Comment on "Solving the riddle of the bright mismatches: Labeling and effective binding in oligonucleotide arrays"

E. Carlon and T. Heim
Interdisciplinary Research Institute c/o IEMN, Cité Scientifique Boite Postale 60069, F-59652 Villeneuve d'Ascq, France

J. Klein Wolterink and G. T. Barkema
Institute for Theoretical Physics, University of Utrecht, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands
(Received 31 May 2005; published 26 June 2006)

In a recent paper [Phys. Rev. E **68**, 011906 (2003)], Naef and Magnasco suggested that the "bright" mismatches observed in Affymetrix microarray experiments are caused by the fluorescent molecules used to label RNA target sequences, which would impede target-probe hybridization. Their conclusion is based on the observation of "unexpected" asymmetries in the affinities obtained by fitting microarray data from publicly available experiments. We point out here that the observed asymmetry is due to the inequivalence of RNA and DNA, and that the reported affinities are consistent with stacking free energies obtained from melting experiments of *unlabeled* nucleic acids in solution. The conclusion of Naef and Magnasco is therefore based on an unjustified assumption.

DOI: 10.1103/PhysRevE.73.063901 PACS number(s): 87.15.−v, 82.39.Pj

In a recent paper [1] Naef and Magnasco investigated the problem of bright mismatches (MMs) in Affymetrix DNA microarrays. In these arrays a perfect matching (PM) 25 nucleotides probe is accompanied by a MM one in which a single nucleotide in the central position is modified. The MM probes in Affymetrix chips are introduced with the purpose of estimating the contributions from nonspecific hybridization. One of the problems with this approach is that of the so-called "bright" mismatches, for which the fluorescence intensity measured from a MM probe is higher than that from the corresponding PM probe. This is seemingly at odds with basic thermodynamics, as a perfectly matching duplex is more stable than one containing a mismatch. The analysis of experimental data shows that the occurrence of bright mismatches is rather frequent. It was found to occur in 30% of the probes in the Affymetrix Human HGU95a microarray [2].

The authors of Ref. [1] analyze a series of microarray experiments performed by Affymetrix. In this set of experiments the targets are single stranded RNA molecules where some of the pyrimidines (the U and C bases) carry a biotin, while the probes are single stranded DNA molecules. After the hybridization step is terminated the solution containing the nonhybridized target is washed off from the array. Fluorescent labels, attached to streptadivin molecules, are then added. The biotin is a strong binding site for the streptadivin during this staining step. Although Naef and Magnasco use "biotin" and "fluorescent label" as synonyms in Ref. [1], it is important to emphasize that there are no fluorescent labels during the hybridization step in Affymetrix experiments.

In Ref. [1] the brightness of PM probes is fitted using affinities $A_{l,i}$ in which $l=A,C,T$, or G denotes the base type and $i=1,2,\ldots,25$ the position along the probe. The resulting fit values are plotted in their Fig. 3. Notably, there are differences between affinities for A,T and C,G bases, which were interpreted in Ref. [1] as due to the presence of biotin in some bases of the RNA strand. We quote from Ref. [1]: "… An unexpected aspect of the above fits is the asymmetry of A versus T (and G versus C) affinities, which goes against the zeroth order energetic consideration that A-T and T-A bonds (or G-C and C-G) would contribute equally to the binding … The obvious culprits are the fluorescent labels…".

We strongly disagree with this interpretation of the data. There is no "symmetry" between A-T and T-A, since one nucleotide is part of an RNA strand and the other of DNA. To start with, RNA does not have a T (thymine) but a U (uracil), and the authors compare an A-T binding with a U-A binding; and although the usual naming of a C-G and a G-C binding suggests symmetry, also this symmetry is broken by the different backbones of RNA and DNA.

The thermodynamics of RNA/DNA duplexes in solution has been investigated in a series of experiments in which the melting temperatures of short duplexes, of about 20 nucleotides, are measured (see, e.g., [3]). These measurements provide estimates of the differences in enthalpy ΔH and entropy ΔS between a duplex and the two separate strands. As for other types of duplexes, e.g., DNA/DNA and RNA/RNA, it turns out that ΔH and ΔS can be well approximated by sums of sequence-dependent local terms taking into account the contribution of hydrogen bonds and stacking interactions between neighboring bases. As a consequence the free-energy difference, or hybridization free-energy, $\Delta G=\Delta H-T\Delta S$ is also well approximated by a sum of local terms. The latter are given in Table I [3]. Note the asymmetries between the free-energy parameters when interchanging nucleotides between DNA and RNA strands.

In order to compare the hybridization free energies in solution with the affinities reported in Ref. [1], which are single nucleotide dependent, we fix a nucleotide on the probe strand and average the values in Table I over all possible neighbors in the 3′ and 5′ direction. For instance for a T in the DNA strand we define

$$\Delta G_{\mathrm{T}} \equiv \frac{1}{4} \sum_{\gamma\in\{A,T,G,C\}} \left[\Delta G\begin{pmatrix} r\mathrm{A}\gamma' \\ d\mathrm{T}\gamma \end{pmatrix} + \Delta G\begin{pmatrix} r\gamma'\mathrm{A} \\ d\gamma\mathrm{T} \end{pmatrix} \right], \quad (1)$$

where γ' is the nucleotide in the RNA strand complementary to γ. These free energies and the corresponding binding affinities are given in Table II.

 ©2006 The American Physical Society

TABLE I. The stacking free-energy parameters ΔG°_{37} for RNA/DNA hybrids measured in solution at a salt concentration 1 M NaCl and $T=37^\circ$ [3]. The upper strand is RNA (with orientation $5'-3'$) and lower strand DNA (orientation $3'-5'$).

Sequence	$-\Delta G_{37}$ (kcal/mol)	Sequence	$-\Delta G_{37}$ (kcal/mol)
rAA dTT	1.0	rAC dTG	2.1
rAG dTC	1.8	rAU dTA	0.9
rCA dGT	0.9	rCC dGG	2.1
rCG dGC	1.7	rCU dGA	0.9
rGA dCT	1.3	rGC dCG	2.7
rGG dCC	2.9	rGU dCA	1.1
rUA dAT	0.6	rUC dAG	1.5
rUG dAC	1.6	rUU dAA	0.2

Note that Eq. (1) gives lower affinities for A compared to T and for G compared to C, in qualitative agreement with the data of Ref. [1] in Table II. While in Ref. [1] these differences were argued to prove that biotin affects the binding, our analysis clearly shows that these differences (or asymmetries as referred to in [1]) are intrinsic properties of unbiotinylated RNA/DNA duplexes in solution.

It is not surprising that the effective affinities measured in Ref. [1] are smaller than the binding free energies obtained from Eq. (1). The affinities of Ref. [1] are obtained by fitting the measured fluorescent signals of the microarray to a Langmuir model (Eq. (1) in [1]). The fluorescence measured in the microarray experiment is not solely determined by the binding free energy between an isolated probe and a specific target, but it is also influenced by many other effects such as polydispersity in probe and target lengths, secondary structure formation in probes and targets, and hybridization between targets in solution. We thus do not expect that the affinities of Ref. [1] should agree quantitatively with the binding free energies in solution. As the neglected processes compete with the hybridization of a probe with a complementary target, it is to be expected that the difference in effective affinities of Ref. [1] are lower than their solution counterparts.

TABLE II. Column 2: Single nucleotide free energies obtained from Eq. (1) expressed in kcal/mole. Column 3: The same free energies to which the average value is subtracted. Column 4: The log 10 affinities derived from the data of column 3 by $A_\gamma=(\langle\Delta G\rangle-\Delta G_\gamma)/(RT\ln 10)$, with $RT=0.63$ kcal/mole. Column 5: Effective affinities for the middle bases as given in Fig. 3 of Ref. [1].

γ	$-\Delta G_\gamma$	$\langle\Delta G\rangle-\Delta G_\gamma$	A_γ	Ref. [1]
C	4.00	1.09	0.75	0.20
G	3.50	0.59	0.40	0.02
T	2.40	−0.51	−0.35	−0.01
A	1.75	−1.16	−0.80	−0.20

Certainly, Ref. [1] does not show that fluorescent labels (or, to be precise, the biotin linker) interfere with binding, or are the cause of bright mismatches.

We acknowledge financial support from the Van Gogh Programme d'Actions Intégrées (PAI) 08505PB of the French Ministry of Foreign Affairs and NWO Grant No. 62403735.

[1] F. Naef and M. O. Magnasco, Phys. Rev. E **68**, 011906 (2003).
[2] F. Naef, D. A. Lim, N. Patil, and M. Magnasco, Phys. Rev. E **65**, 040902(R) (2002).
[3] N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama and M. Sasaki, Biochemistry **34**, 11211 (1995).