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 ofquinone 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 [letPRD], Ref. 4, sdxCDAB, Refs. 4 and 12, and sodA, Ref. 13) or activates others (e.g. cydAB, Ref. 1 and pfIA, 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′-TATTA3t-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 pfIA promoter (14). A subsequent homology search that included ArcA-P-protected promoter regions of cydAB, pfIA, glhA, lldPRD, sdxCDAB, 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-17), 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-buffereed (100 mM, pH7.4)uria-Bertani (LB) medium that was supplemented with 20 mM

Received for publication, December 9, 2003
Published, JBC Papers in Press, January 7, 2004, DOI 10.1074/jbc.M3134G4200

‡ 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.

This paper is available online at http://www.jbc.org
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 Using 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 µl 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 was used for normalizing the concentration of cDNAs 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 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 µl 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 was used for normalizing the concentration of cDNAs 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).

**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 sets, the ORF hybridization signals were passed through three steps of data filtration: (i) a coefficient of variation (CV) of ≥ 0.8; (ii) a mutant-to-wild type signal ratio larger than 2, positive or negative (log₂[MT:WT] > 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′-nGGTTAATTAn-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 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).

**TABLE I**

| Operon   | Putative ArcA-P binding site |
|----------|-----------------------------|
| **aldA** | 1) GTTAATTACGTAAC            |
| **cydAB** | 2) GATGATGATATCATTA          |
| **gltA**  | 3) GAAATGATGATATCATTA        |
| **icdA**  | 4) GATGATGATGATATCATTA       |
| **lpdA**  | 5) GATGATGATGATATCATTA       |
| **pfIA**  | 6) GATGATGATGATATCATTA       |
| **sodA**  | 7) GATGATGATGATATCATTA       |
| **traY**  | 8) GATGATGATGATATCATTA       |

*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.
Probing the Arc Modulon of E. coli

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 $Z = 0$. $\mu_i$ 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). $\sigma_i$ denotes the standard deviation below or above $\mu_i$. 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 the ArcA-P binding sites present in the promoter regions of the eleven ArcA-P-controlled input operons (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 $\mu_i$ 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, $\sigma_i$, of these input sites and ranked the genomic hits with a $Z$ value $\geq +3.5$ into four standard deviation groups relative to $\mu_i$ (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 ldPRD (matrix score: $\mu_i + 1.37\sigma_i$), cydAB ($\mu_i + 1.37\sigma_i$), and sodA ($\mu_i + 1.17\sigma_i$). Three of these intergenic operons were used to create the weight matrix (Table I). The second standard deviation group (Fig. 2, A and B, Table II, rows 4–15) contains 12 potential ArcA-P recognition sites that score between $\mu_i$ 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 xylFGRH operon. Of the 11 intergenic sites, 4 are localized in front of input operons: trcY ($\mu_i + 0.79\sigma_i$), cydAB ($\mu_i + 0.54\sigma_i$), glcDEFGB ($\mu_i + 0.42\sigma_i$), and pflA ($\mu_i + 0.50\sigma_i$), as well as sites in front of operons that are not connected with respiratory metabolism such as cst/ithA ($\mu_i + 0.40\sigma_i$) and insA_5/upsc ($\mu_i + 0.02\sigma_i$) (Table II). The third standard deviation group comprises 193 putative ArcA-P binding sites that score between $\mu_i$ 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 ($\mu_i - 0.38\sigma_i$), lppDA ($\mu_i - 0.33\sigma_i$, $\mu_i - 0.65\sigma_i$), 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 ($\mu_i - 1.01\sigma_i$, $\mu_i - 1.93\sigma_i$), gltAldhCDAB ($\mu_i - 1.51\sigma_i$, $\mu_i - 1.67\sigma_i$), 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 $\pm 200$ 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 A/T-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 or more boxes. 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-
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 independently (1, 4, 25). The results showed an 11-fold negative effect of ArcA-P controlled operon whose footprinted promoter sequence was used to construct the weight matrix.

Position analysis of matrix-identified ArcA-P recognition sites. Probing the Arc Modulon of E. coli 12591

| Operon | Function | Matrix score | Location of the ArcPb-P box | In vitro effect of ArcPb-P | In vitro binding of ArcPb-P |
|--------|----------|--------------|-----------------------------|--------------------------|----------------------------|
| 1)  | tPRD⁵ | t-Lactate permease, regulator and t-lactate dehydrogenase | μ₁ + 1.37σ₁ | 118 → 133 | Negative | Yes (15) |
| 2)  | cydAB⁶ | Cytochrome d oxidase | μ₁ + 1.37σ₁ | 570 → 585 | Positive | Yes (15) |
| 3)  | sodK⁶ | Mn-superoxide dismutase | μ₁ + 1.17σ₁ | 63 → 78 | Negative (13) | Yes (13) |
| 4)  | tpxyG⁹ | Thioredoxin-linked thiol peroxidase/α-Ala-β-L-Glu epimerase | μ₁ + 0.90σ₁ | 16 → 31 | Negative on | N.K.⁴ |
| 5)  | phnC⁶ | Phosphonate transporter subunit I | μ₁ + 0.60σ₁ | 56 → 71 | N.K. | N.K. |
| 6)  | cydAB⁶ | Cytochrome d oxidase | μ₁ + 0.54σ₁ | 592 → 607 | Positive (1) | Yes (15) |
| 7)  | glyDEFG⁶ | Glycolate catabolism | μ₁ + 0.42σ₁ | 89 → 104 | Negative (10) | Yes (10) |
| 8)  | cisthA² | N-acetylornithine-δ-ammonotransferase/exonuclease III | μ₁ + 0.46σ₁ | 36 → 51 | N.K. | N.K. |

⁴ Location of the box is in bp upstream of the start codon.
⁵ ArcA-P-controlled operon whose footprinted promoter sequence was used to construct the weight matrix.
⁶ The single box is located between 2 oppositely transcribed operons. Its location is in bp from the start codon of each operon.
⁷ 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).
such that they represent all candidate target operons that score above \(\mu_i - 2\sigma_i\). Under the implemented non-induced growth conditions, only \(\text{fimB}\) (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 \(\text{arcA}^-\) and \(\text{arcA}^+\) 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 \(2^{0.8}\) was installed (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.

![Diagram](image)

**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 \(\text{csgD, ompT, fur}\) and \(\text{fimB}\) 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 \(2^{0.8}\). See text for details.
on the arcΔ/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 $i_i > 2$ (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 as 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 true for numerous other operons whose induction conditions are unknown, leading to an underreckoning of the Arc modulon. Fortuitously, the matrix-screening profile (available at
| Operon   | Function of operon                                                                 | Effect of ArcA-P | p value (Student’s t test) | Matrix score | Location of ArcA-P box |
|----------|-----------------------------------------------------------------------------------|------------------|----------------------------|--------------|------------------------|
| agoCDI   | PTS system; N-acetylgalactosamine-specific IIIC component 1, IID component 1, and putative galactosamine-6-P isomerase | −2.4            | 0.0144 µ−1.93σ          | 104 → 119    | 3                      |
| aldA'    | NAD-linked aldehyde dehydrogenase                                                  | −6.9            | 0.0075 µ−0.69σ          | 44 → 59      | 4                      |
| atoSC    | Two-component signal transduction system (sensor AtoS, response regulator AtoC)    | −2.3            | 0.0174 µ−1.37σ          | −1298 → −1313| 5                      |
| b1012    | Putative oxidoreductase Fe-S subunit and unknown.                                  | −3.0            | 0.0209 µ−1.50σ          | 62 → 77      | 6                      |
| b1671-ydhU| Putative oxidoreductase Fe-S subunit and unknown.                                 | −2.6            | 0.0260 µ−1.80σ          | 344 → 359    | 7                      |
| b2099    | Unknown                                                                           | −2.6            | 0.0461 µ−1.99σ          | 75 → 90      | 8                      |
| b2146    | Putative oxidoreductase                                                           | −2.1            | 0.0008 µ−1.56σ          | 42 → 57      | 9                      |
| b2651    | Unknown                                                                           | −2.6            | 0.0137 µ−1.63σ          | 76 → 81      | 10                     |
| b2670    | Unknown                                                                           | −2.1            | 0.0154 µ−1.81σ          | 404 → 419    | 11                     |
| b4256    | Unknown                                                                           | −2.5            | 0.0297 µ−0.99σ          | 542 → 557    | 12                     |
| cadB     | Transport of lysine/cadaverine                                                    | −2.1            | 0.0062 µ−1.62σ          | 247 → 262    | 13                     |
| caiT     | Probable carnitine/betaine transporter involved in osmoprotection                | +2.2            | 0.0326 µ−1.35σ          | 190 → 205    | 14                     |
| celB     | PEP-dependent phosphotransferase enzyme II for cellobiose, arbutin, and salicin  | −2.3            | 0.0076 µ−1.33σ          | −71 → −86    | 15                     |
| cni      | Cytochrome c terminal oxidase                                                      | +5.2            | 0.0187 µ−0.54σ          | 592 → 607    | 16                     |
| cysM     | Cysteine synthase B, O-acetylsersinesulphodihydrolyse B                            | −2.2            | 0.0424 µ−1.63σ          | 116 → 131    | 17                     |
| dniB     | Transcriptional regulator for nitrite reductase (cytochrome c552)                | +3.1            | 0.0275 µ−1.89σ          | 260 → 275    | 18                     |
| fabF     | 3-Oxoacyl-[acyl-carrier-protein] synthase II                                       | +3.3            | 0.0043 µ−1.63σ          | 288 → 303    | 19                     |
| fabI     | 3-Hydroxyacyl-CoA dehydrogenase, and thiolase I                                   | −2.6            | 0.0007 µ−1.40σ          | 64 → 79      | 20                     |
| fabD     | Acetyl-CoA synthetase, long-chain fatty acid CoA ligase                            | −2.6            | 0.0037 µ−0.57σ          | 61 → 76      | 21                     |
| fepE     | Ferric enterobactin transport                                                     | −2.9            | 0.0566 µ−1.92σ          | 204 → 219    | 22                     |
| flgE     | Flagellar basal body component                                                    | −2.3            | 0.0226 µ−1.74σ          | −329 → −344  | 23                     |
| flgMN    | Flagellar motor switching and energy and flagellar switch protein                 | −2.2            | 0.0419 µ−0.70σ          | −950 → −965  | 24                     |
| ftsZ     | Cell division; forms circumferential ring; tubulin-like GTP-binding protein and GTPase | +2.5            | 0.0127 µ−1.82σ          | 319 → 334    | 25                     |
| glcC'    | Transcriptional activator for the glc operon                                      | −2.6            | 0.0489 µ+0.42σ          | 89 → 104     | 26                     |
| glcDF    | Glycollate oxidase subunit D, and glycolate oxidase iron-sulfur subunit            | −7.1            | 0.0009 µ+0.42σ          | 147 → 162    | 27                     |
| glcG     | Unknown                                                                           | −9.6            | 0.0094 µ−1.93σ          | 137 → 152    | 28                     |
| glhA     | Citrate synthase                                                                  | −5.2            | 0.0048 µ−1.67σ          | 456 → 471    | 29                     |
| hdhA     | NAD-dependent 7α-hydroxyketoester dehydrogenase (dehydroxylation of bile acids)  | −2.4            | 0.0097 µ−1.59σ          | 54 → 69      | 30                     |
| icdA'    | NADP-specific isocitrate dehydrogenase                                            | −2.7            | 0.0248 µ−0.38σ          | 95 → 110     | 31                     |
| iclP     | α-Ketoglutarate permease                                                          | −2.4            | 0.0044 µ−1.90σ          | 110 → 125    | 32                     |
| lldPRD   | 1-lactate permease, regulator, and dehydrogenase                                  | −11.4           | 0.0002 µ+1.37σ          | 118 → 133    | 33                     |
| manX     | Mannose-specific PTS enzyme IIA                                                   | +2.9            | 0.0267 µ−1.96σ          | 167 → 182    | 34                     |
| mdh      | Malate dehydrogenase                                                             | −3.1            | 0.0106 µ−0.58σ          | 250 → 265    | 35                     |
| metK     | Methionine adenosyltransferase (EC 2.5.1.6)                                       | +2.4            | 0.0447 µ−1.34σ          | 408 → 423    | 36                     |
| ndh      | Respiratory NADH dehydrogenase                                                    | +3.5            | 0.0264 µ−1.71σ          | 141 → 156    | 37                     |
| ndd      | Nucleoside diphosphate kinase                                                     | −3.4            | 0.0128 µ−1.66σ          | 96 → 111     | 38                     |
| nikABCDE  | Nickel transport system (periplasmic binding protein, membrane proteins NikB and NikC, ATP-binding proteins NikD and NikE) | −2.5            | 0.0205 µ−1.57σ          | 113 → 128    | 39                     |
| phnIJK    | Phosphonate metabolism; carbon-phosphorus lyase subunits PhnI and PhnJ, and ATP binding component of phosphonate transport | −3.9            | 0.0296 µ−1.72σ          | −792 → −807  | 40                     |
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 arcΔ 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/xylene (Fig. 5A). In the presence of casein amino acids (Fig. 5B), csgD (regulator of curli biosynthesis, _μ_ = 1.87σ) and ompT′ (outer membrane endoprotease, _μ_ = 1.06σ) were respectively 17-fold and 15-fold activated by ArcA-P, whereas fur (regulator of ferric uptake, _μ_ = 1.61σ) and fimB (switcher of type 1 fimbriae, _μ_ = 1.13σ) 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**

Efficacy 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 ( _μ_ = 1.67σ, _μ_ = 1.51σ) (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 ( _μ_ = 0.05σ, _μ_ = 1.53σ) (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 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 _μ_ = 2σ (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 _arcΔ_ 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 _arcΔ_ cultures, converted to cDNA, and hybridized to oligonucleotide-based microarrays. A total of 4,345 ORFs (Fig. 7,
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 \textit{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 $\mu_{t} - 2\sigma_{t}$) 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 \textit{fur} (encoding a regulator of ferric uptake) as an ArcA-P target, is particularly worthy of note as it regulates the expression of \textit{sodA} (13). \textit{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

### Table IV

| Functional group                          | Microarray | Microarray + weight matrix |
|-------------------------------------------|------------|-----------------------------|
| Amino acid biosynthesis and metabolism    | Activated  | Activated                   |
|                                           | Repressed  | Repressed                  |
| Biosynthesis of cofactors, prosthetic groups, and carriers | 1          | 6                           | cytM, urbA |
| Carbon compound catabolism                | 3          | 7                           | metK, glcC, glcDF, glcD, uxaC |
| Central intermediary metabolism           | 6          | 11                          | pga, agaCD, hdaA, phnIJK, pheH |
| Cell processes (including adaptation, protection) | 6          | 3                           | catT, flaZ |
| Cell structure                            | 3          | 10                          | cpr, flcE, flpMN |
| DNA replication, recombination, modification, repair | 6          | 4                           | |
| Energy metabolism                         | 13         | 23                          | cydAB, dniR, ndh, aldA, gltA, tcaA, mdh, sdiCDAB |
| Fatty acid and phospholipid metabolism    | 1          | 3                           | fabF, fadBA, fadD |
| Hypothetical, unclassified, unknown       | 20         | 68                          | b1012, b2099, b2631, b2670, b4256, ydaM, yiiZ, yiiF |
| Membrane proteins                         | 0          | 1                           | |
| Nucleotide biosynthesis and metabolism    | 3          | 4                           | ndk |
| Other known genes                         | 1          | 1                           | surA |
| Phage, transposon, or plasmid             | 4          | 1                           | tolB |
| Putative enzymes                          | 6          | 26                          | yefJ, b1671-ydhU, b2146, yheR, yiiN |
| Putative regulatory proteins              | 1          | 11                          | atoSC, ydeA |
| Putative transport proteins               | 3          | 14                          | yhaU |
| Regulatory function                       | 1          | 2                           | |
| Transcription, RNA processing, and degradation | 4          | 2                           | |
| Translation, post-translational modification | 42         | 1                           | |
| Transport and binding proteins            | 9          | 19                          | manX, sdaC, treB, cadB, celC, fepE, hgpP, lldPRD, nikABCDE, rhaT |
| Others                                    | 1          | 17                          | |
| 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 $\mu_{t} - 2\sigma_{t}$. 
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 adaptional response extends beyond redox metabolism: fltMN and fltE 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 catT, ndh, nikA, manX, and sdcA 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, ρ^p and σ^p 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 Bhattacharyya, Divyen Patel, and Jan Huang for help with the research; and Drs. Thomas Ferenci and Cathy Lee for insightful comments on the manuscript.

REFERENCES

1. Cotter, P. A., and Gunsalus, R. P. (1993) FEMS Microbiol. Lett. 70, 31–36
2. Fu, H.-A., Iuchi, S., and Lin, E. C. C. (1991) Mol. Gen. Genet. 226, 209–213
3. Iuchi, S., and Lin, E. C. C. (1993) Mol. Microbiol. 9, 9–15
4. Iuchi, S., Arstarkov, A., Dong, J. M., Taylor, J. S., and Lin, E. C. C. (1994) J. Bacteriol. 176, 1685–1701
5. 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.
6. 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
7. Iuchi, S., and Lin, E. C. C. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 1888–1892
8. Iuchi, S., Cameron, D. C., and Lin, E. C. C. (1989) J. Bacteriol. 171, 868–873
9. Georgellis, D., Kwon, O., and Lin, E. C. C. (2001) Science 292, 2314–2316
10. Pellicer, M. T., Fernandez, C., Badia, J., Aguilar, J., Lin, E. C. C., and Baldom, L. (1999) J. Biol. Chem. 274, 1745–1752
11. Chao, G., Shen, J., Tseng, C. P., Park, S.-J., and Gunsalus, R. P. (1997) J. Bacteriol. 179, 4289–4304
12. Park S.-J., Tseng C. P., and Gunsalus, R. P. (1995) Mol. Microbiol. 15, 473–482
13. Tardat, B., and Touati, D. (1991) Mol. Microbiol. 5, 455–465
14. Drapal, N., and Sawers, G. (1995) Mol. Microbiol. 16, 597–607
15. Lynch, A. S., and Lin, E. C. C. (1996) J. Bacteriol. 178, 6238–6249
16. Strohmaier, H., Noiges, R., Kotschan, S., Sawers, G., Hogenauer, G., Zechner, E. L., and Koraimann, G. (1996) J. Mol. Biol. 277, 369–316
17. Collins, S. D., Allen, C., and Sherratt, D. J. (1996) Mol. Microbiol. 26, 521–530
18. Lee, Y. S., Han, J. S., Jean, Y., and Hwang, D. S. (2001) J. Biol. Chem. 276, 9917–9923
19. Berg, O. G., and von Hippel, P. H. (1998) J. Mol. Biol. 193, 723–775
20. Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) Nat. Biotechnol. 16, 566–571
21. Link, A. J., Phillips, D., and Church, G. M. (1997) J. Bacteriol. 179, 6222–6237
22. Sarmentos, P., Sylvestre, J. E., Contente, S., and Cashel, M. (1983) Cell 33, 1337–1346
23. 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
24. Lawrence, G. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F., and Wootton, J. C. (1995) Science 262, 208–214
25. Shen, J., and Gunsalus, R. P. (1997) Mol. Microbiol. 26, 223–236
26. Cunningham, L., Georgellis, D., Green, J., and Guest, J. R. (1998) FEMS Microbiol. Lett. 169, 493–498
27. McGuire A. M., De Wulf, P., and Lin, E. C. C. (1999) Mol. Microbiol. 32, 217–221
28. Robinson, K., McGuire, A. M., and Church, G. M. (1998) J. Mol. Biol. 284, 241–254
29. Jeon, Y., Lee, Y. S., Han, J. S., Kim, J. B., and Hwang D. S. (2001) J. Biol. Chem. 276, 40873–40878
30. Miguel Ramalo-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C., and Melton, D. A. (2002) Science 295, 597–600
31. Yuen, T., Wurmbach, E., Pfeffer, R. L., Ebersole, B. J., and Sealfon S. C. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 13471–13476
32. Wootton, J. C. (1993) Nat. Genet. 5, 189–200
33. Chen, Y., and Sun, S.-P., Lin, E. C. C., and Gunsalus, R. P. (1995) J. Biol. Chem. 270, 10510–10516
34. 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.
35. Pellicer, M. T., Lynch, A. S., De Wulf, P., Boyd, D., Aguilar, J., Lin, E. C. C. (1999) Mol. Gen. Genet. 261, 170–176