SSCP primer design based on single-strand DNA structure predicted by a DNA folding program

David A. Nielsen*, Alexey Novoradovsky[1] and David Goldman[2]

Sections of Molecular Genetics and [1]Human Neurogenetics, [2]Laboratory of Neurogenetics, NIAAA, NIH, Bethesda, MD 20892-0001, USA

ABSTRACT

To predict alterations in single-strand DNA mobility in non-denaturing electrophoretic gels, Zuker's RNA folding program was modified. Energy files utilized by the LRNA RNA folding algorithm were modified to emulate folding of single-strand DNA. Energy files were modified to disallow G•T base pairing. Stacking energies were corrected for DNA thermodynamics. Constraints on loop nucleotide sequences were removed. The LRNA RNA folding algorithm using the DNA fold energy files was applied to predict folding of PCR generated single-strand DNA molecules from polymorphic human ALDH2 and TPH alleles. The DNA-Fold version 1.0 program was used to design primers to create and abolish SSCP mobility shifts. Primers were made that add a 5′ tag sequence or alter complementarity to an internal sequence. Differences in DNA secondary structure were assessed by SSCP analysis and compared to single-strand DNA secondary structure predictions. Results demonstrate that alterations in single-strand DNA conformation may be predicted using DNA-Fold 1.0.

INTRODUCTION

Single-strand conformational polymorphism (SSCP) analysis allows identification of polymorphic sites in DNA (1). In this method, amplified DNA is denatured and separated in a non-denaturing polyacrylamide gel. DNA strands are resolved based on both conformation and size. Single nucleotide differences can alter electrophoretic mobility allowing identification of polymorphic alleles (1). Conformations are determined by several interactions including folding of complementary sequences forming hairpins, pseudoknots (2) and triple helices (3).

SSCP analysis is highly efficient for detecting nucleotide differences (4). For example in this paper we will use a naturally occurring polymorphism detected in a survey of human trypto-phan hydroxylase (TPH) (5). This A to C transversion, referred to as *TPH^L*, has been associated with suicidal behavior (6). However, certain substitutions are not readily discernible by SSCP. One variant in this category is the *ALDH2^2* variant which acts in a dominant fashion to cause a deficiency of mitochondrial aldehyde dehydrogenase (ALDH2). *ALDH2* is a G to A transition

at base pair 94 of exon 12 which results in a lysine to glutamate substitution at amino acid 487 (7,8).

A DNA folding program was developed and used to assist in the design of primers that would yield DNA molecules with altered SSCP mobility. Our starting point was the widely used LRNA algorithm by Jaeger, Turner and Zuker which folds linear RNA sequences (9,10). Folding parameters are based on RNA structure analysis and can correctly predict the majority of secondary structures in an RNA molecule (10). The LRNA algorithm uses a recursive method to predict suboptimal and optimal RNA structures. In addition, constraints can be imposed on allowed secondary structures at specific regions as determined from experimental data. Tertiary structures, triple helix formation (3) and pseudoknots (2) are not predicted by this program. The output can be used by SQUIGGLES [Genetics Computer Group (GCG), University of Wisconsin] (11), as well as by several other programs, to produce graphic representations of secondary structure. As described below, this program was modified to fold DNA and used to design modified primers that detect the ALDH2 variant by SSCP analysis. To further assess the utility of the DNA-Fold version 1.0 program, the SSCP mobility difference between two TPH alleles was abolished.

MATERIALS AND METHODS

Polymerase chain reaction

Polymerase chain reactions were performed on exon 12 of ALDH2 (7) and on TPH for the intron corresponding to mouse intron 7 (5). Amplifications with primer 1 and either primer 2a, 2b, 2c, 2d or 2e (see Table 1) yields DNA fragments 135, 143, 135, 135 and 135 bp in length, respectively. Amplifications with primer 3 and either primer 4a or 4b yields fragments 173 bp in length. Amplifications were performed with 100 ng human DNA, 100 nM of each primer, 250 µM each of dCTP, dGTP, dTTP and dATP, 50 mM KCl, 1.5 mM MgCl$_2$, 0.001% gelatin, 10 mM Tris, pH 8.3, 3.6 µCi [α-^{33}P]dCTP and 2 U AmpliTaq (Perkin-Elmer-Cetus) in 25 µl (14). Samples were amplified for 30 cycles, each cycle consisting of 20 s at 95°C, 20 s at 52°C and 30 s at 72°C, followed by 5 min at 72°C.

SSCP analysis

Amplified DNA (2 µl) was diluted with 18 µl 95% formamide, 10 mM NaOH, 0.05% bromophenol blue and 0.05% xylene
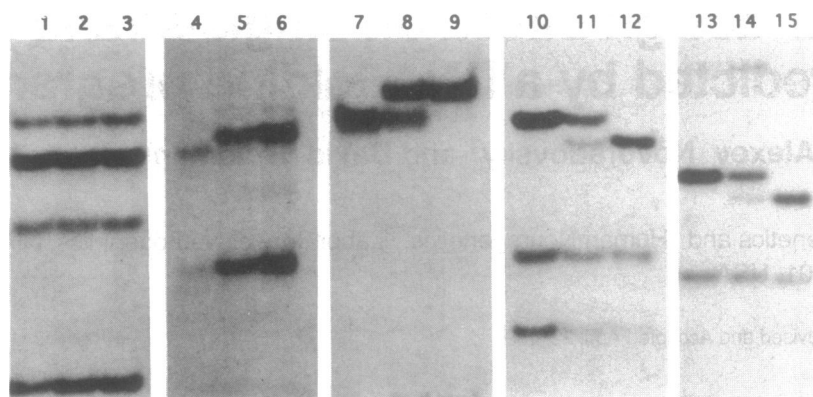

Figure 1. Creating SSCP mobility shifts for *ALDH2¹* and *ALDH2²*. Lanes 1–3, DNA amplified with primers 1 and 2a; lanes 4–6, primers 1 and 2b; lanes 7–9, primers 1 and 2c; lanes 10–12, primers 1 and 2d. Lanes 13–15, primers 1 and 2e; lanes 1, 4 , 7, 10 and 13 are the genotype *ALDH2¹/¹*; lanes 2, 5, 8, 11 and 14 are *ALDH2¹/²* and lanes 3, 6, 9, 12 and 15 are ALDH2²/².

cyanole and incubated for 2 min at 100°C. Four μl of this denatured DNA was loaded per lane and electrophoresed on a 5% non-denaturing polyacrylamide gel in a sequencing apparatus (1). Electrophoresis was carried out at 20°C for 17 h at 200 V. Gels were dried and autoradiography was performed at –70°C.

Table 1.

Primer	Sequence
1	5′-CAAATTACAGGGTCAACTGCT-3′
2a	5′-CCACACTCACAGTTTTCACTT-3′
2b	5′-ACTGAAGTCCACACTCACAGTTTTCACTT-3′
2c	5′-CCACACTAACAGTTTTCACTT-3′
2d	5′-CCACACTGACAGTTTTCACTT-3′
2e	5′-CCACACTCAAAGTTTTCACTT-3′
3	5′-TTGTTTCTTTATTTGATTAGTGT-3′
4a	5′-AGTTCATGGCAGGTATCTCTGAA-3′
4b	5′-AGTTCATGGCAGGTATCTGTGAA-3′

The tag sequence in primer 2b and the mismatches in primer 2c, 2d, 2e and 4b are underlined.

Computer analysis

The LRNA algorithm was run on a Convex supercomputer in a UNIX shell. The CT file was converted and displayed using the SQUIGGLES program of the GCG software package (11). The energy files for use in the DNA-Fold program are available via anonymous FTP at helix.nih.gov in the directory dna-fold. Primers 3 and 4a were designed using Oligo v.4 (National Biosciences, Inc.).

RESULTS AND DISCUSSION

Basis of the DNA-Fold, version 1.0 program

To understand folding of single-strand DNA molecules, others have used RNA folding programs such as that of Zuker (9,10). However, DNA structures predicted by RNA folding programs are unsatisfactory for several reasons. In DNA, the C-5 methyl
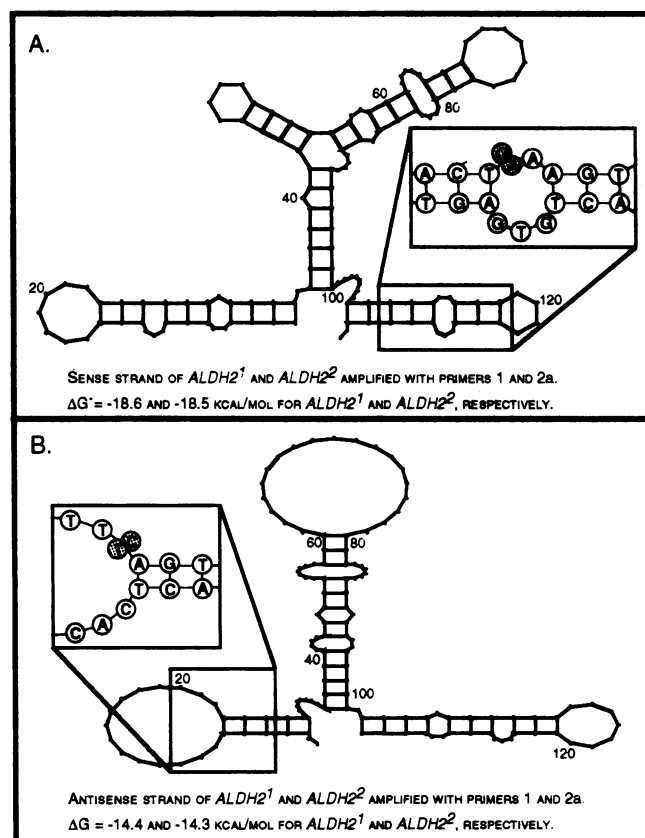

Figure 2. Predicted secondary structure for *ALDH2¹* and *ALDH2²* single-strand DNA amplified with primers 1 and 2a. (A) Sense strand. The G (*ALDH2¹*) to A (*ALDH2²*) mutation at position 114 is located in a bulge of the duplex region displayed to the right. This region is exploded to show the location of the transition. (B) Antisense strand. The C (*ALDH2¹*) to T (*ALDH2²*) mutation at position 114 is located in a loop to the right.

group of thymine prevents pairing with guanine; in RNA, guanine can pair with uracil. Therefore, DNA structures predicted with RNA folding programs will have mismatched G·T base pairs. Furthermore, the thermodynamics of nucleotide interactions in
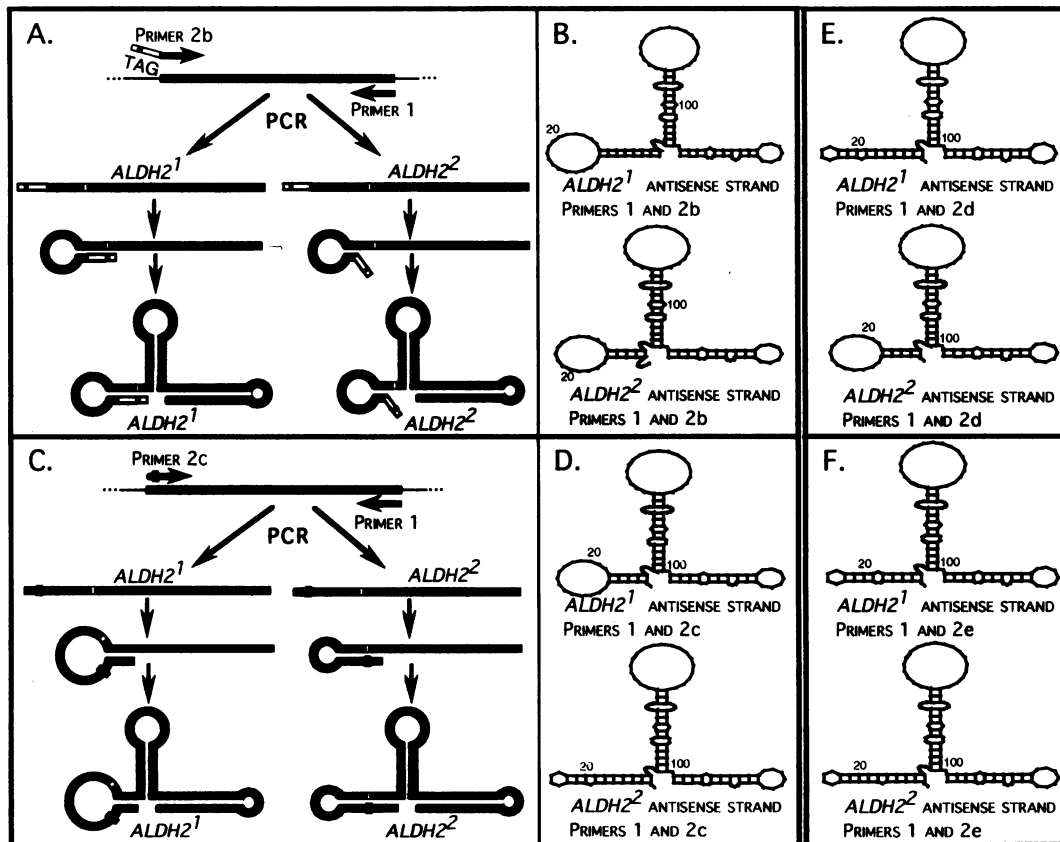
Figure 3. Predicted folding and secondary structure of $ALDH2^1$ and $ALDH2^2$ amplified with modified primers. Position of the polymorphism is highlighted. (A) Model for differential folding of $ALDH2^1$ and $ALDH2^2$ antisense PCR products generated with primers 1 and 2b, the tagged primer. (B) Secondary structures predicted by DNA-Fold 1.0 for $ALDH2^1$ ($\Delta G° = -19.1$ kcal/mol) and $ALDH2^2$ ($\Delta G° = -14.3$ kcal/mol) antisense strands obtained by amplification with primers 1 and 2b corresponding to the folding scheme in (A). (C) Model for differential folding of $ALDH2^1$ and $ALDH2^2$ antisense PCR products generated with primers 1 and 2c, a substitution primer. (D) Predicted secondary structures of $ALDH2^1$ ($\Delta G° = -14.6$ kcal/mol) and $ALDH2^2$ ($\Delta G° = -19.0$ kcal/mol) antisense strands obtained by amplification with primers 1 and 2c corresponding to the folding scheme in (C). (E) and (F) Predicted secondary structures of $ALDH2^1$ and $ALDH2^2$ antisense strands, obtaining by amplification with primer 1 and substitution primers 2d (E) and 2e (F). The $\Delta G°$ for $ALDH2^1$ and $ALDH2$ antisense strands equals -19.7 and -14.7 kcal/mol, respectively in (E) and -16.5 and -19.0 kcal/mol, respectively in (F).
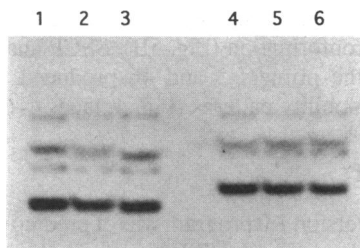

Figure 4. Abolition of the TPH SSCP with a substitution primer. Lanes 1–3, DNA amplified with primers 3 and 4a. Lanes 4–6, Primers 3 and 4b. Lanes 1 and 4 are genotype $TPH^{L/L}$, lanes 2 and 5 $TPH^{L/U}$ and lanes 3 and 6 are $TPH^{U/U}$.

RNA are not the same as in DNA. To circumvent these limitations, energy parameters of the Zuker program were modified as follows (see Materials and Methods): (i) free energy decreases ($\Delta G°$) contributed by G·U base pairings were deleted in the dangle.dat, stack.dat, tstack.dat energy files. (ii) The stacking energy file (stack.dat) was modified to reflect free energy decreases for DNA (15). (iii) The energy data for tetraloop nucleotides (tloop.dat) contribution to $\Delta G°$ were deleted. The

tetraloop energies had been derived from RNA loops occurring in nature (16). However, single-strand DNA folding is an *in vitro* phenomenon not subject to evolutionary pressures exerted on RNA loops. (iv) Loop (loop.dat) and miscloop (miscloop.dat) energy files remain unchanged. This program for folding single-strand DNA molecules contains the LRNA algorithm and modified energy files and is called DNA-Fold version 1.0.

Mobility differences in the ALDH2 alleles not detected by SSCP analysis

The $ALDH2^2$ allele exhibited the same SSCP mobility as the $ALDH2^1$ allele when published primer sequences (Table 1, primers 1 and 2a) were used (12). Individuals with $ALDH2^{1/1}$, $ALDH2^{1/2}$ and $ALDH2^{2/2}$ genotypes were evaluated (Fig. 1, lanes 1–3). No difference in SSCP mobility was observed. Secondary structures predicted by DNA-Fold 1.0 are shown in Figure 2. When primers 1 and 2a are used, the predicted folded DNA structure is the same regardless of whether there is G114 as in $ALDH2^1$ or A114 as in $ALDH2^2$.

As shown in Figure 1, lanes 1–3, complementary sense and antisense strands do not co-migrate. The predicted secondary
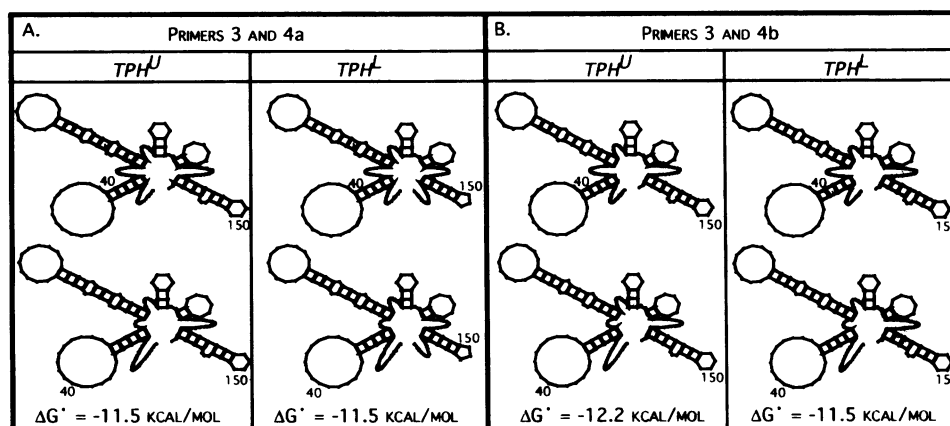
Figure 5. Secondary structures of the TPH^L and TPH^U alleles predicted by DNA-Fold. (A) Predicted structures of TPH^L and TPH^U sense strands with primers 3 and 4a. (B) Predicted secondary structures of TPH^L and TPH^U sense strands with primers 3 and 4b.

structures of complementary single-strands are not mirror images (Fig. 2) nor does their folding result in the same free energy decrease. Although there is reciprocal replacement of nucleotides, the free energy decreases for the nearest-neighbor interactions are completely different. Furthermore, since nucleotides are chiral, the complementary structures are not enantiomers. For example, right-handed helical regions would remain right-handed. Also, the molecular weights of the complementary strands are different. Other interactions involved in the formation of tertiary structure are non-identical in the sense and antisense strands.

Altered primer sequences based on predicted single-strand DNA secondary structure

Because the $ALDH2^2$ polymorphic site was close to the downstream primer 2a, a modified primer could cause alternative folding of the two ALDH2 alleles to yield distinct SSCP patterns. Using DNA-Fold 1.0, four variations of the downstream primer were designed, including a tagged primer and three single-base substitution primers. DNA-Fold predicted that a 5' 8 nt tag [to generate primer 2b (Table 1)] would create a stem–loop due to complementarity to nucleotides 111–118 of the $ALDH2^1$ antisense strand (Fig. 3A, antisense strand displayed). This stem–loop would be disrupted by the $ALDH2^2$ substitution. Altered SSCP mobility predicted on the basis of secondary structure was observed (Fig. 1, lanes 4–6). Curiously, electrophoretic patterns for the $ALDH2^{1/2}$ and $ALDH2^{2/2}$ are similar (Fig. 1B, lanes 5 and 6). The tagged primer preferentially amplifies the $ALDH2^2$ allele in heterozygotes (Fig. 1, lane 4). Because the stem region is longer in $ALDH2^1$ than in $ALDH2^2$ this region could inhibit amplification of the $ALDH2^1$ allele.

Three substitution primers, 2c, 2d and 2e, were designed with a single base internal mismatches to $ALDH2^1$ (Table 1). These mismatches do not significantly alter hybridization properties, but they do modify predicted secondary structures (Fig. 3D–F). In accordance with DNA structure predictions, SSCP patterns of DNA amplified with primer 1 and either primer 2c, 2d or 2e reveal mobility differences for the ALDH2 alleles (Fig. 1, lanes 7–15).

Alteration of primer sequence based on predicted secondary structure can abolish altered SSCP mobility

The majority of single nucleotide substitutions alter SSCP mobility. Therefore, we wanted to further test the predictive value of the DNA-Fold program by designing primers to abolish an SSCP difference. Primers 3 and 4a yielded the SSCP pattern shown in Figure 4, lanes 1–3, for the TPH genotypes. Folding of this fragment with the DNA-Fold program produced alternate allele-specific secondary structures for the TPH^U and TPH^L alleles (Fig. 5A, sense strand displayed) in agreement with the SSCP results. Two sense strand secondary structures, each with the same free energy decrease, were predicted for each TPH allele.

Using DNA-Fold 1.0, the downstream primer was redesigned to abolish SSCP mobility differences between TPH^U and TPH^L. Because DNA-Fold predicted similar secondary structures with the substituted primer, we reasoned that these PCR products should have similar SSCP mobilities. The new secondary structure with primer 4b (Table 1) should place the polymorphic nucleotide in a single-strand region where the polymorphism would not alter conformation (Fig. 5B). SSCP analysis of DNA amplified with the primers 3 and 4b produced, as predicted, identical SSCP mobility patterns (Fig. 4, lanes 4–6).

CONCLUSION

The DNA-Fold version 1.0 program was of predictive value in the rational design of primers for SSCP analysis. This program can be used to design primers that yield altered secondary structures for known polymorphic alleles. It may be useful for predicting the sensitivity of SSCP analysis for identifying substitutions in particular sequences. We used DNA-Fold predictively to modify primers to both create and abolish SSCP mobility differences. It may be advantageous to predict if polymorphisms in particular sequence motifs would alter structure.

Several other algorithms besides the LRNA algorithm have been written to predict RNA secondary structure (e.g. refs 17–29). Similar modifications of these programs could be made to simulate single-strand DNA folding. It would be of value to attempt folding of DNA with these programs modified to fold DNA to identify which program would be optimal for predicting

mobility changes in SSCP analysis. An advantage of the DNA-Fold program is that it uses the LRNA algorithm that it is widely distributed and its associated energy files are accessible to modification. The FOLD-A algorithm (24) by Nussinov and Pieczenik has been used to fold single-strand DNA but has several limitations including an inability able to find minimal free energy structures.

ACKNOWLEDGEMENTS

REFERENCES

1 Orita,M., Iwahana,I, Kanazawa,H., Hayashi,K. and Sekiya,T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2766–2770.
2 Pleij,C.W.A., Rietveld,K. and Bosch,L. (1985) *Nucleic Acids Res.*, **13**, 1717–1731.
3 Häner,R. and Dervan,P.B. (1990) *Biochemistry*, **29**, 9761–9765.
4 Dean,M., White,M.B., Amos,J., Gerrard,B., Stewart,C., Khaw,K.T. and Leppert,M. (1990) *Cell*, **61**, 863–870.
5 Nielsen,D.A., Dean,M. and Goldman,D., (1992) *Am. J. Human. Genet.*, **51**, 1366–1371.
6 Nielsen,D.A., Goldman D., Virkkunen,M., Tokola,R., Rawlings,R. and Linnoila,M. (1994) *Arch. Gen. Psychiatry*, **51**, 34–38.
7 Hsu,L.C., Tani,K., Fujiyoshi,T., Kurachi,K. and Yoshida,A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3771–3775.
8 Impraim,C., Wang,G. and Yoshida,A. (1982) *Am. J. Hum. Genet.* , **34**, 837–841.
9 Jaeger,J.A. Turner,D.H. and Zuker,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7706–7710.
10 Jaeger,J.A. Turner,D.H. and Zuker,M. (1989) *Methods Enzymol.*, **183**, 281–306.
11 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
12 Goedde,H.W., Agarwal,D.P., Harada,S., Rothhammer, F., Whittaker,J.O. and Lisker,R. (1986) *Am. J. Hum. Genet.*, **38**, 395–399.
13 Stoll,J. and Goldman,D. (1991) *J. Neurosci. Res.* **28**, 457–465.
14 Saiki,R.K., Gelfand,D.H., Stoffel,S., Sharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
15 Breslauer,K.J., Frank,R., Blöcker,H. and Marky,L.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3746–3750.
16 Tuerk,C., Gauss,P., Thermes,C., Groebe,D.R., Gayle,M., Guild,N., Stormo,G., d'Aubenton-Carafa,Y., Uhlenbeck,O.C., Tinoco,I.,Jr., Brody,I.N. and Gold,L. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1364–1368.
17 Jordan,B.R. (1971) *J. Theor. Biol.*, **34**, 363–378.
18 Nussinov,R. and Jacobson,A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6309–6313.
19 Zuker,M. and Stiegler,P. (1981) *Nucleic Acids Res.*, **9**, 133–148.
20 Jacobson,A.B., Good,L., Simoneetti,J. and Zuker,M. (1984) *Nucleic Acids Res.*, **12**, 45–51.
21 Martinez,H.M. (1984) *Nucleic Acids Res.*, **12**, 323–334.
22 Mironov,A.A., Dyakonova,L.P. and Kister,A.E. (1984) *Mol. Biol. (USSR)*, **18**, 1686–1694.
23 Nussinov,R. and Pieczenik,G. (1984) *J. Theor. Biol.*, **106**, 245–259.
24 Solovyov,V.V., Zharkikh,A.A., Kolchanov,N.A. and Ratner,V.A. (1984) *FEBS Lett.*, **165**, 72–78.
25 Hogeweg,P. and Hesper,B. (1984) *Nucleic Acids Res.*, **12**, 67–74.
26 Le,S.Y., Chen,J.H., Braun,M.J., Gonda,M.A. and Maizel,J.V. (1988) *Nucleic Acids Res.*, **16**, 5153–5168.
27 Abrahams, J.P., Van Den Berg,M., Van Batenberg, E. and Pleij,C.W.A. (1990) *Nucleic Acids Res.*, **18**, 3035–3044.
28 Gultyaev,A.P. (1991) *Nucleic Acids Res.*, **19**, 2489–2494.
29 Chen,J.-H., Le,S.-Y. and Maizel,J.V. (1992) *CABIOS*, **8**, 243–248.