Oligonucleotide microarrays: widely applied—poorly understood

Alex E. Pozhitkov, Diethard Tautz and Peter A. Noble

Abstract
Microarray technology, which has been around for almost two decades, now provides an indispensable service to the biomedical research community. Soaring demand for high-throughput screening of genes potentially associated with cancer and other diseases, as well as the increased need for identifying microorganisms, have substantially opened up the application of this technology to many fields of science, including new ones such as array-based comparisons of whole genomes. Yet, despite this significant progress, the fundamental understanding of the pillars of this technology, have been largely unexplored, in particular for oligonucleotide-based microarrays. In fact, most of the current approaches for the design of microarrays are based on 'common-sense' parameters, such as guanine-cytosine content, secondary structure, melting temperature or possibility of minimizing the effects of nonspecific hybridization. However, recent experiments suggest that these are inadequate. Here we discuss these results, which challenge the basic principles and assumptions of oligonucleotide microarray technology. It is clear that more systematic physicochemical studies will be required to better understand the hybridization and dissociation behaviour of oligonucleotides.

Keywords: nonspecific hybridization; oligonucleotide arrays; stringent wash; probe design; thermodynamics; kinetics

INTRODUCTION
DNA microarrays are regarded as an enabling technology because they allow scientists to address previously intractable problems and to uncover potentially novel gene targets, perhaps those under-lying genetic causes of many human diseases [1]. They are also applicable to many other areas of biomedical science, including biological defense, and environmental monitoring. The utilization of micro-array technology has significantly increased since 1995, with the largest increase in gene expression studies, and to a lesser extent, microbial identification (Figure 1).

Two array platforms are currently used: glass [2–6] and microfluidic station [7–9]. In addition, several platforms are currently under development: microbead [10, 11], electronic [12–14], cantilever arrays [15] and surface plasmon resonance arrays [16]. These platforms share the common attribute that a sensor detects a signal from target sequences hybridized to immobilized oligonucleotide probes. The intensity of this signal provides a measure of the amount of bound nucleic acids in a sample, which mostly depends on probe-target binding affinities and the concentration of nucleic acid in solution.

All microarray experiments rely upon probe specificity and probe sensitivity, and a unifying algorithm that interprets the multiple signals coming from array probes. Probe specificity is of particular importance because one has to be able to adequately distinguish between closely related nucleic acid targets, which may differ by only a few nucleotides. The oligonucleotide probes must possess adequate sensitivity to nucleic acid targets

Corresponding author. Peter A. Noble, Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA. Tel: 206-685-7583; Fax: 206-685-3836; E-mail: panoble@washington.edu

Alexander Pozhitkov is a Research Assistant Professor at University of Southern Mississippi and is studying hybridization and dissociation of nucleic acids on oligonucleotide arrays.
Diethard Tautz is currently a Professor in the Institute for Genetics at the University of Cologne and will soon move to the Max-Planck Institute for Evolutionary Biology in Ploen, Germany, as the head of the section 'Evolutionary Genetics'.
Peter A. Noble is a Research Assistant Professor at University of Washington and works with the other authors in developing a statistically robust approach to quantifying nucleic acid targets in complex target mixtures.
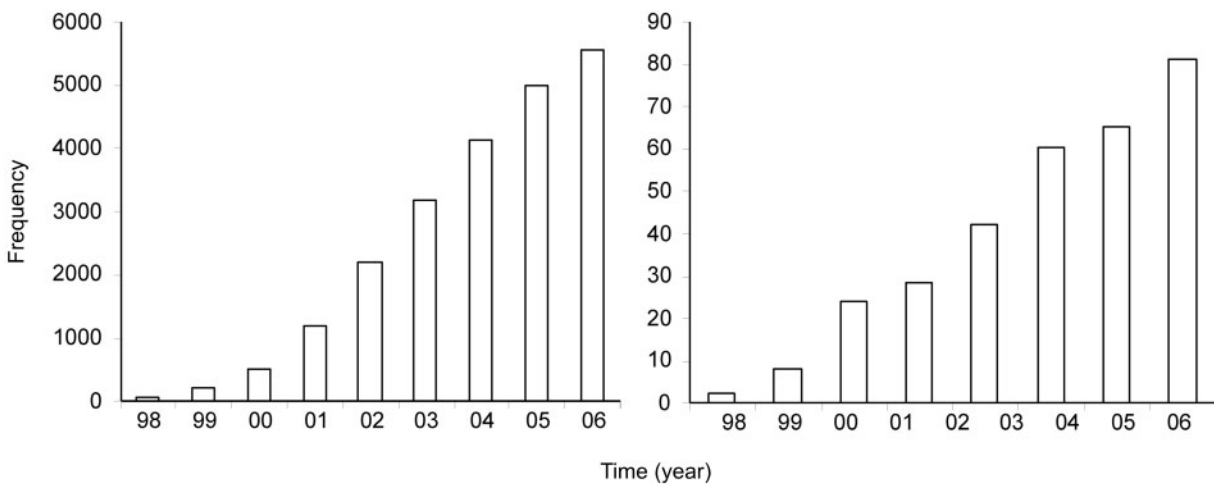
Figure I: Histogram of publications by time for gene expression (left panel) and microbial identification (right panel) array studies based on citations from the Web of Science.

in order to provide a high signal-to-noise ratio, else the signal will be drowned by background noise. Also, an adequate relationship between target concentration and observed signal must exist in order to detect target fold changes upon different treatments. The image acquisition and array software is the integrative component of a microarray experiment because it not only interprets signal from multiple probes that vary in terms of their specificities and sensitivities, but it also plays a pivotal role in calling and/or quantifying a gene target, or several gene targets, within a pool of potentially similar targets.

Recent advances in physicochemistry experiments reveal major problems with the principles of probe design and associated software as well as the meaning of probe specificity in mixed target samples. These problems cast doubt on the downstream interpretation of algorithms used to detect and quantify gene transcripts in biological samples having multiple targets present simultaneously. These findings might also help explain some of the concerns raised by researchers about the repeatability, reproducibility and comparability of microarray technologies [17–19], especially when their results are compared to those of conventional approaches [20]. Comparison of gene expression measurements across different microarray platforms has recently shown that although highly expressed genes yield consistent results, low expression genes remain a major problem [21]. Two-dye platforms might potentially be more tolerant to artefacts, such as sticky oligonucleotides, since those artefacts should be in a two-dye platform. However, this expectation

was not confirmed in the respective comparative studies [21, 22]. For these reasons, we focus this article on single-dye platforms. Moreover, single-dye platforms offer much more flexibility for biological studies since they allow the possibility to study gene expression dynamics resulting from multiple treatments compared to two-dye platforms, which mostly assess up and down regulation between a control and a single treatment.

Subsequently, we discuss recent developments in the understanding of probe specificity and probe sensitivity, and the challenge of interpreting array data using existing approaches. These developments will be placed in the perspective of physicochemistry in order to help elucidate the link between the observed (raw) signals on an array and their biological meaning.

ARRAY PROBE DESIGN

In this section, a brief survey of available probe design approaches is presented. In the next section, we will show that a critical evaluation of these approaches is needed since predicted signal intensities of probes are poorly correlated to what is actually observed.

Earlier work suggested that high specificity probes could be designed by following simple rules [23]. For example, if the composition of target and nontarget sequences was known in a sample, the probes should be designed to have <75% sequence similarity to nontarget sequences, and in the case of marginal similarity between target and nontarget sequences,
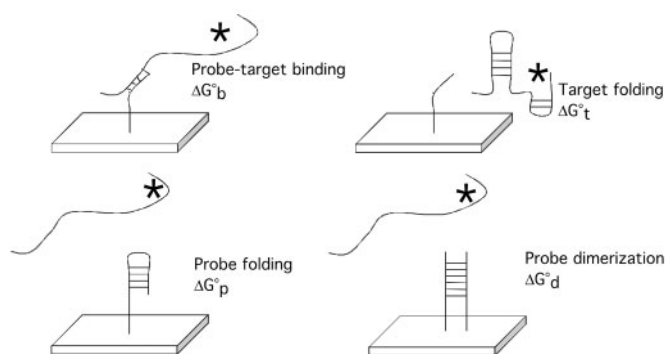
Figure 2: Presumptive processes and ΔG° terms that influence signal intensities. Each panel depicts a labelled (*) target and an immobilized oligonucleotide probe on an array. Adapted from Pozhitkov *et al*. [57].

the probes should be designed to have less than 15 contiguous (complementary sequence) bases. Problems with this approach are: (i) that one often does not know the composition of nontarget sequences in a sample, so it is not possible to determine the degree of similarity nor the number of contiguous bases; and (ii) that a comprehensive study that analysed 13 independent hybridizations to high densities arrays revealed that neither simple nor advanced rules were sufficient to successfully predict hybridizations [6].

An alternative probe design approach is to predict duplex formation by their thermodynamic properties [24–26]. Although duplex formation has been established for nucleic acid targets in solution [27, 28] by the nearest-neighbor model [29], duplex formation on surface-immobilized DNA oligonucleotides is less well understood, presumably due to the complex factors affecting the kinetics and thermodynamics of target capture. Some factors affecting duplex formation on DNA microarrays include: probe density, microarray surface composition, and the stabilities of oligonucleotide–target duplexes, intra- and inter- molecular self-structures, and RNA secondary structures [30–32].

To better understand the contribution of various factors to duplex formation, Matveeva *et al*. [30] proposed that hybridization on an array can be explained by several overlapping processes which include (i) the affinity of a target to bind to a probe, (ii) the formation of stem-loop structures of a probe, (iii) the formation of secondary structure (loops and helices) of a target and (iv) probe-to-probe dimerization (Figure 2), with each process characterized by its own Gibbs free energy value. In the case of increased binding for process (i), one would expect signal intensity of a duplex to increase, while in the case of increased binding for processes (i) to (iv), one would expect signal intensity to decrease because either the target or probe are involved in side reactions that do not contribute to duplex formation.

It should be noted that in addition to thermodynamic prediction, various models have also been examined to account for observed signal intensity based on probe-target binding affinities. For example, we have previously proposed a probabilistic *ad hoc* model that selects probes based on the position of mismatches in the probes while considering the possibility of single nucleotide outloops [33]. Another study by Naef and Magnasco [34] and Mei *et al*. [26] both described an *ad hoc* model that examined the affinity of a probe to a target based on the sum of position-dependent base-specific contributions. Zhang *et al*. [35] described an *ad hoc* model that considered position-dependent nearest-neighbor effects. Held *et al*. [36] examined the effects of free energies of RNA–DNA duplex formation. Wu and Irizarry [37] developed a model that considered both stochastic and deterministic aspects of probe-target hybridizations. The unifying features of these studies are: (i) they are based on the analysis of multiple probes targeting mRNA transcripts (i.e. expression data), (ii) with exception of Held *et al*. [36], they only considered single base-pair mismatches that occurred in the middle of the duplex (position 13- of 25-mers) and (iii) they assumed that binding of various RNA targets was independent and noncompetitive.

As of writing this article, there are more than 19 published papers dealing with designing oligonucleotide probes for DNA arrays studies [38–56]; more than half ($n = 13$) of these articles are published in bioinformatic journals, and few ($n = 6$ of 19) actually verified the quality of the design using experimental data.

CONUNDRUM 1: CURRENT PROBE DESIGN PARAMETERS ARE INCORRECT

In a recent study [57], we have evaluated the relationship between duplex signal intensity values and the overall Gibbs free energy of binding $\left(\Delta G_{Ob}^0\right)$ for ribosomal rRNA targets used for species identification. We studied all four free energy binding terms (i.e., ΔG_b^0, ΔG_p^0, ΔG_t^0 and ΔG_d^0) shown in Figure 2, as well as the effects of secondary structure calculated from aligned (constrained) versus not aligned (free form) sequence on ΔG_{Ob}^0 and ΔG_t^0. In our study, we took care to minimize the occurrence of nonspecific hybridizations by using known *in vitro* transcribed targets. Further, we have examined the effects of secondary structure for both probes and targets, since structure is known to affect target capture and detection of rRNA [58, 59]. Finally, only one RNA target was hybridized to one array while the other studies (see previous section) examined the binding of multiple RNA targets to one array. Multiple targets might confound interpretation of the array data since they all contribute to the extent of nonspecific hybridizations.

Very poor (or no) correlation and low (or non-significant) R^2-values were found for the linear relationships between expected and actual signal intensities, even when constrained and free-form secondary structures of the target RNA were considered. Principal component and artificial neural network analyses (with cross-validation [60]) were also not able to find any consistent patterns in the data. These findings suggest that software programs that are commonly used to calculate Gibbs free energy terms in order to predict signal intensity values, are not very useful for selecting probes for oligonucleotide arrays, and that other factors such as surface density of the probes [31] and/or brush effects [32] might have greater effects on signal intensity values than previously anticipated.

Comparable studies with mRNA targets were conducted by Mei *et al.* [26] and Naef and Magnasco [34] and both found also that current models were inadequate. As mentioned above, *ad hoc* models were proposed to remedy the situation (Zhang *et al.* [35], Held *et al.* [36] and Wu and Irizarry [37]). These studies improved predictions of the published Affymetrix datasets, however it remains unclear if their findings are applicable to other Affymetrix microarrays and other platforms. In truth, none of the models satisfactorily predicted the signal intensities of probes on oligonucleotide microarrays since there were still significant disagreements between actual and predicted values.

In summary, there is little evidence supporting the notion that the known thermodynamic parameters accurately predict signal intensity values of duplexes on oligonucleotide DNA arrays. As a consequence, current thermodynamic criteria [30, 61, 62] are of highly questionable use for designing oligonucleotide probes.

CONUNDRUM 2: MISMATCH PROBES ARE NOT ADEQUATE CONTROLS FOR NONSPECIFIC HYBRIDIZATION

The most popular oligonucleotide gene chip is offered by Affymetrix, which uses 25-mer probe sets that are specific to known target genes. Each probe set usually consists of 10 to 20 perfect match (PM) probes that are complementary to different positions along a gene sequence. In their earlier chip versions, Affymetrix designed for each PM probe corresponding mismatch (MM) probes that were similar to the PM probe, except that it had one mismatched nucleotide located in the centre of the probe. The rationale for including MM probes in a probe set is that their intensities were thought to account for nonspecific signals that affect both PM and MM probes in the same way.

Published studies conducted in solution have shown that a single base-mismatch in oligonucleotide probes can stabilize or destabilize a duplex depending on the identity of the mismatch, its position in the helix, and its neighbouring base pairs [63]. In our study, we have examined the effects of MM type (i.e. A, C, G or T) and type of neighbouring nucleotide on intensity values on DNA arrays and found that both factors had significant effects on the normalized intensity values and that there are interactions among combinations of the two factors [57]. The importance of these findings is that it affirms the notion that all factors (individually and combinations) affect the intensity of signals. Still, even taking these factors into account, it is not yet possible to predict the hybridization characteristics of a given MM probe.

It is important to note that labelling of the target can also affect observed signal intensity values. For example, fluorescent labels have been shown to interfere with probe-target binding [34]. In cases

where there is not enough label in target sequences, the targets that bind a probe but not shine, whereas in the case when there is too much label incorporated in the targets, they will not shine because they will not bind to the probes.

In summary, the type of mismatch, composition of neighbouring bases and fluorescent labels can affect the stability of duplexes on DNA arrays—but *in different ways* from what has been inferred from experiments conducted in solution. Although these findings provide a first start to understand nucleic acid hybridizations on oligonucleotide arrays, they do not yet provide a comprehensive framework that can be used for the design of oligonucleotide probes. Clearly, more empirical research is needed into this direction.

CONUNDRUM 3: WASHING OFF SPECIFIC TARGETS OVER THE NONSPECIFIC ONES

Hybridization occurs not only between specific (perfect match) probe-target pairs but also between nonspecific pairs containing mismatches. Therefore, the observed signal intensity from a single array spot might represent a combination of perfect match and nonspecific targets hybridized to the same probe [64]. This situation seriously compromises the quality of data generated from array experiments, affecting target identification and quantification in complex mixtures. The most commonly used method thought to eliminate or minimize nonspecific hybridization is to perform a stringent wash, i.e. rinse the array with low-salt buffer at constant temperature (isothermal wash). It is widely believed that once the isothermal wash has been performed, nonspecific duplexes are washed away and the observed signal is the 'true' signal of the specific (perfect match) duplexes.

The washing step is a nonequilibrium process because the dissociated targets are washed away. The quality of the washing ultimately affects downstream analysis and microarray interpretation. Several studies [65–68] have suggested that scanning through a range of stringencies in real time would improve discrimination of specific and nonspecific targets. These studies were driven by observation rather than physicochemical interpretation of the results. Subsequent studies revealed image acquisition problems, i.e. experimental results that were affected by multiple overlapping processes [69–71]

(which made modelling not possible), and basic assumptions not to be supported by either theory or experiments [8, 9].

As defined by the Arrhenius equation and first-order kinetics [9, 72], dissociation of duplexes on an array is determined by the activation energy and pre-exponential coefficient. Dissociation experiments that examined the activation energies and pre-exponential coefficients of specific and nonspecific duplex pairs found no significant difference between those parameters [9]. Further analysis revealed that in ~20% of cases, nonspecific duplexes dissociated slower than specific duplexes. These findings indicate that for such duplexes, one would observe non-specific signal rather than specific signal after performing the wash. These results suggest that the premise that nonspecific duplexes dissociate faster than specific duplexes is not supported. This premise is widely believed in the biomedical literature presumably due to the work of Ikuta *et al.* [72], whose study was based on only two specific and six nonspecific duplex pairs.

In summary, since the observed signal intensity on an array spot (i.e. probe) is a composite of all specific and nonspecific targets bound to that probe, one should not expect that stringent washing protocols is always to 'clean up' the signal. Second, although 80% of the probes are correctly washed away, the remaining 20% do not have any sequence-specific characteristics that can be used to recognize them as being deviant.

NONTARGET SIGNAL PRODUCES 'GHOST GENES'

In order to assess the effect of the above-mentioned challenges on the results of a microarray experiment, we assessed the effect of hybridizing a single defined target to a commercial Zebrafish microarray (in triplicate) and processed the data using Affymetrix GCOS software (Pozhitkov *et al.*, manuscript in preparation). Intriguingly, the statistical methods implemented in the GCOS software not only identified the specific gene, but incorrectly called multiple nontarget genes with high confidence. Such genes may be called 'ghost genes'. One potential source for the incorrect calls was found to be the threshold settings. Although adjusting these settings had a significant effect on the number of false-positive calls, we were not able to fully discriminate some of these nontarget signals from the specific

gene regardless of the statistical test used. Evidently, this would not be apparent in an experiment where complex RNA target mixtures are used. Although these observations are currently based on a single experiment only, the results are in line with the problems outlined above. It is anticipated that the noise in a microarray experiment to be at least partially ascribed to signals from 'ghost genes'.

CONCLUSIONS

It seems that there is currently no way to design an oligonucleotide microarray such that the probes have fully predictable hybridization characteristics. One solution to the problem may be to use multiple probes per target, in the hope that the average of the signal might be more reliable. Given that many results obtained from microarrays can often be validated with other methods, this 'black box' approach seems to be reasonable. However, this is scientifically not very satisfying, given that it implicates that results can also be wrong. The problem is even worse when one considers the use of microarrays for microbial identification. Here one has to often rely on relatively few species–specific probes, i.e. the 'black box' approach that gives a reasonable validity to mRNA experiments does not apply. Microbial profiling of biomedical samples using high-density microarrays containing thousands of probes is therefore associated with very high uncertainty. The results discussed here stress the importance of further experimental approaches for understanding the physics of oligonucleotide arrays, and how these affect the interpretation of raw intensities.

Key Points

- Current thermodynamic parameters do not accurately predict hybridization.
- Mismatch probes do not adequately account for nonspecific binding in mixed target samples.
- Sometimes a stringent wash rinses away specific signal while retaining nonspecifics.
- Current algorithms generate nonexistent entities when they process array data.

References

1. National Center for biotechnology Information. 2007. A science primer. http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html

2. DeSantis TZ, Stone CE, Murray SR, *et al*. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. *FEMS Microbiol Lett* 2005;**245**: 271–278.

3. Wilson WJ, Strout CL, DeSantis TZ, *et al*. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probe* 2002;**16**: 119–127.

4. Pozhitkov A, Bailey KD, Noble PA. Development of a statistically robust quantification method for microorganisms in mixtures using oligonucleotide microarrays. *J Microbiol Methods* 2007;**70**:292–300.

5. Palmer C, Bik EM, Eisen MB, *et al*. Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res* 2006;**34**:e5.

6. Pozhitkov A, Noble PA, Domazet-Loso T, *et al*. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* 2006; **34**:e66.

7. Wick LM, Rouillard JM, Whittam TS, *et al*. On-chip nonequilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. *Nucleic Acids Res* 2006;**34**:e26.

8. Pozhitkov A, Stedtfeld RG, Hashsham SA, *et al*. Effects of nucleic acid concentration on nonequilibrium thermal dissociation curves. (in review).

9. Pozhitkov A, Stedtfeld RG, Hashsham SA, *et al*. Revision of the nonequilibrium dissociation and stringent washing approaches for microbial identification studies using oligonucleotide DNA arrays. *Nucleic Acids Res* 2007; **35**:e70.

10. Brenner S, Johnson M, Bridgham J, *et al*. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol* 2000;**18**:630–4.

11. Brenner S, Williams SR, Vermaas EH, *et al*. *In vitro* cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci USA* 2000;**97**:1665–70.

12. Barlaan EA, Sugimori M, Furukawa S, *et al*. Electronic microarray analysis of 16S rDNA amplicons for bacterial detection. *J Biotechnol* 2005;**115**:11–21.

13. Zimmermann K, Eiter T, Scheiflinger F. Consecutive analysis of bacterial PCR samples on a single electronic microarray. *J Microbiol Methods* 2003;**55**:471–4.

14. Barlaan EA, Furukawa S, Takeuchi K. Detection of bacteria associated with harmful algal blooms from coastal and microcosm environments using electronic microarrays. *Environ Microbiol* 2007;**9**:690–702.

15. McKendry R, Zhang JY, Arntz Y, *et al*. Multiple label-free biodetection and quantitative DNA-binding assays on a nanomechanical cantilever array. *Proc Natl Acad Sci USA* 2002;**99**:9783–8.

16. Phillips KS, Cheng Q. Recent advances in surface plasmon resonance based techniques for bioanalysis. *Anal Bioanal Chem* 2007;**387**:1831–40.

17. Draghici S, Khatri P, Eklund AC, *et al*. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* 2005;**22**:101–9.

18. Frantz S. An array of problems. *Nat Rev Drug Discov* 2005;**4**: 362–3.

19. Marshall E. Getting the noise out of gene arrays. *Science* 2004;**306**:630–1.

20. Miklos GL, Maleszka R. Microarray reality checks in the context of a complex disease. *Nat Biotechnol* 2004;**22**: 615–21.

21. Kuo WP, Liu F, Trimarchi J, *et al*. A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat Biotechnol* 2006;**24**:832–40.

22. Shi L, Reid LH, Jones WD, *et al*. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;**24**:1151–61.

23. Kane MD, Jatkoe TA, *et al*. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 2000;**28**:4552–7.

24. Binder H, Preibisch S. Specific and nonspecific hybridization of oligonucleotide probes on microarrays. *Biophys J* 2005;**89**:337–52.

25. Loy A, Schulz C, Lucker S, *et al*. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". *Appl Environ Microbiol* 2005;**71**:1373–86.

26. Mei R, Hubbell E, Bekiranov S, *et al*. Probe selection for high-density oligonucleotide arrays. *Proc Natl Acad Sci USA* 2003;**100**:11237–42.

27. Marky LA, Breslauer KJ. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 1987;**26**:1601–20.

28. Doktycz MJ, Morris MD, Dormady SJ, *et al*. Optical melting of 128 octamer DNA duplexes. *J Biol Chem* 1995; **270**:8439–45.

29. SantaLucia J, Allawi HT, Seneviratne A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 1996;**35**:3555–62.

30. Matveeva OV, Shabalina SA, Nemtsov VA, *et al*. Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Res* 2003;**31**:4211–7.

31. Peterson AW, Heaton RJ, Georgiadis RM. The effect of surface probe density on DNA hybridization. *Nucleic Acids Res* 2001;**29**:5163–8.

32. Halperin A, Buhot A, Zhulina EB. Brush effects on DNA chips: thermodynamics, kinetics, and design guidelines. *Biophys J* 2005;**89**:796–811.

33. Pozhitkov AE, Tautz D. An algorithm and program for finding sequence specific oligo-nucleotide probes for species identification. *BMC Bioinform* 2002;**3**:9.

34. Naef F, Magnasco MO. Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays. *Phys Rev E* 2003;**68**:011906.

35. Zhang L, Miles MF, Aldape KD. A model of molecular interactions on short oligonucleotide microarrays. *Nat Biotechnol* 2003;**21**:818–21.

36. Held GA, Grinstein G, Tu Y. Modeling of DNA microarray data by using physical properties of hybridization. *Proc Natl Acad Sci USA* 2003;**100**:7575–80.

37. Wu Z, Irizarry RA. Stochastic models inspired by hybridization theory for short oligonucleotide microarrays. In: *RECOMB'04, 2004*. San Diego, CA, USA.

38. Kreil DP, Russell RR, Russell S. Microarray oligonucleotide probes. *Methods Enzymol* 2006;**410**:73–98.

39. Navin N, Grubor V, Hicks J, *et al*. PROBER: oligonucleotide FISH probe design software. *Bioinformatics* 2006;**22**: 2437–8.

40. Lin FM, Huang HD, Chang YC, *et al*. Database to dynamically aid probe design for virus identification. *IEEE T Inf Technol Biomed* 2006;**10**:705–13.

41. Chou CC, Lee TT, Chen CH, *et al*. Design of microarray probes for virus identification and detection of emerging viruses at the genus level. *BMC Bioinform* 2006;**7**:232.

42. Leiske DL, Karimpour-Fard A, Hume PS, *et al*. A comparison of alternative 60-mer probe designs in an in-situ synthesized oligonucleotide microarray. *BMC Genomics* 2006;**7**:72.

43. Shin SY, Lee IH, Zhang BT. Microarray probe design using epsilon–multi-objective evolutionary algorithms with thermodynamic criteria. *Lect Notes Comp Sci* 2006;**3907**:184–95.

44. Putonti C, Chumakov S, Mitra R, *et al*. Human-blind probes and primers for dengue virus identification – austive analysis of subsequences present in the human and 83 dengue genome sequences. *FEBS J* 2006;**273**:398–408.

45. Chung WH, Rhee SK, Bae JW, *et al*. Design of long oligonucleotide probes for functional gene detection in a microbial community. *Bioinformatics* 2005;**21**:4092–100.

46. Wu TJ, Huang YH, Li LA. Optimal word sizes for dissimilarity measures and estimation of the degree of dissimilarity between DNA sequences. *Bioinformatics* 2005; **21**:4125–32.

47. Wernersson R, Nielsen HB. OligoWiz 2.0 – integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res* 2005;**33**(Suppl. S):W611–5.

48. He ZL, Wu LY, Li XY, *et al*. Empirical establishment of oligonucleotide probe design criteria. *Appl Environ Microbiol* 2005;**71**:3753–60.

49. Nordberg EK. YODA: selecting signature oligonucleotides. *Bioinformatics* 2005;**21**:1365–70.

50. Rimour S, Hill D, Militon C, *et al*. GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics* 2005;**21**:1094–103.

51. Lee IH, Kim S, Zhang YT. Multi-objective evolutionary probe design based on thermodynamic criteria for HPV detection. *Lect Notes in Artif Intel* 2004;**3157**:742–50.

52. Yoo SM, Keum KC, Yoo SY, *et al*. Development of DNA microarray for pathogen detection. *Biotechnol Bioprocess Eng* 2004;**9**:93–9.

53. Rahmann S. Rapid large-scale oligonucleotide selection for microarrays. *IEEE Computer Society Bioinformatics Conference*. 2002. pp. 54.

54. Lipson D, Webb P, Yakhini Z. Designing specific oligonucleotide probes for the entire *S. cerevisiae* transcriptome. *Lect Notes Comput Sci* 2002;**2452**:491–505.

55. Talla E, Tekaia F, Brino L, *et al*. A novel design of whole-genome microarray probes for *Saccharomyces cerevisiae* which minimizes cross-hybridization. *BMC Genomics* 2003; **4**:38.

56. Li XY, He ZL, Zhou JZ. Selection of optimal oligonucleotide probes for microarrays using multiple criteria,

global alignment and parameter estimation. *Nucleic Acids Res* 2005;**33**:6114–23.

57. Pozhitkov A, Noble PA, Domazet-Loso T, *et al*. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nuceic. Acids Res* 2006; **34**:e66.

58. Chandler DP, Newton GJ, Small JA, *et al*. Sequence versus structure for the direct detection of 16S rRNA on planar oligonucleotide microarrays. *Appl Environ Microbiol* 2003;**69**: 2950–8.

59. Mathews DH, Burkard ME, Freier SM, *et al*. Predicting oligonucleotide affinity to nucleic acid targets. *RNA-Publ RNA Soc* 1999;**5**:1458–69.

60. Noble PA, Tribou E. Neuroet: an easy-to-use artificial neural network for ecological and biological modeling. *Ecol Mod* 2007;**203**:87–98.

61. Tanaka F, Kameda A, Yamamoto M, *et al*. Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. *Nucleic Acids Res* 2005;**33**:903–11.

62. Luebke KJ, Balog RP, Garner HR. Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. *Nucleic Acids Res* 2003;**31**:750–8.

63. Kierzek R, Burkard ME, Turner DH. Thermodynamics of single mismatches in RNA duplexes. *Biochemistry* 1999;**38**: 14214–23.

64. Zhang Y, Hammer DA, Graves DJ. Competitive hybridization kinetics reveals unexpected behavior patterns. *Biophys J* 2005;**89**:2950–9.

65. Kelly JJ, Siripong S, McCormack J, *et al*. DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. *Water Res* 2005;**39**:3229–38.

66. Siripong S, Kelly JJ, Stahl DA, *et al*. Impact of prehybridization PCR amplification on microarray detection of nitrifying bacteria in wastewater treatment plant samples. *Environ Microbiol* 2006;**8**:1564–74.

67. ElFantroussi S, Urakawa H, Bernhard AE, *et al*. Direct profiling of environmental microbial populations by thermal dissociation analysis of native ribosomal RNAs hybridized to oligonucleotide microarrays. *Appl Environ Microbiol* 2003; **69**:2377–82.

68. Liu WT, Mirzabekov AD, Stahl DA. Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ Microbiol* 2001;**3**:619–29.

69. Pozhitkov A, Noble PA. Comment on discrimination of shifts in soil microbial communities using nonequilibrium thermal dissociation and gel pad array technology. *Environ Sci Technol* 2007;**41**:1797–8.

70. Pozhitkov A, Noble PA. High variability in melting profiles from gel pad arrays. *Environ Microbiol* 2007;**9**:1865.

71. Pozhitkov A, Chernov B, Yershov G, *et al*. Evaluation of gel-pad oligonucleotide microarray technology using artificial neural networks. *Appl Environ Microbiol* 2005;**71**: 8663–76.

72. Ikuta S, Takagi K, Wallace RB, *et al*. Dissociation kinetics of 19 base paired oligonucleotide-DNA duplexes containing different single mismatched base-pairs. *Nucleic Acids Res* 1987;**15**:797–811.