DNA Base Bulge vs Unmatched End Formation in Probe-based Diagnostic Insertion/Deletion Genotyping: Genotyping the UGT1A1 (TA)<sup>n</sup> Polymorphism by Real-Time Fluorescence PCR

Nicolas von Ahsen,* Michael Oellerich, and Ekkehard Schütz

**Background:** Gilbert syndrome is a clinically inconsequential entity of mild unconjugated hyperbilirubinemia caused by an A(TA)<sub>n</sub>TAA insertion polymorphism (UGT1A1*28) in the promoter region of the gene coding for the enzyme UDP-glucuronosyltransferase 1 (EC 2.4.1.17; UGT1A1). Present methods for genotyping this polymorphism are laborious.

**Methods:** Hybridization probes were designed complementary to the wild type (TA)<sub>6</sub> and to alleles with (TA)<sub>7</sub> and (TA)<sub>8</sub> repeats in the promoter region. Melting points were measured in samples representing all currently known alleles with (TA)<sub>5</sub> to (TA)<sub>8</sub> repeats. Probe melting points were predicted with a thermodynamic nearest-neighbor model for Watson-Crick paired probes. The dominant secondary structures resulting from probe hybridization were predicted by thermodynamic free energy calculations. Alternatively samples were genotyped based on amplicon size resolved by high-resolution polyacrylamide gel electrophoresis.

**Results:** Only short probes (22–24 bases) could be successfully used for genotyping this locus because of the very low stability of this TA repeat. Assays based on (TA)<sub>6</sub> or (TA)<sub>8</sub> genotype-compatible hybridization probes effectively discriminated five to eight TA repeats. The consecutive use of two different detection probes was necessary for better discrimination of some heterozygous genotypes. All results were in concordance with the alternative genotyping method. Of 100 investigated Caucasians (50 males, 50 females), 9 (9%) were homozygous for the (TA)<sub>7</sub> allele.

**Conclusions:** The presented method for genotyping the (TA)<sub>n</sub> promoter polymorphism of the UGT1A1 gene with the LightCycler has the potential to genotype all currently known (TA)<sub>n</sub> repeats in a single assay and is sensitive toward possible new genotypes. Our findings also show that thermodynamic calculations are of practical value for the design of hybridization probe assays for the genotyping of insertion/deletion polymorphisms.

© 2000 American Association for Clinical Chemistry

Gilbert syndrome (icterus juvenilis Meulengracht) is a prevalent, benign, and clinically inconsequential entity of mild unconjugated hyperbilirubinemia (1). The prevalence is 12.4% in male Caucasians and 4.8% in female Caucasians (2). Gilbert syndrome is caused by impaired hepatic glucuronidation of bilirubin by UDP-glucuronosyltransferase 1 (UGT1A1; EC 2.4.1.17). Most cases of this disorder are associated with a homozygous TA insertion in a TATA motif located in the promoter region of the UGT1A1 gene (3, 4). Alleles carrying the A(TA)<sub>7</sub>TAA genotype were assigned the abbreviation UGT1A1*28 as opposed to the wild type with the A(TA)<sub>6</sub>TAA genotype (5). Subsequently, five to eight TA repeats have been found in different populations, and there is an inverse correlation between promoter activity and TA repeat length (3). Throughout this report, we will designate genotypes as (TA)<sup>n</sup>.

Methods for genotyping this insertion polymorphism are laborious and time-consuming (3, 4, 6). Recently, two methods were described that allow for higher sample throughput. One method, based on denaturing HPLC (7), cannot discriminate different homozygous genotypes in one analysis. The other uses the melting temperature (T<sub>m</sub>) of the PCR product as an indirect estimate of strand length and thus of the (TA)<sup>n</sup> promoter genotype (8).

Department of Clinical Chemistry, Georg-August-University, Robert-Koch-Strasse 40, 37075 Goettingen, Germany.

*Author for correspondence. Fax 49-551-39-8551; e-mail nahsen@gwdg.de.

Received August 23, 2000; accepted October 4, 2000.

1 Nonstandard abbreviations: UGT1A1, UDP-glucuronosyltransferase 1; (TA)<sup>n</sup>, number of TA repeats in the UGT1A1 promoter TATA box; T<sub>m</sub>, melting temperature; and PAGE, polyacrylamide gel electrophoresis.

Clinical Chemistry 46:12
In general, GC as well as TA tandem repeats in DNA are intrinsically unstable because of hairpin formation, replication slippage, and probably also unequal sister chromatid exchange. GC and TA tandem repeats are involved in the pathogenesis of various diseases, and the understanding of their molecular turnover is increasing (9). Therefore, the easy and reliable detection of such nucleotide repeat polymorphisms is of importance for present genotyping methods. The suitability of the LightCycler™ (Roche Biochemica) for hybridization probe genotyping to detect insertions or deletions has been demonstrated (10–12). In these situations and in contrast to the detection of single or multiple mismatches, other secondary structures such as bulges, loops, or unmatched ends often result from mismatched probe hybridization. Unpaired bases, so-called bulges, are formed in nucleic acid helices when one or more consecutive bases do not pair to bases on the other strand. If two or more bases remain unpaired, the structure is called a bulge loop. These secondary structures influence DNA structure and stability (13) in ways different from that of single mismatches.

Because of its high TA content, the UGT1A1 locus requires an optimized probe design to achieve safe discrimination of genotypes in this DNA region of low stability. We have shown that the $T_m$ observed from genotyping of single nucleotide polymorphisms can be predicted by a thermodynamic nearest-neighbor model (14). The thermodynamic parameters (enthalpy and entropy) that describe the destabilization of a nearest-neighbor DNA doublet by bulge or loop formation have not been published, whereas the free energies ($\Delta G^\circ$) for most of the different secondary structures are known. This is useful for prediction of stability trends resulting from mismatched probe hybridization and for assessment of the resulting secondary structures.

We present a PCR method for genotyping the UGT1A1 (TA)$_n$ polymorphism using hybridization probes on the LightCycler. Patterns resulting from probe hybridization on insertion/deletion polymorphisms are discussed based on thermodynamic considerations. The method was compared with the results from genotyping by PCR amplicon size resolution using polyacrylamide gel electrophoresis (PAGE).

**Materials and Methods**

**PRIMER AND PROBE SELECTION**

A primer set was designed for amplification of a 186-bp fragment from the UGT1A1 gene promoter of the (TA)$_n$ genotype (GenBank Accession No. D87674): forward primer UGT1A1-for (5'-GTC ACC TGA AGT CAA AC-3'); reverse primer UGT1A1-rev (5'-CAG CAT GGG ACA CCA CTG-3'). The polymorphic site was covered by a (TA)$_7$ genotype complementary detection probe: UGT1A1(TA)$_7$ probe (5'-GCC ATA TAT ATA TAT ATA TAA G-3' fluoroscein). The hybridization of this probe with different (TA)$_n$ genotypes is depicted in Fig. 1. The choice of a (TA)$_7$ probe should allow detection of all known genotypes, i.e., five to eight TA repeats. To avoid probe slipping, a tight hybridization was enforced by the GCC sequence at the 5' terminus (Fig. 1a). For discrimination between (TA)$_7$ and (TA)$_8$ genotypes, the probe contained a 3' G. The anchor probe, UGT1A1-anchor (6 bp downstream; 5'-Cy5.5-AGG AAC TCT TGG CAG GA-3'-PHO) was 5'-labeled with the Cy5.5 dye (Amersham Pharmacia Biotech), which exhibits an emission maximum at 694 nm, comparable to the LC-Red705 dye (Roche Biochemicals). The 3' end of the UGT1A1-anchor primer is shown below.

**Fig. 1.** Scheme of the hybridization resulting from genotyping the UGT1A1*28 mutation in the UGT1A1 promoter using a (TA)$_7$ probe.

Nucleotides not present in the wild-type (TA)$_8$ allele are in uppercase letters. Note that the TA loops in panels b, c, and d could also shift to adjacent TA/AT positions. (a), in the presence of the (TA)$_7$ allele, the probe is completely Watson-Crick paired, $\Delta G^\circ = -19.31$ kcal/mol (b), the wild-type (TA)$_8$ allele produces a 4-bp unmatched end, $\Delta G^\circ = -14.6$ kcal/mol (l), or a 2-bp bulge loop formation in the hybridized probe, $\Delta G^\circ = -15.05$ kcal/mol (ll). (c), hybridization with the (TA)$_8$ allele produces a 2-bp unmatched end, $\Delta G^\circ = -16.0$ kcal/mol (l), or a 2-bp bulge loop formation in the antisense strand, $\Delta G^\circ = -16.5$ kcal/mol (ll). (d), for a (TA)$_9$ allele, the combination of a double and a single mismatch destabilizes the binding of the probe, producing a 6-bp unmatched end, $\Delta G^\circ = -13.1$ kcal/mol (l), or a 4-bp bulge loop formation in the hybridized probe, $\Delta G^\circ = -12.41$ kcal/mol (ll).

1a gccatatatatatatataTAtaag 1c-1 ..tatataTataag
aacggtatatatatatatATttcat ..atatATATttc
1a-1 ..tatataTataag 1b-1 ..tatatataag
atatatatctc ..atatatatcttcatcc
1b-1 ..tatataTataag 1c-1 ..tatataTataag
atatATATttc ..atatATATttc
1c-1 ..tatataTataag 1c-1-1 ..tatataTataag
atatATATttc ..atatATATttc
1d-1 ..tatataTataag 1d-1 ..tatataTataag
atat.ttataatc ..atatatatctt
was phosphorylated. Alternatively, the following probes were used as detection probes together with the UGT1A1-anchor probe: UGT1A1(TA)₆ probe (5'-TGC CAT ATA TAT ATA TAT AAG T-3’/fluorescein); or UGT1A1(TA)₈ probe (5'-GCC ATA ATA TAT ATA TAT ATA TAT AAG-3’/fluorescein). All primers and probes were from MWG-Biotech.

GENOMIC DNA EXTRACTION
Genomic DNA was extracted as described previously (15). The protocol yielded a genomic DNA concentration of ~0.75 ng/μL; 1 μL of this solution was used in the PCR assay.

SITE-DIRECTED MUTAGENESIS
The (TA)₅ and (TA)₈ alleles have not been detected in the samples investigated to date. They were therefore constructed by site-directed mutagenesis using a megaprimer approach. A sample with the (TA)₇ genotype was cloned into a TOPO TA vector (Invitrogen). Plasmid DNA was then amplified with a proofreading enzyme mixture (Expand™ High Fidelity PCR system; Roche Biochemica) using mutagenic primers that introduced the (TA)₅ or (TA)₈ repeats into the sequence. The products were cloned into a TOPO TA vector, and the expected mutation was confirmed by sequencing (sequenase cycle sequencing kit; Amersham) on an automated DNA sequencer (Licor 4200; Licor). Melting curves of heterozygous samples with the (TA)₅ and (TA)₈ alleles were acquired after PCR amplification of homozygous genomic DNA to which diluted plasmid DNA had been added.

PCR PROTOCOL
PCR reactions were carried out in a final volume of 10 μL in LightCycler glass capillaries. The reaction mixture consisted of 1 μL of DNA solution, 1 U of Taq DNA polymerase (Roche Biochemica), 1 μL of 10× PCR buffer (Roche Biochemica), 0.2 mmol/L each dATP, dCTP, dGTP, and dTTP (Roche Biochemica), 2.5 mmol/L MgCl₂, 0.2 mmol/L each dATP, dCTP, dGTP, and dTTP (Roche Biochemica), 0.1 mg/mL bovine serum albumin (New England BioLabs), and 50 mL/L dimethyl sulfoxide (Sigma). Amplification primers and hybridization probes concentrations were as follows: 0.5 μmol/L UGT1A1-for, 0.5 μmol/L UGT1A1-rev, 0.1 μmol/L UGT1A1(TA)₇ probe, and 0.3 μmol/L UGT1A1-anchor. PCR-grade water was added to 10 μL. The fluorometer gain setting was 30 in channel 3. The cycling program consisted of 30 s of initial denaturation at 95 °C and 45 cycles of 95 °C for 0 s, 50 °C for 5 s, and 72 °C for 5 s, with the maximum ramp rate. The analytical melting program was 95 °C for 30 s and 29 °C for 30 s, increasing to 55 °C at a ramp rate of 0.1 °C/s, with continuous fluorescence acquisition. Use of the minimal possible ramp rate minimized the difference between the experimental Tₘ and the true equilibrium Tₘ as calculated with the nearest-neighbor model. Amplification and detection were complete within 30 min.

Each assay included a DNA control with the (TA)₅/ (TA)₇ genotype and a contamination control. Genotypes of control DNA were confirmed by cycle sequencing of samples initially genotyped with the LightCycler method. In all cases, sequencing confirmed the genotype expected from the hybridization probe assay.

GENOTYPING OF THE UGT1A1 (TA)₅ POLYMORPHISM BY HIGH-RESOLUTION PAGE
Genomic DNA was amplified for 40 PCR cycles using the same protocol as described above but without addition of detection probe and anchor. After PCR, the dNTPs (0.02 mmol/L each), Taq DNA polymerase (0.1 U), and the UGT1A1-rev primer (0.05 μmol/L) 5’ labeled with IRD-800 dye were added to each tube. The samples were centrifuged and amplified for an additional 10 PCR cycles. The PCR product was mixed with 3 μL of gel-loading solution (Sigma), and 1 μL was loaded on a sequencing gel (Licor 4200) together with size calibrators comprising five to eight TA repeats. Genotypes were assigned to samples by comparison to size calibrators.

THERMODYNAMIC CALCULATIONS
Thermodynamic predictions for Watson-Crick paired oligonucleotides were calculated using the MeltCalc software add-in for Excel. This software is free for noncommercial use and can be downloaded from http://www.meltcalc.de (14, 16). Alternatively, the HyTher software (http://jsl1.chem.wayne.edu) (17–19) may be used. For our PCR, we assumed 220 mmol/L Na⁺ equivalents and 0.1 μmol/L hybridizing strand concentration for duplex stability calculations. ΔG°₃₇ for secondary structure stability assessment was calculated under standard conditions (1 mol/L NaCl). See the Appendix for additional details.

Several DNA motifs frequently occur when probes hybridize to targets. These include so-called dangling ends, which occur on both ends of a completely hybridized probe; penultimate and terminal mismatches may also occur. Recently, the thermodynamic parameters for DNA sequences with dangling ends were investigated (20). Accordingly, certain dangling ends can contribute as much to duplex stability as an additional Watson-Crick base pair. The MeltCalc 2.0 software and the HyTher program both calculate double strand stability, including dangling end effects. Currently, only the HyTher program can account for penultimate and terminal mismatches, but it uses as yet unpublished nearest-neighbor parameters for the calculation of these situations. Thermodynamic calculations were carried out using both software programs. Calculations with MeltCalc included a 5’ dangling end and no contribution of the 3’ unmatched end. Calculations with HyTher included a 5’ dangling end and a 3’ terminal/penultimate mismatch. For completely Watson-Crick paired probes, 5’ and 3’ dangling ends were considered.
Results

ASSAY PERFORMANCE

Only short probes (22 and 24 bases) were successfully used for genotyping this locus because of the low stability of this TA repeat. Online PCR during amplification was not possible because the probes do not hybridize at the annealing temperature of the PCR. Successful amplification was evident from the melting curve display after PCR.

SECONDARY STRUCTURE CONSIDERATIONS

A homozygous (TA)₇ genotype with no mismatch under the UGT1A1(TA)₇ probe displayed a single melting peak at 46.2 °C (Fig. 1a and Table 1).  

\[ \Delta G^{0}_{37} = -14.6 \text{ kcal/mol} \]  

(calculation described in the Appendix). Hybridization with the wild-type DNA leads to formation of a 4-bp unmatched end (\[ \Delta G^{0}_{37} = -16.0 \text{ kcal/mol} \]), or of a 2-bp bulge loop (\[ \Delta G^{0}_{37} = -15.05 \text{ kcal/mol} \]), and a lower \( T_m \) of 39.9 °C. A (TA)₈ genotype produces a 2-bp unmatched end (\[ \Delta G^{0}_{37} = -16.0 \text{ kcal/mol} \]) or a 2-bp bulge loop (\[ \Delta G^{0}_{37} = -16.5 \text{ kcal/mol} \]), and a \( T_m \) of 42.2 °C. Hybridization with a (TA)₆ genotype leads to formation of a 4-bp unmatched end (\[ \Delta G^{0}_{37} = -13.1 \text{ kcal/mol} \]) and a lower \( T_m \) of 35.4 °C. The calculation for (TA)₉ is provided in the Appendix. There was good agreement between observed melting points and those predicted with the thermodynamic nearest-neighbor model (Table 1). The consideration of terminal mismatches did not improve the prediction compared with the observed \( T_m \)s using the LightCycler in this specific situation. The mean prediction error for the 12 cases (3 probes and 4 alleles) was 0.4 ± 1.1 °C (mean ± SD) without and 1.5 ± 1.2 °C with terminal mismatch calculations. For the reasons discussed above, \( T_m \) calculations are not possible at present for probe hybridizations that are destabilized by bulge or loop formation.

GENOTYPE ASSESSMENT

The \( T_m \)s for the different probes are summarized in Table 1. The CV was <1.5% for all run-to-run data shown in Table 1. Examples for the genotyping of homozygous samples are shown in Fig. 2. The \( T_m \) difference between matched and mismatched probes must be sufficient to allow discrimination of heterozygous mutations by two melting peaks. Initial genotyping should be performed with the UGT1A1(TA)₇ probe. This resolves samples with the common (TA)₆/(TA)₇ genotype and several other rare heterozygous genotypes (Fig. 3) as well as those homozygous for (TA)₆ or (TA)₇ alleles. Some genotypes are not well resolved with the UGT1A1(TA)₇ probe, e.g., the (TA)₇/(TA)₈ genotype (Fig. 3). If the \( T_m \) difference is too small, the UGT1A1(TA)₈ probe is recommended.

**Table 1. Observed vs predicted \( T_m \)s in diagnostic genotyping of insertion/deletion polymorphisms.**

<table>
<thead>
<tr>
<th>Probe</th>
<th>Genotype</th>
<th>Observed*</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT1A1(TA)₇</td>
<td>(TA)₆</td>
<td>35.4 ± 0.30</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>(TA)₇</td>
<td>39.9 ± 0.58</td>
<td>39.0</td>
</tr>
<tr>
<td></td>
<td>(TA)₈</td>
<td>46.2 ± 0.58</td>
<td>46.4</td>
</tr>
<tr>
<td></td>
<td>(TA)₉</td>
<td>42.2 ± 0.55</td>
<td>41.3</td>
</tr>
<tr>
<td>UGT1A1(TA)₈</td>
<td>(TA)₆</td>
<td>33.9 ± 0.28</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>(TA)₇</td>
<td>37.3 ± 0.31</td>
<td>39.0</td>
</tr>
<tr>
<td></td>
<td>(TA)₈</td>
<td>41.1 ± 0.51</td>
<td>41.3</td>
</tr>
<tr>
<td></td>
<td>(TA)₉</td>
<td>47.0 ± 0.33</td>
<td>47.7</td>
</tr>
<tr>
<td>UGT1A1(TA)₉</td>
<td>(TA)₆</td>
<td>39.0 ± 0.31</td>
<td>37.9</td>
</tr>
<tr>
<td></td>
<td>(TA)₇</td>
<td>46.2 ± 0.40</td>
<td>47.6</td>
</tr>
<tr>
<td></td>
<td>(TA)₈</td>
<td>41.1 ± 0.33</td>
<td>40.5</td>
</tr>
<tr>
<td></td>
<td>(TA)₉</td>
<td>39.8 ± 0.36</td>
<td>40.5</td>
</tr>
</tbody>
</table>

* Observed temperatures are the mean ± SD of four different analyses.
small, a single broad melting peak is produced, as illustrated by the (TA)$_6$/(TA)$_8$ genotype (Fig. 3). However, resulting melting curves are clearly differentiated from samples with other genotypes by their shapes and $T_m$s. Such samples must be reanalyzed using the UGT1A1(TA)$_8$ probe. The (TA)$_6$ probe is useful only for confirmation of a (TA)$_6$ genotype, but it is not needed for clinical genotyping. A general strategy for the unambiguous genotyping of all known (TA)$_n$ repeat genotypes using the UGT1A1(TA)$_7$ (7TA) and UGT1A1(TA)$_8$ (8TA) probes is outlined in Fig. 4. The resolution of melting curves improves with a lower ramp rate for melting curve acquisition and with a smaller number of samples processed in the assay. This should be considered when insufficient resolution of heterozygotes is suspected.

One hundred DNA samples (50 male and 50 female) were anonymously provided by our blood bank from local blood donors. Genotyping was performed using the UGT1A1(TA)$_7$ probe in the LightCycler assay. Allelic frequencies were comparable to those reported for a Caucasian population (3). The homozygous (TA)$_7$ genotype that confers a susceptibility to Gilbert syndrome was found in 6% of males and 12% of females, whereas 54% of males and 34% of females carried the homozygous wild-type allele.

**METHOD COMPARISON WITH HIGH-RESOLUTION PAGE**

Enough DNA material was left from 71 of these samples so that they could also be genotyped by high-resolution PAGE. Results were 100% concordant to the hybridization probe assay.

**Discussion**

When single nucleotide polymorphisms are genotyped with a hybridization probe assay, a well-defined mismatch occurs under an otherwise Watson-Crick paired probe. The relative stability of such mismatched DNA strands as judged by their $T_m$s in a LightCycler hybridization assay can be accurately predicted by a thermodynamic nearest-neighbor model (14–19). However, as of December 1999, mutations of the small deletion or small insertion type accounted for 22% of the 20 000 entries in the Human Gene Mutation Database (http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html) (21). The placement of hybridization probes over such a polymorphism may produce numerous nearest-neighbor interactions, such as single, multiple, and tandem mismatches, as well as single-base bulges or bulge loops (Fig. 1, b–d). Tandem mismatch refers to two immediately adjacent opposing unpaired nucleotides in each strand (an internal 2×2 loop). Multiple mismatches refers to several single mismatches (an internal 1×1 loop) interrupted by Watson-Crick paired nucleotides. The base pairing of a probe hybridizing over an insertion/deletion polymorphism will be such that a minimum free energy results. Therefore, in contrast to what is shown in Fig. 1 of the report by Nauck et al. (10), a single-base bulge will form rather than a loop because the binding of four G bases is energetically much more favorable than loop formation. TA repeats have a low stability, and the destabilizing effect of the polymorphism on the probe is weak. We were unable to achieve sufficient $T_m$ discrimination with longer detection probes (>32 bp, $T_m$ >55°C) at this locus. In our method, the shortest probe possible was chosen, which just covered the investigated region. Thus, probe hybridization is maximally destabilized by the insertion/deletion polymorphism.

In accordance with previous observations (14), we found good agreement between observed and predicted melting temperatures (Table 1). The $T_m$ calculated on the assumption that an unmatched end is not significantly contributing to the probe stability is a good estimate of the
experimental melting temperature (Table 1). In the case of a (TA)_6 genotype hybridized with the (TA)_7 probe, base pairing may be possible in the 6-bp unmatched end (Fig. 1d-I); however, this can produce unfavorable steric interactions of internal mismatches (22). Similar considerations are possible for the (TA)_6 and (TA)_8 probe. The genotyping of samples with five to eight TA repeats is possible with a single properly designed probe. Some rare allelic combinations require the use of a second probe, as outlined in Fig. 4. In Caucasian populations, the (TA)_6 and (TA)_7 alleles are most prevalent; consequently, the (TA)_7 probe is the best choice for genotyping because (TA)_6 and (TA)_7 alleles are well resolved. Repeated elements in the genome are unstable and prone to insertions and deletions because of unequal crossing over in meiosis (3, 9). It is possible that insertions/deletions other than those reported to date will be found once larger samples from different ethnic groups are screened with sensitive methods. LightCycler genotyping has the potential of detecting previously unknown or unexpected base substitutions (11, 14, 23). Hybridization probe-based methods could identify the presence of putative new alleles such as (TA)_4 or (TA)_9. We anticipate that nonspecific methods using an indirect measure of the (TA)_n repeat number by PCR product T_m (8) will give wrong genotyping results in the presence of certain allelic combinations. This is a point of concern because the presence of a heterozygous (TA)_n/(TA)_9 genotype was recently reported in an Italian patient with Gilbert syndrome (24). For example, a (TA)_n/(TA)_9 heterozygous individual might be misclassified as homozygous (TA)_7 genotype by these indirect methods.

In conclusion, the presented method has the potential to genotype all currently known (TA)_n repeats in a single assay and is sensitive toward possible new genotypes. Our findings also show that thermodynamic calculations are of practical value for the design of hybridization probe assays for the genotyping of insertion/deletion polymorphisms.

We thank Prof. Victor W. Armstrong for helpful comments on this manuscript. The skillful technical assistance of Sandra Hartung and Reiner Andag is gratefully acknowledged. The MelCalc software is copyrighted by E. Schütz and N. von Ahsen.

Appendix

Free energy was calculated according to the following formula, where ΔG^0_{37}(i) is the free energy at 37 °C for the 10 possible nearest neighbors and n_i is the number of occurrences of each nearest neighbor (19):

$$\Delta G^0_{37 \text{ total}} = \sum n_i \Delta G^0_{37}(i) + \Delta G^0_{37}(5' \text{ init}) + \Delta G^0_{37}(3' \text{ init})$$

Because DNA base bulges in a duplex have properties similar to RNA molecules (13), we used the known free energy penalties for the latter: a ΔG^0_{37 bulge} of ±3.8 kcal/mol for a 1-bp bulge, +2.8 kcal/mol for a 2-bp bulge, +3.2 kcal/mol for a 3-bp bulge, and +3.6 kcal/mol for a 4-bp bulge (25).

The penalty for bulge formation adds to the nearest-neighbor stacking pair free energy. The more negative the stacking free energy of a nearest-neighbor pair, the less likely it is that a base bulge will be inserted at that site (13). The formation with the lowest free energy is considered the dominant secondary structure. Resulting free energies for different secondary structures within ±0.5 kcal/mol of each other are within the experimental error. In these situations, the dominant secondary structure cannot be predicted.

Example

Hybridization of the 22mer (TA)_7 probe on the (TA)_6 wild-type strand can lead to different secondary structures (Fig. 1, b-I and b-II). The Watson-Crick pairing of the first 18 bp (Fig. 1b-I) produces a ΔG^0_{37 total} of −14.6 kcal/mol. Alternatively (Fig. 1b-II), a 2-bp bulge loop can form between a TA/AT nearest-neighbor pair. The resulting 20-bp Watson-Crick paired duplex has a ΔG^0_{37 total} of −17.85 kcal/mol, to which the loop penalty of +2.8 kcal/mol must be added. The ΔG^0_{37 total} of −15.05 kcal/mol for this secondary structure is not significantly different from the ΔG^0_{37 total} assuming unmatched end formation. Therefore, for this oligonucleotide the dominant secondary structure cannot be predicted.

Similar considerations led us to the postulated secondary structures depicted in Fig. 1c (unmatched end vs 2-bp bulge loop formation) and Fig. 1d (unmatched end vs 4-bp bulge loop formation).

References


