*4642–4652* *Nucleic Acids Research, 2006, Vol. 34, No. 16* *Published online 8 September 2006*
*doi:10.1093/nar/gkl542*

# Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome

**David C. Grainger*, Douglas Hurd[1], Martin D. Goldberg and Stephen J. W. Busby**

School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK and
[1]Oxford Gene Technology, Begbroke Science Park, Sandy Lane, Yarnton, Oxford OX5 1PF, UK

Received June 26, 2006; Revised July 11, 2006; Accepted July 12, 2006

## ABSTRACT

**The *Escherichia coli* chromosome is condensed into an ill-defined structure known as the nucleoid. Nucleoid-associated DNA-binding proteins are involved in maintaining this structure and in mediating chromosome compaction. We have exploited chromatin immunoprecipitation and high-density microarrays to study the binding of three such proteins, FIS, H-NS and IHF, across the *E.coli* genome *in vivo*. Our results show that the distribution of these proteins is biased to intergenic parts of the genome, and that the binding profiles overlap. Hence some targets are associated with combinations of bound FIS, H-NS and IHF. In addition, many regions associated with FIS and H-NS are also associated with RNA polymerase.**

## INTRODUCTION

In eukaryotic organisms, DNA is compacted by histones, which organize DNA into nucleosomes that are further organized into the higher order structures that form chromatin. Recently, the use of chromatin immunoprecipitation (ChIP), in combination with high-density microarrays (ChIP-chip), has allowed the distribution of some histones to be mapped across the *Saccharomyces cerevisiae* genome (1,2). These studies showed that, although histones associate with both coding and non-coding parts of the genome, they are depleted at promoters. Consistent with this, Nagy and co-workers (3) were able to fractionate the yeast genome on the basis of histone density. Eukaryotic chromatin is remodelled by transcription (4) and this remodelling can involve both the removal and the recruitment of histones (5,6).

By comparison, the mechanisms used by prokaryotes to organize their chromosomal DNA are poorly understood. Bacterial DNA forms a highly condensed, yet undefined, structure called the nucleoid, that is thought to be ordered by supercoiling, macromolecular crowding, RNA and nucleoid-associated proteins, the prokaryotic equivalent of histones (7–9). Extensive studies with *Escherichia coli* have shown that many of these nucleoid-associated proteins compact chromosomal DNA (10–13) and can also function as transcription factors (14,15). Thus, factor for inversion stimulation (FIS), histone-like nucleoid structuring protein (H-NS) and the integration host factor (IHF) are nucleoid-associated proteins, whose individual subunits are present at 60 000, 20 000 and 12 000 copies, respectively, in rapidly growing *E.coli* cells (16). All three proteins bind to A:T rich DNA targets and alter DNA topology. For H-NS, these DNA sites are highly degenerate but some DNA-binding specificity is displayed by FIS and IHF. Upon binding, FIS bends DNA by between 50° and 90° (17), whilst IHF induces a bend of 160° (18). H-NS is thought to oligomerize at intrinsically curved DNA sequences and, once bound, may induce further DNA bending (19). In addition to their role as chromosome shaping proteins, FIS, H-NS and IHF function as gene-specific transcription factors. For example, H-NS represses the transcription of many non-essential genes (20,21). Similarly, during rapid growth, FIS represses many promoters (22), though its best-characterized role is as an activator at the *rrn* operons (23).

Little is known about the genome-wide distribution of the various nucleoid proteins. This is illustrated by the fact that only 63, 36 and 55 targets for FIS, H-NS and IHF respectively are listed in the Ecocyc database [www.ecocyc.org (24)]. There have been several attempts to study the distribution of FIS, H-NS and IHF across bacterial genomes. Robison and co-workers (25) used bioinformatics to search the *E.coli* genome for DNA sequences that resemble the binding sites for FIS, H-NS and IHF. The study identified >10 000 binding sites for each factor and, although <10% of the *E.coli* genome is non-coding, >23% of the predicted targets for each protein were in non-coding DNA. Other attempts to study the distribution of nucleoid-associated proteins have relied on comparing the transcriptomes of wild-type and mutant cells (20,22,26). The distribution of the protein under study is then inferred from changes in transcription resulting from a mutation in its gene. However, such approaches are unable to distinguish primary from secondary effects, and they

*To whom correspondence should be addressed. Tel: +44 121 414 5435; Fax: +44 121 414 5925; Email: d.grainger@bham.ac.uk

yield no information about the distribution of nucleoid proteins between coding and non-coding DNA.

The aim of this study was to measure directly the binding of FIS, H-NS and IHF across the entire *E.coli* genome using a ChIP-chip analysis. We show that, in contrast to eukaryotic histone proteins, the binding of these proteins is biased towards non-coding parts of the genome. We identified overlap in the DNA-binding profiles of FIS and H-NS, and many regulatory regions bound with these two proteins were also associated with RNA polymerase. Some binding sites for nucleoid proteins were found in coding DNA and, for FIS and H-NS, some of these sites are in regions of the genome that are highly transcribed.

## MATERIALS AND METHODS

### Bacterial growth conditions

Experiments were performed using *E.coli* strain MG1655 grown to mid-log phase ($OD_{650}$ 0.3–0.4) at 37°C in M9 minimal media supplemented with fructose. Our choice of a minimal and defined growth condition allowed us to monitor the effect of adding specific reagents to *E.coli* cultures. Thus, for induction of the *lac* operon, 2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added and cross-linking with formaldehyde was initiated at different time points. Salicylic acid was added to a final concentration of 5 mM and cross-linking was initiated after 40 min.

### Chromatin immunoprecipitation

Bacterial cells were treated with formaldehyde, harvested and lysed, and nucleoprotein was prepared as described by Grainger *et al*. (27). Immunoprecipitation was then performed using monoclonal antibodies against the β subunit of RNA polymerase (Neoclone, Madison, USA) and H-NS (donated by Jay Hinton) or rabbit polyclonal antibodies against FIS [donated by Akira Ishihama, see Azam *et al*. (16)] and IHF (donated by Steve Goodman). Cross-reactivity of the anti-H-NS antibody with StpA was not observed (M. D. Goldberg, unpublished data). Immunoprecipitated DNA samples or total cell nucleoprotein samples were purified and labelled with Cy5 or Cy3, without amplification, as described previously (27).

### PCR analysis of immunoprecipitated DNA

Quantitative PCR analyses were performed in real-time using an Applied Biosystems 7700 sequence detector. Relative occupancy values shown in Figure 6 were calculated by determining the immunoprecipitation efficiency of the target DNA sequence in samples from cultures either with or without IPTG. As internal controls, we used segments of the *rrl* and *mtlA* genes. The fold enrichment of target sequences, relative to the control regions, is shown as occupancy units and error bars represent the standard deviation of the mean of three independent experiments. Primer pairs used in this analysis were *lacZ* up (5′-aatcgccttgcagcac-3′) and *lacZ* down (5′-gccattcgccattca-3′), *lacY* up (5′-cgccgttttactcttttcg-3′) and *lacY* down (5′-gcaggaaacgccaataacat-3′), *lacA* up (5′-atatgtgcgaaggcttaccg-3′) and *lacA* down (5′-aattgcggcctatatggatg-3′), *rrl* up (5′-ctacggtgctgaagcaacaa-3′) and *rrl* down (5′-cgaagttacggcaccatttt-3′), *mtlA* up (5′-cgcgacagcaaacataagaa-3′) and *mtlA* down (5′-gttcggtaaccaccctgttg-3′).

### Microarray analysis of immunoprecipitated DNA

Microarrays (Oxford Gene Technology) were designed and produced specifically to analyse DNA obtained from ChIP experiments with *E.coli* MG1655 (27). Arrays consisted of 21 321 60mer oligonucleotide probes that match MG1655 sequences at intervals of ∼160 bp. Unamplified DNA obtained from immunoprecipitations and control MG1655 genomic DNA was differentially labelled and hybridized to the microarray as described previously (27). Values shown are the average of two independent experiments.

### ChIP-chip data analysis

The Cy5/Cy3 intensity ratio was calculated for each spot and plotted against the corresponding position on the *E.coli* MG1655 chromosome. A cut-off was chosen and all probes that had an intensity ratio greater than this value in both experiments were selected as a target. Adjacent target probes were merged, the target position being defined by centre of the probe with the highest average intensity ratio. Targets were judged to be in non-coding DNA if the peak centre fell within 200 bp of a non-coding DNA sequence. Leeway of 200 bp was allowed because, for some non-coding sequences, the nearest probe was at the 5′ end of an adjacent gene and because peak locations are only accurate to ∼200 bp (27). Note that, in this study, we refer to DNA encoding stable RNA as 'coding' DNA and that the ribosomal RNA operons were treated as a single entity. To estimate the FDR one dataset was randomized with respect to the probe positions. The number of probes passing the cut-off was then re-calculated. This process was repeated 50 times to determine the average number of probes that passed the cut-off after randomization. When a stringent cut-off was used the number of probes passing the cut-off post randomization was ∼1% of the number that originally passed the cut-off.

To investigate the relationship between DNA sequence and the binding of the various factors we ranked probes on the microarray in order of their % A:T content. Probes with the same A:T content were grouped and the average Cy5/Cy3 ratio was calculated, using our ChIP-chip data for either FIS, H-NS, IHF or, as a control, RNA polymerase. The % A:T content of each group of probes was then plotted against the average signal intensity (Figure 3). Because groups of probes with a very high or very low % A:T content were small they could not be used to calculate a meaningful average signal intensity they were excluded from the analysis (this applied to <1% of the probes on the microarray).

### Phenol–chloroform fractionation of *E.coli* nucleoprotein

Cells were grown to mid-log phase in minimal media and nucleoprotein was cross-linked with 1% formaldehyde. Cells were then washed three times with Tris-buffered saline and lysed by sonication, further sonication was used to fragment the genomic DNA. Cell debris was removed by centrifugation and the sample was split into two aliquots. The first aliquot was retained as a control and the second was

treated with phenol–chloroform to remove DNA fragments cross-linked to large amounts of protein. The control sample and the phenol–chloroform treated sample were then decross-linked and processed as described for immunoprecipitation experiments (27) before being labelled with Cy3 and Cy5, respectively, and hybridized to the microarray.

## RESULTS

### DNA fragments associated with FIS, H-NS, IHF and RNA polymerase *in vivo*

Previously, we analysed the distribution, *in vivo*, of RNA polymerase and different transcription factors bound to the *E.coli* chromosome (27). DNA targets were immunoprecipitated using antibodies directed against the different proteins. To analyse this DNA, we used microarrays of ~22 000 60mer oligonucleotides that correspond to evenly spaced sections of the *E.coli* genome. These arrays allowed us to measure binding in both coding and non-coding sections of the *E.coli* chromosome. We reasoned that these arrays would be suitable for studying the distribution of nucleoid-associated proteins.

Our first aim was to generate ChIP-chip data for FIS, H-NS, IHF and RNA polymerase under the same conditions and thus create a 'snapshot' of their binding profiles. Thus, *E.coli* strain MG1655 was grown aerobically in M9 minimal media to an $OD_{650}$ of 0.4. Cells were treated with formaldehyde, lysed and their DNA was sonicated, yielding DNA fragments of ~500–1000 bp. Antibodies directed against either FIS, H-NS, IHF or the β subunit of RNA polymerase were then used to precipitate DNA fragments associated with each protein. After purification, immunoprecipitated DNA fragments were labelled with Cy5. A control sample, labelled with Cy3, was generated from an aliquot of the total cross-linked nucleoprotein, reserved before immunoprecipitation. The Cy5-labelled samples were then each mixed with the control Cy3-labelled sample and hybridized to the microarray. After washing and scanning, the Cy5/Cy3 signal intensity ratio was calculated for each probe. All experiments were done in duplicate and replicates had a correlation co-efficient of >0.9.

The complete dataset for each experiment is listed in Supplementary Table 1, which shows the Cy5/Cy3 ratio for each probe on the microarray as a function of its position on the genome. Figure 1 shows an overview of the profiles for FIS, H-NS and IHF. It is clear that each of these proteins binds to many different targets across the *E.coli* chromosome, generating a wide range of signal intensities. Thus, first, we searched for the targets for each protein that are listed in the current Ecocyc database.

For the FIS experiment, the measured Cy5/Cy3 ratio ranges from 0.02 to 81.5. Discrete peaks, which we attribute to specific FIS-binding loci, were evident at 51 of the 63 targets listed in Ecocyc (*pdxA, lpdA, queA, glnQ, pheV, rnpB, nirB, mtlA, gyrB, bglG, fumB* and *pheU* were not identified). The signal intensities of the peaks observed at these 51 targets ranged from 0.43 (*aldB*) to 35.8 (*hns*). Figure 2A illustrates data for FIS binding at *aldB* and *hns*.

For H-NS, the Cy5/Cy3 ratio varied from 0.01 to 76.10. We identified signals for H-NS binding at 30 of the 36 targets listed by Ecocyc (we failed to identify *lacZ, bolA, smtA, hlyE,*

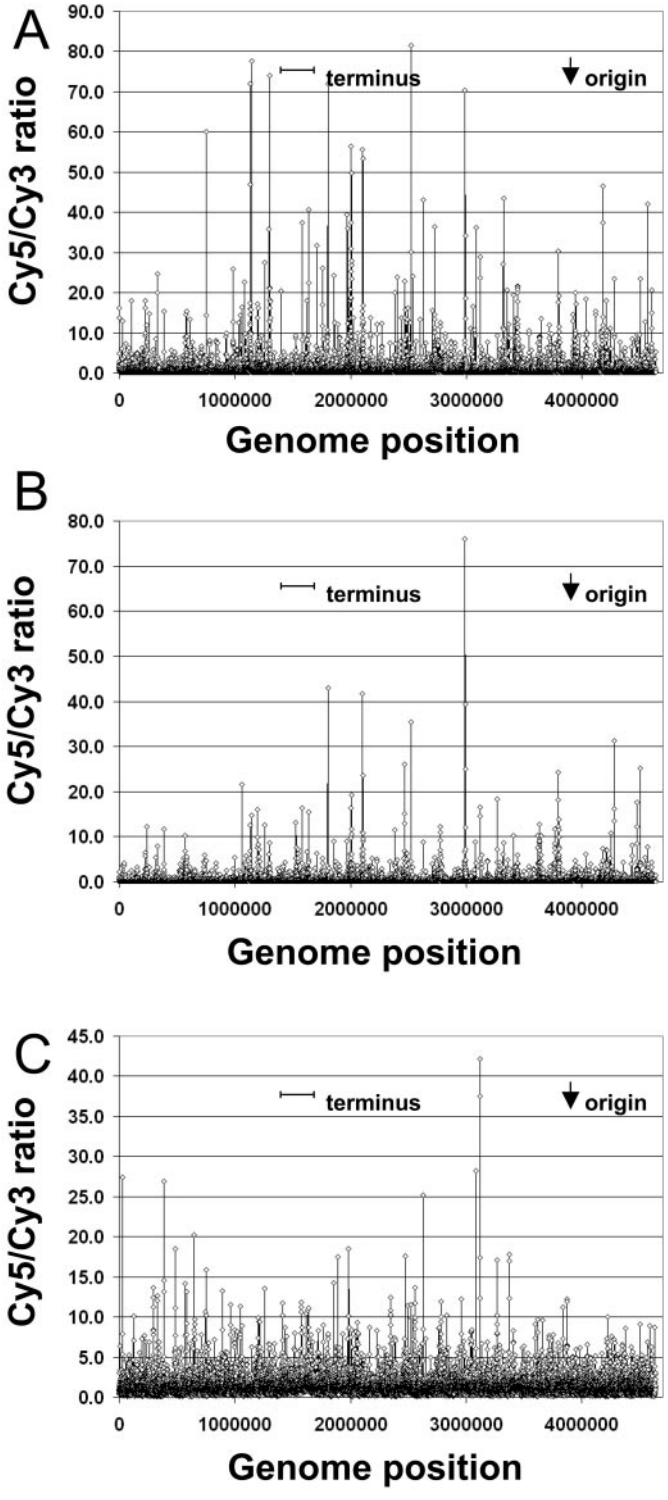


**Figure 1.** Genome-wide DNA-binding profiles of FIS, H-NS and IHF. The figure shows results from ChIP-chip experiments that measure profiles of FIS (**A**), H-NS (**B**) and IHF (**C**) binding across the *E.coli* genome. Binding signals (*y*-axis) are plotted against the location on the 4.64 Mb *E.coli* chromosome (*x*-axis). The positions of the origin and termini of replication are shown.

*flhD* and *nirB*) and the peak intensities ranged from 0.39 (*adiA*) to 16.4 (*fliC*) (Figure 2B). Binding of H-NS at the well-characterized *leuO* and *proV* loci is shown in Supplementary Figure 1.

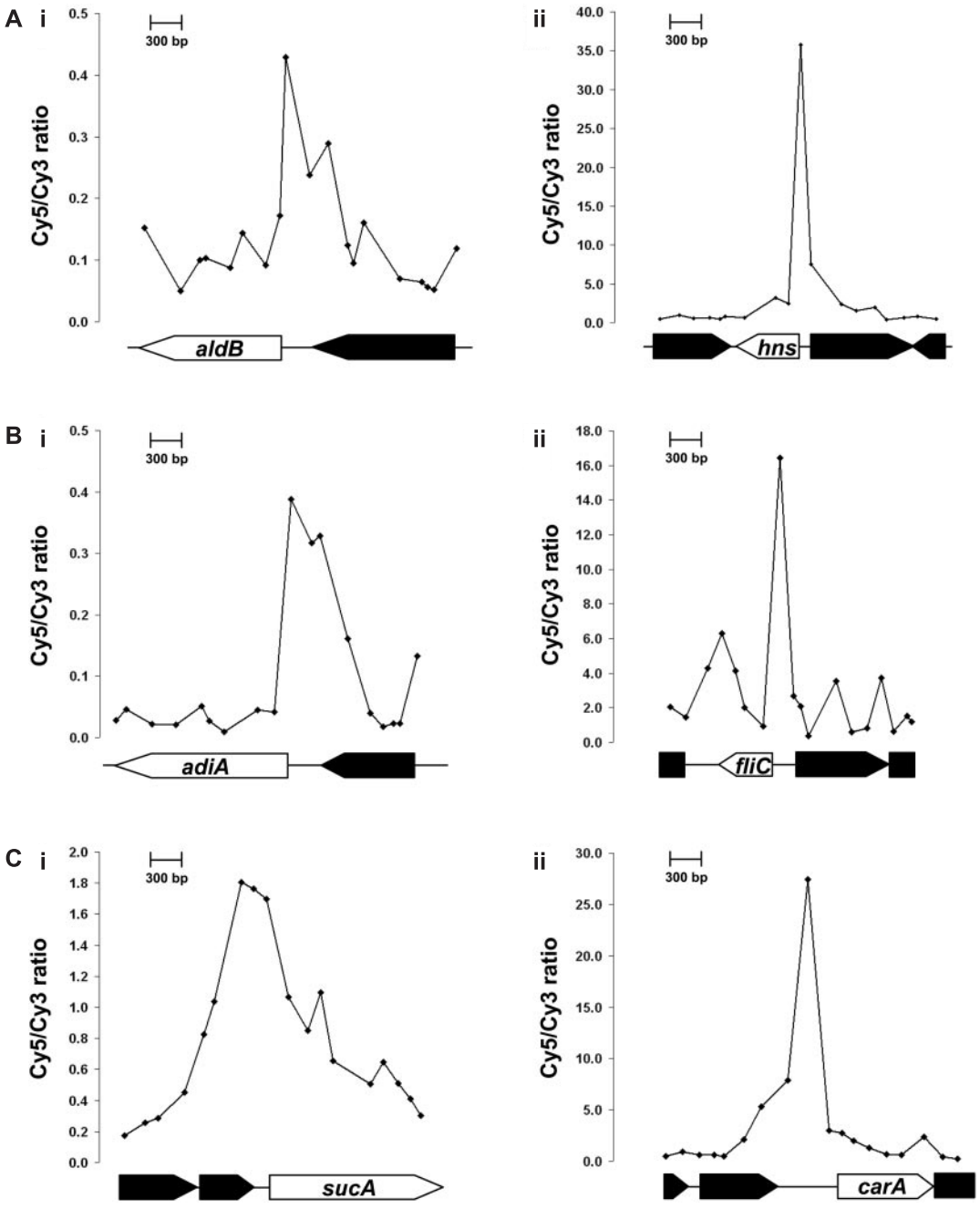


**Figure 2.** Association of FIS, H-NS and IHF with known target sites. The figure shows results from ChIP-chip experiments that measure profiles of FIS, H-NS and IHF binding across segments of the *E.coli* genome that are illustrated schematically below the profile. Binding signals (*y*-axis) are plotted against the genomic location (*x*-axis) for two known targets for each factor. Data for FIS binding at *adlB* and *hns* are shown in (**A**), H-NS binding at *adiA* and *fliC* are shown in (**B**) and IHF binding at *sucA* and *carA* are shown in (**C**).

For the experiment with IHF, the signal ranges from 0.10 to 42.13. We observed peaks, signifying IHF binding, at 51 of the 55 targets listed by the Ecocyc database (we failed to identify *gcd, hemA, glcD* and *mtr*). The peak intensities at these targets ranged from 1.8 (*sucA*) to 27.4 (*carAB*) (Figure 2C).

Our aim was to generate an unbiased list of targets for each factor and to measure their distribution between coding and non-coding DNA. Using a stringent cut-off (selected to give a false discovery rate, FDR, of ~1%) we identified 224, 100 and 135 targets for FIS, H-NS and IHF, respectively. These targets are listed in Supplementary Tables 2–4. Although non-coding sequences account for <10% of the *E.coli* genome, ~50% of the targets for each protein are in non-coding DNA. This distribution was not substantially changed when a more relaxed cut-off was set (corresponding to the value of the lowest scoring known target for each nucleoid protein).

FIS, H-NS and IHF are known to bind to A:T rich targets. Thus, with the data from each experiment, we calculated the average of the measured signal intensities for probes grouped

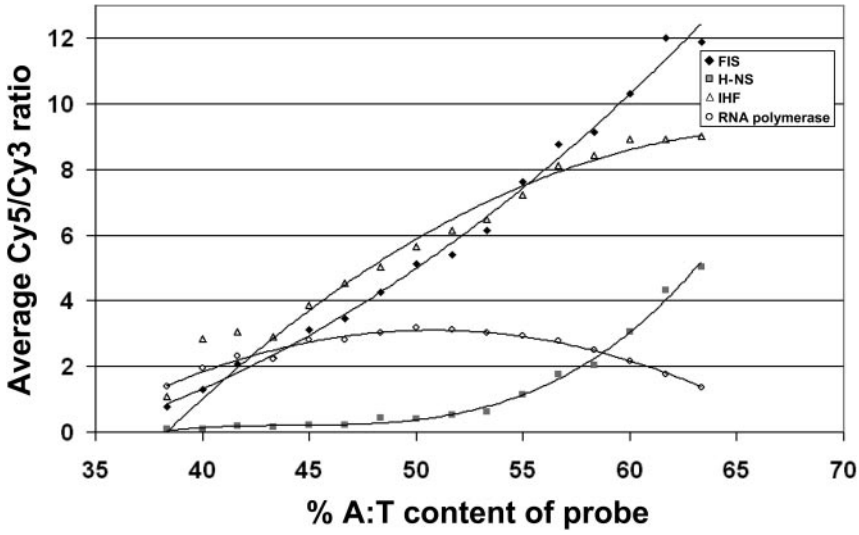


**Figure 3.** Relationship between the A:T content of microarray probes and binding signal for RNA polymerase, FIS, H-NS and IHF. The figure shows averaged ChIP-chip binding signals for RNA polymerase, FIS, H-NS and IHF at groups of microarray probes (*y*-axis) sorted by their A:T content (*x*-axis).

according to their % A:T content. Figure 3 shows a plot of the average binding signal for each protein as a function of probe A:T content. The graph shows that, as expected, there is an overall positive correlation between the A:T content of probes and the binding of FIS, H-NS and IHF. As a control, data from an experiment in which we had measured the chromosome-wide distribution of RNA polymerase was included. These data show that, in contrast to FIS, H-NS and IHF, there is no overall positive correlation between the A:T content of probes and the binding of RNA polymerase (Figure 3). Note that we had run the RNA polymerase experiment in order to compare the binding of RNA polymerase to the binding of FIS, H-NS and IHF (see below). This experiment gave the expected result, reported by Grainger *et al*. (27), with the largest signals for RNA polymerase binding located at genes encoding stable RNA and proteins essential for rapid growth (full data in Supplementary Table 1).

### Analysis of FIS, H-NS, IHF and RNA polymerase binding in non-coding DNA

Although <10% of the *E.coli* genome is non-coding, our results show that ~50% of the targets for FIS, H-NS and IHF are in these regions. To analyse further this binding, we aligned the ChIP-chip datasets for FIS, H-NS, IHF and RNA polymerase, removed data from probes that correspond to coding DNA, and scaled the datasets to have the same average Cy5/Cy3 ratio. We then applied the same stringent cut-off (FRD ~1%) to each dataset and counted the number of probes that passed the cut-off for different combinations of the various factors studied. We observed a clear correlation between the binding of FIS and H-NS. Thus, ~60% of the non-coding DNA loci associated with FIS were also bound with H-NS and vice versa (e.g. see Figure 4A). Additionally, ~50% of the non-coding targets for FIS and/or H-NS were also associated with RNA polymerase (Figure 4B). Overlap between the binding of IHF and FIS/H-NS was less frequent; only ~30% of the targets for FIS or H-NS were also targets for IHF (Figure 4C). A small number of targets passed the cut-off for FIS, H-NS and IHF (Figure 4D) and, in ~50% of these cases, RNA polymerase was also found to be associated (Figure 4E). Note that, at the regulatory regions of *ygfE* (Figure 4B) and *yahA* (Figure 4E), RNA polymerase is bound but appears not to move into the coding part of the gene.

### Analysis of FIS, H-NS and IHF binding in coding DNA

Datasets were aligned and analysed as in the previous section, except that results from probes corresponding to non-coding sections of the genome were discarded. Our expectation was that most binding targets would fall in transcriptionally silent parts of the genome and this was the case for IHF (e.g. see Figure 5A). However, we noted that a significant proportion of probes that pass the cut-off for FIS or H-NS binding (39 and 31%, respectively) correspond to locations at which RNA polymerase is bound. Thus, in contrast to IHF, FIS and H-NS may bind in both transcriptionally active (Figure 5B) and transcriptionally silent (Figure 5C) open reading frames. Note that the analysis also identified large sections of DNA that do not appear to be associated with high levels of FIS, H-NS, IHF or RNA polymerase (Figure 5D). We suggest that these regions may be associated with other factors.

To study directly the effect of transcription on the binding of FIS, H-NS and IHF, we measured their association with the *lac* operon genes, before and after induction with IPTG. To do this, antibodies directed against FIS, H-NS, IHF or the RNA polymerase β subunit were used to immunoprecipitate nucleoprotein samples isolated from cells that had been treated with formaldehyde, either before or after induction. Real-time PCR was used to quantify the relative amounts of different targets in the immunoprecipitates.

Figure 6A illustrates the time course of changes in RNA polymerase binding to targets in the *lacZ, lacY* and *lacA* genes after addition of IPTG. This shows that RNA polymerase is recruited to the *lac* operon within 60 s of inducer addition, and that RNA polymerase requires <2 min to transcribe the operon. This is consistent with rates of *lac*

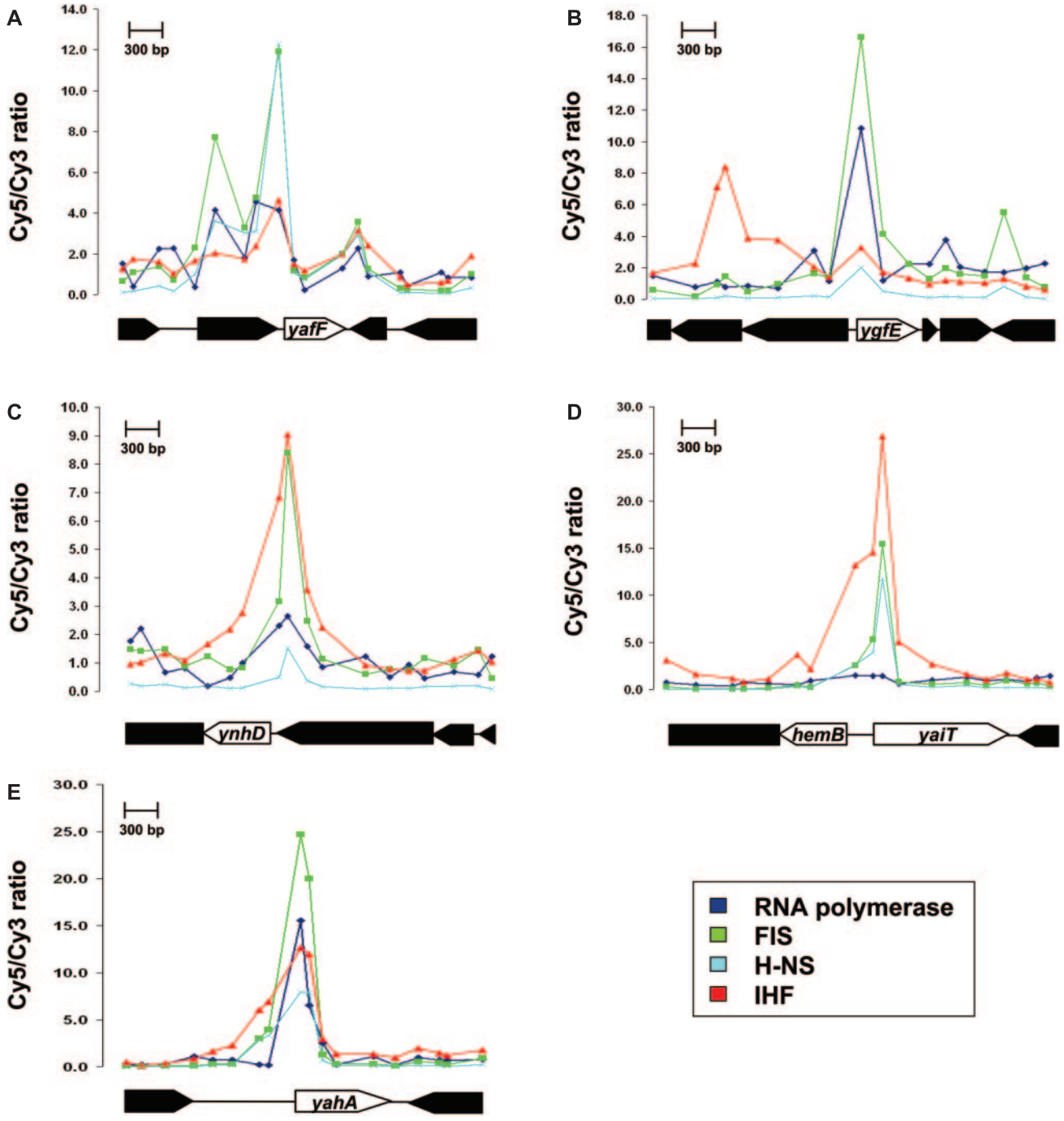


**Figure 4.** (**A**–**E**) Overlap of FIS, H-NS, IHF and RNA polymerase binding signals at some promoters. The figure shows the results from ChIP-chip experiments that measure profiles of binding of RNA polymerase, FIS, H-NS and IHF across the *E.coli* genome. Binding signals (*y*-axis) are plotted against the genomic location (*x*-axis) and data for six intergenic regions are shown. The binding signals for RNA polymerase, FIS, H-NS and IHF are illustrated in blue, green, cyan and red, respectively.

induction and transcription determined by measuring mRNA chain elongation (28). We then used the same approach to measure changes in the binding of IHF, H-NS and FIS, at the same three targets, 4 min after the addition of IPTG [note that the 5056 bp *lac* operon contains 209, 75 and 28 predicted DNA targets for IHF, H-NS and FIS, respectively (25)]. Results in Figure 6B show that, whilst transcription of the *lac* genes results in loss of IHF, the levels of associated FIS and H-NS increase.

### Changes in FIS and H-NS binding triggered by genome-wide changes in transcription

To investigate further the association of FIS and H-NS with transcribed regions of the *E.coli* chromosome, we measured the effects of salicylic acid on their distribution. Recall that the growth of *E.coli* is inhibited by salicylic acid, with concomitant down-regulation of highly transcribed genes and activation of genes in the *mar* (multiple antibiotic resistance) operon (27). Hence, ChIP-chip datasets for FIS and H-NS

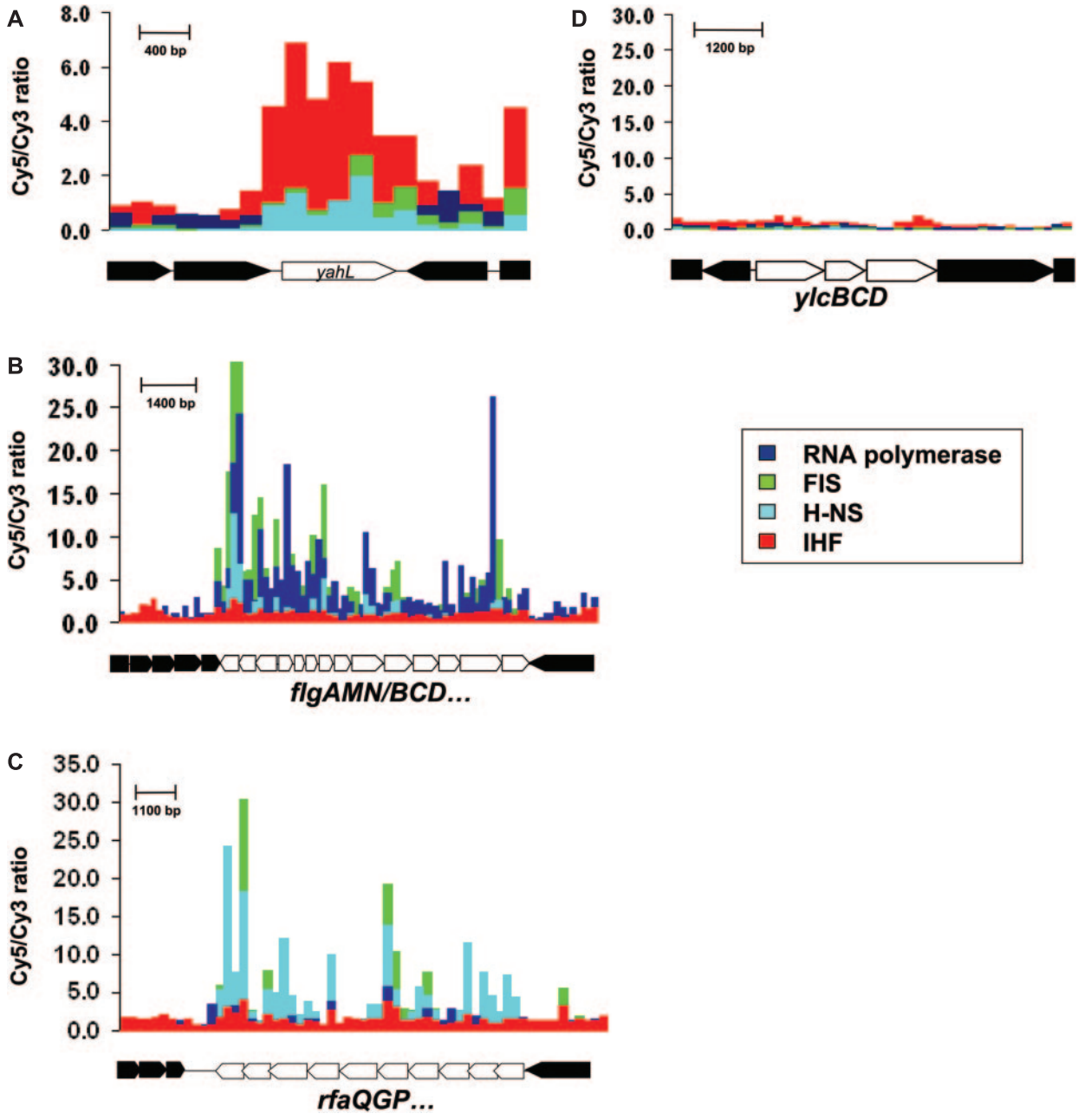


**Figure 5.** Association of FIS, H-NS, IHF and RNA polymerase with coding sections of the genome. The figure shows results from ChIP-chip experiments that measure profiles of binding of RNA polymerase, FIS, H-NS and IHF across the *E.coli* genome. Binding signals (*y*-axis) are plotted against the genomic location (*x*-axis) and data for six coding regions are shown. The binding signals for RNA polymerase, FIS, H-NS and IHF are illustrated in blue, green, cyan and red, respectively.

binding in MG1655 cells that had been treated with salicylic acid were obtained, and the complete results are included in Supplementary Table 1. To measure salicylic acid-induced changes in FIS and H-NS binding, these data were compared to the datasets illustrated in Figure 1A and B.

Despite complications caused by changes in Dps levels induced by salicylic acid, we observed a clear correlation between changes in the binding of RNA polymerase, FIS and H-NS at genes regulated in response to stress. Figure 7 shows salicylic acid induced changes in the binding of FIS and H-NS at the *mar* and *flg* operons, which are loci where transcription is up- and down-regulated by salicylic acid, respectively (27). Changes in the distribution of FIS and H-NS induced by salicylic acid correlate to changes in transcription. Thus, FIS and H-NS binding increases across the *mar* operon (Figure 7A), whilst binding across genes in the *flg* operon decreases (Figure 7B).

## Physical fractionation of the *E.coli* genome

Nagy *et al.* (3) reported that phenol–chloroform/aqueous two-phase mixtures could be used to fractionate formaldehyde cross-linked nucleoprotein complexes from *S.cerevisiae*. Thus, regions of the *S.cerevisiae* genome that are densely

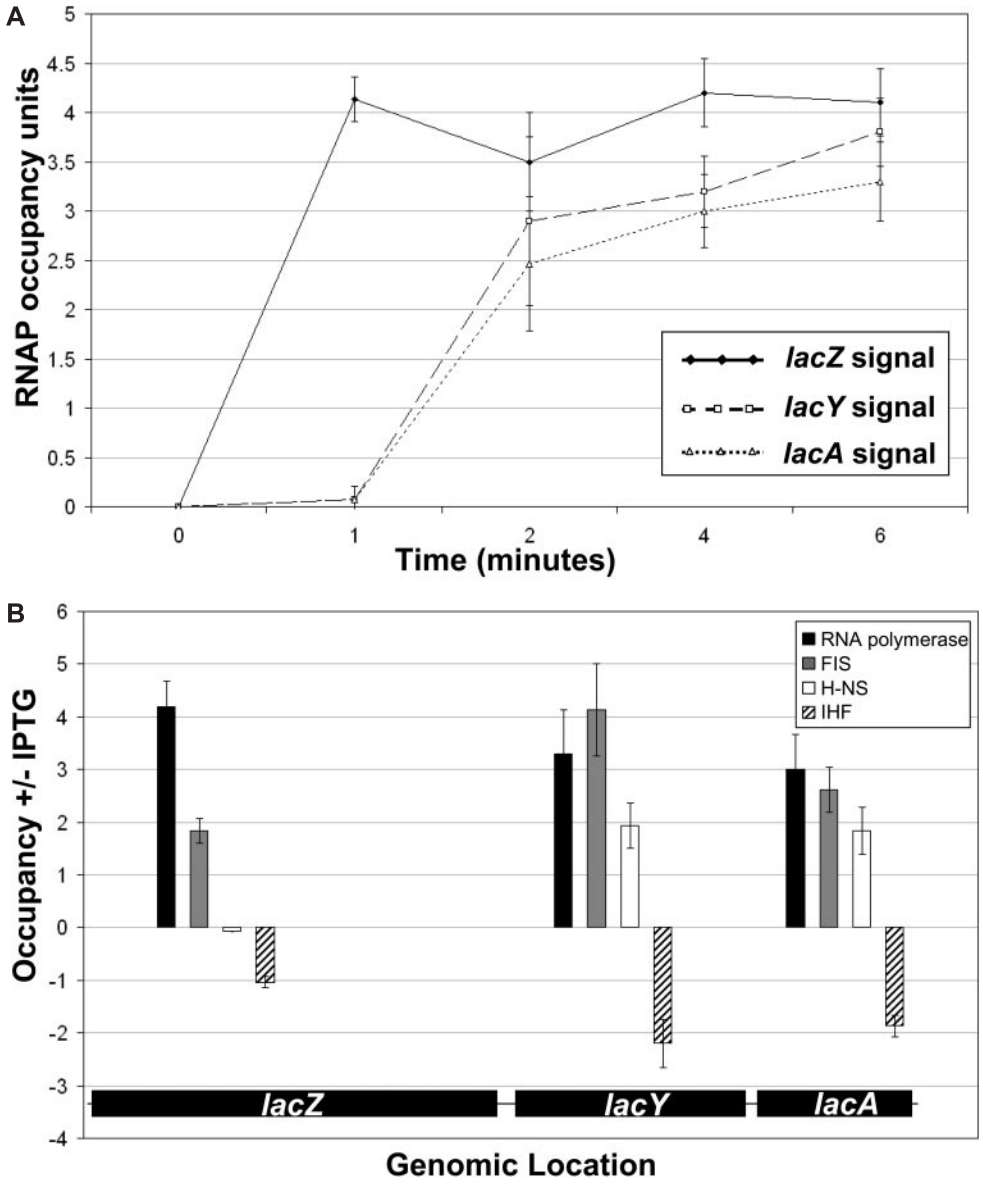


**Figure 6.** Effect of transcription on the binding of nucleoid-associated proteins to the *lac* operon. (**A**) RNA polymerase occupancy of *lacZ* (closed diamonds), *lacY* (open squares) or *lacA* (open triangles) measured by ChIP at intervals (shown on the *x*-axis) after the addition of IPTG to growing *E.coli* cultures. (**B**) RNA polymerase, FIS, IHF and H-NS binding to targets within *lacZ*, *lacY* and *lacA* measured by ChIP, 4 min after the addition of IPTG. Results are shown as ratios of the signal generated ± IPTG.

covered with histones are partitioned into phenol–chloroform, whilst distinct transcribed regions and promoters remain in the aqueous phase. Since our data suggested that nucleoid-associated proteins and RNA polymerase are organized differently in *E.coli*, we applied phenol–chloroform partition to total cross-linked nucleoprotein from growing MG1655 cells. To facilitate microarray analysis, DNA fragments collected from the aqueous phase were labelled with Cy5, whilst non-partitioned total DNA was Cy3-labelled as a control. The Cy5-labelled sample was mixed with the control Cy3-labelled sample and hybridized to the microarray. After washing and scanning, the Cy5/Cy3 signal intensity ratio was calculated for each probe. The complete dataset for this experiment is included in Supplementary Table 1. Scrutiny of the results shows that many DNA fragments associated with RNA polymerase are removed from the aqueous phase (Figure 8A). Thus, there is a correlation between RNA polymerase binding and partition into the phenol–chloroform phase. A limited correlation is also seen with fragments that bind FIS (Figure 8B). In contrast, fragments associated with H-NS and IHF are evenly partitioned (Figure 8C and D).

## DISCUSSION

This paper presents the first direct genome-wide analysis of the *in vivo* DNA-binding profiles of three *E.coli* nucleoid-associated proteins that are involved in chromosome folding and transcription regulation. Consistent with these roles, we observed binding at hundreds of targets across the *E.coli*

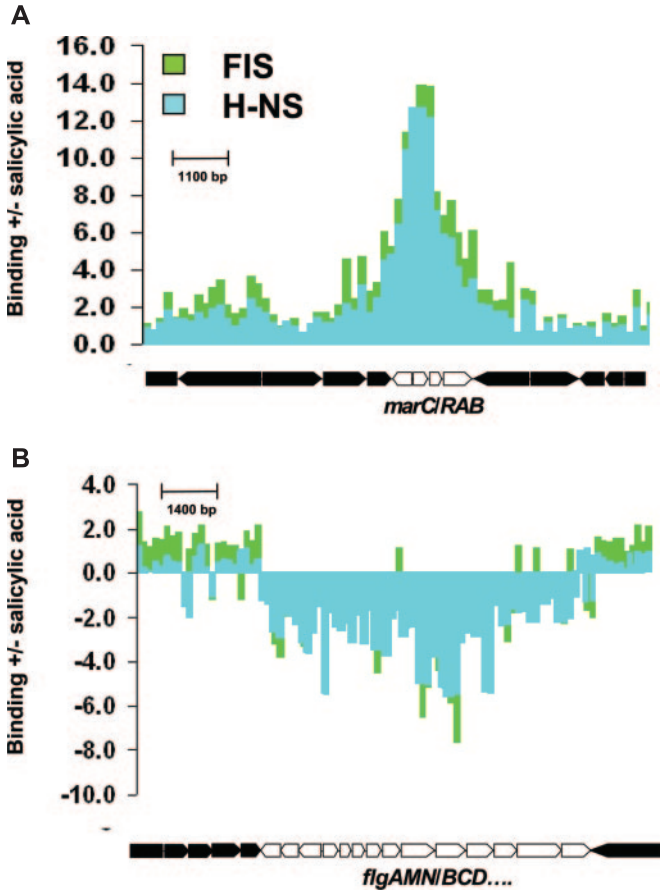


**Figure 7.** Changes in FIS and H-NS binding induced by salicylic acid. The figure shows results from ChIP-chip experiments that measure changes in the binding profiles of FIS and H-NS across the genome of *E.coli* cells that were treated with salicylic acid. The figure shows data for the *mar* operon (**A**) and the *flg* operon (**B**), which are activated and repressed by salicylic acid, respectively.

genome, both in coding and in non-coding DNA (Figure 1). All of the nucleoid-associated proteins tested were found to bind preferentially to A:T rich DNA, which is indicative of an association with specific A:T rich binding sites and a general association with promoter regions. DNA curvature, which is linked the % A:T content of the DNA, is thought to influence levels of H-NS binding. However, precise predictions concerning DNA curvature cannot be made on the basis of DNA sequence alone since, *in vivo*, curvature is likely to be modulated by DNA-binding proteins.

Discrete peaks for FIS, H-NS or IHF binding were evident at >80% of the previously established targets for each factor, and the range of signal intensities at these targets varied by up to 100-fold. We attribute this variation to differences in the occupation of each target by its cognate factor, but differences in the efficiency of formaldehyde cross-linking and immunoprecipitation may also contribute. The frequency and proximity of DNA targets for each of the factors, in combination with the continuum of signal intensities, made it difficult to differentiate between background noise and binding signals in many parts of the genome. To assess the distribution of each factor, we applied both a stringent and a relaxed cut-off to each of our datasets. Irrespective of the cut-off criterion adopted, ~50% of the targets for FIS, H-NS and IHF were found to be to be located in non-coding sections of the genome. Since <10% of the *E.coli* genome is non-coding, this represents a substantial bias towards intergenic regions. Binding of nucleoid proteins to defined regions of the chromosome may correspond to the location of barriers to topological domains (8,9).

Many of the DNA targets that we identified as being associated with H-NS also bind FIS and vice versa. Note that 8 of the 36 targets for H-NS, currently listed by the Ecocyc database, are also targets for FIS. Interestingly, RNA polymerase is also associated with many of these promoters and may be trapped, as predicted by Herring *et al.* (29). Recall that there are several well-documented examples of promoters being co-regulated antagonistically by FIS and H-NS, and RNA polymerase trapping by H-NS has been demonstrated *in vitro* (30). FIS and H-NS are most abundant during logarithmic growth (16) and thus may co-regulate the activity of many promoters in response to growth rate.

Approximately half of the binding targets for FIS, H-NS and IHF were found within genes, and binding at many of these locations may be linked to chromosome organization and DNA compaction. This was expected, but we were surprised to find that some intragenic targets for FIS and H-NS locate to sections of the genome associated with RNA polymerase. This may be indicative of RNA polymerase trapping or may occur because FIS and H-NS are removed from transcribed regions less efficiently than other proteins. However, our observation that FIS and H-NS binding to some regions is enhanced by transcription suggests that FIS and H-NS may associate with transcribed DNA. This association could be driven by changes in DNA accessibility and conformation, or by other proteins. Note that, in their comprehensive study of protein–protein interactions in *E.coli*, Butland *et al.* (31) found that FIS and H-NS co-purify with RNA polymerase subunits and ribosomes. Intriguingly, in rapidly growing *E.coli* cells, RNA polymerase, FIS and H-NS can form foci, and it is possible that these are locations in the cell where highly transcribed genes cluster (32,33). We speculate that FIS and H-NS participate in maintaining the co-localization of transcribed regions. Interestingly, the association of abundant DNA-binding proteins with transcribed DNA is not without precedent. For example, the *Drosophila* histone variant H3.3 has been shown to be recruited to transcribed DNA (6,34,35).

Our results underscore the differences in the organization of eukaryotic and prokaryotic nucleoprotein. In contrast to histones, the nucleoid-associated proteins that we tested do not appear to coat and compact large tracts of DNA, and their binding is biased to non-coding parts of the genome. Thus, RNA polymerase binding, rather than the binding of FIS, H-NS and IHF, drove the fractionation of *E.coli* nucleoprotein in phase separation experiments. Consistent with these observations, Zimmerman (36) found recently that removal of RNA from purified *E.coli* nucleoids had more noticeable effects on nucleoid compaction than the removal of proteins such as FIS, H-NS and IHF, suggesting that the main role of these proteins is regulatory. Additionally, Wade *et al.* (37) used LexA binding to probe the accessibility of DNA sites for LexA *in vivo* throughout the *E.coli* genome, concluding that bacterial nucleoprotein is 'permissive' to LexA binding and therefore fundamentally

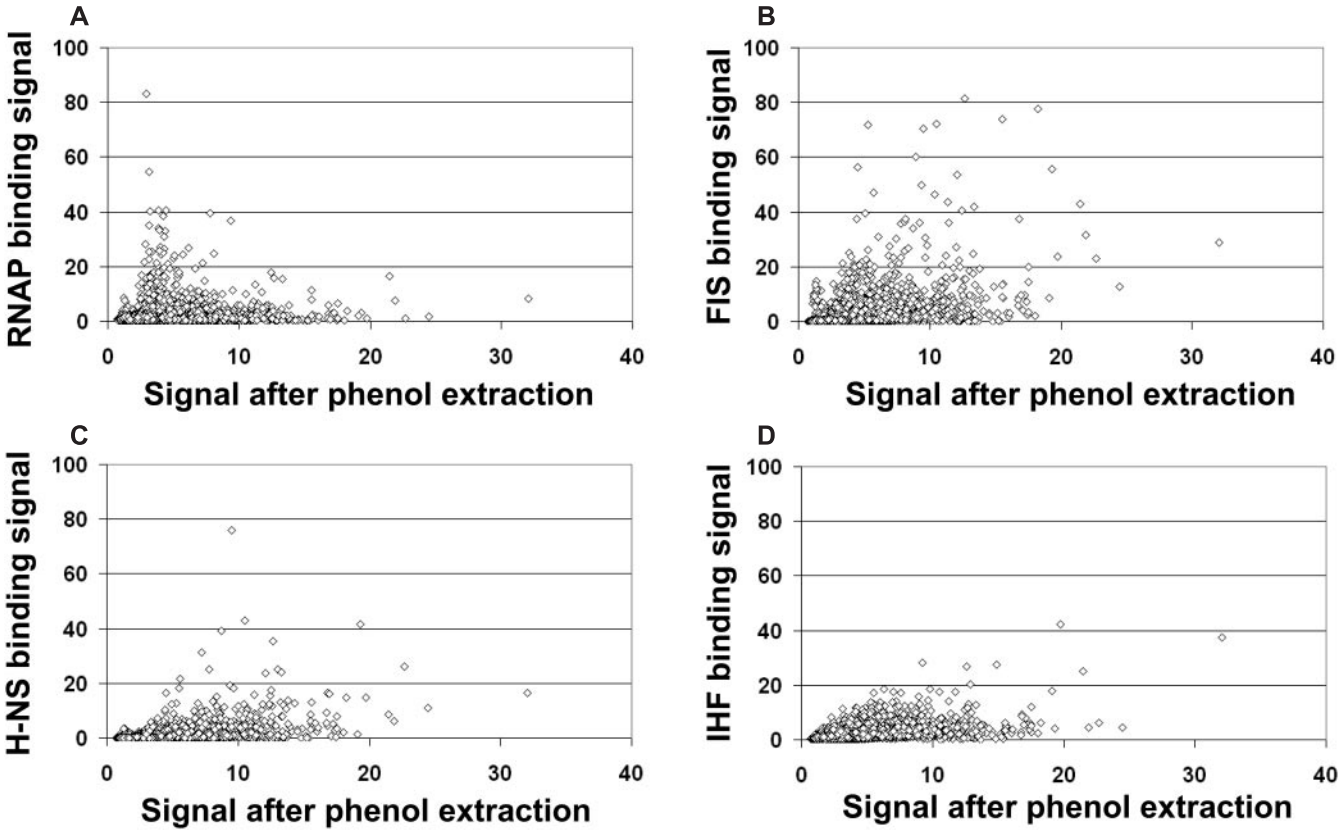


**Figure 8.** Fractionation of the *E.coli* genome. The figure illustrates the results of an experiment in which phenol–chloroform extraction was used to fractionate formaldehyde cross-linked *E.coli* nucleoprotein. A microarray was used to compare DNA present in the aqueous phase, following phenol–chloroform extraction, to a control 'input' DNA sample. The panel shows the relationship between signal intensity at each probe and the binding of RNA polymerase (**A**), FIS (**B**), H-NS (**C**) or IHF (**D**) at that locus measured by ChIP-chip.

different to the nucleoprotein of higher organisms. Further studies of H-NS distribution in *E.coli* (T. Oshima, manuscript in preparation), and similar studies of the *S.typhimurium* nucleoid (38,39), should enhance our understanding of bacterial chromosome organization and the role of nucleoid-associated proteins.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Jay Hinton, Steve Goodman and Akira Ishihama for supplying reagents and Joseph Wade, Sacha Lucchini, Jay Hinton, Taku Oshima and Naotake Ogasawara for helpful discussions. This work was supported by a Wellcome Trust programme grant. Funding to pay the Open Access publication charges for this article was provided by our Wellcome Trust programme grant.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Sekinger,E.A., Moqtaderi,Z. and Struhl,K. (2005) Intrinsic histone–DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell*, **18**, 735–448.
2. Yuan,G.C., Liu,Y.J., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J. and Rando,O.J. (2005) Genome-scale identification of nucleosome positions in *S.cerevisiae*. *Science*, **309**, 626–630.
3. Nagy,P.L., Cleary,M.L., Brown,P.O. and Lieb,J.D. (2003) Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl Acad. Sci. USA*, **100**, 6364–6369.
4. Eberharter,A. and Becker,P.B. (2002) Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep.*, **3**, 224–229.
5. Schwabish,M.A. and Struhl,K. (2004) Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.*, **24**, 10111–10117.
6. Wirbelauer,C., Bell,O. and Schubeler,D. (2005) Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev.*, **19**, 1761–1766.
7. Drlica,K. (1987) The nucleoid. In Neidhardt,F. (ed.), *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*. ASM Press, Washington DC, Vol. 1, pp. 91–103.
8. Dame,R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.*, **56**, 858–870.
9. Travers,A. and Muskhelishvili,G. (2005) Bacterial chromatin. *Curr. Opin. Genet. Dev.*, **15**, 507–514.
10. 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.
11. Dame,R.T., Wyman,C. and Goosen,N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.*, **28**, 3504–3510.

12. Ali,B.M., J, Amit,R., Braslavsky,I., Oppenheim,A.B., Giladi,O. and Stavans,J. (2001) Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc. Natl Acad. Sci. USA*, **98**, 10658–10663.
13. Shoko,D., Yan,J., Johnson,R.C. and Marko,J.F. (2005) Low-force DNA condensation and discontinuous high-force decondensation reveal loop-stabilising function of the protein Fis. *Phys. Rev. Lett.*, **95**, 208101.
14. Williams,R.M. and Rimsky,S. (1997) Molecular aspects of the *E.coli* nucleoid protein, H-NS: a central controller of gene regulatory networks. *FEMS Microbiol. Lett.*, **156**, 175–185.
15. Schneider,D.A., Ross,W. and Gourse,R.L. (2003) Control of rRNA expression in *Escherichia coli*. *Curr. Opin. Microbiol.*, **6**, 151–156.
16. Azam,T.A., Iwata,I., 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.
17. Pan,C.Q., Finkel,S.E., Cramton,S.E., Feng,J., Sigman,D.S. and Johnson,R.C. (1996) Variable structure of Fis–DNA complexes determined by flanking DNA–protein contacts. *J. Mol. Biol.*, **264**, 675–695.
18. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF–DNA complex: a protein induced DNA U-turn. *Cell*, **87**, 1295–1306.
19. Spurio,R., Falconi,M., Brandi,A., Pon,C.L. and Gualerzi,C.O. (1997) The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *EMBO J.*, **16**, 1795–1805.
20. Hommais,F., Krin,E., Laurent-Winter,C., Soutourina,O., Malpertuy,A., Le Caer,J.P., Danchin,A. and Bertin,P. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol. Microbiol.*, **40**, 20–36.
21. Dorman,C.J. (2004) H-NS: a universal regulator for a dynamic genome. *Nature Rev. Microbiol.*, **2**, 391–400.
22. Kelly,A., Goldberg,M.D., Carroll,R.K., Danino,V., Hinton,J.C. and Dorman,C.J. (2004) A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica serovar Typhimurium*. *Microbiology*, **150**, 2037–2053.
23. Paul,B.J., Ross,W., Gaal,T. and Gourse,R. (2004) rRNA transcription in *Escherichia coli*. *Annu. Rev. Genet.*, **38**, 794–770.
24. Keseler,I.M., Collado-Vides,J., Gama-Castro,S., Ingraham,J., Paley,S., Paulsen,I.T., Peralta-Gil,M. and Karp,P.D. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.*, **33**, 334–337.
25. Robison,K., McGuire,A. and Church,G.A. (1998) Comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.*, **284**, 241–254.
26. Mangan,M.W., Lucchini,S., Danino,V., Croinin,T.O., Hinton,J.C. and Dorman,C.J. (2006) The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica serovar Typhimurium*. *Mol. Microbiol.*, **59**, 1831–1847.
27. Grainger,D.C., Hurd,D., Harrison,M., Holdstock,J. and Busby,S.J.W. (2005) Studies of the distribution of cAMP receptor protein and RNA polymerase along the *Escherichia coli* chromosome. *Proc. Natl Acad. Sci. USA*, **102**, 17693–17698.
28. Vogel,U., Sorensen,M., Pedersen,S., Jensen,K.F. and Kilstrup,M. (1992) Decreasing transcription elongation rate in *Escherichia coli* exposed to amino acid starvation. *Mol. Microbiol.*, **6**, 2191–200.
29. Herring,C.D., Raffaelle,M., Allen,T.E., Kanin,E.I., 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.
30. 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**, 2146–50.
31. Butland,G., Peregrin-Alvarez,J.M., Li,J., Yang,W., Yang,X., Canadien,V., Starostine,A., Richards,D., Beattie,B., Krogan,N. *et al.* (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, **433**, 531–537.
32. Azam,T.A., Hiraga,S. and Ishihama,A. (2000) Two types of localisation of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells*, **5**, 613–626.
33. Cabrera,J.E. and Jin,D.J. (2003) The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol. Microbiol.*, **50**, 1493–1505.
34. Schwartz,B.E. and Ahmad,K. (2005) Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev.*, **19**, 804–814.
35. Daury,L., Chailleux,C., Bonvallet,J. and Trouche,D. (2006) Histone H3.3 deposition at E2F-regulated genes is linked to transcription. *EMBO Rep.*, **7**, 66–71.
36. 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.
37. 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.
38. Navarre,W.W., Porwollik,S., Wang,Y., McClelland,M., Rosen,H., Libby,S.J. and Fang,F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science*, **313**, 236–238.
39. Lucchini,S., Rowley,G., Goldberg,M.D., Hurd,D., Harrison,M. and Hinton,J.C. (2006) H-NS mediates the silencing of laterally-acquired genes in bacteria. *PloS Pathogens,* in press.