DNA Research **13**, 141–153 (2006) doi:10.1093/dnares/dsl009

Escherichia coli Histone-Like Protein H-NS Preferentially Binds to Horizontally Acquired DNA in Association with RNA Polymerase

Taku Oshima,[1,†] Shu Ishikawa,[1,†] Ken Kurokawa,[1,†] Hirofumi Aiba,[2] and Naotake Ogasawara[1,*]

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5, Takayama, Ikoma, Nara 630-0192, Japan[1] and Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi 464-8601, Japan[2]

(Received 27 August 2006; revised 30 August 2006; published online 17 October 2006)

Abstract

Heat-stable nucleoid-structuring protein (H-NS) is one of the main nucleoid proteins expressed in exponentially growing *Escherichia coli* cells. In addition to a role in nucleoid organization, H-NS functions as a pleiotropic regulator of gene expression. The genome-wide distribution of H-NS, compared with the distribution of RNA polymerase and transcriptionally active genes, was investigated using a high-density oligonucleotide chip. The new approach utilized in this study revealed that H-NS binds specifically to approximately 250 loci, covering >1000 genes, to maintain transcriptional inactivation. RNA polymerase was detected in >65% of H-NS binding sites with low or no transcriptional activity, indicating that the association of RNA polymerase to promoter regions is a general mode of transcription repression by H-NS. This study also revealed that most H-NS bound DNA have been horizontally acquired, which indicates that repression of inappropriate gene expression by H-NS plays an important role in the diversification of the *E. coli* genome. This study presents a comprehensive assessment of the distribution of H-NS within the *E. coli* genome, sheds light on the mechanism underlying the transcriptional regulation by H-NS, and provides new insight into bacterial genome evolution.

Key words: histone-like protein H-NS; RNA polymerase; horizontal genes; high-density tiling chip; ChIP-chip analysis

1. Introduction

Genomic DNA in bacterial cells is a platform for the action of proteins involved in DNA transactions and is folded into a compact structure called a nucleoid. DNA supercoiling, macromolecular crowding and nucleoid-associated proteins contribute to nucleoid compaction, although the detailed architecture is not yet fully understood.[1] Among the bacterial nucleoid-associated proteins, a heterogeneous family of basic, low molecular weight, and abundant DNA-binding proteins, often called histone-like proteins, play fundamental roles in nucleoid organization.[2]

The *Escherichia coli* genome encodes at least seven related histone-like proteins. The extensively characterized heat-stable nucleoid-structuring protein (H-NS) is one of the main histone-like proteins expressed at high levels during exponential growth.[3] It is a small (136 amino acids) and abundant (>20 000 copies per cell) protein, consisting of three structural domains. The C-terminal domain of H-NS (residues 90–136) binds DNA in a relatively sequence-independent manner and recognizes intrinsically curved DNA. The N-terminal domain of H-NS (residues 1–64) is a dimerization domain, and the H-NS dimer formed through this domain bridges two DNA segments.[4] H-NS has the capacity to form larger oligomers, possibly through

Communicated by Katsumi Isono

* To whom correspondence should be addressed. Tel. +81-743-72-5430, Fax. +81-743-72-5439, E-mail: nogasawa@bs.naist.jp

† These authors contributed equally to this work.

interactions between the flexible linker domain connecting the dimerization and DNA-binding domains. The lateral interaction of H-NS dimers allows the extension of H-NS binding from high-affinity nucleation sites to flanking low affinity sequences.[5] Recent atomic force microscopy experiments revealed that, after the initial binding of H-NS to DNA at distant sites, the H-NS assembles into patches of oligomers bridging the segments of DNA.[6] The bridging of DNA segments by H-NS plays a role in the compaction of DNA and higher order nucleoid organization. This is supported by the finding that overexpression of H-NS leads to abnormally compacted nucleoids in *E. coli*.[7]

In addition to a role in nucleoid organization, H-NS acts as a pleiotropic regulator of gene expression. Comparison of gene expression profiles of wild type and *hns*-inactivated *E. coli* cells by microarray analysis demonstrates that the expression level of about 200 genes is directly or indirectly altered in the *hns* mutant.[8] Approximately one-third of these genes encode cell envelope components or proteins involved in adaptation to changes in environmental conditions. This suggests that H-NS directly or indirectly regulates adaptation to different environmental conditions. A number of H-NS-responsive promoters contain regions of intrinsic DNA curvature located upstream and/or downstream of the transcription start point, and the lateral binding of H-NS at preferential binding sites that extend over the promoter region is thought to be sufficient for repression of transcription.[5] A number of complex mechanisms underlying the inhibition of initiation or elongation of transcription have been proposed. Bending of the promoter sequence by RNA polymerase induces the bridging of two flanking H-NS bound sequences at the *rrnB* P1 promoter, resulting in the trapping of RNA polymerase in a non-productive state.[6,9] Repression by H-NS at the *bgl* operon occurs at two levels, repression of promoter activity through upstream binding and blocking of transcription elongation by downstream binding.[10] Recent *in vitro* experiments showed that DNA wrapped around RNA polymerase associates with the sigma 70 (σ^D) subunits and induces looping of the *hdeAB* promoter sequence when H-NS is bound to upstream and downstream regions, resulting in blockage of transition from the open to the elongation complex of RNA polymerase.[11] In addition to a role in repression of transcription, H-NS may act as an activator of transcription,[12,13] although there is no direct evidence that H-NS has the capacity to act as a direct activator of transcription. H-NS is also involved in the maintenance of local superhelicity of genome DNA by forming a topological barrier and thereby inhibiting the expression of superhelicity-sensitive promoters.[14]

Thus, extensive studies on *E. coli* H-NS since its first detailed biochemical characterization in 1984[15] have revealed important principles of action in DNA compaction and regulation of gene expression. Our understanding of the action of H-NS is still fragmented and a comprehensive assessment of the action of H-NS in bacterial cells is still lacking. A wide variety of eukaryotic studies have used ChIP-chip analysis, a powerful new tool that employs genome-wide high-resolution DNA tiling chips (oligonucleotide chips) for the fine mapping of transcribed regions in combination with chromatin immunoprecipitation, for monitoring the distribution of genome associated proteins *in vivo*.[16,17] A limited number of similar studies in bacteria have been reported. Very recently, ChIP-chip analyses using different oligonucleotide tiling chip formats have been employed to investigate the genome-wide distribution of RNA polymerase and the transcriptional regulators, LexA, Crp and MelR.[18–21] During preparation of the present report, H-NS distribution on the *Salmonella* genome has appeared.[22]

In this study, we designed a custom Affymetrix oligonucleotide chip consisting of 50 085 25 mer probe sets that match protein coding sequences (>100 amino acids) at 45–55 bp intervals and 82 023 probe sets at 2–3 bp intervals, for both strands of the remaining sequences. ChIP-chip analysis was performed using this custom oligonucleotide chip to map H-NS and RNA polymerase associated sequences in *E. coli* cells grown in LB medium. We also determined transcriptionally active regions in wild type and *hns*-inactivated cells using the same high-density chip. These new approaches revealed that H-NS binds specifically to about 250 loci in the genome, covering >1000 genes in total, to maintain transcriptional inactivation. H-NS binding sites containing intrinsically curved A-tract sequences were identified at a high density, consistent with the preferential binding of H-NS to curved DNA. In >65% of H-NS binding sites, the binding of β and σ^D subunits of RNA polymerase was clearly associated with low or no transcriptional activity, indicating that repression of transcription by H-NS involves the binding of RNA polymerase. Furthermore, our genome-wide analysis of H-NS binding sites revealed the preferential association of H-NS with DNA predicted to be horizontally acquired into the *E. coli* genome. We propose that the main role of H-NS is to repress the inappropriate expression of horizontally acquired genes under normal growth conditions, allowing their stable transmission and diversification of the *E. coli* genome. Thus, the genome-wide approach described in this study represents a comprehensive assessment of the genomic distribution of H-NS, sheds light on the mechanism underlying the transcriptional

regulation by H-NS, and provides new insight into bacterial genome evolution.

2. Materials and Methods

2.1. Array design

The *E. coli* K-12 MG1655 genome sequence and other genetic information was retrieved from NCBI GenBank (http://ncbi.nlm.nih.gov/), accession number U00096, and used for the design of probes according to the Affymetrix guidelines. Probe sequences were selected for protein coding regions (3871 regions) and inter-genic regions (5236 regions) using different criteria. Apart from the *rrnA* operon, rRNA genes were excluded from probe design (16S, 23S, 5S). The tRNA genes and protein coding genes with lengths of <100 amino acids were treated as inter-genic regions. As a result, 50 085 25 mer sequences on the coding strand were selected for protein coding regions at 45–55 bp intervals, and 82 023 sequences on both strands were selected for inter-genic regions at 2–3 bp intervals. Thus, 132 108 probe sets (perfect match and mismatch probes) were synthesized on the chip. Chip probe information is available on our web site (http://genome.naist.jp/bacteria/array/ecol.html).

2.2. Bacterial strains

W3110 was used as a wild-type strain of *E. coli* K-12. The Δ*hns* strain (RM597) is a derivative of W3110.[23] Both strains were provided by Dr. K. Ito (University of Tokyo). The *hns* gene was tagged at the 3′-terminus with the 3 × FLAG sequence using the epitope tagging procedure.[24] The 3 × FLAG sequence followed by the Kanamycin resistant cassette was PCR amplified from the plasmid pSUB11[24] using primers, 5′- TGAGCAAGGTAAATCCCTCGACGATTTCCTGATCAAGCAAGACTACAAAGACCATGACGG -3′ and 5′- GGACAATAAAAAATCCCGCCGCTGGCGGGATTTTAAGCAACATATGAATATCCTCCTTAG -3′, and inserted into the 3′ end of the chromosomal *hns* gene of the *E. coli* strain BW25113(pKD46) using homologous sequences added to the primers. Proper addition of the 3 × FLAG tag was confirmed by PCR amplification using appropriate primers. The *hns*-3 × FLAG allele was transduced to W3110 by P1 transduction. Expression of the 3 × FLAG H-NS protein in W3110 was confirmed by western blotting using the anti-flag antibody (data not shown).

2.3. High-resolution transcriptome analysis

An aliquot (0.25 ml) of an overnight liquid culture of wild-type and Δ*hns* cells in LB medium at 37°C was inoculated into 50 ml of LB medium (pH7.4) in a 300 ml flask. When the cells, grown aerobically by rotationally shaking at 200 r.p.m. at 37°C, reached an OD_{600} of 0.5, 10 ml of culture was mixed with 20 ml of RNA protect (Qiagen) and the cells were collected by centrifugation and stored at −80°C. Total RNA was purified from collected cells using the RNeasy mini kit according to the manufacturer's instruction (Qiagen). Synthesis of cDNA, terminal labeling, and hybridization with the oligonucleotide chip was performed following the Affymetrix instruction manual (Affymetrix). Briefly, cDNA was synthesized from 10 μg of total RNA using random primers and reverse transcriptase (Superscript III; Invitrogen), followed by purification using Qiaquick purification columns (Qiagen) and digestion with DNaseI (GE Healthcare Bioscience). Then cDNA fragments were terminally labeled with biotin-ddUTP using ENZO BioArray Terminal Labeling Kit (Enzo Life Sciences). Hybridization with the oligonucleotide chip was performed for 16 h at 42°C, followed by washing, staining and scanning using the GeneChip Instrument System, according to the manufacturer's instructions (Affymetrix). PCR amplification of the cDNA was avoided to allow the coding strands of each transcript to be distinguished.

To compensate for the differences in hybridization efficiency of each 25 mer probe on the chip, we divided the hybridization intensities of cDNA synthesized from total RNA by those of the total genome DNA. Genomic DNA was purified from the collected cells using the DNeasy mini kit (Qiagen) and 1 μg was digested with DNaseI (GE Healthcare Bioscience) for 15 min at 37°C, followed by end-labeling and hybridization with the oligonucleotide chip as described above.

2.4. ChIP-Chip analysis

Cross-linking and purification of DNA–protein complexes were carried out according to Grainger et al.[19] Immunoprecipitation of the complex, recovery of precipitated DNA fragments, and amplification and end-labeling were performed using the protocol specified for high-resolution DNA tiling chip analysis.[17] The antibodies W0004 and W0023 (NeoClone), against the σ^D and the β RNA polymerase subunits, respectively, and the Pan mouse immunoglobulin beads, #110.22 (Invitrogen), were used for ChIP-chip analysis of RNA polymerase binding sites. The antibody against the FLAG tag (Sigma-Aldrich) and Dynal Dynabeads protein A beads, #100.02 (Invitrogen) were used for H-NS analysis. Pan mouse immunoglobulin beads and protein A beads were mixed with antibodies at 4°C for several hours and washed twice with 0.5 ml of cold TBS containing 5 mg/ml BSA, before use.

E. coli cells were grown to an OD_{600} of 0.5, as described for RNA preparation, then 10 ml of culture was transferred to a tube containing formaldehyde (final concentration of 1%) and incubated at room temperature

for 30 min. The cross-linking reaction was terminated by the addition of 1.5 ml of 3 M glycine, and cells were collected by centrifugation at 3000 r.p.m. for 15 min, followed by washing with TBS and lysis buffer (10 mM Tris–HCl pH 8.0, 20% sucrose, 50 mM NaCl, 10 mM EDTA). Cells were lysed by incubation with 20 mg/ml Lysozyme in lysis buffer at 37°C for 30 min and stored at −20°C until use. Lysed cells were dissolved in 4 ml of IP buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS) and incubated at 37°C for 10 min with PMSF at a final concentration of 1 mg/ml. The sample was then sonicated 10 times for 1 min at 1 min intervals on ice using an Astrason XL2020 sonicator. Cell debris was removed by centrifugation at 15 000 r.p.m. for 30 min. The supernatant fraction (whole cell extract) was mixed with the magnetic beads bound to the antibody and incubated at 4°C overnight. The magnetic beads were washed twice with 750 μl then 500 μl of IP buffer while rotating at 4°C for 5 min each. The beads were further washed with 500 μl each of IP salt buffer (IP buffer containing 500 mM NaCl), IP wash buffer (10 mM Tris–HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet P-40, 0.5% Sodium deoxycolate) then finally in TE buffer. The DNA–protein complex bound to the beads was recovered by eluting with 100 μl of elution buffer (50 mM Tris–HCl pH 7.5, 10 mM EDTA, 1% SDS). Protein was removed by a 20 min incubation at 65°C, followed by an incubation at 42°C for 2 h then 65°C for 6 h with 5 mg/ml proteinase K. DNA fragments free of DNA–protein cross-links were purified using the Qiagen PCR clean up kit (Qiagen) and eluted with 100 μl of Buffer EB supplied in the kit. Recovered DNA fragments were amplified according to the random DNA amplification method described by Katou et al.[17] using the primers, PF 43 and PF 44.[20] The PCR reaction was performed using 28 cycles for the H-NS analysis and 34 cycles for the RNA polymerase subunit analysis, using Phusion high-fidelity DNA polymerase (Bio-Rad). Amplified DNA fragments were terminally labeled and hybridized with the oligonucleotide chip as described for the transcript mapping.

To compensate for positional differences in the efficiency of DNA recovery, most likely due to differences in the complexity and strength of local DNA–protein interactions, we normalized the hybridization intensities of DNA recovered from the immuno-precipitated complexes by dividing the values with those determined from DNA isolated from the whole cell extract of the control strain, W3110. Following cross-linking and sonication, a 100 μl aliquot of supernatant from the W3110 cells was incubated at 65°C for 20 min for de-crosslinking, followed by phenol–chloroform treatment and ethanol precipitation. PCR amplification for 28 cycles followed by terminal labeling was performed using a 10 ng sample of DNA.

2.5. Data analysis

We developed a program to extract and manipulate raw intensity data (.CEL files), and to display results along the genome coordinate of strain MG1655. The genome sequence of strain W3110 used in this study has only 8 bp differences, 12 IS insertions and 1 phage deletion when compared with the MG1655 strain, although the sequence between *rrnD* and *rrnE* is inverted.[25]

We subtracted signal intensities of mismatch probes from those of perfect match probes for use as signal intensities.[26] Probes with a negative value for signal intensity were excluded from further analysis. For transcript mapping, cDNA signal intensities were divided by the signal intensities of genome DNA, to normalize the differences in hybridization efficiency of probes on the chip (transcriptional activity signals).[27] For the ChIP-chip analysis, the signal intensities of DNA in the immuno-precipitated fraction were divided by those determined for DNA isolated from the whole cell extract fraction before precipitation (binding signals, enrichment in the chromatin immuno-precipitated fraction).[17] The distribution of transcriptional activity and protein binding signals along the genome coordinate was visualized and transcriptionally active regions, H-NS binding sites and RNA polymerase binding sites were assessed manually. All experiments were carried out more than two times, and average signal intensities were used for the analysis.

2.6. Factorial correspondence analysis of codon usage

The codon usage pattern has been determined for all protein coding genes in the *E. coli* genome.[28] Principal component analysis (PCA) was performed against the frequency of codon usage of all genes and the clustering results were projected onto a two-dimensional field.

3. Results

3.1. Genome-wide ChIP-chip analysis of H-NS binding regions

We performed ChIP-chip analysis using 3 × FLAG-tagged H-NS sequences to construct a map of H-NS binding sites on the whole *E. coli* genome as described in Materials and Methods. A strain of *E. coli,* with the native *hns* gene in-frame with the 3 × FLAG sequence at the 3′ terminus, was constructed by epitope tagging[24] and used to isolate DNA fragments associated with H-NS *in vivo*. The addition of the tag to H-NS had no significant effects on the growth of *E. coli* cells and repression of transcription of the *lysU* gene by H-NS[23] (data not shown). We developed a program to manipulate raw hybridization data and to display final results

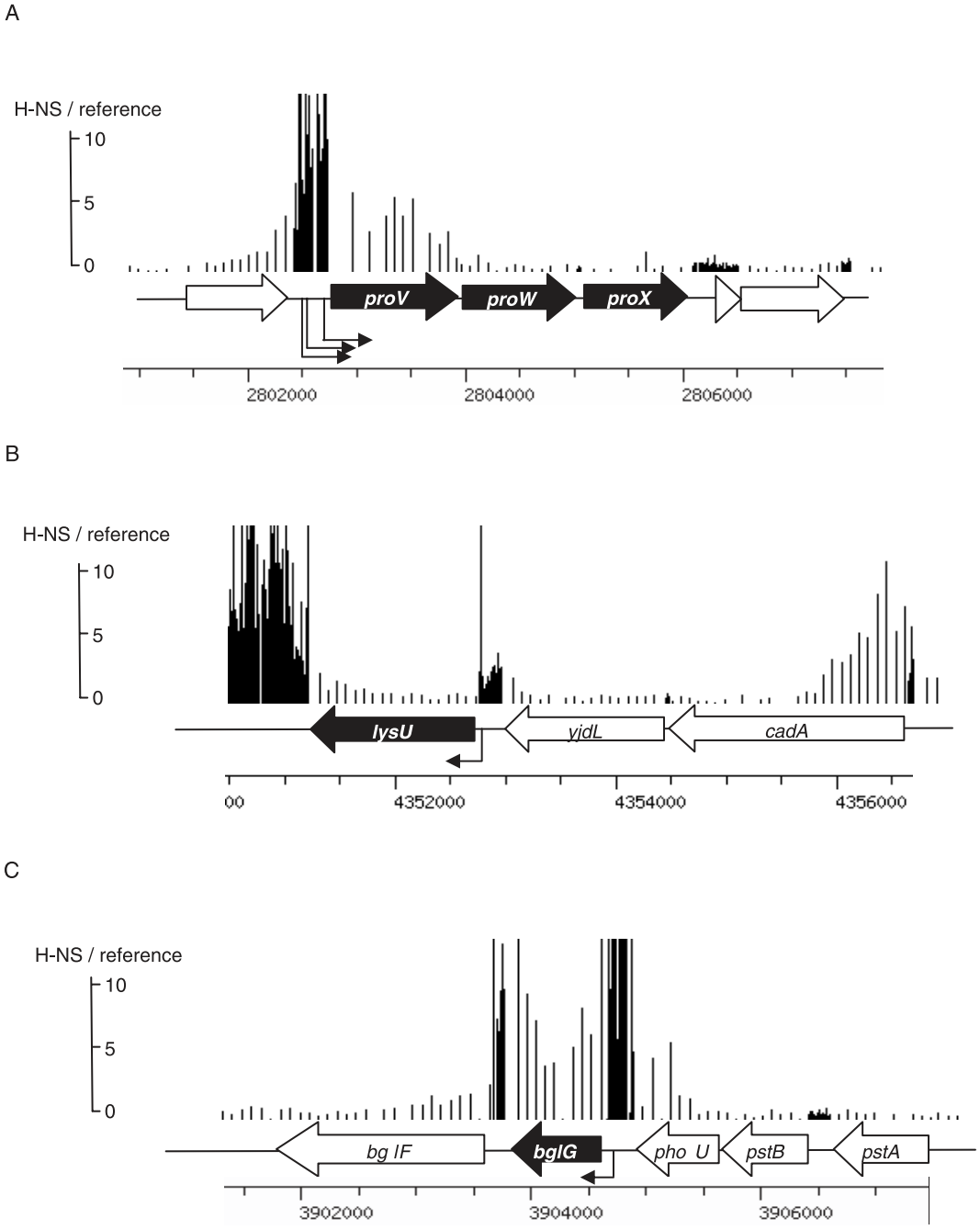

Figure 1. H-NS binding profiles around genes regulated by H-NS. H-NS binding signals around the *proVWX* operon (A), the *lysU* gene (B), and the *bglG* gene (C). Vertical bars indicate the relative hybridization intensity of DNA from immuno-precipitated fraction compared with the DNA from whole cell extracts (enrichment in the chromatin immuno-precipitated fraction). The horizontal line at the bottom indicates the chromosomal coordinates in bp. The arrangement of genes (thick arrows) and known promoters (thin arrows) are shown schematically above the coordinate. The relative hybridization intensities for each 25 mer probe on the chip are shown at their corresponding coordinates.

along the *E. coli* genome coordinate, and H-NS binding sites were assessed manually. As the results, H-NS was found to specifically bind to about 250 loci along the *E. coli* genome, covering >1000 genes (Supplementary Table S1 and Figure S1; see also the authors' web site: http://genome.naist.jp/bacteria/array/ecol.html). Previously reported H-NS binding sites for *proV*, *lysU* and *bglG* genes were clearly detected among these loci (Fig. 1A–C).[10,23,29–32] In 15 loci, H-NS covers a wide region >10 kb, including several well-known H-NS targets (see Fig. 2A).[11,33] As H-NS binds to intrinsically curved and phased A-tract sequences,[5,34] we investigated the association of the density of A-tract sequences ($AnTm$ sequences, where $4 \leq (n + m)$)[35] with the peak position of H-NS binding signals in selected regions (Fig. 2A and B). As expected, peaks of H-NS binding

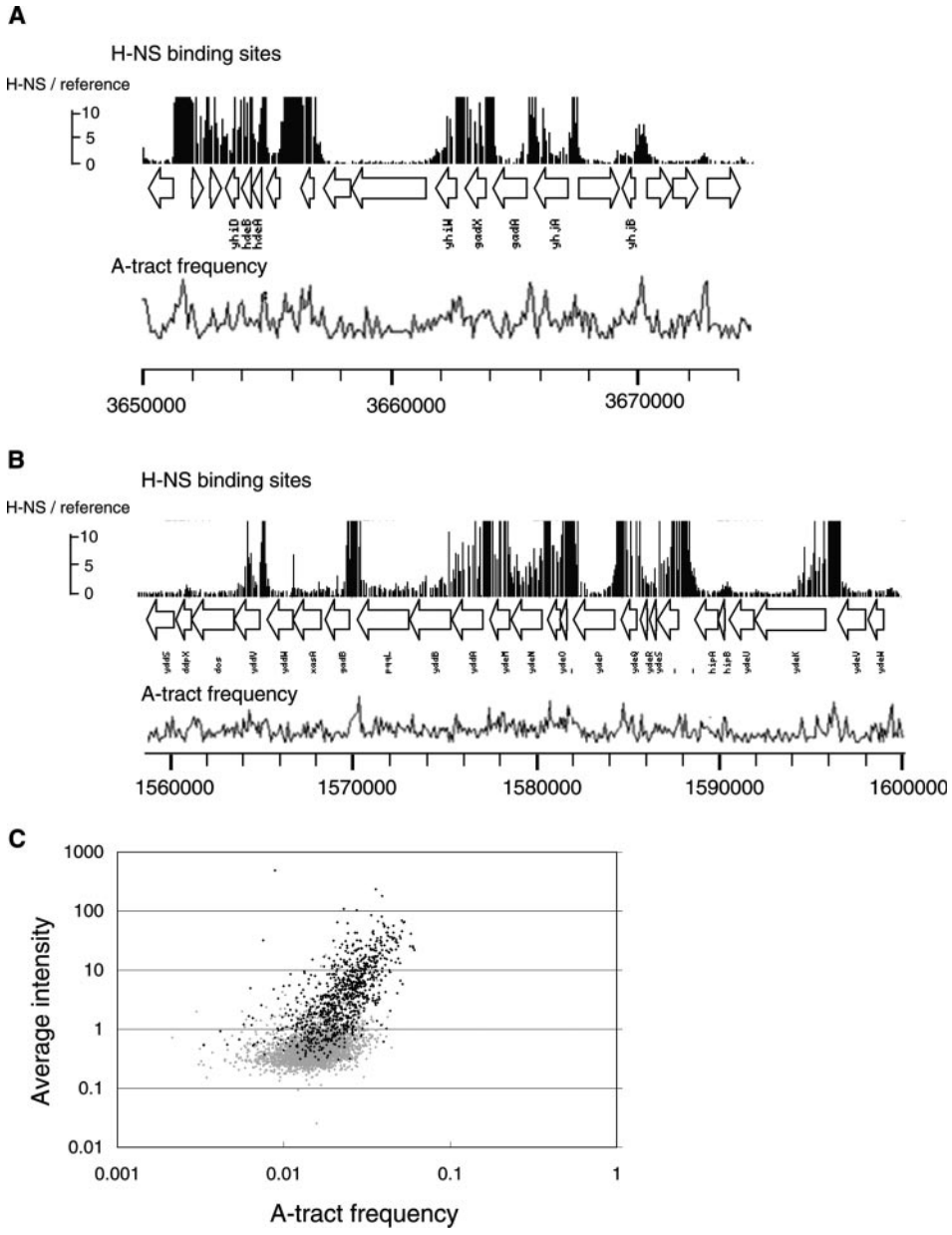

Figure 2. Correlation between H-NS binding sites and A-tract sequences. H-NS binding signals (vertical bars), arrangement of genes (thick arrows), frequency of A-tract sequences (number of A-tract sequences in 100 bp window was plotted by 50 bp steps; line graph) and chromosomal coordinates (horizontal line) are shown for the *gadA* (A) and *gadB* (B) loci. The genome-wide correlation between the average peak heights of H-NS binding signals and the density of A-tract sequences (number of A-tract sequences per 1 kb) in coding sequences of each gene is shown in panel C.

sites correlated with A-tract sequences. A genome-wide analysis showed a strong correlation between the average signal strength of H-NS binding of probes in each protein coding sequences and the density of A-tract sequences in the sequence (Fig. 2C).

3.2. High-resolution transcriptome analysis of wild type and hns-inactivated cells

To examine whether the H-NS binding detected by ChIP-chip analysis is related to gene expression, the transcriptionally active regions of the *E. coli* genome were mapped in wild type and *hns*-inactivated (Δ*hns*) cells using the same high-density oligonucleotide chip as described in Materials and Methods (Supplementary Figure S1).

A comparison of transcriptional activity with H-NS binding sites in wild-type cells revealed that the transcriptional activity of H-NS bound regions is generally low, as is observed, for example, in the fimbria-related gene cluster, *yadCKLM-htrE-ecpD-yadN* where H-NS binds to whole coding sequences (Fig. 3B)

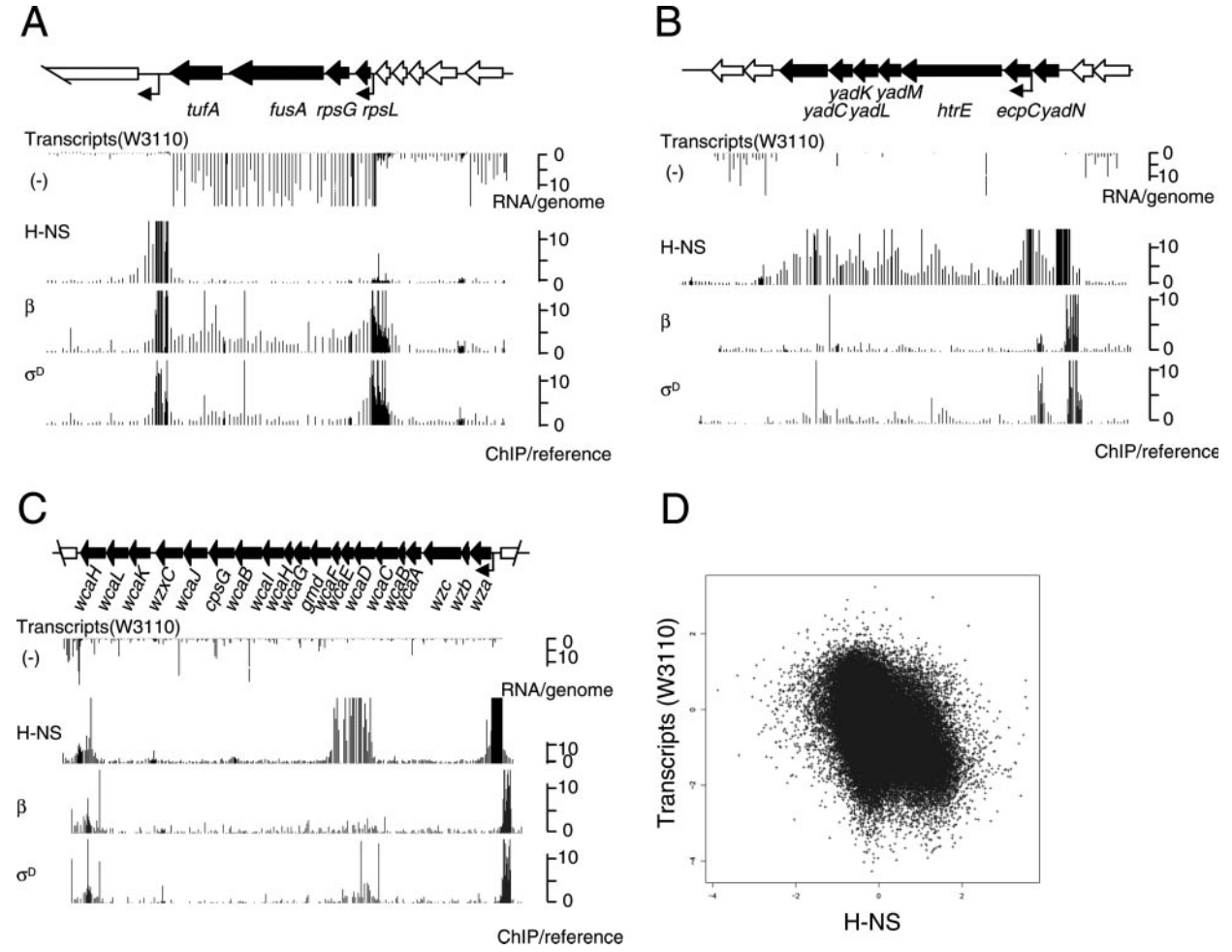

Figure 3. Comparison of transcriptional activity and H-NS binding signals. Transcription activity signals and H-NS, β and σ^D binding signals are shown for selected regions; the *rpsL* locus encoding ribosomal proteins and translation factors (A), the *yadN-yadC* locus densely covered by H-NS (B), and the *wza-wcaH* locus encoding colonic acid biosynthesis related genes and silenced by partial H-NS binding (C). Transcriptional activity signals are calculated by dividing the hybridization intensity of cDNA with that for genome DNA. The genome-wide correlation between transcriptional activity signals and H-NS binding signals for each 25mer probe on the chip is shown in panel D.

and in the LPS synthesis related genes, *wza-wzb-wzc-wcaA-H* where promoter region is covered by H-NS (Fig. 3C). By comparison, transcriptional active loci such as the ribosomal protein operon (Fig 3A) generally have no bound H-NS. In confirmation of these findings, a scatter plot of signal intensities of H-NS binding and transcriptional activity for each probe showed a reverse correlation (Fig. 3D), demonstrating that the transcriptional activity of H-NS bound sites is significantly lower than the transcriptional activity of unbound regions.

A comparison of transcriptional profiles in wild type and *hns*-inactivated cells showed that derepression of transcription in the mutant cells occurs at about 65% of the H-NS bound loci (Supplementary Table S1). Typical examples of derepression of transcription by H-NS inactivation are shown in Fig. 4. In some H-NS repressed genes, other factors are required to activate expression, even when H-NS is removed from the cell. For example, activation of transcription of the *bgl* promoter is dependent upon the activator protein, CRP, in addition to the removal of bound H-NS.[32]

A reduction in transcriptional activity was observed in several loci in the Δ*hns* mutant, especially for *malEFG* and *malK-lamB-malM* (Supplementary Figure S2). This may result from a reduction in local superhelicity in the Δ*hns* mutant, as proposed by Hardy and Cozzarelli.[14] Consistent with this finding, the *malEFG* and *malK-lamB-malM* loci are flanked by H-NS bound sequences (Supplementary Figure S2), and supercoiling is essential for the formation and stability of the initiation complex at the divergent *malE* and *malK* promoters.[36]

3.3. Genome-wide ChIP-chip analysis of RNA polymerase associated regions

A positive role for RNA polymerase in H-NS associated transcription repression has been demonstrated for the *rrnB* P1 and *hdeAB* promoters.[6,9,11]

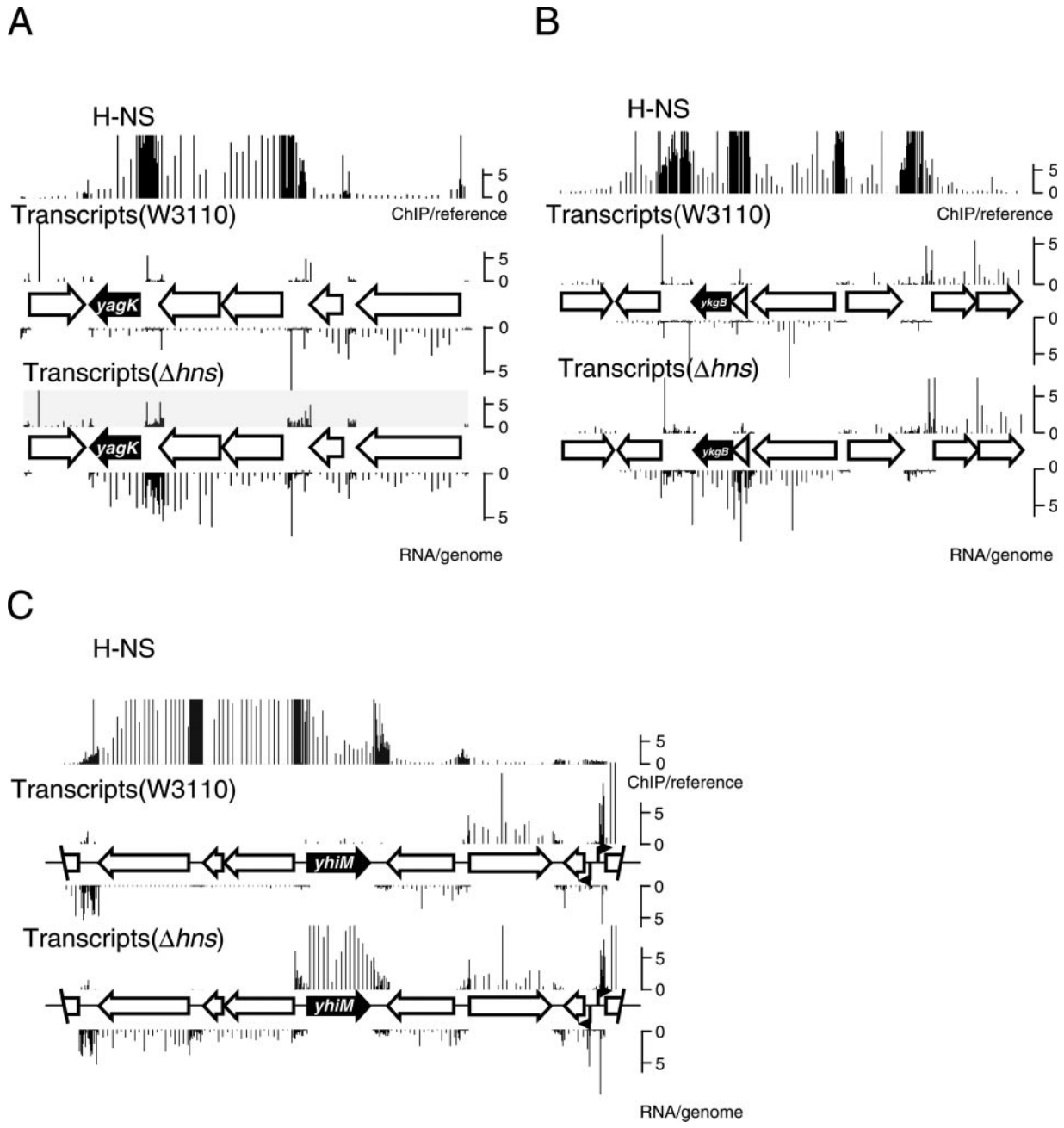

Figure 4. Derepression of transcriptional activity of H-NS bound genes in the *hns-inactivated* cells. H-NS binding signals in wild type (W3110) cells and transcriptional activity signals in wild type (W3110) and *hns-inactivated* (Δ*hns*) cells are shown for selected regions; the *yagK* (A), the *ykgB* (B) and the *yhiM* (C) loci. Transcriptional activities of Watson and Crick strands are shown above and below the ORF map, respectively.

To examine the degree of association between RNA polymerase and H-NS binding, we performed ChIP-chip analysis to map RNA polymerase binding sites, using monoclonal antibodies against the β and σ^D subunits (Supplementary Figure S1). In about 90 highly transcribed regions, the β subunit gave broad peaks throughout the operon, while the σ^D subunit gave a peak at the promoter region (Fig. 3A). It is likely that these signals correspond to the closed or stable open complexes at the promoters (σ^D signal) and elongation complexes in coding sequences (β signal). In addition to signals in highly expressed regions, we detected signal spikes for the β and σ^D subunits at the 5′ end of non-expressed or weakly expressed genes or operons (Fig. 3B and C). Furthermore, we found that such RNA polymerase signals overlapped with the H-NS peaks in >65% of the H-NS binding sites (Supplementary Table S1). The association of RNA polymerase and H-NS binding signals is especially evident in clustered H-NS binding sites (Fig. 5A). Scatter plots (Fig. 5B and C)

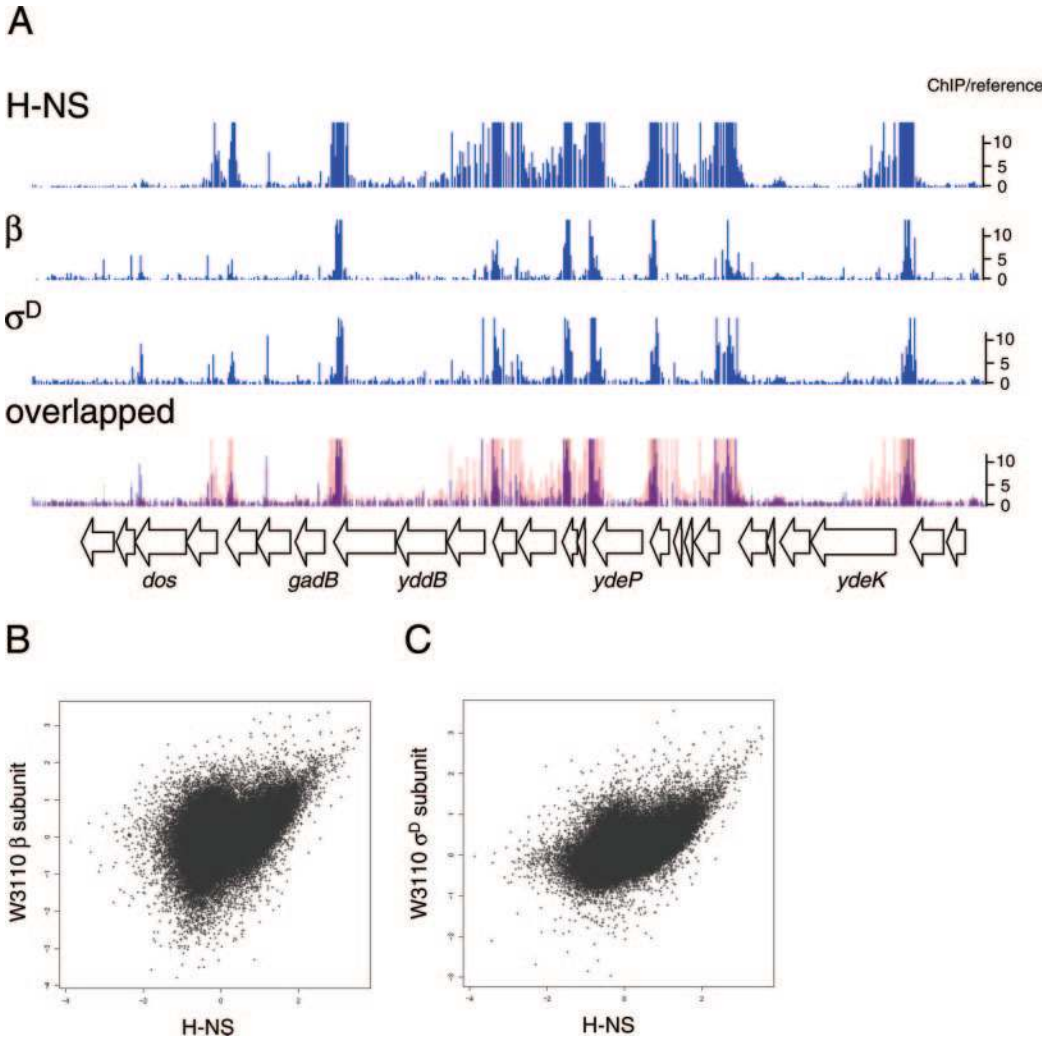

Figure 5. Overlapping of H-NS and RNA polymerase biding sites. As a typical example, the H-NS, β and σ^D binding signals and their merged image (overlapped) at the *gadB* locus are shown in panel (A). The H-NS binding signal is shown by the red bar in the merged image. The genome-wide correlation between the H-NS and β binding signals (B) and the H-NS and σ^D binding signals (C) for each 25mer probe on the chip are also indicated.

demonstrate the partial, but significant, association of RNA polymerase and H-NS along the genome. These results strongly suggest that the trapping of RNA polymerase in an inactive form, with respect to initiation or elongation of transcription, is a genome-wide phenomenon. The signals observed for the β and σ^D RNA polymerase subunits were essentially at the same loci in both wild type and Δ*hns* mutant cells (Supplementary Figure S3). These results suggest that the binding of RNA polymerase promotes the formation of higher order structure of the H-NS–DNA complex *in vivo*, as is observed *in vitro*.[9,11]

3.4. Correlation between H-NS binding sites and horizontally acquired DNA

A positive correlation appears to exist between horizontally acquired genetic elements and the presence of nucleoid-associated proteins.[3] A comparative genome hybridization analysis revealed the presence of highly divergent loci in the *E. coli* genome.[37] Fukiya et al.[37] demonstrated that many of the fimbria-related gene clusters in the *E. coli* strain W3110 are absent in a number of other *E. coli* and *Shigella* strains, suggesting that they have been acquired or deleted in a limited lineage. In this study we found these divergent loci to be silenced by H-NS (Supplementary Figure S1). In addition, 11 prophage and prophage-like sequences were partly or completely covered by bound H-NS, with the exception of one sequence (Supplementary Figure S1). These observations highlight the need for a comprehensive examination of the relationship between H-NS binding sites and horizontally acquired DNA.

On the basis of codon usage profiles, Lawrence and Ochman[38] proposed that 706 genes have been horizontally transferred into the *E. coli* genome after divergence

from the *Salmonella* lineage. Similarly, through a comprehensive comparison of bacterial genome sequences, Nakamura et al.[39] suggested that *E. coli* has horizontally acquired 1038 genes. A comparison of probable horizontally acquired genes and H-NS associated genes listed in Supplementary Table S1 shows that among 1047 genes found to be associated with H-NS, 644 (62%) were proposed to be horizontally acquired by Lawrence and Ochman[38] and/or Nakamura et al.[39] (Supplementary Table S2 and Figure S4). Furthermore, among the 403 remaining H-NS bound genes, 333 are contiguous to horizontally transferred genes bound by H-NS. These results strongly suggest that H-NS binds preferentially to horizontally acquired sequences. In the studies by Lawrence and Ochman[38] and Nakamura et al.[39], 70% and 54%, respectively, of predicted foreign genes are bound by H-NS. Fig. 6 shows a factorial correspondence analysis of codon usage of H-NS bound genes and predicted horizontally transferred genes, compared with genes of the entire *E. coli* genome, and clearly demonstrates the preferential binding of H-NS to horizontally acquired genes.

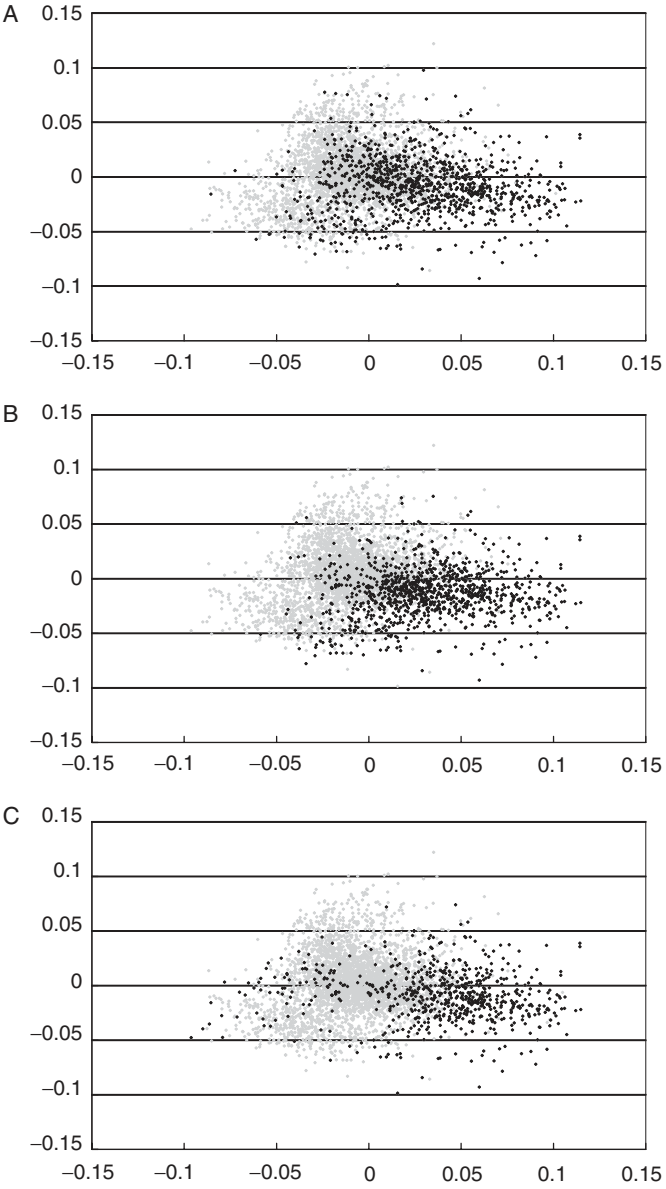

Figure 6. Factorial correspondence analysis of codon usage of H-NS bound and horizontally acquired genes. Genes are mapped in the two-dimensional field by PCA analysis of codon usage patterns. The gray dots indicate whole *E. coli* genes. The black dots in panels (A–C) indicate H-NS bound genes, and horizontally acquired genes predicted by Lawrence and Ochman,[38] and Nakamura et al.,[39] respectively.

4. Discussion

The genome-wide distribution of H-NS, RNA polymerase and transcriptionally active genes in exponentially growing *E. coli* cells was assessed using a custom-designed tiling chip. The results clearly demonstrate that H-NS specifically binds to about 250 loci in the *E. coli* genome, covering >1000 genes in total, to maintain the genes in a transcriptionally inactive state. The H-NS binding sites contain intrinsically curved A-tract sequences, reported to be preferentially bound by H-NS.[5] In clustered H-NS binding regions, the H-NS binding signals extend from highly curved A-tract regions to flanking sequences, suggesting a two-step binding model for H-NS binding[5] that involves nucleation at a high-affinity sequence followed by expansion into the flanking region.

The distribution of RNA polymerase in the *E. coli* genome has been investigated by two independent groups using the ChIP-chip method employing the β subunit antibody. Grainger et al.[19] reported that RNA polymerase associates predominantly with about 90 highly expressed transcriptional units in cells rapidly growing in LB medium. In addition, using rifampicin to trap RNA polymerase at functional promoters, they demonstrated that RNA polymerase associates with >500 functional promoters. In similar experiments in which rifampicin was added to cells growing in M9 minimal medium, Herring et al.[20] identified about 1100 RNA polymerase binding sites. In this report, we performed ChIP-chip analysis of RNA polymerase binding sites in cells rapidly growing in LB medium, using antibodies against the β and σ^D subunits, and compared the distribution of RNA polymerase binding sites with the distribution of transcriptionally active genes and bound H-NS. Consistent with the finding of Grainger et al.,[19] strong signals for RNA polymerase were detected upstream of highly transcribed genes. A preliminary analysis of transcription start sites and transcription units from mapping data, and RNA polymerase distribution data, revealed a strong correlation with known transcription initiation sites and operon structures listed in the RegulonDB database[40] (see our

web site http://genome.naist.jp/bacteria/array/ecol.html). A detailed analysis, including a comparison of promoters mapped by Grainger et al.[19] and Herring et al.[20], is in progress and will be published elsewhere.

Signals for both the β and σ^D subunits of RNA polymerase were also detected at the 5′ end of non-expressed or weakly expressed genes or operons, and were associated with the binding of H-NS. Consistent with our results, Grainger et al.[41] also reported a strong correlation between the distribution of bound H-NS and RNA polymerase. Together, these results suggest that the trapping of RNA polymerase in an inactive form, which thereby prevents the initiation or elongation of transcription, provides a general mode of transcription repression by H-NS. Shin et al.[11] proposed that weak DNA-binding proteins such as H-NS, with a binding constant in the micromolar range, would not have the capacity to bend short stretches of the DNA helix to create loops and thus require association with the σ^D subunit of RNA polymerase. Such a higher order H-NS-RNA polymerase–DNA complex would be formed throughout the genome.

A close relationship between H-NS and horizontally acquired genes was identified in this study. The role of horizontally acquired genes in the evolution of bacterial genomes remains unknown and the mechanisms underlying the regulation of expression, and the mode of integration of these genes into the regulatory network of the host cell, is not understood. The unregulated expression of horizontally acquired genes would be disadvantageous to the growth of the host cell and result in elimination of the population harboring the newly acquired genes or the deletion of deleterious sequences from the genome. The nucleoid-associated protein has been suggested as a candidate cellular regulator to control incoming genetic elements.[3] Our findings strongly suggest that the main role of H-NS is to maintain horizontally acquired genes in a transcriptionally inactive state under conditions leading to inappropriate expression. H-NS dependent repression would play an important role in the stable transmission of horizontally acquired genetic elements and, as a result, lead to the diversification and adaptation of *E. coli* genome to diverse growth conditions. During the preparation of this report, an independent report on the selective silencing of foreign DNA with low GC content by H-NS in *Salmonella* proposed that H-NS provides a bacterial defense against foreign DNA, enabling the acquisition of DNA from exogenous sources while avoiding detrimental consequences from the unregulated expression of newly acquired genes.[22]

About half of the genes suggested by Lawrence and Ochman[38] and/or Nakamura et al.[39] to be horizontally acquired are not bound by H-NS. Although horizontally acquired genes with low GC content are reported to be selectively silenced by H-NS in *Salmonella*,[22] we found no obvious difference in GC content or density of A-tracts between H-NS bound and unbound foreign genes in *E. coli* (data not shown). Parts of these genes are contiguous with H-NS bound sequences and probably silenced by H-NS indirectly (see Fig. 3B). It is of interest to note that although most fimbria-related gene clusters are bound by H-NS, three clusters conserved between all *E. coli* and *Shigella* strains examined[37] are not bound by H-NS. This finding suggests that H-NS binds to recently acquired genes. It is possible that horizontally acquired genes not bound to H-NS are controlled by other nucleoid-associated proteins or have integrated into a specific regulatory pathway of the host. A detailed comparison of H-NS distribution between *E. coli* and *Salmonella* will provide further insight into the role of H-NS in genome evolution.

In summary, our genome-wide approach using ChIP-chip analysis not only provided a comprehensive assessment of the genomic distribution of bound H-NS and modes of transcriptional regulation, but also revealed an avenue for furthering the understanding of the structure, function and evolution of the bacterial nucleoid.

Acknowledgements: We thank Katsuhiko Shirahige, Takehiko Itho, Tetsuya Hayashi, Tohru Tobe, David Grainger and Steve Busby, for helpful discussions. We thank Eiko Nakamoto for technical assistance. This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas Systems Genomics from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Supplementary Data: Supplementary data are available online at http://dnaresearch.oxfordjournals.org

References

1. Zimmerman, S. B. 2006, Cooperative transitions of isolated *Escherichia coli* nucleoids: implications for the nucleoid as a cellular phase, *J. Struct. Biol.*, **153**, 160–175.
2. Dame, R. T., Luijsterburg, M. S., Krin, E., Bertin, P. N., Wagner, R., and Wuite, G. J. 2005, DNA bridging: a property shared among H-NS-like proteins, *J. Bacteriol.*, **187**, 1845–1848.
3. Dorman, C. J. 2004, H-NS: a universal regulator for a dynamic genome, *Nat. Rev. Microbiol.*, **2**, 391–400.
4. Rimsky, S. 2004, Structure of the histone-like protein H-NS and its role in regulation and genome superstructure, *Curr. Opin. Microbiol.*, **7**, 109–114.
5. Rimsky, S., Zuber, F., Buckle, M., and Buc, H. 2001, A molecular mechanism for the repression of transcription by the H-NS protein, *Mol. Microbiol.*, **42**, 1311–1323.
6. Dame, R. T., Wyman, C., Wurm, R., Wagner, R., and Goosen, N. 2002, Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1, *J. Biol. Chem.*, **277**, 2146–2150.
7. Spurio, R., Durrenberger, M., Falconi, M., La Teana, A., Pon, C. L., and Gualerzi, C. O. 1992, Lethal overproduction of the *Escherichia coli* nucleoid protein

H-NS: ultramicroscopic and molecular autopsy, *Mol. Gen. Genet.*, **231**, 201–211.
8. Hommais, F., Krin, E., Laurent-Winter, C., et al. 2001, Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS, *Mol. Microbiol.*, **40**, 20–36.
9. Schroder, O. and Wagner, R. 2000, The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex, *J. Mol. Biol.*, **298**, 737–748.
10. Dole, S., Nagarajavel, V., and Schnetz, K. 2004, The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli bgl* operon downstream of the promoter, *Mol. Microbiol.*, **52**, 589–600.
11. Shin, M., Song, M., Rhee, J. H., et al. 2005, DNA looping-mediated repression by histone-like protein H-NS: specific requirement of Esigma70 as a cofactor for looping, *Genes Dev.*, **19**, 2388–2398.
12. Johansson, J., Dagberg, B., Richet, E., and Uhlin, B. E. 1998, H-NS and StpA proteins stimulate expression of the maltose regulation in *Escherichia coli*, *J. Bacteriol.*, **180**, 6117–6125.
13. Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A., et al. 1999, Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon, *J. Bacteriol.*, **181**, 7500–7508.
14. Hardy, C. D. and Cozzarelli, N. R. 2005, A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure, *Mol. Microbiol.*, **57**, 1636–1652.
15. Spassky, A., Rimsky, S., Garreau, H., and Buc, H. 1984, H1a, an *E. coli* DNA-binding protein which accumulates in stationary phase, strongly compacts DNA in vitro, *Nucleic Acids Res.*, **12**, 5321–5340.
16. Mockler, T. C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S. E., and Ecker, J. R. 2005, Applications of DNA tiling arrays for whole-genome analysis, *Genomics*, **85**, 1–15.
17. Katou, Y., Kaneshiro, K., Aburatani, H., and Shirahige, K. 2006, Genomic approach for the understanding of dynamic aspect of chromosome behavior, *Methods Enzymol.*, **409**, 389–410.
18. Grainger, D. C., Overton, T. W., Reppas, N., et al. 2004, Genomic studies with *Escherichia coli* MelR protein: applications of chromatin immunoprecipitation and microarrays, *J. Bacteriol.*, **186**, 6938–6943.
19. 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. USA*, **102**, 17693–17698.
20. Herring, C. D., Raffaelle, M., Allen, T. E., et al. 2005, Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays, *J. Bacteriol.*, **187**, 6166–6174.
21. Wade, J. T., Reppas, N. B., Church, G. M., and Struhl, K. 2005, Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites, *Genes Dev.*, **19**, 2619–2630.
22. Navarre, W. W., Porwollik, S., Wang, Y., et al. 2006, Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*, *Science*, **313**, 236–238.
23. Ito, K., Oshima, T., Mizuno, T., and Nakamura, Y. 1994, Regulation of lysyl-tRNA synthetase expression by histone-like protein H-NS of *Escherichia coli*, *J. Bacteriol.*, **176**, 7383–7386.
24. Uzzau, S., Figueroa-Bossi, N., Rubino, S., and Bossi, L. 2001, Epitope tagging of chromosomal genes in *Salmonella*, *Proc. Natl Acad. Sci. USA*, **98**, 15264–15269.
25. Hayashi, K., Morooka, N., Yamamoto, Y., et al. 2006, Highly accurate genome sequences of *Escherichia coli K-12* strains MG1655 and W3110, *Mol. Syst. Biol.*, **2**, E1–E5.
26. Kampa, D., Cheng, J., Kapranov, P., et al. 2004, Novel RNAs identified from an in-depth analysis of the transcriptome of Human chromosome 21 and 22, *Genome Res.*, **14**, 331–342.
27. David, L., Huber, W., Granovskaia, M., et al. 2006, A high-resolution map of transcription in the yeast genome, *Proc. Natl Acad. Sci. USA*, **103**, 5320–5325.
28. Kanaya, S., Kudo, Y., Nakamura, Y., and Ikemura, T. 1996, Detection of genes in *Escherichia coli* sequences determined by genome projects and prediction of protein production levels, based on multivariate diversity in codon usage, *Comput. Appl. Biosci.*, **12**, 213–25.
29. Higgins, C. F., Dorman, C. J., Stirling, D. A., et al. 1988, A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*, *Cell*, **52**, 569–584.
30. Owen-Hughes, T. A., Pavitt, G. D., Santos, D. S., et al. 1992, The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression, *Cell*, **71**, 255–265.
31. Ueguchi, C. and Mizuno, T. 1993, The Escherichia coli nucleoid protein H-NS functions directly as a transcriptional repressor, *EMBO J.*, **12**, 1039–1046.
32. Schnetz, K. and Wang, J. C. 1996, Silencing of the *Escherichia coli bgl* promoter: effects of template supercoiling and cell extracts on promoter activity in vitro, *Nucleic Acids Res.*, **24**, 2422–2428.
33. Giangrossi, M., Zattoni, S., Tramonti, A., De Biase, D., and Falconi, M. 2005, Antagonistic role of H-NS and GadX in the regulation of the glutamate decarboxylase-dependent acid resistance system in *Escherichia coli*, *J. Biol. Chem.*, **280**, 21498–21505.
34. Zuber, F., Kotlarz, D., Rimsky, S., and Buc, H. 1994, Modulated expression of promoters containing upstream curved DNA sequences by the *Escherichia coli* nucleoid protein H-NS, *Mol. Microbiol.*, **12**, 231–240.
35. Tolstorukov, M. Y., Virnik, K. M., Adhya, S., and Zhurkin, V. B. 2005, A-tract clusters may facilitate DNA packaging in bacterial nucleoid, *Nucleic Acids Res.*, **33**, 3907–3918.
36. Richet, E. and Raibaud, O. 1991, Supercoiling is essential for the formation and stability of the initiation complex at the divergent *malE*p and *malK*p promoters, *J. Mol. Biol.*, **218**, 529–542.
37. Fukiya, S., Mizoguchi, H., Tobe, T., and Mori, H. 2004, Extensive genomic diversity inpathogenic *Escherichia coli*

and *Shigella* Strains revealed by comparative genomic hybridization microarray, *J. Bacteriol.*, **186**, 3911–3921.
38. Lawrence, J. G. and Ochman, H. 1998, Molecular archaeology of the *Escherichia coli* genome, *Proc. Natl Acad. Sci. USA*, **95**, 9413–9417.
39. Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. 2004, Biased biological functions of horizontally transferred genes in prokaryotic genomes, *Nat. Genet.*, **36**, 760–766.
40. Salgado, H., Gama-Castro, S., Peralta-Gil, M., et al. 2006, RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions, *Nucleic Acids Res.*, **34**, D394–D397.
41. 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**, 4642–4652.