Prediction of DNA single-strand conformation polymorphism: analysis by capillary electrophoresis and computerized DNA modeling

Donald H. Atha*, Wojciech Kasprzak1, Catherine D. O’Connell and Bruce A. Shapiro2

Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA, 1Intramural Research Support Program (IRSP) SAIC-Frederick and 2Laboratory of Experimental and Computational Biology, The NCI Center for Cancer Research, NCI-Frederick, Frederick, MD 21702, USA

Received July 24, 2001; Revised and Accepted September 28, 2001

ABSTRACT

We have analyzed previously three representative p53 single-point mutations by capillary-electrophoresis single-strand conformation polymorphism (CE-SSCP). In the current study, we compared our CE-SSCP results with the potential secondary structures predicted by an RNA/DNA-folding algorithm with DNA energy rules, used in conjunction with a computer analysis workbench called STRUCTURELAB. Each of these mutations produces measurable shifts in CE migration times relative to wild type. Using computerized folding analysis, each of the mutations was found to have a conformational difference relative to wild type, which accounts for the observed differences in CE migration. Additional properties exhibited in the CE electropherograms were also explained using the computerized analysis. These include the appearance of secondary peaks and the temperature dependence of the electrophoretic patterns. The results yield insight into the mechanism of SSCP and how the conditions of this measurement, especially temperature, can be optimized to improve the sensitivity of the SSCP method.

INTRODUCTION

There is an increasing need in DNA diagnostics for more efficient methods of detecting mutations associated with disease. Analysis by single-strand conformational polymorphism (SSCP) provides an efficient means to screen these mutations before the costly and time consuming task of sequencing is begun. The SSCP method is run under specific electrophoretic conditions in which the conformational changes in single-stranded DNA, which result from single-point mutations, can be detected as shifts in migration time (1). Improved methods are needed to predict the sensitivity of SSCP in detecting different point mutations and to optimize SSCP conditions (2). To this end, we have analyzed p53 single-point mutations by capillary-electrophoresis (CE-SSCP) (1) and have compared our results with structures predicted by DNA-folding analysis using DNA energy rules (3–8). These results yield insight into the mechanism of SSCP and how the conditions of this measurement, especially temperature, can be optimized to improve the sensitivity of the SSCP method.

Our analysis is based on a hypothesis that similar structures, i.e. structures with similar stems, will most likely show similar behavior in CE-SSCP runs. In this study we do not attempt to correlate the relative speed differences from the CE-SSCP data (mutations versus wild type) with particular structural elements (substructures), but rather try to show general agreement with the CE data. In other words, we show that as the CE-SSCP distinguishes between the wild type sequences, sense and antisense, and their respective mutations, so do the solution spaces of the predicted secondary structures. We have also tackled a more complex problem of selecting specific secondary structures that could match the CE data by looking at the overall differences in their topologies.

MATERIALS AND METHODS

PCR amplification of p53 mutations

The preparation of single-point p53 mutations was described previously (2). Briefly, genomic DNA was isolated from cell lines known to contain human p53 mutations and amplified with exon-specific fluorescent-labeled PCR primers. The cell lines were obtained from the American Type Culture Collection (Rockville, MD) and contain point mutations in p53 exon 7 as shown in Table 1. The fluorescent-labeled primers labeled with FAM (5-carboxyfluorescein), 5′ primer, JOE (2′,7′-dimethoxy-4′,5′-dichloro-6-carboxyfluorescein) and 3′ primer were obtained from Applied Biosystems (Foster City, CA) using the specific primer sequences for exon 7 purchased

*To whom correspondence should be addressed at: NIST, Biotechnology Division, Building 227, Room A-243, Gaithersburg, MD 20899-8311, USA.
Tel: +1 301 975 3092; Fax: +1 301 975 8503; Email: donald.atha@nist.gov

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
from Clontech Laboratories, Inc. (Palo Alto, CA). (Certain commercial equipment, instruments and materials are identified in this paper in order to specify an experimental procedure as completely as possible. In no case does this identification of particular equipment or materials imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material, instrument, or equipment is necessarily the best available for the purpose.)

PCR products were obtained each containing complementary 139 bp nucleotide sequences corresponding to one of the three mutations or wild type. These sequences (exon 7) were used for the computerized analysis described below. The PCR products were screened for homogeneity by agarose gel electrophoresis and diluted 10-fold in H₂O (2).

**CE-SSCP analysis**

CE-SSCP analysis of p53 samples was described previously (1). Briefly, fluorescent-labeled PCR samples were prepared for electrophoresis by combining 10.5 µl deionized formamide with 0.5 µl 0.3 M NaOH, 1 µl water, 1 µl of PCR sample (diluted 1:10) and 0.5 µl of GENESCAN-500 TAMRA (6-carboxy-tetramethyl-rhodamine)-labeled internal size standard. (The accepted SI unit of concentration, mol/l, has been represented by the symbol M in order to conform to the conventions of this journal.) The mixture was heated for 2 min at 95°C to separate the complementary DNA strands and chilled on ice. SSCP separations were performed using the Beckman P/ACE™ Model 5510 CE equipped for laser excitation at 488 nm and detection at 560 nm and a Perkin Elmer/ Applied Biosystems PRISM™ Model 310 Genetic Analyzer modified to evaluate mutations at subambient temperatures. All separations on both instruments were performed using the Perkin Elmer, Applied Biosystems PRISM™ capillary and polymer system [41 cm × 75 µm capillary, 3% (w/w) GENESCAN™ polymer containing 10% (w/w) glycerol in 1x TBE]. This capillary and polymer system was chosen because previous studies demonstrated its high resolution and reproducibility for the detection of sequence-induced mobility differences (9). Samples were electrokinetically injected (10 s, 3 kV) and separated at 10–13 kV. Data were collected and analyzed using Beckman System Gold™ and Perkin Elmer/ Applied Biosystems PRISM™ and GENESCAN™ software, version 2.0.2 (1). Relative standard deviations for the migration times ranged from 0.01 to 0.05% as reported previously (10).

**Computerized analysis of single-stranded DNA**

We have used RNAstructure 3.5, the Dynamic Programming Algorithm (DPA) implementation, as our primary tool for the prediction of the secondary structures of DNA (3). Results presented below are based on the top 10% of the solutions (in terms of free energy of the structures) at the preset temperature of 37°C and the salt concentration of 1.0 M Na⁺ and (0.0 M Mg²⁺) for the folding of a single-stranded DNA. Another DPA implementation, MFOLD 3.1 (3–5), was also used with DNA energy rules in the structure melting simulations, with 0.1 M Na⁺ concentration (and 0.0 M Mg²⁺) (reasons for the differences are explained in the Discussion) at temperatures ranging from 5 to 100°C with 5°C increments (with an exception for 37°C, instead of 35°C).

The analysis of the results was performed with the help of STRUCTURELAB, an RNA/DNA structure analysis computer workbench developed in our laboratory at the NCI-Frederick (6). Stem Trace was the main tool in STRUCTURELAB used for exploration of structural similarities and differences in the solution spaces for the sense and antisense families of sequences (7).

Stem Trace is a two-dimensional plot of all unique helical stems from a solution space of structure conformations of a given sequence, presenting an orthogonal view to the usually employed stem histograms (dot-plots). Displayed information is color-coded to reflect the cumulative frequency of occurrence of stems, and it retains an explicit depiction of all the individual conformations used as inputs to it. As well as providing a visual representation of a secondary structure solution space for a DNA/RNA sequence, Stem Trace is also an active graphical user interface to the underlying data. A set of functions associated with every trace performs searching, sorting, scaling, thresholding and, through connections with other STRUCTURELAB tools, the drawing and labeling of structures based on the data extracted from a specific trace. The sample secondary structures of DNA were drawn from the associated traces. Energy dot-plots (not shown) were also used to complement the Stem Trace analysis.

**RESULTS**

**Capillary electrophoresis**

Figure 1 shows CE-SSCP data in which the migration times for three representative p53 exon 7 point mutations are compared with wild type. As described previously, both strands of each PCR product show distinct shifts in migration times relative to wild type (1,10). This demonstrates the ability of the CE polymer system to separate the conformational differences that exist in these strands due to single point mutations. The two separate DNA strands labeled with JOE and FAM are individually tracked in the CE-SSCP data obtained using the PRISM™ 310. Both strands of each mutant show shifts in the direction and magnitude of migration relative to wild type. In addition, various isofoms (secondary peaks), which indicate semi-stable conformations, are observed, e.g., Namalwa, FAM-labeled sense strand. The data were obtained at 35°C.

A plot of Colo 320 data, obtained as a function of temperature (1), is shown in Figure 2. Shifts in the migration times of the primary peaks relative to wild type are shown for each temperature. As observed previously, the magnitude and direction

Table 1. Human p53 exon 7 mutations

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Type of mutation</th>
<th>Position of mutation</th>
<th>Base pair change</th>
</tr>
</thead>
<tbody>
<tr>
<td>H 596</td>
<td>Point mutation</td>
<td>14 060 (245)</td>
<td>G–T</td>
</tr>
<tr>
<td>Colo 320</td>
<td>Point mutation</td>
<td>14 069 (248)</td>
<td>C–T</td>
</tr>
<tr>
<td>Namalwa</td>
<td>Point mutation</td>
<td>14 070 (248)</td>
<td>G–A</td>
</tr>
</tbody>
</table>

*American Type Culture Collection (Rockville, MD).
*Nucleotide position with respect to GenBank Locus HSP53G, accession number 54156, amino acid position is in parenthesis. Adapted from Atha and co-workers (1,2).
of these shifts are different for the sense and antisense strands. In addition, the magnitude of these shifts is reduced at elevated temperatures until, ≥50°C, the differences relative to wild type are reduced to experimental error.

As shown in Figure 3, CE-SSCP data for the Colo 320 mutation were obtained as a function of temperature. As described previously (1), the data show the ability of the CE to separate conformational differences that exist between complementary strands of single-stranded DNA. The double-stranded DNA, which migrates slightly faster, was used to align the data to correct for the regular changes in polymer viscosity that occur with temperature. Although significant differences were not observed in the electropherogram of the double-stranded DNA, as the temperature of separation was increased from 30 to 45°C, the single strands of DNA showed stepwise increases in migration time until, at 50°C, the difference in migration times for the sense and antisense stands were essentially eliminated.

**Computational analysis**

As we said earlier, the Stem Trace tool produces two-dimensional plots of all the stems from a solution space of structure conformations. Stems are defined by triplets (5’–3’ stem_size), and every unique triplet (predicted stem) is assigned a new position on the y-axis. In case of the raw, unsorted, stem traces, shown in Figure 4, the order of the y-axis entries is a function of the order in which consecutive structures (sets of stems) are fed into the plot. In other words, a new y-axis entry appears only when a unique stem triplet is submitted to the trace, and it is assigned the next available y-value. As a result, the new entries appear to form an ‘envelope’ in the raw (unsorted) stem traces shown in Figure 4. In this study, consecutive structures (i.e. sets of stems) of increasing free energies (more positive) from a DPA-predicted solution space are plotted along the x-axis. Thus, a vertical ‘slice’ through a stem trace corresponds to one structure-worth of stems, whereas the pixels denoting
highly conserved stems (i.e. stems present in all or many of the suboptimal solutions) may form continuous or nearly continuous horizontal lines. Multiple solution spaces generated for each of the mutated sequences are presented side-by-side, separated by the vertical lines. Such are all the stem traces shown in this study (Figs 4–6, 9 and 10). Stem traces combining multiple solution spaces make the direct examination of which stems are strongly preserved across the solution spaces (such as stems 5 and 6, in Figs 5–10) rather straightforward and give immediate visual clues of the differences in the compared solution spaces. To combine the above discussion on the interpretation of the y-axis and x-axis entries in the stem trace plots, let us observe that, as the first solution space plotted is always that of the wild type sequences (sense and antisense), only the stems appearing above the red horizontal lines in Figure 4 can be identified as truly unique to the other solution spaces, relative to that of the wild type. The presence of strongly preserved stems or groups of stems among the unique parts of the solution spaces indicates strong structural motifs distinguishing them from the wild type structures.

It is important to keep in mind, however, that because of the stem-coding scheme, based on exact triplet values, even relatively small differences between stems (1 bp difference in length, for example) are depicted as separate trace entries, which, if not carefully examined, may lead to erroneous conclusions as to diversity in the solution spaces. While the raw (unsorted) stem traces, shown in Figure 4, allow the user to quickly identify new stems (present for the first time in the solution space) added as new y-axis entries, they make it harder to spot the closely related solutions. The 5′-sorted traces, i.e. sorted in the increasing order of the 5′ values in the stem triplets for all the solution spaces plotted in a trace, bring the similar stems closer to each other. We have added annotations to the 5′-sorted stem traces, shown in Figures 5 and 6, to indicate important examples of such stems.

In the case of the DPA solution spaces used in this analysis, the optimal solutions are the leftmost entries in each solution space trace, followed by the suboptimal solutions. Color-coding of individual stems in the stem traces shown is based on the cumulative (for all solution spaces) measure of their frequency. It is worth keeping in mind that stems with a relatively low cumulative frequency, yet strongly preserved in their own solution spaces are of particular interest in the comparisons.

Stem trace results for the sense and antisense sequences
Solution space differences and similarities, illustrated in the stem traces of the sense and antisense sequences, show a general agreement with the CE data (see the CE results in Fig. 1, and the solution space traces in Figs 4–6). In the case of the sense sequences (Figs 4 and 5), we can see a measure of similarity between the solution spaces for the wild type and the Namalwa mutation with enough variability between them to potentially account for the differences in the CE-SSCP results.
relative to the smaller and the larger Namalwa sense peaks (in blue, in Fig. 1). The Namalwa sense stems, highlighted in yellow in Figure 4, occur at a much higher frequency in its solution space than in the wild type, but they are not unique to it. Thus, we would expect some, but not necessarily very radical, structural differences between them and, consequently, resolvable CE migration patterns. The solution spaces of the H 596 and Colo 320 mutations are more distinct from that of the wild type and from each other. They share less stems in their solution spaces with the wild type and have unique, high frequency patterns (clusters of stems), also highlighted in yellow in Figure 4. Such distinct stem trace 'signatures' could indicate clearly distinct CE migration patterns. Thus, the stem trace patterns of the sense solution spaces discussed here generally agree with the distinct migration patterns of both of the mutations relative to the wild type and relative to each other observed in the CE-SSCP.

In the case of the antisense sequences (Figs 4 and 6) the distinguishing differences generally involve fewer unique stems than in the sense solution spaces. These are highlighted in green in Figure 4. All the antisense solution spaces share a core of highly conserved stems clearly visible below the red line in Figure 4 (also, keep in mind the closely related stems, marked on Fig. 6). Despite a fairly large solution space with a few moderately frequent unique stems in the Namalwa antisense case (e.g. the stems highlighted in green in Fig. 4, or stems 3 and 4 in Fig. 6), most of them occur just once. Exact matches between Namalwa and wild type can also be found, thus pointing to a potential agreement with the CE-SSCP results. Similarly, the solution spaces for H 596 and Colo 320 antisense mutations are dominated by the stems conserved across all the solution spaces. The strongly preserved unique ('signature') stems, highlighted in green in Figure 4 and marked as 3 and 4M (H 596) and 3M and 4 (Colo 320) in Figure 6, distinguish the two solution spaces from that of the antisense wild type. These stems may be responsible for the migration pattern differences indicated by the CE-SSCP.

Figure 5. 5'-Sorted stem trace of the sense sequences solution spaces produced by RNAstructure 3.5 DPA, run with DNA energy rules at 37°C and with salt concentration of 1.0 M Na⁺ and 0.0 M Mg²⁺. Secondary structures, illustrated in other figures and labeled based on this stem trace’s color code, are marked with the dotted, white, vertical lines and arrows at the bottom (vertical collections of stems correspond to all the stems present in one secondary structure). Stems 3–6 are marked as references to the associated secondary structure drawings. Multiple-linked black arrows point to examples of closely related stems differing in length by 1 bp. Color-linked arrows point to resulting linear substructure combinations differing only in the placement of a single nucleotide (G) bulge loop at position 129 or 130.
Again, we can see a general agreement between the stem trace patterns for the antisense solution spaces and the CE data.

**Secondary structure comparisons for the sense sequences**

Wild type versus Namalwa (G to A at nucleotide position 85—relative to the 5′ end of the fragment). The temperature-dependent electropherogram for 35°C (Fig. 1), a condition closest to that used by the RNAstructure 3.5 program, shows a near overlap of the smaller of the two associated peaks in the Namalwa graph (blue) with the wild type peak (blue, panel below). The larger of the two Namalwa peaks (blue) clearly lags behind the wild type, as it does in the electropherograms for other temperatures. The double peak in the Namalwa graph may suggest a coexistence of two isoforms (sense strand) in the CE solution.

When we compare this data against the secondary structure solution spaces for the wild type and Namalwa sense sequences, we can see that the Namalwa space contains two solutions, the first and the sixth suboptimal (Fig. 4), that are similar to each other and distinct from the majority of the solutions in the Namalwa solution space. We can also see that the optimal wild type and the optimal Namalwa structures are
comprised of the same stems, the only difference being the mutated base (G to A at position 85) inside an internal loop. This could be a match for the nearly overlapping wild type peak and the smaller Namalwa peak. However, it is also possible that a larger structural difference should be sought to explain the small difference between the discussed peaks. As can been seen in the same electropherogram, the near overlap of the sense peaks we are discussing here is not as good as in the case of wild type and Namalwa antisense peaks (in green in Fig. 1). In this case (antisense), we can find two nearly identical topologies differing by a nucleotide (second in the wild type solution space and first in Namalwa solution space, as marked in Fig. 6, secondary structures shown in Fig. 8A and D). The sixth suboptimal structure in the Namalwa sense solution space is a potential candidate, with one less stem overall and one alternative stem relative to the optimal solution.
Either one of the two Namalwa sense solutions fits within the constraints of the CE data. On the other hand, the dominant Namalwa sense peak may correspond to a dominant secondary structure containing the most frequent stems, exemplified in our drawings by the second suboptimal Namalwa structure shown in Figure 7E.

Wild type versus H 596 (G to T at nucleotide position 75) and Colo 320 (C to T at nucleotide position 84). The temperature-dependent electropherogram for 35°C (Fig. 1) indicates a distinct migration pattern in which the H 596 sense (blue) migrates slightly slower than the wild type sense (in blue in the bottom panel). The same electropherogram shows very clear results for the wild type sense and Colo 320 sense (blue), with the latter migrating much faster than the wild type. The data point towards mutation conformations distinct from the wild type and from each other.

The key differences between the mutation H 596, the wild type and Colo 320 are the effects of the mutations on stems 3 and 4 (marked in Fig. 5). Mutation H 596 extends stem 4 by 1–4 bp and tightens the hairpin loop associated with it (shrinking its size from 7 to 5 nt). A representative structure, the optimal H 596 sense solution, was selected because it contains the most frequent stems (as shown in Figure 7B). In the majority of its solutions, mutation Colo 320 leads to an extension of stem 3 by 2 bp to the total of 6 bp in length, and creation of a stable stem 3M. Stem 4 is 3 bp long and encloses the 7 nt hairpin loop. Overall, however, the representative high frequency Colo 320 structure topology is similar to that of H 596, as can be seen in comparison of stems comprising the second suboptimal solutions for both of these mutations (see Fig. 4). Hence, the problem with the two representative structures mentioned above is that they do not appear to be sufficiently different from each other. Comparable levels of differences in the secondary structures are predicted for the antisense H 596 and Colo 320 mutations (see Discussion and Fig. 8B and C), but their CE-SSCP results do not show such radical migration speed differences. Therefore, we propose another conformation, present in four out of 13 suboptimal solutions of the Colo 320, and represented by the fifth suboptimal solution, marked in Figure 5 and shown in Figure 7C. Its topology is quite different from that of the wild type and the H 596 mutation, as it disrupts the high frequency stems 3, 4, 5 and 6. Perhaps such a radically different conformation could better explain the CE-SSCP results. In general, however, it is much harder to select specific secondary structures based solely on the relative dissimilarities.
Structure comparisons for the antisense sequences

Wild type versus Namalwa (C to T at nucleotide position 55—relative to the 5′ end of the fragment). The temperature-dependent electropherogram for 35°C shows an almost complete overlap of peaks of the wild type and the Namalwa antisense structures (green peaks in Fig. 1). Using this result as the guiding constraint, we have selected the second suboptimal structure in the wild type antisense solution space and the optimal structure in the Namalwa antisense solution space as the best matching representatives, marked in Figure 6 and shown in Figure 8A and D.

Wild type versus H 596 (C to A at nucleotide position 65) and Colo 320 (G to A at nucleotide position 56). Relative to the wild type peak in the CE-SSCP data, the H 596 and Colo 320 antisense peaks indicate a clear, although not very strong, pattern of trailing the wild type (green peaks in Fig. 1). The data suggest mutation conformations distinct from the wild type, but relatively similar to each other. As in the case of their sense counterparts, the key stems separating mutations H 596 and Colo 320 antisense from the antisense wild type are stems 3 and 4 and their mutated versions (marked in Fig. 6). The effects of mutations are analogous to those described for the sense sequences. We have selected two representative secondary structures, shown in Figure 8B and C, which contain the best preserved stems overall, as well as the best preserved stems in their respective solution spaces. Despite the differences associated with stems 3, 3M, 4 and 4M, their overall topologies are relatively close to each other and not very dissimilar from that of the selected wild type antisense structure. As such, they seem to agree well with the CE-SSCP data.

DISCUSSION

We have used DNA-folding analysis to compare data obtained by CE-SSCP to gain insight into the mechanism of SSCP and to determine to what extent it can be helpful in the prediction of electrophoretic patterns obtained by this method. We have found that several properties exhibited in the electropherograms using CE-SSCP can be explained using DNA-folding analysis. These include shifts in migration time, the appearance of secondary peaks and the temperature dependence of the electrophoretic patterns, which vary depending on the mutation.

In general, we have treated our computational analysis as a problem of multiple constraints satisfaction, the constraints being provided by the CE electropherograms and the predicted secondary structure solution spaces. We have used the CE data plots as the primary, guiding constraints, and tried to find matching solutions within the top 10%, energetically speaking, of the secondary structure solution spaces. Given indications from the literature that the optimal solutions produced by DPA algorithms (attempting to find a structure with the lowest free energy) do not always correspond to biologically viable structures, we did not automatically give more credence to the optimal solutions (11–14). Without ignoring the optimals as potential candidates satisfying the CE constraints, we have been looking for the dominant representatives, i.e. structures with the stems occurring most frequently in the solution spaces.

As there are some differences in the implementation of the solution space sampling (tracebacks) and energy rules between RNAstructure 3.5 and MFOLD 3.1, we have used both of them to verify persistence of certain key stems, marked on the selected drawings. It is important to note that when MFOLD 3.1 is applied to DNA, its NEWTEMP module is used to generate the energy tables for the specified temperature and salt concentration ‘on the fly’, from the reference tables (for the default conditions), based on the older MFOLD 2.3 rules. RNAstructure 3.5 utilizes the latest publicly available DNA energy rules for the default conditions (8), but it does not permit the prediction of secondary structures under a range of temperatures and salt concentrations.

The folding results based on the MFOLD 2.3 energy rules yielded the same highly conserved stems 3, 3M, 4, 4M, 5 and 6, as did RNAstructure 3.5. However, the MFOLD 2.3 results were substantially different in the overall topologies and were harder, if at all possible, to fit within the SSCP constraints. For example, MFOLD 2.3 solution spaces are totally identical for the wild type sense and Namalwa sense sequences under a range of temperatures, whereas with RNAstructure 3.5 this is not the case. At the same time many of the simulated melting experiments that we have performed for all the mutations and the wild type, based on MFOLD 2.3 predictions, agree well with the CE-SSCP-based graphs of the relative mutation speed differences and their variability over a range of temperatures. We say this based on the assumption that, as the secondary structures gradually open up with the rise in the temperature, and the wild type and the mutated structures begin to resemble each other more closely, their migration speeds converge. In Figure 9 we present a stem trace illustrating one of the studied cases, comparing the wild type sense and Colo 320 sense solution spaces between 30 and 70°C. We have used a lower salt concentration (0.1 M Na+) to bring the predicted ‘melting’ (i.e. opening up of the structures) into agreement with the CE-SSCP data. As can be seen in the plot of migration time differences relative to the wild type (Fig. 2), the major drop in the speed of Colo 320 occurs between 45 and 50°C. The stem trace shows a very stable solution space with little variability, when compared with the wild type, for Colo 320 up to 45°C. Above that temperature there is a dramatic drop in the number of stems and an increase in the variability of the solution spaces. The wild type, on the other hand, appears to lose its stability (i.e. persistence of stems in the solution space) at ~37°C. These very striking differences in the variability of the solution spaces of wild type and Colo 320 at 37 and 40°C, correspond to the highest relative speed differences in the CE scans (Fig. 2). Thus, on top of the topological differences, thermal stability of the structure may play a role in the CE-SSCP results. Figure 10 shows a stem trace comparing the wild type and Colo 320 antisense solution spaces predicted for the same range of temperatures and for the same salt concentrations. In this case, the solution spaces show smaller differences in their variability and thermal stability, with respect to each other, than in the case of the sense sequences. A slightly higher stability of the Colo 320 antisense structures persists until 40–45°C, and then at 60°C its melting pattern converges with that of the wild type antisense. The CE-SSCP data also indicate much smaller differences in the relative migration speeds between the antisense wild type and Colo 320.
diminishing further >45°C, in comparison with those measured for their sense equivalents.

All three point mutations produced conformational differences both in CE data and modeling results. Apparently, each of these mutations can have a stabilizing or de-stabilizing effect on semi-stable regions in the structure. These conformational changes result in changes in the average speed at which the DNA strands can pass through the molecular sieve polymer used in CE. We have not yet been able to correlate the magnitude of this change in speed (shift in migration time relative to wild type) with the conformational changes we observe using computerized modeling. Based on our comparison of CE and

Figure 9. 5′-Sorted stem trace of the wild type sense and Colo 320 sense solution spaces produced by an MFOLD 3.1 DPA, run with DNA energy rules (corresponding to an older version 2.3), with salt concentrations of 0.1 M Na⁺ and 0.0 M Mg²⁺. Presented solution spaces were calculated for a range of temperatures from 30 to 70°C in 5°C increments (with an exception for 37°C, instead of 35°C). Stems 1 and 2 are unique to MFOLD’s solution spaces. Stems 3, 3M, 4, 5 and 6 are the same as those predicted by RNAstructure 3.5. Melting of the structures can be seen as the increasing sparseness of the stem traces with increasing temperature.

Figure 10. 5′-Sorted stem trace of the wild type antisense and Colo 320 antisense solution spaces produced by an MFOLD 3.1 DPA, run with DNA energy rules (corresponding to an older version 2.3), with salt concentrations of 0.1 M Na⁺ and 0.0 M Mg²⁺. Presented solution spaces were calculated for a range of temperatures from 30 to 70°C in 5°C increments (with an exception for 37°C, instead of 35°C). Stems 1 and 2 are unique to MFOLD’s solution spaces. Stems 3 (wild type, light green), 3M (Colo 320, dark green), 4, 5 and 6 are the same as those predicted by RNAstructure 3.5. Melting of the structures can be seen as the increasing sparseness of the stem traces with increasing temperature.
modeling results, the speed of CE migration is usually very sensitive to minor changes in conformation. Apparently, even minor substructures can act as ‘hooks’ or ‘drags’ which have a large but unpredictable effect on migration rate. On the other hand, the relatively large changes that we observe by modeling are not always reflected by large differences in migration time with respect to the wild type, which can be seen in the case of the wild type sense and the H 596 sense structures (Fig. 7A and B) and the related electropherograms (Fig. 1). Apparently some of these conformational effects average out to produce little net effect on the migration time. Secondary peaks are observed in the CE electropherograms due to isoform structures, which vary with temperature. As discussed, we have also observed suboptimal conformations in our modeling, which are similarly temperature dependent. These isoforms, as well as the differences in conformation between the sense and antisense strands are reduced at elevated temperatures in both the modeling and the experimental CE data.

The measurements by CE and analysis by molecular modeling produce an insight into how solution conditions could be optimized to maximize the SSCP effect (i.e. temperature). By modeling we have observed the high variability that is seen experimentally in the detectability of point mutations using SSCP, although we cannot yet predict the magnitude and direction of these changes in migration based on the relative structural differences. An examination of the effect of the position of the point mutation, using additional mutations, would yield a better insight into the mechanism of SSCP. With such information, primers could be designed at alternative sequences to amplify forward and reverse strand products with improved shifts in migration time. Insights into the effects of these conformational changes could be obtained by three-dimensional structural analysis. However, given the structural variability in the two-dimensional solution spaces and the complexities of three-dimensional modeling, correlation of our CE experimental data with three-dimensional DNA structures goes beyond the scope of this study. On the other hand, the use of structure-specific probes such as the 5′-nuclease TaqExo used to determine DNA secondary structures in the *Mycobacterium tuberculosis* katG gene and hepatitis C (HCV) cDNA would be very helpful in such an analysis (14).

ACKNOWLEDGEMENTS

We thank Michael Wenz for his help and expertise in CE measurements and Joe Hubbard for helpful discussions and critical reading of this manuscript. We thank John Owens for his help with preparation of the figures. This publication has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000.

REFERENCES