Identification of Base Pairs in Single-Nucleotide Polymorphisms by MutS Protein-Mediated Capillary Electrophoresis

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2006

Impact Factor: 5.64 · DOI: 10.1021/ac0520386 · Source: PubMed

2 AUTHORS:

Andrei P Drabovich
University of Toronto, University Health Net…
36 PUBLICATIONS **978** CITATIONS
SEE PROFILE

Sergey N Krylov
York University
167 PUBLICATIONS **3,624** CITATIONS
SEE PROFILE

Anal. Chem. **2006,** *78,* 2035−2038

Identification of Base Pairs in Single-Nucleotide Polymorphisms by MutS Protein-Mediated Capillary Electrophoresis

Andrei P. Drabovich and Sergey N. Krylov*

Department of Chemistry, York University, Toronto, Ontario M3J 1P3, Canada

Single-nucleotide polymorphisms (SNPs) are widespread genomic variations, which are associated with serious health disorders and drug resistance. Multiple clinical applications and studies of global population genetics require fast and informative analysis of SNPs. Most of conventional methods sense the presence of the SNP but cannot identify the base pair in it. Here we report simple identification of base pairs in SNPs without DNA sequencing. Our approach is based on the unique ability of MutS protein to bind different single-nucleotide mismatches in DNA with different affinities. Conceptually, the DNA in question is mixed with reference DNA, melted, and reannealed. If the DNA in question has an SNP, the products of reannealing will have two different single-nucleotide mismatches, which provide a base-pair-specific signature of the SNP. The products of reannealing are mixed with MutS, equilibrated, and separated by equilibrium capillary electrophoresis of equilibrium mixtures with MutS in the run buffer. The pattern of migration times of DNAs with mismatches is used for unequivocal identification of the base pair in the SNP. In addition to its ability to identify base pairs in SNPs, the new analytical approach is fast, simple, highly sensitive, and requires no quantitation. It will find applications in studies of heterogeneity of base pairs in known SNPs in large human populations.

Single-nucleotide polymorphism (SNP) is an abundant form of genomic variation, which involves the variability of a single base pair with a frequency of the least copious allele of no less than 1% of the entire population.[1,2] SNPs have been found to cause health disorders and drug resistance.[3,4] Fast and informative analysis of SNPs is important for clinical applications and studies of global population genetics. SNP mapping is considered to be the next logical step in the human genome project.[5−8]

Conventional SNP assays utilize two major principles to reveal polymorphisms: (i) finding the sequence or another physical-chemical property of single-stranded DNA and (ii) detecting mismatches in double-stranded DNA. The first principle is realized in direct DNA sequencing,[9] single-strand conformation polymorphism,[10] and mass spectrometry.[11] In the second approach, the DNA in question is mixed with a reference allele, melted, and reannealed. If the SNP is present, the reannealed DNA will contain single-nucleotide mismatches. Conventional methods to detect mismatches include the following: (i) electrochemical analyses,[12,13] (ii) separation methods, such as denaturing gradient gel electrophoresis, temperature gradient gel electrophoresis, and denaturating HPLC,[14] (iii) mismatch modifying methods such as restriction fragment length polymorphism,[15,16] and (iv) methods using fluorescently labeled hybridization probes specific for a certain mismatch.[17,18] The published mismatch-based methods cannot, however, easily and reliably distinguish between all sets of single-nucleotide mismatches formed with six possible SNP variants.

Here, we report the ability to distinguish all six sets of single-nucleotide mismatches and, therefore, to identify the exact base pair in the SNP without DNA sequencing. Our approach is based on the unique ability of MutS (DNA mismatch binding protein from *Thermus aquaticus*) to bind different mismatches with different affinities.[19,20] MutS is a part of the DNA repair machinery. It tightly binds double-stranded DNA with single-nucleotide mismatches and has very weak affinity to DNA without mismatches.

This work was inspired by the insight that the ability of MutS to bind different mismatches with different affinities could be used

* To whom correspondence should be addressed. E-mail: skrylov@yorku.ca.

(1) Brookes, A. J. *Gene* **1999**, *234*, 177−186.
(2) Crawford, D. C.; Nickerson, D. A. *Annu. Rev. Med.* **2005**, *56*, 303−320.
(3) Craig, D. W.; Stephan, D. A. *Expert Rev. Mol. Diagn.* **2005**, *5*, 159−170.
(4) McCarthy, J. J.; Hilfiker, R. *Nat. Biotechnol.* **2000**, *18*, 505−508.
(5) Miller, R. D.; Phillips, M. S.; Jo, I.; et al. *Genomics* **2005**, *86*, 117−126.
(6) Syvanen, A. C. *Nat. Genet.* **2005**, *37*, S5−10.
(7) Hinds, D. A.; Stuve, L. L.; Nilsen, G. B.; Halperin, E.; Eskin, E.; Ballinger, D. G.; Frazer, K. A.; Cox, D. R. *Science* **2005**, *307*, 1072−1079.
(8) Syvänen, A.-C.; Taylor, G. *Hum. Mut.* **2004**, *23*, 401−405.
(9) Altshuler, D.; Pollara, V. J.; Cowles, C. R.; Van Etten, W. J.; Baldwin, J.; Linton, L.; Lander, E. S. *Nature* **2000**, *407*, 513−516.
(10) Andersen, P. S.; Jespersgaard, C.; Vuust, J.; Christiansen, M.; Larsen, L. A. *Hum. Mutat.* **2003**, *21*, 455−465.
(11) Griffin, T. J.; Smith, L. M. *Trends Biotechnol.* **2000**, *18*, 77−84.
(12) Patolsky, F.; Lichtenstein, A.; Willner, I. *Nat. Biotechnol.* **2001**, *19*, 253−257.
(13) Deng, J. Y.; Zhang, X. E.; Mang, Y.; Zhang, Z. P.; Zhou, Y. F.; Liu, Q.; Lu, H. B.; Fu, Z. J. *Biosens. Bioelectron.* **2004**, *19*, 1277−1283.
(14) Petersen, D. C.; Laten, A.; Zeier, M. D.; Grimwood, A.; Rensburg, E. J.; Hayes, V. M. *Hum. Mutat.* **2002**, *20*, 253−259.
(15) Till, B. J.; Burtner, C.; Comai, L.; Henikoff, S. *Nucleic Acids Res.* **2004**, *32*, 2632−41.
(16) Nollau, P.; Wagener, C. *Clin. Chem.* **1997**, *43*, 1114−1128.
(17) Livak, K. J. *Genet. Anal.-Biomol. Eng.* **1999**, *14*, 143−149.
(18) Pont-Kingdon, G.; Lyon, E. *Nucleic Acids Res.* **2005**, *33* (10), e89.
(19) Whitehouse, A.; Deeble, J.; Parmar, R.; Taylor, G. R.; Markham, A. F.; Meredith, D. M. *Biochem. Biophys. Res. Commun.* **1997**, *233*, 834−837.
(20) Brown, J.; Brown, T.; Fox, K. R. *Biochem. J.* **2001**, *354*, 627−633.

10.1021/ac0520386 CCC: $33.50
Published on Web 01/25/2006

for distinguishing mismatches in the analysis of SNPs. MutS was previously used for SNP analyses by classical electrophoretic mobility shift assays (EMSAs) in slab gel electrophoresis.[21−23] Classical EMSA is conducted under nonequilibrium conditions—the electrophoresis run buffer does not contain DNA or protein. EMSA cannot separate MutS−DNA complexes with different mismatches in DNA, because all these complexes have identical charge-to-size ratios. To facilitate the distinction between different mismatches in this work, we used gel-free capillary electrophoresis (CE) under quasi-equilibrium conditions for MutS−DNA interaction. Quasi-equilibrium was achieved through adding MutS to the electrophoresis run buffer.

Gel-free CE was proven to be an excellent platform for studying protein−DNA interactions.[24] However, separation of DNA molecules in gel-free electrophoresis is difficult because charge-to-size ratios of different DNA molecules are similar. We recently proposed the use of DNA-binding proteins in CE instead of gels to facilitate highly efficient separation of DNA molecules. Single-stranded DNA binding protein (SSB) was found to be very useful in quantitative hybridization analyses of DNA and RNA.[25,26] We also used SSB in a CE-based analysis of PCR products.[27] In the current work, we further support this general idea by applying MutS protein to the identification of base pairs in SNPs. The differences in affinities between MutS and DNA with different mismatches are monitored in our case with equilibrium capillary electrophoresis of equilibrium mixtures (ECEEM).[28,29] In addition to its ability to identify base pairs in SNPs, the new analytical approach is simple, fast, and sensitive enough to be used as a tool for routine SNP analysis. While it can hardly be used for SNP discovery, the new method can facilitate highly efficient scoring of the occurrence rate of base pairs in known SNPs in large human populations.

EXPERIMENTAL SECTION

Chemicals and Materials. Thermostable DNA mismatch binding protein (MutS) from *T. aquaticus* was purchased from InterSciences (Markham, ON, Canada). DNA molecules (41 bases) were custom-synthesized by IDT (Coralville, IA). Four forward DNA strands were labeled with a 6-carboxyfluorescein functional group at the 5′-end. The structures of double-stranded DNA homo- and heteroduplexes are shown in Table 1. Fused-silica capillaries were purchased from Polymicro (Phoenix, AZ). Buffer components were obtained from Sigma-Aldrich (Oakville, ON, Canada). All solutions were made using the Milli-Q-quality deionized water and filtered through a 0.22-μm filter (Millipore, Nepean, ON, Canada).

Table 1. Formation of Matching and Mismatching Sequences for SNP Analysis Using the Reference Sequence with the A−T Base as an Example

SNP status	Sequence mixed with the reference	Sequences after mixing, melting, and re-annealing of the sequence in question with the reference sequence	
		Matching sequences	Mismatching sequences
No SNP	5′—A—3′ 3′—T—5′	5′—A—3′ 3′—T—5′	
SNP 1	5′—T—3′ 3′—A—5′	5′—A—3′ 5′—T—3′ 3′—T—5′ 3′—A—5′	5′—A—3′ 5′—T—3′ 3′—A—5′ 3′—T—5′
SNP 2	5′—C—3′ 3′—G—5′	5′—A—3′ 5′—C—3′ 3′—T—5′ 3′—G—5′	5′—A—3′ 5′—C—3′ 3′—G—5′ 3′—T—5′
SNP 3	5′—G—3′ 3′—C—5′	5′—A—3′ 5′—G—3′ 3′—T—5′ 3′—C—5′	5′—A—3′ 5′—G—3′ 3′—C—5′ 3′—T—5′

Double-Stranded DNA Preparation. The general sequence of the forward strands was 6-carboxyfluorescein-5′-CTTCTGCCCGCCTCCTTCCTTCCAACCTTCATC**X**GCCACCC-3′ with X being one of the four bases. These strands (final concentration of 500 nM) were mixed with a slight molar excess of the respective reverse strands in the buffer containing 10 mM Tris-HCl (pH 8.0) and 50 mM NaCl. Duplexes were annealed by slowly cooling from 95 to 20 °C. The uniformity of duplexes was verified by CE.

Capillary Electrophoresis. A P/ACE MDQ instrument (Beckman-Coulter, Fullerton, CA) was used for CE analyses. The apparatus employed laser-induced fluorescence detection with a 488-nm line of an argon ion laser for fluorescence excitation. An uncoated fused-silica capillary was used with the following dimensions: 50 cm length × 75 μm i.d. × 350 μm o.d. The length from the injection end to the detection window was 40 cm. The capillary was maintained at 37 ± 0.1 °C. Electrophoresis was run in 50 mM Tris−acetate (pH 8.2) with 2 mM $MgCl_2$ buffer by an electric field of 600 V/cm with positive polarity at the injection end. The samples were injected into the capillary by a pressure pulse of 5 s × 0.4 psi (2.8 kPa); the length of the corresponding sample plug was 16 mm. The capillary was rinsed with the run buffer solution for 1 min prior to each run. At the end of each run, the capillary was rinsed with 100 mM NaOH for 1 min, followed by a rinse with deionized water for 1 min.

Equilibrium Mixture and ECEEM Separation. The final concentrations of MutS dimers and DNA duplexes were 140 and 40 nM, respectively, in electrophoresis run buffer. Every mixture was incubated for 10 min at 37 ± 0.1 °C and then subjected to ECEEM with 140 nM MutS in the electrophoresis run buffer.

RESULTS AND DISCUSSION

In Table 1, we show the example of mismatch formation in SNP analysis. If the reference sequence is mixed, melted, and reannealed with the sequence in question, four outputs are possible. If the sequence in question is identical to the reference sequence, which corresponds to the No-SNP case, the result will be a single matching sequence. If the sequence in question has a different base pair, the result will be two matching sequences and two sequences with single-nucleotide mismatches. Three SNPs (SNP 1, SNP 2, SNP 3) are possible, depending on the base pair in the examined sequence. To identify this base pair, the mismatches are distinguished by CE separation in the presence of MutS protein, which mediates the mobility of different DNA to

(21) Stanislawska-Sachadyn, A.; Paszko Z.; Kluska, A.; Skasko, E., Sromek, M.; Balabas, A.; Janiec-Jankowska, A.; Wisniewska, A.; Kur, J.; Sachadyn, P. *Mutat. Res.* **2005**, *570*, 97−103.
(22) Lishanski, A.; Ostrander, E. A.; Rine, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2674−2678.
(23) Li, C. Z.; Long, Y. T.; Lee, J. S.; Kraatz, H. B. *Chem. Commun.* **2004**, *5*, 574−575.
(24) Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2002**, *124*, 13674−13675.
(25) Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2003**, *125*, 13451−13454.
(26) Al-Mahrouki, A. A.; Krylov, S. N. *Anal. Chem.* **2005**, *77*, 8027−8030.
(27) Drabovich, A.; Krylov, S. N. *J. Chromatogr., A* **2004**, *1051*, 171−175.
(28) Chu, Y. H.; Avila, L. Z.; Biebuyck, H. A.; Whitesides, G. M. *J. Med. Chem.* **1992**, *35*, 2915−2917.
(29) Drabovich, A.; Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 11224−11225.

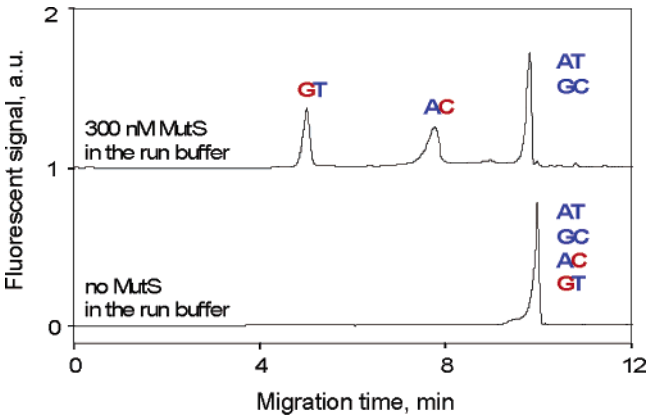

Figure 1. DNA duplexes with mismatches (GT and AC) separated from complimentary DNA (AT and GC) in the presence of 300 nM MutS in the run buffer (top curve). No separation occurs in the absence of MutS (bottom curve).

different extents. When MutS is not involved in the analysis, a single peak is observed for homo- and heteroduplexes (Figure 1, lower trace). When MutS is involved in analysis, however, heteroduplexes with different mismatches are clearly distinguishable (Figure 1, upper trace).

The concept of using MutS for base-pair identification in SNPs was proven experimentally with a design identical to that depicted in Table 1. Four fluorescently labeled DNA sequences shown in column 2 can be practically obtained by PCR amplification with one of the primers labeled with 6-carboxyfluorescein at the 5′-end. In this work, we used chemically synthesized and labeled DNA instead. By annealing single DNA strands, we obtained four short double-stranded (ds) DNA sequences with forward strands fluorescently labeled. The reference sequence (blue) of dsDNA was mixed with itself and every one of the three sequences in question (red). The temperature was first raised to melt DNA and then lowered to reanneal it. The resulting four mixtures contained sequences shown in columns 3 and 4 of the table. The No-SNP mixture contained only the sequence with the AT match. The mixture for SNP 1, SNP 2, and SNP 3 contained identical AT and TA matching sequences but different mismatches: AA and TT for SNP 1, AG and CT for SNP 2, and AC and GT for SNP 3. MutS was added to the four mixtures and equilibrated with DNA.

The four mixtures were then analyzed by CE with fluorescence detection. The resulting four electropherograms are shown in Figure 2. As we expected, the No−SNP mixture generated a single major peak, while SNPs resulted in a maximum of three major peaks. All four electropherograms also contained a minor peak with a migration time of ~4.3 min. The major peaks were assigned to specific mismatches by comparing their migration times with those of six individual DNAs, which had single base-pair mismatches (see the rightmost column in Table 1). All mismatched DNA had different migration times, thus, leading to unique patterns for SNPs with different base pairs. The minor peak at 4.3 min was observed for the analysis of MutS−DNA equilibrium mixtures independent of the presence of mismatches in DNA and the presence of MutS in the run buffer. It was not observed, however, for MutS or DNA sampled separately and analyzed in the MutS-free run buffer. These results suggest that the minor peak corresponds to a MutS−DNA complex with low affinity (high K_d) and low dissociation rate constant. This hypothesis is supported by a previously reported observation of nonspecific low-affinity binding of MutS to homoduplex DNA.[30,31] Since the migration time of the minor peak is identical for the four electropherograms, it can serve as a left internal reference. The rightmost peak of matching DNA can serve as the right internal reference. The two references facilitate unequivocal discrimination between the patterns of the two peaks corresponding to DNA with mismatches.

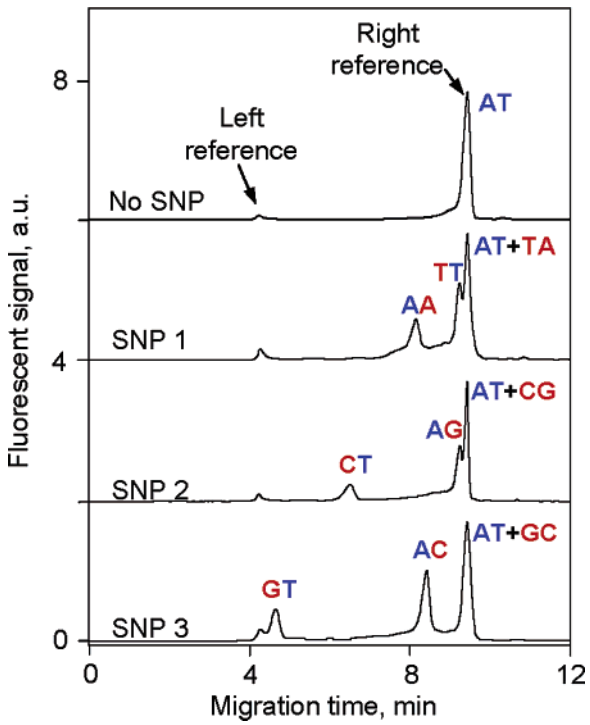

Figure 2. Base-pair specific analysis of three SNPs for the AT base pair in the reference sequence. Sets of mismatches are identical to those depicted in Table 1. Concentration of MutS in equilibrium mixture and electrophoresis buffer was 140 nM.

To prove that the method is generic, we have conducted similar experiments for the remaining three combinations of mismatches (Figure 3). Clearly distinct patterns of migration times of peaks were observed for the three sets of mismatches under conditions identical to those described above. Thus, we proved that MutS could facilitate generic base-pair-specific analysis of SNPs.

ECEEM was used in this work for the identification of base pairs in SNPs.[28,29] ECEEM is a kinetic capillary electrophoresis method,[32] in which quasi-equilibrium between interacting species (MutS and DNA in this case) is maintained during the separation. Conceptually, the equilibrium mixture of MutS with matching and mismatching sequences is prepared and injected into the capillary. To maintain dynamic quasi-equilibrium between MutS and DNA during electrophoresis, MutS is also added to the run buffer, and capillary is prefilled with this buffer prior to each run. During the separation, DNA resides a fraction of time within the complex with MutS and a fraction of time as a free molecule. The higher the affinity of DNA to MutS (lower K_d), the more time DNA

(30) Wang, H.; Yang, Y.; Schofield, M. J.; Du, C.; Fridman, Y.; Lee, S. D.; Larson, E. D.; Drummond, J. T.; Alani, E.; Hsieh, P.; Erie, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14822−14827.
(31) Sixma, T. K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 47−52.
(32) Petrov, A.; Okhonin, V.; Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 17104−17110.

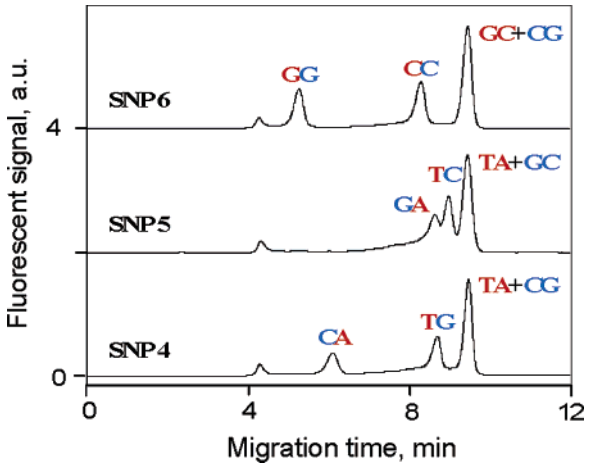

Figure 3. MutS-based analysis of three additional sets of mismatches (compare with Figure 2). The six sets of mismatches shown in Figures 2 and 3 cover all possible variations of mismatches. Experimental conditions are identical to those used in Figure 2.

resides within the complex and the closer its migration time is to that of the complex, $t_{MutS\cdot DNA}$. In contrast, the lower the affinity (higher K_d), the more time DNA spends as a free molecule and the closer its migration time is to that of free DNA, t_{DNA}. The observed migration time of DNA in ECEEM depends on K_d and the concentration of MutS, [MutS], in the following way:

$$\frac{1}{t} = \frac{1}{t_{MutS\cdot DNA}} \frac{[MutS]}{[MutS] + K_d} + \frac{1}{t_{DNA}} \frac{K_d}{[MutS] + K_d}$$

In other words, a mismatch with higher affinity to MutS experiences a greater mobility shift than a mismatch with lower affinity. It has been previously reported that affinity of MutS to dsDNA ranged from 0.19 μM for heteroduplexes with the GT mismatch to 15 μM for DNA without mismatches.[20] Because of the large range of affinities, in No-SNP cases, a single peak is observed, while SNP presence gives rise to three peaks: a single peak for two matching sequences and two peaks for two mismatching sequences. While migration times of all matching sequences must be similar, migration times of two peaks of mismatching sequences can differ. The unique feature of the method is that it relies on pattern analysis of two peaks to identify each SNP type. For example, migration times of DNA with TT and AG mismatches are very close; however, the peaks of DNA with AA and CT mismatches allow unambiguous identification of the exact type of SNP (Figure 2). The concentration of MutS, for which patterns of peaks are distinct for SNPs with different base pairs, can be found experimentally. Such patterns can be then used for routine base-pair-specific analyses of SNPs.

We estimated the limit of detection (LOD) of our method. The mass LOD was 5 million molecules while the concentration LOD was 1 nM. It should be noted, however, that in SNP analysis by our method, LOD is not a critical issue as all samples are amplified by polymerase chain reaction (PCR) to levels that exceed the LOD by at least 2 orders of magnitude. In addition, the DNA in question and the reference DNA are mixed in similar amounts to ensure that the peaks have comparable areas. Nevertheless, the issues of LOD and dynamic range may become important in analyses of SNP in tumor tissues, where both normal and mutated cells are present but at different ratios.

Our method is based purely on the pattern analysis of migration times of peaks. This approach provides us with two important advantages. First, no quantitative analysis of the peaks is required. Second, it is not imperative that all peaks be separated. The analysis of the migration time patterns requires two reference peaks, whose migration times do not change between different SNPs. The MutS−DNA system itself provides us with two ideal references, left reference and right reference. The two references frame the time window, in which the migration times of the peaks of mismatched DNA can vary. It is known that the K_d value for the interaction of MutS with DNA could be dependent not only on a type of mismatch but also on the position of the mismatch in the sequence and potentially on the sequence in the vicinity of the mismatch.[33] Affinities may also depend on separation conditions and lengths of DNA fragments. At the extreme, the order of affinities for mismatches may change. There were reports of the CC mismatch having the lowest affinity to MutS,[20,32] while another work identified the AA mismatch as the one with the lowest affinity.[19] Our results suggest that under our conditions MutS had the weakest affinity for TT and AG mismatches. Moreover, the affinity of MutS to the GA mismatch was different from that to the AG mismatch, confirming that affinity depends on the sequence of DNA.[33] All these facts suggest that the concentration of MutS, which allows the best distinction between the peak patterns, may slightly vary from DNA to DNA. This parameter, therefore, can be used to optimize the pattern distinction.

The method in its current form is applicable to DNA fragments with single SNPs only. This limitation can, however, be easily overcome as all DNA is amplified in PCR prior to the analysis and can be made as short as we need. The gene with multiple SNPs can be, thus, split into fragments with single SNPs and analyzed using our approach. Short DNA fragments also help to separate complexes of MutS with heteroduplexes from those with homoduplexes. However, our data and other sources[21] suggest that longer DNA (120 and 500 base pairs, respectively) can be analyzed as well.

To conclude, this work proves the feasibility of reliably and easily distinguishing base pairs in SNPs without DNA sequencing. The approach is based on the unique property of the MutS protein to bind different mismatches with different affinities. Despite a great variation between the K_d values for different mismatches, a single concentration of MutS can facilitate unequivocal identification of base pairs in SNPs. As long as separation conditions are optimized for a specific SNP, they can be used without changes for fast scoring of this SNP in large human populations (>10 000 samples).

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Council of Canada.

Received for review November 16, 2005. Accepted December 13, 2005.

AC0520386

(33) Joshi, A.; Rao, B. J. *J. Biosci.* **2001**, *26*, 595−606.