Competitive hybridization models

Vera Cherepinsky,[1,*] Ghazala Hashmi,[2] and Bud Mishra[3,4]
[1]*Department of Mathematics and Computer Science, Fairfield University, Fairfield, Connecticut 06824, USA*
[2]*BioArray Solutions, Ltd., Warren, New Jersey 07509, USA*
[3]*Courant Institute of Mathematical Sciences, NYU, New York, New York 10012, USA*
[4]*NYU School of Medicine, New York, New York 10016, USA*
(Received 14 April 2009; revised manuscript received 13 July 2010; published 9 November 2010)

Microarray technology, in its simplest form, allows one to gather abundance data for target DNA molecules, associated with genomes or gene-expressions, and relies on hybridizing the target to many short probe oligonucleotides arrayed on a surface. While for such multiplexed reactions conditions are optimized to make the most of each individual probe-target interaction, subsequent analysis of these experiments is based on the implicit assumption that a given experiment yields the same result regardless of whether it was conducted in isolation or in parallel with many others. It has been discussed in the literature that this assumption is frequently false, and its validity depends on the types of probes and their interactions with each other. We present a detailed physical model of hybridization as a means of understanding probe interactions in a multiplexed reaction. Ultimately, the model can be derived from a system of ordinary differential equations (ODE's) describing kinetic mass action with conservation-of-mass equations completing the system. We examine pairwise probe interactions in detail and present a model of "competition" between the probes for the target—especially, when the target is effectively in short supply. These effects are shown to be predictable from the affinity constants for each of the four probe sequences involved, namely, the match and mismatch sequences for both probes. These affinity constants are calculated from the thermodynamic parameters such as the free energy of hybridization, which are in turn computed according to the nearest neighbor (NN) model for each probe and target sequence. Simulations based on the competitive hybridization model explain the observed variability in the signal of a given probe when measured in parallel with different groupings of other probes or individually. The results of the simulations can be used for experiment design and pooling strategies, based on which probes have been shown to have a strong effect on each other's signal in the *in silico* experiment. These results are aimed at better design of multiplexed reactions on arrays used in genotyping (e.g., HLA typing, SNP, or CNV detection, etc.) and mutation analysis (e.g., cystic fibrosis, cancer, autism, etc.).

DOI: 10.1103/PhysRevE.82.051914 PACS number(s): 87.15.R−, 87.15.Qt, 87.90.+y

I. BACKGROUND

Recognition of a target nucleic acid and analysis of its composition can be carried out by hybridization based on complementary base pairing with a suitably designed much shorter probe oligonucleotide. In essence, the presence of one of several possible known "messages" in the target is detected by checking if a population of identical targets in solution binds, under suitable thermodynamic conditions, to the probe molecules encoding a sequence, designed to be complementary to a message. Furthermore, a more precise quantitative answer can be obtained if other "control" probes are also mixed in with the designed probe in a well-controlled proportion and sharing similar thermodynamic properties.

Many recent advances in genome analysis, detection of polymorphisms, molecular karyotyping, and gene-expression analysis have relied on our abilities to conduct high-throughput multiplexed hybridization involving thousands or millions of probes on a surface (e.g., gene-chips and microarrays) and then, interpret the resulting assay readings (see [1] and references therein). Thus, the reliability of the final computational interpretation of the data depends on understanding the errors due to unintended interactions among targets and probes, as probes and targets are multiplexed.

In particular, we focus on a mathematical analysis of "competitive hybridization," a phenomenon that has been observed in experimental data [2–4], but not adequately explained. In the following simple example of this phenomenon, a target consisting of two possibly distinct messages m_A and m_B can be characterized by separately hybridizing the target with either a mixture of specific probes pm_A and control probes mm_A or a mixture of specific probes pm_B and control probes mm_B, respectively. In either case the ratio of specific signal to the control signal, obtained from each separate experiment, indicates how often either message is present. On the other hand, contrary to one's expectations, if the two messages were queried by ratios of the respective signals in a multiplexed experiment consisting of all four probes pm_A, mm_A, pm_B, and mm_B, one finds these ratios to differ from their values in the earlier experiments and by amounts that cannot simply be explained by the statistical noise. In particular, if one of the ratio values decreases severely, the resulting false negative errors will yield a catastrophic failure of the entire multiplexed assay. Clearly, the situation worsens precipitously as the number of multiplexed probes is increased to any realistic number. Furthermore, it becomes important to ask whether such a multiplexed assay

*Corresponding author; Present address: MACS Department, Fairfield University, 1073 North Benson Road, Fairfield, CT 06824; vcherepinsky@fairfield.edu

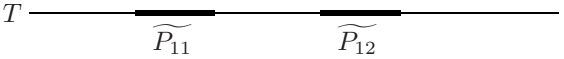
can be rescued by judicious choice of the selected probes and the thermodynamic parameters.

This paper is organized as follows: Sec. II contains the experimental setup of the system being modeled; Sec. III includes the dynamics, yielding the systems of algebraic equations that hold at equilibrium; Sec. IV introduces a change of variables for notational convenience; Sec. V contains the system reduction; Sec. VI presents two additional models for comparison; Sec. VII contains a discussion of thermodynamic parameters; Sec. VIII discusses the observed competition among probes; Sec. IX contains experimental validation of the models; and Sec. X gives the conclusion. In addition, Supplementary Materials are provided [5], containing details of derivations and a glossary of biological terms used in this paper.

II. SETUP

More specifically, we consider the following experimental setup (describing the BeadChip microarray format, see [3,6,7]). Probes are bound to encoded microparticles ("beads") whose sizes are relatively large compared to the size of the probes. We assume that there are thousands of copies of the same probe attached to a single bead, and that the beads are spaced on a planar surface far enough apart in order to ensure that a single target strand may only hybridize to probes on a single bead. Thus, for all intents and purposes, this assumption implies that the only possible complexes involve one target and one probe. The targets are obtained from a longer DNA, by PCR amplification with two primers to select clones of a region that are subjected to further characterization. (Note that, while this setup motivated our investigation, the theory is not limited to this specific microarray format and will be applicable to other schemes with suitable parameter changes. See Supplementary Materials [[5], Sec. X] for further discussion.)

Let T be a target with a single region perfectly complementary to probe P_{11} and another region perfectly complementary to probe P_{12}.



Let P_{01} differ from P_{11} in one base (i.e., the Hamming distance between P_{01} and P_{11} equals to 1, $H(P_{01},P_{11})=1$).

If P_{11} and P_{01} are the only probes present, we can expect that when we compare the concentration of the P_{11} probes bound to T (denoted $[TP_{11}]$) to the concentration of the P_{01} probes bound to T (denoted $[TP_{01}]$), the resulting ratio will be large, i.e.,

$$\frac{[TP_{11}]}{[TP_{01}]} \gg 1,$$

since their free energies are chosen to satisfy $\Delta G(P_{01}) < \Delta G(P_{11})$. P_{01} clearly "competes" with P_{11} for the target T.

Consider yet another probe, P_{02}, that differs from P_{11} in one base as well $[H(P_{11},P_{02})=1]$, but at a location different from the one in P_{01} [so that $H(P_{01},P_{02})=2$]. Then P_{02} also competes with P_{11}, but not as much with P_{01}, since $H(P_{01},P_{02})=2$. Thus, in the presence of P_{02}, we expect $[TP_{11}]/[TP_{01}]$ to decrease, since $[TP_{01}]$ does not decrease much, but $[TP_{11}]$ does. However, in the presence of all four probes P_{11}, P_{01}, P_{12}, and P_{02}, the analysis of the resulting "mutual competitions" poses a nontrivial problem.

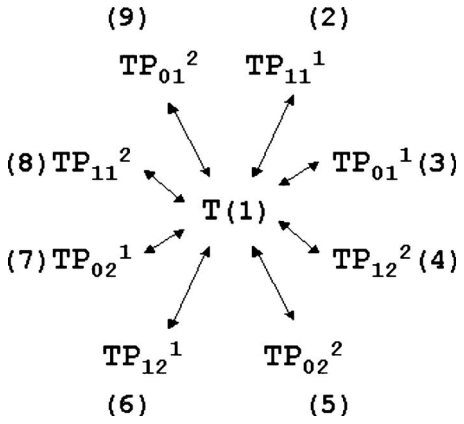


FIG. 1. State transition diagram—full model.

III. DYNAMICS

Here, we present a mathematical model to analyze the dynamics involved in the setup described in the previous section. In particular, we assume that the steric effects prevent multiple probes from hybridizing to a single target strand (as probes are bound to large beads). We begin by considering the situation where all four probes P_{11}, P_{01}, P_{12}, and P_{02} are present (the full model).

A. Full model

We may observe a target strand T in one of the following nine *possible states* (see Fig. 1),

(1) T
(Target is unbound.)
(2) TP_{11}^{1}, (3) TP_{01}^{1}, (4) TP_{12}^{2}, (5) TP_{02}^{2}
(Target is bound by "specific" hybridization.)
(8) TP_{11}^{2}, (9) TP_{01}^{2}, (6) TP_{12}^{1}, (7) TP_{02}^{1}
(Target is bound by "nonspecific" hybridization.)

Bound target states are represented using form TP_{ij}^{k}, where $j \in \{1,2\}$ is the index of the probe, i is the match/mismatch flag

$$i = \begin{cases} 1 & \text{for matched probe,} \\ 0 & \text{for mismatch probe,} \end{cases}$$

and $k \in \{1,2\}$ is the binding site. States within each category are numbered "left-to-right" with respect to the location on the target. We use the term specific hybridization to denote probe j hybridizing to its assigned location on the target (i.e., binding site $k=j$), and nonspecific hybridization to denote probe j hybridizing to any other location on the target (i.e., binding site $k \neq j$).

The set of reversible reactions operating between unbound and bound states in Fig. 1 can be written as shown below, where the forward and backward reaction rates are indicated with $k_{i,j}$ and $k_{j,i}$, respectively. While the reaction

rates themselves are difficult to compute, the ratios $K_i^j = k_{i,j}/k_{j,i}$, also known as affinity constants, may be computed from purely thermodynamic considerations, and are sufficient for equilibrium analysis.

$$T + P_{11} \underset{k_{2,1}}{\overset{k_{1,2}}{\rightleftarrows}} TP_{11}^1 \qquad T + P_{12} \underset{k_{6,1}}{\overset{k_{1,6}}{\rightleftarrows}} TP_{12}^1$$

$$T + P_{01} \underset{k_{3,1}}{\overset{k_{1,3}}{\rightleftarrows}} TP_{01}^1 \qquad T + P_{02} \underset{k_{7,1}}{\overset{k_{1,7}}{\rightleftarrows}} TP_{02}^1$$

$$T + P_{12} \underset{k_{4,1}}{\overset{k_{1,4}}{\rightleftarrows}} TP_{12}^2 \qquad T + P_{11} \underset{k_{8,1}}{\overset{k_{1,8}}{\rightleftarrows}} TP_{11}^2$$

$$T + P_{02} \underset{k_{5,1}}{\overset{k_{1,5}}{\rightleftarrows}} TP_{02}^2 \qquad T + P_{01} \underset{k_{9,1}}{\overset{k_{1,9}}{\rightleftarrows}} TP_{01}^2$$

We are interested in behavior at *equilibrium*, and thus perform a steady-state analysis. (The underlying structure of the reaction network theory and other possible stability states admitted by such dynamical systems have been studied extensively by Feinberg and others; see, e.g., [8–12]). Equations for laws of mass action corresponding to these reactions are written concisely as

$$k_{1,2}[T][P_{11}] = k_{2,1}[TP_{11}^1]$$

$$\Rightarrow K_1^2 \equiv \frac{k_{1,2}}{k_{2,1}} = \frac{[TP_{11}^1]}{[T][P_{11}]} \tag{1}$$

for the first reaction and similarly for the remaining reactions,

$$K_1^3 = \frac{[TP_{01}^1]}{[T][P_{01}]}, \tag{2}$$

$$K_1^4 = \frac{[TP_{12}^2]}{[T][P_{12}]}, \tag{3}$$

$$K_1^5 = \frac{[TP_{02}^2]}{[T][P_{02}]}, \tag{4}$$

$$K_1^6 = \frac{[TP_{12}^1]}{[T][P_{12}]}, \tag{5}$$

$$K_1^7 = \frac{[TP_{02}^1]}{[T][P_{02}]}, \tag{6}$$

$$K_1^8 = \frac{[TP_{11}^2]}{[T][P_{11}]}, \tag{7}$$

$$K_1^9 = \frac{[TP_{01}^2]}{[T][P_{01}]}. \tag{8}$$

The full dynamical system is presented in Supplementary Materials [[5], Sec. I], together with the derivation of how it leads, at equilibrium, to the system of Eqs. (1)–(8).

The affinity constants K_1^j for $j \in \{2, \ldots, 9\}$, appearing in Eqs. (1)–(8), are computed from probe sequence data according to Eq. (54), as shown in Sec. VII.

Furthermore, the following conservation rules must hold:

$$[P_{11}]_0 = [P_{11}] + [TP_{11}^1] + [TP_{11}^2], \tag{9}$$

$$[P_{01}]_0 = [P_{01}] + [TP_{01}^1] + [TP_{01}^2], \tag{10}$$

$$[P_{12}]_0 = [P_{12}] + [TP_{12}^1] + [TP_{12}^2], \tag{11}$$

$$[P_{02}]_0 = [P_{02}] + [TP_{02}^1] + [TP_{02}^2], \tag{12}$$

$$[T]_0 = [T] + [TP_{11}^1] + [TP_{01}^1] + [TP_{12}^2] + [TP_{02}^2] + [TP_{11}^2] + [TP_{01}^2] + [TP_{12}^1] + [TP_{02}^1]. \tag{13}$$

Note that in Eqs. (9)–(13), for each species X, $[X]_0$ is a free parameter that denotes the initial concentration, and $[X]$ denotes the equilibrium concentration.

Consider the system of Eqs. (1)–(8) and conservation rules [Eqs. (9)–(13)]. It consists of

(i) 13 polynomial equations (some quadratic, others linear) in

(ii) 13 unknowns: $[T]$, $[TP_{11}^1]$, $[TP_{01}^1]$, $[TP_{12}^2]$, $[TP_{02}^2]$, $[TP_{12}^1]$, $[TP_{02}^1]$, $[TP_{11}^2]$, $[TP_{01}^2]$, $[P_{11}]$, $[P_{01}]$, $[P_{12}]$, and $[P_{02}]$, with

(iii) five free parameters: $[P_{11}]_0$, $[P_{01}]_0$, $[P_{12}]_0$, $[P_{02}]_0$, and $[T]_0$.

This algebraic system, when solved, yields the equilibrium concentrations. From these computed concentrations, we evaluate the "match-to-mismatch ratio" (or the "discrimination signal") for each probe,

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{full}} = \left(\frac{[TP_{11}^1] + [TP_{11}^2]}{[TP_{01}^1] + [TP_{01}^2]}\right)_{\text{full model}}$$

and

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\text{full}} = \left(\frac{[TP_{12}^2] + [TP_{12}^1]}{[TP_{02}^2] + [TP_{02}^1]}\right)_{\text{full model}}.$$

In order to examine the effects of competition between probes P_{11} and P_{12} on the signals for each of them, we compare this situation with the one where only P_{11} and P_{01} are present without P_{12} or P_{02}, and vice versa. In the rest of the paper, we refer to the model introduced in this section as the *full model* and compare its performance with two partial models, one consisting of P_{11}, P_{01}, and T only (referred to as *model I*) and the other consisting of P_{12}, P_{02}, and T only (referred to as *model II*).

B. Partial model—model I

This model consists of two probes P_{11}, P_{01}, and the target T only. We proceed as before by setting up and solving an algebraic system of equations to evaluate the match-to-mismatch ratio for probe with index $j=1$ (in the absence of probe with index $j=2$),

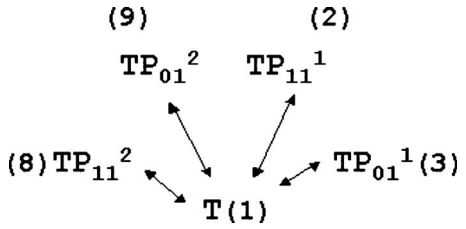


FIG. 2. State transition diagram—model I.

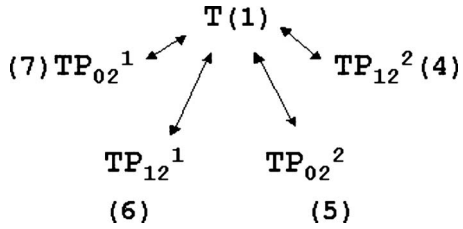


FIG. 3. State transition diagram—model II.

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\mathrm{I}} = \left(\frac{[TP_{11}^1]+[TP_{11}^2]}{[TP_{01}^1]+[TP_{01}^2]}\right)_{\mathrm{I}}.$$

1. Possible states

Here, only five states (shown in Fig. 2) are possible for the target T:

(1) T

(Target is unbound.)

(2) TP_{11}^1, (3) TP_{01}^1

(Target is bound by specific hybridization.)

(8) TP_{11}^2, (9) TP_{01}^2

(Target is bound by nonspecific hybridization.)

The set of reversible reactions operating between unbound and bound states in Fig. 2 can be written as shown below.

$$T + P_{11} \underset{k_{2,1}}{\overset{k_{1,2}}{\rightleftarrows}} TP_{11}^1 \qquad T + P_{11} \underset{k_{8,1}}{\overset{k_{1,8}}{\rightleftarrows}} TP_{11}^2$$

$$T + P_{01} \underset{k_{3,1}}{\overset{k_{1,3}}{\rightleftarrows}} TP_{01}^1 \qquad T + P_{01} \underset{k_{9,1}}{\overset{k_{1,9}}{\rightleftarrows}} TP_{01}^2$$

2. Equilibrium equations

At equilibrium (see Supplementary Materials [[5], Sec. II] for details), these reactions are characterized by mass-action Eqs. (1), (2), (7), and (8). In addition, we have conservation rules [Eqs. (9) and (10)] for $[P_{11}]$ and $[P_{01}]$, respectively, and a new conservation rule for $[T]$,

$$[T]_0 = [T] + [TP_{11}^1] + [TP_{01}^1] + [TP_{11}^2] + [TP_{01}^2]. \quad (14)$$

Thus, in this case, the algebraic system consists of

(i) seven polynomial equations: Eqs. (1), (2), (7)–(10), and (14), in

(ii) seven unknowns: $[T]$, $[TP_{11}^1]$, $[TP_{11}^2]$, $[TP_{01}^1]$, $[TP_{01}^2]$, $[P_{11}]$, and $[P_{01}]$, with

(iii) three free parameters: $[P_{11}]_0$, $[P_{01}]_0$, and $[T]_0$.

Note that, for comparison with the full model, the free parameters need to be scaled to retain the same initial target-to-probe ratio.

C. Partial model—model II

This model consists of two probes P_{12}, P_{02}, and the target T only. We proceed as before by solving an appropriate algebraic system of equations to evaluate the match-to-mismatch ratio for probe with index $j=2$ (in the absence of probe with index $j=1$),

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\mathrm{II}} = \left(\frac{[TP_{12}^2]+[TP_{12}^1]}{[TP_{02}^2]+[TP_{02}^1]}\right)_{\mathrm{II}}.$$

1. Possible states

Here, as well, only five states (shown in Fig. 3) are possible for the target T:

(1) T

(Target is unbound.)

(4) TP_{12}^2, (5) TP_{02}^2

(Target is bound by specific hybridization.)

(6) TP_{12}^1, (7) TP_{02}^1

(Target is bound by nonspecific hybridization.)

The set of reversible reactions operating between unbound and bound states in Fig. 3 can be written as shown below.

$$T + P_{12} \underset{k_{4,1}}{\overset{k_{1,4}}{\rightleftarrows}} TP_{12}^2 \qquad T + P_{12} \underset{k_{6,1}}{\overset{k_{1,6}}{\rightleftarrows}} TP_{12}^1$$

$$T + P_{02} \underset{k_{5,1}}{\overset{k_{1,5}}{\rightleftarrows}} TP_{02}^2 \qquad T + P_{02} \underset{k_{7,1}}{\overset{k_{1,7}}{\rightleftarrows}} TP_{02}^1$$

2. Equilibrium equations

At equilibrium (see Supplementary Materials [[5], Sec. III] for details), these reactions are characterized by mass-action Eqs. (3)–(6). In addition, we have conservation rules [Eqs. (11) and (12)] for $[P_{12}]$ and $[P_{02}]$, respectively, and a new conservation rule for $[T]$,

$$[T]_0 = [T] + [TP_{12}^2] + [TP_{02}^2] + [TP_{12}^1] + [TP_{02}^1]. \quad (15)$$

Thus, in this case, the algebraic system consists of

(i) seven equations: Eqs. (3)–(6), (11), (12), and (15), in

(ii) seven unknowns: $[T]$, $[TP_{12}^2]$, $[TP_{12}^1]$, $[TP_{02}^2]$, $[TP_{02}^1]$, $[P_{12}]$, and $[P_{02}]$, with

(iii) three free parameters: $[P_{12}]_0$, $[P_{02}]_0$, and $[T]_0$.

As above (Sec. III B), the parameters need to be scaled.

In practice, once the *exact nucleotide sequences* of T, P_{11}, P_{01}, P_{12}, and P_{02} are determined from the needs of the biological assay, we can compute K_1^j explicitly using Eq. (54), and then solve for the unknowns in all three setups:

(i) full model,

(ii) model I, and

(iii) model II.

The computed target and probe concentrations can be used to evaluate the match-to-mismatch ratios (discrimination signals) for each probe under full and partial models. With these ratio values, we are ready to compare the models in order to discern the effects of competition. To see what effect the presence of probe 2 plays on the signal of probe 1, we compare the discrimination signal of probe 1 under the full model to the discrimination signal of probe 1 under partial model, model I,

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{full}} \text{ vs } \left(\frac{TP_{11}}{TP_{01}}\right)_{\text{I}}.$$

Similarly, to discern the effect of probe 1 on the signal of probe 2, we compare

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\text{full}} \text{ vs } \left(\frac{TP_{12}}{TP_{02}}\right)_{\text{II}}.$$

IV. CHANGE OF VARIABLES

As stated above, in order to obtain the discrimination signals for the two probes under full and partial models, we need to solve the algebraic systems of equations for the equilibrium target and probe concentrations. This is facilitated by renaming the variables and constant parameters in each system of equations. It turns out that each system reduces to a single equation in one unknown (the concentration of free target, $[T]$). The new notation makes this easier to see.

Starting with the full model, we rename the equilibrium target concentrations as variables X_n, with the subscript n corresponding to target state; further, we rename free probe concentrations as variables Y_m,

$$X_1 = [T]$$

$$X_2 = [TP_{11}^1] \qquad X_6 = [TP_{12}^1] \qquad Y_1 = [P_{11}]$$

$$X_3 = [TP_{01}^1] \qquad X_7 = [TP_{02}^1] \qquad Y_2 = [P_{01}]$$

$$X_4 = [TP_{12}^2] \qquad X_8 = [TP_{11}^2] \qquad Y_3 = [P_{12}]$$

$$X_5 = [TP_{02}^2] \qquad X_9 = [TP_{01}^2] \qquad Y_4 = [P_{02}]$$

For convenience, we also relabel initial probe and target concentrations,

$$a_0 = Y_1^0 = [P_{11}]_0, \quad b_0 = Y_2^0 = [P_{01}]_0,$$

$$c_0 = Y_3^0 = [P_{12}]_0, \quad d_0 = Y_4^0 = [P_{02}]_0,$$

$$e_0 = X_1^0 = [T]_0.$$

The other constant parameters (the affinity constants) remain in their symbolic form: K_1^2, K_1^3, K_1^4, K_1^5, K_1^6, K_1^7, K_1^8, K_1^9.

We now consider the systems of equations under the three models.

A. Full model

Equations (1)–(13) are rewritten in terms of $\{X_i, Y_j\}$ as follows (see Supplementary Materials [[5], Sec. VI] for the derivation).

$$\begin{aligned}
X_2 &= K_1^2 X_1 Y_1 \\
X_3 &= K_1^3 X_1 Y_2 \\
X_4 &= K_1^4 X_1 Y_3 \\
X_5 &= K_1^5 X_1 Y_4 \\
X_6 &= K_1^6 X_1 Y_3 \\
X_7 &= K_1^7 X_1 Y_4 \\
X_8 &= K_1^8 X_1 Y_1 \\
X_9 &= K_1^9 X_1 Y_2 \\
a_0 &= X_2 + X_8 + Y_1 \\
b_0 &= X_3 + X_9 + Y_2 \\
c_0 &= X_4 + X_6 + Y_3 \\
d_0 &= X_5 + X_7 + Y_4 \\
e_0 &= X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9.
\end{aligned} \tag{16}$$

B. Model I

Under the new notation, the system of algebraic equations under model I involves variables X_1, X_2, X_3, X_8, X_9, Y_1, and Y_2, and constant parameters K_1^2, K_1^3, K_1^8, K_1^9, a_0, b_0, and e_0. Its Eqs. (1), (2), (7)–(10), and (14) become

$$\begin{aligned}
X_2 &= K_1^2 X_1 Y_1 \\
X_3 &= K_1^3 X_1 Y_2 \\
X_8 &= K_1^8 X_1 Y_1 \\
X_9 &= K_1^9 X_1 Y_2 \\
a_0 &= X_2 + X_8 + Y_1 \\
b_0 &= X_3 + X_9 + Y_2 \\
e_0 &= X_1 + X_2 + X_3 + X_8 + X_9.
\end{aligned} \tag{17}$$

C. Model II

Under the new notation, the system of algebraic equations under model II involves variables X_1, X_4, X_5, X_6, X_7, Y_3, and Y_4, and constant parameters K_1^4, K_1^5, K_1^6, K_1^7, c_0, d_0, and e_0. Its Eqs. (3)–(6), (11), (12), and (15) become

$$\begin{aligned}
X_4 &= K_1^4 X_1 Y_3 \\
X_5 &= K_1^5 X_1 Y_4 \\
X_6 &= K_1^6 X_1 Y_3 \\
X_7 &= K_1^7 X_1 Y_4 \\
c_0 &= X_4 + X_6 + Y_3 \\
d_0 &= X_5 + X_7 + Y_4 \\
e_0 &= X_1 + X_4 + X_5 + X_6 + X_7.
\end{aligned} \tag{18}$$

Note that with the exception of the conservation rules for $[T]$ [i.e., the last equations in Eqs. (16)–(18)] under the different models, we have

$$(16) = (17) \cup (18).$$

V. SYSTEM REDUCTION

After algebraic manipulation, each of the systems in Eqs. (16)–(18) reduces to a single polynomial equation in X_1.

(This result can also be obtained using Gröbner bases (see, e.g., [13] and [[14], Ch. 4]), although this was not the method used by the authors.) The values of the remaining unknowns are given symbolically in terms of X_1 and the constant parameters, as shown below (see Supplementary Materials [[5], Sec. VII] for the details of the derivation). For clarity, we introduce the following short-hand notation for the frequently-appearing sums of affinity constants,

$$s_{28} \equiv K_1^2 + K_1^8 \quad s_{39} \equiv K_1^3 + K_1^9$$

$$s_{46} \equiv K_1^4 + K_1^6 \quad s_{57} \equiv K_1^5 + K_1^7.$$

A. Full model

Let $z \equiv X_1$. The system of equations in Eq. (16) reduces to the fifth degree polynomial equation

$$e_0 = z\left[1 + a_0\frac{s_{28}}{1+s_{28}z} + b_0\frac{s_{39}}{1+s_{39}z} + c_0\frac{s_{46}}{1+s_{46}z} + d_0\frac{s_{57}}{1+s_{57}z}\right]. \tag{19}$$

The remaining unknowns are then given in terms of X_1,

$$\begin{aligned} X_2 &= \frac{a_0K_1^2X_1}{1+s_{28}X_1} \quad X_8 = \frac{a_0K_1^8X_1}{1+s_{28}X_1} \quad Y_1 = \frac{a_0}{1+s_{28}X_1} \\ X_3 &= \frac{b_0K_1^3X_1}{1+s_{39}X_1} \quad X_9 = \frac{b_0K_1^9X_1}{1+s_{39}X_1} \quad Y_2 = \frac{b_0}{1+s_{39}X_1}, \end{aligned} \tag{20}$$

$$\begin{aligned} X_4 &= \frac{c_0K_1^4X_1}{1+s_{46}X_1} \quad X_6 = \frac{c_0K_1^6X_1}{1+s_{46}X_1} \quad Y_3 = \frac{c_0}{1+s_{46}X_1} \\ X_5 &= \frac{d_0K_1^5X_1}{1+s_{57}X_1} \quad X_7 = \frac{d_0K_1^7X_1}{1+s_{57}X_1} \quad Y_4 = \frac{d_0}{1+s_{57}X_1}. \end{aligned} \tag{21}$$

Since roots of the fifth degree polynomial in Eq. (19) cannot be expressed symbolically in a closed form, we must resort to a purely numerical approach. This yields the desired match-to-mismatch ratios (or discrimination signals) for probe 1,

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{full}} = \frac{a_0}{b_0}\frac{s_{28}}{s_{39}}\frac{1+s_{39}z}{1+s_{28}z} \tag{22}$$

and for probe 2,

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\text{full}} = \frac{c_0}{d_0}\frac{s_{46}}{s_{57}}\frac{1+s_{57}z}{1+s_{46}z} \tag{23}$$

where z solves Eq. (19).

B. Model I

Let $x \equiv X_1$. The system of equations in Eq. (17) reduces to the cubic polynomial equation

$$e_0 = x\left[1 + a_0\frac{s_{28}}{1+s_{28}x} + b_0\frac{s_{39}}{1+s_{39}x}\right]. \tag{24}$$

Note that we are intentionally using a different symbol for X_1 here as the equation to be solved differs from Eq. (19). The remaining unknowns under model I, namely, X_2, X_3, X_8, X_9, Y_1, and Y_2, are then given symbolically in terms of X_1 and the constant parameters using Eqs. (20). The discrimination signal for probe 1 under model I is given by

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{I}} = \frac{a_0}{b_0}\frac{s_{28}}{s_{39}}\frac{1+s_{39}x}{1+s_{28}x}, \tag{25}$$

where x solves Eq. (24).

C. Model II

Let $y \equiv X_1$. The system of equations in Eq. (18) reduces to the cubic polynomial equation

$$e_0 = y\left[1 + c_0\frac{s_{46}}{1+s_{46}y} + d_0\frac{s_{57}}{1+s_{57}y}\right]. \tag{26}$$

Note again that we are using a different symbol for X_1 here to avoid confusion with the variables used in Eqs. (19) and (24). The remaining unknowns under model II, namely X_4, X_5, X_6, X_7, Y_3, and Y_4, are then given symbolically in terms of X_1 and the constant parameters using Eqs. (21). The discrimination signal for probe 2 under model II is given by

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\text{II}} = \frac{c_0}{d_0}\frac{s_{46}}{s_{57}}\frac{1+s_{57}y}{1+s_{46}y}, \tag{27}$$

where y solves Eq. (26).

It is precisely the observed differences in the discrimination signal for probe 1 under model I [Eq. (25)] and the discrimination signal for probe 1 under full model, in the presence of probe 2 [Eq. (22)], as well as the corresponding differences in probe 2 signals under model II [given by Eq. (27)] and under full model [Eq. (23)], that constitute the phenomenon of competition among probes that we wish to examine.

VI. ADDITIONAL MODELS

Next, for the purpose of comparison, we consider two additional models: *simple model*, where the target has exactly one region for the probe to hybridize with, and *extended full model*, where the target has three regions for hybridization and the multiplexed assay involves three pairs of match and mismatch probes. In particular, while the simple model allows us to understand how just the mismatch probe should be designed optimally, the extended full model gives us insight into the extent to which a system of three or more multiplexed probe pairs can be designed by considering only two probe pairs at a time.

A. Simple model

In a situation where the probe can hybridize with exactly one region on the target, we have three possible target states to model: unbound, bound to match probe in the region of interest, and bound to mismatch probe in the region of interest (see Fig. 4).

Here, the match-to-mismatch ratio of interest is

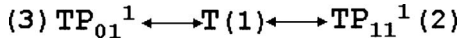


FIG. 4. State transition diagram—simple model.

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{simp}} = \left(\frac{[TP^1_{11}]}{[TP^1_{01}]}\right)_{\text{simp}}.$$

1. Possible states

(1) T
(Target is unbound.)
(2) TP^1_{11}, (3) TP^1_{01}
(Target is bound by specific hybridization.)

The set of reversible reactions operating between unbound and bound states in Fig. 4 can be written as shown below.

$$T + P_{11} \underset{k_{2,1}}{\overset{k_{1,2}}{\rightleftarrows}} TP^1_{11}$$

$$T + P_{01} \underset{k_{3,1}}{\overset{k_{1,3}}{\rightleftarrows}} TP^1_{01}$$

2. Equilibrium equations

At equilibrium (see Supplementary Materials [[5], Sec. IV] for details), these reactions are characterized by mass-action Eqs. (1) and (2).

We augment the above equations with linear constraints corresponding to the conservation rules for $[P_{11}]$, $[P_{01}]$, and $[T]$,

$$[P_{11}]_0 = [P_{11}] + [TP^1_{11}], \tag{28}$$

$$[P_{01}]_0 = [P_{01}] + [TP^1_{01}], \tag{29}$$

$$[T]_0 = [T] + [TP^1_{11}] + [TP^1_{01}]. \tag{30}$$

Finally, we gather the system of equations to be solved, with the change of variables introduced in Sec. IV,

$$\begin{aligned} X_2 &= K_1^2 X_1 Y_1 \\ X_3 &= K_1^3 X_1 Y_2 \\ a_0 &= X_2 + Y_1 \\ b_0 &= X_3 + Y_2 \\ e_0 &= X_1 + X_2 + X_3. \end{aligned} \tag{31}$$

This system consists of
(i) five equations [shown in Eq. (31)], in
(ii) five unknowns: X_1, X_2, X_3, Y_1, and Y_2, with
(iii) three free parameters: a_0, b_0, and e_0.

Let $w \equiv X_1$. The system of equations in Eq. (31) reduces to the cubic polynomial equation

$$e_0 = w\left[1 + a_0\frac{K_1^2}{1+K_1^2 w} + b_0\frac{K_1^3}{1+K_1^3 w}\right]. \tag{32}$$

(Note the similarity with Eq. (24) under model I, in Sec. V B.) The remaining unknowns X_2, X_3, Y_1, and Y_2 are given symbolically in terms of X_1,

$$\begin{aligned} X_2 &= \frac{a_0 K_1^2 X_1}{1+K_1^2 X_1} \quad Y_1 = \frac{a_0}{1+K_1^2 X_1} \\ X_3 &= \frac{b_0 K_1^3 X_1}{1+K_1^3 X_1} \quad Y_2 = \frac{b_0}{1+K_1^3 X_1}. \end{aligned} \tag{33}$$

If w solves Eq. (32), then the match-to-mismatch ratio for probe 1 under simple model is given by

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{simp}} = \frac{a_0}{b_0}\frac{K_1^2}{K_1^3}\frac{1+K_1^3 w}{1+K_1^2 w}$$

$$= \frac{a_0}{b_0}\frac{K_1^2}{K_1^3}\frac{K_1^3 + \frac{1}{w}}{K_1^2 + \frac{1}{w}} \tag{34}$$

$$= \frac{a_0}{b_0}\frac{K_1^2}{K_1^3}\frac{1 + \frac{1}{K_1^3 w}}{\frac{1}{K_1^3}\left(K_1^2 + \frac{1}{w}\right)}. \tag{35}$$

3. Analysis: Choosing initial concentrations

According to Eq. (34), if $w \gg 1$ then ratio $\sim(a_0/b_0)$. On the other hand, from Eq. (35) it follows that if $w \sim 1/K_1^2$, then the ratio simplifies to

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{simp}} \sim \frac{1}{2}\frac{a_0}{b_0}\left(1 + \frac{K_1^2}{K_1^3}\right).$$

(See Supplementary Materials [[5], Sec. VIII] for details.) Thus, for very large and very small values of w, the match-to-mismatch ratio under simple model depends on the initial concentration of the probes and on their thermodynamic parameters.

The initial target concentration $[T]_0 = e_0$ is a free parameter. The question of how it should be chosen in order to optimize the observed discrimination signal at equilibrium merits further investigation.

Discrimination is lowest when target is initially in excess (the prevailing condition in many microarray experiments), because then even the mismatch probe, while interacting more weakly with the target than the match probe, will capture large amounts of target and generate a large signal. This corresponds to the $w \gg 1$ case discussed above, and yields ratio $\sim a_0/b_0$, where a_0 and b_0 are the initial concentrations of the matched probe and the mismatched probe, respectively. These two parameters are usually chosen to be equal, i.e., $a_0 = b_0$. Thus, in this situation, we *cannot* distinguish match signal from mismatch signal.

Conversely, discrimination is highest in the target-depleted setting; in the extreme case, a single target molecule would have to select the match probe over the mismatch probe, producing infinite discrimination but at the expense of a very weak signal; with such low signal strength, the detected intensities would be drowned out by noise.

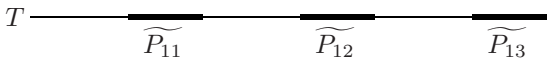
This appears to imply that under usual experimental conditions (i.e., target initially in excess), no competition effects

should be observed, and yet experiments reveal the presence of competition [4]. This apparent paradox is resolved by restricting attention to only those target molecules (a small fraction of the total) that are able to interact with the immobilized probes. The constraint $([T]_0)_{\text{eff}} < \Sigma[P_{ij}]_0$ frequently holds for the effective initial target concentration, which explains why competition effects, revealed in the model only for $[T]_0 < \Sigma[P_{ij}]_0$, are also observed in practice. (See Supplementary Materials [[5], Sec. IX] for a more detailed discussion.)

Preferably, multiplexed analysis should thus be carried out under conditions of slight effective target depletion so as to maximize discrimination while retaining an acceptable signal intensity to facilitate experimental measurements.

B. Extended full model

The final mathematical model (extended full model) involves multiplexed hybridization of a single target with three different probes and is used to verify that the effects suggested by pairwise probe analysis extend to probe triples correctly.



In this scheme, we consider one target, three possible binding sites, and three probe pairs, one for each binding site, as shown in the figure.

1. Possible states

We consider the following states:
(1) T (Target is unbound.)
(2) TP^1_{11}, (3) TP^1_{01}, (4) TP^2_{12},
(5) TP^2_{02}, (6) TP^3_{13}, (7) TP^3_{03}
(Target is bound by specific hybridization: P_{ij} hybridizes to site j.)
(8) TP^2_{11}, (9) TP^2_{01}, (10) TP^3_{11},
(11) TP^3_{01}, (12) TP^1_{12}, (13) TP^1_{02},
(14) TP^3_{12}, (15) TP^3_{02}, (16) TP^1_{13},
(17) TP^1_{03}, (18) TP^2_{13}, (19) TP^2_{03}
(Target is bound by nonspecific hybridization: P_{ij} hybridizes to site k, $k \neq j$.)

The state transition diagram for this model is not shown, as it involves 19 states and is cumbersome to display. The state numbering (the assigned values of states $l=l(i,j,k)$) for all 18 bound target states TP^k_{ij}, where

$$i \in \{0,1\}, \quad \text{probe } j \in \{1,2,3\},$$

$$\text{and site } k \in \{1,2,3\}, \tag{36}$$

grouped by site k, is summarized in Table I(b).

The state interaction can be easily inferred from Table I(a), which shows the set of reversible reactions operating between unbound and bound states: each line represents three reactions, one each for sites $k=1$, 2, and 3, with the corresponding state numbers listed in Table I(b).

TABLE I. State transition table—full extended model.

	Site k	1	2	3
$T+P_{11} \rightleftarrows TP^k_{11}$	States l:	**(2)**	**(8)**	**(10)**
$T+P_{01} \rightleftarrows TP^k_{01}$		**(3)**	**(9)**	**(11)**
$T+P_{12} \rightleftarrows TP^k_{12}$		**(12)**	**(4)**	**(14)**
$T+P_{02} \rightleftarrows TP^k_{02}$		**(13)**	**(5)**	**(15)**
$T+P_{13} \rightleftarrows TP^k_{13}$		**(16)**	**(18)**	**(6)**
$T+P_{03} \rightleftarrows TP^k_{03}$		**(17)**	**(19)**	**(7)**
(a) Reversible reactions		(b) State numbering		

2. Equilibrium equations

At equilibrium (see Supplementary Materials [[5], Sec. V] for details), the set of reversible reactions operating between unbound and bound target states [shown in Table I(a)] is characterized by a system of 18 mass-action equations

$$K^l_1 = \frac{[TP^k_{ij}]}{[T][P_{ij}]} \tag{37}$$

for values of i, j, and k in Eq. (36), where values of l are specified in Table I(b).

Mass conservation rules add the following linear constraints,

$$[P_{ij}]_0 = [P_{ij}] + \sum_{k=1}^{3} [TP^k_{ij}] \quad \text{for } i \in \{0,1\}, \ j \in \{1,2,3\}, \text{ and} \tag{38}$$

$$[T]_0 = [T] + \sum_{i,j,k} [TP^k_{ij}] \tag{39}$$

Analogously to the change of variables introduced in Sec. IV, we let X_n represent the equilibrium concentration of target in state n,

$$X_1 = [T]$$

$$X_2 = [TP^1_{11}] \quad X_8 = [TP^2_{11}] \quad X_{14} = [TP^3_{12}]$$

$$X_3 = [TP^1_{01}] \quad X_9 = [TP^2_{01}] \quad X_{15} = [TP^3_{02}]$$

$$X_4 = [TP^2_{12}] \quad X_{10} = [TP^3_{11}] \quad X_{16} = [TP^1_{13}]$$

$$X_5 = [TP^2_{02}] \quad X_{11} = [TP^3_{01}] \quad X_{17} = [TP^1_{03}]$$

$$X_6 = [TP^3_{13}] \quad X_{12} = [TP^1_{12}] \quad X_{18} = [TP^2_{13}]$$

$$X_7 = [TP^3_{03}] \quad X_{13} = [TP^1_{02}] \quad X_{19} = [TP^2_{03}]$$

let Y_m represent free probe concentrations,

$$Y_1 = [P_{11}] \quad Y_3 = [P_{12}] \quad Y_5 = [P_{13}]$$

$$Y_2 = [P_{01}] \quad Y_4 = [P_{02}] \quad Y_6 = [P_{03}]$$

and have the constant parameters

$$K_1^l, \quad l = 2, \ldots, 19,$$

$$Y_1^0 = [P_{11}]_0, \quad Y_2^0 = [P_{01}]_0,$$

$$Y_3^0 = [P_{12}]_0, \quad Y_4^0 = [P_{02}]_0,$$

$$Y_5^0 = [P_{13}]_0, \quad Y_6^0 = [P_{03}]_0,$$

$$X_1^0 = [T]_0.$$

Note that some indices have changed relative to the notation in Sec. IV since under extended full model we are now considering three possible binding sites k.

We obtain the following simplified system of equations,

$$X_l = K_1^l X_1 Y_n \quad \text{for } l \in \{2, \ldots, 19\},$$
$$\text{where } n \in \{1, \ldots, 6\} \text{ depends on } l(i,j,k) \tag{40}$$

$$Y_n^0 = Y_n + \sum_{l \in f^{-1}(n)} X_l \quad \text{for } n \in \{1, \ldots, 6\}$$
$$\text{(probe conservation)} \tag{41}$$

$$X_1^0 = \sum_{l=1}^{19} X_l$$
$$\text{(target conservation)}. \tag{42}$$

In Eq. (41) we have written $f^{-1}(n)$ to denote the set of states involving probe Y_n, so that, according to Table I, we have

$$f^{-1}(1) = \{2,8,10\} \quad f^{-1}(4) = \{5,13,15\}$$

$$f^{-1}(2) = \{3,9,11\} \quad f^{-1}(5) = \{6,16,18\}$$

$$f^{-1}(3) = \{4,12,14\} \quad f^{-1}(6) = \{7,17,19\}.$$

The simplified algebraic system of equations in this case consists of

(i) 25 equations: 18 quadratic Eqs. (40), 6 linear probe conservation rules [Eq. (41)], and one linear target conservation rule [Eq. (42)], in

(ii) 25 unknowns: $X_1, \ldots, X_{19}$ and $Y_1, \ldots, Y_6$, with

(iii) seven free parameters: $Y_1^0, \ldots, Y_6^0$, and X_1^0.

Again, for clarity we introduce new short-hand notation for frequently appearing sums of affinity constants. Let $s_{i,j,k} \equiv K_1^i + K_1^j + K_1^k$ and let $v \equiv X_1$. After algebraic manipulation (see Supplementary Materials [[5], Sec. VIIE] for details), the system of equations [Eqs. (40)–(42)] reduces to a single seventh degree polynomial equation

$$X_1^0 = v\left[1 + Y_1^0 \frac{s_{2,8,10}}{1 + s_{2,8,10}v} + Y_2^0 \frac{s_{3,9,11}}{1 + s_{3,9,11}v} + Y_3^0 \frac{s_{4,12,14}}{1 + s_{4,12,14}v} + Y_4^0 \frac{s_{5,13,15}}{1 + s_{5,13,15}v} + Y_5^0 \frac{s_{6,16,18}}{1 + s_{6,16,18}v} + Y_6^0 \frac{s_{7,17,19}}{1 + s_{7,17,19}v}\right]. \tag{43}$$

The remaining unknowns are given in terms of X_1 as shown below. Note that vector notation is used for convenience.

$$\begin{pmatrix} X_2 \\ X_8 \\ X_{10} \\ Y_1 \end{pmatrix} = \begin{pmatrix} K_1^2 X_1 \\ K_1^8 X_1 \\ K_1^{10} X_1 \\ 1 \end{pmatrix} \frac{Y_1^0}{1 + s_{2,8,10}X_1}, \tag{44}$$

$$\begin{pmatrix} X_3 \\ X_9 \\ X_{11} \\ Y_2 \end{pmatrix} = \begin{pmatrix} K_1^3 X_1 \\ K_1^9 X_1 \\ K_1^{11} X_1 \\ 1 \end{pmatrix} \frac{Y_2^0}{1 + s_{3,9,11}X_1}, \tag{45}$$

$$\begin{pmatrix} X_4 \\ X_{12} \\ X_{14} \\ Y_3 \end{pmatrix} = \begin{pmatrix} K_1^4 X_1 \\ K_1^{12} X_1 \\ K_1^{14} X_1 \\ 1 \end{pmatrix} \frac{Y_3^0}{1 + s_{4,12,14}X_1}, \tag{46}$$

$$\begin{pmatrix} X_5 \\ X_{13} \\ X_{15} \\ Y_4 \end{pmatrix} = \begin{pmatrix} K_1^5 X_1 \\ K_1^{13} X_1 \\ K_1^{15} X_1 \\ 1 \end{pmatrix} \frac{Y_4^0}{1 + s_{5,13,15}X_1}, \tag{47}$$

$$\begin{pmatrix} X_6 \\ X_{16} \\ X_{18} \\ Y_5 \end{pmatrix} = \begin{pmatrix} K_1^6 X_1 \\ K_1^{16} X_1 \\ K_1^{18} X_1 \\ 1 \end{pmatrix} \frac{Y_5^0}{1 + s_{6,16,18}X_1}, \tag{48}$$

$$\begin{pmatrix} X_7 \\ X_{17} \\ X_{19} \\ Y_6 \end{pmatrix} = \begin{pmatrix} K_1^7 X_1 \\ K_1^{17} X_1 \\ K_1^{19} X_1 \\ 1 \end{pmatrix} \frac{Y_6^0}{1 + s_{7,17,19}X_1}. \tag{49}$$

Finally, the discrimination signals for probes 1, 2, and 3 under extended full model are given by

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{ext}} = \frac{Y_1^0}{Y_2^0} \frac{s_{2,8,10}}{s_{3,9,11}} \frac{1 + s_{3,9,11}v}{1 + s_{2,8,10}v}, \tag{50}$$

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\text{ext}} = \frac{Y_3^0}{Y_4^0} \frac{s_{4,12,14}}{s_{5,13,15}} \frac{1 + s_{5,13,15}v}{1 + s_{4,12,14}v}, \tag{51}$$

$$\left(\frac{TP_{13}}{TP_{03}}\right)_{\text{ext}} = \frac{Y_5^0}{Y_6^0} \frac{s_{6,16,18}}{s_{7,17,19}} \frac{1 + s_{7,17,19}v}{1 + s_{6,16,18}v}, \tag{52}$$

where v solves Eq. (43).

VII. THERMODYNAMIC PARAMETERS

A. Nearest-neighbor model

All models of hybridization discussed in the preceding sections treat the dynamics in terms of kinetic mass-action

reactions and ignore both the mixing properties of the molecules and the exact physics of hybridization except for simply acknowledging that the thermodynamic parameters depend on base-pair composition. The process of hybridization involves the formation of base pairs between Watson-Crick-complementary bases (see, e.g., [15,16]). Specifically, *base pairing* of two single-stranded DNA molecules is determined by the fact that A (adenine) is complementary to T (thymine), and C (cytosine) is complementary to G (guanine). Such base pairing is due to the formation of hydrogen bonds between the complementary bases; this interaction is characterized primarily by the composition of the interacting strands.

The hybridization process is also characterized by another physical interaction, *base stacking*, which has been shown to depend on the sequence rather than the composition of the strands. As base stacking depends on short-range interactions, it is thought to be adequately described by the nearest-neighbor (NN) model (pioneered by Zimm [17] and Tinoco *et al.* [18,19]).

Under the NN model, the stability of a given base pair is determined by the identity and orientation of the neighboring base pairs. Thus, each thermodynamic parameter of the hybridization process, namely the change in enthalpy (ΔH), entropy (ΔS), and free energy (ΔG), is calculated as a sum of the contributions from each nearest-neighbor pair along a strand, corrected by symmetry and initiation parameters. As the enthalpy and entropy terms are assumed to be independent of temperature, they are computed from the sequence data as follows (see e.g., [[15], Ch. 3], [[16], Ch. 20 and 23]).

$$\Delta H = \sum_x \Delta H_x + \Delta H(\text{init}) + \Delta H(\text{sym}),$$

$$\Delta S = \sum_x \Delta S_x + \Delta S(\text{init}) + \Delta S(\text{sym}),$$

where the terms ΔH_x and ΔS_x are tabulated for all ten possible NN dimer duplexes, as are the initiation and symmetry terms ([20,21]). The free energy computation is analogous,

$$\Delta G = \sum_x \Delta G_x + \Delta G(\text{init}) + \Delta G(\text{sym}) \tag{53}$$

with the initiation and symmetry terms tabulated. The values ΔG_x for the dimer duplexes have been reported at 25 °C [20] and at 37 °C [21].

The initiation parameters account for the differences between duplexes with terminal $A \cdot T$ and duplexes with terminal $G \cdot C$. The additional "symmetry" parameter accounts for the maintenance of the $C2$ symmetry of self-complementary duplexes [16].

The ten distinct dimer duplexes arise as follows. Below we list all sixteen possible dimers, identifying the equivalent ones,

$$\boxed{\frac{AA}{TT}} \quad \frac{AC}{TG} \equiv \frac{GT}{CA} \quad \frac{AG}{TC} \equiv \frac{CT}{GA} \quad \boxed{\frac{AT}{TA}}$$

$$\boxed{\frac{CA}{GT}} \quad \frac{CC}{GG} \equiv \frac{GG}{CC} \quad \boxed{\frac{CG}{GC}} \quad \boxed{\frac{CT}{GA}}$$

$$\boxed{\frac{GA}{CT}} \quad \boxed{\frac{GC}{CG}} \quad \boxed{\frac{GG}{CC}} \quad \boxed{\frac{GT}{CA}}$$

$$\boxed{\frac{TA}{AT}} \quad \frac{TC}{AG} \equiv \frac{GA}{CT} \quad \frac{TG}{AC} \equiv \frac{CA}{GT} \quad \frac{TT}{AA} \equiv \frac{AA}{TT}$$

Here, following standard notation (see, e.g., [22]), we write $\frac{AG}{TC}$ to denote $5'$-AG-$3'$ Watson-Crick base-paired with $3'$-TC-$5'$.

Since the NN model was first introduced, there have been a number of studies, both experimental and theoretical, on NN thermodynamics, for DNA, RNA, and RNA/DNA hybrid duplexes (see [20,23–40]). We are using DNA sequences, and thus focused our attention on papers addressing DNA parameters. For short oligonucleotide probes (17-mer DNA probes, in our case), the solution parameters are applicable. Thus, we chose the simplest accepted set of thermodynamic parameters in solution, namely, the consensus parameters from the 1998 paper by SantaLucia [21]. (This study directly compares NN thermodynamic parameters from seven different laboratories [23,26,20,31,32,34,35], and compares them also with oligonucleotide parameters compiled in a study of 108 oligonucleotide duplexes from the literature, reported in [36]. For a comprehensive overview of the field, see [41,42], and references therein.)

When computing ΔG for mismatches, we take the simplest approach by dropping the contribution from the mismatched dimers in Eq. (53). Computations for non-specific hybridization are handled analogously. (For more details, see Supplementary Materials [[5], Sec. X].)

B. Affinity constants

Using appropriate sequence data, the free energy of hybridization (or Gibbs free energy) for target state l due to stacking interactions is computed according to Eq. (53). Since ΔG denotes the tendency for the hybridization reaction to reach equilibrium, it also satisfies

$$\Delta G = -RT \ln K_1^l,$$

where $R = 0.001987$ kcal/(mol K) is the gas constant, T is the temperature (in kelvins), and K_1^l is the affinity constant. Thus, using the value of ΔG computed in Eq. (53), the affinity constant for state l is given by

$$K_1^l = \exp[-\Delta G/RT]. \tag{54}$$

With the affinity constant values computed, we are ready to find the discrimination signals (namely, the "ratios of perfect match to mismatch values") for each probe at any given initial target and probe concentrations.

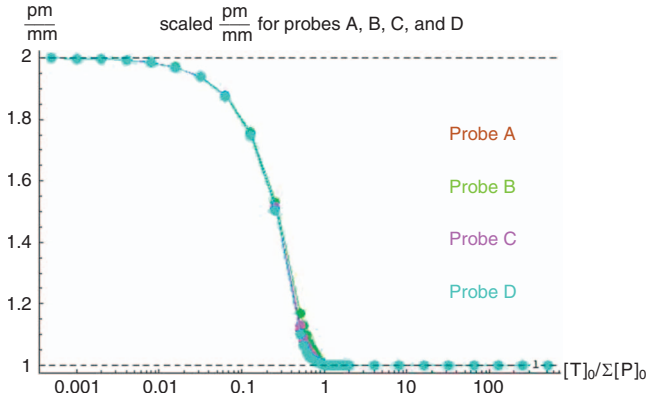


FIG. 5. (Color) Normalized discrimination, or universal Δ-plot.

VIII. OBSERVED COMPETITION AMONG PROBES

Our investigation of the competition effect was initially prompted by a laboratory-observed discrepancy in the signals from probes used in cystic fibrosis genotyping experiments, with 17-mer probes based on CFTR exon 11 sequence [4,3]. In this paper, those probes are denoted as $A=A327$, $B=B354$, $C=C381$, $D=D359$, and $E=E286$, where the number indicates the starting location of the probe sequence in the CFTR exon 11 sequence.

As shown in Secs. III A–III C, we can compute the equilibrium TP concentrations, and thus the corresponding discrimination signals, from the initial target and probe concentrations.

A. Universal Δ-plot

To display the results of the computation and to describe the principal effects of competitive hybridization in a graphical manner, we consider a plot of discrimination (match-to-mismatch ratio) as a function of the molar ratio, $[T]_0/\Sigma[P]_0$, of the initial target concentration and the sum of initial probe concentrations. As discussed in Sec. VI A, for a given probe pair, discrimination will be highest at low molar ratio values, and lowest when target is initially in excess. While the value of maximum discrimination is specific to the probe sequence, the *shape* of the curve is not, as illustrated by the "normalized discrimination" curve in Fig. 5.

Figure 5 shows the match-to-mismatch ratio for each probe pair normalized by the respective sequence-specific affinities, as a function of the molar ratio $[T]_0/\Sigma[P]_0$. Since each affinity constant, computed from sequence-specific NN interactions, indicates the degree to which a duplex is stabilized, normalizing by the respective affinity constant effectively takes out the dependence on the specific sequence, making the resulting Δ-curve sequence-independent.

This result makes the Δ-plot a valuable tool to study competition effects, in that competitive hybridization manifests itself in the form of a *shift* of the Δ-plot for a single pm/mm probe pair in the presence of other probe pairs.

B. Competition: Pairwise analysis

We have computationally simulated the hybridization process for a large number of target/probe sequences used in practice, and observed a difference in pm/mm ratio for probe 1 under partial model (P_1+T) vs full model (P_1+P_2+T). A similar effect was observed for probe 2. These experiments indicated that the *direction* of the shift depends on the affinity constants and can be empirically characterized to be a function of the products of the affinity constants of the perfect match and mismatch probes.

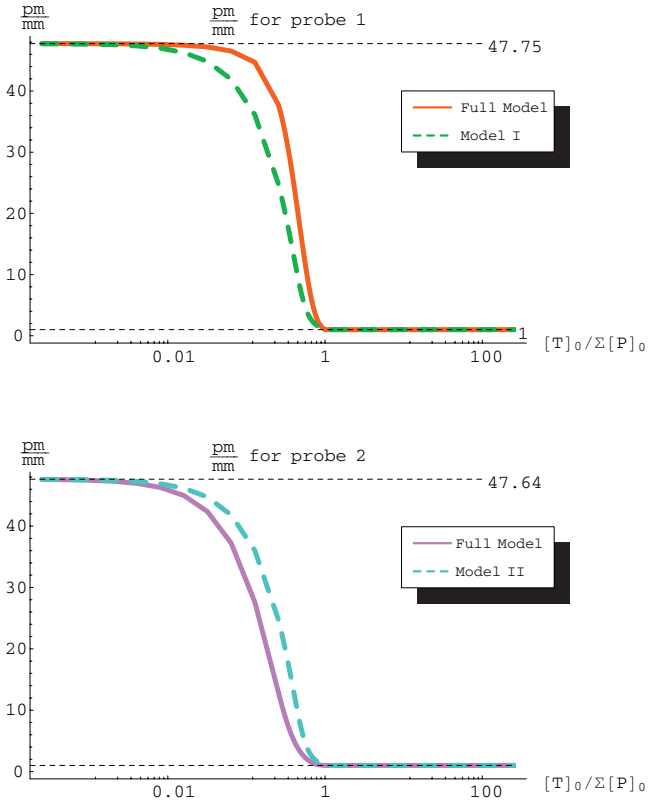


FIG. 6. (Color) pm/mm ratios for probe A (top graph) and probe B (bottom graph), plotted against scaled initial target concentration.

For instance, we examined the behaviors of exon 11 probes A and B (treated as probes 1 and 2, respectively) under the full hybridization model, discussed in Sec. III A, and under partial hybridization models (Secs. III B and III C), as illustrated in Fig. 6. We observe the following:

(1) Ratio $[TP_{A,pm}]/[TP_{A,mm}]$ for probe A (that is, probe pair $\{P_{A,pm}, P_{A,mm}\}$) shifts *up* in the presence of probe B (that is, probe pair $\{P_{B,pm}, P_{B,mm}\}$).

(2) Symmetrically, ratio $[TP_{B,pm}]/[TP_{B,mm}]$ for probe B shifts *down* in the presence of probe A.

A similar discrimination curve-shift effect was observed for all other probe pairs examined. In other words, we examined a number of probes that, in a laboratory setting, showed a different signal in a multiplexed reaction than they did individually. When considered pairwise, in all such probe pairs, the simulated discrimination signal curves exhibited symmetric curve shifts analogous to those displayed in Fig. 6 for the probes A and B.

This suggests the following questions for separate study:

(i) How can the shift direction be predicted?

(ii) How does it depend on the sequences of the probe pairs in question?

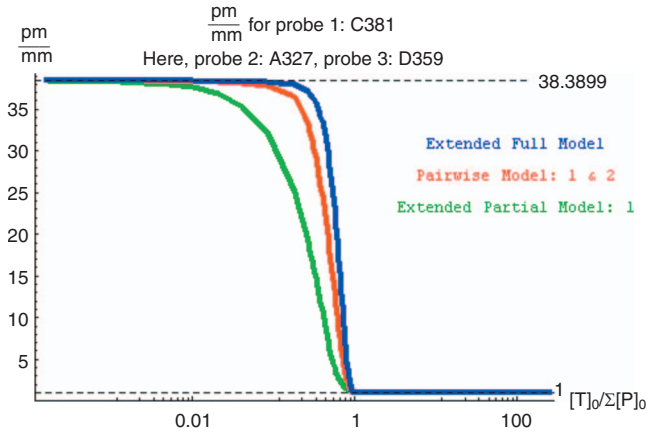


FIG. 7. (Color) Example: pm/mm ratio for probe 1 under three models.

C. Competition: Triple analysis

We have examined several probe triples, drawn from a set used in a multiplexed hybridization reaction. For each triple, and each probe in that triple alternately treated as "probe 1," we simulated the hybridization process and compared the discrimination signals for probe 1 under the following scenarios: alone, paired with another probe from the triple, and grouped with both remaining probes from the triple. Bringing in each additional probe generally caused a shift in the discrimination signal plot. We aimed to address the following question: to what extent can the signal shift for probe 1 in a multiplexed reaction involving all three probes be accurately predicted from pairwise analysis?

We found that in cases where the shift directions are the same, the effect is transitive, and pairwise analysis suffices (as illustrated in Fig. 7). However, in cases where shift directions are opposite for the two pairs, the conclusion cannot be drawn from pairwise analysis, as it is not clear which one dominates. To illustrate the transitivity of the effect, we present the following example.

Example

Let probe *C*381 from exon 11 be probe 1, probe *A*327 be probe 2, and probe *D*359 be probe 3, with the alternates used in the experiments. Pairwise computational analysis indicates that: *A*327 improves the signal for *C*381 and *D*359 improves the signal for *A*327. Transitivity implies that as *D*359 boosts the signal for *A*327, that would, in turn, automatically improve the signal for *C*381 as well. This conclusion was verified using the extended model, as described in detail in Sec. VI B. Recall that the setup for this model includes three probes (each with an alternate) and three possible binding sites on the target for each probe; the "perfect match" for each probe is designed to match the corresponding binding site on the target. In this example, we compared the ratio curves for probe 1 from the full and partial models with the curve from the extended model, as shown in Fig. 7.

Note that, in Fig. 7, the pm/mm ratio curve for *C*381 in the presence of both *A*327 and *D*359 (the blue curve) lies above both the red curve (the ratio for *C*381 in the presence of *A*327) and the green curve (the ratio for *C*381 alone with the target). This indicates that for a given initial target concentration, that is, a given x-coordinate, the pm/mm ratio for *C*381 (the y-coordinate) goes up in the presence of *A*327, which is consistent with pairwise analysis; the ratio increases further when *D*359 is added to the mix, confirming the transitivity of the effect.

TABLE II. Part I: Each probe pair alone.

Probes	PM/MM ratio
A	21.3
B	2.9
C	116.0
D	1.9
E	39.3

TABLE III. Part II: Two probe pairs together. PM/MM Ratio shown for P1 in the presence of P2.

	P2:	A	B	C	D	E
P1		Ratio	Ratio	Ratio	Ratio	Ratio
A		X	9.6	12.1	11.5	27.1
B		5.6	X	2.5	6.0	4.6
C		54.0	43.2	X	56.9	39.9
D		2.3	2.5	2.3	X	2.4
E		26.6	29.4	21.1	27.3	X

IX. EXPERIMENTAL VALIDATION (LABORATORY MEASUREMENTS)

Results presented in the preceding sections are based entirely on simulations, which need to be validated with experi-

TABLE IV. Part III: Three probe pairs together. PM/MM Ratio shown for P1 in the presence of P2&P3.

	P2&P3:	A&B	A&C	A&D	A&E	B&C
P1		Ratio	Ratio	Ratio	Ratio	Ratio
A		X	X	X	X	13.6
B		X	4.0	5.1	5.5	X
C		52.3	X	55.7	59.9	X
D		2.4	2.2	X	2.1	1.8
E		18.6	38.3	25.3	X	24.5
	P2&P3:	B&D	B&E	C&D	C&E	D&E
P1		Ratio	Ratio	Ratio	Ratio	Ratio
A		7.4	7.6	8.2	15.7	14.6
B		X	X	2.9	3.8	4.0
C		41.9	55.4	X	X	57.9
D		X	2.1	X	1.7	X
E		32.7	X	38.9	X	X

TABLE V. Bead count—part I: Each probe pair alone.

Probes	No. beads
A	285
A′	242
B	264
B′	223
C	286
C′	259
D	197
D′	169
E	229
E′	221

mental data from laboratory measurements. We report these data here.

A. Laboratory measurements—raw data

This section contains laboratory measurements from CFTR Exon 11 matrix experiments [4] with probes labeled

TABLE VI. Bead count—part II: Two probe pairs together.

Probes	No. beads	Probes	No. beads
A	143	A	123
A′	141	A′	135
B	153	C	174
B′	109	C′	174
Probes	No. beads	Probes	No. beads
A	93	A	93
A′	69	A′	80
D	83	E	113
D′	59	E′	107
Probes	No. beads	Probes	No. beads
B	108	B	270
B′	127	B′	218
C	134	D	270
C′	97	D′	168
Probes	No. beads	Probes	No. beads
B	78	C	150
B′	61	C′	134
E	82	D	142
E′	83	D′	95
Probes	No. beads	Probes	No. beads
C	119	D	82
C′	133	D′	69
E	144	E	89
E′	121	E′	77

TABLE VII. Bead count—part III: Three probe pairs together.

Probes	No. beads	Probes	No. beads
A	126	A	58
A′	118	A′	77
B	100	B	93
B′	103	B′	74
C	134	D	93
C′	127	D′	68
Probes	No. beads	Probes	No. beads
A	129	A	146
A′	142	A′	155
B	131	C	179
B′	127	C′	188
E	135	D	170
E′	124	D′	129
Probes	No. beads	Probes	No. beads
A	139	A	199
A′	120	A′	164
C	146	D	192
C′	134	D′	147
E	195	E	212
E′	113	E′	200
Probes	No. beads	Probes	No. beads
B	127	B	163
B′	130	B′	135
C	151	C	172
C′	157	C′	162
D	127	E	170
D′	118	E′	148
Probes	No. beads	Probes	No. beads
B	170	C	196
B′	158	C′	175
D	170	D	154
D′	142	D′	154
E	201	E	166
E′	154	E′	172

A, *B*, *C*, *D*, and *E* (see Sec. VIII). Hybridization experiments were performed with perfect match and mismatched versions of these five probes, signal intensity was collected for each combination, and the discrimination signal (PM/MM ratio) was calculated. Each probe [or, more precisely, probe pair $(P_{\cdot,pm}, P_{\cdot,mm})$] was tested individually and in combination with other probes.

Experimental results from Part I (each probe pair alone) are shown in Table II, results from Part II (two probe pairs together, showing PM/MM ratios for P1 in the presence of

TABLE VIII. Direction of shift of PM/MM ratio for probe P1 in the presence of probe P2. x_I shows x-coordinate for one probe (corresponding to Part I); x_{II} shows x-coordinate for two probes (corresponding to Part II).

P1/P2	Sim	Exp	Error	x_I	x_{II}
A/B	Down	Down		0.381	0.368
A/C	Down	Down		0.381	0.331
A/D	Up	Down	X	0.381	0.660
A/E	Down	Up	X	0.381	0.511
B/A	Up	Up		0.412	0.368
B/C	Down	Down		0.412	0.431
B/D	Up	Up		0.412	0.217
B/E	Up	Up		0.412	0.660
C/A	Down	Down		0.368	0.331
C/B	Up	Down	X	0.368	0.431
C/D	Up	Down	X	0.368	0.385
C/E	Down	Down		0.368	0.388
D/A	Up	Up		0.549	0.660
D/B	Up	Up		0.549	0.217
D/C	Up	Up		0.549	0.385
D/E	Up	Up		0.549	0.633
E/A	Up	Down	X	0.446	0.511
E/B	Down	Down		0.446	0.660
E/C	Up	Down	X	0.446	0.388
E/D	Up	Down	X	0.446	0.633

P2) are shown in Table III, and results from Part III (three probe pairs together, showing PM/MM ratios for P1 in the presence of P2 and P3) are shown in Table IV.

B. Experimental setup

To understand the data reported in Sec. IX A and to explain how simulations were rescaled to allow a valid comparison with experimental data, we present here some of the details of the experimental setup.

Cystic fibrosis is one of the most common autosomal recessive disorders affecting Caucasians with a carrier frequency among Ashkenazi Jewish and non-Jewish individuals of 1 in 25 to 1 in 29. About 900 mutations have been identified in this gene. However, common mutations occur in greater frequency in different populations. These mutations are located on several exons within the gene. Exon 11 has several mutations located within 300 to 400 nucleotides. This study describes the competitive hybridization of five probe pairs designed to identify five of these mutations in exon 11 sequences. The experiments were performed with each probe pair alone or in combination with other probes.

Materials and Methods. Perfect match and mismatched probes for five mutations were synthesized with 5′ amine modification and coupled on bead surface. Genomic DNA was used to amplify 240 bp long PCR products from Exon 11. Forward primer was modified with a Cy3 label at 5′ end. After amplification PCR products were processed to produce single stranded target. Hybridization was carried out in buffer composed of: 1.125 M tetramethyl-ammonium chloride (TMAC), 18.75 mM Tris-HCL (pH 8.0) and 0.375% SDS. Fifteen μL of hybridization mixture containing buffer and ssDNA was added on the chip surface and incubated at 55°C for 15 min. Each probe pair was tested individually and in combination with all other probes.

Bead counts for each probe in each of the experimental groupings are shown in the following tables: Table V shows the numbers of beads used for each probe pair examined alone (this corresponds to the measured PM/MM ratios shown in Table II). Table VI shows bead counts for two probe pairs examined together (corresponding to PM/MM ratios in Table III). Finally, Table VII shows bead counts for three probe pairs examined together (corresponding to PM/MM ratios in Table IV).

C. Comparison of simulations with experiments

Tables VIII and IX show the direction of shift of the PM/MM ratio for probe pair P1 in the presence of a single other probe pair P2 (Table VIII) and in the presence of two other probe pairs P2 and P3 (Table IX), computed based on competitive hybridization models ("Sim" column) and measured experimentally ("Exp" column). The experimental shift direction was obtained from the measured PM/MM ra-

TABLE IX. Direction of shift of PM/MM ratio for probe P1 in the presence of probes P2 and P3. x_I shows x-coordinate for one probe (corresponding to Part I); x_{III} shows x-coordinate for three probes (corresponding to Part III).

P1/P2P3	Sim	Exp	Error	x_I	x_{III}
A/BC	Down	Down		0.381	0.284
A/BD	Down	Down		0.381	0.434
A/BE	Down	Down		0.381	0.255
A/CD	Down	Down		0.381	0.208
A/CE	Down	Down		0.381	0.237
A/DE	Up	Down	X	0.381	0.180
B/AC	Down	Up	X	0.412	0.284
B/AD	Up	Up		0.412	0.434
B/AE	Down	Up	X	0.412	0.255
B/CD	Down	Down		0.412	0.248
B/CE	Up	Up		0.412	0.211
B/DE	Down	Up	X	0.412	0.202
C/AB	Down	Down		0.368	0.284
C/AD	Down	Down		0.368	0.208
C/AE	Down	Down		0.368	0.237
C/BD	Down	Down		0.368	0.248
C/BE	Down	Down		0.368	0.211
C/DE	Up	Down	X	0.368	0.197
D/AB	Up	Up		0.549	0.434
D/AC	Up	Up		0.549	0.208
D/AE	Up	Up		0.549	0.180
D/BC	Down	Down		0.549	0.248
D/BE	Up	Up		0.549	0.202
D/CE	Down	Down		0.549	0.197
E/AB	Up	Down	X	0.446	0.255
E/AC	Up	Down	X	0.446	0.237
E/AD	Up	Down	X	0.446	0.180
E/BC	Up	Down	X	0.446	0.211
E/BD	Up	Down	X	0.446	0.202
E/CD	Down	Down		0.446	0.197

tios shown in Tables II–IV. To ensure valid comparison of simulations with experimental data, the PM/MM ratios were computed using initial probe concentrations (such as parameters a_0 and b_0 in partial model I) set to reflect initial probe concentrations used in experiments, as calculated from the tabulated bead counts (shown in Tables V–VII).

We note that, since in the experimental setup the numbers of beads varied dramatically from experiment to experiment, as well as from probe to probe within a given experiment (and even between a given probe and its alternate), adjusting initial probe concentrations accordingly for simulations resulted in comparing PM/MM ratios (y-values in Figs. 6 or 7) for *different* x-values, which differs from the intended interpretation and the analysis we used in discussing the probe triple example in Fig. 7. The x-coordinates used in simulations are shown in the last two columns of Tables VIII and IX.

We see disagreements between simulations and experiments (marked by 'X') in approximately 1/3 of the cases, both in pairwise and triple analysis.

More specifically, in pairwise analysis (Table VIII) we have $C(5,2)=10$ pairs {P1,P2}, yielding 20 data points (signal for P1 in the presence of P2). The error rate is 7 out of 20. 4 of these involve probe E (once as P2, three times as P1). In triple analysis (Table IX) we have $C(5,3)=10$ triples {P1,P2,P3}, yielding 30 data points (signal for P1 in the presence of P2 and P3). The error rate is 10 out of 30. 9 of these involve probe E (four times in the P2/P3 pair, five times as P1). The majority of disagreements involves subsets containing probe E, the capture sequence for which is much closer to the 5′ end of the target than in the other probes. We suspect a potential interaction between probe E and 20-bp-long tags at the ends of the target sequence, unaccounted for in the competitive hybridization models. If probe E is ex-

cluded from consideration, in pairwise analysis this leaves $C(4,2)=6$ pairs (12 data points), which brings the error rate down to 3 out of 12; analogously, in triple analysis this leaves $C(4,3)=4$ pairs (12 data points), which brings the error rate down to 1 out of 12.

X. CONCLUSION

Microarrays use is ubiquitous in molecular biology. Yet, the field is still far from fully understanding interactions among the many multiplexed probes involved. The mathematical models of competitive hybridization presented in this paper go a long way toward advancing that understanding. These models were validated (see Sec. IX); simulations implementing them were largely in agreement with experimental data.

Pairwise analysis of probe interactions based on competitive hybridization models can be performed *in silico* as a step in experiment design. Results from this analysis can lead to probe selection (choosing which of the several candidate probes should be used) and pooling strategies (choosing sets of probes that should be used together in the same experiment) in many applications ranging from genotyping to pathogen identification.

The work presented here lays out a solid foundation for understanding pairwise probe interactions. Discovering the precise mechanisms for applying competitive hybridization models to improving experiment design forms a fertile ground for future research.

ACKNOWLEDGMENTS

The authors thank Michael Seul for posing the questions motivating the original investigation as well as for many helpful discussions. The authors also wish to acknowledge the support of the members of the NYU Bioinformatics group and our colleagues at Cold Spring Harbor Laboratory and at Fairfield University, who offered constructive criticisms, insights, suggestions, and encouragement. V.C. was supported by NSF and NYU. B.M. was supported by a Department of Energy grant, a National Cancer Institute grant, two NSF ITR grants, an NSF EMT grant, a NYSTAR grant, an NYNBIT grant, a NIST grant, and a DARPA BioComp grant.

[1] J. Quackenbush, Yearb. Med. Inform. **2006**, 91 [http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/imia-yearbook/imia-yearbook-2006/issue/839/manuscript/6447.html].
[2] F. Naef, D. A. Lim, N. Patil, and M. Magnasco, Phys. Rev. E **65**, 040902 (2002).
[3] L. Edelmann, G. Hashmi, Y. Song, Y. Han, R. Kornreich, and R. J. Desnick, Genet. Med. **6**, 431 (2004).
[4] G. Hashmi (private communication).
[5] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevE.82.051914 for details of derivations and a glossary of biological terms used in this paper.
[6] A. X. Li, M. Seul, J. Cicciarelli, J. C. Yang, and Y. Iwaki, Tissue Antigens **63**, 518 (2004).
[7] G. Hashmi, T. Shariff, M. Seul, P. Vissavajjhala, K. Hue-Roye, D. Charles-Pierre, C. Lomas-Francis, A. Chaudhuri, and M. E. Reid, Transfusion **45**, 680 (2005).
[8] M. Feinberg and F. J. M. Horn, Chem. Eng. Sci. **29**, 775 (1974).
[9] M. Kork and M. Feinberg, Chem. Eng. Sci. **48**, 4143 (1993).
[10] M. Feinberg, Chem. Eng. Sci. **44**, 1819 (1989).
[11] Martin Feinberg, "Lectures on chemical reaction networks." (1979), Notes of lectures given at the Mathematics Research Center, University of Wisconsin-Madison.
[12] *Dynamics and Modeling of Reactive Systems*, edited by Warren E. Stewart, W. Harmon Ray, and Charles C. Conley, Proceedings of an Advanced Seminar Conducted by the Mathematics Research Center, The University of Wisconsin–Madison, 1979 (Academic Press, New York, 1980).
[13] B. Sturmfels, Notices of the AMS **52**, 2 (2005) [http://www.ams.org/notices/200510/index.html] and click on title "What is a Groebner Basis" under Communications to download.
[14] B. Mishra, *Algorithmic Algebra, Texts and Monographs in Computer Science* (Springer-Verlag, New York, 1993) .
[15] Charles R. Cantor and Cassandra L. Smith, *Genomics: The Science and Technology Behind the Human Genome Project* (John Wiley & Sons, Inc., New York, 1999).
[16] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules* (Freeman, San Francisco, 1980).
[17] D. M. Crothers and B. H. Zimm, J. Mol. Biol. **9**, 1 (1964).
[18] H. DeVoe and I. Tinoco, J. Mol. Biol. **4**, 500 (1962).
[19] D. M. Gray and I. Tinoco, Biopolymers **9**, 223 (1970).
[20] K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, Proc. Natl. Acad. Sci. U.S.A. **83**, 3746 (1986).
[21] J. SantaLucia, Jr., Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[22] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules* (Freeman, San Francisco, 1980).
[23] O. Gotoh and Y. Tagashira, Biopolymers **20**, 1033 (1981).
[24] R. L. Ornstein and J. R. Fresco, Biopolymers **22**, 1979 (1983).
[25] P. Otto, J. Mol. Struct.: THEOCHEM **188**, 277 (1989).
[26] A. V. Vologodskii, B. R. Amirikyan, Y. L. Lyubchenko, and M. D. Frank-Kamenetskii, J. Biomol. Struct. Dyn. **2**, 131 (1984).
[27] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[28] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner, Proc. Natl. Acad. Sci. U.S.A. **83**, 9373 (1986).
[29] M. Aida, J. Theor. Biol. **130**, 327 (1988).
[30] R. S. Quartin and J. G. Wetmur, Biochemistry **28**, 1040 (1989).
[31] S. G. Delcourt and R. D. Blake, J. Biol. Chem. **266**, 15160 (1991).
[32] M. J. Doktycz, R. F. Goldstein, T. M. Paner, F. J. Gallo, and A.

S. Benight, Biopolymers **32**, 849 (1992).
[33] R. F. Goldstein and A. S. Benight, Biopolymers **32**, 1679 (1992).
[34] J. SantaLucia, Jr., H. T. Allawi, and P. A. Seneviratne, Biochemistry **35**, 3555 (1996).
[35] N. Sugimoto, S. Nakano, M. Yoneyama, and K. Honda, Nucleic Acids Res. **24**, 4501 (1996).
[36] H. T. Allawi and J. SantaLucia, Jr., Biochemistry **36**, 10581 (1997).
[37] D. M. Gray, Biopolymers **42**, 795 (1997).
[38] C. Rennie, H. A. Noyes, S. J. Kemp, H. Hulme, A. Brass, and D. C. Hoyle, BMC Genomics **9**, 317 (2008).
[39] J. Hooyberghs, P. Van Hummelen, and E. Carlon, Nucleic Acids Res. **37**, e53 (2009).
[40] D. J. Fish, M. T. Horne, G. P. Brewood, J. P. Goodarzi, S. Alemayehu, A. Bhandiwad, R. P. Searles, and A. S. Benight, Nucleic Acids Res. **35**, 7197 (2007).
[41] J. SantaLucia and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[42] Victor A. Bloomfield, Donald M. Crothers, and Ignacio Tinoco, *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, California, 2000).