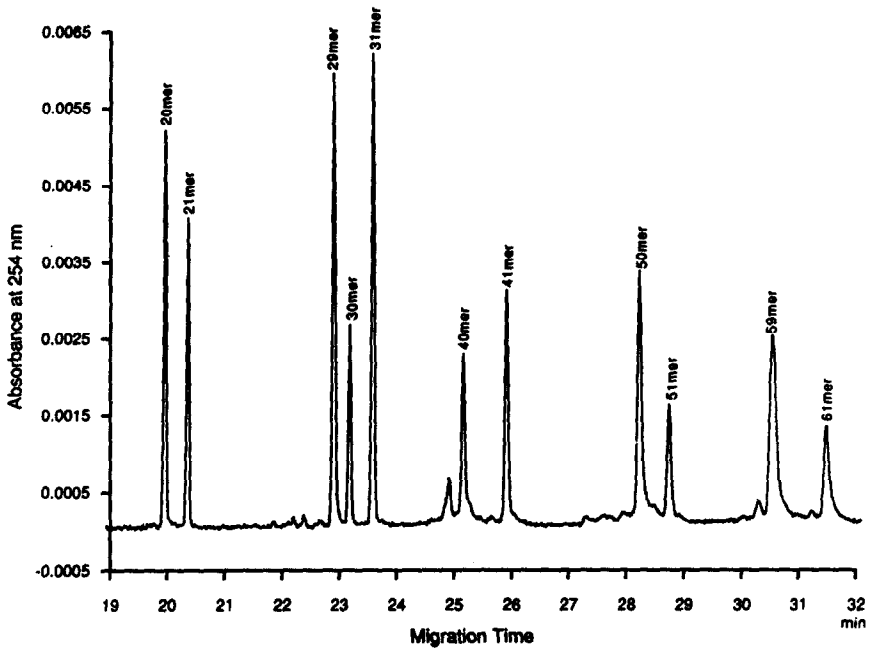

Fig. 1. Separation of heterogeneous oligo-DNA standards. The actual sequences of oligo-DNAs and CE separation conditions are summarized in Table I and under Experimental, respectively.

TABLE I

HETEROGENEOUS OLIGO-DNA STANDARD USED IN THE EXPERIMENTS

For experimental conditions, see Experimental.

Oligomer	Sequence
20-mer	5'-AAGAAAGCAGCAACTGCATT-3'
21-mer	5'-AGTACGTGCCAGTGCCAGCGT-3'
29-mer	5'-ATTACAGTTCATCGTGCACAGCTTTCTGA-3'
30-mer	5'-AGTTCATCGTGCTCAAATTTCTGATCATCG-3'
31-mer	5'-GATCCGTTTCAGGTCTGATTGAAAGCGGTAC-3'
40-mer	5'-AGAGCATCAGATCACCTGGGACCCCATCGATGGACGCGTT-3'
41-mer	5'-ATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTAC-3'
50-mer	5'-CCTCCGGAAGTGGACGTTGACGACGAGCCAGAAGAAGAATAAGGATCCGG-3'
51-mer	5'-TATTTATGGAGTTCAGACACTCAGCACAGCAGGGTCCTGAGCTTATATAAT-3'
59-mer	5'-CGGATGAGCTAAGGGATGAGGATGAGAGCAAGAGCAAAAGAAGGAGCAACAACTTCATC-3'
61-mer	5'-TCGAGATGAAGTTGTTGCTCCTTCTTTTGCTCTTGCTCTCATCCTCATCCCTTAGCTCATC-3'

migration time, as shown in Fig. 2, which makes it critical to use an internal migration standard such as Orange G, as Guttman *et al.* used [10] to compensate for the results obtained in different runs or to determine the migration coefficients in each run based on multiple oligo-DNA markers in a sample mixture. We used the latter method to describe the migration properties of the heterogeneous oligo-DNA by expanding the Gauss least-squares method for observation functions, which can give information on the exact separation in a run.

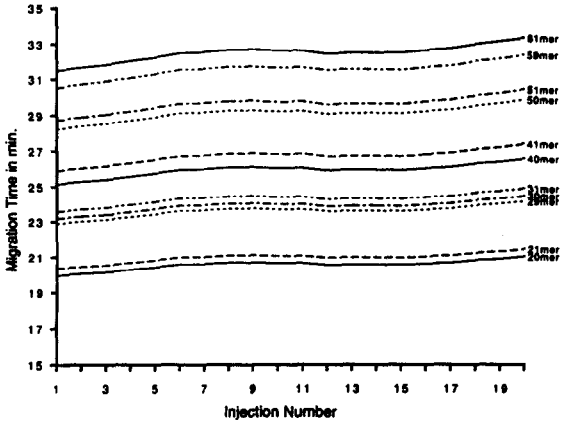

Fig. 2. Migration time shift during the first twenty successive injections. A newly prepared eCAP gel U100P capillary was subjected to the continuous separation of the heterogeneous oligo-DNA standard. The sample was freshly prepared every three injections as the peak height decreased to less than one third of that for the first injection owing to the inherent nature of electrokinetic injection. The electrolytes in 4-ml anodic and cathodic vials were changed every five injections. Other experimental conditions are summarized under Experimental. After 100 injections, no serious decrease in resolution was observed.

The results for the migration coefficient of each base analysed by this method are given in Table II. The calculated migration coefficients (*a*, *g*, *t* and *c*) varied with the bases, but in more than 20 successive runs they consistently showed that $t > g > a > c$ immigration coefficients and C > A > G > T in migration speed within the range of relative standard deviations 1.8% (G)–2.5% (T). Guttman *et al.* [10] observed a relative migration order of A > C > G > T, which is slightly different from our results. The reason for this is not clear, but the difference in oligo-DNA type and length may be critical as Guttman *et al.* mainly used homogeneous oligo-DNAs of 12–18-mers whereas we used heterogeneous 20–61-mers.

Attempts at averaging or normalization were made to improve the accuracy of the analysis, but no significant improvement in the probable error was achieved, as is shown in the last two rows of Table II.

The calculated migration time based on the migration coefficients agreed well with the actual migration times, as is shown in Table III, although some of the sequences gave a 0.16 min maximum difference after nearly 30 min of migration (even in this instance the deviation was as small as 0.5%). The reason for the variation

TABLE II

MIGRATION COEFFICIENTS FOR TWENTY SUCCESSIVE RUNS

a, *g*, *t* and *c* are migration coefficients for A, G, T and C, respectively, and *k* is a constant determined in each run or after averaging or normalization. The average migration coefficients ("Average") were calculated by averaging the migration times of each peak obtained by twenty successive runs, before Gauss least-squares analysis of observation functions. For the "Normalized" values, the migration times of each peak in twenty successive runs were corrected by assuming the migration time of the 61-mer to be 32 min before the analysis. The experimental conditions are summarized under Experimental.

Expt. No.	*a*	*g*	*t*	*c*	*k*	ΔM [a]	Δ/C [b]
1st	0.229	0.331	0.348	0.162	14.889	0.077	0.475
6th	0.238	0.343	0.360	0.166	15.345	0.079	0.476
11th	0.240	0.345	0.362	0.168	15.803	0.079	0.470
16th	0.240	0.345	0.363	0.169	15.353	0.076	0.450
20th	0.246	0.345	0.372	0.173	15.618	0.079	0.457
Average	0.238	0.343	0.361	0.168	15.309	0.078	0.464
Normalized	0.234	0.338	0.355	0.165	15.078	0.076	0.461

[a] ΔM = probable error of the migration time in minutes.
[b] Δ/C = probable error as the number of C.

remains unclear. Although in the set of experiments shown in Table III we used eleven heterogeneous oligo-DNAs as markers, as few as six markers gave equally good results.

We applied this technique to synthesized heterogeneous oligo-DNAs to examine the migration times, and found out that the calculated results agreed well with actual migration time in most instances. By using the migration coefficients obtained, we can easily predict that oligo-dC_{40}, for example, may migrate faster than oligo-dT_{20} as the expected migration times are 21.68 and 22.18 min, respectively.

Effects of secondary structure on migration time

One of the interesting applications of this approach is the determination of the migration

TABLE III

COMPARISON OF ACTUAL AND CALCULATED MIGRATION TIMES

A, G, T and C represent the numbers of dA, dG, dT and dC in each DNA fragment; t_M is the actual migration time and t_M (calc.) is the calculated migration time based on the Gauss least-squares method. Δ is the difference between the actual and calculated migration times.

Oligomer	A	G	T	C	t_M (min)	t_M (calc.) (min)	Δ (min)
20-mer	9	4	3	4	20.57	20.58	−0.01
21-mer	4	7	4	6	20.98	21.11	−0.13[a]
29-mer	7	5	10	7	23.61	23.47	0.14[a]
30-mer	7	5	11	7	23.90	23.83	0.07
31-mer	7	9	9	6	24.32	24.32	0.00
40-mer	10	11	7	12	25.95	26.00	−0.05
41-mer	17	7	11	6	26.72	26.73	−0.01
50-mer	16	17	6	11	29.12	28.96	0.16[a]
51-mer	15	11	15	10	29.65	29.74	−0.09[a]
59-mer	24	19	7	9	31.53	31.58	−0.05
61-mer	8	9	25	19	32.49	32.50	−0.01

[a] Differences of more than 0.5-mer C.

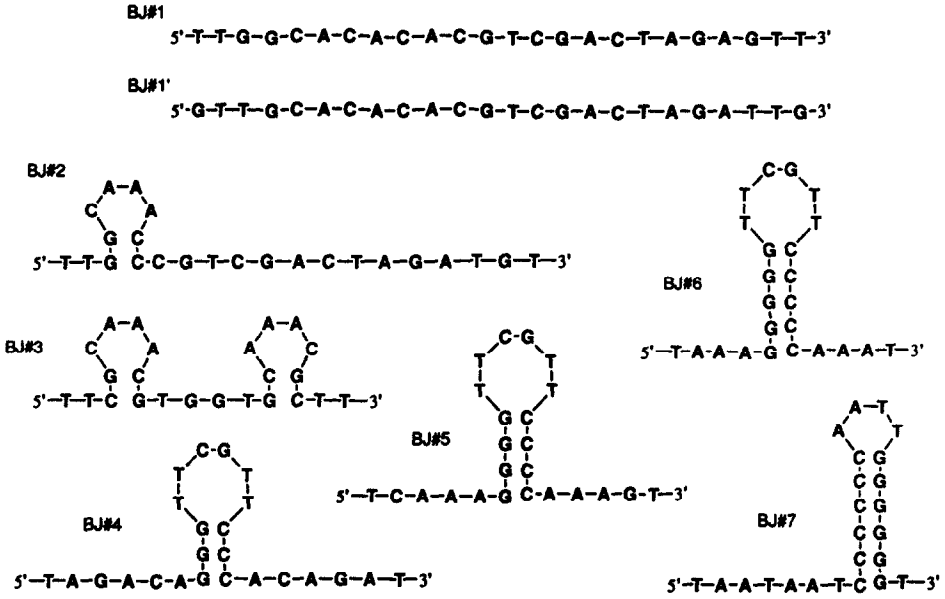

Fig. 3. Heterogeneous oligo-DNA used in the analysis for the effect of secondary structure on migration time. Most probable secondary structures are illustrated.

time of structural isomers of oligo-DNAs which may or may not have a secondary structure, because in this mathematical treatment the effect of such a structure was not taken into account. To minimize the effects of the difference in base compositions and the effect of the bases of the 5′- and 3′-termini, we designed and synthesized the 24-mer DNAs, which may or may not have secondary structure. All of them contained six of all four bases, having T at both ends except for BJ#1′, as shown in Fig. 3.

Contrary to the expectation that all these sequences would show the same migration time in the presence of 7 *M* urea, most of the sequences (except BJ#1 and #4) were separated from each other, as summarized in Table IV and

TABLE IV

MIGRATION TIMES OF DNA FRAGMENTS THAT MAY HAVE SECONDARY STRUCTURES

Each DNA fragment was run with the heterogeneous standard marker as described under Experimental. As there were small migration time drifts of the DNA standards, correction was done by assuming the migration time of the 31-mer in the standard to be 24.32 min [t_M (corr.)]. Migration coefficients were then calculated for every DNA fragment and were used for the calculation of estimated migration time [t_M (calc.)]. Δ is the difference between the corrected and calculated migration times.

Sequence	t_M (corr.) (min)	t_M (calc.) (min)	Δ (min)
BJ#1	23.13	22.83	0.30
BJ#1′	23.22	22.85	0.37
BJ#2	23.19	22.85	0.34
BJ#3	22.92	22.87	0.05
BJ#4	23.10	22.88	0.22
BJ#5	22.95	22.82	0.13
BJ#6	22.37	22.86	−0.49
BJ#7	20.76	22.88	−2.12

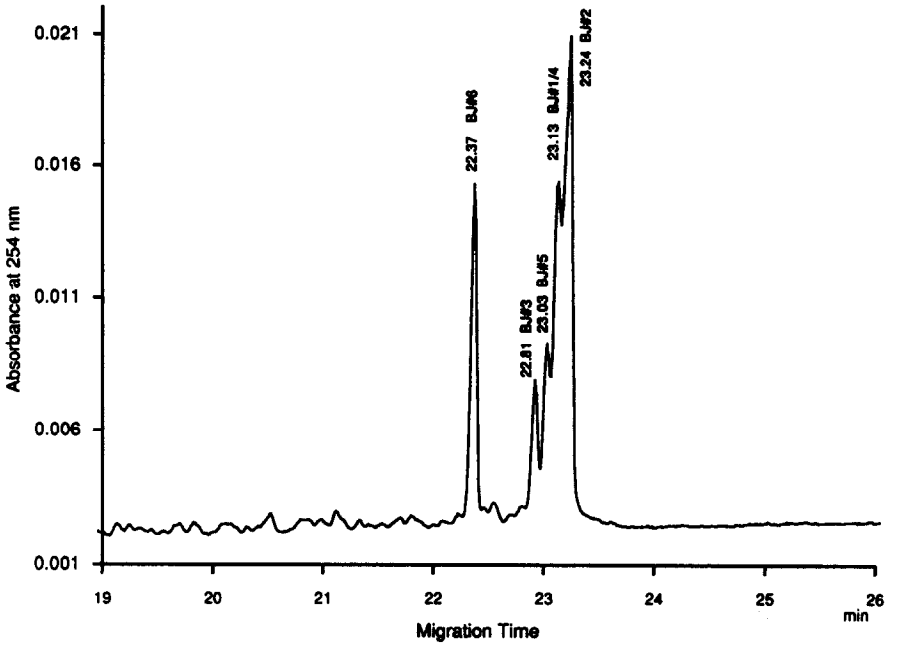

Fig. 4. Separation of heterogeneous oligo-DNAs that may have secondary structures in the presence of 7 *M* urea. The actual sequences #1–#6 are summarized in Fig. 3 with possible secondary structures. Separation conditions as in Fig. 1.

Fig. 4. This strongly suggested the existence of secondary structures. The deviation of the migration time of BJ#1 (linear oligo-DNA) from the calculated value was fairly large (0.30 min). One of the reasons for this is that different bases at the 3′- and 5′-termini must have different effects on the migration time. In fact, G at both termini (BJ#1′) had a 0.09 min longer migration time than that with T at both ends (BJ#1). The sequence that had a short G and C base pair stem (BJ#2) showed a slightly longer migration time than the linear sequence (BJ#1). As the difference was as small as 0.06 mins, it was not significant.

The parallel relationship between the decrease in migration time and the increase in the G and C base pair number in the sequences (Fig. 5) strongly suggested that the secondary structure due to hydrogen bonds or base stacking between the G and C bases plays an important role in maintaining the secondary structure even under such circumstances as in the presence of 7 *M* urea. The effect of having five or six G and C base pairs (BJ#6, 7) was so strong that the migration time was reduced by *ca.* 2% and 10% compared with linear oligo-DNA or calculated migration time, respectively.

Two loops in an oligo-DNA (BJ#3) seemed to have a greater effect on the migration time shift than one loop (BJ#2), which were clearly separated from each other as shown in Fig. 4.

Konrad and Pentoney [11] reported similar findings of an increment in mobility in cases where there was intermolecular base pairing at the 3′-end of single-stranded DNA molecules. However, this phenomenon may be generalized to most DNA fragments which have secondary structure, as we observed a decrease in migration time of oligo-DNA fragments that have a loop structure in the middle of the chain. They also reported the compression phenomena in DNA sequencing (Fig. 2 in ref. 11), which may be easily understood by assuming the existence of a loop and subsequent un-looping along with the decrease in the chain length from the 3′-terminus. As the effect of the decrease in migration time was as large as 10% of the calculated migration time in some instances (BJ#7), it may

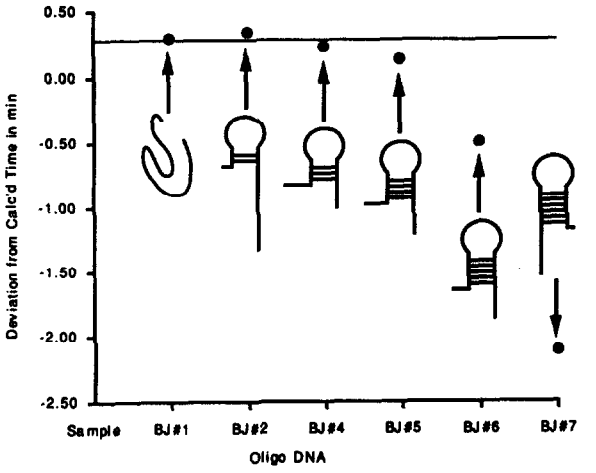

Fig. 5. Migration times and secondary structure. The results in Table IV are visualized by plotting the deviation of the actual migration time from calculated migration time *vs.* the number of possible hydrogen bonds. Although we are not sure whether there were hydrogen bonds or base stacking between the G and C bases to maintain the secondary structure in the presence of 7 *M* urea in the case of the small number of the G and C base pairs, the decrease in migration time and the increment of the number of G and C base pairs are directly correlated. This strongly suggests that hydrogen bonds or base stacking plays an important role in affecting the migration time, probably via secondary structure.

explain the peak compression phenomena. In addition, this predicts the apparent chain length reversal along with the decrease in the actual chain length in DNA sequencing experiments (data not shown).

CONCLUSIONS

By using the Gauss least-squares method for observation functions, we can describe the apparent migration phenomena with probable errors of less than 0.5-mer C with a limited number of oligo-DNA standards (at least six). This method gives separation information for each run as the migration coefficients are calculated every time based on a run. This statistical treatment showed that the migration time of heterogeneous oligo-DNA is the function of the chain length and the base composition. Typical migration coefficients are 0.165 (c), 0.234 (a), 0.338 (g), 0.355 (t) and 15.078 (k) under the present experimental conditions. The existence of secondary structures even in the presence of 7 *M* urea was strongly suggested from the experiments with specially designed oligo-DNAs which may have secondary structures. By changing the urea concentration or other constituents of the gel, such as the Mg^{2+} concentration, it may be possible to detect secondary structures of different magnitudes. As this migration time analysis based on the Gauss least-squares method for observation functions can given an indication of whether there is secondary structure or not, one of the most important applications of this technique may be point mutation detection following the polymerase chain reaction–single strand conformational polymorphism (PCR-SSCP) method on conventional slab gels [13]. As gene regulation via the triple helix is becoming an increasingly important subject of investigation [14,15], CGE may be a useful technique for evaluating such structures if appropriate gels and experimental conditions are developed.

REFERENCES

1 J.W. Jorgenson and K.D. Lukacs, *Anal. Chem.*, 53 (1981) 1298–1302.
2 S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya and T. Ando, *Anal. Chem.*, 56 (1984) 111–113.
3 W.G. Kuhr, *Anal. Chem.*, 62 (1990) 403R–414R.
4 W.G. Kuhr and C.A. Monnig, *Anal. Chem.*, 64 (1992) 389R–407R.
5 A.S. Cohen and B.L. Karger, *J. Chromatogr.*, 397 (1987) 409–417.
6 A.S. Cohen, D.R. Nagarian, A. Polus, A. Guttman, J.A. Smith and B.L. Karger, *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 9660–9663.
7 A. Guttman, A.S. Cohen, D.N. Heiger and B.L. Karger, *Anal. Chem.*, 62 (1990) 137–141.
8 J.A. Luckey, H. Drossman, A.J. Kostichka, D.A. Mead, J.D'Cunha, T.B. Norris and L.M. Smith, *Nucleic Acids Res.*, 18 (1990) 4417–4421.
9 Y. Baba, T. Matsuura, K. Wakamoto, Y. Morita, Y. Nishitsu and M. Tsuhako, *Anal. Chem.*, 64 (1992) 1221–1225.
10 A. Guttman, R.J. Nelson and N. Cooke, *J. Chromatogr.*, 593 (1992) 297–303.
11 K.D. Konrad and S.L. Pentoney, Jr., *Electrophoresis*, in press.
12 M. Ichinose, *Gosaron*, Baifukan, Tokyo, 1953 (in Japanese).
13 M. Orita, H. Iwahana, H. Kanazawa, K. Hayashi and T. Sekiya, *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 2766–2770.
14 K. Yoon, C.A. Hobbs, J. Koch, M. Sardaro, R. Kutny and A.L. Weis, *Proc. Natl. Acad. Sci. U.S.A.*, 89 (1992) 3840–3844.
15 R.W. Roberts and D.M. Crothers, *Science*, 258 (1992).