0273-2289/04/118/0215/$25.00

Genomewide Expression Analysis in Amino Acid-Producing Bacteria Using DNA Microarrays

Tino Polen and Volker F. Wendisch*

Institute of Biotechnology 1, Research Center Jülich, Leo-Brandt-Str, D-52428 Jülich, Germany, E-mail: v.wendisch@fz-juelich.de

Received June 2, 2003; Accepted August 1, 2003

Abstract

DNA microarray technology has become an important research tool for biotechnology and microbiology. It is now possible to characterize genetic diversity and gene expression in a genomewide manner. DNA microarrays have been applied extensively to study the biology of many bacteria including *Escherichia coli*, but only recently have they been developed for the Gram-positive *Corynebacterium glutamicum*. Both bacteria are widely used for biotechnological amino acid production. In this article, in addition to the design and generation of microarrays as well as their use in hybridization experiments and subsequent data analysis, we describe recent applications of DNA microarray technology regarding amino acid production in *C. glutamicum* and *E. coli*. We also discuss the impact of functional genomics studies on fundamental as well as applied aspects of amino acid production with *C. glutamicum* and *E. coli*.

Index Entries: Genomewide gene expression analysis; DNA chips; DNA microarrays; cluster analysis; global regulatory mechanisms; amino acid production; *Corynebacterium glutamicum*; *Escherichia coli*.

Introduction

Amino acids are currently produced by either chemical synthesis (e.g., D,L-methionine); enzymatic catalysis (e.g., L-aspartate); extraction (e.g., L-arginine); or, most important, fermentation processes (reviewed in ref. *1*). For the fermentative production of L-glutamate and L-lysine, mainly *Corynebacterium glutamicum* strains are used, whereas for the production of L-phenylalanine and L-threonine, mainly *Escherichia coli* strains are used *(1)*. In addition to the production of amino acids, *E. coli* is employed for the

*Author to whom all correspondence and reprint requests should be addressed.

production of proteins (reviewed in ref. 2) and may find application for the production of 1,3-propanediol *(3)*, 1,2-propanediol *(4)*, succinic acid *(5,6)*, lactic acid *(7,8)*, ethanol *(9)*, and adipic acid *(10)*.

C. glutamicum was isolated as a natural L-glutamate producer, and high-producing strains for the production of L-glutamate, as well as of L-lysine and other amino acids, have been generated through classic mutation and selection *(1,11)*. With the advent of molecular biology methods for this organism, targeted metabolic pathway engineering became possible. Besides the molecular elucidation of amino acid biosynthesis and central metabolic pathways on the genetic and biochemical level, carbon flux analyses have allowed researchers to gain a quantitative understanding of corynebacterial central metabolism *(12)*. Furthermore, these studies have led to the rational improvement of *C. glutamicum* strains by "metabolic design" for the production of D-pantothenate, L-isoleucine, L-valine, L-threonine, and L-lysine *(12)*. Now that the *C. glutamicum* genome sequence has been determined *(13,14)*, genetic methods can be performed based on information on the whole genome. Thus, not only for the well-studied *E. coli* but also for *C. glutamicum* are genome-based methods applicable for strain development. Genomewide gene expression analyses with DNA microarrays allow the unraveling of global regulatory mechanisms and hold promise for targeted improvement of known pathways of central metabolism and amino acid biosynthesis.

Genomewide gene expression analysis is a cornerstone of systems biology that combines the global analysis of transcript abundance *(15,16)*; RNA turnover *(17)*; protein abundance *(18)*; protein modification and turnover *(19)*; interactions of proteins with other proteins *(20)*, RNA *(21)*, DNA *(22)*, or small molecules *(23)*; metabolite concentrations *(24)*; and fluxes *(12)* with mathematic modeling *(25)*, which should ultimately lead to a predictive understanding of the whole cell. Very few studies have combined at least some of the methods mentioned, such as transcriptomics and proteomics *(26–30)* or proteomics and enzyme activities *(31)*.

The technological aspects as well as the impact of DNA microarray analysis of *E. coli* on studies of gene expression, physiology, and pathogenicity *(32–37)* have been reviewed. The present review of DNA microarray analysis of *E. coli* and *C. glutamicum* focuses on studies related to amino acid biosynthesis, regulation by amino acids, and amino acid production.

Construction and Use of Whole-Genome *C. glutamicum* and *E. coli* DNA Microarrays

For the analysis of genomewide expression patterns, several techniques are currently in use: DNA microarrays *(15,16)*; proteome analyses *(18)*, which also allow detection of posttranslational modifications of proteins and which are established for *C. glutamicum (38,39)*; differential display *(40)*; serial analysis of gene expression (SAGE) *(41)*; and massively

parallel signature sequencing (MPSS) *(42)*. The latter methods involve cloning and sequencing of expressed RNAs, whereas DNA microarray experiments rely on rapid nucleic acid hybridization. It should be noted that different RNA levels indicate differential gene expression, but it remains to be determined whether they are owing to transcriptional control or regulated RNA degradation. DNA microarrays are based on either single-stranded DNA oligonucleotides *(16)* or double-stranded DNA (dsDNA) fragments *(15)* generated by polymerase chain reaction (PCR) amplification of each gene of an organism *(32,37,43)* or, in rare cases, of inserts from a random DNA library *(44)*. Whereas for *E. coli* both types of DNA microarray are in use *(36,45,46)*, so far genomewide expression analysis in *C. glutamicum* has been performed only using DNA microarrays with dsDNA *(26,47–49)*. The commercial *E. coli* oligonucleotide-based DNA microarray covers more than 1350 intergenic regions *(50)* and thus allows comprehensive monitoring of abundances of small noncoding RNAs, which recently have been identified to have roles in a great variety of processes, including transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability and translation, and protein degradation and translocation *(51)*.

PCR Product-Based Whole-Genome Arrays of C. glutamicum *and* E. coli

The laboratories of Pat Brown and David Botstein *(15,37,52)* pioneered the generation of PCR product-based whole genome microarrays and their use for gene expression analysis for *Saccharomyces cerevisiae*. Subsequently, genomewide expression analysis was established for model bacteria such as *E. coli* (e.g., see refs. *30* and *53–57*) but also for *C. glutamicum (26,47–49)* and the related *Mycobacterium tuberculosis (58)*.

For *E. coli*, PCR product-based genome arrays were made in several laboratories using either nylon membranes *(53)* or glass slides *(54,56,57,59)*; they can also be purchased (Sigma-Genosys, Woodlands, TX; Affymetrix, Santa Clara, CA). For making *E. coli* DNA microarrays in our laboratory *(57)*, the *E. coli* ORFmer primer set (Sigma-Genosys) was used to generate 4108 whole gene length PCR products from genomic DNA of 4290 genes. These and a number of negative controls, such as 100 *C. glutamicum* genes, as well as hybridization controls, such as *E. coli* genomic DNA, were spotted onto glass slides.

In *C. glutamicum*, as in other bacteria, first preliminary tests of parallel transcript profiling of a few genes were performed *(15,60–62)*. Now, whole-genome DNA microarrays have been developed for *C. glutamicum (26,47–49)* and allow one to perform genomewide gene expression analysis as well as comparative genomics experiments. The *C. glutamicum* whole-genome DNA microarray used in our laboratory comprises 3673 PCR products covering 2860 of the 2994 genes (506 in duplicate) described for the genome according to NCBI NC 003450 and 284 further putative coding sequences (23 in duplicate). Additionally, 100 spots of *C. glutamicum* genomic DNA

are present as normalization controls and 16 spots of λ DNA, 16 spots of *E. coli*, and 1 spot of the *E. coli aceK* gene as negative controls.

RNA Isolation and Fluorescent Labeling, Whole-Genome DNA Microarray Hybridization, and Fluorescence Scanning

In general, only minor modifications exist among the methods that have been developed for DNA microarray analysis of several microorganisms *(26,37,46)*. Typically, total RNA of sufficient quality and quantity can be purified after mechanical disruption of *C. glutamicum* or *E. coli* cells *(26,46)*. Quenching to minimize RNA degradation, a problem much more pronounced in bacteria than in eukarya, can be achieved by rapid cooling or by chemical treatment, such as with azide. A protocol for the selective isolation and purification of bacterial mRNA is also available *(46,63)*. Bacterial RNA can be labeled green or red fluorescently in a reverse transcription reaction primed with random dNTP hexamers, either directly using, e.g., Cy3-dUTP or Cy5-dUTP or indirectly using aminoallyl-dUTP followed by reacting the modified cDNA with monoreactive Cy3 or Cy5 dyes *(46)*. After purification of labeled cDNAs from dNTPs not incorporated, hybridization to the whole-genome DNA microarrays is performed in humid chambers. Hybridization periods may vary between 5 and 24 h, and stringent washing is performed under low-salt conditions *(46)*. Several commercial fluorescence scanners can be used to determine fluorescence at 635 and 532 nm with a resolution down to 5 μm. The recorded fluorograms can be exported to a number of file formats containing the fluorescence information of both fluorophors for each pixel.

Storage and Analysis of Raw Data From C. glutamicum *and* E. coli *Whole-Genome DNA Microarray Hybridizations*

Whole-genome DNA microarray experiments massively generate data, and to store primary data rather than derived values in a searchable format, suitable databank systems are needed. Several commercial (e.g., Axon Acuity, Lion arraySCOUT, Silicon Genetics GeneSpring) and academic solutions *(64–67)* are available. Additionally, with some basic knowledge and snippets of code freely available on the World Wide Web, one can compile custom-made database applications with the advantages of free choice of modular components at reasonable cost. A very low-price network solution can be assembled using mySQL running on a Linux host providing a Web server, combined with CGI programs (e.g., self-made Java or Kylix applications), generating dynamic Web browser forms to query, display, and analyze stored microarray data. Figure 1 summarizes some features of such a network solution. Raw primary data including position on the array and number of pixels representing the spot or the local background, and derived data such as fluorescence ratios, are stored in a relational mySQL database. All relevant experimental information (e.g., strain,

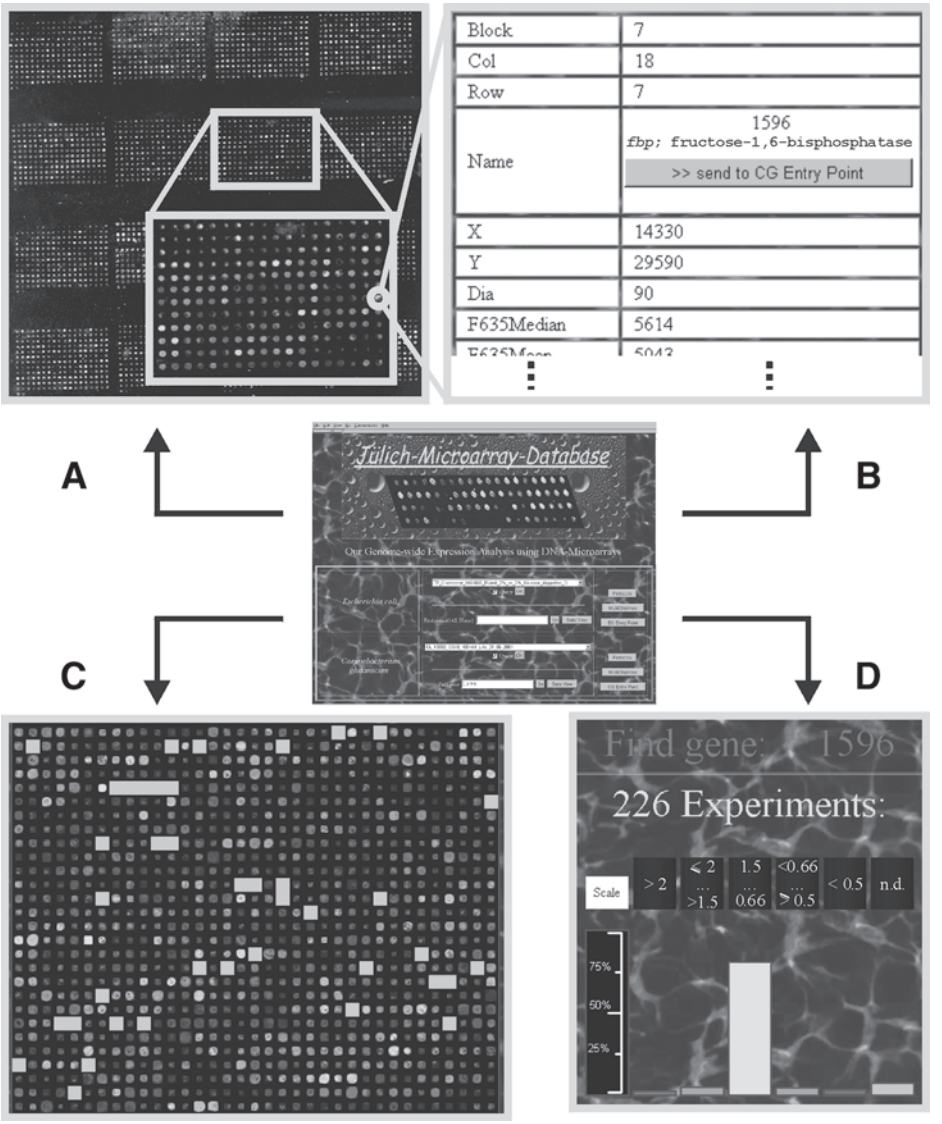

Fig. 1. Example of a relational database for storage and analysis of whole-genome DNA microarray data based on freely available database management system mySQL. A fluorogram of genomewide expression analysis using a whole-genome *C. glutamicum* DNA microarray **(A)** linked to information on raw data as well as background information **(B)** is shown. **(C)** Relative RNA levels of operons can be visualized as a genome map image. **(D)** Expression data of individual genes can be searched in all experiments, grouped (ratios <0.5 or between 0.5 and 0.66 or between 0.66 and 1.5 or between 1.5 and 2 or >2), and summarized graphically.

media, preculture conditions, growth curve, number of generations for which a balanced condition has been maintained, or time point of harvest after a particular stimulus) for each experiment are stored in an accompanying data table. Additionally, the fluorogram of the DNA microarray

hybridization (Fig. 1A) is stored, and hybridization spots on all visual representations are linked to the raw data and further available information about that gene in a Web-based manner (Fig. 1B). Genome map images *(55)* representing hybridization signals of each gene arranged according to the position on the genome (Fig. 1C) can also be generated and stored. To compare expression of a particular gene throughout the experimental conditions represented in the database, expression changes are summarized, grouped (<0.5, <0.66 and >0.5, between 0.66 and 1.5, >1.5 and <2, >2), and graphically represented (Fig. 1D).

Relative RNA levels of a gene are calculated from the net fluorescence intensities obtained in a DNA microarray hybridization experiment. It is important that relative RNA levels be calculated only from fluorescence signals clearly exceeding background noise, and when neither fluorescence signal shows a signal-to-noise ratio greater than a defined threshold (e.g., threefold), the signals should be considered too low to derive a relative RNA level for that gene. In fluorescence images scanned at high resolution (5 or 10 µm), hybridization spots are represented typically by at least 100 pixels. Calculations taking the variance of pixel information for hybridization spots into account, such as the ratio of medians, offer robust means to calculate relative RNA levels from hybridization images.

Bioinformatics Tools

Genomewide expression analyses pose the challenge of carefully controlling all experimental parameters starting from the precultivation to the final data analysis. The problem of noise in such analyses is obvious and needs to be addressed for interpretation of the DNA microarray results. Several statistical approaches have been established to identify statistically significant gene expression changes (reviewed, in ref. *68*). Among them, the two-sample *t*-test has been used to identify significant gene expression changes in genomewide expression analysis (e.g., *see* refs. *26*, *57*, and *69*). Clearly, it is a prerequisite to include more than two independent replicates in a genomewide expression analysis, and these replicates must be obtained from independent cultivations to take biologic as well as experimental noise into account.

A versatile tool to identify and visualize genomewide expression patterns of many different experimental conditions was introduced by Eisen et al. *(70)*. By hierarchical clustering of gene expression data, genes as well as experimental conditions are sorted according to similarities in gene expression patterns. This and similar analyses (reviewed in refs. *32* and *71*) are well suited for identifying operons based on experimentally determined expression (for *E. coli* s.) *(72)*, as can be demonstrated by hierarchical clustering of a set of 220 *C. glutamicum* whole-genome DNA microarray experiments from our laboratory (Fig. 2). Although none of the experimental conditions were aimed at analysis of the regulation of arginine biosynthesis (specifically, neither experiments using arginine auxotrophic mutants nor

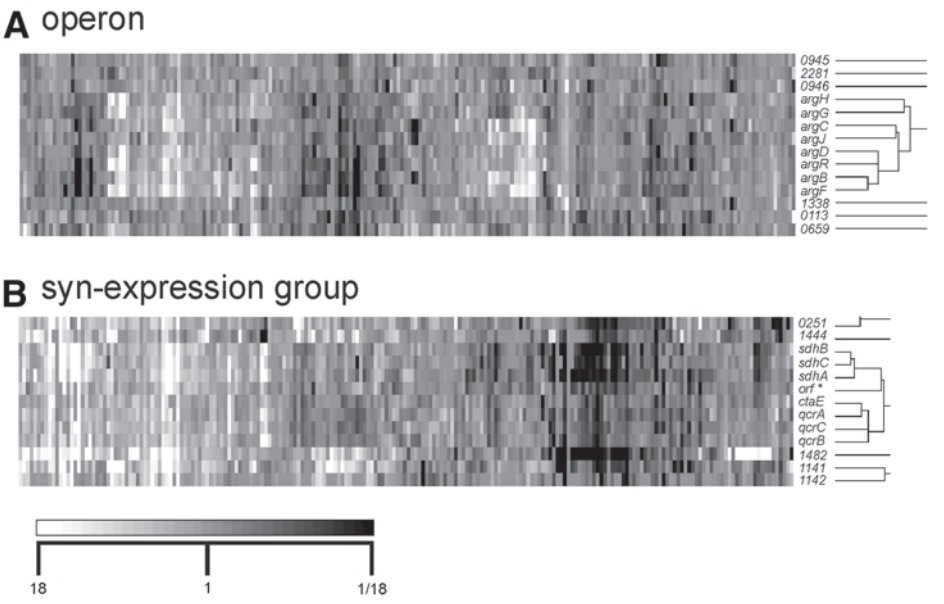

Fig. 2. Hierarchical cluster analysis of 220 global gene expression experiments using whole-genome *C. glutamicum* DNA microarrays. Two details show coexpression of the putative *argCJBDFRGH* operon or cluster **(A)** and of a syn-expression group comprising two operons separated by about 1.4 Mbp: the succinate dehydrogenase operon *sdhCAB* with an adjacent open reading frame and the *ctaE-qcrCAB* operon encoding subunit III of cytochrome aa_3 and the three subunits of the cytochrome bc_1 complex **(B)**. Cluster analysis was performed on 736 genes that were reliably detected in more than 180 of the 220 experiments and that in one or more of them showed an at least twofold change in RNA level. Genes referred to by gene name or NCgl number of NC 003450 are in lines and experiments in columns. The scale depicts the gray color coding for relative RNA levels.

experiments using arginine as nitrogen or carbon source or other medium component were included), the subtle gene expression changes of the putative *argCJBDFRGH* operon of *C. glutamicum* revealed that expression of each gene within this group is more similar to each other than to any other gene (Fig. 2A). It becomes evident as well that some operons physically distant on the genome are exhibiting very similar expression patterns and form syn-expression groups *(73)*. It remains to be shown whether syn-expression groups, such as that comprising the *C. glutamicum ctaE-qcrCAB (74)* and *sdhCAB* operons (Fig. 2B), reflect regulation by a common regulatory protein and thus are part of the same regulon and/or whether they respond to a common stimulus and thus are part of the same stimulon. With the help of pattern recognition as realized by clustering global gene expression data, one can readily deduce hypotheses on global gene regulation in bacteria such as *E. coli* and *C. glutamicum* and help to plan experiments aimed at their verification (or falsification).

Applications and Perspectives

Comparative Genomics or Genomotyping of C. glutamicum *and* E. coli *Strains*

By hybridization of labeled genomic DNA to DNA microarrays it is possible to identify differences among genomes. Oligonucleotide DNA microarrays allow detection of genomic differences down to the single-nucleotide level, i.e., the identification of single-nucleotide polymorphisms *(75)*. Behr et al. *(76)* employed DNA microarray analyses for the identification of genome differences of the different *M. tuberculosis* Bacille Calmette-Guerin strains used for vaccination throughout the world. This study demonstrated that DNA microarrays based on PCR products can be used efficiently for the detection of gene deletions or amplifications although they offer a resolution of only about 1 kb. Although this might pertain to a lesser extent to *C. glutamicum*, so-called wild-type laboratory strains of *E. coli* K-12 differ considerably *(77–79)*. To illustrate this fact as well as the ease with which genomic differences are detected by DNA microarray analysis, Fig. 3A depicts signal differences in a comparison of genomic DNA from *E. coli* wild-type strains MG1655 *(80)* and LJ110 *(81)*. A deletion of 14 contiguous genes (b3709–b3723) in *E. coli* LJ110 at 84 min on the chromosomal map was readily detected, at least in our stock. PCR using primers within genes b3708 and b3724, respectively, resulted in a 2.5-kb PCR product rather than the 16.7 kilobases expected for MG1655, thus verifying the DNA microarray results. Similarly, using *C. glutamicum* whole-genome DNA microarrays, it was possible to detect chromosomal rearrangements such as deletions at the gene level. In a comparison of genomic DNA from *leuA* deletion strains and their respective isogenic parent strains (WT, WTΔ*leuA*, MH20-22B, and MH20-22BΔ*leuA*), the absence or presence of the *leuA* gene was readily and specifically detected (Fig. 3B).

In future applications, it will be important to identify genomic differences between amino acid-producing strains obtained by repeated mutageneses and selections and the respective wild-type strain. The deletions or amplifications identified in such a manner will be tested individually regarding their relevance for amino acid production. Introduction of these mutations into the wild-type strain, alone and in combination, should give rise to well-defined, high-level amino acid-producing strains. This approach, also named "inverse metabolic engineering" *(82)* or "genome breeding," recently led to a first success *(83)* when the combinatory introduction of a limited number of previously known, beneficial mutations present only in L-lysine production strains into the *C. glutamicum* wild-type strain endowed the resulting defined strain with the capability to produce ~100 g/L of L-lysine.

Whole-genome DNA microarrays can also be used to determine which genes confer beneficial or detrimental traits, such as growth (dis)advantages or improved/decreased product yields. This form of parallel gene-trait mapping consists of a genome-altering step to generate a

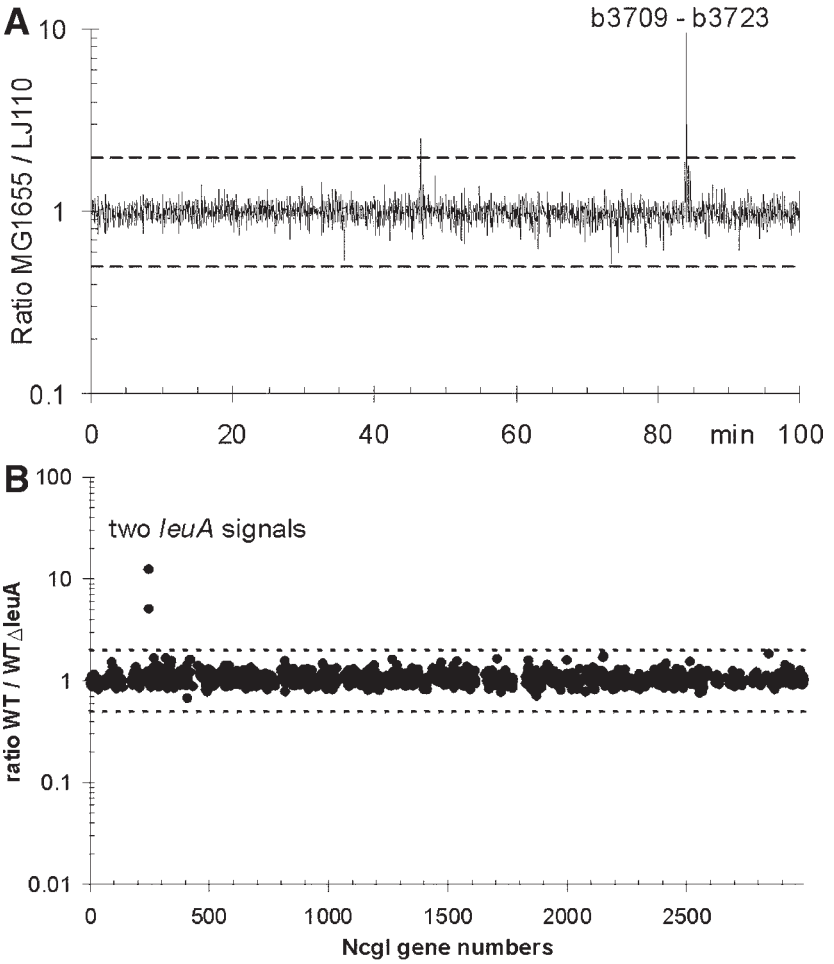

Fig. 3. Comparative genomics of *E. coli* **(A)** and *C. glutamicum* **(B)** strains. (A) Deletion of a 16-kbp chromosomal region in *E. coli* LJ110 detected by hybridization signals in a DNA microarray experiment comparing genomic DNA from *E. coli* MG1655 and *E. coli* LJ110. Ratios of hybridization signals are shown according to the 100-min chromosome of *E. coli* MG1655. Dotted lines indicate ratios of hybridization signals of 2 and 1/2, respectively. (B) The ratios of hybridization signals in DNA microarray experiments comparing genomic DNA from *C. glutamicum* ATCC 13032 (WT) to WTΔ*leuA* are shown according to the respective gene numbers in NCBI NC 003450. Dotted lines indicate ratios of hybridization signals of 2 and 1/2, respectively. The absence or presence of the *leuA* gene was detected with both of the two *leuA* PCR products that cover different parts of the *leuA* gene.

pool of mutants, such as by deletion mutagenesis or by transforming a strain with a genomic library, followed by an enrichment of a subpopulation containing the trait-conferring gene(s) by selection and using DNA microarrays to identify the enriched or counterselected genes *(84)*. This method can readily be applied, e.g., for the identification of conditionally essential genes *(85)* or genes conferring antibiotic resistance *(86)*.

However, whereas the diagnostic readout of such experiments by DNA microarray analysis is straightforward, the challenge for the biotechnologist will be to work out schemes that allow to select and compare subpopulations with improved product yields.

Global Gene Regulation of Nitrogen Assimilation

Ammonium assimilation occurs in reactions of glutamate and glutamine biosynthesis. It is evident that under amino acid production conditions, a high flux of ammonium assimilation must be maintained. Under nitrogen-sufficient conditions, assimilation of ammonium in *E. coli* occurs via glutamate dehydrogenase, but when nitrogen is scarce, ammonium is assimilated in reactions of glutamine synthetase and glutamate synthase *(87)*. Enteric bacteria perceive external nitrogen limitation as a drop in the intracellular glutamine concentration *(88)*. As a consequence, transcription of genes under control of nitrogen regulatory protein C (NtrC) is activated *(87)*. DNA microarray experiments revealed that NtrC directly controls expression of 50 genes in 16 operons *(55)*. The products of NtrC-activated genes are involved in assimilation of ammonium into the intermediates glutamate and glutamine or yield these intermediates catabolically or spare the requirement for them. Among the products of NtrC-activated genes are the ammonia transporter (AmtB), glutamine synthetase (encoded by glnA), and several amino acid and peptide permeases as well as catabolic enzymes. Moreover, NtrC activates transcription of the genes encoding the regulatory proteins NtrB, NtrC, GlnK, and Nac. Whereas NtrC activates transcription of σ^{54}-dependent genes, the nitrogen assimilation control (Nac) protein activates transcription of 25 σ^{70}-dependent genes in 9 operons *(55)*. Thus, in *E. coli*, transcription of 75 genes in 25 operons is controlled by NtrC.

In *C. glutamicum*, which is able to use ammonium, allantoin, glutamate, glutamine, ornithine, and urea as nitrogen sources *(89)*, expression of nitrogen-controlled genes is regulated by the global repressor protein AmtR *(90)*. AmtR controls expression of the *amt-ocd-soxA* operon containing the ammonium transporter gene *amt*, the *amtB-glnK-glnD* operon with the second ammonium transporter gene and genes for a P*II*-like protein and a uridylyltransferase, the *gltBD* operon for glutamate synthase, and the *glnA* gene for glutamine synthetase *(89)*. A binding motif of AmtR has been identified, and its occurrence within the *C. glutamicum* genome predicts that the regulon of the central nitrogen regulator AmtR is small and contains less than 20 members *(89)*, although DNA microarray experiments have not addressed this issue directly.

Global Gene Regulation by Amino Acids

E. coli and *C. glutamicum* are able to grow on a number of amino acids as nitrogen and/or carbon source. Thus, it is not surprising that global regulatory systems control gene expression in response to the intracellular concentration of a particular amino acid. In *E. coli*, the leucine-responsive regulatory protein Lrp is a global transcriptional regulator *(91)*. The branched-chain amino acid L-leucine is a coregulator that, depending on the target promoter, potentiates,

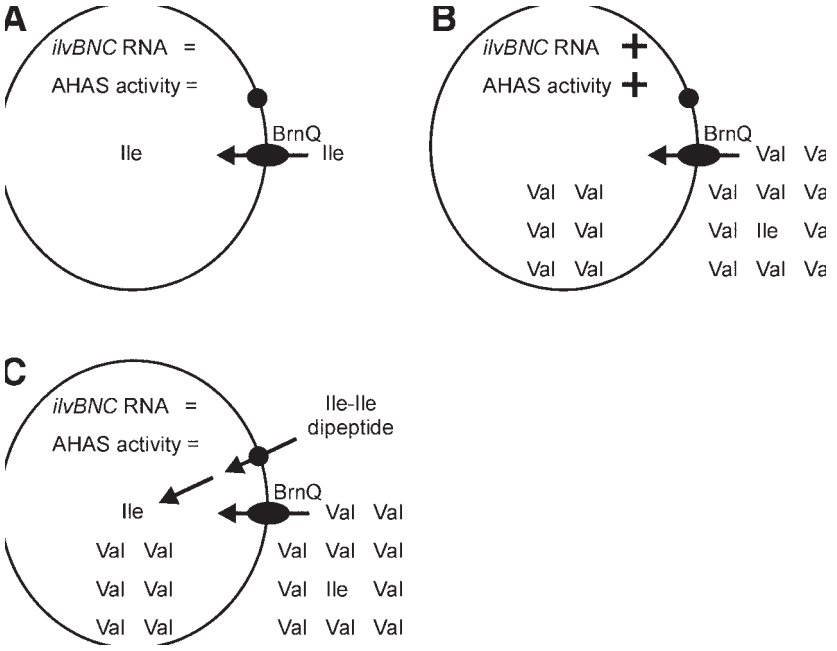

Fig. 4. Effect of L-valine on *C. glutamicum* Val1. The valine production strain *C. glutamicum* Val1 contains an *ilvA* deletion and, thus, needs supplementation with isoleucine, which is taken up via the branched-chain amino acid transporter BrnQ. In comparison to growth without added L-valine **(A)**, *C. glutamicum* Val1 shows growth inhibition, an increased *ilvBNC* RNA level, an increased acetohydroxy acid synthase (AHAS) protein level, and increased AHAS activity during growth in the presence of valine **(B)**. When isoleucine is supplemented in the form of isoleucyl-isoleucine dipeptide, which is taken up by a dipeptide or oligopeptide transport system, L-valine exerts no effects on growth, *ilvBNC* RNA level, AHAS protein level, or AHAS activity **(C)**.

antagonizes, or is without effect on Lrp activity *(91)*. Through DNA microarray analyses, more than 400 genes were found to be Lrp responsive, and of these about one-half changed expression only when leucine was present in the medium *(92,93)*. Surprisingly, among them many genes known to be induced in the stationary growth phase were found *(93)*.

In *C. glutamicum*, Lrp-responsive gene expression has not yet been studied. However, effects of the addition of the branched-chain amino acid L-valine to the growth medium on global gene expression have been characterized using DNA microarray analysis *(26)*. Metabolic pathway engineering of the *C. glutamicum* wild-type strain ATCC 13032, which does not overproduce valine, led to a valine production strain *(94)*. Growth of this valine-producing strain as well as of the L-lysine-producing strain MH20-22B was inhibited by valine added to the growth medium, whereas growth of the wild type was unaffected *(26,95)*. DNA microarray analysis of the valine stress response of the valine-producing strain identified gene expression changes indicating an intracellular isoleucine limitation. In comparison to the absence of valine (Fig. 4A), growth in the presence of valine

resulted in increased expression of the branched-chain amino acid biosynthesis genes *ilvBNC* (Fig. 4B), which was confirmed by proteome analysis and determination of enzyme activities, and of an isoleucine tRNA synthetase gene. Subsequently, it was shown that isoleucine limitation as a consequence of the addition of valine was linked to the *ilvA* deletion in the isoleucine auxotrophic valine-producing strain and could be overcome by the addition of high isoleucine concentrations. Finally, the fact that supplementation of the valine-producing strain with the dipeptide isoleucylisoleucine relieved the inhibitory effect of valine (Fig. 4C) identified competition for uptake of isoleucine by the carrier BrnQ *(96)*, which transports all branched-chain amino acids, as the cause of valine inhibition *(26)*. Based on the surprising finding that valine increased *ilvBN* expression, it was shown that the addition of external valine stimulated valine production by the valine-producing strain. This indicates that the activity of the *ilvBN*-encoded acetohydroxyacid synthase activity may still be a limiting factor for valine production in this valine-producing strain.

In *E. coli* the tryptophan stimulon and the regulon of the tryptophan-activated *trp* repressor TrpR have been characterized by DNA microarray analysis *(54)*. In response to externally added tryptophan, 30 genes changed expression. Of these, the *trpEDCBA*, *aroH*, *mtr*, *trpR*, and *aroL* operons belong to the TrpR regulon, whereas the operons *aroF-tyrA*, *aroL*, *aroG*, and *aroP* are known to be controlled by TyrR *(97)*. These genes are involved in the biosynthesis of chorismate, the key intermediate of aromatic amino acid biosynthesis; in the biosynthesis of tryptophan; and in the uptake of aromatic amino acids *(98)*. Only the *tnaAB* operon, which is responsible for tryptophan degradation *(98)*, showed increased expression when tryptophan was added to the medium.

The effects of other aromatic amino acids on the growth medium have not been published. Preliminary results indicate that the addition of low concentrations of phenylalanine, which only have a mild effect on growth, result in quite a number of gene expression changes (unpublished results). The affected genes were members of the TyrR regulon as well as genes that likely changed expression owing to indirect effects, indicating parallels to the response of *E. coli* to tryptophan.

Global gene expression changes as a consequence of the production of aromatic amino acids such as phenylalanine or other aromatic compounds such as *p*-hydroxybenzoic acid in *E. coli*, however, still await investigation.

Effects of Intermediates and of Byproducts of Amino Acid Biosynthesis on Global Gene Expression

An optimal producer strain should be designed not to accumulate any intermediates and not to form any significant byproducts. However, whereas this goal is achievable theoretically, most production strains are expected to show higher intracellular concentrations of at least some intermediates in comparison with the nonproducing parent strain as well as the formation of byproducts. Accumulation of intermediates, such as

ketobutyrate *(99)*, or byproducts, such as acetate *(57)*, might result in perturbation of growth and production. Thus, the overall productivity of such a process may critically depend on avoidance of the formation of perturbing intermediates or byproducts. It is conceivable that production strains obtained by undirected mutagenesis and selection for increased productivity may have acquired traits that endow them with some form of resistance to the perturbations by increased intermediate or byproduct concentrations.

Recently, we determined that the addition of shikimate, an intermediate of the aromatic amino acid biosynthesis pathway, to the *E. coli* wild-type strain LJ110 did not change global gene expression (unpublished results). This is consistent with the fact that shikimate is taken up into *E. coli (100)* but does not perturb growth.

A more severe problem is the formation of acetate as a byproduct of many aerobic production processes in *E. coli (101)*. Acetate exerts negative effects on growth and productivity *(101)* independent of the fact that its formation acidifies the medium *(57)*, which, of course, is readily counteracted by monitoring and controlling the pH in the production process. Fed-batch processes have been developed that allow minimization of acetate formation during production *(102)*. Nevertheless, simply the presence of low concentrations of acetate at neutral external pH results in global gene expression changes *(57)*. The adaptive response to acetate involved genes in three categories. First, the RNA levels for chemotaxis and flagellum genes increased. Accordingly, the expression of chromosomal reporter gene fusions to selected flagella genes as well as swimming motility as the phenotypic manifestation increased after adaptation to acetate. Second, the expression of many genes that are involved in the uptake and utilization of carbon sources decreased, indicating some kind of catabolite repression by acetate. Third, the expression of at least some genes of the general stress response increased, but the respective increases were more pronounced after short-term exposure to acetate than for the long-term adaptive response to acetate. The gene expression changes after adaptation to acetate were not caused solely by uncoupling or osmotic effects but represented specific characteristics of the long-term response of *E. coli* to either compound. It is obvious that simply the presence of acetate reprograms carbon metabolism and that increased synthesis of flagella is costly for *E. coli (57)*. Extending this analysis, the genes that change expression not only when acetate is present but, rather, when acetate is produced as a byproduct remain to be identified.

Direct Comparison of Global Gene Expression of an Amino Acid-Producing Strain With Its Nonproducing Parent Strain

Classic strain optimization by mutation and selection was successful from an industrial point of view, but not very rewarding for the microbiologist attempting to understand amino acid production. Then, based on the knowledge of the most important pathways in the conversion of the

carbon source to the wanted product, targeted metabolic engineering allowed improvement of amino acid production on a rational basis. This rational approach came with a price, though, because it was reductionistic and did not take into account any influence of the majority of the genes within the genomes of *E. coli* or *C. glutamicum* (i.e., 2000–3500 genes), gene products, or their activities. With the diagnostic power of *E. coli* and *C. glutamicum* whole-genome DNA microarrays, it appears possible to widen the rational metabolic engineering approach from its traditionally constricted focus to the full genome level.

To date, a comparison of global gene expression differences between an amino acid-producing strain and its nonproducing parent strain has not been published. Obviously, owing to the classic strain optimization in addition to the desired mutations improving productivity irrelevant mutations have accumulated and will have an impact on genomewide expression patterns. Additionally, many genes likely changed expression not as a prerequisite of amino acid production but, rather, as a direct or indirect consequence. Thus, although very promising, gene expression analysis of amino acid-producing strains will only yield novel target genes for strain development if care is taken to distinguish expression changes leading to improved amino acid production from those arisen as a consequence of improved amino acid production.

Clearly, the impact of genomewide expression analyses on our understanding of amino acid production by *C. glutamicum* or *E. coli* from a whole-cell point of view is just emerging. However, we expect that DNA microarray analysis will have a central role as enabling technology for the optimization of amino acid production.

Future Directions

We can now use *C. glutamicum* and *E. coli* whole-genome DNA microarrays to characterize global gene regulatory mechanisms on the RNA level, and subsequently we can determine whether changed RNA levels are owing to transcriptional regulation and/or differential RNA stability. However, to gain a complete view of global regulation, we must follow a more comprehensive approach by complementing transcriptomics with proteomics and biochemistry to study regulation of translation, protein folding, degradation, or modification and with metabolic flux analysis and metabolomics to study in vivo activities and allosteric control of enzymes.

We expect that DNA microarrays will be a cornerstone in a comprehensive experimental approach to increase our knowledge of fundamental and applied aspects of the physiology of the amino acid-producing bacteria. They hold promise for enabling strain and process optimization for amino acid production with *C. glutamicum* or *E. coli*.

Acknowledgments

We would like to thank Hermann Sahm for continuous support and our colleagues Doris Rittmann, Georg Sindelar, Christian Lange,

Takeru Ishige, Andrea Veit, Malgorzata Krause, Andreas Krug, Sandra Knebel, and Corinna Stansen for their work in various projects. We also gratefully acknowledge the support of the European Union (QLK3-2000-00497); the German Ministry of Science BMBF (Netzwerk Genomforschung an Mikroorganismen); the Deutsche Bundesstiftung Umwelt (Verbund Biokatalyse); DSM Biotech, Jülich; and Degussa AG, Düsseldorf.

References

1. Eggeling, L., Pfefferle, W., and Sahm, H. (2001), in *Basic Biotechnology*, Ratledge, C., ed., Cambridge University Press, Cambridge, UK, pp. 281–302.
2 Swartz, J. R. (2001), *Curr. Opin. Biotechnol.* **12(2),** 195–201.
3. Nakamura, C. E., Gatenby, A. A., Hsu, A. K.-H., et al. (2000), US patent 6,013,494.
4 Altaras, N. E. and Cameron, D. C. (1999), *Appl. Environ. Microbiol.* **65(3),** 1180–1185.
5. Donnelly, M. I., Millard, C. S., Clark, D. P., Chen, M. J., and Rathke, J. W. (1998), *Appl. Biochem. Biotechnol.* **70–72,** 187–198.
6 Vemuri, G. N., Eiteman, M. A., and Altman, E. (2002), *J. Industrial Microbiol. Biotechnol.* **28(6),** 325–332.
7 Chang, D. E., Jung, H. C., Rhee, J. S., and Pan, J. G. (1999), *Appl. Environ. Microbiol.* **65(4),** 1384–1389.
8 Zhou, S., Causey, T. B., Hasona, A., Shanmugam, K. T., and Ingram, L. O. (2003), *Appl. Environ. Microbiol.* **69(1),** 399–407.
9 Underwood, S. A., Zhou, S., Causey, T. B., Yomano, L. P., Shanmugam, K. T., and Ingram, L. O. (2002), *Appl. Environ. Microbiol.* **68(12),** 6263–6272.
10 Niu, W., Draths, K. M., and Frost, J. W. (2002), *Biotechnol. Prog.* **18(2),** 201–211.
11 Sahm, H., Eggeling, L., Eikmanns, B., and Kramer, R. (1996), *Ann. NY Acad. Sci.* **782,** 25–39.
12. Sahm, H., Eggeling, L., and de Graaf, A. A. (2000), *Biol. Chem.* **381(9–10),** 899–910.
13 Tauch, A., Homann, I., Mormann, S., et al. (2002), *J. Biotechnol.* **95(1),** 25–38.
14 Kalinowski, J., Bathe, B., Bartels, D., et al. (2003), *J. Biotechnol.* **104,** 5–25.
15 Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995), *Science* **270(5235),** 467–470.
16 Lockhart, D. J., Dong, H., Byrne, M. C., et al. (1996), *Nat. Biotechnol.* **14(13),** 1675–1680.
17 Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S., and Cohen, S. N. (2002), *Proc. Natl. Acad. Sci. USA* **99(15),** 9697–9702.
18 Pedersen, S., Bloch, P. L., Reeh, S., and Neidhardt, F. C. (1978), *Cell* **14(1),** 179–190.
19 Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002), *Nat. Biotechnol.* **20(3),** 301–305.
20 Gavin, A. C., Bosche, M., Krause, R., et al. (2002), *Nature* **415(6868),** 141–147.
21 Takizawa, P. A., DeRisi, J. L., Wilhelm, J. E., and Vale, R. D. (2000), *Science* **290(5490),** 341–344.
22 Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001), *Nature* **409(6819),** 533–538.
23 Zhu, H., Bilgin, M., Bangham, R., et al. (2001), *Science* **293(5537),** 2101–2105.
24 Raamsdonk, L. M., Teusink, B., Broadhurst, D., et al. (2001), *Nat. Biotechnol.* **19(1),** 45–50.
25 Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002), *Nature* **420(6912),** 190–193.
26 Lange, C., Rittmann, D., Wendisch, V. F., Bott, M., and Sahm, H. (2003), *Appl. Environ. Microbiol.* **69(5),** 2521–2532.
27 Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999), *Mol. Cell. Biol.* **19(3),** 1720–1730.
28 Yoshida, K., Kobayashi, K., Miwa, Y., et al. (2001), *Nucleic Acids Res.* **29(3),** 683–692.
29 Jurgen, B., Hanschke, R., Sarvas, M., Hecker, M., and Schweder, T. (2001), *Appl. Microbiol. Biotechnol.* **55(3),** 326–332.

30 Yoon, S. H., Han, M. J., Lee, S. Y., Jeong, K. J., and Yoo, J. S. (2003), *Biotechnol. Bioeng.* **81(7),** 753–767.
31 Peng, L. and Shimizu, K. (2003), *Appl. Microbiol. Biotechnol.* **61(2),** 163–178.
32 Rhodius, V., Van Dyk, T. K., Gross, C., and LaRossa, R. A. (2002), *Annu. Rev. Microbiol.* **56,** 599–624.
33 Khodursky, A. B., and Bernstein, J. A. (2003), *Trends Genet.* **19(3),** 113–115.
34 Conway, T. and Schoolnik, G. K. (2003), *Mol. Microbiol.* **47(4),** 879–889.
35 Schoolnik, G. K. (2002), *Adv. Microb. Physiol.* **46,** 1–45.
36 Picataggio, S. K., Templeton, L. J., Smulski, D. R., and LaRossa, R. A. (2002), *Methods Enzymol.* **358,** 177–188.
37 Ye, R. W., Wang, T., Bedzyk, L., and Croker, K. M. (2001), *J. Microbiol. Methods* **47(3),** 257–272.
38 Hermann, T., Pfefferle, W., Baumann, C., et al. (2001), *Electrophoresis* **22(9),** 1712–1723.
39 Schaffer, S., Weil, B., Nguyen, V. D., Dongmann, G., Gunther, K., Nickolaus, M., Hermann, T., and Bott, M. (2001), *Electrophoresis* **22(20),** 4404–4422.
40 Liang, P. and Pardee, A. B. (1992), *Science* **257(5072),** 967–971.
41 Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995), *Science* **270(5235),** 484–487.
42 Brenner, S., Johnson, M., Bridgham, J., et al. (2000), *Nat. Biotechnol.* **18(6),** 630–634.
43 van Hal, N. L., Vorst, O., van Houwelingen, A. M., Kok, E. J., Peijnenburg, A., Aharoni, A., van Tunen, A. J., and Keijer, J. (2000), *J. Biotechnol.* **78(3),** 271–280.
44 Hayward, R. E., Derisi, J. L., Alfadhli, S., Kaslow, D. C., Brown, P. O., and Rathod, P. K. (2000), *Mol. Microbiol.* **35(1),** 6–14.
45 Rhodius, V. A. and LaRossa, R. A. (2003), Curr Opin Microbiol. **6(2)**, 114–119.
46 Khodursky, A. B., Bernstein, J. A., Peter, B. J., Rhodius, V., Wendisch, V. F., and Zimmer, D. P. (2003), *Methods Mol. Biol.* **224,** 61–78.
47 Muffler, A., Bettermann, S., Haushalter, M., Horlein, A., Neveling, U., Schramm, M., and Sorgenfrei, O. (2002), *J. Biotechnol.* **98(2–3),** 255–268.
48 Ishige, T., Krause, M., Bott, M., Wendisch, V. F., and Sahm, H. (2003), *J. Bacteriol.* **185(15),** 4519–4529.
49 Gerstmair, R., Wendisch, V. F., Schnicke, S., Ruan, H., Farwick, M., Reinscheid, D., and Eikmanns, B. J. (2003), *J. Biotechnol.* **104(1–3),** 99–122.
50 Tjaden, B., Haynor, D. R., Stolyar, S., Rosenow, C., and Kolker, E. (2002), *Bioinformatics* **18(Suppl. 1),** S337–S344.
51 Storz, G. (2002), *Science* **296(5571),** 1260–1263.
52 DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997), *Science* **278(5338),** 680–686.
53 Richmond, C. S., Glasner, J. D., Mau, R., Jin, H., and Blattner, F. R. (1999), *Nucleic Acids Res.* **27(19),** 3821–3835.
54 Khodursky, A. B., Peter, B. J., Cozzarelli, N. R., Botstein, D., Brown, P. O., and Yanofsky, C. (2000), *Proc. Natl. Acad. Sci. USA* **97(22),** 12,170–12,175.
55 Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., Bender, R. A., and Kustu, S. (2000), *Proc. Natl. Acad. Sci. USA* **97(26),** 14,674–14,679.
56 Wei, Y., Lee, J. M., Richmond, C., Blattner, F. R., Rafalski, J. A., and LaRossa, R. A. (2001), *J. Bacteriol.* **183(2),** 545–556.
57 Polen, T., Rittmann, D., Wendisch, V. F., and Sahm, H. (2003), *Appl. Environ. Microbiol.* **69(3),** 1759–1774.
58 Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., and Schoolnik, G. K. (1999), *Proc. Natl. Acad. Sci. USA* **96(22),** 12,833–12,838.
59 Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiura, M., Wanner, B. L., Mori, H., and Mizuno, T. (2002), *Mol. Microbiol.* **46(1),** 281–291.
60 de Saizieu, A., Certa, U., Warrington, J., Gray, C., Keck, W., and Mous, J. (1998), *Nat. Biotechnol.* **16(1),** 45–48.
61 Loos, A., Glanemann, C., Willis, L. B., O'Brien, X. M., Lessard, P. A., Gerstmeir, R., Guillouet, S., and Sinskey, A. J. (2001), *Appl. Environ. Microbiol.* **67(5),** 2310–2318.

62 Hayashi, M., Mizoguchi, H., Shiraishi, N., Obayashi, M., Nakagawa, S., Imai, J., Watanabe, S., Ota, T., and Ikeda, M. (2002), *Biosci. Biotechnol. Biochem.* **66(6),** 1337–1344.
63 Wendisch, V. F., Zimmer, D. P., Khodursky, A., Peter, B., Cozzarelli, N., and Kustu, S. (2001), *Anal Biochem.* **290(2),** 205–213.
64 Brazma, A., Parkinson, H., Sarkans, U., et al. (2003), *Nucleic Acids Res.* **31(1),** 68–71.
65 Diehn, M., Sherlock, G., Binkley, G., et al. (2003), *Nucleic Acids Res.* **31(1),** 219–223.
66 Gollub, J., Ball, C. A., Binkley, G., et al. (2003), *Nucleic Acids Res.* **31(1),** 94–96.
67 Conway, T., Kraus, B., Tucker, D. L., Smalley, D. J., Dorman, A. F., and McKibben, L. (2002), *Biotechniques* **32(1),** 110, 112–114, 116, 118, 119.
68 Pan, W. (2002), *Bioinformatics* **18(4),** 546–554.
69 Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehle, M. M., Paegle, E. S., and Hatfield, G. W. (2000), *J. Biol. Chem.* **275(38),** 29,672–29,684.
70 Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998), *Proc. Natl. Acad. Sci. USA* **95(25),** 14,863–14,868.
71 Tamames, J., Clark, D., Herrero, J., Dopazo, J., Blaschke, C., Fernandez, J. M., Oliveros, J. C., and Valencia, A. (2002), *J. Biotechnol.* **98(2–3),** 269–283.
72 Sabatti, C., Rohlin, L., Oh, M. K., and Liao, J. C. (2002), *Nucleic Acids Res.* **30(13),** 2886–2893.
73 Niehrs, C. and Pollet, N. (1999), *Nature* **402(6761),** 483–487.
74 Niebisch, A. and Bott, M. (2001), *Arch. Microbiol.* **175(4),** 282–294.
75 Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P., and Collins, F. S. (1996), *Nat Genet.* **14(4),** 441–447.
76 Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., and Small, P. M. (1999), *Science* **284(5419),** 1520–1523.
77 Jishage, M. and Ishihama, A. (1997), *J. Bacteriol.* **179(3),** 959–963.
78 Ochman, H. and Jones, I. B. (2000), *EMBO J.* **19(24),** 6637–6643.
79 Perkins, J. D., Heath, J. D., Sharma, B. R., and Weinstock, G. M. (1993), *J. Mol. Biol.* **232(2),** 419–445.
80 Blattner, F. R., Plunkett, G. 3rd, Bloch, C. A., et al. (1997), *Science* **277(5331),** 1453–1474.
81 Zeppenfeld, T., Larisch, C., Lengeler, J. W., and Jahreis, K. (2000), *J. Bacteriol.* **182(16),** 4443–4452.
82 Bailey, J. E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W. A., and Tsai, P. S. (1996), *Biotechnol. Bioeng.* **52,** 109–121.
83 Ohnishi, J., Mitsuhashi, S., Hayashi, M., Ando, S., Yokoi, H., Ochiai, K., and Ikeda, M. (2002), *Appl. Microbiol. Biotechnol.* **58(2),** 217–223.
84 Cho, R. J., Fromont-Racine, M., Wodicka, L., Feierbach, B., Stearns, T., Legrain, P., Lockhart, D. J., and Davis, R. W. (1998), *Proc. Natl. Acad. Sci. USA* **95(7),** 3752–3757.
85 Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2001), *Proc. Natl. Acad. Sci. USA* **98(22),** 12,712–12,717.
86 Gill, R. T., Wildt, S., Yang, Y. T., Ziesman, S., and Stephanopoulos, G. (2002), *Proc. Natl. Acad. Sci. USA* **99(10),** 7033–7038.
87 Reitzer, L. (2003), *Annu. Rev. Microbiol.* **57,** 155–176.
88 Ikeda, T. P., Shauger, A. E., and Kustu, S. (1996), J Mol Biol. **259(4),** 589–607.
89 Burkovski, A. (2003), *Arch. Microbiol.* **179(2),** 83–88.
90 Jakoby, M., Nolden, L., Meier-Wagner, J., Kramer, R., and Burkovski, A. (2000), *Mol. Microbiol.* **37(4),** 964–977.
91 Calvo, J. M. and Matthews, R. G. (1994), *Microbiol. Rev.* **58,** 466–490.
92 Hung, S. P., Baldi, P., and Hatfield, G. W. (2002), *J. Biol. Chem.* **277(43),** 40,309–40,323.
93 Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O., and Matthews, R. G. (2002), *Proc. Natl. Acad. Sci. USA* **99(21),** 13,471–13,476.
94 Radmacher, E., Vaitsikova, A., Burger, U., Krumbach, K., Sahm, H., and Eggeling, L. (2002), *Appl. Environ. Microbiol.* **68(5),** 2246–2250.
95 Eggeling, L., Morbach, S., and Sahm, H. (1997), *J. Biotechnol.* **56(Pt. 7),** 167–182.
96 Tauch, A., Hermann, T., Burkovski, A., Kramer, R., Puhler, A., and Kalinowski, J. (1998), *Arch. Microbiol.* **169(4),** 303–312.

97 Yang, J., Wang, P., and Pittard, A. J. (1999), *J. Bacteriol.* **181(20),** 6411–6418.
98. Pittard, A. J. (1996), in Escherichia coli *and* Salmonella: *Cellular and Molecular Biology*, Neidhardt, F. C., ed., ASM Press, Washington, DC, pp. 458–484.
99 LaRossa, R. A., Van Dyk, T. K., and Smulski, D. R. (1987), *J. Bacteriol.* **169(4),** 1372–1378.
100 Whipp, M. J., Camakaris, H., and Pittard, A. J. (1998), *Gene* **209(1–2),** 185–192.
101 Kirkpatrick, C., Maurer, L. M., Oyelakin, N. E., Yoncheva, Y. N., Maurer, R., and Slonczewski, J. L. (2001), *J. Bacteriol.* **183(21),** 6466–6477.
102. Lee, S. Y. (1996), *Trends Biotechnol.* **14(3),** 98–105.