DNA multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison

Daniel J. Fish[1],*, M. Todd Horne[1,2], Greg P. Brewood[1], Jim P. Goodarzi[1], Saba Alemayehu[1], Ashwini Bhandiwad[1], Robert P. Searles[3] and Albert S. Benight[1,2]

[1]Portland Bioscience, Inc., [2]Departments of Chemistry and Physics, Portland State University, Portland and [3]Spotted Microarray Core, Oregon Health and Science University, Beaverton, OR, USA

ABSTRACT

Hybridization intensities of 30 distinct short duplex DNAs measured on spotted microarrays, were directly compared with thermodynamic stabilities measured in solution. DNA sequences were designed to promote formation of perfect match, or hybrid duplexes containing tandem mismatches. Thermodynamic parameters ΔH°, ΔS° and ΔG° of melting transitions in solution were evaluated directly using differential scanning calorimetry. Quantitative comparison with results from 63 multiplex microarray hybridization experiments provided a linear relationship for perfect match and most mismatch duplexes. Examination of outliers suggests that both duplex length and relative position of tandem mismatches could be important factors contributing to observed deviations from linearity. A detailed comparison of measured thermodynamic parameters with those calculated using the nearest-neighbor model was performed. Analysis revealed the nearest-neighbor model generally predicts mismatch duplexes to be less stable than experimentally observed. Results also show the relative stability of a tandem mismatch is highly dependent on the identity of the flanking Watson–Crick (w/c) base pairs. Thus, specifying the stability contribution of a tandem mismatch requires consideration of the sequence identity of at least four base pair units (tandem mismatch and flanking w/c base pairs). These observations underscore the need for rigorous evaluation of thermodynamic parameters describing tandem mismatch stability.

INTRODUCTION

Microarray technology has found wide ranging utility in high-throughput, highly parallel assessment of DNA target sequences via sequence specific DNA hybridization. Provided sufficient quantitative benchmarks can be achieved, multiplex assays performed on microarrays could become an enabling cornerstone heralding the era of personalized medicine, i.e. medical diagnosis and prognosis based on unique genotypes of individual humans. Intensities observed on DNA microarrays are dictated by hybridization behavior of the manifold sequences (and duplexes they form) existent in a multiplex reaction environment. The relationship between hybridization intensities observed for duplex complexes formed on microarrays and the thermodynamic stability of the same duplexes in solution is not precisely known. This lack of knowledge poses a serious concern for effective implementation of sequence design strategies, since most assay design approaches rely heavily on sequence specific thermodynamic stability parameters evaluated from solution measurements. Without exception, all sequence stability predictions currently employ the nearest-neighbor model (1).

Current microarray assay design approaches utilize solution derived parameters to assemble sets of sequences that have desirable sequence dependent stabilities for use in microarray-based multiplex hybridization reactions. An underlying premise in this approach is that the thermodynamic rules for nearest-neighbor base pairing, and therefore predictions for short duplex DNAs, are essentially the same in solution as they are on microarrays. It is usually assumed that microarray intensities scale in a linear fashion with DNA duplex stability. However, the use of thermodynamic parameters evaluated from solution

---

*To whom correspondence should be addressed. Tel: (503) 725-2350; Fax: (503) 725-2305; Email: djf@pdxbio.com

measurements for predicting microarray hybridization intensities has not been critically tested.

An additional complicating factor associated with hybridization reactions is cross-hybridization, i.e. the formation of unintended duplexes. In a multiplex microarray environment, where many duplexes are hybridized simultaneously, the potential for cross-hybridization is significant (2). The propensity for cross-hybridization depends on a number of factors, including the number (and relative concentrations) of strands present and the relative stability of hybrid duplex complexes comprised primarily of a mixture of Watson–Crick (w/c) perfect match and tandem mismatch base pairs in the same duplex (herein, the term 'tandem mismatch' refers to two or more contiguous mismatched bases). Although, the effects tandem mismatches may have on hybridization experiments is not clearly understood, their potential effects on experimental results should not be discounted, as the number of possible tandem mismatches that can form in a duplex increases exponentially with the number of bases. Thus, thermodynamic stabilities of tandem mismatches and associated effects of mismatch formation on duplex stability are expected to play an important role in the study of DNA multiplex hybridization.

The aim of this study was to critically examine the relation between DNA hybridization thermodynamics measured in solution and hybridization intensities measured on microarrays for the same DNA sequences. The possibility of estimating semi-quantitative values of thermodynamic parameters for DNA duplex stability from microarray intensity data was explored. Effects of duplex length, number of mismatches and mismatch position on solution stability and microarray hybridization intensity were investigated. It is widely assumed that, for perfect match duplexes, solution stabilities and microarray intensities are linearly related, and thus the rules of predicting duplex stability on both microarrays and in solution are essentially the same. Although such a linear relationship has not been definitively established, results of this study support this assumption for both perfect match and mismatch duplexes. Previous studies have reported on relationships between calculated thermodynamic stabilities in solution and microarray hybridization intensities (3–5). In these studies, model-based calculations using published nearest-neighbor thermodynamic parameters were employed to calculate duplex thermodynamic stability, but because sequence dependent thermodynamic stability parameters for tandem mismatches have not been evaluated, such calculations did not explicitly account for thermodynamic contributions from tandem mismatches in short DNA duplexes. Here, we present the first direct comparison of measured thermodynamic melting parameters with hybridization intensities, for the same duplexes, measured on microarrays.

MATERIALS AND METHODS

DNA sequences

On each microarray, 10 different probe sequences were spotted at multiple sites. These sequences were designed to have nearly equal thermodynamic stability and exhibit minimum cross-hybridization with one another. Three different sets of 10 DNA target strands, uniquely defined by their sequences to form different types of duplex complexes with 10 probe sequences, were examined. This resulted in three sets (types) of 10 DNA duplexes, i.e. probe-target pairs, with lengths varying from 26–32 bp. Type 1 was designed to form perfect match duplexes with specific probe sequences and not cross-hybridize, when all perfect match target strands are simultaneously present. Each duplex in Type 1 contains only perfect-matched (w/c) base pairs with the exception of duplex 5.1, which has an A/C mismatch near the middle (Table 1, column 1). Type 2 duplexes are derivatives of the first set, the difference being that duplexes in Type 2 contain tandem mismatches replacing w/c base pairs near the center (Table 1, column 2). Duplexes in set 3 have multiple tandem mismatches in different arrangements in each of the sequences (Table 1, column 3). The 30 duplexes comprising Types 1, 2 and 3 were melted in solution and their transition thermodynamic parameters were directly measured by differential scanning calorimetry (DSC). Multiple hybridization experiments on microarrays were also performed. A total of 63 distinct multiplex hybridization experiments were conducted, each containing different combinations of target strands hybridized to the same probe set, forming the duplexes shown in Table 1.

Solution melting experiments

For solution melting experiments, DNA samples were re-suspended in 85 mM Na^+ buffer solution (75 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA). Measurements of the excess heat capacity ΔC_p versus temperature were made using a Nano DSC^{TM} (Calorimetry Sciences Corp. differential scanning calorimetry). Data analysis was performed using the Cpcalc^{TM} (Calorimetry Sciences Corp.) routine and produced the transition enthalpy (ΔH°) and entropy (ΔS°). Free energy (ΔG°) was calculated using the formula $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ with $T = 298$ K. For all melting experiments, DNA concentrations ranged from 75–130 μM. Complete DSC melting curves collected for all 30 duplexes are included as Supplementary Data.

Microarray hybridization

A total of 63 microarray hybridization experiments were performed. In each experiment, all 10 probes were hybridized with 10 distinct targets simultaneously. Probes were re-suspended at a strand concentration of 50 μM in Micro Spotting Solution^{TM} spotting buffer (ArrayIt). For printing, three microliter aliquots were placed in individual wells of a 384-well plate. Each probe was spotted 10 times on aldehyde slides (Telechem) using a PixSys^{TM} 5500XL microarray printer (Genomic Solutions) with one Chip-Maker^{TM} 3 pin (TeleChem). Probes were fixed to the slides by placing them in a desiccator in the dark for at least 24 h.

Target strands labeled with Cy-3 were hybridized to the arrays at a concentration of 10 μM in 5X SSC, 0.1% SDS at 30°C for 16 h. Arrays were then washed three times for

Table 1. Duplexes in Sets 1, 2 and 3

Set 1 duplexes	Set 2 duplexes	Set 3 duplexes
1.1 5'GATTGTGAGTAATCATACACATTGATAAA 3'CTAACATCATTAGTATGTGTAACTATTT	1.2 5'GATTGTGAGTAATCATACACATTGATAAA 3'CTAACATCATTA<u>TGTA</u>GTGTAACTATTT	1.3 5'GATTGTGAGTAATCATACACATTGATAAA 3'<u>AG</u>AACATCATTAGTATGTGTAACTATTT
2.1 5'TAAGATGTAGATCTAACTATAGAAGATT 3'ATTCTACATCTAGATTGATATCTTCTAA	2.2 5'TAAGATGTAGATC<u>TAA</u>CTATAGAAGATT 3'ATTCTACATCTAT<u>TGA</u>GATATCTTCTAA	2.3 5'<u>TT</u>AAGATGTAGATCTAACTATAGAAGATT 3'ATTCTACATCTAGATTGATATA<u>GGC</u>TAA
3.1 5'TAGTGAAGGAGTCTAGACTATAGTTTATT 3'ATCACTTCCTCAGATCTGATATCAAATAA	3.2 5'TAGTGAAGGAGT<u>CTAGA</u>CTATAGTTTATT 3'ATCACTTCCTCA<u>TTGAA</u>GATATCAAATAA	3.3 5'T<u>AGT</u>GAAGGAGTCTAGACTA<u>TAGTTT</u>ATT 3'ATCACTTCCTCAGATCTGATATCAAATAA
4.1 5'GAAAAAGTGTAATGGAAAATGTTGAGTA 3'CTTTTTCACATTAACCTTTACAACTCAT	4.2 5'GAAAAAGTGTAATGGAAAATGTTGAGTA 3'CTTTTTCACATTAA<u>GG</u>TTTACAACTCAT	4.3 5'GAAAAAGTGTAATGGAAAATGTTGAGTA 3'C<u>AG</u>TTTCACATTACA<u>GTT</u>TACAAA<u>GC</u>AT
5.1 5'TGATGTAATGACCTAAATCCAAAGATTGTGT 3'ACTACATTACTGGATTTAGGCTTTCTAACACA	5.2 5'TGATGTAATGACCT<u>AAA</u>TCCAAAGATTGTGT 3'ACTACATTACTGGA<u>TTTAGG</u>C<u>TTT</u>CTAACACA	5.3 5'<u>TG</u>ATGTAATGACCT<u>AAA</u>TCCAAA<u>AG</u>ATTGTGT 3'AAGACATTA<u>AGG</u>GA<u>T</u>AAAGGCTTT<u>AG</u>AACACA
6.1 5'TTATGAAGCAACGAAATTAATGAGAA 3'AATACTTCGTTGCTTTAATTACTCTT	6.2 5'TTATGAAGCAACGAAATTAA<u>T</u>GAGAA 3'AATACTTCGTTGC<u>A</u>TTAATTACTCTT	6.3 5'TTATGAAGC<u>AA</u>CGAAATTAATGA<u>GA</u> 3'AATACTTCGTTGCTTTAATTACTA<u>GG</u>
7.1 5'AAGAAAGATTAGGACATGAGATTATG 3'TTCTTTCTAATCCTGTACTCTAATAC	7.2 5'AAGAAAGATTAGGACATGAGATTATG 3'TTCTTTCTAATCA<u>G</u>TGTACTCTAATAC	7.3 5'AAGAAAGATTAGGACA<u>T</u>GAGAT<u>TAT</u>G 3'TT<u>AGT</u>TGAATCGAGTAA<u>GC</u>TAATAC
8.1 5'TTAGTTAGTATACGGAAACTGTTAGTTA 3'AATCAATCTATGCCTTGACAATCAAT	8.2 5'TTAGTTAGTATACG<u>G</u>AAACTGTTAGTTA 3'AATCAATCTATGA<u>GG</u>TTGACAATCAAT	8.3 5'TT<u>AGT</u>TAGTATAC<u>GG</u>AAACTGTTAGTTA 3'AA<u>GA</u>AAAT<u>AG</u>ATGA<u>AG</u>TTGACAA<u>GA</u>AAT
9.1 5'TAGTGTAGTAACGGGAAATCTAAAGTGT 3'ATCACATCATTGCCCTTTAGATTTCACA	9.2 5'TAGTGTAGTAACGGGAAATCTAAAGTGT 3'ATCACATCATTGCCCTTTAGA<u>TTT</u>CACA	9.3 5'TAGTGTAGTAACGGGAAATCTA<u>AA</u>GTGT 3'ATCACATCATTGCCCTTTAGAGAGAACA
10.1 5'TTATGAAATTATGAAATAGTGTAGAT 3'AATACTTTAATACTTATCACATCTA	10.2 5'TTATGAAATTAT<u>GAAA</u>TAGTGTAGAT 3'AATACTTTAATA<u>AGTT</u>ATCACATCTA	10.3 5'TTATGAAATTATGAAATA<u>GTGT</u>AGAT 3'AATACTTTAATA<u>AG</u>TTATCACATCTA

Duplexes in Set 1: all perfect-matches except 5.1, which contains one single base-pair mismatch. Duplexes in Set 2: each contains from 2 to 5 mismatches near the middle. Duplexes in Set 3: each contains multiple mismatches and/or mismatches near the ends.

5 min each at room temperature in 0.6X SSC/1.0% SDS followed by 0.06X SSC/0.1% SDS, and 0.06X SSC. Arrays were scanned using a ScanArray[TM] 4000XL (PerkinElmer), equipped with a 543 nm laser at 5 μm resolution. Images were quantified using ImaGene[TM] 5.6 (BioDiscovery). Mean intensity for each spot was calculated after trimming of the lowest and highest 5% of the signals.

Actual concentration of spotted probes on the surface is not known and is difficult to determine. Recently, microarray hybridization experiments conducted under similar conditions, at different concentrations of added target, indicated that target was in excess of probe concentration with clear resolution of mismatch duplexes (6). A similar set of conditions was assumed to prevail in the present study.

RESULTS

General observations

In each microarray experiment, individual probes were designed to hybridize with unique targets, forming duplexes with thermodynamic stability (ΔG°), which was determined from solution measurements. Each bound probe/target pair is associated with a unique mean probe-spot intensity. The mean signals associated with each probe/target complex were corrected for background, normalized by signal mean and averaged over the entire chip. Results were then compared with the solution thermodynamic parameters measured for each duplex.

In analyzing microarray intensity data, three sources of signal variation are typically encountered: chip-to-chip signal variation, signal variation within each chip and signal variation between different print runs. Being aware of the potential sources for signal variation, initial data analysis was performed on results obtained from a subset of 11 microarray experiments that were shown to have relatively small variations of each type. These 11 experiments were identified by constructing a correlation coefficient matrix for the full data set to compare replicate probe signals for all the data (data not shown). The sets of data with the highest correlation coefficients, larger than 0.8, comprised the 'primary data set'. The correlation coefficient matrix also showed that two print runs were statistically robust and gave consistent replicate data. Given the current experimental variability of the spotted microarray process, this analysis demonstrates the importance of careful statistical analysis of spot-to-spot reproducibility and chip-to-chip variability for microarray data when quantitative results are desired. The complete set of results from all 63 experiments is referred to as the 'full data set'. A direct comparison of microarray results from the primary and full data sets is shown in Figure 1a. Thermodynamic parameters measured in solution by DSC are displayed in Table 2. Results from microarray experiments for both the primary and full data sets are shown in Table 3.

A comparison between duplex free energy, $\Delta G^\circ(25^\circ C)$, and microarray mean (over multiple probe spot replicates) hybridization intensities for the primary data set is shown

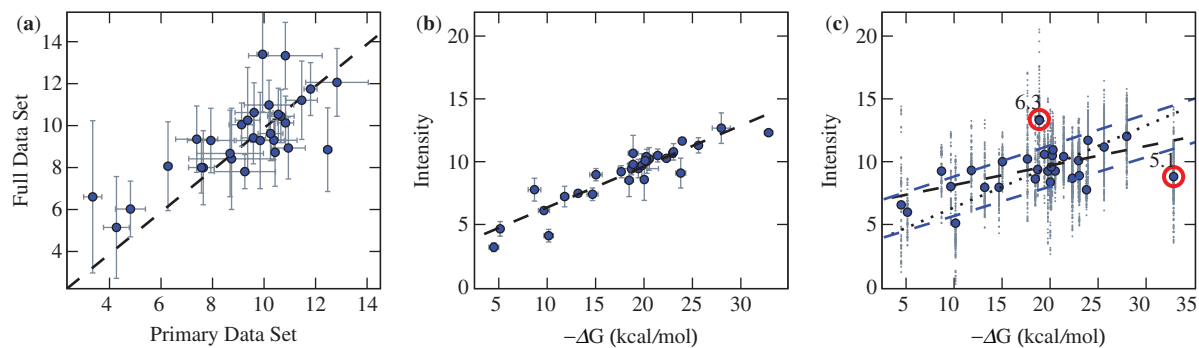**Figure 1.** Comparisons of microarray hybridization and solution melting. (**a**) Plot of intensities (normalized by signal mean) of the full data set collected from 63 experiments versus the primary microarray data set comprised of results from 11 experiments. The dashed line shows the best linear fit to the data ($R = 0.75$). (**b**) Plot of the primary data set of relative microarray intensities versus the measured free energy, $\Delta G°(25°C)$. The dashed line is the best linear fit to the data ($R = 0.92$). (**c**) Plot of the full data set of the relative microarray intensity versus the measured free energy, $\Delta G°(25°C)$. The dotted line is the same line as in (b) obtained for the best fit to the primary data set. The black dashed line is the best linear fit to the full data set ($R = 0.53$). Upper and lower dashed lines depict the parameter window bounded by the line $y = m \cdot \Delta G° + b$, with $m = 0.250 \pm 0.007$ and $b = 4.86 \pm 1.50$.

Table 2. Thermodynamic parameters measured in solution

Duplexes	Free energy		Enthalpy		Entropy	
	$-\Delta G°$ kcal/mol	σ^2	$-\Delta H°$ kcal/mol	σ^2	$-\Delta S°$ kcal/k·mol	σ^2
1.1	22.18	0.34	194.20	2.14	0.58	0.01
1.2	13.05	0.16	154.30	2.61	0.47	0.01
1.3	18.57	0.19	170.40	3.39	0.51	0.01
2.1	22.82	0.37	201.55	1.13	0.60	0.00
2.2	10.01	0.49	138.30	0.57	0.43	0.00
2.3	14.90	0.64	155.85	1.06	0.47	0.00
3.1	25.48	0.36	209.20	0.99	0.62	0.00
3.2	11.66	0.42	158.80	2.48	0.49	0.01
3.3	19.27	2.08	171.33	12.01	0.51	0.04
4.1	23.95	0.52	202.45	4.02	0.60	0.01
4.2	19.88	0.11	185.13	2.98	0.55	0.01
4.3	18.32	0.45	181.40	1.09	0.55	0.00
5.1	32.72	0.44	274.40	3.76	0.81	0.01
5.2	19.63	0.28	202.75	0.35	0.61	0.00
5.3	9.52	0.58	161.65	2.05	0.51	0.00
6.1	21.27	0.49	177.20	2.47	0.52	0.01
6.2	17.50	0.31	159.50	1.22	0.48	0.00
6.3	18.74	0.51	165.50	2.14	0.49	0.01
7.1	22.90	0.23	194.70	1.56	0.58	0.01
7.2	20.40	0.37	183.55	2.79	0.55	0.01
7.3	8.55	0.65	142.35	1.06	0.45	0.00
8.1	23.80	0.06	195.60	1.41	0.58	0.00
8.2	20.09	0.12	182.28	2.38	0.54	0.01
8.3	3.16	0.51	93.75	2.62	0.30	0.01
9.1	27.84	0.94	220.13	6.32	0.65	0.02
9.2	20.16	0.34	192.63	1.82	0.58	0.01
9.3	18.73	0.51	165.20	3.11	0.49	0.01
10.1	19.98	0.52	185.97	1.89	0.56	0.00
10.2	14.54	0.57	154.90	1.41	0.47	0.00
10.3	5.00	0.13	79.80	1.13	0.25	0.00

Table 3. Microarray intensities of the primary data set and the full set of 63 experiments (arbitrary units)

Duplexes	Primary set		Full set	
	average Int	σ^2	average Int	σ^2
1.1	10.37	0.15	8.69	1.62
1.2	7.59	0.13	7.87	1.77
1.3	9.53	0.47	9.48	1.25
2.1	10.78	0.31	10.17	1.28
2.2	4.22	0.50	5.21	2.43
2.3	9.07	0.49	10.18	1.04
3.1	11.41	0.61	11.52	1.87
3.2	7.33	0.82	9.38	1.59
3.3	9.56	0.74	10.69	1.43
4.1	9.20	1.19	7.87	0.81
4.2	8.68	1.66	8.67	2.42
4.3	8.63	1.30	8.74	2.07
5.1	12.42	0.16	9.22	2.00
5.2	9.80	0.94	9.19	2.30
5.3	6.22	0.15	8.34	2.13
6.1	10.60	0.52	10.42	1.26
6.2	9.31	0.48	11.01	2.52
6.3	9.89	0.22	13.68	2.76
7.1	10.89	0.66	9.66	1.48
7.2	10.33	0.90	9.68	0.93
7.3	7.88	0.91	8.99	1.54
8.1	11.75	0.25	11.83	1.26
8.2	10.50	0.29	11.60	1.25
8.3	3.29	0.35	6.67	3.65
9.1	12.78	1.22	12.24	1.62
9.2	10.14	0.96	11.57	1.18
9.3	10.77	1.43	13.59	1.60
10.1	10.20	0.48	9.62	1.29
10.2	7.52	0.50	8.41	1.22
10.3	4.76	0.58	6.09	1.34

in Figure 1b. The best linear fit, $y = mx + b$, had $m = 0.33$ and $b = 3.18$, and a correlation coefficient $R = 0.93$. Note that these data were derived for a collection of duplexes containing both perfect matches and tandem mismatches. The linear trend is independent of whether the duplexes consist of perfectly matched base pairs or

contain tandem mismatches. Data from all 63 experiments involving the same DNA duplexes (full data set) were analyzed in a similar manner. Averages from the full data set were fit to a linear model, $y = mx + b$, with $m = 0.17$, $b = 6.56$ providing the best fit. This line is plotted in Figure 1c along with the original line of correlation

from Figure 1b. For each linear regression, reduced χ^2 values χ_p^2 (in relation to the primary data set) and χ_f^2 (in relation to the full data set) were calculated, with the following results. For the best fit to the primary data set: $\chi_p^2 = 1.76$ (Figure 1b), and for the best fit to the full data set: $\chi_f^2 = 1.51$ (Figure 1c).

By averaging the lines of best fit for both the primary and the full data set, the following linear relation between free energy in solution, ΔG°(solution), and average (mean-normalized) microarray intensity, I_{avg}, was obtained:

$$I_{avg} = -m\Delta G^\circ(\text{solution}) + b,$$

with $m = 0.25 \pm 0.07$, $b = 4.86 \pm 1.50$. Any line within this parameter window will have a reduced χ^2 $1 < \chi_f^2 < 4$ in relation to the full data set (this includes of the best linear fit for the primary data set, which has $\chi_f^2 = 2.30$). This relation enables estimation of the relative microarray intensity as a function of either measured or predicted thermodynamic free energy of DNA duplex formation.

Signal variability

The overall comparison shown in Figure 1c reveals a linear relation between duplex stability in solution (ΔG°) and microarray hybridization intensity. Solid circles denote averages of the measured hybridization intensities versus experimentally determined free energy, ΔG°(25°C), of the full data set. It is clear from the figure that significant variations in microarray intensities exist. Much of this signal variability is likely due to experimental variations on the microarray, since a closer linear fit was found for the subset of experiment with less experimental variation (Figure 1b). Additional sources of increased variability might include competition of targets for probe spots via cross-hybridization depending on position and type of tandem mismatches.

Due to effects that different experimental conditions can have on fluorescence intensities, the slope of this line could vary under different conditions. For example, it is known that wash criteria (stringency and wash time) can result in significant scaling of the overall observed intensities, which could result in a linear relation with greater slope. In the experiments performed here, all hybridized chips were washed under the same conditions (see 'Materials and methods' section) to ensure consistency of intensity data.

Deviations from linearity

The linear behavior of the experimental data for mismatch duplexes implies that, in general, the formation of mismatched bases affects both ΔG° and microarray intensity in equal proportion. However, significant deviations from the linear relation are observed in at least two cases. As shown in Figure 1c, two anomalous duplexes, 6.3 and 5.1 (circled in red), fall well outside the linear trend established by the other molecules. Duplex 5.1 forms a 32 bp duplex (Table 1), which is longer than any other duplex, and is also the only duplex with a single base pair mismatch. Duplex 6.3 contains three mismatched bases on the labeled end.

For duplex 6.3, there appears to be a relatively greater destabilizing effect in solution than on hybridization intensity (microarray binding). There is evidence in the literature that higher than expected microarray intensities can arise from a duplex with terminal mismatches due to the so-called 'bright mismatch' phenomena (3,7–14). This effect produces anomalous behavior like that observed for duplex 6.3, in which microarray intensity is greater than would be expected in comparison to stability in solution. Although this explanation for the outlying behavior for duplex 6.3 is plausible and consistent, further experimental validation will be required before generality of the effect can be established.

The behavior observed for duplex 5.1 appears to be the result of duplex length (32 versus 26 bp). While the longer duplex is more stable in solution it does not produce proportionally higher hybridization intensity on the microarray. Certainly, a longer duplex is expected to exhibit greater stability in solution, which would result in shifting the corresponding data point to the right in Figure 1c. Hybridization efficiency on microarrays is also known to be affected by duplex length (3,15,16), but an adequate physical explanation for this observation has not been provided. Increased duplex length might complicate strand orientation resulting in weaker binding or increased likelihood of non-specific binding. Such factors could contribute to relative differences in microarray hybridization intensities.

Linearity and mismatch number

Sequences of DNA duplexes examined in experiments are grouped into Type 1, 2 or 3 according to the number and type of mismatches that form in each duplex. Type 1 duplexes contain only perfect-match duplexes (with the exception of duplex 5.1). Duplexes in Type 2 each contain from 2 to 5 mismatches near the middle, while duplexes in Type 3 duplexes contain multiple mismatches interspersed throughout the sequence or mismatches on or near the ends. Correlations between free energy and microarray hybridization intensities were determined for each duplex set. As shown in Figure 2, the relation between ΔG° and intensity is progressively more linear as duplex stability decreases. For the primary data set, duplexes with the most mismatches (Type 3) exhibited the highest degree of linearity (correlation coefficient $R = 0.93$, $\chi_p^2 = 1.58$), while duplexes with no mismatches (Type 1) exhibited the lowest degree of linearity (correlation coefficient $R = 0.73$, $\chi_p^2 = 1.07$).

These results suggest that, while a linear relation between microarray intensity and ΔG° was observed over the measured range of stabilities (Table 2), an increasing degree of nonlinearity occurs for more stable (larger ΔG°) duplexes. Thus, longer and more stable duplexes tend to fall outside the linear dynamic range displayed by the less stable duplexes.

An alternate description of the relation between stability in solution and microarray hybridization can be achieved using a Langmuir model analysis. This approach is not unprecedented and has been previously applied to the analysis of microarray hybridization data (3,4).
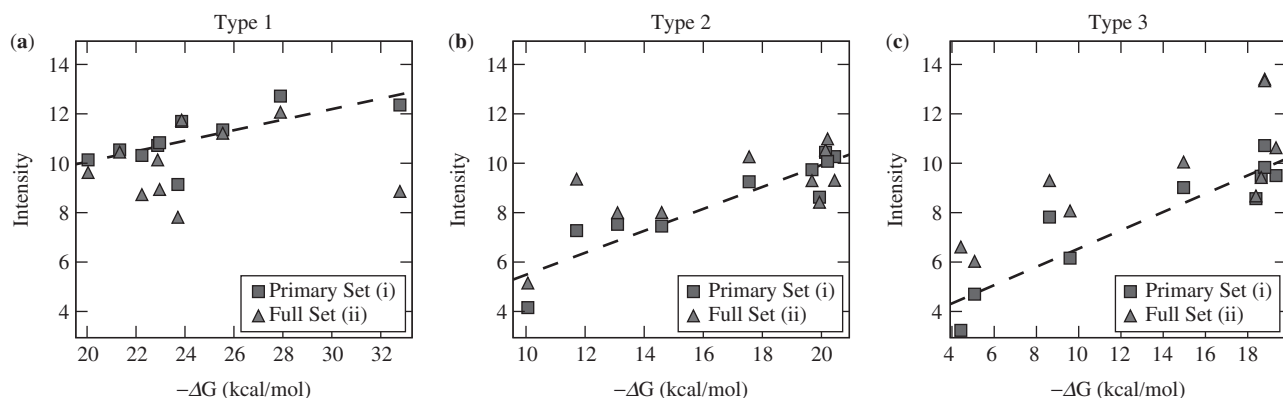
Figure 2. Plots of intensity (normalized by signal mean) versus the measured free energy, $\Delta G°(25°C)$, for each type of duplex set. (**a**) Type 1 duplexes containing perfect match w/c base pairs. Correlation coefficients: $R = 0.88$ (primary set), $R = 0.26$ (full set); (**b**) Type 2 duplexes containing mismatches in the center. Correlation coefficients: $R = 0.92$ (primary set), $R = 0.71$ (full set); (**c**) Type 3 duplexes containing multiple mismatches. Correlation coefficients: $R = 0.93$ (primary set), $R = 0.78$ (full set). The best linear fit to the primary data set is shown as a dashed line in each case.

A parameterized, Langmuir model describing the relation between $K^{eq} = \exp(-\Delta G°/RT)$ and microarray intensity can be expressed as:

$$\text{Intensity} = L(\Delta G°) = \frac{abK^{eq}}{[1 + aK^{eq}]}$$

where K^{eq} is the equilibrium constant. Model parameters (a and b) were fit to the data using a minimization routine (e.g. least squares optimization). Assuming physical values for all parameters, the plot of a Langmuir function, intensity versus ($\Delta G°$), provided the Langmuir isotherm, a well-known characterization of surface adsorption and reaction kinetics. If signal values are below saturation, it is reasonable to assume that microarray intensity, as a function of $\Delta G°$, will follow a Langmuir isotherm, especially for highly stable duplexes.

As shown in Figure 3, such a Langmuir function can be parameterized to fit both the primary data set and the full data set reasonably well. Best fit parameters for the Langmuir model were $a = 0.12$, $b = 13.34$ for the primary set, and $a = 0.13$, $b = 13.76$ for the full data set. Averaging these curves, the parameter window $a = 0.13 \pm 0.02$, $b = 13.55 \pm 0.75$ was obtained which can be used to estimate microarray intensity from free energy. However, because critical model parameters (probe density and target concentration) are not precisely known, these fits obtained should be viewed as demonstrative. The primary purpose of this study is to define the general (empirical) relationship between solution thermodynamic stability and microarray hybridization intensity. Moreover, our aim is to identify sources of deviations from the linear trend. Certainly, an in-depth model analysis of the appropriate mechanism for transforming from solution thermodynamics to microarray hybridization intensities will be required for establishment of an accurate quantitative metric, but such an analysis is well beyond the scope of this work.

Effect of mismatched bases

In order to analyze effects of mismatched bases on $\Delta G°$ in solution, and their concomitant effects on microarray
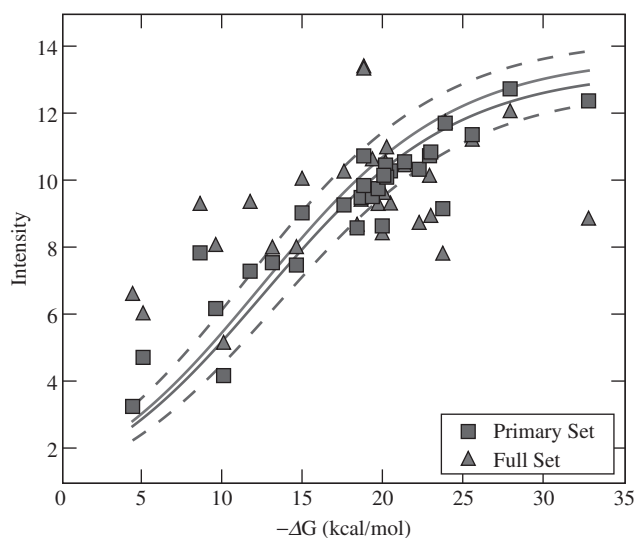

Figure 3. Plot of best fits of a Langmuir model to both the primary data set (squares) and the full data set (triangles). Model parameters were $a = 0.12$, $b = 13.34$ for the primary set, and $a = 0.13$, $b = 13.76$ for the full data set. Upper and lower dashed curves depict the parameter window bounding the Langmuir function with $a = 0.125 \pm 0.020$, $b = 13.55 \pm 0.75$.

hybridization intensities, experimental values obtained for each type of duplex were directly compared. In Figure 4, solution and microarray results are displayed in separate subplots. In each subplot, results are grouped according to probe number (1–10), and each type of duplex (Type 1— perfect match, Type 2 and Type 3) is represented as a histogram. Several interesting points arise from this analysis.

First, it is clear from the upper subplot in Figure 4 that in solution, all duplexes with multiple mismatches are found to be less stable than their perfect match counterparts. This contrasts results from microarray experiments (lower subplot in Figure 4), where several mismatch duplexes exhibit higher intensities than the perfect match. There are several sequences (1–4,6,9) for which Type 3

duplex microarray intensity is comparable to, or higher than the perfect match intensity. These all have multiple mismatches within four base pairs of either the free end or the tethered end of the duplex. This observation is consistent with previous comments regarding terminal mismatches with unexpectedly high intensities on the microarray. Solution results also show that duplexes with mismatches on the end have a higher stability than those with mismatches in the middle, but none are more stable than the perfect match.

For all duplexes, the ratios of mismatch duplex intensity (Types 2 and 3) to perfect match intensity (Type 1) and the ratios of mismatch duplex free energy, ΔG°(mm) to perfect match free energy, ΔG°(pm) were determined. Absolute differences in free energies, $\Delta\Delta G$(kcal/mol), were also calculated for each duplex. These are summarized in Table 4. Results reveal that effects of mismatches

on microarray intensities are significantly less pronounced than on ΔG° values in solution (Figure 4). Several intensities for Type 3 (mismatch) duplexes are greater than the perfect match on the microarray, while in solution none of the mismatch duplex stabilities exceeds those of the perfect matches. In solution, $\Delta\Delta G$ was found to range from 2.51 to 13.82 kcal/mol for Type 2 and from 2.54 to 23.20 kcal/mol for Type 3. The average percent change in microarray intensities was 8% for Type 2 and only 4% for Type 3.

Mismatches in the nearest-neighbor model

The thermodynamic transition parameters obtained in these experiments provided an opportunity to compare predictions of the nearest-neighbor model to measured stabilities of short duplexes containing tandem mismatches.

For each set of duplexes, the difference in free energy ($\Delta\Delta G = \Delta G$(mm) $- \Delta G$(pm)) between Type 1 (perfect match) and Type 2 and between Type 1 and Type 3 (mismatch) duplexes was predicted using the nearest-neighbor model. Because the duplex sequences of molecules in each set are the same, i.e. Type 2 and Type 3 duplex sequences are derivatives of Type 1, differences between thermodynamic parameters for perfect match and mismatch duplexes should reflect only contributions (perturbations) due to mismatched bases. Although thermodynamic parameters for single base mismatches are known (17), parameters for tandem mismatches are currently unavailable, due in part to the extremely large number of possible mismatches of a given length (greater than one). Consequently, approximate strategies using the nearest-neighbor model have been developed to predict the thermodynamic stability of mismatch duplexes (17).

The most commonly employed approximate method assumes that tandem mismatches make no contribution to thermodynamic stability of the duplex, i.e. nearest-neighbor stacking interactions are fully disrupted by tandem mismatches. Therefore, mismatched regions are considered to be disordered loops, and the thermodynamic effects of these looped regions are accounted for solely by a loop entropy term, $\Delta G^\circ_{\text{loop}}(n)$, which depends
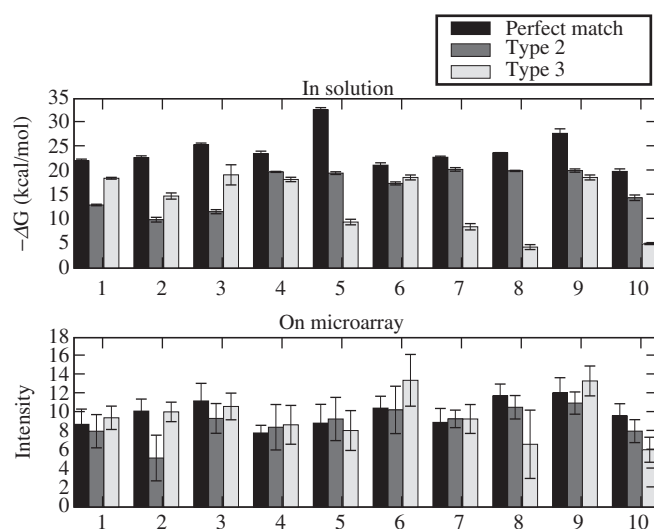

Figure 4. Charts showing relative effect of mismatches on stability and microarray intensity. Results are grouped according to probe number, and each type of duplex: Type 1 (perfect match), Types 2 and 3 are represented by a different histograms. Solution results are shown in the upper subplot, and microarray results are shown in the lower subplot.

Table 4. Ratios and differences of mismatch and perfect match data

| Seq	Solution (ΔG°)				Array (Int.)			
	Type2/PM %	Type3/PM %	Type2–PM kcal/mol	Type3–PM kcal/mol	Type2/PM %	Type3/PM %	Type2–PM kcal/mol	Type3–PM kcal/mol
1	0.59	0.84	9.13	3.61	0.91	1.09	−0.82	0.79
2	0.44	0.65	12.81	7.93	0.51	1.00	−4.96	0.01
3	0.46	0.76	13.82	6.21	0.81	0.93	−2.14	−0.83
4	0.83	0.76	4.06	5.63	1.10	1.11	0.80	0.87
5	0.60	0.29	13.09	23.20	1.00	0.90	−0.02	−0.88
6	0.82	0.88	3.77	2.54	1.06	1.31	0.60	3.26
7	0.89	0.37	2.51	14.36	1.00	0.93	0.01	−0.67
8	0.84	0.18	3.71	19.45	0.98	0.56	−0.23	−5.16
9	0.72	0.67	7.68	9.11	0.95	1.11	−0.67	1.35
10	0.73	0.25	5.44	14.98	0.87	0.63	−1.21	−3.53
Average	0.69	0.57	7.60	10.70	0.92	0.96	−0.86	−0.48
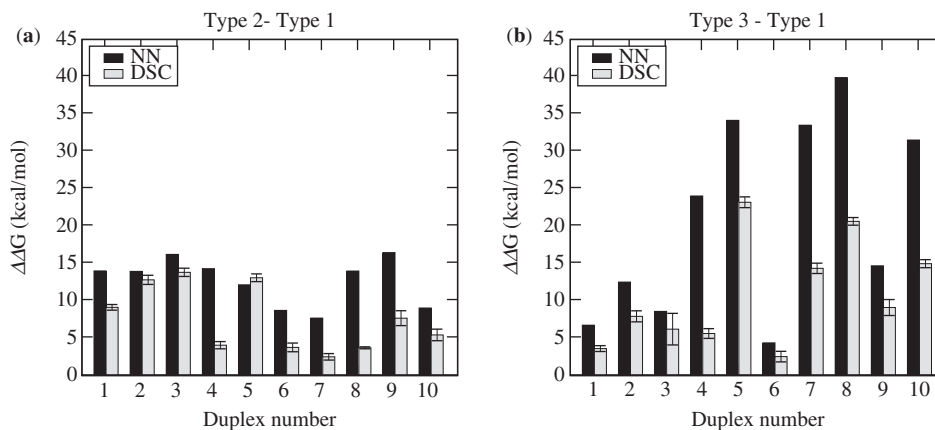
Figure 5. Comparison of the difference, $\Delta\Delta G = \Delta G^\circ(\text{mm}) - \Delta G^\circ(\text{pm})$, in free energies between perfect match and mismatch duplexes predicted using the nearest-neighbor model and experimentally measured by DSC. (**a**) Differences in ΔG° between Type 2 and Type 1 duplexes. (**b**) Differences in ΔG° between Type 3 and Type 1 duplexes. Dark bars (NN) depict $\Delta\Delta G$ predicted with the nearest-neighbor model, assuming $\Delta G = 0$ for tandem mismatches (see text). Light bars (DSC) show $\Delta\Delta G$ as measured by DSC.

on the length (n) of contiguous mismatched bases comprising the loop (17). The thermodynamic contribution of a mismatch region (for $n > 1$) is thus given by the sum of the flanking single base pair mismatch parameters and the loop entropy term $\Delta G^\circ_{\text{loop}}$ (n).

For example, consider the duplex sequence containing a mismatch, $S = \text{C}\underline{\text{GTAA}}/\text{GA}\underline{\text{CG}}\text{T}$ that is part of a larger DNA fragment such as the Type 2 or Type 3 mismatch duplexes, where only the sequence region containing the mismatch differs from the sequence of the perfect match. The calculated nearest-neighbor free energy of this segment is given by,

$$\Delta G^\circ(S) = \Delta G^\circ\left(\frac{\text{C}\underline{\text{G}}}{\text{G}\underline{\text{A}}}\right) + \Delta G^\circ_{loop}(3) + \Delta G^\circ\left(\frac{\underline{\text{A}}\text{A}}{\underline{\text{G}}\text{T}}\right),$$

where $\Delta G^\circ(\text{C}\underline{\text{G}}/\text{G}\underline{\text{A}})$ and $\Delta G^\circ(\underline{\text{A}}\text{A}/\underline{\text{G}}\text{T})$ are the nearest-neighbor free energy parameters for a single base pair mismatch bounded by an intact w/c base pair that have been tabulated and reported (17). Following this strategy, the middle term, $\Delta G^\circ(\underline{\text{GTA}}/\underline{\text{ACG}})$, corresponding to the free energy of the $\underline{\text{GT}}/\underline{\text{AC}}$ and $\underline{\text{TA}}/\underline{\text{CG}}$ tandem mismatches, is assigned a value of $\Delta G^\circ_{\text{loop}}(3)$. Note that two duplexes of Type 3 (1.3 and 6.3), contain a tandem mismatch on the end, which is suspected to stabilize the duplex, but thermodynamic parameters are not known for loops on the end of a duplex (17). Thus, for duplexes 1.3 and 6.3, published entropy terms for non-terminal loops with the relevant mismatch lengths were employed.

Clearly, this method does not consider possible sequence dependent thermodynamic contributions of tandem mismatches to DNA duplex stability and the positive sign of the loop free energy term, $\Delta G^\circ_{\text{loop}}(n)$, for $n > 2$ presumes that larger numbers of tandem mismatches are increasingly destabilizing (17). Consequently, this treatment seriously discounts the possibility that tandem mismatches might, depending on the sequence, contribute substantial stability to the duplex, in some cases equal to a significant fraction of the stability of w/c base pairs. Even so, thermodynamic stabilities of short duplex DNAs

containing tandem mismatches are often estimated just as described, ignoring potential sequence dependent contributions to stability possibly inherent in some tandem mismatches.

Using analogous expressions for the specific sequences involved, and the published nearest-neighbor stacking stability parameters and tabulated $\Delta G^\circ_{\text{loop}}(n)$ values, the differences, $\Delta\Delta G$, between the perfect match duplexes (Type 1) and their corresponding mismatch duplexes (Types 2 and 3) were predicted. The predicted differences were then compared directly with the same differences obtained from experimentally measured free energies. Results are summarized in Figure 5, where predicted differences are plotted as histograms (dark bars) along with DSC-measured quantities (light bars). Note that these are plots of $\Delta\Delta G$ for each mismatch compared to the corresponding perfect match duplex. Thus, the greater the $\Delta\Delta G$ value, the lower the predicted (or measured) stability of the mismatch compared to the perfect match duplex. Nearest-neighbor calculations give somewhat different values when compared to the experimentally measured thermodynamic parameters. This probably results from neglecting enthalpic stabilizing contributions in mismatched regions of the duplex. A taller dark (predicted) bar than light (experimental) bar indicates the mismatch duplex is predicted to be more unstable compared to the perfect match, than experimentally observed. The differences between nearest-neighbor predictions and DSC measurements ranged from -0.94 to 10.28 kcal/mol with an average difference of 5.05 kcal/mol for $\Delta\Delta G$(Type 2–Type 1), and from 1.94 to 19.25 kcal/mol with an average difference of 10.19 kcal/mol for $\Delta\Delta G$(Type 3–Type 1).

In many cases, especially for duplexes with several contiguous mismatches (i.e. Type 3), the nearest-neighbor calculations gave significantly different results compared to the experimentally measured thermodynamic parameters. This apparent deficiency of the predictive power of the nearest-neighbor model (in its current rendition) for tandem mismatches or loops, underscores the need to

obtain a more quantitative understanding of the thermodynamic contributions of tandem mismatches to duplex stability. Until such parameter values are evaluated, the general applicability of the nearest-neighbor model in characterizing sequence dependent stabilities of mismatch duplexes should be viewed as somewhat questionable.

Thermodynamic analysis

DSC experiments provided direct measurements of the transition enthalpy (ΔH°) and entropy (ΔS°) for each duplex. From these parameters, ΔG° values were determined and used in comparisons with microarray intensities. The obvious advantage of DSC measurements is that derived thermodynamic parameters are directly evaluated and not dependent on assumptions regarding the nature of the melting transition, i.e. that it occurs in a two-state manner. For this reason, ΔG° values determined from DSC measurements and used in subsequent analysis were deemed more reliable than ΔG° values calculated from the nearest-neighbor model parameters, particularly for duplexes with multiple mismatches. In addition to the observed correlation between ΔG° values measured in solution and microarray hybridization intensities, for both perfect match and duplexes containing tandem mismatches, the evaluated thermodynamic parameters provide a means for gaining additional insight into specific features of sequence dependent stability of tandem mismatches.

Sequences examined in DSC melting experiments were not designed specifically for a comprehensive study of the sequence dependent thermodynamic properties of mismatch duplexes. The distribution of mismatches is hardly heterogeneous and nearly all tandem mismatches examined are of the GA/AG type, which is known to exhibit significant stability (18). Base pairs flanking the mismatches also vary. Although, incomplete in the sense that all sequence possibilities are not included, more subtle features of tandem mismatch stability can be deduced from the limited data that were obtained.

For the sets of duplexes examined, the only difference between Type 1 and Type 2 or Type 1 and Type 3, for a given probe number (1–10), is the introduction of tandem mismatches in various positions. Therefore, thermodynamic effects of substituting w/c base pairs with tandem mismatches in the same duplex sequence can be directly assessed. Relative differences of measured ΔH° and ΔS° values for each probe number, and for each mismatch type (Types 2 and 3) are plotted in Figure 6. As shown, a direct correlation was found (correlation coefficient $R = 0.997$) between the changes in enthalpy, $\Delta\Delta H = \Delta H^\circ(mm) - \Delta H^\circ(pm)$, and entropy, $T\Delta\Delta S = T[\Delta S^\circ(mm) - \Delta S^\circ(pm)]$, relative to the perfect match, regardless of the number, type or mismatch position (at $T = 25^\circ C$). The best linear fit $y = mx + b$ to the data has parameters $m = 0.83$ and $b = -1.22$, and is shown as a dashed line in Figure 6. The line $y = x$ is shown as a dotted line for reference. When comparing mismatched duplexes to the corresponding perfectly matched duplexes, it is clear that the loss of stabilizing interactions is not fully balanced by a gain in degrees of freedom of
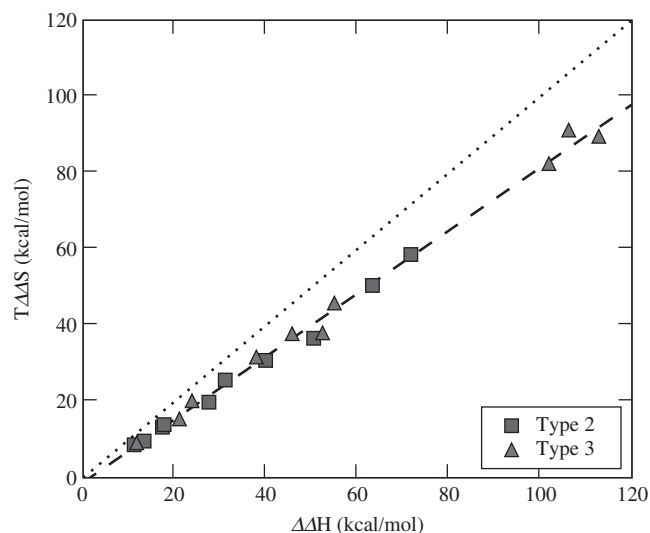

Figure 6. Comparison of change in enthalpy ($\Delta\Delta H$), relative to the perfect match and change in entropy ($T\Delta\Delta S$, $T = 25^\circ C$), relative to the perfect match, as measured by DSC, for both Type 2 (squares) and Type 3 (triangles) duplexes. The dashed line that fits the data has a slope $m = 0.83$ and intercept $b = -1.22$, with a linear correlation coefficient, $R = 0.997$. For reference, the dotted line $y = x$ is also shown.

the system. This comparison for tandem mismatches expands previous results in the literature reporting 'free energy compensation' (cf. 19) for perfect match duplexes or duplexes containing only single base pair mismatches.

Although, changes in ΔH° and ΔS° are closely correlated, a more detailed analysis reveals minor variations in individual parameter values. In Figure 7a, $\Delta\Delta H$ and $T\Delta\Delta S$ are plotted versus the percentage of mismatched bases in each duplex. As expected, the change in both ΔH° and $T\Delta S^\circ$ increase with the relative number of mismatches present in each duplex. The actual difference between $\Delta\Delta H$ and $T\Delta\Delta S$ also increases with the relative number of mismatches, and is plotted versus mismatch percentage in Figure 7b. This, combined with the fact that $\Delta\Delta H > T\Delta\Delta S$ for each duplex, suggests that the destabilizing effects of tandem mismatches is more enthalpic than entropic, as the number of mismatches increases. This observation is consistent with maintenance of duplex structure with the substitution of w/c base pairs by tandem mismatches (20).

Duplexes 6.2, 7.2 and 10.2 all contain a GA/AG mismatch near the middle (Table 1, column 2), while duplexes 1.3 (and 3.3) each contain a GA/AG (AG/GA) mismatch at (or near) the end (Table 1, column 3). For these specific sequences, differences in $\Delta\Delta H$ and $T\Delta\Delta S$ indicate effects of relative position and nearest-neighbor flanking sequences on thermodynamics of duplex formation. The change in entropy and enthalpy for these duplexes (relative to the perfect match) is shown in Figure 8. Note that the increasing trend among Type 2 duplexes ($7.2 < 6.2 < 10.2$) cannot be due to GC content alone, since duplexes 6.2 and 7.2 have similar percentage of GC (23.1 and 26.9%, respectively). However, examination of the base pairs flanking the tandem mismatch
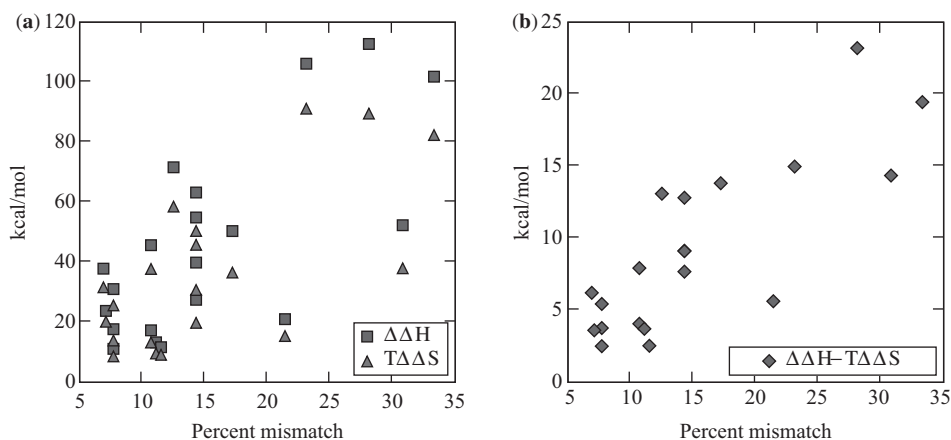
Figure 7. Change in thermodynamic parameters with increased fraction of mismatched bases. (**a**) Change in enthalpy, $\Delta\Delta H$ (squares), and change in entropy, $T\Delta\Delta S$, $T = 25°C$ (triangles), as a function of mismatch percent in the duplex (number of mismatches/duplex length). (**b**) Differences between $\Delta\Delta H$ and $T\Delta\Delta S$ as a function of mismatch percent.
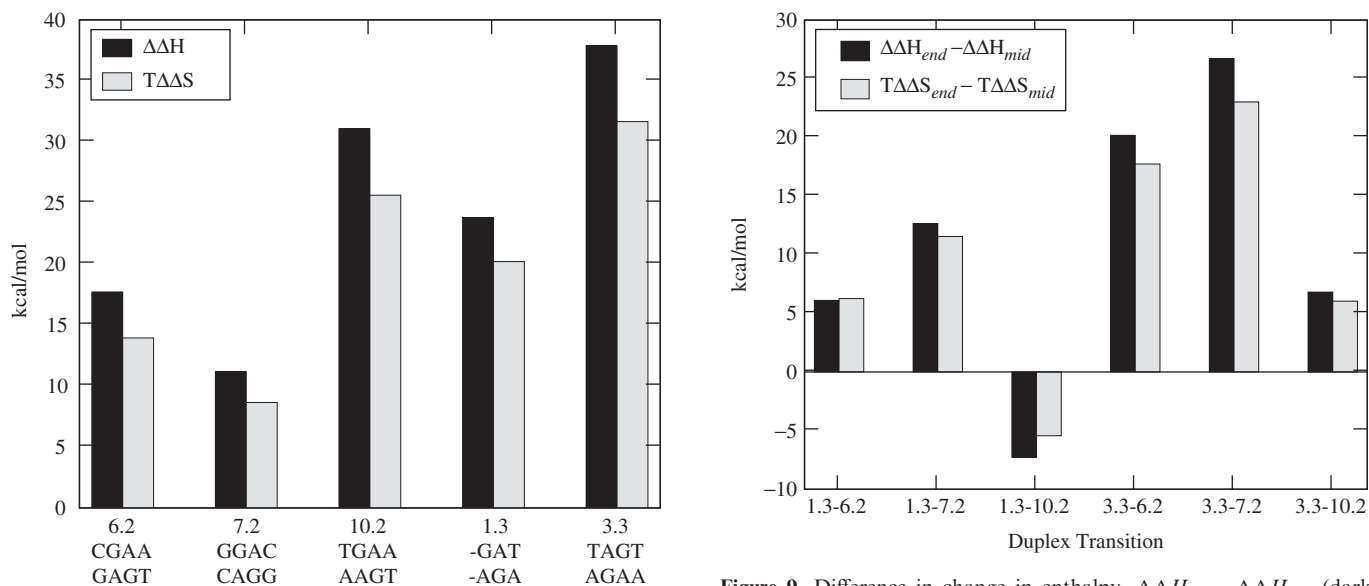

Figure 8. Change in enthalpy, $\Delta\Delta H$ (dark), and change in entropy, $T\Delta\Delta S$, $T = 25°C$ (light), for five specific duplexes. Type 2 duplexes (6.2, 7.2 and 10.2) have a GA/AG mismatch in the middle; Type 3 duplexes (1.3 and 3.3) have a GA/AG mismatch near the end.

Figure 9. Difference in change in enthalpy, $\Delta\Delta H_{end} - \Delta\Delta H_{mid}$ (dark) and change in entropy, $T\Delta\Delta S_{end} - T\Delta\Delta S_{mid}$ (light), for moving a tandem mismatch from the middle to near the end of the duplex, i.e. from Type 2 duplex (mismatch in the middle: 6.2, 7.2 or 10.2) to Type 3 duplex (mismatch near the end: 1.3 or 3.3).

in each case reveals that the flanking base pairs in 7.2 are both G/C, in 6.2 one is a G/C and the other an A/T and in 10.2 both are A/T base pairs. Thus, the relative stability of flanking base pairs accounts for the change in entropy and enthalpy for these three duplexes.

In a similar manner, the change in energies for the two Type 3 duplexes (1.3 and 3.3) can also be correlated with the combined stabilities of the base pairs that border the tandem mismatch (duplex 1.3 has a GA/AG mismatch on the 5′ end, flanked by a 3′T/A base pair; duplex 3.3 has an AG/GA mismatch on the 5′ penultimate end, flanked on both sides by a T/A base pair). Previous results have been reported that implicate the role of flanking base pairs in mismatch stability (21–23). The results shown in Figure 8 support these findings and directly show the influence of

flanking w/c base pairs on the stability of tandem mismatches.

In order to emphasize these results, the differences between $\Delta\Delta H$ (and $T\Delta\Delta S$) for these duplexes are shown in Figure 9. These values represent the effective difference in $\Delta H°$ and $T\Delta S°$ upon replacing a tandem mismatch in the middle of the duplex with one near the end. As expected from the data in Figure 8, the differences in $\Delta\Delta H$ (and $T\Delta\Delta S$) decrease with decreasing stability of the flanking base pairs. The relative cost of replacing a tandem mismatch in the middle with one on the end (6.2, 7.2, 10.2 → 1.3) follows the following order: $\Delta\Delta H(1.3) - \Delta\Delta H(7.2) > \Delta\Delta H(1.3) - \Delta\Delta H(6.2) > \Delta\Delta H(1.3) - \Delta\Delta H(10.2)$ (similarly for $T\Delta\Delta S$). The relative cost of replacing a tandem mismatch in the middle

with one on the penultimate end (6.2, 7.2, 10.2 → 3.3) also follows the same order. In both cases, the relative order of the differences between mismatches in the middle and on the ends is maintained. This observation indicates that the sequence dependent thermodynamics of flanking w/c base pairs have a greater influence than mismatch position on tandem mismatch stability.

Together these results indicate that an accurate characterization of the thermodynamic stability of a GA/AG tandem mismatch requires consideration of at least four base pair units: the tandem mismatch and two flanking (w/c) base pairs. This observation raises several questions. In the quest to evaluate general sets of thermodynamic parameters that enable predictions of thermodynamic stability of duplexes containing tandem mismatches: is the nearest-neighbor model valid or even appropriate? What is the minimum number of base pair units required for specification of general thermodynamic stability of tandem mismatches? To address these questions, investigations are in progress to assess, among other things, general applicability and utility of the nearest-neighbor-model to characterize sequence dependent thermodynamics of tandem mismatches.

CONCLUSION

For 30 short DNA duplexes (26–32 bp), thermodynamic parameters for duplex melting, directly measured in solution by DSC, were compared to microarray intensities measured in 63 DNA multiplex hybridization experiments performed on microarrays. The DNA molecules used in experiments were a collection of perfect match duplexes and duplexes containing tandem mismatches. Overall, a linear relation between duplex free energy, ΔG° and microarray hybridization intensity was found, but exceptions exist that can be attributed to variations in both duplex length and position of tandem mismatches.

Others have reported similar findings of an approximately linear relationship between DNA thermodynamic stability, ΔG° and microarray intensity (3–5). These studies employed a Langmuir binding model to compare microarray hybridization intensity data for short DNAs with sequence dependent thermodynamic stabilities *calculated* using the nearest-neighbor parameters. The design and results of our study contrast these studies in several ways. First, a low-density microarray was purposefully designed so as to precisely direct specific hybridization events, while minimizing cross-hybridization, and avoiding the use of statistical methods inherent in the analysis of very large data sets. Microarray intensity data used in our analysis was subject to minimal statistical manipulation and reported values are actual averages of background corrected signal intensities. A second unique feature of our study bears reiteration. Rather than rely on nearest-neighbor model parameters to calculate thermodynamic transition parameters, these parameters for all duplexes were directly measured in DSC experiments. Others have reported the superior advantages of using DSC for thermodynamic parameter evaluation as opposed to model dependent optical melting curve procedures (24).

Consequently, relationships found between microarray hybridization and thermodynamic stabilities are model-independent, and constitute the first direct comparison of measured ΔG° values and microarray intensities. These data also afford the opportunity to test the accuracy of the nearest-neighbor model in predicting thermodynamic stabilities of tandem mismatch duplexes. Significant discrepancies were observed.

A detailed analysis of the measured thermodynamic parameters for a few specific sequences revealed that the relative effect of a tandem GA/AG mismatch on duplex stability depends explicitly on the flanking w/c base pairs. At least four base pair units must be considered when describing the thermodynamic stability of tandem mismatches. Apparently this is the case regardless of where the tandem mismatch resides in the duplex, i.e. in the middle or on the end.

Finally, this study begins to establish a basis for definition of a quantitative metric relating solution thermodynamic parameters and microarray hybridization intensities. The purpose of establishing such a metric will be to enable evaluation of sequence dependent thermodynamic stability parameters of mismatch duplex complexes in high-throughput fashion on microarrays. Evaluation of these sequence-dependent parameters is also essential for accurate and effective assay design and interpretation. Such parameters will certainly provide a central component in the development of diagnostic technologies for individualized genotyping, diagnosis of genetic disorders and the eventual realization of personalized medicine.

REFERENCES

1. SantaLucia,J.Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci.*, **95**, 1460–1465.
2. Horne,M.T., Fish,D.J. and Benight,A.S. (2006) Statistical thermodynamics and kinetics of DNA multiplex hybridization reactions. *Biophys. J.*, **91**, 4133–4153.
3. Wick,L.M., Rouillard,J.M., Whittam,T.S., Gulari,E., Tiedje,J.M. and Hashsham,S.A. (2006) On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. *Nucleic Acids Res.*, **34**, e26.
4. Carlon,E. and Heim,T. (2006) Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays. *Physica A*, **362**, 433–449.
5. Held,G.A., Grinstein,G. and Tu,Y. (2003) Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl Acad. Sci. USA*, **100**, 7575–7580.
6. Fish,D.J., Horne,M.T., Searles,B.P., Brewood,G.P. and Benight,A.S. (2007) Multiplex SNP discrimination. *Biophys. J.*, **92**, L89–L91.

7. Fotin,A.V., Drobyshev,A.L., Proudnikov,D.Y., Perov,A.N. and Mirzabekov,A.D. (1998) Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Res.*, **26**, 1515–1521.

8. Timofeev,E. and Mirzabekov,A. (2001) Binding specificity and stability of duplexes formed by modified oligonucleotides with a 4096-hexanucleotide microarray. *Nucleic Acids Res.*, **29**, 2626–2634.

9. Naef,F., Lim,D.A., Patil,N. and Magnasco,M. (2002) DNA hybridization to mismatched templates: a chip study. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **65**(4 Pt 1), 040902.

10. Naef,F. and Magnasco,M.O. (2003) Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **68**(1 Pt 1), 011906.

11. Carlon,E., Heim,T., Wolterink,J.K. and Barkema,G.T. (2006) Comment on solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **73**(6 Pt 1), 063901.

12. Zhang,L., Wu,C., Carta,R. and Zhao,H. (2006) Free energy of DNA duplex formation on short oligonucleotide microarrays. *Nucleic Acids Res.*, **35**, e18.

13. Binder,H. and Preibisch,S. (2005) Specific and nonspecific hybridization of oligonucleotide probes on microarrays. *Biophys. J.*, **89**, 337–352.

14. Letowski,J., Brousseau,R. and Masson,L. (2004) Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J. Microbiol. Meth.*, **57**, 269–278.

15. Stillmanm,B.A. and Tonkinson,J.L. (2001) Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal. Biochem.*, **295**, 149–157.

16. Nuwaysir,E.F., Huang,W., Albert,T.J., Singh,J., Nuwaysir,K., Pitas,A., Richmond,T., Gorski,T., Berg,J.P. *et al.* (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.*, **12**, 1749–55.

17. SantaLucia,J.Jr and Hicks,D.J. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.

18. Ebel,S., Lane,A.N. and Brown,T. (1992) Very stable mismatch duplexes: structural and thermodynamic studies on tandem G.A mismatches in DNA. *Biochemistry*, **31**, 12083–12086.

19. Petruska,J. and Goodman,M.F. (1995) Enthalpy-entropy compensation in DNA melting thermodynamics. *J. Biol. Chem.*, **270**, 746–750.

20. Chou,S-H. and Chin,K-H. (2001) Solution structure of a DNA helix incorporating four consecutive non-Watson-Crick base-pairs. *J. Mol. Bio.*, **312**, 769–781.

21. Qu,X., Ren,J., Riccelli,P.V., Benight,A.S. and Chaires,J.B. (2003) Enthalpy/entropy compensation: influence of DNA flanking sequence on the binding of 7-amino actinomycin D to its primary binding site in short DNA duplexes. *Biochemistry*, **42**, 11960–11967.

22. Ke,S. and Wartell,R.M. (1995) Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragment. *Biochemistry*, **34**, 4593–4600.

23. Allawi,H.T. and SantaLucia,J.Jr (1998) Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA. *Biochemistry*, **37**, 2170–2179.

24. Tikhomirova,A., Beletskaya,I.V. and Chalikian,T.V. (2006) Stability of DNA duplexes containing GG, CC, AA, and TT Mismatches. *Biochemistry*, **45**, 10563–10571.