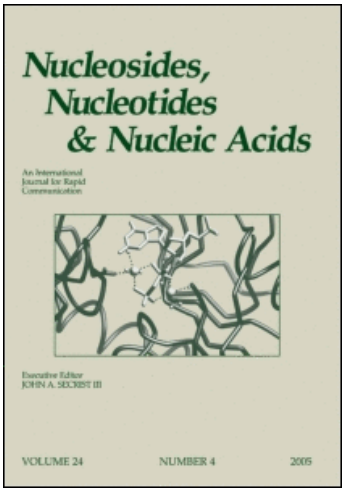

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

ChipcheckII—Predicting Binding Curves for Multiple Analyte Strands on Small DNA Microarrays

Karsten Siegmund[a]; Carolin Ahlborn[a]; Clemens Richert[a]
[a] Institute for Organic Chemistry, University of Karlsruhe, Karlsruhe, Germany

To cite this Article Siegmund, Karsten , Ahlborn, Carolin and Richert, Clemens(2008) 'ChipcheckII—Predicting Binding Curves for Multiple Analyte Strands on Small DNA Microarrays', Nucleosides, Nucleotides and Nucleic Acids, 27: 4, 376 — 388
To link to this Article: DOI: 10.1080/15257770801944147
URL: http://dx.doi.org/10.1080/15257770801944147

Nucleosides, Nucleotides and Nucleic Acids, 27:376–388, 2008

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257770801944147

Taylor & Francis
Taylor & Francis Group

CHIPCHECKII—PREDICTING BINDING CURVES FOR MULTIPLE ANALYTE STRANDS ON SMALL DNA MICROARRAYS

Karsten Siegmund, Carolin Ahlborn, and Clemens Richert

Institute for Organic Chemistry, University of Karlsruhe, Karlsruhe, Germany

□ *Incomplete binding, saturation, and cross-hybridization between partially complementary strands complicate the parallel detection of nucleic acids via DNA microarrays. Treating the competing equilibria governing binding to microarrays requires computational tools. We have developed the web-based program ChipCheckII that calculates total hybridization matrices for target strands interacting with probes on small DNA microarrays. The program can be used to compute the extent of cross-hybridization and other phenomena affecting fidelity of detection based on sequences, quantities of strands, and hybridization conditions as inputs. Enthalpy and entropy of duplex formation are generated locally with UNAfold, including those for complexes that are partially matched. Simulated binding versus temperature curves for portions of a commercial genome chip demonstrate the extent to which cross-hybridization can complicate DNA detection. ChipCheckII is expected to aid nucleic acid chemists in developing high fidelity DNA microarrays.*

Keywords DNA; microarray; hybridization; binding equilibria; simulation

INTRODUCTION

Nature evolved DNA as carrier of genetic information. Today's biomedicine frequently employs short DNA strands as hybridization probes, however. Perhaps the most demanding application for DNA, in terms of sequence-selective hybridization, is the massively parallel detection of complementary strands in hybridization experiments involving microarrays.[1] Microarrays are most commonly used for gene expression profiling,[2] but also for cancer diagnosis[3] and SNP genotyping.[4,5] Hundreds of thousands of duplexes have to form on high density DNA microarrays in the presence of competing strands with very similar sequences. This pushes the selectivity

Received 19 January 2007; accepted 20 January 2007.
The authors wish to thank Dr. M. Bauer, Dr. U. Plutowski, and Prof. U. Steiner for helpful discussion. This work was supported by DFG (FOR 434, grant RI 1063/4).
Address correspondence to Clemens Richert, Institute for Organic Chemistry, University of Karlsruhe (TH) 76131 Karlsruhe, Germany. E-mail: cr@rrg.uka.de

of molecular recognition to its physical limits, and there is little doubt that processes happening on the surface during hybridization are complex,[6,7] and that microarray experiments can suffer from substantial drawbacks.[8] Among these are cross-hybridization,[9] caused by residual affinity of target strands with partial complementarity, as well as false negative results, caused by low stability of A/T-rich duplexes. Further, reproducibility is limited, especially when different microarray systems are used.[10,11] Finally, differences in hybridization kinetics can bias results.[12,13]

So, while DNA microarrays are now widely used for biological and biomedical applications,[14] obtaining reliable, quantitative results will require a careful correction for phenomena caused by competitive binding equilibria and differences in duplex stability between different probe/target duplexes.[15] Simple approaches for correction, such as subtracting signals for mismatched probes as background, have been described as unsatisfactory.[16,17] The development of high fidelity microarrays, therefore, calls for computational tools that allow one to detect weaknesses of possible microarray set-ups and to eventually correct computationally for physical limitations of the molecular recognition process.

It has been reported that thermodynamic data for duplex stability calculated via 'nearest neighbor' approaches correlate well with the thermodynamic data for duplexes formation on microarrays.[18] We developed the program ChipCheck that computes the total hybridization matrices for small DNA microarrays, using either experimental or calculated thermodynamic data for the binding equilibria (Figure 1).[19] ChipCheck does not use cutoffs to reduce the number of equilibria in the calculations, so that

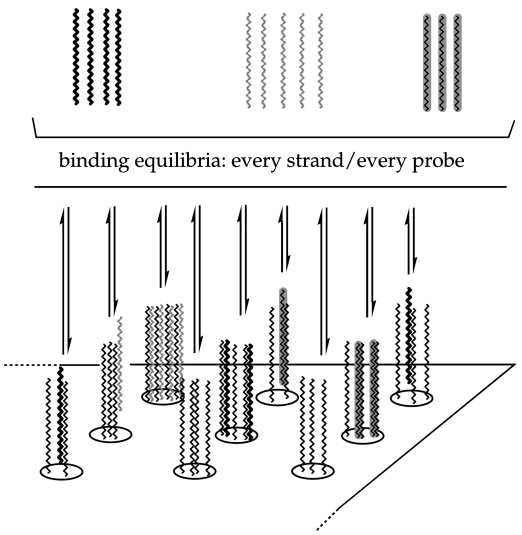

FIGURE 1 The total hybridization matrix approach employed in ChipCheck calculations. Wavy lines in different shades of gray represent DNA target strands of different sequence.

the full extent of aggregate low-level binding events is represented. Further, ChipCheck can be freely accessed on our internet server. Other software for simulation of binding phenomena on microarrays exists,[20–23] as well as computer algorithms for quantitation and visualization,[24] statistical analysis,[25–27] extraction of features,[28] the different sources of noise,[29,30] detection of spot inhomogeneities and background,[31] and eliminating saturation artifacts,[32] but these programs do not compute total hybridization matrices, nor do they focus on competitive equilibria.

For our earlier version of ChipCheck, free enthalpies (ΔG) had to be provided for every possible duplex of the virtual experiment. This required lengthy retrieval of data from servers producing nearest neighbor data, making the ChipCheck version 1.0 unattractive for users wishing to avoid this effort. Further, the binding equilibria were calculated for a single temperature only, making it necessary to perform many rounds of calculations to generate binding curves and, thus, optimized conditions for hybridization. Here, we present ChipCheckII, where the generation of thermodynamic data for all duplexes involved occurs in fully automated fashion. ChipCheckII accesses the program UNAFold,[33] and its advanced algorithm for predicting folding and duplex thermodynamics, for the required ΔH and ΔS values. As a result, no more than sequences and concentrations (as well as hybridization conditions) have to be entered. Further, ChipCheckII readily generates binding versus temperature data, from which optimized incubation conditions may be gleaned.

RESULTS AND DISCUSSION

Simulating Binding of Target Strands at one Temperature

To start a calculation of the DNA distribution on a microarray, the sequences of target strands (in solution) and the probe strands (on the spots of the microarray) must be entered in two text boxes of the data entry mask of ChipCheckII. The sequences must be entered in a FASTA format,[34] in which the concentration of each strand is given in the FASTA comment next to a short name describing that sequence. Additionally, temperature, the volume of the solution, and the salt concentrations have to be entered. For the given experimental conditions, ΔH and ΔS are calculated for a specific temperature using the program UNAFold.[33] So, while the first version of ChipCheck required ΔG values to be entered for every binding equilibrium and temperature as a user input to calculate the equilibrium constant, ChipCheckII has been extended to generate value pairs of ΔH and ΔS and to then accept these in the input file. The thermodynamic data is stored in a local file, whose content is displayed on a separate HTML page, where it may also be edited. From this, ΔG is calculated via the Gibbs-Helmholtz equation ($\Delta G = \Delta H - T\,\Delta S$) for every temperature

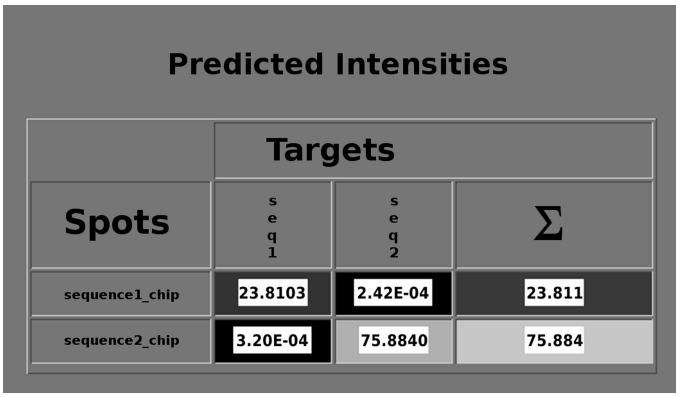

FIGURE 2 ChipCheckII output display for a minimal microarray, consisting of two spots, incubated with two target sequences at a temperature of 35°C. Each row represents a single spot on the chip, while each column corresponds to one sequence in solution. The last column (Σ = total signal) shows the sum of all hybridization to the DNA of a spot, corresponding to signal intensity observed in a microarray experiment. The lighter the spot, the greater the extent to which binding occurred.

chosen. The concept of an input file containing the thermodynamic data was retained to enable the use of experimental enthalpies and entropies, for example, from solution experiments instead of the ones calculated by a nearest-neighbor approximation. With all necessary data generated, the main algorithm then calculates the distribution of all target strands on the respective spots. This algorithm arrives at the final distribution through iterations employing the multidimensional Newton method.[19] Figure 2 shows an output for a very small model microarray, consisting of only two spots and two different sequences in solution.

Simulation of Target Strand Binding for a Subset of Spots on an Affymetrix Microarray

We previously employed ChipCheck to calculate DNA distributions on oligonucleotide arrays with target strands of equal length.[19] Arrays with 100 spots were shown to require no more than a few seconds computational time. To demonstrate that other setups, more common in the biomedical community, can be treated successfully, a calculation was run with a subset of the sequences of a commercial microarray (Arabidopsis ATH1 Genome Array, Affymetrix, Inc., Santa Clara, CA, USA),[35] featuring 25mer probe sequences for the detection of a total of 25,000 target sequences. From the probe sequences of this chip, one set of sequences with significant similarity were chosen for a ChipCheckII simulation. Figure 3 shows the result of this calculation involving 385 spots and 17 target sequences in solution. The sequences and conditions employed are detailed in the Supporting Information. (The supporting information can be obtained online at http://chip.chemie.uni-karlsruhe.de/sup-inf/ or from the authors.) The

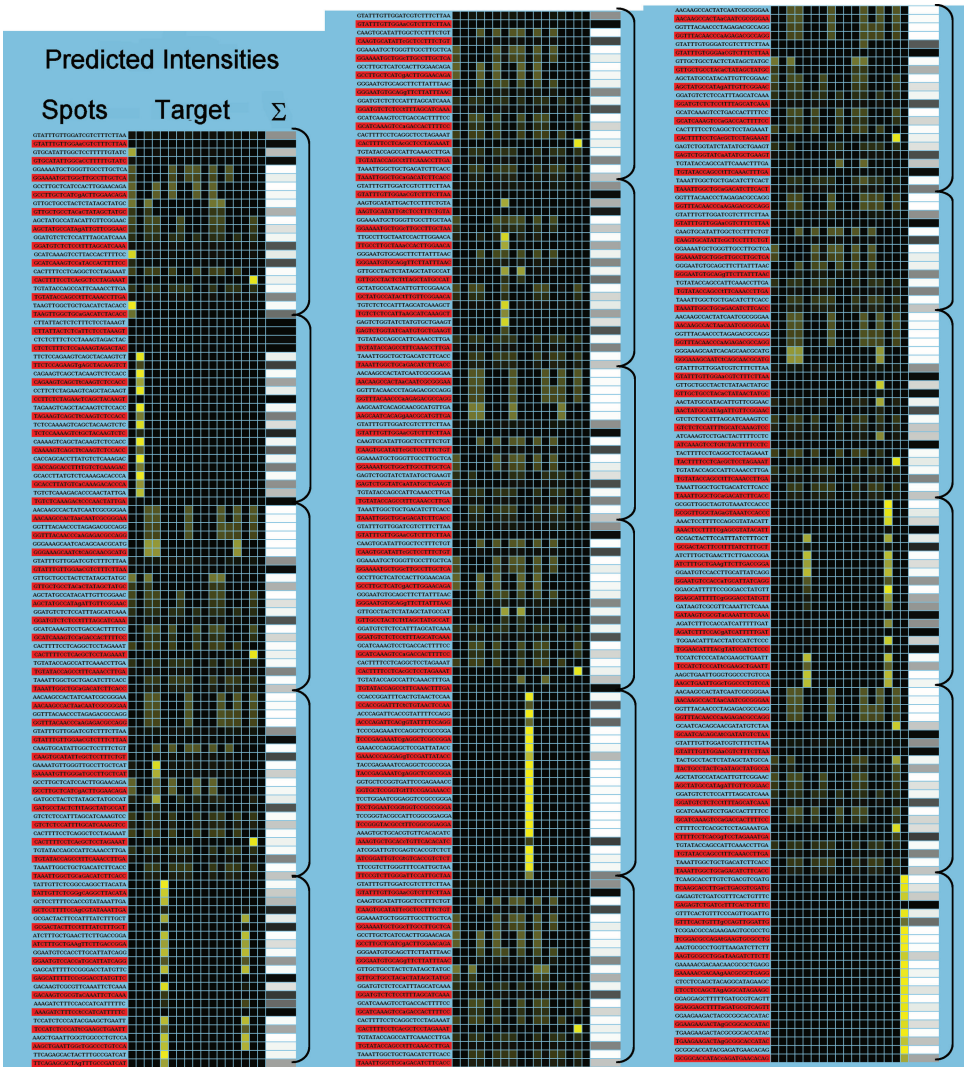

FIGURE 3 ChipCheckII output showing the binding of target DNA to the spots of a Affymetrix ATH1 genome array. Spots for the (multiple) detection of 17 transcripts of several hundred nucleotides each are shown. The probes detecting different regions of the same target are indicated by a bracket on the right-hand side. Perfect match probes are on light-blue lines, and mismatched probes are on red lines. The simulation was performed with a 5 pM concentration of target strands, a value typical for spike-in strands. For an ideal chip, one would expect a repetitive pattern of light and dark spots similar to the one observed for the area labeled by the second bracket. This pattern should "move" from the left-most bracketed column (first target sequence) to the next one, and so on.

computational time on the current ChipCheck server was under 10 seconds for calculating the distribution at a single temperature. Generating the data for the binding curves in the temperature range of 0°C to 100°C required less than 5 minutes (time for generating graphs not included).

It can be discerned that the pattern of light (mostly covered) and dark (poorly covered) spots is not as regular as one could wish for a microarray detecting an equimolar mixture of transcripts via multiple probe spots. Also, the extent of cross-hybridization observed is as worrisome as the number of false negative results (spots that fail to bind the target sequence, even though they are fully complementary to the target). This suggests that for problematic sequences, it may be difficult to achieve reliable detection on high density microarrays. We would like to emphasize that the results are not representative, as the sequences chosen include similar and in some cases identical sequences on different spots that are used on this microarray. General conclusions on the expected fidelity of this microarray will require simulation on the entire sequence space for this chip.

Binding versus Temperature Curves

Optimization of microarray experiments requires judicious choice of sequences and experimental conditions. Perhaps the most difficult-to-choose parameter is the exact temperature of incubation. While the dependence of duplex stability on ionic strength, G/C content, and oligonucleotide length is generally known, it is not as clear what the best temperature for the most universally stringent incubation is that does not lead to massive false negative results. Identifying such a temperature based on simulations with the earlier version of ChipCheck required manually entering many different sets of ΔG values and starting a long series of individual calculations, together with manually managing the results. The same is true for binding curves generated to calibrate simulation parameters, based on experimental data from calibration experiments.

ChipCheckII can automatically perform calculations at different temperatures for a given set of sequences and experimental conditions. The fully automated generation of entire binding curves for each possible duplex formed on a given spot leads to binding/melting curves such as those shown in Figure 4a. In addition, ChipCheckII generates a graph showing the overlay of the binding curves for all spots and target sequences of a given example. An example for this type of graph, again for the simplest case possible, is shown in Figure 4b.

The results of calculations for a more complex example are shown in Figure 5. Here, the hybridization onto a microarray consisting of 100 spots with 25mer probe strands was simulated. Between one and six target strands of several hundred nucleotides in length were employed at different concentrations. The sequences are a subset of those employed in a microarray study on gene expression in Drosophila.[36]

In the simplest case, where slightly above one equivalent of one target strand was employed, so that full coverage of all complementary probes at low temperature could be expected, very little cross-hybridization was

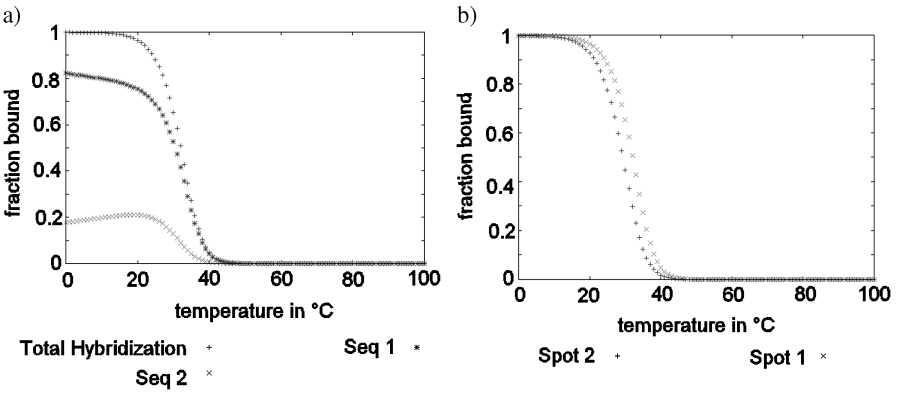

FIGURE 4 Binding versus temperature curves for a minimal microarray consisting of two spots, each featuring $4\cdot \times 10^{-15}$ mol of an undecamer probe oligonucleotide (GCAGTCAACCA or ACAGTCAACCA) and two target strands in solution at an initial concentration of 1.33×10^{-10} M (TGGTTGACTGC, black asterisks, and TGGTTGACTGT, gray diagonal crosses) in 30 μL buffer containing 1 M NaCl. a) Binding curves for an individual spot. The extent to which different target strands bind is shown, together with the total signal to be expected for a non-sequence-specific detection method (total hybridization, black crosses). Note that the extent to which the mismatched strand contributes to binding (gray diagonal crosses) reaches a maximum at approx. 25°C. b) Overview of binding curves for both spots of the microarray (total signal).

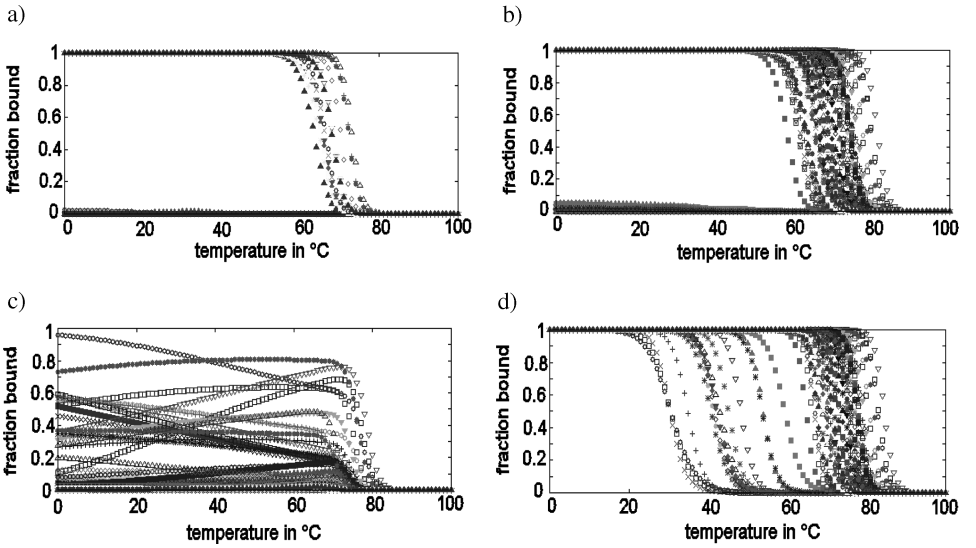

FIGURE 5 Binding versus temperature curves for a microarray consisting of 100 spots with 4×10^{-15} mol 25mer probe oligonucleotides covering potential 8 target transcripts of the Drosophila genome. The volume of the hybridization solution was set to 30 μL. The curves are the total binding occurring on a given spot, irrespective of what targets bind to what extent. This is the equivalent of the total signal one can expect via fluorescence intensity, if the same fluorophore is used for labeling all targets. a) One target in solution at an initial concentration of 1.87×10^{-9} M, that is, approximately 1.03 equivalents to the fully complementary probe sequences on the surface. b) Six targets, again at an initial concentration of 1.87×10^{-9} M each. c) Six targets at an initial concentration of 1.87×10^{-10} M, that is, approximately 0.1 equivalents to the complementary probes. d) Six targets at an initial concentration of 1.87×10^{-8} M, that is, 10 equivalents to the complementary probes.

observed (Figure 5a). The midpoints of the binding curves are spread over a range of approximately 10°C, so that reasonably selective detection can be expected, assuming that a single mismatch in the interior of the duplex will cause 10–20°C melting point depression. When six of the potential eight target strands were employed at 1.03 equivalents (Figure 5b), modest cross-hybridization was again observed, even at low temperatures. The binding curves for the complementary probes was spread over a significantly larger temperature interval, though, (approximately 20 °C), so that it would seem more difficult to find universally stringent conditions without risking significant cross-hybridization. When substoichiometric amounts of six targets were used, so that the probe strands competed for a limited supply of target in solution, the different spots attracted vastly different amounts of their respective target, and cross-hybridization became significant (Figure 5c). In the most challenging case in terms of cross-hybridization, namely a 10-fold excess of the six target strands, all probe-displaying spots attracted target strands, including the 14 spots that lack a target they were designed for (Figure 5d). In fact, at room temperature, all spots on the microarray are fully saturated, due to cross-hybridization. Coverage versus temperature curves, such as the one shown in Figure 5d may be particularly useful for identifying problematic probe sequences that one might want to change prior to fabrication of a microarray. These give binding curves close to those for the fully complementary probes. The data also shows how much cross-hybridization depends on the experimental conditions.

An interesting observation was made for some experimental conditions. In simulations where the target strands were present at a large excess, the extent to which partially mismatched target strands bound to the probe sequences on the surface depended on the temperature. In a number of cases, the selectivity increased when the temperature dropped well below the melting point of the duplex, that is, the fully matched target sequence bound more strongly relative to the mismatched competitors than at what would usually be considered more stringent conditions (higher temperature). The resulting maxima in the binding of mismatched targets (compare Figure 5c) appear to be the result of the subtle changes in enthalpy/entropy compensation for duplexes with PM and MM sequences, affecting the ΔG values. A particularly striking example is given in Figure S1 of the Supporting Information. It displays curves with multiple maxima, obtained with a minute amount of incubation solution and ten target strands, for two of which no probe strands were present. It is interesting to ask whether this phenomenon is a computational artifact or whether it can also be observed experimentally. If it is real, it might have practical consequences, as the selectivity of affinity-selections, not just on microarrays, should increase with decreasing temperature in certain cases.

DISCUSSION

Simulating the results of hybridization experiments involving microarrays is important to identify physical limits of sequence-selective detection. These can lead to substantial limitations. The number of spots on commercial microarrays continues to increase,[37] and so does the density of spots. Increasingly demanding applications come into focus for the experimentalist,[38] such as genotyping single nucleotide polymorphisms, where a single nucleobase has to be detected unambiguously in a lengthy sequence. Selective, quantitative detection is severely complicated by homologies between genes, and biases introduced during amplification. The difficulties of achieving high fidelity results under these circumstances are substantial, even for the most advanced commercial platforms.[39] The use of decorated nucleic acids as probes may lead to higher fidelity. Their binding to target strands has been predicted successfully, using ChipCheckII.[40]

ChipCheckII, despite the added capabilities, does not allow for the simulation of high density microarrays with very high spot number, as it relies on the most rigorous approach of considering every binding equilibrium, including those with very weak binding, whose aggregate effect ever so often produces the high "noise" levels well known to practitioners in the field. This approach necessarily limits the size of the system that can be treated. ChipCheckII is expected to be useful for those who wish to develop (or check) focused microarrays of moderate size, as well as those who wish to study fundamental phenomena underlying the limited fidelity of molecular recognition on microarrays. Examples, such as the one shown in Figure 3, show that even for seemingly optimized platforms, severe false positives and false negatives may result from unchecked gene chip experiments.

Of course, any simulation is only as good as the thermodynamic parameters it uses. New approaches for treating the thermodynamics of duplex formation of microarray continue to appear,[41] and future versions of ChipCheck may benefit from these advances. Currently, ChipCheckII assumes that a target strand will form only one duplex and that every additional strand that can hybridize to it will do so in a competing equilibrium. While this is a reasonable assumption for typical microarrays, more complicated scenarios, involving (partial) duplex formation between target strands that do not interfere with binding of one of these strands are conceivable and may be accommodated in future versions of the program. This will necessarily increase the CPU time for simulations, an effect that might be compensated by advances in computer technology. Currently, the UNAFold calculations are the limiting factor in the simulations. We estimate that in its current form, microarray settings featuring $\leq$3000 spots and the same number of target strands are the upper limit for ChipCheckII calculations that our server, operating with an AMD 2000+ chip set, may

compute in less than 1 day. Whether the calculations will run stably for the resulting approximately 10′000′000 equilibria has not been tested thus far.

The use of ChipCheckII is not limited to oligonucleotide microarrays. Other forms of competing equilibria for which the thermodynamics of complex formation are known, may also be simulated, possibly including oligosaccharide[42] or protein microarrays.[43] Further, the program also may be eventually used to calculate the complexes formed when binding processes compete in solution only, without immobilization of either of the complex-forming partners. We have performed exploratory work in this direction, the results of which suggest that such calculations are feasible.

In summary, the web-based program ChipCheckII simulates small to medium size DNA microarray experiments based on sequence data, concentrations, and hybridization parameters. The program treats competitive binding equilibria occurring on the surface of DNA microarrays. It can calculate entire on-chip melting profiles for hybridization. ChipCheckII has been set up to retrieve thermodynamic parameters for binding equilibria locally from the program UNAfold.[33] It is not necessary for the sequences to be fully complementary or to have the same length. The results of ChipCheckII simulations are made available as text files and in graphic form. For equilibria at a single temperature, the results are displayed in form of a HTML table. The binding curves are made available as PNG graphics or PDF. The program ChipCheckII can be accessed at http://chipcheck.chemie.uni-karlsruhe.de/ and can be made available for download upon request.

MATERIALS AND METHODS

General

ChipCheckII utilizes algorithms described for the earlier version.[19] Briefly, the equation for the law of mass action is used, together with the equations for the conservation of mass. For every individual spot of a microarray (index i), the mole fractions of bound DNA and unbound, single stranded DNA in solution (index j) are compared to the initial amounts of DNA, resulting in the following equations:

$$K_{ij} = \frac{\tilde{x}_{ij}}{x_i c_j} \text{ law of mass action on a single spot}$$

$$\sum_j \tilde{x}_{ij} + x_j = 1 \text{ definition of mole fraction on a single spot}$$

$$\sum_i (n_i \tilde{x}_{ij}) + V_{cj} = V c_{oj} \text{ starting concentration}$$

Here, V is the volume, n_i is the molar amount of strands on spot i, K_{ij} is the equilibrium constant for a binding event, c_j is the concentration of single stranded DNA in solution, x_{ij} is the mole fraction of bound DNA (j) on spot i, x_i is the mole fraction of single stranded DNA on spot i, and co_j is the initial concentration of DNA in the sample. These equations can be combined into a single equation:

$$\tilde{x}_{ij} = \frac{K_{ij}\left[co_j - \left(V^{-1}\sum_i n_i\tilde{x}_{ij}\right)\right]}{1 + \left[\sum_j K_{ij}\left(C_{oj} - V^{-1}\sum_i n_i\tilde{x}_{ij}\right)\right]}$$

This combined equation calculates the mole fraction of each possible DNA duplex between a target strand (index j) in solution and an immobilized probe strand (index i) on the microarray.

Software Used

ChipCheckII itself, as well as algorithms producing whole binding curves via ChipCheckII were written in the scripting language Perl 5 under i386 Debian GNU/LINUX. User interactions are accomplished by running ChipCheckII via ASCII input files and the cgi-bin interface of an Apache web server. Thermodynamic data for DNA duplexes are calculated locally using UNAFold.[33] Other software producing thermodynamic data may be used and the ΔH and ΔS values entered directly, as described for the earlier version of ChipCheck.

REFERENCES

1. a) Fodor, S.P.A.; Read, J.L.; Pirrung, M.C.; Stryer, L.; Lu, A.T.; Solas, D. Light-directed spatially addressable parallel chemical synthesis. *Science* **1991**, 251, 767–773. b) Debouck, C.; Goodfellow, P. N.DNA microarrays in drug discovery and development. *Nat. Genetics* **1999**, 21, 48–50. c) Pirrung, M.C. How to make a DNA chip. *Angew. Chem.* **2002**, 114, 1327–1341; *Angew. Chem. Int. Ed.* **2002**, 41, 1276–1289.
2. Duggan, D.J.; Bittner, M.; Chen, Y.; Meltzer, P.; Trent, J.M. Expression profiling using cDNA microarrays. *Nature Genetics* **1999**, 21, 10–14.
3. Ragoussis, J.; Elvidge, G. Affymetrix GeneChip system: moving from research to the clinic. *Expert Rev. Mol. Diagn.* **2006**, 6, 145–152.
4. Matsuzaki, H.; Dong, S.; Loi, H.; Di, X.; Liu, G.; Hubbell, E.; Law, J.; Berntsen, T.; Chadha, M.; Hui, H.; Yang, G.; Kennedy, G.C.; Webster, T.A.; Cawley, S.; Walsh, P.S.; Jones, K.W.; Fodor, S.P.A.; Mei, R. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat. Methods* **2004**, 1, 109–111.
5. Gewin, V. Array of possibilites opens up in genotyping. *Nature* **2005**, 435, 1159.
6. Naef, F.; Lim, D.A.; Patil, N.; Magnasco, M. DNA hybridization to mismatched templates: a chip study. *Phys. Rev. E.* **2002**, 65.040902/1–040902/4.
7. Bishop, J.A.; Blair, S.M.; Chagovetz, A.M. A competitive kinetic model of nucleic acid surface hybridization in the presence of point mutants. *Biophys. J.* **2006**, 90, 831–840.
8. Marshall, E. Getting the noise out of gene arrays. *Science* **2004**, 306, 630–631.

9. Plutowski, U.; Richert, C. A direct glimpse of cross-hybridization: background-passified microarrays that allow mass spectrometric detection of captured oligonucleotides. *Angew. Chem.* **2005**, 117, 627–631; Angew. *Chem. Int. Ed.* **2005**, 44, 621–625.
10. Kuo, W.P.; Jenssen, T.-K.; Butte, A.J.; Ohno-Machado, L.; Kohane, I.S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **2002**, 18, 405–412.
11. Tan, P.K.; Downey, T.J.; Spitznagel, E.L.Jr.; Xu, P.; Fu, D.; Dimitrov, D.S.; Lempicki, R.A.; Raaka, B.M.; Cam, M.C. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res.* **2003**, 31, 5676–5684.
12. Wang, Y.; Wang, X.; Guo, S.W.; Ghosh, S. Conditions to ensure competitive hybridization in two-color microarray: a theoretical and experimental analysis. *Biotechniques* **2002**, 32, 1342–1346.
13. Zhang, Y.; Hammer, D.A.; Graves, D.J. Competitive Hybridization Kinetics Reveals Unexpected Behavior Patterns. *Biophys. J.* **2005**, 89, 2950–2959.
14. Matsuzaki, H.; Dong, S.; Loi, H.; Di, X.; Liu, G.; Hubbell, E.; Law, J.; Berntsen, T.; Chadha, M.; Hui, H.; Yang, G.; Kennedy, G. C.; Webster, T. A.; Cawley, S.; Walsh, P. S.; Jones, K. W.; Fodor, S. P. A.; Mei, R. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat. Methods* **2004**, 1109–111.
15. Halperin, A.; Buhot, A.; Zhulina, E.B. Brush effects on DNA chips: thermodynamics, kinetics and design guidelines. *Biophys. J.* **2005**, 89, 796–811.
16. Irizarry, R.A.; Bolstad, B.M.; Collin, F.; Cope, L.M.; Hobbs, B.; Speed, T.P. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **2003**, 31, e15.
17. Wu, C.L.; Carta, R.; Zhang, L. Sequence dependence of cross-hybridization on short oligo microarrays. *Nucleic Acids Res.* **2005**, 33, e84.
18. Held, G.A.; Grinstein, G.; Tu, Y. Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 7575–7580.
19. Siegmund, K.; Steiner, U.E.; Richert, C. CHIPCHECK—A program predicting total hybridization equilibria for DNA binding to small oligonucleotide microarrays. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 2153–2162.
20. Li, C.; Wong, W.H. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **2001**, 98, 31–36.
21. Chu, T.M.; Weir, B.; Wolfinger, R. A systematic statistical linear modeling approach to oligonucleotide array experiments. *Math. Biosci.* **2002**, 176, 35–51.
22. Lemon, W.J.; Palatini, J.J.T.; Krahe, R.; Wright, F.A. Theoretical and experimental comparisons of gene expression indexes for oligonucleotide arrays. *Bioinformatics* **2002**, 18, 1470–1476.
23. Zhang, Y.; Hammer, D.A.; Graves, D.J. Competitive hybridization kinetics reveals unexpected behavior patterns. *Biophys. J.* **2005**, 89, 2950–2959.
24. Strehlow, D. Software for quantitation and visualization of expression array data. *Biotechniques* **2000**, 29, 118–121.
25. Lazaridis, E.N.; Sinibaldi, D.; Bloom, G.; Mane, S.; Jove, R. A simple method to improve probe set estimates from oligonucleotide arrays. *Math. Biosci.* **2002**, 176, 53–58.
26. Zhang, L.; Wang, L.; Ravindranathan, A.; Miles, M.F. A new algorithm for analysis of oligonucleotide arrays: Application to expression profiling in mouse brain regions. *J. Mol. Biol.* **2002**, 317, 225–235.
27. Pan, W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **2002**, 18, 546–554.
28. Schadt, E.E.; Li, C.; Ellis, B.; Wong, W.H. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J. Cell. Biochem.* **2001**, S37, 120–125.
29. Albers, C.J.; Jansen, R.C.; Kok, J.; Kuipers, O.P.; vanHijum, S.A. SIMAGE: simulation of DNA-microarray gene expression data. *BMC Bioinformatics* **2006**, 7, 205.
30. Nykter, M.; Aho, T.; Ahdesmäki, M.; Ruusuvuori, P.; Lehmussola, A.; Yli-Harja, O. Simulation of microarray data with realistic characteristics. *BMC Bioinformatics* **2006**, 7, 349.
31. Balagurunathan, Y.; Dougherty, E.R. Simulation of cDNA microarrays via a parameterized random signal model. *J. Biomed. Opt.* **2002**, 7, 507–523.
32. Hsiao, L.L.; Jensen, R.V.; Yoshida, T.; Clark, K.E.; Blumenstock, J.E.; Gullans, S.R. Correcting for signal saturation errors in the analysis of microarray data. *Biotechniques* **2002**, 32, 330.
33. Markham, N.R.; Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* **2005**, 33, W577–W581.

34. (a) Pearson, W.R. Computer analysis of sequence data, Part II. *Methods Mol. Biol.* **1994**, 24, 307–331. (b) Pearson,W. R., Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **1990**, 183, 63–98.
35. Further information currently available on the internet at http://www.affymetrix.com/products/arrays/specific/arab.affx
36. Zinke, I.; Schütz, C.S.; Katzenberger, J.D.; Bauer, M.; Pankratz, M.J. Nutrient control of gene expression in Drosophila: microarray analysis of starvation and sugar-dependent response. *EMBO J.* **2002**, 21, 6162–6273.
37. Pennisi, E. The ultimate gene gizmo: humanity on a chip. *Science* **2003**, 302, 211.
38. Service, R.F. The race for the $ 1000 genome. *Science* **2006**, 311, 1544–1546.
39. Waldman, M. The chips are down. *Nature* **2006**, 444, 256–257.
40. Ahlborn, C.; Siegmund, K.; Richert, C. Isostable DNA. *J. Am. Chem. Soc.* **2007**, 129, 15218–15232.
41. Binder, H. Thermodynamics of competitive surface adsorption on DNA microarrays. *J. Phys. Condensed Matter* **2006**, 18, S491–S523.
42. Selected references: (a) Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M. Global analysis of protein activities using proteome chips. *Science* **2001**, 293, 2101–2105. (b) Flitsch, S.L.; Ulijn, R.V. Sugars tied to the spot. *Nature* **2003**, 421, 219–220. (c) Byon, H.R.; Hong, B.J.; Gho, Y.S.; Park, J.W.; Choi, H.C. Pseudo 3D single-walled-carbon nanotube film for BSA-free protein chips. *ChemBioChem* **2005**, 6, 1331–1334.
43. (a) See e.g.: De Smet, L.C.P.M.; Stork, G.A.; Hurenkamp, G.H.; Sun, Q.Y.; Topal, H.; Vronen, P.J.; Sieval, A.B.; Wright, A.; Visser, G.M.; Zuilhof, H.; Sudholter, E. Covalently attached saccharides on silicon surfaces. *J. Am. Chem. Soc.* **2003**, 125, 13916–13917. (b) Biskup, M.B.; Müller, J.U.; Weingart, R.; Schmidt, R.R. New methods for the generation of cyrbohydrate arrays on glass slides and their evaluation. *ChemBioChem* **2005**, 6, 1007–1015.