Biotechnology and Bioprocess Engineering **2005**, 10: 432-443

Design, Optimization and Validation of Genomic DNA Microarrays for Examining the *Clostridium acetobutylicum* Transcriptome

Keith V. Alsaker, Carlos J. Paredes, and Eleftherios T. Papoutsakis*

Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA

Abstract Microarray technology has contributed significantly to the understanding of bacterial genetics and transcriptional regulation. One neglected aspect of this technology has been optimization of microarray-generated signals and quality of generated information. Full genome microarrays were developed for *Clostridium acetobutylicum* through spotting of PCR products that were designed with minimal homology with all other genes within the genome. Using statistical analyses it is demonstrated that signal quality is significantly improved by increasing the hybridization volume, possibly increasing the effective number of transcripts available to bind to a given spot, while changes in labeled probe amounts were found to be less sensitive to improving signal quality. In addition to Q-RT-PCR, array validation was tested by examining the transcriptional program of a mutant (M5) strain lacking the pSOL1 178-gene megaplasmid relative to the wildtype (WT) strain. Under optimal conditions, it is demonstrated that the fraction of false positive genes is 1% when considering differentially expressed genes and 7% when considering all genes with signal above background. To enhance genomic-scale understanding of organismal physiology, using data from these microarrays we estimated that 40~55% of the *C. acetobutylicum* genome is expressed at any time during batch culture, similar to estimates made for *Bacillus subtilis*.

Keywords: gram-positive, solvent-production, gene-expression, sporulation, heat-shock, butanol, degenerate strain, extrachromosomal, number-of-expressed-genes, Spo0A, AbrB, diffusion limitation, microarray design

INTRODUCTION

The advent of DNA-microarray technology has resulted in the generation of an enormous amount of new information about microbial genetics and transcriptional regulation [1-4]. An important aspect of the DNA microarray technology is that large amounts of information can be generated for organisms whose genome is sequenced but whose genetics has not been as extensively studied as *Escherichia coli* or *Bacillus subtilis*. An example is the industrially relevant microbe *Clostridium acetobutylicum* that was used extensively in the first half of the 20th Century for production of the solvents acetone, butanol, and ethanol. Biochemical production of these chemicals fell out of favor with the rise of the petrochemical industry [5]. Today there is a significant interest in studying this organism not only because of the renewed interest in production of chemicals through renewable resources, but because *C. acetobutylicum* is a strong candidate as a model organism for all clostridia, including many pathogenic, clinically important strains. We have previously constructed and used targeted cDNA microarrays with ca. 25% genome coverage to study asporogenic and nonsolventogenic *C. acetobutylicum* strains [6], recombinant strains featuring overexpression of the stationary phase transcriptional regulator Spo0A [7] and the chaperone GroESL [8], product-concentration driven changes [9], and the effects of butanol stress on cellular transcription [7,10]. One of the most important aspects of these studies was that many of the cellular programs studied with partial genome arrays, like motility and fatty acid synthesis, had not been extensively examined in any clostridia and it is unlikely that these programs would have been *a priori* hypothesized to be affected by seemingly unrelated cellular events like sporulation and the stress response to solvents.

To continue our understanding of clostridial genomics we undertook the design, construction and validation of full genome microarrays for *C. acetobutylicum*, an organism with ca. 73% A+T genome. Basic production of glass microarrays and optimization of printing conditions has been extensively covered [11]. However, a neglected aspect of microarray construction is the optimization of signal quality. The quality of the data generated from

***Corresponding author**
Tel: +1-847-491-7455 Fax: +1-847-491-3728
e-mail: e-paps@northwestern.edu

microarrays can only be as good as the microarray itself and the methodology used to normalize data and identify differentially expressed genes. We previously demonstrated the advantages of using a nearest neighbor approach for normalization and identification of differentially expressed genes [12]. While constructing *C. acetobutylicum* arrays we took several approaches to optimize the generated signal, namely through reduction of nonspecific binding of probes with spotted material and optimization of signal quality with probe quantity and hybridization volume. We also undertook a thorough validation of these microarrays with the aim of further improving the quality of outcomes through understanding of gene- and probe-specific issues that affect the data. Finally, we use data from these arrays to examine interesting questions in the context of microbial physiology in the genomic era, foremost among which is the number of genes expressed during key stages of this well-known ("classical") batch fermentation.

MATERIALS AND METHODS

PCR Primer Design

The probes for the full genome array were designed using the PRIMEGENS [13] program to select regions of each open reading frame (ORF) that has minimal sequence similarity with any other ORF. PRIMEGENS first carries out a BLAST [14] comparison for each target ORF against all other genomic ORFs to identify possible homologous sequences with a similarity of 90% or greater. Then it performs optimal sequence alignment between the target ORF and each of its homologous ORFs using dynamic programming. Based on these alignments, a gene-specific fragment is selected and a pair of suitable primers to amplify such fragment is automatically generated by PRIMER3 [15]. After several trial runs, we determined that the best results were obtained using a maximum expectation value (e-value) of 1×10^{-15} and a minimum fragment size of 150 nucleotides. Whenever a new set of suitable primers could not be generated from the non-homology zone of the gene determined by PRIMEGENS, the 100 nucleotides upstream and downstream of it were also used in the search for a suitable new pair of primers. PCR primers for 632 genes designed for a previous set of microarrays [6] were also used for this set of microarrays.

PCR Amplification

In order to minimize genomic DNA in the spotted PCR product and thus minimize nonspecific hybridization of cDNA probes, a two-step PCR amplification [16] was used to dilute the amount of chromosomal DNA in the final PCR product. For the first reaction in a total volume of 50 μL, 90~100 ng of *C. acetobutylicum* chromosomal DNA were combined with 1.25 units AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA, USA), PCR buffer (Applied Biosystems), 2.5 mM $MgCl_2$, 0.63 μM of each primer, and 0.66 mM dNTPs (Applied Biosystems). Fragments were amplified in a GeneAmp PCR System 9700 (Applied Biosystems) by 32 cycles of 94°C for 50 sec, the primer-specific melting temperature for 50 sec, and 72°C for 60 sec. A second PCR reaction was performed using 2 μL of the first PCR reaction as the template, and fragments were amplified as previously described but with a final volume of 100 μL and using 35 amplification cycles. Products in each step were verified using a Ready-To-Run Separation Unit (Amersham Biosciences, Piscataway, NJ, USA). PCR products from the second amplification reaction were purified by loading samples on a Millipore $MultiScreen_{96}$ Cleanup Plate (Bedford, MA, USA) following manufacturer's instructions. Purity and products sizes were verified by running 1 μL of samples on a 1.5% agarose gel. Quantification was performed by running approximately 12 samples per 96 well plate on gels with serial dilutions of ϕX174-HaeIII digest (New England Biolabs, Beverly, MA, USA). After the gel was imaged using a Kodak DC-120 digital camera (Rochester, NY, USA) and quantitated with Kodak 1D 2.0.2 software, linear regression was performed on the 603 bp fragment of the ϕX174-HaeIII lanes to relate DNA quantity and signal intensity. Results from quantitation of some samples of a single plate were used to evaluate all samples from the same plate. For PCR products that did not meet the quantitation benchmark of 1.2 μg total, the reaction was repeated under less stringent PCR reannealing conditions or PCR primers were redesigned and products reamplified.

For microarray negative controls, portions of the *C. pasteurianum* genes *dhaC* (primers 5′-CCCCTGCCCTG CAA-3′ and 5′-TGCAGAAGAGACGAA-3′) and *mop* (5′-AAATTGTAAAACTAGGAGGAATTAATTATG-3′ and 5′-ACTATGTTGGTTATTAAGTGAAAACAAT-3′) were PCR amplified as described above. Also used as negative controls were SpotReport Alien PCR Products (Stratagene, La Jolla, CA, USA). Negative controls represented ca. 2% of all spots in the complete microarrays.

Microarray Printing

PCR products were transferred from 96 well plates to 384-well V-bottom polystyrene plates (Genetix, Dorset, UK). Samples were concentrated in a rotary SpeedVac until dried, then redissolved in either a 35% dimethyl sulfoxide (DMSO) solution (Sigma, St. Louis, MO, USA) or Pronto Hybridization Solution (Corning, NY, USA) to a final concentration of 0.1 or 0.2 μg/μL. For long-term storage, plates were covered with Corning Costar Thermowell Sealing Tape and stored at -20°C.

Microarrays were printed (University of Wisconsin Gene Expression Center, Madison, WI, USA) with a Genomic Solutions GeneMachines OmniGrid arrayer (Ann Arbor, MI, USA) on Corning UltraGAPS slides and then UV cross-linked at 300 mJ. A Pronto Universal Microarray Hybridization kit (Corning) was used to reduce slide autofluorescence via sodium borohydride treatment [17] and to block nonspecific binding on the glass surface. Both unblocked and blocked slides were stored in a ni-

trogen-purged dessiccator. It was found that slides could be stored under nitrogen-purged desiccation with no increase in background signal for over four months if the slides were unblocked and over one month if blocked (data not shown).

RNA Sampling and Isolation

Cell pellets were collected by centrifuging 3 to 10 mL of cultures at 5,000×g for 10 min, 4°C. Cells were then lysed by resuspending in 200 μL SET buffer (25% sucrose, 50 mM EDTA [pH 8.0], 50 mM Tris [pH 8.0]) with 20 mg/mL lysozyme (Sigma) and incubation at 37°C for 4 min. One milliliter of ice cold TRIzol (Invitrogen, Carlsbad, CA, USA) was added and the sample was vortexed. Samples were immediately stored at -85°C and used within one month to minimize RNA degradation. To isolate and purify the RNA, samples were thawed at room temperature; 500 μL of sample was diluted with an equal amount of ice cold TRIzol and then purified following manufacturer's instructions. The RNA was resuspended in 20~30 μL RNase-free water and quantitated in a spectrophotometer by measuring absorbance at 260 and 280 nm. Samples were only used if the A_{260}/A_{280} ratio was greater than 1.8. RNA integrity was also verified by running 0.5 μL of sample on a 1.0% agarose gel. Purified RNA were stored at -85°C and used within one week of purification.

cDNA Probe Labeling

Since reverse transcriptases do not efficiently incorporate dye-labeled dNTPs into nascent cDNA, an indirect labeling approach was followed where cDNA was first made in the presence of nucleotides and aminoallyl dUTP, purified, and then the aminoallyl dUTP groups were coupled to a monoreactive dye (Cy3 or Cy5). This protocol was adapted from one developed by The Institute for Genomic Research (http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml). Briefly, total RNA (25 μg) was mixed with 15 μg random hexamers (Roche, Indianapolis, IN, USA) and the volume brought up to 18.5 μL with RNase-free water. Samples were incubated at 70°C for 10 min and then snap-frozen in a dry ice/ethanol bath for 30 sec. Deoxynucleotide triphosphates (Invitrogen; 0.6 mM dATP, 0.4 mM dCTP, 0.4 mM dGTP, and 0.36 mM dTTP), 0.24 mM aminoallyl dUTP (Sigma), 400 units SuperScript II reverse transcriptase (Invitrogen), First Strand Buffer (Invitrogen), and 6 μM dithiothreitol were added to a final volume of 30 μL. The samples were incubated for 3 h to overnight at 42°C, and the reaction was stopped by addition of 0.5 M EDTA (pH 8.0). The remaining RNA was degraded by addition of NaOH (final concentration 11 mM), incubation at 65°C for 15 min, and neutralization with HCl (final concentration 11 mM). The cDNA was purified by bringing up the volume to 450 μL with sterile water and concentrating three times on a Millipore YM-30 column. The samples were dried in a rotary SpeedVac, redissolved in 4.5 μL of a 0.1 M sodium carbonate buffer (pH 9.0), and combined with 54 nmol NHS-ester Cy3 or Cy5 monofunctional dye (Amersham Biosciences) dissolved in an equal volume of DMSO. The mixture was incubated in the dark for 1 h. To purify the labeled cDNA, 35 μL of a 100 mM sodium acetate solution (pH 5.2) was added before purifying on a Qiagen Qiaquick PCR purification column (Valencia, CA, USA). cDNA quantity, quality, and dye incorporation was measured (A_{260}, A_{280}, A_{550} [Cy3 probe], and A_{649} [Cy5 probe]) on a spectrophotometer. Only samples with a dye incorporation of at least 10 dye molecules per 1,000 nucleotides were hybridized on microarrays. Samples were stored at -20°C for up to 3 weeks until microarray hybridization.

Microarray Hybridizations

Oppositely labeled microarray probe pairs were mixed together, dried in a rotary SpeedVac, and redissolved in Pronto Universal Hybridization Solution. The samples were incubated at 95°C for 5 min and loaded on the microarray under a LifterSlip (Erie Scientific, Portsmouth, NH, USA). The microarrays were placed in a Corning hybridization chamber with 50 μL 10×SSC (1×SSC is 0.15 M NaCl and 0.015 M sodium citrate) to maintain humidity. Microarrays were hybridized for 14~16 h in a 42°C water bath. Following hybridization, microarrays were washed in TeleChem (Sunnyvale, CA, USA) wash buffers A (42°C, 5 min), B (room temperature, 5 min), and C (room temperature, 2 min) and water. Slides were then spun dry at 240×g for 10 min, scanned with an Agilent Scanner model G2656BA (Wilmington, DE, USA), and quantitated with GenePix software (Molecular Devices, Union City, CA, USA).

Microarray Analysis

Microarray expression data were normalized using a segmental nearest neighbor logarithmic expression ratio-of-the mean (SNN-LERM) approach that has been previously described [12] and coded in MATLAB (MathWorks, Natick, MA, USA). Three self-self hybridizations were used to calibrate the number of genes per interval, number of intervals and gamma factor needed to run the SNN-LERM algorithm with the microarrays in this set of experiments. The number of genes per interval (200), the number of intervals (25) and the gamma factor (2.0) used in previous experiments with smaller genome-coverage microarrays [12] were also applicable for the microarrays described here (data not shown).

Normalized ratios were grouped by expression similarity using an average linkage clustering algorithm in Cluster [18]. Data was visualized by heat maps (or 'Eisen plots') in Treeview [18].

Strains

C. acetobutylicum ATCC 824 (American Type Culture Collection, Manassas, VA, USA) is the wildtype (WT) strain. Degenerate strain M5 lacks the pSOL1 megaplasmid [19].

Table 1. Primers used for Q-RT-PCR in this study

Gene Name	Gene Accession Number	Left Primer	Right Primer
abrB	CAC0310	TGGTGAGCAAATCATTTTAAAGAAA	CCTTCTTTTAATTCTTCAAGACAATGC
nadE	CAC1050	GCCCCTATACCTCCAAGCAG	GCACATGCACCTTCCACTTT
sigG	CAC1696	CATGACGATGGTGATGCAAT	CGCTTAGCTTTTTCATAGCCTCT
spo0A	CAC2071	CCGCTGTTGGACAGGATAAA	TGACCTTTTCTGCTCTGAATTTG
sigF	CAC2306	ATGGGACTCGTAAAGGCTGTT	CAAAAATCTTTTTATCTCCCCCATA
pullulanase	CAC2679	TTCTCCACTGTGGCGTAGAGTT	TCTCTAAGATCCCAATCTATCCAATTT
hsp18	CAC3714	ATGGAAATGCAGGCTTCAAA	CAACAATGTCATCTCTTTTTGCAT
aad	CAP0162	AGAAAATGGCTCACGCTTCA	GCAATGCCACTAGGAATATTGTG
adc	CAP0165	GAAAGCGGACAGGCTATTCC	GCTTTGGATACCCGAGCTTT

Growth Conditions and Maintenance

Strains were stored at -85°C in clostridial growth medium (CGM) [20] containing 15% glycerol. Strains were grown in an anaerobic chamber (Forma Scientific, Marietta, OH, USA) in CGM supplemented with 80 g/L glucose or on agar-solidified 2×YTG plates. WT cultures were inoculated with single colonies from plates at least 4 days old and heat shocked at 70~80°C for 10 min. Liquid cultures of M5 were inoculated with single colonies from a plate not more than two days old without heat shocking.

Analytical Methods

Cell growth was determined by absorbance measurement at 600 nm (A_{600}) in a spectrophotometer. Linearity of measurements was ensured by diluting samples to an absorbance below 0.40 with the appropriate media. Culture supernatants were analyzed for glucose, acetate, butyrate, acetoin, ethanol, acetone, and butanol on a Waters high-performance liquid chromatography system (Milford, MA, USA) [21,22].

Housekeeping Genes Determination

The data from six previously published microarray hybridizations [6] comparing the expression profiles of *C. acetobutylicum* and strain M5 (lacking the pSOL1 megaplasmid) were used to find potential candidates for housekeeping genes to be used in Q-RT-PCR. The subset of common genes present in at least four of the six slides was used as a starting set. For each slide, the log ratios and log mean intensities of each gene were sorted, and percentile ranks of 0 and 100 were assigned respectively to the lowest and highest values. For each gene the mean rank and standard deviation across slides was calculated for both the log ratio and the log mean intensity. Since genes with relatively low expression levels have high error associated with intensity measurement and genes with relatively high expression may also be inaccurate due to saturation or nonspecific binding effects, genes with a mean rank (either log ratio or log mean intensity) between 0 and 15 or 85 and 100 were discarded. Of the remaining genes, the best candidate housekeeping genes had a relatively small standard deviation of log ratio or log mean intensity ranking; in other words, these genes had a relatively consistent pattern of gene expression. Selected as putative housekeeping genes were CAC1050 (*nadE*, an NH_3-dependent NAD^+-synthetase) and CAC 2679 (pullulanase).

Quantitative Reverse Transcription Polymerase Chain Reaction (Q-RT-PCR)

Two micrograms of total RNA was reverse transcribed in a total volume of 100 μL using a Taqman (Applied Biosystems) kit following the manufacturer instructions. One microliter of the reverse transcription reaction was then mixed with 1 μM of each gene specific primer (Table 1), SYBR Green (Applied Biosystems), and sterile water up to 25 μL. Samples were run on an iCycler iQ Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) following the manufacturer's directions. To reduce data variability, samples for each gene were run in triplicate on the same 96 well plate along with two negative control wells per gene (defined as all Q-RT-PCR components with 1 μL water substituted for the reverse transcription product) to verify that true signals were above background. For each gene, a PCR standard curve was generated in order to calculate the PCR efficiency and to verify that all real time PCR results were linear. Housekeeping genes (selected as described above) were run on each plate and used to normalize the calculated expression ratios (M5/WT).

RESULTS AND DISCUSSION

Effect of Microarray Spotting Solution and Contact Time on Spot Diameters

PCR products spotted on glass microarrays are usually redissolved in spotting solutions at concentrations in the range of 0.1~0.5 μg/μL [23-25]. While solutions like 3× SSC are convenient for redissolving PCR products, the small volumes required for maintaining the proper concentration will often evaporate away by the end of printing. DMSO solutions of 25~50% were introduced [11] as an improvement; it is not only a denaturant of double stranded PCR products, but its hygroscopic properties

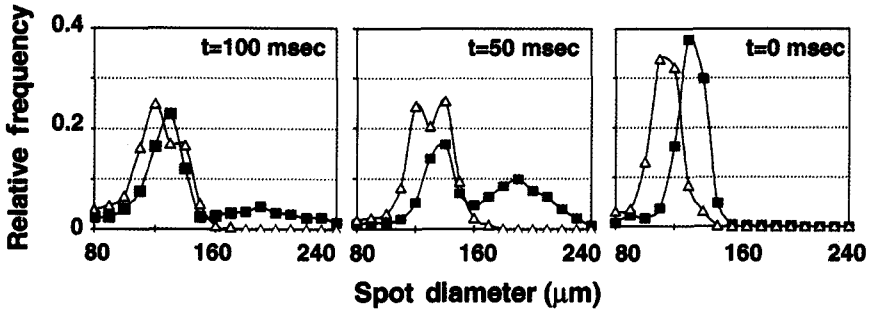

Fig. 1. Histograms showing typical results of spot size distribution as a function of contact time (t=100, 50, and 0 msec) and spotting solution. PCR products were redissolved in either Pronto Spotting Solution (closed squares) or 35% DMSO (open triangles). Results are from a single microarray containing three subarrays with all three contact times.

minimizes evaporation within the plate holding the spotting DNA solutions (spotting plate). A product recently introduced by Corning, the Pronto Spotting Solution, is marketed as a minimally evaporating alternative to DMSO. The ability to print more arrays is very desirable considering the cost and labor involved in making PCR products for an entire bacterial genome, but the spotting solution must perform similarly (in terms of signal-to-noise ratio, spot diameter, and ability to withstand repeated freeze thaws) or better to currently used solutions like DMSO.

Two 384-well plates, each containing identical PCR products from 192 randomly selected *C. acetobutylicum* genes, were generated as described in Materials and Methods. One plate was redissolved in 35% DMSO and the other in Pronto Spotting Solution; in each case PCR products were redissolved to a final concentration of at least 0.2 μg/μL. Three separate subarrays were printed from both plates on the same glass slides with contact times (the length of time a microarray pin touches the glass slide during the printing process) of 0, 50, and 100 msec. Within each subarray there were ten spots from each well in the two separate microarray spotting plates. By printing the arrays in this manner the effects of contact time and spotting solution could be tested simultaneously with the same hybridization solution. For each condition (spotting solution and contact printing time) 1,920 spots (192 genes×10 replicates) were quantitated. Spots for which the quantitation software could not detect signal were omitted from the analysis.

The frequencies of various spot sites from a typical single microarray are shown in Fig. 1, which demonstrates that increased contact time generally leads to increased spot diameter. Contact times of 50 and 100 msec for the Pronto spotting solution results in a spot size distribution that "tails" off towards 240 μm. Such a distribution is not acceptable for high density microarrays due to the risk of spots touching each other; increasing the distance between spots would reduce the total number of genes spotted on the array. At a 0 msec contact time, the distributions become narrower. At a 100 msec contact time the mean spot diameters for the Pronto- and DMSO-dissolved spots are, respectively, 141.4 and 119.5 μm; when the contact time is decreased to 0 msec the mean spot diameters decrease to 129.5 and 110.0 μm. While the Pronto Spotting Solution results in larger spot diameters, the increase relative to the DMSO spots is small enough to use in high density arrays printed with 0 msec contact time.

Increased Hybridization Volume Results in Improved Signal Quality

Under standard static hybridization conditions, the movement of labeled cDNA molecules is by diffusion alone. Even during the course of a 12~20 h hybridization, a microarray spot is likely to only hybridize complementary cDNAs in the local neighborhood of a few hundred microns [26]. While some devices have been demonstrated [27,28] to introduce fluid circulation across the microarray (and some of which are commercially available [Affymetrix, http://www.affymetrix.com; Agilent, http://www.agilent.com]), it is probably a reasonable assumption that most researchers still perform microarray hybridizations statically. One way to increase the probability that a transcript will bind to a spot is to increase the distance (which is typically considerably shorter than the diffusion distance of a labeled cDNA molecule [26,29,30]) between the coverslip and slide by increasing the total hybridization volume. For a 22×60 mm coverslip (total surface area 1,320 mm^2) the volume of hybridization fluid that can fit between the coverslip and slide is approximately 35 μL or 35 mm^3; therefore the distance between coverslip and slide is approximately 25 μm. LifterSlips are commercially available coverslips with Teflon "rails" on two opposite sides. The volume underneath a 22×60 mm LifterSlip is approximately 75 μL, corresponding to a height of 57 μm. Since this distance is still smaller than the distance a cDNA can diffuse within 12~20 h, it was hypothesized that hybridizations performed under a LifterSlip would result in improved hybridization signal quality. However, another logical way to improve signal quality may be to simply increase the amount of probe hybridizing on the microarray. One approach to examining this issue could be to hybridize the same probe concentrations under different volumes; with this approach the cDNA molecules under different hybridization conditions would have a statistically equal likelihood of entanglement and ability to bind with other

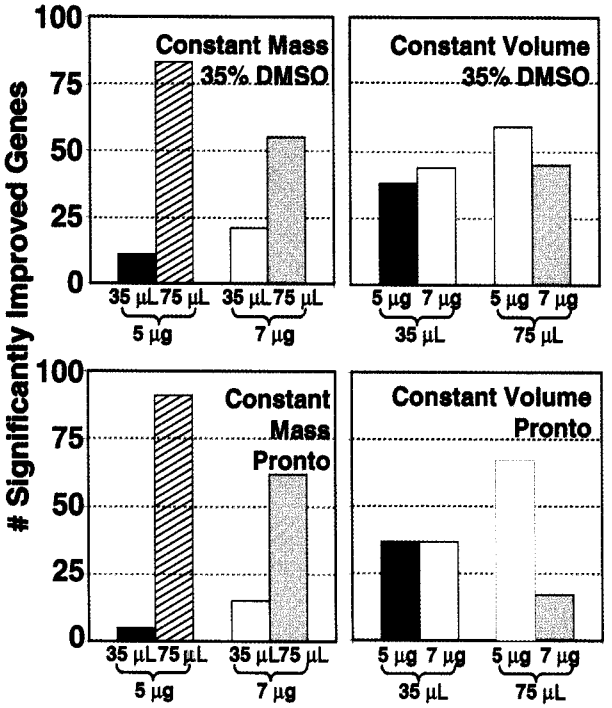

Fig. 2. Comparison of the number of genes with improved microarray signal quality between microarray hybridization conditions (probe mass or hybridization volume varied). Each test case represents separate one-way ANOVA comparisons of Z-test statistic scores between identical genes in two different conditions (see text). In each condition 192 genes were tested, but the number of genes without improved signal quality is not shown. While the results are shown for a single emission channel (Cy3 dye), results are similar in the second channel (Cy5 dye, data not shown).

cDNA molecules. However, cDNA probes are typically relatively dilute under hybridization conditions (0.1~0.2 µg/µL). We, thus, hypothesized such cDNA-cDNA interactions under hybridization conditions are relatively unimportant and chose to simultaneously vary the probe quantity and hybridization volume without keeping the probe concentration constant. To simultaneously examine the effects of probe quantity and hybridization volume on signal quality, four microarray hybridizations were performed on test microarrays where the amount of probe hybridization volume were varied: 5 µg probe-35 µL hybridization volume; 5 µg-75 µL; 7 µg-35 µL; and 7 µg-75 µL. Cy3- and Cy5-labeled cDNA pools were created by reverse transcribing identical pools of RNA taken from exponential, transitional, and stationary stages of *C. acetobutylicum* cultures to improve the chance that transcripts would be present. Also, the cDNA from individual reactions were combined so that the dye incorporation for probe hybridized on all microarrays would be equal. After quantitation, spots that were not located by the software (low signal intensity compared to background) were excluded. For each remaining spot in each channel, the Z-test statistic

$$z = \frac{\bar{y}_{spot} - \bar{y}_{background}}{\sqrt{\frac{\sigma^2_{spot}}{n_{spot}} + \frac{\sigma^2_{background}}{n_{background}}}}$$

was calculated where $\bar{y}_{spot}$ and $\bar{y}_{background}$ are, respectively, the average intensity of a spot and the average background intensity of a spot (the area taken for background calculation is a region, excluding the spot itself and other local spots, three diameters surrounding the spot); σ_{spot} and $\sigma_{background}$ are standard deviations of signal intensity; and n_{spot} and $n_{background}$ are the number of pixels in the spot and background. While the Z-test statistic can be interpreted as an indicator that the spot intensity is greater than background intensity, it is used here to compare signal-to-background ratios between spots; with a higher Z-test statistic, there is higher confidence that a spot intensity is greater than background. Z values were averaged for replicates of genes spotted from the same spotting solutions.

If we compare genes spotted in 35% DMSO and hybridized with 5 µg probe (Fig. 2), we see that 83 genes out of 192 in the 75 µL hybridization have significantly improved signal quality compared to the 35 µL hybridization; the 35 µL hybridization has only 11 genes with improved signal quality compared to the 75 µL hybridization. When we compare spots from the same spotting solution hybridized with 7 µg probe, we find that 55 genes in the 75 µL hybridization have significantly improved signal quality compared to the 35 µL hybridization which has only 21 significantly improved genes. When the same comparison is made (constant probe quantity) with genes spotted in the Pronto hybridization solution, the same trend of improved signal quality in the larger (75 µL) hybridization volume still holds. For example, when the amount of probe is held constant at 5 µg, 91 genes have significantly better signal quality in the 75 µL case compared to only 5 in the 35 µL case.

The same method can be used to compare signal quality in hybridizations with constant volume, or to compare signal quality when probe quantity is varied. When signal quality from genes spotted in 35% DMSO and hybridized at 75 µL is compared, it is found in the 5 µg probe case that 59 genes have improved signal quality compared to the 7 µg case which has 45 genes with improved signal quality. This trend (improved signal quality with 5 µg hybridized probe in a 75 µL volume) is consistent with spots in the Pronto hybridization solution (5 µg probe: 67 improved genes; 7 µg probe; 17 improved genes) but the same trend is not consistent at the smaller hybridization volume. For genes spotted in 35% DMSO and hybridized in 35 µL volume, 44 genes are significantly better in the 7 µg case compared to 38 in the 5 µg case. When comparing genes spotted in the Pronto solution and hybridized in 35 µL volume, 37 genes have significantly improved signal quality in both the 5 µg and 7 µg hybridizations. These results indicate that signal quality can be significantly improved when the hybridization volume is increased from 35 to 75 µL. At this larger volume

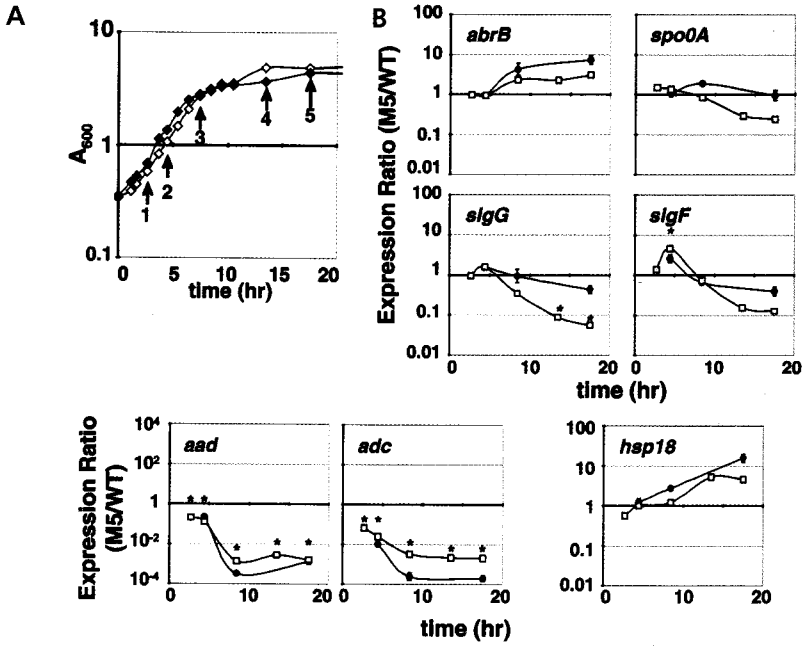

Fig. 3. Growth kinetics for the WT (open diamonds) and M5 (closed diamonds) strains (A). Time points for which microarray (1-5) and Q-RT-PCR (2, 3, and 5) analysis was performed are indicated by numbers directly below the symbols. Comparison of M5/WT expression ratios using microarray analysis (open squares) and Q-RT-PCR (closed circles) (B). Asterisks next to microarray data points indicate differential expression to a 95% confidence interval. Error bars on the Q-RT-PCR data indicate the standard error of the mean. Q-RT-PCR ratios are averaged from results calculated using *nadE* (CAC1050) and pullulanse (CAC2657) as housekeeping genes. Note the *aad* and *adc* graphs have a different scale than the other graphs.

it seems that signal quality can be improved further by, counterintuitively, reducing the amount of hybridizing probe from 7 to 5 μg. However, at lower hybridization volumes signal quality is not necessarily improved by reducing the amount of probe hybridized on the microarray. While these results are shown for scannings made in the Cy3 channel, the same trends were also observed simultaneously in the Cy5 channel (data not shown).

Signal Quality is Retained When Spotted PCR Product Concentration is Decreased

Repeated microarray printings may allow PCR product-dissolving solvents to evaporate. This may change the PCR product concentration and thus the spotting morphology; this issue may be offset by increasing the final PCR product volume. On the other hand, if PCR product concentration becomes too low, the quality of signal may be diminished by reducing the signal-to-noise ratio. Manufacturer guidelines suggest redissolving PCR products in the Pronto hybridization solution to a final concentration of 0.15 μg/μL. To test how signal quality changes under diluted conditions, we examined signal quality from spotted PCR products when the PCR product concentration is reduced from 0.2 μg/μL to 0.1 μg/μL. A microarray spotting plate was made where 192 wells contained PCR products redissolved in 6 μL Pronto spotting solution; 96 wells had a final concentration of ~0.2 μg/μL and 96 wells had a final concentration of ~0.1 μg/μL. While there was a different set of 96 genes in each concentration condition, the genes in each set were selected at random; thus it is assumed that each of the two 96 gene subsets has an equal intensity distribution. The microarray printing quality between the two sets of genes was the same as determined by average spot diameter (data not shown). Six slides were hybridized on the same day with 5, 7, or 9 μg of both Cy3- and Cy5-labeled probe (n=2 slides) from the same cDNA pool. Microarrays were scanned in a random order with identical laser and PMT settings. After the slides were quantitated and Z-test statistics were calculated for spots with signal intensities above background, a two-way ANOVA test was used to compare signal quality as a function of spotted PCR-product concentration and probe quantity. The calculated P-value for the effect of PCR-product concentration versus signal quality was 0.145, indicating that statistically there was no difference in signal quality when the PCR-product concentration is decreased from 0.2 to 0.1 μg/μL. This indicates that redissolving the spotted PCR products at a final concentration of 0.1 μg/μL would give as good signal quality as 0.2 μg/μL.

Microarray Validation: Comparison to Genotype Data and Quantitative Reverse Transcriptase Polymerase Chain Reaction (Q-RT-PCR)

Microarray validation ensures data can be reliably trusted; errors such as misprimed PCR products, incor-

rectly located PCR products, and microarray spotting plate contamination can lead to false information. Microarray data confirmation with a second, independent technique such as quantitative reverse transcriptase polymerase chain reaction (Q-RT-PCR) or Northern blots improves the data reliability.

C. acetobutylicum genes are located on a chromosome (3739 ORFs) and an extrachromosomal megaplasmid pSOL1 (178 ORFs) [31] which contains the solventogenic genes *aad, ctfA, ctfB,* and *adc* [32]. Repeated transfers of *C. acetobutylicum* during growth results in strain degeneration whereby the pSOL1 megaplasmid is lost. Various strains of *C. acetobutylicum*, including M5 [19] and DG1 [33], have been created in which pSOL1 has been lost and are neither able to produce solvents nor, surprisingly, sporulate. Analysis of these strains through microarray analysis would increase the understanding as to how clostridial strains produce solvents and sporulate, which genes on the pSOL1 megaplasmid are required for sporulation, and when these genes are transcribed. The M5 strain was previously analyzed on a partial genome microarray containing approximately 25% of the *C. acetobutylicum* genome [6]. Analysis of M5 versus WT is particularly helpful when used for microarray validation since up to 178 transcripts are present only in WT but not M5.

WT and M5 cultures were grown in static flasks and RNA and supernatant samples were taken through out exponential, transitional, and stationary phases of growth (Fig. 3(A)). Growth (Fig. 3(A)) and metabolite kinetics (not shown) for both the WT and M5 strains were similar to previously published results [6]. Microarray analysis was performed on five time points (Fig. 3(A), numbered 1-5), and Q-RT-PCR was performed on three time points (numbers 2, 3, and 5) from both the WT and M5 strains. Rank invariant [34] or housekeeping genes are often used to normalize expression levels of Q-RT-PCR data. Since candidate genes for the entire *C. acetobutylicum* genome were not known *a priori*, data from a previous partial genome microarray analysis of M5 (hybridized against WT) [6] were used to generate a list of candidate genes based on (1) rank invariant expression ratio and (2) rank invariant average spot intensity. The two genes selected as housekeeping genes have accession numbers CAC1050 (*nadE*, annotated as being an NH_3-dependent NAD^+-synthetase) and CAC2679 (pullulanase). PCR plates were run in pairs where samples were prepared simultaneously and plates run immediately after each other; for each pair, one plate contained housekeeping gene CAC1050 and the other CAC2679. Expression ratios (M5/WT) were calculated twice using each housekeeping gene; since the normalized data resulted in similar trends (data not shown), the data generated by normalizing to each housekeeping gene was averaged together (Fig. 3(B)) to generate an average expression ratio.

Since M5 lacks the pSOL1 megaplasmid and the microarray spotted material was designed to minimize nonspecific hybridization, it was expected that all expressed pSOL1 genes would have greater expression in the WT channel. A subset of pSOL1 genes was selected for which expression ratios (M5/WT) could be calculated for at least four out of five time points and were identified as being differentially expressed to a 95% confidence interval for at least one time point (n=102). Of the 509 ratios in this data set (one ratio could not be calculated due to signal intensities close to background), only 6 (1%) had higher expression in M5 than WT, and the highest false positive indicated a minimal 1.9-fold difference (Fig. 4(A)). Now we consider the case (Fig. 4(B)) where the differential expression condition is dropped and all pSOL1 genes with expression ratios available at four out of five time points are examined (n=144). Fig. 4(B) shows that of the 706 ratios (14 ratios were not calculable due to low signal intensities) calculated from the averaged data, 50 (7%) have positive ratios indicating higher expression in M5 (a false positive). The actual ratios of the false positives are relatively low compared to true positives. The highest false positive indicates M5 having expression 3.3-fold higher than WT at a single time point; by contrast, 453 ratios indicate expression of at least 3-fold higher in WT than M5. The results indicate that relative patterns of gene expression generated from these arrays are reliable whether or not differential expression filters are applied. In an attempt to understand the source of these errors (and thus filter out such false positives in future experiments), we first examined spot signal intensities of all occurrences of false positives. Of the 50 false positives, 19 came from 12 genes in which the spot signal intensity was less than 300 on a scale of 0 to 65,535. At signal intensities of such a low magnitude, there is inherently a higher degree of error in the ratio calculation and these genes may not even be expressed (as discussed later). Next, we used BLAST [14] searches of the spotted PCR products to look for cross-homologies with chromosomal genes. Seven more false positives occurred from two genes whose PCR products contained homologous regions with chromosomal genes; the 187 bp PCR product from CAP0011 contained a 60 bp match with a chromosomal gene, and the 383 bp PCR product from CAP0022 contained stretches of 60 bp matches over a 72 bp region of DNA and another 64 bp matches over a 79 bp region. Finally, four additional false positives from a single gene (CAP0161) were attributed to a misprimed PCR product. These last three genes will be excluded in future experiments; their probe-design was retained from our earlier partial genome array [6] and was not designed by the PRIMEGENS program.

As expected, the solvent formation genes *aad* (CAP0162) and *adc* (CAP0165) were expressed higher in WT at all five microarray time points (Fig. 3(B)); expression at each time point was differential with a 95% confidence. *aad* and *adc* expression has previously been shown [35] by Northern analysis to occur primarily in transitional and stationary phases of culture; these culture phases correspond with the last three microarray time points, where expression ratios indicate much higher expression in WT. Expression of both *aad* and *adc* were up to 400-fold higher in WT as determined by microarrays (Fig. 3(B)). Q-RT-PCR results also indicate higher transcription levels of *aad* and *adc* in WT; expression of *adc* is up

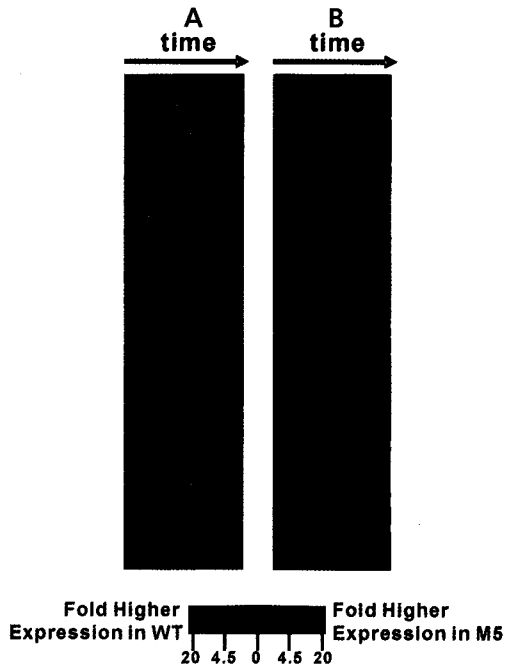

Fig. 4. Comparison of pSOL1 gene ratios of expression (M5/WT); genes with data available in four out of five time points were selected by having differential expression for at least one time point (A) or regardless of differential expression (B). Grey rectangles indicate no measurable ratio for the time point.

to 4900-fold higher in WT (Fig. 3(B)).

Since M5 exhibits a nonsporulating phenotype it would be expected that transcription of genes related to control of growth stage-specific expression would be affected. In *B. subtilis, abrB* has been shown to be the transcriptional regulator of transitional phase genes [36] and whose expression is repressed by the SpoOA protein [37,38], which is a stationary phase specific transcriptional regulator. The *C. acetobutylicum* genome contains three genes whose proteins greatly resemble the AbrB protein (CAC0310, CAC1941, and CAC3647), making it difficult to design a PCR product-specific target for each individual transcript [7]. Reporter studies of the three genes have shown that gene CAC0310 is expressed significantly higher than the other two candidate *abrB* genes [39]. For this reason, microarray probes were designed specifically for the CAC0310 *abrB* gene with the understanding that it will pick up nonspecific signals from the other two probes; probes specific for the other candidate *abrB* genes were not included on the microarray. Comparison of the microarray and Q-RT-PCR generated ratios show similar patterns (Fig. 3(B)). During exponential phase *abrB* is expressed equally in WT and M5. In transitional phase (time point 3) expression is higher in M5 (Q-RT-PCR: 4.3-fold higher; microarray: 2.3-fold higher), and this trend continues through stationary phase. It is then expected that *abrB* overexpression in M5 may be due to decreased expression of the *spo0A* gene which has been shown in *B. subtilis* [37], *C. beijerinckii* [40], and *C. acetobutylicum* [35] to regulate expression of stationary phase-specific genes. This hypothesis is confirmed only by microarray data, which indicates a decrease in *spo0A* expression in stationary phase (4.1-fold higher expression in WT at time point 5) but is not statistically differentially expressed. Interestingly, this pattern is not confirmed by Q-RT-PCR data, which shows an induction of expression at the transitional time point 3 (1.88-fold higher expression in M5) and nearly identical expression at the stationary time point 5. The *spo0A* Q-RT-PCR results were confirmed by additional Q-RT-PCR experiments (data not shown). We have previously used and validated the same *spo0A* microarray probe successfully [6,7] and it is not clear here why these results do not match.

sigG (CAC1696) and *sigF* (CAC2306) encode sporulation-related sigma factors specifically for, respectively, late forespore and late mother cell genes [41]. Since M5 does not sporulate, it would be expected that these genes would be expressed higher in WT than M5. This hypothesis is confirmed by the microarray and Q-RT-PCR data. According to the microarrays *sigG* is differentially expressed in the stationary phase time points 4 and 5 (18-fold higher in WT at time point 5) and, although not statistically differentially expressed, *sigF* is expressed at higher levels in WT (8.0-fold higher in WT at time point 5). Q-RT-PCR data show the same pattern but with smaller fold differences; at time point 5, expression of *sigG* is 2.3-fold higher in WT and *sigF* is 2.5-fold higher in WT.

hsp18 (CAC3714) is annotated as a small heat-shock protein. Previous results have shown that translation of the gene is induced by butanol stress [7,10], and protein expression is increased during pH-induced solventogenic chemostat cultures [42]. Both microarrays and Q-RT-PCR results indicate higher expression in M5 during stationary phase time points 4 and 5; at time point 5 expression is 4.6-fold higher as determined by microarrays and 15.5-fold higher by Q-RT-PCR analysis.

Estimation of the Number of Genes Expressed in WT and M5 Cultures

Generally it is not known which and what fraction of an organism's genes is utilized during various culture conditions, and such information can provide significant insight into microbial physiology. In order to estimate the number of genes expressed during these *C. acetobutylicum* culture, the intensity values of each channel (WT and M5) and for each gene were calculated from the log mean intensities and the normalized log ratios. Given that strain M5 lacks all of the pSOL1 megaplasmid genes, any intensity measured in the M5 channel of the pSOL1 genes cannot represent real gene expression. Hence, a distribution (Fig. 5) of the intensities of such pSOL1 genes yields an estimation of the minimum intensity that a gene should have to be considered as expressed. It can be seen (Fig. 5) that 90% of the pSOL1 genes in the M5 channel show an intensity less than or equal to 300 units (on a scale from 0 to 65,535). In contrast, only 47% of

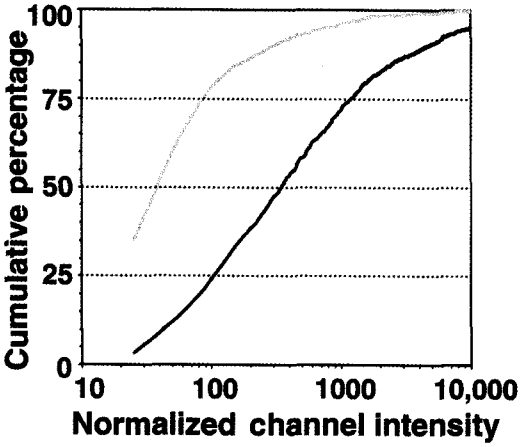

Fig. 5. Cumulative distribution of normalized intensities of pSOL1 genes for the strains WT (black line) and M5 (gray line).

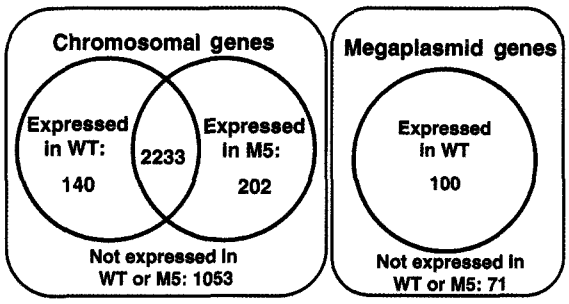

Fig. 6. Venn diagrams showing the distribution of expressed chromosomal and megaplasmid genes in the WT and M5 strains.

the pSOL1 genes in the WT channel have an intensity of 300 units or less, which implies that 53% of the pSOL1 genes plotted are actually expressed. Therefore this cutoff method can provide an estimate of the actual number of expressed genes.

Table 2 shows the number of expressed genes through the culture using a threshold of 300 intensity units. In both strains the evolution of genes expression follows the same trend. The number of expressed genes peaks in time point 2 (midexponential phase) whereby around 55% of the genes are expressed and then decreases until the end of the culture whereby only 40% of the genes are expressed.

An array-based approach for estimating the number of expressed genes in exponential phase *B. subtilis* cultures has also been proposed [43]. Using single channel macroarray results, Eymann and coworkers selected genes for which spot intensities minus the spot intensity standard deviation exceeded 20% of the background intensity [43]. Using this approach, it was estimated that 2,515 of 4,100 *B. subtilis* genes (61%) are expressed during exponential phase, a value that is similar to the results presented here. These microarray-based methods select within a population the genes that are most likely to be expressed. However, these methods are less likely to identify rarely expressed transcripts (<1 copy/cell).

The WT and M5 strains have major phenotypic and genotypic differences. To understand these differences further, we used the threshold results to determine in what strains each gene was expressed in at least one time point (Fig. 6). The completed microarrays contained targets for 3,628 of the 3,739 chromosomal genes and 171 of the 178 pSOL1 genes (this takes into account that 3 pSOL1 genes have been omitted from further analysis as previously described). A total of 2,233 chromosomal genes were expressed in both strains while 140 and 202 were expressed only in, respectively, the WT and M5 strains. Over 1,000 genes were not expressed in either strain, although these may include low expressed (<1 copy/cell) transcripts. Using the KEGG Pathway Database [44], some of these genes are included in pathways that are not expected to be expressed in either strain due to culture conditions; these pathways include fructose and mannose metabolism (10 genes), galactose metabolism (7 genes), and starch and sucrose metabolism (7 genes) and suggests that this type of analysis is reasonably accurate. Genes expressed in WT but not M5 include several chemotaxis genes (CAC0117-CAC0121) and potassium transport genes (CAC3679-CAC3682). Among the megaplasmid genes, 100 were expressed in the WT strain and 71 genes were not expressed. Complete lists of expressed genes and microarray data can be found at http://www.papoutsakisresearch.northwestern.edu.

Table 2. The estimated number of expressed genes during the time course of WT and M5 static flask cultures. Percentages are based on the total number of genes on the microarray (3,628 chromosomal genes, 171 megaplasmid genes), and the percentage of M5 expressed genes excludes all megaplasmid genes

Time Point	Time (hr)	# Genes Expressed		% Genes Expressed	
		WT	M5	WT	M5
1	2.7	2,046	1,997	54%	55%
2	4.4	2,086	2,046	55%	56%
3	7.4	2,028	1,973	53%	54%
4	13.5	1,922	1,926	51%	53%
5	17.5	1,518	1,489	40%	41%

CONCLUSION

In this study we showed that microarray design, construction, and use to achieve complete and reliable data can be optimized and evaluated by (1) using microarray probes that target nonhomologous genetic regions, (2) increasing the hybridization volume (as compared to increasing the amount of hybridizing probe), and (3) comparison of signal intensities and ratios between the *C. acetobutylicum* wildtype strain with the M5 mutant that lacks 178 genes. The optimized microarrays showed that, to a 95% confidence, 40~56% of each strain's genes are expressed at any given time and that sporulation in the M5 strain was probably arrested in the preseptational phase. These microarrays will be useful in future studies for further interpreting the transcriptional basis of *C. acetobutylicum*'s solventogenic, sporulation, and stress-response programs.

Acknowledgements The authors thank Iwona Spath, Erik Vanderlip, and Ryan Sillers for helping construct the microarrays, acknowledge the use of Keck Biophysics Facility equipment, and thank Prof. J.W. Peters (Montana State University) for providing *C. pasteurianum* genomic DNA. This work was supported by a National Science Foundation grant (BES-0418157) and an NIH/NIGMS Biotechnology Training Grant (T32-GM08449) fellowship to K.V.A.

REFERENCES

[1] Ye, R. W., T. Wang, L. Bedzyk, and K. M. Croker (2001) Applications of DNA microarrays in microbial systems. *J. Microbiol. Meth.* 47: 257-272.

[2] Conway, T. and G. K. Schoolnik (2003) Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Mol. Microbiol.* 47: 879-889.

[3] Rhodius, V. A. and R. A. LaRossa (2003) Uses and pitfalls of microarrays for studying transcriptional regulation. *Curr. Opin. Microbiol.* 6: 114-119.

[4] Dharmadi, Y. and R. Gonzalez (2004) DNA microarrays: experimental issues, data analysis, and application to bacterial systems. *Biotechnol. Progr.* 20: 1309-1324.

[5] Jones, D. T. and D. R. Woods (1986) Acetone-butanol fermentation revisited. *Microbiol. Rev.* 50: 484-524.

[6] Tomas, C. A., K. V. Alsaker, H. P. J. Bonarius, W. T. Hendriksen, H. Yang, J. A. Beamish, C. J. Parades, and E. T. Papoutsakis (2003) DNA-array based transcriptional analysis of asporogenous, non-solventogenic *Clostridium acetobutylicum* strains SKO1 and M5. *J. Bacteriol.* 185: 4539-4547.

[7] Alsaker, K. V., T. R. Spitzer, and E. T. Papoutsakis (2004) Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *J. Bacteriol.* 186: 1959-1971.

[8] Tomas, C. A., N. E. Welker, and E. T. Papoutsakis (2003) Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and large changes in the cell's transcriptional program. *Appl. Environ. Microbiol.* 69: 4951-4965.

[9] Tummala, S. B., S. G. Junne, C. J. Paredes, and E. T. Papoutsakis (2003) Transcriptional analysis of product-concentration driven changes in cellular programs of recombinant *Clostridium acetobutylicum* strains. *Biotechnol. Bioeng.* 84: 842-854.

[10] Tomas, C. A., J. A. Beamish, and E. T. Papoutsakis (2004) Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *J. Bacteriol.* 186: 2006-2018.

[11] Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush (2000) A concise guide to cDNA microarray analysis. *Biotechniques* 29: 548-562.

[12] Yang, H., H. Haddad, C. Tomas, K. Alsaker, and E. T. Papoutsakis (2003) A segmental nearest neighbor normalization and gene identification method gives superior results for DNA-array analysis. *P. Natl. Acad. Sci. USA* 100: 1122-1127.

[13] Xu, D., G. S. Li, L. Y. Wu, J. Z. Zhou, and Y. Xu (2002) PRIMEGENS: robust and efficient design of gene-specific probes for microarray analysis. *Bioinformatics* 18: 1432-1437.

[14] Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* 215: 403-410.

[15] Rozen, S. and H. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. pp 365-386. In: S. Krawetz, S. Misener, (eds.). *Methods in Molecular Biology*. Humana Press, Totowa, NJ.

[16] Richmond, C. S., J. D. Glasner, R. Mau, H. F. Jin, and F. R. Blattner (1999) Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* 27: 3821-3835.

[17] Raghavachari, N., Y. P. Bao, G. S. Li, X. Y. Xie, and U. R. Müller (2003) Reduction of autofluorescence on DNA microarrays and slide surfaces by treatment with sodium borohydride. *Anal. Biochem.* 312: 101-105.

[18] Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA.* 95: 14863-14868.

[19] Clark, S. W., G. N. Bennett, and F. B. Rudolph (1989) Isolation and characterization of mutants of *Clostridium acetobutylicum* ATCC 824 deficient in acetoacetyl-coenzyme A:acetate/butyrate:coenzyme A transferase (EC 2.8.3.9) and in other solvent pathway enzymes. *Appl. Environ. Microbiol.* 55: 970-976.

[20] Wiesenborn, D. P., F. B. Rudolph, and E. T. Papoutsakis (1988) Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. *Appl. Environ. Microbiol.* 54: 2717-2722.

[21] Tummala, S. B., N. E. Welker, and E. T. Papoutsakis (2003) Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. *J. Bacteriol.* 185: 1923-1934.

[22] Buday, Z., J. C. Linden, and M. N. Karim (1990) Improved acetone butanol fermentation analysis using subambient HPLC column temperature. *Enzyme Microb. Tech.* 12: 24-27.

[23] Wei, Y., J. M. Lee, C. Richmond, F. R. Blattner, J. A.

Rafalski, and R. A. LaRossa (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J. Bacteriol.* 183: 545-556.

[24] Schena, M., D. Shalon, R. Heller, A. Chai, P. O. Brown, and R. W. Davis (1996) Parallel human genome analysis: microarray-based expression monitoring of 1,000 genes. *Proc. Natl. Acad. Sci. USA* 93: 10614-10619.

[25] Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly (2003) Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Biol. Chem.* 278: 7540-7552.

[26] Worley, J., K. Bechtol, S. Penn, D. Roach, D. Hanzel, M. Trounstine, and D. Barker (2000) A systems approach to fabricating and analyzing DNA microarrays. In: M. Schena, editor. *Microarray Biochip Technology*. Eaton Publishing, Natick, MA, USA.

[27] Yuen, P. K., G. Li, Y. Bao, and U. R. Muller (2003) Microfluidic devices for fluidic circulation and mixing improve hybridization signal intensity on DNA arrays. *Lab Chip* 3: 46-50.

[28] Adey, N. B., M. Lei, M. T. Howard, J. D. Jensen, D. A. Mayo, D. L. Butel, S. C. Coffin, T. C. Moyer, D. E. Slade, M. K. Spute, A. M. Hancock, G. T. Eisenhoffer, B. K. Dalley, and M. R. McNeely (2002) Gains in sensitivity with a device that mixes microarray hybridization solution in a 25-micron-thick chamber. *Anal. Chem.* 74: 6413-6417.

[29] Gadgil, C., A. Yeckel, J. J. Derby, and W. S. Hu (2004) A diffusion-reaction model for DNA microarray assays. *J. Biotechnol.* 114: 31-45.

[30] Borden, J. R., C. J. Paredes, and E. T. Papoutsakis (2005) Diffusion, mixing, and associated dye effects in DNA-microarray hybridizations. *Biophys. J.* In press.

[31] Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183: 4823-4838.

[32] Cornillot, E., R. V. Nair, E. T. Papoutsakis, and P. Soucaille (1997) The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *J. Bacteriol.* 179: 5442-5447.

[33] Stim-Herndon, K. P., R. Nair, E. T. Papoutsakis, and G. N. Bennett (1996) Analysis of degenerate variants of *Clostridium acetobutylicum* ATCC 824. *Anaerobe* 2: 11-18.

[34] Tseng, G. C., M. K. Oh, L. Rohlin, J. C. Liao, and W. H. Wong (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* 29: 2549-2557.

[35] Harris, L. M., N. E. Welker, and E. T. Papoutsakis (2002) Northern, morphological, and fermentation analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 184: 3586-3597.

[36] Strauch, M. A., G. B. Spiegelman, M. Perego, W. C. Johnson, D. Burbulys, and J. A. Hoch (1989) The transition state transcription regulator *abrB* of *Bacillus subtilis* is a DNA-binding protein. *EMBO J.* 8: 1615-1621.

[37] Strauch, M., V. Webb, G. Spiegelman, and J. A. Hoch (1990) The Spo0A protein of *Bacillus subtilis* is a repressor of the *abrB* gene. *Proc. Natl. Acad. Sci. USA* 87: 1801-1805.

[38] Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick (2003) The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* 50: 1683-1701.

[39] Scotcher, M. C., F. B. Rudolph, and G. N. Bennett (2005) Expression of *abrB310* and *sinR*, and effects of decreased *abrB310* expression on the transition from acidogenesis to solventogenesis, in *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 71: 1987-1995.

[40] Ravagnani, A., K. C. Jennert, E. Steiner, R. Grunberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, J. G. Morris, and M. Young (2000) Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Mol. Microbiol.* 37: 1172-1185.

[41] Helmann, J. D. and C. P. Moran Jr (2002) RNA polymerase and sigma factors. pp 289-312. In: A. L. Sonenshein, J. A. Hoch, R. Losick, (eds.). *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*. ASM Press, Washington, D.C., USA

[42] Schaffer, S., N. Isci, B. Zickner, and P. Dürre (2002) Changes in protein synthesis and identification of proteins specifically induced during solventogenesis in *Clostridium acetobutylicum*. *Electrophoresis* 23: 110-121.

[43] Eymann, C., A. Dreisbach, D. Albrecht, J. Bernhardt, D. Becher, S. Gentner, L. T. Tam, K. Büttner, G. Buurman, C. Scharf, S. Venz, U. Völker, and M. Hecker (2004) A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* 4: 2849-2876.

[44] Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32: D277-D280.

[Received April 15, 2005; accepted July 23, 2005]