Competitive Hybridization Model
CIMS Technical Report: TR2008-913

Vera Cherepinsky,[1, *] Ghazala Hashmi,[2] Michael Seul,[2] and Bud Mishra[3, 4, †]

[1]*Department of Mathematics and Computer Science, Fairfield University, Fairfield, CT 06824, USA*
[2]*BioArray Solutions, Ltd., Warren, NJ 07509, USA*
[3]*Courant Institute of Mathematical Sciences, NYU, New York, NY 10012, USA*
[4]*NYU School of Medicine, New York, NY 10016, USA*
(Dated: August 28, 2008)

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 gives 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. The model is formulated as a system of ordinary differential equations (ODE's) describing kinetic mass action and conservation-of-mass equations completing the system.

We examine pair-wise probe interactions in detail and present a model of "competition" between the probes for the target—especially, when target is 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 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 are 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.).

Keywords: physical models of DNA, microarray design, multiplexed hybridization analysis, competitive hybridization, steady-state models, chemical reactions.

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. 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, but not adequately explained. In the following simple example of this phenomenon, a target consisting of possibly two 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 can be rescued by judicious choice of the selected probes and the thermodynamic parameters.

II. SETUP

More specifically, we consider the following experimental setup: Probes are bound to encoded microparticles (e.g., "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.

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

$$T \underset{\widetilde{P_{11}} \qquad \widetilde{P_{12}}}{\rule{6cm}{0.4pt}}$$
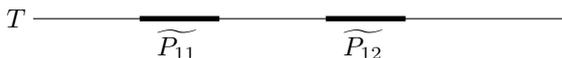
Let P_{01} differ from P_{11} in one base (i.e, the Hamming distance between P_{01} and P_{11} equals to 1, $H(P_{01}, P_{11}) = 1$).

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} ($H(P_{01}, P_{02}) = 2$). 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 non-trivial problem.

III. DYNAMICS

A mathematical model to analyze the dynamics involved in a setup like the earlier one is described below. As before, we assume that the steric effects prevent multiple probes from hybridizing to a single target strand (as probes are bound to large beads).

A. Full Model

We may observe a target strand T in one of the following nine *possible states*:

(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 "non-specific" hybridization.)

Bound target states have form TP_{ij}^{k}, where $j \in \{1, 2\}$ is the probe index,

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

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.

1. State Transition Diagram
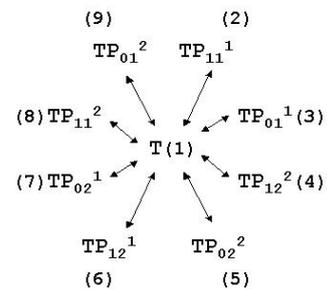
The set of reversible reactions operating between unbound and bound states can be written as shown below, where the forward and backward reaction rates are

indicated with $k_{i,j}$ and $k_{j,i}$, respectively. While the reaction rates themselves are difficult to compute, the ratios (affinity constants, $K_i^j = k_{i,j}/k_{j,i}$) may be computed from purely thermodynamic considerations, and are sufficient for the "equilibrium analysis."

$$T + P_{11} \xrightleftharpoons[k_{2,1}]{k_{1,2}} TP_{11}^1 \qquad T + P_{12} \xrightleftharpoons[k_{6,1}]{k_{1,6}} TP_{12}^1$$

$$T + P_{01} \xrightleftharpoons[k_{3,1}]{k_{1,3}} TP_{01}^1 \qquad T + P_{02} \xrightleftharpoons[k_{7,1}]{k_{1,7}} TP_{02}^1$$

$$T + P_{12} \xrightleftharpoons[k_{4,1}]{k_{1,4}} TP_{12}^2 \qquad T + P_{11} \xrightleftharpoons[k_{8,1}]{k_{1,8}} TP_{11}^2$$

$$T + P_{02} \xrightleftharpoons[k_{5,1}]{k_{1,5}} TP_{02}^2 \qquad T + P_{01} \xrightleftharpoons[k_{9,1}]{k_{1,9}} TP_{01}^2$$
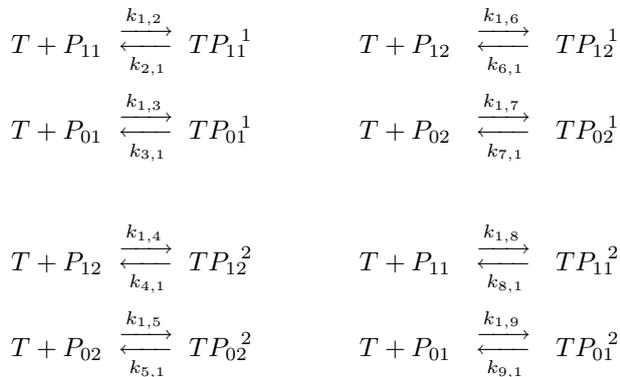
We wish to perform a stationary analysis, where these reactions are allowed to run to *equilibrium*. We begin by assuming that all complexes can be distinguished and writing down the ODE's (ordinary differential equations) describing the dynamics of the system as follows.

$$\frac{d[T]}{dt} = k_{2,1}[TP_{11}^1] + k_{3,1}[TP_{01}^1] + k_{4,1}[TP_{12}^2] + k_{5,1}[TP_{02}^2] + k_{6,1}[TP_{12}^1] + k_{7,1}[TP_{02}^1] + k_{8,1}[TP_{11}^2] + k_{9,1}[TP_{01}^2]$$
$$- k_{1,2}[T][P_{11}] - k_{1,3}[T][P_{01}] - k_{1,4}[T][P_{12}] - k_{1,5}[T][P_{02}]$$
$$- k_{1,6}[T][P_{12}] - k_{1,7}[T][P_{02}] - k_{1,8}[T][P_{11}] - k_{1,9}[T][P_{01}] \tag{1}$$

$$\frac{d[TP_{11}^1]}{dt} = k_{1,2}[T][P_{11}] - k_{2,1}[TP_{11}^1] \tag{2}$$

$$\frac{d[TP_{01}^1]}{dt} = k_{1,3}[T][P_{01}] - k_{3,1}[TP_{01}^1] \tag{3}$$

$$\frac{d[TP_{12}^2]}{dt} = k_{1,4}[T][P_{12}] - k_{4,1}[TP_{12}^2] \tag{4}$$

$$\frac{d[TP_{02}^2]}{dt} = k_{1,5}[T][P_{02}] - k_{5,1}[TP_{02}^2] \tag{5}$$

$$\frac{d[TP_{12}^1]}{dt} = k_{1,6}[T][P_{12}] - k_{6,1}[TP_{12}^1] \tag{6}$$

$$\frac{d[TP_{02}^1]}{dt} = k_{1,7}[T][P_{02}] - k_{7,1}[TP_{02}^1] \tag{7}$$

$$\frac{d[TP_{11}^2]}{dt} = k_{1,8}[T][P_{11}] - k_{8,1}[TP_{11}^2] \tag{8}$$

$$\frac{d[TP_{01}^2]}{dt} = k_{1,9}[T][P_{01}] - k_{9,1}[TP_{01}^2] \tag{9}$$

Let
$$\vec{X} = (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9)^T \tag{10}$$
$$= \bigg([T],$$
$$[TP_{11}^1], [TP_{01}^1], [TP_{12}^2], [TP_{02}^2],$$
$$[TP_{12}^1], [TP_{02}^1], [TP_{11}^2], [TP_{01}^2] \bigg)^T$$

Note that at equilibrium,

$$\frac{d\vec{X}}{dt} = \vec{0}. \tag{11}$$

Applying (11) to equations (2)–(9) yields

$$k_{1,2}[T][P_{11}] = k_{2,1}[TP_{11}^1]$$
$$\implies K_1^2 \equiv \frac{k_{1,2}}{k_{2,1}} = \frac{[TP_{11}^1]}{[T][P_{11}]} \tag{12}$$

$$k_{1,3}[T][P_{01}] = k_{3,1}[TP_{01}^1]$$
$$\implies K_1^3 \equiv \frac{k_{1,3}}{k_{3,1}} = \frac{[TP_{01}^1]}{[T][P_{01}]} \tag{13}$$

$$k_{1,4}[T][P_{12}] = k_{4,1}[TP_{12}^2]$$
$$\implies K_1^4 \equiv \frac{k_{1,4}}{k_{4,1}} = \frac{[TP_{12}^2]}{[T][P_{12}]} \tag{14}$$

$$k_{1,5}[T][P_{02}] = k_{5,1}[TP_{02}^{\ 2}]$$
$$\implies K_1^5 \equiv \frac{k_{1,5}}{k_{5,1}} = \frac{[TP_{02}^{\ 2}]}{[T][P_{02}]} \qquad (15)$$

$$k_{1,6}[T][P_{12}] = k_{6,1}[TP_{12}^{\ 1}]$$
$$\implies K_1^6 \equiv \frac{k_{1,6}}{k_{6,1}} = \frac{[TP_{12}^{\ 1}]}{[T][P_{12}]} \qquad (16)$$

$$k_{1,7}[T][P_{02}] = k_{7,1}[TP_{02}^{\ 1}]$$
$$\implies K_1^7 \equiv \frac{k_{1,7}}{k_{7,1}} = \frac{[TP_{02}^{\ 1}]}{[T][P_{02}]} \qquad (17)$$

$$k_{1,8}[T][P_{11}] = k_{8,1}[TP_{11}^{\ 2}]$$
$$\implies K_1^8 \equiv \frac{k_{1,8}}{k_{8,1}} = \frac{[TP_{11}^{\ 2}]}{[T][P_{11}]} \qquad (18)$$

$$k_{1,9}[T][P_{01}] = k_{9,1}[TP_{01}^{\ 1}]$$
$$\implies K_1^9 \equiv \frac{k_{1,9}}{k_{9,1}} = \frac{[TP_{01}^{\ 1}]}{[T][P_{01}]} \qquad (19)$$

and applying it to equation (1) yields

$$k_{2,1}[TP_{11}^{\ 1}] + k_{3,1}[TP_{01}^{\ 1}] + k_{4,1}[TP_{12}^{\ 2}] + k_{5,1}[TP_{02}^{\ 2}] + k_{6,1}[TP_{12}^{\ 1}] + k_{7,1}[TP_{02}^{\ 1}] + k_{8,1}[TP_{11}^{\ 2}] + k_{9,1}[TP_{01}^{\ 2}]$$
$$= [T]\big(k_{1,2}[P_{11}] + k_{1,3}[P_{01}] + k_{1,4}[P_{12}] + k_{1,5}[P_{02}] + k_{1,6}[P_{12}] + k_{1,7}[P_{02}] + k_{1,8}[P_{11}] + k_{1,9}[P_{01}]\big) \qquad (20)$$

Equation (20) is a linear combination of (12), …, (19), and hence provides no additional information. Observe that

$$\frac{d[T]}{dt} = -\frac{d}{dt}\big\{[TP_{11}^{\ 1}] + [TP_{01}^{\ 1}] + [TP_{12}^{\ 2}] + [TP_{02}^{\ 2}]$$
$$+ [TP_{12}^{\ 1}] + [TP_{02}^{\ 1}] + [TP_{11}^{\ 2}] + [TP_{01}^{\ 1}]\big\}$$
$$\text{or } (1) = -\sum_{j=(2)}^{(9)}\{\text{equation } j\}$$

The constants K_1^j for $j \in \{2,\ldots,9\}$, appearing in equations (12)–(19), can be computed from probe sequence data. For each j,

$$\Delta G_{\text{total}} = -RT \ln K_1^j,$$

where R is the gas constant and T is the temperature (in degrees Kelvin). Thus, we have

$$K_1^j = \exp\left[-\Delta G_{\text{total}}/RT\right], \qquad (21)$$

where

$$\Delta G_{\text{total}} = -(\underbrace{\Delta g_i}_{\text{initiation}} + \underbrace{\Delta g_{\text{symm}}}_{\text{symmetry}}) + \sum_x \underbrace{\Delta g_x}_{\text{sequence data}}.$$

This notation and form follows [1]. Since a more recent paper by SantaLucia [2] presents the calculation of ΔG_{total} in a slightly different format (see equation (136)), both versions are available in the implementation of our model.

The described model can now be used to predict equilibrium concentrations of complexes TP_{ij} $\{i \in \{0,1\}, j \in \{1,2\}\}$:
- K_1^j can be computed from (21), where ΔG_{total} is computed based on sequence information.

- The following conservation rules must hold:

$$[P_{11}]_0 = [P_{11}] + [TP_{11}^{\ 1}] + [TP_{11}^{\ 2}] \qquad (22)$$
$$[P_{01}]_0 = [P_{01}] + [TP_{01}^{\ 1}] + [TP_{01}^{\ 2}] \qquad (23)$$
$$[P_{12}]_0 = [P_{12}] + [TP_{12}^{\ 1}] + [TP_{12}^{\ 2}] \qquad (24)$$
$$[P_{02}]_0 = [P_{02}] + [TP_{02}^{\ 1}] + [TP_{02}^{\ 2}] \qquad (25)$$

$$[T]_0 = [T] + [TP_{11}^{\ 1}] + [TP_{01}^{\ 1}] + [TP_{12}^{\ 2}] + [TP_{02}^{\ 2}] + [TP_{11}^{\ 2}] + [TP_{01}^{\ 2}] + [TP_{12}^{\ 1}] + [TP_{02}^{\ 1}] \qquad (26)$$
$$= [T] + ([P_{11}]_0 - [P_{11}]) + ([P_{01}]_0 - [P_{01}]) + ([P_{12}]_0 - [P_{12}]) + ([P_{02}]_0 - [P_{02}])$$

Note that in these equations, for each species X,

$[X]_0$ is a free parameter that denotes the initial con-

centration, and $[X]$ denotes the equilibrium concentration.

Consider the system consisting of equations (12)–(19) and the conservation rule equations (22)–(26). We have a system of

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

- 13 unknowns: $X_1, \ldots, X_9$ (see (10)) and $[P_{11}]$, $[P_{01}]$, $[P_{12}]$, $[P_{02}]$, with

- 5 free parameters: $[P_{11}]_0$, $[P_{01}]_0$, $[P_{12}]_0$, $[P_{02}]_0$, and $[T]_0$.

Therefore, this algebraic system, when solved, yields the equilibrium concentrations. From these computed concentrations, we can 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 should now 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 will refer to the model introduced in this section as the *Full Model* and will 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 solving the algebraic system of equations to evaluate the "match-to-mismatch ratio" for probe with index $j = 1$ (in the absence of probe with index $j = 2$):

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

1. Possible States

We consider the following states:

(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 "non-specific" hybridization.)
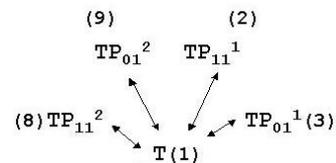
2. State Transition Diagram

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

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

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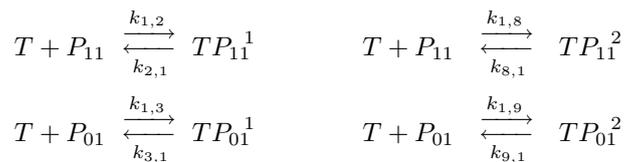
3. Dynamics

The following are the ODE's describing the dynamics of the system.

$$\begin{aligned}
\frac{d[T]}{dt} \;=\;& k_{2,1}[TP_{11}{}^1] - k_{1,2}[T][P_{11}] \\
&+ k_{3,1}[TP_{01}{}^1] - k_{1,3}[T][P_{01}] \\
&+ k_{8,1}[TP_{11}{}^2] - k_{1,8}[T][P_{11}] \\
&+ k_{9,1}[TP_{01}{}^2] - k_{1,9}[T][P_{01}] \quad (27)
\end{aligned}$$

$$\frac{d[TP_{11}{}^1]}{dt} \;=\; k_{1,2}[T][P_{11}] - k_{2,1}[TP_{11}{}^1] \quad (28)$$

$$\frac{d[TP_{01}{}^1]}{dt} \;=\; k_{1,3}[T][P_{01}] - k_{3,1}[TP_{01}{}^1] \quad (29)$$

$$\frac{d[TP_{11}{}^2]}{dt} \;=\; k_{1,8}[T][P_{11}] - k_{8,1}[TP_{11}{}^2] \quad (30)$$

$$\frac{d[TP_{01}{}^2]}{dt} \;=\; k_{1,9}[T][P_{01}] - k_{9,1}[TP_{01}{}^2] \quad (31)$$

Note that equations (28)–(31) are identical to equations (2), (3), (8), and (9) in the original system in section III A, while equation (27) differs from (1), since it now involves only the states with probes P_{11} and P_{01}.

At equilibrium, $\frac{d[\cdot]}{dt} = 0$ for all substances, i.e., T, TP_{11}^1, TP_{11}^2, TP_{01}^1, and TP_{01}^2, yielding:

$$K_1^2 = \frac{[TP_{11}^1]}{[T][P_{11}]} \qquad (32)$$

$$K_1^3 = \frac{[TP_{01}^1]}{[T][P_{01}]} \qquad (33)$$

$$K_1^8 = \frac{[TP_{11}^2]}{[T][P_{11}]} \qquad (34)$$

$$K_1^9 = \frac{[TP_{01}^1]}{[T][P_{01}]} \qquad (35)$$

Since nothing else has changed in the thermodynamics, K_1^j computed from (21) are the same as before for $j \in \{2, 3, 8, 9\}$, and we have the following conservation rules:

$$[P_{11}]_0 = [P_{11}] + [TP_{11}^1] + [TP_{11}^2] \qquad (36)$$

$$[P_{01}]_0 = [P_{01}] + [TP_{01}^1] + [TP_{01}^2] \qquad (37)$$

$$\begin{aligned} [T]_0 &= [T] + [TP_{11}^1] + [TP_{01}^1] \\ &\quad + [TP_{11}^2] + [TP_{01}^2] \qquad (38) \\ &= [T] + ([P_{11}]_0 - [P_{11}]) + ([P_{01}]_0 - [P_{01}]) \end{aligned}$$

In this case, we have

- 7 polynomial equations: (32)–(35), (36), (37), and (38), in

- 7 unknowns: $[T]$, $[TP_{11}^1]$, $[TP_{11}^2]$, $[TP_{01}^1]$, $[TP_{01}^2]$, and $[P_{11}]$, $[P_{01}]$, with

- 3 free parameters $[P_{11}]_0$, $[P_{01}]_0$, and $[T]_0$.

Note that, for comparison with full model, the free parameters will 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 the algebraic system of equations to evaluate the "match-to-mismatch ratio" for probe with index $j = 2$ (in the absence of probe with index $j = 1$):

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

1. Possible States

We consider the following states:

(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 "non-specific" hybridization.)
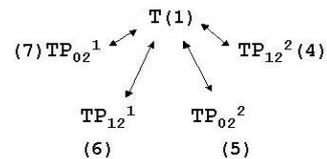
2. State Transition Diagram

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

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

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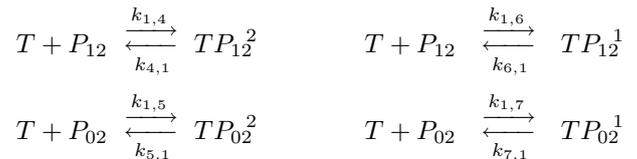
3. Dynamics

The following are the ODE's describing the dynamics of the system.

$$\begin{aligned} \frac{d[T]}{dt} &= k_{4,1}[TP_{12}^2] - k_{1,4}[T][P_{12}] \\ &\quad + k_{5,1}[TP_{02}^2] - k_{1,5}[T][P_{02}] \\ &\quad + k_{6,1}[TP_{12}^1] - k_{1,6}[T][P_{12}] \\ &\quad + k_{7,1}[TP_{02}^1] - k_{1,7}[T][P_{02}] \qquad (39) \end{aligned}$$

$$\frac{d[TP_{12}^2]}{dt} = k_{1,4}[T][P_{12}] - k_{4,1}[TP_{12}^2] \qquad (40)$$

$$\frac{d[TP_{02}^2]}{dt} = k_{1,5}[T][P_{02}] - k_{5,1}[TP_{02}^2] \qquad (41)$$

$$\frac{d[TP_{12}^1]}{dt} = k_{1,6}[T][P_{12}] - k_{6,1}[TP_{12}^1] \qquad (42)$$

$$\frac{d[TP_{02}^1]}{dt} = k_{1,7}[T][P_{02}] - k_{7,1}[TP_{02}^1] \qquad (43)$$

Note that equations (40)–(43) are identical to equations (4), (5), (6), and (7) in the original system in section

III A, while equation (39) differs from (1), since it now involves only the states with probes P_{12} and P_{02}.

At equilibrium, $\frac{d[\cdot]}{dt} = 0$ for all substances, i.e., T, TP_{12}^2, TP_{12}^1, TP_{02}^2, and TP_{02}^1, yielding:

$$K_1^4 = \frac{[TP_{12}^2]}{[T][P_{12}]} \tag{44}$$

$$K_1^5 = \frac{[TP_{02}^2]}{[T][P_{02}]} \tag{45}$$

$$K_1^6 = \frac{[TP_{12}^1]}{[T][P_{12}]} \tag{46}$$

$$K_1^7 = \frac{[TP_{02}^1]}{[T][P_{02}]} \tag{47}$$

Again, since nothing else has changed in the thermodynamics, K_1^j computed from (21) are the same as before for $j \in \{4, 5, 6, 7\}$, and we have the following conservation rules:

$$[P_{12}]_0 = [P_{12}] + [TP_{12}^2] + [TP_{12}^1] \tag{48}$$

$$[P_{02}]_0 = [P_{02}] + [TP_{02}^2] + [TP_{02}^1] \tag{49}$$

$$\begin{aligned}
[T]_0 &= [T] + [TP_{12}^2] + [TP_{02}^2] \\
&\quad + [TP_{12}^1] + [TP_{02}^1] \\
&= [T] + ([P_{12}]_0 - [P_{12}]) + ([P_{02}]_0 - [P_{02}])
\end{aligned} \tag{50}$$

In this case, we also have

- 7 equations: (44)–(47), (48), (49), and (50), in

- 7 unknowns: $[T]$, $[TP_{12}^2]$, $[TP_{12}^1]$, $[TP_{02}^2]$, $[TP_{02}^1]$, and $[P_{12}]$, $[P_{02}]$, with

- 3 free parameters $[P_{12}]_0$, $[P_{02}]_0$, and $[T]_0$.

As above (section III B), the parameters will 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, and then solve for the unknowns in all three setups:

- Full Model,

- Model I, and

- Model II.

With these computed ratio values, we are ready to evaluate and compare the models in order to discern the effects of competition:

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

and

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

IV. CHANGE OF VARIABLES

A. Full Model

In order to simplify the algebraic system of equations, we rename the unknown variables as follows (see (10)):

$$\begin{aligned}
X_1 &= [T] \\
X_2 &= [TP_{11}^1] & X_6 &= [TP_{12}^1] & Y_1 &= [P_{11}] \\
X_3 &= [TP_{01}^1] & X_7 &= [TP_{02}^1] & Y_2 &= [P_{01}] \\
X_4 &= [TP_{12}^2] & X_8 &= [TP_{11}^2] & Y_3 &= [P_{12}] \\
X_5 &= [TP_{02}^2] & X_9 &= [TP_{01}^2] & Y_4 &= [P_{02}]
\end{aligned}$$

The constant parameters in the system are initially left in their symbolic forms.

$$\begin{aligned}
&K_1^2, \quad K_1^3, \quad K_1^4, \quad K_1^5, \quad K_1^6, \quad K_1^7, \quad K_1^8, \quad K_1^9, \\
&a_0 = [P_{11}]_0, \quad b_0 = [P_{01}]_0, \\
&c_0 = [P_{12}]_0, \quad d_0 = [P_{02}]_0, \\
&e_0 = [T]_0.
\end{aligned}$$

Equations (12)–(19) and (22)–(26) can now be rewritten in terms of $\{X_i, Y_j\}$ as follows.

$$\left.\begin{aligned}
[TP_{11}^1] &= K_1^2[T][P_{11}] \\
\implies X_2 &= K_1^2 X_1 Y_1 \\
[TP_{01}^1] &= K_1^3[T][P_{01}] \\
\implies X_3 &= K_1^3 X_1 Y_2 \\
[TP_{12}^2] &= K_1^4[T][P_{12}] \\
\implies X_4 &= K_1^4 X_1 Y_3 \\
[TP_{02}^2] &= K_1^5[T][P_{02}] \\
\implies X_5 &= K_1^5 X_1 Y_4 \\
[TP_{12}^1] &= K_1^6[T][P_{12}] \\
\implies X_6 &= K_1^6 X_1 Y_3 \\
[TP_{02}^1] &= K_1^7[T][P_{02}] \\
\implies X_7 &= K_1^7 X_1 Y_4 \\
[TP_{11}^2] &= K_1^8[T][P_{11}] \\
\implies X_8 &= K_1^8 X_1 Y_1 \\
[TP_{01}^2] &= K_1^9[T][P_{01}] \\
\implies X_9 &= K_1^9 X_1 Y_2 \\
[P_{11}]_0 &= [P_{11}] + [TP_{11}^1] + [TP_{11}^2] \\
\implies a_0 &= X_2 + X_8 + Y_1 \\
[P_{01}]_0 &= [P_{01}] + [TP_{01}^1] + [TP_{01}^2] \\
\implies b_0 &= X_3 + X_9 + Y_2 \\
[P_{12}]_0 &= [P_{12}] + [TP_{12}^1] + [TP_{12}^2] \\
\implies c_0 &= X_4 + X_6 + Y_3 \\
[P_{02}]_0 &= [P_{02}] + [TP_{02}^1] + [TP_{02}^2] \\
\implies d_0 &= X_5 + X_7 + Y_4 \\
[T]_0 &= [T] + [TP_{11}^1] + [TP_{01}^1] \\
&\quad + [TP_{12}^2] + [TP_{02}^2] \\
&\quad + [TP_{11}^2] + [TP_{01}^2] \\
&\quad + [TP_{12}^1] + [TP_{02}^1] \\
\implies e_0 &= X_1 + X_2 + X_3 + X_4 + X_5 \\
&\quad + X_6 + X_7 + X_8 + X_9
\end{aligned}\right\} \tag{51}$$

B. Model I

Now, we consider the system of algebraic equations representing the concentrations at equilibrium and involving unknown 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. Thus, in a manner analogous to that derived for the full model in the previous section, we may rewrite the equations (32), (33), (34), (35), (36), (37), and (38) in terms of $\{X_i, Y_j\}$, as shown below.

$$\left.\begin{aligned}
X_2 &= K_1^2 X_1 Y_1 \\
X_3 &= K_1^3 X_1 Y_2 \\
X_8 &= K_1^8 X_1 Y_1 \\
X_9 &= K_1^9 X_1 Y_2 \\
a_0 &= [P_{11}]_0 = X_2 + X_8 + Y_1 \\
b_0 &= [P_{01}]_0 = X_3 + X_9 + Y_2 \\
e_0 &= [T]_0 = X_1 + X_2 + X_3 + X_8 + X_9
\end{aligned}\right\} \quad (52)$$

C. Model II

Next, we consider the system of algebraic equations representing the concentrations at equilibrium and involving unknown 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. Once again we may rewrite the equations (44), (45), (46), (47), (48), (49), and (50) in terms of $\{X_i, Y_j\}$, as shown below.

$$\left.\begin{aligned}
X_4 &= K_1^4 X_1 Y_3 \\
X_5 &= K_1^5 X_1 Y_4 \\
X_6 &= K_1^6 X_1 Y_3 \\
X_7 &= K_1^7 X_1 Y_4 \\
c_0 &= [P_{12}]_0 = X_4 + X_6 + Y_3 \\
d_0 &= [P_{02}]_0 = X_5 + X_7 + Y_4 \\
e_0 &= [T]_0 = X_1 + X_4 + X_5 + X_6 + X_7
\end{aligned}\right\} \quad (53)$$

Note that with the exception of the conservation rules for $[T]$ (i.e., the last equations in (51), (52), and (53)) under the different models, we have

$$(51) = (52) \cup (53).$$

V. SYSTEM REDUCTION

A. Model I

Starting with (52), we may obtain the following linear equalities:

$$Y_1 = a_0 - X_2 - X_8 \qquad (54)$$
$$Y_2 = b_0 - X_3 - X_9 \qquad (55)$$

Furthermore, since

$$\frac{X_2}{X_8} = \frac{K_1^2 X_1 Y_1}{K_1^8 X_1 Y_1} = \frac{K_1^2}{K_1^8} \implies X_8 = \frac{K_1^8}{K_1^2} X_2 \quad (56)$$

$$\frac{X_3}{X_9} = \frac{K_1^3 X_1 Y_2}{K_1^9 X_1 Y_2} = \frac{K_1^3}{K_1^9} \implies X_9 = \frac{K_1^9}{K_1^3} X_3 \quad (57)$$

we may simplify to obtain

$$\begin{aligned}
X_2 &= K_1^2 X_1 Y_1 = K_1^2 X_1 (a_0 - X_2 - X_8) \\
&= K_1^2 X_1 \left(a_0 - X_2 - \frac{K_1^8}{K_1^2} X_2 \right) \\
&= K_1^2 X_1 \left(a_0 - X_2 \left[1 + \frac{K_1^8}{K_1^2} \right] \right) \\
&= K_1^2 X_1 a_0 - K_1^2 X_1 X_2 \left[1 + \frac{K_1^8}{K_1^2} \right] \\
&= K_1^2 X_1 a_0 - X_1 X_2 \left[K_1^2 + K_1^8 \right] \\
X_2 &+ X_1 X_2 \left[K_1^2 + K_1^8 \right] = a_0 K_1^2 X_1 \\
&\therefore \boxed{X_2 = \frac{a_0 K_1^2 X_1}{1 + X_1 \left(K_1^2 + K_1^8 \right)}}
\end{aligned} \quad (58)$$

and

$$\begin{aligned}
X_3 &= K_1^3 X_1 Y_2 = K_1^3 X_1 (b_0 - X_3 - X_9) \\
&= K_1^3 X_1 \left(b_0 - X_3 - \frac{K_1^9}{K_1^3} X_3 \right) \\
&= K_1^3 X_1 \left(b_0 - X_3 \left[1 + \frac{K_1^9}{K_1^3} \right] \right) \\
&= K_1^3 X_1 b_0 - K_1^3 X_1 X_3 \left[1 + \frac{K_1^9}{K_1^3} \right] \\
&= K_1^3 X_1 b_0 - X_1 X_3 \left[K_1^3 + K_1^9 \right] \\
X_3 &+ X_1 X_3 \left[K_1^3 + K_1^9 \right] = b_0 K_1^3 X_1 \\
&\therefore \boxed{X_3 = \frac{b_0 K_1^3 X_1}{1 + X_1 \left(K_1^3 + K_1^9 \right)}}
\end{aligned} \quad (59)$$

We also obtain, from (56) and (58),

$$\begin{aligned}
X_8 &= \frac{K_1^8}{K_1^2} X_2 = \frac{K_1^8}{K_1^2} \frac{a_0 K_1^2 X_1}{1 + X_1 \left(K_1^2 + K_1^8 \right)} \\
&= \boxed{\frac{a_0 K_1^8 X_1}{1 + X_1 \left(K_1^2 + K_1^8 \right)} = X_8}
\end{aligned} \quad (60)$$

and from (57) and (59),

$$\begin{aligned}
X_9 &= \frac{K_1^9}{K_1^3} X_3 = \frac{K_1^9}{K_1^3} \frac{b_0 K_1^3 X_1}{1 + X_1 \left(K_1^3 + K_1^9 \right)} \\
&= \boxed{\frac{b_0 K_1^9 X_1}{1 + X_1 \left(K_1^3 + K_1^9 \right)} = X_9}
\end{aligned} \quad (61)$$

Thus, equations (58), (59), (60), and (61) express X_2, X_3, X_8, and X_9, respectively, in terms of X_1.

Now, from (54), (56), and (58), we derive

$$
\begin{aligned}
Y_1 &= a_0 - X_2 - X_8 = a_0 - X_2\left(1 + \frac{K_1^8}{K_1^2}\right) \\
&= a_0 - \frac{a_0 K_1^2 X_1}{1 + X_1\left(K_1^2 + K_1^8\right)}\left(1 + \frac{K_1^8}{K_1^2}\right) \\
&= a_0 - \frac{a_0 X_1\left(K_1^2 + K_1^8\right)}{1 + X_1\left(K_1^2 + K_1^8\right)} \\
&= a_0\left[\frac{1 + X_1\left(K_1^2 + K_1^8\right) - X_1\left(K_1^2 + K_1^8\right)}{1 + X_1\left(K_1^2 + K_1^8\right)}\right] \\
&= \frac{a_0}{1 + X_1\left(K_1^2 + K_1^8\right)} \\
\therefore \quad & \boxed{Y_1 = \frac{a_0}{1 + X_1\left(K_1^2 + K_1^8\right)}}
\end{aligned}
\tag{62}
$$

Similarly, we derive

$$
\begin{aligned}
Y_2 &= b_0 - X_3 - X_9 = b_0 - X_3\left(1 + \frac{K_1^9}{K_1^3}\right) \\
&= b_0 - \frac{b_0 K_1^3 X_1}{1 + X_1\left(K_1^3 + K_1^9\right)}\left(1 + \frac{K_1^9}{K_1^3}\right) \\
&= b_0 - \frac{b_0 X_1\left(K_1^3 + K_1^9\right)}{1 + X_1\left(K_1^3 + K_1^9\right)} \\
&= b_0\left[\frac{1 + X_1\left(K_1^3 + K_1^9\right) - X_1\left(K_1^3 + K_1^9\right)}{1 + X_1\left(K_1^3 + K_1^9\right)}\right] \\
&= \frac{b_0}{1 + X_1\left(K_1^3 + K_1^9\right)} \\
\therefore \quad & \boxed{Y_2 = \frac{b_0}{1 + X_1\left(K_1^3 + K_1^9\right)}}
\end{aligned}
\tag{63}
$$

Thus, equations (62) and (63) express Y_1 and Y_2 in terms of X_1.

A final simplification yields a univariate rational function only in X_1 equating to a constant e_0:

$$
\begin{aligned}
e_0 &= X_1 + X_2 + X_3 + X_8 + X_9 \quad \text{(by (52))} \\
&= X_1 + X_1\frac{a_0 K_1^2}{1 + X_1\left(K_1^2 + K_1^8\right)} \\
&\quad + X_1\frac{b_0 K_1^3}{1 + X_1\left(K_1^3 + K_1^9\right)} \quad \text{(by (58),(59))} \\
&\quad + X_1\frac{a_0 K_1^8}{1 + X_1\left(K_1^2 + K_1^8\right)} \\
&\quad + X_1\frac{b_0 K_1^9}{1 + X_1\left(K_1^3 + K_1^9\right)} \quad \text{(by (60),(61))}
\end{aligned}
$$

or

$$
\begin{aligned}
e_0 &= X_1\Bigg[1 + a_0\frac{K_1^2 + K_1^8}{1 + X_1\left(K_1^2 + K_1^8\right)} \\
&\quad + b_0\frac{K_1^3 + K_1^9}{1 + X_1\left(K_1^3 + K_1^9\right)}\Bigg]
\end{aligned}
\tag{64}
$$

Since the terms $\left(K_1^2 + K_1^8\right)$ and $\left(K_1^3 + K_1^9\right)$ appear frequently, in order to express the preceding equations in a simpler form, we introduce short-hand notations shown below. Let

$$
s_{28} \equiv K_1^2 + K_1^8, \quad s_{39} \equiv K_1^3 + K_1^9, \quad \text{and } x \equiv X_1.
$$

In the simplified from, the equation (64) becomes

$$
x\left(1 + a_0\frac{s_{28}}{1 + s_{28}x} + b_0\frac{s_{39}}{1 + s_{39}x}\right) = e_0
\tag{65}
$$

$$
x\left(\frac{(1 + s_{28}x)(1 + s_{39}x) + a_0 s_{28}(1 + s_{39}x) + b_0 s_{39}(1 + s_{28}x)}{(1 + s_{28}x)(1 + s_{39}x)}\right) = e_0
\tag{66}
$$

$$
x\left((1 + s_{28}x)(1 + s_{39}x) + a_0 s_{28}(1 + s_{39}x) + b_0 s_{39}(1 + s_{28}x)\right) = e_0(1 + s_{28}x)(1 + s_{39}x),
\tag{67}
$$

or

$$
(s_{28}s_{39})x^3 + (s_{28} + s_{39} + s_{28}s_{39}[a_0 + b_0 - e_0])x^2 + (1 + s_{28}[a_0 - e_0] + s_{39}[b_0 - e_0])x - e_0 = 0
\tag{68}
$$

Now the cubic polynomial equation (68) must be solved for the unknown $x = X_1$, and then the solution can be substituted into (58)–(63) in order to solve for the rest of the variables. We may obtain the solutions in their symbolic form using Mathematica ([4]) as the three possible roots may be easily expressed in radicals. More to the

point, we only need to solve for

$$
\boxed{\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{I}}} = \left(\frac{[TP_{11}^1] + [TP_{11}^2]}{[TP_{01}^1] + [TP_{01}^2]}\right)_{\text{I}} = \frac{X_2 + X_8}{X_3 + X_9}
$$

$$
= \left(\frac{a_0 K_1^2 x}{1 + s_{28}x} + \frac{a_0 K_1^8 x}{1 + s_{28}x}\right)\Bigg/\left(\frac{b_0 K_1^3 x}{1 + s_{39}x} + \frac{b_0 K_1^9 x}{1 + s_{39}x}\right)
$$

or

$$\left(\frac{TP_{11}}{TP_{01}}\right)_{\mathrm{I}} = \left(\frac{a_0 s_{28} x}{1 + s_{28} x}\right) \Big/ \left(\frac{b_0 s_{39} x}{1 + s_{39} x}\right)$$

$$= \boxed{\frac{a_0}{b_0} \frac{s_{28}}{s_{39}} \frac{1 + s_{39} x}{1 + s_{28} x}} \tag{69}$$

$$= \frac{a_0}{b_0} \frac{s_{28}}{s_{39}} \frac{s_{39} + 1/x}{s_{28} + 1/x}$$

where x is a solution of (68).

B. Model II

As before, starting with (53), we may obtain the following linear equalities:

$$Y_3 = c_0 - X_4 - X_6 \tag{70}$$
$$Y_4 = d_0 - X_5 - X_7 \tag{71}$$

Since

$$\frac{X_4}{X_6} = \frac{K_1^4 X_1 Y_3}{K_1^6 X_1 Y_3} = \frac{K_1^4}{K_1^6} \implies X_6 = \frac{K_1^6}{K_1^4} X_4 \tag{72}$$

$$\frac{X_5}{X_7} = \frac{K_1^5 X_1 Y_4}{K_1^7 X_1 Y_4} = \frac{K_1^5}{K_1^7} \implies X_7 = \frac{K_1^7}{K_1^5} X_5 \tag{73}$$

we obtain

$$X_4 = K_1^4 X_1 Y_3 = K_1^4 X_1 \left(c_0 - X_4 \left[1 + \frac{K_1^6}{K_1^4}\right]\right)$$

$$= K_1^4 X_1 c_0 - X_1 X_4 \left(K_1^4 + K_1^6\right)$$

$$\therefore \boxed{X_4 = \frac{c_0 K_1^4 X_1}{1 + X_1 \left(K_1^4 + K_1^6\right)}} \tag{74}$$

and

$$X_5 = K_1^5 X_1 Y_4 = K_1^5 X_1 \left(d_0 - X_5 \left[1 + \frac{K_1^7}{K_1^5}\right]\right)$$

$$= K_1^5 X_1 d_0 - X_1 X_5 \left(K_1^5 + K_1^7\right)$$

$$\therefore \boxed{X_5 = \frac{d_0 K_1^5 X_1}{1 + X_1 \left(K_1^5 + K_1^7\right)}} \tag{75}$$

Furthermore, from (72) and (74), we obtain

$$X_6 = \frac{K_1^6}{K_1^4} X_4 = \frac{K_1^6}{K_1^4} \frac{c_0 K_1^4 X_1}{1 + X_1 \left(K_1^4 + K_1^6\right)}$$

$$= \boxed{\frac{c_0 K_1^6 X_1}{1 + X_1 \left(K_1^4 + K_1^6\right)} = X_6} \tag{76}$$

and from (73) and (75),

$$X_7 = \frac{K_1^7}{K_1^5} X_5 = \frac{K_1^7}{K_1^5} \frac{d_0 K_1^5 X_1}{1 + X_1 \left(K_1^5 + K_1^7\right)}$$

$$= \boxed{\frac{d_0 K_1^7 X_1}{1 + X_1 \left(K_1^5 + K_1^7\right)} = X_7} \tag{77}$$

Thus, equations (74), (75), (76), and (77) express X_4, X_5, X_6, and X_7, respectively, in terms of X_1.

From (70), (72), and (74), we derive

$$Y_3 = c_0 - X_4 - X_6 = c_0 - X_4 \left(1 + \frac{K_1^6}{K_1^4}\right)$$

$$= c_0 - \frac{c_0 K_1^4 X_1}{1 + X_1 \left(K_1^4 + K_1^6\right)} \left(1 + \frac{K_1^6}{K_1^4}\right)$$

$$= c_0 - \frac{c_0 X_1 \left(K_1^4 + K_1^6\right)}{1 + X_1 \left(K_1^4 + K_1^6\right)}$$

$$= c_0 \left[\frac{1 + X_1 \left(K_1^4 + K_1^6\right) - X_1 \left(K_1^4 + K_1^6\right)}{1 + X_1 \left(K_1^4 + K_1^6\right)}\right]$$

$$= \frac{c_0}{1 + X_1 \left(K_1^4 + K_1^6\right)}$$

$$\therefore \boxed{Y_3 = \frac{c_0}{1 + X_1 \left(K_1^4 + K_1^6\right)}} \tag{78}$$

Similarly, we derive

$$Y_4 = d_0 - X_5 - X_7 = d_0 - X_5 \left(1 + \frac{K_1^7}{K_1^5}\right)$$

$$= d_0 - \frac{d_0 K_1^5 X_1}{1 + X_1 \left(K_1^5 + K_1^7\right)} \left(1 + \frac{K_1^7}{K_1^5}\right)$$

$$= d_0 - \frac{d_0 X_1 \left(K_1^5 + K_1^7\right)}{1 + X_1 \left(K_1^5 + K_1^7\right)}$$

$$= d_0 \left[\frac{1 + X_1 \left(K_1^5 + K_1^7\right) - X_1 \left(K_1^5 + K_1^7\right)}{1 + X_1 \left(K_1^5 + K_1^7\right)}\right]$$

$$= \frac{d_0}{1 + X_1 \left(K_1^5 + K_1^7\right)}$$

$$\therefore \boxed{Y_4 = \frac{d_0}{1 + X_1 \left(K_1^5 + K_1^7\right)}} \tag{79}$$

Putting it all together, we derive the univariate rational equation for X_1.

$$e_0 = X_1 + X_4 + X_5 + X_6 + X_7 \qquad \text{(by (53))}$$

$$= X_1 + X_1 \frac{c_0 K_1^4}{1 + X_1 \left(K_1^4 + K_1^6\right)}$$

$$+ X_1 \frac{d_0 K_1^5}{1 + X_1 \left(K_1^5 + K_1^7\right)} \quad \text{(by (74),(75))}$$

$$+ X_1 \frac{c_0 K_1^6}{1 + X_1 \left(K_1^4 + K_1^6\right)}$$

$$+ X_1 \frac{d_0 K_1^7}{1 + X_1 \left(K_1^5 + K_1^7\right)} \quad \text{(by (76),(77))}$$

or

$$e_0 = X_1 \left[1 + c_0 \frac{K_1^4 + K_1^6}{1 + X_1 \left(K_1^4 + K_1^6\right)} \right.$$

$$\left. + d_0 \frac{K_1^5 + K_1^7}{1 + X_1 \left(K_1^5 + K_1^7\right)}\right] \tag{80}$$

As before, we abbreviate the terms $(K_1^4 + K_1^6)$ and $(K_1^5 + K_1^7)$ by short-hand notation, shown below. Let

$$s_{46} \equiv K_1^4 + K_1^6, \quad s_{57} \equiv K_1^5 + K_1^7, \quad \text{and} \quad y \equiv X_1.$$

Note that, in order to avoid confusion, we have intro-

duced a different abbreviation for X_1 (i.e., y) intentionally since the equation to be solved in this case differs from (68). Then (80) can be expressed as

$$y \left(1 + c_0 \frac{s_{46}}{1 + s_{46}y} + d_0 \frac{s_{57}}{1 + s_{57}y}\right) = e_0$$

$$y \left(\frac{(1 + s_{46}y)(1 + s_{57}y) + c_0 s_{46}(1 + s_{57}y) + d_0 s_{57}(1 + s_{46}y)}{(1 + s_{46}y)(1 + s_{57}y)}\right) = e_0$$

$$y \left((1 + s_{46}y)(1 + s_{57}y) + c_0 s_{46}(1 + s_{57}y) + d_0 s_{57}(1 + s_{46}y)\right) = e_0(1 + s_{46}y)(1 + s_{57}y),$$

or

$$(s_{46}s_{57})y^3 + (s_{46} + s_{57} + s_{46}s_{57}[c_0 + d_0 - e_0]) y^2 + (1 + s_{46}[c_0 - e_0] + s_{57}[d_0 - e_0]) y - e_0 = 0 \tag{81}$$

Again, the cubic polynomial equation (81) must be solved for $y = X_1$, and then the solution can be substituted into (74)–(79) for the rest of the variables. Actually, we only need

$$\boxed{\left(\frac{TP_{12}}{TP_{02}}\right)_{\mathrm{II}}} = \left(\frac{[TP_{12}^2] + [TP_{12}^1]}{[TP_{02}^2] + [TP_{02}^1]}\right)_{\mathrm{II}} = \frac{X_4 + X_6}{X_5 + X_7}$$

$$= \left(\frac{c_0 K_1^4 y}{1 + s_{46}y} + \frac{c_0 K_1^6 y}{1 + s_{46}y}\right) \bigg/ \left(\frac{d_0 K_1^5 y}{1 + s_{57}y} + \frac{d_0 K_1^7 y}{1 + s_{57}y}\right)$$

or

$$\left(\frac{TP_{12}}{TP_{02}}\right)_{\mathrm{II}} = \left(\frac{c_0 s_{46}y}{1 + s_{46}y}\right) \bigg/ \left(\frac{d_0 s_{57}y}{1 + s_{57}y}\right)$$

$$= \boxed{\frac{c_0}{d_0} \frac{s_{46}}{s_{57}} \frac{1 + s_{57}y}{1 + s_{46}y}} \tag{82}$$

$$= \frac{c_0}{d_0} \frac{s_{46}}{s_{57}} \frac{s_{57} + 1/y}{s_{46} + 1/y}$$

where y solves (81).

C. Full Model

As noted in section IV, the system (51) of equations for the Full Model is simply the union of the systems (52) and (53) for models I and II, respectively, with the exception of the conservation rule for $[T]$, i.e., the equation for X_1. Therefore, while the equation for X_1 itself must be handled separately, the derivations from sections V A and V B can be duplicated to obtain equations for all the variables in terms of X_1. For convenience, we gather the

resulting equations in one place, as shown below.

$$\boxed{X_2 = \frac{a_0 K_1^2 X_1}{1 + X_1 \left(K_1^2 + K_1^8\right)}} \qquad \text{(see (58))} \quad (83)$$

$$\boxed{X_3 = \frac{b_0 K_1^3 X_1}{1 + X_1 \left(K_1^3 + K_1^9\right)}} \qquad \text{(see (59))} \quad (84)$$

$$\boxed{X_4 = \frac{c_0 K_1^4 X_1}{1 + X_1 \left(K_1^4 + K_1^6\right)}} \qquad \text{(see (74))} \quad (85)$$

$$\boxed{X_5 = \frac{d_0 K_1^5 X_1}{1 + X_1 \left(K_1^5 + K_1^7\right)}} \qquad \text{(see (75))} \quad (86)$$

$$\boxed{X_6 = \frac{c_0 K_1^6 X_1}{1 + X_1 \left(K_1^4 + K_1^6\right)}} \qquad \text{(see (76))} \quad (87)$$

$$\boxed{X_7 = \frac{d_0 K_1^7 X_1}{1 + X_1 \left(K_1^5 + K_1^7\right)}} \qquad \text{(see (77))} \quad (88)$$

$$\boxed{X_8 = \frac{a_0 K_1^8 X_1}{1 + X_1 \left(K_1^2 + K_1^8\right)}} \qquad \text{(see (60))} \quad (89)$$

$$\boxed{X_9 = \frac{b_0 K_1^9 X_1}{1 + X_1 \left(K_1^3 + K_1^9\right)}} \qquad \text{(see (61))} \quad (90)$$

$$Y_1 = \frac{a_0}{1 + X_1\left(K_1^2 + K_1^8\right)} \qquad \text{(see (62))} \quad (91)$$

$$Y_2 = \frac{b_0}{1 + X_1\left(K_1^3 + K_1^9\right)} \qquad \text{(see (63))} \quad (92)$$

$$Y_3 = \frac{c_0}{1 + X_1\left(K_1^4 + K_1^6\right)} \qquad \text{(see (78))} \quad (93)$$

$$Y_4 = \frac{d_0}{1 + X_1\left(K_1^5 + K_1^7\right)} \qquad \text{(see (79))} \quad (94)$$

It remains to derive the univariate equation in X_1. Since the terms $(K_1^2 + K_1^8)$, $(K_1^3 + K_1^9)$, $(K_1^4 + K_1^6)$, and $(K_1^5 + K_1^7)$ appears frequently in the following derivation, as in the previous sections, we abbreviate these terms with the short-hand notation given below. As we did in sections V A, and V B, let

$$s_{28} \equiv K_1^2 + K_1^8, \qquad s_{39} \equiv K_1^3 + K_1^9,$$
$$s_{46} \equiv K_1^4 + K_1^6, \qquad s_{57} \equiv K_1^5 + K_1^7,$$

and let

$$z \equiv X_1.$$

Note again that a different symbol for X_1 has to be employed to avoid confusion with the variables used in equations (68) and (81).

$$
\begin{aligned}
e_0 &= X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 \\
&\quad \text{(by (51))} \\
&= z + z\frac{a_0 K_1^2}{1 + s_{28}z} + z\frac{b_0 K_1^3}{1 + s_{39}z} \\
&\quad + z\frac{c_0 K_1^4}{1 + s_{46}z} + z\frac{d_0 K_1^5}{1 + s_{57}z} \quad \text{(by (83)–(86))} \\
&\quad + z\frac{c_0 K_1^6}{1 + s_{46}z} + z\frac{d_0 K_1^7}{1 + s_{57}z} \\
&\quad + z\frac{a_0 K_1^8}{1 + z s_{28}} + z\frac{b_0 K_1^9}{1 + z s_{39}} \quad \text{(by (87)–(90))} \\
&= z\left[1 + a_0\frac{K_1^2 + K_1^8}{1 + s_{28}z} + b_0\frac{K_1^3 + K_1^9}{1 + s_{39}z} \right. \\
&\quad \left. + c_0\frac{K_1^4 + K_1^6}{1 + s_{46}z} + d_0\frac{K_1^5 + K_1^7}{1 + s_{57}z}\right] \quad (95)
\end{aligned}
$$

or

$$
\begin{aligned}
e_0 &= z\left[1 + \frac{a_0 s_{28}}{1 + s_{28}z} + \frac{b_0 s_{39}}{1 + s_{39}z} \right. \\
&\quad \left. + \frac{c_0 s_{46}}{1 + s_{46}z} + \frac{d_0 s_{57}}{1 + s_{57}z}\right] \quad (96)
\end{aligned}
$$

or

$$
\begin{aligned}
(1 + s_{28}z)(1 + s_{39}z)(1 + s_{46}z)(1 + s_{57}z)e_0 &= \\
= \quad z[(1 + s_{28}z)(1 + s_{39}z)(1 + s_{46}z)(1 + s_{57}z) &+ a_0 s_{28}(1 + s_{39}z)(1 + s_{46}z)(1 + s_{57}z) \\
+ b_0 s_{39}(1 + s_{28}z)(1 + s_{46}z)(1 + s_{57}z) &+ c_0 s_{46}(1 + s_{28}z)(1 + s_{39}z)(1 + s_{57}z) \\
+ d_0 s_{57}(1 + s_{28}z)(1 + s_{39}z)(1 + s_{46}z)] & \quad (97)
\end{aligned}
$$

Since we now have a 5th order polynomial equation in z to solve, and since its roots cannot be expressed symbolically in a closed form, we must resort to a purely numerical approach. Nonetheless, the match-to-mismatch ratio signals can be obtained in terms of z.

$$
\begin{aligned}
\left(\frac{TP_{11}}{TP_{01}}\right)_{\text{full}} &= \left(\frac{[TP_{11}^1] + [TP_{01}^2]}{[TP_{01}^1] + [TP_{01}^2]}\right)_{\text{full}} = \frac{X_2 + X_8}{X_3 + X_9} \\
&= \frac{a_0}{b_0}\frac{s_{28}}{s_{39}}\frac{1 + s_{39}z}{1 + s_{28}z} = \frac{a_0}{b_0}\frac{s_{28}}{s_{39}}\frac{s_{39} + 1/z}{s_{28} + 1/z} \\
&\quad \text{(see (69))} \quad (98)
\end{aligned}
$$

and

$$
\begin{aligned}
\left(\frac{TP_{12}}{TP_{02}}\right)_{\text{full}} &= \left(\frac{[TP_{12}^2] + [TP_{12}^1]}{[TP_{02}^2] + [TP_{02}^1]}\right)_{\text{full}} = \frac{X_4 + X_6}{X_5 + X_7} \\
&= \frac{c_0}{d_0}\frac{s_{46}}{s_{57}}\frac{1 + s_{57}z}{1 + s_{46}z} = \frac{c_0}{d_0}\frac{s_{46}}{s_{57}}\frac{s_{57} + 1/z}{s_{46} + 1/z} \\
&\quad \text{(see (82))}, \quad (99)
\end{aligned}
$$

where z solves (97).

VI. ADDITIONAL MODELS

Next, for the purpose of comparison, we will 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

We consider a situation where the target has exactly one region for the probe to hybridize with. Thus, we have three possible states to model: unbound targets, targets bound to "match" probes in the region of interest, and lastly, targets bound to "mismatch" probes in the region of interest—all other possible hybridization states are ignored.

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

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

1. Possible States

We consider the following three states:

(1) T

(Target is unbound.)

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

(Target is bound by "specific" hybridization.)

2. State Transition Diagram

(3) $TP_{01}^1 \longleftrightarrow T$ (1) $\longleftrightarrow TP_{11}^1$ (2)

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

$$T + P_{11} \underset{k_{2,1}}{\overset{k_{1,2}}{\rightleftharpoons}} TP_{11}^1$$
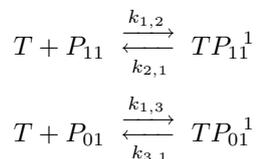
$$T + P_{01} \underset{k_{3,1}}{\overset{k_{1,3}}{\rightleftharpoons}} TP_{01}^1$$

3. Dynamics

The following are the ODE's (ordinary differential equations) describing the dynamics of the system.

$$\frac{d[T]}{dt} = k_{2,1}[TP_{11}^1] - k_{1,2}[T][P_{11}]$$
$$+ k_{3,1}[TP_{01}^1] - k_{1,3}[T][P_{01}] \quad (100)$$

$$\frac{d[TP_{11}^1]}{dt} = k_{1,2}[T][P_{11}] - k_{2,1}[TP_{11}^1] \quad (101)$$

$$\frac{d[TP_{01}^1]}{dt} = k_{1,3}[T][P_{01}] - k_{3,1}[TP_{01}^1] \quad (102)$$

At equilibrium, once again, $\frac{d[\cdot]}{dt} = 0$ for all substances, which yields:

$$K_1^2 = \frac{[TP_{11}^1]}{[T][P_{11}]}$$

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

We augment the above equations with the linear constraints corresponding to the conservation rules.

$$T: \quad [T] + [TP_{11}^1] + [TP_{01}^1] = [T]_0 = e_0$$
$$P_{11}: \quad [P_{11}] + [TP_{11}^1] = [P_{11}]_0 = a_0$$
$$P_{01}: \quad [P_{01}] + [TP_{01}^1] = [P_{01}]_0 = b_0$$

Finally, we gather the system of equations to be solved, with the appropriate change of variables.

$$
\begin{array}{ll}
X_1 = [T] & X_1 + X_2 + X_3 = e_0 \\
X_2 = [TP_{11}^1] & X_2 = K_1^2 X_1 Y_1 \\
X_3 = [TP_{01}^1] & X_3 = K_1^3 X_1 Y_2 \\
Y_1 = [P_{11}] & X_2 + Y_1 = a_0 \\
Y_2 = [P_{01}] & X_3 + Y_2 = b_0
\end{array}
$$

After simplification, we have

$$X_2 = K_1^2 X_1 (a_0 - X_2)$$
$$= K_1^2 X_1 a_0 - K_1^2 X_1 X_2$$
$$\implies \boxed{X_2 = \frac{a_0 K_1^2 X_1}{1 + K_1^2 X_1}}$$

$$X_3 = K_1^3 X_1 (b_0 - X_3)$$
$$= K_1^3 X_1 b_0 - K_1^3 X_1 X_3$$
$$\implies \boxed{X_3 = \frac{b_0 K_1^3 X_1}{1 + K_1^3 X_1}}$$

Analogously,

$$\boxed{Y_1 = \frac{a_0}{1 + K_1^2 X_1}}$$

$$\boxed{Y_2 = \frac{b_0}{1 + K_1^3 X_1}}$$

Finally, we get the following equation involving rational functions in one variable X_1.

$$X_1 = e_0 - X_2 - X_3 = e_0 - a_0 \frac{K_1^2 X_1}{1 + K_1^2 X_1} - b_0 \frac{K_1^3 X_1}{1 + K_1^3 X_1} \tag{103}$$

Simplifying equation (103) for X_1, we have the following equation with $w \equiv X_1$.

$$
\begin{aligned}
e_0 &= w + a_0 \frac{K_1^2}{1 + K_1^2 w} w + b_0 \frac{K_1^3}{1 + K_1^3 w} w \\
&= \boxed{w \left(1 + a_0 \frac{K_1^2}{1 + K_1^2 w} + b_0 \frac{K_1^3}{1 + K_1^3 w} \right) = e_0}
\end{aligned} \tag{104}
$$

We may solve (104) for w numerically or symbolically (e.g., in Mathematica). Writing the "match-to-mismatch ratio" in terms of the roots of the above equation, we get

$$
\begin{aligned}
\left(\frac{TP_{11}}{TP_{01}} \right)_{\text{simp}} &= \frac{X_2}{X_3} = \frac{a_0 K_1^2 X_1}{1 + K_1^2 X_1} \frac{1 + K_1^3 X_1}{b_0 K_1^3 X_1} \\
&= \frac{a_0 K_1^2}{K_1^2 + \frac{1}{X_1}} \frac{K_1^3 + \frac{1}{X_1}}{b_0 K_1^3} \\
&= \frac{a_0}{b_0} \cdot \frac{K_1^2}{K_1^3} \cdot \frac{K_1^3 + \frac{1}{X_1}}{K_1^2 + \frac{1}{X_1}} \tag{105} \\
&= \frac{a_0}{b_0} \cdot \frac{K_1^2}{K_1^3} \cdot \frac{1 + \frac{1}{X_1 K_1^3}}{\frac{K_1^2}{K_1^3} + \frac{1}{X_1 K_1^3}}. \tag{106}
\end{aligned}
$$

According to (105), if $X_1 \gg 1$ then we have ratio $\equiv (a_0/b_0)$. On the other hand, if $K_1^2/K_1^3 \sim \frac{1}{X_1 K_1^3}$, i.e., $K_1^2 \sim \frac{1}{X_1}$, then the ratio simplifies to the following, indicating that the ratio depends on the initial concentration of the probes and their thermodynamic parameters.

$$
\begin{aligned}
\left(\frac{TP_{11}}{TP_{01}} \right)_{\text{simp}} &\sim \frac{a_0}{b_0} \frac{K_1^2}{K_1^3} \frac{(X_1 K_1^3 + 1)/X_1 K_1^3}{2/X_1 K_1^3} \\
&= \frac{1}{2} \frac{a_0}{b_0} \frac{1}{X_1 K_1^3} \left(X_1 K_1^3 + 1 \right) \tag{107} \\
&= \frac{1}{2} \frac{a_0}{b_0} \left(1 + \frac{1}{X_1 K_1^3} \right) \\
&= \frac{1}{2} \frac{a_0}{b_0} \left(1 + \frac{K_1^2}{K_1^3} \right) \tag{108}
\end{aligned}
$$

We need to further investigate what should be the proper initial target concentration $[T]_0 = e_0$ in order to optimize the observed discrimination (match-to-mismatch ratio) at equilibrium.

- Discrimination is lowest in the presence of excess amounts of target, 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 $X_1 \gg 1$ case discussed

above, yielding ratio $\equiv \frac{a_0}{b_0}$, where $a_0 =$ the initial concentration of the matched probe and $b_0 =$ the initial concentration of the mismatched probe. As discussed in section A 1, these two parameters are usually set 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.

Preferably, multiplexed analysis should thus be carried out under conditions of slight 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 can be used to verify that the effects suggested by pairwise probe analysis extend to probe triples correctly.

In this scheme, we will 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_{11}^1, (3) TP_{01}^1, (4) TP_{12}^2,
(5) TP_{02}^2, (6) TP_{13}^3, (7) TP_{03}^3
(Target is bound by "specific" hybridization;
   P_{ij} hybridizes to site j.)
(8) TP_{11}^2, (9) TP_{01}^2, (10) TP_{11}^3,
(11) TP_{01}^3, (12) TP_{12}^1, (13) TP_{02}^1,
(14) TP_{12}^3, (15) TP_{02}^3, (16) TP_{13}^1,
(17) TP_{03}^1, (18) TP_{13}^2, (19) TP_{03}^2
(Target is bound by "cross-hybridization";
   P_{ij} hybridizes to site k, $j \neq k$.)

2. State Transition Diagram

The state transition diagram for this model is not shown, as it involves 19 states and is cumbersome to display. The state interaction can be easily inferred from (109).

The set of reversible reactions operating between unbound and bound states can be written as shown below, where K_i^j denotes the affinity constant for going from state $i(=1)$ to state $j \in \{2, 3, \ldots, 19\}$.

$$
\begin{cases}
T + P_{11} \rightleftarrows TP_{11}^k \\
T + P_{01} \rightleftarrows TP_{01}^k \\
T + P_{12} \rightleftarrows TP_{12}^k \\
T + P_{02} \rightleftarrows TP_{02}^k \\
T + P_{13} \rightleftarrows TP_{13}^k \\
T + P_{03} \rightleftarrows TP_{03}^k
\end{cases}
$$

Site k	1	2	3
States l:	(2)	(8)	(10)
	(3)	(9)	(11)
	(12)	(4)	(14)
	(13)	(5)	(15)
	(16)	(18)	(6)
	(17)	(19)	(7)

$$(109)$$
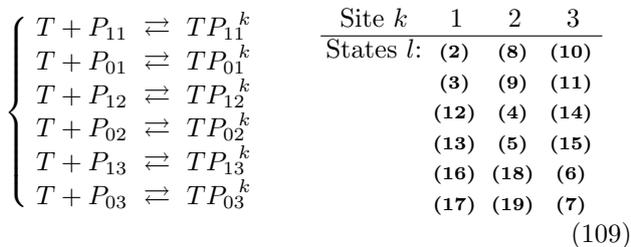
3. Dynamics

The following ODE's describe the dynamics of the system.

$$\frac{d[TP_{ij}^k]}{dt} = k_{1,l}[T][P_{ij}] - k_{l,1}[TP_{ij}^k] \qquad (110)$$

$$\text{(18 eqns)}$$

where

$$i \in \{0, 1\}, \ \text{probe } j \in \{1, 2, 3\},$$
$$\text{site } k \in \{1, 2, 3\}, \ \text{and } l(i, j, k) \text{ is given in (109)}.$$

$$
\begin{aligned}
\frac{d[T]}{dt} &= -\sum_{i=0}^{1}\sum_{j=1}^{3}\sum_{k=1}^{3} k_{1,l(i,j,k)}[T][P_{ij}] + \sum_{i,j,k} k_{l,1}[TP_{ij}^k] \\
&= \sum_{i=0}^{1}\sum_{j=1}^{3}\sum_{k=1}^{3} \left\{ k_{l,1}[TP_{ij}^k] - k_{1,l}[T][P_{ij}] \right\} \qquad (111)
\end{aligned}
$$

At equilibrium, $\frac{d\vec{X}}{dt} = \vec{0}$, where $\vec{X} = (X_1, \ldots, X_{19})^T$ and

$$
\begin{aligned}
&X_1 = [T] \\
&X_2 = [TP_{11}^1] \quad X_8 = [TP_{11}^2] \quad X_{14} = [TP_{12}^3] \\
&X_3 = [TP_{01}^1] \quad X_9 = [TP_{01}^2] \quad X_{15} = [TP_{02}^3] \\
&X_4 = [TP_{12}^2] \quad X_{10} = [TP_{11}^3] \quad X_{16} = [TP_{13}^1] \\
&X_5 = [TP_{02}^2] \quad X_{11} = [TP_{01}^3] \quad X_{17} = [TP_{03}^1] \\
&X_6 = [TP_{13}^3] \quad X_{12} = [TP_{12}^1] \quad X_{18} = [TP_{13}^2] \\
&X_7 = [TP_{03}^3] \quad X_{13} = [TP_{02}^1] \quad X_{19} = [TP_{03}^2]
\end{aligned} \qquad (112)
$$

Applying this equilibrium condition to (110) yields

$$k_{1,l}[T][P_{ij}] = k_{l,1}[TP_{ij}^k]$$

or

$$\boxed{K_1^l \equiv \frac{k_{1,l}}{k_{l,1}} = \frac{[TP_{ij}^k]}{[T][P_{ij}]}} \qquad (113)$$

while (111) becomes the sum of the previous 18 equations and thus provides no additional information.

Mass conservation rules add the following linear constraints:

$$[P_{ij}]_0 = [P_{ij}] + \sum_{k=1}^{3} [TP_{ij}^k] \qquad (114)$$

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

$$[T]_0 = [T] + \sum_{i,j,k} [TP_{ij}^k] \qquad (115)$$

Recall the notation for the target and target-probe complex concentrations introduced in (112). For simplification, we rename the remaining variables (probe concentrations):

$$
\begin{aligned}
Y_1 &= [P_{11}] & Y_3 &= [P_{12}] & Y_5 &= [P_{13}] \\
Y_2 &= [P_{01}] & Y_4 &= [P_{02}] & Y_6 &= [P_{03}]
\end{aligned}
$$

and the constant parameters:

$$
\begin{aligned}
&K_1^l, \quad l = 2, \ldots, 19 \\
&a_0 = [P_{11}]_0, \quad b_0 = [P_{01}]_0, \\
&c_0 = [P_{12}]_0, \quad d_0 = [P_{02}]_0, \\
&e_0 = [P_{13}]_0, \quad f_0 = [P_{03}]_0, \\
&g_0 = [T]_0.
\end{aligned}
$$

Finally, we obtain the following simplified equations:

$$
\begin{aligned}
K_1^l[T][P_{ij}] &= [TP_{ij}^k] \\
\implies \quad &\boxed{X_l = K_1^l X_1 Y_n}, \qquad (116) \\
&\text{where } n \text{ depends on } l(i, j, k) \\
Y_n^0 &= Y_n + \sum_{l \in f^{-1}(n)} X_l \\
\implies \quad &\boxed{Y_n = Y_n^0 - \sum_{l \in f^{-1}(n)} X_l} \qquad (117) \\
&\text{(probe conservation)}
\end{aligned}
$$

$$\boxed{X_1^0 = \sum_{l=1}^{19} X_l} \qquad (118)$$

$$\text{(target conservation)}$$

In these equations we have written $f^{-1}(n)$ to denote the set of states involving probe Y_n, so that, according to (109), we have

$$
\begin{aligned}
f^{-1}(1) &= \{2, 8, 10\} & f^{-1}(4) &= \{5, 13, 15\} \\
f^{-1}(2) &= \{3, 9, 11\} & f^{-1}(5) &= \{6, 16, 18\} \\
f^{-1}(3) &= \{4, 12, 14\} & f^{-1}(6) &= \{7, 17, 19\}
\end{aligned}
$$

Reducing the equations further, we get:

$$n = 1 \begin{cases} X_2 = K_1^2 X_1 Y_1 \\ X_8 = K_1^8 X_1 Y_1 = \dfrac{K_1^8}{K_1^2} X_2 \\ X_{10} = \dfrac{K_1^{10}}{K_1^2} X_2 \end{cases} \qquad n = 4 \begin{cases} X_5 = K_1^5 X_1 Y_4 \\ X_{13} = \dfrac{K_1^{13}}{K_1^2} X_5 \\ X_{15} = \dfrac{K_1^{15}}{K_1^2} X_5 \end{cases}$$

$$n = 2 \begin{cases} X_3 = K_1^3 X_1 Y_2 \\ X_9 = \dfrac{K_1^9}{K_1^3} X_3 \\ X_{11} = \dfrac{K_1^{11}}{K_1^3} X_3 \end{cases} \qquad n = 5 \begin{cases} X_6 = K_1^6 X_1 Y_5 \\ X_{16} = \dfrac{K_1^{16}}{K_1^6} X_6 \\ X_{18} = \dfrac{K_1^{18}}{K_1^6} X_6 \end{cases}$$

$$n = 3 \begin{cases} X_4 = K_1^4 X_1 Y_3 \\ X_{12} = \dfrac{K_1^{12}}{K_1^4} X_4 \\ X_{14} = \dfrac{K_1^{14}}{K_1^4} X_4 \end{cases} \qquad n = 6 \begin{cases} X_7 = K_1^7 X_1 Y_6 \\ X_{17} = \dfrac{K_1^{17}}{K_1^7} X_7 \\ X_{19} = \dfrac{K_1^{19}}{K_1^7} X_7 \end{cases}$$

Now, we consider the equations where $l \in \{2, 8, 10\}$ and $n = 1$:

$$\begin{aligned} Y_1 &= Y_1^0 - (X_2 + X_8 + X_{10}) \\ X_2 &= K_1^2 X_1 Y_1 = K_1^2 X_1 \{ Y_1^0 - (\underbrace{X_2 + X_8 + X_{10}}) \} \end{aligned}$$

$$\begin{bmatrix} X_2 + X_8 + X_{10} \\ = X_2 + \frac{K_1^8}{K_1^2} X_2 + \frac{K_1^{10}}{K_1^2} X_2 \\ = \frac{X_2}{K_1^2} [K_1^2 + K_1^8 + K_1^{10}] \\ = X_2 s_{2,8,10} / K_1^2, \\ \text{where } \boxed{s_{i,j,k} = K_1^i + K_1^j + K_1^k} \end{bmatrix}$$

$$\begin{aligned} &= K_1^2 X_1 Y_1^0 - K_1^2 X_1 \frac{X_2}{K_1^2} s_{2,8,10} \\ &= K_1^2 X_1 Y_1^0 - X_1 X_2 s_{2,8,10} \\ X_2 + X_1 X_2 s_{2,8,10} &= K_1^2 X_1 Y_1^0 \\ \implies X_2 (1 + X_1 s_{2,8,10}) &= K_1^2 X_1 Y_1^0 \\ \therefore X_2 &= K_1^2 Y_1^0 \frac{X_1}{1 + s_{2,8,10} X_1} \end{aligned}$$

Let

$$\boxed{t(2, 8, 10) = \frac{Y_1^0}{1 + s_{2,8,10} X_1}}$$

Then, we have

$$\boxed{X_2 = K_1^2 X_1 t(2, 8, 10)} \tag{119}$$

$$\boxed{X_8 = K_1^8 X_1 t(2, 8, 10)} \tag{120}$$

$$\boxed{X_{10} = K_1^{10} X_1 t(2, 8, 10)} \tag{121}$$

Similarly, we obtain

$$\boxed{\begin{pmatrix} X_3 \\ X_9 \\ X_{11} \end{pmatrix} = \begin{pmatrix} K_1^3 \\ K_1^9 \\ K_1^{11} \end{pmatrix} X_1 t(3, 9, 11)} \tag{122}$$

where $\boxed{t(3, 9, 11) = \dfrac{Y_2^0}{1 + s_{3,9,11} X_1}}$

$$\boxed{\begin{pmatrix} X_4 \\ X_{12} \\ X_{14} \end{pmatrix} = \begin{pmatrix} K_1^4 \\ K_1^{12} \\ K_1^{14} \end{pmatrix} X_1 t(4, 12, 14)} \tag{123}$$

where $\boxed{t(4, 12, 14) = \dfrac{Y_3^0}{1 + s_{4,12,14} X_1}}$

$$\boxed{\begin{pmatrix} X_5 \\ X_{13} \\ X_{15} \end{pmatrix} = \begin{pmatrix} K_1^5 \\ K_1^{13} \\ K_1^{15} \end{pmatrix} X_1 t(5, 13, 15)} \tag{124}$$

where $\boxed{t(5, 13, 15) = \dfrac{Y_4^0}{1 + s_{5,13,15} X_1}}$

$$\boxed{\begin{pmatrix} X_6 \\ X_{16} \\ X_{18} \end{pmatrix} = \begin{pmatrix} K_1^6 \\ K_1^{16} \\ K_1^{18} \end{pmatrix} X_1 t(6, 16, 18)} \tag{125}$$

where $\boxed{t(6, 16, 18) = \dfrac{Y_5^0}{1 + s_{6,16,18} X_1}}$

and

$$\begin{pmatrix} X_7 \\ X_{17} \\ X_{19} \end{pmatrix} = \begin{pmatrix} K_1^7 \\ K_1^{17} \\ K_1^{19} \end{pmatrix} X_1 t(7,17,19) \qquad (126)$$

where $\boxed{t(7,17,19) = \dfrac{Y_6^0}{1 + s_{7,17,19}X_1}}$

After this manipulation we have equations for all X_j's in terms of X_1, $j \neq 1$. Next, we obtain equilibrium probe concentrations:

$$\begin{aligned} Y_1 &= Y_1^0 - (X_2 + X_8 + X_{10}) \\ &= Y_1^0 - s_{2,8,10}Y_1^0 X_1 \frac{1}{1 + s_{2,8,10}X_1} \\ &= Y_1^0 \left\{ 1 - \frac{X_1 s_{2,8,10}}{1 + s_{2,8,10}X_1} \right\} = Y_1^0 \frac{1}{1 + s_{2,8,10}X_1} \\ \therefore \quad & \boxed{Y_1 = \frac{Y_1^0}{1 + s_{2,8,10}X_1} = t(2,8,10)} \qquad (127) \end{aligned}$$

Similarly, we have

$$\boxed{Y_2 = t(3,9,11)} \qquad (128)$$

$$\boxed{Y_3 = t(4,12,14)} \qquad (129)$$

$$\boxed{Y_4 = t(5,13,15)} \qquad (130)$$

$$\boxed{Y_5 = t(6,16,18)} \qquad (131)$$

$$\boxed{Y_6 = t(7,17,19)} \qquad (132)$$

It remains to get the univariate polynomial equation for X_1—"the main equation."

$$\begin{aligned} X_1^0 &= \sum_{l=1}^{19} X_l \\ &= X_1 + (X_2 + X_8 + X_{10}) + (X_3 + X_9 + X_{11}) + (X_4 + X_{12} + X_{14}) \\ &\quad + (X_5 + X_{13} + X_{15}) + (X_6 + X_{16} + X_{18}) + (X_7 + X_{17} + X_{19}) \\ &= X_1 + (K_1^2 + K_1^8 + K_1^{10})X_1 t(2,8,10) + (K_1^3 + K_1^9 + K_1^{11})X_1 t(3,9,11) \\ &\quad + (K_1^4 + K_1^{12} + K_1^{14})X_1 t(4,12,14) + (K_1^5 + K_1^{13} + K_1^{15})X_1 t(5,13,15) \\ &\quad + (K_1^6 + K_1^{16} + K_1^{18})X_1 t(6,16,18) + (K_1^7 + K_1^{17} + K_1^{19})X_1 t(7,17,19) \\ &= X_1 \big(1 + s_{2,8,10}t(2,8,10) + s_{3,9,11}t(3,9,11) + s_{4,12,14}t(4,12,14) \\ &\quad + s_{5,13,15}t(5,13,15) + s_{6,16,18}t(6,16,18) + s_{7,17,19}t(7,17,19) \big) \end{aligned}$$

Therefore,

$$\begin{aligned} X_1^0 = X_1 \bigg\{ &1 + Y_1^0 \frac{s_{2,8,10}}{1 + s_{2,8,10}X_1} + Y_2^0 \frac{s_{3,9,11}}{1 + s_{3,9,11}X_1} + Y_3^0 \frac{s_{4,12,14}}{1 + s_{4,12,14}X_1} \\ &+ Y_4^0 \frac{s_{5,13,15}}{1 + s_{5,13,15}X_1} + Y_5^0 \frac{s_{6,16,18}}{1 + s_{6,16,18}X_1} + Y_6^0 \frac{s_{7,17,19}}{1 + s_{7,17,19}X_1} \bigg\} \end{aligned} \qquad (133)$$

which is a 7th order polynomial in X_1. As in other models, (133) can be solved for X_1 numerically (e.g., in Mathematica).

VII. OBTAINING THERMODYNAMIC PARAMETERS

A. Nearest-Neighbor Model

The model of hybridization discussed so far treats the dynamics in terms of kinetic mass-action reactions and ignores both the mixing properties of the molecules and the exact physics of hybridization except for simply acknowledging that the thermodynamics parameters depend on base-pair composition. Recall that the pro-

cess of hybridization involves the formation of base pairs between Watson-Crick–complementary bases. Namely, 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; thus, this interaction is characterized primarily by the composition of the interacting strands. Another physical interaction, *base stacking*, characterizes the hybridization process, and it has been shown to depend on the sequence rather than the composition of the strands. As base stacking depends on the short-range interactions, it is thought to be adequately described by the Nearest-Neighbor (NN) model.

In the NN model, it is assumed that 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, such as 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 some symmetry parameters and choice of initial values. As the enthalpy and entropy terms may be assumed to be independent of temperature, they can be computed as follows ([1], [2]):

$$\Delta H = \sum_x \Delta H_x + \Delta H(\text{init}) + \Delta H(\text{sym}) \quad (134)$$

$$\Delta S = \sum_x \Delta S_x + \Delta S(\text{init}) + \Delta S(\text{sym}) \quad (135)$$

where the terms ΔH_x and ΔS_x are tabulated for all ten possible NN dimer duplexes, as are the initiation and symmetry terms. The free energy computation is analogous:

$$\Delta G = \sum_x \Delta G_x + \Delta G(\text{init}) + \Delta G(\text{sym}) \quad (136)$$

with the initiation and symmetry terms tabulated. The values ΔG_x for the dimer duplexes have been tabulated at 25°C ([1]) and at 37°C ([2]). Since ΔG depends on the temperature, the values ΔG_x for the dimer duplexes can be easily calculated from the corresponding ΔH_x and ΔS_x parameters by

$$\Delta G_x(T) = \Delta H_x - T\Delta S_x \quad (137)$$

The ten distinct dimer duplexes arise as follows. Following the notation of Breslauer *et al.* ([1]), we denote each dimer duplex with a "slash-sign" separating antiparallel strands, e.g., AG/TC denotes 5'-AG-3' Watson-Crick base-paired with 3'-TC-5'. Alternately, $\frac{AG}{TC}$ is equivalent to AG/TC. The table below lists all sixteen ($= |\{A, T, C, G\}|^2 = 4^2$) possible dimers, identifying the equivalent ones.

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

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

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

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

Since our simulations involve oligonucleotide probes, we used the parameters for the initiation of duplex formation drawn from the results in the 1998 paper of SantaLucia ([2]). There, two different initiation parameters were introduced to 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 ([5]).

The table of parameters used in our simulations, drawn from [2], is duplicated in Table I for convenience. The

TABLE I: Unified oligonucleotide ΔH, ΔS, and ΔG NN parameters in 1M NaCl. [Reproduced with permission from [2] (Copyright 1998, National Academy of Sciences, U.S.A.)]. ΔG is computed at 37°C.

Interaction	ΔH kcal/mol	ΔS cal/K·mol	ΔG kcal/mol
AA/TT	-7.9	-22.2	-1.00
AT/TA	-7.2	-20.4	-0.88
TA/AT	-7.2	-21.3	-0.58
CA/GT	-8.5	-22.7	-1.45
GT/CA	-8.4	-22.4	-1.44
CT/GA	-7.8	-21.0	-1.28
GA/CT	-8.2	-22.2	-1.30
CG/GC	-10.6	-27.2	-2.17
GC/CG	-9.8	-24.4	-2.24
GG/CC	-8.0	-19.9	-1.84
Init. w/term. $G \cdot C$	0.1	-2.8	0.98
Init. w/term. $A \cdot T$	2.3	4.1	1.03
Symmetry correction	0	-1.4	0.43

following example illustrates how the free energy can be computed according to (136) using the values from Table I.

1. Example

```
5'  C-G-A-A-G-T  3'
    * * * * * *
3'  G-C-T-T-C-A  5'
```
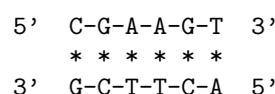
$$\begin{aligned}
\Delta G &= \Delta G(CG/GC) + \Delta G(GA/CT) \\
&\quad + \Delta G(AA/TT) + \Delta G(AG/TC) \\
&\quad + \Delta G(GT/CA) + \Delta G(\text{init. w}/G \cdot C) \\
&\quad + \Delta G(\text{init. w}/A \cdot T) + 0 \\
&= -2.17 - 1.30 - 1.00 - 1.28 - 1.44 + 0.98 + 1.03 \\
&= -5.18 \text{ kcal/mol}
\end{aligned}$$

Since the duplex is not self-complementary, $\Delta G(\text{sym}) = 0$.

B. Affinity Constants

We further recall that at equilibrium, the affinity constants K_1^j are given by (21), replicated here for convenience, as described in section III A:

$$K_1^j = \exp\left[-\Delta G/RT\right],$$

and ΔG due to stacking interactions is calculated as above. Also, we note that with the affinity constant values computed, we are ready to compute the "ratios of perfect match to mismatch values" for a particular initial target and probe concentrations.

VIII. OBSERVED COMPETITION AMONG PROBES

As discussed in sections III A, III B, and III C, we can compute the equilibrium TP concentrations, and thus the corresponding discrimination signal (match-to-mismatch ratio), from the initial target and probe concentrations.

To display the results of the computation and to describe the principal effects of competitive hybridization in a graphical manner, 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 section 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 Figure 1.

Figure 1 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
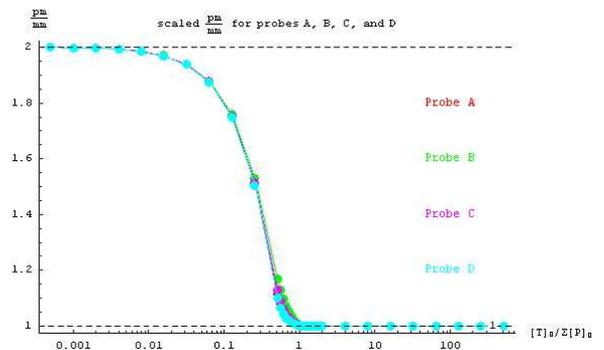
FIG. 1: Normalized discrimination, or universal Δ-plot

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.

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.

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 section III A, and under partial hybridization models (sections III B and III C), as illustrated in Figure 2. We observe the following:

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

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

3. We address the following questions: How can the shift direction be predicted? How does it depend on the sequences of the probe pairs in question?

A. Heuristic Development

Our empirical study was conducted as follows. Let us consider two probes, each having associated with it the pair $\{P_{\cdot,pm}, P_{\cdot,mm}\}$. For each probe, the pm/mm ratio shifts *up* or *down* in the presence of the other probe. The direction of the shift was determined to be a function of the relative sizes of the affinity constants K, where cross-bound states can be neglected. For a given probe, let $K_T$pm, $K_T$mm denote the affinity constants for *This* probe's binding site with pm and mm, respectively; let
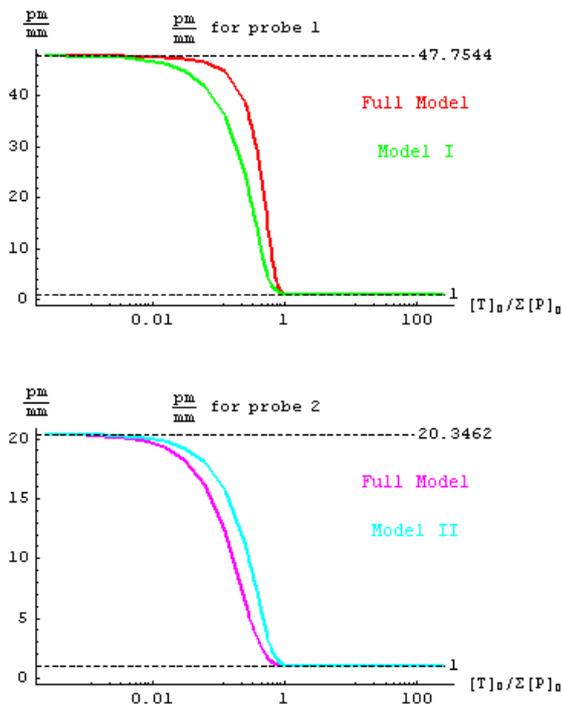
FIG. 2: pm/mm ratios for probe A (top graph) and probe B (bottom graph), plotted against scaled initial target concentration.

FIG. 3: Competition effect binary function on exon 11 probes.

FIG. 4: Competition effect binary function on exon 11 probes, shown with the separatrix $y = -x$.

$K_O$pm, $K_O$mm be the *Other* probe's affinity constants with pm and mm.

Let us view the competition effect as a binary function on the space of affinity constants ($+1$ for *up*, -1 for *down* shift) and consider the projection of the affinity constant space

$$\mathbb{R}^4 = \{K_T\text{pm}, K_T\text{mm}, K_O\text{pm}, K_O\text{mm}\}$$

onto the plane $\mathcal{L}$ with axes $\log(K_T\text{pm}/K_O\text{pm})$ and $\log(K_T\text{mm}/K_O\text{mm})$. On this plane, the competition effect function values can be clearly separated by the line $x + y = 0$. This condition holds for physical exon 11 probes, as shown in Figures 3 and 4.

The empirically determined condition can be described by the following logically equivalent statements:

pm/mm ratio shifts *up*

$$\Longleftrightarrow \quad y < -x$$
$$\Longleftrightarrow \quad \log(K_T\text{mm}/K_O\text{mm}) < -\log(K_T\text{pm}/K_O\text{pm})$$
$$\Longleftrightarrow \quad K_T\text{pm}\,K_T\text{mm} < K_O\text{pm}\,K_O\text{mm} \tag{138}$$

Thus, the signal for *This* probe improves whenever (138) holds.

In order to test the heuristic computationally, we generated more points for the competition effect function by perturbing existing probe sequences in one base and pairing one actual exon 11 probe with one perturbed probe.

The results of this empirical investigation of the competition effect on these probe pairs are graphically presented in Figure 5.
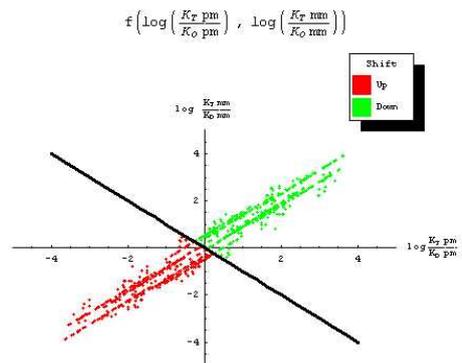
FIG. 5: Testing the heuristic computationally: each probe pair contains one actual exon 11 probe and one perturbed probe.

IX.   EXPERIMENTAL VALIDATION

In order to further verify the performance of our heuristic, we proposed the following experiments. The pm/mm ratios should be measured for the probes as listed below under Partial model (i.e., the probe and its alternate are present alone with the target) and Full model (i.e., the specified probes, each with an alternate, are present with the target) for:

- Actual probe pairs:

    - AB, AC, AD, BC, BD, CD
    - In each case, both probes should be used alternately as *This* and as *Other* probe.

- Actual/Perturbed probe pairs that show a change of shift direction in simulation:

This	*Other*			
A	D_5A		D_5T	
D	A_2G	A_6G	A_8G	A_14C

- Actual/Perturbed probe pairs (to be used as controls) that do not show change:

This	*Other*	
A	D_5C	D_6T
D	A_2C	A_2T

The remarkable consistency with which our heuristic conforms with the results of the simulation suggests that the heuristic can be used reliably in place of the simulation to predict the competition effect, i.e., the direction of the shift. This predictive power can be used in experiment design (e.g., for HLA typing).

1.   Example

Let $A = C381$, $B = A327$, and $C = D359$ from exon 11, with the alternates used in the experiments. Pairwise computational analysis indicated that: $A327$ improves the signal for $C381$ and $D359$ improves the signal for $A327$. Our heuristic implies that $D359$ automatically improves the signal for $C381$. This conclusion was tested using the extended model, as described in detail in section 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 the first probe from the Full and Partial models with the curve from the Extended model, as shown in Figure 6.

Note that, in Figure 6, the pm/mm ratio curve for $C381$ in the presence of both $A327$ and $D359$ (the blue curve) lies above both the red curve (the ratio for $C381$
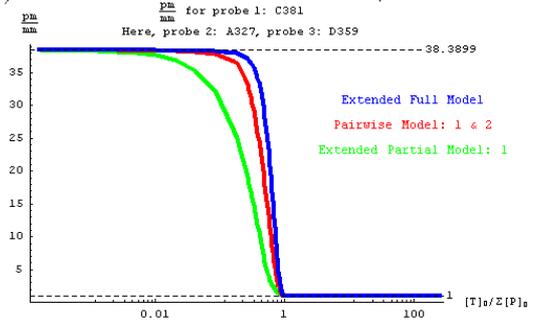
FIG. 6: Example: pm/mm ratios for three probes

in the presence of $A327$ alone) and the green curve (the ratio for $C381$ alone with the target). This indicates that for a given initial target concentration, i.e., a given point on the $x-$axis, the pm/mm ratio for $C381$ goes up in the presence of $A327$, which is consistent with pairwise analysis; the ratio increases further when $D359$ is added to the mix, confirming the heuristic prediction.

X.   CONCLUSION

In this paper we present mathematical models of the competitive probe-target hybridization process. Simulations based on the implementations of these models and the heuristic developed and presented in section VIII A generate results that are in agreement with experimental results observed in the laboratory. Prediction of competition effects based on *in silico* experiments can be used for the design of better biological experiments. Possible applications include experiment design for genotyping and mutation analysis.

Address for communication:

Vera Cherepinsky
MACS Department
Fairfield University
1073 North Benson Road
Fairfield, CT 06824
email: vcherepinsky@mail.fairfield.edu

[1] K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, PNAS USA **83**, 3746 (1986).

[2] J. SantaLucia, Jr., PNAS USA **95**, 1460 (1998).

[3] J. SantaLucia, Jr., H. T. Allawi, and P. A. Seneviratne, Biochemistry **35**, 3555 (1996).

[4] S. Wolfram, *The Mathematica Book.* (Cambridge University Press, 1999), 4th ed.

[5] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules.* (Freeman, San Francisco, 1980).

[6] S. Bommarito, N. Peyret, and J. SantaLucia, Jr., Nucleic Acids Research **28**, 1929 (2000).

[7] N. Peyret, P. A. Seneviratne, H. T. Allawi, and J. SantaLucia, Jr., Biochemistry **38**, 3468 (1999).

[8] H. T. Allawi and J. SantaLucia, Jr., Biochemistry **37**, 9435 (1998).

[9] H. T. Allawi and J. SantaLucia, Jr., Nucleic Acids Research **26**, 2694 (1998).

[10] H. T. Allawi and J. SantaLucia, Jr., Biochemistry **37**, 2170 (1998).

[11] H. T. Allawi and J. SantaLucia, Jr., Biochemistry **36**, 10581 (1997).

APPENDIX A: DETAILS OF MODEL IMPLEMENTATION

1. Choice of initial concentration parameters

The pm/mm ratios at equilibrium are computed from systems of equations corresponding to a particular model (equations (12)–(19) and (22)–(26) for Full Model; equations (32)–(35) and (36)–(38) for Model I; and equations (44)–(47) and (48)–(50) for Model II). In all these systems, the second set, made up of conservation equations, involves the initial concentration constants. The solutions will depend on the initial probe and target concentrations. There are several complications that stem from the choice of these parameters.

a. Initial probe concentrations

To measure competition effects inherent to the probes, the initial probe concentrations must be equal for different probes. If that is not the case, the observed results will be biased by the unequal starting amounts of probe material.

b. Initial target concentration

The choice of the initial target concentration must be made carefully as well. If $[T]_0 > \sum [P_{ij}]_0$, no competition effect will be observed since there is plenty of target to go around—each probe will get as much target as it needs.

The idea of "competition" described in this manuscript (see section II for the initial discussion) is based on the fact that a given target molecule, which can hybridize to either probe 1 or probe 2, or to a variant (alternate) of either of the probes, is more likely at equilibrium to end up hybridized to the probe that "holds it the strongest," i.e., the one with the most negative ΔG of hybridization. On the mass-action scale, this means that a higher proportion of the total target concentration will end up in a complex with the most "attractive" probe, or the best matching probe. However, this argument implicitly assumes that the targets are in short supply and the probes are competing for them.

Initial conditions for actual experiments performed in the experimental laboratory frequently use a different setup: the initial amount of the target is in huge excess over the probes. This appears to imply that under such conditions no competition effects should be observed—target-probe complexes should be formed for each of the probes, as there is plenty of target to go around! The affinity of the target to a particular probe does not enter the equation. And yet, experiments reveal the presence of competition. For a discussion of how this apparent paradox is resolved, see section A 2.

c. Scaling initial concentrations for comparison

To allow meaningful comparison among match-to-mismatch ratios for a given probe under different models, the initial concentration parameters must be scaled. If that is not done, much more sophisticated post-processing will be required to interpret the differences in the pm/mm ratio values. Scaling the parameters *a priori* also allows for the ratio curves to be plotted on the same set of axes and for the changes to be interpreted as "shifts" of the ratio curves.

2. Accuracy of entered parameters

The amount of target initially placed in the reaction chamber, together with chamber volume, is usually used to compute the initial target concentration. However, the value of $[T]_0$ computed in this manner may not be accurate. There are steric hindrances in the system. Probes are physically attached to large (relative to probe size) beads, and placed in the reaction chamber. The target molecules, which are much longer than probes, are free to float around the chamber. In order to interact with the probes, the target molecules must diffuse through the chamber. Only a small fraction of the target molecules placed in the reaction chamber end up close enough to the probe molecules to interact (i.e., hybridize) with them. Thus, while the amount of target the experimenter places

in the reaction chamber may significantly exceed the total amount of probes, the constraint

$$([T]_0)_{\text{effective}} < \sum [P_{ij}]_0$$

frequently holds for the *effective* initial target concentration, that is, the concentration of target molecules that diffused sufficiently far to reach the probes and participate in the hybridization reaction. This explains why competition effects, which are observed in the model only when $[T]_0 < \sum [P_{ij}]_0$, are also observed in practice.

Theoretically, one can compute the effective initial target concentration from the initial probe concentration, temperature, and measured pm/mm ratio. If the ratios are obtained for each probe separately, no competition effects will be present; hence the simple model of hybridization, described in section VI A can be used. Furthermore, if this data is obtained for a sequence of physical $[T]_0$'s, it may also be possible to observe a functional relationship between the physical $[T]_0$ and the effective $[T]_0$.

Suppose that for a given probe, a hybridization experiment is performed involving that probe, its alternate, and the target, and the concentrations of the matched probe-target complex (denoted by X_2) and the mismatched probe-target complex (denoted by X_3) at equilibrium are measured; the pm/mm ratio at equilibrium can be computed from the values of X_2 and X_3. The outcome of the experiment can be predicted *in silico* by the simple model. Recall that during the discussion of the dynamics of the simple model in section VI A, equation (105) for the pm/mm ratio was obtained; this equation is repeated here for convenience:

$$\left(\frac{TP_{11}}{TP_{01}} \right)_{\text{simp}} = \frac{a_0}{b_0} \cdot \frac{K_1^2}{K_1^3} \cdot \frac{K_1^3 + \frac{1}{X_1}}{K_1^2 + \frac{1}{X_1}}$$

Equation (105) can be used to solve for the equilibrium concentration of the free target (denoted by X_1) in terms of ratio $= \left(\frac{TP_{11}}{TP_{01}} \right)_{\text{simp}}$, which is, in turn, given in terms of the measured quantities X_2 and X_3:

$$\text{ratio} \; = \; \frac{X_2}{X_3} \tag{A1}$$

$$X_1 \; = \; \frac{\text{ratio} - (a_0/b_0) \cdot (K_1^2/K_1^3)}{K_1^2 \left((a_0/b_0) - \text{ratio} \right)} \tag{A2}$$

Finally, the effective initial target concentration can be obtained from the conservation rule

$$e_0 = ([T]_0)_{\text{effective}} = X_1 + X_2 + X_3, \tag{A3}$$

where X_1 is given in (A2). It is worth noting that this computation requires the values of X_2 and X_3, and not just their ratio. This brings up the additional complication of converting the measured quantities (intensities) into the same units as the computed quantities (concentrations), which is discussed in detail in section A 3.

3. Interpreting the results

In the laboratory, to obtain the concentration of a particular complex, one measures instead the total intensity of the fluorophores attached to the molecules of the complex. This intensity is a function of the concentration of the substance in question. The form of this function is generally assumed to be linear in a certain range, growing nonlinear outside the said range. Since the experimental laboratory measurements are in the units of intensity, and the model predicts concentrations of the substances at equilibrium, direct comparison of the *in silico* and laboratory data does not make sense. However, one can make the argument that since the primary interest is not in the concentrations of individual substances but rather in their ratios (the pm/mm ratios), the intensity-to-concentration scaling factor cancels out. The investigation described in this manuscript has relied on this assumption.

Nevertheless, one should be careful to verify that the quantities in question do indeed fall into the "linear" range of the intensity function. Should that prove not to be the case, it would no longer be appropriate to treat intensities and concentrations interchangeably; a more careful analysis of this "unit translation" would be prudent.

Furthermore, some of the analysis discussed here, in particular in section A 2, requires individual concentration values, making it necessary to formulate the relationship between intensity and concentration explicitly. Information required to obtain the function in question includes the intensity of a single fluorophore, the number of fluorophores attached to each target molecule, and the details of how the experimentally measured results are scaled (i.e., the post-processing of the scanned data).

APPENDIX B: FUTURE IMPROVEMENTS

1. Choice of alternate sites

All models discussed in this paper, with the exception of Simple Model, allow alternate binding sites for each probe. In the current formulation, those alternate sites are hard-wired to be the matching sites for the other probes involved. This choice of alternate sites fits in with the idea of how competition between probes works, and was convenient to implement, since the portions of the target in question were already stored as the complementary sequences for the other probes. As an added convenience, it also allowed the implementation to avoid string matching, since all the necessary string matching was done as a pre-processing step.

However, it would be more realistic to choose the alternate binding site(s) for each probe based on the sequence of that probe as well as that of the target. One possible approach to selecting potential "alternate sites" for a given probe could be the following. One could generate

a landscape of affinity constants ($K12$) and/or melting temperatures (T_m) by convolving the given probe with the long target, i.e., by shifting the probe along the target and computing the quantity of interest at each such alignment, and then threshold it, only keeping the "peaks" as the alternate sites.

2. Thermodynamics of mismatches

The current implementation of all hybridization models discussed in this paper computes the thermodynamic parameters of hybridization based on the NN model, making use of the parameters for all possible *matching* dimer duplexes, as described in section VII A. Recall that in all these hybridization models, for each probe there is an alternate probe, almost identical to the matching one (in all examples considered, the alternate (mismatching) probe differs from the matching one in only one base). Thus, it is necessary to make regular computations of thermodynamic parameters for target-probe pairs where mismatches occur. Further, more severe mismatches occur when "cross-terms" are considered, where a probe hybridizes to the "wrong" location on the target.

a. Current implementation

The simplest way to deal with such mismatches, and the one used in the current implementation, is to ignore the contributions of all mismatched dimers to the summation term (recall equation (136) for ΔG) when the mismatch occurs in the middle of the probe, and to omit the helix initiation parameter contribution if the mismatch occurs on the end of the probe. A single base mismatch in a probe automatically guarantees that the probe is not self-complementary (in the Watson-Crick sense); thus, if the original probe was self-complementary, the contribution from the symmetry term is omitted as well.

One should also consider the situation where a matching probe P is *almost* self-complementary, with only one base violating the property. In that case, replacing the offending base appropriately would generate a self-complementary mismatch probe P'. In the computation of ΔG for the hybridization of P' with the target T, the mismatched dimer contributions will be ignored, as discussed above, but the contribution of the symmetry term will be included.

It is also possible for two strategically placed mismatches to turn a matching self-complementary probe into a mismatched self-complementary probe. Thus, the test for self-complementarity should be performed on each probe sequence from scratch, rather than being inferred from the self-complementarity status of the original probe and the editing changes.

b. More detailed treatment of mismatches

Thermodynamic contributions of different mismatched dimers have been studied as well (see [6], [7], [8], [9], [10], and [11]). These studies showed that different internal mismatches have different effects on the thermodynamic parameters of hybridization—some even stabilize the resulting duplex. One can make use of these available parameters to treat mismatches in much more detail. However, one must be careful to keep in mind that the parameters for the internal (and some terminal) mismatches were derived using the stabilizing effect of neighboring matching base-pairs. As a result, these parameters may not have the same additive properties as the parameters for matching NN dimers. In any case, potential improvements in the accuracy of the resulting thermodynamic parameters must be weighed against the loss of speed due to more involved computations.