Title

The architecture of ArgR-DNA complexes at the genome-scale in Escherichia coli

Permalink

https://escholarship.org/uc/item/67j0z94r

Journal

Nucleic Acids Research, 43(6)

ISSN

0305-1048

Authors

Cho, Suhyung
Cho, Yoo-Bok
Kang, Taek Jin
et al.

Publication Date

2015-03-31

DOI

10.1093/nar/gkv150

Peer reviewed

The architecture of ArgR-DNA complexes at the genome-scale in *Escherichia coli*

Suhyung Cho[1,2], Yoo-Bok Cho[1], Taek Jin Kang[3], Sun Chang Kim[1,2], Bernhard Palsson[4,5,6] and Byung-Kwan Cho[1,2,*]

[1]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea, [2]KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea, [3]Department of Chemical and Biochemical Engineering, Dongguk University-Seoul, Seoul 100-715, Republic of Korea, [4]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA, [5]Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA and [6]Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark

ABSTRACT

DNA-binding motifs that are recognized by transcription factors (TFs) have been well studied; however, challenges remain in determining the *in vivo* architecture of TF-DNA complexes on a genome-scale. Here, we determined the *in vivo* architecture of *Escherichia coli* arginine repressor (ArgR)-DNA complexes using high-throughput sequencing of exonuclease-treated chromatin-immunoprecipitated DNA (ChIP-exo). The ChIP-exo has a unique peak-pair pattern indicating 5′ and 3′ ends of ArgR-binding region. We identified 62 ArgR-binding loci, which were classified into three groups, comprising single, double and triple peak-pairs. Each peak-pair has a unique 93 base pair (bp)-long (±2 bp) ArgR-binding sequence containing two ARG boxes (39 bp) and residual sequences. Moreover, the three ArgR-binding modes defined by the position of the two ARG boxes indicate that DNA bends centered between the pair of ARG boxes facilitate the non-specific contacts between ArgR subunits and the residual sequences. Additionally, our approach may also reveal other fundamental structural features of TF-DNA interactions that have implications for studying genome-scale transcriptional regulatory networks.

INTRODUCTION

Transcription factors (TFs) are ubiquitous regulatory proteins found across all domains of life that determine gene expression by controlling the distribution of RNA polymerase (RNAP) molecules on promoter sites (1). TFs recognize and bind to specific DNA sequences in response to various environmental conditions and govern transcriptional activation or repression of the genes via promoter-associated RNAP (2). Therefore, the determination of TF-binding site (TFBS) with consensus DNA sequence motif is critical to understand the regulatory mechanism and role of TFs in transcription (3). In bacterial genomes, the TF-binding consensus sequences are generally between 12 and 30 base pairs (bp) in length, and are often structured as direct repeats or palindromes spaced with a fixed number of random nucleotides (4,5).

Furthermore, the location of the TFBS determines whether the TFs interfere with or support the association of RNAP to a particular promoter. For example, TFBS in the vicinity of the core promoter elements, the start of the coding region, or the activator-binding site can inhibit transcription by preventing the access of RNAP to those genomic regions (3). Interestingly, TFs often exert regulatory functions such as transcriptional activation and repression even at distal locations by causing topological changes in the structures of the genome such as DNA looping or bending (6–8). Among the bacterial TFs, cAMP receptor protein (CRP) and arginine repressor (ArgR) are particularly interesting from a DNA structure point of view. CRP bends the DNA by at least 90° at the site of interaction with DNA, thereby contributing to transcriptional regulation. The association of hexameric ArgR complex induces DNA bending with the angle of ~70−90° apparently centered at its binding motif (9–11). Genome-scale studies for mapping of TFBS have been performed using chromatin immunoprecipitation (IP) coupled with microarray (ChIP-chip) or sequencing (ChIP-seq) for various bacterial TFs (7,12–18). These studies, however, have not revealed the broad changes in genome topology and motif recognition mechanism by ArgR *in vivo*.

Here, we describe *in vivo* architecture of how DNA wraps around the hexameric ArgR complex on a genome-scale. The comprehensive determination of ArgR target genes by

analysis of unique peak-pair pattern of ChIP-exo demonstrates that the sharp DNA bending (70–90º) at the TFBS facilitates the non-specific contacts between ArgR subunits and residual sequences of TFBS. This approach provides a foundation to determine direct regulon members and *in vivo* architecture of TFs and DNA complexes to elucidate a mechanistic understanding of transcriptional regulatory networks.

MATERIALS AND METHODS

Bacterial strains and growth

All strains used are *Escherichia coli* K-12 MG1655 and its derivatives. The strain harboring ArgR-8myc was constructed as described previously with the tagging primers, AACGGTTTCACAGTCAAAGACC TGTACGAAGCGATTTTAGAGCTGTTCGACC AGGAGCTTGTCGGATCCAGTCTTCGTGAT and GCAGGGGGTTGAGAGGGATAAGCAACATTTTC CCCGCCGTCAGAAACGACGGGGCAGAGAAATT CCGGGGATCCGTCGACC [19]. A Glycerol stock of the strain was inoculated into 3 ml Luria broth supplemented with 150 µg kanamycin and cultured overnight at 37°C with constant agitation. The cultured cells were inoculated with 1:100 dilution into 50 ml of the fresh M9 medium containing 2 g/l-glucose in either the presence or absence of 1 g/l-arginine and continued to be grown at 37°C until reaching an appropriate cell density (OD600 ≈ 0.5).

ChIP-exo

Cultured cells (50 ml) were cross-linked with 1% formaldehyde at room temperature for 30 min. 2 ml of 2.5 M glycine was added to quench the unused formaldehyde. After washing three times with 50 ml of ice-cold Tris-buffered saline (TBS), the washed cells were resuspended in 0.5 ml of lysis buffer composed of 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1 µg/ml RNaseA, protease inhibitor cocktail and 1 kU Ready-Lyse lysozyme (Epicentre, Madison, WI, USA), and then incubated at 37°C for 30 min [20]. The cells were then treated with 0.5 ml of 2× immunoprecipitation (IP) buffer (100 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 2%(v/v) Triton X-100 and protease inhibitor cocktail), followed by incubation on ice for 30 min. The lysate was sonicated in an ice bath using Sonic Dismembrator Model 500 (four times for 20 s each, output level, 2.5 W). Size distribution of the fragmented DNAs was confirmed using agarose gel electrophoresis (200–400 bp) after removing cell debris by centrifugation. The cross-linked DNA-ArgR complexes in the supernatant were then subjected to IP by adding 10 µl of Anti-myc (9E10) (Santa Cruz, Dallas, TX, USA). For mock-IP control, 2 µg of normal mouse IgG (Santa Cruz) was added into the supernatant in parallel. They were then incubated overnight at 4°C with constant rotation. The cross-linked DNA-protein and antibody complexes were selectively captured by adding 50 µl of Dynabeads Pan Mouse IgG magnetic beads (Invitrogen, Grand Island, NY, USA). Next, DNAs were end-polished using T4 DNA polymerase (NEB, Ipswich,

MA, USA), ligated with the annealed adaptor 1 (5′-Phospho-AACTGCCCCGGGGTTGCTCTTCCGATCT and 5′-OH-AGATCGGAAGAGC-OH), nick-repaired using phi29 polymerase (NEB), and digested with λ exonuclease (NEB) as illustrated in the Supplementary Figure S1 [21]. Then, protein-DNA complexes were reverse-cross-linked by heating at 65°C overnight and proteins were degraded by 8 µg of protease K (Invitrogen). The purified DNAs were denatured at 95°C and extended by P1 primer (5′-OH-GTGACTGGAGTTCAGACGTGTGCTCTTCC GATCT), further ligated with the annealed adaptor 2 (5′-OH-ACACTCTTTCCCTACACGACGCTCTTCCGAT CT and 5′-OH-AGATCGGAAGAGCGTCGTGTAGG GAAAGAGTGTAG). The ligated DNA products were purified using Qiagen polymerase chain reaction (PCR) purification kit and were PCR-amplified by P2 primer (5′-OH-AATGATACGGCGACCACCGAGATCTAC ACTCTTTCCCTACACGACGCTCTTCCGATCT) and P3 primer (5′-OH-CAAGCAGAAGACGGCATACGA GAT<u>NNNNNN</u>GTGACTGGAGTTCAGACGTGT). The degenerate sequence (the underlined 6Ns) in the P3 primer indicates the index sequence for the Illumina next-generation sequencing (Illumina, San Diego, CA, USA). The PCR-amplified DNA products were separated on a 2% agarose gel and the amplicons were excised from the gel and extracted using QIAquick gel purification columns.

Real-time quantitative PCR

To measure the enrichment of the ArgR-binding DNA in chromatin IP samples, real-time quantitative PCR (qPCR) was performed. 1 µl of IP or mock-IP DNA was used with specific primers to the previously identified ArgR binding regions (*gltB* promoter) and non-binding regions (*aroH* gene) [17]. The primer sequences for *gltB* were 5′-AAGCTT GCCATTTGACCTGT and 5′-TCCTTTTCGCATCGGT TAAT, the ones for *aroH* were 5′-TCCTCTCGCCAGAC AAAAAT and 5′-TCAAACTCGTGCAGCGTATC. A reaction mixture of 1 µl of IP of mock-IP DNA, 1 µl of 10 µM primers of each region, 15 µl of SYBR mix (Biorad, Hercules, CA, USA) and 13 µl of ddH2O was prepared on ice. All real-time qPCR reactions were conducted in triplicate. The samples were cycled for 15 s to 94°C, for 30 s to 54°C and for 30 s to 72°C (total 40 cycles) in Thermal Cycler (Biorad). The threshold cycle (C_t) values were calculated automatically by the iCycler iQ optical system software (Bio-Rad). Normalized C_t (ΔC_t) values for each sample were calculated by subtracting the C_t value obtained for the mock-IP DNA from the C_t value for the IP-DNA ($\Delta C_t = C_{t,IP} - C_{t,mock}$).

Next-generation sequencing

Prior to the high-throughput sequencing, the sequencing libraries for ChIP-exo were cloned into TOPO vector (Invitrogen) and several colonies were subjected to Sanger sequencing to confirm the adapter sequences and inserted DNA length of the sequencing library. Then, the sequencing libraries were quantified using Qubit®2.0 fluorometer (Invitrogen) and Experion™ system (Bio-Rad), and se-

quenced using Illumina Miseq® V2 (Supplementary Figure S2).

Read mapping and data processing

All sequencing reads from ChIP-exo experiments were mapped to *E. coli* MG1655 reference genome (NC_000913) using CLC Genomics Workbench5 with the length fraction of 0.9 and the similarity of 0.99 (Supplementary Table S1). To capture target protein binding sites from ChIP-exo data, corresponding genomic position of mapped reads start position (MRSP) was counted and stored for visual inspection using in-house scripts.

Motif searching

The motif search and sequence logo was completed using the BioProspector, MEME Suite ver. 4.9.128, and WebLogo 3.

Raw experimental data

All raw data files can be downloaded from Gene Expression Omnibus through accession number GSE60546.

RESULTS

Immunoprecipitation (IP) of ArgR-DNA complexes

ArgR is a transcription factor involved in arginine biosynthesis and metabolism in *E. coli*. The high concentration of cellular arginine enhances ArgR affinity for specific genomic regions and concurrently modulates the transcription of the related genes. Cellular arginine facilitates the formation of the ArgR hexamer. Consequently, the presence of arginine is essential for ArgR hexamer to bind its binding sites with high affinity for the transcriptional regulation of its regulon members (22). We used the genome-wide ChIP-exo method on the *E. coli* K-12 MG1655 strain harboring myc-tagged ArgR protein to probe the ArgR-binding sites at single nucleotide resolution *in vivo* (17,21). Since ArgR responds to the concentration of exogenous L-arginine, the cells were grown in M9 minimal media either in the presence (+ARG) or absence (−ARG) of the amino acid. Prior to the genome-wide ChIP-exo assay, we first examined the enrichment of ArgR proteins on the promoter of *gltBDF* operon in the IP ArgR-DNA complexes under the experimental conditions (Figure 1a). A cross-linking experiment was performed at mid-log phase, followed by lysis, DNA shearing, and IP using anti-myc antibody and then purification of DNA fragments. Quantitative PCR was performed to confirm the enrichment of ArgR-binding regions in the immunoprecipitated DNA (IP-DNA) samples by using primers that amplified the previously known ArgR-binding region. ArgR negatively regulates the *gltBDF* operon, which encodes one of the two main ammonia assimilation pathways in *E. coli* (23). As a negative control, we examined the level of ArgR enrichment on the promoter region of *aroH*, which is involved in the biosynthesis of aromatic amino acids (24). The occupancy level of ArgR at the promoter region of *gltBDF* operon was ∼60-fold higher

than *aroH* under both +ARG and −ARG growth conditions (Figure 1a). This result is in good agreement with the previous ChIP-chip results (17), demonstrating that ArgR-bound DNA fragments were selectively enriched under the experimental conditions.

Determination of genome-wide ArgR-binding loci using ChIP-exo

The direct analysis of *in vivo* ArgR-binding across the *E. coli* genome, previously described using ChIP-chip experiments, revealed a total of 61 unique ArgR-binding regions. This study demonstrated that integration of the ChIP-chip with transcriptome analysis determines the ArgR regulon along with its transcriptional regulatory network overarching the amino acid metabolism (17). Although a partially conserved 18-bp-long imperfect palindrome sequence was inferred as the consensus ArgR-binding motif from the previous ChIP-chip study, we were unable to elucidate the interaction between ArgR hexamer and the neighboring sequences of the ArgR-binding motif due to the limitation of peak resolution. Therefore, we employed ChIP-exo assay (Supplementary Figure S1), which sequentially performs exonuclease trimming, end polishing, blunt-ended and nick-repairing of the IP-DNA followed by high-throughput sequencing (Figure 1b) (21). To this end, we modified the ChIP-exo method for the Illumina sequencing platforms. The high-quality sequencing reads from the +ARG and −ARG samples were uniquely mapped to the *E. coli* reference genome (NC_000913), separately, resulting in identification of ArgR-binding sites in the genome-wide landscape (Figure 1c). In case of the +ARG sample, ArgR-binding occupancy was increased in the identified binding regions (over 90% loci), in comparison to the –ARG sample (Supplementary Figure S3), which is consistent with the previous ChIP-chip result (17). Overall, the genome-wide ChIP-exo profile exhibits a pattern similar to the ChIP-chip profile; but, we observed ∼100-fold higher signal-to-noise (S/N) ratio with ChIP-exo profile.

The ChIP-exo method enabled the identification of the precise location of the ArgR-binding genomic regions, which are represented by the two peaks (hereafter, referred to as a peak-pair), one from the top strand and the other from the bottom strand (Figure 1b). The additional exonuclease treatment digested the ArgR-bound DNA up to the first nucleotide point of cross-linking between DNA and ArgR in the 5′ to 3′ direction. Thus, these peak-pairs allowed us to identify ArgR-binding locations, which are strand-specific for the interaction between DNA and ArgR. From this data set, a total of 62 unique ArgR-binding locations were identified (Supplementary Table S2).

The ChIP-exo profiles represented complete coverage of the 15 ArgR-binding regions, which had been characterized by *in vitro* DNA-binding experiments and *in vivo* mutational analysis (25). The previous ChIP-chip assays determined a total of 64 ArgR-binding regions, including two divergent promoter regions (17). From the comparative analysis of the ChIP-chip data with the ChIP-exo data, a majority of them (90%) were identified simultaneously; however, a few exceptions were observed, such as *asnT*, *yoeI*, *yqaE*, *plsC*, *atpI* and *phnN* promoters (Figure 1d).
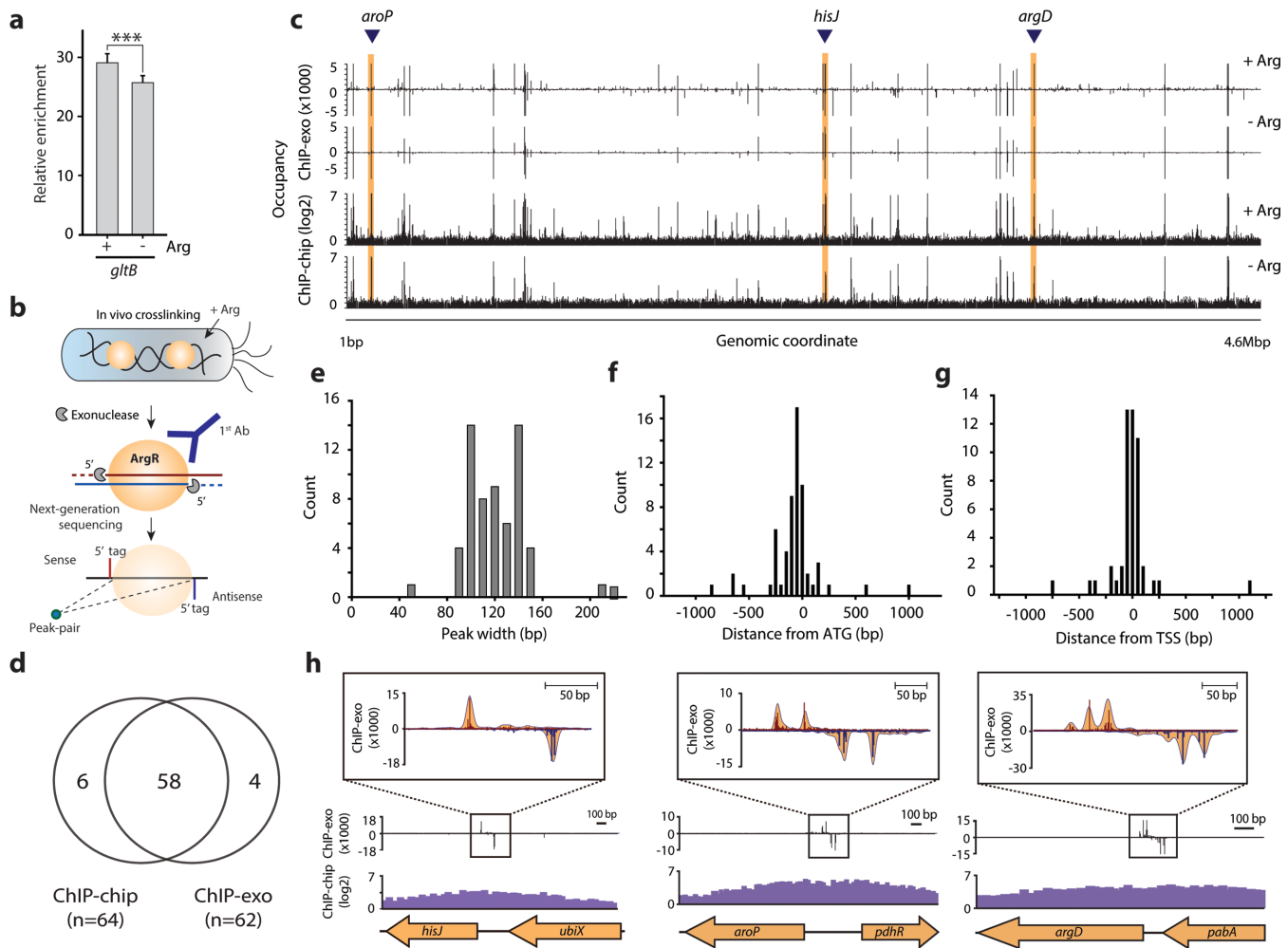
Figure 1. Identification of ArgR-binding regions using ChIP-exo. (**a**) Association of ArgR was measured by qPCR with promoter regions of *gltB* and *aroH* in the presence and absence of arginine. Relative occupancy on *Y*-axis represents the ratio of the IP DNA with specific anti-myc antibody and normal IgG. ***$P < 0.0005$ (two-tailed Student's *t*-test). (**b**) In ChIP-exo, the DNA is crosslinked with ArgR and fragmented followed by the IP by the protein-specific antibody. Subsequently, exonuclease treatment trimmed non-crosslinked DNAs are introduced and then intact DNA regions protected by ArgR are sequenced from 5′ end tag. (**c**) Genome-wide locations of ArgR across the *E. coli* K12 MG1655 genome were identified using ChIP-exo in the presence and absence of arginine, which were compared with ChIP-chip data. (**d**) The ArgR binding TUs identified from ChIP-exo and ChIP-chip data. (**e–g**) The distributions of peak-width between the most forward and reverse peak locations and distances of ArgR-binding locations from start codon (ATG) and transcription start site (TSS). (**h**) Examples of binding patterns at the upstream regions of *aroP*, *hisJ* and *argD* in ChIP-exo are demonstrated with ChIP-chip result.

These exceptions were attributed to low occupancy level (∼1.10) measured by ChIP-chip, which was significantly lower than other regions (∼2.78) (Supplementary Table S2). Thus, exonuclease treatment may eliminate contamination of non-ArgR-bound non-specific DNA fragments with the detection of DNA fragments that are weakly bound by ArgR (21). Additionally, ChIP-exo profiles exhibited four new ArgR-associations from the upstream regions of *proV*, *mltA*, *yhcC* and *ygaW*, which encode a subunit of glycine-betaine/proline ABC transporter, one of six methionine tRNAs, predicted Fe-S oxidoreductase and L-alanine exporter, respectively (Supplementary Table S2). All newly identified ArgR-binding regions were confirmed by electrophoretic mobility shift assays (EMSA) (Supplementary Figures S4 and S5).

The average distance between peaks at the extremities was 116 bp, which indicates a better peak resolution than ChIP-chip analysis (Figure 1e). The high resolution of ArgR-binding location led us to infer its mode of regulation. Based upon the position of 84% and 76% of ArgR-binding peaks found at the upstream sites of translation start codon and within ±100 bp at the vicinity of transcription, ArgR regulates most of the genes in its regulon at the transcriptional level (Figure 1f and g). Taken together, ChIP-exo profiles show low background and enhanced signals, leading to the attainment of bona fide ArgR-binding locations with high resolution.

Analysis of unique ArgR-binding peak-pair pattern

We found that the ArgR-binding signals are often composed of multiple peak-pairs using ChIP-exo analysis. The

presence of such multiple peaks indicates that the interaction between ArgR and the cognate DNA sequence is more complicated than previously thought; that it was based upon the simple DNA binding motif composed of a pair of palindromic sequences (9,11,26). For quantitative analysis of the ChIP-exo profiles, we determined 5′ end positions of mapped reads (MRSP$_{exo}$) at each genomic position. The MRSP$_{exo}$ provides strand-specific first point of cross-linking site between DNA and the ArgR at top and bottom strands, which may directly provide structural information of the complex. For instance, we found single, double and triple peak-pairs from the promoter regions of *hisJ*, *aroP* and *argD*, which are responsible for the ATP-dependent histidine transport, active transport of three aromatic amino acids across *E. coli* inner membrane and amination steps in lysine, ornithine and arginine biosynthesis, respectively (Figure 1h) (27–29).

We sought to analyze the characteristics of the different multiplicities of ArgR at different binding sites. First, to analyze genome-wide multiple peak-pair patterns, the MRSP$_{exo}$ signals of individual ArgR-binding regions were visualized as heatmaps using the values ranging from −150 to +150 bp from the center position. The heatmaps were categorized into three classes of ArgR-binding regions based on the number of peak-pairs (Figure 2a, Supplementary Table S3). From the 63 unique ArgR-binding loci, we identified 21 sites (∼33%) with a single peak-pair. Significant portions of ArgR-binding loci (∼67%) were composed of double (25 sites) and triple peak-pairs (17 sites) (Figure 2b, Supplementary Table S3). MRSP$_{exo}$ at the single peak were enriched between −150 and +150 bp from the center of forward and reverse single peak-pair (F1-R1). Double and triple peak-pairs are composed of F1-R1 and F2-R2; and F1-R1, F2-R2 and F3-R3, respectively (Figure 2c). In cases of double and triple peak-pairs, the signals were enriched from the center of F1-R2 and F1-R3 between −150 and +150 bp, respectively. Thus, the complex interaction between ArgR and the cognate DNA is a genome-wide pattern.

Next, we calculated the distance between forward and reverse peaks from each peak-pair category. Surprisingly, the pitch had a uniform distance of 93 bp (±2) between symmetrically arranged peaks of the peak-pair (F1-R1, F2-R2 and F3-R3), regardless of the number of the peak-pair (Figure 2d). In addition, the distance between each peak-pair was approximately 20 bp (Figure 2e), suggesting that the ArgR binds to the cognate DNA in similar manner (i.e. sequence specific binding) but different conformation according to the number of binding events between ArgR and DNA.

We next examined if the number of peak-pairs show direct correlation at the loci with the ArgR-binding occupancy in the ChIP-chip data (17). Indeed, we observed an increase in occupancy between single, double and triple peak-pairs, whose median values were 1.56, 3.34 and 4.08, respectively, indicating a positive correlation due to the number of cross-linking sites between ArgR protein and DNA sequence (Figure 2f). The ChIP-chip or ChIP-seq signal intensities at the ArgR-binding sites serve as a good indicator of the different binding occupancies of ArgR (30). Furthermore, the multiple peak-pairs are a direct conse-

quence of various topological structures of ArgR-DNA complexes. It was proposed that the association of hexameric ArgR complex induces sharp DNA bend by an angle of ∼70−90° (9–11), which covers a region of approximately four helical turns through only one side of the DNA helix (26,31). Despite *in vitro* experimental evidence supporting such a steric-hindrance model, our results argue that the bending angle and region covered by ArgR complex *in vivo* is variable.

In vivo organization of the ArgR-DNA complexes

The hexameric ArgR complex binds to the specific DNA motif composed of a pair of imperfect palindromic sequences that are connected by a fixed length spacer sequence (2 or 3 bp) (26). To examine if the multiple peak-pairs are the consequence of the presence of multiple ArgR-binding motifs, we inferred a *de novo* position-specific weight matrix (PSWM) for ArgR using MEME, which is a bioinformatics tool that identifies overrepresented motifs in multiple unaligned sequences (32). The DNA motifs were screened from the sequences for peak pairs of the three categories. All peak-pairs contained the 39-bp long ArgR-binding motif comprising two 18 bp palindromic sequences with three nucleotides as a spacer, however the multiple ArgR-binding motifs were not observed in double and triple peak-pairs (Figure 3a). Thus, we speculated that the multiple peak-pairs in our ChIP-exo profiles did not originate due to the interaction between ArgR subunits with the multiple binding motifs. Instead, we hypothesize that the multiple peak-pairs are the consequence of the single binding motif serving as an anchor for the confined non-specific interaction with neighboring sequences by the ArgR subunits. This hypothesis is further supported by the fact that the distance between forward and reverse peak (∼93 bp) is longer than the 39-bp long ArgR-binding motif.

To investigate this hypothesis, we determined the location of the ArgR-binding motif (i.e. two ArgR boxes connected by 3-bp spacer) between each paired peak. A total of 122 individual peak-pairs were identified from the 63 ArgR-binding loci (Figure 3b, Supplementary Table S4). Interestingly, these peak-pairs were classified into three groups based upon the location of the two ARG boxes in the DNA sequence between forward and reverse peak (i.e. left, middle and right position). In the first group (34 peak-pairs), the two ARG boxes are located at 6.7 bp on average from the left end of the DNA sequence. In the second (47 peak-pairs) and third group (41 peak-pairs), the two ARG boxes were located at 26.9 and 47.3 bp from the left end, respectively. The respective distance between the left ends of each group were 20.2 and 20.4 bp. These unique peak-pair patterns suggest that the crosslinking positions detected from ChIP-exo are correlated with the interaction between a multimeric ArgR complex and its binding region. It is known that two monomeric ArgR subunits bind one ARG box. Thus, two ARG boxes of 39-bp in length are occupied by four monomeric ArgR subunits through interaction with only one side of the DNA helix that is equivalent to a region of about four helical turns (31). Note that a hexameric ArgR complex, which is the functionally active form for regulating the target genes, is composed of two trimeric
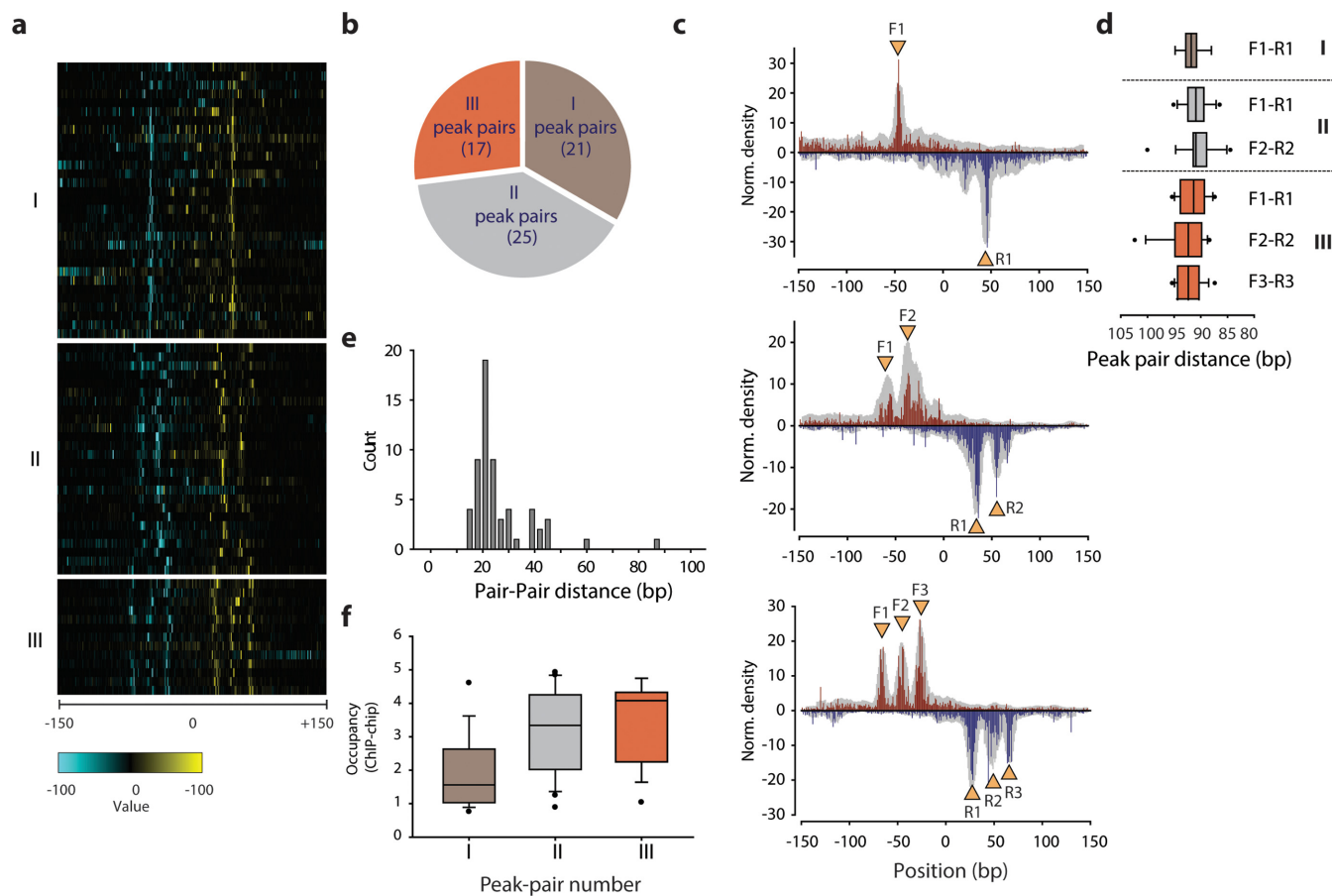
Figure 2. Characterization of ArgR-binding according to ChIP-exo peak-pair. (**a**) The heatmaps indicate the occupancy of ArgR from the center between the most forward and reverse position of peak-pair. (**b**) ArgR-binding regions are categorized into three groups. (**c**) The occupancies of ArgR at ±150 bp from the center of detected forward end (5′ end) and reverse end (3′ end) indicated from the average of values normalized as relative height in each binding point according to the number of peak-pair. Average MRSP calculated by the moving average of 10 bp is expressed as a gray color. (**d**) The distance between forward and reverse peak in a peak-pair of different modes is shown by box-plot. (**e**) The distribution of distances at the positions where multiple peak-pairs are present. (**f**) The peak-pair number is correlated with the occupancies of ArgR in ChIP-chip data.

ArgR complexes depending on the allosteric effect of arginine (33,34). However, our data show a difference in the sequence length of ArgR-binding region (∼39 bp) between *in vitro* experiments and the protected region (∼93 bp) by *in vivo* ChIP-exo experiment.

Thus, we propose three ArgR-binding modes based upon the participation of the remaining two monomeric ArgR subunits in the interaction with the residual DNA region (Figure 3c). For modes α and γ, four monomeric ArgR subunits from the extreme left or right positions bind to the two ARG boxes, and the remaining two monomeric ArgR subunits interact non-specifically with the residual DNA (Figure 3c (α) and (γ)). The interaction between two ARG boxes and four monomeric ArgR subunits, which bends the DNA by an angle of ∼ 70−90° (9–11), may permit the contact of two monomeric ArgR subunits with the residual DNA. For mode β, four monomeric ArgR subunits at the center position hold the ArgR-binding motif by bending DNA. Each ArgR subunit at the extreme left and right positions interacts with the residual DNA sequences non-specifically (Figure 3c (β)), which does not require an additional binding motif or identical length of sequence with the ARG box.

Furthermore, the N-terminal domain of ArgR carries a basic charge that interacts with the negatively charged DNA (35).

To test this hypothesis, we screened the additional motif or a single ARG box from the DNA sequences of nonspecific contact region using the MEME tool. No significant DNA motifs were found from residual sequences of the mode α, β and γ. For example, the upstream region of *hisJQMP* operon containing ARG boxes participates in binding and stabilizing ArgR interaction (36). This site is ∼90 bp positioned away from ARG boxes (37). Thus, the binding of four monomeric ArgR subunits to ARG boxes facilitates DNA-bending that mediates non-specific contacts between ArgR subunits and the ArgR-binding region.

Next, we elucidated the structural difference between single, double and triple peak-pairs. The previous gel-retardation experiments suggested that one ArgR hexamer binds to the two palindromic ARG boxes (31). Consistent with this, our data imply that the ArgR-binding regions can bind to one of the three modes (Figure 3d). Thus, the number of peak-pairs can be determined by the binding accessibility of ArgR to the ARG boxes that results in regulat-
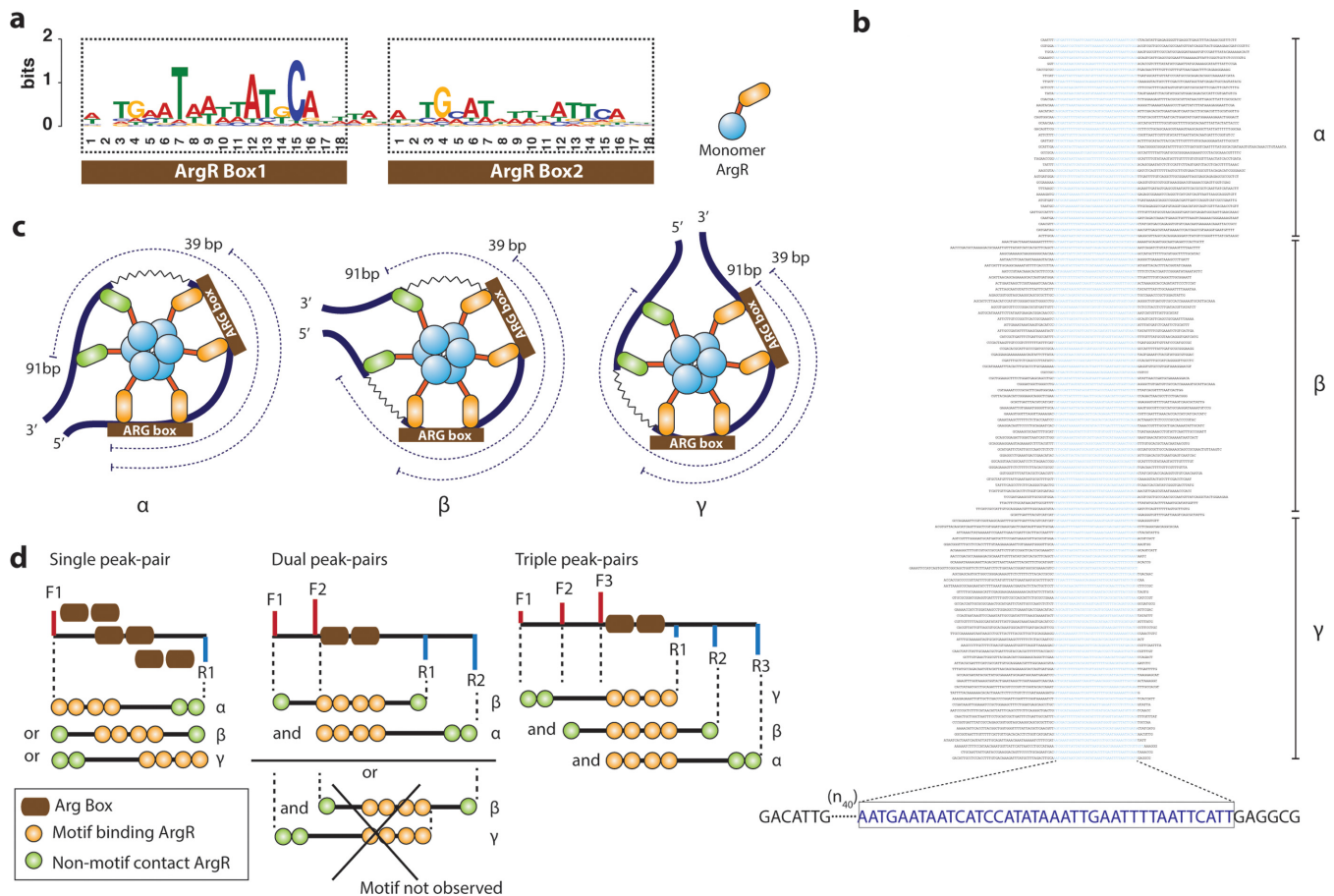
Figure 3. Determination of DNA-wrapping architecture by ArgR. (**a**) ArgR-binding motifs found in the sequences. (**b**) The motif locations are aligned with the extracted sequences between peak-pair. (**c**) Models of DNA bending architecture formed by homo-hexamer ArgR with the location of motifs. The motif and non-motif contact between ArgR monomers and DNA are indicated by orange and green, respectively. (**d**) The potential ArgR binding modes according to the peak-pair numbers are schematically displayed with motif position.

ing the bending angle (∼70–90°). For example, the higher ArgR-binding accessibility can induce the lower bending angle, resulting in a greater chance of non-specific contact for generating the multiple peak-pairs. These diverse binding patterns agree well with the fact that the imperfect ArgR consensus sequences are important for increasing the range of the arginine concentration *in vivo* to regulate genes in a large regulon (38).

Interaction between ArgR and RNA polymerase

In general, the ArgR represses transcription by steric exclusion of RNAP from the promoter regions (26,29,39). To determine this interaction, we compared the ArgR-binding sites with the −10 and −35 promoter elements occupied by RNAP. We classified the interactions between ArgR and RNAP into three unique modes based on their binding locations. For instance, ArgR binds to the promoter region of the *hisJQMP* operon, which is occupied by RNAP for transcriptional initiation (36). 34 genes showed overlap of binding location of ArgR with RNAP, henceforth referred to as the overlapped mode (O) (Figure 4a). In the genes of *aroP* and *yaaU*, which encode an aromatic amino acid permease and an uncharacterized member of the major facilitator

superfamily (MFS) of transporters, the ArgR-binding loci were determined at the upstream (U) and downstream (D) sites from RNAP-binding region, respectively (Figure 4b and c). We determined 11 such genes as having the upstream and downstream modes, respectively (Figure 4d). The relative binding locations of ArgR to the TSS positions (upstream, downstream and overlapped) were not directly correlated with the number of peak-pairs and transcriptional activity (17) (Figure 4d). Altogether, the binding of ArgR does not simply exclude the RNAP for the transcriptional repression, but instead the transcriptional regulation by ArgR is likely mediated by the combinatorial effect of DNA-bending at the ARG boxes, the ArgR-binding positions, the interaction with other TFs, and the number of peak-pairs (23,37).

DISCUSSION

In conclusion, we describe *in vivo* DNA-wrapping modes around the hexameric ArgR complex induced by DNA-bending at the ARG boxes and non-specific contacts on a genome-wide scale. ArgR is a hexameric transcriptional regulator, which controls the transcription of genes involved in arginine biosynthesis, utilization and transport, as well as
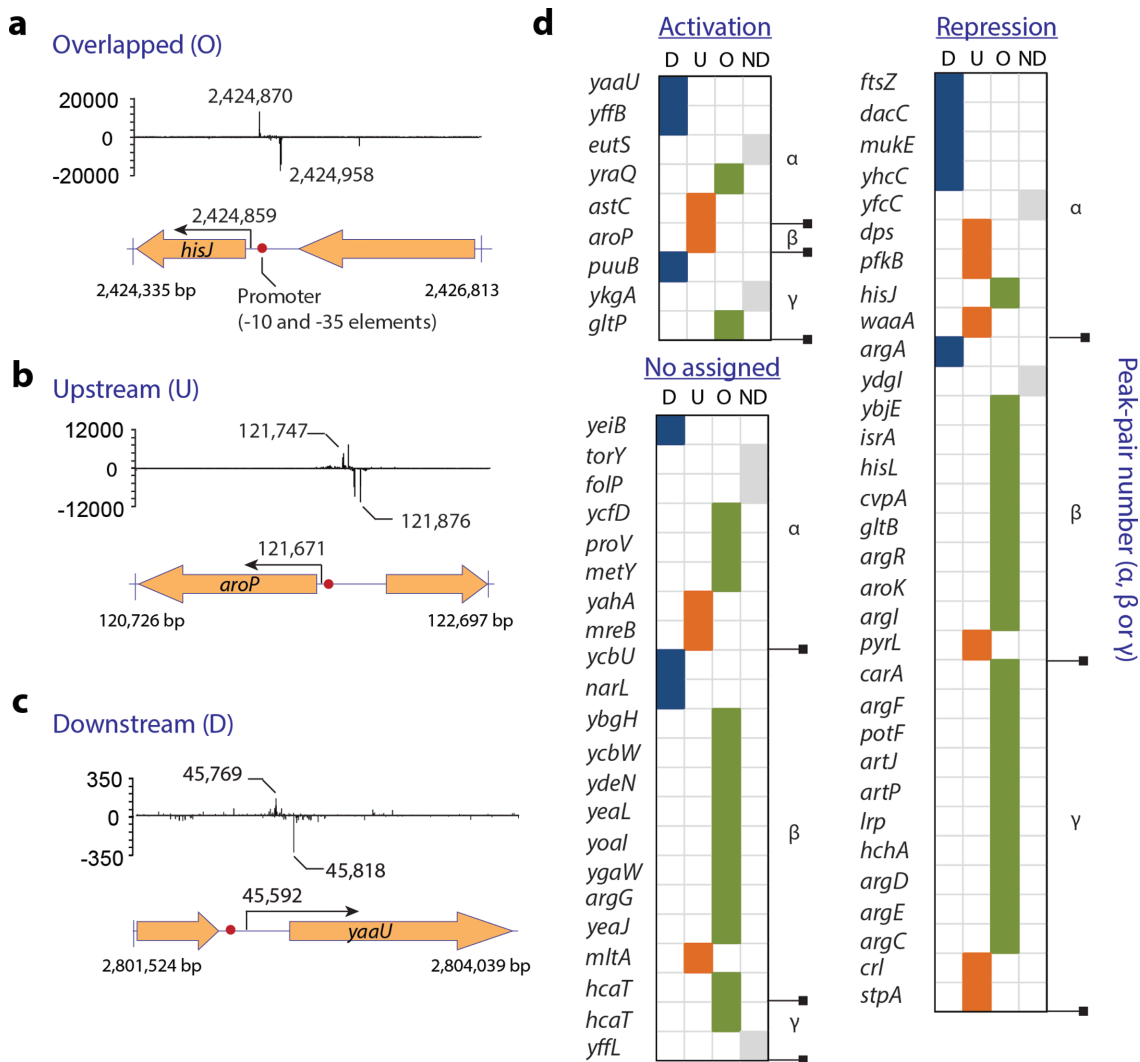
Figure 4. Transcriptional regulation by the position of ArgR relative to the promoters. (**a**) ArgR-binding region at the upstream of *hisJ* overlaps with the promoter occupied by RNAP. (**b**) In the case of the *aroP*, the ArgR is located in the upstream region of the promoter. (**c**) The ArgR binds to the downstream site of the *yaaU* promoter. (**d**) The positions of ArgR from the promoter are categorized according to the gene regulation and peak-pair number. The abbreviations, D, U, O and ND indicate the downstream, upstream, overlapped and non-detected position, respectively. The change of gene expression by arginine addition was obtained from (17).

histidine transport (17,36). In the presence of L-arginine, the hexameric ArgR complex binds to specific DNA sequences called ARG boxes, which consist of a pair of imperfect palindromic sequences. The two palindromes are connected by a fixed-length spacer sequence (2 or 3 bp), resulting in the ArgR-binding site totaling 39 bp in length (26). It has been proposed that the association of hexameric ArgR complex with two ARG boxes bends DNA by an angle of ∼70−90° apparently centered between the pair of palindromes (9–11). Additionally, it was postulated that the hexameric ArgR complex covers a region of about four helical turns through only one side of the DNA helix (26,31). Despite *in vitro* experimental evidence supporting such a steric-hindrance model, the mode of interaction of hexameric ArgR-DNA complex *in vivo* is unclear. Our ChIP-exo data indicated comprehensive ArgR-DNA interactions at high-resolution with successful removal of false positives, resulting in a clearer snapshot of *in vivo* ArgR-binding

events than in a previous study (17). The ArgR-binding data showing the unique DNA sequences (93 ± 2 bp) defined by peak-pairs were classified into three modes comprising multiple peak-pairs (93 bp-long for each peak-pair and 20-bp-long interval between peak-pairs). Moreover, we discovered that 67% of ArgR-binding regions contain multiple peak-pairs where one broad peak was shown in the previous ArgR ChIP-chip data (17). Furthermore, the peak-pairs were grouped into three modes defined by the location of the two ARG boxes (left, middle, right). The sharp DNA bending (70−90°) can be induced by specific interaction between four monomeric ArgR subunits and two ARG boxes. Subsequently, the interaction facilitates non-specific contacts between residual monomeric ArgR subunits and DNA sequences. These findings along with results of RNAP-binding loci suggest that the transcriptional regulation by hexameric ArgR complex is likely mediated by the combinatorial effect of DNA-bending at the ARG boxes, the

ArgR-binding positions, the interaction with other TFs and the non-specific contacts between ArgR and neighboring sequences. ChIP-exo data significantly contributed to elucidating protein-DNA binding mechanisms at the genome-scale through the recognition of accurate protein-binding sites. In the future, this technology will support fundamental information for various transcription factors to understand the bacterial transcription regulatory network.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Ishihama,A. (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Ann. Rev. Microbiol.*, **54**, 499–518.
2. Ishihama,A. (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.*, **34**, 628–645.
3. van Hijum,S.A., Medema,M.H. and Kuipers,O.P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.*, **73**, 481–509.
4. Rodionov,D.A. (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem. Rev.*, **107**, 3467–3497.
5. van Helden,J., Rios,A.F. and Collado-Vides,J. (2000) Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Res.*, **28**, 1808–1818.
6. Semsey,S., Virnik,K. and Adhya,S. (2005) A gamut of loops: meandering DNA. *Trends Biochem. Sci.*, **30**, 334–341.
7. Cho,B.K., Knight,E.M., Barrett,C.L. and Palsson,B.O. (2008) Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res.*, **18**, 900–910.
8. Perez-Martin,J., Rojo,F. and de Lorenzo,V. (1994) Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol. Rev.*, **58**, 268–290.
9. Lu,C.D., Houghton,J.E. and Abdelal,A.T. (1992) Characterization of the arginine repressor from *Salmonella typhimurium* and its interactions with the *carAB* operator. *J. Mol. Biol.*, **225**, 11–24.
10. Thompson,J.F. and Landy,A. (1988) Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res.*, **16**, 9687–9705.
11. Szwajkajzer,D., Dai,L., Fukayama,J.W., Abramczyk,B., Fairman,R. and Carey,J. (2001) Quantitative analysis of DNA binding by the *Escherichia coli* arginine repressor. *J. Mol. Biol.*, **312**, 949–962.
12. Kahramanoglou,C., Seshasayee,A.S., Prieto,A.I., Ibberson,D., Schmidt,S., Zimmermann,J., Benes,V., Fraser,G.M. and Luscombe,N.M. (2011) Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.*, **39**, 2073–2091.
13. Grainger,D.C., Hurd,D., Harrison,M., Holdstock,J. and Busby,S.J. (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 17693–17698.
14. Grainger,D.C., Hurd,D., Goldberg,M.D. and Busby,S.J. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*, **34**, 4642–4652.
15. Singh,S.S., Singh,N., Bonocora,R.P., Fitzgerald,D.M., Wade,J.T. and Grainger,D.C. (2014) Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.*, **28**, 214–219.
16. 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.
17. Cho,B.K., Federowicz,S., Park,Y.S., Zengler,K. and Palsson,B.O. (2012) Deciphering the transcriptional regulatory logic of amino acid metabolism. *Nat. Chem. Biol.*, **8**, 65–71.
18. Cho,B.K., Federowicz,S.A., Embree,M., Park,Y.S., Kim,D. and Palsson,B.O. (2011) The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res.* **39**, 6456–6464.
19. Cho,B.K., Knight,E.M. and Palsson,B.O. (2006) PCR-based tandem epitope tagging system for *Escherichia coli* genome engineering. *Biotechniques*, **40**, 67–72.
20. Cho,B.K., Zengler,K., Qiu,Y., Park,Y.S., Knight,E.M., Barrett,C.L., Gao,Y. and Palsson,B.O. (2009) The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.*, **27**, 1043–1049.
21. Rhee,H.S. and Pugh,B.F. (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, **147**, 1408–1419.
22. Van Duyne,G.D., Ghosh,G., Maas,W.K. and Sigler,P.B. (1996) Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. *J. Mol. Biol.*, **256**, 377–391.
23. Paul,L., Mishra,P.K., Blumenthal,R.M. and Matthews,R.G. (2007) Integration of regulatory signals through involvement of multiple global regulators: control of the *Escherichia coli gltBDF* operon by Lrp, IHF, Crp, and ArgR. *BMC Microbiol.*, **7**, 2.
24. Hudson,G.S., Rellos,P. and Davidson,B.E. (1991) Two promoters control the *aroH* gene of *Escherichia coli*. *Gene*, **102**, 87–91.
25. Keseler,I.M., Mackie,A., Peralta-Gil,M., Santos-Zavaleta,A., Gama-Castro,S., Bonavides-Martinez,C., Fulcher,C., Huerta,A.M., Kothari,A., Krummenacker,M. *et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.*, **41**, D605–D612.
26. Charlier,D., Roovers,M., Van Vliet,F., Boyen,A., Cunin,R., Nakamura,Y., Glansdorff,N. and Pierard,A. (1992) Arginine regulon of *Escherichia coli* K-12. A study of repressor-operator interactions and of *in vitro* binding affinities versus *in vivo* repression. *J. Mol. Biol.*, **226**, 367–386.
27. Cosgriff,A.J. and Pittard,A.J. (1997) A topological model for the general aromatic amino acid permease, AroP, of *Escherichia coli*. *J. Bacteriol.*, **179**, 3317–3323.
28. Ledwidge,R. and Blanchard,J.S. (1999) The dual biosynthetic capability of N-acetylornithine aminotransferase in arginine and lysine biosynthesis. *Biochemistry*, **38**, 3019–3024.
29. Caldara,M., Minh,P.N., Bostoen,S., Massant,J. and Charlier,D. (2007) ArgR-dependent repression of arginine and histidine transport genes in *Escherichia coli* K-12. *J. Mol. Biol.*, **373**, 251–267.
30. Jothi,R., Cuddapah,S., Barski,A., Cui,K. and Zhao,K. (2008) Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.*, **36**, 5221–5231.
31. Tian,G., Lim,D., Carey,J. and Maas,W.K. (1992) Binding of the arginine repressor of *Escherichia coli* K12 to its operator sites. *J. Mol. Biol.*, **226**, 387–397.
32. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
33. Miller,C.M., Baumberg,S. and Stockley,P.G. (1997) Operator interactions by the *Bacillus subtilis* arginine repressor/activator, AhrC: novel positioning and DNA-mediated assembly of a transcriptional activator at catabolic sites. *Mol. Microbiol.*, **26**, 37–48.
34. Holtham,C.A., Jumel,K., Miller,C.M., Harding,S.E., Baumberg,S. and Stockley,P.G. (1999) Probing activation of the prokaryotic arginine transcriptional regulator using chimeric proteins. *J. Mol. Biol.*, **289**, 707–727.
35. Grandori,R., Lavoie,T.A., Pflumm,M., Tian,G., Niersbach,H., Maas,W.K., Fairman,R. and Carey,J. (1995) The DNA-binding domain of the hexameric arginine repressor. *J. Mol. Biol.*, **254**, 150–162.

36. Caldara,M., Charlier,D. and Cunin,R. (2006) The arginine regulon of *Escherichia coli*: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation. *Microbiology*, **152**, 3343–3354.

37. Krin,E., Laurent-Winter,C., Bertin,P.N., Danchin,A. and Kolb,A. (2003) Transcription regulation coupling of the divergent *argG* and *metY* promoters in *Escherichia coli* K-12. *J. Bacteriol.*, **185**, 3139–3146.

38. Sunnerhagen,M., Nilges,M., Otting,G. and Carey,J. (1997) Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. *Nat. Struct. Biol.*, **4**, 819–826.

39. Maas,W.K. (1994) The arginine repressor of *Escherichia coli*. *Microbiol. Rev.*, **58**, 631–640.