THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

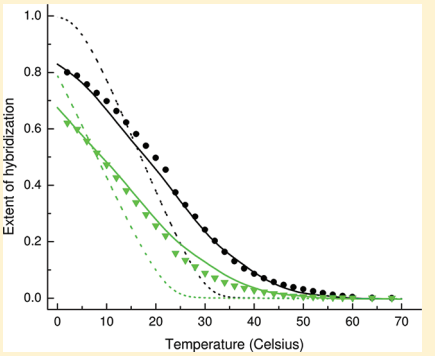
Accurate Prediction of Binding Thermodynamics for DNA on Surfaces

Arnold Vainrub*,† and B. Montgomery Pettitt*,‡

†Center for Biomedical Engineering, University of Texas Medical Branch, Galveston, Texas 77550-1156, United States

‡Department of Chemistry, University of Houston, Houston, Texas 77204-5003, United States

ABSTRACT: For DNA mounted on surfaces for microarrays, microbeads, and nanoparticles, the nature of the random attachment of oligonucleotide probes to an amorphous surface gives rise to a locally inhomogeneous probe density. These fluctuations of the probe surface density are inherent to all common surface or bead platforms, regardless of whether they exploit either an attachment of presynthesized probes or probes synthesized in situ on the surface. Here, we demonstrate for the first time the crucial role of the probe surface density fluctuations in the performance of DNA arrays. We account for the density fluctuations with a disordered two-dimensional surface model and derive the corresponding array hybridization isotherm that includes a counterion screened electrostatic repulsion between the assayed DNA and probe array. The calculated melting curves are in excellent agreement with published experimental results for arrays with both presynthesized and in situ synthesized oligonucleotide probes. The approach developed allows one to accurately predict the melting curves of DNA arrays using only the known sequence-dependent hybridization enthalpy and entropy in solution and the experimental macroscopic surface density of probes. This opens the way to high-precision theoretical design and optimization of probes and primers in widely used DNA array-based high-throughput technologies for gene expression, genotyping, next-generation sequencing, and surface polymerase extension.



INTRODUCTION

A thermodynamic theory of hybridization for surface-tethered oligonucleotide probes is of importance for the design and optimization of microarray, microbead, and all surface-mounted DNA technologies. These technologies using surface-attached DNA probes enable high-throughput DNA analysis at the genome-wide level and find their widest applications in genomics, including sequencing[1] and screening for individual genome variations.[2]

In this contribution, we propose a novel theoretical framework for the hybridization thermodynamics of DNA arrays. Although for solution hybridization assays, calculations of the melting temperature and melting curve are routine tools in drafting of almost any experimental protocol (for example, design of PCR primers), the same approach has not been as successful for arrayed DNA probes because of the complexity of hybridization at a surface.[3−7]

In solution, the thermodynamics of DNA hybridization is quantitatively predicted using the nearest-neighbors (NN) model.[8,9] This model is implemented in numerous software packages (e.g., MELTING[10]) that allow calculation based on the DNA sequence of the melting temperature and melting curve for a complementary (or few mismatch) probe−target DNA pair as a function of their concentration and the hybridization buffer (content of mono- and bivalent metal cations and denaturing agent). The NN model has become an indispensable tool for the design of various DNA assays that include polymerase chain reaction (PCR) primers, Förster resonance energy transfer (FRET) probes, and molecular beacons.

In contrast to hybridization in solution, for DNA probes arrayed on a surface, our understanding of hybridization is not satisfactory.[5−7,11−14] Briefly, although some key factors affecting the hybridization near a surface have been identified, such as the surface density of probes,[11,13,15] effect of spacers,[12] and surface material and charge,[16] the previously proposed hybridization (non-Langmuir) isotherms[16−20] do not succeed in quantitatively predicting DNA melting curves accurately in terms of the thermodynamic stability in solution (NN model thermodynamics) and surface conditions.[3,4,21] Effects of probe length dispersion due to synthesis problems can affect isotherms as well.[3]

Here, we identify the local fluctuations of probe density as a key factor in DNA hybridization thermodynamics. We note that the structure of the amorphous (glassy) substrate surface dictates the seemingly random nature of attachment sites available. The distribution and occupancy of those sites consequently determine the surface fields and the observed effects for both metallic and dielectric materials. Taking into account local inhomogeneities at the time of synthesis frozen into the surface dictated by its preparation, each probe on the surface may be surrounded by a different pattern and number of surrounding neighbor probes. The surface densities of occupied site fluctuations arise as a result of random attachment of oligonucleotide probes to the reactive groups on the surface in the process of array

Received: August 23, 2011
Revised: September 30, 2011
Published: October 05, 2011

synthesis/fabrication. For example, on commonly used glass surfaces, a typical mean distance between SiOH reactive groups is 0.4 nm. Oligonucleotide probes covalently attach by the functionalized end to only a small fraction of the randomly distributed reactive SiOH groups forming an array with average surface densities of $10^{12}-10^{13}$ cm^{-2} corresponding to mean interprobe distances from 10 to 3 nm, respectively. Therefore, one expects local fluctuations at the time of synthesis to be frozen into the random (nonequilibrium) pattern dictated by the surface preparation such that each probe on the surface may have a different pattern and number of surrounding neighbor probes. Thus, the local surface density of probes strongly fluctuates on the nanometer size scale. We consider the effect this has on hybridization.

We note that others have considered inhomogeneities in available probe positions in kinetic studies.[22] Those studies did not consider an a priori inhomogeneous distribution with explicit statistics, but rather one with a number of depleted sites for hybridization (sites already occupied) and thus unavailable for further kinetic trapping. We also do not consider the parts of sequence space where probe cross-hybridization in very dense arrays may interfere.[23] Here, we consider the equilibrium thermodynamics associated with a nonuniform distribution of non cross-hybridizing probes whose attachment sites were frozen when the glassy or noncrystalline substrate was formed. We wish to explore the question of how the surface structure for attachments affects equilibrium hybridization isotherms.

THEORY

To account for the fluctuations of probe surface density, we start with a DNA array hybridization isotherm derived previously.[19,20] This isotherm describes the electrostatic repulsion between DNA targets and a geometrically regular probe array arising due to the large negative charge of the DNA phosphate backbone in terms of a uniform probe surface density, i.e., in a mean-field approximation. It assumes that the repulsion energy is

$$V_r = wN_PZ_T(Z_P + \theta Z_T) \quad (1)$$

Here N_P is the surface density of probes, Z_P and Z_T are the lengths of probe and target expressed in the number of bases, θ is the extent of hybridization ($0 \le \theta \le 1$), and w is the electrostatic interaction parameter that depends on hybridization solution salinity. This leads to the hybridization isotherm[19,20]

$$C_0 = \frac{\theta}{1-\theta}\exp\left(\frac{\Delta H_0 - T\Delta S_0}{RT}\right)\exp\left(\frac{wN_PZ_T(Z_P + \theta Z_T)}{RT}\right) \quad (2)$$

Here C_0 is the concentration of assayed DNA targets, and ΔH_0 and ΔS_0 are the enthalpy and entropy of double helix formation in solution, which can be taken from the NN model.

In the isotherm eq 2, the repulsion energy of eq 1 scales with the hybridizing target charge Z_T and the average surface charge density $N_P(Z_P + \theta Z_T)$, which is the sum of the probes charge N_PZ_P and hybridized targets charge $N_P\theta Z_T$. This means that all the probes are assumed to be equivalent. Given the expected surface coverage fluctuations, this is not reasonable since the short-range target repulsion depends on the local surrounding, i.e., the number, location, and hybridization state of neighboring arrayed probes around a given probe hybridizing to DNA targets, and not on an average probe density. Indeed, the number n of probes that effectively repel a target is small due to a short Debye screening length of 1 to 0.33 nm at typically used hybridization buffer ionic strengths from 0.1 to 1 M, respectively. Therefore, relative fluctuations of n (known to be on the order of $1/\sqrt{n}$) are large, and thus fluctuations of the *local* probe density are important.

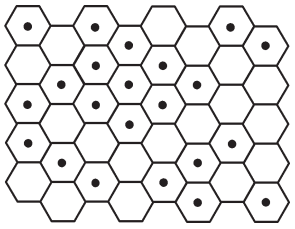


Figure 1. Mapping the probe array onto a hexagonal lattice. Dark circles depict random lattice cells containing a probe.

We model the random fluctuations of local probe surface density by mapping the surface onto a hexagonal two-dimensional (2D) lattice and assume a random occupation of sites (and consequently a random number of near neighbors) as shown in Figure 1.

The size of the lattice cell is chosen to allow any site to contain only one b-DNA double helix of radius r = 1 nm, and thus the surface area of a hexagonal cell is $s = 2\sqrt{3}r^2$. On this lattice, each probe can have m = 0, 1, 2, 3, 4, 5, or 6 neighboring probes. For a purely random probe attachment, the probability p_m that the probe has m neighbors is simply

$$p_m = \frac{6!p^m(1-p)^{6-m}}{m!(6-m)!} \quad (3)$$

Here $p = sN_P$ is the probability for a hexagonal cell to be occupied by the probe. Now the local surface density N_m fluctuates depending on the number of probes, m, in the six surrounding cells and $N_m = m/6s$. We apply the isotherm eq 2 to each of the seven subsets m = 0, 1, ...6 and express the melting curve $\theta_m = \theta_m(T)$ from eq 2

$$T = \frac{\Delta H_0 + wN_mZ_P(Z_P + \theta_m Z_T)}{\Delta S_0 + R\ln[C_0(1-\theta_m)/\theta_m]} \quad (4)$$

Equation 4 describes the hybridization yield θ_m for the probes that have m neighbors. The hybridization yield for the entire system is then given by

$$\theta = \sum_{m=0}^{6} p_m\theta_m = \sum_{m=0}^{6}\frac{6!p^m(1-p)^{6-m}}{m!(6-m)!}\theta_m \quad (5)$$

RESULTS

For comparison of the theory with experiments, we use two sets of published data. First, the melting curves were measured for arrays with presynthesized oligonucleotide probes tethered to a gold surface by pyrrol electropolymerization.[3] Three types of the spotted probes are (P9) 5′-(T_{10})-GCC TTG ACG-3′ ($\Delta H_0 = -314.8$ kJ/mol, $\Delta S_0 = -864.8$ J/mol·K, $T_m = 44.4$ °C), (P12) 5′-(T_{10})-GCC TTG ACG ATA-3′ ($\Delta H_0 = -400$ kJ/mol, $\Delta S_0 = -1108$ J/mol·K, $T_m = 51$ °C), and (P14) 5′-(T_{10})-GCC TTG ACG ATA CA-3′ ($\Delta H_0 = -470.7$ kJ/mol, $\Delta S_0 = -1299.6$ J/mol·K, $T_m = 56.8$ °C). Here in brackets we include the hybrid

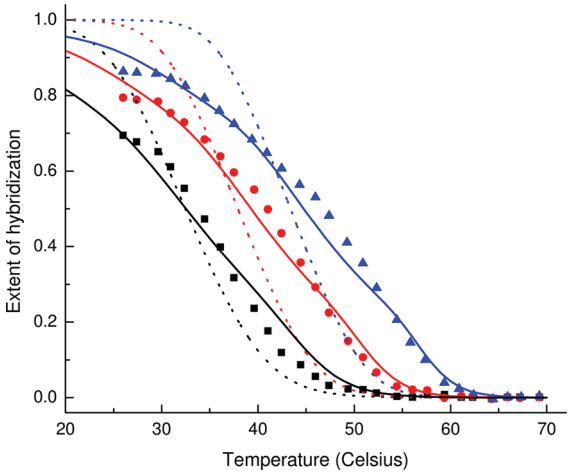


Figure 2. Theoretical and experimental[3] melting curve comparison on gold. Experiment: 9-mer (squares), 12-mer (circles), and 14-mer (triangles) oligonucleotides arrayed on a gold surface. Theory with surface fluctuations eq 5 (solid lines) and homogeneous density of probes eq 2 (dashed lines).

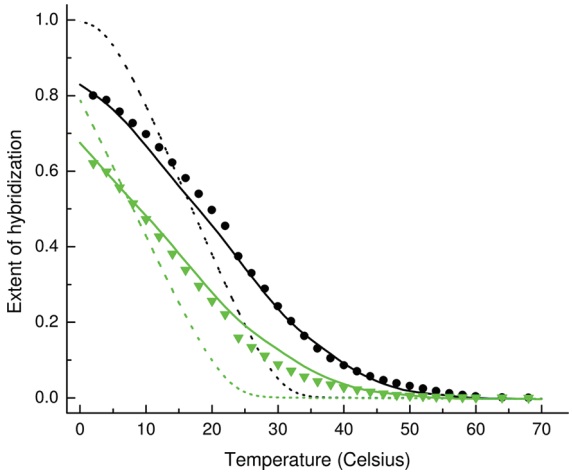


Figure 3. Theoretical and experimental[21] melting curves on glass. Experiment: perfect match (circles) and single mismatch (triangles) 19-mer oligonucleotides synthesized on glass surface. Theory: fluctuating eq 5 (solid lines) and homogeneous eq 2 (dashed lines) density of probes.

formation enthalpy ΔH_0, entropy ΔS_0, and melting temperature in solution $T_m = \Delta H_0/(\Delta S_0 + R\cdot\ln C_0)$ calculated using the NN method.[10] The array was hybridized with complementary target (T) 5′-ACA AAA TGA TTC TGA ATT AGC TGT ATC GTC AAG GCA CTC T-3′ at a concentration C_0 = 250 nM in phosphate buffer with 0.35 M NaCl. The 40-mer target is longer than a probe. The last T base distal from the 5′-end of the (T_{10}) probe linker is complementary to the fifth target A base from the 3′-end and thus increases the length of hybrid by one base pair. This was taken into account in the reference calculation. In addition, the hybridized target has a 4-mer dangling 3′-end near the surface contributing to the repulsive charge. Probe (T_{10}) linkers also provide a repulsive layer. As a result, for the P9 probe, we take $Z_P = 9 + 10 = 19$ and $Z_T = 10 + 4 = 14$, and similarly for the other probes, e.g., for P12 $Z_P = 22$ and $Z_T = 17$. The electrostatic interaction parameter is $w = 4 \times 10^{-16}$ J·m^2/mol in hybridization buffer with 1 M NaCl.[19] Since the probe surface density was only crudely estimated experimentally as $N_P = 10^{13}$ cm^{-2},[24] we slightly vary it in calculations; our calculated melting curves correspond to $N_P = 6 \times 10^{12}$ cm^{-2}.

Figure 2 shows the experimental[3] and calculated melting curves. The presented random surface density fluctuation theory (eq 5) is in excellent agreement with experiment, predicting even the shape change of the curve. In Figure 2 the scaling factors (0.69 for P9, 0.8 for P12, and 0.87 for P14 probes) are applied to the experimental melting curves[3] that were published with arbitrary normalization to θ = 1 at 26 °C, although the hybridization is not saturated at this temperature. For comparison, results for a regular array (no probe density fluctuations) isotherm eq 2 are also presented in Figure 2. The results demonstrate that accord with experiment is markedly improved by accounting for the fluctuations in surface coverage.

Next we compare our theory with the experiments[21] for an array with 19-mer probes synthesized in situ on the glass surface using a photogenerated acid method.[25] Again we expect a random, but not uniform probe distribution on the surface. From a large set of the measured equilibrium melting curves for 640 probe and 64 target sequences on a microfluidic chip with 3968 cells, we use the results for the 19-mer target 5′-ACA AGG ACC ATG ACA ACG A-3′, with complementary perfect match probe PM 5′-T CGT TGT CAT GGT CCT TGT (T_{10})-3′ ($\Delta H_0 = -561.7$ kJ/mol, $\Delta S_0 = -1.542$ kJ/mol·K, $T_m = 56.3$ °C) and single mismatch probe SM 5′-T CGT TGT CAA GGT CCT TGT (T_{10})-3′ ($\Delta H_0 = -504.6$ kJ/mol, $\Delta S_0 = -1.393$ kJ/mol·K, T_m = 51.1 °C). In brackets are shown ΔH_0, ΔS_0 and T_m calculated with the NN method for the experimental 1 M NaCl buffer and target concentration C_0 = 3 nM, but corrected as described in ref 21 for the presence of the denaturing agent formamide, 25% by volume. In calculations we used $Z_T = 19$ and, because of the (T_{10}) linker, $Z_P = 19 + 10 = 29$. Using $N_P = 1.5 \times 10^{17}$ probes/m^2, we calculated the melting curves shown in Figure 3. Again, the probe density fluctuation theory eq 5 (solid lines) is in excellent agreement with experiment, which is a considerable improvement over the assumption of a homogeneous probe density eq 2 (dashed lines).

DISCUSSION AND CONCLUSION

The physical interpretation of the melting curves in Figures 2 and 3 is as follows: At high temperature, only a subset of probes with small numbers of neighbors m is hybridized. In particular, the rise of the melting curve starts closer to the solution melting temperature T_m corresponding to the subset of surface probes with $m = 0$. At lower temperatures, the probe subsets with $m = 1$, 2, 3, 4, 5, and 6 neighboring probes subsequently start to be hybridized. When m increases, the thermodynamic stability and melting temperature of the hybrid decrease because of stronger electrostatic repulsion of the hybridized target from a larger number of neighboring probes.

In summary, we have demonstrated the thermodynamic effects of fluctuations of the local surface probe density. The calculated melting curves are in excellent agreement with microarray experiments. The calculation uses only known NN model DNA thermodynamics in solution as a reference state, and the mean density of probes on the surface. The derived isotherm accurately predicts the hybridization/melting thermodynamics for DNA arrays and thus can be widely used in the design and optimization of microarrays, microbeads, next-generation

sequencing, surface-supported PCR, and biosensors of various kinds.

AUTHOR INFORMATION

Corresponding Author
*E-mail: arvainru@utmb.edu (A.V.); pettitt@uh.edu (B.M.P.).

ACKNOWLEDGMENT

A.V. thanks the NSF (grant 939048) and the Detection and Food Safety Center at Auburn University (Grant USDA-NIFA 20053439415674A), and B.M.P. thanks the Robert A. Welch Foundation (Grant E-1028) and the NIH (Grant GM066813-05) for partial financial support.

REFERENCES

(1) Mardis, E. R. *Annu. Rev. Genomics Hum. Genet.* **2008**, *9*, 387–402.
(2) Fan, J. B.; Chee, M. S.; Gunderson, K. L. *Nat. Rev. Genet.* **2006**, *7*, 632–644.
(3) Fiche, J. B.; Buhot, A.; Calemczuk, R.; Livache, T. *Biophys. J.* **2007**, *92*, 935–946.
(4) Fuchs, J.; Fiche, J. B.; Buhot, A.; Calemczuk, R.; Livache, T. *Biophys. J.* **2010**, *99*, 1886–1895.
(5) Gong, P.; Levicky, R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5301–5306.
(6) Levicky, R.; Horgan, A. *Trends Biotechnol.* **2005**, *23*, 143–149.
(7) Wong, I. Y.; Melosh, N. A. *Biophys. J.* **2010**, *98*, 2954–2963.
(8) Breslauer, K. J.; Frank, R.; Blocker, H.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3746–3750.
(9) SantaLucia, J.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
(10) Le Novere, N. *Bioinformatics* **2001**, *17*, 1226–1227.
(11) Forman, J. E.; Walton, I. D.; Stern, D.; Rava, R. P.; Trulson, M. O. In *Molecular Modeling of Nucleic Acids*; Leontis, N. B., Santa Lucia, J., Eds.; Oxford University Press: New York, 1998; Vol. 682, p 206−228.
(12) Guo, Z.; Guilfoyle, R. A.; Thiel, A. J.; Wang, R. F.; Smith, L. M. *Nucleic Acids Res.* **1994**, *22*, 5456–5465.
(13) Peterson, A. W.; Heaton, R. J.; Georgiadis, R. M. *Nucleic Acids Res.* **2001**, *29*, 5163–5168.
(14) Binder, H.; Kirsten, T.; Hofacker, I. L.; Stadler, P. F.; Loeffler, M. *J. Phys. Chem. B* **2004**, *108*, 18015–18025.
(15) Chen, S.; Phillips, M. F.; Cerrina, F.; Smith, L. M. *Langmuir* **2009**, *25*, 6570–6575.
(16) Vainrub, A.; Pettitt, B. M. *Chem. Phys. Lett.* **2000**, *323*, 160–166.
(17) Halperin, A.; Buhot, A.; Zhulina, E. B. *Biophys. J.* **2004**, *86*, 718–730.
(18) Halperin, A.; Buhot, A.; Zhulina, E. B. *J. Phys.: Condens. Matter* **2006**, *18*, S463–S490.
(19) Vainrub, A.; Pettitt, B. M. *Phys. Rev. E* **2002**, *66*, 041905.
(20) Vainrub, A.; Pettitt, B. M. *J. Am. Chem. Soc.* **2003**, *125*, 7798–7799.
(21) Vainrub, A.; Deluge, N.; Zhang, X.; Zhou, X.; Gao, X. *Methods Mol. Biol.* **2007**, *382*, 393–403.
(22) Hagan, M.; Chakraborty, A. *J. Chem. Phys.* **2004**, *120*, 4958–4968.
(23) Irving, D.; Gong, P.; Levicky, R. *J. Phys. Chem. B* **2010**, *114*, 7631–7640.
(24) Guedon, P.; Livache, T.; Martin, F.; Lesbre, F.; Roget, A.; Bidan, G.; Levy, Y. *Anal. Chem.* **2000**, *72*, 6003–6009.
(25) Gao, X. L.; Gulari, E.; Zhou, X. C. *Biopolymers* **2004**, *73*, 579–596.