J. Biochem. 120, 320-325 (1996)

Single-Strand Conformation Polymorphism (SSCP) Can Be Explained by Semistable Conformation Dynamics of Single-Stranded DNA[1]

Yoichi Nakabayashi[2] and Koichi Nishigaki[3]

Department of Environmental Chemistry and Department of Functional Materials Science, Saitama University, 255 Shimo-Okubo Urawa, Saitama 338

Received for publication, March 1, 1996

The cause of separation in gel electrophoresis between highly homologous ss (single-stranded) DNAs as observed in SSCP (single-strand conformation polymorphism) was pursued. Advancing our previous explanation [*J. Biochem.* 99, 663-671 (1986)], the mobility difference of ssDNAs was correlated to differences in their dynamic conformation (not to differences in their most stable structure), focusing on point-substituted sites. The contribution of semistable conformation dynamics was considered to be critical. Putative factors which may influence mobility (*i.e.*, length of ssDNAs, location of substituted sites, and types of substitutions) were experimentally examined and critically discussed. Understanding of these phenomena should yield improvements in various techniques, such as SSCP, and in evaluation of the solution structures of DNA.

Key words: conformation dynamics, point mutation, SSCP, ssDNA, solution structure.

The solution structures of nucleic acids are not yet completely understood, although they are thought to have a direct relationship with function (*1*). Recently, there have been considerable advances in studies on unusual and semistable structures (*2-4*). Under physiological conditions, semistable structures of nucleic acids may participate, more or less, in biological functions. For example, the behavior of ssDNA in solution is important in relation to both hybridization and separation techniques. Surprisingly, highly homologous ssDNAs (a few point-substitutions) in solution were found to be separable by gel electrophoresis by Nishigaki *et al.* (*5*) (*i.e.*, basic discovery of SSCP). Independently, Orita *et al.* developed a technique to detect point-mutations occurring in ssDNAs, naming the technique single-strand conformation polymorphism (SSCP) (*6*). Irrespective of the mechanism of separation, this technique has been widely accepted owing to its great utility. Various efforts to improve the technique have been made (*7-13*). The effects of length of DNA fragments, position of mutation site, types of base substitutions, *etc.*, on the separation have been investigated (*10-13*). However, consistent results have not been obtained. We have conducted further studies, which substantially support and extend our previous explanation (*5*). That is, the structural dynamics of ssDNAs, including semistable conformations, can explain the mobility differences between highly homologous ssDNAs.

THEORETICAL

The solution structure of ssDNA (or RNA) can be approximated as a hybrid of helices and coils, if the tertiary structures can be neglected. The helix, which is mostly sustained by Watson-Crick base-pairings, is interposed by mismatches, non-Watson-Crick base-pairings, bulges and/or loops (*1*). Therefore, most such helices are smaller and far less stable than those of tRNAs. Thus, intramolecular interactions are supposed to occur freely between any two parts of a DNA molecule and then the other parts in the remaining regions interact with each other and so on. This idea leads to a dynamical model for the solution structure of DNA: semi- and unstable structures, easy to unfold, interconvert into other possible structures within a relatively short period (see Fig. 1; a more theoretical paper regarding this is in preparation by the authors). In short, variously altering conformations of a DNA molecule, most of which are semi- and unstable, contribute in ensemble to the macroscopic mobility. What is important is that the spectrum (diversity) of conformations is determined by the sequence of DNA and that even a single-base substitution leads to alteration of the conformational spectrum. Therefore, we can correlate the difference in conformational spectrum with that in observed electrophoretic mobility between a pair of single-base-substituted DNAs.

MATERIALS AND METHODS

Preparation of ssDNAs—Point substitution series of ssDNAs were prepared by the following method: PCRs were performed using the appropriate primers as shown in Fig. 2 to obtain ds (double stranded) DNAs. For series C, jumping PCR which uses two types of templates with overlapping ends (5291-5341 and 5329-5401) was adopt-

[1] This study was supported by a Grant-in-Aid (#06454661) from the Ministry of Education, Science, Sports and Culture of Japan.
[2] Present address: Japan International Cooperation Agency (JICA) 1-1 Nishi-shinjuku 2-chome, Shinjuku-ku, Tokyo 163-04.
[3] To whom correspondence should be addressed.
Abbreviations: ds, double-stranded; ss, single-stranded; SSCP, single strand conformation polymorphism.

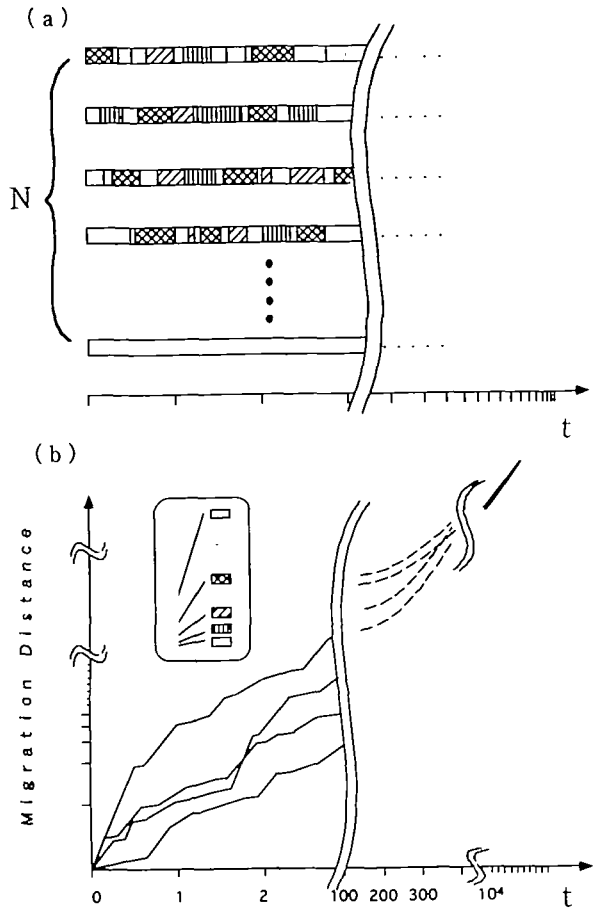

Fig. 1. **Schematic drawing of conformational dynamics and migration in gel of ssDNAs.** (a) Each pattern, hatched or stippled, represents an individual conformation of an ssDNA and the length of a pattern is proportional to the duration of the conformation. Each of N molecules is shown to experience a different course of conformation history (random process). (b) Each conformation has its own mobility in gel (shown by the insert). Therefore, in a short period, individual molecules migrate quite differently (left in b) but, in a long period, the statistical effect based on the law of large numbers renders their migration distances identical (rightward in b).

ed. Prior to electrophoresis, dsDNAs were boiled for 10 min and then rapidly chilled to generate ssDNAs (plus/minus equimolar mixtures). To obtain A–D series of plus-excess or minus-excess ssDNA solutions, asymmetric PCR, in which excess primers of one type become the seeds for ssDNA, was used. The reaction buffer for PCR contained: 1 fmol of template DNA, 10 pmol of (each) primers, 100 μM (each) dNTPs and 1 unit of Taq DNA polymerase in 50 μl of PCR buffer [50 mM KCl, 2.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 9.0), and 0.1% Triton X-100] and PCR was usually carried out with the cycle of denaturation (94°C, 30 s), annealing (45°C, 50 s), and extension (65°C, 50 s).

Electrophoresis—Gel electrophoreses of ssDNAs were performed at a constant temperature (6°C). Gels of 5% (w/v) acrylamide in 0.5×TBE buffer [1×TBE; 90 mM Tris-base, 92 mM boric acid, and 2.5 mM EDTA (pH 8.3)] containing 5% (v/v) glycerin and of 18 cm in length were used.

Computing—Stabilities of secondary structures were calculated using a computer program HYBRI, which was derived from our previous program PCRAna (*14*). The small region (*ca.* 20 nts in length) which contains the substituted base was regarded as "a primer" and the remaining regions of an ssDNA as "a template" in the original program, PCRAna. This enables us to obtain the free energies for all possible local secondary structures which involve the point-substituted site. The algorithm and thermodynamic parameters used were the same as those used in the previous work (*14*). Briefly, since the complete thermodynamic parameter set for DNA cannot be obtained at present due to the lack of loop entropy parameters for DNA, the thermodynamic parameters used for calculating these structures were those obtained by Salser (*15*) for RNA, which are known to provide a good approximation for the estimation of helix-coil transitions of DNAs (*16*). Since we neglected, in calculation, the possible secondary structures for loops and regarded them as random coils for the sake of simplicity, only loop parameters for bulge and internal loops were used. (The effect of this simplification is discussed later.) The program SEQL (*17*), which was used for obtaining secondary structures, is available from the Computer Center of Kyoto University.

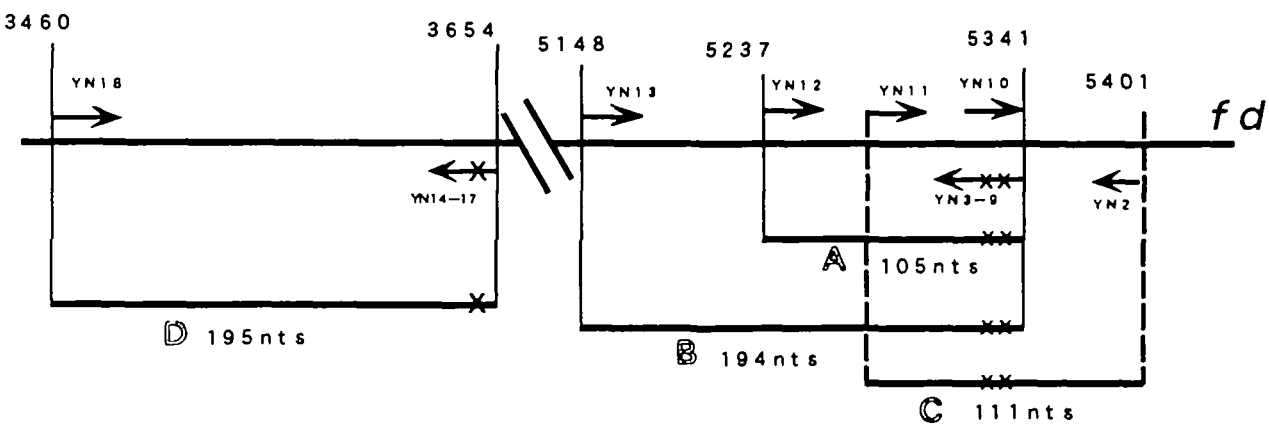

Fig. 2. **Point-substituted ssDNA series.** The ssDNA series, A (7 members of 105 nts), B (7 of 194 nts), C (7 of 111 nts), and D (4 of 195 nts), were prepared as described in the text using fd DNA. Nucleotide numbering of fd DNA is the same as in GenBank. The primers used were forward primers; YN10 (dAGACTCTTTTG), YN11 (dAAGAAGTATTGC), YN12 (dTAAGGCCGATAG), YN13 (dCAGACGGTTGAG), and YN18 (dTGAATGCGGTACTTG) and reverse primers; YN2 (dTTAGACAGGAAC), YN3-9 (dAGCAAAAGAGTCTXYCCA), and YN14-17 (dGTAATTCTGTCZAGA). The primers of YN3-9 and YN14-17 contain substitution site(s), X/Y and Z at 5328/5327 and 3643, respectively. Phage type ssDNA is a plus strand by definition.

RESULTS AND DISCUSSION

Four sets of point-substitution series, A-D, of single-stranded DNAs (see Fig. 2) were prepared and subjected to gel electrophoresis (Fig. 3). Each band in Fig. 3 was assigned to be a plus or minus strand or a double strand by referring to the results of asymmetric PCR for the corresponding DNA fragments (data not shown). Evidently, there are mobility differences among DNA fragments of each series. As discussed in detail later, there are parallel correlations in mobility among the three series of A-C, which have a common substitution site and common flanking regions.

Differences in Conformation Dynamics—ssDNAs in solution interact intramolecularly as shown in Fig. 4. If there are any differences in dynamically changing conformations between mutually point-substituted ssDNAs (Fig. 4, a and b), they must have resulted from those secondary structures which involve the point-substituted site (the hatched or stippled boxes in Fig. 4, a and b) since the other interactions (shown by broken lines in Fig. 4) are equivalent for both the ssDNAs. As, based on the theory introduced in "THEORETICAL," the mobility difference should be reduced to the difference in conformation dynamics, we calculated using HYBRI the free energy contributions of the local structures of each sequence which include a point-substituted site and obtained the most stable conformers and their ΔG values. What is important is that difference-generating conformers thus obtained give not the global minimum but only a local minimum at most. In other words, the global minimum conformation, which may be the same for each sequence, will not explain the real mobility difference. Figure 5a shows an experimental result obtained for series A minus strands. There is a good correlation between the mobility of an ssDNA expressed in r_m [$r_m = 2(m - m_s)/(m_f - m_s)$ where m_f and m_s denote the fastest mobility and the slowest one in a particular series, respectively], and the free energy change of the representative difference-generating conformer. Exact calculation for this problem is impossible at present because of its complexity and the low reliability of the thermodynamic parameters. Therefore, we had to introduce an approximation to solve this problem. Since it is evident that the most stable structure has the greatest influence on the entire mobility, and if a similar, though weaker, tendency can be expected for the remaining structures (which was computer-experimentally supported, although there is no clear rationale for it at present), the above approximation would be reasonable. Similar correlations were obtained for series B and D [data not shown; series C was omitted due to its ambiguity. A few points in series B and D which behaved to reduce the correlation, were further analyzed to see whether or not they might have resulted from the approximation adopted in the program, HYBRI. Since this program is designed to search even very unstable structures at the cost of narrowing the range of direct interactions to within less than 100 nts or so (*14*), it may fail to take into account more global contributions. To compensate for this effect, we employed a program (SEQL) which has an opposite principle to ours (*17*): this program cuts off less stable structures at the early stage of calculation in order to obtain a global minimum in conformation space (note that both programs (HYBRI and SEQL) were equipped with RNA parameters, and therefore, if applied to DNA, are only approximately correct). When we computed the minimum free energies of local structures which surround the point-substituted sites (100-200 nucleotides in length), in both the plus and minus

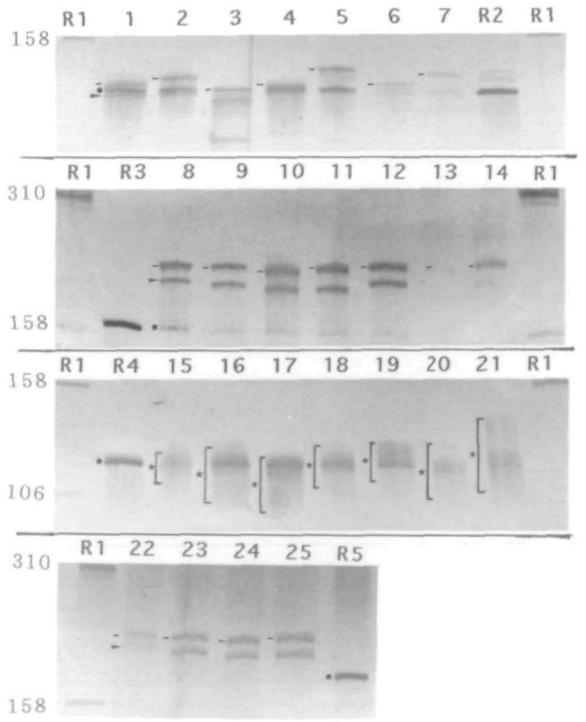

Fig. 3. **Electrophoretic results of point-substituted ssDNA series.** In series A (lanes 1-7), B (8-14), and C (15-21), lanes 1, 8, and 15 are the wild type (XY=GA) and the rest are (XY=): GT (2, 9, and 16); GG (3, 10, and 17); GC (4, 11, and 18); AA (5, 12, and 19); TA (6, 13, and 20); CA (7, 14, and 21). In series D (lanes 22-25), lane 25 is the wild type (Z=C); lane 22, Z=A; lane 23, Z=T; lane 24, Z=G. R_1 is *Hae*III restriction fragments of fd ssDNA as molecular weight markers. R_2, R_3, R_4, and R_5 are non-denatured wild-type dsDNAs of A, B, C, and D series, respectively. The symbols used are as follows: minus strand (−), plus strand (▶), double strand (●) and the center of a smear band (*). Band assignment of plus/minus strands was done by reference to the results of asymmetric PCR (data not shown). ds bands can be seen as thick bands on the background of smears (lanes 15-21).

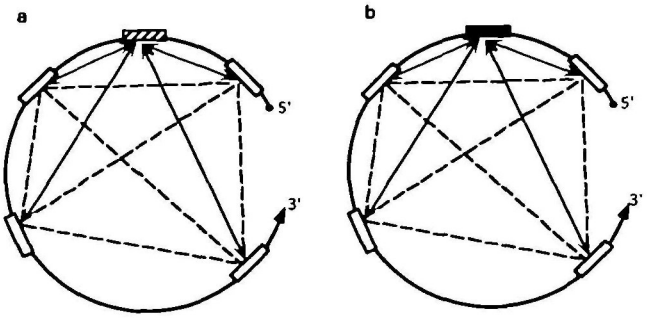

Fig. 4. **Intramolecular interactions which cause differences in mobility.** For both molecules (left and right), intramolecular interactions (shown by solid and broken lines) are the same except those interactions (solid lines) which involve the substitution site (hatched or stippled).

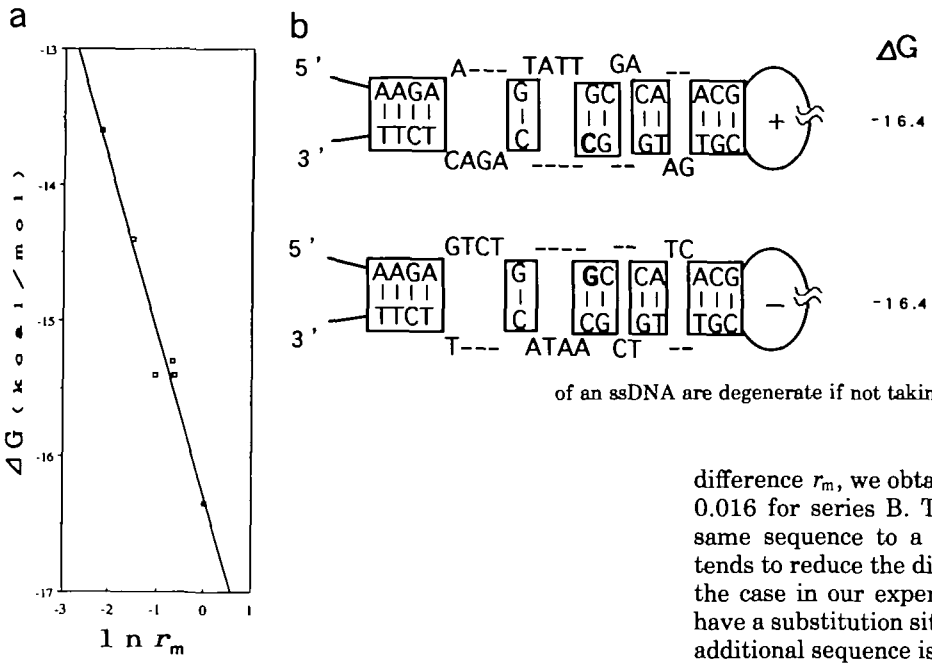

Fig. 5. **Mobility *vs.* stability and calculated stable structures of ssDNAs.** a: Relative mobility difference (r_m) against stability (ΔG) of the most influential subconformer of an ssDNA is plotted for each member of series A. r_m is defined in the text. The GG- (fastest) and the AA- (slowest) substituted for XY were used throughout this series. b: One example of the most influential subconformers for plus/minus strands in series A is depicted. The RNA parameters were used as mentioned in the text. Note that ΔG values for the plus and minus strands of an ssDNA are degenerate if not taking the sequence effect at loops into account.

strands of D series, those points which reduced the correlation were shifted toward a better correlation, while the others were not much influenced. This observation indicates that although semistable conformations are principally responsible for the phenomenon investigated here, the stability values that we have calculated remain to be improved. Therefore, although at present the theoretical approach using computer calculation is applicable not to obtaining exact values but to evaluating relative values, the overall scheme supported by them was consistent with the experimental results.] These results support the hypothesis that mobility differences of point substituted ssDNAs in gel electrophoresis result from differences in their dynamical structures, not from a difference in their most stable structures.

In Fig. 5b, the locally most stable conformations which were calculated with HYBRI are shown. Evidently, these conformations are not remarkably stable ones and by no means correspond to the global minimum. Notwithstanding, semistable structures, which are far less durable during the experimental time scale [minute to hour; direct evidence for this fact was obtained by a kinetic experiment in which restriction enzyme *Hae*III cleaved semistable secondary structures within this time scale (*18*)] and have rather small contributions to the global conformation dynamics, must have worked as a separation impetus. In this sense, the static image which assigns a particular stable conformation to each ssDNA and regards it as stable during the experimental time scale (current notions of SSCP often appear to be based on this image) is not correct. So, the term conformational polymorphism in SSCP should mean polymorphism not in stable conformations but in difference-determining semistable conformations.

Factors Affecting Mobilities—As shown in Fig. 3, there are much greater differences in mobility among the members of series A than those of series B which have an additional sequence (89 nts) to that of series A (105 nts). Transforming the mobility differene into relative mobility difference r_m, we obtained $r_m = 0.053$ for series A and $r_m = 0.016$ for series B. This may imply that addition of the same sequence to a series of point-substituted ssDNAs tends to reduce the difference in mobility. We think this is the case in our experiment, since the ssDNAs used here have a substitution site near one end of each strand and the additional sequence is linked remotely to the opposite end, resulting in weaker interactions between the substitution site and the additional sequence due to the loop entropy. This provides a rational explanation for the seemingly contradictory reports that larger fragments (354 nts) resolved better than smaller ones (176 nts) by Fan *et al.* (*10*), whereas Hayashi had suggested that the smaller ones (100-300 bp) are more informative (separate better) than the larger (300-450 bp) (*19*). Recently, Ushijima *et al.* also reported that ssDNAs of less than about 250 nts were informative (*13*). Actually, Fan *et al.* used DNAs which were overlapping in the middle part of the larger DNA (see Fig. 6) so that additional sequences at both sides of the larger DNA had additional interactions with the point-substituted site, affording different semistable structures which can generate differences in mobility, while the samples treated by Hayashi were heterogeneous from this viewpoint. Therefore, we cannot predict separation effect only from the length of ssDNAs; the context of a substituted site is important.

A dramatic positional effect appeared in this experiment as presented by series C, where every band became a smear (which was quite reproducible), whereas the corresponding dsDNAs always gave sharp bands (partly shown in Fig. 3 lane R4). Smears may be interpreted as indicating the presence of multiple rather stable conformations in ssDNAs, since those stable conformations tend to be distributed unequally in each molecule within a short period. It is noteworthy that when measured at the center of a smear band (star-marks in Fig. 3), series C ssDNAs kept a similar mobility-relationship to those of series A and B, supporting the notion that the final mobility of ssDNA is determined by combining the mobility difference and the basal common mobility. That is, if we can assign mobilities to each conformer, we can compute mobility differences quantitatively (papers aiming toward this goal: Refs. *4* and *20*). Anyway, provided that the PCR products used here for series C have the expected sequences (which cannot be confirmed currently), the smear phenomenon observed in series C is strong evidence that a single sequence has a range of mobility, corresponding to multiple conformers

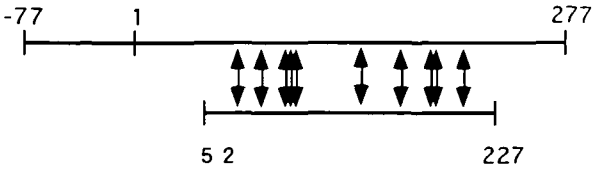

Fig. 6. **Point mutation sites in ssDNAs tested in Ref. *10*.** Two types of fragments (large and small) are shown with sequence numbers of the lacI gene of *E. coli*. Arrows indicate the sites at which point mutations, single for each strand, occurred.

(conformation polymorphism).

Series D, which has similar length (195 nts) and similar positioning of the substitution site to series B, presented distinct mobility differences from series B, where the G-substituted molecule (lane 10), for example, is significantly more mobile than the T-substituted one (lane 9), whereas in series D, the G- (lane 24) and the T-substituted (lane 23) molecules have comparable mobility. Since the context (or flanking sequences) of a substituted site is critical in determining mobility differences, this result is quite understandable. On the other hand, such a generalization as the suggestion that purine-rich strands are more sensitive to changes in gel electrophoresis (*11*) has no rationale here; it was experimentally ruled out here by the observation that purine/pyrimidine equimolar strands (the plus/minus strands of series A in Fig. 3; the wild-type plus strand contains 52 purines and 53 pyrimidines) were not the same in sensitivity (lanes 1-7 in Fig. 3).

The most stable subconformer that contains a point-substituted site is, stochastically, not highly stable in general and at most semistable, as shown in Fig. 5b. Nevertheless, it is the main cause of the difference in mobility. Conventionally, loop entropy is calculated by taking account of only loop size and the sequence is neglected. [Empirically, such simplification combined with the classical RNA parameters of Salser (*15*) is known to give a good approximation which does not contradict experimental results (*14*). This may be partly because the present level of knowledge does not require such improved parameters as those of Turner *et al.* (*21*).] Therefore, as shown in Fig. 5b, plus and minus strands in a corresponding secondary structure, just as shown in this figure, take the same value of free energy change. Every secondary structure taken by a plus strand can be taken by the complementary minus strand with the same stability, at least in the current calculation! Consequently, if the tertiary structure effect is negligibly small, a pair of plus and minus strands should have had the same mobility. As this is not the case (see Fig. 3), either the tertiary structure effect or the sequence effect at loops must be significant. The tertiary structure effect is based on interactions between nucleotides rather remotely located along the sequence, while the sequence effect at loops is generated by nucleotide-nucleotide interactions occurring nearby along the sequence and by nucleotide-solvent interactions. This suggests that the sequence effect at loops may contribute substantially to the separation between plus and minus strands, since the interaction frequency of two moieties depends on their relative distance. The sequence effect here discussed should include non-Watson-Crick base pairings and other local nucleotide interactions such as GNRA (*22*) and those observed in oligonucleotides (*4*). In this sense too, study of the semistable structures of DNA/RNA is essential to an understanding of nucleic acids.

In conclusion, we have emphasized the importance of a dynamic image of ssDNAs in solution (semistable structures) which should be helpful in improving the SSCP technique and in understanding the solution structures of DNAs.

The authors are grateful to Dr. Yuzuru Husimi for critical discussions.

REFERENCES

1. Saenger, W. (1984) *Principles of Nucleic Acids Structure*, pp. 1-556, Springer-Verlag, New York
2. Wells, R.D., Amirhaeri, S., Blaho, J.A., Collier, D.A., Hanvey, J.C., Hsieh, W.-T., Jaworski, A., Klysik, J., Larson, J.E., Mclean, M.J., Wohlrab, F., and Zacharias, W. (1988) Unusual DNA structures and the probes used for their detection in *Unusual DNA Structures* (Wells, R.D. and Harvey, S.C., eds.) pp. 1-21, Springer-Verlag, New York
3. Jaeger, J.A., SantaLucia, J., Jr., and Tinoco, I., Jr. (1993) Determination of RNA structure and thermodynamics. *Annu. Rev. Biochem.* **62**, 255-287
4. Nishigaki, K., Miura, T., Tsubota, M., Sutoh, A., Amano, N., and Husimi, Y. (1992) Structural analysis of nucleic acids by precise denaturing gradient gel electrophoresis: I. Methodology. *J. Biochem.* **111**, 151-156
5. Nishigaki, K., Husimi, Y., and Tsubota, M. (1986) Detection of differences in higher order structure between highly homologous single-stranded DNAs by low-temperature denaturant gradient gel electrophoresis. *J. Biochem.* **99**, 663-671
6. Orita, M., Iwahara, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989) Detection of polymorphism of human DNA by gel electrophoresis as single strand conformation polymorphism. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770
7. Sarkar, G., Yoon, H.-S., and Sommer, S.S. (1992) Screening of mutations by RNA single-strand conformation polymorphism (rSSCP): Comparison with DNA-SSCP. *Nucleic Acids Res.* **20**, 871-878
8. Lo, Y.-M.D., Patel, P., Mehal, W.Z., Fleming, K.A., Bell, J.I., and Wainscoat, J.S. (1992) Analysis of complex genetic systems by ARMS-SSCP application to HLA genotyping. *Nucleic Acids Res.* **20**, 1005-1009
9. Savov, A., Angelicheva, D., Jordanova, A., Eigel, A., and Kalaydjieva, L. (1992) High percentage acrylamide gels improve resolution in SSCP analysis. *Nucleic Acids Res.* **20**, 6741-6742
10. Fan, E., Levin, D.B., Glickman, B.W., and Logan, D.M. (1993) Limitations in the use of SSCP analysis. *Mutat. Res.* **288**, 85-92
11. Glavac, D. and Dean, M. (1993) Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Human Mutat.* **2**, 404-414
12. Tokue, Y., Sugano, K., Saito, D., Noda, T., Ohkura, H., Shimosato, Y., and Sekiya, T. (1994) Detection of novel mutations in the *gyrA* gene of *Staphylococcus aureus* by nonradioisotopic single-strand conformation polymorphism analysis and direct DNA sequencing. *Antimicrob. Agents Chemother.* **38**, 428-431
13. Ushijima, T., Hosoya, Y., Suzuki, T., Sofuni, T., Sugimura, T., and Nagao, M. (1995) A rapid method for detection of mutations in the lacI gene using PCR-single strand conformation polymorphism analysis. *Mutat. Res.* **334**, 283-292
14. Sakuma, Y. and Nishigaki, K. (1994) Computer prediction of general PCR products based on dynamical solution structures of DNA. *J. Biochem.* **116**, 736-741
15. Salser, W. (1977) Globin mRNA sequences: Analysis of base pairing and evolutionary implications. *Cold Spring Harbor Symp. Quant. Biol.* **42**, 985-1002
16. Nishigaki, K., Husimi, Y., Masuda, M., Kaneko, K., and Tanaka, T. (1984) Strand dissociation and cooperative melting of double-stranded DNAs detected by denaturant gradient gel electrophoresis. *J. Biochem.* **95**, 627-635
17. Kanehisa, M. and Goad, W.B. (1982) Pattern recognition in

nucleic acid sequences. II. An efficient method for finding locally stable secondary structures. *Nucleic Acids Res.* **10**, 265-278
18. Nishigaki, K., Kaneko, Y., Wakuda, H., Husimi, Y., and Tanaka, T. (1985) Type II restriction endonucleases cleave single-stranded DNAs in general. *Nucleic Acids Res.* **13**, 5747-5760
19. Hayashi, K. (1991) PCR-SSCP: A simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods Appl.* **1**, 34-38
20. Abrams, E.S. and Stanton, V.P., Jr. (1992) Use of denaturing gradient gel electrophoresis to study conformational transition in nucleic acids. *Methods Enzymol.* **212**, 71-104
21. Turner, D.H., Sugimoto, N., and Freier, S.M. (1988) RNA structure prediction. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167-192
22. SantaLucia, J., Jr., Kierzek, R., and Turner, D.H. (1992) Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science* **256**, 217-219