Published online February 13, 2006

Nucleic Acids Research, 2006, Vol. 34, No. 3 ***e26***
doi:10.1093/nar/gnj024

On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes

Lukas M. Wick, Jean Marie Rouillard[3], Thomas S. Whittam, Erdogan Gulari[3], James M. Tiedje[1] and Syed A. Hashsham[1,2,*]

National Food Safety and Toxicology Center, [1]Center for Microbial Ecology, [2]Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI, USA and [3]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA

Received November 22, 2005; Revised and Accepted January 26, 2006

ABSTRACT

Nucleic acid hybridization serves as backbone for many high-throughput systems for detection, expression analysis, comparative genomics and re-sequencing. Specificity of hybridization between probes and intended targets is always critical. Approaches to ensure and evaluate specificity include use of mismatch probes, obtaining dissociation curves rather than single temperature hybridizations, and comparative hybridizations. In this study, we quantify effects of mismatch type and position on intensity of hybridization signals and provide a new approach based on dissociation rate constants to evaluate specificity of hybridized signals in complex target mixtures. Using an extensive set of 18mer oligonucleotide probes on an *in situ* synthesized biochip platform, we demonstrate that mismatches in the center of the probe are more discriminating than mismatches toward the extremities of the probe and mismatches toward the attached end are less discriminating than those toward the loose end. The observed destabilizing effect of a mismatch type agreed in general with predictions using the nearest neighbor model. Use of a new parameter, specific dissociation temperature ($T_{d\text{-}w}$, temperature of maximum specific dissociation rate constant), obtained from probe–target duplex dissociation profiles considerably improved the evaluation of specificity. These results have broad implications for hybridization data obtained from complex mixtures of nucleic acids.

INTRODUCTION

Microarrays represent the apex of nucleic acids hybridization-based methods in terms of throughput and number of applications. In addition to their widespread use in gene expression studies (1,2), they are also important in comparative genomics (3–5), microbial gene detection (6–13), single nucleotide polymorphism analysis and re-sequencing (14–17). Specificity of hybridization signal is critical in all cases. It is especially important in applications involving hybridization of samples with complex uncharacterized background [e.g. microbial detection (18–21)]. Specificity at the probe design level is provided by selecting oligonucleotide probes that are perfect matches (i.e. identical in complementary sequence) to their intended targets, differ significantly compared with all other sequences, and have a narrow range of melting temperature (T_m). Although it is well known that attachment of probes to surfaces can dramatically influence the hybridization behavior of duplexes (14,22–25), most probe design for microarrays employs rules established with data collected from hybridization carried out in solution (26,27).

At the experimental level, specificity is provided by hybridization conditions that minimize hybridization of mismatch targets to the probes (10,28) or by conducting hybridization over a range of temperature (29–35). Another common approach is to design a set of mismatch (MM) probes that may be compared in their signal intensity with the perfect match (PM) probes. Data analysis techniques then evaluate the specificity provided by the above approaches. Comparison of dissociation temperature [$T_{d\text{-}50}$, the temperature at which 50% of the initial probe-target duplex is washed off, (30,31)] or signal intensity for PM probe compared with MM probe are the common examples at this level. Since probe-target duplexes containing one or more mismatches are less stable than perfect match probe-target duplexes, these MM duplexes

*To whom correspondence should be addressed. Tel: +1 517 355 8241; Fax: +1 517 355 0250; Email: hashsham@egr.msu.edu

can be identified because of their lower dissociation temperature, $T_{d\text{-}50}$ (the temperature at which 50% of the initial signals are washed out) compared with the PM duplex (29–34).

The purpose of this study was to: (i) assess the difference in hybridization behavior of PM and MM probes in terms of their predicted T_m and experimentally determined $T_{d\text{-}50}$, and (ii) determine if alternative parameters obtained from the non-equilibrium dissociation curves (namely, the dissociation rate constant—k_d and the maximum specific dissociation temperature—$T_{d\text{-}w}$) are more discriminatory than dissociation temperature, $T_{d\text{-}50}$. To this end, we used *in situ* synthesized microfluidic chips containing an extensive set of 18mer probes to obtain $T_{d\text{-}50}$ and $T_{d\text{-}w}$ for a number of gene targets. We compared experimental variation in signal intensities and $T_{d\text{-}w}$ between a variety of perfect match and mismatch probes with predictions of the nearest neighbor model.

MATERIALS AND METHODS

Oligonucleotide arrays

Sequences of 11 genes from *Escherichia coli* strain O157:H7 RIMD 0509952 (36) (*aroE*, *clpX*, *cstA*, *glyA*, *lysP*, *rpoS*, *mdh*, *stx1*, *stx2*, *eae* and *uidA*) and one gene from *Shigella flexneri* 2A 2457T (*virA*) were used to design 935 PM 18mer oligonucleotide probes. For 352 PM probes targeting *aroE*, *clpX*, *cstA*, *glyA*, *lysP*, *rpoS*, *mdh* and *virA* genes, three single mismatch 18mer probes created randomly with respect to both position and type of mismatch were also designed resulting in a total of 1056 MM probes. For 578 PM probes, additional 18mer MM probes with a single mismatch in the center (position 9) were designed. Furthermore, 20, 25, 35 and 45mer probes for the *aroE* gene and 20mer probes for the *stx1*, *stx2*, *eae* and *uidA* genes were added. These probes were synthesized *in situ* on microfluidic chips by Xeotron (Houston, TX, now part of Invitrogen, Carlsbad, CA) (37). Briefly, the glass-silicon chip surface was first derivatized with an N-(3-triethoxysilylpropyl)-4hydroxybutyramide linker (Gelest, Morrisville, PA) and then a spacer consisting of Ts and C18 spacers for an effective length of 12 bp was directly synthesized on the linker's hydroxyl group using the phosphoramidite chemistry. The oligonucleotides were synthesized on top of this spacer with an estimated density of 1 molecule per 200 square angstroms.

DNA and target preparation

Fragments of ~600 bp including the sequences targeted by the oligos on the chip were amplified from DNA of *E.coli* strain O157:H7 RIMD 0509952 (36) (*aroE*, *clpX*, *cstA*, *glyA*, *lysP*, *rpoS*, *mdh*, *stx1*, *stx2*, *eae* and *uidA*) and from DNA of *S.flexneri* 2A 2457T (*virA*). A purified PCR product mixture (6.8×10^{-14} mol of each amplicon) was aminoallyl-dUTP labeled (Sigma, St Louis, MO) with the Klenow-based Invitrogen DNA labeling system (Invitrogen, Carlsbad, CA) as described previously (38). For replicate measurements on the same chip, aliquots from the same pool of labeled DNA were used. For different chips, aliquots from different pools were used. Genomic DNA was prepared as described previously (38). The Cy3 dye was used as the fluorescent label in all experiments.

Microarray hybridization

The *in situ* synthesized chips were prehybridized, hybridized and washed in a M-2 microfluidic station (Xeotron Corporation, Houston, TX, now part of Invitrogen, Carlsbad, CA) at a flow rate of 500 µl/min. Hybridization buffer was 6× SSPE, 35% formamide, 0.4% Triton X-100 for hybridizations of only PCR products and 6× SSPE, 25% formamide, 0.4% Triton X-100 for hybridizations of samples containing genomic DNA. Chips were prehybridized with 6× SSPE, 0.2% Triton X-100 and then with hybridization buffer for 2 min each. All SSPE buffers were made from a stock of 18× SSPE, which is 2.7 M NaCl, 180 mM Na_2PO_4, 18 mM Na_2EDTA (pH adjusted to 6.6 with HCl). Labeled target was suspended in 50 µl hybridization buffer, denatured at 95°C for 3 min, cooled on ice for 1 min, filtered through a 0.22 µm Costar spin filter and then hybridized to the chip for 14–15 h at 20°C. Since the residual prehybridization buffer in the Xeotron chip is ~50 µl, the final hybridization volume was ~100 µl. After hybridization the chip was washed at 20°C with hybridization buffer, with 6× SSPE, 0.2% Triton X-100, with 1× SSPE, 0.2% Triton X-100 and finally with 6× SSPE for 2.2 min each. The chip was scanned with a GenePix 4000B laser scanner (Axon Instruments, Union City, CA). All solutions were filtered through a 0.22 µm filter to prevent clogging of the microfluidic channels. The high stringency wash buffer was degassed under vacuum.

Melting curve profiles

To create a dissociation profile, a hybridized chip was washed at 25°C with high stringency wash buffer (20 mM NaCl, 10 mM Na_2PO_4, 5 mM Na_2EDTA, pH adjusted to 6.6 with HCl) for 1.4 min and then scanned. Cycles of washing and scanning were repeated manually at 1°C intervals until 60°C was reached. At the end of this series, the chip was stripped further by washing with distilled water (three times each for 2.2 min at 60°C).

Data acquisition

Hybridization signal intensities were extracted with GenePix 5.0 software (Axon Instruments, Union City, CA), yielding values between 0 and 65 535 arbitrary units (a.u.). For each dissociation temperature, a background value was determined as the median of the 95% empty spots with the lowest signals on the array and subtracted from each signal at the corresponding temperature. Background values were between 50 and 80 a.u. If a spot signal after background subtraction was less than three times the standard deviation of the background, it was set to 3 SD of the background.

Data flagging

Bad curves were excluded from analysis by flagging them when one or more of the following conditions were satisfied: (i) the initial signal was <200 a.u. or saturated; (ii) the R^2 of the linear fit to calculate $T_{d\text{-}50}$ was below 0.9; (iii) the $T_{d\text{-}50}$ value was outside the points used for the linear fit; (iv) the signal at $T_{d\text{-}50}$ or at $T_{d\text{-}w}$ was <5 times the standard deviation of the background signal intensity; (v) the interpolated $T_{d\text{-}w}$ was more than 1°C different from the actually used wash temperature yielding maximum k_d; (vi) the coefficient of variation for

the mean $T_{d\text{-}50}$ or $T_{d\text{-}w}$ from the five replicate experiments exceeded 5%.

Five replicate hybridizations (two on one chip and three on a second chip) of PCR product mixtures were used to calculate the mean of $T_{d\text{-}50}$, $T_{d\text{-}w}$ and $\log_2$(PM/MM) ratios for each probe. Probes flagged in any of the five replicates were excluded from $T_{d\text{-}50}$ and $T_{d\text{-}w}$ analysis. For the $\log_2$ (PM/MM) ratios, all those PM–MM pairs were analyzed that had an initial PM signal intensity higher than 128 times three standard deviations of the background (between 700 and 2600 a.u. depending on chip). Following this protocol, $\log_2$ ratios up to nearly seven, which is about the 95th percentile observed for the probe set, can be determined relatively accurately for all probes. Multiple pairwise comparisons were made with JMP4 (SAS Institute).

Nearest neighbor calculations

Melting temperatures (T_m), free energy changes (ΔG^0), enthalpy changes (ΔH^0) and entropy changes (ΔS^0) were calculated with the nearest neighbor model using published parameters for PM and MM duplexes assuming a two-state model (26,27). Contributions of the dangling end in computing the above parameters were neglected. For calculating T_m, a probe concentration of 10 nM was assumed. For calculation of ΔG^0 in the hybridization buffer, 43°C instead of the actual 20°C degrees was used since the buffer contained 35% formamide, which is expected to destabilize duplexes in a way equivalent to increasing the temperature by 21–25°C (33,39).

Data analysis

Dissociation curve parameters were computed using Excel (Microsoft, Redmond, WA). Median of signal intensities at 26, 27 and 28°C was used as the initial signal and median of signal intensities at temperatures from 56 to 60°C was used as the final signal for a dissociation curve. The difference between the initial and final signal was defined as 'range'. Dissociation curves were normalized by setting the initial value to 1, the end point to 0 and scaling all other values linearly. For hybridizations with PCR products, dissociation curves were linear in the central range of the hybridization temperature. Therefore, the best linear fit through nine points of each dissociation curve was used to calculate $T_{d\text{-}50}$ as the temperature where 50% of the range was washed off, i.e. where the normalized signal intensity equals 0.5. For hybridizations with genomic DNA, a linear fit through three points closest to the 50% value of the initial signal was used to interpolate $T_{d\text{-}50}$.

The hybridization of a target to a probe can be described with the following equation:

$$[\text{Probe}] + [\text{Target}] \underset{k_a}{\overset{k_d}{\leftrightarrows}} [\text{Probe} \bullet \text{Target}]. \quad \mathbf{1}$$

The above equation assumes there are no effects of probe–probe interactions, target–target interactions, or secondary structures of probes and targets. For a non-equilibrium dissociation curve where the target is washed away instantaneously, the association constant, k_a, becomes zero and the decrease in hybridization signal intensity (I) with time (dt) at a given temperature can be described by the following first order equation:

$$dI = -Ik_d dt. \quad \mathbf{2}$$

In the integrated form, Equation 2 can be written as:

$$I_t = I_0 e^{-k_d(t-t_0)} \quad \mathbf{3}$$

where I_t is the intensity after the wash, I_0 is the intensity before the wash, k_d is the temperature-dependent dissociation rate constant and $(t - t_0)$ the washing time. Hence, the dissociation rate constant for each probe–target duplex at each temperature can be obtained by calculating k_d at the corresponding temperature using the following equation:

$$k_d = \frac{\ln I_0 - \ln I_t}{t - t_0} \quad \mathbf{4}$$

The temperature dependence of k_d is expected to follow the Arrhenius equation:

$$k_d = Ae^{\frac{-E_a}{RT}} \quad \mathbf{5}$$

where, A is the Arrhenius constant, E_a the activation energy and R the gas constant.

For calculation of $T_{d\text{-}w}$ (temperature at the maximum measured dissociation rate constant) the signals were transformed by taking their natural logarithm. Slopes at the resulting 'ln(I) versus T' curve yielded the dissociation rate constant (k). To smooth the 'k versus T' curve, k at a given temperature was calculated by taking the linear slope over either five degrees (including two temperature points before and after the given temperature) or nine degrees (including four temperature points before and after the given temperature). For hybridization with PCR products, nine points were used. For hybridization involving genomic DNA, five points were used in order to increase resolution. The temperature at the maximum dissociation rate constant ($T_{d\text{-}w}$) was obtained by interpolating the temperature at which the first derivative (approximated by the slope through three consecutive points) of the 'k versus T' curve was zero.

RESULTS

$T_{d\text{-}50}$ and $T_{d\text{-}w}$ for dissociation curves of PCR products

Of the 1991 non-equilibrium dissociation curves obtained from 935 PM probes (18mer) and 1056 randomly created single MM probes, 597 (30%) were flagged as bad quality in one or more of the five replicate hybridizations of PCR products. The single most important criterion for flagging curves was the initial signal intensity. Of the flagged curves, 76% had an initial intensity below 200 a.u. and 85% were below 500 a.u. Most flagged curves (81%) were from MM probes. Examples for normalized dissociation curves and dissociation rate constants (k_d) for a PM probe and two single-base-pair MM probes are shown in Figure 1. According to the Arrhenius equation, k_d should continually increase with temperature (dashed lines in Figure 1B), but practically it has a maximum. The temperature at which the measured k_d reaches this maximum is a critical point on the dissociation rate curve and is referred as $T_{d\text{-}w}$. It represents the temperature where the

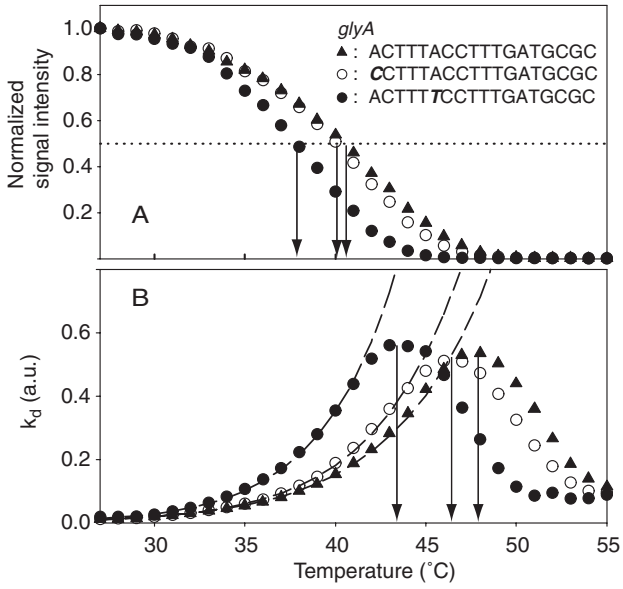

Figure 1. Dissociation profiles from a perfect match (filled triangles) probe and two probes containing a single base pair mismatch (empty and filled circles). (**A**) Normalized signal intensities. Probe sequences are given with the mismatches in bold italics. $T_{d\text{-}50}$ values are indicated by arrows. (**B**) Dissociation rate constants. Dashed lines show the best fit of values measured between 30 and 1°C below the maximum k_d to the Arrhenius equation. $T_{d\text{-}w}$ values are indicated by arrows.

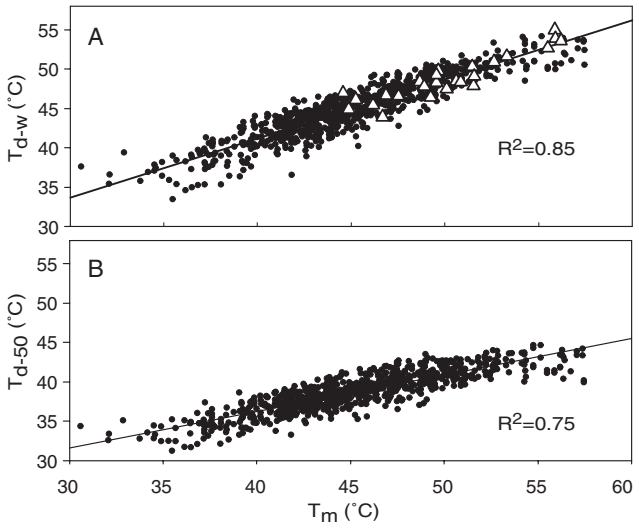

Figure 2. Correlation between T_m and $T_{d\text{-}w}$ (**A**), and T_m and $T_{d\text{-}50}$ (**B**) for PM probes. Black circles: 18mer, not saturated ($n = 819$); open triangles: 18 and 20mer saturated ($n = 25$). Values are means of five experiments. The R^2 value for the linear fit between T_m and T_d is given for not saturated 18mer.

target has essentially been washed off, i.e. the abundance of targets bound to the probe is insufficient to yield a sensible measurement of the real k_d and the Arrhenius equation no longer holds true.

Correlation of $T_{d\text{-}50}$ and $T_{d\text{-}w}$ with nearest neighbor melting temperature (T_m)

T_d values of PM probes obtained by the traditional dissociation curve method ($T_{d\text{-}50}$) and the new maximum rate of dissociation method ($T_{d\text{-}w}$) were compared with T_m values calculated with the nearest neighbor model. The correlation between $T_{d\text{-}w}$ and T_m was better compared with the correlation between $T_{d\text{-}50}$ and T_m (Figure 2). Also $T_{d\text{-}w}$ computed for 18 and 20mer probes with initially saturated signal intensities corresponded well with $T_{d\text{-}w}$ values for non-saturated probes with similar T_m (open triangles in Figure 2).

Effect of mismatch position

For the 352 PM probes that had each three randomly created single mismatch probes, the effect of mismatch position on initial signal intensity and $T_{d\text{-}w}$ was evaluated. Log_2(PM/MM) of the initial signal intensity and $\Delta T_{d\text{-}w}$ (i.e. $T_{d\text{-}w}$ for PM $- T_{d\text{-}w}$ for MM) were calculated and plotted against the position of the mismatch (Figure 3). Both curves have an inverse U-shape, which indicates that the best discrimination between PM and MM is better for mismatches near the center of the probe than for mismatches near the terminal positions. For example, the signal intensity of a probe with a mismatch at position 9 was on average only ~4% of the corresponding PM probe signal intensity. $T_{d\text{-}w}$ for the MM was more than 4°C lower. In contrast, mismatches at the very end of the 18mer had signal intensities that were ~50% the corresponding PM probe signal intensity with $T_{d\text{-}w}$ values only about 1°C lower. The effect of the mismatch position on $\Delta T_{d\text{-}50}$ was similar to that observed for $\Delta T_{d\text{-}w}$ (data not shown).

Effect of mismatch type

For some mismatch positions no significant effect of position on $\log_2$(PM/MM) or on $\Delta T_{d\text{-}w}$ was observed (Tukey-Kramer, $\alpha = 0.05$). Probes with mismatches in these positions [positions 5–11 for $\log_2$(PM/MM) and 4–14 for $\Delta T_{d\text{-}w}$] were grouped by mismatch type and the average $\log_2$(PM/MM) and $\Delta T_{d\text{-}w}$ were calculated for the group. To designate mismatch type, we used the first letter to denote the base in the probe and the second letter to denote the base in the target (Figure 4). Because the targets are the same for PM and MM probes, a C-T mismatch, for example, means that the corresponding PM probe has an A instead of a C at that position resulting in an A-T Watson–Crick pair for the PM probe–target duplex. For a T-C mismatch, the perfect match in the corresponding PM probe–target duplex would be a G-C base pair. Because the effects of mismatches are measured as differences in parameters obtained from MM probes and their corresponding PM probes, a T-C mismatch (compared to a G-C perfect match) is generally more discriminating than a C-T mismatch (compared with a A-T perfect match) because G-C is a stronger bonding base-pair than A-T. Log_2(PM/MM) values were compared with differences in ΔG^0 of the PM-target and MM-target duplexes (Figure 5A) because, on average, the log(signal) of all 18mer PM probes on the chip linearly correlated with ΔG^0 of probe target duplexes (data not shown). Values for $\Delta T_{d\text{-}w}$ were compared with values for ΔT_m (defined as the difference in T_m between PM and MM) (Figure 5B). Overall, there was a good correlation between measured values and values calculated with the nearest neighbor model. However, for both the $\log_2$ ratios as well as the $\Delta T_{d\text{-}w}$ values some of the values are probably an

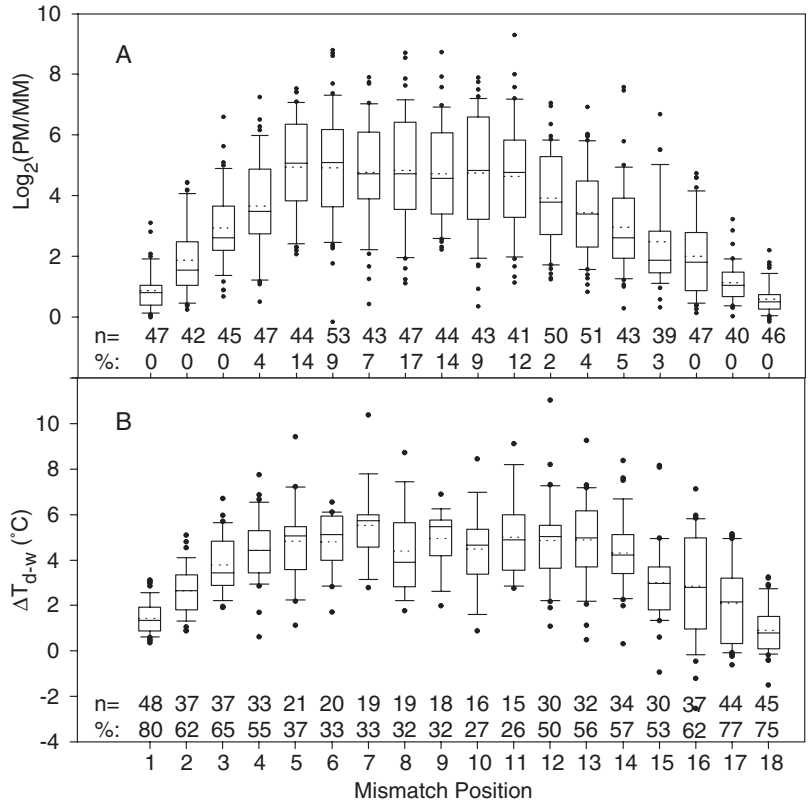

Figure 3. Influence of the position of a single base pair mismatch on the initial signal intensity (**A**) and on $\Delta T_{d\text{-}w}$ (**B**). Boxes indicate the range from the 25th to 75th percentile, whiskers the 10th and the 90th percentile. The median is given as a solid, the mean as a dotted line. Sample sizes are given for each position ($n =$). In (**A**) the percentage of MM probes with signal intensities below 5 SD of background are given (%). In (**B**) '%' shows what percentage of all probes with the MM at that position had good quality dissociation curves and was used for analysis. Note that for positions 5–11, <40% of probes gave good quality measurement of $T_{d\text{-}w}$. Values in this range are likely an underestimation. Probes are attached to the chip at the 3′ end (position 18).

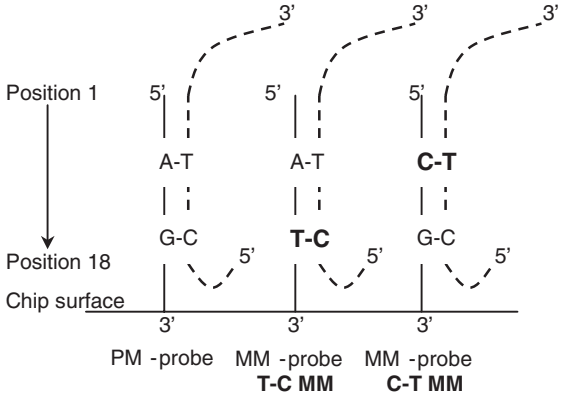

Figure 4. Designation of MM-types. In naming MM-types, the first letter denotes the base in the probe (solid line), the second the one in the target (dashed line). An example for a T-C (middle) and a C-T (right) mismatch with the corresponding PM probe (left) are given.

underestimation. For MM probes with initial intensities less than five times the standard deviation of the background, the actual $\log_2$(PM/MM) ratio is likely to be larger than the measured one. For computing $\Delta T_{d\text{-}w}$, MM probes with low intensities were excluded. Most of these excluded probes would probably have a $\Delta T_{d\text{-}w}$ lower than average. Indeed, mismatch types with higher predicted discrimination were also more likely to have a very low MM signal or to be excluded from $T_{d\text{-}w}$ analysis (Table 1).

Nearly all MM probes had significantly lower $T_{d\text{-}50}$ (92%) and $T_{d\text{-}w}$ (94%) than their corresponding PM probes ($\alpha = 0.05$, two-tailed t-test, five replicate experiments). Also the majority of terminal and penultimate (second to last) mismatches were detected by their dissociation profiles (79% based on $T_{d\text{-}50}$ and 88% based on $T_{d\text{-}w}$) (Table 2). Toward the surface distal 5′ end (position 1) the resolution of near terminal mismatches was better than toward the surface proximal 3′ end (position 18) (Table 2). Based on $T_{d\text{-}w}$, 100% of mismatches at position 2 and 98% of mismatches at position 1 were resolved, whereas only 84% of mismatches between position 14 and 18 (i.e. near the 3′ end) were resolved. Interestingly, most of the unresolved mismatch types were G-T, C-A and C-T mismatches (24 out of 31). These three mismatch types were significantly more likely to be unresolved than the other nine mismatch types [residual analysis of the contingency table, $\alpha = 0.05$ (40)].

Analysis of dissociation curves influenced by cross-hybridization

The data above were obtained from low complexity samples (i.e. twelve 600 bp PCR fragments resulting in a

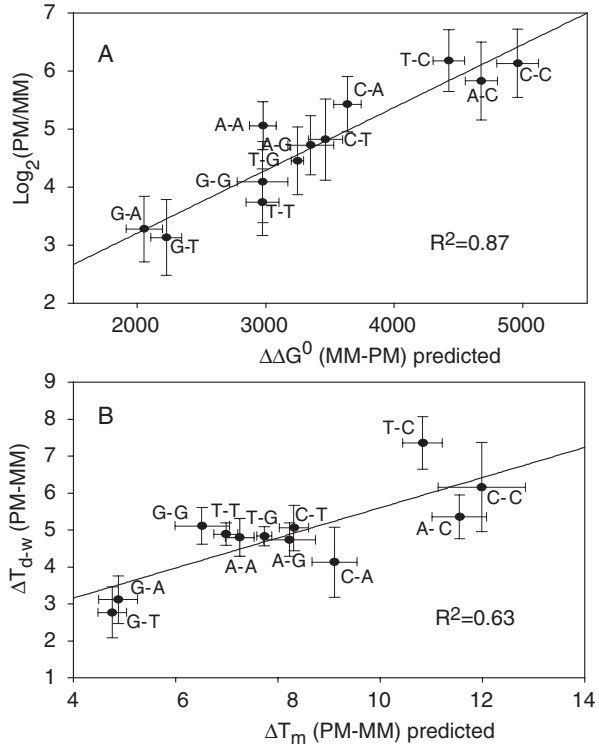

Figure 5. Correlation between predicted and measured effects of mismatch type on initial signal intensity (**A**), and on $\Delta T_{d\text{-}w}$ (**B**). Group means with 95% confidence intervals are given. R^2 values are for the best linear fit between calculated and measured group means. See Table 1 for sample sizes.

Table 1. Sample sizes for data plotted in Figure 5

MM type	Signal intensity (positions 5–11) Sample size	% Low intensity[a]	Melting temperature (positions 4–14) Sample size	% Good quality[b]
C-C	21	24	7	17
A-C	29	24	15	28
T-C	29	24	18	35
C-A	34	15	19	29
C-T	25	12	20	34
A-A	27	7	16	35
T-G	25	8	33	52
G-A	26	4	24	44
G-T	22	5	22	47
G-G	24	8	27	57
A-G	31	3	25	44
T-T	22	5	31	53

[a]Percentage of samples with a signal intensity below 5 SD of the background.
[b]Percentage of MM probes with good quality dissociation curves (i.e. sample size divided by number of all MM probes with given mismatch).

total of ~7 kb); therefore, we observed minimal cross-hybridization, as evident in the overall strong correlation between T_m and $T_{d\text{-}50}$ (Figure 2). When hybridizing more complex samples, e.g. whole genomes containing about 5000 genes (~5 Mb), more substantial cross-hybridization is expected and probes without a perfect match target in the sample can still give high signals, especially at low temperatures. To test the influence of cross-hybridization on dissociation curve behavior, genomic DNA from *E.coli* Sakai spiked with different amounts of PCR products were used for hybridization. Dissociation curves of PM duplexes are normally of a sigmoid shape, but they change to a more linear shape and then to an exponential decay shape with increasing amounts of cross-hybridization (Figure 6). A target melting with a lower $T_{d\text{-}50}$ than expected from a PM target indicates hybridization of mismatched targets. These can be defined targets containing a single nucleotide mismatch (Figure 1A) or less well defined non-specific cross-hybridization signals resulting from a complex background (Figure 6A, C and E). With complex background DNA, it is difficult to determine whether 100% of the signal is due to cross-hybridization, or whether it is a combination of signals due to cross-hybridization and perfect match target. In the case of a combined signal originating from a mixture of non-specific and specific targets, a decision based on the $T_{d\text{-}50}$ alone would lead to a false negative call. $T_{d\text{-}50}$ only can detect cross-hybridization, but does not indicate if the signal is from unspecific background, a single mismatch target or a combination of unspecific and specific targets. $T_{d\text{-}w}$, on the other hand, should be able to detect such differences. Unspecific signals wash off earlier and show a higher k_d than the specific signal in the beginning of the melting curve. After the unspecific signal is washed off, the k_d and $T_{d\text{-}w}$ of the specific signal will still reveal the presence of the specific target. Indeed, most of the $T_{d\text{-}w}$ values obtained from the spiked samples showing cross-hybridization were close to $T_{d\text{-}w}$ values obtained from PM targets alone, whereas $T_{d\text{-}50}$ values were generally much lower (Figure 7). For a few probes with extensive cross-hybridization even the $T_{d\text{-}w}$ may substantially decrease, making it difficult to distinguish the cross-hybridization signal from the specific hybridization signal (Figure 6F). For this probe, with increasing amount of genomic DNA in the sample, the $T_{d\text{-}w}$ shifts more toward lower temperatures, indicating an increasing amount of cross-hybridization (Figure 6F). In contrast, $T_{d\text{-}50}$ has nearly the same low value in all samples containing genomic DNA and cannot detect the different amounts of cross-hybridization (Figure 6E). Generally, the 18mer probes proved too short and unspecific when only unspiked genomic DNA was hybridized.

Another approach to assess the amount of cross-hybridization is by comparing the PM signal with the signal from a MM probe with a single mismatch at the center of the probe. For 578 such PM–MM probe pairs, signals of PM and MM were compared. In the absence of cross-hybridization, signals from MM probes are consistently lower than signals from PM probes. In three of the five replicates using PCR products, none of these MM probes had a signal higher than the PM probes. In one of the five replicates four, in another one five MM probes, had a slightly higher signal, but four of those were low intensity spots (less than twice the background level) on both chips. The situation changes when hybridizing more complex samples. For example, the number of MM probes with a signal higher than the PM probes increased from 19 to 37 to 111 when the ratio of PCR product spiked in genomic DNA (expressed as ng of PCR product to ng of genomic DNA) changed from 400:3400 to 45:3700 to 0:3800, respectively. This shows that cross-hybridization can be very different for PM and MM probes and use of MM probes to estimate cross-hybridization of PM probes may be problematic in complex samples. But the $T_{d\text{-}w}$ for all the MM probes in the spiked samples was lower than for the PM

Table 2. Significant differences in $T_{d\text{-}50}$, $T_{d\text{-}w}$ and initial signal intensities between perfect match and mismatch probes

Mismatch position	Sample size	$T_{d\text{-}50}$[a]			$T_{d\text{-}w}$[a]			Log(PM/MM)[a]		
		PM > MM	PM = MM	PM < MM	PM > MM	PM = MM	PM < MM	PM > MM	PM = MM	PM < MM
1	48	87.5	8.3	4.2	97.9	2.1	0	93.8	4.2	2.1
2	37	97.3	2.7	0	100	0	0	100	0	0
3	37	97.3	2.7	0	100	0	0	100	0	0
6	20	100	0	0	100	0	0	95	5	0
8	19	89.5	10.5	0	100	0	0	100	0	0
14	34	97.1	2.9	0	97.1	2.9	0	97.1	2.9	0
15	30	93.3	3.3	3.3	96.7	0	3.3	100	0	0
16	37	86.5	2.7	10.8	83.8	8.1	8.1	97.3	2.7	0
17	44	75	13.6	11.4	79.5	11.4	9.1	97.7	2.3	0
18	45	60	24.4	15.6	71.1	17.8	11.1	84.4	13.3	2.2

[a]Percentage of PM–MM probe pairs where PM value is significantly higher (PM > MM), not significantly different (PM = MM) or significantly lower (PM < MM) than MM value ($\alpha = 0.05$, two-tailed *t*-test, five replicate experiments). For all other mismatch positions not shown the PM value was higher in 100% of the cases.

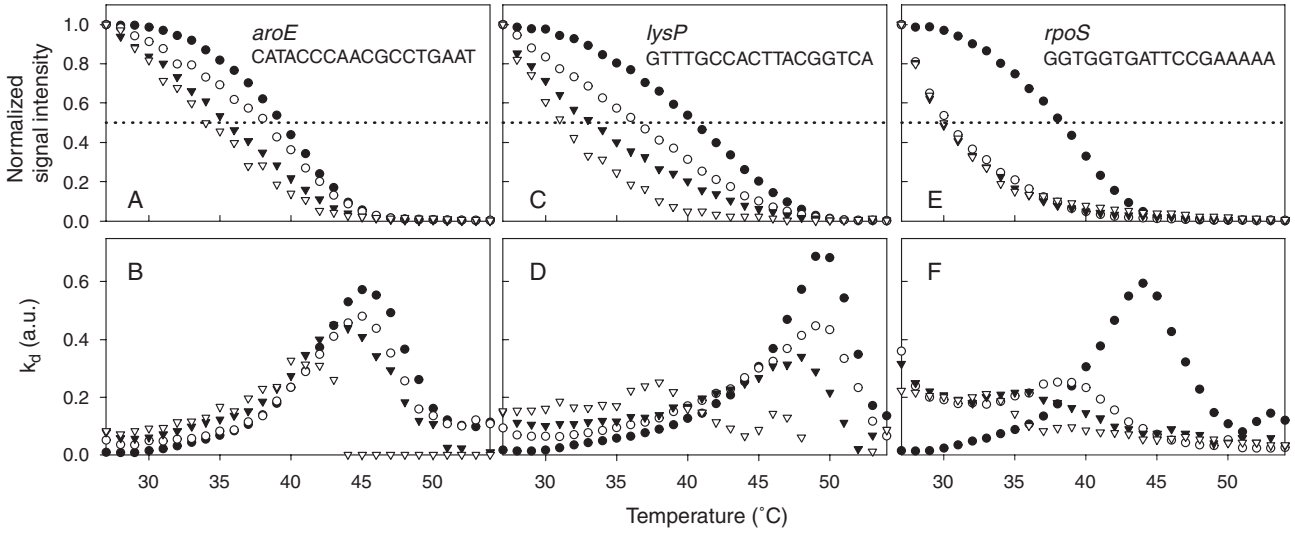

Figure 6. Normalized dissociation curves (top panels **A**, **C**, **E**) and dissociation rate constants k_d (bottom panels **B**, **D**, **F**) of three PM probes showing various degrees of cross-hybridization. Hybridization was carried out with 4.2 μg of PCR-products only (filled circles), 3.4 μg genomic DNA spiked with 400 ng PCR products (empty circles), 3.7 μg genomic DNA spiked with 45 ng PCR products (filled triangles) and 3.8 μg genomic DNA only (empty triangles). Gene targets and probe sequences are given in the upper panels.

probes and $T_{d\text{-}w}$ values for most PM probes were in the range expected from values obtained without cross-hybridization (Figure 7).

DISCUSSION

This study shows that monitoring the dissociation rate constant of non-equilibrium melting curves considerably improves the interpretation of hybridization signals. Specifically, we demonstrate here that the temperature where the measured dissociation rate constant is maximal, $T_{d\text{-}w}$, obtained from non-equilibrium dissociation curves has several advantages over the conventional measure, $T_{d\text{-}50}$. Compared with $T_{d\text{-}50}$, $T_{d\text{-}w}$ had better correlation with T_m over a wide range of temperature (Figure 2). Computation of $T_{d\text{-}w}$ does not depend on initial signal intensity and is thus more reliable even for probes having saturated signals at lower temperatures (Figure 2) or probes with cross-hybridization (Figures 6 and 7). Moreover, the method uses the dissociation rate constant, which is the actual physical parameter measured during a non-equilibrium washing curve. Besides the $T_{d\text{-}w}$ used here, other methods of analyses based on the dissociation rate constant might be useful, such as absolute values of k_d at given temperatures, temperatures where k_d has a certain value or the run of the k_d-curve. For example, the k_d of a single mismatch target starts off at a low level, similar to the k_d from the corresponding perfect match target, but then increases more rapidly with increasing temperatures (Figure 1B). In contrast, the signals from a combination of mismatch and perfect matched targets yielded larger k_d at lower temperatures, but gradually approached the k_d values of the perfect match targets when more and more of the mismatched targets are lost and the fraction of perfect match target contributing to the signal increases (Figure 6B and D). This shows that the dissociation rate constant is a promising parameter in interpreting signals from more complex samples leading to binding of several different targets to one probe. The resolution of such mixed signals gets better the more accurately k_d is determined at each single temperature point. In order to reduce noise, the

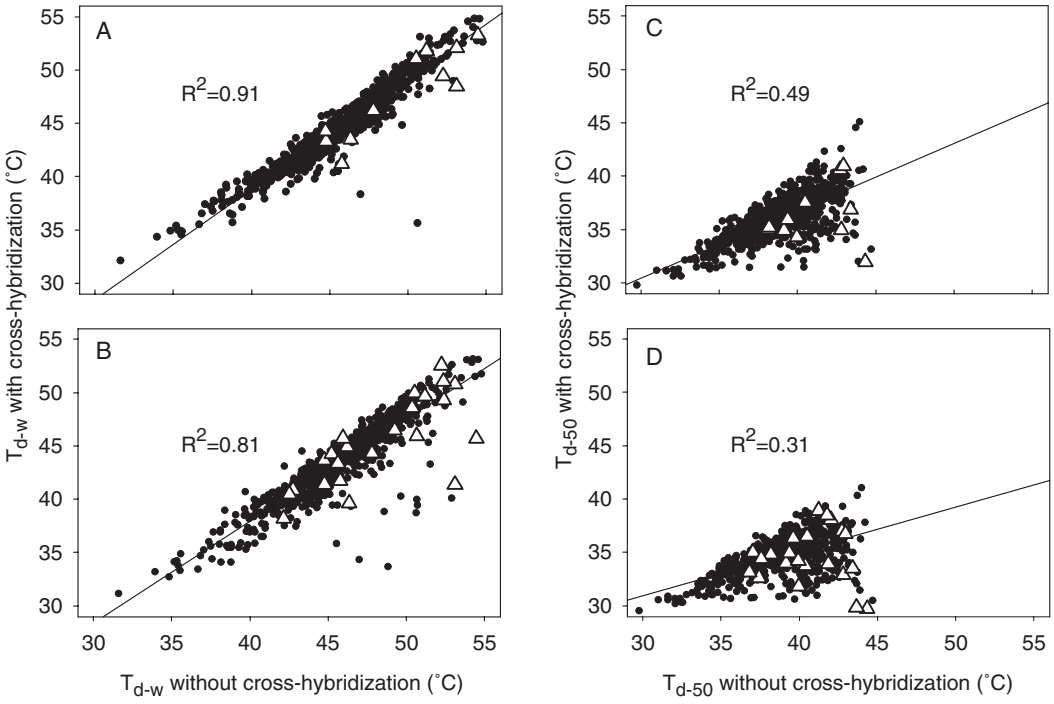

Figure 7. Impact of cross-hybridization on computed $T_{d\text{-}w}$ (**A** and **B**) or $T_{d\text{-}50}$ (**C** and **D**). For cross-hybridization, 3.4 μg genomic DNA was spiked with 400 ng PCR products (A and C) or 3.7 μg genomic DNA was spiked with 45 ng PCR products (B and D). Both were compared with $T_{d\text{-}w}$ or $T_{d\text{-}50}$ obtained using PCR products only ($n = 704$). Values for PM probes where the signal intensity of an MM probe (with MM at position 9) was higher are shown as open triangles.

data were smoothed in this study by calculating the slopes at the 'ln(I) versus T' curve over several temperature points in a linear fit, which of course lowers the resolution. Advanced technical setups (cycles between washing and scanning were carried out manually in this study), producing data with less noise and also improved smoothing methods, should be able to further increase the resolution and the specificity of this method. For some of the mixed probe–target duplexes, a decomposition of the observed curve and a quantification of the contributions of each target to the overall signal should also be achievable.

$T_{d\text{-}w}$ values from PM duplexes correlated well with T_m obtained from the nearest neighbor model, suggesting that parameters obtained from PM duplexes in solution can be applied to duplexes on the array. This is possible despite the fact that T_m values are based on thermodynamic equilibrium and T_d values are based on non-equilibrium dissociation measurements. Thus, the dissociation rate constant is the principal sequence-dependent factor in determining probe–target binding thermodynamics, and the association constant (Equation 1) is rather sequence independent. For equilibrium values on Affymetrix chips containing 25mer, Held *et al*. (41) observed an overall good correlation between measured intensities and free energies calculated with parameters obtained from solutions; however, in contrast, Zhang *et al*. (42) found only a weak correlation between nearest neighbor values fitted to Affymetrix chip data and values obtained from solution.

Effect of mismatch type

For duplexes containing single-base pair mismatches, there was on average a good correlation between predicted and measured values of differences in signal intensities after overnight hybridization and also between differences in dissociation temperatures of MM and PM duplexes. As predicted by the two-state nearest neighbor model, G-T and G-A mismatches were least discriminating and C-C, T-C and A-C mismatches were most discriminating (Figure 5). The same effects of mismatch types were predicted by simulations using the software Oligonucleotide Modeling Platform (OMP, DNA Software, Inc., Ann Arbor, MI) by Lee and co-workers (43). OMP uses a multi-state model instead of the two-state model and should generally produce more accurate predictions. We observed a considerable variability among different probes with the same mismatch type. The use of formamide buffer and the presence of secondary structures or probe–probe interactions may account for the differences in expected and observed log ratios after overnight hybridizations. To calculate thermodynamic parameters in formamide buffer, we assumed that 1% increase in formamide concentration has a destabilizing effect equivalent to an increase in temperature of 0.6–0.7°C (33,39). However, formamide may destabilize some base pairs differently than others (39). Introduction of a mismatch not only affects the probe–target duplex, but may also change the stability of the probe's secondary structures and of probe dimers, thus impacting the availability of MM-probes for target binding, compared with PM probes.

One reason for differences between predicted and measured $\Delta T_{d\text{-}w}$ values is that curves yielding useful parameters are biased, because for several probes the mismatch was so destabilizing that signals were too low to be analyzed accurately. The percentage of excluded probes was generally higher for mismatch types with higher predicted destabilizing effects

(Table 1), further supporting the correlation between nearest neighbor predictions and measured values.

Moreover, some base-pairings might behave very differently depending on next to nearest neighbors or sequence position. For example, some C-T, C-A and G-T mismatches near the 3′ end had the same or even a higher $T_{d\text{-}w}$ than their PM probes, although C-T and C-A are expected to be more destabilizing than several other mismatch types (Figure 5 and Table 1).

Because the nearest neighbor predictions are based on equilibrium values, the variation observed among MM probes might also be based on a larger influence of mismatches on the association constant, compared with PM probes. Ideally, nearest neighbor parameters should be established for non-equilibrium melting curves, so that the melting can be predicted from the probe sequence.

Effect of mismatch position

A major difference between predictions based on solution parameters using the two-state nearest neighbor model and our chip data was that the mismatch position had an effect on differences in initial signal intensities [$\log_2$(PM/MM)] as well as on both $\Delta T_{d\text{-}w}$ and $\Delta T_{d\text{-}50}$. Mismatches in the central part of the probe were more destabilizing than mismatches near the extremities, leading to an inverse U-shaped relationship between mismatch position and destabilizing effect of the mismatch. For 25mer PM duplexes on Affymetrix chips, modeling suggested that Watson–Crick base pairs in the center of a probe contributed more to the stability of the duplex than base pairs toward the ends (42,44). Our study shows an analogous effect for mismatches, which have a greater destabilizing effect when placed around the middle part of a probe than toward the ends of the probe. For hybridizations in solution, the same U-shaped position effect has been found for duplexes containing a single 3-nitropyrrole artificial mismatch at all possible positions in a 15mer (45). Although the OMP software correctly predicts smaller destabilizing effects of mismatches at the three end positions, it does not predict a position effect for positions 4 and higher (43) and might therefore underestimate the importance of mismatch position.

Effect of attachment surface on destabilization

As indicated earlier, the 3′ end of the oligonucleotide probes was attached to the surface of the chip. The data suggest an asymmetry that mismatches near the surface distal 5′ end are more destabilizing than the ones near the surface proximal 3′ end. A similar observation has been reported for 60mer oligonucleotide arrays (22), where maximum discrimination was observed for mismatches ~10 bases away from the distal end, whereas single base pair mismatches within the first 30 bases near the proximal end did not show a considerable destabilizing effect. One hypothesis for this effect is that the duplex and the free dangling ends (i.e. single-stranded target not bound to a probe sequence) nearer to the surface get partly immobilized in the DNA film. In contrast, the more freely moving ends protruding into the solution may dissociate and diffuse away more easily when destabilized by a mismatch.

In conclusion, we have established experimentally that the physical parameter measured during a non-equilibrium melting curve, i.e. the dissociation rate constant, k_d, is very useful for the interpretation of hybridization signals. This analysis method is especially valuable when hybridization of complex samples leads to simultaneous binding of specific and unspecific targets to the same probe.

ACKNOWLEDGEMENTS

This project was funded in part by the MSU foundation, the Michigan Economic Development Corporation (GR-476 PO 085P3000517) and the NIH (1 R01 RR018625-01). Funding to pay the Open Access publication charges for this article was provided by the MSU Foundation, Michigan Economic Development Corporation, and the NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Norton,H. *et al*. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol*., **14**, 1675–1680.
2. Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.
3. Oostlander,A., Meijer,G. and Ylstra,B. (2004) Microarray-based comparative genomic hybridization and its applications in human genetics. *Clin. Genet*., **66**, 488–495.
4. Behr,M.A., Wilson,M.A., Gill,W.P., Salamon,H., Schoolnik,G.K., Rane,S. and Small,P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, **284**, 1520–1523.
5. Ochman,H. and Jones,I.B. (2000) Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J*., **19**, 6637–6643.
6. Bodrossy,L. and Sessitsch,A. (2004) Oligonucleotide microarrays in microbial diagnostics. *Curr. Opin. Microbiol*., **7**, 245–254.
7. Guschin,D., Mobarry,B., Proudnikov,D., Stahl,D., Rittmann,B. and Mirzabekov,A. (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl. Environ. Microbiol*., **63**, 2397–2402.
8. Hashsham,S.A., Wick,L.M., Rouillard,J.-M., Gulari,E. and Tiedje,J.M. (2004) Potential of DNA microarrays for developing parallel detection tools (PDTs) for microorganisms relevant to biodefense and related research needs. *Biosens. Bioelectron*., **20**, 668–683.
9. Wang,D., Coscoy,L., Zylberberg,M., Avila,P.C., Boushey,H.A., Ganem,D. and DeRisi,J.L. (2002) Microarray-based detection and genotyping of viral pathogens. *PNAS*, **99**, 15687–15692.
10. Loy,A., Lehner,A., Lee,N., Adamczyk,J., Meier,H., Ernst,J., Schleifer,K.-H. and Wagner,M. (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol*., **68**, 5064–5081.
11. Wilson,K.H., Wilson,W.J., Radosevich,J.L., DeSantis,T.Z., Viswanathan,V.S., Kuczmarski,T.A. and Andersen,G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes. *Appl. Environ. Microbiol*., **68**, 2535–2541.
12. Rhee,S.-K., Liu,X., Wu,L., Chong,S.C., Wan,X. and Zhou,J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl. Environ. Microbiol*., **70**, 4303–4317.
13. Peplies,J., Lau,S.C.K., Pernthaler,J., Amann,R. and Glockner,F.O. (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton. *Environ. Microbiol*., **6**, 638–645.
14. Guo,Z., Guilfoyle,R.A., Thiel,A.J., Wang,R. and Smith,L.M. (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res*., **22**, 5456–5465.
15. Drmanac,S., Kita,D., Labat,I., Hauser,B., Schmidt,C., Burczak,J.D. and Drmanac,R. (1998) Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nat. Biotechnol*., **16**, 54–58.

16. Hacia,J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genet*., **21** (Suppl.), 42–47.
17. Hacia,J.G. and Collins,F.S. (1999) Mutational analysis using oligonucleotide microarrays. *J. Med. Genet*., **36**, 730–736.
18. Zhou,J. and Thompson,D.K. (2002) Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotechnol*., **13**, 204–207.
19. Zhou,J. (2003) Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Microbiol*., **6**, 288–294.
20. Chandler,D.P. and Jarrell,A.E. (2004) Automated purification and suspension array detection of 16S rRNA from soil and sediment extracts by using tunable surface microparticles. *Appl. Environ. Microbiol*., **70**, 2621–2631.
21. Chandler,D.P. and Jarrell,A.E. (2005) Taking arrays from the lab to the field: trying to make sense of the unknown. *BioTechniques*, **38**, 591–600.
22. 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*. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol*., **19**, 342–347.
23. Southern,E., Mir,K. and Shchepinov,M. (1999) Molecular interactions on microarrays. *Nature Genet*., **21**, 5–9.
24. Shchepinov,M., Case-Green,S. and Southern,E. (1997) Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. *Nucleic Acids Res*., **25**, 1155–1161.
25. Peterson,A.W., Heaton,R.J. and Georgiadis,R.M. (2001) The effect of surface probe density on DNA hybridization. *Nucleic Acids Res*., **29**, 5163–5168.
26. SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct*., **33**, 415–440.
27. SantaLucia,J.Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *PNAS*, **95**, 1460–1465.
28. Peplies,J., Glockner,F.O. and Amann,R. (2003) Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. *Appl. Environ. Microbiol*., **69**, 1397–1407.
29. Mobarry,B., Wagner,M., Urbain,V., Rittmann,B. and Stahl,D. (1996) Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria [Erratum (1997) Appl. Environ. Microbiol. 63 815.]. *Appl. Environ. Microbiol*., **62**, 2156–2162.
30. Raskin,L., Stromley,J.M., Rittmann,B.E. and Stahl,D.A. (1994) Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol*., **60**, 1232–1240.
31. Liu,W.-T., Mirzabekov,A.D. and Stahl,D.A. (2001) Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ. Microbiol*., **3**, 619–629.
32. Li,E.S.Y., Ng,J.K.K., Wu,J.-H. and Liu,W.-T. (2004) Evaluating single-base-pair discriminating capability of planar oligonucleotide microchips using a non-equilibrium dissociation approach. *Environ. Microbiol*., **6**, 1197–1202.
33. Urakawa,H., Noble,P.A., El Fantroussi,S., Kelly,J.J. and Stahl,D.A. (2002) Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. *Appl. Environ. Microbiol*., **68**, 235–244.
34. Urakawa,H., El Fantroussi,S., Smidt,H., Smoot,J.C., Tribou,E.H., Kelly,J.J., Noble,P.A. and Stahl,D.A. (2003) Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Appl. Environ. Microbiol*., **69**, 2848–2856.
35. El Fantroussi,S., Urakawa,H., Bernhard,A.E., Kelly,J.J., Noble,P.A., Smidt,H., Yershov,G.M. and Stahl,D.A. (2003) Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl. Environ. Microbiol*., **69**, 2377–2382.
36. Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.G., Ohtsubo,E., Nakayama,K., Murata,T. *et al*. (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res*., **8**, 11–22.
37. Gao,X., LeProust,E., Zhang,H., Srivannavit,O., Gulari,E., Yu,P., Nishiguchi,C., Xiang,Q. and Zhou,X. (2001) A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res*., **29**, 4744–4750.
38. Wick,L.M., Qi,W., Lacher,D.W. and Whittam,T.S. (2005) Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *J. Bacteriol*., **187**, 1783–1791.
39. Blake,R. and Delcourt,S. (1996) Thermodynamic effects of formamide on DNA stability. *Nucleic Acids Res*., **24**, 2095–2103.
40. Everitt,B.S. (1977) *The Analysis of Contingency Tables*. Chapman and Hall, London.
41. Held,G.A., Grinstein,G. and Tu,Y. (2003) Modeling of DNA microarray data by using physical properties of hybridization. *PNAS*, **100**, 7575–7580.
42. Zhang,L., Miles,M.F. and Aldape,K.D. (2003) A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol*., **21**, 818–821.
43. Lee,I., Dombkowski,A.A. and Athey,B.D. (2004) Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray. *Nucleic Acids Res*., **32**, 681–690.
44. Mei,R., Hubbell,E., Bekiranov,S., Mittmann,M., Christians,F.C., Shen,M.-M., Lu,G., Fang,J., Liu,W.-M., Ryder,T. *et al*. (2003) Probe selection for high-density oligonucleotide arrays. *PNAS*, **100**, 11237–11242.
45. Guo,Z., Liu,Q. and Smith,L.M. (1997) Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nat. Biotechnol*., **15**, 331–335.