Full Length Research Paper

Structural organization of co-regulated genes in *Escherichia coli*

Ali Azam Talukder[1-3, 5]*, Saidul Alam[2], Md Ershaduzzaman[4], Sam Khairul Bashar[3] and Mamoru Yamada[5]

[1]Department of Microbiology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh.
[2]Department of Botany, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh.
[3]Department of Life Sciences, North South University, Plot 15, Block B, Bashundhara R/A, Dhaka 1229, Bangladesh.
[4] Bangladesh Livestock Research Institute, Savar, Dhaka 1341, Bangladesh.
[5]Department of Biological Chemistry, Yamaguchi University, Yamaguchi 753-8515, Japan.

A promoter-protein fusion library was constructed and analyzed previously which covered about 8% promoter-proximal genes from *Escherichia coli* in respect to their reading frames, protein productivity as well as expressional regulations under normal and various environmental stress conditions (Talukder et al., 1994, Bioscience Biochemistry and Biotechnology, 58: 117-120). In this study, we further analyzed 13 significantly responsive genes. β-galactosidase expression levels were varied among the examined clones, indicating that the library having different ranges of promoters from stronger to weaker. DNA database analysis revealed that out of 13 genes studied here, 11 and 2 genes were found to known and unknown or new genes, respectively. Moreover, new findings in case of three genes having interesting structural organizations were recorded this time. The present study supports all previous data to conclude that about 5 - 10% E. coli genes are co-regulated under the control of complex regulatory circuits.

Key words: Gene expression, RNA-polymerase, gene organization.

INTRODUCTION

Escherichia coli possesses about 4,300 genes and is expected to have more than several hundred genes proximal to their own promoters in operon organization (Sambrook et al., 1989; Glover, 1992; Blattner et al., 1997). Most of the genes in the organism have been predicted by computer-aided analysis after the determination of complete genomic information like DNA sequences. These predicted genes should thus be proven by identification of gene products in addition to the elucidation of their functions and expressional regulation under various conditions.

A very useful phage library of *E. coli* was constructed previously (Kohara et al., 1987). In that library, each clone was physically mapped and its genomic location was also determined. Sequencing and analysis of the whole *E. coli* genome have already been accomplished

based on such order phage libraries (Blattner et al., 1997; Feist et al., 2007; Mizoguchi et al., 2007). In addition, DNA microarray technique is being used for transcriptional analysis of some genomes (Weber and June, 2002). These strategies, however, do not seem to be followed by functional and biochemical analysis of specific genes of interest. To clarify the entire genomic organization and metabolic regulation of *E. coli*, useful approaches to analysis of gene product, gene location and regulation of its expression, as well as gene structure organization are to be established.

Another mini-library of the *E. coli* genomic genes was also constructed, in which each clone was with part of a genomic gene fused in frame with the lacZ gene in addition to its promoter and operator. This library represented about 8% promoters of *E. coli* genome (Talukder et al., 1994, 1996, 2005a). Analysis of this type of library enables to identify reading frames, genomic locations, structural organization, confirmation of their protein productivity and to examine their expressional regulation under various environmental conditions. DNA sequence

---
*Corresponding author. E -mail: tunubj@yahoo.com. Tel: +8802-966-2444. Fax: +8802-779-1052.

analysis, open reading frame identification of protein product and expressional regulation under various environmental stress conditions were accomplished previously (Talukder et al., 1994, 1996, 2005a). Those earlier reports showed that out of 77 independent genes from the library, 13 were expressed significantly under various environmental conditions and 6 genes were co-transcribed by more than one σ-factors of RNA polymerrase holoenzyme (Talukder et al., 2005a). Among them, 5 genes were found to be negatively regulated by RpoS, encoded σs which is responsible for stationary phase specific gene expression and two of them, named ssnA and ves, were extensively analyzed (Talukder et al., 1996; Yamada et al., 1999, 2002). The gene ssnA was shown to be involved in cell death at the stationary phase and ves was expressed at a low temperature at the beginning of the stationary phase (Yamada et al., 1999, 2002). Furthermore, 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 σ32 subunit of RNA polymerase and adenylate cyclase, respectively (Yura et al., 1984; Talukder et al., 2006). In addition to gene expressional regulation, open reading frames (ORFs) identification, promoter position and gene structural organizations of the cloned genes of this library are yet to be performed. Therefore, in this study, an attempt was made to analyze the co-regulation on gene expression by measuring β-galactosidase level as well as structural organization of the 13 responsive genes based on expressional pattern and nucleotide sequence information that was used for homology searching, gene mapping, ORFs positioning and locating the promoters' position as well (Talukder et al., 2005a).

MATERIALS AND METHODS

Bacterial strains and plasmid

The *E. coli* K-12 strains used in this study were W3110, wild type; MC1000, araD139 Δ(ara-leu)7697 ΔlacX74 galU galK rpsL (Casadaban and Cohen, 1980); JM103, Δ(lac-pro)thi strA supE endA sbcB hsdR4 F' traD36 proAB+ lacIqZΔM15 (Messing, 1983). Plasmid pMC1396 was used as a plasmid vector for construction of *E. coli* promoter library as described below.

Construction of *E. coli* promoter library

The promoter library [pYU series of 14 plasmids, such as, pYU3, pYU7, pYU12, pYU14, pYU16, pYU27, pYU34, pYU66, pYU70, pYU80, pYU82, pYU87 (internal control), pYU92 and pYU100] was kindly provided by Professor Mamoru Yamada, Yamaguchi University, Japan, which was constructed earlier by Talukder et al. (1994). In brief, *E. coli* W3110 genomic DNA was partially digested with Sau3AI and approximately 0.5 - 1.0 kb DNA fragments were isolated and inserted by ligation into the BamHI site in front of the lacZ gene on pMC1396 vector, lacking its promoter and the first 8 amino acid residues (Figure 1) (Casadaban et al., 1984). The lacZ gene is expressed to produce the protein product only by supplying gene of interest and the gene corresponding to the N-terminus of its
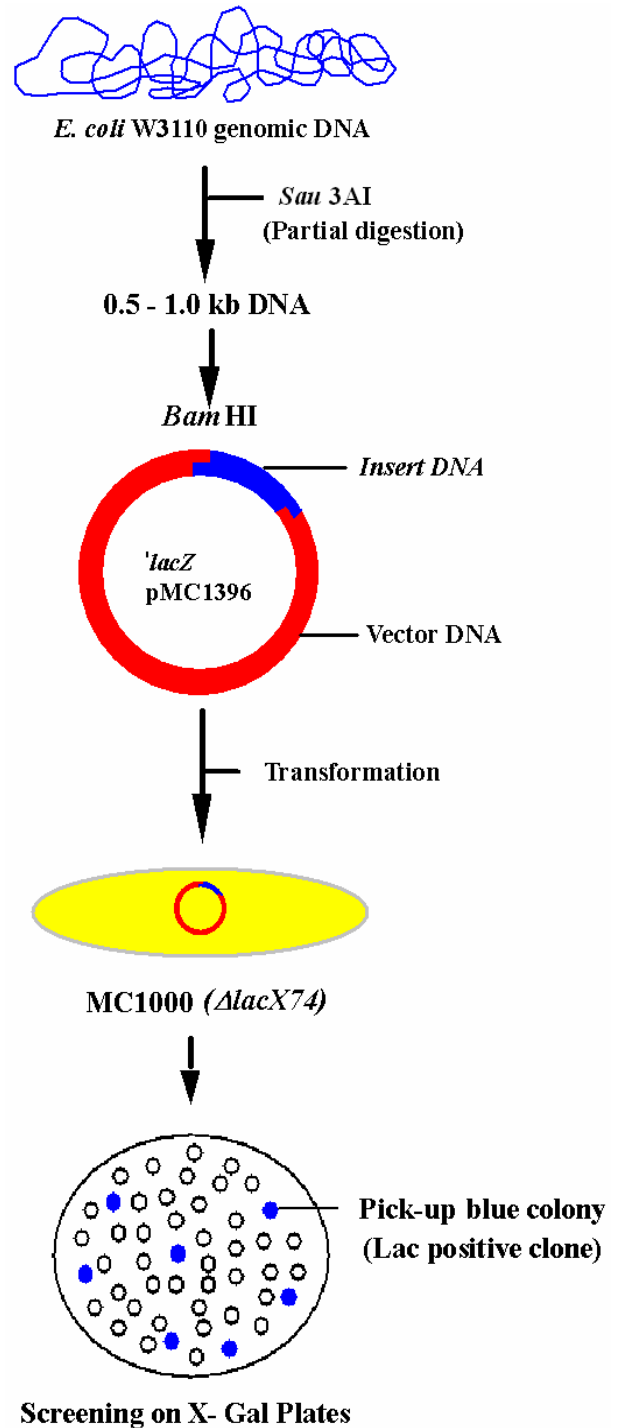

Figure 1. Construction strategy of *E. coli* promoter-protein fusion library. *E. coli* W3110 genomic DNA was partially digested with Sau3AI and approximately 0.5 - 1.0 kb DNA fragments were isolated and inserted by ligation into the BamHI site in front of the lacZ gene on pMC1396 plasmid vector (Talukder et al., 1994).

product. When a fragment encompassing from the promoter to part a genomic DNA fragment consisting of the promoter of a specific of coding region of a promoter-proximal gene was inserted and the

gene was situated in frame to the lacZ gene, an active β-galactosidase was produced which in turn makes colony blue on agar plates containing X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase). Thus, such blue color formation indicated that a gene encoding a protein was cloned and fused with lacZ gene. After nucleotide sequencing, its exact reading frame was determined by comparison to that of the lacZ gene.

Out of 4,500 colonies grown on the various agar plates, 106 were obtained as Lac+ (positive clone). Among the 106, 29 colonies were found to be duplicate or false positive were confirmed by DNA sequence information. Finally, 77 independent pYU plasmid series clones encoding a protein were selected for further analysis (Talukder et al., 1994, 2005a). Among them, 13 significantly responsive clones were used here for further analysis.

Bacterial culture development

LB-agar plates containing 100 mg/l of Ampicillin were streaked with MC1000 cells harboring each clone plasmid and incubated at 37°C overnight. Then an individual colony from each clone was transferred to 3 ml of fresh LB broth and incubation was carried out at 37°C with 100-rotation per minute (rpm) in water bath shaker. Incubation was further continued until the growth reached to an early log phase with an OD at 600 nm of 0.25 - 0.30 (that is, 0.25 - 0.30 × 108 cells/ml).

The fresh culture was then diluted 1,000-fold into 3 ml fresh LB medium and was grown overnight at 37°C in a water bath shaker with 100 rpm. Cells were collected in a tube and growth was stopped with chilling treatment keeping the culture in an ice bucket for a period of at least 30 min.

β-galactosidase assay

Cell culture was diluted whenever required to adjust the cell concentration to an OD at 600 nm (A600) of 0.28 - 0.80 for the preparation of sample for enzyme assay. β-galactosidase activity was measured following the method described by Miller (1992).

1 ml of reaction sample for enzyme assay was prepared by mixing 0.1 ml culture sample with 0.9 ml Z-buffer for each clone. Then, one drop of 0.1% Sodium Dodisyl Sulphate (SDS) was added to the reaction sample and mixed with vortexing, which facilitates the cell lysis and was allowed for 5 min pre-incubation at 37°C. Then, the reaction was allowed to start by adding 200 μl ONPG (Oortho-nitrophenyl-β-D-galactopyranoside) as substrate of β-galactosidase. Reaction was continued until the development of characteristic yellow color of O-nitrophenol. Finally, the reaction was stopped by adding 500 μl of 1 M Na2CO3.

$$\beta\text{-galactosidase activity (Miller Unit)} = \frac{OD_{420} \times (1.75 \times OD_{550}) \times 1000}{T \times V \times OD_{600}}$$

Where,
OD_{420} is the absorbance of the yellow o-nitrophenol.
OD_{550} is the scatter from cell debris, which, when multiplied by 1.75 approximates the scatter observed at 420 nm.
OD_{600} reflects cell density in the reaction mixture.
T = Reaction time in minutes.
V = volume of culture used in the assay in ml.

The Units give the change in A420/min/ml of cells/OD600.

Gene structural organization of the *E. coli* genomic genes

Nucleotide sequence around the junction with the lacZ gene in each library clone was determined previously by the dideoxy-chain determination method (Sanger et al., 1977; Talukder et al., 1994, 2005a). Putative gene structural organizations of 13 clones were determined on the basis of nucleotide sequences of the cloned genomic genes and also their deduced amino acid sequences were compared with those listed in the DDBJ, EMBL, GeneBank, SWISS-PROT and NBRF-PIR databases on computer network [http://www.genome.naist.jp/bacteria/array.ecol.html or Genobase (*E. coli* database) or http://www.genome.wisc.edu (*E. coli* K12, University of Wisconsin, USA, Refseq: NC000913, Gene bank U00096)].

Open reading frames (ORF) of each of 13 clones were determined by comparison of that of the lacZ gene. Therefore, the proper ORF was determined easily, because nucleotide sequences around the junction between an inserted gene (fragment of W3110) and the lacZ gene were available. These nucleotide sequence information were used to determine the amino acid sequencing. The position of amino acid residues fused with lacZ is represented by numbers. For example pYU3 encoded SsnA protein and this protein have ORF containing 442 amino acid residues and it's 95 amino acid residues fused with the LacZ protein (Yamada et al., 1999). So, fused product is either SsnA-LacZ (protein) or ssnA-lacZ (gene).

Possible promoter sequences of respective genes analyzed here were predicted by using available computer aided program, SELEX. Promoter positions as well as promoter sequences of significantly responsive cloned genomic genes were further confirmed by comparison of consensus sequences for σ70 (-35 region having consensus sequence TTGACA and -10 region having consensus sequence TATAAT) and for σ32 (consensus sequences, -35 region CCTTGAA and -10 region CCCAT-TA, underlines indicated conserved sequence) subunits of RNA polymerase holoenzyme in *E. coli*. Other minor factors specific-promoter sequences were predicted as described previously (Talukder et al., 2005a). Sequences homologous to the 22 base pair cAMP-CRP binding consensus sequences (-62 region AA-TGTGA and -40 region TCACA-TT) were explored for the gene in pYU34, according to the responsiveness of glucose starvation (Talukder et al., 2005a).

Total sizes of each of 13 ORFs were determined on the basis of the total amino acid residues. Each ORF was calculated from the start codon to the stop codon. Structures, orientations and organizations of ORFs and their upstream promoter positions were designed according to the information explained above. A typical example of gene structural organization for the gene in pYU3 (Figure 6).

RESULTS

Expressional regulation of the cloned genomic genes

Plasmid-harboring each of 13 clones was grown to LB medium. Cultures were then diluted 1000-fold into the fresh LB medium and the cultures were continued to grown overnight at 37°C. Sample collection, preparation and measurement of enzyme activity were described in materials and methods. Table 1 shows the β-galactosidase activity of 14 clones. Enzyme production levels were varied from 100- to 90,000 Millar Units. The clone in pYU87 was used as an internal control, because this clone gave constant level of enzyme even it was exposed by different environmental stresses (Talukder et al., 2005a). All 13 clones examined in this study were found to be four distinct groups.

The clone pYU92 showed a very high level of β-galactosidase activity of 90,000 Miller Units (Figure 2, red

Table 1. Expressional regulation of the *E. coli* promoter-proximal genes and information of their products.

Clone number	β-Gal level [a]	Fused gene	Function/Gene product [b]	Accession number [c]
pYU3	500	ssnA	Related to cell death	EG13062
pYU7	13000	sohB	Suppressor of hyr	EG10956
pYU12	7000	mpl	UDP-N-acetylmuramate: L- alanyl -gamma- D-glutamyl-meso-diaminopimelate ligase	EG12440
pYU14	900	ydjR	Cold shock regulon	EG13994
pYU16	230	sdaA	L-serine deaminase	EG10930
pYU27	13000	hslV	Heat shock regulon	EG11676
pYU34	5000	cheA	Chemotactic response	EG10146
pYU66	460	truA	Pseudouridine synthase	EG10454
pYU70	100	yjgL	Possible heat shock protein	d
pYU80	150	amr	Antisense of mrdA	d
pYU82	210	lpd	Lipoamide dehydrogenase	EG10543
pYU87	450	thrA	Aspartokinase I-homoserine dehydrogenase I	EG10998
pYU92	90000	yihL	Unknown	EG11835
pYU100	150	yaiW	Unknown	EG13608

[a]Number represents the total β-galactosidase activity in Miller Units. The values are averages of three independent experiments; [b]Functions/gene product name were identified from the nucleotide sequence information by homology searching.; [c]Accession number was collected from the *E. coli* gene bank; [d]Data were collected from Talukder et al. (2005a).

bar). A quite higher expression of β-galactosidase was found in the clones pYU7 (13000), pYU12 (7000), pYU27 (13000) and pYU34 (5000) ranging from 5,000 Units to 13,000 Units (blue bars). Pink bars represent an inter-mediate level of β-galactosidase activity which was found from the three clones, pYU3 (500), pYU14 (900) and pYU66 (460). The rest five clones, pYU16 (230), pYU70 (100), pYU80 (150), pYU82 (210) and pYU100 (150) showed very low level of β-galactosidase activity. This wide differences β-galactosidase level by various clones from the library is a clear indication that the library covering quite different types of promoters from the very stronger to the weaker. This fact, however, ultimately relates the various binding affinity levels of these pro-moters to the RNA polymerase. It is also noted that the RNA polymerase binding affinity depends on the mat-ching percentage of the promoter consensus sequences of the specific σ factor to the promoter recognition sequences. Because, all 13 genes analyzed here have different promoter recognition sequences (Talukder et al., 2005a and Figure 4).

In addition to β-galactosidase activity, DNA sequence information was used to identify the open reading frame as well as function or gene products of 13 cloned geno-mic genes are also presented in Table 1. All 13 cloned genes in pYU3, pYU7, pYU12, pYU14, pYU16, pYU27, pYU34, pYU66, pYU70, pYU80, pYU82, pYU92 and pYU100 are fused with ssnA (releted to cell death), sohB (suppressor of hyr), mpl (UDP-N -acetylmuramate: L - alanyl -gamma -D- glutamyl-meso-diaminopimelate ligase), ydjR (cold shock regulon), sdaA (L-serine deaminase), hslV (heat shock regulon), cheA (chemo-tactic response), truA (pseudouridine synthase), yjgL (possible heat shock protein), amr (antisense of mrdA), lpd (lipoamide dehydrogenase), yihL (unknown) and yaiW (unknown) genes, respectively. The product of the gene in pYU87 which was used as an internal control in this study, identified as thrA (aspartokinase I-homo-serine dehydrogenase I). Out of 13 genes analyzed here, func-tions or products of 11 genes in pYU3, pYU7, pYU12, pYU14, pYU16, pYU27, pYU34, pYU66, pYU70, pYU80 and pYU82 and two genes in pYU92 and pYU100 were known and unknown or new genes, respectively. The functions of rest two genes in pYU92 (yihL) and pYU100 (yaiW) are yet to be determined. The two unknown gene products were not reported previously (Figure 3).
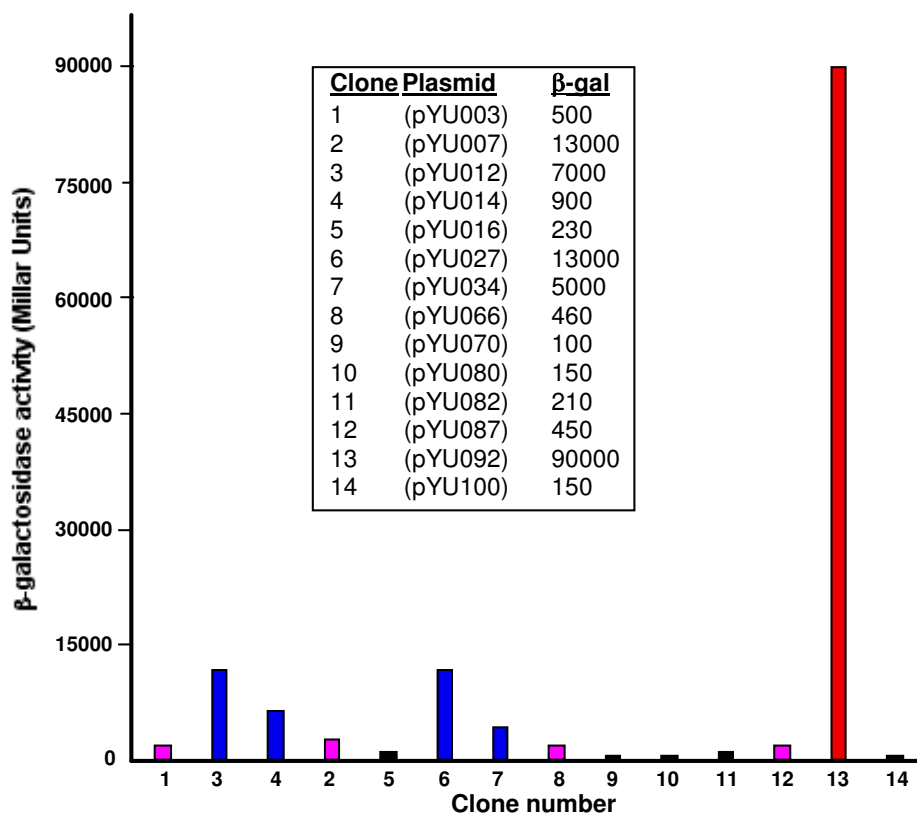
Figure 2. β-galactosidase activity of MC1000 cells harboring each of 14 plastid clones, which were grown aerobically at 37°C for over night under normal condition. Plastid pYU87 was used as an internal control. Cell cultures and enzyme activities were measured as described in materials and methods.

Database searching and structural organization of the *E. coli* cloned genomic genes

Putative structural organization of the cloned genes and the adjacent genes was deduced based on β-galactosidase expressional patterns, nucleotide sequence information which was used for database searching, prediction of open reading frames, as well as the promoter consensus sequences, directions and positions (Figure 3).

Open reading frames of all 13 genes were predicted previously and also in this study by computer aided analysis except pYU80, because the gene in pYU80 encoded antisense strand of mrdA (Blatner et al., 1997, Talukder et al., 2005a). All genes analyzed here were fused with the exact frame of lacZ. The number mentioned bellow represented the position of amino acid residues fused with lacZ in pYU3 (ssnA, 95 amino acids), pYU7 (sohB, 135 amino acids), pYU12 (mpl, 63 amino acids), pYU 14 (ydjR, 188 amino acids), pYU16, (sdaA, 221 amino acids), pYU27 (hslV, 86 amino acids), pYU34 (cheA, 219 amino acids), pYU66 (truA, 257 amino acids) pYU70 (yjgL, 7 amino acids), pYU80 (amr, 17 amino

acids), pYU82 (lpd, 343 amino acids), pYU87 (thrA, 766 amino acids), pYU92 (yihI, 25 amino acids) and pYU100 (yaiW, 59 amino acids). These data were used for structural analysis, especially for estimate the size, position and organization of ORFs (Figure 3).

The Selex aided analysis as well as conservedness of specific promoter consensus sequences study revealed that all genes have their own promoters (Figure 3). A σ70 dependent promoter sequences (-35 region TTGACA and -10 region TATAAT) were found in all responsive genes analyzed here, with different matching frequencies (Talukder et al., 2005a). In addition to σ70 dependent promoter, a σ32 dependent promoter sequences (-35 region GGTTTAT and -10 region CCCTTGTG) were also found for yigL in pYU70 (Figure 4). The putative σ70 and σ32 promoter sequences for pYU70 are well matching with the consensus sequences, being 75 and 70% respectively (Figure 4). Whereas, σ70 promoter sequence for ydjR in pYU14 are relatively weaker in matching with the consensus sequence (Talukder et al., 2005a). One possible reason was the reduction of β-galactosidase activity in pYU70 (100 Units) which might be due to the binding competition between σ70 and σ32
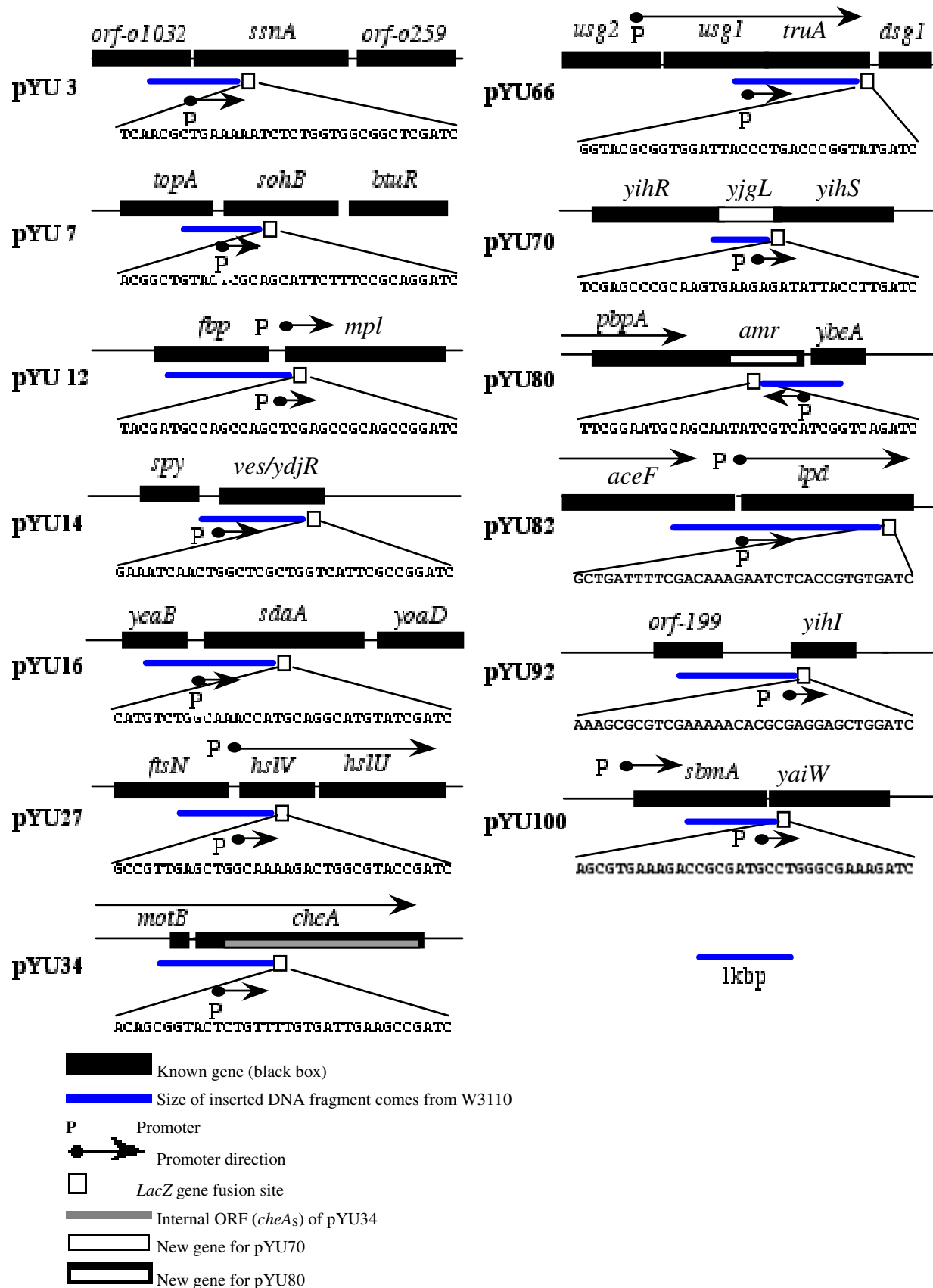
Figure 3. Putative structural organizations of the *E. coli* cloned genomic genes.

pYU7 (*sohB*)

                                                                                              −35
CGCAAAATGGATACTTTGTCATACTTTCGCTGCAATAACATCTCTGC<u>GAGACG</u>GCTTAACAT
        −10
GCCTGTTG<u>TAAACT</u>GTGAGCCAAAGCGTTGTTTAACCAAGGTGGGGACTCGTG

pYU34 (*cheA*)

                                                                                          −35
CGTTTTGCAGGAAACCACGCATCGATGGAAAACCTGCTCGATGAAGCCAGA<u>CGAGGTGAGAT</u>
   cAMP-CRP   −10
<u>GCAACTCAACA</u>CCGACA<u>TTATCA</u>ATCTGTTTTTGGAAACGAAGGACATCATG

pYU70 (*yjgL*)

                                                            −35   −35
TATTAAGGCATCCGGCTAATCTGTAAGCAATAACAACGGTTTATGCGGTGCGCGT<u>TTTTCA</u>C
   −10   -10
CCTTCGCCAGATTAT<u>TATATT</u>GGCAATTGCCTGCTGTTCGAGCCCGCAAGTG

pYU80 (*amr*)

                                                          −35
CGACGTTTATAGCCTTTAACTTCGACACCCGGAAAACGGTACTGA<u>TTGACG</u>GCAAAGCGAGC
        −10
TACTTG<u>TACTTC</u>GGTCAGGTTAGTTTTCACCGGAATAGAGGTGAAACGGTGTG

Figure 4. One hundred and fifteen base-pair nucleotide sequences at the 5'-flanking regions of 4 responsive genes and their possible promoter sequences. Consensus Sequences for σ70 = − 35 region (TTGACA) – 10 region (TATAAT); σ32 = − 35 region (CCTTGAA) – 10 region (CCATTA); cAMP-CRP = − 62 region (AATGTGA) – 40 region (TCACATT); Ribosome Binding Site (RBS) = GGAGG; Translational Start Codon = ATG or GTG; (Underlines indicate conserved sequences).

RNA polymerase complexes to the promoter sequences (Figure 4). On the other hand, the gene in pYU34 possessed a possible cAMP-CRP binding sequence, CGAGGTGAGATGCAACTCAACA which was overlapped with the σ70 promoter sequence (−35 region GTGAGA). Because, it was reported previously that the gene in pYU34 was controlled negatively by cAMP recaptor protein CRP (Talukder et al., 2005a).

Among the 13 genes studied here, only five genes in the clones pYU12, pYU27, pYU34, pYU66 and pYU82 had their own promoters (Figure 3). The remaining eight genes' promoters might have been elucidated in this study based on their expressional patterns as well as the nucleotide sequence information (Figures 3). Among the eight genes, six known genes in pYU3 (ssnA), pYU7 (sohB), pYU14 (ves/ydjR), pYU16 (sdaA), pYU92 (yihl) and pYU100 (yaiW) were confirmed having only their open reading frames, but no information about their promoters was available. Both ORF/antisense strand and promoter positions for remaining two genes in pYU70

(yigL) and pYU80 (amr) could not be predicted in early reports. This study has unearthed the location of promoter sequences in all eight genes along with the ORF sequences in case of two unknown genes (Figure 4). Out of 13 genes analyzed, three were found to have interesting structural organizations. Among them, one in pYU34 was known gene and rest two were unknown genes and their detailed structural organizations are shown in Figure 5. It was reported earlier that an intragenic promoter was there in the cheA gene, which might had been associated with a short product named CheAS which is shown in Figure 5 (Sanatinia et al., 1995). But, this claim was not confirmed by any means. This present analysis has revealed the existence of a new promoter located at the upstream region of CheAS. DNA fragment size in pYU34 was not enough to cover the upstream promoter to transcribe such a product (CheAS). In addition, β-galactosidase expression also is in consistent with the presence of this new promoter discovered in this study. The two unknown genes in pYU70 and pYU80
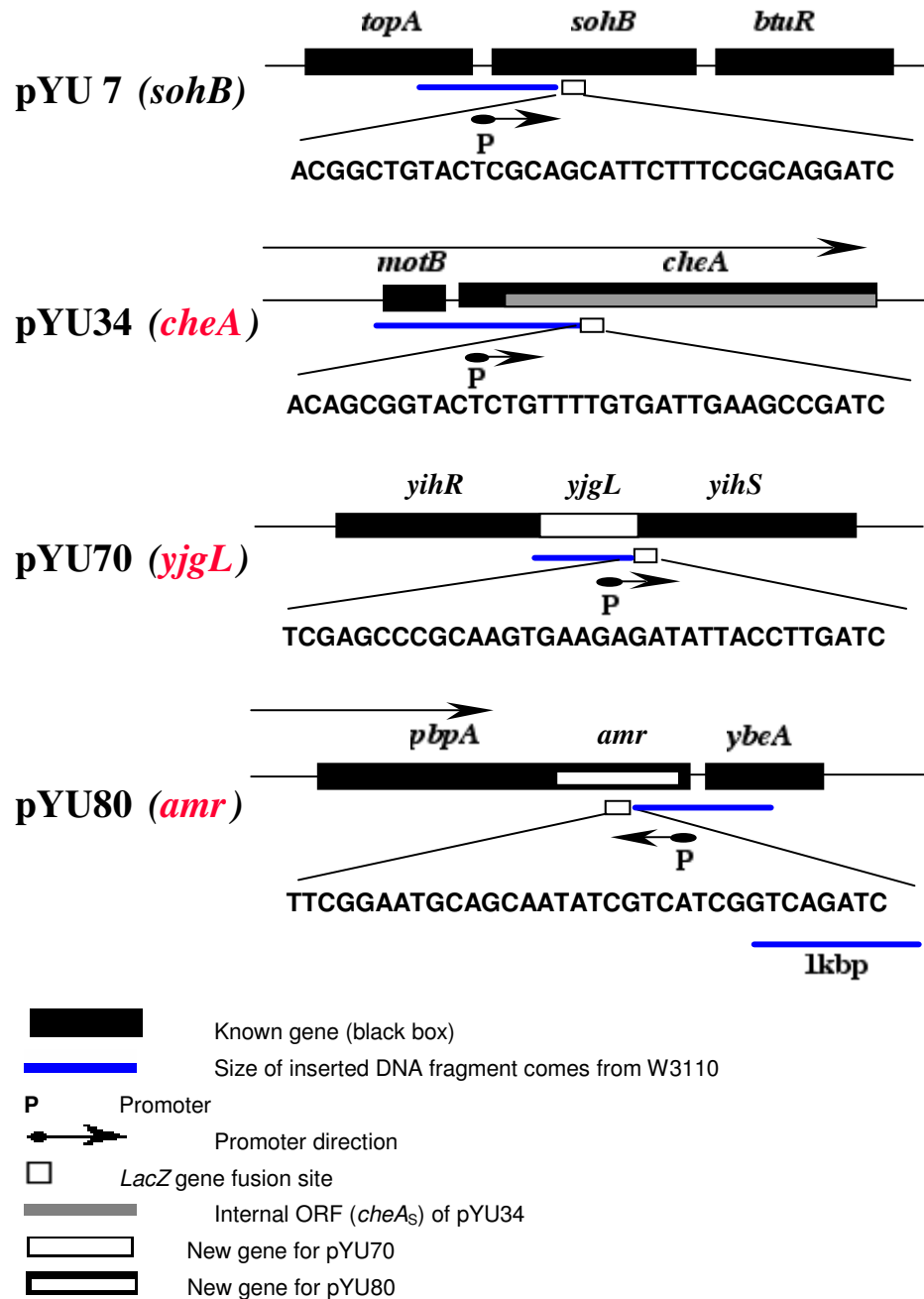
Figure 5. Structural organization of the three interesting cloned genomic genes of *E. coli* mini-library.

were located beside the known genes of yihR and pbpA, respectively (Figure 5). The both genes might encode previously uncharacterized proteins, which should be characterized. Because, the open reading frame of the gene in pYU70 was not listed in database. On the other hand, the insert of pYU80 was found to be located on the antisense strand at the 3' region of the gene amr.

The nucleotide sequences of the genomic gene-lacZ

fusions and analysis with the database, as well as the results of expressional regulation of the cloned genomic genes were used to deduce the putative structural organizations. The inserted genomic DNA fragment sizes (blue lines) fused with the lacZ gene (small open boxes) was determined previously by restriction enzymemapping (Talukder et. al., 2005a). Open and closed boxes indicate the unknown and known genes, respectively. The promo-
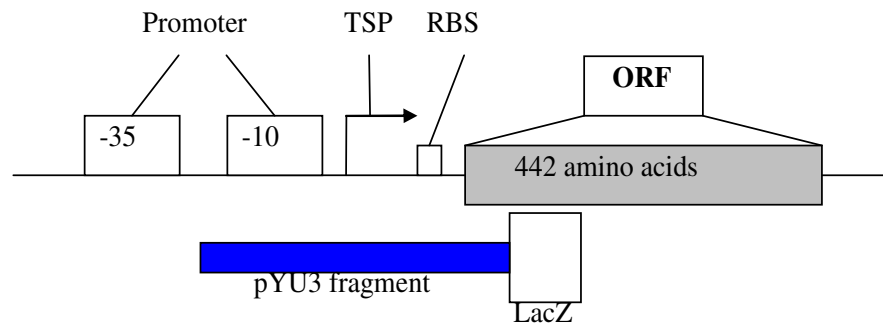
Figure 6. Gene structural organization for the gene in pYU3; TSP = Transcription start point, RBS = Ribosome binding site; ORF = Open reading frame.

ter (P) positions of the genes and their transcription directions are shown by closed circles and arrows, respectively. The promoters over and under the genes were identified previously and predicted in this study, respectively. Thirty base pair DNA sequences from the LacZ fusion of each of 13 independent clones are shown on the bottom from the LacZ fusion site (small open boxes). The overlapping closed and shadowed boxes in pYU34 represent the possible overlapping genes, which seem to have different translational start points. Over-lapping closed and open boxes in pYU80 represent the possible overlapping genes, which may be transcribed in the opposite direction (antisense). The 5'-portion of the usg2 and aceF genes in pYU66 and pYU82 are not shown.

In each gene, a sequence homologous to the σ70-promoter consensus sequences (-35 sequence, TTGACA and -10 sequence, TATAAT) and a putative initiation codon (blue color) with a possible ribosome-binding sequence (red color) at an appropriate distance, which are followed by the open reading frame in frame with the lacZ gene, were predicted from the available computer aided databases. The gene in pYU7 is used as an internal control. Among the four genes analyzed here, the initiation codons of pYU70 and pYU80 were predicted in databases. In the remaining two genes of pYU7 and pYU34, their initiation codons for translation were reported previously. Sequences homologous to the σ32-promoter consensus sequences (-35 region CCTTGAA and -10 region CCCAT-TA indicated by violet color) and 22 base pair cAMP-CRP binding consensus sequences (-62 region AA-TGTGA and -40 region TCACA-TT indicated by green color) were explored for the genes in pYU70 and pYU34, respectively (Yura et al., 1984). Underlines indicate conserved sequences. The black color with underlines represents possible σ70-dependent promoter sequences (Miller, 1992 and Gosset et al., 2004).

The gene in pYU7 was used as an internal control is this figure. The promoters over and under the genes were identified previously and predicted in this study, respectively. Thirty base pair DNA sequences from the LacZ fusion of each of 4 independent clones were shown on the bottom from the LacZ fusion site (small open boxes). The overlapping closed and shadowed boxes in pYU34 represent the possible overlapping genes cheAL and cheAS, respectively, which seem to have different translational start sites. Overlapping closed and open boxes in pYU80 represent the possible overlapping genes amr and antisense strand of mrdA, respectively, which may be transcribed in the opposite direction (antisense).

DISCUSSION

Global expressional regulation of *E. coli* genomic genes is the most widely studied topic in molecular genetics (Chuang et al., 1993a; Blattner et al., 1997, Richmond et al., 1999, Weber and June, 2002; Talukder et al., 2005a, Park et al., 2007). Yet, systematic biochemical and functional analyses along with DNA sequence analysis are to be integrated for further elucidation of the co-regulated genes under complex regulatory system in *E. coli* and other organisms as well. In this present study, an attempt was made to check the expressional regulation in terms of β-galactosidase production and to find out promoter position as well as structural organization of 13 cloned genomic genes from *E. coli*. It was found that 13 genes in pYU3, pYU7, pYU12, pYU14, pYU16, pYU27, pYU34, pYU66, pYU70, pYU80, pYU82, pYU92 and pYU100 were expressed well under normal laboratory condition. Another gene in pYU87 was used as an internal control. The β-galactosidase activities varied from 100 to 90000 Miller Units (Figure 2). This fact indicates that the library covers various types of promoters from weaker to stronger. Moreover, these genes were also found to respond significantly under various environmental stresses including anaerobiosis, heat shock, osmotic shock and starvation of glucose, phosphate or ammo-nium (Talukder et al., 1994, 1996, 2005a, Yamada et al., 1999, 2002). It is to note here that the promoters regulated by either negative or positive factors, some-times refer to as transcription factors, a group of DNA-binding proteins which activate or repress transcription, may be included

in this library. Because, in addition to RNA polymerase holoenzyme, several regu-lators are often involved to bind the promoter regions of the respon-sive genes to be either induced or repressed promoter activity (Ishihama, 1997, 1999; Maeda et al., 2000a, 2000b). Previous reports showed that the gene ssnA in the clone pYU3 was found to be negatively regulated by rpoS, a gene encoded stationary phase specific sigma factor, σ38 and to be involved in cell death at stationary phase (Talukder et al., 1996; Yamada et al., 1999) and the clone in pYU27 encoded heat shock specific gene hslV controlled positively by heat shock specific σ factor σ32 (Chuang et al., 1993b; Talukder et al., 2005b). Therefore, it is concluded that all 13 genes analyzed here possess at least one promoter in their upstream regulatory regions (Figures 3 and 4).

It is important to note that the analysis with a multi-copy plasmid vector pMC1396 may not be suitable for a specific factor-dependent regulation like LacI on the lacoperon, because the regulatory factor molecules become limited in the presence of the cognate operator on the plasmid vector. On the other hand, in the cases of common regulators like cAMP-CRP for many different genes (about 300 genes in *E. coli* under the control of cAMP receptor protein, CRP), a multi-copy plasmid may be suitable 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 DNA (Gosset et al., 2004). Moreover, this approach with a multi-copy plasmid also allows us to analyze relatively weak promoters like pYU70 or pYU80 (Table 1). Further detailed study is necessary to elucidate the functional analysis of these two weak promoters. About 30% clones in the 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 (Talukder et al., 1994, 1996, 2005a; Gutierrez et al., 1987).

The complete *E. coli* genome was sequenced 12 years before (Blattner et al., 1997). It is expected that all genes should have been known at least based on the nucleotide sequence information. However, out of 13 genes analyzed in this study, two new genes in pYU70 and pYU80 were not be identified by computer analysis. Open Reading Frame and antisense strand of these newly identified genes in pYU70 and pYU80, respectively, were found to be too small that they might have been considered as out of ORF category in computer aided genome analysis earlier (Blattner et al., 1997). This misleading identification or failure in locating the small ORF sequences of the gene in pYU70 (yjgL) was caused by its position in between the two known genes of yihR and yihS (Figure 5). Therefore, an assumption can be made that several dozen of inter- or intra- genic genes with small sizes might have resided in the *E. coli* genome. This assumption should be confirmed as early as possible by further detailed study. Another gene in pYU80 is an antisense of mrdA, which can make special interest to usefulness of such library for genome analysis for other organisms.

Open reading frames and protein productivity of 7 genes, which were only predicted by computer analysis, have been confirmed by this study. This present result suggests that most of the genes analyzed here possess their own promoters (Figure 3). At the same time, it can be assumed that there is a probability for the genes with relatively a low β-galactosidase activity might not have their own promoter because of the read-through transcription from a promoter on the vector used for construction of the library (Table 1). It is further supported by the fact that *E. coli* has its characteristic mechanism for transcription of more than 4000 genes by 1000 promoters. Therefore it is interesting to know that how single promoter transcribes more than four genes in *E. coli*. Occurrence of own promoter in 7 genes out of 13, has been suggested by this study. Other 6 genes were characterized or predicted for their own promoter previously (Talukder et al., 2005a). The mRNA start point and promoter sequences have to be confirmed by further detailed study for these two new genes in the clones pYU70 and pYU80.

Previous studies showed that about 10 - 20% of *E. coli* genes having more than one promoter, which were also responded to various environmental stress conditions like anaerobiosis, heat shock, cold shock, pH shock, osmotic shock or different starved conditions. (Gutierirrez et al., 1987, Katayama and Nagata, 1990, Jenkins et al., 1991, Hengge-Aronis et al., 1993, Black and Maier, 1995, Blattner et al., 1997, Talukder et al., 2005a, 2005b). Sometimes, promoter positions as well as sequences are overlapping each other. These sequences are also the targeted places by binding of other transcription factors rather than RNA polymerase holoenzyme. Therefore, complex transcriptional regulatory networks for co-regulated genes would generally exist by binding compe-tition among the various groups of transcription factors to the similar or overlapping promoter sequences. In addition to promoter position, the structural organization of the *E. coli* genomic genes in this present study reveals that 4 - 8% genes are with interesting structural organi-zations in aspects of intragenic overlapping sequences (pYU34), intergenic new ORF (pYU70) and mRNA coding antisense strand (pYU80). The present study along with previous reports concluded that about 5 - 10% coregulated genes in *E. coli* under the control of complex regulatory networks.

REFERENCES

Black LK, Maier RJ (1995). IHF- and RpoN-dependent regulation of hydrogenase expression in Bradyrhizobium japonicum. Mol. Microbiol. 16: 405-413.
Blattner FR, Plunkett IIIG, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997). The complete genome sequence of *Escherichia coli* K-12. Science

277: 1453-1474.

Casadaban MJ, Cohen SN (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138: 179-207.

Casadaban MJ, Chou J, Cohen SN (1984). *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 SE, Daniels DL, Blattner FR (1993a). Global regulation of gene expression in *Escherichia coli*. J. Bacteriol. 175: 2026-2036.

Chuang SE, Burland V, Plunkett IIIG, Daniels DL, Blattner FR (1993b). Sequence analysis of four new heat-shock genes constituting the hslTS/ibpAB and hslVU operons in *Escherichia coli*. Gene 134: 1-6.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Broadbelt LJ, Hatzimanikatis V, Palsson BO (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3: 121.

Glover DM (1992). In "Gene cloning: The mechanisms of DNA manipulation," In Brammer WJ, Edidin M, Chapman and Hall (eds), New York.

Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MHJr (2004). Transcription Analysis of Crp-Dependent Catabolite Control of Gene Expression in *Escherichia coli*. J. Bacteriol. 186: 3516-3524.

Gutierrez C, Barondess J, Manoil C, Beckwith J (1987). The use of transposon TnphoA 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, Fischer D (1993). Osmotic regulation of rpoS-dependent genes in *Escherichia coli*. J. Bacteriol. 175: 259-265.

Ishihama A (1997). Promoter selectivity control of *Escherichia coli* RNA polymerase. In Eckstein F, Lilley D (eds). Nucleic Acids and Molecular Biology, Vol. 11: pp. 53-70. Mechanism of Transcription. Springer-Verlag, Heidelberg, Germany.

Ishihama A (1999). Modulation of the nucleoid, the transcription apparatus and the translation machinery in bacteria for stationary phase survival. Genes Cells 3: 135-143.

Jenkins DE, Auger EA, 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, 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.

Kohara Y, Akiyama K, 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.

Maeda H, Fujita N, Ishihama A (2000a). Competition among seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res. 28: 3497-3503.

Maeda H, Jishage M, Nomura T, Fujita N, Ishihama A (2000b). Two extracytoplasmic function sigma subunits, σE and σFecI, of *Escherichia coli*: promoter selectivity and intracellular levels. J. Bacteriol. 182: 1181-1184.

Messing J (1983). New M13 vectors for cloning. Methods Enzymol. 101: 20-79.

Miller JH (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.

Mizoguchi H, Mori H, Fujio T (2007). *Escherichia coli* minimum genome factory. Biotechnol. Appl. Biochem. 46: 157-167.

Park YS, Seo SW, Hwang S, Chu HS, Ahn JH, Kim TW, Kim DM, Jung GY (2007). Design of 5'-untranslated region variants for tunable expression in *Escherichia coli*. Biochem. Biophys. Res. Commun. 27: 136-141.

Richmond CS, glasner JD, Mau R, Jin H, Blattner FR (1999). Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res. 27: 3821-3835.

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanatinia H, Kofoid EC, Morrison TB, Parkinson JS (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, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Talukder AA, Hossain MA, Yamada M, Ishihama A (2006). Nucleoids dynamics in *Escherichia coli*: A growth phase dependent process. Bangladesh J. Microbiol. 23: 81-88.

Talukder AA, Yanai S, Yamada M (2005a). Analysis of reading frame and expressional regulation of randomly selected promoter-proximal genes in *Escherichia coli*. J. Gen. Appl. Microbiol. 51: 93-103.

Talukder AA (2005b). Survival and death in bacterial. In Yamada M (eds). Structure and composition of *Escherichia coli* nucleotides. 2:pp 77-101. Research Signpost, Kerala, India.

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

Talukder AA, Yanai S, 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. Biotech. Biochem. 58: 117-120.

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

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

Yamada M, Talukder AA, 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.

Yura T, Tobe T, Ito K, 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.