Review

Prokaryotic genome regulation: A revolutionary paradigm

By Akira ISHIHAMA*1

(Communicated by Tasuku HONJO, M.J.A.)

Abstract: After determination of the whole genome sequence, the research frontier of bacterial molecular genetics has shifted to reveal the genome regulation under stressful conditions in nature. The gene selectivity of RNA polymerase is modulated after interaction with two groups of regulatory proteins, 7 sigma factors and 300 transcription factors. For identification of regulation targets of transcription factors in Escherichia coli, we have developed Genomic SELEX system and subjected to screening the binding sites of these factors on the genome. The number of regulation targets by a single transcription factor was more than those hitherto recognized, ranging up to hundreds of promoters. The number of transcription factors involved in regulation of a single promoter also increased to as many as 30 regulators. The multi-target transcription factors and the multi-factor promoters were assembled into complex networks of transcription regulation. The most complex network was identified in the regulation cascades of transcription of two master regulators for planktonic growth and biofilm formation.

Keywords: transcription regulation, genome regulation, transcription factor, regulation network, genomic SELEX, Escherichia coli

1. Introduction

In the early stage of molecular biology, Escherichia coli served as a model organism of biochemical, biophysical, molecular genetic and biotechnological studies. Most of our current molecular-level knowledge of biological systems was obtained using this best-characterized prokaryote. With the advance in DNA sequencing technology, the complete genome sequence has been determined for a number of different E. coli strains. From the complete genome sequence, the whole set of protein-coding sequences on the E. coli genome has been predicted,1,2 even though the molecular functions of gene products remain unidentified for about half of the genes even for this best-characterized model prokaryote. At present, however, no short-cut theoretical procedure is available to identify functions of uncharacterized individual genes and proteins only from DNA sequences. In parallel with the genome sequencing, a variety of high-throughput techniques have been developed and employed to reveal the expression of the whole set of genes on the genome (the transcriptome) under a given culture condition. The high-throughput microarray has made a break-through for providing transcription patterns of the whole set of genes of the bacterial genome (for reviews see Lockhart and Winzeler, 2002; Steinmetz and Davis, 2004).3,4 On the proteomic level, the high-resolution two-dimensional PAGE system coupled with mass spectrometry (MS) has also elucidated the genome expression patterns at protein level (the proteome) (for reviews see Pandey and Mann, 2000; Han and Lee, 2006).5,6 In combination with the accumulated knowledge of the regulation of a large number of individual genes in E. coli, the transcriptome, proteome, metabolome and interactome data have been assembled to construct comprehensive models of the regulation of E. coli genome.7–11 At present, however, the mechanism how the genome expression pattern is determined and modulated remains unsolved. In this article I will overview the recent progress of our studies on the regulation of genome transcription focusing on the regulatory roles and networks of all transcription factors in a single model organism E. coli.
2. The model of genome regulation

Bacteria constantly monitor extracellular physicochemical conditions, so that they can respond by modifying their genome expression pattern. In bacteria, transcription initiation is the major step of regulation of the genome expression even through mRNA synthesis is also regulated at the step of transcription elongation and termination. mRNA degradation is also subject to control and, furthermore, increasing data indicate the involvement of translational control of mRNA through various mechanisms, including the interference of mRNA translation by regulatory RNAs and proteins.

The RNA polymerase holoenzyme or transcriptase of *E. coli* is composed of a multi-subunit core enzyme with subunit composition $\alpha_2\beta\beta'\omega$, and one of seven species of the $\sigma$ subunit with promoter recognition activity (Fig. 1A).$^{12-15}$ The gene selectivity of RNA polymerase holoenzyme is further modulated after interaction with a total of about 300 species of the transcription factor (Table 1).$^{14-16}$ The growing *E. coli* cells contain only about 2,000 molecules of the core enzyme per genome equivalent of DNA,$^{14}$ which is less than the total number of about 4,500 genes on the *E. coli* K-12 genome. Thus, the pattern of genome transcription is determined by the distribution of a limited number of RNA polymerase within the genome. One of the important research subjects of post-genome sequence era is to reveal the mechanism how the distribution of RNA polymerase within the genome is determined and modulated in response to environmental conditions.

Sometime ago we proposed that the pattern of genome transcription is altered through modulation of the gene selectivity of RNA polymerase after interactions with two groups of the regulatory proteins, *i.e.*, seven species of the $\sigma$ factor and a...
total of about 300 species of the transcription factor (Table 1 and Fig. 1A). The set of promoters recognized by RNA polymerase holoenzyme is determined by the species of associated factor. Within a group of the promoters recognized by one factor, the order of transcription level is determined primarily by the strength of promoter. The promoter strength is, however, subject to modulation by the second set of regulatory proteins, herein referred to transcription factors, which associates with the target DNA, usually located near promoters, and modulate transcription level from the promoters. The DNA-binding transcription factors interact with DNA-bound RNA polymerase subunits, together forming the transcription apparatus. Generally transcription factors are composed of two domains, one functioning as the sensor for external and internal signals and the other interacting DNA targets. In prokaryotes, the helix-turn-helix motif is the most common element in the DNA-binding domain. Based on the type of DNA-binding motifs and the organization of functional domains, we classified E. coli transcription factors into 63 families (Table 2). One group of transcription factors, known as negative regulators or repressors, is active by binding to target operators in the absence of low-molecular weight effectors, known as inducers. Promoters under the control of repressors are inactive in the presence of repressors, but become active once the repressors are dissociated from target DNA after association with the inducers. On the other hand, another group of transcription factors, known as positive regulators or activators, require interaction with effector ligands to function. Repressors and activators are inter-convertible depending on the position of DNA binding relative to promoters. Generally repressor-type transcription factors bind upon or downstream of promoters to interfere with the binding of RNA polymerase to promoter or its elongation along template, but in several cases, upstream-bound transcription factors repress transcription initiation by interfering with promoter escape due to strong protein-protein contact with RNA polymerase. On the contrary, activators bind upstream and in a few specific cases, downstream of promoters, for support of stable association of RNA polymerase to promoters (class-I transcription factors) or of promoter DNA opening (class-II transcription factors) (Fig. 1B). Noteworthy is that a single and the same transcription factor functions as both a repressor and an activator depending on the site of DNA binding relative to promoters.

Complete genome sequence allowed the prediction of full repertoire of the transcription factors in E. coli (Table 2). Approximately 290 species of the transcription factor are sequence-specific DNA-binding proteins. When bound to target DNA sites, these proteins interact directly with RNA polymerase subunits to function. In order to facilitate the frequent and quick exchange of RNA polymerase-interacting transcription factors, the affinity of protein-protein interaction between RNA polymerase and transcription factors must be weak. The binding of transcription factors at specific target sites near promoter is necessary for effective protein-protein interaction by increasing the local concentration of pairing proteins at promoter region. Besides these DNA-binding regulators, about 20–30 species of transcription factors associate directly with
Table 2. DNA-binding transcription factors in *Escherichia coli*

| Family | No. members | Member protein |
|--------|-------------|----------------|
| AidB   | 1           | AidB           |
| AlaS   | 1           | AlaS           |
| AlpA   | 1           | AlpA           |
| AraC   | 29          | Ada, AdiY, AppY, AraC, CelD, EnvY, EutR, FeaR, GadW, GadX, MarA, MelR, RhaR, RhaS, Rob, SoxS, XyIR, YbcM, YdeO, YdhP, YeaM, YfiE, YgiV, YidL, YgfO, YkgA, YkgD, YpdC, YqhC |
| ArgR   | 1           | ArgR           |
| ArsR   | 2           | ArsR, YgaV     |
| AsnC   | 3           | AsnC, Lrp, YbaO |
| BirA   | 1           | BirA           |
| BolA   | 1           | BolA           |
| CadC   | 3           | CadC, YqeH, YqeI |
| CaiF   | 1           | CaiF           |
| CdaR   | 1           | CdaR           |
| CheY   | 1           | MqsR           |
| ChtB   | 4           | ChtB, CrrR, DctR, DcuR |
| Crl    | 1           | Crl            |
| Crp    | 3           | Crp, Fnr, YeiL |
| Csp    | 1           | CspA           |
| DeeR   | 14          | AgaR, DeeR, DeeT, FrvR, FucR, GatR, GlpR, SgcR, SrlR, UlaR, YafY, YdfE, YffR, YihW |
| DicC   | 1           | DicC           |
| DnaA   | 1           | DnaA           |
| DtxR   | 1           | MntR           |
| Fis    | 1           | Fis            |
| FliC   | 1           | FliC           |
| FliD   | 1           | FliD           |
| Fur    | 2           | Fur, Zur       |
| GntR   | 23          | CsiR, DgoR, ExuR, FadR, FarR, FrLR, GlcC, LetR, MecR, NanR, PaaX, PdhR, PhtM, UxuR, YdeC, YdfH, YegW, Yghl, YidP, YiuP, YihL, YjsM, YjyR |
| GutM   | 1           | GutM           |
| IclR   | 7           | IclR, KdgR, MhpR, YglI, YfaX, YsaL, YjkI |
| IleR   | 1           | YgfA           |
| LexA   | 1           | LexA           |
| LuxR   | 12          | BgIJ, CsgD, GadE, MalT, RcsA, SdiA, UvrY, YahA, YhiB, YjgQ, YkgK, YqeH |
| LysR   | 46          | AbgR, AliR, AllS, Cbi, CynR, CysB, Dan, DmrlR, DsdC, GcvA, HcaR, IciA, IlvY, LenO, LrhA, LysR, MetR, MurR, Nac, NhaR, OxyR, PerR, PssR, QsaE, QsaD, TdcA, XapR, YafC, YahB, YbbO, YbeF, YbhD, YcaN, YcgZ, Ycdl, YdfL, YefY, YemY, YevE, YgfI, YhaJ, YjhC, YiuU, YidZ, YncL, Ynfl, YqfL |
| LytR   | 2           | YehT, YpdB     |
| MarR   | 3           | EmrR, MarR, SlyA |
| MerR   | 6           | CepR, MrrA, SoxR, ZntR, YcfQ, YegE |
| MetJ   | 1           | MetJ           |
| ModE   | 1           | ModE           |
| MtlR   | 2           | MtlR, YggD     |
| NadR   | 1           | NadR           |
| NagC   | 3           | Mlc, NagC, YphH |
| NikR   | 1           | NikR           |

Continued on next page.
the RNA polymerase in the absence of DNA.\textsuperscript{15,17} This group of transcription factors, referred to class-III and class-IV factors (Fig. 1A), are associated with \(O\) and \(O'\) subunits even during transcription elongation and control RNA chain elongation, attenuation and termination (Fig. 1B). There is a tight correlation between the mode of transcription action and the contact subunit (class-I, II, III and IV).\textsuperscript{27,28} Once we found this rule, we developed a quick identification system of RNA polymerase-transcription factor interaction sites by using a chemical nuclease-protease FeBABE.\textsuperscript{31–33}

At present, about two thirds of the estimated 300 transcription factors in \textit{E. coli} have been linked to at least one regulation target gene in the genome. Surprisingly the regulatory roles have been left unidentifed for about 100 species of the transcription factor even for this best-characterized model organism \textit{E. coli} (Table 2). Furthermore, even for about 200 species of the known transcription factor, only a single or a fraction of regulation targets have been identified and analyzed, but the whole sets of regulation targets have not been identified for these transcription factors. At present, however, we have purifed a total of 270 transcription factors, and have so far performed the Genomic SELEX screening for a total of 200 transcription factors.

| Family | No. members | Member protein |
|--------|-------------|----------------|
| Nlp    | 1           | Nlp            |
| NarL   | 9           | EvgA, FimZ, NarL, NarP, RcsB, UhpA, UvrY, YgeK, YhjB |
| NrdR   | 1           | NrdR           |
| NsrR   | 2           | IscR, NsrR     |
| NtrC   | 4           | AtoC, GluG, HydG, QseF |
| OgrK   | 1           | OgrK           |
| OmpR   | 14          | ArcA, BaeR, BasR, CpxR, CreB, CusR, KdpE, LsrR, OmpR, PhoB, PhoP, QseB, RstA, TorR |
| OraA   | 1           | OraA           |
| PadR   | 1           | YgfJ           |
| PhaN   | 1           | PaaX           |
| PutA   | 1           | PutA           |
| RfaH   | 1           | RfaH           |
| RpiR   | 4           | HexR, RpiR, YfeT, YfbH |
| RtcR   | 1           | RtcR           |
| SorC   | 2           | IdnR, YdeW    |
| TdcR   | 1           | TdcR           |
| TetR   | 13          | AcrR, BetI, EnvR, FabR, GusR, NefR, PutR, Tdk, YbhH, YbjB, YcfQ, YjiC, YjgJ |
| TrpR   | 1           | TrpR           |
| TyrR   | 8           | DhaR, FliA, HyfR, NorR, PrpR, PspF, TyrR, YgeV |
| Xre    | 8           | DicA, HipB, PunR, YdcX, YfgA, YfgB, YiaG, YjgT |
| AT*    | 11          | ChpBI, DinJ, HicB, HigA, MaaE, MpsA, PrfB, RelB, RnlB, YafN, YefM |

A total of 288 transcription factors can be classified into 63 families on the basis of DNA-binding motifs (Ishihama, 2012). At least one regulation target has been identified for 202 factors, shown in bold, while regulatory functions have not been identified for other 82 putative transcription factors, shown in italic. * AT, antitoxin (these low-molecular weight proteins carry DNA-binding activity even though they do not have known DNA-binding motif). Up to the present time, we have purified a total of 270 transcription factors, and have so far performed the Genomic SELEX screening for a total of 200 transcription factors.

3. Regulation modes of transcription factors

Combination of the microarray-based high-throughput technology and the ordinary molecular genetic analysis allows the identification of whole sets of genes whose expression depends on the functions of each of the transcription factors.\textsuperscript{34} For instance, the transcriptome pattern has been analyzed for various \textit{E. coli} strains growing under various stress conditions such as alterations in nutrients.\textsuperscript{35–38} at
increased or decreased temperature, upon exposure to oxidative stress, after addition of polyamines and metals, under anaerobic or acidic conditions, and within biofilms. Microarray analyses have also been performed for a number of E. coli mutants, each lacking a specific transcription factor such as ArcA, CRP, EvgA, Fis, FNR, IHF, LexA, LrhA, Lrp, ModE, NalL and NalP, NsrR and SdiA or over-producing a specific transcription factor such as MarA, SoxS and Rob.

Microarray technology produces gene expression data of E. coli on a genome scale for an endless variety of conditions. The gene set affected by depletion of one specific regulator gene or after overproduction of one specific transcription factor, however, does not represent the regulation targets under the direct control of the test transcription factor but instead includes large amounts of genes, which are affected indirectly due to the change in the expression level of direct target genes. Generally the direct targets represent only minor fractions of the genes detected by Microarray analysis, because often the genes for transcription factors are under the control of other transcription factors, together forming cascades of the transcription factor network.

The active conformation of transcription factors is generally a homo-dimer or homo-multimeric oligomer. In concert with the symmetrical conformation of transcription factors, their binding sites on DNA often include palindromic sequences. The regulation of target promoter by a transcription factor depends on the intracellular concentration of the transcription factor and its affinity to the target DNA site. The affinity of protein-protein association increases upon binding to DNA. Cooperative binding to DNA targets reduces the noise arisen by binding of non-specific proteins and increases the sensitivity for regulation. The DNA-binding activity of transcription factors is controlled by either interaction with effector ligands or covalent modification such as protein phosphorylation. The environmental conditions and/or cellular metabolic states influence both the activity of transcription factors through these two pathways and the intracellular concentration of transcription factors. In the case of prokaryotic transcription, transcription factors themselves sense changes in extracellular environmental conditions and/or intracellular metabolic states. For a small set of regulatory systems, two functions are mediated by two different proteins, i.e., sensors and response regulators. Of 300 species of transcription factors in E. coli, about 10% are involved in this mode of two-component system. The sensor kinases monitor environmental conditions and auto-phosphorylate at their conserved His residues while the receiver domain of the response regulators are phosphorylated at their conserved Asp residues by the sensor kinase to function as transcription factors. Overall the link between changes in environmental conditions and genome transcription involves signal-transduction pathways through the generation of effectors for modulation of the transcription factor activities or a cascade of protein phosphorylation of transcription factors.

Post-translational modification by reversible acetylation of transcription factors is a means of regulating gene expression in eukaryotes. Acetyl coenzyme A (AcCoA), the key molecule in central metabolism, functions as an acetyl donor by donating its acetyl group to lysine residues located on the surface of proteins. In bacteria, the global impact of protein acetylation not yet well understood. Recently, however, protein acetylation is also involved in regulation of a number of bacterial transcription factor such as E. coli RcsB. Protein acetylation of RNA polymerase was also indicated at the contact site of subunit with class-I transcription factors.

As in the case of σ factors, the intracellular concentrations of transcription factors are also subject to growth condition- or growth phase-dependent control. Using specific antibodies and quantitative immune-blot analysis, the intracellular concentrations have been determined for more than 150 species of transcription factors in E. coli (Kori, A. and Ishihama, A., unpublished). Except for about 10 species of the global regulator and the bifunctional nucleoid proteins with both architectural and regulatory functions (see below), the levels of transcription factors are less than 100 molecules per cell under steady-state of cell growth under laboratory culture conditions.

4. Regulation targets of DNA-binding transcription factors

4.1. Search in vitro for regulation targets: Genomic SELEX screening. The regulation targets under the direct control of a test transcription factor can not be identified simply relying on the comparison of transcriptomes or proteomes between wild-type and mutants lacking the test transcription factor because the majority of genes thus detected represents the set of genes indirectly affected (see
One short-cut approach for the identification of the promoters, genes and operons under the direct control of a test transcription factor is to determine the binding sites of the test transcription factor on the genome. Identification of the connections between transcription factors and DNA-binding sites represents a major bottleneck for modeling transcriptional regulatory networks. Thus the first step of a bottom-up approach toward understanding the regulatory network is to make the connection list of all the transcription factors and their DNA recognition motifs in the genome.

For quick search of DNA sequences that are recognized by DNA-binding proteins, the elegant SELEX (systematic evolution of ligands by exponential enrichment) system was developed, in which DNA-protein complexes were isolated from mixtures of a test DNA-binding protein and synthetic oligonucleotides of all possible sequences followed by sequencing of protein-bound DNA fragments. Typically, the starting DNA library used for screening contained $4^n$ different sequences, where $n$ represents the length of nucleotide residues of the DNA probes. Upon increase in the chain length, however, the number of probe species increase and as a result, it becomes difficult to solve all the long-sized probes at the effective concentration needed for protein binding. To overcome the solubility problem, mixtures of genome DNA fragments can be used in place of synthetic oligonucleotide mixtures because the binding sites of test transcription factors are located on the genome. In order to search for regulation targets by hitherto uncharacterized putative transcription factors as well as to identify the whole set of targets by known transcription factors, we have then developed an improved method of ‘Genomic SELEX’ (Fig. 2) and initiated a systematic search for DNA sequences recognized by each of all 300 species of the DNA-binding transcription factor from E. coli. For determination of the
sequences of protein-bound SELEX DNA fragments, two procedures are employed, SELEX-clos (cloning and sequencing) and SELEX-chip (mapping by tilling array consisting of 22,000 species of 60 b-long oligonucleotide probe aligned at 160 bp intervals along the *E. coli* genome) (Fig. 2). Up to the present time, the newly developed ‘Genomic SELEX’ has been successfully employed for identification of the recognition and binding sequences of about 200 species of *E. coli* transcription factors (Table 1), of which the results of target screening have been published for AllR,77) AscG,78) BasR,79) CitB,80) Cra,78),81) CRP,82) Dan (renamed from YgiP),83) H-NS,84) LenO,84),85) NemR (renamed from YdhM),86) PdhR,87) RstA,88) RutR (renamed from YcdC),89) and TyrR.90) After repetition of Genomic SELEX, DNA sequences with high affinity to transcription factors are enriched and thus in SELEX-clos method, the proportion of plasmid clones carrying SELEX sequences with high affinity to the test transcription factor increases, thereby providing an list of the affinity order to the test factors. On the other hand, the whole set of factor-binding sequences can be obtained by SELEX-chip method (Fig. 2). Since the low level peaks are unreliable, the number of factor-binding peaks changes, depending on the setting of cut-off level of background pattern without protein addition. Combination of the SELEX-clos and SELEX-chip patterns provides not only the more reliable set of regulation targets by the test transcription factor but also the order of binding affinity between the predicted targets. The fraction of known targets successfully identified by the Genomic SELEX screening varies depending on the test transcription factors, mainly because the current databases *E. coli* transcription factors such as RegulonDB include regulation targets with different levels of accuracy, some being predicted in silico simply based on the presence of sequences similar to the recognition sequence by test transcription factors but without experimental confirmation.

The Genomic SELEX is a powerful experimental system but has potential pitfall. For instance, in order for Genomic SELEX to work in the search of regulation targets by the hitherto uncharacterized regulators, the conditions under which the test transcription factors are active need to be known before experiments are conducted. Since most of the uncharacterized putative transcription factors are considered to be needed for expression of the genes for response to as yet unidentified environmental stresses in nature. In the absence of required effector ligands such as inducers and co-repressors or specific reaction conditions, Genomic SELEX screening yields mixtures of non-specific sequences. In these cases, one possible approach to identify specific effectors or conditions for activation of transcription factors, the phenotype microarray (PM) may be useful, in which the growth of *E. coli* mutants lacking the genes for test transcription factors can be examined under up to 2,000 different conditions to monitor the utilization of various C, N, P and S sources, survival at different pH ranges or different osmolarity, and the sensitivity to various drugs and chemicals.91)

### 4.2. Search in silico for regulation targets

Recognition in silico for transcription regulatory signals in bacterial genomes is still a difficult problem of bioinformatics because of the lack of algorithms capable of making reliable predictions. The initial computer analysis of transcription factor-binding sequences produces a huge number of false positives. However, once the list of recognition and association sequences by transcription factors are established after Genomic SELEX, the consensus sequence can be deduced, which can afterward be used for in silico search of additional targets using the whole genome sequence. Comparative analysis of multiple genomes is one approach for confirmation of the transcription factor-DNA binding site interactions.92),93) The comparative approach is based on the assumption that sets of co-regulated genes are conserved in related bacteria. Computational methods of phylogenetic footprinting have been applied to the *E. coli* genome, allowing the discovery of many novel transcription factor-binding sites.94),95) Clustering of phylogenetic footprintings has generated DNA motif models for both unknown transcription factors and many previously characterized transcription factors, altogether yielding the sets of regulons.96),97)

### 4.3. Search in vivo for regulation targets: NIP-chip system

Traditional methods in molecular genetics have been successfully employed to identify only a fraction of the transcription regulatory interactions.98) Modern high-throughput methods such as chromatin immuno-precipitation coupled with promoter microarrays (ChIP-chip) have been developed to rapidly associate a number of transcription factors with their cognate binding sites in the yeast genome,99)–101) providing the genome-scale interaction necessary to model the regulatory network. Initial efforts of the application of ChIP-chip to prokaryotes have been made for identification of the localization on the *E. coli* genome of individual components of the transcription apparatus such as
RNA polymerase, CRP, Fis, IHF and H-NS, NsrR, RutR, and Lrp. Genomic SELEX screening allows the identification of whole set of potential binding sites for one specific transcription factor, while the actual binding sites of the test transcription factor under a given culture condition can be identified by ChIP-chip analysis. All these successful attempts were made for identification of the binding sites of abundant DNA-binding proteins such as nucleoid proteins. As in the case of Genomic SELEX with uncharacterized transcription factors, the growth conditions under which the transcription factors are present at the level enough for detection by immune-purification and the test transcription factors are functional. The expression level of most transcription factors in E. coli under laboratory culture conditions is too low for reliable detection with ChIP-chip analysis (Ishihama et al., in preparation).

For ChIP-chip analysis, cells must be treated with a reagent, typically formaldehyde, which creates covalent crosslinks between proteins and genome DNA. An antibody specific for a protein of interest is then used to immmuno-precipitate protein-bound DNA fragments, which are subsequently labeled in an amplification reaction and hybridized to DNA microarrays for mapping the protein-bound DNA fragments. Initially the ChIP-chip system was developed with yeast and animal cultured cells and formaldehyde treatment was performed for 15–20 min. Formaldehyde, a highly toxic carbonyl compound, reacts as an electrophile with the side-chains of arginine and lysine, resulting in the formation of glycation end-products, and causes protein-protein and protein-DNA cross-links in vivo. As a stress response to formaldehyde treatment, the distribution of transcription factors changes even during formaldehyde treatment. Moreover, the cross-linked proteins to the E. coli genome are gradually digested during formaldehyde treatment (Ishihama, A. et al., unpublished). Attempts are therefore being made to improve the ChIP-chip system to minimize the time down to a few minutes and concentration of formaldehyde treatment for application to prokaryotes. We propose the improved method as NIP (nucleoid immunoprecipitation)-chip system (Ishihama, A. et al., in preparation).

5. Transcription factor-binding sites on the genome

In sharp contrast with the eukaryotic genomes, non-coding sections are limited in the prokaryotic genomes. In the case of E. coli genome, for instance, more than 90% DNA sequence is used for coding whereas non-coding sequences occupy only less than 10% (1, 2). Transcription factors so far analyzed tend to bind to the non-coding intergenic regions. Even for the bifunctional nucleoid proteins such as IHF and Fis, approximately 50% are bound in vivo within intergenic regions as detected by ChIP-chip analysis. After extensive Genomic SELEX search of the binding sites by more than 200 species of E. coli transcription factors so far examined, the binding preference for coding regions has been identified only for a specific set of transcription factors, implying ORF (open reading frame)-associated transcription factors may play an as yet unidentified regulatory role(s) (Ishihama, A. et al., in preparation).

The spacing between transcription and translation start sites in the E. coli genome mostly ranges up to 50 nucleotides, but a small number of E. coli genes carry longer untranslated flanking sequences ranging up to about 300 nucleotides upstream from the translation start codon. Recently these regions have been indicated to encode small peptides or small RNA with regulatory functions. The distance between transcription factor-binding sites and transcription initiation sites is various, ranging approximately from +200 to −100. The determination of transcription factor-binding sites relative to promoters contributes better understanding of regulatory modes of the respective promoters. Among the transcription factors that bind to non-coding intergenic regions, functional binding sites for a transcription factor is present in both upstream and downstream of transcription initiation sites. Generally positive factors bind upstream from promoter −10 while negative factors binds downstream from promoter −35. One reliable but simple criterion for this classification of transcription factors into activated and repressed subsets is the location of their binding sites relative to that of the RNA polymerase-binding site (or promoter).

For determination of the regulatory signals associated with each E. coli promoter, a collection of about 2,000 promoter assay vectors has been constructed, in which about 500 bp-long DNA fragment upstream of the translation initiation codon was isolated from each gene and inserted into GRP promoter assay vector carrying two-fluorescent protein reporters. The initiation codon of promoter fragment was sealed to the initiation codon of GFP-coding sequence while another fluorescent protein RFP was fused to a reference promoter lacUV5 in the
same vector. The involvement of test transcription factors in regulation of the target promoters can be easily confirmed by measuring GFP/RFP ratio in mutants lacking the factor gene or after over-expression of the test factor.77),86)

Transcription factors are generally functional when bound at either orientation relative to the RNA polymerase binding site,112) possibly because transcription factors form symmetric oligomers or induce DNA looping so as to make contact with either the flexible alpha or sigma subunits of the RNA polymerase.14),15),17),113)

5.1. Single-target transcription factors. In the classic molecular genetic studies, prokaryotic promoters were believed to be regulated by a single specific regulatory protein, either a repressor and an activator, as originally identified in the lac operon regulation by LacI repressor.116) Accordingly each of a large number of “specific” or “local” transcription factors have been believed to regulate the expression of one specific gene or a small number of transcription units.88),117) After Genomic SELEX screening, however, most of the E. coli transcription factors were found to regulate multiple promoters, and most of the E. coli promoters were indicated to be under the control of multiple transcription factors.15) Among a total of more than 200 transcription factors examined, the single-target transcription factors are very rare, ranging approximately less than 20, including BetI (betaine inhibitor) (Fig. 3A), NorR (NO reduction and detoxification regulator) (Fig. 3B), NanR (N-acetyl-neuraminic acid regulator) and UlaR (utilization of L-ascorbate operon regulator).

5.2. Multi-target transcription factors. Until recently only a small number (about 10–20) of transcription factors were believed to be “global” regulators, which influence the expression of a large number of transcription units that belong to different metabolic pathways, thereby exhibiting pleiotropic phenotypes.25),118),119) After Genomic SELEX screening of transcription factors with known regulatory roles, however, the number of regulation targets were found to be more than those hitherto identified or predicted,15),16) ranging from one specific (in the case of single-target transcription factors as noted above) to more than 1,000 targets (see below). This finding raised a criticism over the classic classification of transcription factors into a larger number of “specific” (local) regulators and a small number of “global” regulators. After the Genomic SELEX screening, it is now difficult to discriminate 300 transcription factors simply into two groups, “specific (local)” and “global” regulators. Instead a linear gradient is formed with respect to the number of regulation targets.

A set of promoters, genes or operons have been found to be controlled by one and the same transcription factor, altogether forming the “regulon”. The regulons under the control of multi-target regulators include a large number of genes or operons. The genes organized in one regulon are often a member of other regulons, altogether forming complex and hierarchic networks of transcription factors (see below).

5.3. Global regulators for carbon metabolism: CRP (cAMP receptor protein) and Cra (catabolite repressor activator). Carbon availability in the environment influences the expression pattern of a number of genes in E. coli in various ways. cAMP receptor protein CRP, also called catabolite gene activator protein CAP, was the first purified transcription activator,20) and is the best-characterized global regulator involved in the regulation of genes for transport and utilization of carbon sources.121)–123) CRP is a dual regulator, acting as an activator or a repressor depending on the position of CRP binding relative to promoters.122) In the absence of glucose, cAMP is synthesized, which associates CRP for its conversion into the active regulator in transcription. The functional CRP protomer is composed of two molecules of CRP, each being associated with cAMP. Binding of cAMP to its N-terminal domain leads to activate the C-terminal DNA-binding domain,124),125) of which the characteristic helix-turn-helix (H-T-H) motif is responsible for interaction with CRP-box consisting of a palindromic TGTGAnnnnnnTCACA sequence associated with target promoters.120) When CRP binds DNA, it induces DNA bending of about 87°.127)–129) The DNA-bound CRP is the first transcription factor, that was identified to directly interact with the promoter-bound RNA polymerase for function.27),28),30)

The total number of known target promoters under the direct control of cAMP-CRP is reaching to 100.11) After Genomic SELEX searching, however, a total of 378 promoters have been identified as the potential targets (Fig. 3D; and Fig. 4A; Table 3).82) The CRP regulon includes a large number of the genes encoding enzymes and transport systems of sugars. Unexpected findings are that the major role of CRP is the control of the genes for uptake carbon sources and for the metabolism downstream of glycolysis, including TCA cycle and aerobic respiration (Fig. 4C). Most of the transporter genes for carbon sources are under the control of CRP.
In addition to CRP, a number of the genes for both glycolysis and gluconeogenesis are under the control of catabolite repressor activator (Cra), initially characterized as FruR (fructose repressor). Cra, a member of GalR-LacI family, consists of two functional domains, an N-terminal DNA-binding domain with H-T-H motif and a C-terminal inducer-binding and subunit-subunit contact domain. Cra controls transcription of the genes in major pathways of carbon and energy metabolism by playing a key role to modulate the direction of carbon flow through the different metabolic pathways of energy metabolism, but independently of cAMP-CRP. After genomic SELEX screening, we found the regulation targets of Cra are at least 178 (Fig. 3C and Fig. 4A), more than the number 23 that were identified previously and listed in the database (Fig. 4A; Table 3). Cra was found to play as an activator of most of the genes encoding enzymes for gluconeogenesis.
genesis, TCA cycle, and glyoxylate shunt pathway, and as a repressor of the genes encoding Entner-Doudoroff pathway and glycolysis (Fig. 4C). Derepression of the glycolysis genes takes place when the repressor Cra is inactivated after interaction with inducers such as D-fructose-1-phosphate and D-fructose-1,6-bisphosphate. In the absence of these inducers, Cra recognizes and binds to Cra box consisting of TGAAACGTTTCA palindromic sequence. In the presence of glucose, the intracellular concentration of the inducers increase, which interact with Cra to prevent its binding to the target operons. On the other hand, the genes activated by Cra is subject to regulation through the control of Cra level.

Genomic SELEX screening revealed that a set of genes are controlled by both CRP and Cra (Fig. 4B). The decision which regulator operates under a given condition is determined by the intracellular concentrations of respective effectors, cAMP and phosphorylated fructose.

5.4. Global regulators for nitrogen metabolism: RutR (regulator of pyrimidine utilization), LeuO (leucine biosynthesis regulator) and Lrp (leucine-responsive regulatory protein). RutR was originally identified as a repressor of the rut
identified as a regulator of the genes involved in leucine biosynthesis.\textsuperscript{136} Genomic SELEX screening indicated the presence of at least 140 LeuO-binding sites on the \textit{E. coli} genome (Table 3).\textsuperscript{85} Interestingly 133 LeuO-binding sites (95\%) were found to overlap with the binding sites of H-NS, the universal silencer of stress-response genes including the foreign genes such as phage genes. This finding indicates that one important biological role of LeuO is anti-silencing of H-NS-mediated repression of some toxic genes. In fact, a set of stress-response genes including cryptic chaperone/usher-type fimbriae operons are under the control of antagonistic interplay between LeuO and H-NS.\textsuperscript{84}

LeuO is also a transcription factor sensing leucine level and is believed to regulate the genes for amino acid transport, biosynthesis and catabolism,\textsuperscript{137,138} similar to the role of CRP in carbohydrate metabolism. More recently Lrp has been suggested to be involved in regulation of the genes for not only amino acid metabolism but also nutrient transport, pilus synthesis and even carbon metabolism in particular those expressed in stationary phase. In agreement with these observations, we identified as many as 506 genes as regulation targets of Lrp by Genomic SELEX screening (Table 3) (Shimada \textit{et al.}, in preparation). In good concert with the sensing role of Lrp of leucine availability, a number of the genes for nitrogen metabolism and the genes for components of translation system appear to be under the direct control of Lrp. In addition, a variety of stress-response genes that respond to the nutrient availability are also included in the list of Lrp targets.

### 5.5. Global regulators for energy metabolism: FNR (fumarate nitrate reduction) and Dan (DNA-binding protein under anaerobic conditions)

FNR, initially named for the mutant defect in “fumarate and nitrate reduction”, is another global transcription factor of the CRP/FNR superfamily. FNR plays a key role in the metabolic transition from aerobic to anaerobic growth through the regulation of a number of genes.\textsuperscript{139,140} As in the case of CRP, FNR has an N-terminal sensory domain, an internal dimerization domain, and a C-terminal H-T-H DNA-binding domain. Generally, FNR activates the genes involved in anaerobic metabolism, but it also regulates transcription of a number of genes with other functions, such as acid resistance, chemotaxis, and cell structure. The intracellular concentration of FNR stays constant under both anaerobic and aerobic growth, but its activity is regulated directly by oxygen. The sensory domain of FNR contains five

| TF | Genomic SELEX (A) | RegulonDB (B) | A/B |
|----|------------------|---------------|-----|
| [A] Global regulators for carbon metabolism | | | |
| Cra | 178 | 23 | 7.7 |
| CRP | 378 | 150 | 2.5 |
| [B] Global regulators for nitrogen metabolism | | | |
| LeuO | 140 | 6 | 23.3 |
| Lrp | 506 | 40 | 12.7 |
| [C] Nucleoid-associated global regulators | | | |
| Fis | 1,269 | 95 | 13.4 |
| H-NS | 987 | 72 | 13.7 |
| IHF | 813 | 80 | 10.2 |
| Rob (CbpB) | 916 | 15 | 61.1 |

FNR has an N-terminal sensory domain, an internal dimerization domain, and a C-terminal H-T-H DNA-binding domain. Generally, FNR activates the genes involved in anaerobic metabolism, but it also regulates transcription of a number of genes with other functions, such as acid resistance, chemotaxis, and cell structure. The intracellular concentration of FNR stays constant under both anaerobic and aerobic growth, but its activity is regulated directly by oxygen. The sensory domain of FNR contains five
Cys residues, four of which are essential for linking the [4Fe-4S] cluster. Under anaerobiosis, FNR is activated by forming a [4Fe-4S] cluster that causes a conformational change and dimerization of the protein but upon exposure to O₂, FNR is inactivated via oxidation of [4Fe-4S] cluster into [2Fe-2S]. The activated FNR conformation is able to bind the FNR-box sequence consisting of a palindromic TTGATNNNNATCAA sequence.

A systematic search for the regulation targets by Dan (DNA-binding protein under anaerobic conditions, renamed from YgiP, by using the genomic SELEX indicated a total of more than 700 binding sites within the *E. coli* genome. At low concentrations, Dan binds at various sites and enhances the sensitivity of associated DNA to nucleolytic digestion because of Dan-induced local opening of DNA. At high concentrations, Dan covers the entire DNA surface as observed by AFM and protected the DNA from nucleolytic digestion. The intracellular level of Dan is very low under aerobic conditions, leaving it hitherto unidentified as a nucleoid protein, but increased more than 100-fold to the level as high as those of nucleoid proteins HU and IHF under hypoxic and anaerobic culture conditions. Dan is a novel nucleoid protein of *E. coli* under the anaerobic condition. As in the cases of other nucleoid proteins, Dan plays dual roles in both maintenance of the nucleoid architecture and expression of the nucleoid function under the anaerobic condition. One regulation target of Dan is the ttd operon encoding L-tartrate dehydratase and the L-tartrate:succinate antiporter. An *E. coli* mutant lacking dan showed retarded growth under anaerobic conditions. In the case of FNR, there are four Cys residues within a limited region Dan of 310 residues in length.

**5.6. Nucleoid proteins as global regulators:**

**IHF (integration host factor) and Fis (factor for inversion stimulation).** In the *E. coli* nucleoid, two groups of the nucleoid protein exist, universal nucleoid proteins (UNPs) that always stay in the nucleoid; and growth phase-specific nucleoid proteins (GNPs) that appear only at specific phases of cell growth. IHF, a member of universal nucleoid proteins (GNPs), was originally found to be required for the site-specific recombination of phage λ with the *E. coli* genome. IHF is a heterodimer consisting of the two subunits, *IhfA* (HimA) and *IhfB* (HimD, Hip), that share about 25% amino acid identity. IHF is highly abundant during all the growth phases, thus being classified into UNP. The intracellular concentration of IHF ranges from 6,000 dimers per cell at the log phase and to 3,000 dimers in stationary phase.

By using Genomic SELEX screening, a total of 813 IHF-binding sites were identified on the *E. coli* genome (Table 3) (Ishihama et al., in preparation). The list of IHF-binding targets supports its dual role model, i.e., an architectural role for DNA supercoiling and DNA duplex destabilization and a regulatory role of genome functions controlling processes such as DNA replication, recombination, and the expression of a number of genes. IHF binds tightly to DNA regions of about 40 bp carrying the 13-bp consensus sequence with A/T-rich elements upstream of the core consensus sequence. The structure of IHF bound to DNA has been solved, showing that IHF makes only a few contacts with the minor groove. Thus the DNA recognition specificity is due to the sequence-dependent structural parameters of the DNA, where A/T-rich regions play an important role. The bend angle induced by IHF is approximately 160°. In transcription regulation, IHF acts to facilitate the formation of the loop around promoter for conversion into active conformation. The binding to low-affinity sites and introduction of sharp bends in the promoter DNA promote the formation of initiation complex for transcription.

Fis is a member of the growth phase-specific nucleoid proteins (GNPs) associated with the growing cell nucleoid as Dps in stationary-phase cells and Dan in cells growing under anaerobic conditions. Under optimal growth conditions, Fis is the dominant nucleoid protein, reaching to the concentration of as high as 60,000 copies in a single log-phase cell and plays an essential role for maintenance of the nucleoid competent for transcription of the growth-related genes. Genomic SELEX screening identified a total of as many as 1,269 Fis-binding sites in both intergenic spacers and open reading frames on the *E. coli* genome (Fig. 3E; Table 3), implying its involvement in regulation of a large number of genes that are expressed in growing cells. Expression of fis is regulated by several systems and at different levels. At the transcription level, Fis is autoregulated, induced by high supercoiling levels, and regulated by both growth rate-dependent and stringent control systems. Transcription of fis is also regulated by the availability of the nucleotide triphosphate CTP, the initiation nucleotide of fis RNA synthesis. DksA, an RNA polymerase-interacting transcription factor, inhibits transcription of fis by increasing the
sensitivity to ppGpp, another RNA polymerase-interacting nucleotide transcription factor.\(^2\) The GNP group of the nucleoid proteins carries dual functions, playing an architectural role for folding the genome DNA into the nucleoid structure and its maintenance and a regulatory role in genome functions such as transcription, replication, DNA inversion and transposition, and phage integration-excision. As a transcriptional regulator, Fis regulates the expression of a number of genes involved in translation (rRNA, tRNA and r-protein genes), virulence, biofilm formation, energy metabolism, stress response, central intermediary metabolism, amino acid biosynthesis, transport, cell structure, carbon compound metabolism, amino acid metabolism, nucleotide metabolism, motility, and chemotaxis.\(^1\) Accordingly microarray analysis indicated that transcription of approximately 21% of genes is modulated directly or indirectly by Fis, while ChiP-chip analysis indicated that Fis binds to 894 DNA regions in the genome.\(^2\) A core binding site of Fis is as long as 15 bp with partial dyad symmetry commonly presents an AT-rich sequence. Once bound to DNA, Fis bends the DNA between 40° and 90°. This bending stabilizes the DNA looping to regulate transcription and to promote DNA compaction.\(^2\)

In stationary-phase cells, Fis decreases to nearly imperceptible level, and thus Fis was identified as a growth condition-specific nucleoid protein (GNP).\(^2\) The positions of Fis binding on the genome are occupied by Dps (DNA-binding protein under starved conditions), another GNP protein. Dps becomes the major nucleoid protein produced only in starved stationary-phase cells\(^2\) and plays a protecting role of the genome in resting \textit{E. coli} cells from environmental stresses such as high levels of toxic iron.\(^2\) The binding sites of all these multiple factors are carried out under various stressful conditions, multiple promoters could be identified in a single gene or operon of \textit{E. coli}.

Among the set of promoters under the control of a single and the same sigma factor, the level of transcription varies depending on the culture conditions or the growth phase. For this control of the promoter strength recognized by the same sigma factor, multiple species of the transcription factor are involved. The current promoter data bases indicate that approximately 50% of the \textit{E. coli} promoters is under the control of one specific regulator while other 50% genes are regulated by more than two transcription factors.\(^2\) After genomic SELEX search, however, we found that most of the \textit{E. coli} promoters carry the binding sites for multiple transcription factors,\(^2\) each factor monitoring a different environmental condition or a metabolic state. The involvement of multiple transcription factors may be employed for the fine tuning system of genome transcription. For instance, the expression of genes encoding metabolic enzymes is controlled by metabolites in the metabolic cycle the enzymes participate, each metabolite being monitored by a specific transcription factor. Likewise the promoters for the genes involved in construction of cell structures are controlled by environmental conditions and factors, each being monitored by a different transcription factor. The binding sites of all these multiple factors are located in a single and the same promoter.

### 6. Multi-factor promoters: involvement of multiple transcription factors for regulation of single promoters

The number of genes or operons with multiple transcription initiation sites (and thus multiple promoters) is increasing after detailed analysis of transcription regulation of the stress-response genes\(^2\),\(^2\) and \textit{in silico} analysis of \textit{E. coli} genome with newly developed programs for search of promoters.\(^2\) Often each promoter of the same gene or operon is recognized by a different sigma factor, and thus it is difficult to have a chance of detecting all potential promoters under a single culture condition.\(^2\) If experiments for mRNA detection are carried out under various stressful conditions, multiple promoters could be identified in a single gene or operon of \textit{E. coli}.

Among the set of promoters under the control of a single and the same sigma factor, the level of transcription varies depending on the culture conditions or the growth phase. For this control of the promoter strength recognized by the same sigma factor, multiple species of the transcription factor are involved. The current promoter data bases indicate that approximately 50% of the \textit{E. coli} promoters is under the control of one specific regulator while other 50% genes are regulated by more than two transcription factors.\(^2\) After genomic SELEX search, however, we found that most of the \textit{E. coli} promoters carry the binding sites for multiple transcription factors,\(^2\) each factor monitoring a different environmental condition or a metabolic state. The involvement of multiple transcription factors may be employed for the fine tuning system of genome transcription. For instance, the expression of genes encoding metabolic enzymes is controlled by metabolites in the metabolic cycle the enzymes participate, each metabolite being monitored by a specific transcription factor. Likewise the promoters for the genes involved in construction of cell structures are controlled by environmental conditions and factors, each being monitored by a different transcription factor. The binding sites of all these multiple factors are located in a single and the same promoter.

#### 6.1. Search for promoter-specific transcription factors

The most typical examples of the multi-factor promoter system are the promoters for the genes encoding the master regulator FlhCD for flagella formation and the master regulator CsgD for biofilm formation (Fig. 5A). The complexity of these two multi-factor promoters reflects the two opposite behaviors of bacterial survival, \textit{i.e.}, planktonic growth as single cells and biofilm formation as bacterial community, in stressful conditions in nature. After Genomic SELEX screening of regulation targets for more than 200 transcription factors, we realized more than 10 transcription factors bind within a narrow region of the promoter of \textit{csgD} encoding the master regulator of biofilm formation.\(^2\) In order to identify the whole set of transcription factors involved in the regulation of \textit{csgD} promoter, we have developed ‘Promoter-Specific Transcription Factor’ (PS-TF) screening system \textit{in vitro} (Ishihama, A. et al., in preparation). To mixtures of \textit{csgD} promoter and reference promoters, each of 300 purified transcription factors
were added and after incubation, subjected to mixed gel shift assays. To our surprise, as many as 30 transcription factors were found to specifically bind the \textit{csgD} promoter but not to other promoters (Fig. 5B), indicating that about 30 transcription factors participate in regulation of the \textit{csgD} promoter. This finding indicates that as far as the number of transcription factors is concerned, transcription regulation in prokaryotes is more complex than in eukaryotes.

6.2. Control of bacterial habits between single planktonic growth and biofilm formation.

Under laboratory culture conditions rich in nutrients and oxygen, bacteria exhibit single-cell planktonic growth habit. In stressful conditions in nature, however, surface-associated communities of bacteria, “biofilm”, play a key role in bacterial survival. Biofilm development can be divided into several distinct stages: attachment of cells to a surface, association of cells onto the surface-attached cell aggregates, and growth of the cells into a sessile biofilm (Fig. 5A). Biofilms tend to develop on a surface of plastic materials in nature or on tissues in host animals. The initial reversible interaction between a bacterial cell and a solid surface is mediated by non-specific physical interactions. This transient attachment is
reinforced by adhesins that are located on the bacterial cell surface or on cellular appendages such as pili and fimbriae, leading to irreversible attachment of the bacterial cell to the surface.\textsuperscript{158} The second stage of biofilm development involves the multiplication of bacterial cells on the surface and the concomitant synthesis of extracellular polysaccharide matrixes. The matrix holds the bacterial cells together in a mass and firmly attaches the bacterial mass to the underlying surface. In addition to providing a structural scaffold for the biofilm colony, the matrix also contributes to biofilm-mediated antimicrobial resistance, either by acting as a diffusion barrier, or by binding directly to antimicrobial agents and preventing their access to the biofilm cells.\textsuperscript{139} The pathway of biofilm formation is under a complex network of transcription factors. As noted above, a total of 20–30 transcription factors were found to be directly involved in regulation of the promoter for \textit{csgD} encoding the master regulator of biofilm formation (Fig. 5B).\textsuperscript{156,157} The expression of these primary transcription factors that directly regulate the \textit{csgD} promoter are under the control of secondary transcription factors. Various environmental factors and conditions affect the \textit{csgD} expression via a set of transcription factors.

Among the transcription factors involved in \textit{csgD} regulation, we identified FlhDC, the master regulator of flagella formation. FlhDC represses the \textit{csgD} promoter. On the other hand, CsgD was found to repress the genes for flagella formation.\textsuperscript{156} These observations altogether indicate that the two pathways of bacterial habits, planktonic growth and biofilm formation, are tightly interconnected each other by repressing their master regulators (Fig. 5C). Furthermore, downstream of both regulation cascades, the genes for sigma factors are included, \textit{i.e.}, the \textit{rpoF} gene in the pathway of flagella formation and the \textit{rpoE} gene in the pathway of biofilm formation. The formation of new sigma factors renders the respective pathway into irreversible cascade.

7. Hierarchic networks of transcription factors

Transcription factors and their regulation target genes and operons are generally located near each other in the genome. Such distance constraints are considered to be arisen from the horizontal gene transfer.\textsuperscript{93,161} The Genomic SELEX search supports the prediction that transcription factors and their regulated genes tend to evolve concurrently. The regulator-target sets were then interconnected through cross-talks between regulators and targets. The transcription factor network involved in regulation of single promoters can be connected to yield the interaction network consisting of a number of signaling pathways.\textsuperscript{162} These interacting pathways construct an intricate network. This network integrates diverse extracellular and intracellular signals to ensure the regulated expression of appropriate genes in the genome at proper time and proper level. The signals in one pathway is often transferred into another pathway.

The cross-talk in signal transduction among various signaling pathways has been recognized, particularly among the two-component systems (TCSs) consisting of two components, \textit{i.e.}, sensor His kinase and response regulator.\textsuperscript{63,64} \textit{E. coli} harbors a total of about 36 pairs of TCS. Sensor kinases monitor external factors and conditions, self-phosphorylate His residues in their receiver domains, and then transfer phosphoryl residue to Asp residues of response regulators to function. A single response regulator is often trans-phosphorylated by sensor kinases organized in different TCS pathways.\textsuperscript{64} Cross-talks take place not only through the sharing of the same targets between different transcription factors but also during signal transduction pathway such as recognition of the same external signals by two different sensors. Comprehensive microarray analysis of a set of 36 TCSs mutants also indicated high-correlation for gene expression among deletion mutants.\textsuperscript{63} Deletion of one TCS mutant often influence transcription pattern under the control of other TCSs, implying the sharing of same regulation targets between two TCSs.

Conclusion

The regulation targets of each of approximately 300 species of the transcription factor, the second-step regulator involved in the functional differentiation of RNA polymerase, in \textit{Escherichia coli} are more than those listed in databases. In this study, we identified the regulatory roles and regulation targets for most of transcription factors from a single model organism \textit{E. coli}. Regulatory interactions in \textit{E. coli} can now be recognized to be more complicated than those hitherto understood and probably as complex as those in eukaryotes, involving the multi-factor promoters and the multi-target regulators, altogether forming hierarchic regulation networks.
Acknowledgements

The author acknowledges Tomohiro Shimada, Jun Teramoto and Hiroshi Ogawara for experimental support and discussion, and Ayako Kori and Kayoko Yamada for technical and secretarial assistance. The research was supported by Grants-in-Aid for Scientific Research Priority Area (17076016) and Scientific Research (A) (21241047) and (B) (18310133) from MEXT (Ministry of Education, Culture, Sports, Science and Technology of Japan), and MEXT-Supported Program for the Strategic Research Foundation at Private Universities 208-2012 (S0801037).

References

1) Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.B., Blattner, F.R., Chadhuri, R.R., Glasner, J.D., Mori, H., Horiiuchi, T., Keseler, I.M., Kosuge, T., Perna, N.T., Plunkett III, G., Rled, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D.S. and Wanner, B.L. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Res. 34, 1–9.

2) Hayashi, K., Morokca, N., Yamamoto, Y., Fujita, K., Iseno, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H. and Horiiuchi, T. (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol. Syst. Biol. 2, 2006.0007.

3) Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. Nature 405, 827–836.

4) Steinmetz, L.M. and Davis, R.W. (2004) Maximizing the potential of functional genomics. Nat. Rev. Genet. 5, 190–201.

5) Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. Nature 405, 837–846.

6) Han, M-J. and Lee, S.Y. (2006) The *Escherichia coli* proteome: past, present and future prospects. Microbiol. Mol. Biol. Rev. 70, 362–439.

7) Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. and Teichmann, S.A. (2004) Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. 14, 283–291.

8) Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: Understanding the cell’s functional organization. Nat. Rev. Genet. 5, 101–113.

9) Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Pepeleo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J., Martinez-Antonio, A. and Collado-Vides, J. (2006) RegulonDB (version 5.0): *Escherichia coli* K12 transcriptional regulatory network, operon organization, and growth conditions. Nucleic Acids Res. 34, D394–D397.

10) Baumbach, J., Wittkop, T., Rademacher, K., Rahmann, S., Brinkrolf, K. and Tauch, A. (2006) CoryneRegNet 3.0—an interactive systems biology platform for the analysis of gene regulatory networks in corynebacteria and *Escherichia coli*. J. Biotechnol. 129, 279–289.

11) Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, P., Contreras-Moreira, B., Segura-Salazar, S., Muniz-Rascado, L., Martinez-Flores, I., Salgado, H., Bonavides-Martinez, C., Abrue-Godger, C., Rodriguez-Penagos, C., Miranda-Rios, J., Morett, E., Merino, E., Huerta, A.M., Trevino-Quintanilla, L. and Collado-Vides, J. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res. 36, D120–D124.

12) Helman, J. and Chamberlin, M. (1988) Structure and function of bacterial sigma factors. Annu. Rev. Biochem. 57, 859–872.

13) Gross, C.A., Lonetto, M. and Losick, R. (1992) *Bacterial sigma factors*. In Transcriptional Regulation (eds. McKnight, S.L. and Yamamoto, K.R.). Cold Spring Harbour Press, New York, pp. 129–176.

14) Ishihama, A. (2000) Functional modulation of *Escherichia coli* RNA polymerase. Annu. Rev. Microbiol. 54, 499–518.

15) Ishihama, A. (2010) Prokaryotic genome regulation: Multi-factor promoters, multi-target regulators and hierarchic networks. FEMS Microbiol. Rev. 34, 628–645.

16) Ishihama, A. (2012) Transcription factors and transcriptional apparatus in bacteria. In Encyclopedia of Systems Biology (eds. Dubitzky, W., Wolkenhauer, O., Cho, K.-H. and Yokota, H.). Springer (in press).

17) Ishihama, A. (2009) The nucleoid: an overview. In EcoSal—*Escherichia coli* and Salmonella: Cellular and Molecular Biology (eds. Böck A., Curtiss III, R., Kaper, J.B., Karp, P.D., Neidhardt, F.C., Nystrom, T., Slauch, J.M., Squires, C.L., USSery, D.). ASM Press, Washington.

18) Ishihama, A. (1997) Adaptation of gene expression in stationary phase bacteria. Curr. Opin. Gen. 7, 582–588.

19) Ishihama, A. (1999) Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. Genes Cells 4, 130–143.

20) Gruber, T.M. and Gross, C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. 57, 441–466.

21) Paget, M.S. and Helman, J.D. (2003) The sigma 70 family of sigma factors. Genome Biol. 4, 203.

22) Gourse, R.L., Ross, W. and Rutherford, S.T. (2006) General pathway for turning on promoters transcribed by RNA polymerase containing alternative sigma subunits. Mol. Microbiol. 63, 1290–1306.
23) Typas, A., Becker, G. and Hengge, R. (2007) The molecular basis of selective promoter activation by the sigma S subunit of RNA polymerase. Mol. Microbiol. 63, 1296–1306.

24) Yamamoto, K. and Ishihama, A. (2003) Two different modes of transcription repression of the Escherichia coli acetate operon by IclR. Mol. Microbiol. 47, 183–194.

25) Perez-Rueda, E. and Collado-Vides, J. (2000) The repertoires of DNA-binding transcription regulators in Escherichia coli K-12. Nucleic Acids Res. 28, 1838–1847.

26) Babu, M.M. and Teichmann, S.A. (2003) Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res. 31, 1234–1244.

27) Ishihama, A. (1992) Role of the RNA polymerase alpha subunit in transcription activation. Mol. Microbiol. 6, 3283–3288.

28) Ishihama, A. (1993) Protein-protein communication within the transcription apparatus. J. Bacteriol. 175, 2483–2489.

29) Ebright, R.H. (1993) Transcription activation at class I CAP-dependent promoters. Mol. Microbiol. 8, 797–802.

30) Busby, S. and Ebright, R.H. (1999) Transcription activation by catabolite activator protein. J. Mol. Biol. 293, 199–213.

31) Murakami, K., Kimura, M., Owens, J.T., Meares, C.F. and Ishihama, A. (1997) The two alpha subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. Proc. Natl. Acad. Sci. U.S.A. 94, 1709–1714.

32) Owens, J.T., Miyake, R., Murakami, K., Chmura, A.J., Fujita, N., Ishihama, A. and Meares, C.F. (1998) Mapping the sigma70 subunit contact sites on Escherichia coli RNA polymerase with a sigma70-conjugated chemical protease. Proc. Natl. Acad. Sci. U.S.A. 95, 6021–6026.

33) Meares, P.C., Datwyler, S.A., Schmidt, B.D., Owens, J. and Ishihama, A. (2003) Principles and methods of affinity cleavage in studying transcription. Methods Enzymol. 371, 92–106.

34) Bulyk, M.L. (2006) DNA microarray technologies for measuring protein-DNA interactions. Curr. Opin. Biotechnol. 17, 422–430.

35) Khodursky, A.B., Peter, B.J., Cozzarelli, N.R., Botstein, D., Brown, P.O. and Yanofsky, C. (2000) DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 97, 12170–12175.

36) Oh, M.-K., Rohlin, L., Kao, K.C. and Liao, J.C. (2002) Global expression profiling of acetategrown Escherichia coli. J. Biol. Chem. 277, 13173–13183.

37) Kao, K.C., Tran, L.M. and Liao, J.C. (2005) A global regulatory role of gluconeogenic genes in Escherichia coli revealed by transcription network analysis. J. Biol. Chem. 280, 36079–36087.

38) Overton, T.W., Griffiths, L., Patel, M.D., Hobman, J.L., Penn, C.W., Cole, J.A. and Constantinidou, C. (2006) Microarray analysis of gene regulation by oxygen, nitrate, nitrite, FNR, NarL and NarP during anaerobic growth of Escherichia coli: new insights into microbial physiology. Biochem. Soc. Trans. 34, 104–107.

39) Phadtare, S. and Inouye, M. (2004) Genome-wide transcriptional analysis of the cold shock response in wild-type and cold-sensitive, quadruple-csp deletion strains of Escherichia coli. J. Bacteriol. 186, 7007–7014.

40) Zheng, M., Wang, X., Templeton, L.J., Smulski, D.R., LaRossa, R.A. and Storz, G. (2001) DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. J. Bacteriol. 183, 4562–4570.

41) Terni, Y., Higashi, K., Taniguchi, S., Shigemasa, A., Nishimura, K., Yamamoto, K., Kashiwagi, K., Ishihama, A. and Igarashi, K. (2007) Enhancement of the synthesis of RpoN, Cra, and H-NS by polyamines at the level of translation in Escherichia coli cultured with glucose and glutamate. J. Bacteriol. 189, 2359–2368.

42) Lee, L.J., Barrett, J.A. and Poole, R.K. (2005) Genome-wide transcriptional response of chemostat-cultured Escherichia coli to zinc. J. Bacteriol. 187, 1124–1134.

43) Salmon, K., Hsing, S.-P., Mekjian, K., Baldi, P., Hatfield, G.W. and Gunsalus, R.P. (2003) Global gene expression profiling in Escherichia coli K12: the effects of oxygen availability and FNR. J. Biol. Chem. 278, 29837–29855.

44) Constantinidou, C., Hobman, J.L., Griffiths, L., Patel, M.D., Penn, C.W., Cole, J.A. and Overton, T.W. (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as Escherichia coli K12 adapts from aerobic to anaerobic growth. J. Biol. Chem. 281, 4802–4815.

45) Masuda, N. and Church, G.M. (2003) Regulatory network of acid resistance in Escherichia coli. Mol. Microbiol. 48, 699–712.

46) Ren, D., Bedzyk, L.A., Thomas, S.M., Ye, R.W. and Wood, T.K. (2004) Gene expression in Escherichia coli microfilms. Appl. Microbiol. Biotechnol. 64, 515–524.

47) Liu, X. and de Wulf, P. (2004) Probing the ArcA-P modulon of Escherichia coli by whole genome transcriptional analysis and sequence recognition profiling. J. Biol. Chem. 279, 12588–12597.

48) Zheng, D., Constantinidou, C., Hobman, J.L. and Minchin, S.D. (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. Nucleic Acids Res. 32, 5874–5893.

49) Masuda, N. and Church, G.M. (2002) Escherichia coli gene expression responsive to levels of the response regulator EnvA. J. Bacteriol. 184, 6225–6234.

50) Bradley, M.D., Beach, M.B., de Koning, P.J., Pratt, T.S. and Osuna, R. (2007) Effects of Fis on Escherichia coli gene expression during different
growth stages. Microbiology **153**, 2922–2940.
51) Arfin, S.M., Long, A.D., Ito, E.T., Tolleri, L., Richle, M.M., Paegle, E.S. and Hatfield, G.W. (2000) Global gene expression profiling in *Escherichia coli* K12: the effects of integration host factor. J. Biol. Chem. **275**, 29672–29684.

52) Fernandez de Henestriza, A.R., Ogi, T. and Aoyagi, S. (2000) Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. Mol. Microbiol. **35**, 1560–1572.

53) Lehnen, D., Blumer, C., Polen, T., Wackwitz, B., Wendisch, V.F. and Unden, G. (2002) LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in *Escherichia coli*. Mol. Microbiol. **45**, 521–532.

54) Hung, S.-P., Baldi, P. and Hatfield, G.W. (2002) Global gene expression profiling in *Escherichia coli* K12: the effects of leucine-response regulatory protein. J. Biol. Chem. **277**, 40303–40323.

55) Tao, H., Hasona, A., Do, P.M., Ingram, L.O. and Shanmuga, K.T. (2005) Global gene expression analysis revealed an unsuspected deco operon under the control of molybdate sensor, ModE, protein in *Escherichia coli*. Arch. Microbiol. **184**, 225–233.

56) Bodenmüller, D.M. and Spiro, S. (2006) The gieB (nsrR) gene of *Escherichia coli* encodes a nitric oxide-sensitive transcriptional regulator. J. Bacteriol. **188**, 874–881.

57) Rankin, L.D., Bodenmüller, D.M., Partridge, J.D., Nishino, S.F., Spain, J.C. and Spiro, S. (2008) *Escherichia coli* NarR regulates a pathway for the oxidation of 3-nitrotyramine to 4-hydroxy-3-nitrophenylacetate. J. Bacteriol. **190**, 6170–6177.

58) Wei, Y., Lee, J.-M., Smluski, D.R. and Larossa, R.A. (2001) Global impact of *sdiA* amplification revealed by comprehensive gene expression profiling of *Escherichia coli*. J. Bacteriol. **183**, 2265–2272.

59) Martin, R.G. and Rosner, J.L. (2002) Genomics of the marA/sea/rob regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. Mol. Microbiol. **44**, 1611–1624.

60) Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation. Annu. Rev. Biophys. Biomol. Struct. **28**, 29–56.

61) Brent, R. and Ptashne, M. (1981) Mechanism of action of the lexA gene product. Proc. Natl. Acad. Sci. U.S.A. **78**, 1202–1208.

62) Bunschuh, R., Hayot, F. and Jayaprakash, C. (2003) The role of dimerization in noise reduction of simple genetic networks. J. Theor. Biol. **220**, 261–269.

63) Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiuara, M., Wanner, B.L., Mori, H. and Mizuno, T. (2002) Transcriptional analysis of two-component response regulatory system mutants of *Escherichia coli* K-12. Mol. Microbiol. **46**, 281–291.

64) Yamamoto, K., Hirao, K., Oshima, T., Aiba, H., Utsumi, R. and Ishihama, A. (2005) Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*. J. Biol. Chem. **280**, 1440–1456.

65) Kouzarides, T. (2000) Acetylation: a regulatory modification to rival phosphorylation? EMBO J. **19**, 1176–1179.

66) Sterner, D.E. and Berger, S.L. (2000) Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. **64**, 435–459.

67) Yang, X.-J. and Seto, E. (2008) Lysine acetylation: condihed crosstalk with other posttranslational modification. Mol. Cell **31**, 449–461.

68) Spange, S., Wagner, T., Heinzl, T. and Kramer, O.H. (2009) Acetylation of non-histone proteins modulates cellular signaling at multiple levels. Int. J. Biochem. Cell Biol. **41**, 185–198.

69) Thao, S., Chen, C.-S., Zhu, H. and Escalante-Semerena, J.C. (2010) N-Lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity. PLoS ONE **5**, e15123.

70) Lima, B.P., Antelmann, H., Gronau, K., Chi, B.K. and Becker, D. (2011) Involvement of protein acetylation in glucose-induced transcription of a stress-responsible promoter. Mol. Microbiol. **81**, 1190–1204.

71) Jishage, M. and Ishihama, A. (1997) Variation in RNA polymerase sigma subunit synthesis in *Escherichia coli* intracellular levels of four species of sigma subunit under various growth conditions. J. Bacteriol. **178**, 5447–5451.

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

73) Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of DNA molecules that bind specific ligands. Nature **346**, 818–822.

74) Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249**, 505–510.

75) Singer, B.S., Shtatland, T., Brown, D. and Gold, L. (1997) Libraries for genomic SELEX. Nucleic Acids Res. **25**, 781–786.

76) Shimada, T., Fujita, N., Maeda, M. and Ishihama, A. (2005) Systematic search for the Cra-binding promoters using genomic SELEX systems. Genes Cells **10**, 907–918.

77) Hasegawa, A., Ogasawara, H., Kori, A., Teramoto, J. and Ishihama, A. (2008) The transcription regulator AllR senses both allantoin and glyoxylate and controls a set of genes for degradation and utilization of purines. Microbiology **154**, 3366–3378.

78) Ishida, U., Kori, A. and Ishihama, A. (2009) Participation of regulator AesG of the β-glucoside utilization operon in regulation of the propionate catabolism operon. J. Bacteriol. **191**, 6130–6144.

79) Ogasawara, H., Shimohara, S. and Ishihama, A. (2012) Regulation targets of metal-response BasS-
92) Tan, K., Moreno-Hagelsieb, G., Colado-Vides, J. and Storino, G.D. (2001) A comparative genomics approach to prediction of new members of regulons. Genome Res. 11, 566–584.

93) Tan, K., McCue, L.A. and Stormo, G.D. (2007) Making connections between novel transcription factors and their DNA motifs. Genome Res. 15, 312–320.

94) McCue, L., Thompson, W., Carmack, C., Ryan, M.P., Liu, J.S., Derbishire, V. and Lawrence, C.E. (2001) Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. Nucleic Acids Res. 29, 774–782.

95) McCue, L.A., Thompson, W., Carmack, C.S. and Lawrence, C.E. (2002) Factors influencing the identification of transcription factor binding sites by cross-species comparison. Genome Res. 12, 1523–1532.

96) Van Nimwegen, E., Zavolan, M., Rajewsky, N. and Siggia, E.D. (2002) Probabilistic clustering of sequences: Inferring new bacterial regulons by comparative genomics. Proc. Natl. Acad. Sci. U.S.A. 99, 7323–7328.

97) Qin, Z.S., McCue, L.A., Thompson, W., Mayerhofer, L., Mayerhofer, L., Lawrence, C.E. and Liu, J.S. (2003) Identification of co-regulated genes through Bayesian clustering of predicted regulatory binding sites. Nat. Biotechnol. 21, 435–443.

98) Martinez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. Curr. Opin. Microbiol. 6, 482–489.

99) Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wison, C.J., Bell, S.P. and Young, R.A. (2000) Genome-wide location and function of DNA-binding proteins. Science 290, 2306–2309.

100) Lee, T.L., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K. and Young, R.A. (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298, 799–804.

101) Harbison, C.T., Gordon, D.B., Lee, T.L., Rinaldi, N.J., Mac Issac, K.D., Danford, T.W., Hannett, N., Tagne, J.B., Reynolds, D.B., Yoo, J., Hennings, E.G., Zeitlinger, J., Pokholok, D.K., Kellis, M., Rolfe, P.A., Takusagawa, K.T., Lander, E.S., Gifford, D.K., Fraenkel, E. and Young, R.A. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99–104.

102) Herring, C.D., Raffaele, M., Allen, T.E., Kanin, E.L., Landick, R., Ansari, A.Z. and Palsson, B.O. (2005) Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. J. Bacteriol. 187, 6166–6174.

103) Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J. and Bushby, S.J.W. (2005) Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase on the E. coli
chromosome. Proc. Natl. Acad. Sci. U.S.A. 102, 496–510.

104) Grainger, D.C., Hurd, D., Goldberg, M.D. and Busby, S.J.W. (2006) Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. Nucleic Acids Res. 34, 4642–4650.

105) Cho, B.K., Knight, E.M., Barrett, C.L. and Passon, B.O. (2006) Genome-wide analysis of Fis-binding in Escherichia coli indicates a causative role for A-/AT-tracts. Genome Res. 18, 900–910.

106) Shimada, T., Ishihama, A., Busby, S.J. and Grainer, D.C. (2008) The Escherichia coli RutR transcription factor binds at targets within genes as well as intergenic regions. Nucleic Acids Res. 36, 3950–3955.

107) Cho, B.K., Barrett, C.L., Knight, E.M., Park, Y.S. and Palsson, B.O. (2008) Genome-scale reconstruction of the Lrp regulatory network in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 105, 19462–19467.

108) Orlando, V., Strutt, H. and Paro, R.O. (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods 11, 205–214.

109) Huyen, N.T.T., Eiamphungporn, W., Maeder, U., Liebeke, M., Lalk, M., Hecker, M., Helmann, J.D. and Antelmann, H. (2009) Genome-wide responses to carbonyl electrophiles in Bacillus subtilis: control of the thiol-dependent formaldehyde dehydrogenase AdhA and cysteine proteinase YraA by the MerR-family regulator YraB (AdhR). Mol. Microbiol. 71, 876–894.

110) Collado-Vides, J., Magasanik, B. and Gralla, J.D. (1991) Control site location and transcriptional regulation in Escherichia coli. Microbiol. Rev. 55, 371–394.

111) Shimada, T., Makinoshima, H., Ogawa, Y., Miki, T., Maeda, M. and Ishihama, A. (2004) Classification and strength measurement of stationary-phase promoters by use of a newly developed promoter cloning vector. J. Bacteriol. 186, 7112–7122.

112) Lobell, R.B. and Schleif, R.F. (1991) AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. J. Mol. Biol. 218, 45–54.

113) Barnard, A., Wolfe, A. and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organization to produce different regulatory outcomes. Curr. Opin. Microbiol. 7, 102–108.

114) Teichmann, S.A. and Babu, M.M. (2002) Conservation of gene co-regulation in prokaryotes and eukaryotes. Trends Biotechnol. 20, 407–410.

115) Harari, O., del Val, C., Romeo-Zaliz, R., Shin, D., Hung, H., Groisman, E.A. and Zvir, I. (2008) Identifying promoter features of co-regulated genes with similar network motifs. BMC Bioinform. 10 (Suppl. 4), S1.

116) Reznikoff, W.S. (1992) The lactose operon—controlling elements: a complex paradigm. Mol. Microbiol. 6, 2419–2422.

117) Wei, G.H., Liu, D.P. and Lian, C.C. (2004) Charting gene regulatory networks: strategies, challenges and perspectives. Biochem. J. 381, 1–12.

118) Gottesman, S. (1984) Bacterial regulation: global regulatory networks. Annu. Rev. Genet. 18, 415–441.

119) Gutierrez-Rois, R.M., Rosenblueth, D.A., Loza, J.A., Huerta, A.M., Glasner, J.D., Blattner, F.R. and Collado-Vides, J. (2003) Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles. Genome Res. 13, 2435–2443.

120) Zubay, G., Schwartz, D. and Beckwith, J. (1970) Mechanism of activation of catabolite-sensitive genes: a positive control system. Proc. Natl. Acad. Sci. U.S.A. 66, 104–110.

121) Harman, J.G. (2001) Allosteric regulation of the cAMP receptor protein. Biochim. Biophys. Acta 1547, 1–17.

122) Kolb, A., Busby, S., Buc, H., Garges, S. and Adhya, S. (1993) Transcriptional regulation by cAMP receptor protein of Escherichia coli. Annu. Rev. Biochem. 62, 749–795.

123) Krueger, S., Gregurick, S., Shi, Y., Wang, S., Wladowksi, B.D. and Schwarz, F.P. (2003) Entropic nature of the interaction between promoter and bound CRP mutants and RNA polymerase. Biochemistry 42, 1958–1968.

124) Kim, J., Adhya, S. and Garges, S. (1992) Allosteric changes in the cAMP receptor protein of Escherichia coli hinge reorientation. Proc. Natl. Acad. Sci. U.S.A. 89, 1770–1773.

125) Passner, J.M. and Steitz, T.A. (1997) The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. Proc. Natl. Acad. Sci. U.S.A. 94, 2843–2847.

126) Berg, O.G. and von Hippel, P.H. (1988) Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. J. Mol. Biol. 200, 709–723.

127) Parkinson, G., Wilson, C., Gunasekera, A., Ebrigt, Y.W., Ebrigt, R.E. and Berman, H.M. (1996) Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface. J. Mol. Biol. 304, 847–859.

128) Pyles, E.A. and Lee, J.C. (1998) Escherichia coli cAMP receptor protein-DNA complexes. 2. Structural asymmetry of DNA bending. Biochemistry 37, 5201–5210.

129) Lin, S.H. and Lee, J.C. (2003) Determinants of DNA bending in the DNA-cyclic AMP receptor protein complexes in Escherichia coli. Biochemistry 42, 4809–4818.

130) Geerse, R.H., van der Pluijm, J. and Postma, P.W. (1989) The repressor of the PEP-fructose phosphotransferase system is required for the transcription of the pps gene of Escherichia coli. Mol. Gen. Genet. 218, 348–352.

131) Ramseier, T.M. (1996) Cra and the control of carbon flux via metabolic pathways. Res. Microbiol. 147, 489–493.
132) Saier, M.H. and Ramsheier, T.M. (1996) The catabolite repressor/activator (Cra) protein of enteric bacteria. J. Bacteriol. 176, 3411–3417.

133) Negre, D., Bonod-Bidau, C., Gouyon, G., Delage, G., Cozzzone, A.J. and Cortay, J.C. (1996) Definition of a consensus DNA-binding site for the Escherichia coli pleiotropic regulatory protein FruR. Mol. Microbiol. 21, 257–266.

134) Loh, K.D., Gyaneshwar, P., Papadimitriou, E.M., Fong, R., Kim, K.S., Pareles, R., Zhou, Z., Inwood, W. and Kustu, S. (2006) A previously undescribed pathway for pyrimidine catabolism. Proc. Natl. Acad. Sci. U.S.A. 103, 5114–5119.

135) Ma, Z., Gong, S., Richard, H., Tucker, D.L., Conway, T. and Foster, J.W. (2002) Collaborative regulation of Escherichia coli glutamate-dependent acid resistance by two AraC like regulators, GdcA and GdcW. J. Bacteriol. 184, 7001–7012.

136) Hertzberg, K.M., Gemmill, R., Jones, J. and Calvo, J.M. (1980) Cloning of an EcoR1-generated fragment of the leucine operon of Salmonella typhimurium. Gene 8, 810–814.

137) Calvo, J.M. and Matthews, R.G. (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in Escherichia coli. Microbiol. Rev. 58, 466–490.

138) Newman, E.B. and Lin, R. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in E. coli. Annu. Rev. Microbiol. 49, 747–775.

139) Spiro, S. and Guest, J.R. (1990) FNR and its role in oxygen-regulated gene expression in Escherichia coli. FEMS Microbiol. Rev. 6, 399–428.

140) Iuchi, S. and Lin, E.C. (1993) Adaptation of Escherichia coli to redox environments by gene expression. Mol. Microbiol. 9, 9–15.

141) Oshima, T. and Biville, F. (2006) Functional identification of yyiP as a positive regulator of the tdtR-ttdB-yyiE operon. Microbiology 152, 2129–2135.

142) Nash, H.A. and Robertson, C.A. (1981) Purification and properties of the Escherichia coli protein factor required for lambda integrative replication. J. Biol. Chem. 256, 9246–9253.

143) Azam, T.A., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J. Bacteriol. 181, 6361–6370.

144) Goodrich, J.A., Schwartz, M.L. and McClure, W.R. (1990) Searching for and predicting the activity of sites for DNA binding proteins; compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). Nucleic Acids Res. 18, 4993–5000.

145) Swinger, K.K. and Rice, P.A. (2007) Structure-based analysis of HU-DNA binding. J. Mol. Biol. 365, 1005–1016.

146) Sugimura, S. and Crothers, D.M. (2006) Stepwise binding and bending of DNA by Escherichia coli integration host factor. Proc. Natl. Acad. Sci. U.S.A. 103, 18510–18514.

147) Ball, C.A., Osuna, R., Ferguson, K.C. and Johnson, P.C. (1992) Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli. J. Bacteriol. 174, 8042–8056.

148) Almiron, M., Link, A.J., Furlog, D. and Kolter, D. (1992) A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. Genes Dev. 6, 2624–2654.

149) Travers, A., Schneider, R. and Muskheli-Shvili, G. (2001) DNA supercoiling and transcription in Escherichia coli: The Fis connection. Biochimie 83, 213–217.

150) Walker, K.A., Mallik, P., Pratt, T.S. and Osuna, R. (2004) The Escherichia coli Fis promoter is regulated by changes in the levels of its transcription initiation nucleotide CTP. J. Biol. Chem. 279, 50818–50828.

151) Mallik, P., Paul, B.J., Rutherford, S.T., Gourse, R.L. and Osuna, R. (2006) DksA is required for growth phase-dependent regulation, growth rate-dependent control, and stringent control of fis expression in Escherichia coli. J. Bacteriol. 188, 5775–5782.

152) Nair, S. and Finkel, S.E. (2004) Dps protects cells against multiple stressed during stationary phase. J. Bacteriol. 186, 4192–4198.

153) Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Toboadá, B., Jimenez, V., Salgado, H., Juarez, K., Contreras-Moreira, B., Huerta, A.M., Collado-Vides, J. and Morett, E. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. PLoS ONE 4, e7526.

154) Saldago, H., Gama-Castro, S., Martinez-Antonio, A., Diaz-Peredo, E., Sanchez-Solano, F., Perez-Rueda, E., Bonavides-Martinez, C. and Collado-Vides, J. (2004) RegulonDB 4.0: transcriptional regulation, operon organization and growth conditions in Escherichia coli. Nucleic Acids Res. 32, D303–D306.

155) Ogasaawara, H., Yamada, K., Kori, A., Yamamoto, K. and Ishihama, A. (2010) The E. coli csgD promoter: Interplay between eight transcription factors. Microbiology 156, 2470–2483.

156) Ogasaawara, H., Yamamoto, K. and Ishihama, A. (2010) Regulatory role of MlrA in transcription activation of csgD, the master regulator of biofilm formation in Escherichia coli. FEMS Microbiol. Lett. 312, 160–168.

157) Sauer, K., Camper, A.K., Ehrich, G.D., Costerton, J.W. and Davies, D.G. (2002) Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. J. Bacteriol. 184, 1140–1154.

158) Mah, T.F. and O’Toole, G.A. (2001) Mechanisms of biofilm resistance to antimicrobial agents. Trends Microbiol. 9, 34–39.

159) Ogasaawara, H., Yamamoto, K. and Ishihama, A.
Profile

Akira Ishihama was born in 1938 and started his research career in 1960 with studies on bacterial gene transcription at Nagoya University, Institute of Molecular Biology. While he stayed as a postdoctoral research associate at Albert Einstein College of Medicine, New York, from 1967 to 1969, he identified the subunit composition of DNA-dependent RNA polymerase or the transcriptase from Escherichia coli, the model prokaryote. After returning to Kyoto University, Virus Research Institute in 1970, he succeeded the reconstitution in vitro of RNA polymerase from isolated individual subunits, and then identified the subunit assembly sequence in vitro and in vivo. He also determined the intracellular concentration of RNA polymerase, which is maintained at a constant level under the autogenous regulation system. In 1984, he moved to the National Institute of Genetics as Professor and Head of Department of Molecular Genetics, and from 1994, he served as School of Genetics Professor of the Graduate University for Advanced Studies. During this period, his research subject shifted to the functional modulation of RNA polymerase through molecular interaction with two groups of regulatory protein, sigma factors and transcription factors. He identified the set of promoters recognized by each of seven species of the RNA polymerase sigma factor. He also determined the intracellular concentration of each sigma subunit under various growth conditions. One of his marked contributions in this period is the finding of transcription regulation through of direct protein-protein interaction between transcription factors and RNA polymerase subunits. After the complete sequence of E. coli genome was established, he initiated the project of identification of the regulation targets for all 300 species of the transcription factor from E. coli. In 2004, he was invited from Hosei University to set up the Department of Frontier Bioscience and then devoted himself as the Department Head to construct the Faculty of Applied Chemistry and Bioscience. The ultimate purpose of his current research is to reveal the regulatory roles of all transcription factors from a single organism.