J. Gen. Appl. Microbiol., **51**, 93–103 (2005)

Full Paper

Analysis of reading frame and expressional regulation of randomly selected promoter-proximal genes in *Escherichia coli*

Ali Azam Talukder,* Satoshi Yanai, and Mamoru Yamada

Department of Biological Chemistry, Faculty of Agriculture, Yamaguchi University, Yamaguchi 753–8515, Japan

(Received December 26, 2003; Accepted December 28, 2004)

The expression of seventy-seven randomly cloned genes of *Escherichia coli* was examined following a variety of treatments including heat shock, glucose starvation, phosphate starvation, ammonium starvation or osmotic shock, with the aid of *lacZ* reporter gene protein fusions on multicopy plasmids. Two of 77 genes (*amr* and *yigL*) had not previously been identified as protein encoding open-reading frames (ORFs) in annotations of the *E. coli* genome database. Thirteen genes exhibited significant changes in expression in response to at least one of the treatments, and six of them appeared to be controlled by more than one σ (sigma) factor of RNA polymerase. This study thus allows us not only to identify the reading frame of the genomic genes but also to support the hypothesis earlier proposed that a significant proportion of genes in *E. coli* are involved in adaptations to various stresses to which the organism is likely to be exposed in the environment.

Key Words——*Escherichia coli*; gene expression; genomic gene; global regulation; LacZ protein fusion

Introduction

Escherichia coli possesses about 4,300 genes (Blattner et al., 1997) and is expected to have more than several hundred genes proximal to their own promoters in operon organization. Most of the genes in the organism are just predicted by computer-aided analysis after the genomic information has completely been determined. These predicted genes should thus be proven by identification of gene products in addition to elucidation of their functions and expressional regulation. We have constructed a mini-library of the *E. coli* genomic genes, in which each clone has part of a genomic gene fused in frame with the *lacZ* gene in addition to its promoter and operator (Talukder et al., 1994). Analysis of the library allows us to identify reading frames of the cloned genes, confirm their protein productivity and examine their expressional control. From the library, 5 genes were found to be negatively regulated by RpoS, responsible for gene expression as a σ factor in the stationary phase (Talukder et al., 1996) and two of them, named *ssnA* and *ves*, were extensively analyzed. *ssnA* was shown to be involved in cell death, and *ves* was expressed at a low temperature at the beginning of the stationary phase (Yamada et al., 1999, 2002).

Here, we have analyzed about 10% promoter-proximal genes in *E. coli* in respect to their reading frames, protein productivity and expressional changes under various conditions. The genes responsive to heat shock and glucose starvation were then further analyzed using mutant strains of *rpoH* and *Δcya* which encode a heat shock promoter specific σ^{32} subunit of RNA polymerase and adenylate cyclase, respectively. With these results and by comparison with data reported previously, we showed the possible promoter

*Address reprint requests to: Dr. Ali Azam Talukder, Graduate School of Science, Department of Science, Nagoya University, Furou-cho, Chikusa-ku, Nagoya 464–8602, Nagoya, Japan.

Tel: +81–(0)–52–789–2507 Fax: +81–(0)–52–789–2963

E-mail: azam@biol1.bio.nagoya-u.ac.jp

organization of the responsive genes.

Materials and Methods

Bacterial strains and plasmids. The *E. coli* K-12 strains used in this study were W3110, wild type; MC1000, *araD139* Δ(*ara-leu*)*7697* Δ*lacX74 galU galK rpsL* (Casadaban and Cohen, 1980); JM103, Δ(*lac-pro*) *thi strA supE endA sbcB hsdR4* F′ *traD36 proAB*$^+$ *lacI*^{q}*ZΔM15* (Messing, 1983); KY1436, F$^-$*araD* Δ*lac rpsL rpoH165 zhf50*::*Tn10* (Yura et al., 1984); CA8306, Hfr Hayes Sm^s B1$^-$Δ*cya* provided by J. Beckwith; YU379, MC1000 *rpoH*; and YU230, MC1000 Δ*cya*. The last two strains were constructed by P1 transduction (Miller, 1992). The library analyzed here was constructed as described previously (Talukder et al., 1994).

DNA manipulations, sequencing and gene mapping. Conventional recombinant DNA techniques (Sambrook et al., 1989) were used. Restriction enzymes, T4 DNA ligase, *Taq* DNA polymerase and the nucleotide sequencing kit (TaKaRa Shuzo, Kyoto, Japan) were used according to the specifications of the manufacturer. Nucleotide sequence around the junction with the *lacZ* gene in each clone was determined by the dideoxy-chain termination method (Sanger et al., 1977) and their deduced amino acid sequences were compared with those listed in the EMBL, GenBank, SWISS-PROT and NBRF-PIR databases on a computer network. The *E. coli* ordered phage library (Kohara et al., 1987) was hybridized with the DNA fragments as a probe, and the detection was carried out by using the ECL kit (Amersham, Buckinghamshire, England).

Bacterial growth and enzyme assay. Cells harboring each clone plasmid were grown to late exponential phase at 37°C in LB medium (1% Bacto tryptone, 0.5% yeast extract, 0.5% NaCl) containing ampicillin (100 μg/ml). The cells were then diluted 100-fold with MOPS minimal medium (0.4% glucose, 10 mM NH_4Cl, 1.32 mM K_2HPO_4, 0.53 mM $MgCl_2$, 0.28 mM K_2SO_4, 0.01 mM $FeSO_4$, 0.5 mM $CaCl_2$, 50 mM NaCl, 4 mM Tricine, 40 mM MOPS [3-(*N*-Morpholino) propanesulfonic acid], pH 7.2, 0.002% leucine, and 0.002% thiamine) and grown for 16 h at 37°C or 30°C. The cells were harvested, washed twice and resuspended in MOPS minimal medium. For starvation experiments, one component in the MOPS minimal medium was decreased as follows, 0.13 mM K_2HPO_4, 2.4 mM NH_4Cl or 0.04% glucose for phosphate, ammonium, or glucose starvation, respectively. For osmotic shock, the washed cells were resuspended in the LB medium containing 200 mM NaCl. Successively, incubation was done at 37°C for 4 h in the different starvation experiments and for 16 h in the osmotic shock. For heat shock treatment, the washed cells were resuspended in LB medium and incubated for 1 h at 42°C as a heat shock or at 30°C as a control or for 0–2 h in time course experiments. Samples were then withdrawn at the time indicated and cooled to prevent further growth by immersion in an ice bucket. After lysed or dead cells were removed by washing and centrifugation, cooled cell suspension was adjusted to an OD_{600} of 0.28–0.80 and subjected to β-galactosidase assay (Miller, 1992). β-Lactamase assay was done as described previously (Katayama and Nagata, 1990).

Results

Identification of genes encoding proteins and their reading frames

In each plasmid clone in the library constructed, 0.5–1.0 kb genomic DNA fragments of the *E. coli* W3110 were inserted in front of the *lacZ* gene lacking its promoter and the first 8 amino acid residues on pMC1396 [Fig. 1, for details see reference (Casadaban et al., 1980)]. When a fragment encompassing from the promoter to part of coding region of a promoter-proximal gene was inserted and the gene was situated in frame to the *lacZ* gene, β-galactosidase was produced which in turn makes the colony blue on plates containing X-gal. Thus, such blue color formation indicated that a gene encoding a protein was cloned, and after nucleotide sequencing, its reading frame was determined by comparison of that of the *lacZ* gene. As a result, out of 4,500 colonies grown on the plates, 106 were obtained as Lac$^+$. Among the 106, 29 colonies were found to be duplicate or false positive. Finally, we isolated 77 clones all of which should encode a protein, and determined the nucleotide sequence around the junction between an inserted gene and the *lacZ* gene. By this procedure, the reading frames of 75 genes were found to agree with those published or listed in databases, and the remaining two were identified as new genes (Table 1). One of the new genes occurs between *yihR* and *yihS* and the other between *pbpA* and *ybeA*. Out of them, 31 genes were first confirmed by this analysis, whose open reading frames had only been predicted.

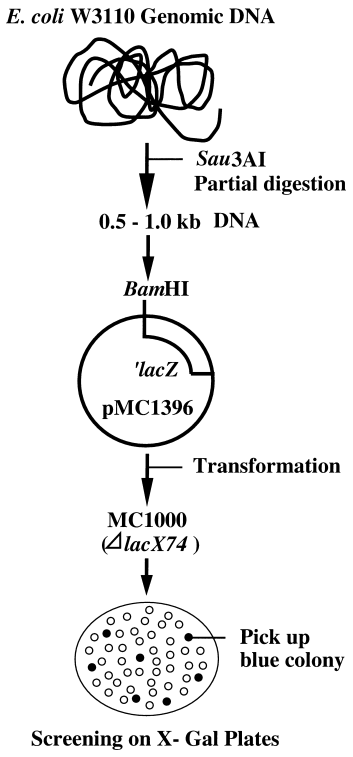

Fig. 1. Construction of an *E. coli* promoter library.

E. coli W3110 genomic DNA was partially digested with *Sau*3AI, and 0.5–1.0 kb DNA fragments were isolated and inserted by ligation into the *Bam*HI site in front of the *'lacZ* gene on pMC1396 (for details see reference, Talukder et al., 1994).

On the basis of the size of cloned DNA fragments and the gene organization around each analyzed gene, which could be visualized on databases, it was expected that at least 71 genes analyzed here possess their own promoters (data not shown). Interestingly, an intragenic promoter was found in the *cheA* gene, which may be responsible for a short product, $CheA_S$ (see Fig. 4). Because, *cheA* is transcribed from a promoter upstream of the *motA* gene and produces overlapping products, $CheA_L$ and $CheA_S$ (Sanatinia et al., 1995) but the upstream promoter was not included in the inserted DNA fragment in pYU34 (data not shown).

Expression of the genomic genes cloned under various conditions

Cells harboring each clone were subjected to heat or osmotic shock treatments in LB medium and phosphate, glucose or ammonium starvation in MOPS minimal medium (Fig. 2). The effects of these treatments were examined by measuring β-galactosidase activity, which was done more than three times for each fusion gene and evaluated by the following statistical procedure. The standard error averaged 16% and was less than 20% in 85% of the cases and less than 25% in 90% of the cases. Nearly the same standard error was obtained from assay of a control enzyme, β-lactamase, which is encoded by *bla* on the plasmid bearing the fusion gene (data not shown). We thus regarded the induction ratios of >1.5 or <0.75 as significant and those of >2.0 or <0.5 as highly significant. In addition to eleven highly significant clones, two significant clones with the ratios of >1.5 were further analyzed in this study.

Heat shock

Six clones pYU3, pYU14, pYU27, pYU66, pYU70 and pYU82, were found to be responsive to 1-h heat shock treatment in LB (Fig. 2). Of these, β-galactosidase activities from pYU27, pYU66, pYU70 and pYU82 at 42°C were about 1.6- to 3-fold higher than those at 30°C, whereas the activities from the remaining two at 42°C were less than 50% of those at 30°C. The responsive 6 clones were then examined using the *rpoH* mutant as a host strain (Fig. 3A). All clones in the *rpoH* background exhibited an opposite tendency to those in the wild type, whereas no significant difference between the wild type and *rpoH* background was observed in pYU68 (*yafH*) as a control.

We also carried out the time course experiment (Fig. 3,B and C). In the wild type background, β-galactosidase activity from pYU27 dramatically increased within 10 min after the temperature shift and the activity in pYU66 became maximal at 60 min and those in pYU70 and pYU82 continued to increase at least for 120 min. The activities from the remaining two decreased within 30 min at 42°C compared to those at 30°C. In the *rpoH* background, the activities from pYU14 and pYU27 gradually increased and decreased, respectively after the temperature shift compared to those at 30°C. These results suggest that the expression of these 6 genes is directly or indirectly dependent on σ^{32}. The cells harboring any of the other pYUs showed slight or no change.

Nucleotide sequences of the genes revealed that the genes in pYU14 and pYU70 might encode previously uncharacterized proteins, and the open reading frame of the last one has not been listed in databases. The gene in pYU70 seems to be a new heat shock gene, *yigL* encoding a small protein consisting of 25 amino acid residues because its genomic location

Table 1. Expression of the *Escherichia coli* promoter-proximal genes under various treatments and information of their products.

Clone number	β-Gal activity[a]	30 bp sequences from the LacZ fusion[b]	Fused gene	Fusion site[c]	Function/Gene Product	Accession number
pYU 1	27,000	ATTGCTGACTTCAATATGATTGAAGAAGGCGATC	*ydaO*	40	Unknown	EG13357
pYU 3	400	TCAACGCTGAAAAATCTCTGGTGGCGGCTCGATC	*ssnA*	95	Related to cell death	EG13062
PYU 6	1,000	CAGCAGGTGATTATGCCGCCTCAGCCGATTGATC	*yaeY*	161	Unknown	EG13511
pYU 7	10,000	ACGGCTGTACTCGCAGCATTCTTTCCGCAGGATC	*sohB*	135	Suppressor of *hyr*	EG10956
pYU 12	5,700	TACGATGCCAGCCAGCTCGAGCCGCAGCCGGATC	*mpl*	63	UDP-*N*-acetylmuramate:L-alanyl-γ-D-glutamyl-meso-diaminopimelate ligase	EG12440
PYU 13	410	GGTCAGCGAGTGCCGCCGATTGGCATTACCGATC	*ytfJ*	36	Unknown	EG12510
PYU 14	700	GAAATCAACTGGCTCGCTGGTCATTCGCCGGATC	*ydjR*	188	Unknown	EG13994
PYU 15	190	GCGGTGATTGGTGCCGGGATTATTCGCTATGATC	*yigM*	131	Unknown	EG11471
pYU 16	180	CATGTCTGGCAAACCATGCAGGCATGTATCGATC	*sdaA*	221	L-Serine deaminase	EG10930
PYU 18	540	GTTGTGGCGTGTGATAAAAACGGCAACATCGATC	*gcvP*	631	Glycine dehydrogenase	EG11810
PYU 19	980	ATGCAGATGTTGCAGGAAACCTTGTGGCCGGATC	*mrp*	225	Putative ATPase	EG10611
pYU 20	10,000	GCTAAAGATATGACCTGCCAGGAATTTATTGATC	*hdeB*	44	Unknown	EG11399
pYU 24	3,100	GTCACAACCGGTGGCGCGGGTGGGCAGATTGATC	*ygdL*	160	Unknown	EG13084
PYU 25	42,000	AAAAGCGAGGATGCTATGTCTACTCAACTTGATC	*ydiH*	33	Unknown	EG13967
pYU 27	10,000	GCCGTTGAGCTGGCAAAAGACTGGCGTACCGATC	*hslV*	86	Heat shock regulon	EG11676
pYU 32	33,000	GAAGACTATACCCCGTATACCTTCACCAAAGATC	*yeiP*	98	Unknown	EG12035
pYU 33	17,000	CGCGACATTGAAGAAGATAAGGCACCCGCCGATC	*yfgA*	55	Unknown	EG10015
pYU 34	3,600	ACAGCGGTACTCTGTTTTGTGATTGAAGCCGATC	*cheA*	219	Chemotactic response	EG10146
pYU 35	800	ACCGTAGGGTTTATTCGCCACCTGCCGCACGATC	*hflX*	262	Putative GTPase	EG10437
pYU 36	1,200	GGCGTAGACGGCATCTTCGTCGGCCCCAGCGATC	*yhaF*	179	Unknown	EG10016
pYU 37	880	GGCGACAACGTGGCGATGGTGATTAACGGCGATC	*yfeY*	100	Unknown	EG14166
pYU 38	6,300	ATCGCACCTGAACTTTACTTCCGCGAAGGCGATC	*ysgA*	84	Unknown	EG14321
pYU 39	450	GAAGCGCTGGAGATGATGGGCTCAGGCCGCGATC	*yagQ*	95	Unknown	EG13556
pYU 40	13,000	CAAGTCACCACGACTGTGATGGCGCGTTACGATC	*tus*	5	Ter DNA binding protein	EG11038
pYU 43	3,200	AAGCTATTTCATTCAGATGCTGTGCCGATTGATC	*trkA*	120	Major K^+ transporter	EG11019
pYU 46	9,400	GAGCAAACCGCCGATGGCCCGCGCAAGCGGGATC	*rnpB*	165*	RNase P	EG30069
pYU 47	11,000	ATGGATGGTCCTGCCAACCTGCGTCATTTTGATC	*yafV*	26	Unknown	EG13334
pYU 48	4,700	ACCGGATGCGGTGTTCACGCCTTTACTTACGATC	*fbp*	181	Fructose-bisphosphatase	EG10283
pYU 49	2,000	ATTCTTTCCGCCCCGGCATCGCAACACGAAGATC	*yfiQ*	109	Unknown	EG14224
pYU 51	82,000	AAATTCCTCCATCCGGAGATCGTAACCGTAGATC	*treR*	281	Repressor	EG12202
pYU 53	65	ACAATTACCCTTACCTTCCAGTCCGGGACCGATC	*acrF*	99	Lipoprotein	EG10267
pYU 54	130	GATAACGGTATGGTGGTCTTGCTGGTTTCTGATC	*ptrA*	60	Protease III	EG10786
pYU 55	240	GTGGCGCGTTACGGCAACTTCGTCACTTACGATC	*ccmH*	94	Required for synthesis of c-type cytochrome	EG12052
pYU 56	400	GATGGTCGCCCGTTACATGTGATTTCTGGCGATC	*sufI*	258	Suppressor of *ftsI*	EG11376
pYU 57	240	GACTTAAAACGTCACGACACCATTGGTATTGATC	*purM*	86	Phosphoribosylaminoimidazole synthetase	EG10798
pYU 58	540	ACCAACTACCTCATTTACACCCAAACCAGCGATC	*pldA*	67	Detergent-resistant Phospholipase A activity	EG10738
pYU 59	340	CGCCGCAACCGGTACATGGACTGAAGAAGAGATC	*agaS*	20	Putative tagatose-6-phosphate ketose/aldose isomerase	EG12767
pYU 60	5,300	GAGTTTGAAGAAATTTGCGCCCGCGGTCTGGATC	*carB*	197	Carbamoylphosphate synthase, heavy subunit	EG10135

pYU	61	240	GCGCTGCAACAACTCACCTGGACACCGGTCGATC	*malM*	55	Periplasmic protein	EG10559
pYU	63	170	ACCTCCTGGAGTCTGGTTGAGGCGTCAATGGATC	*ydcV*	169	Unknown	EG13765
pYU	64	720	GGTCTGGCGTTCAACGTTGTCGCGAAAGGCGATC	*ydbA*	1742	Unknown	EG11307
pYU	65	9,100	GCAGGTCTGGGTCAGGAAATGGGCGAGCCGGATC	*ycjK*	97	Unknown	EG13908
pYU	66	360	CTGGTCGCGGTGGATTACCCTGACCGGTATGATC	*truA*	257	Pseudouridine synthase I	EG10454
pYU	67	130	GAATTTACGGTAGCCGCCGTCAGCCACATCGATC	*yjfK*	62	Unknown	EG11459
pYU	68	140	AGAAGCCTGCCGGATGGCGAATGATTTCCAGATC	*yafH*	144	Unknown	EG13145
pYU	69	280	CCGGACAACGTGCGGTTGATTCGTGATTACGATC	*glnL*	225	Protein kinase/phosphatase	EG10387
pYU	70	80	TCGAGCCCGCAAGTGAAGAGATATTACCTTGATC	*yjgL*	7	Possible heat shock protein	This study
pYU	72	260	AACCTGCCGTTAACAGCGGCGGATTTATATGATC	*alsK*	97	Allose kinase	EG11956
pYU	74	11,000	CTGGTGATTGTGGCGGCAGATATCGAAGTGGATC	*fruA*	171	Fructosephosphotransferase II	EG10336
pYU	76	520	GAGGAATGGGTGCATCTTCGGATTGAGGAAGATC	*ydaU*	102	Unknown	EG13363
pYU	78	1,400	GCTATCGGCAATTCCATTCTTGATGCTTACGATC	*purH*	305	Phosphoribosylaminoimidazole-carboxamide formyltransferase	EG10795
pYU	79	80	GACATTGACATTGCCGATGCGCTGCCGGTGGATC	*yddS*	256	Unknown	EG13790
pYU	80	95	TTCGGAATGCAGCAATATCGTCATCGGTCAGATC	*amr*	17	Antisense of *mrdA*	This study
pYU	81	140	GTTCCGACCAACCGTCCAATGATTCGTAAAGATC	*secA*	422	Translocation ATPase for protein export	EG10936
pYU	82	110	GAAGTTATCGCCGGTAAGAAACACTACTTCGATC	*lpd*	343	Lipoamide dehydrogenase	EG10543
pYU	83	1,200	GCCGGTATTCAGGCGGCAAAACGCACCTGGGATC	*moaC*	71	MPT synthesis	EG11666
pYU	84	35,000	AGCGATATTACCAAAGCGGATTTCCTCTCAGATC	*katE*	155	Catalase hydroperoxidase III	EG10509
pYU	85	300	AAAACGTTGGCGCAGATTGGTATCACGACCGATC	*fliD*	349	Fook-associated protein 2	EG10841
pYU	86	400	AACTTGCTGGCGGCGCTGGGCTGCCAGTTTGATC	*ego*	133	Essential for aerobic growth	EG13806
pYU	87	350	CGCGTGAAGATTGCCGAAGTGGATGGTAATGATC	*thrA*	766	Aspartokinase I-homoserine dehydrogenase I	EG10998
pYU	88	6,400	GCGCGCTTCAAGCGCCCGGTCGTGCCTGGCGATC	*fabZ*	109	3R-hydroxymyristoyl acyl carrier protein dehydrase	EG11284
pYU	89	330	CTGGCGGGGCATATCGGGCATACGCTTGCCGATC	*yhcI*	160	Unknown	EG12815
pYU	90	2,400	AGTGATGAGTCATCGTGGCATCTCGTTGAAGATC	*tfaD*	21	λ tail fiber gene homolog	EG13640
pYU	91	200	ACCAATAAGCAGGTTCAGCAGCCAGTGATGGATC	*ftsZ*	337	Initiation of septation	EG10347
pYU	92	70,000	AAAGCGCGTCGAAAAACACGCGAGGAGCTGGATC	*yihI*	25	Unknown	EG11835
pYU	93	110	CCGTCTGTTAAATCATGGCCCAACGGTCTTGATC	*yddH*	24	Unknown	EG13779
pYU	95	8,000	ATTGTCGGGCCGGGAGCCAAAGGCACCATTGATC	*glpX*	134	Glycerol metabolism	EG11517
pYU	96	62,000	CCGGACCCGGACGTCCCGGAAGCTCGCGTCGATC	*speA*	71	Biosynthetic decarboxylase	EG10959
pYU	97	4,200	TGCGTGCTGGTATATCTGCGTCACGCGGAAGATC	*yggV*	130	Unknown	EG12982
pYU	98	95	GTGCTGTTTCTCTCATTCTGTCGGATTAACGATC	*yhjS*	438	Unknown	EG12263
pYU	99	490	AGCTTGGGTTGCAATGCTGCGGTGCAACCCGATC	*ybgP*	25	Unknown	EG13312
pYU	100	100	AGCGTGAAAGACCGCGATGCCTGGGCGAAAGATC	*yaiW*	59	Unknown	EG13608
pYU	101	220	GCTGACCTTGAGCAGAGCGAAGCCATCGGCGATC	*hyaE*	85	Processing of HyaA and HyaB	EG10472
pYU	103	1,300	GTGAAAATCCGCATTGAGTGCCACATGCCGGATC	*rus*	65	Resolvase	EG20283
pYU	104	3,000	TTCCGAACATCTGCAAATTGGCATTCGGGCGATC	*amn*	460	AMP nucleosidase	EG10039
pYU	105	500	CCTTATTGTGGCAATTGGCCTTGTTGGTCTGATC	*secG*	20	Protein export	EG12095
pYU	106	2,000	GACATCTCGGAACTGGTTCGTCAGGGGCAAGATC	*cafA*	102	Cell division and growth	EG11299

[a] Cells harboring each clone were grown at 37°C for 16 h and β-galactosidase activity was then measured as described in MATERIALS AND METHODS. Number represents the total β-galactosidase activity (Miller units). The values are averages of at least three independent experiments.

[b] Thirty base pair DNA sequences from the LacZ fusion of each of 77 independent clones.

[c] Number represents the position of amino acid residue fused with LacZ.

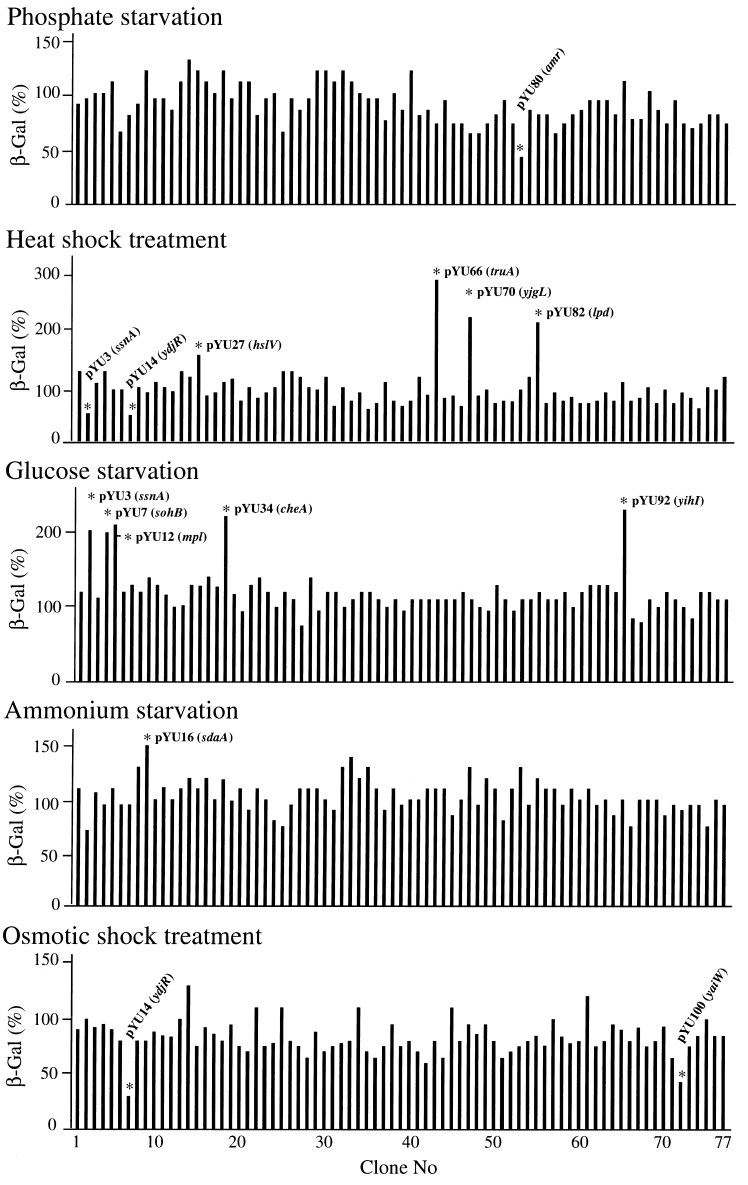

Fig. 2. Expressional regulation of the cloned genomic genes under different environmental conditions.

Cell treatment and β-galactosidase assay were as described in MATERIALS AND METHODS. β-Galactosidase activity is expressed as a percentage of that without the treatment. The values are averages of at least three independent experiments.

(97.8 min and is located between two genes, *yihR* and *yihS*), promoter position (Fig. 4) and induction profile are different from those of heat shock genes reported previously (Chuang et al., 1993a, b). It may be possible that in addition to σ^{32}, some unknown factor is involved in the *yigL* expression. The genes in pYU27 and pYU82 are *hslV* and *lpd* that had been characterized as heat shock genes (Chuang et al., 1993b; Richmond et al., 1999). The latter gene and *truA* (*hisT*) in pYU66 encode lipoamide dehydrogenase and pseudouridine synthase I, respectively.

We found a possible σ^{32}-dependent promoter (−35 sequence, CGTAAAT and −10 sequence, ACCATCCA) and σ^{70}-dependent promoter (−35 sequence, CTAACG and −10 sequence, GATAAC) for *ydjR* in pYU14 and a possible σ^{32}-dependent promoter (−35 sequence, GGTTTAT and −10 sequence, CCCTTGTG) and σ^{70}-dependent promoter (−35 sequence, TTTTCA and −10 sequence, TATATT) for *yigL* in pYU70. Promoters for *ssnA*, *hslV*, *lpd* and *truA* have been reported (Fig. 4). The putative σ^{32}- and σ^{70}-promoter sequences for *yigL* are well matching with

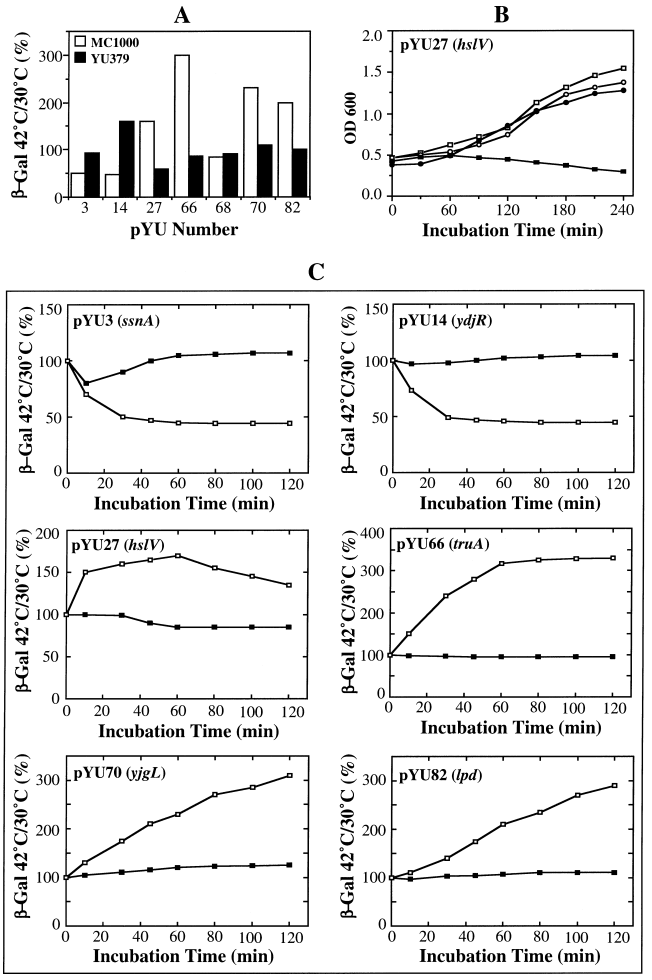

Fig. 3. Effect of *rpoH* on the clones responsive to heat shock.

(A) β-Galactosidase activity of MC1000 (open columns) or the isogenic *rpoH* (closed columns) cells harboring each plasmid clone, which were grown aerobically at 42°C for 1 h. The activity is expressed as a percentage of that of control at 30°C. Plasmid pYU68 was used as a control clone. (B) Representative growth curves of cells harboring pYU27. MC1000 and *rpoH* cells harboring the clone grown at 42°C are represented by open and closed squares, respectively, and at 30°C by open and closed circles, respectively. The growth curves of other clones were nearly the same as those from pYU27. (C) Time course of β-galactosidase expression of MC1000 (open squares) or *rpoH* (closed squares) after shifting the temperature from 30°C to 42°C. Other details are shown in MATERIALS AND METHODS. Incubation time (min) and rate (percentage) of β-galactosidase activity (Miller units) at 42°C to that at 30°C are shown by horizontal and vertical axes, respectively.

the cannonical sequences, being 70 and 75%, respectively. Whereas, both σ^{32}- and σ^{70}-promoter sequences for *ssnA* and *ydjR* are relatively weakly matching with the cannonical sequences (see Fig. 4 for detail). One possible reason could be that the reduction of β-galactosidase activity in pYU3 and pYU14 might be due to binding competition between σ^{32}- and σ^{70}-RNA polymerase complexes to the promoter sequences, which are overlapping (Fig. 4).

Glucose starvation

Five clones pYU3, pYU7, pYU12, pYU34 and pYU92, were selected as clones responsive to glucose starvation (Fig. 2). Some of such genes are known to be regulated by cAMP-CRP; therefore, β-galactosidase activity from these clones was measured in the *Δcya* background (Table 2). The activities from pYU12 and pYU92 in the presence of cAMP were about 1.7- and 1.9-fold higher, respectively, than those in the absence of cAMP under the conditions with or without addition of glucose. The activities from pYU3 and pYU34 were decreased in the presence of cAMP to be less than 50% under both conditions with or without glucose. The activity from pYU7 was not changed by the addition of cAMP. Therefore, the 5 genes appear to be classified as follows: (a) cAMP-dependent positive control in the cases of the genes in pYU12 and pYU92, which possess a possible cAMP-CRP binding sequence, TCAGGTGAATCGCGCCAGCAAA and GATTGTAAGACCCGTTAAGGGT, respectively, similar to the consensus sequence upstream from a possible or the reported σ^{70} promoter (Fig. 4), (b) cAMP-dependent negative control for the genes in pYU3 and pYU34, which possess a possible cAMP-CRP binding sequence, TGCGGCAACTGCGCTCAGTTCT and CGAGGTGAGATGCAACTCAACA, respectively, overlapping with the σ^{70}-promoter sequence (Fig. 4), and (c) a cAMP-independent and glucose-dependent negative control for the gene in pYU7. However, the last gene has a possible cAMP-CRP binding sequence, AACTGTGAGCCAAAGCGTTGTT overlapping with a possible -10 region of σ^{70}-promoter sequence, TAAACT (Fig. 4). Our results suggest that the previously characterized *sohB* and *cheA* genes, encoding periplasmic protease and chemotactic signaling protein, respectively, are regulated by glucose and cAMP-CRP.

Phosphate starvation

Out of 77 clones, only pYU80 showed a highly significant reduction in β-galactosidase activity under phosphate starvation (Fig. 2). Notably, the gene *amr* in pYU80 was found to be located on antisense strand at

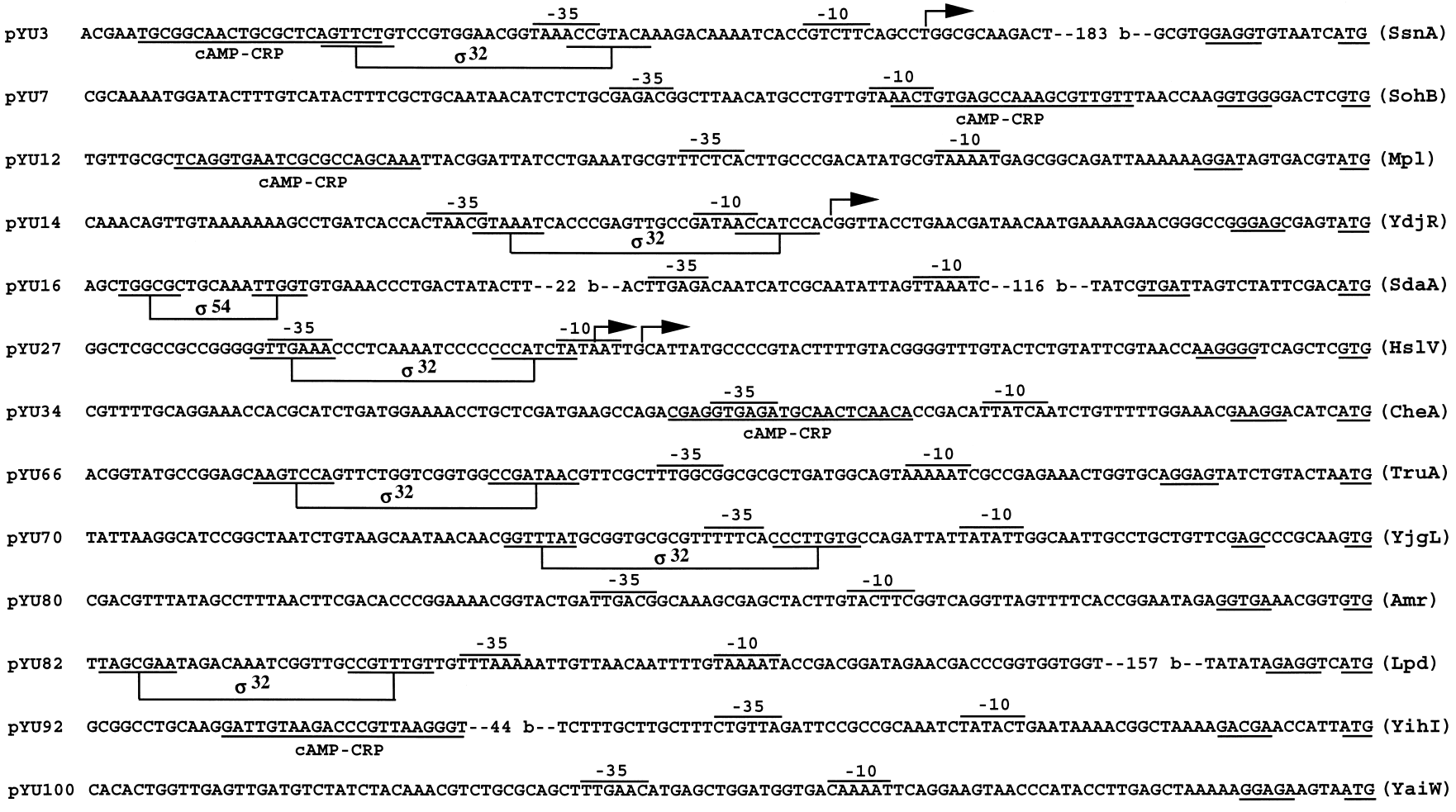

Fig. 4. Nucleotide sequences at the 5′-flanking regions of 13 responsive genes to various treatments tested and their possible promoter sequences.

In each gene, a sequence homologous to the σ^{70}-promoter consensus sequences (−35 sequence, TTGACA and −10 sequence, TATAAT) and a putative initiation codon with a possible ribosome-binding sequence at an appropriate distance, which are followed by the open reading frame in frame with the *lacZ* gene, were predicted. The initiation codons of *ssnA* (Yamada et al., 1999) and *ves* (Yamada et al., 2002) have been determined by amino acid sequencing of their products, and the initiation codons of pYU70 and pYU80 were predicted in databases. Sequences homologous to the σ^{32}-promoter consensus sequences (−35 sequence, CCTTGAA and −10 sequence, CCCAT-TA), σ^{54}-promoter consensus sequences (−24 sequence, TGGCAC and −12 sequence, TTGCA) and 22 base pair cAMP-CRP binding consensus sequences (−62 sequence, AA-TGTGA and −40 sequence, TCACA-TT) were explored for each gene according to the responsiveness (underlines indicate conserved sequences). Over lines and underlines represent possible σ^{70}-dependent promoter sequences and cAMP-CRP binding sequences, respectively (Gosset et al., 2004; Miller, 1992). Another two underlines at the right side show the possible ribosome-binding sequences and initiation codons. Possible σ^{32}- and σ^{54}-promoter sequences (Black and Maier, 1995; Miller, 1992) are represented by underlines with a bracket. The nucleotide sequences of the unknown genes analyzed in this study have been submitted to the DDBJ database.

Table 2. Effect of cAMP or glucose on gene expression.

Clone no. (gene)	β-Galactosidase activity (Miller units)[a]					
	wild type (−cAMP)		*Δcya* (−cAMP)		*Δcya* (+cAMP)	
	+G	−G	+G	−G	+G	−G
pYU 12 (*mpl*)	6,000	12,500	3,300	6,500	5,700	11,000
pYU 92 (*yihI*)	65,000	150,000	27,000	55,000	50,000	100,000
pYU 3 (*ssnA*)	310	600	900	1,600	450	800
pYU 34 (*cheA*)	3,000	5,500	6,000	10,000	2,700	4,500
pYU 7 (*sohB*)	11,600	22,000	11,000	20,000	10,500	19,000

[a] MC1000 (wild type) and its isogenic *Δcya* mutant were used as a host strain. Cells harboring each clone were grown in MOPS minimal medium in the presence or absence of 1 mM cAMP (+cAMP or −cAMP) or glucose (+G or −G) at 37°C. β-Galactosidase activity was then measured as described in MATERIALS AND METHODS. Reported values are averages of at least three independent experiments.

the 3′ side of the *pbpA* gene encoding penicillin-binding protein 2 (Asoh et al., 1986). The gene may thus be regulated negatively by the starvation. Nucleotide sequence revealed that the gene possesses a possible σ^{70}-dependent promoter, but no similar sequence to PhoB-binding consensus sequence (Kim et al., 2000) was seen around the promoter (Fig. 4).

Ammonium starvation

The ammonium starvation significantly influenced the expression of the gene *sdaA* encoding L-serine deaminase, in pYU16 to increase in β-galactosidase activity. It has strong possible σ^{54}- (−24 sequence, TGGCGC and −12 sequence, TTGGT) and σ^{70}-(−35 sequence, TTGAGA and −10 sequence, TTAAAT) dependent promoter sequences as shown in Fig. 4. Thus, the gene may be transcribed by RNA polymerase associated with the σ^{54} subunit under nitrogen-limited conditions.

Osmotic shock

We checked osmotic shock conditions including the treatment time from 1 to 16 h and the medium osmolarity from 200 to 530 mM of NaCl in different media. In all cases, the average activity of β-galactosidase from 30 randomly selected clones was reduced to be 65 to 82% of those without the osmotic shock (data not shown). Thus, the mild osmolarity containing 0.2 M NaCl in LB medium, which showed 82% in the average activity, was used for our experiments (Fig. 2). Since the activity of β-lactamase encoded by the same clone plasmid was about 85% in average activity under the same osmotic condition (data not shown). Recently, DNA macroarray technology has been used to monitor early transcriptional alterations in mRNA levels of 3,290 genes of *E. coli* in response to an osmotic upshift imposed by the addition of 0.4 M NaCl (Weber and June, 2002). Under this condition, more than 70% of genes were repressed. Therefore, the reduction of β-galactosidase activity from most clones may not be due to instability of LacZ-fusion protein in cells.

Two highly significant clones, pYU14 and pYU100, showed β-galactosidase activity less than 50% of those without osmotic shock (Fig. 2). It is not obvious whether OmpR is involved in the regulation because its consensus sequence is not available (Huang et al., 1994). Gutierrez et al. (1987) identified 12 nonessential genes subject to osmotic shock, whose expression except for *ompC* and *ompF* was shown to be independent of OmpR.

Discussion

From 77 genes corresponding to about 10% of the promoter-proximal genes in *E. coli*, we found two new genes which have not been identified by computer analysis. Both genes have relatively small ORFs like those which may be out of the ORF category in the genome analysis (Blattner et al., 1997). So it is presumed that several dozen genes small in size remain to be identified. Reading frames and protein productivity of 31 genes, which were only predicted by computer analysis, were confirmed by this study. We also suggested that most genes analyzed here possess their own promoters, on the basis of the size of cloned

DNA fragments and gene organization around each analyzed gene in database. Of course, genes with very low β-galactosidase activity in Table 1 might not have their own promoter because of the readthrough transcription from a promoter on the vector. Occurrence of the promoters of 71 genes was suggested by this study and those of the remaining genes had been characterized or predicted from nucleotide sequences. For the former genes, further detailed experiments including determination of the mRNA start site or promoter sequence is required.

We found that 13 genes in pYU3, pYU7, pYU12, pYU14, pYU16, pYU27, pYU34, pYU66, pYU70, pYU80, pYU82, pYU92 and pYU100 are regulated under various environmental stresses including heat shock, osmotic shock and starvation of glucose, phosphate or ammonium. Further study revealed that the gene, *ssnA*, in pYU3 was found to be negatively regulated by *rpoS* (Talukder et al., 1996) and to be involved in cell death at the stationary phase (Yamada et al., 1999), and that the gene, *ves*, in pYU14 was shown to be expressed at a relatively low temperature (Yamada et al., 2002). Thus, it is expected that these data and procedures will allow us to uncover interesting genes and regulation. Analysis with a multi-copy plasmid vector may be not suitable for a specific factor-dependent regulation like LacI on the *lac* operon because the regulatory factor molecules become limited in the presence of the cognate operator on the plasmid. Whereas, in the cases of common regulators like cAMP-CRP for many different genes, a multi-copy plasmid may be applicable because a lot of such regulator molecules may be enough to cover the cognate operator on the plasmid as well as many cognate operators on the genome. This approach with a multi-copy plasmid also allows us to analyze relatively weak promoters. About 30% of the clones in our library including 5 of the responsive genes have weak promoters of less than 300 units of β-galactosidase activity, which could be close to the background level if a single-copy plasmid was used.

It could be estimated that the library is equivalent to about 10% of the *E. coli* promoter-proximal genes on the basis of the expected gene numbers and gene organization as an operon. The library consists of random clones because the genes were found to be scattered on the circular genome (data not shown). Expressional analysis under the various conditions revealed that 16% (13 clones) seem to be subject to expressional control under the conditions tested. The rate of clones responsive to heat shock (7%), glucose starvation (6%), osmotic shock (2.6%), or stationary phase (6% [Talukder et al., 1996]) was relatively higher than that of genes responsive to phosphate starvation (1.2%) or ammonium starvation (1.2%). The rates of genes responsive to various conditions appeared to be nearly the same as those indicated previously by mRNA analysis (Chuang et al., 1993a; Weber and June, 2002).

Out of the 13 responsive genes, 6 were found to be sensitive to more than one treatment (Fig. 2). The results from such treatments and nucleotide sequencing suggest that 7 genes may be dependent on more than one σ-factor (Fig. 4), in addition to σ^{70} (Talukder et al., 1996). Consistently, Blattner et al. (1997) suggested that about 20% of *E. coli* genes have more than one promoter. Interestingly, 4 genes may be transcriptionally regulated by cAMP-CRP. It was reported that out of 30 genes responsive to glucose starvation, 11, 6, and 5 also responded to heat shock, oxidative stress and osmotic shock, respectively (Jenkins et al., 1991), and that most of the 18 genes induced osmotically were also controlled by σ^S (Hengge-Aronis et al., 1993). Therefore, the genes under more than one σ-dependent promoter or responsive to multiple stresses would generally exist.

Acknowledgments

We thank A. Kato, O. Adachi and K. Matsushita for their helpful discussion, and Y. Kohara, T. Yura and J. Beckwith for phage clones or bacterial strains.

References

Asoh, S., Matsuzawa, H., Ishino, F., Strominger, J. L., Matsuhashi, M., and Ohta, T. (1986) Nucleotide sequence of the *pbpA* gene and characteristics of the deduced amino acid sequence of penicillin-binding protein 2 of *Escherichia coli* K12. *Eur. J. Biochem.*, **160**, 231–238.

Black, L. K. and Maier, R. J. (1995) IHF- and RpoN-dependent regulation of hydrogenase expression in *Bradyrhizobium japonicum*. *Mol. Microbiol.,* **16**, 405–413.

Blattner, F. R., Plunkett III G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.

Casadaban, M. J. and Cohen, S. N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia*

coli. *J. Mol. Biol.*, **138**, 179–207.

Casadaban, M. J., Chou, J., and Cohen, S. N. (1980) In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. *J. Bacteriol.*, **143**, 971–980.

Chuang, S.E., Daniels, D. L., and Blattner, F. R. (1993a) Global regulation of gene expression in *Escherichia coli*. *J. Bacteriol.*, **175**, 2026–2036.

Chuang, S. E., Burland, V., Plunkett III, G., Daniels, D. L., and Blattner, F. R. (1993b) Sequence analysis of four new heat-shock genes constituting the *hslTS/ibpAB* and *hslVU* operons in *Escherichia coli*. *Gene*, **134**, 1–6.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W. A., and Saier, M. H., Jr. (2004) Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J. Bacteriol.*, **186**, 3516–3524.

Gutierrez, C., Barondess, J., Manoil, C., and Beckwith, J. (1987) The use of transposon Tn*phoA* to detect genes for cell envelope proteins subject to a common regulatory stimulus: Analysis of osmotically regulated genes in *Escherichia coli*. *J. Mol. Biol.*, **195**, 289–297.

Hengge-Aronis, R., Lang, R., Henneberg, N., and Fischer, D. (1993) Osmotic regulation of *rpoS*-dependent genes in *Escherichia coli*. *J. Bacteriol.*, **175**, 259–265.

Huang, K.-J., Schieberl, J. L., and Igo, M. M. (1994) A distant upstream site involved in the negative regulation of the *Escherichia coli ompF* gene. *J. Bacteriol.*, **176**, 1309–1315.

Jenkins, D. E., Auger, E. A., and Matin, M. (1991) Role of *rpoH*, a heat shock regulator protein, in *Escherichia coli* carbon starvation protein synthesis and survival. *J. Bacteriol.*, **173**, 1992–1996.

Katayama, T. and Nagata, T. (1990) Inhibition of cell growth and stable DNA replication by overexpression of the *bla* gene of plasmid pBR322 in *Escherichia coli*. *Mol. Gen. Genet.*, **223**, 353–360.

Kim, S. K., Kimura, S., Shinagawa, H., Nakata, A., Lee, K. S., Wanner, B. L., and Makino, K. (2000) Dual transcription regulation of the *Escherichia coli* phosphate-starvation-inducible *psiE* gene of the phosphate regulon by PhoB and the cyclic AMP (cAMP)-cAMP receptor protein complex. *J. Bacteriol.*, **182**, 5596–5599.

Kohara, Y., Akiyama, K., and Isono, K. (1987) The physical map of the whole *E. coli* chromosome: Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, **50**, 495–508.

Messing, J. (1983) New M13 vectors for cloning. *Methods Enzymol.*, **101**, 20–79.

Miller, J. H. (1992) A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

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

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanatinia, H., Kofoid, E. C., Morrison, T. B., and Parkinson, J. S. (1995) The smaller of two overlapping *cheA* gene products is not essential for chemotaxis in *Escherichia coli*. *J. Bacteriol.*, **177**, 2713–2720.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Talukder, A. A., Yanai, S., and Yamada, M. (1994) Analysis of products of the *Escherichia coli* genomic genes and regulation of their expression: An applicable procedure for genomic analysis of other microorganisms. *Biosci. Biotechnol. Biochem.*, **58**, 117–120.

Talukder, A. A., Yanai, S., Nitta, T., Kato, A., and Yamada, M. (1996) RpoS-dependent regulation of genes expressed at late stationary phase in *Escherichia coli*. *FEBS Lett.*, **386**, 177–180.

Weber, A. and June, K. (2002) Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA macroarrays. *J. Bacteriol.*, **184**, 5502–5507.

Yamada, M., Talukder, A. A., and Nitta, T. (1999) Characterization of the *ssnA* gene, which is involved in the decline of cell viability at the beginning of stationary phase in *Escherichia coli*. *J. Bacteriol.*, **181**, 1838–1846.

Yamada, M., Nagamitsu, H., Izu, H., Nakamura, K., and Talukder, A. A. (2002) Characterization of the *ves* gene, which is expressed at a low temperature in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.*, **4**, 163–169.

Yura, T., Tobe, T., Ito, K., and Osawa, T. (1984) Heat shock regulatory gene (*htpR*) of *Escherichia coli* is required for growth at high temperature but is dispensable at low temperature. *Proc. Natl. Acad. Sci. USA*, **81**, 6803–6807.