Oligonucleotide microarrays: widely applied—poorly understood

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

Advance Access publication date 20 July 2007

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 underlying 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 microarray 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...
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,
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 processes (i) to (iv), 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 ($\Delta G^0_{\text{Ob}}$) for ribosomal rRNA targets used for species identification. We studied all four free energy binding terms (i.e., $\Delta G^0_{\text{Ob}}, \Delta G^0_{\text{Pf}}, \Delta G^0_t$ and $\Delta G^0_{\text{Ad}}$) shown in Figure 2, as well as the effects of secondary structure calculated from aligned (constrained) versus not aligned (free form) sequence on $\Delta G^0_{\text{Ob}}$ and $\Delta G^0_t$.

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 nonspecific 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 implies 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.

Funding
Supported in part by a grant from the US National Oceanic and Atmospheric Administration (NAO4NOS4260203) to A.E.P., National Institutes of Health grant 1U01DE014955-01 from NIH/NIDCR, R-82945801 from EPA-CEER-GOM, the Royalty Research Fund and the Provost Bridge Funding Program to P.A.N.

References


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


