Statistical Thermodynamics and Kinetics of DNA Multiplex Hybridization Reactions

M. T. Horne,*‡ D. J. Fish,‡ and A. S. Benight*†‡
*Departments of Chemistry, †Physics and Mathematics, Portland State University, Portland, Oregon; and ‡Portland Bioscience, Portland, Oregon

ABSTRACT A general analytical description of the equilibrium and reaction kinetics of DNA multiplex hybridization has been developed. In this approach, multiplex hybridization is considered to be a competitive multichannel reaction process: a system wherein many species can react both specifically and nonspecifically with one another. General equations are presented that can consider equilibrium and kinetic models of multiplex hybridization systems comprised, in principle, of any number of targets and probes. Numerical solutions to these systems for both equilibrium and kinetic behaviors are provided. Practical examples demonstrate clear differences between results obtained from more common simplex methods, in which individual hybridization reactions are considered to occur in isolation; and multiplex hybridization, where desired and competitive cross-hybrid reactions between all possible pairs of strands are considered. In addition, sensitivities of the hybridization process of the perfect match duplex, to temperature, target concentration, and existence of sequence homology with other strands, are examined. This general approach also considers explicit sequence-dependent interactions between targets and probes involved in the reactions. Sequence-dependent stabilities of all perfect match and mismatch duplex complexes are explicitly considered and effects of relative stability of cross-hybrid complexes are also explored. Results reveal several interdependent factors that strongly influence DNA multiplex hybridization behavior. These include: relative concentrations of all probes and targets; relative thermodynamic stability of all perfect match and mismatch complexes; sensitivity to temperature, particularly for mismatches; and amount of sequence homology shared by the probe and target strands in the multiplex mix.

INTRODUCTION

Historically, theoretical and experimental studies of the equilibrium melting (or hybridization), and kinetics of short duplex DNAs have primarily focused on ''simplex'' reactions wherein two short single strands, whose sequences are perfectly complementary, anneal to form a perfectly matched Watson-Crick (w/c) basepaired duplex. Simplex melting experiments of short duplex DNAs, with well-defined sequences, and theoretical descriptions of the melting process, have been an active subject of study for over 25 years (1–11). Systematic studies of the melting stability of short duplex DNAs with different sequences have provided evaluations of nearest-neighbor thermodynamic stability parameters that enable prediction of the thermodynamic stability of a short duplex from its basepair sequence (1,8,9). These parameter sets are commonly used in the design of probes and primers having desired hybridization and stability properties for diagnostic assays. Similarly, kinetic studies of the annealing (or hybridization) of two perfectly matched single strands to form a duplex, have been performed and have provided analytical descriptions of the simplex hybridization process. These have provided evaluations of the kinetic rate constants, and other parameters, describing hybridization kinetics of short DNAs (12–23).

Although parameter values determined from kinetic and equilibrium analysis of simplex reactions have been quite useful in facilitating improved predictions of the sequence-dependent melting and kinetic behaviors of homogeneous solutions of short duplex DNAs and their relative sequence-dependent stability, more accurate parameters alone are insufficient to provide realistic descriptions of multiplex hybridization reactions, i.e., when more than one duplex is present in the same solution. There are several additional, essential components of multiplex hybridization reactions that must be considered, which make them remarkably different from more common simplex reactions. These are important considerations as multiplex hybridization forms the basis of many modern nucleic acid diagnostic reactions.

Nucleic acid diagnostic assays based on multiplex hybridization have the potential to revolutionize genomic research and genetic medicine (24–31). Multiplex assays can be performed on microarrays or in solution via various iterations of the polymerase chain reaction (32–42). These assays offer never-before-imagined capabilities for systematic high throughput screening, discrimination, and analysis of large numbers of DNA (and RNA) sequences. Despite the vast and important applications of multiplex hybridization in nucleic acid diagnostics, there have been relatively few studies aimed at gaining a better understanding of the actual hybridization reactions that can occur in mixtures containing more than two strands.

As stated above, the majority of studies of melting (or hybridization) reactions of short duplex DNAs have been performed primarily on equimolar mixtures of two single

Submitted June 8, 2006, and accepted for publication July 26, 2006.

Address reprint requests to A. S. Benight, Tel.: 503-725-9513; E-mail: abenight@pdx.edu.

0006-3495/06/12/4133/21 $2.00

doi: 10.1529/biophysj.106.090662

strands whose sequences are perfectly complementary. Exceptions are studies of solutions containing a single species of a linear or circular self-complementary single strand for the purpose of evaluating DNA sequence-dependent stability parameters, and examining intramolecular hairpin stability (43–45), and studies performed for the purpose of investigating the thermodynamic parameters of single basepair mismatches (1,9,46), bulged loops (47–49), or multiple strand mixtures, such as triplexes and quadruplex (50). Melting studies of mixtures of more than two strands in solution, which can form both perfect match and mismatch duplex complexes, have suggested that sequence-dependent interactions in tandem mismatches might contribute to the stability of mismatch hybrid duplex complexes (51). For this and other reasons described herein, when multiple strands meant to form perfectly matched duplexes are present together in a reaction mix, it cannot be presumed that each pair of strands will form the appropriate duplex, exclusively, in the same way they would in isolation, in the absence of the influence from other single strands (and duplexes). Several additional essential features of multiplex hybridization must be considered. These considerations are associated with the significant probability of formation of mismatch hybrids (cross-hybridization) brought about by sequence homology with other strands and sequence-dependent stability of mismatch basepairs.

Existing nucleic acid platforms and analysis procedures continue to be improved and new technology platforms promising higher sensitivity and more highly reproducible results are being developed to the point of providing reliable quantitative measurements (16,17,52–54). Along with these developmental improvements, it is essential that an analytical foundation be established that will enable formulation and support of robust and realistic models of multiplex hybridization. In turn, these will facilitate enhanced design, analysis, and optimization of nucleic acid multiplex diagnostic assays.

Several authors have recognized the necessity of considering multiple hybridization reactions simultaneously and demonstrated, in several simple cases, generally dramatic effects associated with the potential of multiple two-strand interactions (55,56). As might be expected, these studies have pointed out that interactions between each probe:target pair cannot be assumed to occur as separate events. These studies suggest the essential importance of considering competitive reactions in a multiplex hybridization reaction mixture in a multichannel-systems manner.

For the case of hybridization on microarrays, a formal approach to modeling the hybridization of targets with probes affixed to a surface involves two distinct phases (16,22, 56,57), termed ''transport'' and ''reaction.'' The transport phase involves diffusion of target strands across the probe surface. The reaction phase involves the reaction (binding and dissociation) kinetics between targets and probes at the surface. For the latter, once targets have diffused to the interaction zone, reaction kinetics dominates the process, and effects of sequence-dependent interactions become quite significant. If the target concentration is sufficiently in excess of the probe concentration, to the point where encounters between all targets and probes are equally likely, then the entire process is dominated by the reaction phase. The development presented here considers only the reaction phase of the multiplex hybridization process.

Model systems studied so far have discovered complications associated with competition and cross-hybridization between two and three strands in the same reaction, and anticipated added complexities associated with competition between multiple-strand interactions in multiplex mixes (55,56). However, general descriptions of the equilibrium and kinetic behaviors of DNA multiplex hybridization reactions, containing any number of probes and targets, have not been presented. To model multiplex hybridization reactions in a completely general way, we have developed a systems-analysis approach involving collective consideration of multiple reactions and the simultaneous solution for individual products of specific reactions. In the development presented here, multiplex hybridization is considered to be a competitive, multichannel, reaction process: a system wherein many species can react both specifically and nonspecifically with one another. General equations are presented supporting both equilibrium and kinetic models of multiplex hybridization systems comprised, in principle, of any number of targets and probes. Practical examples demonstrate clear differences between simplex and multiplex hybridization and sensitivity of the hybridization process of perfect-match duplexes to temperature, target concentration, and existence of sequence homology with other strands.

This article is presented in the following sections. In Theoretical Approach, our general theoretical approaches are presented. Both equilibrium and kinetic models are described, and their analytical solutions are provided. Example Calculations contains the results of a number of comparisons of examples of simplex and multiplex calculations, and their sensitivity to temperature, sequence homology, and stability of mismatch hybrid duplexes. In Comparisons with Previous Work, comparisons to relevant published works are provided. The final section (Conclusions) summarizes our major findings and provides conclusions.

THEORETICAL APPROACH

Equilibrium statistical thermodynamics

Consider a population of DNA single strands at constant temperature and pressure. For any molecular configuration adopted by these strands, i.e., single-strand intramolecular hairpin or two-stranded duplex structure, the number of microstates is given by the Gibb's factor for that configuration. For a specific configuration, i, at approximately constant solution volume, the Gibb's factor or statistical weight is given by $sw_i = e^{-\frac{\Delta G_i^0}{RT}}$, where ΔG_i^0 is the standard-state free energy of the i^{th} configuration. The standard-state

free energy is given in terms of the differences of the standard-state chemical potentials of the configuration and the unstructured (melted) single strands from which it formed. For example, for a duplex $p_i t_j$ formed from a probe p_i and a target t_j, the standard-state free energy is given by

$$\Delta G^0_{p_i t_j} = \mu^0_{p_i t_j} - \mu^0_{p_i} - \mu^0_{t_j},$$

where $\mu^0_{p_i t_j}$, $\mu^0_{p_i}$, and $\mu^0_{t_j}$ are the standard-state chemical potentials for the probe:target complex, and the individual probe and target strands, respectively.

In this development, the reference state of the single strands is defined as the unstructured, unstacked (melted) single strand. The reference state for the duplex can be defined in a number of ways. For example, as the nucleation complex of two single strands in approximately the same orientation and volume as the fully duplex state, in the absence of hydrogen-bonding or stacking interactions between the strands. The system partition function Z is the sum of the statistical weights for all possible configurations, i.e., $Z = \sum_i sw_i$. The probability of an occurrence of configuration i is then given by $P = sw_i/Z$.

Consider a multiplex system wherein a set of probes, p_i, is designed to specifically hybridize with particular targets, t_j. Implicit in such a specificity requirement is the potential for error. This is necessarily so because, conceivably, every probe and target strand may form either a perfectly matched w/c duplex or a ''degenerate'' structure, i.e., a nonperfect match hybrid duplex, or an intramolecular hairpin. The standard-state free energies that define the statistical weights, and therefore the population of error states, depend explicitly on the differences in their respective standard-state chemical potentials.

Chemical system

To model the ''system'' behavior of multiplex hybridization reactions, a system of equilibrium reactions is assumed (see Table 1) where, for example, $p_i t_j \sigma$ is the ensemble of duplex states that form from single strands p_i and t_j with equilibrium constant $K^{eq}_{p_i t_j}$. Similar definitions apply for the other reactions. This formulation also considers, through the index σ, the possibility of microstates within each configuration, such that $x\sigma$ represents a specific microstate of the configuration, e.g., $p_i t_j \sigma$ could represent an overlap duplex state of a configuration involving single strands p_i and t_j, where the two strands might not be perfectly aligned at their ends. If these microstates are not considered (as in formulation of the kinetic model described later), $\sigma = 1$.

TABLE 1 Chemical reactions of microarray hybridization

Reaction	Equilibrium constant	Rate constants
$p_i + t_j \rightleftharpoons p_i t_j \sigma$	$K^{eq}_{p_i t_j}$	$k^f_{p_i t_j} k^r_{p_i t_j}$
$p_i + p_j \rightleftharpoons p_i p_j \sigma$	$K^{eq}_{p_i p_j}$	$k^f_{p_i p_j} k^r_{p_i p_j}$
$t_i + t_j \rightleftharpoons t_i t_j \sigma$	$K^{eq}_{t_i t_j}$	$k^f_{t_i t_j} k^r_{t_i t_j}$
$t_i \rightleftharpoons t_i^{hp} \sigma$	$K^{eq}_{t_i^{hp}}$	$k^f_{t_i^{hp}} k^r_{t_i^{hp}}$
$p_i \rightleftharpoons p_i^{hp} \sigma$	$K^{eq}_{p_i^{hp}}$	$k^f_{p_i^{hp}} k^r_{p_i^{hp}}$

For the above equilibrium system, strand conservation for the probe and target strands is given by

$$C_{p_i} = C^0_{p_i} - \sum_{j,\sigma} C_{p_i t_j \sigma} - \sum_{k,\sigma} C_{p_i p_k \sigma} - \sum_{\sigma \neq 1} C_{p_i \sigma}, \quad (1a)$$

$$C_{t_i} = C^0_{t_i} - \sum_{j,\sigma} C_{p_i t_j \sigma} - \sum_{k,\sigma} C_{t_i t_k \sigma} - \sum_{\sigma \neq 1} C_{t_i \sigma}, \quad (1b)$$

where $C^0_{p_i}$ and $C^0_{t_i}$ are the total concentrations of the probe and target strands, respectively.

Now, the concentrations of the different reaction products are given in terms of the equilibrium constants for their formation:

$$C_{p_i t_j \sigma} = K^{eq}_{p_i t_j \sigma} C_{p_i} C_{t_j} = C_{p_i} C_{t_j} exp\left(-\frac{\Delta G^0_{p_i t_j \sigma}}{RT}\right), \quad (2a)$$

$$C_{t_i t_j \sigma} = K^{eq}_{t_i t_j \sigma} C_{t_i} C_{t_j} = C_{t_i} C_{t_j} exp\left(-\frac{\Delta G^0_{t_i t_j \sigma}}{RT}\right), \quad (2b)$$

$$C_{p_i p_j \sigma} = K^{eq}_{p_i p_j \sigma} C_{p_i} C_{p_j} = C_{p_i} C_{p_j} exp\left(-\frac{\Delta G^0_{p_i p_j \sigma}}{RT}\right), \quad (2c)$$

$$C_{p_i \sigma} = K^{eq}_{p_i \sigma} C_{p_i} = C_{p_i} exp\left(-\frac{\Delta G^0_{p_i \sigma}}{RT}\right), \quad (2d)$$

$$C_{t_i \sigma} = K^{eq}_{t_i \sigma} C_{t_i} = C_{t_i} exp\left(-\frac{\Delta G^0_{t_i \sigma}}{RT}\right). \quad (2e)$$

Define the following:

$$K^{eq}_x = \prod_\sigma K^{eq}_{x\sigma}.$$

With these expressions, Eqs. 1a and 2b become

$$C^0_{p_i} = C_{p_i} + \sum_j K^{eq}_{p_i t_j} C_{p_i} C_{t_j} + \sum_j K^{eq}_{p_i p_j} C_{p_i} C_{p_j} + K^{eq}_{p_i} C_{p_i}, \quad (3a)$$

$$C^0_{t_i} = C_{t_i} + \sum_j K^{eq}_{t_i t_j} C_{t_i} C_{t_j} + \sum_j K^{eq}_{t_i p_j} C_{t_i} C_{p_j} + K^{eq}_{t_i} C_{t_i}. \quad (3b)$$

For simplicity, if target:target, probe:probe, and intramolecular states in single-strand targets and probes are ignored, i.e., $K^{eq}_{t_i t_j} = K^{eq}_{p_i p_j} = K^{eq}_{t_i} = K^{eq}_{p_i} = 0$, Eqs. 3a and 3b can be combined to form the following system of coupled, nonlinear equations:

$$C_{t_i}\left(1 + \sum_j \frac{K^{eq}_{t_i p_j} C^0_{p_i}}{1 + \sum_l K^{eq}_{p_j t_l} C_{t_l}}\right) + C^0_{t_i} = 0, \quad (4a)$$

$$C_{p_i}\left(1 + \sum_j \frac{K^{eq}_{p_i t_j} C^0_{t_i}}{1 + \sum_l K^{eq}_{t_j p_l} C_{p_l}}\right) + C^0_{p_i} = 0. \quad (4b)$$

If there are N_P probes and N_T targets, and l goes from 1 to N_T, then Eq. 4a or 4b represents a system of $N_P + N_T$ nonlinear

coupled equations. Given input values of $C^0_{p_i}$ and $C^0_{t_i}$, and the equilibrium constants, $K^{eq}_{t_i p_i}$ and $K^{eq}_{p_j t_i}$, a solution of this system produces values of C_{t_i}, which can in turn be used to evaluate the corresponding values of C_{p_i}. These are then combined with the particular equilibrium constants, Eq. 2, to determine the duplex concentrations, $C_{p_i t_j}$.

Multiplex hybridization kinetics

Consider a multicomponent system comprised of two species types, i.e., probes p_i, and targets t_j. Each p_i can react with every t_j forming a $p_i t_j$ complex or can react with each of its counterparts of the same species, forming $p_i p_j$ and $t_i t_j$ complexes (for simplicity, we ignore hairpin formation as its inclusion does not affect the results that follow). Each of these individual reactions occurs with forward rate constants $k^f_{p_i t_j}$, $k^f_{p_i p_j}$, and $k^f_{t_i t_j}$ and reverse rate constants, $k^r_{p_i t_j}$, $k^r_{p_i p_j}$, and $k^r_{t_i t_j}$ respectively (see Table 1). The forward rates constants are assumed to be the same for all species, i.e., $k^f_{p_i t_j} = k^f_{p_i p_j} = k^f_{t_i t_j} \equiv k^f$.

It should be noted that this development concerns only the ''reaction'' kinetics of the hybridization process. That is, all targets are available for interactions with all probes. It is implicitly assumed that there are no diffusional barriers to the reaction process. Under the umbrella of this assumption, in combination with the above assignments, the fundamental rate equations are as follows:

$$\frac{d}{d\tau}C_{p_i} = \sum_{j=1}^{M}(k^r_{p_i t_j}C_{p_i t_j} - k^f C_{p_i}C_{t_j}) + \sum_{j=1}^{N}(\delta_{ij}+1)(k^r_{p_i p_j}C_{p_i p_j} - k^f C_{p_i}C_{p_j}), \quad (5a)$$

$$\frac{d}{d\tau}C_{t_j} = \sum_{i=1}^{N}(k^r_{p_i t_j}C_{p_i t_j} - k^f C_{p_i}C_{t_j}) + \sum_{i=1}^{M}(\delta_{ji}+1)(k^r_{t_i t_j}C_{t_i t_j} - k^f C_{t_i}C_{t_j}); \quad (5b)$$

and

$$\frac{d}{d\tau}C_{p_i p_j} = k^f C_{p_i}C_{p_j} - k^r_{p_i p_j}C_{p_i p_j}, \quad (6a)$$

$$\frac{d}{d\tau}C_{p_i t_j} = k^f C_{p_i}C_{t_j} - k^r_{p_i t_j}C_{p_i t_j}, \quad (6b)$$

$$\frac{d}{d\tau}C_{t_i t_j} = k^f C_{t_i}C_{t_j} - k^r_{t_i t_j}C_{t_i t_j}. \quad (6c)$$

Here, C_{p_i} is the concentration of the i^{th} probe ($i \le N$), C_{t_j} is the concentration of the j^{th} target ($j \le M$), and $C_{p_i p_j}$, $C_{p_i t_j}$, and $C_{t_i t_j}$ are the concentrations of the probe:probe ($i, j \le N$), probe:target ($i \le N$, $j \le M$), and target:target ($i, j \le M$) duplexes. The factor ($\delta_{ab}+1$) is either 1 or 2, depending on whether $a \ne b$ or $a = b$, respectively. After ignoring repetition due to symmetries (C_{pp} and C_{tt} are symmetric), a system of $d = (1/2)(N+M)(N+M+5)$ equations is formed.

Observe that by combining Eqs. 5 and 6, the rate equations for the single-strand probe and target concentrations can be written as

$$\frac{d}{d\tau}C_{p_i} = -\sum_{j=1}^{M}\frac{d}{d\tau}C_{p_i t_j} - \sum_{j=1}^{N}\frac{d}{d\tau}C_{p_i p_j} - \frac{d}{d\tau}C_{p_i p_i},$$

$$\frac{d}{d\tau}C_{t_j} = -\sum_{i=1}^{N}\frac{d}{d\tau}C_{p_i t_j} - \sum_{i=1}^{M}\frac{d}{d\tau}C_{t_i t_j} - \frac{d}{d\tau}C_{t_j t_j}.$$

If, for $i \le N$ and $j \le M$, the functions S_{p_i} and S_{t_j} are defined as

$$S_{p_i} = C_{p_i} + \sum_{j=1}^{M}C_{p_i t_j} + \sum_{j=1}^{N}C_{p_i p_j} + C_{p_i p_i}, \quad (7a)$$

$$S_{t_j} = C_{t_j} + \sum_{i=1}^{N}C_{p_i t_j} + \sum_{i=1}^{M}C_{t_i t_j} + C_{t_j t_j}, \quad (7b)$$

then

$$\frac{d}{d\tau}S_{p_i} = \frac{d}{d\tau}C_{p_i} + \sum_{j=1}^{M}\frac{d}{d\tau}C_{p_i t_j} + \sum_{j=1}^{N}\frac{d}{d\tau}C_{p_i p_j} + \frac{d}{d\tau}C_{p_i p_i} = 0,$$

$$\frac{d}{d\tau}S_{t_i} = \frac{d}{d\tau}C_{t_j} + \sum_{i=1}^{N}\frac{d}{d\tau}C_{p_i t_j} + \sum_{i=1}^{M}\frac{d}{d\tau}C_{t_i t_j} + \frac{d}{d\tau}C_{t_j t_j} = 0.$$

Thus, the functions S_{p_i} and S_{t_j} are constants with respect to the time of reaction for the system, Σ. In fact, these functions form a set of $N+M$ independent constants. This allows the reduction of the system Σ to the smaller system obtained by performing the following substitution.

Let a solution to Σ be given with initial conditions $C_{p_i} = C^0_{p_i}$, $C_{t_j} = C^0_{t_j}$, $C_{p_i p_j} = 0$, $C_{p_i t_j} = 0$, and $C_{t_i t_j} = 0$. Then for all time, τ, $S_{p_i}(\tau) = C^0_{p_i}$ and $S_{t_j}(\tau) = C^0_{t_j}$. According to the definitions in Eqs. 7a and 7b, and suppressing the variable τ,

$$C_{p_i} = C^0_{p_i} - \sum_{j=1}^{M}C_{p_i t_j} - \sum_{j=1}^{N}C_{p_i p_j} - C_{p_i p_i}, \quad (8a)$$

$$C_{t_j} = C^0_{t_j} - \sum_{i=1}^{N}C_{p_i t_j} - \sum_{i=1}^{M}C_{t_i t_j} - C_{t_j t_j}. \quad (8b)$$

For convenience, denote the right-hand sides of Eqs. 8a and 8b by $C^0_{p_i} - S^p_i$ and $C^0_{t_j} - S^t_j$, respectively. Then Eqs. 6a–6c can be replaced by

$$\frac{d}{d\tau}C_{p_i p_j} = k^f(C^0_{p_i} - S^p_i)(C^0_{p_j} - S^p_j) - k^r_{p_i p_j}C_{p_i p_j}, \quad (9a)$$

$$\frac{d}{d\tau}C_{p_i t_j} = k^f(C^0_{p_i} - S^p_i)(C^0_{t_j} - S^t_j) - k^r_{p_i t_j}C_{p_i t_j}, \quad (9b)$$

$$\frac{d}{d\tau}C_{t_i t_j} = k^f(C^0_{t_i} - S^t_i)(C^0_{t_j} - S^t_j) - k^r_{t_i t_j}C_{t_i t_j}. \quad (9c)$$

Together with Eqs. 8a and 8b, these define an equivalent system of $N+M$ fewer equations.

Since the functions S_{p_i} and S_{t_j} are constant along solutions to Σ, Eqs. 8a and 8b form a system of $N+M$ independent nonlinear equations, which is satisfied by the solution to Σ with initial conditions of

$$(C_p, C_t, C_{pp}, C_{pt}, C_{tt}) = (C_p^0, C_t^0, 0, 0, 0).$$

By continuity, any equilibrium point of Σ also satisfies this system of equations. Thus, the search for equilibria of the system Σ should begin with a search for solutions to this system. The following relations are seen to hold at equilibrium points of Σ,

$$k^f C_{p_i} C_{p_j} = k^r_{p_i p_j} C_{p_i p_j}, \tag{10a}$$

$$k^f C_{p_i} C_{t_j} = k^r_{p_i t_j} C_{p_i t_j}, \tag{10b}$$

$$k^f C_{t_i} C_{t_j} = k^r_{t_i t_j} C_{t_i t_j}; \tag{10c}$$

or, if $K^{eq}_{xy} = k^f / k^r_{xy}$, where x and y are either p_i or t_j, then

$$K^{eq}_{p_i p_j} C_{p_i} C_{p_j} = C_{p_i p_j}, \tag{11a}$$

$$K^{eq}_{p_i t_j} C_{p_i} C_{t_j} = C_{p_i t_j}, \tag{11b}$$

$$K^{eq}_{t_i t_j} C_{t_i} C_{t_j} = C_{t_i t_j}. \tag{11c}$$

Thus, any equilibrium solution to Σ must be contained in the set of common solutions to both systems of Eqs. 8 and 11. These equations comprise the formal foundation of the so-called equilibrium model for the hybridization of probes and targets. In summary, the basic expressions are

$$C_{p_i} = C^0_{p_i} - \sum_{j=1}^{M} C_{p_i t_j} - \sum_{j=1}^{N} C_{p_i p_j} - C_{p_i p_i}, \tag{12a}$$

$$C_{t_j} = C^0_{t_j} - \sum_{i=1}^{N} C_{p_i t_j} - \sum_{i=1}^{M} C_{t_i t_j} - C_{t_j t_j}, \tag{12b}$$

$$K^{eq}_{p_i p_j} = \frac{C_{p_i p_j}}{C_{p_i} C_{p_j}}, \tag{12c}$$

$$K^{eq}_{p_i t_j} = \frac{C_{p_i t_j}}{C_{p_i} C_{t_j}}, \tag{12d}$$

$$K^{eq}_{t_i t_j} = \frac{C_{t_i t_j}}{C_{t_i} C_{t_j}}. \tag{12e}$$

Examination of these expressions reveals they are identical to those in Eqs. 3a and Eq. 3b derived from equilibrium considerations alone. Solutions of these determine equilibrium concentrations for each species.

Recall that $p_i t_j$ represents the duplex formed from single strands p_i and t_j, with similar definitions for the other reactions in Table 1 (neglecting hairpins in single strands). Following are alternate forms of the expressions in Eq. 5:

$$\begin{aligned}\frac{dC_{t_i}}{d\tau} = &\sum_{\sigma \neq 1} [k^r_{t_i\sigma} C_{t_i\sigma} - k^f_{t_i\sigma} C_{t_i\sigma}] \\ &+ \sum_{\sigma,j} [(k^r_{t_i p_j \sigma} C_{t_i p_j \sigma} - k^f_{t_i p_j \sigma} C_{t_i} C_{p_j})] \\ &+ \sum_{\sigma,j} [(k^r_{t_i t_j \sigma} C_{t_i t_j \sigma} - k^f_{t_i t_j \sigma} C_{t_i} C_{t_j})],\end{aligned} \tag{13a}$$

$$\begin{aligned}\frac{dC_{p_i}}{d\tau} = &\sum_{\sigma \neq 1} [k^r_{p_i\sigma} C_{p_i\sigma} - k^f_{p_i\sigma} C_{p_i\sigma}] \\ &+ \sum_{\sigma,j} [k^r_{p_i t_j \sigma} C_{p_i t_j \sigma} - k^f_{p_i t_j \sigma} C_{p_i} C_{t_j}] \\ &+ \sum_{\sigma,j} [k^r_{p_i p_j \sigma} C_{p_i p_j \sigma} - k^f_{p_i p_j \sigma} C_{p_i} C_{p_j}].\end{aligned} \tag{13b}$$

Likewise, the following are alternate forms of expressions in Eq. 6:

$$\frac{dC_{p_i t_j}}{d\tau} = \sum_{\sigma} [k^f_{p_i t_j \sigma} C_{p_i} C_{t_j} - k^r_{p_i t_j \sigma} C_{p_i t_j \sigma}], \tag{14a}$$

$$\frac{dC_{t_i t_j}}{d\tau} = \sum_{\sigma} [k^f_{t_i t_j \sigma} C_{t_i} C_{t_j} - k^r_{t_i t_j \sigma} C_{t_i t_j \sigma}], \tag{14b}$$

$$\frac{dC_{p_i p_j}}{d\tau} = \sum_{\sigma} [k^f_{p_i p_j \sigma} C_{p_i} C_{p_j} - k^r_{p_i p_j \sigma} C_{p_i p_j \sigma}]. \tag{14c}$$

Initial conditions for the system are $C_{xy\sigma} = 0$ and $C_{x\sigma} = 0$ (i.e., all strands are initially in single-strand conformations) and $C_z = C^0$ (i.e., initial single-strand concentrations are given), where x, y, z represent p_i or t_j. If misaligned, overlapping states are ignored, then $\sigma = 1$, in the expressions in Eq. 13. Assuming that $k^f_x = k^f$ for all states x, then $k^r_x = k^f [K^{eq}_x]^{-1}$, in analogy with the expressions in Eqs. 5 and 6.

It should be noted that, with little difficulty, the preceding development can be extended to include hairpin formation in single strands. To do this, the expressions in Eq. 5 are amended to include the effect of hairpin formation, and rate equations for hairpin configurations are added to the set of expressions in Eq. 6. As a result, Eqs. 12a and 12b include additional terms, and the expressions in Eqs. 12c–12e are expanded. This extension has been omitted in the above discussion for the sake of simplicity.

Further, the system in Eq. 5 allows for potential interactions between distinct species of probes, i.e., the formation of probe:probe duplexes is not precluded. Thus, the models developed above are not specifically restricted to microarray hybridization, but can also be applied to multichannel simulations of DNA hybridization in solution. Of course, elimination of probe:probe interactions from the simulation can be achieved by the appropriate assignment of rate constants (i.e., $k^f_{p_i p_j} = 0$ for all i and j).

Numerical models

Equilibrium model

Estimates of the solutions to the system of Eq. 12 can be obtained using any standard numerical optimization program. For example, the nonlinear least-squares optimizer provided by MatLab's Optimization Toolbox (58), as well as a similar algorithm provided by the software OCTAVE (59) have been employed with comparable degrees of success. These algorithms were applied to the above equations, written as a vector-valued function F on R^{N+M} as follows

(hairpin reactions are not considered here, but can be included without difficulty):

$$F_i = C_{p_i} + C_{p_i}\left(\sum_{j=1}^{M} C_{t_j}K^{eq}_{p_it_j} + \sum_{j=1}^{N} C_{p_j}K^{eq}_{p_ip_j}\right) + C_{p_i}C_{p_i}K^{eq}_{p_ip_i} - C^0_{p_i},$$

$$F_j = C_{t_j} + C_{t_j}\left(\sum_{i=1}^{N} C_{p_i}K^{eq}_{p_it_j} + \sum_{i=1}^{M} C_{t_i}K^{eq}_{t_it_j}\right) + C_{t_j}C_{t_j}K^{eq}_{t_jt_j} - C^0_{t_j}.$$

To numerically find the roots of this system of equations, an initial guess (or seed) must first be made that approximates the true solutions. If seed values are not chosen appropriately, the algorithm can terminate without returning acceptable values for concentrations of probes and targets. To overcome this difficulty, an iterative process was developed that reseeds the algorithm until an acceptable level of error is reached in the approximate solutions. Using this process, the equilibrium concentrations of hybridization systems of N probes and M targets with N and M as high as 1000 have been successfully computed. On a 3.6 GHz Pentium-IV machine with 1 Gigabyte of RAM, the algorithm required ~45 min to complete a 1000 × 1000 simulation.

For more accurate values, equilibrium concentrations calculated from the Kinetic Model (see next section) can be used as seed values for the equilibrium algorithm. Unfortunately, the kinetic algorithm involves a much larger number of equations, and due to hardware limitations this method was not feasible for systems higher than 80 × 80. However, for smaller systems, the combination of the two algorithms produces highly accurate results in a reasonable amount of computing time.

Kinetic model

The set of differential expressions in Eqs. 5 and 6 above (again, hairpin reactions are not considered) comprise a model of the kinetics for the reaction phase of microarray hybridization. An ODE solver written in FORTRAN was implemented to find solutions to this set of equations for kinetic simulations up to $N = M = 80$. The number of equations (when $N = M$) grows like N^2, and the size of the Jacobian matrix grows like N^4, making simulations of larger systems highly memory-intensive. In Fig. 1, the computation times for the two types of simulations (*top*, kinetic model; *bottom*, equilibrium model) are shown. Kinetic simulations up to $N = M = 35$ and equilibrium simulations up to $N = M = 100$ were run on a 3.6 GHz Pentium-IV CPU and 1 Gigabyte of RAM, running Windows XP. Similar results were obtained when the simulations were carried out on a 1.8 GHz/512 RAM machine.

Flow chart of the calculation scheme

The ultimate aim of developing this analytical foundation is to have the ability to diagnose each and every hybridization reaction that can conceivably occur in a multiplex reaction scheme. The equilibrium and kinetic models ultimately lead

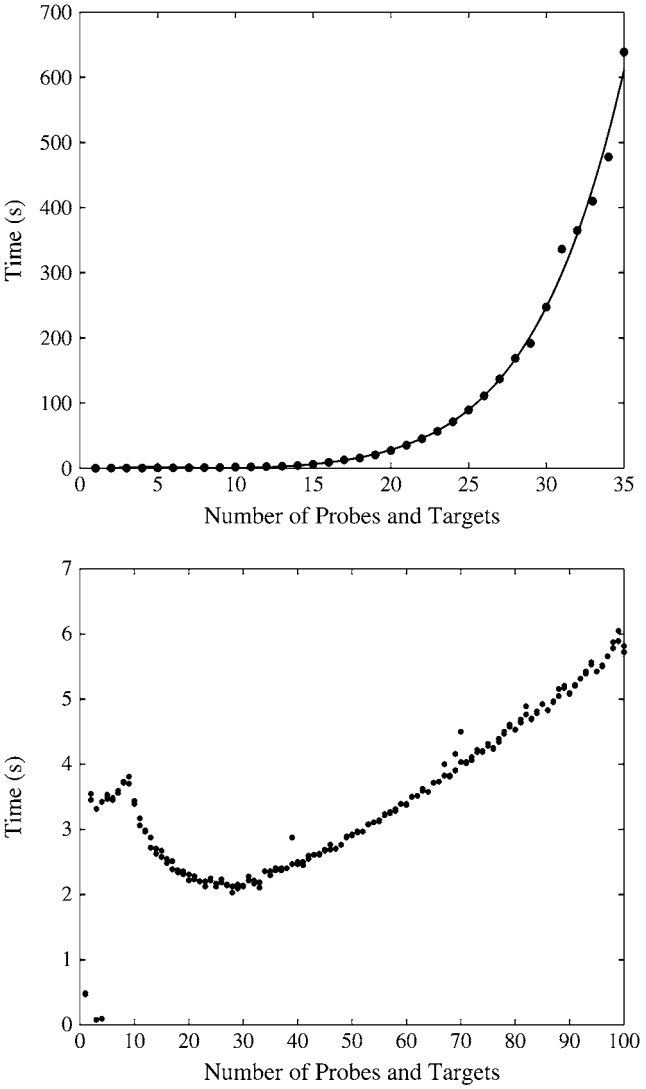

FIGURE 1 Computation time versus number of probes and targets for the kinetic model (*top*) and equilibrium model (*bottom*).

to quantitative predictions of the concentrations of probe: target complexes. Flow charts diagramming steps in the calculational schemes for both the kinetic and equilibrium models are shown in Fig. 2, *a* and *b*.

Although some aspects of these models are very similar, the output produced by the two models is quite different. The kinetic model produces a time series of all concentrations involved in hybridization reactions, while the equilibrium model produces only the equilibrium values of these concentrations (i.e., $\tau = \infty$). Ideally, the kinetic model could be applied to any system, but computational complexity and hardware limitations prohibit this option for very large systems. On the other hand, for large systems, full knowledge of the concentration levels at all times is perhaps unnecessary, and concentrations at different times before equilibrium and the final equilibrium values may be sufficient for many desired applications. In general, combined use of both models

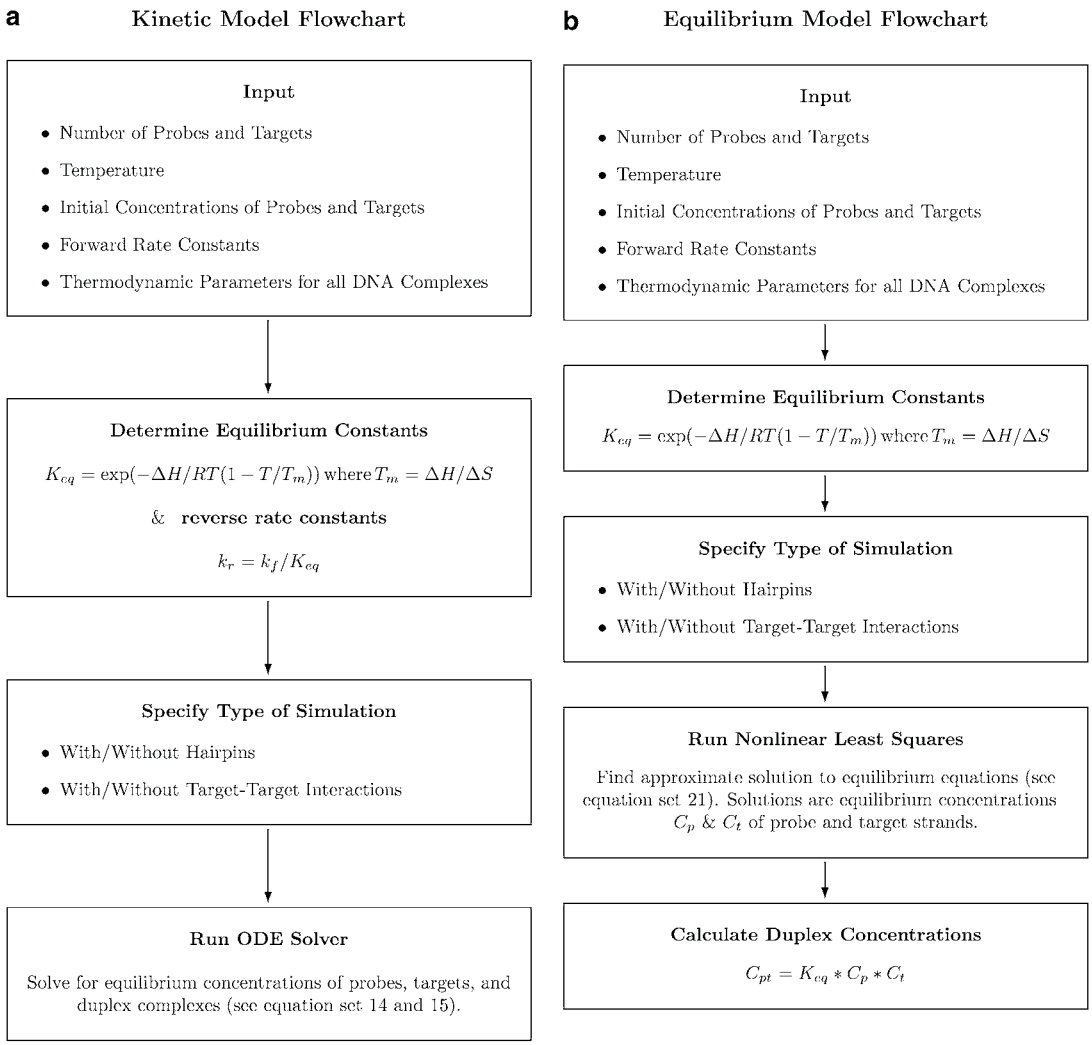

FIGURE 2 Flowcharts of the model calculations. (*a*) Flowchart of the kinetic model calculation. (*b*) Flowchart of the equilibrium model calculation.

is probably most appropriate and expected to produce the best results.

EXAMPLE CALCULATIONS

Model system: definition and rationale

Our general models can, in principle, be applied to multiplex hybridization of any number of probes and targets. In the examples that follow, a simple model system was defined to demonstrate features of the calculation and test sensitivity to a variety of experimental and model parameters. Although described methods are generally applicable to systems of arbitrary dimension, for simplicity, details of a system comprised of three probes and three targets (3×3 system) are considered. Although relatively simple, this system demonstrates the features of added complexity of multiplex hybridization and provides adequate resolution to reveal typical behaviors. Two sets of sequences, set I and set II, were the subject of model calculations. In this regard, this system is an extension of the 2×2 competitive hybridization analysis that has been reported by others (55). An added feature of our system is the inclusion of the specific sequence-dependent stability of perfect-match and cross-hybrid duplexes, and corresponding consideration of the effects of sequence homology on the hybridization process. As seen for set I in Fig. 3 *a*, there is no sequence homology between the three probes or targets comprising the set. In contrast, for set II, two probes P1 and P2 share 50–54% sequence homology, while the sequence P3 has only 25% homology with P1 and 30% homology with P2. All model probe:target duplexes contain 24 basepairs.

Calculation of duplex thermodynamics and stability

For each duplex that can form from any pair of the three probes and three targets in sets I and II, thermodynamic transition parameters, ΔH, ΔS, and ΔG, used in kinetic and

a

P1:	5′ - GCCGGGCTAATAAGGTAGAGATCA - 3′
T1:	3′ - CGGCCCGATTATTCCATCTCTAGT - 5′
P2:	5′ - TTTACCGCGTACGATCGAGTTCTG - 3′
T2:	3′ - AAATGGCGCATGCTAGCTCAAGAC - 5′
P3:	5′ - AGGCTTAACGGTTCAGCTCACAAC - 3′
T3:	3′ - TCCGAATTGCCAAGTCGAGTGTTG - 5′

P1:	5′ - GCCGGGCTAATAAGGTAGAGATCA - 3′
T2:	3′ - AAATGGCGCATGCTAGCTCAAGAC - 5′
P2:	5′ - TTTACCGCGTACGATCGAGTTCTG - 3′
T1:	3′ - CGGCCCGATTATTCCATCTCTAGT - 5′
P3:	5′ - AGGCTTAACGGTTCAGCTCACAAC - 3′
T1:	3′ - CGGCCCGATTATTCCATCTCTAGT - 5′

P1:	5′ - GCCGGGCTAATAAGGTAGAGATCA - 3′
T3:	3′ - TCCGAATTGCCAAGTCGAGTGTTG - 5′
P2:	5′ - TTTACCGCGTACGATCGAGTTCTG - 3′
T3:	3′ - TCCGAATTGCCAAGTCGAGTGTTG - 5′
P3:	5′ - AGGCTTAACGGTTCAGCTCACAAC - 3′
T2:	3′ - AAATGGCGCATGCTAGCTCAAGAC - 5′

b

P1:	5′ - GCCGGGCTAATAAGGTAGAGATCA - 3′
T1:	3′ - CGGCCCGATTATTCCATCTCTAGT - 5′
P2:	5′ - GAAACCGAGTAGAGGTAGAGATCA - 3′
T2:	3′ - CTTTGGCTCATCTCCATCTCTAGT - 5′
P3:	5′ - TTTCTTAACGGTGCCGCCAGATCA - 3′
T3:	3′ - AAAGAATTGCCACGGCGGTCTAGT - 5′

P1:	5′ - GCCGGGCTAATAAGGTAGAGATCA - 3′
T2:	3′ - CTTTGGCTCATCTCCATCTCTAGT - 5′
P2:	5′ - GAAACCGAGTAGAGGTAGAGATCA - 3′
T1:	3′ - CGGCCCGATTATTCCATCTCTAGT - 5′
P3:	5′ - TTTCTTAACGGTGCCGCCAGATCA - 3′
T1:	3′ - CGGCCCGATTATTCCATCTCTAGT - 5′

P1:	5′ - GCCGGGCTAATAAGGTAGAGATCA - 3′
T3:	3′ - AAAGAATTGCCACGGCGGTCTAGT - 5′
P2:	5′ - GAAACCGAGTAGAGGTAGAGATCA - 3′
T3:	3′ - AAAGAATTGCCACGGCGGTCTAGT - 5′
P3:	5′ - TTTCTTAACGGTGCCGCCAGATCA - 3′
T2:	3′ - CTTTGGCTCATCTCCATCTCTAGT - 5′

FIGURE 3 Sequences used in model calculations. Mismatched sequences are underlined. (*a*) Set I is comprised of three probe and target sequences. The probe and target sequences are independent and do not share any homology with one another. (*b*) Set II sequences are comprised of three probe and target sequences. Probes P1 and P2 are 50% and 54% homologous with target strands T2 and T1, respectively. Probes P1 and P2 share 25 and 30% homology, respectively, with target T3.

equilibrium model calculations, were determined from published sequence-dependent thermodynamic parameters (9,10,46,60–64). For each pair of strands all overlap complexes were considered and their corresponding thermodynamic quantities were calculated for the particular duplex sequence at each overlap. Of these states at different alignments, the one having the lowest calculated free energy was selected. For this state, the total helix-coil transition thermodynamics were calculated from the sum of appropriate nearest-neighbor sequence-dependent parameters, where available (65). For example, consider the following hybrid duplex sequence and its decomposition into nearest-neighbor components of the enthalpy, ΔH (mismatches are underlined):

$$\Delta H\begin{pmatrix}\text{AGCGATGA}-\\ -\text{CAATAATT}\end{pmatrix}=\Delta H\begin{pmatrix}\text{AG}\\ \text{-C}\end{pmatrix}+\Delta H\begin{pmatrix}\text{G}\underline{\text{C}}\\ \text{C}\underline{\text{A}}\end{pmatrix}+\Delta H\begin{pmatrix}\underline{\text{CG}}\\ \underline{\text{AA}}\end{pmatrix}+\Delta H\begin{pmatrix}\underline{\text{G}}\text{A}\\ \underline{\text{A}}\text{T}\end{pmatrix}+\Delta H\begin{pmatrix}\text{AT}\\ \text{TA}\end{pmatrix}+\Delta H\begin{pmatrix}\text{T}\underline{\text{G}}\\ \text{A}\underline{\text{A}}\end{pmatrix}+\Delta H\begin{pmatrix}\underline{\text{G}}\text{A}\\ \underline{\text{A}}\text{T}\end{pmatrix}+\Delta H\begin{pmatrix}\text{A-}\\ \text{TT}\end{pmatrix}+\Delta H_{\text{init}}\begin{pmatrix}\text{A}\\ \text{T}\end{pmatrix}+\Delta H_{\text{init}}\begin{pmatrix}\text{G}\\ \text{C}\end{pmatrix}.$$

This duplex contains eight nearest-neighbor interactions, including single-base 5′ dangling ends. The nearest-neighbor dependent parameters, $\Delta H\begin{pmatrix}\text{AG}\\ -\text{C}\end{pmatrix}$, $\Delta H\begin{pmatrix}\text{G}\underline{\text{C}}\\ \underline{\text{C}}\text{A}\end{pmatrix}$, $\Delta H\begin{pmatrix}\underline{\text{CG}}\\ \underline{\text{AA}}\end{pmatrix}$, $\Delta H\begin{pmatrix}\underline{\text{G}}\text{A}\\ \underline{\text{A}}\text{T}\end{pmatrix}$, $\Delta H\begin{pmatrix}\text{AT}\\ \text{TA}\end{pmatrix}$, etc. for the appropriate sequences and interactions as tabulated in Tables 2–4 were utilized. Experimentally derived parameters for w/c perfect matched doublets, single base dangling ends, and single basepair mismatches were available from the published literature (46). Nearest-neighbor tandem mismatches were assigned values as described in Eq. 15 (see below). The actual parameter values employed are summarized in Tables 2–4. In addition, two initiation factors, $\Delta H_{\text{init}}(\text{A}\quad\text{T})$ and $\Delta H_{\text{init}}(\text{G}\quad\text{C})$ were assigned depending on the particular identities of the end basepairs. Values of the initiation thermodynamic parameters that were employed are

$$\Delta H_{\text{init}}\begin{pmatrix}A\\ T\end{pmatrix}=2.3\text{kcal/mol},\ \Delta S_{\text{init}}\begin{pmatrix}A\\ T\end{pmatrix}=4.1\text{kcal}\ K^{-1}\ \text{mol}^{-1},$$

$$\Delta H_{\text{init}}\begin{pmatrix}G\\ C\end{pmatrix}=0.1\text{kcal/mol},\ \Delta S_{\text{init}}\begin{pmatrix}G\\ C\end{pmatrix}=-2.8\text{kcal}\ K^{-1}\ \text{mol}^{-1}.$$

Furthermore, recall the formulas for total free energy, $\Delta G = \Delta H T\Delta S$, and $T_{\text{m}} = \Delta H/\Delta S$, $\Delta G = \Delta H(1 - T/T_{\text{m}})$.

For strands resulting in hybrid complexes containing tandem mismatches, quantitative prediction of their stabilities

TABLE 2 Nearest-neighbor thermodynamic parameters for w/c doublets (61–64)

W/C doublet	Enthalpy(cal/mol)	Entropy (cal/kmol)
AA	−7900	−22.2
AC	−8400	−22.4
AG	−7800	−21.0
AT	−7200	−20.4
CA	−8500	−22.7
CC	−8000	−19.9
CG	−10,600	−27.2
GA	−8200	−22.2
GC	−9800	−24.4
TA	−7200	−21.3

is a little less certain. Although there are nearest-neighbor parameters for single basepair mismatches for nearly all of the possible nearest-neighbor combinations (Table 4 below) (10), a parameter set for tandem mismatches does not currently exist. In some of the example calculations that were performed, the influence of the relative thermodynamic stability of tandem mismatches was investigated. For the purpose of example, tandem mismatches were assigned thermodynamic parameter values that were a fraction of the corresponding w/c basepair doublet values, i.e., the free-energy of a mismatch basepair doublet in a tandem mismatch complex was assigned according to

$$\Delta G_{\mathrm{MM}} = \kappa \Delta G_{\mathrm{PM}} = \kappa(\Delta H_{\mathrm{PM}} - T\Delta S_{\mathrm{PM}}), \qquad (15)$$

where ΔG_{PM}, ΔH_{PM}, and ΔS_{PM} are the free energy, enthalpy, and entropy, respectively, for melting a hydrogen-bonded w/c basepair doublet. The factor κ was introduced as a means of scaling values of thermodynamic parameters of mismatch basepairs in tandem mismatches as a relative fraction of the stability of w/c perfect matches. In examples that were performed, tandem mismatches were treated in two ways. They were either assumed to be minimal, $\kappa = 0$, or assigned a value of $\kappa = 0.5$. Although consideration of tandem mismatches in this manner is clearly an oversimplified generalization, it provided a convenient means of universally weighting non-w/c tandem mismatch pair interactions differently than w/c basepairs, and discerning potential effects of tandem mismatch stability on multiplex hybridization.

TABLE 3 Sequence-dependent parameters for dangling ends (60)

Dangling end	Enthalpy (cal/mol)	Entropy (cal/Kmol)	Dangling end	Enthalpy (cal/mol)	Entropy (cal/Kmol)
TA/-T	−6900	−20.0	CG/G-	−3200	−10.4
AC/-G	−6300	−17.1	AG/-C	−3700	−10
CA/G-	−5900	−16.5	AT/-A	−2900	−7.6
GT/-A	−4200	−15.0	CC/G-	−2600	−7.4
CT/G-	−5200	−15.0	-C/AG	−2100	−3.9
GC/-G	−5100	−14.0	TG/A-	−1600	−3.6
TG/-C	−4900	−13.8	GA/-T	−1100	−1.6
AG/T-	−4100	−13.1	AA/T-	−500	−1.1
-C/TG	−4400	−13.1	TA/A-	−700	−0.8
CT/-A	−4100	−13.0	TT/-A	−200	−0.5
CC/-G	−4400	−12.6	-C/CG	−200	−0.1
AT/T-	−3800	−12.6	AA/-T	200	2.3
CG/-C	−4000	−11.9	CA/-T	600	3.3
-C/GG	−3900	−11.2	TT/A-	2900	10.4
GG/-C	−3900	−10.92	AC/T-	4700	14.2
TC/-G	−4000	−10.9	TC/A-	4400	14.9

Theoretical results

Single-channel versus multichannel hybridization

Our approach adds several new features to modeling of the reaction-phase of the multiplex hybridization process. For a multiplex mixture comprised of any numbers of probes and targets, duplexes formed between perfect match probes and targets, as well as all cross hybrids formed from mismatched probes and targets, i.e., cross hybrids, are all considered collectively in a multichannel, system approach. There are multiple reaction channels available to every probe and target. In addition, sequence-specific effects are considered for both perfect match duplexes and mismatched duplexes containing some amount of basepairing. Consideration of the potential fractional stability of tandem mismatches and associated stronger weighting of mismatch complexes is also included.

The multichannel reaction model is inherently more complex than the more common single channel approach, where hybridization reactions between perfect match probe:target pairs are considered in isolation. Practically, when the single channel model is employed, potential cross-reactions between mismatched probe:target pairs are either screened and filtered according to their degree of sequence homology, or ignored entirely (66). For the model system comprised of three probes and three targets that will be examined in detail, the single-channel calculation (as so defined) considers independently the hybridization behaviors of only the three perfect match probe:target complexes. With added complexity the multichannel model considers the interdependent behaviors of nine different duplexes, i.e., three perfectly matched, and six cross-hybrids that contain mismatches. As will be seen, both the kinetic and equilibrium behaviors of the three perfect match probe:target duplexes are influenced significantly by added considerations inherent in the multichannel model.

Plots in Fig. 4, *a–d*, show the results of the single channel (*top panels*) and multichannel (*bottom panels*) calculations for a simple 3 × 3 system. Kinetic curves (duplex concentration versus time) that are displayed were calculated for set I (Fig. 4, *a* and *c*) and set II (Fig. 4, *b* and *d*) at 37°C for the single-channel and multichannel models. Plots in Fig. 4 also show effects of assigning differential stability to tandem mismatches and essentially giving more weight to cross-hybrids. Plots in Fig. 4, *a* and *b*, were obtained by assuming the free energy of tandem mismatches as 0, i.e., $\kappa = 0$, while those in Fig. 4, *c* and *d*, were obtained assuming a universal constant factor, $\kappa = 0.5$ (to scale tandem-mismatch free energies relative to normal w/c basepairs).

TABLE 4 Sequence-dependent thermodynamic parameters for single basepair mismatches (72)

MM	Enthalpy (cal/mol)	Entropy (cal/Kmol)	MM	Enthalpy (cal/mol)	Entropy (cal/Kmol)	MM	Enthalpy (cal/mol)	Entropy (cal/Kmol)
GC/GG	−6000	−15.8	CC/GT	−800	−4.5	GA/GT	1600	3.6
CT/GT	−5000	−15.8	AC/CG	−700	−3.8	CA/GC	1900	3.7
CG/GG	−4900	−15.3	AG/TA	−700	−2.3	AA/TC	2300	4.6
GC/TG	−4400	−12.3	CA/GG	−700	−2.3	GC/CT	2300	5.4
CG/GT	−4100	−11.7	AA/TG	−600	−2.3	AA/GT	3000	7.4
AG/GC	−4000	−13.2	GA/CG	−600	−1.0	GG/CT	3300	10.4
AG/TG	−3100	−9.5	TA/GT	−100	−1.7	CA/AT	3400	8.0
AC/AG	−2900	−9.8	AC/TC	0	−4.4	CC/CG	3600	8.9
CT/GG	−2800	−8.0	TA/TT	200	−1.5	GT/TG	4100	9.5
AT/TT	−2700	−10.8	AC/GG	500	3.2	AA/AT	4700	12.9
AT/TG	−2500	−8.3	AG/CC	600	−0.6	CC/AG	5200	14.2
GT/CT	−2200	−8.4	AC/TT	700	0.2	CC/TG	5200	13.5
CC/GC	−1500	−7.2	GA/AT	700	0.7	GA/CC	5200	14.2
CG/TC	−1500	−6.1	AG/TT	1000	0.9	AC/TA	5300	14.6
TT/AG	−1300	−5.3	TT/AC	1000	0.7	GG/TT	5800	16.3
AT/TC	−1200	−6.2	AA/TA	1200	1.7	CA/CT	6100	16.4
AG/AC	−900	−4.2	TA/CT	1200	0.7	AA/CT	7600	20.2

Examination and comparison of the plots in Fig. 4 reveals specific features of the model calculations and how they are influenced by sequence homology among the probes and targets, and tandem mismatch stability. First, compare the curves in Fig. 4, *a* and *b* ($\kappa = 0$). Fig. 4 *a* is for set I, where there is no sequence homology between the different probe and target sequences. In the top panel for the single-channel calculation, three curves are shown corresponding to the three perfect match duplexes. These curves exhibit typical exponential behavior. Plots in the bottom panel correspond to the multichannel calculation where colored lines are the perfect matches and gray lines are the mismatched cross-hybrids. Although there are nine curves generated in the multichannel calculation for set I, the curves for mismatches

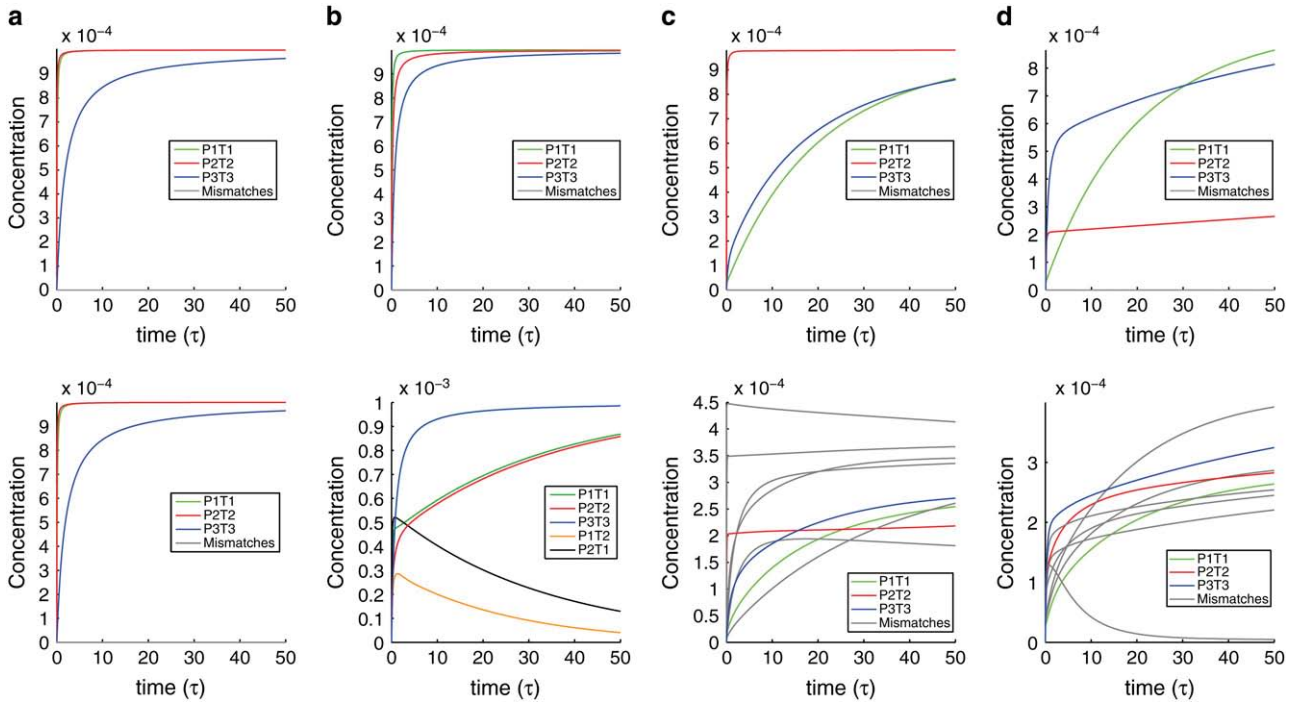

FIGURE 4 Kinetic plots (time versus concentration) for the 3 × 3 systems comprised of sequence sets I and II (Fig. 4) at constant temperature, 37°C. Curve assignments are given in the individual figures. (*a*) Set I. Single-channel (*top*) and multichannel calculation (*bottom*). (*b*) Set II. Single-channel (*top*) and multichannel calculation (*bottom*). Curves in panels *a* and *b* were generated assuming a minimal contribution of tandem mismatches to cross-hybrid duplex stability ($\kappa = 0$). (*c*) Set I. Single-channel (*top*) and multichannel (*bottom*) calculations assuming tandem mismatch stability is comparable to w/c basepair doublet stability ($\kappa = 0.5$). (*d*) Set II. Single-channel (*top*) and multichannel (*bottom*) calculations. All curves were generated assuming negligible stability for tandem mismatches ($\kappa = 0.5$).

(or cross-hybrids) lie along the horizontal axis and are not visible. As seen by comparison of the curves in the upper and lower panels of Fig. 4 *a*, for set I, kinetic curves for the perfect match duplexes are essentially identical in both the single-channel and multichannel calculations, and display the typical exponential behavior. The comparison in Fig. 4 *a* shows when there is no sequence homology between the different probe and target strands, and there is no extra weighting for tandem mismatches ($\kappa = 0$), the single-channel and multichannel results for the perfect match duplexes are indistinguishable.

Calculated kinetic curves for set II are shown in the upper and lower panels in Fig. 4 *b*. Recall that, in set II, probes 1 and 2 share at least 50–54% sequence homology with targets 2 and 1, respectively, and probe 3 is 25% homologous with targets 1 and 2. Kinetic curves in the upper panel of Fig. 4 *b* are from the single-channel calculation and exhibit typical exponential behavior.

In the multichannel calculation for set II, behavior of the curves displayed in the lower panel of Fig. 4 *b* contrasts to what was observed for set I. In the multichannel calculation, when $\kappa = 0$, kinetic curves for the perfect match duplexes of set II differ considerably from those obtained in the single-channel calculation. In this case, concentrations of the perfect match complexes, P1:T1 and P2:T2, do not increase as rapidly as in the single-channel calculation (*upper panel*, Fig. 4 *b*) and take longer to level off to equilibrium values. Apparently, due to the reduction of available single strands from competition with cross-hybrids, particularly at early times, the perfect match complexes reach equilibrium more slowly, and achieve lower equilibrium values at comparable times. This is evident from the kinetic curves in Fig. 4 *b* for the dominant mismatch complexes, P1:T2 and P2:T1 (*black* and *orange lines* in the *lower panel* of Fig. 4 *b*), which increase at early times attaining concentrations comparable with the perfect matches, then decrease exponentially to their equilibrium concentrations. The extent to which the competition for resources affects formation of the perfect match hybrids depends directly on the amount of sequence homology among the strands. Note that, for set II, the perfect-match complexes, P1:T1 and P2:T2, approach equilibrium much more slowly than the P3:T3 complex. Consistent with this scenario, the P3:T3 complex suffers less dramatic effects of cross-hybridization. The P3 and T3 strands share only 25% homology with the other strands. Consequently, more of them are available to form the perfect-match complex resulting in much faster approach to equilibrium than the other perfect match complexes. The curves for the P1:T2 and P2:T1 cross-hybrids (*black* and *orange lines* in plot in *lower panel*, Fig. 4 *b*) bracket the values for the perfect matches P1:T1 and P2:T2 at very early times, but then decrease exponentially with time to reach equilibrium. Some of these curves demonstrate a stark departure from the kinetic behavior seen thus far. Apparently, competition from mismatches reduces the amount of corresponding perfect matches containing the same strands. In Fig. 4 *b* the curves for the P1:T2 and P2:T1 mismatches are the only ones visible on the scale of perfect matches. In summary, when tandem mismatches have marginal stability ($\kappa = 0$), cross-hybrid formation is largely dominated by the amount of sequence homology shared by the probe and target strands.

When tandem mismatches have a significant fraction of the stability of normal w/c basepairs, $\kappa = 0.5$ (Fig. 4, *c* and *d*), results are much more dramatic. For sets I and II, in the single-channel calculation, assigning more stability to tandem mismatches has no effect, since mismatch cross-hybrids are not considered. As a result, for the single-channel calculations, when $\kappa = 0.5$ for sets I and II (*upper plots* in Fig. 4, *c* and *d*), plots are identical to those in Fig. 4, *a* and *b*, respectively. However, in the multichannel calculation, when $\kappa = 0.5$, cross-hybrids are given more stability. Because of increased competition from mismatches, all the perfect matches reach lower equilibrium values much more slowly than in the single-channel model calculation. When $\kappa = 0.5$, some mismatches are actually found to reach higher concentrations than the perfect matches (which seems unlikely). Note that all perfect-match complexes reach lower equilibrium values than observed in the single-channel calculation. The scale of the plot is different in order to help visualize the qualitative behavior. Plots in the bottom panel of Fig. 4 *c* for set I underscore the significant effects of tandem mismatch stability on the multiplex hybridization process, even when no sequence homology exists between the probes and targets. In effect, stable cross-hybrids compete for the perfect-match probe and target strands. As a consequence of this competition and resulting resource depletion, the perfect-match complexes approach equilibrium much more slowly in the multichannel calculation compared to the single-channel calculation.

For set II, when $\kappa = 0.5$ (*lower panel* of Fig. 4 *d*), competition from additional cross-hybrids formed with the aid of sufficient sequence homology acts to decrease levels of the perfect matches and the most significant cross-hybrids, compared to what was observed in the bottom panel of Fig. 4 *b*. In a dynamic sense, sequence homology and heightened stability of tandem mismatches acts to increase concentrations of all mismatch cross-hybrids. Only one hybrid complex (P3:T1) displays contrasting behavior, first attaining levels of the perfect-match duplexes, then decreasing exponentially to equilibrium. Note that there are only two mismatch complexes observable in Fig. 4 *b*, *bottom panel*. This enhanced cross-hybridization acts to decrease rates of hybridization for perfect match complexes (which must compete for resources with the cross-hybrids). Under this environment the equilibrium concentrations of the perfect match duplexes are also depressed.

In summary, assigning significant stability to tandem mismatches ($\kappa = 0.5$) has a much larger effect on the kinetic behaviors and distributions of equilibrium concentrations of the perfect-match duplexes than sequence homology, which

also enhances cross-hybridization. Although probably somewhat unrealistic in that the magnitude of tandem-mismatch interactions is surely not so large in all cases (as has been assumed), this example helps us discern the interrelated effects of sequence homology and tandem-mismatch stabilization. It further demonstrates relative sensitivity of the calculations to influences of different potential sources of cross-hybridization. Examples in Fig. 4, *b* and *d*, also reveal a different type of kinetic behavior of some mismatch complexes. In contrast to the typical exponential behavior, these curves increase rapidly then decrease exponentially. This behavior was further revealed when the temperature-dependence of the kinetic behavior for the perfect-match and mismatch complexes present in the 3 × 3 system was examined.

Effects of temperature

Effects of temperature on the kinetics of two-strand complex formation are depicted in Fig. 5. Each of the nine plots shown

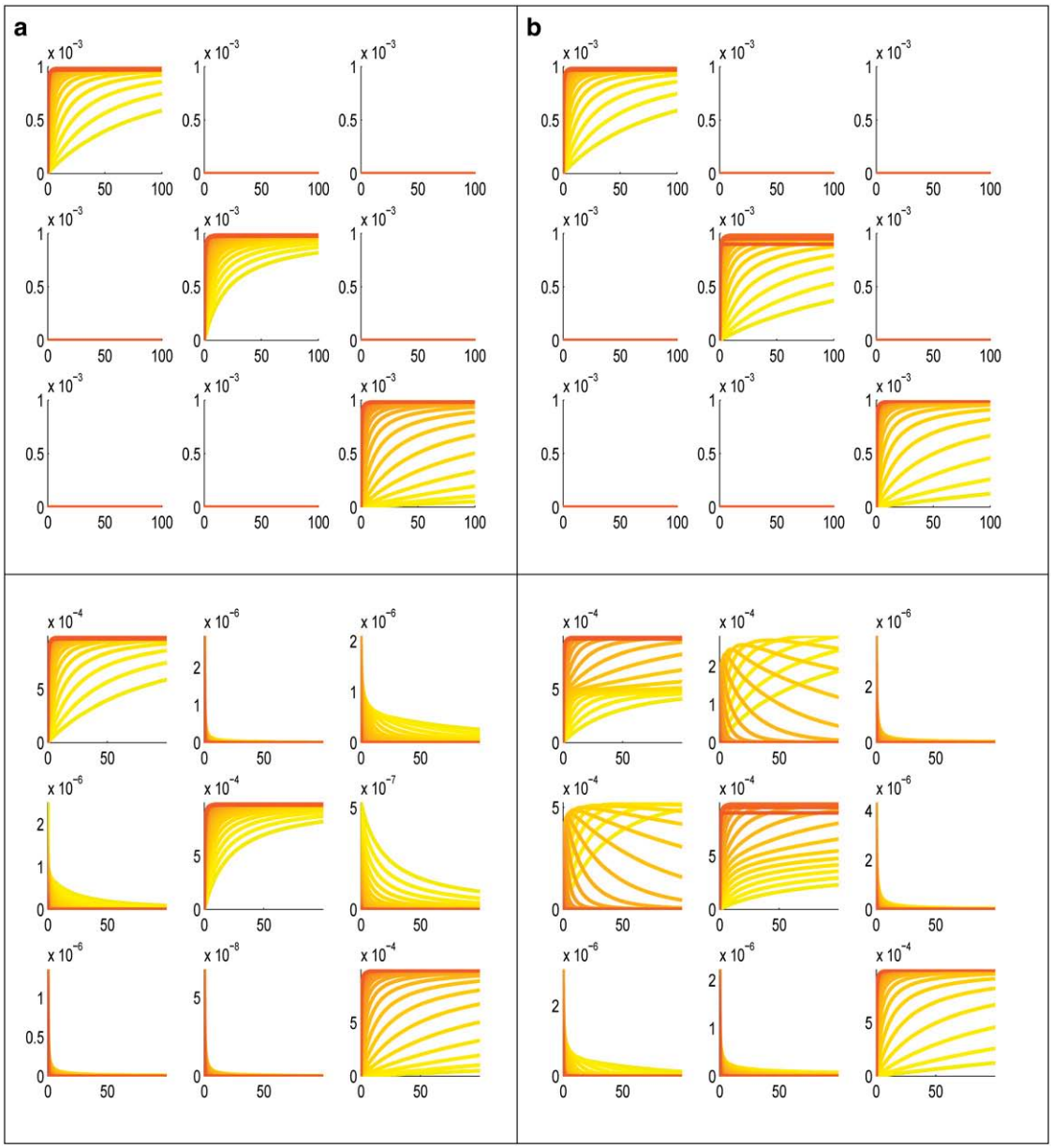

FIGURE 5 Kinetic plots (time versus concentration) for the 3 × 3 systems comprised of sequence sets I and II (Fig. 4) as a function of temperature from 37 (*yellow*) to 97°C (*red*). (*a*) Set I. Single-channel calculation (*top panel*) and multichannel calculation (*bottom panel*). Columns are probes, rows are targets. Thus, the upper-left plot in either panel corresponds to the P1:T1 complex, while the third column, second-row plot is for the P2:T3 complex. (*b*) Set II. Single-channel calculation (*top panel*) and multichannel calculation (*bottom panel*). All curves were generated assuming negligible stability for tandem mismatches ($\kappa = 0$).

in the panels of these figures are of duplex concentrations versus time for the different probe:target complexes (probes are rows, targets are the columns) at different temperatures. Predictions from the single-channel (*top*) and multichannel (*bottom*) model calculations are shown. Plots on the diagonal are for the perfect matches, and off-diagonal plots are cross-hybrids. For example, the plot in the upper-left corner is for the P1:T1 complex (concentration versus time of the perfect-match complex formed from probe 1 and target 1). The plot in the first row, second column corresponds to the P1:T2 complex, and is the concentration as a function of time of the cross-hybrid complex formed from probe 1 and target 2. Likewise, the plot in the second row, third column is for the P2:T3 complex, the concentration versus time of the cross-hybrid formed from probe 2 and target 3. In each case, temperature was varied over the 60° range from 310 K (*yellow*) to 370 K (*red*). Note that the scales in the plots of Fig. 5 are not all equal. Different scales were used to emphasize and compare qualitative behaviors.

Results of the calculation for set I (no sequence homology) are shown in Fig. 5 *a*. For this set of sequences, results from the single-channel calculation are unremarkable. Plots display familiar exponential behavior with the concentration of complexes increasing with increased temperature. Cross-hybrids are not considered in the single-channel calculation, thus, only plots along the diagonal appear at the top of Fig. 5, *a* and *b*. In the multichannel calculation, cross-hybrids (off-diagonal plots) are significant and display essentially the opposite behavior of the perfect-match complexes. As temperature is increased (from *yellow* to *red*), the mismatches initially at high concentrations (in some cases at or near the perfect match concentrations) decrease exponentially with increasing time.

As shown in Fig. 5 *b*, results are different for the set II sequences that share from 25 to 54% homology with one another. For the single-channel calculation, results are essentially identical to what was observed for set I. However, in the multichannel calculation, sequence homology has a significant effect on the cross-hybrids containing that homology. In particular, the curves in Fig. 5 *b* for the P1:T2 and P2:T1 that share at least 50% homology are considerably different than seen in Fig. 5 *a*, where the sequences share no sequence homology. In essence, the kinetic behavior and level of cross-hybridization seen in these plots indicates the level of sequence homology present in the complexes.

A notable observation emerges from this analysis. The kinetic plots for the P1:T2 and P2:T1 complexes obtained from the multichannel calculation display two distinctly different behaviors. Similar observations were also made for these mismatches in the multichannel calculation (*bottom panels* of Fig. 4, *b* and *d*). These contrasting behaviors are referred to as ''overdamped'' and ''critically damped'' because of their similarity with the behavior of damped harmonic oscillators. For overdamped behavior, the concentration of hybridized complexes first increases at early times, then decreases exponentially with time, and levels off smoothly to some equilibrium value. In contrast, for critically damped behavior (what is typically observed) the hybridized concentration increases with time exponentially and converges to a constant value. In certain cases, depending on the temperature, both types of behavior can be observed for the same complex. The transition from critically damped to overdamped behavior seems to occur as temperature is increased, relative to the melting temperature T_m, for that particular complex. It can be shown that the maximum value of an overdamped duplex (one that increases at first, then decreases to an equilibrium value) is an unstable local equilibrium for the subsystem consisting of only that duplex and its probe:target pair. The exact mechanism underlying these observations, and when each type of kinetic behavior might be expected, is currently under investigation (D. J. Fish and A. S. Benight, unpublished).

In summary, effects of temperature on hybridization kinetics of mismatch complexes are manifest in two distinguishable kinetic regimes, termed ''critically damped'' and ''overdamped''. Which regime is occupied depends strongly on the relative stability of the hybrid duplex, i.e., the melting temperature of the complex relative to the assay temperature, and homology (cross-reactivity) with other sequences of the system.

Effects of differential strand concentrations

The influence of strand concentration and sequence homology on both the single channel and multichannel calculations, for sets I and II, is shown in Fig. 6. To explore the interdependent effects of concentration and sequence homology, three different scenarios were considered. In the first case (Fig. 6 *a*), concentration of target 1 was set to be 100 times higher than that for targets 2 and 3. In the second case (Fig. 6 *b*), concentration of target 2 was 100 times higher than targets 1 and 3. And in the third case (Fig. 6 *c*), concentration of target 3 was 100 times higher than targets 1 and 2. In each case, probe concentrations were equal to the concentration of the two lesser concentrated targets. On the plots in Fig. 6, multichannel calculations for set I (*left*) and set II (*right*) are displayed. Upper curves, provided for comparison, are the same in Fig. 6, *a–c*, and correspond to the situation in the multichannel calculation when concentrations of all strands are equal. In all cases, temperature was 315 K.

When all probe and target concentrations are equal for set I (*top left*, Fig. 6, *a–c*), the cross-hybrids all display the same decaying behavior in time, and are similar. The P1:T1 and P2:T2 perfect-match complexes form faster than P3:T3, and reach the equilibrium concentrations faster. When there is sequence homology among the probes and targets (set II), the situation is more complicated. Even at equal concentrations of probes and targets, sequence homology of the P1:T2 and P2:T1 complexes in set II results in considerable cross-hybridization. Note the upper left and right off-diagonal plots for these complexes, which are comparable with the perfect-

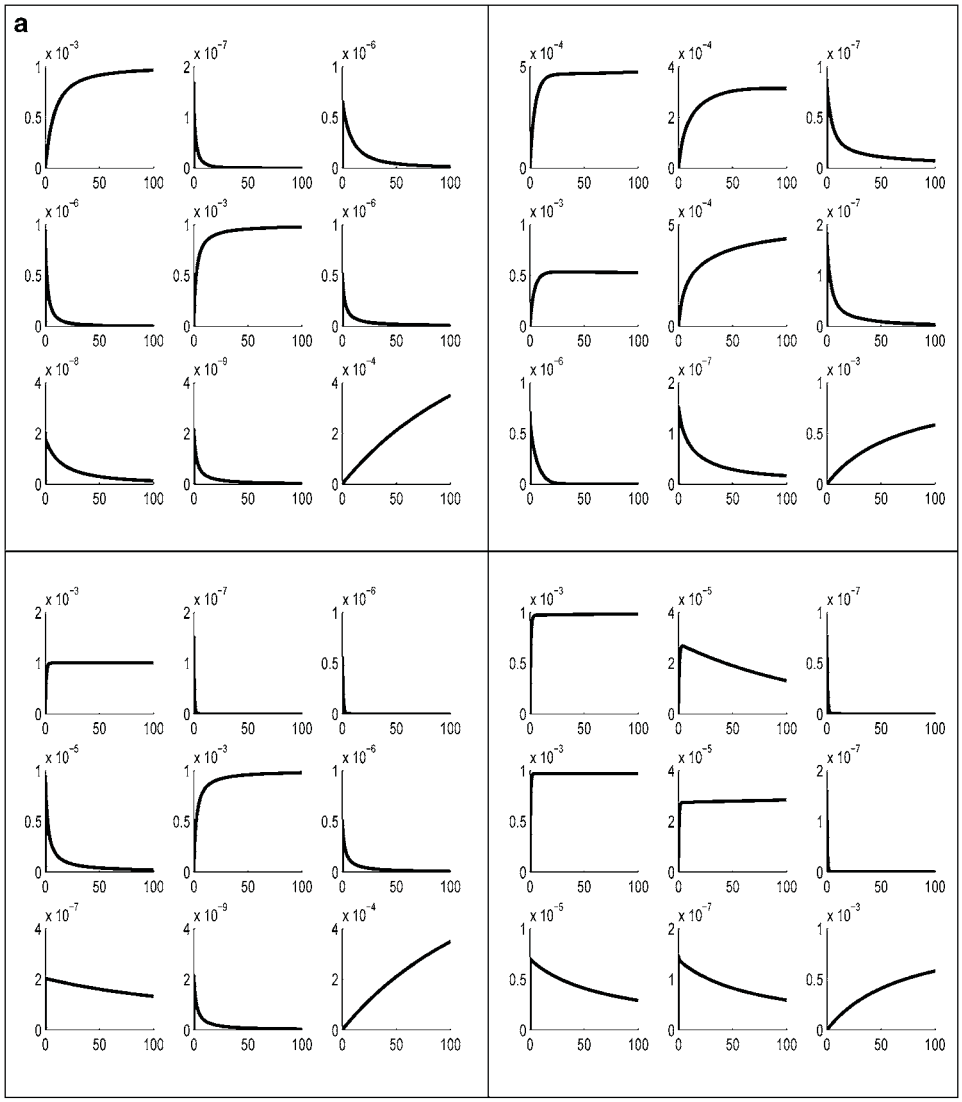

FIGURE 6 Kinetic plots (time versus concentration) for the 3 × 3 systems comprised of sequence sets I and II (Fig. 3) as a function of differential target concentrations at constant temperature, 42°C. Results from the multichannel calculation for set I are on the left, set II on the right. In the panels that are shown there are nine different plots. In these plots, columns are targets and rows are probes. Thus, the plot in the second column, third row corresponds to the P3:T2 complex, etc. Top panels are results when concentrations of the target and probe strands are all equivalent and are replicated in each figure for comparison. Bottom panels correspond to results obtained in the multichannel calculation for set I (*left*) and set II (*right*) when: (*a*) concentration of target 1 is 100 times that of the other target and probe strands, i.e., [T1] = 100 [T2] = 100 [T3]; (*b*) [T2] = 100 [T1] = 100 [T3]; and (*c*) [T3] = 100 [T1] = 100 [T2].

match complexes P1:T1 and P2:T2. At a 100-times' greater concentration for T1 compared to T2 and T3 (Fig. 6 *a*, *lower right*), the P2:T1 cross-hybrid is greatly enhanced, but not the P1:T2 cross-hybrid. For formation of the perfect-match complexes, the P1:T1 perfect match reaches the highest equilibrium value as does the P2:T1 cross-hybrid. In contrast, the P1:T2 cross-hybrid and P2:T2 perfect-match reach much lower equilibrium values. Here again is a case of resource depletion affecting the outcome of the process. The higher concentration of T1 depletes the amount of P1 and P2 through formation of the P1:T1 and P2:T1 complexes. As a consequence of this resource depletion, there is less of P1 to form the P1:T2 cross-hybrid or P2 to form the P2:T2 perfect match complex, resulting in their relatively lower (compared to P1:T1 and P2:T1) equilibrium concentrations.

The situation is similar when [T2] = 100 [T1] = 100 [T3] (Fig. 6 *b*). For set I, the remarkable effect is to speed up duplex formation with increased equilibrium concentration of the P2:T2 perfect-match complex and the P3:T2 cross-hybrid complex. For set II, the P2:T2 perfect-match complex and the P2:T1 and P1:T2 cross-hybrids reach the highest equilibrium values. The P3:T1 and P3:T2 cross-hybrids also increase. These results are entirely analogous to those displayed in Fig. 6 *a*, except the complexes dominated by T2 (P1:T2 and P2:T2) have the highest rates of formation and reach the highest equilibrium values. The P3:T2 cross-hybrid also increases. Meanwhile, the P1:T1 and P2:T1 complexes, although they form faster, reach lower equilibrium values. Here, depletion of P1 and P2, through formation of the P1:T2 cross-hybrid and P2:T2 perfect match, respectively, driven by the relatively higher concentration of T2, results in lower equilibrium values of the P1:T1 and P2:T2 complexes. The above interpretations are validated by the plots in Fig. 6 *c*. Here, T3=100 [T1]=100 [T2]. Recall, in this case, T3 and P3 share only 25% homology with the other probes and targets in set II. As observed for the other situations, the effect of increasing T3 is to increase the rate of formation and final equilibrium value of the P3:T3 perfect-

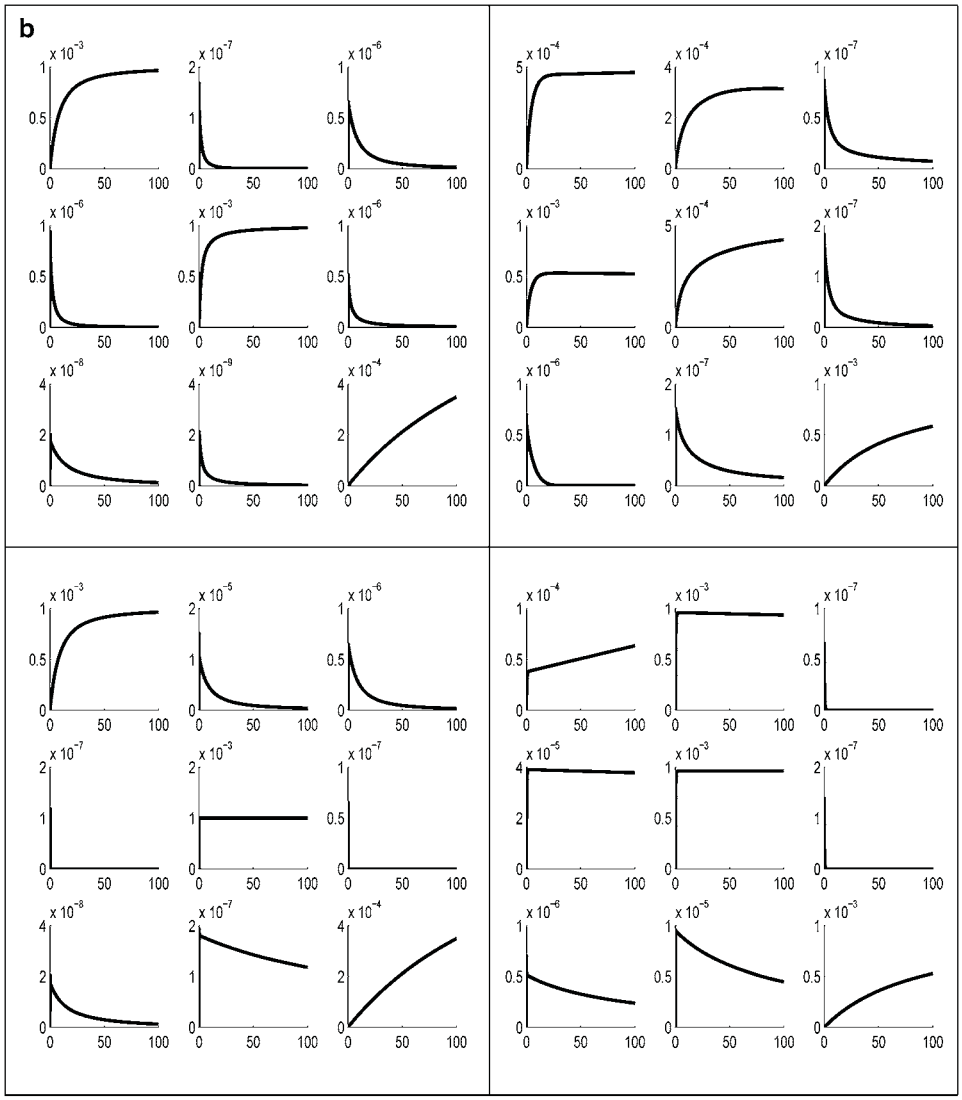

FIGURE 6 Continued

match complex. However, there is apparently little effect on formation of the cross-hybrids (regardless of sequence homology). Effects of resource depletion due to cross-hybrid formation are not observed for the higher (more nonhomologous) T3 target concentration.

Results of increasing the concentration of one of the target strands can be summarized as follows. Sequence homology alone has the largest effect on the cross-hybrid kinetics, but higher concentration of one strand enhances formation of perfect-match and cross-hybrid complexes. Formation of these complexes, driven by the strand in excess concentration, then acts to deplete the strands it binds from the reaction, thus reducing formation of other cross-hybrids or even perfect-match complexes.

COMPARISONS WITH PREVIOUS WORK

In similar ways from different perspectives, several authors have approached the problem of defining hybridization error (41,61–74). The following provides brief descriptions and comparisons of some of these approaches, and how results generated from them compare with ours. In some cases, our results are identical; for others, some differences are observed.

Quantification of degeneracy and hybridization error

Rose and co-workers introduced the concept of hybridization error (67–72). Their approach is based on definition of the ''computational incoherence.'' However, their definition did not take into account certain effects of concentration. It will be seen that consideration of concentration effects can have significant consequences on the hybridization error. This is clearly evident from the plots shown below in Fig. 7, where the calculated error determined with and without concentration considerations is plotted versus temperature (explained below). For multiplex hybridization reactions, the probability of error was defined (68,72) as

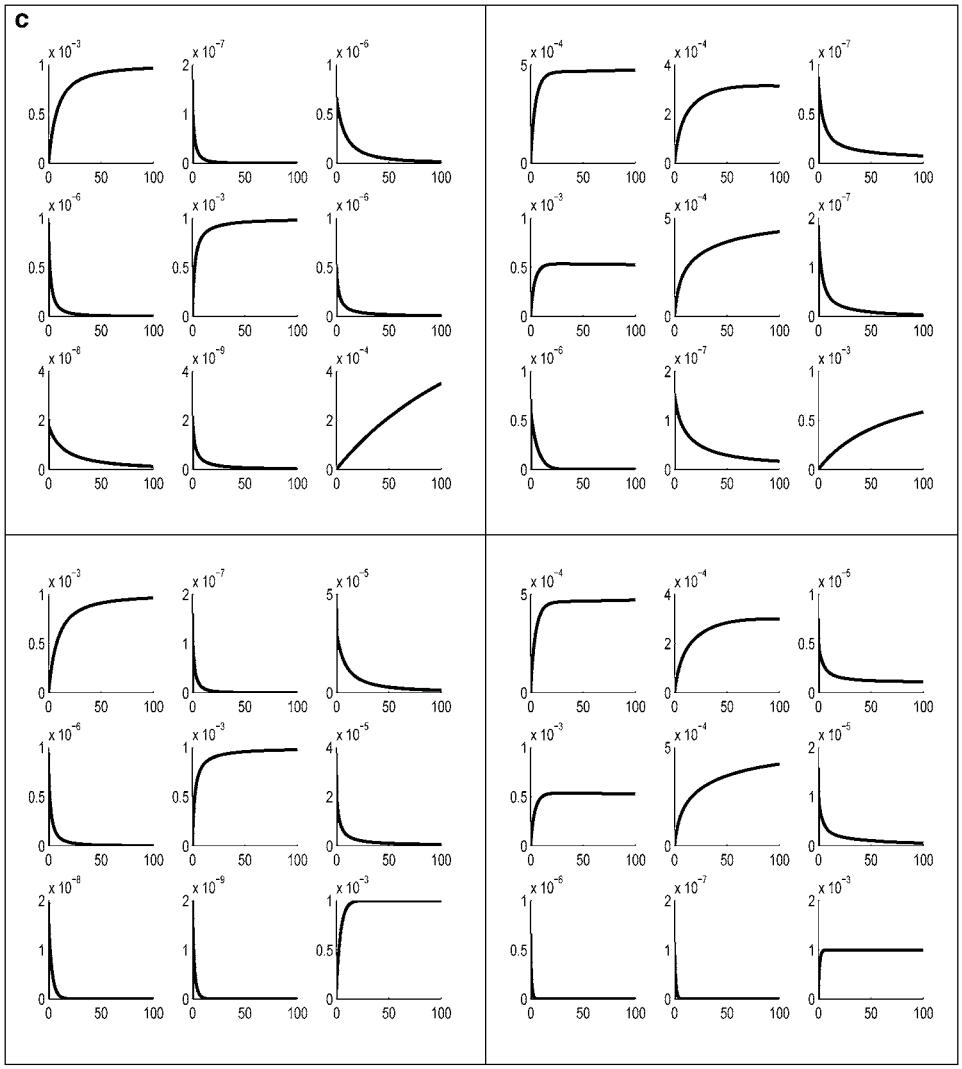

FIGURE 6 Continued

$$P_{\text{error}} = \frac{\text{concentration of error configurations}}{\text{concentration of all configurations}}. \quad (16)$$

A description of different estimates of error made by Rose and co-workers (61,71) follows. Denoting a duplex configuration formed from single strands p_i and t_j as $p_i t_j$, defining E as the set of all error duplexes, and invoking the definition in Eq. 16, the error may be written as

$$P_{\text{error}} = \frac{\sum\limits_{\mathrm{p_i t_j} \in \mathrm{E}} C_{\mathrm{p_i t_j}}}{\sum\limits_{\mathrm{i,j}} C_{\mathrm{p_i t_j}}} = \frac{\sum\limits_{\mathrm{p_i t_j} \in \mathrm{E}} C_{\mathrm{p_i}} C_{\mathrm{t_j}} K_{\mathrm{p_i t_j}}}{\sum\limits_{\mathrm{i,j}} C_{\mathrm{p_i}} C_{\mathrm{t_j}} K_{\mathrm{p_i t_j}}} = \frac{\sum\limits_{\mathrm{i \neq j}} C_{\mathrm{p_i}} C_{\mathrm{t_j}} K_{\mathrm{p_i t_j}}}{\sum\limits_{\mathrm{i,j}} C_{\mathrm{p_i}} C_{\mathrm{t_j}} K_{\mathrm{p_i t_j}}}, \quad (17)$$

where $K_{\mathrm{p_i t_j}} = \exp\left(\frac{-\Delta G^0_{\mathrm{p_i t_j}}}{\mathrm{RT}}\right)$ is the equilibrium constant for the reaction involving formation of the duplex i, j. A further assumption imposed by these authors is that the concentration of single-strand species is small relative to the concentration of hybridized strands, so that $C_i << C_i^\circ$. Invoking this assumption in Eq. 17 leads to a simplified expression, as in

$$P_{\text{error}} = \frac{\sum\limits_{\mathrm{p_i t_j} \in \mathrm{E}} K_{\mathrm{p_i t_j}}}{\sum\limits_{\mathrm{i,j}} K_{\mathrm{p_i t_j}}} = \frac{\sum\limits_{\mathrm{i \neq j}} K_{\mathrm{p_i t_j}}}{\sum\limits_{\mathrm{i,j}} K_{\mathrm{p_i t_j}}}. \quad (18)$$

Furthermore, the following assumptions are also invoked (61,71):

1. Probes are present in equal and excess concentrations of targets.
2. Total equilibrium constants for target:target, probe:probe, and partial overlap configurations are small relative to the full length (designed) target:probe pairs.
3. Equilibrium constants of all mismatch target-probe pairs are small relative to the perfect-match equilibrium constant, i.e., $K_{\mathrm{p_i t_j}} \ll K_{\mathrm{p_k t_k}}$ for all k and $i \neq j$.
4. Equilibrium constants of hairpin formation for each probe and corresponding target are roughly equal, i.e., $K^{\mathrm{hp}}_{\mathrm{t_i}} \approx K^{\mathrm{hp}}_{\mathrm{p_i}}$.

With these additional assumptions, Eq. 18 becomes

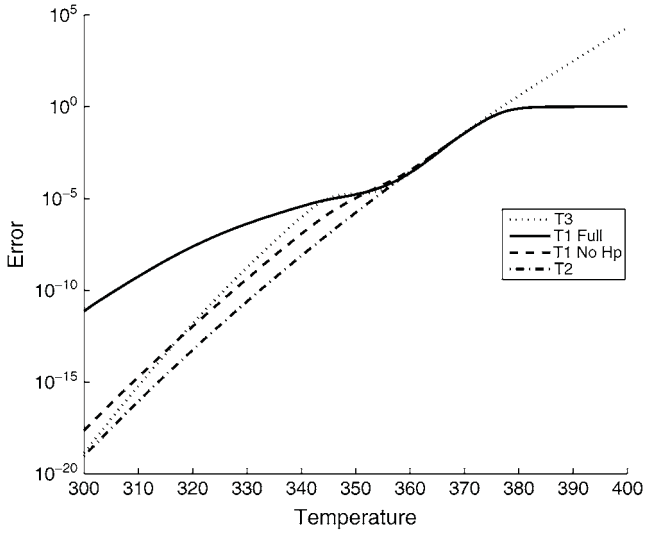

FIGURE 7 Comparison of estimates on hybridization error. Plots of the predicted error obtained from Eq. 17 of the text, under four different assumptions as a function of temperature from 300 to 400 K. (*Solid line*, full model; *dashed line*, full model minus hairpins; *dot-dashed line*, T2, Eq. 18; and *dotted line*, T3, Eq. 19.)

$$P_{\text{error}} = \frac{\sum_{\text{i}} \left(\dfrac{1 + \text{K}_{\text{p}_\text{i}}^{\text{hp}}}{\left(1 + \text{K}_{\text{p}_\text{i}}^{\text{hp}}\right)^2 + \text{C}_\text{t}\text{K}_{\text{p}_\text{i}\text{t}_\text{i}}} \sum_{\text{j}\neq\text{i}} \dfrac{\text{K}_{\text{p}_\text{i}\text{t}_\text{j}}}{1 + \text{K}_{\text{t}_\text{j}}^{\text{hp}}} \right)}{\sum_{\text{i}} \left(\dfrac{\text{K}_{\text{p}_\text{i}\text{t}_\text{j}}}{\left(1 + \text{K}_{\text{p}_\text{i}}^{\text{hp}}\right)^2 + \text{C}_\text{t}\text{K}_{\text{p}_\text{i}\text{t}_\text{j}}} \right)}, \qquad (19)$$

where C_t is the average strand concentration of each target species. A comparison of the different forms of the errors given in Eqs. 4a and 4b is shown in Fig. 7, where error is plotted versus temperature.

Results of a comprehensive treatment of error using Eq. 17 and concentration values obtained from the equilibrium model (described above) are shown by the solid line in the Fig. 7. The dashed line shows the error when concentrations are computed using the equilibrium model neglecting hairpins. The dot-dashed line is a plot of Eq. 18, which assumes the concentration of single-strand species is small relative to the concentration of hybridized strands. The dotted line corresponds to Eq. 19. Note, in the region from 360 to 375 K, all the error functions are essentially equivalent. In contrast, outside this range they differ, in some cases, considerably.

Simultaneous solutions of the equilibrium model

Siegmund and co-workers (75) reported solutions to the Equilibrium Model equations essentially as given in Eqs. 10 and 11. With direct applications to microarray sequence design and analysis, their ''hybridization calculator'' has been packaged in a suite of publicly available routines called CHIPCHECK. CHIPCHECK combines multidimensional and optimized Newton's methods (75) to solve the system of coupled nonlinear equations essentially as written in Eqs. 4a and 4b. Although our approach is practically identical to that of CHIPCHECK, there are several somewhat subtle differences worth noting. First, our approach is more generally complete in that the ''standard'' version of CHIPCHECK does not consider the possibility of either hairpin formation in single strands, or target:target duplexes (although these authors have advertised availability of a β-version that includes the potential for target:target duplexes). Second, in our approach both the probe and target concentrations are generally considered. In CHIPCHECK, target concentrations are considered to be in solution, but probe concentrations are considered strictly in terms of the number of probe molecules attached to a fixed volume (probe density) on the surface. Probe concentration is defined according to attachment of the number of probes per volume, V, on the surface. That is, the authors define total target concentration in the same way we do, i.e., as $C_{\text{t}_\text{i}}^{\circ}$, but probe concentration is defined as $C_{\text{p}_\text{i}}^{\circ} = n_\text{i}/V$. Using Eqs. 4a and 4b, and the equilibrium constants, $K_{\text{t}_\text{i}\text{p}_\text{j}}^{\text{eq}} = K_{\text{ij}}$, yields

$$C_{\text{t}_\text{i}} \left(1 + \frac{1}{V} \sum_{\text{j}} n_\text{j} \frac{K_{\text{ij}}}{1 + \sum_{\text{l}} K_{\text{jl}} C_{\text{t}_\text{l}}} \right) + C_{\text{t}_\text{i}}^{\circ} = 0,$$

which is in precise agreement with Eq. 7 of CHIPCHECK (75). After appropriate consideration of the probe concentrations, results of our approach can be compared directly to those obtained by CHIPCHECK. Since the two models are essentially the same it should be expected that, given the same input, they would produce identical results. Not surprisingly, as the comparisons described below indicate, there are only subtle differences. Although these differences are small, they can perhaps be partly explained by the inherent variance in robustness of the fitting algorithms that were employed.

Comparison 1

The first comparison concerned an 8×8 microarray with initial probe concentration $= 4 \times 10^{-15}$ mol/L, initial target concentration 2×10^{-7} mol/L, and temperature 298.15 K. Appropriate ΔG values for the sequences were supplied by CHIPCHECK. Two simulations were run. The first simulation ignored target:target duplexes, while the second included them. Duplex concentrations obtained from CHIPCHECK and from both our equilibrium and kinetic model calculations were normalized to unit length, and then compared by determining the norm of the difference σ, between results. For the first comparison (ignoring target:target duplexes), this norm was $\sigma = 6.0318908 \times 10^{-4}$ for the equilibrium model, and $\sigma = 6.0315873 \times 10^{-4}$ for the kinetic model. When target:target duplexes were included in the calculations, the error was $\sigma = 1.111409345 \times 10^{-2}$, and $\sigma = 1.111407571 \times 10^{-2}$ for the equilibrium and kinetic models, respectively. Evidently, there is very little

difference between results obtained by CHIPCHECK and our equilibrium and kinetic models for the 8×8 system.

Comparison 2

In the second comparison, duplex concentration was calculated for ''example 3'' on the CHIPCHECK website (75), in which 42 target-strands were hybridized to 100-probes spots. In this case, the normalized difference was $\sigma = 2.05869 \times 10^{-5}$ for the equilibrium model (ignoring target: target duplexes). For this larger system, differences between the CHIPCHECK results and our model calculations were even smaller than those observed for the 8×8 systems. It is not surprising that results of CHIPCHECK and our model calculations are in close agreement, as the same parameters were employed for both calculations. These comparisons show the validity of predictions of duplex equilibrium concentrations obtained from both our equilibrium and kinetic models, and show equivalence with concentrations predicted by the equilibrium model used in CHIPCHECK.

Statistical mechanical approaches to probe design

There have been several reports (11,46) on developments of multistate models for predicting hybridization and melting of duplex DNAs. As employed, their definition of ''multistate'' requires clarification and should be distinguished from the term ''multiplex'' as used herein. The aforementioned multistate models were developed with a particular focus on consideration of intramolecular structure (hairpin) formation in single strands and the relative influence of hairpin stability on duplex melting and hybridization. Essentially, in the multistate model, every pair of strands defines a system. For each system, the partition function is constructed considering multiple configurations that could be adopted by the two strands of the system. These include intramolecular hairpins that can form in the individual single strands, intermolecular dimers of the probe or target strands, and the bimolecular duplex. This is quite different from multiplex behavior as described here. In the multistate models described above, even though multiple configurational states for each pair of strands are considered, the collective behaviors of all possible reactions that can occur between different strands in a multiplex environment are not. In these approaches, hybridization of multiple strands is assumed to occur in a single-channel manner. Thus, the multistate approach, although perhaps more accurate for hybridization and melting of two-strand systems, does not provide for realistic design and analysis of multiplex hybridization reactions.

Kinetics of multiplex hybridization

A number of experimental and theoretical studies of the kinetics of short DNA oligomer hybridization have been reported (12–14,16–23). For the most part, these theoretical approaches were developed along the lines of the single-channel model. As has been reported, kinetic curves of duplex concentration versus time obtained from the single-channel model display typical exponential approaches to equilibrium values. Our results for the single-channel model calculations (described above) are entirely consistent with those that have been reported.

Despite the apparent need, relatively few analytical treatments of multiplex hybridization have been developed. Limited studies that have been performed considered inherent complexities of multiplex reactions, such as competition between different probes and targets, and effects of resource depletion on hybridization of perfect-match complexes (55,56). Two studies are particularly pertinent, whose results are in several ways comparable with ours using the multi-channel model (demonstrated in our 3×3 examples given above). A brief summary of these results and some comparisons follow.

In the first study by Graves and co-workers (55), the different reactions that can occur within a simple system, comprised of two probes and two targets, were described. It was assumed that the target concentration was relatively large compared to the probe concentration, and an expression was derived for the dependence of duplex concentrations on time. Kinetic equations for the general cases of one or two targets competing for one or two probe sites were approximated using the Mathematica T software package. Simulations were used to examine the influence of different values for the forward and reverse kinetic rate constants. The model did not consider the thermodynamic stability of duplexes, hairpin formation, or target:target duplex formation.

Probe concentration was defined in terms of the surface area and probe density, in a manner similar to the way the authors of CHIPCHECK (75) define probe concentrations. Diffusional barriers were not considered, i.e., all probes were assumed to be accessible for target interactions (as we have also assumed). For the case of a single target interacting with two probes (one a perfect match, the other a mismatch), results showed that target molecules do not partition between perfect and mismatch complexes in proportion to their dissociation constants. Once hybridized in a mismatch complex, it is virtually impossible for most of the targets to find their perfect matches. Furthermore, when it is possible for targets to interact with more than one probe on the surface, simple exponential approaches to equilibrium, as seen for a single duplex, are not observed. Relative abundance of two hybrid duplexes that can form simultaneously, changes dramatically with time; so much so, that mismatches can be temporarily present in relatively greater amounts than the perfect match. For this reason, the time over which a hybridization reaction is allowed to occur before washing, fixing, and analysis may be critical. The actual time duration of the hybridization kinetics is uncertain and must await calibration from experimental hybridization studies conducted as a function of time.

Results of this study indicate that if the mismatch and perfect match duplex equilibria are similar, they will exist in similar concentrations. This, combined with the situation when mismatch targets are in much higher concentration than perfect-match targets, makes cross-hybridization between mismatched probes and targets especially problematic. As we have shown, the relative sequence homology and stability of mismatch complexes can also affect the time-dependence.

The authors also provided analysis of the wash step, which we did not consider in our analysis. From values of the binding constants for duplexes and the value of the wash volume, approximate values for an optimum wash time were estimated. This points out the significance of the wash step.

In a subsequent study, Bishop et al. (56) reported analysis with a more sophisticated kinetic model that included diffusion of target strands and their binding reaction kinetics with probes on a surface. The analysis considered both mass transport and kinetic reactions. In this approach, the overall rate of hybridization was assumed to be determined by the integrated action of two independent processes: transport and reaction. Transport regards the delivery of targets to the surface as governed by Fick's law (neglecting convection). Once targets reach the reaction zone, hybridization is governed by reaction hybridization kinetics (the part of the process considered in our development). A major finding of the study showed that at high concentrations of target, the overall reaction was dominated by hybridization reaction kinetics (mass action transport was not a large factor). Not surprisingly, at low target concentrations, mass transport was the dominant component of the reaction.

In our development, we have only addressed the reaction kinetics portion of the process, and although the reported reaction kinetic portion of the model is similar to ours, theirs does not consider the potential for hairpin or target:target duplex formation. Furthermore, energetics of duplex complexes were assigned in an ad hoc manner. The multiplex scenario considered was for one probe interacting with two targets. Their analysis found a two-phase process. At early times, where amounts of bound targets are much lower than available probes, both matched and mismatched species bind independently. At later times, when the amount of bound complexes is comparable to the amount of free probes, perfect-match complexes gradually displaced mismatched species from the probes. This displacement was attributed to relatively higher stability of the perfect-match duplex, which causes a significantly lower value of its dissociation constant compared to the mismatch. Since competition is considered in the model, higher relative concentrations of the mismatches cause longer times to reach equilibrium. In early times, mismatches are the dominant hybridized species because they exist in higher concentration. In later stages, the rate of mismatch displacement by the perfect match is dominated by the rate of mismatch dissociation.

These studies of simple model systems have clearly demonstrated the importance of considering the effects of competitive hybridization in a multiplex mixture, and have laid important groundwork for our investigations. There are several new and distinguishing features in our approach: first, effects of sequence homology; second, explicit consideration of potential effects of mismatch stability; and third, the model development is general, in that any system of arbitrary size can be similarly treated (limited only by hardware capabilities). As summarized below, our results are consistent with those that have been previously reported, and build on the foundation established by previous studies.

CONCLUSIONS

We have developed a general analytical description of the equilibrium and reaction kinetics of DNA multiplex hybridization reactions. Generally applicable mathematical expressions are presented that thoroughly describe the equilibrium and kinetic behaviors of any multiplex system consisting of any numbers of probes and target strands. Numerical solutions to this system for both equilibrium and kinetic behaviors are provided. Although diffusional barriers (and resulting consequences), have not been implemented in the development, we have assumed (as have others) that targets are equally accessible for reactions with all probes. Testing of this assumption, proper parameterization, and demonstrated precision and accuracy improvements provided by our multiplex model approach must await completion of the acquisition and analysis of experimental multiplex hybridization data (M. T. Horne and A. S. Benight, unpublished).

In addition to being completely general in our approach we have considered explicit sequences of targets and probes involved in the reactions. Incorporation of sequence-dependent stability of all perfect-match and mismatch-duplex complexes is a major feature. Potentially significant influence of tandem-mismatch stability and resultant relative stabilization of cross-hybrid complexes was also explored. Again, the actual importance and quantitative sequence dependence of tandem-mismatch thermodynamics must await evaluation of a generally applicable database of sequence-dependent parameters for tandem mismatches.

In summary, results of our study reveal four interdependent factors that strongly influence multiplex hybridization behaviors. These are:

1. Relative concentrations of all probes and targets. As the numbers of probes and targets increase, hybridization kinetics increase but mismatch complexes can effectively compete with perfect matches and dominate perfect matches through consequences of resources depletion achieved through several means. For example, if the concentration of one or more targets is significantly higher, occurrence of mismatch states preferred by sequence homology, or increases in relative stability of tandem mismatches, acts to deplete the amount of perfect-match probe available for correct hybridization.

2. Relative thermodynamic stability of all perfect-match and mismatch complexes is critical to determining the type of diagnostic behavior being displayed.
3. Sensitivity of the calculation to temperature, particularly for the mismatches, suggests a diagnostic metric. At relatively early times, perfect matches increase as do mismatches. At intermediate times, the behavior of perfect matches and mismatches diverge. That is, perfect matches proceed along the critically damped coordinate, increasing at first, then exponentially approaching a constant equilibrium value; mismatches proceed along the overdamped coordinate where their concentrations initially increase, then decay exponentially to equilibrium values. The transition from one regime to another and how this relates to sequence homology and stability of mismatches is not yet totally clear. What is clear is that the predicted behavior, if experimentally verified, provides a new metric for analysis and diagnosis of sequence-dependent hybridization.
4. Amount of sequence homology shared by the probe and target strands in the multiplex mix, strongly influences both equilibrium and kinetic hybridization behavior.

We thank the reviewers for instructional comments. We also gratefully acknowledge Dr. Greg Brewood for careful reading of the manuscript and useful suggestions.

This work was supported in part by grant No. R44GM064299 from the Small Business Innovation Research program of the National Institutes of Health.

REFERENCES

1. Hall, T., P. Pancoska, P. Riccelli, K. Mandell, and A. Benight. 2001. Sequence context and thermodynamic stability of a single basepair mismatch in short deoxyoligonucleotide duplexes. *J. Am. Chem. Soc.* 123:11811–11812.
2. Hillen, W., T. Goodman, A. Benight, R. Wartell, and R. D. Wells. 1981. High resolution experimental and theoretical thermal denaturation studies on small overlapping restriction fragments containing the *Escherichia coli* lactose genetic control region. *J. Biol. Chem.* 256: 2762–2766.
3. Wartell, R. M., and A. S. Benight. 1985. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* 126: 67–107.
4. Breslauer, K., R. Frank, H. Blöcker, and L. Marky. 1986. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA.* 83:3746–3750.
5. Marky, L. A., and K. J. Breslauer. 1982. Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules. *Biopolymers.* 21:2185–2194.
6. Marky, L. A., and K. J. Breslauer. 1987. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers.* 26:1601–1620.
7. Breslauer, K. 1994. Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods Mol. Biol.* 26:347–372.
8. Owczarzy, R., P. Vallone, F. J. Gallo, T. M. Paner, M. Lane, and A. Benight. 1997. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers.* 44:217–239.
9. SantaLucia, J., Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA.* 95:1460–1465.
10. SantaLucia, J., Jr., and D. J. Hicks. 2004. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33:415–440.
11. Dimitrov, R., and M. Zuker. 2004. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.* 87:215–226.
12. Anshelevich, V., A. Vologodskii, A. Lukashin, and M. Frank-Kamenetskii. 1984. Slow relaxational processes in the melting of linear biopolymers: a theory and its application to nucleic acids. *Biopolymers.* 23:39–58.
13. Frank-Kamenetskii, M. 1997. Biophysics of DNA molecule. *Phys. Rep.* 288:13–60.
14. Ikuta, S., K. Takagi, R. Wallace, and K. Itakura. 1987. Dissociation kinetics of 19 basepaired oligo-DNA duplexes containing different single mismatched basepairs. *Nucleic Acids Res.* 15:797–811.
15. Dai, H., M. Meyer, S. Stepaniants, M. Ziman, and R. Stroughton. 2002. Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Res.* 30:e86.
16. Erickson, D., D. Li, and U. Krull. 2003. Modeling of DNA hybridization kinetics of spatially resolved biochips. *Anal. Biochem.* 317: 186–200.
17. Koval, V. V., O. V. Gnedenko, Y. D. Ivanov, O. S. Federova, A. I. Archakov, and D. G. Knorre. 1999. Real-time oligonucleotide hybridization kinetics monitored by resonant mirror technique. *IUBMB Life.* 48:317–320.
18. Nelson, J., and I. Tinoco, Jr. 1982. Comparison of the kinetics of ribooligonucleotide, deoxyribooligonucleotide and hybrid oligonucleotide double-strand formation by temperature jump kinetics. *Biochemistry.* 21:5289–5295.
19. Wang, J. Y., and K. Drlica. 2003. Modeling hybridization kinetics. *Math. Biosci.* 183:37–47.
20. Wang, S., A. E. Friedman, and E. T. Kool. 1995. Origins of high sequence selectivity: a stopped flow kinetics study of DNA/RNA hybridization by duplex- and triplex-forming oligos. *Biochemistry.* 34:9774–9784.
21. Wetmur, J., and N. Davidson. 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* 31:349–370.
22. Chan, V., D. Graves, and S. McKenzie. 1995. The biophysics of DNA hybridization with immobilized oligonucleotide probes. *Biophys. J.* 69:2243–2255.
23. Livshits, M., and A. Mirzabekov. 1996. Theoretical analysis of the kinetics of DNA hybridization with gel-immobilized oligos. *Biophys. J.* 71:2795–2801.
24. Duggan, D., M. Bittner, Y. Chen, P. Meltzer, and J. Trent. 1999. Expression profiling using cDNA microarrays. *Nature Genet. Suppl.* 21:10–14.
25. Eisen, M., P. Spellman, P. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA.* 95:14863–14868.
26. Fodor, S., R. Rava, X. Huang, A. Pease, C. Holmes, and C. Adams. 1993. Multiplexed biochemical assays with biological chips. *Nature.* 364:555–556.
27. Khan, J., L. Saal, M. Bittner, Y. Chen, J. Trent, and P. Meltzer. 1999. Expression profiling in cancer using cDNA microarrays. *Electrophoresis.* 20:223–229.
28. Lipshutz, R., D. Morris, M. Chee, E. Hubbell, M. Kozal, N. Shah, N. Shen, R. Yang, and S. Fodor. 1995. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques.* 19:442–447.
29. O'Donnell-Maloney, M., C. Smith, and C. Cantor. 1996. The development of microfabricated arrays for DNA sequencing and analysis. *Trends Biotechnol.* 14:401–407.
30. Szallasi, Z. 1999. Genetic network analysis in light of massively parallel biological data acquisition. *Pac. Symp. Biocomput.* 4:5–16.
31. Chee, M., R. Yang, E. Hubbell, A. Berno, X. Huang, D. Stern, J. Winkler, D. Lockhart, M. Morris, and S. Fodor. 1996. Accessing genetic information with high-density DNA arrays. *Science.* 274:610–614.
32. Stockton, J., J. S. Ellis, M. Saville, J. P. Clewley, and M. Zambon. 1998. Multiplex PCR for typing and subtyping influenza and respiratory syncytial viruses. *J. Clin. Microbiol.* 26:2990–2995.

33. Tettelin, H., D. Radune, S. Kasif, H. Khouri, and S. L. Salzberg. 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics*. 62:500–507.
34. Elnifro, E. M., A. M. Ashshi, R. J. Cooper, and P. E. Klapper. 2000. Multiplex PCR: optimization and application in diagnostic virology. *Clin. Microbiol. Rev*. 13:559–570.
35. Southern, E., K. Mir, and M. Shchepinov. 1999. Molecular interactions on microarrays. *Nat. Genet. Suppl*. 21:5–9.
36. Shalon, D., S. Smith, and P. Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*. 6:639–645.
37. Schena, M., D. Shalon, R. Davis, and P. Brown. 1995. Quantitative monitoring of gene expression patterns with a complimentary DNA microarray. *Science*. 270:467–470.
38. Shoemaker, D., D. Lashkari, D. Morris, M. Mittmann, and R. W. Davis. 1996. Quantitative phenotype analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet*. 14:450–456.
39. Pease, A., D. Solas, E. Sullivan, M. Cronin, C. Holmes, and S. Fodor. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA*. 91:5022–5026.
40. Ramsay, G. 1998. DNA chips: state-of-the-art. *Nat. Biotechnol*. 16:40–44.
41. Chen, J., R. Wu, P. Yang, J. Huang, Y. Sher, M. Han, W. Kao, P. Lee, T. Chiu, F. Chang, Y. Chu, C. Wu, and K. Peck. 1996. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with calorimetry detection. *Genomics*. 51:313–324.
42. Cheung, V., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, and G. Childs. 1999. Making and reading microarrays. *Nat. Genet*. 21:15–19.
43. Paner, T., M. Amaratunga, M. Doktycz, and A. Benight. 1990. Analysis of melting transitions of the DNA hairpins formed from the oligomer sequences *d*[GGATAC(*x*)4GTATCC] where $x = $ A, T, G and C. *Biopolymers*. 29:1715–1734.
44. Paner, T., P. Riccelli, R. Owczarzy, and A. Benight. 1996. Studies of DNA dumbbells. VI. Analysis of optical melting curves of dumbbells with a 16 basepair duplex stem and end-loops of variable size and sequence. *Biopolymers*. 39:779–793.
45. Vallone, P. M., T. M. Paner, J. Hilario, M. Lane, B. Faldasz, and A. Benight. 1999. Melting studies of short DNA hairpins: influence of loop sequence and adjoining basepair identity of hairpin thermodynamic stability. *Biopolymers*. 50:425–442.
46. SantaLucia, J., Jr., H. Allawi, and P. Seneviratne. 1996. Improved nearest neighbor parameters for predicting DNA duplex stability. *Biochemistry*. 35:3555–3562.
47. Ke, S., and R. M. Wartell. 1995. Influence of neighboring basepairs on the stability of single base bulges and basepairs in a DNA fragment. *Biochemistry*. 34:4593–4600.
48. LeBlanc, D. A., and K. M. Morden. 1991. Thermodynamic characterization of deoxyribooligonucleotide duplexes containing bulges. *Biochemistry*. 30:4042–4047.
49. Turner, D. H. 1992. Bulges in nucleic acids. *Curr. Opin. Struct. Biol*. 2:334–337.
50. Merkina, E. E., and K. R. Fox. 2005. Kinetic stability of intermolecular DNA quadruplexes. *Biophys. J*. 89:365–373.
51. Riccelli, P., T. S. Hall, P. Pancoska, K. E. Mandell, and A. S. Benight. 2003. DNA sequence context and multiplex hybridization reactions: melting studies of heteromorphic duplex DNA complexes. *J. Am. Chem. Soc*. 125:141–150.
52. Fixe, F., H. M. Branz, N. Louro, V. Chu, D. M. F. Prazeres, and J. P. Conde. 2005. Electric-field assisted immobilization and hybridization of DNA oligomers on thin-film microchips. *Nanotechnology*. 16:2061–2071.
53. Park, P. J., Y. A. Cao, S. Y. Lee, J.-W. Kim, M. S. Chang, R. Hart, and S. Choi. 2004. Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference. *J. Biotechnol*. 112:225–245.
54. Kooperberg, C., T. G. Fazzio, J. J. Delrow, and T. Tsukiyama. 2002. Improved background correction for spotted DNA microarrays. *J. Comput. Biol*. 9:55–66.
55. Zhang, Y., D. A. Hammer, and D. J. Graves. 2005. Competitive hybridization kinetics reveals unexpected behavior patterns. *Biophys. J*. 89:2950–2959.
56. Bishop, J., S. Blair, and A. Chagovetz. 2006. A competitive model of nucleic acid surface hybridization in the presence of point mutants. *Biophys. J*. 90:831–840.
57. Vijayendran, R. A., F. S. Ligler, and D. E. Leckband. 1999. A computational reaction-diffusion model for the analysis of transport-limited kinetics. *Anal. Chem*. 71:5405–5412.
58. MatLab. V. 7.1. 2005. The MathWorks, Natick, MA.
59. OCTAVE. 2005. GPL open source software. http://www.octave.org: http://www.octave.org.
60. Bommarito, S., N. Peyret, J. SantaLucia Jr. 2000. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res*. 28:1929–1934.
61. Allawi, H., J. Sant, and J. SantaLucia Jr. 1997. Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry*. 36:10581–10594.
62. Allawi, H., J. Sant, and J. SantaLucia Jr. 1998. Nearest neighbor thermodynamics of internal A·C mismatches in DNA: sequence dependence and pH effects. *Biochemistry*. 37:9435–9444.
63. Allawi, H., J. Sant, and J. SantaLucia Jr. 1998. Nearest neighbor thermodynamics of internal G·A mismatches in DNA. *Biochemistry*. 37:2170–2179.
64. Allawi, H., J. Sant, and J. SantaLucia Jr. 1998. Thermodynamics of internal C·T mismatches in DNA. *Nucleic Acids Res*. 37: 2694–2701.
65. Le Novere, N. 2001. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics*. 17:1226–1227.
66. Nicolas, P., and J. SantaLucia Jr. 1999. HYTHER, V. 1.0. Wayne State University, Detroit, MI.
67. Rose, J., R. Deaton, D. Francescetti, M. Garzon, and S. E. J. Stevens. 1999. A statistical mechanical treatment of error in the annealing biostep of DNA computation. *In* Proceedings of the Genetic Evolutionary Computing Conference. Morgan–Kauffman, San Francisco, CA. 1829–1834.
68. Rose, J., and R. Deaton. 2001. Simulations of statistical mechanical estimates of hybridization error. *In* Preliminary Proceedings of the 6th International Meeting of DNA-Based Computers. Leiden Center for Natural Computing, Leiden, The Netherlands.
69. Rose, J., R. Deaton, M. Hagiya, and A. Suyama. 2002. The fidelity of the TAG-antiTAG system. *In* DNA Computing, 7th International Workshop on DNA-Based Computers (DNA7). Springer LNCS, Springer Verlag, NY. 2340:138–149.
70. Deaton, R., H. B. Chen, and J. Rose. 2003. A software tool for generating non-crosshybridizing libraries of DNA oligonucleotides. *In* DNA Computing, 8th International Workshop on DNA-Based Computers (DNA8). Springer LNCS, Springer Verlag, NY. 2568: 252–261.
71. Rose, J. A., M. Hagiya, and A. Suyama. 2003. The fidelity of the tag-antitag system 2: reconciliation with the stringency picture. *In* Proceedings of the Congress on Evolutionary Computing. Dec. 2003. Canberra, Australia. 2740–2747.
72. Rose, J. A., R. J. Deaton, and A. Suyama. 2004. Statistical thermodynamic analysis and design of DNA-based computers. *Natur. Comput*. 3:443–459.
73. Nakao, M., Y. K. Okuji, and M. Kaneshisa. 2000. Quantitative estimation of cross-hybridization in DNA microarrays based on a liner model. *Genome Informatics*. 11:231–232.
74. Ben-Dor, A., R. Karp, B. Schwikowski, and Z. Yakhini. 2000. Universal DNA tag systems: a combinatorial design scheme. *J. Comput. Biol*. 7:503–519.
75. Siegmund, K. H., U. Steiner, and C. Richert. 2003. CHIPCHECK: a program predicting total hybridization equilibria for DNA binding to small oligonucleotide microarrays. *J. Chem. Inf. Comput. Sci*. 43:2153–2162.