Breakdown of thermodynamic equilibrium for DNA hybridization in microarrays

J. Hooyberghs,[1,2,3] M. Baiesi,[2] A. Ferrantini,[2] and E. Carlon[2]
[1]*Flemish Institute for Technological Research (VITO), Boeretang 200, B-2400 Mol, Belgium*
[2]*Institute for Theoretical Physics, KULeuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium*
[3]*Hasselt University, Campus Diepenbeek, B-3590 Diepenbeek, Belgium*
(Received 24 August 2009; published 13 January 2010)

Test experiments of hybridization in DNA microarrays show systematic deviations from the equilibrium isotherms. We argue that these deviations are due to the presence of a partially hybridized long-lived state, which we include in a kinetic model. Experiments confirm the model predictions for the intensity vs free-energy behavior. The existence of slow relaxation phenomena has important consequences for the specificity of microarrays as devices for the detection of a target sequence from a complex mixture of nucleic acids.

DOI: 10.1103/PhysRevE.81.012901 PACS number(s): 87.15.H−, 82.39.Pj

DNA hybridization (the binding of two strands to form a double helix) in bulk solution has been extensively studied in the past [1]. In this Brief Report we discuss hybridization in DNA microarrays. Microarrays are high throughput devices which have been widely used to measure the activity of genes at a genome wide level. In a microarray single stranded DNAs are arrayed on a solid substrate in spots, each containing a specific sequence. Hybridization takes place between surface-bound sequences (referred to as *probes*) and sequences in solution (*targets*) carrying a fluorophore. The amount of hybridized target is obtained from the emitted fluorescence from a given spot. Although hybridization in microarrays has attracted some interest in recent years its physical properties are still poorly understood (for reviews on the topic see, e.g., [2]). We demonstrate here that, contrary to a widespread belief, in DNA microarrays equilibration times may largely exceed typical experimental times. These claims are based on experimental results and are corroborated by the analysis of a kinetic model.

The experimental setup is shown in Table I and extends that of Ref. [3]. A single sequence is present in solution: either a 30 or a 25 mer. The surface probe sequences are perfect matching, with one or two mismatches. The mismatches can be of different nature and they are at different positions along the sequence [4]. In total there are 1006 different probe sequences [3]. Custom arrays containing spots with the probe sequences of Table I were purchased from Agilent Technologies. We used 15 K slides which accommodate 15 replicas of the 1006 sequences. The analysis is performed on the median intensities over the replicas. The standard Agilent protocol (except for target fragmentation) and Agilent buffers [5] were used. The temperature is 65 °C.

TABLE I. Target and probe sequences used in the experiments. All sequences have a 5′ to 3′ orientation. Target sequences have a 20 mer poly(A) stretch attached to their 3′ end, which terminates with a Cy3 fluorophore. Probes have a 30 mer poly(A) stretch at their 3′ end, which is covalently linked to the microarray surface.

	Target sequences in solution
1.	CTTTGTCGAGCTGGTATTTGGAGAACACGT
2.	TCGAGCTGGTATTTGGAGAACACGT
	Probes at the microarray surface
PM	ACGTGTTCTCCAAATACCAGCTCGACAAAG
	ACGTGATCTCCAAATACCAGCTCGACAAAG
1MM	ACGTGCTCTCCAAATACCAGCTCGACAAAG
	ACGTGGTCTCCAAATACCAGCTCGACAAAG
	…
2MM	ACGTGATCTCCCAATACCAGCTCGACAAAG
	…

Table I shows a plot of I/c (the intensity divided by the target concentration) vs $\Delta\Delta G$ for four experiments at different concentrations using the setup of Table I (for the 30-mer target). The variable $\Delta\Delta G \equiv \Delta G_{\mathrm{PM}} - \Delta G$ is the difference in hybridization free energies between a given sequence and the perfect match (PM) sequence. It is calculated from the nearest-neighbor parameters obtained from the analysis of microarray data, as discussed in Ref. [3] (the nearest-neighbor model assumes that ΔG can be written as a sum of dinucleotide terms [1]). Figure 1 shows that the intensity is proportional to the concentration for four orders of magnitude in I. In the low c limit, equilibrium thermodynamics predicts that

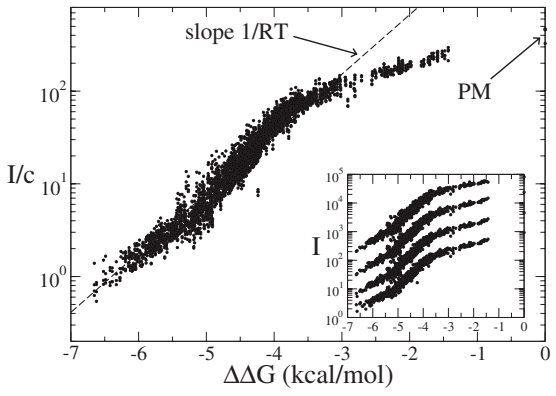

FIG. 1. Plot of I/c as a function of $\Delta\Delta G$ for four different experiments at concentrations $c=2$, 10, 50 and 250 pM. The hybridization time is of 17 h. The "collapse" of the four data sets into a single curve demonstrates that $I \propto c$ in the whole intensity range. Deviations from the equilibrium isotherm $I \propto e^{-\Delta G/RT}$ are observed at high intensities. Inset: plot of each individual concentration.

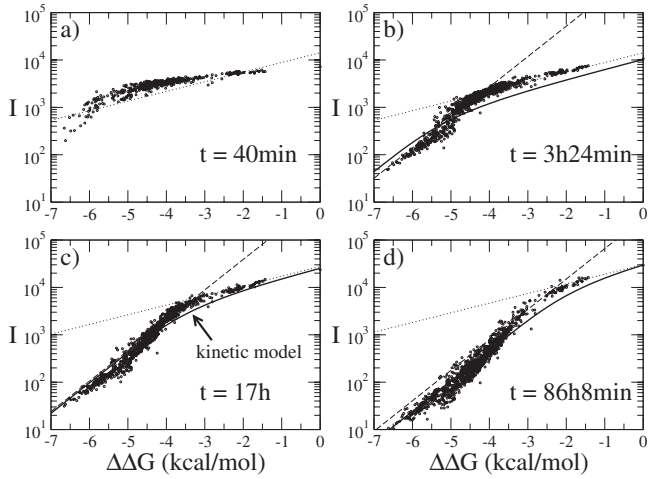

FIG. 2. Plot of intensity vs $\Delta\Delta G$ four experiments at different times with a 30-mer target and $c=50$ pM. Dashed lines have slope $1/RT$, dotted lines have slope γ/RT with $\gamma=0.32$. Solid lines are obtained from the solution of Eqs. (2) and (3) using as parameters $k_1=10^5$ M^{-1} s^{-1}, $k_2=1$ s^{-1}, $\gamma=0.32$, and $\Delta G_{PM}=-14.5$ kcal/mol.

$$I = Ace^{-\Delta G/RT} \tag{1}$$

where A sets the intensity scale, R is the gas constant and T the temperature. Equation (1) is obtained from the $c\to 0$ limit of the Langmuir isotherm (which was used in microarray data analysis [6–10]) but also from isotherms in which electrostatic effects are taken into account [11] (electrostatic effects were experimentally observed at low ionic strengths [12]). In the latter ΔG contains a contribution from electrostatic interactions. For both isotherms, in the $I\propto c$ regime one expects a linear dependence of log I on ΔG (or $\Delta\Delta G$). Figure 1 shows that the experimental data are only in partial agreement with Eq. (1), which is drawn as a dashed line in the figure.

We then extended the analysis at different hybridization times. Figure 2 shows a plot of I vs $\Delta\Delta G$ for a 30-mer target at four different times and for a concentration of 50 pM (the 17 h hybridization data are those already shown in Fig. 1). Once the desired hybridization time has been reached the experiment is stopped, the microarray washed and scanned to measure the emitted fluorescence from every spot. Experiments at different hybridization times thus require different slides. As the hybridization time increases, a larger fraction of the data aligns along a line with a slope $1/RT$, which shows that the observed deviations from Eq. (1) are due to the breakdown of thermodynamic equilibrium. Surprisingly, full equilibrium has not been reached here even after 86 h. In Figs. 2(b)–2(d) the data for log I align along two slopes: in the equilibrium regime the slope is $1/RT$ [Eq. (1)], in the nonequilibrium regime the slope is smaller and appears to be constant in the course of time [13]. Hybridization data for the shorter target sequence (25 mer) are shown in Fig. 3. The agreement with Eq. (1) is over three orders of magnitude in the intensity scale at times >3 h. Hence equilibration is much faster for the shorter sequences. The experimental setup allows a detection of nonequilibrium effects as deviation from the $1/RT$ line without the need, in principle, of a time series analysis.

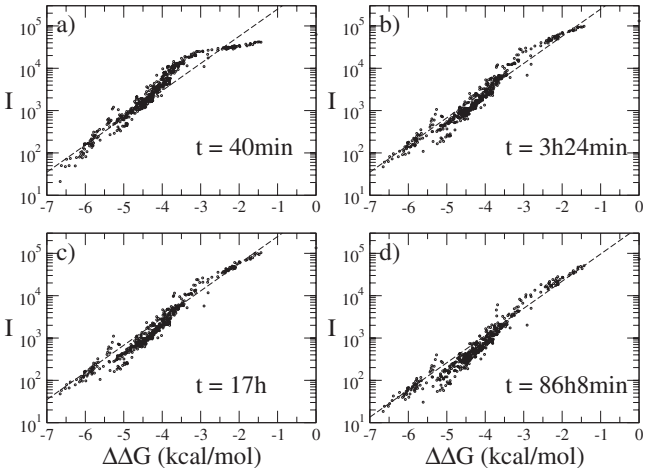

FIG. 3. As in Fig. 2 for a target sequence of length 25. The target concentration is $c=500$ pM.

Hybridization of oligonucleotides in solution is usually described as a two-state process. However, as will be shown, a two-state process cannot be reconciled with the experimental data shown in Fig. 2. During manufacturing the probes are tethered to the surface and can form a dense layer that slows down hybridization. Some of these effects have been discussed in Refs. [14,15]. The typical distance between probes is 10 nm, and the length of a fully stretched 30-mer duplex is 10 nm and its thickness of 2 nm. Probe sequences in the experiment have also a poly(A) 30-mer spacer (see Table I). Therefore a single target molecule can interact with more than one probe. Taking this into account, we have extended the two-state hybridization model with an additional intermediate state (Fig. 4). Indicating with θ_1 and θ_2, the fraction of partially and fully hybridized probes on a microarray spot, the kinetics of these reactions is given by

$$\frac{d\theta_1}{dt} = ck_1(1-\theta_1-\theta_2) + k_{-2}\theta_2 - (k_{-1}+k_2)\theta_1, \tag{2}$$

$$\frac{d\theta_2}{dt} = k_2\theta_1 - k_{-2}\theta_2, \tag{3}$$

where c is the target concentration in solution and k_1, k_{-1}, k_2, and k_{-2} the four rates involved (see Fig. 4). For simplicity we have assumed that at most a single target molecule can bind to a given probe. θ_1 is the average occupation fraction over several configurations in which partial binding can occur at different positions of the probe sequence.

The rate constants, using a two-state model description, have been measured in several microarray experiments. In Ref. [16] the hybridization of a common target sequence to a perfect match probe and to a probe containing one mismatch were considered. The following rates were measured (at 45 °C): $k_1^{(PM)}=19\cdot 10^4$ M^{-1} s^{-1}, $k_1^{(MM)}=21\cdot 10^4$ M^{-1} s^{-1}, $k_{-1}^{(PM)}=12\cdot 10^{-4}$ s^{-1}, and $k_{-1}^{(MM)}=29\cdot 10^{-4}$ s^{-1}. While there is more than a factor two of difference in the detachment rates, the attachment rates differ only by 10%. These results are in

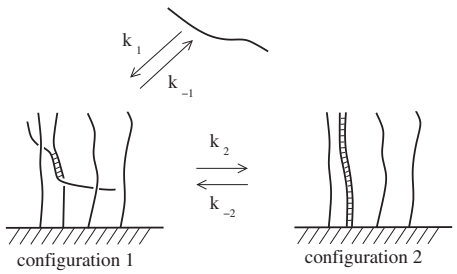

FIG. 4. The three-state model for hybridization in DNA microarrays is specified by the four rate constants.

agreement with observation for kinetic behavior in bulk solution [18]. The probes in our experiment differ by at most two nucleotides out of 30. We will then take k_1 as sequence independent. Consider now the reaction rate k_2. In the partially hybridized state the target strand binds over a stretch of nucleotides with one probe sequence (primary contact) but it can also bind to a second neighboring probe (secondary contact). The rate limiting step is the unbinding from secondary contacts and the strand contraction so that the target probe can overcome steric hindrance and wind up into a fully formed helix all along its length. k_2 will depend on target length and probe length and density. We will assume k_2 to be the same for the probes of Table I. The reverse rates are then fixed by the thermodynamics relations,

$$k_{-1} = k_1 e^{\Delta G'/RT}, \quad k_{-2} = k_2 e^{(\Delta G - \Delta G')/RT} \tag{4}$$

where $\Delta G'$ and ΔG are the free-energy differences between configurations 1 and 2, and the unbound state, respectively. Next we link $\Delta G'$ to ΔG. Weak total binding (small $|\Delta G|$) caused by the presence of multiple mismatches should also correspond to weak partial binding (small $|\Delta G'|$). As a simple approximation we will assume that the two free energies are monotonically linked as

$$\Delta G' \approx \gamma \Delta G \quad (\gamma < 1). \tag{5}$$

The model is thus characterized by k_1, k_2 and γ.

To gain some more insight we consider the limit of fast equilibration for Eq. (2). First we obtain the equilibrium value for θ_1 by setting the right-hand sides of Eqs. (2) and (3) to zero in the limit $ce^{-\Delta G/RT} \ll 1$. We then solve Eq. (3) replacing for θ_1 its equilibrium value $\theta_1^{(\mathrm{eq})} = ce^{-\Delta G'/RT}$. Setting the initial condition $\theta_2(0)=0$ we get

$$\theta_2(t) = ce^{-\Delta G/RT}(1 - e^{-t/\tau}), \tag{6}$$

$$\tau^{-1} = k_{-2} = k_2 e^{(\Delta G - \Delta G')/RT} = k_2 e^{(1-\gamma)\Delta G/RT}. \tag{7}$$

The relaxation time, τ, depends on ΔG: weakly bounded sequences (small $|\Delta G|$) equilibrate faster than strongly bounded ones (large $|\Delta G|$). For fast equilibrating sequences ($\tau \ll t$) one recovers Eq. (1) from Eq. (6); for sequences with long equilibration times $\tau \gg t$ we expand Eq. (6) to lowest order in t/τ. With this approximation we find that for a given time t,

$$\theta_2(t) = \begin{cases} ce^{-\Delta G/RT} & |\Delta G| \ll |\Delta G^*| \\ ctk_2 e^{-\gamma \Delta G/RT} & |\Delta G| \gg |\Delta G^*| \end{cases} \tag{8}$$

where ΔG^* is a crossover free energy that depends on time and is obtained by setting $\tau = t$ in Eq. (7). After hybridization the slides undergo washing steps according to the standard Agilent protocol, which are expected to remove weakly bound target molecules from the slide and have been also included in thermodynamics models of arrays [19]. In the present setup there is only one target sequence in solution and washing is likely to affect the partial hybridized state. In this case one can assume that the measured intensity is given by $I \approx A\theta_2$ (in the model the typical free energies of the partially hybridized states are such that $\theta_1 \ll \theta_2$). Equation (8) reproduces the two slopes in the log I vs $\Delta\Delta G$ plots as seen in the experiments (Fig. 1). It shows that the nonequilibrium regime is characterized by a slope equal to γ/RT.

The solid lines in Figs. 2(b)–2(d) are plots of the intensity $I = A\theta_2$ obtained from the solution of Eqs. (2) and (3). The parameters used are given in the caption of Fig. 2. For the choice of parameters given, the fast equilibration limit [Eq. (6)] approximates very well the full solution of Eqs. (2) and (3). There is a reasonable agreement between the experiments and the kinetic model. We note though that the crossover between the two regimes is somewhat sharper in experiments. In addition, the experimental I vs $\Delta\Delta G$ data show a slight sigmoidal trend which is not present in the kinetic model. Within the two-state model kinetics, and using the assumption k_1 as sequence independent, one arrives to a solution similar to Eq. (8) although with $\gamma = 0$. Therefore the two-state model cannot account for a second finite slope as observed in the experiments. Note that the fit of the kinetic model to the data requires also an estimate of ΔG_{PM}, as the method of Ref. [3] does not provide absolute ΔG for a sequence but only $\Delta\Delta G$, i.e., differences in free energies with respect to a perfect match hybridization. In addition, in the fit we adjusted the constant $A(I = A\theta_2)$ as we note in the experimental data a global decrease of the intensity scale. This is probably due to some degradation of the fluorophores or of the target and probe strands (as depurination, the loss of purines, which alters the binding energies of the involved strands). The overall decrease in intensity occurs both for the 25 mer and the 30 mer sequences. In the latter the effect is somewhat stronger, especially at longer hybridization times.

Equation (7) predicts that the relaxation time is a function of ΔG, $\Delta G'$, and k_2. We can use this equation to compare the ratio between the times for a $L=30$ and $L=25$ targets. Consider the same probe sequence hybridizing to the two targets. Assuming $\Delta G'(L=30) \approx \Delta G'(L=25)$ and using as estimate of the difference in binding energies between $\Delta G(L=30) - \Delta G(L=25) \approx -2.5$ kcal/mol [17], we get from Eq. (7) a decrease in a factor $\exp(2.5/RT) \approx 40$ in the relaxation time. In addition one also expects $k_2(L=25) > k_2(L=30)$ which decreases the relaxation time even further. The similarities between Figs. 2(c) and 3(a) suggest that the relaxation-time ratio between 25 and 30 mers is approximately 25, which is the same order of magnitude just obtained from Eq. (7). Since typical biological experiments involve target strands of lengths 30–50, the breakdown of equilibrium shown here

may occur in many different microarrays platforms and in biological experiments, and could involve even longer relaxation times.

Summarizing: we showed that hybridization in DNA microarrays under standard conditions is characterized by relaxation times which may largely exceed the experimental time. In the nonequilibrium regime the intensities are distributed as $e^{-\gamma\Delta G/RT}$, with $\gamma<1$. This is equivalent to introducing an effective temperature $T_{\mathrm{eff}}=T/\gamma>T$. Interestingly effective temperatures were used as adjustable phenomenological parameters to fit biological microarray data [6,9]. This work provides an insight on the origin of these.

The breakdown of equilibrium implies lower specificity of the microarrays as devices for the detection of a desired sequence from a complex mixture. To see this consider a probe at the microarray surface and two sequences at equal concentration in solution: one perfect matching with the probe and one with a mismatch. In the equilibrium regime the two sequences hybridize to the probe with a probability ratio $e^{(\Delta G_{\mathrm{PM}}-\Delta G_{\mathrm{MM}})/RT}\approx 0.05$, where we have used a typical value $\Delta G_{\mathrm{MM}}-\Delta G_{\mathrm{PM}}\approx 2$ kcal/mol [3] and a temperature of $T=65$ °C. In the nonequilibrium regime, due to the presence of a factor $\gamma<1$ in the exponential the ratio is about 0.4 (taking $\gamma=0.32$). Therefore in the nonequilibrium regime a significant fraction of a measured signal may be due to hybridization to noncomplementary targets, a phenomenon known as cross hybridization. For an optimal functioning of the microarrays it is then desirable to work under equilibrium conditions [8]. Several parameters may influence the relaxation time as temperature, salt and buffer conditions. The experimental setup discussed in this Brief Report provides a good test of equilibrium (single line vs broken line in a I vs $\Delta\Delta G$ plot) and can be used to investigate the best working conditions for hybridization.

We thank Karen Hollanders (VITO) for help with the experiments. We acknowledge financial support from Research Foundation-Flanders (FWO) Grant No. G.0311.08 and from KULeuven Grant No. OT/07/034A.

[1] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids Structures, Properties and Functions* (University Science Books, Mill Valley, 2000).
[2] D. J. Graves, Trends Biotechnol. **17**, 127 (1999); R. Levicky and A. Horgan, *ibid.* **23**, 143 (2005); A. Halperin, A. Buhot, and E. B. Zhulina, J. Phys.: Condens. Matter **18**, S463 (2006); H. Binder, *ibid.* **18**, S491 (2006).
[3] J. Hooyberghs, P. Van Hummelen, and E. Carlon, Nucleic Acids Res. **37**, e53 (2009).
[4] Up to five nucleotides to the edges to avoid terminal mismatches and such that the minimal distance between mismatches in the same strand is of five nucleotides.
[5] The buffer is described in US Patent 6753145.
[6] G. A. Held, G. Grinstein, and Y. Tu, Proc. Natl. Acad. Sci. U.S.A. **100**, 7575 (2003).
[7] F. Naef and M. O. Magnasco, Phys. Rev. E **68**, 011906 (2003).
[8] G. Bhanot *et al.*, Biophys. J. **84**, 124 (2003).
[9] E. Carlon and T. Heim, Physica A **362**, 433 (2006).
[10] T. Naiser, J. Kayser, T. Mai, W. Michel, and A. Ott, Phys. Rev. Lett. **102**, 218301 (2009).
[11] A. Vainrub and B. M. Pettitt, Phys. Rev. E **66**, 041905 (2002); A. Halperin, A. Buhot, and E. B. Zhulina, Clin. Chem. **50**, 2254 (2004).
[12] P. Gong and R. Levicky, Proc. Natl. Acad. Sci. U.S.A. **105**, 5301 (2008).
[13] We note that the slope in the nonequilibrium regime of Fig. 1(a) is somewhat smaller than that of Figs. 1(b)–1(d). This is probably due to the protocol followed: at time $t=0$ both the solution containing the target molecules and the array are at room temperature. They are then placed into an oven for the duration of the experiment at a temperature $T=65$ °C. Short time behavior could then be influenced by the initial room-temperature hybridization.
[14] A. W. Peterson, R. J. Heaton, and R. M. Georgiadis, Nucleic Acids Res. **29**, 5163 (2001).
[15] M. F. Hagan and A. K. Chakraborty, J. Chem. Phys. **120**, 4958 (2004).
[16] M. Glazer *et al.*, Anal. Biochem. **358**, 225 (2006).
[17] We estimated $\Delta G_{\mathrm{PM}}(L=30)=-14.5$ kcal/mol from the fits of Fig. 2, using extensivity we expect $\Delta G_{\mathrm{PM}}(L=25)\approx 0.83\Delta G_{\mathrm{PM}}(L=30)=-12$ kcal/mol.
[18] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules* (W. H. Freeman and Company, New York, 1980).
[19] G. A. Held, G. Grinstein, and Y. Tu, Nucleic Acids Res. **34**, e70 (2006); C. J. Burden, Y. Pittelkow, and S. R. Wilson, J. Phys.: Condens. Matter **18**, 5545 (2006).