

Thermodynamic Behavior of Short Oligonucleotides in Microarray Hybridizations Can Be Described Using Gibbs Free Energy in a Nearest-Neighbor Model

Stefan Weckx,^{†,‡} Enrico Carlon,^{§,||,⊥} Luc De Vuyst,[‡] and Paul Van Hummelen^{*,†,‡,#}

MicroArray Facility, VIB, Herestraat 49, B-3000 Leuven, Belgium, Interdisciplinary Research Institute, Cité Scientifique BP 60069, F-59652 Villeneuve d'Ascq, France, Ecole Polytechnique Universitaire de Lille, Cité Scientifique, F-59655 Villeneuve d'Ascq, France, Institute for Theoretical Physics, K. U. Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium, Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium, and MicroArray Facility, K. U. Leuven, Herestraat 49, B-3000 Leuven, Belgium

Received: July 4, 2007; In Final Form: September 21, 2007

While designing oligonucleotide-based microarrays, cross-hybridization between surface-bound oligos and non-intended labeled targets is probably the most difficult parameter to predict. Although literature describes rules-of-thumb concerning oligo length, overall similarity, and continuous stretches, the final behavior is difficult to predict. The aim of this study was to investigate the effect of well-defined mismatches on hybridization specificity using CodeLink Activated Slides and to study quantitatively the relation between hybridization intensity and Gibbs free energy (ΔG), taking the mismatches into account. Our data clearly showed a correlation between the hybridization intensity and ΔG of the oligos over 3 orders of magnitude for the hybridization intensity, which could be described by the Langmuir model. As ΔG was calculated according to the nearest-neighbor model, using values related to DNA hybridizations in solution, this study clearly shows that target–probe hybridizations on microarrays with a three-dimensional coating are in quantitative agreement with the corresponding reaction in solution. These results can be interesting for some practical applications. The correlation between intensity and ΔG can be used in quality control of microarray hybridizations by designing probes and corresponding RNA spikes with a range of ΔG values. Furthermore, this correlation might be of use to fine-tune oligonucleotide design algorithms in a way to improve the prediction of the influence of mismatching targets on microarray hybridizations.

Introduction

In transcriptome research, microarrays are used to indicate genes expressed by a given cell type or organism. The overall expression pattern is in most cases compared to other patterns to reveal differentially expressed genes between different states or experimental conditions. Microarray technology is based on nucleotide–nucleotide hybridizations, where one molecule (the probe) is attached to the surface of a slide, and the other molecule (the fluorescently labeled target) moves freely in solution and will bind to its complementary probe. To obtain a microarray with high sensitivity and specificity, the design of probe sequences is an important step. A good sensitivity results in intensities proportional to the real amount of labeled target in the hybridization solution. To obtain a good specificity, homology should be high for the intended target sequence but low for all other target sequences in solution to avoid cross-hybridization of non-intended similar sequences.

Previous work described in the literature presents hybridization studies with respect to sensitivity, specificity, cross-hybridization, mismatches, oligo length, and number of oligos

per gene.^{1–5} Kane et al.¹ demonstrated that 50-mer oligonucleotides have a comparable sensitivity as PCR-derived probes of 300–400 nucleotides and that 50-mers show good specificity as long as the similarity with non-intended sequences is <75% or in the absence of a stretch of 15 perfectly matching bases. A comparable similarity was reported by Hughes et al.,² who performed hybridizations with permutations-containing 60-mer oligonucleotides and found that a similarity between probe and target of less than 77% resulted in a lack of signal. Although Tiquia et al.⁴ showed a 10-fold higher hybridization intensity for 60-mers as compared to 30-mers, according to the results of Relógio et al.,³ the length of oligonucleotides can be decreased to 30-mers with the preservation of a good balance between sensitivity and specificity. These results provide a set of rules-of-thumb to be taken into account in the oligo design process: oligonucleotides should be at least 30 bp long, should have a maximal similarity of 75% with non-intended similar sequences, and may not possess perfect matching stretches longer than 15 bp with these non-intended sequences.

Whereas the former results were based on the outcome of hybridization experiments only, thermodynamics has been taken into account in several other studies.^{6–19} Indeed, hybridization of nucleotide molecules in solution is a well-understood biophysical process driven by thermodynamics.²⁰ Experiments over the past two decades on oligonucleotide melting have provided precise measurements of thermodynamic parameters, such as the binding free energy (ΔG), involved in the process.²¹ These experiments have validated a simple model known as

* Corresponding author. Tel.: +32 16 347939; fax: +32 16 347940; e-mail: paul.vanhummelen@vib.be.

[†] MicroArray Facility, VIB.

[‡] Research Group of Industrial Microbiology and Food Biotechnology.

[§] Interdisciplinary Research Institute c/o IEMN.

^{||} Universitaire de Lille.

[⊥] Institute for Theoretical Physics.

[#] MicroArray Facility, K. U. Leuven.

```

5-AGCTTTCATTTACCTAAATTA AAAAC TTTTATATAA TTTTGTCAATATTTAAACATAATCTTGCATATT-3
1# 3-TTTTGGAAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
2# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
3# 3-TTTTGGAAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
4# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
5# 3-TTTTGGAAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
6# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
7# 3-TTTTGGAAAATATTTAAACAGTTATAA-5-TAG-C6amino-----
8# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
9# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
10# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
11# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----
12# 3-TTTTGGAAAATATATTTAAACAGTTATAA-5-TAG-C6amino-----

```

Figure 1. Oligo set 1 (GC content = 10%, $T_m = 47^\circ\text{C}$) as an example for the eight oligo sets, containing 12 oligos with 0, 1, 2, or 3 MM at well-chosen positions. The mismatches are highlighted using a black background. The position of the 12-mer universal spacer is indicated by TAG. The uppermost sequence is the fluorescently labeled 70-mer target sequence. Apart from differences in the sequences, all the other oligo sets have mismatches positioned as in the example given here.

TABLE 1: Overview of GC Content, T_m , and Binding Free Energy of the Eight Oligo Sets Used (Determined for PM Oligo)^a

	perfect matching oligo (5′–3′)	GC (%)	T_m (°C)	ΔG_{45} (kcal/mol)
set 1	AATATTGACAAAATTATATAAAAAGTTTTT	10.0	48	24.14
set 2	AAATATAACGTATCCTGCGTGTACACAT	33.3	56	30.76
set 3	CCAAATTGCGCAATTTCCGCCACCATGTCAG	50.0	63	36.98
set 4	TCCATGTCAGCCTTGGTTGACATGTAGCGT	50.0	63	36.61
set 5	CGCAAAAAGTTCAAAAAGCAGCGGCTGCTCA	50.0	63	37.88
set 6	TCATCATGGCTTGTCCTCCCTTTCACATA	50.0	63	35.71
set 7	CGGCCACACCCCTGATGGCGTTACCCAT	66.7	70	41.07
set 8	CCCGTGGCACTCCGGCCATGGCGACCGGC	80.0	75	45.67

^a T_m is calculated with a simple empirical formula (<http://www.promega.com/biomath/>), while ΔG is calculated from the NN model²² at a hybridization temperature of 45 °C.

the nearest-neighbor (NN) model, where ΔG is calculated as the sum of ΔG values for each couple of two neighboring bases in a sequence.²¹ The free energy for mismatching bases in DNA sequences has been investigated as well, and ΔG values are described that can be used to calculate the overall ΔG of a mismatch-containing DNA sequence.^{22–23} Although there is a general consensus of the validity of the NN model for perfect-matching duplexes, the case of mismatches has been more debated. Indeed, some studies reported that a single mismatch may have an effect that goes beyond the nearest-neighboring bases.²⁴ However, these effects have not yet been systematically quantified.

Although previous studies reported on the influence of ΔG on hybridization intensity,^{15,17,19} and on the use of ΔG to select suitable oligos,^{10,16} it has been subject to debate as to whether these thermodynamic parameters, determined for DNA sequences in solution, apply for microarray hybridizations where one of the molecules is attached to a surface and hence prone to steric hindrance. Fotin et al.⁶ performed one of the first studies to quantify molecular interactions in a gel pad microarray. They found a convincing linear relationship between microarray hybridization free energies ($\Delta G(\text{chip})$) and the corresponding free energies in solution ($\Delta G(\text{solution})$). Various thermodynamics-inspired approaches have also been invoked in the analysis of microarray data (see the review of Halperin et al.²⁵ and references therein), which focused in particular on Affymetrix GeneChips. These approaches relied on the so-called Langmuir isotherm, but there is no consensus as to whether one can correlate $\Delta G(\text{chip})$ and $\Delta G(\text{solution})$. Pozhitkov et al.¹⁸ reported a poor statistical relationship between hybridization intensity and $\Delta G(\text{solution})$ and also stated that microarray-based hybridizations using mismatch-containing oligos did differ from solution-based hybridizations.

These somewhat controversial results call for more experimental investigations aimed at a better understanding and at a quantification of molecular interactions in microarrays. The aim of the present study was to assess the effect of well-defined mismatches on hybridization specificity using a spotted mi-

croarray consisting of eight oligo sets with varying GC content and to study quantitatively the relation between hybridization intensity and $\Delta G(\text{solution})$, calculated from the NN model, taking mismatches into account.

Materials and Methods

Oligo Sets. The genomic sequence of *Lactobacillus plantarum* was used to search for eight 30-mer sequences with a GC content ranging from 10 to 80% (Table 1). From each oligo, 11 additional oligos were derived by introducing 1, 2, or 3 mismatches (MM) at well-defined places, allowing evaluation of the mismatch position and inter-mismatch distance. The mismatches were of the type AA, CC, GG, or TT. These 12 oligos made up one oligo set. For each set, a 70-mer complementary target sequence was chosen, consisting of the 30-bp region of the 30-mer plus 20 bp up- and downstream of the 30-mer (Figure 1 and Supporting Information). To study the influence of a spacer sequence between the oligo and the slide surface, a related oligo set was derived containing an additional 12-mer universal spacer at the 5′ end. This universal spacer had no homology to any other sequence in these sets and was the same for all oligos. The 192 oligos were obtained from Eurogentec (Seraing, Belgium) with 5′ amino modification. The eight 70-mer target sequences were obtained from the same company with a 5′ Cy3 modification and were dissolved in TE buffer pH 8.0 until a final concentration of 100 μM .

Microarray Production. The 192 oligos were diluted to a concentration of 25 pmol/ μL in 100 mM sodium-phosphate buffer, 0.000675% sarcosyl (Sigma-Aldrich, Milwaukee, WI). A MicroGrid II spotter (Genomic Solutions, Huntingdon, U.K.) was used to spot the oligos using 10K pins on CodeLink Activated Slides (GE Healthcare, Bucks, U.K.). Each oligo was spotted 12 times on each slide using the same pin. After spotting, the slides were placed in a chamber containing saturated NaCl for a minimum of 24 h. The active groups on the slides were blocked with preheated blocking solution (50 mM ethanolamine, 0.1 M Tris, pH 9.0) at 55 °C for 30 min. Slides were rinsed

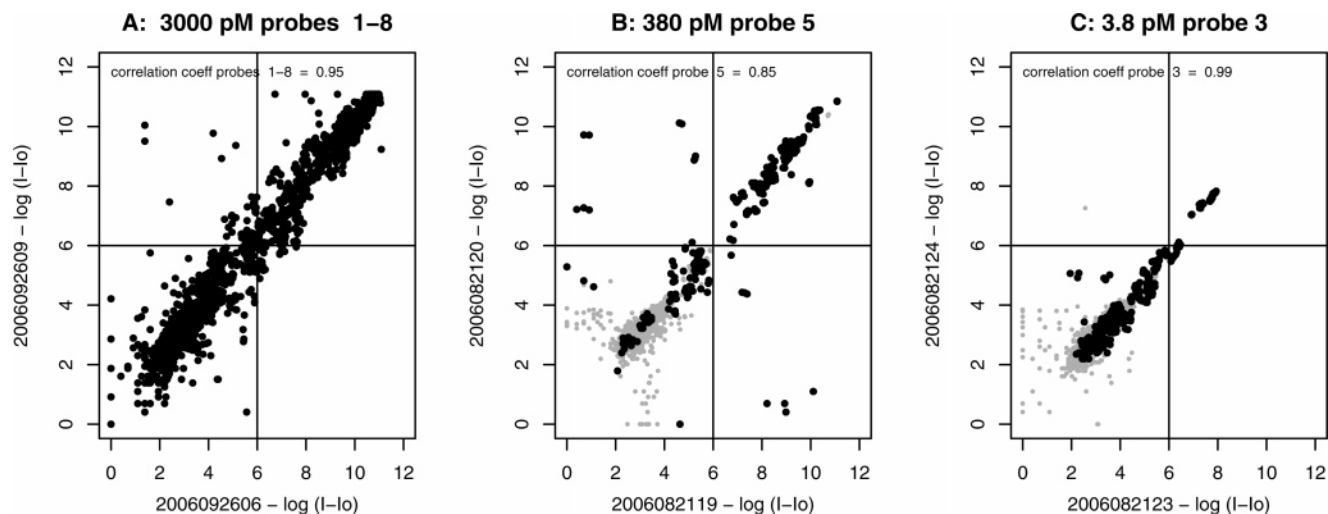


Figure 2. Reproducibility plots of background-corrected log intensities of replicate hybridizations. The data points in the plots represent the average value for all 12 spots per oligo. Panel A: using all eight targets at a concentration of 3000 pM; panel B: only using target 5 (in black), concentration = 380 pM; and panel C: only using target 3 (in black), concentration = 3.8 pM.

twice with ultrapure water and washed with preheated 4x SSC-0.1% SDS at 55 °C for 30 min in a shaking water bath. Slides were rinsed with ultrapure water and dried by centrifugation at 800 rpm for 3 min.

Hybridizations. Hybridization mixtures containing one or several 70-mer targets were prepared in 210 μ L of hybridization buffer (GE Healthcare) containing 50% formamide (Sigma-Aldrich). The hybridization mixtures were denatured at 96 °C for 3 min and immediately put on ice for at least 5 min. Hybridization mixtures were kept at 42 °C for 5 min and subsequently centrifuged at 12 000 rpm for 5 min. Slides were hybridized at 45 °C during 12 h and washed in 1x SSC/0.2% SDS for 10 min, 0.1x SSC/0.2% SDS for 10 min, and 0.1x SSC for 4 min, using the Lucidea Slidepro (LSP) hybridization station (GE Healthcare). Slides were dried by the LSP using isopropyl alcohol and air. Subsequently, slides were scanned using the Agilent scanner (Agilent Technologies, Palo Alto, CA) at 10 μ m, and images were analyzed using ArrayVision v7 (GE Healthcare).

Data Processing. Spot intensity was regarded as above background if the intensity was larger than the sum of the local background signals and twice the background's standard deviation. The average oligo intensity and average background were calculated when at least 10 of the 12 replicate spots had an intensity above background. The average value for the 12 replicate spots was calculated taking a confidence interval of 95% into account.

The Gibbs free energy was calculated from the NN model using the experimental stacking free energies obtained from DNA melting experiments in solution, as reported by Peyret et al.²² It is known²² that terminal mismatches (e.g., mismatches close to double helix edges) are more stable than internal mismatches, although this effect has not yet been quantified. The values for oligos 10 and 12, containing a mismatch in the third nucleotide at the 3' and 5' ends (Figure 1), should therefore be seen as an approximation.

Langmuir Model Applied to Microarray Hybridizations. The simplest thermodynamic model for hybridization in DNA microarrays is the Langmuir model,²⁵ which predicts a spot intensity as given by eq 1

$$I = I_0 + \frac{Ace^{\Delta G/RT}}{1 + ce^{\Delta G/RT}} \approx I_0 + Ace^{\Delta G/RT} \quad (1)$$

where I_0 is the non-specific background signal, c is the target concentration in solution, T is the experimental temperature, R is the universal gas constant, and ΔG is the binding free energy. In the limit of small concentrations and weak free energies, $ce^{\Delta G/RT} \ll 1$, the denominator in eq 1 can be neglected. In this limit, the intensity is proportional to the target concentration (a measure of the gene expression level), which is usually assumed when analyzing microarray data. In the opposite limit, the spot saturates; there are no longer free single stranded probes available for hybridization, and the intensity attains a maximal value $I = I_0 + A$. Usually, one assumes that the hybridization is a two-state process, in which either target and probe are unbound or fully bound so that the description of the process is restricted to two thermodynamic states.

Results

Hybridizations and Reproducibility. In total, 18 hybridizations were performed using nine different target mixtures. One mixture contained all eight labeled target molecules simultaneously at a concentration of 3000 pM. The other mixtures contained targets 3–6 separately at a final concentration of 380 or 3.8 pM. Each of the mixtures was prepared and hybridized in duplicate.

The reproducibility of the hybridizations is illustrated in Figure 2. To calculate the correlation coefficient for the 3000 pM plot, spots with an intensity close to the maximal detectable intensity by the scanner were excluded. The hybridizations using all targets at a concentration of 3000 pM showed a correlation coefficient of 88%. The coefficients for the other hybridizations ranged from 85 to 99% (data not shown).

GC Content and Spacer. The bar charts in Figure 3 give a detailed view on the effect of the GC content showing log-transformed hybridization intensities (in relative fluorescent units or RFU) from a 3000 pM hybridization for six representative types of oligos. Figure 3A shows the intensities for the perfect matching (PM) oligo (oligo 1) of all oligo sets. The intensities for all oligos except sets 1 and 2 were at the maximal detection limit of the scanner. Figure 3B shows the intensities for oligo 3 containing 1 MM in the middle of the oligo. The intensities for the low GC oligo sets 1 and 2 are low (only one out of four was above the background cutoff), the intensities for the oligo sets with a GC content of 50% (sets 3–6) were not at the maximal level, and the intensities of the high GC oligos were

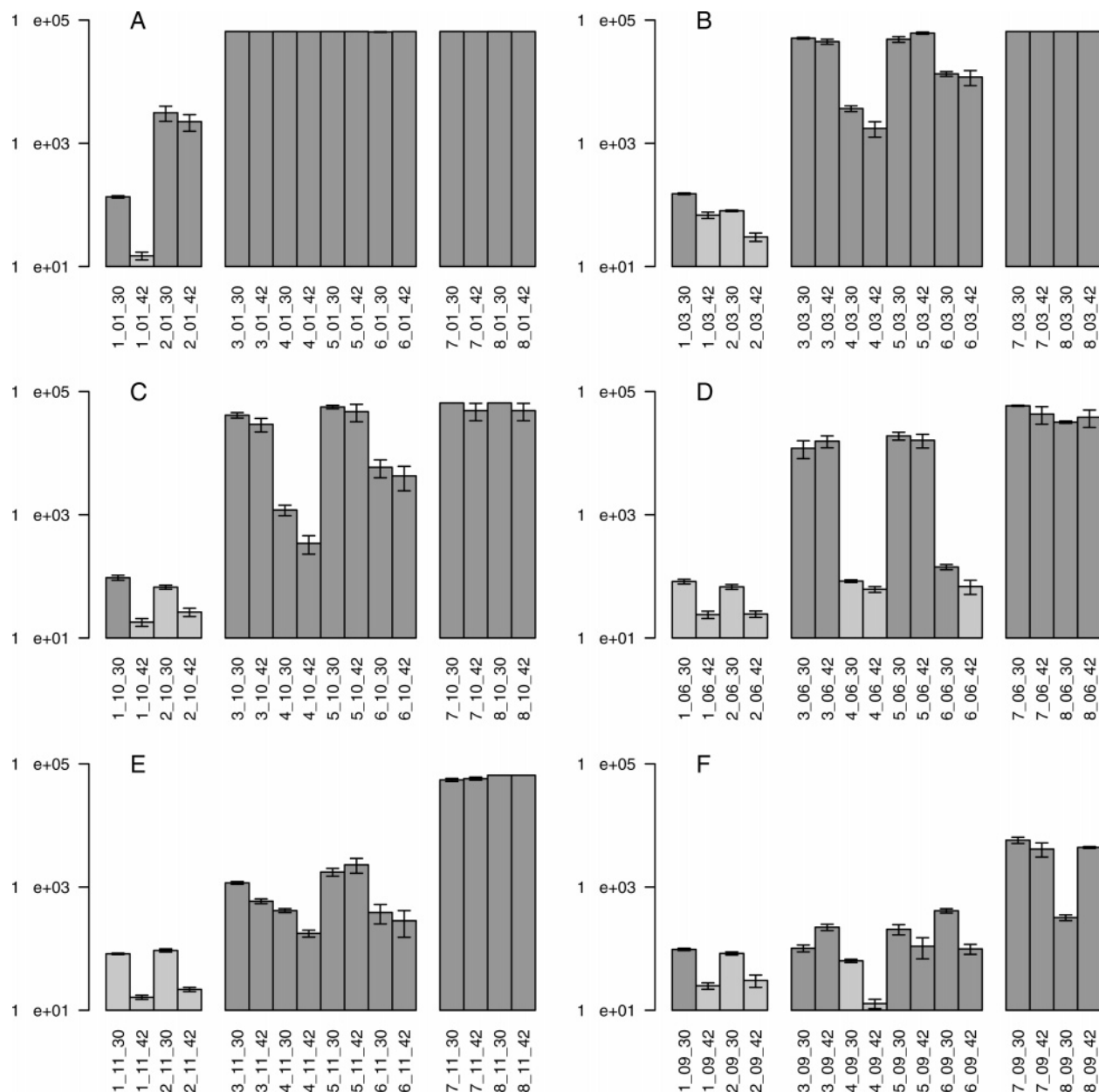


Figure 3. Bar charts showing log intensities for six types of oligos. Each bar chart shows the log intensity of one type of oligo (30-mer and 42-mer) in all eight oligo sets. The average of 12 spots gives the height of the bars, and the flags indicate a 95% confidence interval. Light gray bars denote intensities that are below the background cutoff, defined as the average background signal for all 12 spots plus twice the standard deviation of this average background signal. The name of the oligos is composed of three numbers, separated by an underscore: the first point is the oligo set number, the second is the number of the oligo in each set (see also Figure 1 and the Supporting Information), and the third gives the length of the oligo: oligo length 42 stands for the 30-mer oligo with 12-mer spacer attached. (A) Oligo 1, PM; (B) oligo 3, 1 MM in the middle; (C) oligo 10, 2 MM and perfect matching stretch of 23 bp; (D) oligo 6, 2 MM with a perfect matching stretch of 16 bp in between; (E) oligo 11, 2 MM in the middle; and (F) oligo 9, 3 MM.

at the maximal detection level. Figure 3C–E shows the intensities of three different types of oligos containing 2 MM. Oligo 10 (Figure 3C) had the 2 MM close to each other (3 bp in between) at the 3' end, the free end of the oligo. In oligo 6 (Figure 3D), the 2 MM were separated by a perfect matching stretch of 16 bp, and oligo 11 had the 2 MM close to each other (3 bp in between) near the middle of the oligo. All three bar charts show the same tendencies: the oligos with a low GC content (oligo sets 1 and 2) had low intensities, of which most were below the background cutoff. The oligos with a high GC content (oligo sets 7 and 8) had high intensities, nearly at the maximal level. The intensities of the oligos with a GC content of 50% showed a good variation in specificity. From

these data, it became clear that oligo sets 3 and 5 showed overall higher intensities than oligo sets 4 and 6. Also, the intensities for oligo 10, containing a stretch of 23 PM bases was higher than the intensities for oligo 6 containing a stretch of 16 PM bases, which were in turn higher than the intensities for oligo 11, in which the 2 MM were close to each other in the middle of the oligo. Oligo 9 (Figure 3F), containing 3 MM, had intensities at the background level. Oligos with a high GC content were below the maximal level but still considerably above the intensities of the other oligos.

All eight oligo sets consisted of a subset of 30-mer oligos and a subset of the 30-mer oligos with a 12-mer universal spacer at the 5' end. As becomes clear from the bar charts, the

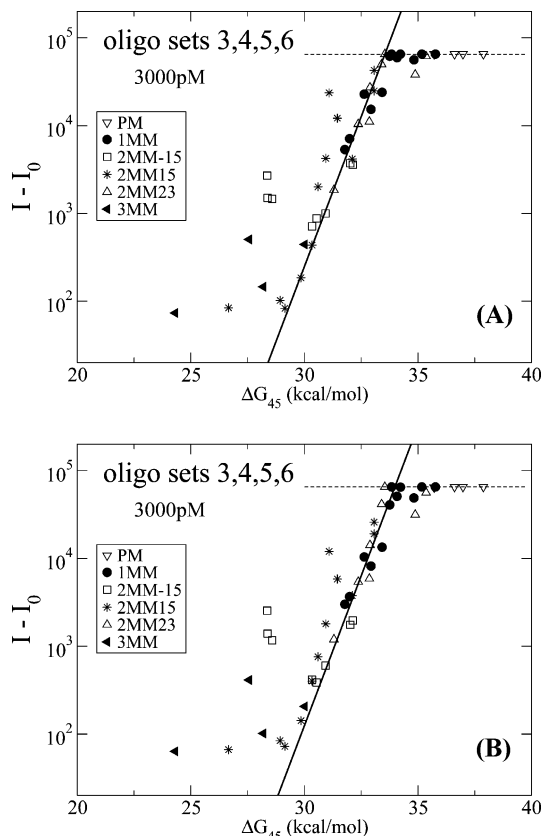


Figure 4. Two graphs (A and B) representing log-scale background-corrected hybridization intensities as a function of the binding free energy (ΔG) for two hybridizations where all eight targets were used at a concentration of 3000 pM. Only the data points for the oligo sets with a GC content of 50% are plotted. The data points in each graph represent the average of 12 replicate spots of the oligo. The six different kinds of oligos regarding the amount of mismatches and their relative positions are indicated by the six symbols as represented in the legend, where PM stands for the perfect matching oligo 1, 1 MM for oligos 3–5 containing 1 mismatch, 2 MM-15 for oligos with 2 MM with an inter-mismatch distance smaller than 15 (oligos 2 and 11), 2 MM15 for oligos 6–8 with an inter-mismatch distance of 15 bases, 2 MM23 for oligos 10 and 12 with an inter-mismatch distance of 23 bases, and 3 MM for oligo 9 containing 3 MM. The solid line has a slope equal to $1/RT$, where T is the experimental temperature ($T = 45^\circ\text{C} = 318\text{K}$). The horizontal dashed line corresponds to saturation level.

intensities of the 30-mers with universal spacer were slightly lower than the intensities of the 30-mer, although these differences were in most cases within the 95% confidence intervals.

Relationship between Binding Free Energy and Hybridization Intensity: Langmuir Model. Figures 4–6 show log-transformed background-corrected hybridization intensities for the 30-mer oligonucleotides (in RFU, y-axis) as a function of the Gibbs free energy (ΔG , x-axis), calculated using the NN model based on ΔG values for hybridization in solution. In what follows, if not explicitly stated, we use ΔG as shorthand for $\Delta G(\text{solution})$. The free energy is obtained from ΔH and ΔS parameters as described by Peyret et al.²² using the formula $\Delta G = \Delta H - T\Delta S$, where $T = 318\text{K}$ (45°C). To evaluate the Langmuir model in a non-saturating environment, a line with slope $1/RT = 1.6\text{ mol/kcal}$ (with $T = 318\text{K}$ and $R = 0.002\text{ kcal/(mol K)}$) was plotted as a guide to the eye.

Figure 4 contains two graphs, based on data of two different hybridizations at 3000 pM, in which all eight targets were present. Only the data points for the oligo sets with a GC content of 50% were plotted (oligo sets 3–6). PM oligos (oligo 1 in

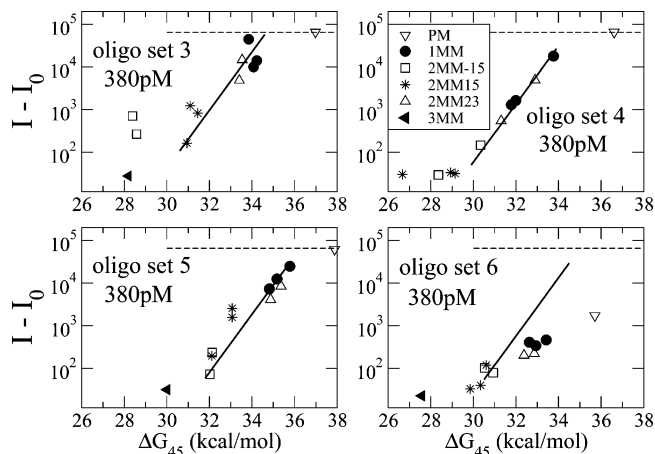


Figure 5. Graph representing log-scale background-corrected hybridization intensities as a function of the binding free energy (ΔG) for four hybridizations with a target concentration of 380 pM. Each of the four panels represents data from separate hybridizations, each for one of the four oligo sets with a GC content of 50%. The data points in each graph represent an average of 12 replicate spots of the oligo. The six different kinds of oligos regarding the amount of mismatches and their relative positions are indicated by the six symbols as represented in the legend, where PM stands for the perfect matching oligo 1, 1 MM for oligos 3–5 containing 1 MM, 2 MM-15 for oligos with 2 MM with an inter-mismatch distance smaller than 15 (oligos 2 and 11), 2 MM15 for oligos 6–8 with an inter-mismatch distance of 15 bases, 2 MM23 for oligos 10 and 12 with an inter-mismatch distance of 23 bases, and 3 MM for oligo 9 containing 3 MM. The solid line has a slope of $1/RT$. The horizontal dashed line corresponds to saturation level.

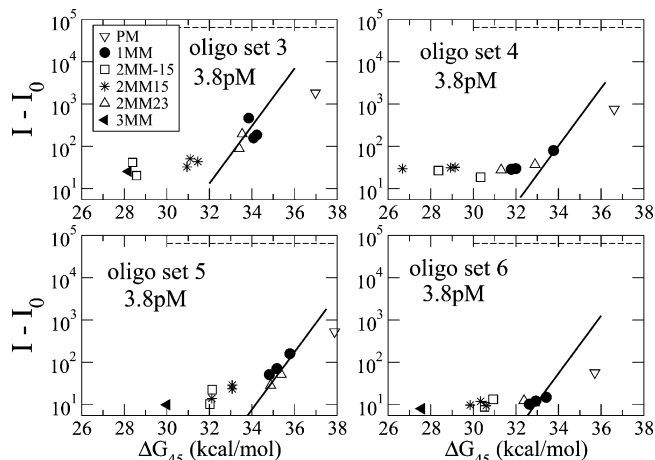


Figure 6. Graph representing log-scale background-corrected hybridization intensities as a function of the binding free energy (ΔG) for four hybridizations with a target concentration of 3.8 pM. Each of the four panels represents data from separate hybridizations, each for one of the four oligo sets with a GC content of 50%. The data points in each graph represent the average of 12 replicate spots of the oligo. The six different kinds of oligos regarding the amount of mismatches and their relative positions are indicated by the six symbols as represented in the legend, where PM stands for the perfect matching oligo 1, 1 MM for oligos 3–5 containing 1 MM, 2 MM-15 for oligos with 2 MM with an inter-mismatch distance smaller than 15 (oligos 2 and 11), 2 MM15 for oligos 6–8 with an inter-mismatch distance of 15 bases, 2 MM23 for oligos 10 and 12 with an inter-mismatch distance of 23 bases, and 3 MM for oligo 9 containing 3 MM. The solid line has a slope of $1/RT$. The horizontal dashed line corresponds to saturation level.

Figure 1) were all at the maximal detection level. Also, the data points of the 1 MM (1 MM, oligos 3–5 in Figure 1) oligos had high intensities, spread out over two log intervals, until saturation. The same spread is found for the 2 MM oligos with

2 MM close to each other, having a perfect matching stretch of 23 bp (2 MM23, oligos 10 and 12 in Figure 1). The oligos with 2 MM and a perfect matching stretch of 15 bp (2 MM15, oligos 6–8 in Figure 1) did not reach the maximal level and were spread out over three log intervals. The oligos with 2 MM and a perfect matching stretch less than 15 bp (2 MM-15, oligo 11 in Figure 1) were around $I = 10^3$. The oligos with 3 MM (3 MM, oligo 9 in Figure 1) had low intensities, some below the background cutoff (see also Figure 4). All oligos had a ΔG value between 24 and 37.5 kcal/mol. Most of the data followed nicely the line with slope $1/RT$, representing a Langmuir isotherm. Only the data points for the PM oligos and some data points for the 1 MM and 2 MM23 oligos deviated from the line because of signal saturation. In addition, in the middle of the intensity range, some data points for the 3 MM and 2 MM-15 oligos deviated from that line. Comparing both graphs reveals that the two distinct hybridizations had a high reproducibility.

Figure 5 contains data of four hybridizations using targets 3–6 individually at a concentration of 380 pM. Also here, the PM and 1 MM oligos still had high intensities close to or at saturation. The data points of the 2 MM23 oligos were spread over 3 log intervals (10^2 to 10^5), where the highest values were similar to the intensities of the 1 MM oligos. The 2 MM15 and 2 MM-15 oligos ranged over two log intervals, with the lower limit around or under the background cutoff (see also Figure 4). The 3 MM oligos had lower intensities between 10^1 and 10^2 . All data points representing an intensity higher than 10^2 fitted the line with slope $1/RT$, representing the Langmuir isotherm. Also here, some outliers were observed for the 2 MM-15 oligos.

Figure 6 contains data of the hybridizations with a target concentration of 3.8 pM. The four graphs represent data from separate hybridizations of the four targets 3, 4, 5, or 6. The overall intensities were much lower as compared to Figures 5 and 6, with most of the data points having an intensity around background. Only for oligo sets 3 and 5, were the data points for the 1 MM and 2 MM23 oligos above background and in reasonable agreement with the Langmuir isotherm.

Discussion

When designing oligonucleotides to construct a microarray, high sensitivity and specificity of the oligonucleotides should be achieved. During the design process, cross-hybridization of non-intended target molecules should be taken into account, and in most cases avoided. Therefore, it is important to be able to have insight on the effect of mismatches on the overall hybridization, influencing both specificity and sensitivity. As DNA–DNA hybridization is a thermodynamically driven process, which already has been studied in detail, models such as the NN model are extensively described.²¹ The use of this model to approach the thermodynamics of microarray hybridizations is a subject of discussion since one basic assumption is totally different: whereas the NN model is related to DNA molecules in solution,²¹ in the case of microarrays one of the nucleotide molecules is attached to a surface, so hybridization might be liable to steric hindrance or other factors, resulting in a behavior of the molecules that does not fit the NN model. However, several studies reported a correlation between signal intensity and ΔG ,^{15,17,19} but Pozhitkov et al.¹⁸ claimed that this is not the case for mismatch-containing oligonucleotides and that there is only a poor correlation between measured intensities and ΔG .

To study the effect of well-defined mismatches on hybridization sensitivity and specificity, to evaluate whether the NN

model is valid for microarray hybridizations, and to investigate a possible relationship between intensity and ΔG , calculated using thermodynamic parameters determined for DNA–DNA hybridizations in solution, we designed eight oligo sets with varying GC content. Each set consisted of 12 30-mers with and without a 12-mer spacer, containing 0, 1, 2, or 3 MM (Figure 1 and Supporting Information). Each of the 192 oligos was spotted 12 times on CodeLink Activated Slides, where each oligo was spotted by four different pins. The microarray was hybridized with different amounts of 70-mer labeled target sequences, simulating hybridizations with high and low expressed genes.

A first and obvious observation was that the GC content was an important parameter in the sensitivity of oligonucleotides in general and for mismatch-containing oligonucleotides in particular (Figure 3). At a low GC content of 10–33% (oligo sets 1 and 2), intensities were overall lower as compared to the other oligo sets, even for the perfect matching oligo 1 (Figure 3, all bar charts). This means that oligonucleotides with this low GC content were not sensitive enough, resulting in suboptimal hybridization of the target. At a high GC content of 67–90% (oligo sets 7 and 8), almost all oligos showed intensities at the maximum detection level (Figure 3, all bar charts). The high GC content made the oligos less specific for mismatches; the strong hydrogen bonds between the high amount of cytosine and guanine overruled almost completely the unfavorable energy for mismatching bases. This makes oligos with a high GC content not specific and will also result in non-proper detection of the actual target amount.

A more balanced relation between sensitivity and specificity was observed for oligo sets 3–6 with a GC content of 50%, where two effects became clear: the effect of the amount of mismatches and positions of those mismatches and the effect of the ΔG free energy. For these oligo sets, we observed that 1 MM (Figure 3B) did not affect the intensity in such a way that it dropped under the background cutoff at the hybridization and washing conditions that were used. This phenomenon has also been described for hybridizations using the Affymetrix platform.²⁶ For the oligos with 2 MM, the relative positions of these mismatches played a crucial role: the closer the mismatches and the longer the remaining perfect matching stretch of the oligo, the higher the intensity was (Figure 3C–E). These results are in agreement with earlier studies^{1,2} dealing with longer oligonucleotides. Oligo 9, containing 3 MM, showed in all oligo sets a clearly lower intensity, which suggests that, given this concentration, at least 3 MM were needed to avoid non-intended targets to hybridize under the hybridization and wash conditions used. Comparing the intensities for the four oligo sets 3–6 (Figure 3B–E), it is notable that the intensities of oligo sets 3 and 5 were always higher as compared to the intensities of oligo sets 4 and 6. This could be explained by the ΔG free energy, which was higher for oligo sets 3 and 5 (Table 1).

Data presented in Figures 4–6 clearly suggest a correlation between hybridization intensity and ΔG , which can be described by the Langmuir model, and this for all three concentrations used. The ΔG value was calculated using the NN model, taking specific values for mismatches into account. The resulting ΔG values ranged between 24 and 37.5 kcal/mol. Consequently, we can state that, from a thermodynamics point of view, the hybridization intensity is solely dependent on ΔG , hybridization temperature, and concentration, as indicated by eq 1 (the proportionality constant A is unknown). Our results thus confirmed that the NN model, used to calculate ΔG (solution),

is strongly correlated to the actual hybridization free energy $\Delta G(\text{chip})$ where one of the molecules is attached to a surface.

In an early experiment on gel pad microarrays, Fotin et al.⁶ indeed reported a linear correlation between the two Gibbs free energies: $\Delta G(\text{chip}) = a\Delta G(\text{solution}) + b$, where the parameters were found to be $a = 1.1 \pm 0.2$ and $b = -3.2 \pm 0.4$ kcal/mol, for hybridizations to 8-mer sequences. In our experiments, an increase or decrease of the calculated values for ΔG by a global constant b , representing the potential differences between conditions in solution as compared to conditions on a microarray, does not affect the Langmuir model since b can be omitted by redefining the parameter A in eq 1 as $A' = A \exp(b/RT)$. Hence, a value of the additive constant b cannot be determined from the experimental data presented here, as the prefactor A is unknown. There are several effects that may lead to an overall constant rescaling of free energies, for instance, denaturants as formamide or DMSO in microarray hybridization buffers are most likely to produce an overall shift of the free energies (as is the effect of monovalent salt on ΔG^{21}). These effects would influence the parameter b in the previous formula.

The data presented in this paper allow a determination of the relative differences in ΔG values. The fact that the experimental data, when plotted in a logarithmic scale, follow a slope that equals $1/RT$ implies that the factor a in the previous formula equals 1. This is in agreement with the result of Fotin et al.⁶ on gel pad microarrays for which $a = 1.1 \pm 0.2$ kcal/mol. This conclusion should, however, not be taken as a general statement valid for all types of microarrays. The slides used in these experiments (CodeLink Activated Slides) have a three-dimensional polymeric coating to which the oligo probes are attached. As a result, the probes are positioned further away from the surface, thereby giving the hybridization a solution phase behavior, as described by Dorris et al.⁸ This three-dimensional coating also explains the fact that the 12-mer spacer did not increase the signal intensities, as also observed earlier by Fotin et al.,⁶ suggesting that all 30 bases were fully available for hybridization without the use of a spacer.⁸ Other studies showing a beneficial effect of a spacer^{27,28} used two-dimensional coatings. More in particular, recent analysis of Affymetrix GeneChips data^{12,13,26} showed that the intensities versus ΔG plots tend to follow straight lines as seen in Figures 4–6, although with a slope different from $1/RT$, as expected from the experimental temperature value. This suggests a rescaling of ΔG on the GeneChips as compared to the spotted microarrays due to steric hindrance, or other surface-induced effects, thereby suggesting that the factor a in the previous formula relating $\Delta G(\text{chip})$ to $\Delta G(\text{solution})$ will be smaller than 1.

It is previously described that terminal mismatches are more stable than internal mismatches.²³ As could be observed from our data, the intensities of oligos 10 and 12 indeed systematically deviated from the expected thermodynamic behavior. Although this effect has not been quantified yet, we found that by empirically subtracting a value of 2.5 kcal/mol from the total Gibbs free energy to compensate for this effect, the resulting hybridization free energy became close to the values that could be expected from the Langmuir model. It should be pointed out that the empirical correction is constant for all oligos 10 and 12, labeled as 2 MM23 in Figures 4–6.

As mentioned in the Introduction, the NN model might have some limitations when applied to oligos containing mismatches. Although Peyret et al.²² provided data to calculate ΔG for mismatch-containing oligos, Hall et al.²⁴ reported evidence that the influence of the flanking sequence context goes beyond the nearest-neighbors. Given these findings, it is likely that two

neighboring mismatches on the same sequence influence each other and result in a combined effect. As there is no precise quantification of these effects yet, the NN parameters of Peyret et al.²² were used. The combined effects could be responsible for some of the data spread in Figures 4–6. However, we do not expect substantial aberrations due to these effects since in most of our oligos, mismatches are separated by stretches of at least eight bases (except probes 10, 11, and 12).

Finally, Figures 5 and 6 show tendencies similar to the ones in Figure 4 concerning the relation between the intensities and the presence of mismatches: the PM oligos had, as expected, the highest intensities, followed by the 1 MM oligos and the 2 MM23 oligos. The other oligos with 2 MM, 2 MM15 and 2 MM-15, showed lower intensities that were above or under the background cutoff, depending on the concentration of target used. The oligos with 3 MM showed low intensities for all concentrations. These data confirm that the rules-of-thumb from previous studies, concerning global identity and length of consecutive base stretches, also apply for short 30-mer oligonucleotides: a consecutive stretch of 23 bp gave overall higher intensities as compared to oligos with a 15 bp stretch, which on their turn had overall higher intensities than oligos with 2 MM without a 15 bp stretch.

On the basis of microarray hybridizations using mismatch-containing oligos, spotted on slides with a three-dimensional surface coating, we showed that the fluorescence intensities fit the Langmuir model and that this relationship is valid over 3 orders of magnitude for the hybridization intensity. These results, besides having their own importance, can be interesting for some practical applications. For instance, the correlation between intensity and ΔG can be used in quality control of microarray hybridizations. As described by van Bakel and Holstege,²⁹ the best approach to determine the performance of microarray experiments is by adding different, well-known amounts of external control RNA, or spikes, to the RNA of biological samples, where the microarray contains probes specific for these external RNA sequences. The current set of available spikes focuses on different concentration and ratios, mimicking a range of low to high differential expressed genes. However, a more stringent test of the hybridization performance could be done with spikes hybridizing oligo sets similar to those reported in this study (i.e., containing mismatches). The degree to which the Langmuir model is followed could be a measure of the quality of the microarray hybridization.

Another important application could be for oligo design strategies. Current oligo design algorithms select a specific probe to be as different as possible from other homologous sequences. The main criteria are the overall homology and the presence of a perfect matching stretch of a sufficient length. We believe that a more accurate criterion would be to maximize the difference in ΔG between the perfect matching target–probe complex and possible complexes consisting of the same probe but homologous targets, leading to one or more mismatches. As the mismatch free energy depends rather strongly on the identity of the mismatch and of the flanking nucleotides, it could happen that a target with 2 MM has a higher affinity to a given probe sequence as compared to the target with a single mismatch. Hence, selecting probes only on the basis of the number of mismatches with a potential cross-hybridizing target may not be the optimal criterion.

Acknowledgment. This work was financed by SBO Project IWT-030263 of IWT-Vlaanderen. The authors are grateful to Joke Allemeersch for her contributions concerning the R

environment and Kizi Coeck, Ruth Maes, and Tom Bogaert for their assistance with the microarray experiments.

Supporting Information Available: Sequence information of all eight oligo sets. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Kane, M. D.; Jatkoe, T. A.; Stumpf, C. R.; Lu, J.; Thomas, J. D.; Madore, S. J. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res.* **2000**, *28*, 4552–4557.
- (2) Hughes, T. R.; Mao, M.; Jones, A. R.; Burchard, J.; Marton, M. J.; Shannon, K. W.; Lefkowitz, S. M.; Ziman, M.; Schelter, J. M.; Meyer, M. R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **2001**, *19*, 342–347.
- (3) Relogio, A.; Schwager, C.; Richter, A.; Ansorge, W.; Valcarcel, J. Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res.* **2002**, *30*, e51.
- (4) Tiquia, S. M.; Wu, L.; Chong, S. C.; Passovets, S.; Xu, D.; Xu, Y.; Zhou, J. Evaluation of 50mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* **2004**, *36*, 664–675.
- (5) Chou, C. C.; Chen, C. H.; Lee, T. T.; Peck, K. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res.* **2004**, *32*, e99.
- (6) Fotin, A. V.; Drobyshev, A. L.; Proudnikov, D. Y.; Perov, A. N.; Mirzabekov, A. D. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Res.* **1998**, *26*, 1515–1521.
- (7) Dai, H.; Meyer, M.; Stepanians, S.; Ziman, M.; Stoughton, R. Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Res.* **2002**, *30*, e86.
- (8) Dorris, D. R.; Nguyen, A.; Gieser, L.; Lockner, R.; Lublinsky, A.; Patterson, M.; Touma, E.; Sendera, T. J.; Elghanian, R.; Mazumder, A. Oligodeoxyribonucleotide probe accessibility on a three-dimensional DNA microarray surface and the effect of hybridization time on the accuracy of expression ratios. *BMC Biotechnol.* **2003**, *3*, e6.
- (9) Nielsen, H. B.; Wernersson, R.; Knudsen, S. Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays. *Nucleic Acids Res.* **2003**, *31*, 3491–3496.
- (10) Zhang, L.; Miles, M. F.; Aldape, K. D. A model of molecular interactions on short oligonucleotides microarrays. *Nat. Biotechnol.* **2003**, *21*, 818–821.
- (11) Luebke, K. J.; Balog, R. P.; Garner, H. R. Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. *Nucleic Acids Res.* **2003**, *31*, 750–758.
- (12) Hekstra, D.; Taussig, A. R.; Magnasco, M.; Naef, F. Absolute mRNA concentrations from sequence-specific calibration of oligonucleotides arrays. *Nucleic Acids Res.* **2003**, *31*, 1962–1968.
- (13) Held, G. A.; Grinstein, G.; Tu, Y. Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7575–7580.
- (14) Matveeva, O. V.; Shabalina, S. A.; Nemtsov, V. A.; Tsodikov, A. D.; Gesteland, R. F.; Atkins J. F. Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Res.* **2003**, *31*, 4211–4217.
- (15) He, Z.; Wu, L.; Li, X.; Fields, M. W.; Zhou, J. Empirical establishment of oligonucleotide probe design criteria. *Appl. Environ. Microbiol.* **2005**, *71*, 3753–3760.
- (16) Wick, L. M.; Rouillard, J. M.; Whittam, T. S.; Gulari, E.; Tiedje, J. M.; Hashsham, S. A. On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. *Nucleic Acids Res.* **2006**, *34*, e26.
- (17) Liebich, J.; Schadt, C. W.; Chong, S. C.; He, Z.; Rhee, S. K.; Zhou, J. Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. *Appl. Environ. Microbiol.* **2006**, *72*, 1688–1691.
- (18) Pozhitkov, A.; Noble, P. A.; Domazet-Lošo, T.; Nolte, A. W.; Sonnenberg, R.; Staehler, P.; Beier, M.; Tautz, D. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.* **2006**, *34*, e66.
- (19) Zhang, L.; Wu, C.; Carta, R.; Zhao, H. Free energy of DNA duplex formation on short oligonucleotide microarrays. *Nucleic Acids Res.* **2007**, *35*, e18.
- (20) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.
- (21) SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460–1465.
- (22) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; 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*, 3468–3477.
- (23) SantaLucia, J.; Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
- (24) Hall, T. S.; Pancoska, P.; Riccelli, P. V.; Mandell, K.; Benight, A. S. Sequence context and thermodynamic stability of a single base pair mismatch in short deoxyoligonucleotide duplexes. *J. Am. Chem. Soc.* **2001**, *123*, 11811–11812.
- (25) Halperin, A.; Buhot, A.; Zhulina, E. B. On the hybridization isotherms of DNA microarrays: Langmuir model and its extensions. *J. Phys.: Condens. Matter* **2006**, *18*, 463–490.
- (26) Carlon, E.; Heim, T. Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays. *Physica A (Amsterdam, The Netherlands)* **2006**, *362*, 433–449.
- (27) Guo, Z.; Guilfoyle, R. A.; Thiel, A. J.; Wang, R.; Smith, L. M. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* **1994**, *22*, 5456–5465.
- (28) Shchepinov, M. S.; Case-Green, S. C.; Southern, E. M. Steric factors influencing hybridization of nucleic acids to oligonucleotide arrays. *Nucleic Acids Res.* **1997**, *25*, 1155–1161.
- (29) van Bakel, H.; Holstege, F. C. P. In control: Systematic assessment of microarray performance. *EMBO Rep.* **2004**, *5*, 964–969.