 Nucleic Acids Research, 1999, Vol. 27, No. 19 **3821–3835**

Genome-wide expression profiling in *Escherichia coli* K-12

Craig S. Richmond*, Jeremy D. Glasner, Robert Mau, Hongfan Jin and Frederick R. Blattner

Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

Received June 15, 1999; Revised and Accepted July 23, 1999

ABSTRACT

We have established high resolution methods for global monitoring of gene expression in *Escherichia coli*. Hybridization of radiolabeled cDNA to spot blots on nylon membranes was compared to hybridization of fluorescently-labeled cDNA to glass microarrays for efficiency and reproducibility. A complete set of PCR primers was created for all 4290 annotated open reading frames (ORFs) from the complete genome sequence of *E.coli* K-12 (MG1655). Glass- and nylon-based arrays of PCR products were prepared and used to assess global changes in gene expression. Full-length coding sequences for array printing were generated by two-step PCR amplification. In this study we measured changes in RNA levels after exposure to heat shock and following treatment with isopropyl-β-D-thiogalactopyranoside (IPTG). Both radioactive and fluorescence-based methods showed comparable results. Treatment with IPTG resulted in high level induction of the *lacZYA* and *melAB* operons. Following heat shock treatment 119 genes were shown to have significantly altered expression levels, including 35 previously uncharacterized ORFs and most genes of the heat shock stimulon. Analysis of spot intensities from hybridization to replicate arrays identified sets of genes with signals consistently above background suggesting that at least 25% of genes were expressed at detectable levels during growth in rich media.

INTRODUCTION

The complete genome sequences from more than 21 prokaryotes and two eukaryotes have been reported. Of the completed microbial genomes the best studied is the enteric, Gram-negative bacterium *Escherichia coli*. The 4.6 Mbp *E.coli* genome is predicted to encode 4290 open reading frames (ORFs) and at least 115 stable RNAs (1). Although determination of the sequence of every gene in an organism allows a better understanding of the organism's physiological potential, it is just the beginning of a complete description of how the cell works. Despite over 40 years of intensive study, more than 30% of the ORFs which make up the *E.coli* chromosome have no known function. This theme is reiterated in all completely sequenced genomes published to date (for example, 2–6) reinforcing the idea that knowledge of the nucleotide sequence of the complete genome is just the first step to understanding the dynamic nature of gene function that allows the cell to grow, replicate and respond to its environment.

To provide a means of expression analysis in *E.coli* at single gene resolution we have created high density arrays composed of full-length ORF-specific PCR products. This strategy of gene expression analysis was originated by Chuang *et al.* using λ clone spot blots to monitor gene expression in *E.coli* at 10 gene resolution (7). Brown and colleagues have subsequently miniaturized this method and have also brought it to single gene resolution (8,9). This basic approach for large scale expression monitoring has been used in a number of different formats including spotted cDNA microarrays (9–14) and oligonucleotide arrays (15–17). Microarrays of PCR fragments corresponding to most of the ORFs from *Saccharomyces cerevisiae* have been used to study transcriptional changes throughout the cell cycle, during sporulation and in response to a diauxic shift (12,13,18). Methods for global genome comparisons using *Mycobacterium tuberculosis* microarrays have also been reported (19,20).

The ability to determine changes in RNA levels simultaneously for all the genes in a cell is an extremely powerful tool. Regardless of the state of cell growth, one can measure the relative expression levels for each gene under various growth conditions, different genetic states or over a time course during environmental change. These types of studies have begun to identify new sets of genes involved in specific physiological responses. For example, DeRisi *et al.* identified 183 genes whose expression levels change by greater than 4-fold during the diauxic shift in yeast, about half of which have no previously determined function (12). Human cDNA microarrays have been used to study the transcriptional response of human fibroblasts to serum stimulation (21), revealing many similarities between serum stimulation and wound repair.

Analysis of global gene expression data has been difficult due to the large number of data points collected in a single experiment. Eisen *et al.* recently presented a clustering method that allows easy visualization of genes that may be co-regulated (22). Their method clusters genes based on similarities in expression patterns and graphically presents the data using dendograms. Comparison of global expression data from *S.cerevisiae* using their clustering method found that genes having similar function often cluster together, suggesting that ORFs of unknown function may be involved in similar physiological pathways as those genes of known function in the same cluster.

*To whom correspondence should be addressed. Tel: +1 608 262 2534; Fax: +1 608 263 7459; Email: craig@genetics.wisc.edu

To demonstrate the ability of our system to measure changes in gene expression on a global scale and to compare radioactive methods with fluorescence-based methods we have used two classic systems which affect gene expression in *E.coli*. Induction with isopropyl-β-D-thiogalactopyranoside (IPTG) provided a simple test of the methods since we expected only a few gene transcripts to change, while the effect of heat shock allowed global regulatory affects to be observed. Hybridization of genomic DNA probes to high density arrays allowed assessment of signal variation between spots due to factors such as size, amount of DNA spotted and cross-hybridization to members of gene families.

MATERIALS AND METHODS

PCR primer design

PCR primer pairs were designed to amplify each of the 4290 ORFs in the *E.coli* K-12 genome. Primers were designed to amplify each ORF beginning at the putative start codon and ending at the stop codon. All primer pairs used ATG as a start codon and TAA as a stop codon regardless of the start and stop codons found in the genomic sequence. Each primer contains a 13-base non-variable 'adaptamer' sequence at its 5' end followed by 20–25 bases of ORF-specific sequence. The length of the ORF-specific region of each primer was adjusted to achieve a melting temperature of 68–70°C. Adaptamer sequences were included to facilitate directional cloning in later stages of functional analysis (J.D.Glasner, C.S.Richmond, G.Plunkett III, S.Hinsa, B.Bochner and F.R.Blattner, manuscript in preparation). The adaptamer sequences of all N-terminal primers include a *Sap*I restriction endonuclease recognition site and have the sequence 5'-TTGCTCTTCC*ATG* . . .-3'. All C-terminal primers also include a *Sap*I site and have the adaptamer sequence 5'-TTG-CTCTTCG*TTA* . . .-3'. Primers were synthesized then arrayed in 96-well plates (Genosys Biotechnologies) for convenient parallel processing of amplification reactions. These primers are available from Sigma-Genosys Biotechnologies as a complete set or as ORF-specific primer pairs.

PCR amplification

Amplification reactions were performed in two rounds using 96-well plates. For the first round of PCR, 1 U of *Pfu* DNA polymerase (Stratagene) was used in a 25 µl reaction volume containing 20–30 ng *E.coli* MG1655 genomic DNA template, 0.5 µM each primer and 200 µM dNTPs. Reactions were cycled 25 times as follows: 95°C for 15 s, 64°C for 15 s, 72°C for 4 min, with a final cycle of 72°C for 5 min. To generate re-amplified PCR products used in production of high density arrays, the first round PCR products were diluted 500-fold and 1 µl used as template in 100 µl reactions containing 200 µM dNTPs, 2.5 U ExTaq (PanVera) and 0.5 µM each primer. Reactions were cycled 25 times as follows: 95°C for 15 s, 64°C for 15 s, 72°C for 2 min, with a final cycle of 72°C for 5 min.

All PCR products (2 µl of the total reaction volume) were analyzed by electrophoresis on 1.4% agarose gels in 0.5× TBE. Gels were stained with ethidium bromide and digitally imaged using a CCD camera. BioImage™ software was used to determine the size and purity of each PCR product. An amplification was scored as 'successful' if a single product was within 10% of the expected size of the ORF predicted from the genomic sequence.

Unsuccessful PCRs in first round amplification reactions were of three types: (i) no product observed by ethidium bromide staining (693); (ii) multiple products observed (149); and (iii) reactions resulting in single products of unexpected size (38). Reactions that failed were repeated using conditions expected to favor amplification of the single desired product. For example, those reactions yielding no product were amplified using lower annealing temperatures. Reactions that resulted in multiple products were amplified using a lower primer concentration and/or a higher annealing temperature. Approximately 97% of the ORFs in *E.coli* have been successfully amplified using this approach.

Bacterial growth and isolation of total RNA

For all experiments a single colony of *E.coli* K-12 (MG1655) was used to inoculate 60 ml of Luria-Bertini (LB) broth [1% Bacto-tryptone (Difco), 0.5% yeast extract (Difco) and 1% NaCl] in 250 ml Erlenmeyer flasks and grown to an OD A_{600} between 0.8 and 0.9 at 37°C with constant aeration. IPTG treatment was performed by splitting a 60 ml culture in half and adding IPTG (final concentration of 1 mM) to one 30 ml sample (experimental) while the other sample (control) was untreated. Incubation was continued at 37°C for 30 min prior to harvesting cells. Heat shock induction was carried out by splitting a mid-log culture into two 30 ml samples with the control culture kept at 37°C for 7 min while the experimental culture was shifted to a 50°C shaking water bath for 7 min.

Total RNA was isolated from cells using Qiagen RNeasy columns. Cultures were pelleted by brief centrifugation at 5000 *g* and cell pellets flash-frozen on dry ice/ethanol. Cells were resuspended by vortexing in 1 ml of TE (10 mM Tris pH 8.0, 1 mM EDTA) containing 500 µg/ml lysozyme (Boehringer Mannheim) and RNA isolation proceeded following the manufacturer's protocol. To remove contaminating genomic DNA from purified RNA, samples were treated with RQ1 RNAse-free DNAse (Promega) followed by Proteinase K digestion, phenol–chloroform extraction and precipitation with ethanol. Pelleted RNA samples were resuspended in diethyl pyrocarbonate (DEPC)-treated H_2O, quantitated by absorbance at 260 nm, and stored in ethanol at –20°C until further use. For time-course studies cells were grown in rich, defined media (23) to an OD A_{600} of 0.8 and 30 ml samples taken at time 0, 5, 10, 15 and 20 min following transfer from 37 to 50°C. RNA was isolated as described above.

Genomic DNA labeling

Genomic DNA isolated from *E.coli* K-12 was labeled radioactively using nick translation, or fluorescently labeled by random priming using the Klenow fragment of DNA polymerase. Genomic DNA was purified from MG1655 using standard protocols (24) and fragmented by sonication to an average length of 1500 bp. Each radioactive labeling reaction contained 500 ng of fragmented genomic DNA, 0.02 mM nucleotide mix (dGTP, dTTP, dATP), 50 µCi ^{33}P-labeled dCTP, 50 mM Tris pH 7.9, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10 µg/ml BSA and 1.5 U of DNA polymerase I/DNAse I (Life Technologies, Inc.). Reactions were carried out at 15°C for 1 h and stopped by addition of EDTA to 0.05 M. Labeled genomic DNA was purified from unincorporated nucleotides by gel filtration on

Sephadex G-50 columns. For each hybridization 2×10^6 c.p.m./ml of probe was used.

Fluorescence labeling reactions for microarray analysis contained 1 μg of fragmented genomic DNA, 5 μg of random hexamers, 0.5 mM nucleotide mix (dGTP, dATP, dCTP), 0.1 mM Cy3 or Cy5 labeled dUTP (Amersham-Pharmacia), 50 mM Tris–HCl pH 7.4, 10 mM $MgCl_2$, 50 μg/ml BSA, 3 mM DTT and 5 U Klenow (Promega). Reactions were carried out at 25°C for 2 h and stopped by addition of 25 mM EDTA, 0.25% SDS. Labeled DNA was purified and concentrated prior to hybridization using Microcon 30 concentrators (Amicon).

Labeled cDNA synthesis

Random primed cDNA synthesis of total *E.coli* RNA was used to prepare ^{33}P-labeled and fluorescently-labeled probes for array hybridization. For radioactive labeling 5 μg of total RNA was pelleted from ethanol, washed once with 70% ethanol and briefly dried prior to resuspension in 12 μl DEPC-treated H_2O containing 250 ng of random hexamer oligonucleotides (Amersham-Pharmacia). Samples were heated to 70°C for 10 min and chilled on ice prior to probe synthesis. Probe synthesis was carried out at 42°C for 2 h in a 50 μl reaction volume containing 50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.5 mM dATP, 0.5 mM dTTP, 0.5 mM dGTP, 0.05 mM dCTP, 100 μCi ^{33}P-labeled dCTP (1–3000 Ci/mmol; Amersham), 40 U RNAsin (Promega) and 200 U Superscript II reverse transcriptase (Life Technologies, Inc.). Following cDNA reactions RNA template was degraded by incubating for 30 min at 65°C in 0.27 M NaOH and 20 mM EDTA followed by neutralization with HCl and Tris buffer. Unincorporated nucleotides were removed by Sephadex G50 gel filtration chromatography (24).

Fluorescence cDNA labeling was performed essentially as described above. Random hexamers (10 μg) and RNA (20 μg) were mixed in a final volume of 12 μl and treated as above. Probe synthesis was performed at 42°C for 2 h in a 50 μl reaction volume containing 50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.5 mM dATP, 0.5 mM dGTP, 0.5 mM dCTP, 0.2 mM dTTP, 0.1 mM Cy dye-labeled dUTP (Amersham), 40 U RNAsin (Promega) and 200 U Superscript II reverse transcriptase (Life Technologies, Inc.). For all experiments Cy3-dUTP was used to label control samples and Cy5-dUTP used to label experimental samples. Following cDNA reactions the RNA template was degraded by incubating for 30 min at 65°C in 0.27 M NaOH followed by neutralization with HCl and Tris buffer. Labeled cDNA was purified and concentrated prior to hybridization using Microcon 30 concentrators (Amicon).

Preparation of high density arrays

Spot blots of ORF-specific PCR products and controls were printed on 23 × 24 cm nylon membranes (GeneScreen Plus, New England Nuclear) in a 64 × 68 hexagonal grid using a Gilson Model 215 liquid handling robot. Before printing, 2 μl of each re-amplified PCR was diluted with 1 μl of an alkaline denaturing solution (0.75 M NaOH, 30 mM EDTA) and directly spotted on pre-wet membranes. Immediately following spot deposition, membranes were neutralized for 20 min in 20 mM Tris (pH 7.0) and baked for 45 min at 80°C. Membranes were stored dry at room temperature or used immediately for hybridization. Control spots on all arrays include genomic DNA and λ DNA. As negative controls, mock amplification reactions and amplification products from several yeast ORFs having no significant match to *E.coli* genes were used. Whole genome spot blots printed in a similar fashion to those described here are currently available from Sigma-Genosys Biotechnologies.

Microarrays were prepared as described previously (12) using re-amplified PCR products (60 μl reactions) that were precipitated with isopropanol, pelleted and resuspended in 15 μl of 3× SSC prior to spotting.

Radioactive hybridization and raw data analysis

Prior to hybridization, high density arrays were pre-wet in 2× SSC, 0.5% SDS and pre-hybridized for 2 h at 64°C in roller bottles containing 15 ml of 1× hybridization buffer (0.9 M NaCl, 90 mM sodium citrate pH 7.0, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% BSA, 0.5% SDS and 100 μg/ml sheared salmon sperm DNA). Pre-hybridization buffer was removed and replaced with 1× hybridization buffer containing 2×10^6 c.p.m./ml cDNA probe and hybridization continued at 64°C for 40 h. Following hybridization, arrays were washed twice at room temperature and once at 65°C in 0.3 M NaCl, 30 mM sodium citrate, 0.5% SDS followed by one wash at 65°C in 15 mM NaCl, 1.5 mM sodium citrate, 0.5% SDS. Arrays were then sealed in thin polypropylene bags to avoid drying and exposed to a phosphor screen (Molecular Dynamics) for 96 h. Phosphor screens were scanned using a STORM phosphorimaging instrument (Molecular Dynamics) at 100 micron pixel resolution. Images were stored electronically and analyzed using ImageQuant v.4.1 analysis software (Molecular Dynamics). Between successive hybridizations, membranes were stripped by two 30 min incubations in 1% SDS at 100°C and dried for 30 min at 65°C.

To calculate signal intensities a grid of ellipses was drawn and overlaid on the array image and signal intensities for each spot calculated using the volume quantitation method of ImageQuant. Total intensity of all pixels within each ellipse was determined and data saved in spreadsheets. Background correction was achieved by sampling regions outside the array grid and averaging their volumes.

We calculate expression ratios as follows: after correction for background, the percentage of total signal is calculated for each spot as a means of normalization. These values are used to determine the ratio of experimental to control signal. To allow easy comparison of induction and repression ratios, signals that were higher under the control condition (transcription repressed) were used in the numerator of the ratio calculation and then converted to negative values. A threshold of minimum acceptable signal was used to eliminate expression ratios that were extremely high or low due to undetectable signal in control or experimental samples. To determine consistency of ratios across replicate hybridizations a *t*-test was applied. We report only those ratios with values greater than or equal to ±5 and having a 95% confidence interval as determined by the *t*-test. The entire data sets are available at http://www.genetics.wisc.edu

Microarray hybridization and raw data analysis

Escherichia coli microarrays were hybridized in a final volume of 13 μl containing 3× SSC, 0.8 mg/ml salmon sperm DNA and 0.2% SDS. Prior to hybridization samples were heated to 95°C for 2 min and pipetted directly onto microarrays. A cover slip was applied and the arrays hybridized overnight at 64°C in a humidified hybridization chamber. Following hybridization,

slides were washed for 3 min in 0.2× SSC, 0.1% SDS, followed by two washes of 2 min each in 0.2× SSC and rinsed in 0.05× SSC. To remove residual salts, slides were spun at 500 r.p.m. for 5 min prior to scanning on a ScanArray 3000 confocal laser scanner (GSI-Lumonics). Signal intensities for each spot were determined using ScanAlyze software (available at http://rana.stanford.edu/software/). The average fluorescence intensity for each spot was calculated and local background determined as the median pixel intensity in a square surrounding each spot. Following background subtraction, signal intensities were calculated as the percent of total signal as a means of normalization. Ratios and *t*-tests were performed as described above.

RESULTS

Creation of PCR gene set

To begin systematic functional and expression analysis in *E.coli* K-12 a complete set of ORF-specific PCR primers was created. The primer pairs were designed to amplify the complete coding sequences from start to stop codon and included a 13-base, non-variable adaptamer sequence. A high fidelity PCR system was chosen to obtain PCR products for downstream cloning and expression studies. We used *Pfu* polymerase which has 3'→5' proofreading activity and a low error rate (25–27). All PCR products were analyzed by agarose gel electrophoresis and digital image analysis. Amplification of all 4290 ORFs with *Pfu* polymerase was followed by a re-amplification step designed to minimize carry-over of genomic DNA template when printing high density arrays. The re-amplification step used a 500-fold dilution of the *Pfu* PCR product as template and a *Taq*/*Pfu* thermostable DNA polymerase mix to improve yield of large fragments (28). Using this two-step method we have successfully amplified 97% of the 4290 *E.coli* ORFs using the original set of PCR primers. Products printed on nylon membranes represent >90% of the *E.coli* genome while those printed on microarrays represent 95% of the genome. Those ORFs that failed to amplify successfully were eliminated from further analysis during expression profiling.

High density array analysis and validation

To ensure that DNA samples were successfully deposited on nylon membranes and to assess differential hybridization to target genes, we hybridized ^{33}P-labeled MG1655 genomic DNA to each spot blot. The total intensity of all pixels within each spot (corrected for background signal) was determined after scanning of exposed phosphor screens. In a few cases (32 spots) signals close to average were detected for ORFs that were scored as unamplified following electrophoresis of PCR products. These products were likely produced at levels below the limit of ethidium bromide detection. Only nine spots, scored as successful PCRs, were consistently scored as undetectable (defined as having a signal strength less than three standard deviations above the average background) in all hybridizations.

Figure 1 (upper panel) shows the distribution of ^{33}P-labeled genomic DNA hybridization signal intensities as a function of ORF length. We noticed that small genes (<300 bp) often showed lower intensity signals even when the PCR product was abundant. For the largest genes, the hybridization signal also tended to be lower than average. Inspection of the PCR

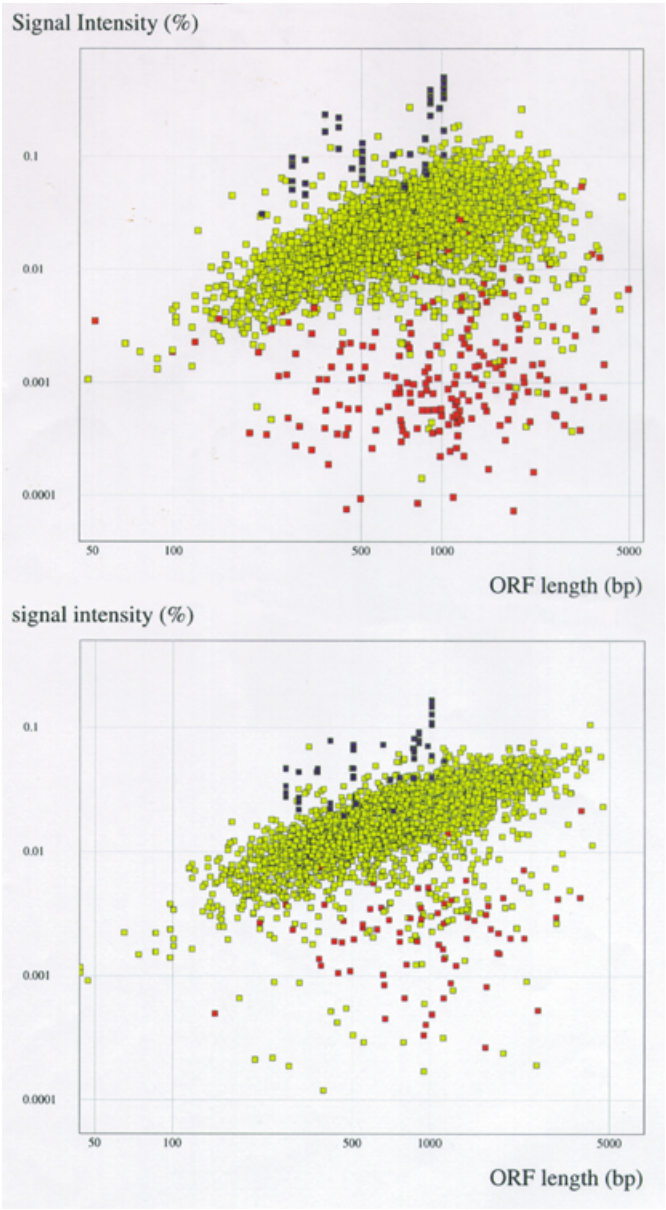


Figure 1. Scatter plot showing the relationship between signal intensities from genomic DNA hybridizations and ORF length. (**Upper panel**) Radioactive hybridization of ^{33}P-labeled *E.coli* genomic DNA to spot blot arrays. Average percent spot intensity values from three hybridizations are plotted against ORF length in bp. (**Lower panel**) Same as upper panel using microarrays and fluorescently labeled genomic DNA. The average percent intensities from two hybridizations are plotted against ORF length in bp. Paralogous ORFs are highlighted in blue and failed PCR products in red.

electrophoresis results reveals that most ORFs >2 kb were amplified less efficiently than smaller genes.

We observed 61 genes with genomic DNA hybridization signals much higher than average (Fig. 1). Since the majority of these spots (45 spots, marked blue in Fig. 1) correspond to ORFs that are members of gene families (paralogs) with high degrees of sequence identity (>70% nucleotide identity over >200 bp) we conclude that the effect is likely due to cross-hybridization between paralogs. There is a general trend for genes with greater numbers of paralogs to have higher signal

Table 1. Genes affected by IPTG

Bnumber[a]	Name[b]	Radioactive[c]	Microarray[d]	Gene product[e]
b0342	lacA	79	43	Thiogalactoside acetyltransferase
b0344	lacZ	62	49	β-D-galactosidase
b0343	lacY	55	86	Galactoside permease
b4119	melA	8	9	α-Galactosidase
b4120	melB	6	7	Melibiose permease II
b3076	ebgA	6	1	Evolved β-D-galactosidase, α subunit
b1441	b1441	6	1	Putative transport; not classified
b1297	b1297	6	1	Putative enzyme; not classified
b0767	b0767	–5	1	ORF; not classified

[a]Unique identifier for *E.coli* genes.
[b]Name of gene if known.
[c]Average ratio calculated from spot blot analysis.
[d]Average ratio calculated from microarray analysis.
[e]Description of gene (if known) as reported by Riley *et al*. (60).

strengths although their intensity also varies depending on the length and degree of similarity between the paralogous sequences. ORFs with sequences that align over >200 bp with >80% sequence identity show the strongest intensity signals. This result is also seen using DNA microarrays (Fig. 1, lower panel). Genomic DNA was labeled separately with Cy3 and Cy5 dUTP, the probes were mixed and hybridized to a single microarray followed by laser scanning and quantification. As with the radioactive hybridizations, signal intensities were higher than average for paralogous ORFs. We also find, as with radioactive methods, that small ORFs tend to have weaker signals than do larger ORFs.

IPTG expression profile

The *lac* operon (*lacZYA*) is one of the best-studied operons in *E.coli*. Addition of lactose or IPTG to *E.coli* cultures in exponential phase growth results in transcriptional induction of the *lac* genes caused by release of LacI-mediated transcriptional repression. Table 1 shows all genes found to be significantly affected by 30 min treatment with 1 mM IPTG as measured by radioactive methods. As expected, genes of the *lac* operon are the most highly induced transcripts. Induction ratios for *lacZ* and *lacY* were near 60-fold while *lacA* had an induction ratio of 79-fold. This result confirms the ability of this method to measure specific changes in gene expression using a complex mixture of labeled cDNA. In addition, *melA* and *melB*, which make up the melibiose operon, are induced 8- and 6-fold, respectively. This is consistent with previous data from our laboratory (7). The *melA* gene encodes α-galactosidase and *melB* encodes a melibiose transport system permease II, both of which are induced by melibiose under the positive control of MelR (29). It is interesting to note that melibiose is also able to induce the *lac* operon (30) although there have been no previous reports of IPTG positively affecting MelR.

Radioactive methods also identified the cryptic β-D-galactosidase gene, *ebgA*, and two ORFs of unknown function, b1441 and b1297, as induced by IPTG. The *ebgA* gene is part of the *ebgRAC* operon thought to have evolved from the *lac* operon (31). We did not, however, see significant induction of *ebgC*, which is thought to be co-transcribed with *ebgA*. Comparison of the *ebgA* and *lacZ* DNA sequences using the Wilbur–Lipman alignment method (32) shows ~50% sequence identity over >2000 bp. To test if this level of sequence similarity was enough to produce cross-hybridization artifacts using radioactive hybridization methods, *in vitro* transcribed RNA from the *lacZ* gene was radioactively labeled using reverse transcriptase and hybridized in the presence of unlabeled total cDNA to an array of 200 spots including *lacZ* and *ebgA*. The results confirm that cross-hybridization may have led to the observed 'induction' of *ebgA* expression (data not shown).

We compared results from radioactive methods to those using *E.coli* microarrays (Fig. 4E and F). Control and experimental RNA samples from the same preparations used for radioactive studies were fluorescently labeled and hybridized in parallel to a single microarray. The averaged results from two hybridizations were compared to those from radioactive hybridizations. Using microarrays, genes of the *lac* and *mel* operons were induced to levels similar to those obtained with radioactive methods (Table 1); however, no significant effect was observed for the two unknown ORFs (b1297 and b1441) or *ebgA*. This suggests that hybridization and wash conditions were stringent enough to avoid the cross-hybridization artifact or that the sensitivity of microarrays is not high enough to detect cross-hybridization of *ebgA*. However, Heller *et al*. report that short regions of sequence identity over the length of the target sequence do result in cross-hybridization artifacts using cDNA microarrays, although the degree of sequence identity for these cross-hybridizing species was not reported (33).

Heat shock response expression profile

The heat shock response is well studied (34–37) and is conserved among many different organisms (36,38). Early studies of heat shock in *E.coli* used two-dimensional gel electrophoresis to identify proteins whose expression was induced due to increased temperature (39–41). More recently, experiments

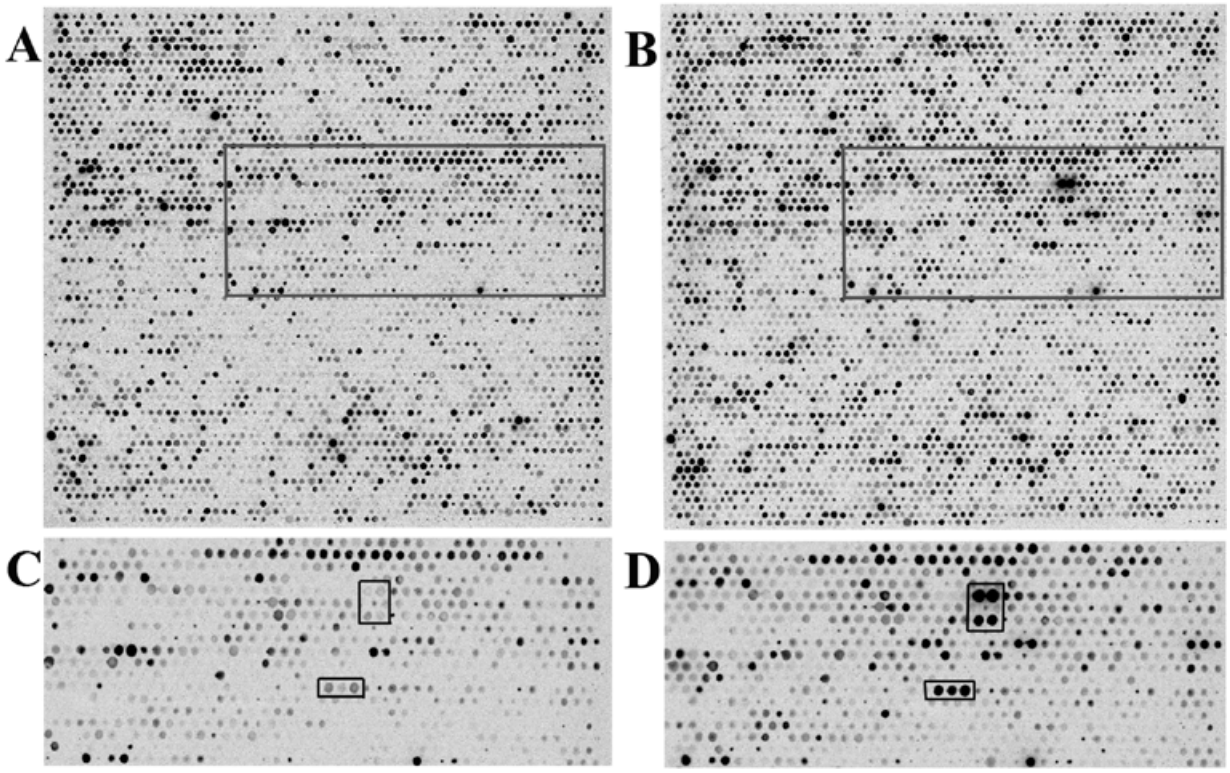


Figure 2. *Escherichia coli* spot blot of heat shock induction. (**A**) *Escherichia coli* array hybridized with ^{33}P-labeled cDNA from RNA isolated from MG1655 grown at 37°C. (**B**) Same blot as in (A) following stripping and re-hybridization with cDNA prepared from RNA isolated from a culture that was heat shock treated at 50°C for 7 min. (**C**) and (**D**) are enlargements from boxed regions in (A) and (B) respectively. Spots indicated by boxes in (C) and (D): upper box includes *ibpAB* and *hslUV*, lower box is *miaA-hfq-hflX*.

with λ clone spot blots have identified additional members of the heat shock stimulon (7).

The heat shock response in *E.coli* is controlled at the transcriptional level by the alternative σ factors *rpoH* (σ32) (42) and *rpoE* (σE) (43,44). At least 51 loci have been identified as members of the heat shock stimulon in *E.coli* (reviewed in 35) although the specific gene sequence affected is known for only 30 of these loci. The remaining 21 loci, although named as heat shock inducible genes in the literature, are as yet functionally or genetically uncharacterized. Heat shock proteins where the biochemical activity of the gene product is known include proteases, chaperones, tRNA synthetases and σ factors.

We used spot blots to identify the genes transcriptionally regulated following a temperature shift from 37 to 50°C using radioactive cDNA probes prepared from total *E.coli* RNA (Figs 2 and 4G). Most genes are apparently unaffected while a specific subset of genes show increased or decreased expression following heat shock treatment (for example, see Fig. 2C and D, boxed regions). This experiment identified 77 genes that are induced and 42 genes repressed by heat shock (Table 2). For 35 ORFs of unknown function, this study provides the first biochemical evidence for their expression and biological role, although for 12 of these ORFs a putative function was previously assigned based on similarities to sequences of known function (1). The up-regulated genes include 23 previously identified members of the heat shock stimulon (Table 2), with an additional three known heat shock genes showing induction just below our cutoff values.

The remaining 19 genes with known function have not previously been described as heat shock inducible. Some of these genes are known to be affected in response to other stress conditions, such as the *cadAB* operon, induced in response to low pH or high lysine concentrations (45); and *cspD*, induced under conditions of nutritional stress (46). Another group encode proteins with functions related to heat shock genes: *rseA*, a negative regulator of σ E (47,48); *clpA*, the ATPase component of the ClpAP protease (49,50) [*clpP* is a heat shock gene (51)]; and *prlC*, a trypsin-like proteinase in *E.coli* (52,53).

Nine of the induced ORFs are members of paralogous gene families. For example, *marA* has a stretch of 369 bp that are 51% identical to *rob*. However, although the *marA* signal was induced 6-fold, *rob* was not, confirming that cross-hybridization at this level of sequence similarity was not significant. Spots corresponding to IS186 elements were the only spots showing evidence for cross-hybridization in this experiment with all members induced to similar levels.

Of the 42 genes repressed, all but six have known biochemical activities in a number of functional categories. Sixteen are involved in purine or pyrimidine ribonucleotide biosynthesis while four are involved in chemotaxis or motility. One very interesting member of this group is the *glyA* gene which encodes serine hydroxymethyltransferase and is responsible

Table 2. Genes significantly affected by heat shock

Bnumber[a]	Name[b]	Radioactive[c]	Microarray[d]	Gene product[e]
b3686	*ibpB	393.4	327.2	Inclusion body protein B; heat shock protein
b3687	*ibpA	288.1	297.4	Inclusion body protein A; heat shock protein
b2592	*clpB	102.2	36.5	Heat shock protein
b0473	*htpG	74.6	33.8	Chaperone Hsp90, heat shock protein
b3400	yrfH	42.9	51.3	ORF, hypothetical protein
b4131	cadA	41.6	4.4	Lysine decarboxylase 1
b4143	*mopA	40.0	37.9	GroEL, chaperone Hsp60
b4142	*mopB	34.3	77.5	GroES, 10 kDa chaperone binds to Hsp60
b3932	*hslV	31.5	16.2	Heat shock protein hslVU, proteasome-related peptidase subunit
b1967	yedU	31.3	30.5	ORF, hypothetical protein
b3401	yrfI	28.8	21.6	ORF, hypothetical protein
b0014	*dnaK	25.2	58.5	Chaperone Hsp70; DNA biosynthesis
b4171	*miaA	23.4	11.9	Δ(2)-Isopentenylpyrophosphate tRNA-adenosine transferase
b4140	b4140	22.5	50.7	ORF, hypothetical protein
b3179	*ftsJ	20.2	9.1	Cell division protein
b3498	prlC	19.7	16.7	Oligopeptidase A
b0582	yi81_2	19.0	27.3	IS186 hypothetical protein
b3399	yrfG	17.9	12.1	Putative phosphatase
b2394	yi81_3	16.4	29.6	IS186 hypothetical protein
b3931	*hslU	16.1	10.3	Heat shock protein hslVU, ATPase subunit
b0015	*dnaJ	15.8	85.3	Chaperone with DnaK; heat shock protein
b3816	corA	15.6	1.0	Mg^{2+} transport, system I
b2614	*grpE	15.0	24.1	Phage λ replication; host DNA synthesis; heat shock protein
b0016	yi81_1	14.1	32.9	IS186 hypothetical protein
b0399	phoB	13.4	28.2	Positive response regulator for pho regulon, sensor is PhoR
b3293	yhdN	13.3	9.5	ORF, hypothetical protein
b0882	clpA	13.2	3.8	ATP-binding component of serine protease
b0017	yi82_1	12.5	22.8	IS186 and IS421 hypothetical protein
b0374	yaiU	12.2	2.7	Putative flagellin structural protein
b3343	yheL	12.0	22.0	ORF, hypothetical protein
b2572	*rseA	10.5	1.0	σE factor, negative regulatory protein
b0966	yccV	10.3	34.3	ORF, hypothetical protein
b0879	ybjZ	10.2	5.2	Putative ATP-binding component of a transport system
b0881	yljA	10.2	6.5	ORF, hypothetical protein
b0400	phoR	10.1	34.5	Positive and negative sensor protein for pho regulon
b4208	cycA	9.9	1.0	Transport of D-alanine, D-serine and glycine
b4141	yjeH	9.8	5.1	Putative transport
b4172	*hfq	9.3	4.5	Host factor I for bacteriophage Q β replication
b0492	ybbN	9.1	9.9	Putative thioredoxin-like protein
b0439	*lon	8.9	20.3	DNA-binding, ATP-dependent protease La; heat shock K-protein
b0316	yahB	8.8	4.3	Putative transcriptional regulator LYSR-type
b4209	ytfE	8.8	13.7	ORF, hypothetical protein
b3022	b3022	8.7	6.7	ORF, hypothetical protein
b0491	ybbM	8.7	2.0	Putative metal resistance protein
b1829	*htpX	8.1	36.1	Heat shock protein, integral membrane protein
b3402	yhgE	8.0	2.4	Putative transport
b1531	marA	7.9	6.2	Multiple antibiotic resistance
b0437	*clpP	7.9	3.3	ATP-dependent proteolytic subunit of clpA–clpP serine protease
b3067	*rpoD	7.9	7.7	RNA polymerase, σ 70

Table 2. *Continued*

b0315	yahA	7.7	28.9	ORF, hypothetical protein
b0210	yafE	7.7	12.2	Putative biotin synthesis protein
b1112	ycfR	7.6	12.2	ORF, hypothetical protein
b0880	cspD	7.5	5.5	Cold-shock protein
b3669	uhpA	7.4	–1.0	Response regulator, positive activator of uhpT transcription
b3685	yidE	7.3	3.9	Putative transport protein
b4132	cadB	7.3	2.0	Transport of lysine/cadaverine
b0209	yafD	7.2	6.7	ORF, hypothetical protein
b1874	cutC	6.8	3.2	Copper homeostasis protein
b1322	ycjF	6.8	19.8	ORF, hypothetical protein
b2573	*rpoE	6.5	1.0	RNA polymerase, σE factor; heat shock and oxidative stress
b1530	marR	6.3	8.5	Multiple antibiotic resistance protein; repressor of *mar* operon
b1060	yceP	6.3	25.5	ORF, hypothetical protein
b2796	sdaC	6.2	2.2	Probable serine transporter
b0438	*clpX	6.2	2.9	ATP-dependent specificity component of clpP serine protease
b2613	yfjD	6.2	9.0	Putative transport protein
b3413	yhgH	6.1	1.9	ORF, hypothetical protein
b0660	ybeZ	5.9	9.6	Putative ATP-binding protein in pho regulon
b4173	*hflX	5.9	3.3	Subunit of protease specific for phage λ cII repressor
b4398	creB	5.9	3.4	Catabolic regulation response regulator
b2193	narP	5.8	4.2	Nitrate/nitrite response regulator (sensor NarQ)
b1593	b1593	5.7	–1.4	ORF, hypothetical protein
b0659	ybeY	5.6	5.8	ORF, hypothetical protein
b0281	intF	5.5	2.8	Putative phage integrase
b1274	*topA	5.5	5.9	DNA topoisomerase type I, Ω protein
b3635	mutM	5.4	12.2	Formamidopyrimidine DNA glycosylase
b3280	yrdB	5.3	1.3	ORF, hypothetical protein
b4239	treC	5.2	5.8	Trehalase 6-P hydrolase
b2779	eno	–5.0	–6.7	Enolase
b1887	cheW	–5.1	–6.0	Positive regulator of CheA protein activity
b2286	nuoC	–5.2	–10.4	NADH dehydrogenase I chain C, D
b1782	yeaF	–5.4	–2.8	ORF, hypothetical protein
b0572	ylcB	–5.5	1.4	Putative resistance protein
b2945	endA	–5.5	–1.2	DNA-specific endonuclease I
b0411	tsx	–5.6	–2.2	Nucleoside channel; receptor of phage T6 and colicin K
b2313	cvpA	–5.7	1.8	Membrane protein required for colicin V production
b1903	b1903	–6.2	–1.0	ORF, hypothetical protein
b1132	ycfC	–6.5	–1.2	ORF, hypothetical protein
b1076	flgE	–6.6	–22.7	Flagellar biosynthesis, hook protein
b0523	purE	–6.6	1.9	Phosphoribosylaminoimidazole carboxylase, catalytic subunit
b2508	guaB	–6.8	–1.8	IMP dehydrogenase
b4117	adiA	–6.9	2.9	Biodegradative arginine decarboxylase
b0973	hyaB	–7.0	1.0	Hydrogenase-1 large subunit
b2476	purC	–7.1	2.2	Phosphoribosylaminoimidazole-succinocarboxamide synthetase
b2297	pta	–7.3	–6.0	Phosphotransacetylase
b1888	cheA	–8.0	–3.3	Sensory transducer kinase
b4115	yjdE	–8.0	2.5	Putative amino acid/amine transport protein, cryptic
b1676	pykF	–8.3	–9.2	Pyruvate kinase I (formerly F), fructose stimulated
b2497	uraA	–8.5	2.2	Uracil transport
b3114	tdcE	–8.6	1.2	Probable formate acetyltransferase 3

Table 2. *Continued*

b1924	fliD	–9.0	–4.8	Flagellar biosynthesis; filament capping protein
b0116	lpdA	–9.0	–1.1	Lipoamide dehydrogenase (NADH)
b0903	pflB	–9.0	–10.0	Formate acetyltransferase 1
b2925	fba	–9.4	–12.5	Fructose-bisphosphate aldolase, class II
b3512	yhiE	–9.4	1.3	ORF, hypothetical protein
b2507	guaA	–10.4	–3.3	GMP synthetase (glutamine-hydrolyzing)
b0033	carB	–10.7	3.2	Carbamoyl-phosphate synthase large subunit
b0522	purK	–10.8	2.0	Phosphoribosylaminoimidazole carboxylase
b2498	upp	–11.8	–1.0	Uracil phosphoribosyltransferase
b0337	codA	–12.0	–1.0	Cytosine deaminase
b2500	purN	–13.2	1.5	Phosphoribosylglycinamide formyltransferase 1
b1131	purB	–13.4	–2.3	Adenylosuccinate lyase
b0336	codB	–15.2	2.1	Cytosine permease/transport
b2551	glyA	–15.9	–8.1	Serine hydroxymethyltransferase
b1062	pyrC	–19.8	–1.0	Dihydro-orotase
b0032	carA	–23.3	2.7	Carbamoyl-phosphate synthetase, glutamine (small) subunit
b0945	pyrD	–24.1	–1.1	Dihydro-orotate dehydrogenase
b4246	pyrL	–26.7	3.9	*PyrBI* operon leader peptide
b4244	pyrI	–73.6	1.3	Aspartate carbamoyltransferase, regulatory subunit
b4245	pyrB	–150.1	3.1	Aspartate carbamoyltransferase, catalytic subunit

[a]Bnumber; unique identifier for *E.coli* genes.
[b]Name of gene if known.
[c]Radioactive; average ratios from four independent radioactive hybridizations to spot blots.
[d]Microarray; average ratios from two independent fluorescent hybridizations to microarrays.
[e]Description of gene (if known) as reported by Riley *et al.* (60).
*Members of the known heat shock stimulon.

for converting serine into glycine with the formation of one-carbon units. We have not been able to reproduce repression of purine and pyrimidine biosynthetic genes due to heat shock using a variety of growth conditions.

In previous experiments using spot blots prepared from λ clones, 26 new heat shock genes were found on 12 different clones (7,34). Due to the low resolution of this method many loci were simply given a generic name (hslA–Z). Ten of the 12 λ clones identified in the previous study encode genes shown in this report to be significantly induced by heat shock (Table 3). The criteria for our tentative correlations are as follows: if only one ORF, identified as induced in the present study, was present on a λ clone, the ORF identified here was given that clone's hsl designation. If λ clones were shown previously to have multiple genes induced by heat shock, the corresponding ORFs were compared based on predicted protein size to determine correspondence. In some cases the hsl genes have been subsequently characterized as heat shock inducible [*miaA* locus (*hslY* and Z) (54), *ftsJ* (*hslN*), *hflB* (*hslL*) (55)] or were recently named based on biochemical characterization [*clpYQ* (*hslUV*) (56)].

We also compared heat shock data from radioactive spot blot analysis to those collected using microarrays (Figs 3 and 4; Table 2). cDNA probes were prepared from RNA isolated at 5 min intervals following temperature shift from 37 to 50°C. Fluorescent probes were hybridized to *E.coli* microarrays comparing time 0 to each of the four subsequent time points. Comparing data from the first 5 min time point using microarrays to data from radioactive hybridizations we see that the expression profiles are comparable for both (Fig. 4G and H). The majority of ORFs identified as induced by radioactive methods are also detected by microarray analysis (Table 2). With microarrays, 62 of the 76 genes identified by radioactive methods were seen to be induced to levels of 3-fold or greater. Using a 5-fold cutoff for microarray data identifies 45 of the 76 genes. We also found an additional subset of genes to be induced to levels greater than 5-fold using microarray methods. In general, we found that genes significantly induced by 5 min remained high throughout the time course although a few genes appeared to be temporally regulated (data not shown).

Transcriptome analysis

Most approaches to global transcription analysis rely on generating ratios of signal intensities between control and experimental samples. The ratio between signals for a particular spot provides a robust measure of change in expression level. However, experimental conditions often result in only a small number of genes with altered levels of expression. Induction and repression ratios identify dramatic changes in transcript abundance but ignore the variations in signal intensities between spots that do not significantly change. To a large degree, these variations in signal reflect the absolute abundance of different transcripts in the cell, but one should be cautious when interpreting absolute signal intensities. Two different genes on an array may have

Table 3. Correlation of heat shock locus (hsl) ORFs with specific genes

hsl gene	λ clone	Protein size (kDa)[a]	Corresponding gene product	Fold induction	Predicted mol. wt
hslA	148	65	nd	nd	nd
hslB	148	46	clpX	6.2	46
hslC	212	80	clpA	13	84
hslD	232/233	nd	b1060[b]	6.3	9.7
hslE	260	60	nd	nd	nd
hslF	260	51	b1321	3.9[c]	52.6
hslG	260	41	b1322[d]	6.7	39.4
hslH	260	39	b1322[d]	6.7	39.4
hslI	265	36	ldhA	3.4[c]	36.5
hslJ	265	14	hslJ[e]	2.7[c]	15
hslK	334	49	nd	nd	nd
hslL	520	70	hflB	4.4[c]	70.7
hslM	520	31	nd	nd	nd
hslN	520	27	ftsJ	20	23
hslO	620	33	b3401	29	32.8
hslP	620	30	b3399	18	27
hslQ	620	24	b3413[f]	6	27.7
hslR	620	18	b3400	43	15.5
hslS	566/567	nd	ibpB[e]	390	16.3
hslT	566/567	nd	ibpA[e]	288	15.8
hslU	538/539	nd	clpY[e]	16	49
hslV	538/539	nd	clpQ[e]	31	21
hslW	648/649	22	b4140	22	14
hslX	652	51	nd	nd	nd
hslY	652	45	hflX[g]	5.9	48
hslZ	652	37	miaA[g]	23	35

nd, not determined.
[a]From Chuang and Blattner (34).
[b]Located between *htrB* and *pyrC* as determined by Southern blotting (34).
[c]Induction ratio is below 5-fold cutoff.
[d]b1322 is very close in size to both *hslG* and *hslH* and therefore exact assignment not possible.
[e]Determined previously.
[f]Clone 621 does not express 24 kDa protein (34). b3413 is only found in clone 620 and not 621.
[g]Previously shown to be heat shock induced (54).

different signals because their transcripts are present at different levels or because they differ in size, base composition, or concentration of DNA in the spot. Multiple replicates of an experiment are necessary before we can be confident that the signal derives from a detectable level of RNA resulting from gene transcription. For example, in four replicates of the heat shock experiment we observed 1023 genes expressed above background level in four out of four replicates of the control sample. Likewise, for 1158 genes we failed to detect a signal above background in any replicate. Of the 1158 genes that were not detected, 75% are of unknown function; in contrast, genes of unknown function constitute only 37% of the genes that are always detected. The number of genes with significant signals in only one or two experiments is large, indicating the necessity to repeat experiments to definitively determine if a gene is expressed or not. In fact the proportion of expressed genes is likely to be an underestimate. In these experiments we required that the hybridization signal be above three standard deviations of the background signal in all trials. In a few individual trials the variation in background signal was substantially greater than others leading to exclusion of a large number of genes. Taking these factors into account we observe that at least 25% of the *E.coli* ORFs were expressed at detectable levels in batch culture at 37°C.

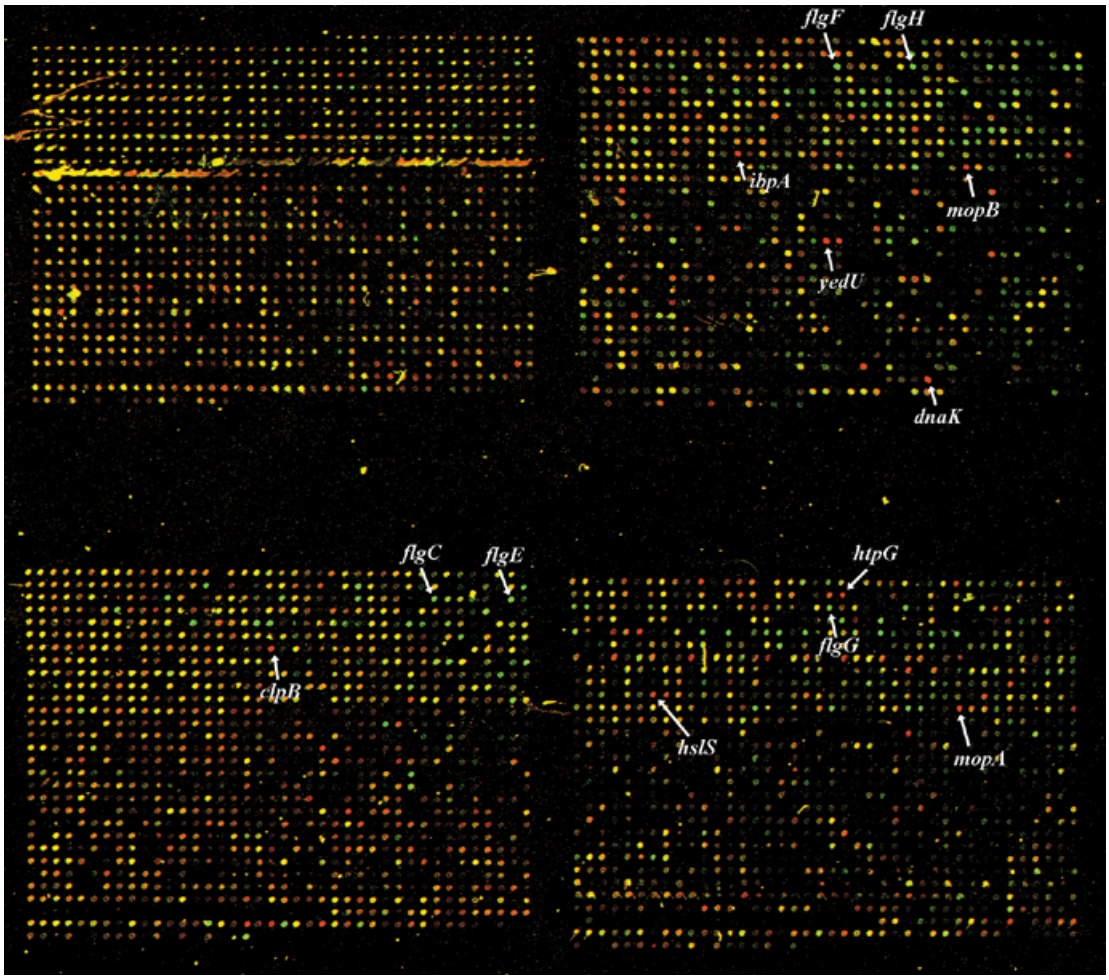


Figure 3. *Escherichia coli* genome microarray. The microarray was hybridized with fluorescently labeled cDNA (Cy3 labeled control and Cy5 labeled heat shock samples) from the first 5 min time point of a heat shock time-course experiment. The actual size of this microarray is 1.8 cm^2. Genes that are induced or repressed appear in the image as red or green spots, respectively. Genes that are expressed at similar levels in both samples appear as yellow or orange spots. Labeled arrows identify examples of genes affected by heat shock.

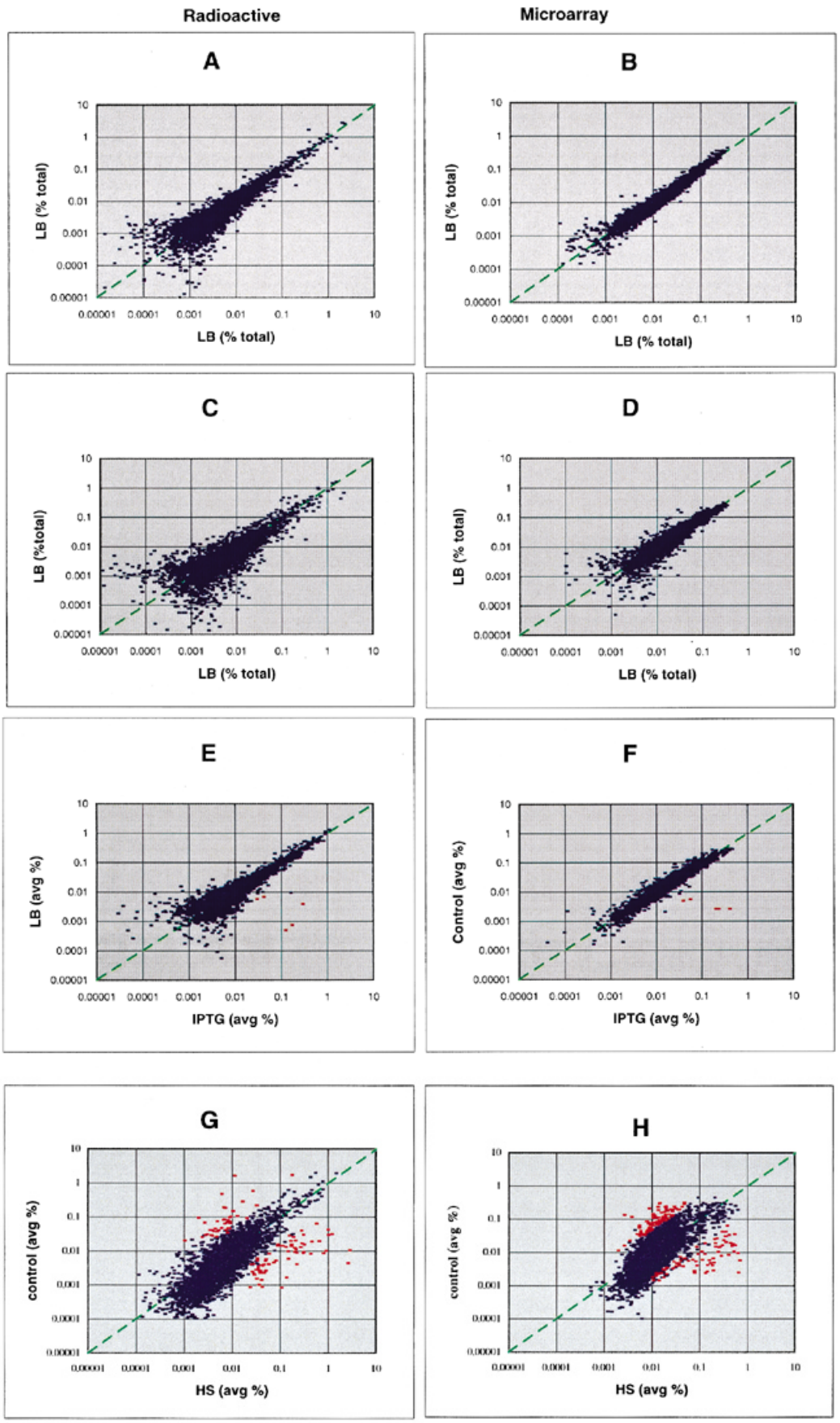
DISCUSSION

We describe the use of high density arrays of amplimers from the complete *E.coli* genome sequence to determine global changes in transcript abundance under various experimental conditions. The changes observed in response to treatment with IPTG and heat shock validate use of this method to detect alterations in transcript abundance from samples of total cellular RNA. We have also compared hybridization of radioactive cDNAs to spot blots on nylon membranes with fluorescence-based hybridization to glass microarrays to demonstrate the reliability and reproducibility of the two methods.

Figure 4 shows multiple comparisons of data collected using radioactive hybridization to spot blots and fluorescence hybridization to microarrays. Hybridization of identical RNA samples (RNA from *E.coli* grown in LB broth to mid-log) allowed for measurement of experimental reproducibility between the two methods. Parallel hybridization of identical samples to the same microarray (Fig. 4B) shows a higher degree of correlation than sequential hybridization to the same spot blot (Fig. 4A). With microarrays only four spots showed ratios greater than 3 with only one of these having a ratio greater than 4. In comparison, spot blots showed 30 spots with ratios greater than 3 and 11 of these with ratios greater than 4. We find that the variation in signal intensities is more apparent for low intensity signals with both methods. Low level signals are more difficult to accurately measure and are impacted to a larger degree by use of average background subtraction (spot blots) as compared to local background subtraction (microarrays).

Hybridization of identical RNA samples to different arrays was used to assess reproducibility of expression analysis when comparing data collected from separate arrays (Fig. 4C and D). For both methods it is apparent that use of different arrays for experimental and control hybridizations is less reproducible than hybridization of both samples to the same array (Fig. 4, compare A and B with C and D). Hybridization to different glass microarrays shows a similar correlation to sequential hybridization of the same RNA to a single spot blot.

The same variation between methods is seen when different RNAs from an experimental condition, known to affect few genes, are compared (Fig. 4E and F). In this comparison RNA


Radioactive
Microarray
A
B
C
D
E
F
G
H
LB (% total)
LB (%total)
LB (avg %)
Control (avg %)
control (avg %)
IPTG (avg %)
HS (avg %)
0.00001
0.0001
0.001
0.01
0.1
1
10

from the IPTG study was used. These RNA samples were isolated independently from a culture that was split for 30 min and maintained at 37°C and would therefore be expected to show more variation than comparison of identical RNA samples. We do see a larger degree of variation when comparing independent preparations of RNA although we only see six spots with ratios greater than 3 using microarrays and 39 spots with ratios greater than 3 with spot blots. A similar degree of variation was seen by Wodicka *et al.* when independently prepared RNA samples were hybridized to separate oligonucleotide arrays (16). It is apparent that all global expression monitoring methods will show more variation when independently isolated RNA samples are used. This will become an even greater problem when different RNA samples from normal and wild-type cells are compared.

Based on our results we conclude that the microarray approach is more reproducible than radioactive hybridization to spot blots. The relative ease and minimum time required to perform expression studies with microarrays make this the preferred method. Radioactive hybridization to spot blots produce similar results but appear to be more subject to variation. One explanation for the higher variability with hybridization to separate membranes is variability in spot deposition. We used a single channel robotic printing device which may be more prone to variation than printing methods for microarrays which rely on capillary action of small volumes and contact printing. The advantage of radioactive methods is their minimal cost, availability of whole genome arrays, and the fact that most laboratories are currently equipped to perform these studies. However, if radioactive methods are used for global expression monitoring the preferred method should be sequential hybridization to the same membrane (Fig. 4A) rather than hybridization of control and experimental samples to separate membranes (Fig. 4C).

In the IPTG experiment described here we observed induction of the well-characterized lactose operon as well as the melibiose operon. Two ORFs of unknown function were seen to be significantly affected by IPTG using spot blots while these same two ORFs show no affect using microarray assays. Induction of *ebgA* by IPTG, as seen on spot blots, contrasted with results reported by Hall and Clarke (57) who found that IPTG treatment does not lead to increased enzymatic activity of EbgA assayed in crude protein extracts. The nucleotide sequences of *lacZ* and *ebgA* share ~50% identity over most of their length and further investigation reveals that *lacZ* cDNA can indeed bind to the *ebgA* PCR product during hybridization to spot blots. Although we did not observe the same affect on *ebgA* using glass microarray methods, Heller *et al.* report that cross-hybridization artifacts using microarray methods can occur due to low levels of sequence identity when they occur over the entire length of a target (33). The potential for cross-hybridization between genes with significant levels of sequence identity must always be considered when evaluating the results of global expression studies. In fact, 556 *E.coli* ORFs contain regions of >200 bp that share a minimum of 50% identity to at least one other ORF in the genome. We currently maintain a database of potential cross-hybridizing gene sequences and correlate these with expression analysis to avoid misinterpretation of potentially artifactual hybridization results (available at http://www.genetics.wisc.edu).

In the heat shock analysis we identified 119 genes significantly affected by growth at 50°C for 7 min when determined by spot blots. More than half of the induced genes (46 of 77) were observed to be similarly affected when assayed using microarray analysis. Genes found to be down regulated using spot blots do not correlate as well when compared to microarray analysis. One explanation of the variation in ratios for these comparisons may be the use of rich defined growth media (23) for RNA samples used in microarray studies compared to growth in LB broth for spot blot analysis. In fact, the observation of repression of nucleotide biosynthetic genes following heat shock was not confirmed in additional heat shock experiments using LB broth, minimal media or rich defined media (not shown).

Correlation analysis of averaged heat shock data showed a very interesting result. As in control studies (IPTG), RNA samples from control and heat shock treated cells were isolated from a single culture that was split and subject to an environmental stimulus prior to RNA isolation. The correlation observed for this study (Fig. 4G and H) was quite different from that observed with IPTG treatment (Fig. 4E and F). Heat shock is pleiotropic effector of gene expression which is apparent in the plot. We reproducibly see a larger amount of signal variation between heat shock and control samples using either monitoring method. Again we see that this variation is diminished when using microarrays but the trend is still there. These results suggest that exposure to heat shock has much greater global effects than we are measuring using our significance tests. There may be a large number of genes subtly affected (either specifically or non-specifically) that we cannot accurately measure within the noise inherent to the method.

We have observed that different sets of genes are affected, both positively and negatively, when cells are subject to the same environmental stimuli in different media or at different times. For example, heat shock in LB broth (spot blots) showed a number of differences compared to heat shock in minimal media (data not shown) or rich defined media (microarrays). The majority of genes previously assigned to the heat shock stimulon, as well as a subset of ORFs with unassigned function, were consistently affected in all heat shock conditions, but a

Figure 4. (Opposite) Comparison of radioactive and fluorescence-based hybridization methods. Percent total signal intensity from radioactive and fluorescence hybridization studies were plotted on a logarithmic scale. (**A**) and (**B**) Identical cDNA probes hybridized sequentially to the same membrane (A) or in parallel to a single microarray (B). (**C**) and (**D**) Identical cDNA probes hybridized to separate spot blots (C) or hybridized to separate microarrays (D). (**E**) and (**F**) Average percent signal intensity from control (LB) and experimental (IPTG) samples hybridized to same membrane (E) or to a single microarray (F). (**G**) and (**H**) Average percent signal intensity from control (37°C) and experimental (50°C, 7 min) samples hybridized sequentially to the same membrane (G) or in parallel to a single microarray (H). Spots in red in (E)–(H) are those that meet our cutoff criteria (see Materials and Methods).

subset of genes responded in one condition but not in the other. In addition, when we compared RNA samples from batch cultures of MG1655 grown in different preparations of LB broth and grown to similar optical density, significant differences in the signal intensities of particular genes were observed. These observations argue for special care in the design of global expression studies. Since many interesting expression experiments will depend on addition or removal of compounds from the growth media we favor the use of a chemically defined medium rather than broth. The medium used for heat shock time course studies is ideal in that it is simple to prepare in bulk and supplement as required (23). Use of such a defined media will also facilitate integration of metabolite abundance information with global transcription analysis.

The ORFs affected by heat shock provide a list of potential players in the cell's response to this stimulus but do not describe how the cell integrates these functions in response to stress. In some cases, however, it is possible to infer a biological role for a gene product based on its inclusion in the set of affected genes. For example, *glyA* is significantly repressed during the heat shock response under all growth conditions tested. The *glyA* gene encodes the enzyme serine hydroxymethyltransferase which converts serine to glycine generating 5,10-methylenetetrahydrofolate, a major source of one-carbon (C1) units in *E.coli*. The next step in catabolism of glycine to CO_2 and ammonia by the enzymes of the *gvc* operon and *lpdA* also results in production of C1 units. (The *lpdA* gene is also down-regulated by heat shock; Table 2.) C1 units are used in a variety of biosynthetic pathways including deoxy-pyrimidine nucleotide biosynthesis and formyl-methionine biosynthesis. Gold has proposed that a possible mechanism of global regulation of protein synthesis may occur in the initiation of protein synthesis (58) for which C1 units are required. It is possible that down-regulation of the *glyA* and *lpdA* genes could lead to a global down-regulation of translation by limiting the supply of C1 units. This would suggest that simple repression of a few genes could lead to the global down-regulation of protein synthesis observed following heat shock (36,39).

Our findings also suggest that there is little change in the overall expression level of most genes following a 50°C heat shock. This observation is in agreement with the findings of Henry *et al.* (59) who found little change in the level of mRNA turnover following heat shock and suggest that the decrease in protein expression is controlled, not at the level of transcriptional repression, but at the translational or post-translational level.

The results presented here underscore the power of genome-wide measurements of transcript abundance. Detecting changes in response to different stimuli will define hypotheses regarding mechanisms of gene regulation and cellular responses for further analyses.

ACKNOWLEDGEMENTS

We wish to thank Valerie Burland and Guy Plunkett III for stimulating discussion during preparation of this manuscript, Brian Peters and Patrick Brown's lab for help in preparing microarrays, Chris Herring for work on *lacZ* and *ebgA* analysis, Nicole Perna for help with paralog analysis and Cathryn Lundberg and Shannon Hinsa for excellent technical assistance. This is Laboratory of Genetics Paper 3542. This work was supported by NIH grant R01 GM35682.

REFERENCES

1. Blattner,F.R., Plunkett,G.III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y. (1997) *Science*, **277**, 1453–1474.
2. Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.F., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.L., Geoghagen,N.S.M. and Venter,J.C. (1996) *Science*, **273**, 1058–1073.
3. Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry,C.E.III, Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Barrell,B.G. *et al.* (1998) *Nature*, **393**, 537–544.
4. Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.F., Dougherty,B.A., Merrick,J.M. *et al.* (1995) *Science*, **269**, 496–512.
5. Himmelreich,R., Hilbert,H., Plagens,H., Pirkl,E., Li,B.C. and Herrmann,R. (1996) *Nucleic Acids Res.*, **24**, 4420–4449.
6. Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S., Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C., Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V., Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Danchin,A. *et al.* (1997) *Nature*, **390**, 249–256.
7. Chuang,S.-E., Daniels.,D.L. and Blattner,F.R. (1993) *J. Bacteriol.*, **175**, 2026–2036.
8. Shalon,D., Smith,S.J. and Brown,P.O. (1996) *Genome Res.*, **6**, 639–645.
9. Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) *Science*, **270**, 467–470.
10. Welford,S.M., Gregg,J., Chen,E., Garrison,D., Sorensen,P.H., Denny,C.T. and Nelson,S.F. (1998) *Nucleic Acids Res.*, **26**, 3059–3065.
11. Khan,J., Simon,R., Bittner,M., Chen,Y., Leighton,S.B., Pohida,T., Smith,P.D., Jiang,Y., Gooden,G.C., Trent,J.M. and Meltzer,P.S. (1998) *Cancer Res.*, **58**, 5009–5013.
12. DeRisi,J.L., Iyer,V.R. and Brown,P.O. (1997) *Science*, **278**, 680–686.
13. Chu,S., DeRisi,J., Eisen,M., Mulholland,J., Botstein,D., Brown,P.O. and Herskowitz,I. (1998) *Science*, **282**, 699–705.
14. Schena,M., Shalon,D., Heller,R., Chai,A., Brown,P.O. and Davis,R.W. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 10614–10619.
15. Zhu,H., Cong,J.P., Mamtora,G., Gingeras,T. and Shenk,T. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14470–14475.
16. Wodicka,L., Dong,H., Mittmann,M., Ho,M.H. and Lockhart,D.J. (1997) *Nature Biotechnol.*, **15**, 1359–1367.
17. Holstege,F.C., Jennings,E.G., Wyrick,J.J., Lee,T.I., Hengartner,C.J., Green,M.R., Golub,T.R., Lander,E.S. and Young,R.A. (1998) *Cell*, **95**, 717–728.
18. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) *Mol. Biol. Cell*, **9**, 3273–3297.
19. Gingeras,T.R., Ghandour,G., Wang,E., Berno,A., Small,P.M., Drobniewski,F., Alland,D., Desmond,E., Holodniy,M. and Drenkow,J. (1998) *Genome Res.*, **8**, 435–448.
20. Behr,M.A., Wilson,M.A., Gill,W.P., Salamon,H., Schoolnik,G.K., Rane,S. and Small,P.M. (1999) *Science*, **284**, 1520–1523.
21. Iyer,V.R., Eisen,M.B., Ross,D.T., Schuler,G., Moore,T., Lee,J.C.F., Trent,J.M., Staudt,L.M., Hudson,J.Jr, Boguski,M.S., Lashkari,D., Shalon,D., Botstein,D. and Brown,P.O. (1999) *Science*, **283**, 83–87.
22. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
23. Neidhardt,F.C., Bloch,P.L. and Smith,D.F. (1974) *J. Bacteriol.*, **119**, 736–747.
24. Sambrook,J., Fritsch,E.F. and Maniatis,T. (eds), (1989) *Molecular Cloning: A Laboratory Manual.* 2nd Edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY.
25. Andre,P., Kim,A., Khrapko,K. and Thilly,W.G. (1997) *Genome Res.*, **7**, 843–852.
26. Lundberg,K.S., Shoemaker,D.D., Adams,M.W., Short,J.M., Sorge,J.A. and Mathur,E.J. (1991) *Gene*, **108**, 1–6.
27. Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) *Nucleic Acids Res.*, **24**, 3546–3551.

28. Barnes,W.M. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 2216–2220.
29. Webster,C., Kempsell,K., Booth,I. and Busby,S. (1987) *Gene*, **59**, 253–263.
30. Miller,J.H. (1980) In Miller,J.H. and Reznikoff,W.S. (eds), *The Operon.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY, pp. 31–88.
31. Stokes,H.W. and Hall,B.G. (1985) *Mol. Biol. Evol.*, **2**, 478–483.
32. Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl Acad. Sci. USA*, **80**, 726–730.
33. Heller,R.A., Schena,M., Chai,A., Shalon,D., Bedilion,T., Gilmore,J., Woolley,D.E. and Davis,R.W. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 2150–2155.
34. Chuang,S.-E. and Blattner,F.R. (1993) *J. Bacteriol.*, **175**, 5242–5252.
35. Gross,C.A. (1996) In Neidhardt,F.C., Curtiss,R.III, Ingraham,J.L., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella Cellular and Molecular Biology*. 2nd Edn. ASM Press, Washington, DC, Vol. 1, pp. 1382–1400.
36. Neidhardt,F.C., VanBogelen,R.A. and Vaughn,V. (1984) *Annu. Rev. Genet.*, **18**, 295–329.
37. Neidhardt,F.C. (1987) *Genes Dev.*, **1**, 109–110.
38. Lindquist,S. (1986) *Annu. Rev. Biochem.*, **55**, 1151–1191.
39. Lemaux,P.G., Herendeen,S.L., Bloch,P.L. and Neidhardt,F.C. (1978) *Cell*, **13**, 427–434.
40. Yamamori,T., Ito,K., Nakamura,Y. and Yura,T. (1978) *J. Bacteriol.*, **134**, 1133–1140.
41. Herendeen,S.L., VanBogelen,R.A. and Neidhardt,F.C. (1979) *J. Bacteriol.*, **139**, 185–194.
42. Grossman,A.D., Erickson,J.W. and Gross,C.A. (1984) *Cell*, **38**, 383–390.
43. Erickson,J.W. and Gross,C.A. (1989) *Genes Dev.*, **3**, 1462–1471.
44. Wang,Q.P. and Kaguni,J.M. (1989) *J. Bacteriol.*, **171**, 4248–4253.
45. Watson,N., Dunyak,D.S., Rosey,E.L., Slonczewski,J.L. and Olson,E.R. (1992) *J. Bacteriol.*, **174**, 530–540.
46. Yamanaka,K. and Inouye,M. (1997) *J. Bacteriol.*, **179**, 5126–5130.
47. Missiakas,D., Mayer,M.P., Lemaire,M., Georgopoulos,C. and Raina,S. (1997) *Mol. Microbiol.*, **24**, 355–371.
48. De Las Penas,A., Connolly,L. and Gross,C.A. (1997) *Mol. Microbiol.*, **24**, 373–385.
49. Katayama,Y., Gottesman,S., Pumphrey,J., Rudikoff,S., Clark,W.P. and Maurizi,M.R. (1988) *J. Biol. Chem.*, **263**, 15226–15236.
50. Gottesman,S., Clark,W.P. and Maurizi,M.R. (1990) *J. Biol. Chem.*, **265**, 7886–7893.
51. Kroh,H.E. and Simon,L.D. (1990) *J. Bacteriol.*, **172**, 6026–6034.
52. Conlin,C.A., Trun,N.J., Silhavy,T.J. and Miller,C.G. (1992) *J. Bacteriol.*, **174**, 5881–5887.
53. Jiang,X., Zhang,M., Ding,Y., Yao,J., Chen,H., Zhu,D. and Muramatu,M. (1998) *J. Biochem. (Tokyo)*, **124**, 980–985.
54. Tsui,H.C., Feng,G. and Winkler,M.E. (1996) *J. Bacteriol.*, **178**, 5719–5731.
55. Herman,C., Thevenet,D., D'Ari,R. and Bouloc,P. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 3516–3520.
56. Missiakas,D., Schwager,F., Betton,J.M., Georgopoulos,C. and Raina,S. (1996) *EMBO J.*, **15**, 6899–6909.
57. Hall,B.G. and Clarke,N.D. (1977) *Genetics*, **85**, 193–201.
58. Gold,L. (1988) *Annu. Rev. Biochem.*, **57**, 199–233.
59. Henry,M.D., Yancey,S.D. and Kushner,S.R. (1992) *J. Bacteriol.*, **174**, 743–748.
60. Riley,M. (1998) *Nucleic Acids Res.*, **26**, 54.