Application of mismatch detection methods in DNA computing

Henkel, C.V.; Rozenberg, G.; Spaink, H.P.; Ferretti, C.; Mauri, G.; Zandron, C.

Application of Mismatch Detection Methods in DNA Computing

Christiaan V. Henkel[1,2], Grzegorz Rozenberg[2], and Herman P. Spaink[1]

[1] Institute of Biology, Leiden University,
Wassenaarseweg 64, 2333 AL Leiden, The Netherlands
`{henkel, spaink}@rulbim.leidenuniv.nl`
[2] Leiden Institute of Advanced Computer Science, Leiden University,
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
`rozenber@liacs.nl`

Abstract. In many implementations of DNA computing, reliable detection of hybridization is of prime importance. We have applied several well-established DNA mutation scanning methods to this problem. Since they have been developed for speed and accuracy, these technologies are very promising for DNA computing. We have benchmarked a heteroduplex migration assay and enzymatic detection of mismatches on a 4 variable instance of 3SAT, using a previously described blocking algorithm. The first method is promising, but yielded ambiguous results. On the other hand, we were able to distinguish all perfect from imperfect duplexes by means of a CEL I mismatch endonuclease assay.

1   Introduction

Biomolecular computing studies the potential of using biological molecules, currently mainly DNA, to perform computations. One focus of investigation is the use of combinatorial libraries of DNA to provide search spaces for parallel filtering algorithms. Many different methods for library generation, solution filtering, and output have been studied experimentally. The formal satisfiability problem has become a sort of benchmark in the implementation of such algorithms [1,2,3].

Computing by blocking is one of the methodologies for molecular computing [4]. The blocking algorithm uses nucleic acid complementarity to remove molecules not representing a solution from the candidate pool. To an initial library of single stranded DNA molecules corresponding to (all) potential solutions, a set of complementary falsifying DNA (blockers) is added. Only those library molecules not representing solutions will combine with a blocker to form a perfect DNA duplex. Library molecules corresponding to solutions should remain single stranded or form a duplex with mismatched basepairs, depending on experimental conditions. The experimental challenge in implementing this algorithm is to very precisely separate perfectly matched molecules from mismatched ones.

The original proposal for the implementation of the blocking algorithm was using PCR inhibition. Falsified molecules were to be made unavailable for DNA polymerase through their association with a blocker molecule, for example peptide nucleic

acid (PNA). This would result in the selective amplification of unblocked DNA. So far, experimental data supporting this method is lacking. However, several fluorescence techniques (resonance energy transfer and correlation spectroscopy) have been successfully employed in combination with the blocking algorithm [5].

Here, we report the use of a heteroduplex migration assay and enzymatic mismatch recognition to implement blocking. In contrast to the fluorescence techniques mentioned before, both rely on DNA duplex structure rather than hybridization kinetics. These techniques are widely used to scan for mutations in molecular biological and clinical laboratories, and are well suited for high-throughput analysis of large numbers of samples [6,7].

During electrophoresis, perfect double stranded (homoduplex) DNA migrates through a gel at a predictable rate, dependent only on the strength of the applied electrical field, gel and buffer conditions and DNA length. However, DNA containing nucleotide mismatches (heteroduplex) and single stranded DNA migrate at anomalous rates, caused by secondary structure formation (ssDNA) or helix distortion (dsDNA). Such structures experience specific, but unpredictable, resistances when migrating through the gel matrix. Their mobility is lower than that of homoduplexes of equal length and as a result bands end up higher on the gel. Several well-established and sensitive mutation detection techniques exploit this effect, such as single strand conformational polymorphism (SSCP), temperature or denaturing gradient gel electrophoresis (TGGE, DGGE) and heteroduplex analysis [8].

Enzymatic mismatch recognition is also widely used in mutation detection [9]. It uses specific endonucleases which recognize and digest the abnormal DNA conformations which result from mismatched nucleotides. We have used the recently discovered CEL I nuclease, purified from celery, for this purpose [10,11].

2   Experimental

2.1   Problem Instance and Algorithm

We have tested mutation detection techniques on the following 4 variable, 4 clause satisfiability (3SAT) problem:

$$F = (\sim a \vee b \vee \sim c) \,\&\, (a \vee \sim b \vee d) \,\&\, (\sim a \vee c \vee \sim d) \,\&\, (b \vee c \vee \sim d) \,, \qquad (1)$$

where a, b, c and d are the 4 variables with values of true (1) or false (0). $\vee$ stands for the OR operation, $\&$ for AND, $\sim$ for negation. Since the clauses are connected by AND, falsifying one clause is sufficient for falsification of the whole formula. For example, falsification of the first clause by $abc = \{101\}$ falsifies the entire formula F.

The blocking algorithm proceeds as follows:

1. synthesize all possible assignments as ssDNA;
2. synthesize blockers representing to falsifying assignments;
3. mix and hybridize;
4. apply mismatch detection method.

Library/blocker combinations that form perfect duplexes correspond to false assignments.

2.2   Sequence Design

To represent the entire solution space to a 4 variable SAT problem, 16 library oligonucleotides were designed. The general structure of the library molecules is:

5' [*start*] [*a*] [*b*] [*c*] [*d*] [*stop*] ,

with *a*, *b*, *c* and *d* sequences representing variables. Two subsequences correspond to the two values these variables can take: ATCACC for *false*, and GTCTGA for *true*. The sequence of any variable thus only depends on its value, not on its identity. *start* and *stop* are invariable sequences (CTTGCA and TTGCAC, respectively), bringing the total length of the molecules to 36 nucleotides. Library molecules are numbered from 0 to 15, after the binary numbers they encode. For example, truth assignment *abcd* = {1010} is represented by oligonucleotide 10:

5' CTTGCA GTCTGA ATCACC GTCTGA ATCACC TTGCAC
[*start*]   [*a* = 1]   [*b* = 0]   [*c* = 1]   [*d* = 0]   [*stop*]

Falsifying oligonucleotides, or blockers, are complementary to the library oligonucleotides:

3' [*start*] [*a*] [*b*] [*c*] [*d*] [*stop*] ,

with *start* = GAACGA, *stop* = AACGTG, *true* = CAGACT and *false* = TAGTGG (all 3' → 5'). Since the falsification of a clause only requires 3 specified variables, and blocker molecules must contain a statement on all 4 variables, 2 blockers need to be designed for every clause. The fourth variable is set to true in one, and to false in the other. (It may be possible to circumvent this encoding complication through the use of redundant blockers, which contain universal nucleotides [12].) The translation of all clauses into blockers is summarized in table 1.

Table 1. Blocker molecules

Clause	Falsified by *abcd*	Blocker molecules
(~*a* ∨ *b* ∨ ~*c*)	1010	A0
	1011	A1
(*a* ∨ ~*b* ∨ *d*)	0100	B0
	0110	B1
(~*a* ∨ *c* ∨ ~*d*)	1001	C0
	1101	C1
(*b* ∨ *c* ∨ ~*d*)	0001	D0
	1001	identical to C0

Constraints on the design of the variable sequences were: GC content < 50%, isothermal melting behaviour (calculated according to [13,14]), no repeats or subsequence complementarity >2 bp, and no self complementarity. The uniform melting behaviour results in a T_m that is in theory identical for all possible library/blocker combinations.

2.3   Oligonucleotides

Oligonucleotides were custom synthesized and labelled at Isogen Bioscience (Maarssen, NL). For detection, library molecules contain a covalent 5' Cy5 label (Amersham Biosciences), blockers a 5' fluorescein (FITC, Molecular Probes). All oligos were purified from 10% denaturing polyacrylamide gels to remove unbound dye. DNA was allowed to diffuse from gel slices by overnight soaking in 0,5 M NH_4Ac, 2 mM EDTA, 0.1% SDS, and recovered by ethanol precipitation. Concentrations were calculated from absorption measurements of the dyes at 494 nm (fluorescein) or 649 nm (Cy5). Molar extinction coefficients of 77,000 cm^{-1} M^{-1} (fluorescein) and 250,000 cm^{-1} M^{-1} (Cy5) were used.

2.4   Duplex Migration Assay

Mixtures of library and blocker molecules were made by combining 5 pmol per oligo in a gel loading buffer consisting of 1x TBE, 3.3% sucrose and 0.033% Orange G. Duplex DNA was formed by heating the mixtures to 95 °C for 5 minutes, and cooling to 4 °C at 0.1 °C $second^{-1}$ in a thermocycler (Biometra TGradient). Gels were prepared from regular acrylamide: bisacrylamide (20:1) or proprietary SequaGel MD (Mutation Detection) acrylamide matrix (National Diagnostics, Atlanta, Georgia, USA). Duplex destabilizing chemicals (urea, ethylene glycol, formamide, or glycerol) were sometimes added to enhance heteroduplex migration effects [15]. Gels were run in 1x TBE at 200 V and 4 °C. Gel images were captures on a Biorad FluorS MultiImager, using UV excitation with 530 nm band pass and 610 nm long pass filters for detection of fluorescein and Cy5 fluorescence, respectively. Digital images were processed in Corel Photopaint 11 (contrast adjustment and greyscale conversion).

2.5   Enzymatic Mismatch Cleavage Assay

Duplexes were prepared as described above, except that hybridization was carried out in 10 mM Tris/HCl pH 8.5. T7 endonuclease I (T7EI) was obtained from New England Biolabs and handled according to the manufacturers recommendations. Reactions containing 5 pmol per oligonucleotide and 1 unit of enzyme were allowed to proceed for up to 150 minutes.

   CEL I enzyme was obtained from Dr Edwin Cuppen (Hubrecht Laboratory, Utrecht, NL), see *http://cuppen.niob.knaw.nl* for a detailed isolation protocol. Several batches of varying activity were used throughout the experiments described in this article. Every lot of CEL I was tested, and for all subsequent experiments quantities were used that gave the effect shown in Fig. 3 after 30 minutes of incubation. Reactions were performed with 5 pmol per oligonucleotide in a 4 µl volume at 45 °C, in a 10 mM $MgSO_4$, 10 mM HEPES pH 7.5, 10 mM KCl, 0.002% Triton X-100, 0.2 µg $µl^{-1}$ BSA buffer. Reactions were stopped by placing samples on ice and adding 4 µl 80% formamide, 100 mM EDTA. Digests were analyzed on 10% TBE/polyacrylamide gels, which were imaged as before. Bands were analyzed using ImageJ software (version 1.31v, *http://rsb.info.nih.gov/ij*).

3   Results

3.1   Heteroduplex Migration

Optimal conditions for the heteroduplex migration assay were determined using several blocking and non-blocking oligo combinations and various gel formulations. 12.5% acrylamide gels supplemented with 20% urea were found to give good separation of duplexes and heteroduplexes and were used for all subsequent experiments (Fig. 1).
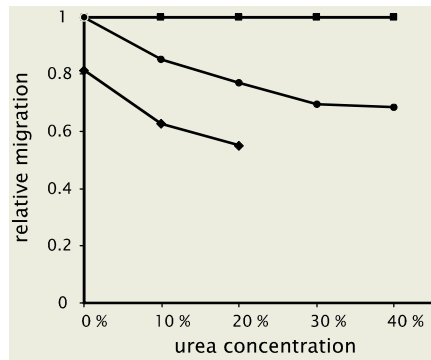

Fig. 1. Optimization of gel formulation. Shown are relative migration distances of 3 blocker/library combinations on 12.5% gels in the presence of progressive concentrations of urea. Migration of the perfect duplex combination C0 + 09 (squares) is set to 1. Other combinations are C0 + 08 (circles) and C0 + 10 (diamonds). The latter combination migrates as separate oligonucleotides on gels with urea concentrations of 30% and higher
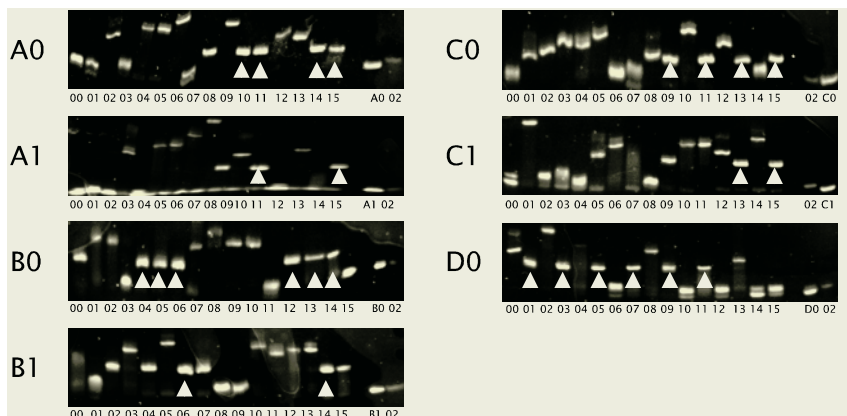

Fig. 2. Heteroduplex migration assay for all blocker/library combinations. Each gel contains the complete library (00-15) of oligonucleotides hybridized to the indicated blocker. The rightmost two lanes were loaded with unhybridized blocker and library 02. Arrows indicate apparent homoduplexes, which can be identified by coinciding fluorescence in both detection channels. The grayscale images in this figure only show total fluorescence

Fig. 2 shows the gel images for all combinations of blockers with library molecules. These images are grayscale superpositions of the 530 BP (showing the blocker fluorescein label) and 610 LP (library Cy5) channels. In RGB stack images, duplexes appear as yellow bands, since they fluoresce in both channels at the same location. Red (Cy5) and green (fluorescein) bands indicate non-hybridizing oligonucleotides. Arrows indicate apparent homoduplexes that can be identified from colour images.

Every blocker should only be able to form a perfect duplex with one of the library oligonucleotides, but Fig. 2 shows up to 6 apparent homoduplexes per blocker. No improvement was found using MD gel matrix or longer gels (not shown). Nonetheless, some solutions to the satisfiability problem can be identified from Fig. 2. Library oligos 00, 02 and 08 ($abcd$ = {0000}, {0010} and {1000}, respectively) do not behave as a homoduplex in any combination.

3.2   Mismatch Endonucleases

Fig. 3 shows the effects of T7EI and CEL I on homoduplex and heteroduplex DNA. Both were incubated for a range of times. In our hands, the T7EI enzyme did not have any discernable effect on any DNA sample (here, 0.2 units were used per reaction; 1 unit per reaction gave identical results), and was therefore not considered for further experiments.
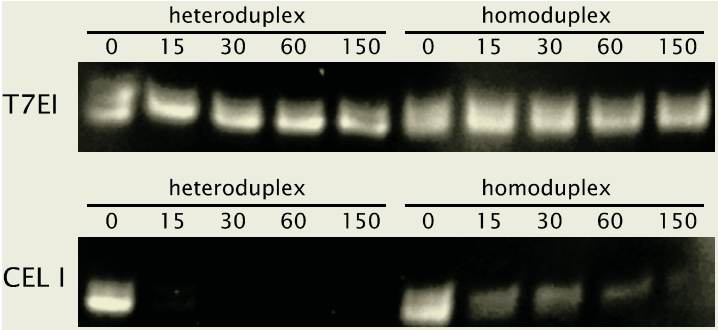

Fig. 3. T7EI & CEL I time series. 3 pmol samples of heteroduplex (C0 + 11) or homoduplex DNA (C0 + 13) were subjected to both endonucleases for up to 150 minutes. Samples were analyzed on a 12% denaturing gel. Images show total (combined 530 BP and 610 LP) fluorescence

CEL I, however, has a clear effect on all samples. The enzyme completely degrades the mismatched DNA within 30 minutes. Perfect dsDNA, although also subject to degradation, is still detectable after 150 minutes of reaction. ssDNA is quickly degraded completely (not shown). To test whether CEL I would successfully pick blocking from non-blocking combinations, all library molecules were incubated with blockers and enzyme (Fig. 4). From these results, satisfying assignments could be identified (summarized in table 2).
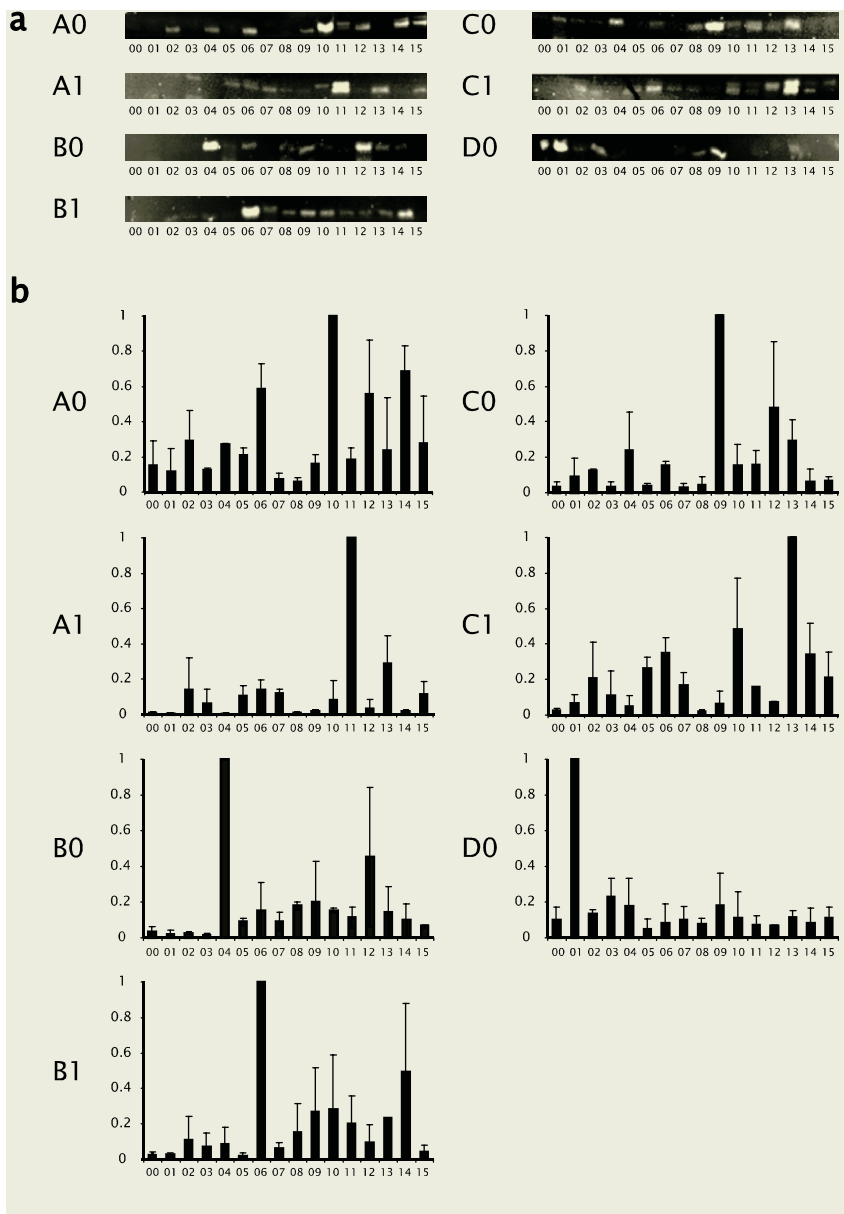
Fig. 4. a Effects of CEL I on all blocker/library combinations. Shown are denaturing gels of complete sets of library oligonucleotides and blockers, incubated with CEL I. **b** Quantified fluorescence from the gels. Fluorescence signals from the 530 BP channel are given relative to the signal of the most intense band on each gel (set to 1). The y-axis shows relative fluorescence, the x-axis the library molecules. Values are averages of two independent experiments. Error bars give standard deviations

Table 2. Apparent solutions to F (eq. 1)

Library molecule	*abcd*	falsified by duplex migration	falsified by CEL I
00	0000		
01	0001	D0	D0
02	0010		
03	0011	D0	
04	0100	B0	B0
05	0101	B0, D0	
06	0110	B0, B1	B1
07	0111	D0	
08	1000		
09	1001	C0, D0	C0
10	1010	A0	A0
11	1011	A0, A1, C0, D0	A1
12	1100	B0	
13	1101	B0, C0, C1	C1
14	1110	A0, B0, B1	
15	1111	A0, A1, C0, C1	

4   Discussion

Differential duplex migration did not provide a suitable test system to distinguish every satisfying solution from non-satisfying ones. There is no general theory describing the effect of anomalous DNA conformations on migration rate, and it was already known that not all mismatches can be detected this way [16,17]. A possible explanation for the ambiguous results reported here is the length of the DNA molecules: heteroduplex migration is generally recommended for DNA 100 – 500 bp in length. Such lengths also accentuate the effect of a single mismatch, which produces a bend in the helix. In addition, the nature of the mismatches studied here may have contributed. A single variable difference between blocker and library is represented by 4 non-matching basepairs at a molecular level. This will probably form a bubble-type mismatch, which may not always be subject to higher gel resistances.

We believe that with careful optimization of the encoding, the use of longer molecules (perhaps in combination with scaling to larger problem instances) and more sophisticated analytical techniques (e.g. capillary electrophoresis), the method holds considerable promise. In particular, duplex migration might be employed as a phenotype for the implementation of evolutionary algorithms in DNA [18].

The CEL I assay gave more consistent results. However, the results are difficult to interpret from visual inspection of single gels, because CEL I also degrades perfect duplexes. This breakdown of homoduplex DNA may be due to equilibrium fraying of the molecules, continuously giving the enzyme a toehold on the duplex. Therefore, longer molecules may also be an option for this method.

Using the blocking algorithm and encoding as reported here, the mismatch endonuclease assay is only useful as an analytical method. Because satisfying library

molecules are fragmented, multiple rounds of selection (as in evolutionary algorithms) cannot be easily implemented. However, several other proteins that bind mismatches (such as MutS [19]) do not destroy the DNA molecule. In future experiments, such proteins may be used in a gel-shift assay [20]. Besides the enzymatic method tested here, chemical cleavage of mismatches [21] could be considered. In conclusion, we believe that the use of mutation detection techniques is an interesting option for DNA based computing.

Acknowledgments

References

1. Faulhammer, D., Cukras, A. R., Lipton, R. J., Landweber, L. F., Molecular computation: RNA solutions to chess problems. Proc. Natl. Acad. Sci. USA **97** (2000) 1385-1389
2. Liu, Q. H., Wang, L. M., Frutos, A. G., Condon, A. E., Corn, R. M., Smith, L. M., DNA computing on surfaces. Nature **403** (2000) 175-179
3. Braich, R. S., Chelyapov, N., Johnson, C., Rothemund, P. W. K., Adleman, L., Solution of a 20-Variable 3-SAT problem on a DNA computer. Science **296** (2002) 499-502
4. Rozenberg, G., Spaink, H., DNA Computing by blocking. Theoretical Computer Science **292** (2003) 653-665
5. Schmidt, K. A., Henkel, C. V., Rozenberg, G., Spaink, H. P., DNA computing using single molecule hybridization detection. Nucl. Acids Res., in press
6. Taylor, G. R., Enzymatic and chemical cleavage methods. Electrophoresis **20** (1999) 1125-1130
7. Kristensen, V. N., Kelefiotis, D., Kristensen, T., Borresen-Dale, A. L., High-throughput methods for detection of genetic variation. Biotechniques **30** (2001) 318-332
8. Nataraj, A. J., Olivos-Glander, I., Kusukawa, N., Highsmith, W. E., Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis **20** (1999) 1177-1185
9. Mashal, R. D., Koontz, J., Sklar, J., Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. Nat. Genet. **9** (1995) 177-183
10. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., Yeung, A. T., Mutation detection using a novel plant endonuclease. Nucl. Acids. Res. **26** (1998) 4597-4602
11. Yang, B., Wen, X., Kodali, N. S., Oleykowski, C. A., Miller, C. G., Kulinski, J., Besack, D., Yeung, J. A., Kowalski, D., Yeung, A. T., Purification, cloning, and characterization of the CEL I nuclease. Biochemistry **39** (2000) 3533-3541
12. Loakes, D., The applications of universal DNA base analogues. Nucl. Acids Res. **29** (2001) 2437-2447
13. SantaLucia, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA **95** (1998) 1460-1465
14. Peyret, N., Seneviratne, P. A., Allawi, H. T., SantaLucia, J., Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A - A, C - C, G - G, and T - T mismatches. Biochemistry **38** (1999) 3468-3477

15. Ganguly, A., Rock, M. J., Prockop, D. J., Conformation-sensitive gel electrophoresis for rapid detection of single- base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **90** (1993) 10325-10329

16. Highsmith, W. E., Jin, Q., Nataraj, A. J., O'Connor, J. M., Burland, V. D., Baubonis, W. R., Curtis, F. P., Kusukawa, N., Garner, M. M., Use of a DNA toolbox for the characterization of mutation scanning methods. I: construction of the toolbox and evaluation of heteroduplex analysis. Electrophoresis **20** (1999) 1186-1194

17. Upchurch, D. A., Shankarappa, R., Mullins, J. I., Position and degree of mismatches and the mobility of DNA heteroduplexes. Nucl. Acids Res. **28** (2000) e69

18. Wood, D., Chen, J., Antipov, E., Lemieux, B., Cedeno, W., A DNA implementation of the Max 1s problem. In: Banzhaf, W., Eiben, A. E., Garzon, M. H., Honavar, V., Jakiela, M., Smith, R. E. (eds.), Proceeding of the genetic and evolutionary computation conference 1999. Morgan Kaufman, San Francisco (1999) 1835-1841

19. Brown, J., Brown, T., Fox, K. R., Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches. Biochem. J. **354** (2001) 627-633

20. Goode, E., Wood, D. H., Chen, J., DNA Implementation of a Royal Road Fitness Evaluation. In: Condon, A. , Rozenberg, G. (eds.), DNA computing, proceedings 6th international meeting on DNA based computers. Springer-Verlag, Berlin Heidelberg (2001) 247-262

21. Bui, C. T., Rees, K., Lambrinakos, A., Bedir, A., Cotton, R. G. H., Site-selective reactions of imperfectly matched DNA with small chemical molecules: applications in mutation detection. Bioorg. Chem. **30** (2002) 216-232