Microarray analysis using disiloxyl 70mer oligonucleotides

Marcus Gry Björklund1, Christian Natanaelsson1, Amelie Eriksson Karlström1, Yong Hao2 and Joakim Lundeberg1,*

1School of Biotechnology, Department of Gene Technology, KTH, Royal Institute of Technology, AlbaNova University Center, SE-10691 Stockholm and 2Oligovation, Dag Hammarskjölds väg 32 B, Uppsala Science Park, 751 83 Uppsala, Sweden

Received August 28, 2007; Revised November 18, 2007; Accepted December 10, 2007

ABSTRACT

DNA microarray technology has evolved dramatically in recent years, and is now a common tool in researchers' portfolios. The scope of the technique has expanded from small-scale studies to extensive studies such as classification of disease states. Technical knowledge regarding solid phase microarrays has also increased, and the results acquired today are more reliable than those obtained just a few years ago. Nevertheless, there are various aspects of microarray analysis that could be improved. In this article we show that the proportions of full-length probes used significantly affects the results of global analyses of transcriptomes. In particular, measurements of transcripts in low abundance are more sensitive to truncated probes, which generally increase the degree of cross hybridization and loss of specific signals. In order to improve microarray analysis, we here introduce a disiloxyl purification step, which ensures that all the probes on the microarray are at full length. We demonstrate that when the features on microarrays consist of full-length probes the signal intensity is significantly increased. The overall increase in intensity enables the hybridization stringency to be increased, and thus enhance the robustness of the results.

INTRODUCTION

DNA microarrays prepared by printing oligonucleotides synthesized in vitro onto a solid surface are widely used research tools for studying gene expression profiles in biological systems (1). Experimental microarray results containing tens of thousands of data points are considered informative, but are not trivial to analyse due to technical and biological variations. Various statistical models have been introduced to optimize microarray fabrication, sample preparation, hybridization and image acquisition, to assure the quality and consistency of microarray products (2–4). Several aspects such as sequence design, probe length and attachment of probes to the solid surface have been explored in order to increase the strength of microarray signals and to improve the reproducibility of microarray data (4). Nonetheless, few studies have dealt with the effect of probe purity on microarray signals.

There are several possible causes and types of impurities in oligonucleotide microarray probes manufactured by in vitro synthesis. First, when synthesis is based on phosphoramidite chemistry the coupling efficiency of each monomer is typically around 99%, so only 50% of the probes produced have full length sequences (5,6). Secondly, capping is not 100% efficient, and n-x sequences are produced on uncapped sites (7). Finally, purines are unstable in the acidic step of each synthesis cycle, which causes cleavage of the oligonucleotides at some of these sites (8–10). The truncated sequences, n-x species and cleaved apurinic fragments are the major impurities in the synthesis products (5,6). Since the truncated sequences do not contain the last 5'-hydrophobic protecting group (the trityl) they can be removed by reversed phase chromatography (11–14). However, the n-x species and some of the depurinated fragments contain the last 5'-end trityl and are therefore not removed by the use of simple reversed phase cartridges (RPC) for purification. To remove these tritylated fragments, highly efficient HPLC gradient systems are required (11–13). However, since such large numbers of oligonucleotides are needed for microarray printing, purification of oligonucleotide probes by HPLC is constrained by throughput, speed and cost considerations. On the other hand, standard RPC-based purification can be run at high throughput cost effectively, but only a fraction of the shorter oligonucleotide
impurities can be removed, therefore the method is not efficient for longer oligonucleotides. In addition, 5'-amino functional groups are often used to covalently attach oligonucleotide probes to the solid surface in microarray printing (15,16). Many of these functional groups contain monomethoxytrityl (MMTr) as the 5' protecting group, which is more stable than the conventional dimethoxytrityl (DMTr) (17). However, in commercial RPC methods detritylation is performed in an aqueous trifluoroacetic acid solution (TFA) (11,14), which does not efficiently detritylate MMTr groups. Thus, no current method offers a high throughput process for manufacturing 5' aminated microarray oligonucleotide probe sets with high purity. Commercial oligonucleotide sets for microarrays are usually supplied as desalted, crude or in a few cases RPC-purified synthesis products.

Kwiatkowski et al. (18) suggested the use of a disiloxyl linker to connect the oligonucleotides to the synthesis support. Such a linker is stable both during the synthesis cycles and in basic conditions, so its use enables the cleavage of apurinic sites before the oligonucleotide is cleaved from the support. The tritylated 5' parts of the apurinic fragments can thus be easily removed by washing, while the full-length oligonucleotide, together with the 3' non-tritylated fragments, remains attached to the synthesis support. The method described in this study was designed to further develop this approach in combination with the use of a reversed phase cartridge system, and thus provide a high throughput parallel manufacturing process to synthesize and purify oligonucleotide sets for microarrays. Oligonucleotides manufactured by the method described in this study were compared in microarray experiments with oligonucleotides obtained using a conventional oligonucleotide synthesis procedure.

MATERIALS AND METHODS

The oligonucleotides

Two sets of 24 70mer oligonucleotides were used in this microarray study: one synthesized and purified using the method described subsequently, and one purchased from Operon Biotechnologies (Huntsville, AL, USA). The oligonucleotides included in these test sets had identical sequences (shown in Table 1), corresponding to sense strands of mRNA transcripts spanning a wide range of expression levels in the human brain, according to a previous microarray experiment using spotted cDNA arrays (data not shown). The 24 oligonucleotides in each set can be broadly divided into three categories, corresponding to weakly, moderately and strongly expressed genes, with a slight bias in numbers of weakly expressed genes (Table 1). Each oligonucleotide contained a 5'-terminus amino link with a C6 spacer to facilitate covalent immobilization to the pre-activated slides.

Ten probes of varying lengths (25mer to 70mer) purchased from Operon Biotechnologies corresponding to part of the human hepatitis C virus NS5A gene were synthesized for use as signal controls in the microarray system. These oligonucleotides also contained a 5'-terminus amino link with a C6 spacer to facilitate covalent immobilization. These oligonucleotide probe sequences were spotted on to the array slides as: 25-, 30-, 35-, 40-, 45-, 50-, 55-, 60-, 65- and 70-mers, respectively.

Synthesis of disiloxyl oligonucleotides

All of the oligonucleotides in the set made in-house were synthesized with CE-phosphoramidites (Glen Research, Sterling, VA, USA) in 0.2 μmol quantities using a model 394 DNA synthesizer, leaving the dimethoxytrityl (DMTr-O-) or monomethoxytrityl (MMTr-NH-) at the 5'-end, and standard chemicals supplied by the synthesizer’s manufacturer (Applied Biosystems Inc., Foster City, CA, USA). The oligonucleotides were aminated at the 5'-end with an amino-C6 phosphoramidite (Glen Research), and synthesized using either Qt DNA Oligonucleotide 1000 A controlled pore glass (CPG) Synthesis Support (Oligovation AB, Sweden), containing a disiloxyl linker connecting the first nucleotide, or (for the oligonucleotides purified using the purchased RPCs, see subsequently) 1000 A CPG with standard ester bonds (Applied Biosystems).

Depurinated fragments were cleaved by incubation in ammonium hydroxide (Merck KGaA, Darmstadt, Germany) for 30 min at 25°C. The CPG was then washed with water to remove the residual fragments before drying with acetonitrile. The oligonucleotides were cleaved from the disiloxyl synthesis support by incubation for 30 min at 25°C with 1-methyl-2-pyrrolidone (NMP, Sigma-Aldrich Inc. St Louis, MO), triethylamine (TEA, Sigma Aldrich Inc.), triethylamine-triethyfluoride (TEA-3HF, Sigma-Aldrich Inc.) mixed 3:2:2:1 (v:v:v) to ensure a basic pH (19, 20). All oligonucleotides and the collected apurinic supernatants were deprotected in ammonium hydroxide for 8–16 h at 55°C.

Purification of disiloxyl oligonucleotides

The disiloxyl oligonucleotides were purified using cartridges consisting of IST Isolute® SPE columns (inner diameter 5 mm, Argonaut Technologies, Foster City, CA, USA) packed (in a slurry with acetonitrile) with 150 mg Qt DNA Oligonucleotide Purification Support (Oligovation AB, Uppsala, Sweden) and an ASPECTM XL4 SPE robotic liquid handler (Gilson Inc., Middleton, WI, USA) connected to pumps providing both positive and negative pressures. The binding buffer contained 10% NaCl and 5% acetonitrile. The non-tritylated fragments were eluted with 5% followed by 14% acetonitrile in 0.1 M triethylammonium acetate (TEAAc). The cartridges were dried by washing them with acetonitrile then passing a stream of nitrogen gas through them for 5–10 min. After drying the oligonucleotides were detritylated using 2% trifluoroacetic acid (TFA) in dichloromethane (DCM), washed with acetonitrile followed by 0.1 M TEAAc then water and eluted, salt-free, using 70% acetonitrile.

HPLC analysis

The purity of both the disiloxyl-purified and the conventional oligonucleotides was analysed by reversed phase HPLC using a 4.6 × 150 mm Zorbax 300SB C18 column.
with a particle size of 3.5 μm (Agilent Technologies). Injected samples were eluted in a gradient of 0–20% B in 20 min, where solvent A: 95% 100 mM TEAAc (pH 7.0); 5% CH₃CN, and solvent B: 10% 100 mM TEAAc (pH 7.0); 90% CH₃CN, at a flow rate of 1 ml min⁻¹. The oligonucleotides were detected at λ = 260 nm. The HPLC data indicated that the purity is enhanced for the disiloxyl-purified probes compared to the conventional probes (data not shown).

### Array manufacture

The two sets of 24 oligonucleotide probes (Operon and disiloxyl probes), and 10 viral control probes were printed in triplicates at a concentration of 20 μM, 0.06% sarcosyl, pH 8.5, and 150 mM sodium phosphate onto Codelink™ slides (GE Healthcare: Chalfont St Giles, UK) using a QArray spotter (Genetix, Hampshire, UK) with SMP 2.5 pins (Telechem, Sunnyvale, CA, USA). After printing, the arrays were incubated overnight in a humid chamber followed by post coupling as recommended by the manufacturer. Briefly, the slides were incubated at 50°C for 30 min in a blocking solution (50 mM ethanalamine, 0.1 M Tris, pH 9 and 0.1% SDS), rinsed twice in dH₂O and washed with pre-warmed (50°C) 4× SSC, 0.1% SDS for 30 min on a shaker (20× SSC contains 175.3 g/l NaCl, 88.2 g/l sodiumcitrate, pH 7; and 10% SDS contains 100 g/l SDS, pH 7.2). The slides were then rinsed in dH₂O and dried by centrifugation for 3 min at 800 rpm.

### Table 1. Sequences of the 24 oligonucleotides

| Oligo ID | Ensembl gene | GeneBank | Sequence |
|----------|--------------|----------|----------|
| H200017531 | ENSG00000145996 | NM_017774 | TGGAAAGCCGCTTGGGAGAGAGCAACAGATGGTTT(G)CAGACAGTAGAAAGATGAAAGACCTCA |
| H200008090 | ENSG00000138337 | BC06427 | GAGGCGGGCCCAAAATAAGTAAGTGAGTAAACTTAGTAAGAAGATGTGGTATGTCCTTCACCAAA |
| H200004383 | ENSG00000100122 | NM_018887 | ACCTCAAGTATTTGCCATCAGTTATTGCTGGAGCTGCCTTTCATTTAGCACTCTACA |
| H200007139 | ENSG00000145386 | NM_001237 | ACCTCAAGTATTTGCCATCAGTTATTGCTGGAGCTGCCTTTCATTTAGCACTCTACA |
| H200020609 | ENSG00000166253 | AK058066 | CACCGTGAAATAGTACAGGTCCTGACAGCTGGGTCCCTGATGGGCCTGGGTGAC |
| H30000600 | ENSG00000181585 | BC030279 | GACCAGATGATATTAAAGAAGCAAGGGAGGAGCTGCTGATTTGGTGGCCTTCAGG |
| H200012632 | ENSG00000156977 | BC006427 | GAGGCCGCCCAGGAAATAGAAGTTGAGCTAGAACTTAGTAAAGAGATGGTTAGTC |
| H200006166 | ENSG00000115204 | NM_02437 | CACCGTGAAATAGTACAGGTCCTGACAGCTGGGTCCCTGATGGGCCTGGGTGAC |
| H30000524 | ENSG00000174749 | AK096689 | GAGAAGCTACAGAACCACACTGTGAGGCGAGATCTACAACAAAAATTTT |
| H300017279 | ENSG00000119979 | AF168713 | TTTCTCCCAACAAAA |
| H300004627 | ENSG00000181563 | BC030279 | TTTCTCCCAACAAAA |
| H200012342 | #N/A | BF026507 | #N/A |
| H300000036 | ENSG00000172335 | 0 | AAAATAACATTGAGACTCCCCAATCTGACACTGGCACCTAATGGAAG |
| H200005324 | ENSG00000174749 | AK096689 | GAAGAAGCTACAGAACCACACTGTGAGGCGAGATCTACAACAAAAATTTT |
| H300012632 | #N/A | BF026507 | #N/A |
| H300000600 | ENSG00000181585 | BC030279 | GACCAGATGATATTAAAGAAGCAAGGGAGGAGCTGCTGATTTGGTGGCCTTCAGG |
| H300006166 | ENSG00000115204 | NM_02437 | CACCGTGAAATAGTACAGGTCCTGACAGCTGGGTCCCTGATGGGCCTGGGTGAC |
| H300005638 | ENSG00000131959 | AF130106 | GAGGCTATGAGAAATAGTACAGGTCCTGACAGCTGGGTCCCTGATGGGCCTGGGTGAC |
| H300009192 | ENSG00000180075 | AB062480 | TGGAGATGATATTAAAGAAGCAAGGGAGGAGCTGCTGATTTGGTGGCCTTCAGG |
| H300015513 | ENSG00000169101 | BC030436 | GAGAAGCTGCTGACAGCTGGGTCCCTGATGGGCCTGGGTGAC |
| H300009506 | ENSG00000180390 | BC01809 | TGGAGATGATATTAAAGAAGCAAGGGAGGAGCTGCTGATTTGGTGGCCTTCAGG |
| H300007192 | ENSG00000169822 | BC01809 | GGGGTTGTTTCCATGATTTGCGAGCTGCTGACAGCTGGATGATCAGC |
| H300007187 | ENSG00000170189 | BC01809 | GGGGTTGTTTCCATGATTTGCGAGCTGCTGACAGCTGGATGATCAGC |
| H300000140 | ENSG00000170135 | 0 | AAAATAACATTGAGACTCCCCAATCTGACACTGGCACCTAATGGAAG |

Where applicable, the Genebank and/or Ensembl IDs are shown. Since Operon switched from basing the designs of their oligonucleotides from Genebank to Ensembl sequence data, IDs are missing for some of the sequences, depending on when the oligos were designed. The two missing sequences are due to their absence from the Operon Microarray Database (OMAD).
Hybridization sample preparation

Total Universal Human Reference and Human Brain Library RNAs were purchased from Stratagene (La Jolla, CA, USA) and Applied Biosystems, respectively. The RNA was either amplified using a Ribonuclease Amplification Kit (Arcturus, Sunnyvale, CA, USA), or not amplified. Approximately 5 μg of each of the amplified RNA, and 20 μg of the non-amplified RNA samples was primed with 5 μg random hexamers (Invitrogen, Carlsbad, CA, USA). The volume was adjusted to 18.4 μl using DEPC-treated water; the RNA was denatured at 70°C for 10 min, and then renatured on ice for 5 min. Reverse-transcription reaction mixture (Invitrogen) and 400 units of Superscript III RT-polymerase were added to yield a final volume of 30 μl containing 1× first-strand buffer (Invitrogen), 0.01 mM DDT (Invitrogen) and 0.5 mM dNTPs (Sigma-Aldrich, St. Louis MO, USA). The ratio of aminoallyl-modified dUTP to dTTP was 4:1 in the dNTP mixture. The samples were incubated at 25°C for 10 min followed by 46°C for 2 h. The CDNA synthesis was halted by adding 3 μl 0.2 M EDTA (pH 8.0).

Template RNA was removed by adding 4.5 μl 1 M NaOH. The samples were incubated at 70°C for 15 min, and then chilled to room temperature, neutralized with 4.5 μl 1 M HCl and purified using the MinElute Reaction Cleanup system (QiaGen) with the provided wash and elution buffers replaced by 80% ethanol and 100 mM NaHCO₃, pH 9.0, respectively. The elution step from the column was repeated, generating a volume of 20 μl. This was mixed with a tenth of the contents of a monofunctional NHS-ester Cy3 or Cy5 dye tube (GE Healthcare), which had been dissolved in DMSO and moulded in an A custom-made Plexiglas rack, with drilled holes, was used to press the silicone firmly to the slide and keep it in place during the reactions.

The microarray was divided into nine sub-arrays for the amplified RNA and two sub-arrays for the non-amplified RNA, which were separated during hybridization by a reusable silicone mask (Elastosile® RT 625 A/B, Wacker-Chemie GmbH, Munich, Germany), moulded in an inverted 96-well plate (Falcon flexible plate, 96 well, U-bottom without lid; Becton Dickinson Labware, Franklin Lakes, NJ, USA) and excised to fit the slide. A custom-made Plexiglas rack, with drilled holes, was used to press the silicone firmly to the slide and keep it in place during the reactions.

The microarray slides were pre-hybridized for 30 min at 42°C or 50°C in a pre-hybridization solution consisting of 5× SSC, 0.1% SDS (Sigma-Aldrich) and 1% BSA (Sigma-Aldrich) to avoid unspecific hybridization to the glass surface. The slides were subsequently washed in water and isopropanol (Sigma-Aldrich) and dried using a slide centrifuge. The labelled sample (Cy3) and reference (Cy5) were pooled and denatured (3 min at 95°C) in a hybridization mixture containing 25% formamide (Sigma-Aldrich), 5× SSC and 0.1% SDS. For the non-amplified RNA the samples and the reference were also labelled in the reverse order. The viral control oligonucleotide was added to the hybridization mixture which contained the amplified RNA to a final concentration of 50 PM. The mixture was applied under a lifter-slip cover slip (Erie Scientific Company, Portsmouth, NH, USA) placed on top of the printed array and hybridized for 18–24 h at 42°C or 50°C (Results) in a water bath. Following hybridization the slides were washed with increasing stringency using 2× SSC and 0.1% SDS at 42°C, followed by 0.1× SSC and 0.1% SDS at room temperature and finally by five repeated washes with 0.1× SSC at room temperature.

For the amplified RNA, two sets of experiments were carried out: (i) a self/self hybridization experiment using Stratagene Universal Reference RNA, and (ii) a Stratagene Universal Reference RNA versus Human Brain Library (Stratagene and Ambion, respectively) hybridization to assess the performance of the two sets of probes in a differential expression analysis. For both experiments, two of the nine available sub-arrays were hybridized with 2.5, 1.25, 0.625 and 0.3125 μg of RNA, respectively and one sub-array with 0.156 μg of RNA. Two arrays were used in the self/self hybridization experiment, one hybridized at 42°C and one at 50°C. A single array was used in the differential expression experiment. For the non-amplified RNA, a differential expression analysis was carried out using two slides to facilitate a dye swap analysis, two of the nine sub-arrays were used in these experiments and the slides were hybridized at 50°C. Following hybridization the arrays were scanned at 10 μm resolution using an Agilent G2565BA scanner (Agilent Technologies, Santa Clara, CA, USA), with the photomultiplier set to 100% for each laser. The acquired images were analysed using the circular gridding algorithm in GenePix Pro 5.1 (Axon Instruments, Barrington, NJ, USA), and the resulting data were imported into the R environment for statistical computing and visualization (19). The intensities were extracted from the median foreground intensity for the 532 nm and 635 nm channels. The features were filtered based on the data from GenePix manual inspection of the slides. Spots that were removed were either not found by the image software or were marked as bad spots due to dust particles and spots that had contact with adjacent spots on the array.

The intensities of signals from features within each array were normalized by print-tip lowess normalization (19). The log2 values of the products of the two normalized intensities (abbreviated A value (½(log₂(F532° F635))) for all features were then calculated and compared. Differences in the A-value distributions from the standard and disiloxyl-purified probe sets, for each of the three slides, were then evaluated by t-tests. The intensities were calculated based on the different sequences and were compared between the two probe types with a Wilcoxon rank test. In addition, linear regression based on the intensities of the viral control 25mers at the different concentrations was used to calculate an intensity cut-off. The 25mer was used since it is probably the least
discriminator of hybridizing to non-viral RNA:s. The coefficients of the linear regression were used to extrapolate the regression line to $x = 0$, which was then used as the noise threshold level.

RESULTS
Experimental design
The aim of this study was to investigate a novel approach for synthesizing oligonucleotides for array-based gene expression with 70 mer probes. Twenty-four different oligonucleotides were manufactured using two different procedures and spotted at equal concentration onto microarrays using an array-on-array design. Each microarray consisted of nine sub-arrays to minimize between-replicate differences between hybridization conditions (Figure 1). In addition, each subarray consisted of a set of triplicate probes to minimize between-replicate differences between hybridization conditions (Figure 1). In addition, each subarray consisted of a set of triplicate probes, each consisting of the same 24 oligonucleotide sequences, in order to compare the performance of the disiloxyl-purified probes within a subarray to that of a conventionally prepared set of 70 mer probes purified by a desalting protocol used by the provider. Furthermore, a control set of probes with varying lengths (from 25 mer to 70 mer with 5-nucleotide increments) of viral origin was co-spotted and used in the analysis to quantify unspecific hybridization. Two sets of experiments were performed using the amplified RNA: a self/self hybridization and a differential expression analysis. The RNA was hybridized at a range of concentrations to the nine sub-arrays and the performance of the two sets of probes was evaluated at two hybridization temperatures. The viral controls were spiked in all RNA mixtures at a constant concentration as described in Material and Methods. For the non-amplified RNA a differential expression experiment was performed, using two slides using two sub-arrays. The arrays were scanned after hybridization, and the intensities of the signals from their features were calculated after an initial filtering, as outlined in the Material and Methods section.

Self-self hybridization
First the hybridization accuracy and efficiency of the two sets of probes was investigated using the self/self hybridization data. For this purpose, the raw intensity data were transformed to log2-values using the product from the two channels (Cy3 and Cy5). The intensities were calculated for both the 42°C and 50°C hybridizations. Initially, the differences between conventional probes and disiloxyl purified probes were investigated using ratio versus total intensity plots (MA plots). The MA plots are depicted in Figure 2, and the total intensity (A-value) distributions are shown in Figure 3A. Differences in A-values can be seen in the MA-plots so a $t$-test was carried out to compare the intensity distributions. The resulting $P$-values for the 42°C and 50°C hybridizations were 1.053e-06 and 2.2e-16, respectively, indicating that there were significant differences between the intensities of the signals obtained with both sets of probes at the two different temperatures. The mean A-values (total intensity) for the two different probe types were 9.55 for the disiloxyl-purified probes and 8.93 for the conventional probes at 42°C. At 50°C the mean intensities were 9.41 and 8.40 for the disiloxyl-purified probes and the conventional probes, respectively, showing that the shift in intensities was larger for the conventional probes.

To elucidate whether the shifts in intensities were biased by the responses of one or a few specific sequences, shifts in the intensities of the signals from each of the 24 sequences in both sets of probes were calculated and compared, using the A-values obtained. The intensities are illustrated in the box plot shown in Figure 3B, in which the intensities for each sequence in both sets are...
of probes are plotted. Each box encompasses values for a specific sequence at all concentrations (0.156, 0.3125, 0.625, 1.25 and 2.5 µg), so each box contains \(9 \times 3 = 27\) values. The plots show that the difference is consistent across every sequence (except one sequence for which the probes were saturated with the hybridized target). To assess the significance of the differences a Wilcoxon rank test was carried out, which yielded a \(P\)-value of <2.2e-16 for both temperatures, corroborating the conclusion that there was a significant difference for every sequence. The results were also consistent for the two tested hybridization temperatures (data not shown). The non-amplified RNA data shows the same pattern as the amplified data shown in a box plot in Figure 3D. In this case the boxes contains all data from the two slides used for the non-amplified RNA experiment. The difference in signal intensity between the two probe types is significant using a Wilcoxon rank test (\(P\)-value = 2.0e -5).

**Differential gene expression**

Differential gene expression profiles were obtained by co-hybridization of amplified Stratagene Universal RNA Reference RNA and Brain RNA Library, which were labelled with Cy5 and Cy3 dyes, respectively, using both sets of probes. For the non-amplified RNA the samples was labelled with Cy5 and CY3 and hybridized to one sub-array, and vice versa for the second sub-array using two slides. The data derived from the amplified RNA experiment were normalized and evaluated using a ratio versus total intensity plot (Figure 4A), in which the probespecific differences are retained. A \(t\)-test between the distributions showed that there was a significant difference between the strength of the signals yielded by the two sets of probes (\(P\)-value <2.2e-16), with mean A-values of 7.87 and 8.90 for the conventional and disiloxyl-purified probes, respectively. This corresponds to a 2-fold difference in signal intensity. The data derived from the non-amplified RNA experiment was also normalized and the mean A-values were 8.93 for the conventional probes and 9.33 for the disiloxyl-purified probes. The difference in mean intensities is significant (\(P\)-value 6.695e-06), and the non-amplified RNA data is concordant with the amplified RNA (Figure 3C). To estimate the amount of cross hybridization on the array we used the amplified RNA data and linearly regressed the intensities from the viral control 25mer sequence (\(P\)-values <2.2e-16 for the

---

**Figure 2.** MA plots for all the features in the amplified RNA data set. M equals the fold-change or ratio of the two intensity channels, and the A-value is the product of the two intensities. (A) Plots derived from the 42°C data set. In the left panel the conventional probes are highlighted in red and all other probes in black. In the right panel the disiloxyl purified probes are shown in green and all other probes in black. (B) Corresponding plots derived from the data set for the array hybridized at 50°C.
intercept and 1.07e-14 for the slope, respectively), which was used as a negative control. Increasingly strong signals were obtained from the control features with increasing amounts of hybridized RNA, although the concentration of the spiked NS5A sequence was the same in all experiments. We then used the intersection at $X=0$ obtained from linear regression of the intensity data for the NS5A 25mer as a cut-off to estimate the degree of cross hybridization (Figure 4B). The 25mer was chosen based on the expectation that the 25mer is the least discriminate oligomer on the array, and therefore an appropriate measurement of unspecific hybridization. The cut-off was set to $A=7.197$ for the hybridization at 50°C. The A-values of 330 features were below this cut-off for the conventional probes, but only five for the disiloxyl-purified probes. The disiloxyl-purified probes corresponding to the sequences with A-values lower than the cut-off for the conventionally purified probes are shown as blue dots in Figure 4A. In addition, we could not observe any dye bias affects in the differential gene expression analysis using the non-amplified RNA (data not shown).

For the amplified material, two of the 24 sequences displayed high fold-change values (M-values): CCNA2 (up-regulated in the Stratagene Universal RNA reference, mean M-value = 1.86) and KIAA1279 (down-regulated in the Stratagene Universal RNA reference, mean M-value = −1.24). The A-values for the two sequences spanned a substantial range, depending on the amount of labelled RNA (as observed for the NS5A data). A plot of the A-values versus the concentration of the hybridized material is shown in Figure 3C for the two genes and each...
of the probe types. The intensities for the KIAA1279-sequence at the 0.156 µg RNA concentration were below the value chosen as the intensity cut-off.

**DISCUSSION**

In this article we demonstrate the importance of probe purity when using microarrays. We show that the amount of full-length sequences within features affect the analysis of differences in gene expression between two samples, due to the effects of the unknown amount of cross-hybridization and low signal intensities. Our data support the hypothesis that the intensity is significantly higher for probes purified using the disiloxyl-purification method than for conventionally purified probes at hybridization temperatures of both 42 and 50°C, implying that they hybridize more specifically since all of the features in the array are full-length. The self-self hybridization data show a tail-like structure at the low intensity region of the ratio versus total intensity plots. This is a common feature of microarray data and is generally treated using various normalization techniques. However, if the print tip lowess normalization is applied and the sample space is quite low (small data set), this normalization procedure is sensitive to outliers, and it can be hard to adjust the data using a lowess normalization. The t-test of the self-self hybridization data showed that there were significant differences between the A-value distributions at both of the tested temperatures between the two compared probe sets. The difference in probe intensities was also significantly different for the non-amplified RNA experiment. The differences in total intensities (A-values) for the amplified RNA experiment at 42 and 50°C were 0.62 and 1.01, respectively. Using the amplified RNA data, sufficient features were used in this test to detect a significant difference in A-values, of ~0.24, which corresponds to an intensity product of the two channels of 1.39 (2 × 0.24), where the maximum product is 65536. Such a difference in gene expression levels is likely to be phenotypically insignificant generally. Nevertheless, there was a clear shift in the intensity distribution of the conventional probes when the hybridization temperature was increased, as shown in Figure 3A, but not for the disiloxyl-purified probes, which corroborates the hypothesis that the purity of the

---

**Figure 4.** (A) MA plot of data from the differential expression experiment using the amplified RNA data (upper panel). The values for the conventional probes are shown in red. The solid line indicates the noise cut-off based on the linear regression of the intensities of the viral control 25mer oligonucleotide. The lower panel shows the corresponding MA plot for disiloxyl-purified probes, in which the values are shown in green. The blue dots indicate the disiloxyl-purified probes with the same sequences and labelled amounts as the conventional probes with A-values that were consistently (in all three replicates) below the cut-off. (B) Intensities for the viral control 25mer oligonucleotides across the tested RNA concentrations. The line indicates fitted values from the linear regression. (C) Intensities for two sequences (for a gene that is up-regulated in the human brain, CCNA2, upper panel and one that is down-regulated, KIAA1279, lower panel) across the tested RNA concentrations in the differential expression experiment using the amplified RNA data. Green and red dots indicate data for disiloxyl- and conventionally purified probes, respectively. The solid black line indicates the noise cut-off intensities determined by the linear regression.
features is a significant contributor to the intensity of signals from a probe. Furthermore, the difference in A-values more pronounced when the amount of hybridizing RNA was increased (data not shown), and the effect was enhanced at the 50°C hybridization temperature. Interestingly, we have also shown that the intensity range of the replicates across different RNA concentrations is higher for the moderately and highly expressed genes for probes purified by both methods. This is also true for the coefficient of variation, which increases with increasing intensities (data not shown). These correlations may be due to the amplification step introducing a 3′ bias in the RNA population, which increases the variability for certain probes depending on the position of the target sequence in the transcript. Our data show that the signal strength generally increases when the oligonucleotide corresponds to a sequence towards the 3′ end of the mRNA (data not shown). The intensity range of the non-amplified RNA does not show the same pattern as the amplified RNA. For the non-amplified RNA the intensity range is more random between different probes shown in Figure 3D. The results also show that the higher overall intensity levels for the disiloxyl-purified probes was not due to increases in intensity for a few transcripts, but rather an overall increase (Figure 3B). The disiloxyl-purified probes gave stronger signals than the conventionally purified probes for all of the sequences, except the PTMA transcript, which was saturated in the scanning procedure. In the differential gene expression experiment we used a linear model to derive a cut-off A-value and to evaluate the numbers of probes that gave A-values below this cut-off. To choose the cut-off for unspecific hybridization using a linear regression model, we took the intersection where the regression line crosses the y-axis. Regression models should be extrapolated cautiously, but since the P-value was low and there was a clear linear relationship between the intensities of the signals from the viral control oligomers and their concentrations, we are confident in using extrapolation in this case. The positive slope in the regression model shown in Figure 4B indicates that there was an unspecific increase in the intensity of the viral controls as the concentration of the human RNA samples increased. The viral probes used in the illustrated experiment were 25-mers, so even though the probes were short, there seems to have been a non-trivial degree of unspecific hybridization, corroborating the conclusion that truncated probes can contribute to non-specific signals, even at rather high hybridization temperatures.

The benefits of using pure probes were also evident in the differentially gene expression experiment, in which the disiloxyl-purified probes yielded signals that were stronger than the cross-hybridization cut-off for more genes than the conventional probes, indicating that the disiloxyl probes are more suitable for detecting low-abundance transcripts. Finally, the consequences of differences in probe purity for identifying differentially expressed genes are obvious. We show that microarray analysis could be enhanced by using full-length probes, due to the associated improvements in intensity and specificity.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Research Council. We are grateful to Annelie Waldén at the KTH array facility for providing arrays. Funding to pay the Open Access publication charges for this article was provided by the Swedish research council.

Conflict of interest statement. None declared.

REFERENCES

1. Schena, M. (2003) Microarray Analysis. John Wiley & Sons, Inc. Hoboken, New Jersey.
2. Chandler, D.P. and Jarrell, A.E. (2005) Taking arrays from the lab to the field: trying to make sense of the unknown. BioTechniques, 38, 591–600.
3. Murphy, D. (2002) Gene expression studies using microarrays: principles, problems, and prospects. Advan. Physiol. Educ., 26, 256–270.
4. Graves, D.J. (1999) Powerful tools for genetic analysis come of age. Trends Biotechnol., 17, 127–134.
5. Reece, C. (2002) The chemical synthesis of oligo- and polynucleotides: a personal commentary. Tetrahedron, 58, 8893–8920.
6. Beauchage, S.L. and Caruthers, M.H. (1996) The chemical synthesis of DNA/RNA. In Hecht, S.M. (ed), Bioorganic Chemistry: Nucleic Acids (Topics in Bioorganic & Biological Chemistry), Oxford University Press, UK.
7. Tentsamani, J., Kubert, M. and Agrawal, S. (1995) Sequence identity of the n-1 product of a synthetic oligonucleotide. Nucleic Acids Res., 23, 1841–1844.
8. Septak, M. (1996) Kinetic studies on depurination and detritylation of CPG-bound intermediates during oligonucleotide synthesis. Nucleic Acids Res., 24, 3053–3058.
9. Tanaka, T. and Letsinger, R.L. (1982) Syringe method for stepwise chemical synthesis of oligonucleotides. Nucleic Acids Res., 10, 3249–3260.
10. Froehler, B.C. and Matteucci, M.D. (1983) Diallylformamidines: depurination resistant N6-protecting group for deoxyadenosine. Nucleic Acids Res., 11, 8031–8036.
11. Gilar, M. and Bouvier, E.S.P. (2000) Purification of crude DNA oligonucleotides by solid-phase extraction and reversed-phase high-performance liquid chromatography. J. Chromatogr., 890, 167–177.
12. Gilar, M. (2001) Analysis and purification of synthetic oligonucleotides by reversed-phase high-performance liquid chromatography and photo diode array and mass spectrometry detection. Anal. Biochem., 298, 196–206.
13. Arghavani, M.B. and Romano, I.J. (1995) A method for the purification of oligonucleotides containing strong intramolecular interactions by reversed-phase high-performance liquid chromatography. Anal. Biochem., 231, 201–209.
14. Ivanetic, K.M., Reid, R.C., Ellison, R., Perry, K., Taylor, R., Reschenberg, M., Mainieri, A., Zhu, D., Argo, J. et al. (1999) Automated purification and quantification of oligonucleotides. BioTechniques, 27, 810–812, 814–818, 820 passim.
15. Cullen, P. and Lorkowski, S. (2002) DNA microarrays in medicine: manufacturing techniques and potential applications. Expert Opin. Ther. Pat., 12, 1783–1794.
16. Heller, M.J. (2002) DNA microarray technology: devices, systems, and applications. Annu. Rev. Biomed. Eng., 4, 129–153.
17. Connolly, B.A. (1987) The synthesis of oligonucleotides containing a primary amino group at the 5′-terminus. Nucleic Acids Res., 15, 311–3139.
18. Kwiatkowski, M., Nilsson, M. and Landegren, U. (1996) Synthesis of full-length oligonucleotides: cleavage of apurinic molecules on a novel support. Nucleic Acids Res., 24, 4632–4638.
19. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol., 5, R80, http://www.bioconductor.org/docs/faq/index.html#Citing%20Bioconductor.

Downloaded from https://academic.oup.com/nar/article-abstract/36/4/1334/2410393
by guest on 26 July 2018