

Probing the ArcA-P Modulon of *Escherichia coli* by Whole Genome Transcriptional Analysis and Sequence Recognition Profiling*

Received for publication, December 9, 2003
Published, JBC Papers in Press, January 7, 2004, DOI 10.1074/jbc.M313454200

Xueqiao Liu and Peter De Wulf‡

From the Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115

The ArcB/ArcA two-component signal transduction system of *Escherichia coli* regulates gene expression in response to the redox conditions of growth. Over the years, genetic screens have led to the identification of about 30 ArcA-P-controlled operons that are involved in redox metabolism. However, the discovery of 3 targets that are not implicated in respiratory metabolism (the *tra* operon for plasmid conjugation, *psi* site for Xer-based recombination, and *oriC* site for chromosome replication) suggests that the Arc modulon may comprise additional operons that are involved in a myriad of functions. To identify these operons, we derived the ArcA-P-dependent transcription profile of *E. coli* using oligonucleotide-based microarray analysis. The findings indicated that 9% of all open reading frames in *E. coli* are affected either directly or indirectly by ArcA-P. To identify which operons are under the direct control of ArcA-P, we developed the ArcA-P recognition weight matrix from footprinting data and used it to scan the genome, yielding an ArcA-P sequence affinity map. By overlaying both methods, we identified 55 new Arc-regulated operons that are implicated in energy metabolism, transport, survival, catabolism, and transcriptional regulation. The data also suggest that the Arc response pathway, which translates into a net global downscaling of gene expression, overlaps partly with the FNR regulatory network. A conservative but reasonable assessment is that the Arc pathway recruits 100–150 operons to mediate a role in cellular adaptation that is more extensive than hitherto anticipated.

In *Escherichia coli*, gene expression in response to changing respiratory conditions of growth is partially mediated by the Arc two-component signal transduction system (1–6), which comprises the transmembrane ArcB sensor kinase and its cytosolic cognate response regulator ArcA (7, 8). Under anaerobic or microaerobic conditions, ArcB undergoes autophosphorylation and then catalyzes the transphosphorylation of ArcA. Under aerobic conditions, oxidized forms of quinone electron carriers in the membrane inhibit the autophosphorylation of ArcB and therefore its mediation of the Arc metabolic response (9). Phospho-ArcA (ArcA-P)¹ represses certain target operons (e.g. *glcDEFGB*, Ref.

10, *icdA*, Ref. 11, *lldPRD* [*lctPRD*], Ref. 4, *sdhCDAB*, Refs. 4 and 12, and *sodA*, Ref. 13) or activates others (e.g. *cydAB*, Ref. 1 and *pflA*, Ref. 14). To date, some 30 operons are known to be controlled by ArcA-P, most of which are involved in respiratory metabolism (6).

The 7-bp sequence 5'-TATTTaa-3' (the lowercase letters are less-conserved nucleotides) was proposed as a putative signature for promoter recognition by ArcA-P, based on DNaseI protection experiments at the *pflA* promoter (14). A subsequent homology search that included ArcA-P-protected promoter regions of *cydAB*, *pflA*, *gltA*, *lldPRD*, *sdhCDAB*, and *sodA*, plus the entire promoter regions of 16 additional operons whose expressions are ArcA-P-controlled, led to the suggestion of 5'-nGTTAATTAn-3' (n is A or T) as the ArcA-P binding consensus (15). This 10-bp consensus proved useful for locating ArcA-P binding sites at two novel targets that are not involved in respiratory metabolism: the *tra* operon for conjugation of resistance plasmid R1 (16), and the *psi* site for Xer-based recombination in plasmid pSC101 (17). The identification of these functionally distinctive targets, and of the Arc-controlled chromosome replication site *oriC* (18), suggests that more (unexpected) operons may be under the transcriptional control of ArcA-P.

To discover these operons and possibly new functions of the Arc system that reach beyond redox metabolism, we undertook two complementary approaches. First, a profile of ArcA-P-dependent gene expression in *E. coli* was obtained using oligonucleotide-based microarray analysis. Second, an ArcA-P recognition weight matrix was derived from footprinted promoter sequences and used to screen the *E. coli* genome to locate potential ArcA-P binding sites. By combining both techniques we aimed to: (i) identify operons that are most likely under the direct control of ArcA-P, (ii) estimate conservatively the number of operons that are Arc-controlled, and (iii) reveal which physiological roles are governed by the Arc pathway in *E. coli*.

EXPERIMENTAL PROCEDURES

ArcA-P Weight Matrix Development and Screening—The method of Berg and von Hippel (19) was used to score the *E. coli* genome sequence with the ArcA-P weight matrix, which was developed from footprinted promoter regions of 11 ArcA-P-controlled operons (Table I) using the AlignACE program (20). The matrix-screening method predicts the affinity of ArcA-P for any genomic 15-bp DNA sequence, based on the sequence statistics of the input promoter regions. Both strands of the *E. coli* K-12 MG1655 genome sequence, obtained from GenBank™ entry U00096, were searched. Near-symmetric sites with high scores in both the forward and reverse direction were counted only once, and the higher of the two scores was used.

Strains and Growth Conditions—*E. coli* wild-type strain MC4100 (F⁻ *araD139* Δ(*argF-lac*) U169 rpsL150 deoC1 relA1 thiA ptsF25 flb-5301 rbsR), and isogenic *arcA::kan* deletion strain ECL5331 were used in the experiments. The *arcA::kan* allele was constructed according to the method of Link *et al.* (21) and then P1 transduced into wild-type strain MC4100, yielding strain ECL5331. To obtain total RNA for use in microarray and quantitative real-time PCR analyses (RT-PCR), both strains were grown anaerobically in 400 ml of MOPS-buffered (100 mM, pH7.4) Luria-Bertani (LB) medium that was supplemented with 20 mM

* This work was supported by United States Public Health Service Grant GM40993 from the NIGMS, National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Biology, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. Tel.: 617-452-4938; Fax: 617-253-8550; E-mail: dewulf@mit.edu.

¹ The abbreviations used are: ArcA-P, phospho-ArcA; MOPS, 4-morpholinepropanesulfonic acid; MES, 4-morpholineethanesulfonic acid; ORF, open reading frame.

of D-xylose. The cells were grown at 37 °C under cap-sealed conditions with constant stirring with a magnetic bar until an optical density (OD₆₀₀) of 0.3–0.4 was reached.

Total RNA Isolation—Upon reaching an OD₆₀₀ of 0.3–0.4, nine independent cultures of the wild-type and mutant strains were pooled, yielding three sets of three cultures per strain (the experimental setup is detailed in Fig. 4). The cultures were cooled down with shaved ice, harvested, and subjected to total RNA isolation using the hot phenol extraction method (22). The obtained RNA was treated with RNase-free DNase I (Invitrogen) in the presence of RNA Recombinant Ribonuclease Inhibitor (RNAout, Invitrogen). The absence of contaminant DNA was then confirmed by RT-PCR, and the quality of the RNA examined using agarose gel electrophoresis. The RNA concentrations were measured at OD₂₆₀, and the three independent batches of RNA (for both strains) were aliquoted and stored at –80 °C for use in experiments.

Quantitative Analysis of Gene Expression by Real-Time PCR of Reverse Transcribed RNA—cDNA was first synthesized from 2 µg of total RNA using the SuperScript™ First-Strand Synthesis System (Invitrogen) according to the manufacturer's instructions. RT-PCR amplification of the cDNAs (50 °C, 2 min; 95 °C, 10 min; 95 °C, 15 s; 60 °C, 1 min; 40 cycles) was performed with SYBR Green I Dye in a reaction mixture containing the SYBR PCR Master Mix (Applied Biosystems), 1 µM of the forward and reverse primers, DEPC-treated water, and either cDNA or genomic DNA (25-µl reaction volume). The dye-labeled PCR products were quantified with a 7700 Sequence Detector (Applied Biosystems). A standard curve for normalizing the concentration of cDNAs was generated by amplifying *envZ* from genomic DNA of the wild-type strain, because *envZ* transcription is not subject to ArcA-P control (data not shown).

Analysis of Gene Expression Using Oligonucleotide-based Affymetrix GeneChip® Microarrays—cDNA for microarray experiments was synthesized from 5–15 µg of total RNA using the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen) and random primers (Promega) according to the manufacturers' instructions. The RNA template was removed from the cDNA by adding 1 N NaOH, followed by cDNA purification using QIAquick columns (Qiagen). 3–7 µg of cDNA were next fragmented with DNase I (Promega) and labeled (1 h, 37 °C) with biotinylated dd-UTP using the Bioarray™ Terminal Labeling kit (Enzo). The labeled cDNA fragments were next hybridized (16 h, 45 °C) to *E. coli* Antisense Genome Oligonucleotide Arrays (Affymetrix). Arrays were washed at 25 °C with 6× SSPE buffer (900 mM NaCl, 60 mM NaH₂PO₄, 6 mM EDTA + 0.01% Tween 20), followed by a stringent wash at 50 °C with 100 mM MES, 100 mM NaCl, and 0.01% Tween 20. The microarrays were then stained with phycoerythrin-conjugated streptavidin (Molecular Probes), and the fluorescence intensities measured after laser confocal scanning (Hewlett-Packard) with Microarray software (Affymetrix). Sample loading and variations in staining were standardized by scaling the average of the fluorescent intensities of all genes to a constant target intensity of 250 for all arrays used. The signal intensity for each open reading frame (ORF) was calculated as the average intensity difference, represented by $[\Sigma(\text{PM-MM})/(\text{number of probe pairs})]$, where PM and MM denote perfect match and mismatch probes.

Statistical Analysis of Microarray Profiles—To determine the reproducibility of the three independent microarray analyses that were performed per strain (Fig. 4), the ORF signals derived from each microarray were correlated using Spotfire DecisionSite (Spotfire Inc.). Because these correlations were excellent (Fig. 6, A–F), we averaged the triple data sets for both the wild-type and mutant strains. To extract statistically relevant information from the data suites, the ORF hybridization signals were passed through three steps of data filtration: (i) a coefficient of variation <0.8; (ii) mutant-to-wild type signal ratio larger than 2, positive or negative ($\log_2[\text{MT:WT}] > \pm 1$); and (iii) signal ratios with a $p < 0.05$ (Student's *t* test). The filtered signals were then ascribed to ORFs whose expressions are (in)directly affected by ArcA-P.

RESULTS

Derivation of the ArcA-P Weight Matrix for Sequence Recognition Profiling—Support for ArcA-P binding at the consensus sequence 5'-nGTTAATTAn-3' (n is A or T, Ref. 16) was previously obtained from footprinting and site-directed mutagenesis experiments at the *aldA* promoter region (23). Nonetheless, the palindromic character of the sequence and its high A/T-content make the identification of functional ArcA-P binding sites difficult in A/T-rich promoter regions. In order to refine the recognition consensus, we searched for a common motif within ten

TABLE I
ArcA-P-controlled operons used to construct the ArcA-P recognition weight matrix

The ArcA-P putative binding sites are the 15-bp fragments identified by AlignACE (20) in the 10 promoter regions of 11 footprinted ArcA-P-controlled operons (27). The base frequency at each nucleotide position within the 15-bp sites was used to calculate the ArcA-P recognition weight matrix (Fig. 1).

Operon	Putative ArcA-P binding site
<i>aldA</i>	1) GTTAATTAACATCAA
	2) GTGAATACATTGTTA
<i>cydAB</i>	1) GTTAACATAAATGTTA
	2) GTTAATTATAATATA
	3) GTTAATAAAAACTCA
	4) GTTAATCATGTTTCA
	5) TATAATTATAAGTTA
<i>glcDEFGB</i> <i>gltA/sdhCDAB</i> ^a	1) GTTAACATAAATGTTA
	2) GTTAATGTAATGAT
<i>icdA</i> <i>lldPRD</i>	1) GTTCACAAAATCATT
	2) GTTAATGATTTGTAA
<i>lpdA</i>	1) GTTAACAAATTTTAA
	2) GTTAACAAATTTTGT
<i>pflA</i>	1) GTTAATTAAGAGGGA
	2) TTTAATTAACGTGTT
<i>sodA</i>	1) GTTAATTAATGATA
	2) GTTCATTATAGTTAA
<i>traY</i>	GTTAAGTAAATGTTA

^a Binding site is located in the promoter region shared by the oppositely transcribed *gltA* and *sdhCDAB* operons.

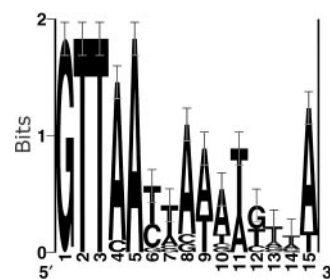


FIG. 1. Sequence logo for the ArcA-P recognition matrix in *E. coli*. The sequence conservation, measured in bits, is shown as the height of a stack of letters at each base position (41). The total sequence conservation is 15.18 ± 0.55 bits.

ArcA-P footprinted promoter regions (400-bp upstream to 100-bp downstream of the start codon) of eleven ArcA-P-controlled operons (Table I). The screen was performed with the program AlignACE (20), which uses the Gibbs sampler algorithm (24). The analyzed operons were *aldA* (23), *cydAB* (1), *glcDEFGB* (10), *gltA* and *sdhCDAB* (4, 15, 25), *icdA* (11), *lldPRD* (4), *lpdA* (26), *pflA* (14), *sodA* (13), and *traY* (16) (Table I). Highly conserved stretches of 15 base pairs were found in the footprinted regions of all input promoters. By determining the base frequency at each position, we converted this sequence into a weight matrix (Fig. 1). The sequence of most conserved bases, 5'-GTTAATTAAATGTTA-3', resembles the previous 10-bp consensus (5'-nGTTAATTAn-3', n is A or T). However, the first nucleotide of the consensus (5'-A/T) turned out to be poorly conserved and is not included in the present motif. On the other hand, the motif is extended by 5 residues at the 3'-end.

Screening the *E. coli* Genome with the ArcA-P Recognition Weight Matrix—Each of the 4,639,206 successive 15-bp stretches of the *E. coli* K-12 MG1655 genome was next scored on both strands for its degree of matching with the ArcA-P weight matrix using the log transformation method of Berg and von Hippel (19). The distribution of the scores of all potential ArcA-P recognition sites was determined and the average of these scores, *i.e.* the genomic mean, assigned a Z score of 0. A

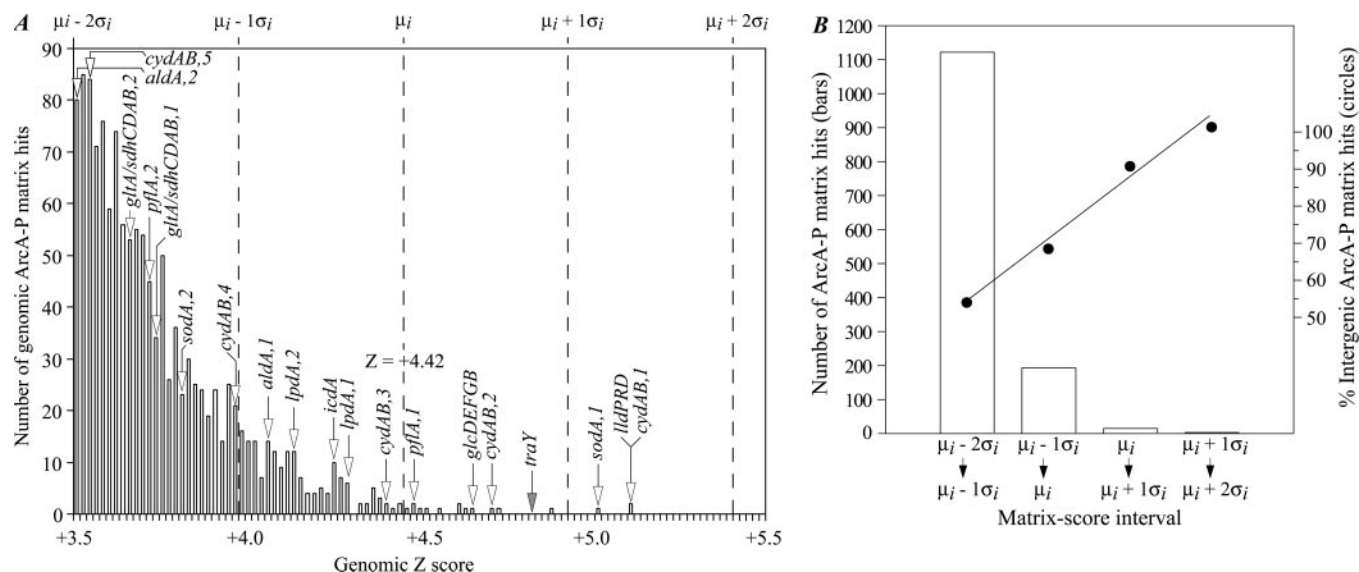


FIG. 2. Score-based distribution of matrix-identified ArcA-P recognition sites. A, distribution of genomic matrix-identified ArcA-P recognition sites that score above $Z = +3.5$. The genomic Z score represents the number of standard deviations above the genomic mean of $Z = 0$. μ_i denotes the mean Z score of the ArcA-P binding sites present in the promoter regions of the eleven ArcA-P-controlled input operons (Table I). σ_i denotes the standard deviation below or above μ_i . The *openhead arrows* indicate the Z-score positions of the recognition sites used to create the ArcA-P weight matrix. The *gray arrow* designates the potential genomic Z-score position of *traY*, an Arc-controlled input operon that is plasmid encoded. B, classification of high scoring matrix hits based on their matrix score and their probability of being located intergenically.

histogram window comprising the best scoring sites ($Z \geq +3.5$) is shown in Fig. 2A. As one would expect, all sites in the promoter regions of the eleven input operons were identified (Table I, operons are marked with an *arrow* in Fig. 2A), because the probability of being a true ArcA-P regulatory site statistically increases with an increasing Z score. The mean score μ_i of the ArcA-P recognition sites in front of the 11 input operons has a Z value of +4.42 (Fig. 2A). We next determined the standard deviation, σ_i , of these input sites and ranked the genomic hits with a Z value $\geq +3.5$ into four standard deviation groups relative to μ_i (Fig. 2, A and B). The first standard deviation group contains potential ArcA-P binding sites that score above $\mu_i + 1\sigma_i$ (Fig. 2, A and B; Table II, rows 1–3). This group comprises only three sites (Fig. 2, A and B, bars), all of which are intergenically located (Fig. 2B, circles) upstream of *lldPRD* (matrix score: $\mu_i + 1.37\sigma_i$), *cydAB* ($\mu_i + 1.37\sigma_i$), and *sodA* ($\mu_i + 1.17\sigma_i$). These three input operons were used to create the weight matrix (Table I). The second standard deviation group (Fig. 2, A and B and Table II, rows 4–15) contains 12 potential ArcA-P recognition sites that score between μ_i and $\mu_i + 1\sigma_i$ (Fig. 2B, bars). Among these, 11 sites (or 92%) are intergenically located immediately upstream of an open reading frame (Fig. 2B, circles). Only the site upstream of *xylH* is located intragenically as it covers the coding sequence of preceding *xylG*. This matrix site, however, may well be functional as it covers the mRNA sequence of the *xylFGHR* operon. Of the 11 intergenic sites, 4 are localized in front of input operons: *traY* ($\mu_i + 0.79\sigma_i$), *cydAB* ($\mu_i + 0.54\sigma_i$), *glcDEFGB* ($\mu_i + 0.42\sigma_i$), and *pflA* ($\mu_i + 0.05\sigma_i$), as well as sites in front of operons that are not connected with respiratory metabolism such as *cst/xthA* ($\mu_i + 0.40\sigma_i$) and *insA_5uspC* ($\mu_i + 0.02\sigma_i$) (Table II). The third standard deviation group comprises 193 putative ArcA-P binding sites that score between μ_i and $\mu_i - 1\sigma_i$ (Fig. 2B, bars). Among these, 132 sites (or 68%) are located in intergenic regions (Fig. 2B, circles). Representatives include 5 sites located in front of input operons *alda* ($\mu_i - 0.69\sigma_i$), *icdA* ($\mu_i - 0.38\sigma_i$), *lpdA* ($\mu_i - 0.33\sigma_i$, $\mu_i - 0.65\sigma_i$), and *cydAB* ($\mu_i - 0.11\sigma_i$) (Fig. 2A). The fourth standard deviation group comprises 1,122 potential ArcA-P boxes that score between $\mu_i - 1\sigma_i$ and $\mu_i - 2\sigma_i$ (Fig. 2B, bars). Among these, 606 sites (or 54%) are located in

intergenic regions (Fig. 2B, circles). Members include 7 sites found in the promoter regions of input operons *cydAB* ($\mu_i - 1.01\sigma_i$, $\mu_i - 1.937\sigma_i$), *gltA/sdhCDAB* ($\mu_i - 1.51\sigma_i$, $\mu_i - 1.67\sigma_i$), *sodA* ($\mu_i - 1.35\sigma_i$), *pflA* ($\mu_i - 1.53\sigma_i$), and *alda* ($\mu_i - 1.98\sigma_i$) (Fig. 2A).

In total, 752 intergenic sites score higher than $\mu_i - 2\sigma_i$. Interestingly, a positive first order relationship (r^2 : 0.99) was found between the matrix score of a site and the probability of it being located intergenically (Fig. 2B, circles). Even more, the hits scoring above $\mu_i - 2\sigma_i$ appear to be located especially in sequence regions ≤ 200 -bp upstream from start codons (Fig. 3A). The matrix, derived from footprinting data, therefore seems to recognize high scoring sites in potential promoter regions, suggesting that most identified sites are very likely to be active ArcA-P binding sites. However, a number of intragenically located sites are expected to be true ArcA-P binding sites (e.g. sites located in front of or covering long transcripts) as both trends (Figs. 2B and 3A) may in part reflect the AT-rich bias of both the ArcA-P recognition sequence and promoter regions. An in-depth discussion of the significance of weight matrix discrimination between non-coding and coding sequences in *E. coli* is found in Robison *et al.* (28).

Putative target operons were also classified based on the number of ArcA-P matrix hits (scoring above $\mu_i - 2\sigma_i$) that are clustered in their promoter regions (Fig. 3B, bars): 148 promoter regions harbor two ArcA-P boxes, and 74 contain three boxes or more. The observed increase in probability that multiple boxes are tightly grouped in intergenic regions (r^2 : 0.98; Fig. 3B, circles) adds to the suggestion that more than the 30 currently known operons may be subject to Arc regulation. Clustered ArcA-P boxes are on average separated by 15 bp, indicating that ArcA-P may act multimerically at some of its target regions. This observation supports the recent finding that ArcA-P may multimerize prior to DNA binding (29).

Measuring the Effect of ArcA-P on Gene Expression using Microarrays—In order to identify which *E. coli* genes are under the regulatory control of ArcA-P, we compared the transcription profile of a wild-type strain to that of an isogenic *arcA* deletion strain (the experimental setup is detailed in Fig. 4). For this purpose, both strains were cultured anaerobically be-

TABLE II
Bona fide and putative ArcA-P target operons containing a potential ArcA-P binding site that scores higher than μ_i

Operon	Function	Matrix score	Location of the ArcA-P box ^a	In vivo effect of ArcA-P	In vitro binding of ArcA-P
1) <i>llPRD</i> ^b	L-Lactate permease, regulator and L-lactate dehydrogenase	$\mu_i + 1.37\sigma_i$	118 → 133	Negative (This study, 4)	Yes (15)
2) <i>cydAB</i> ^b	Cytochrome d oxidase	$\mu_i + 1.37\sigma_i$	570 → 585	Positive (This study, 1)	Yes (15)
3) <i>sodA</i> ^b	Mn-superoxide dismutase	$\mu_i + 1.17\sigma_i$	63 → 78	Negative (13)	Yes (13)
4) <i>tpx/ycjG</i> ^c	Thioredoxin-linked thiol peroxidase/ L-Ala-D/L-Glu epimerase	$\mu_i + 0.90\sigma_i$	16 → 31/ 46 → 61	Negative on <i>tpx</i> (42)	N.K. ^d
5) <i>phnC</i>	Phosphonate transporter subunit I	$\mu_i + 0.60\sigma_i$	56 → 71	N.K.	N.K.
6) <i>cydAB</i> ^b	Cytochrome d oxidase	$\mu_i + 0.54\sigma_i$	592 → 607	Positive (1)	Yes (15)
7) <i>glcDEFGB</i> ^b	Glycolate catabolism	$\mu_i + 0.42\sigma_i$	89 → 104	Negative (10)	Yes (10)
8) <i>cstC/xthA</i> ^c	N- α -Acetylornithine- Δ -aminotransferase/ exonuclease III	$\mu_i + 0.40\sigma_i$	36 → 51/ 395 → 410	N.K.	N.K.
9) <i>ycbU</i>	Putative fimbrial-like protein	$\mu_i + 0.35\sigma_i$	10 → 25	N.K.	N.K.
10) <i>feoA</i>	Ferrous iron uptake system	$\mu_i + 0.34\sigma_i$	278 → 293	N.K.	N.K.
11) <i>xylH</i>	Xylose permease	$\mu_i + 0.24\sigma_i$	172 → 187	N.K.	N.K.
12) <i>pntA/ydgH</i> ^c	Pyridine nucleotide transhydrogenase α subunit (EC 1.6.1.1)/unknown	$\mu_i + 0.12\sigma_i$	300 → 315/ 209 → 224	N.K.	N.K.
13) <i>focA-pflA</i> ^b	Formate transport and pyruvate formate lyase	$\mu_i + 0.05\sigma_i$	211 → 226	Positive (14)	Yes (14)
14) <i>yhjE</i>	Putative transporter	$\mu_i + 0.05\sigma_i$	186 → 201	N.K.	N.K.
15) <i>insA_5/uspC</i> ^c	IS1 protein InsA/universal stress response protein	$\mu_i + 0.02\sigma_i$	308 → 323/ 215 → 230	N.K.	N.K.

^a Location of the box is in bp upstream of the start codon.

^b ArcA-P-controlled operon whose footprinted promoter sequence was used to construct the weight matrix.

^c The single box is located between 2 oppositely transcribed operons. Its location is in bp from the start codon of each operon.

^d N.K., not known. No literature data are available and no effects were observed in our microarray experiments under the implemented growth conditions. See text for more information.

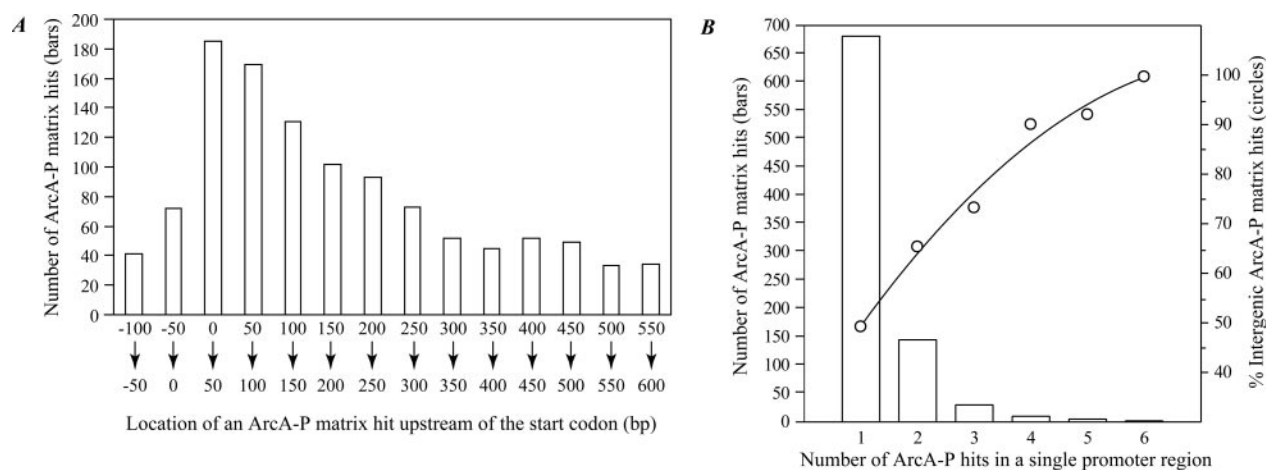


FIG. 3. Position analysis of matrix-identified ArcA-P recognition sites. A, location of high scoring matrix hits within promoter regions. A negative location means that the matrix site is positioned downstream of the start codon within the coding sequence of the candidate target operon. B, classification of high scoring matrix hits clustered in promoter regions (bars), and on the probability that these clusters are located intergenically (circles).

cause only under these conditions does ArcB phosphorylate ArcA. ArcA-P then proceeds to activate or repress its target operons. Furthermore, the growth medium was free of D-glucose (D-xylose as the carbon source) and buffered with MOPS at pH 7.4 to avoid glucose-based catabolite repression or indirect pH effects on gene expression. For each strain, nine independent cultures were grown. These were then pooled into three groups of three, followed by total RNA extraction (Fig. 4). The resulting sets of RNA (three for each strain) were then subjected to reverse transcription and RT-PCR to determine the expression states of *sdhCDAB* and *cydAB*, two operons known to be under the negative and positive control of ArcA-P, respectively (1, 4, 25). The results showed an 11-fold negative effect of ArcA-P on *sdhCDAB* expression, and a 3-fold positive effect on *cydAB* (Fig. 5A), corroborating the experimental design and quality of the isolated RNAs. For each strain, the three sets of reversely transcribed RNAs were then hybridized onto *E. coli* Antisense Genome Oligonucleotide Arrays® (Affymetrix) (Fig.

4). A total of 7,312 hybridization signals (4,345 ORF signals; 2,886 intergenic signals; and 81 internal control signals) were obtained from each array (Fig. 4). When the expression profiles from the three *arcA* Δ arrays (designated MT1, MT2, and MT3) were plotted against each other, intercorrelation coefficients of 0.81, 0.77, and 0.75 were obtained (Fig. 6, A–C). Similarly, the expression profiles from the wild-type strain (designated WT1, WT2, and WT3) gave intercorrelation coefficients of 0.86, 0.94, and 0.86 (Fig. 6, D–F), showing excellent degrees of reproducibility. Consequently, we averaged for each strain the triple data sets, yielding a final expression profile for the mutant and wild-type strain. Before deriving biologically relevant information from these profiles, we validated further the microarray data by analyzing, using RT-PCR, the expression states of 15 additional operons (*metR*, *csdG*, *rhdD*, *tolB*, *fur*, *dcm*, *rpoE*, *phoE*, *fimB*, *ompT*, *secA*, *ompF*, *phoH*, *lon*, and *ompC*; Fig. 5A) in the *arcA* Δ and *arcA*⁺ strains. These operons were selected based solely on the scores of their ArcA-P recognition sites,

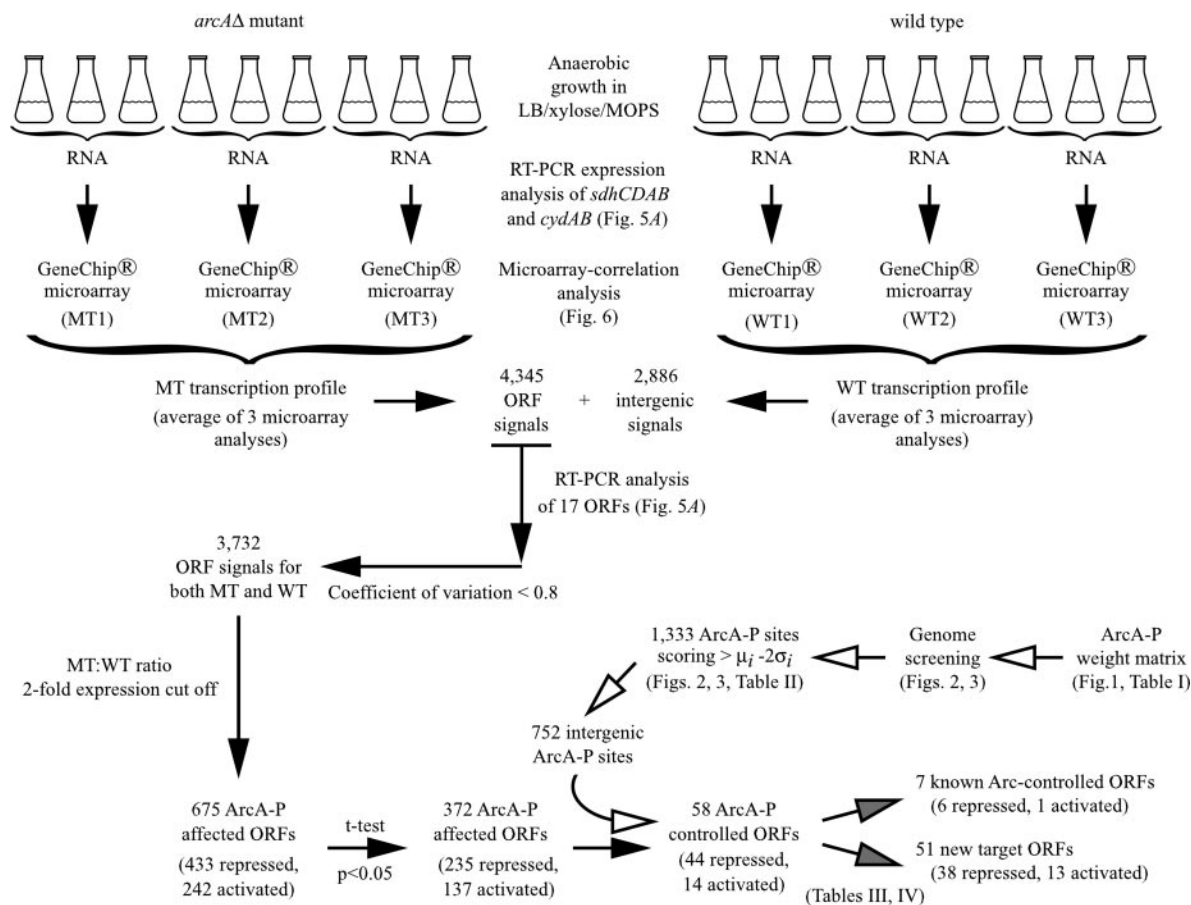


FIG. 4. **Experimental design of this study.** The experimental part underlying the microarray analysis is designated with *solid arrowheads*. The part dealing with weight matrix development and screening is marked with *open arrowheads*, and the part describing the overlay between the microarray and matrix data is shown with *gray arrowheads*.

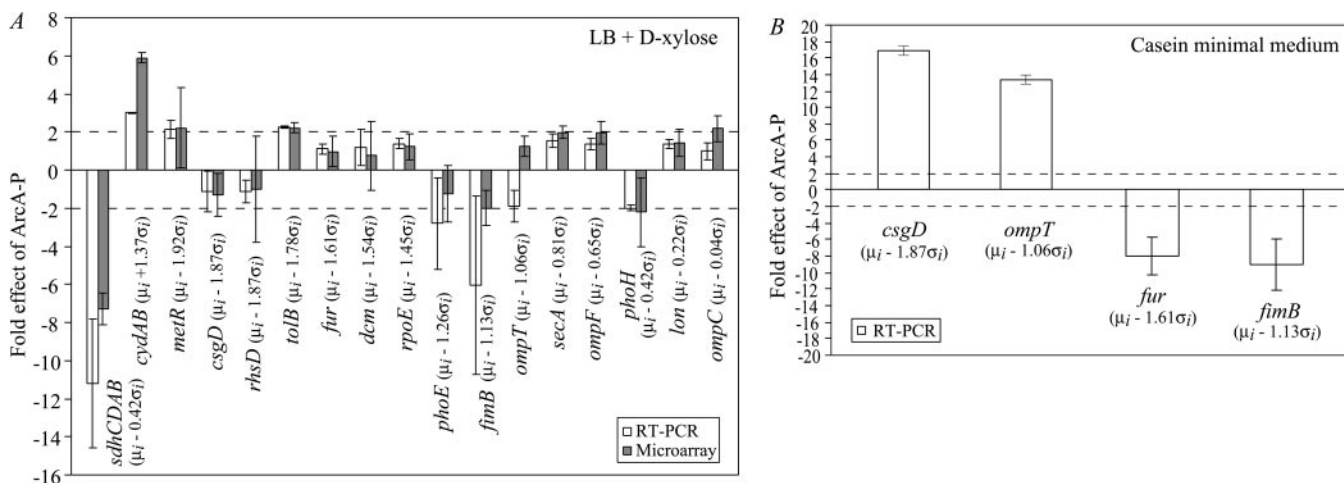


FIG. 5. **Effect of ArcA-P on the expression of selected operons scoring above $\mu_i - 2\sigma_i$.** A, effect of ArcA-P on the expression of 15 operons as measured by RT-PCR (*open bars*), and microarray (*gray bars*). The cells were grown in LB medium supplemented with D-xylose. B, effect of ArcA-P on the expression of *csgD*, *ompT*, *fur* and *fimB* as measured by RT-PCR in cells grown in minimal medium supplemented with 5% casein amino acids. The *dashed lines* designate a noise-to-signal ratio of ± 2 . See text for details.

such that they represent all candidate target operons that score above $\mu_i - 2\sigma_i$. Under the implemented non-induced growth conditions, only *fimB* (encoding the switcher of Type 1 fimbriae) was identified as a potential Arc-controlled target (Fig. 5A). When the RT-PCR expression data were plotted against the data obtained using microarrays, an excellent degree of correlation (r^2 : 0.91) was obtained, validating our microarray transcription profiles (data not shown). Quantitative RT-PCR appeared to be slightly more sensitive (~ 1.5 -fold, data not

shown) than microarray-based hybridization (30, 31), attesting that our microarray approach represents a really stringent means for identifying Arc-controlled operons.

To determine which operons are affected by the Arc signal transduction pathway, we next subjected the *arcAΔ* and *arcA+* array-transcription profiles to three steps of data filtration (Fig. 4). First, the results were subject to a coefficient of variation (C.V.) cut-off of 0.8, thereby eliminating 14% of the ORF signals. Second, a noise-to-signal ratio of $\log_2[\pm 1]$ was installed

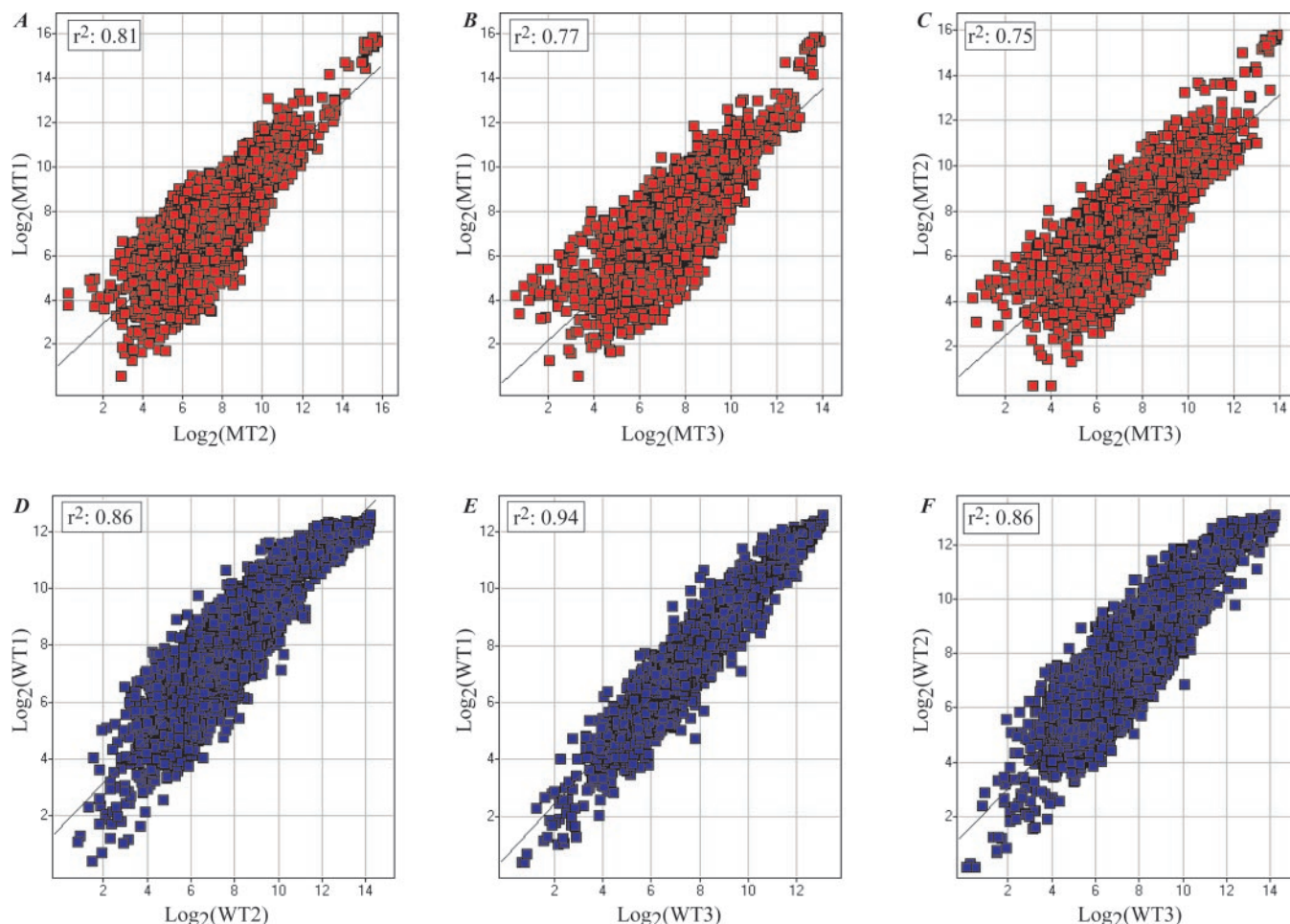


FIG. 6. **Correlation analysis of the microarray data.** A–C, plots depicting the degrees of correlation between hybridization profiles obtained from three independent analyses of *arcAΔ* cDNAs (*MT1*, *MT2*, *MT3*). D–F, plots depicting the degrees of correlation between the hybridization profiles derived from three independent analyses of *arcA*⁺ cDNAs (*WT1*, *WT2*, *WT3*).

on the *arcAΔ/arcA*⁺ expression data, resulting in an additional removal of 82% of the ORF signals. Third, a Student's *t* test with a *p* < 0.05 was run on the remaining ORF signals, yielding a final set of 372 open reading frames (or 9% of the initial ORF hybridization signals) that are significantly affected, directly or indirectly, by ArcA-P (Fig. 7).

Matching the Matrix Screening Data to the Microarray Data Reveals New Members of the Arc Modulon—To determine which microarray-identified ORFs may be under the direct control of ArcA-P, we overlaid the filtered microarray data set with the matrix data set (Fig. 4). Of the 372 Arc-affected ORFs, 58 (or 16%) contained one or multiple matrix box(es) scoring above $\mu_i - 2\sigma_i$ (Table III). Among these, we recognized seven input operons (*aldA*, *cydAB*, *glcDEFGB*, *gltA*, *icdA*, *lldPRD*, and *sdhCDAB*), validating the methodology and outcome under the implemented experimental conditions. As a consequence, 51 operons (38 repressed, 13 activated) encoding a broad variety of functions (Tables III and IV) emerged a potentially new members of the Arc modulon.

As microarray-based profiling is environment-dependent, and matrix screening is not, it is clear that the microarray data set of Arc-controlled targets is incomplete. In fact, some known ArcA-P regulated operons (e.g. *cyoABCD* and *sodA*, the latter was used to construct the matrix), as well as operons containing top-scoring ArcA-P recognition sites (designated by N.K. in Table II, column 5) were not identified in our microarray data set. This is because some of these operons are transcribed only upon external induction (e.g. *cyoABCD* and *sodA*). This may be

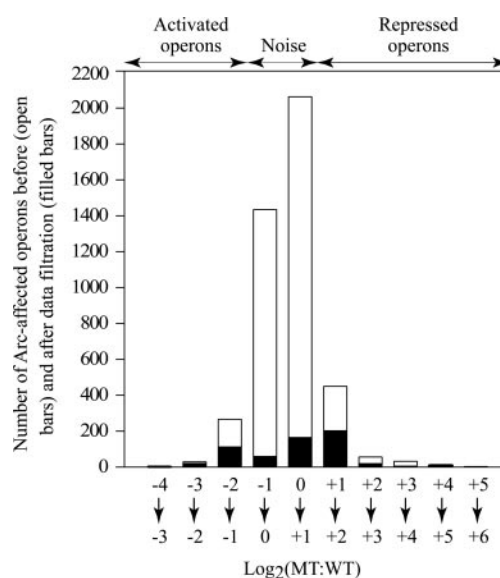


FIG. 7. **Statistical filtration of the microarray data.** Histogram showing the classification of Arc-affected operons based on their expression level before (open bars), and after (filled bars) data filtration. See text for details.

true for numerous other operons whose induction conditions are unknown, leading to an underreckoning of the Arc modulon. Fortunately, the matrix-screening profile (available at

TABLE III
Operons that are most probably under the direct transcriptional control of ArcA-P as determined by microarray analysis and matrix-based promoter profiling

Operon	Function of operon	Effect of ArcA-P ^a	<i>p</i> value (Student's <i>t</i> test)	Matrix score	Location of ArcA-P box ^b
<i>agaCDI</i>	PTS system; <i>N</i> -acetylglucosamine-specific IIC component 1, IID component 1, and putative galactosamine-6-P isomerase	-2.4	0.0144	$\mu_i - 1.93\sigma_i$ $\mu_i - 1.81\sigma_i$	104 → 119 -5 → 10
<i>aldA^c</i>	NAD-linked aldehyde dehydrogenase	-6.9	0.0075	$\mu_i - 0.69\sigma_i$ $\mu_i - 1.98\sigma_i$	44 → 59 33 → 48
<i>atoSC</i>	Two-component signal transduction system (sensor AtoS, response regulator AtoC)	-2.3	0.0174	$\mu_i - 1.37\sigma_i$	-1298 → -1313
<i>b1012</i>	Unknown	-3.0	0.0209	$\mu_i - 1.50\sigma_i$	62 → 77
<i>b1671-ydhU</i>	Putative oxidoreductase Fe-S subunit and unknown	-2.6	0.0260	$\mu_i - 1.80\sigma_i$	344 → 359
<i>b2099</i>	Unknown	-2.6	0.0461	$\mu_i - 1.99\sigma_i$	75 → 90
<i>b2146</i>	Putative oxidoreductase	-2.1	0.0008	$\mu_i - 1.56\sigma_i$	42 → 57
<i>b2651</i>	Unknown	-2.6	0.0137	$\mu_i - 1.63\sigma_i$ $\mu_i - 0.28\sigma_i$	76 → 91 47 → 62
<i>b2670</i>	Unknown	-2.1	0.0154	$\mu_i - 1.81\sigma_i$ $\mu_i - 1.78\sigma_i$ $\mu_i - 1.63\sigma_i$	404 → 419 212 → 227 207 → 222
<i>b4256</i>	Unknown	-2.5	0.0227	$\mu_i - 0.99\sigma_i$	542 → 557
<i>cadB</i>	Transport of lysine/cadaverine	-2.1	0.0062	$\mu_i - 1.62\sigma_i$	247 → 262
<i>caiT</i>	Probable carnitine/betaine transporter involved in osmoprotection	+2.2	0.0326	$\mu_i - 1.38\sigma_i$ $\mu_i - 1.60\sigma_i$ $\mu_i - 0.46\sigma_i$ $\mu_i - 0.09\sigma_i$	190 → 205 189 → 204 72 → 87 50 → 65
<i>celB</i>	PEP-dependent phosphotransferase enzyme II for cellobiose, arbutin, and salicin	-2.3	0.0076	$\mu_i - 1.33\sigma_i$	-71 → -86
<i>crl</i>	Transcriptional regulator of <i>csgA</i> (curli synthesis)	-3.3	0.0477	$\mu_i - 1.81\sigma_i$	47 → 62
<i>cydAB^c</i>	Cytochrome d terminal oxidase	+5.2	0.0187	$\mu_i + 0.54\sigma_i$ $\mu_i - 1.94\sigma_i$ $\mu_i + 1.37\sigma_i$ $\mu_i - 1.01\sigma_i$ $\mu_i - 0.11\sigma_i$	592 → 607 581 → 596 570 → 585 359 → 374 144 → 159
<i>cysM</i>	Cysteine synthase B, <i>O</i> -acetylserinesulphhydrolase B	-2.2	0.0424	$\mu_i - 1.63\sigma_i$	116 → 131
<i>dniR</i>	Transcriptional regulator for nitrite reductase (cytochrome <i>c552</i>)	+3.1	0.0275	$\mu_i - 1.89\sigma_i$ $\mu_i - 1.78\sigma_i$	260 → 275 150 → 165
<i>fabF</i>	3-Oxoacyl-[acyl-carrier-protein] synthase II	+3.3	0.0043	$\mu_i - 1.63\sigma_i$	288 → 303
<i>fadBA</i>	3-Hydroxyacyl-CoA dehydrogenase, and thiolase I	-2.6	0.0007	$\mu_i - 1.40\sigma_i$	64 → 79
<i>fadD</i>	Acetyl-CoA synthetase, long-chain fatty acid CoA ligase	-2.6	0.0037	$\mu_i - 0.57\sigma_i$ $\mu_i - 1.73\sigma_i$	61 → 76 59 → 74
<i>fepE</i>	Ferric enterobactin transport	-2.9	0.0056	$\mu_i - 1.92\sigma_i$	204 → 219
<i>fliE</i>	Flagellar basal body component	-2.3	0.0226	$\mu_i - 1.74\sigma_i$	-329 → -344
<i>fliMN</i>	Flagellar motor switching and energy and flagellar switch protein	-2.2	0.0419	$\mu_i - 0.70\sigma_i$ $\mu_i - 1.96\sigma_i$	-950 → -965 -971 → -986
<i>ftsZ</i>	Cell division; forms circumferential ring; tubulin-like GTP-binding protein and GTPase	+2.5	0.0127	$\mu_i - 1.82\sigma_i$	319 → 334
<i>glcC^c</i>	Transcriptional activator for the <i>glc</i> operon	-2.6	0.0489	$\mu_i + 0.42\sigma_i$	89 → 104
<i>glcDF</i>	Glycolate oxidase subunit D, and glycolate oxidase iron-sulfur subunit	-7.1	0.0009	$\mu_i + 0.42\sigma_i$	147 → 162
<i>glcG</i>	Unknown	-9.6	0.0094	$\mu_i - 1.93\sigma_i$	137 → 152
<i>gltA^c</i>	Citrate synthase	-5.2	0.0048	$\mu_i - 1.67\sigma_i$ $\mu_i - 1.51\sigma_i$	456 → 471 447 → 462
<i>hdhA</i>	NAD-dependent 7 α -hydroxysteroid dehydrogenase (dehydroxylation of bile acids)	-2.4	0.0097	$\mu_i - 1.59\sigma_i$	54 → 69
<i>icdA^c</i>	NADP-specific isocitrate dehydrogenase	-2.7	0.0248	$\mu_i - 0.38\sigma_i$	95 → 110
<i>kgtP</i>	α -Ketoglutarate permease	-2.4	0.0044	$\mu_i - 1.99\sigma_i$	110 → 125
<i>lldPRD^c</i>	L-lactate permease, regulator, and dehydrogenase	-11.4	0.0002	$\mu_i + 1.37\sigma_i$	118 → 133
<i>manX</i>	Mannose-specific PTS enzyme IIAB	+2.9	0.0267	$\mu_i - 1.96\sigma_i$ $\mu_i - 0.71\sigma_i$	167 → 182 11 → 26
<i>mdh</i>	Malate dehydrogenase	-3.1	0.0106	$\mu_i - 0.58\sigma_i$	250 → 265
<i>metK</i>	Methionine adenosyltransferase (EC 2.5.1.6)	+2.4	0.0447	$\mu_i - 1.34\sigma_i$	408 → 423
<i>ndh</i>	Respiratory NADH dehydrogenase	+3.5	0.0264	$\mu_i - 1.71\sigma_i$ $\mu_i - 0.33\sigma_i$ $\mu_i - 1.68\sigma_i$ $\mu_i - 0.40\sigma_i$ $\mu_i - 1.60\sigma_i$ $\mu_i - 1.73\sigma_i$	141 → 156 50 → 65 34 → 49 29 → 44 18 → 33 9 → 24
<i>ndk</i>	Nucleoside diphosphate kinase.	-3.4	0.0128	$\mu_i - 1.66\sigma_i$	96 → 111
<i>nikABCDE</i>	Nickel transport system (periplasmic binding protein, membrane proteins NikB and NikC), ATP-binding proteins NikD and NikE)	-2.5	0.0205	$\mu_i - 1.57\sigma_i$	113 → 128
<i>phnIJK</i>	Phosphonate metabolism; carbon-phosphorus lyase subunits PhnI and PhnJ, and ATP binding component of phosphonate transport	-3.9	0.0296	$\mu_i - 1.72\sigma_i$	-792 → -807

TABLE III—continued

Operon	Function of operon	Effect of ArcA-P ^a	<i>p</i> value (Student's <i>t</i> test)	Matrix score	Location of ArcA-P box ^b
<i>phoH</i>	PhoB-dependent, ATP-binding <i>pho</i> regulon component and induced by P starvation	-2.2	0.0485	$\mu_i - 0.42\sigma_i$	93 → 108
<i>ppa</i>	Inorganic pyrophosphatase	+2.5	0.0039	$\mu_i - 1.71\sigma_i$ $\mu_i - 1.54\sigma_i$	190 → 205 179 → 194
<i>rhaT</i>	L-Rhamnose/H ⁺ symporter	-2.3	0.0138	$\mu_i - 1.35\sigma_i$ $\mu_i - 1.54\sigma_i$ $\mu_i + 1.17\sigma_i$	217 → 232 216 → 231 207 → 222
<i>sdaC</i>	Probable serine transporter	+3.6	0.0418	$\mu_i - 1.33\sigma_i$ $\mu_i - 1.24\sigma_i$	327 → 342 223 → 238
<i>sdhCDAB</i> ^c	Succinate dehydrogenase (cytochrome b556 subunit SdhC, hydrophobic subunits SdhD and SdhB, and flavoprotein subunit SdhA)	-7.3	0.0146	$\mu_i - 1.51\sigma_i$ $\mu_i - 1.67\sigma_i$	247 → 262 238 → 253
<i>surA</i>	Survival protein	+2.2	0.0106	$\mu_i - 1.03\sigma_i$	7 → 22
<i>tolB</i>	Periplasmic protein involved in the TonB-independent uptake of group A colicins	+2.2	0.0145	$\mu_i - 1.78\sigma_i$	47 → 62
<i>treB</i>	Trehalose-specific PTS system enzyme II	+4.3	0.0223	$\mu_i - 1.39\sigma_i$	249 → 264
<i>uxaC</i>	Uronate isomerase	-2.3	0.0024	$\mu_i - 1.90\sigma_i$	115 → 130
<i>wrbA</i>	Trp repressor-binding protein; affects the association between Trp repressor and operators	-3.5	0.0195	$\mu_i - 0.63\sigma_i$	327 → 342
<i>ydeA</i>	Putative resistance/regulatory protein	-2.2	0.0003	$\mu_i - 1.72\sigma_i$	50 → 65
<i>ydeD</i>	Unknown	-2.1	0.0238	$\mu_i - 1.00\sigma_i$	149 → 164
<i>yefJ</i>	Putative creatinase, glycosyltransferase	+3.8	0.0156	$\mu_i - 0.93\sigma_i$ $\mu_i - 1.60\sigma_i$ $\mu_i - 1.87\sigma_i$	513 → 528 116 → 131 14 → 29
<i>ygaM</i>	Unknown	-2.8	0.0128	$\mu_i - 1.63\sigma_i$	29 → 44
<i>yhaU</i>	Putative transport protein	-2.1	0.0174	$\mu_i - 1.95\sigma_i$ $\mu_i - 1.51\sigma_i$	146 → 161 -5 → 10
<i>yheR-kefB</i>	Putative NAD(P)H oxidoreductase, and K ⁺ efflux protein	-2.9	0.0412	$\mu_i - 1.83\sigma_i$	101 → 116
<i>yjcZ</i>	Unknown	-2.8	0.0016	$\mu_i - 1.93\sigma_i$	151 → 166
<i>yjdF</i>	Unknown	-2.3	0.0075	$\mu_i - 1.71\sigma_i$	-52 → -67
<i>yjjN</i>	Putative oxidoreductase	-2.7	0.0025	$\mu_i - 1.87\sigma_i$	130 → 145

^a Transcriptional effect of ArcA-P is derived from the microarray experiments. (-): repressed, (+): activated.

^b Location of the box is in bp upstream of the start codon. A minus sign means that the box covers the coding sequence of the ORF.

^c ArcA-P-controlled operons whose promoter sequences were used to construct the ArcA-P weight matrix.

arep.med.harvard.edu/ecoli_matrices) may give us a clue as to the identity of these cryptic operons. To illustrate the utility of the matrix in this respect, we grew the *arcA*⁺ and *arcA*Δ strains anaerobically in minimal medium supplemented with 5% casein amino acids. RT-PCR analysis was performed on operons whose expression states were analyzed in cells grown in LB/xylose (Fig. 5A). In the presence of casein amino acids (Fig. 5B), *csdG* (regulator of curli biosynthesis, $\mu_i - 1.87\sigma_i$) and *ompT* (outer membrane endoprotease, $\mu_i - 1.06\sigma_i$) were respectively 17-fold and 15-fold activated by ArcA-P, whereas *fur* (regulator of ferric uptake, $\mu_i - 1.61\sigma_i$) and *fimB* (switcher of Type 1 fimbriae, $\mu_i - 1.13\sigma_i$) were, respectively, 7-fold and 9-fold repressed in an ArcA-P-dependent manner. The Arc modulon thus may contain beyond 80 members. Based on the current data, a conservative but reasonable assessment is that the Arc response pathway may recruit 100–150 operons. Modulon subsets are triggered or repressed depending on the growth conditions.

DISCUSSION

Effectiveness of ArcA-P Recognition Weight Matrix Screening—Matrix-based genome analysis has proven to be very meaningful to identify truly functional regulator binding sites in the *E. coli* chromosome (27, 28, 32). The *sdhCDAB* promoter, for instance, contains three ArcA-P protected segments as determined by footprinting experiments (15, 25). However, only one segment, covering the -35 to -10 region, was shown to be functional by genetic analysis (25). By our matrix criterion, this short region contains two ArcA-P boxes ($\mu_i - 1.67\sigma_i$, $\mu_i - 1.51\sigma_i$) (Table I and Fig. 2A), whereas the other two footprinted, non-functional regions contain none. In case of the *pflA* promoter region, four distinct ArcA-P protected segments were found (14). By our weight matrix criterion, two ArcA-P

boxes ($\mu_i + 0.05\sigma_i$, $\mu_i - 1.53\sigma_i$) (Table I and Fig. 2A) were identified only in the one region that showed the highest *in vitro* affinity for ArcA-P. One reason why ArcA-P matrix scanning is so reliable is that the matrix was derived from footprinting experiments. As was shown for the CpxR-P weight matrix, one can anticipate a strong correlation between the matrix score of an ArcA-P recognition site and the *in vitro* affinity of ArcA-P for that site (32).

The advantage of using the weighted ArcA-P matrix (Fig. 1) over the rigid ArcA-P consensus is illustrated by the 46,943 genomic hits that were obtained using the consensus (data not shown), whereas the matrix recognized 1,333 statistically significant binding sites scoring above $\mu_i - 2\sigma_i$ (Fig. 2, A and B). Importantly, all ArcA-P-controlled operons used to create the matrix are distributed uniformly throughout this Z-score window (Fig. 2A). Another indication of the soundness of the matrix screening is the increase in the percentage of ArcA-P recognition sites that are located intergenically with an increasing Z score (Fig. 2B). However, not all of the targets scoring above the cut-off level can be expected to be true binding sites as the set of eleven input operons may not accurately reflect the statistical distribution of all true targets.

Probing the Arc Modulon by Superimposing Matrix- and Microarray-derived Data—To identify true ArcA-P target operons, we complemented the matrix screening with array-based expression analysis. For this purpose, *arcA*⁺ and *arcA*Δ strains were grown anaerobically to activate the Arc signaling pathway. Because growth under aerobic conditions does not trigger the Arc response, Arc-controlled targets should not be identified under such conditions (33). Total RNA was isolated from the anaerobic *arcA*⁺ and *arcA*Δ cultures, converted to cDNA, and hybridized to oligonucleotide-based microarrays. A total of 4,345 ORFs (Fig. 7,

FIG. 8. Score-based distribution of ArcA-P-controlled operons identified in this study. Score-based distribution of ArcA-P recognition sites identified in promoter regions of operons whose expression is most likely under the direct control of ArcA-P.

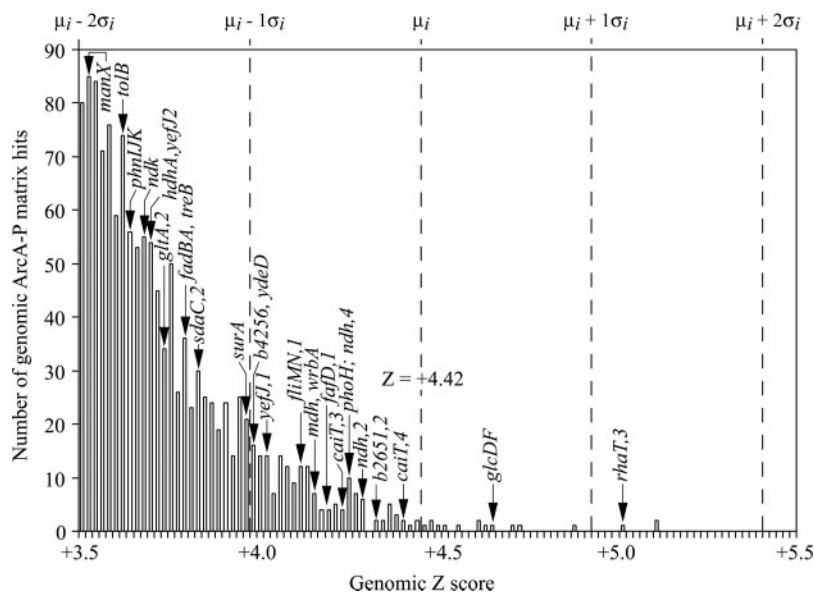


TABLE IV
Functional classification of operons whose expression is significantly affected by ArcA-P

Functional group ^a	Microarray ^b		Microarray + weight matrix ^c	
	Activated	Repressed	Activated	Repressed
Amino acid biosynthesis and metabolism	1	6	0	<i>cysM</i> , <i>wrbA</i>
Biosynthesis of cofactors, prosthetic groups, and carriers	3	0	0	0
Carbon compound catabolism	3	7	<i>metK</i>	<i>glcC</i> , <i>glcDF</i> , <i>glcD</i> , <i>uxaC</i>
Central intermediary metabolism	6	11	<i>ppa</i>	<i>agaCD</i> , <i>hdhA</i> , <i>phnIJK</i> , <i>phoH</i>
Cell processes (including adaptation, protection)	6	3	<i>caiT</i> , <i>ftsZ</i>	0
Cell structure	3	10	0	<i>crl</i> , <i>fliE</i> , <i>flmN</i>
DNA replication, recombination, modification, repair	6	4	0	0
Energy metabolism	13	23	<i>cydAB</i> , <i>dniR</i> , <i>ndh</i>	<i>aldA</i> , <i>gltA</i> , <i>icdA</i> , <i>mdh</i> , <i>sdhCDAB</i>
Fatty acid and phospholipid metabolism	1	3	<i>fabF</i>	<i>fadBA</i> , <i>fadD</i>
Hypothetical, unclassified, unknown	20	68	0	<i>b1012</i> , <i>b2099</i> , <i>b2651</i> , <i>b2670</i> , <i>b4256</i> , <i>ydeD</i> , <i>ygaM</i> , <i>yjzC</i> , <i>yjdF</i>
Membrane proteins	0	1	0	0
Nucleotide biosynthesis and metabolism	3	4	0	<i>ndk</i>
Other known genes	1	1	<i>surA</i>	0
Phage, transposon, or plasmid	4	1	<i>tolB</i>	0
Putative enzymes	6	26	<i>yefJ</i>	<i>b1671-ydhU</i> , <i>b2146</i> , <i>yheR</i> , <i>yjiN</i>
Putative regulatory proteins	1	11	0	<i>atoSC</i> , <i>ydeA</i>
Putative transport proteins	3	14	0	<i>yhaU</i>
Regulatory function	1	2	0	0
Transcription, RNA processing, and degradation	4	2	0	0
Translation, post-translational modification	42	1	0	0
Transport and binding proteins	9	19	<i>manX</i> , <i>sdaC</i> , <i>treB</i>	<i>cadB</i> , <i>celC</i> , <i>fepE</i> , <i>kgtP</i> , <i>lldPRD</i> , <i>nikABCDE</i> , <i>rhaT</i>
Others	1	17	0	0
Total	137	235	14	44

^a Functional annotations follow Blattner *et al.* (35).

^b Operons whose transcription is significantly affected by ArcA-P as measured by microarray analysis. The operons were identified after statistical filtration of the dataset (2-fold expression cut off, coefficient of variation <0.8, and Student's *t* test, *p* < 0.05).

^c Operons affected by ArcA-P (see microarray column) that contain in their promoter sequence an ArcA-P recognition site that scores above $\mu_i - 2\sigma_i$.

open bars) were obtained but data filtration analysis suggested that only 372 ORFs (Fig. 7, *filled bars*) were significantly affected by ArcA-P. Of these, 234 ORFs (or 63%) are repressed, whereas 137 are activated. This finding shows that 9% of all ORFs in *E. coli* (372/4,345) are affected, either directly or indirectly, by ArcA-P and that the Arc-mediated response translates into a net global downscaling of gene expression.

To determine which operons are under the direct control of ArcA-P, we matched the matrix-screening data with the array-derived expression profiles. 58 of the 372 ArcA-P affected operons contained one or more matrix boxes upstream of the start codon, strongly suggesting a direct involvement of ArcA-P in their expression. Importantly, these boxes (marked with an

arrow in Fig. 8) cover the window of statistically significant targets (scoring above $\mu_i - 2\sigma_i$) and overlie the positions of the matrix sites located in the promoters of the eleven input operons (compare Fig. 8 with Fig. 2A). In addition, expression analysis under a different growth condition (casein amino acid medium *versus* D-xylose in LB) revealed an additional four operons, with high-scoring matrix sites, that are most likely under the direct control of ArcA-P. The identification of *fur* (encoding a regulator of ferric uptake) as an ArcA-P target, is particularly worthy of note as it regulates the expression of *sodA* (13). *sodA*, encoding the manganese-containing superoxide dismutase, is induced under conditions of iron deprivation and is also under the transcriptional control of ArcA-P and

FNR (34), suggesting that the actions of Fur, ArcA-P, and FNR are carefully orchestrated during the adaptation to iron deprivation.

Computational analysis of the 62 Arc-controlled operons did not reveal any correlation between the score of a matrix site or its position relative to the start codon in an upstream sequence region, and the degree or type of Arc regulation at that site. Weight matrix profiling alone therefore does not yield decisive information about Arc activity at matrix-identified operons.

ArcA-P Controls a Functionally Diverse Set of Operons—The results described above illustrate that ArcA-P controls a myriad of operons that are involved in a variety of functions. Categorizing the newly identified operons into 21 functional groups (Table IV) according to Blattner *et al.* (35) shows that the Arc adaptational response extends beyond redox metabolism: *fliMN* and *fliE* encode proteins involved in flagella synthesis and switching, respectively; *ftsZ* encodes the protein that forms the cell division ring, *surA* encodes a protein that mediates stress-induced survival, *nikABCDE* encodes the nickel transport system, and *atoSC* is involved in short-chain fatty acid degradation. This broader than expected role is further supported by the recent findings that ArcA-P modulates the expression of virulence factors in *Vibrio cholerae* (36) and of membrane proteins in *Haemophilus influenzae* that are implicated in serum resistance (37). The disparate functions of the newly identified operons suggests that their organization into a single Arc modulon somehow must have created a selective advantage, resulting in a more competent response under suboptimal respiratory growth conditions.

The Arc-activated *cydAB* (1), and Arc-repressed *sodA* (13) and *lpdA* (26), are also under the direct negative control of FNR, a regulator that co-mediates gene expression in response to the redox conditions of growth (38). The identification by matrix screening and microarray analysis of *caiT*, *ndh*, *nikA*, *manX*, and *sdaC* as ArcA-P regulated targets (Table III) extends the overlap between the Arc and FNR networks as these operons were recently shown to be under the direct control of FNR (38). This integration of operons and response circuits seems to be a prevailing denominator as contemporary systems-biology approaches are consistently proving that regulatory pathways intertwine (*e.g.* the CpxA/CpxR, σ^B and σ^H pathways, Ref. 32; the Lrp leucine-response system with the FNR network, Ref. 38, and with stationary phase adaptation, Ref. 39; the oxidative stress, heat shock, and SOS responses, Ref. 40) to maximize cellular survival in the constantly changing environment.

In conclusion, our systems approach serves as a first step toward the long term goal of establishing a functional map of the Arc system, which appears more multifaceted than anticipated. Our findings extend the Arc modulon to beyond 80 members (30 known plus 55 novel operons), but for reasons acknowledged earlier, this number likely is a underestimation. Ad interim, a conservative but reasonable assessment is that around 100–150 operons may be under transcriptional control of ArcA-P. Further molecular characterization of candidate Arc-regulated targets identified in this study will ultimately reveal how and when the Arc system functions in coordinating cellular adaptation in situations of environmental adversity.

Acknowledgements—We thank our mentor Prof. Emeritus Edmund C. C. Lin for his continuous support and tutelage; Drs. Abigail McGuire, Sanchita Bhattacharya, Divyen Patel, and Jan Huang for help with the research; and Drs. Thomas Ferenci and Cathy Lee for insightful comments on the manuscript.

REFERENCES

- Cotter, P. A., and Gunsalus, R. P. (1992) *FEMS Microbiol. Lett.* **70**, 31–36
- Fu, H.-A., Iuchi, S., and Lin, E. C. C. (1991) *Mol. Gen. Genet.* **226**, 209–213
- Iuchi, S., and Lin, E. C. C. (1993) *Mol. Microbiol.* **9**, 9–15
- Iuchi, S., Aristarkhov, A., Dong, J. M., Taylor, J. S., and Lin, E. C. C. (1994) *J. Bacteriol.* **176**, 1695–1701
- Iuchi, S., and Lin, E. C. C. (1995) in *Two-Component Signal Transduction* (James, A. H., and Silhavy, T. J., eds) pp. 223–231, ASM Press, Washington, D. C.
- Lynch, A. S., and Lin, E. C. C. (1996) in *Regulation of Gene Expression in Escherichia coli* (Lin, E. C. C., and Lynch, A. S., eds) pp. 361–376, R. G. Landes Company, Austin, TX
- Iuchi, S., and Lin, E. C. C. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1888–1892
- Iuchi, S., Cameron, D. C., and Lin, E. C. C. (1989) *J. Bacteriol.* **171**, 868–873
- Georgellis, D., Kwon, O., and Lin, E. C. C. (2001) *Science* **292**, 2314–2316
- Pellicer, M. T., Fernandez, C., Badia, J., Aguilar, J., Lin, E. C. C., and Baldom, L. (1999) *J. Biol. Chem.* **274**, 1745–1752
- Chao, G., Shen, J., Tseng, C. P., Park, S.-J., and Gunsalus, R. P. (1997) *J. Bacteriol.* **179**, 4299–4304
- Park S. J., Tseng C. P., and Gunsalus, R. P. (1995) *Mol. Microbiol.* **15**, 473–482
- Tardat, B., and Touati, D. (1991) *Mol. Microbiol.* **5**, 455–465
- Drapal, N., and Sawers, G. (1995) *Mol. Microbiol.* **16**, 597–607
- Lynch, A. S., and Lin, E. C. C. (1996) *J. Bacteriol.* **178**, 6238–6249
- Strohmaier, H., Noiges, R., Kotschan, S., Sawers, G., Hogenauer, G., Zechner, E. L., and Koraimann, G. (1998) *J. Mol. Biol.* **277**, 309–316
- Colloms, S. D., Alén, C., and Sherratt, D. J. (1998) *Mol. Microbiol.* **28**, 521–530
- Lee, Y. S., Han, J. S., Jeon, Y., and Hwang, D. S. (2001) *J. Biol. Chem.* **276**, 9917–9923
- Berg, O. G., and von Hippel, P. H. (1998) *J. Mol. Biol.* **193**, 723–775
- Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) *Nat. Biotechnol.* **16**, 566–571
- Link, A. J., Phillips, D., and Church, G. M. (1997) *J. Bacteriol.* **179**, 6228–6237
- Sarmientos, P., Sylvester, J. E., Contente, S., and Cashel, M. (1983) *Cell* **32**, 1337–1346
- Pellicer, M. T., Lynch, A. S., De Wulf, P., Boyd, D., Aguilar, J., and Lin, E. C. C. (1999) *Mol. Gen. Genet.* **261**, 170–176
- Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F., and Wootton, J. C. (1993) *Science* **262**, 208–214
- Shen, J., and Gunsalus, R. P. (1997) *Mol. Microbiol.* **26**, 223–236
- Cunningham, L., Georgellis, D., Green, J., and Guest, J. R. (1998) *FEMS Microbiol. Lett.* **169**, 403–408
- McGuire A. M., De Wulf, P., and Lin, E. C. C. (1999) *Mol. Microbiol.* **32**, 217–221
- Robison, K., McGuire, A. M., and Church, G. M. (1998) *J. Mol. Biol.* **284**, 241–254
- Jeon, Y., Lee, Y. S., Han, J. S., Kim, J. B., and Hwang D. S. (2001) *J. Biol. Chem.* **276**, 40873–40879
- Miguel Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C., and Melton, D. A. (2002) *Science* **298**, 597–600
- Yuen, T., Wurmbach, E., Pfeffer, R. L., Ebersole, B. J., and Sealfon S. C. (2002) *Nucleic Acids Res.* **30**, e48
- De Wulf, P., McGuire A. M., Liu, X., and Lin, E. C. C. (2002) *J. Biol. Chem.* **277**, 26652–26661
- Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiura, M., Wanner, B. L., Mori, H., and Mizuno, T. (2002) *Mol. Microbiol.* **46**, 281–291
- Hassan, H. M., and Schrum, L. W. (1994) *FEMS Microbiol. Rev.* **14**, 315–323
- Blattner, F. R., Plunkett, G. 3rd, 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**, 1432–1434
- Sengupta N., Paul, K., and Chowdhury, R. (2003) *Infect. Immun.* **71**, 5583–5589
- De Souza-Hart, J. A., Blackstock, W., Di Modugno, V., Holland, I. B., and Kok, M. (2003) *Infect. Immun.* **71**, 163–172
- Salmon, K., Hung, S.-P., Mekjian, K., Baldi, P., Hatfield, G. W., and Gunsalus, R. P. (2003) *J. Biol. Chem.* **278**, 29837–29855
- Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O., and Matthews, R. G. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13471–13476
- Zheng, M., Wang, X., Templeton, L. J., Smulski, D. R., LaRossa, R. A., and Storz, G. (2001) *J. Bacteriol.* **183**, 4562–4570
- Schneider, T. D., and Stephens, R. M. (1990) *Nucleic Acids Res.* **18**, 6097–6100
- Kim, S. J., Han, Y. H., Kim, I. H., and Kim, H. K. (1999) *I.U.B.M.B. Life* **48**, 215–218