

Design and Application of Noncontinuously Binding Probes Used for Haplotyping and Genotyping

Genevieve Pont-Kingdon,^{1*} Rebecca L. Margraf,¹ Kelli Sumner,¹ Alison Millson,¹
Elaine Lyon,^{1,2} and Ekkehard Schütz³

BACKGROUND: Many methods for genotyping use melting temperature (T_m) of sequence-specific probes. Usually the probes hybridize to a continuous stretch of DNA that contains the variant(s). In contrast, hybridization of noncontinuous probes to a template can form bulges. This report generates guidelines for the design of noncontinuous probes.

METHODS: We used software to predict hybridization structures and T_m s from 10 noncontinuous probes and 54 different templates. Predicted T_m s were compared to existing experimental data. The bulging template's sequences (omitted in the probe) ranged in size from 1 to 73 nucleotides. In 36 cases, we compared observed and predicted ΔT_m s between alleles complementary to the probe and mismatched alleles. In addition, using software that predicts effects of bulges, we designed a probe and then tested it experimentally.

RESULTS: The mean differences between predicted and observed T_m s were 0.65 (2.51) °C with the Visual OMP software and 0.28 (1.67) °C with the MeltCalc software. ΔT_m s were within a mean (SD) of 0.36 (1.23) °C (Visual OMP) and -0.01 (1.02) °C (MeltCalc) of observed values. An increase in the size of the template bulge resulted in a decrease in T_m s. In 2 templates, the presence of a variant in the bulge influenced the experimental T_m of 2 noncontinuous probes, a result that was not predicted by the software programs.

CONCLUSIONS: The use of software prediction should prove useful for the design of noncontinuous probes that can be used as tools for molecular haplotyping, multiplex genotyping, or masking sequence variants.

© 2008 American Association for Clinical Chemistry

Many molecular technologies developed for mutation detection rely on the thermodynamic stability of sequence-specific oligonucleotide probes. Sequence variations are detected using probe melting analysis, because probe stability with its template varies with the presence or absence of mismatches. Primers used for PCR and probes that allow identification of specific variants usually hybridize across a continuous stretch of template DNA.

Probes that do not bind continuously on the template DNA have been described recently. Probes that span several loci [loci-spanning probes (LSProbes)⁴] allow molecular haplotyping (1) and multiplex genotyping in single step melting experiments (2). In LSProbes, the sequence of the template between the analyzed loci is partially omitted. On binding of the probe, the nonhybridized template sequence between the analyzed loci creates a bulge. By hybridizing simultaneously on several loci, LSProbes can analyze several variants separated by up to 80 nucleotides (nt). As with continuous probes, mismatches in the hybridized region of the LSProbe lowers the melting temperature (T_m) (temperature at which 50% of duplex strands are separated). This characteristic enables the identification of different variants and haplotypes.

To simplify melting interpretation of other sequence changes detected by the probe, masking probes incorporate a neutralizing mismatch at the position of nontargeted sequence variation (3). The masking probes in this report were designed with a deletion of the nontargeted sequence-variation position from the probe sequence. This design creates a template bulge between all possible alleles when hybridized to the masking deletion probe. Wild-type and masked variant alleles will have only the template bulge and a similar T_m . In contrast, any other sequence variation else-

¹ ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah, USA; ² Department of Pathology, University of Utah, Salt Lake City, Utah, USA; ³ Institute of Veterinary Medicine, Georg-August University, Göttingen, Germany.

* Address correspondence to this author at: Dr. Pont-Kingdon, ARUP Institute for Clinical and Experimental Pathology, ARUP Laboratories, 500 Chipeta Way, Salt

Lake City, UT 84108. E-mail: pontkig@aruplab.com.

Received November 15, 2007; accepted February 7, 2008.

Previously published online at DOI: 10.1373/clinchem.2007.100487

⁴ Nonstandard abbreviations: LSProbes, loci-spanning probes; nt, nucleotide; T_m , melting temperature; OMP, oligonucleotide modeling platform; FRET, fluorescence resonance energy transfer; Hb, hemoglobin.

Table 1. Design of the probes used in this study.^a

Target	(Name) Sequence	Length, nt		
		Total	Details	Reference
Ret protooncogene	cgtagcggcacagctc *1* tcgcacagtgg	26	15 (1) 11	3
	ggagaagtgtctc *3* gagcccgaagacatc	28	13 (3) 15	3
<i>FGFR3</i> gene	cacgttggt *1* gtctCcttgta	20	9 (1) 11	^b
<i>HBB</i> gene	(HbCmatch) ggctCaccac *49*ctccTtaggagtc	24	11 (49) 13	2
	(HbEmatch) ggctTaccac *49*ctccTCaggagtc	24	11 (49) 13	^b
	(HbEshort) gcctTacc *49*ctccTCaggagtc	21	8 (49) 13	^b
<i>ADRB2</i> gene	(46/79-GG) cccaatGgaagcca*22*cgagGaaagg	25	14 (49) 11	1
	(46/79-AC) cccaatAgaagcca*22*cgagCaaagg	25	14 (49) 11	^b
	(-20/46/79-CGG) gcgctCacctg*55*ccaatGgaagcca*22*cgagGaaagg	36	11 (55) 14 (22) 11	1
WIAF1537-1538	1537/1538: atgttatgtCgaaact*13-33-53-73*agaagaaaAtggc	29	16 (13-73) 13	18
<i>MC1R</i> gene	MC1R: gcaggcAgggg*40*gagcccGtcagg	23	12 (40) 11	^b

^a In the sequence, the numbers between the asterisks (*) indicate the number of template nucleotides omitted in the probe. The 4 numbers in the probe 1537/1538 indicate that the probe was hybridized on 4 different templates, each with a different bulge length, e.g., 13, 33, 53, and 73 nt. Capital letters in the sequence indicate variant positions.

^b See this article.

where under the probe will have an additional mismatch with the probe and a lower T_m than the wild-type and masked variant alleles.

The thermodynamic properties of nucleic acid hybridization are well studied (4–13), and software prediction of probe stability is useful in the design of molecular assays (14–16). Several software programs are available to predict the T_m s of continuous probes and primers, the most accurate using the thermodynamic nearest-neighbor model (5, 7, 12). The effect of single and multiple mismatches on T_m are well documented and predicted by several software programs (4–7, 11). Only a few, however, allow for the presence of bulges in either the template or the probe. We have evaluated the Visual OMP (oligonucleotide modeling platform) and the MeltCalc (14) software programs for their prediction of binding of noncontinuous probes on templates with and without mismatches. In addition, prediction of structure between templates and probes, effect of mismatches, size of the template bulge, and presence of variants in the bulge were examined.

Materials and Methods

NONCONTINUOUS PROBES

Ten of the 11 noncontinuous probes (Table 1) used in this study were designed without the aid of software that can incorporate into either template or probe the thermodynamic instability due to bulges. The noncontinuous probes with omission of 1–3 nt of the template sequence were designed to mask nontargeted sequence

variants such as benign polymorphic sites in *RET* protooncogene and the fibroblast growth factor receptor 3 gene (*FGFR3*)⁵ (3). The probes with longer template omissions of 13–73 nt were designed for multiplex genotyping of the beta hemoglobin gene (*HBB*) (2) or haplotyping of the beta 2 adrenergic receptor (*ADRB2*) or of the WIAF 1537–1538 markers on chromosome 21 (1, 17)]. We used the MeltCalc software that incorporates the thermodynamic effect of bulges on T_m predictions to design 1 probe with a 40-nt omission in the bovine *MC1R* melanocortin 1 receptor gene (*MC1R*), which we then tested experimentally.

In addition to probes published previously, a masking deletion probe was designed for the detection of the c.1948A>G mutation in the *FGFR3* gene, which is responsible for more than 99% of cases of thanatophoric dysplasia type II. The probe was a fluorescence resonance energy transfer (FRET) hybridization probe in which the site c.1953 has been omitted (Table 1) to avoid detection of an A>G benign polymorphism present in 4% of the general population. With this probe, the c.1953 A>G polymorphism is undetectable, and the masked polymorphic allele (G) has an identical T_m to the wild-type allele (A) of 47.5 °C, whereas the c.1948G disease-causing allele has a T_m of 56.6 °C (data not shown).

⁵ Genes: *FGFR3*, fibroblast growth factor receptor 3; *HBB*, hemoglobin, beta; *ADRB2*, adrenergic, beta-2-, receptor; *MC1R*, bovine *MC1R* melanocortin 1 receptor.

EXPERIMENTAL T_{ms} AND ΔT_{ms}

Experimental T_{ms} were all obtained by analysis of changes of a fluorescence signal during probe melting. Both adjacent hybridization probes detected by FRET and unlabeled probes detected by double-strand DNA-specific dyes (1–3, 18) have been used. Briefly, all PCR experiments were performed using a LightCycler instrument, and melting curves of hybridization probes were analyzed with the LightCycler data analysis software version 3.5 (Roche Diagnostics). T_{ms} of unlabeled probes were analyzed by high-resolution melting in an HR1 instrument (Idaho Technology) in the presence of the double-strand-specific dye LCGreen^{®Plus+} (Idaho Technology).

For each set of probes and templates, ΔT_{ms} were calculated by subtracting the lowest allele T_m from the highest allele T_m.

SOFTWARE ANALYSIS OF NONCONTINUOUS PROBES

Visual OMP (DNA Software; <http://www.DNAsoftware.com>) simulates hybridization of nucleic acids in solution and provides structural and thermodynamic parameters in intra- or intermolecular situations. We used version 6.0 of the software. Data entry included the sequences of the probe and the templates (complete PCR fragment) or the area of the templates that hybridized with the probe (see Fig. 2 of (2)). Structure settings in the program were adjusted to allow formation of bulges up to 80 nt. After simulation, the optimal structure of the probe hybridized to the template was selected as the predicted structure and T_m value.

MeltCalc is an Excel-based worksheet (<http://www.meltcalc.com>) for thermodynamic calculation and automatic design of hybridization probes. Version 2.95 was used for this study, upgraded from the commercially available 2.0 for introduction of bulges in the template. We entered the sequence of the probe, the “dangling ends” (nts corresponding to the nts found on the template, both in 5′ and 3′ of the probe), the length of the potential template bulge, and the location of the bulge in the probe.

In both programs, the salt conditions (214 mmol/L Na⁺) and oligonucleotide concentrations (0.2 μmol/L for adjacent hybridization probes and 0.5 μmol/L for unlabeled probes) corresponded to the conditions used in our experiments (19). In MeltCalc, a correction for double-strand dye (SYBR-Green I) was used with unlabeled probes.

Results

SOFTWARE PREDICTION OF MELTING TEMPERATURES FOR NONCONTINUOUS PROBES

Experimental T_{ms} from 54 experiments using 9 noncontinuous probes were compared to T_{ms} predicted by both Visual OMP and the MeltCalc software.

Fig. 1A shows the correlation between observed and calculated T_{ms} and Bland-Altman plots (20). Whereas the correlations do not significantly differ, the 95% CI ranges of the predicted values were somewhat different (Visual OMP, ±4.8 °C vs MeltCalc, ±3.3 °C). With both software programs, there is a slight trend toward underestimation of higher T_{ms}; this trend is more pronounced with Visual OMP, mainly owing to the noncontinuous probes that form small template bulges (≤3 nt).

In 36 experiments, 9 of the probes were hybridized with templates of different genotypes or haplotypes, and ΔT_{ms} were analyzed. As shown in Fig. 1B, there was no significant deviation from the line of identity within the given 95% CI of the correlation. The ΔT_{ms} are predicted with better precision (95% CI of predicted value: Visual OMP, ±2.6 °C vs MeltCalc, ±2.1 °C) than the T_m values.

To evaluate if the size of the DNA bulge affected the T_m and ΔT_{ms} predictions, the calculated minus observed values were plotted against the bulge size (Fig. 1C and 1D). Table 2 provides an overview of the calculations and shows that the average differences from both programs are similar for both T_m and ΔT_{ms} predictions. A more detailed evaluation of the T_m data (Fig. 1C) shows a significant trend toward underestimation of the destabilizing effect of bulges with Visual OMP, with a dependency of prediction on the bulge size that reaches statistical significance (significant deviation from slope = 1). If ΔT_{ms} are considered (Fig. 1D), the underestimation of the destabilization by larger bulges becomes substantially smaller, and both programs show almost identical predictions over the investigated range.

We tested the use of the MeltCalc software by designing a noncontinuous probe to analyze the linkage between 2 single-nucleotide polymorphisms separated by 53 nt in the promoter region of the bovine *MC1R* gene (21). The probe must reliably allow differentiation of the 2 linked haplotypes (WT and UU) from the 2 other possible haplotypes (nonlinkage types, U1 and U2). We selected the antisense wild-type probe that predicted a ΔT_m of 12.2 °C (see Supplemental Data Table 1 in the Data Supplement that accompanies the online version of this report at <http://www.clinchem.org/content/vol54/issue6>) between the 2 common haplotypes WT and UU. After experimentation, the observed ΔT_m between the haplotypes was 9.6 °C (SD 0.7 °C), which is within the error range calculated above.

DETECTION OF SEQUENCE VARIANTS IN THE BULGE

As noted above, both Visual OMP and MeltCalc can calculate the predicted T_{ms} and ΔT_{ms} for noncontinuous probes. We used Visual OMP to predict the

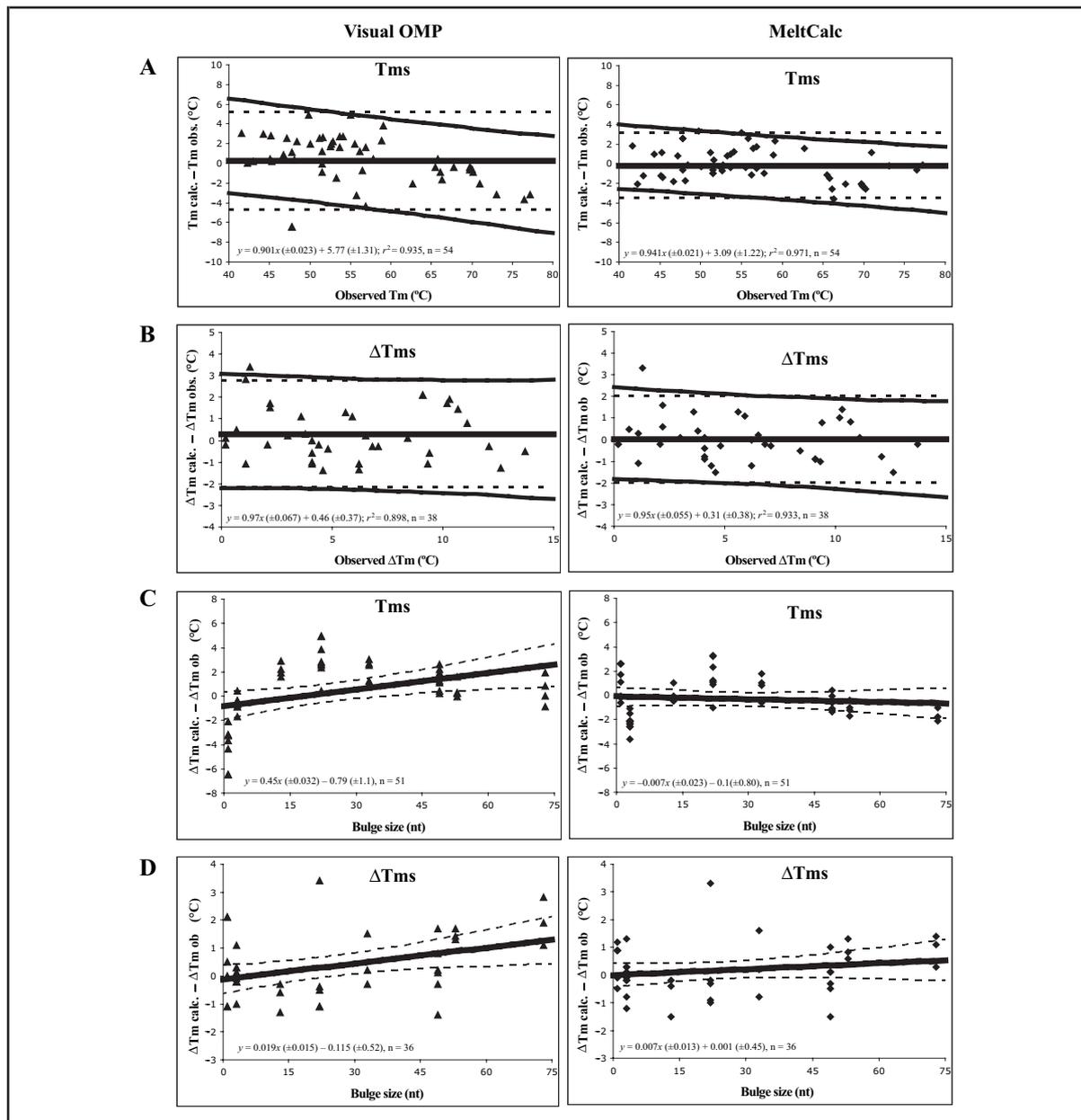


Fig. 1. Comparison of observed and predicted Tms and ΔTms using the Visual OMP and the MeltCalc software programs.

(A), Difference plots of calculated (calc.) minus observed (obs.) Tm vs observed Tm. Regression and correlation were calculated from the untransformed data; the thick solid lines shows the mean, dashed lines the 2SD of the transformed data. Fine solid lines shows the 95% CI of the Tm prediction. (B), Difference plots (calc. minus obs.) vs obs. ΔTm. Other descriptions as in (A). (C), Difference plots of Tms as function of bulge sizes. The linear regression is indicated by the solid line, with the 95% CI of the regression line marked with a dashed line (data from the probe that create 2 bulges in the template were not used, thus decreasing the number of data to n = 51). (D), Difference plots of ΔTms as a function of bulge sizes. Other descriptions as in (C) (2 bulges forming a probe not used, n = 36).

effect of differences in nucleotide sequences in the bulge. Bulged template sequences are not used in MeltCalc, and therefore bulge analysis could not be

performed with this software. With the use of 3 non-continuous probes, 46/79-GG in the *ADRB2* receptor and HbCmatch and HbEmatch in the *HBB* gene

Table 2. Differences between calculated and observed Tms and Δ Tms for each bulge size.

Bulge size, nt	Tms, n	Δ Tms, n	Tms, mean (SD) °C		Δ Tms, mean (SD) °C	
			Visual OMP	MetCalc	Visual OMP	MetCalc
1	8	6	-4.11 (1.60)	1.24 (1.39)	0.41 (1.44)	-0.58 (0.71)
3	11	6	-0.62 (0.5)	-2.24 (0.64)	-0.03 (0.69)	-0.13 (0.88)
13	4	3	2.15 (0.56)	0.03 (0.67)	-0.73 (0.51)	-0.70 (0.70)
22	8	6	3.04 (1.49)	1.51 (1.40)	0.02 (1.70)	0.12 (1.60)
33	4	3	2.35 (0.79)	0.75 (1.00)	0.47 (0.93)	0.33 (1.21)
49	8	6	1.41 (0.84)	-0.61 (0.60)	0.18 (1.04)	-0.18 (0.83)
53	4	3	0.49 (0.76)	-1.08 (0.54)	1.47 (0.21)	0.91 (0.36)
73	4	3	0.45 (1.19)	-1.4 (0.66)	1.93 (0.85)	0.93 (0.57)
Mean	51	36	0.65	0.28	0.36	-0.01
SD	51	36	2.51	1.67	1.23	1.02

(Table 1), on templates with a variant within the template bulge, we observed Tm shifts compared to WT templates (Fig. 2). The 46/79-GG LSProbe was designed to haplotype the 2 variants c.46A>G and c.79C>G of the *ADRB2* gene by binding to both loci and creating a 22-nt bulge of template sequence (1). Fig. 2A shows 5 samples heterozygous for the c.46G/79G (GG) and c.46A/C79C (AC) haplotypes. In 1 sample (black trace), a Tm shift of 2 °C (50.2 vs 48.2 °C) from wild-type samples can be observed. Sequencing showed that this sample had a c.66C>T variant in the 22-nt template bulge (AtC). The second example used 2 LSProbes that simultaneously genotyped 3 loci of the *HBB* gene and clearly identified the 4 possible *HBB* alleles (HbC, HbE, HbS, and WT) (Fig. 2B and (2)). One probe matched to the HbC allele and the other to the HbE allele. Forty nine nucleotides of template sequence are omitted between the HbS/C and the HbE loci. During validation of the assay, we tested 6 samples heterozygous for diverse variants in the 49-nt template bulge. The variants were not detected in 5 of the samples that appeared as wild type, but in one sample a Tm shift from WT was observed. The presence of a Coushatta variant (c.68A>C) on 1 chromosome, confirmed by sequencing, allowed a slight increase of Tm with both the probe matched to the HbC allele and the probe matched to the HbE allele (respectively 0.85 and 0.66 °C). Analysis of the Tm of these probes with the different templates using Visual OMP did not reveal these Tm variations. Our observation of differences in stability of the probes due to template sequence variants in the bulge indicate that in some cases, nontargeted polymorphisms can also be detected, increasing the analysis power of noncontinuous probes.

STRUCTURE PREDICTION BETWEEN PROBE AND TEMPLATE IN THE DESIGN OF NONCONTINUOUS PROBES

We explored the usefulness of using structure prediction provided by software in the design of noncontinuous probes used for multiplex genotyping and/or haplotyping. Both Visual OMP and MeltCalc allow prediction of the structure created by a noncontinuous probe annealed on its template. Visual OMP has the added feature of providing visual representation of the hybridization of the probe with its template.

As shown above, the *HBB*-HbE matched probe allows genotyping of the 4 *HBB* alleles HbE, HbS, HbC, and WT. The optimal structure predicted by Visual OMP between the probe and the template shows the probe hybridized on each side of the 49-nt template bulge (Fig. 3A). Fig. 3B shows data obtained with a probe that was shortened and had only 8 nt homologous to the template region containing the HbE allele. This probe does not distinguish the HbE genotype from WT, whereas the HbC and HbS alleles are clearly distinguished from WT and from each other. Therefore, this “short” probe allowed genotyping of the HbS and HbC alleles, localized 1 nt from each other, but did not allow simultaneous genotyping if the HbE locus that is 59 nt from the HbS locus. The most stable structure predicted for the short probe had a dangling end that does not bind on the HbE locus. A visual representation using Visual OMP is shown in Fig. 3. Hybridization of the template to only the 3' region of the short probe is more stable than a complete probe hybridization, owing to the relatively higher destabilizing effect of the bulge in this short probe compared to longer probes. Both Visual OMP and MeltCalc predicted this occurrence. This prediction was confirmed by the similar Tms observed for the HbE allele and the WT allele.

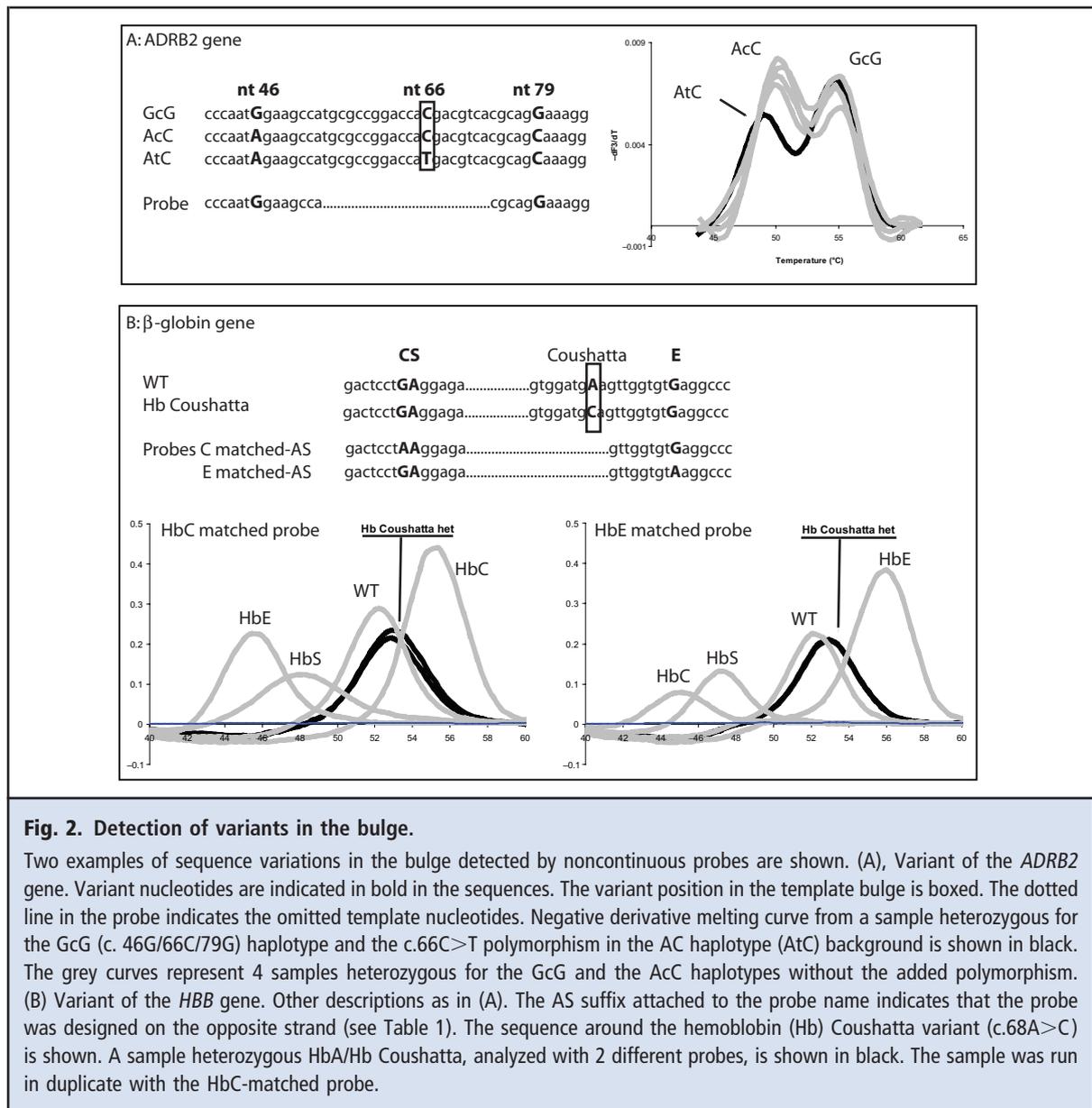


Fig. 2. Detection of variants in the bulge.

Two examples of sequence variations in the bulge detected by noncontinuous probes are shown. (A), Variant of the *ADRB2* gene. Variant nucleotides are indicated in bold in the sequences. The variant position in the template bulge is boxed. The dotted line in the probe indicates the omitted template nucleotides. Negative derivative melting curve from a sample heterozygous for the GcG (c. 46G/66C/79G) haplotype and the c.66C>T polymorphism in the AC haplotype (AtC) background is shown in black. The grey curves represent 4 samples heterozygous for the GcG and the AcC haplotypes without the added polymorphism. (B) Variant of the *HBB* gene. Other descriptions as in (A). The AS suffix attached to the probe name indicates that the probe was designed on the opposite strand (see Table 1). The sequence around the hemoglobin (Hb) Coushatta variant (c.68A>C) is shown. A sample heterozygous HbA/Hb Coushatta, analyzed with 2 different probes, is shown in black. The sample was run in duplicate with the HbC-matched probe.

Both software programs predicted that the short probe would fail to genotype the 3 loci simultaneously, and that additional nucleotides were needed to hybridize with the HbE locus to allow multiplex genotyping.

In the second example, we assessed the structure prediction of both programs in the design of more complex haplotyping probes. We have previously described a noncontinuous probe that spans 3 loci of the *ADRB2* gene and determines the 3 main haplotypes of the gene (22). These haplotypes are a combination of variants at the 3 positions c.1–20 C>T, c.46 G>A, and c.79 G>C. The haplotypes are c.1–20C/46G/79G (CGG), c.1–20T/46G/79C (TGC), or c.1–20T/46A/

79C (TAC). The probe was designed to anneal with the 3 loci and have 2 intervening sequences of 22 and 55 nt (1). The structure predicted by Visual OMP for the probe annealed to a perfectly matched template (CGG) is shown on the left of Fig. 4. The probe had 2 mismatches with TGC (not shown) and 3 mismatches with TAC (Fig. 4 right). Tms observed with the 3 templates allow clear distinction of the 3 haplotypes (Fig. 4, bottom). Both Visual OMP and MeltCalc predicted the probes to dangle and not anneal on the first mismatched nucleotide (C:A), as illustrated by the Visual OMP predicted structure with a TAC template that has a C:A mismatch, a G:T mismatch, and a G:G mismatch

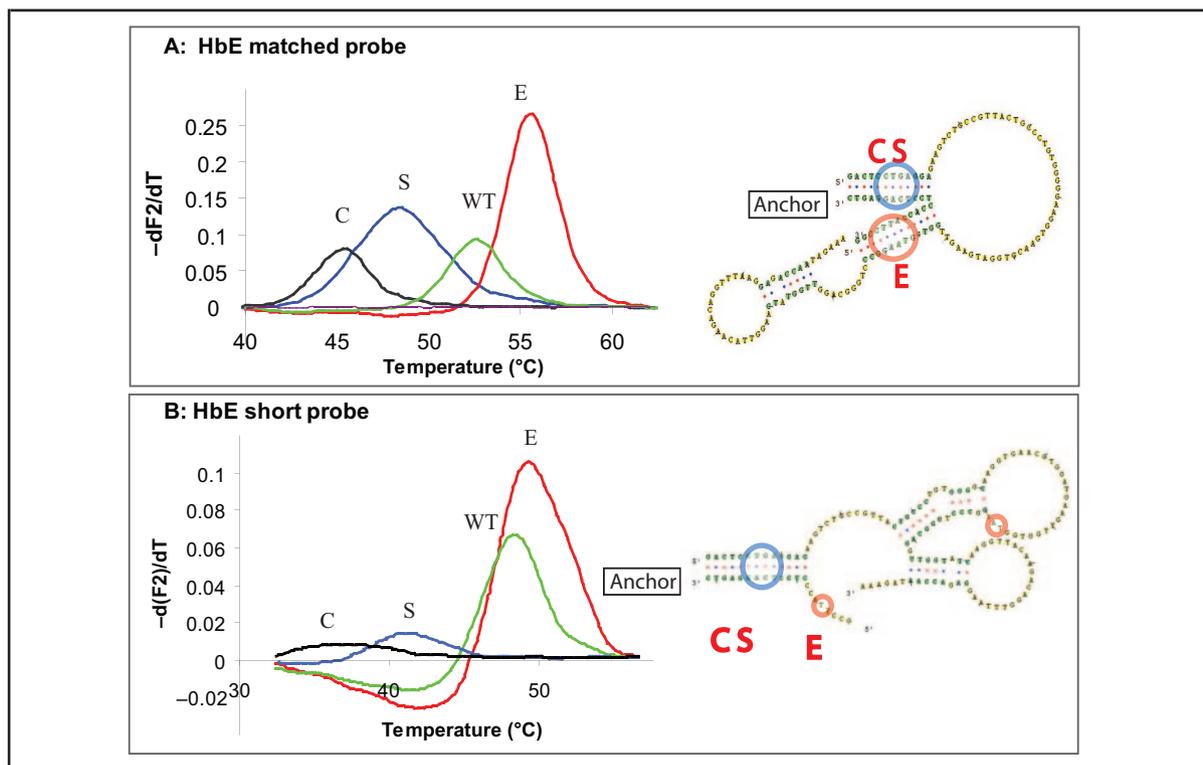


Fig. 3. Influence of the length of hybridization of a noncontinuous probe on the properties of the probe.

(A), Negative derivative melting curves from 4 hemoglobin alleles [HbE (red), WT (green), HbS (blue), and HbC (black)] hybridized with a noncontinuous probe homologous to the HbE allele. Positions of the 3 loci tested and hybridized on the template are indicated with colored circles on the Visual OMP–predicted structure. The HbC and HbS loci are 2 adjacent nucleotides circled in blue, and the HbE locus is circled in orange. Position of the anchor probe, which allows analysis of probe by fluorescence, is indicated. (B), Negative derivative melting curves from the same 4 hemoglobin alleles hybridized with a short probe homologous to the HbE allele. Note that on the Visual OMP the HbE region of the short probe doesn't anneal with the HbE locus on the template.

with the probe. The TGC template that does not have the G:T mismatch has an intermediate T_m . This example shows that even if the probe design does not allow complete hybridization of the noncontinuous probe with all the haplotypes, the differences in probe hybridization stability can still allow unambiguous haplotyping, as predicted by both Visual OMP and MeltCalc.

Discussion

Oligonucleotide probes that hybridize to specific loci have been used extensively for allele recognition. Usually the probe interrogates 1 sequence variant, thus allowing only genotyping, but occasionally several variants under 1 contiguous probe have been reported, enabling short-range haplotyping (23). Recently, probes have been described that, upon template hybridization, force the creation of a bulge in the probe (24, 25)

or in the template sequences (1–3, 26). Probes creating small bulges can be used to detect small deletion variants, or to mask nontargeted sequence changes under the probe. One major application of noncontinuous probes is the direct haplotyping of variants within 1 gene, which was shown to have advantages over statistical methods (27, 28). Molecular haplotyping normally requires extensive, costly, and laborious measures, such as cloning, ligation, and allele-specific PCR (29).

In the design of any probe system (e.g., FRET hybridization probes, molecular beacons, Eclipse probes, unlabeled probes) maximizing the effect of mismatch on probe stability is important and is accomplished by varying the length of the probe, by choosing the strand of DNA exhibiting the most destabilizing mismatch, by designing short probes with a minor groove binder moiety, or by incorporating nonconventional nucleotides such as locked nucleic acids (30). T_m prediction software is available for continuous probe design, but

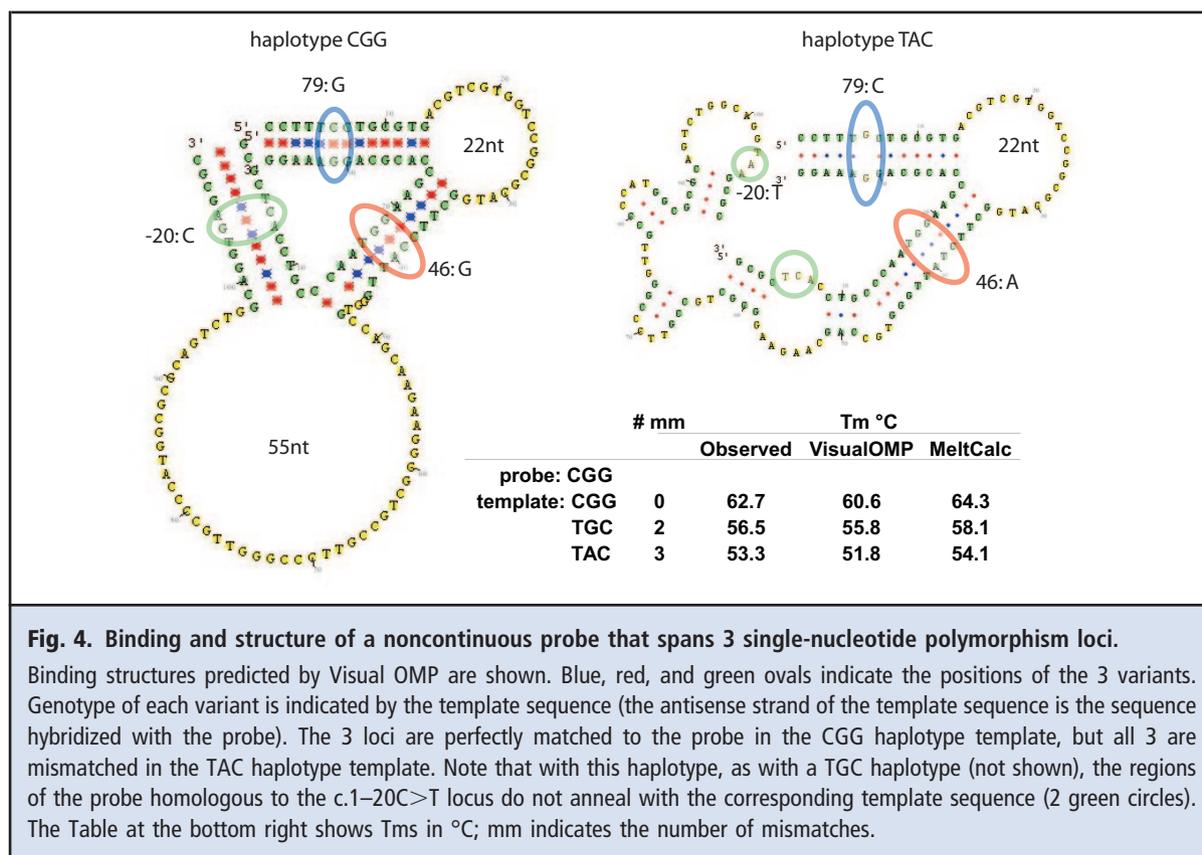


Fig. 4. Binding and structure of a noncontinuous probe that spans 3 single-nucleotide polymorphism loci.

Binding structures predicted by Visual OMP are shown. Blue, red, and green ovals indicate the positions of the 3 variants. Genotype of each variant is indicated by the template sequence (the antisense strand of the template sequence is the sequence hybridized with the probe). The 3 loci are perfectly matched to the probe in the CGG haplotype template, but all 3 are mismatched in the TAC haplotype template. Note that with this haplotype, as with a TGC haplotype (not shown), the regions of the probe homologous to the c.1-20C>T locus do not anneal with the corresponding template sequence (2 green circles). The Table at the bottom right shows Tms in °C; mm indicates the number of mismatches.

these programs cannot be used for the design of noncontinuous probes.

By comparing experimental Tm data from several noncontinuous probes to Tms predicted by software programs capable of incorporating the effect of bulges as well as mismatches, we tested the usefulness of software analysis in the design of noncontinuous probes. Also, because noncontinuous oligonucleotides could be used as PCR primers to avoid polymorphic loci, Tm prediction will help optimization of PCR using noncontinuous primers.

The noncontinuous probes ranged from 20–35 nt, which is in the range observed with continuous probes. The omitted template sequences were positioned in a central area of the probe. We found that in general, 11 nt of probe sequence on each side of the template bulge was sufficient for the probes to span several hybridization regions of the template. This number of nucleotides strictly depends on the nearest-neighbor composition of the duplex-forming area of the probe. In comparing the observed and the predicted Tms with software that uses the thermodynamic NN-model, we showed that a noncontinuous probe Tm prediction is possible with an average error of <1 °C. This is possible because the 2-state melting model can describe the un-

derlying biophysical mechanism. Additionally, the software can predict the Tm difference between loci matched and mismatched with the probes with an average error of <0.5 °C. As expected, the error range is narrower with Δ Tms than Tms because Δ Tms values are relative and depend less on experimental conditions, concentration of probes, and PCR product and bulge effects. For haplotyping and genotyping applications of noncontinuous probes, the Δ Tms rather than the absolute Tms determine how many different haplotypes and genotypes are distinguishable by the probe.

Structure prediction is also very useful in the design of noncontinuous probes. As demonstrated in Fig. 3, data obtained from a probe that will not allow hybridization across the different regions indicates the need for redesign of the probe (i.e., increase the nucleotide length on the dangling end of the probe). On the other hand, a combination of different mismatches and different structures might actually contribute to the differentiation of several haplotypes. The binding of the probe that spans 3 loci, and creates 2 bulges in the template sequence, is different on a matched template (CGG) than on templates with 2 or 3 mismatches (TGC and TAC) (Fig. 4). The presence of both mismatches and structure changes must contribute to the

T_m difference between the CGG and the TGC template, whereas the T_m difference between TGC and TAC is due only to additional mismatches between TAC and the probe, because the structures are the same.

Limitations of LSProbe design for haplotyping include the number of haplotypes that can be distinguished and the maximum number of loci tested. It seems reasonable to incorporate 2 bulges for haplotyping of a 3-loci block, as shown with the ADRB2 20/46/79 probe, because only 4 haplotypes have been described at this locus. In theory, a variant at 3 loci can lead to up to 8 haplotypes, and the probe might not have the ability to distinguish all 8 haplotypes, although use of high-resolution instruments could increase the number of distinguishable haplotypes. For diagnostic purposes, it is usually not necessary to distinguish all possible haplotypes, because some will not naturally occur.

In the majority of the samples that had a known polymorphism in the bulge, the polymorphism did not alter the T_m of the probe, as expected. On the other hand, in the 2 cases in which we observed T_m variations, slight changes in secondary structures inside the bulge could have affected the probe T_m. No secondary structures were predicted in the bulge, and Visual OMP did not predict the effect of the sequence variant. As with continuous hybridization probes or unlabeled

probes, samples with T_ms that differ significantly from the expected T_m (>2SD of the within run precision) should be sequenced for verification.

From our observations, the following guidelines can be proposed for the design of successful noncontinuous probes used for haplotyping and multiplex genotyping: (a) like with continuous probes, the probe should be as short as possible to maximize ΔT_m and the number of alleles they can distinguish; (b) the occurrence of dangling ends should be avoided for the probe to interrogate several loci simultaneously; and (c) the position of the mismatch should be close to the middle of the hybridized stems.

Software prediction that includes the effect of bulges in addition of the effect of mismatches should be very useful in the design of noncontinuous probes, specifically when large (>3 nt) bulges are expected and when multiple loci, genotypes or haplotypes are interrogated.

Grant/Funding Support: This work was supported by the ARUP Institute for Clinical and Experimental Pathology®.

Financial Disclosures: Dr. E. Schütz holds copyrights for the MeltCalc software.

Acknowledgments: We wish to thank Astrid Thuid for help with the Visual OMP software and C. Wittwer for critical reading of the manuscript.

References

1. Pont-Kingdon G, Lyon E. Direct molecular haplotyping by melting curve analysis of hybridization probes: beta 2-adrenergic receptor haplotypes as an example. *Nucleic Acids Res* 2005;33:e89.
2. Pont-Kingdon G, Chou LS, Damjanovich K, Sumner K, Herrmann M, Erali M, Lyon E. Multiplex genotyping by melting analysis of loci-spanning probes: beta-globin as an example. *Biotechniques* 2007;42:193–7.
3. Margraf RL, Mao R, Wittwer CT. Masking selected sequence variation by incorporating mismatches into melting analysis probes. *Hum Mutat* 2006;27:269–78.
4. Allawi HT, SantaLucia J Jr. Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry* 1997;36:10581–94.
5. Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics of internal A.C mismatches in DNA: sequence dependence and pH effects. *Biochemistry* 1998;37:9435–44.
6. Allawi HT, SantaLucia J Jr. Thermodynamics of internal C.T mismatches in DNA. *Nucleic Acids Res* 1998;26:2694–701.
7. Allawi HT, SantaLucia J Jr. Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA. *Biochemistry* 1998;37:2170–9.
8. Bommarito S, Peyret N, SantaLucia J Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res* 2000;28:1929–34.
9. Bonnet G, Tyagi S, Libchaber A, Kramer FR. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc Natl Acad Sci USA* 1999;96:6171–6.
10. Dimitrov RA, Zuker M. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J* 2004;87:215–26.
11. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* 1999;38:3468–77.
12. SantaLucia J Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA* 1998;95:1460–5.
13. Tanaka F, Kameda A, Yamamoto M, Ohuchi A. Thermodynamic parameters based on a nearest-neighbor model for DNA sequences with a single-bulge loop. *Biochemistry* 2004;43:7143–50.
14. Schutz E, von Ahsen N. Spreadsheet software for thermodynamic melting point prediction of oligonucleotide hybridization with and without mismatches. *Biotechniques* 1999;27:1218–22, 24.
15. von Ahsen N, Schutz E. Using the nearest neighbor model for the estimation of matched and mismatched hybridization probe melting points and selection of optimal probes on the lightcycler. In: Meurer S, Wittwer C, Nakagawara K, editors. *Rapid cycle real-time pcr: methods and applications*. Vol. 1. Berlin: Springer-Verlag 2001: 43–56.
16. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406–15.
17. Pont-Kingdon G, Millson A, Lyon E. Molecular haplotyping by real-time PCR. Norfolk, UK: Horizon Scientific Press. In Press.
18. Zhou L, Myers AN, Vandersteen JG, Wang L, Wittwer CT. Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clin Chem* 2004;50:1328–35.
19. von Ahsen N, Wittwer CT, Schutz E. Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg(2+), deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clin Chem* 2001;47:1956–61.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
21. Kriegesmann B, Dierkes B, Leeb T, Jansen S, Brenig B. Two breed-specific bovine MC1-R alleles in Brown Swiss and Saler breeds. *J Dairy Sci* 2001;84:1768–71.
22. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, et al. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci USA* 2000;97:10483–8.
23. Song Y, Araki J, Zhang L, Froehlich T, Sawabe M,

- Arai T, et al. Haplotyping of TNFalpha gene promoter using melting temperature analysis: detection of a novel -856(G/A) mutation. *Tissue Antigens* 2005; 66:284-90.
24. Millson A, Pont-Kingdon G, Page S, Lyon E. Direct molecular haplotyping of the IVS-8 poly(TG) and polyT repeat tracts in the cystic fibrosis gene by melting curve analysis of hybridization probes. *Clin Chem* 2005;51:1619-23.
25. von Ahnen N, Oellerich M, Schutz E. DNA base bulge vs unmatched end formation in probe-based diagnostic insertion/deletion genotyping: genotyping the UGT1A1 (TA)(n) polymorphism by real-time fluorescence PCR. *Clin Chem* 2000;46:1939-45.
26. Schnittger S, Schoch C, Kern W, Mecucci C, Tschulik C, Martelli MF, et al. Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. *Blood* 2005;106:3733-9.
27. Douglas JA, Boehnke M, Gillanders E, Trent JM, Gruber SB. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat Genet* 2001;28:361-4.
28. Thomas S, Porteous D, Visscher PM. Power of direct vs. indirect haplotyping in association studies. *Genet Epidemiol* 2004;26:116-24.
29. McDonald OG, Krynetski EY, Evans WE. Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. *Pharmacogenetics* 2002;12:93-9.
30. Chou LS, Meadows C, Wittwer CT, Lyon E. Unlabeled oligonucleotide probes modified with locked nucleic acids for improved mismatch discrimination in genotyping by melting analysis. *Biotechniques* 2005;39:644, 6, 8 passim.