

# Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12

Ana I. Prieto<sup>1,2</sup>, Christina Kahramanoglou<sup>1,3</sup>, Ruhi M. Ali<sup>2,4</sup>, Gillian M. Fraser<sup>1,\*</sup>,  
Aswin S. N. Seshasayee<sup>2,5,\*</sup> and Nicholas M. Luscombe<sup>2,6,\*</sup>

<sup>1</sup>Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP,

<sup>2</sup>EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD,

<sup>3</sup>MRC-National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK, <sup>4</sup>Centre for Biotechnology, Anna University, Chennai 600 035, <sup>5</sup>National Centre for Biological Sciences, GKVK, Bellary Road, Bangalore 560 065, India and <sup>6</sup>Okinawa Institute of Science and Technology, Tancha, Onna-son, Kunigami-gun, Okinawa 904-0495, Japan

Received October 18, 2011; Revised November 24, 2011; Accepted November 28, 2011

## ABSTRACT

IHF and HU are two heterodimeric nucleoid-associated proteins (NAP) that belong to the same protein family but interact differently with the DNA. IHF is a sequence-specific DNA-binding protein that bends the DNA by over 160°. HU is the most conserved NAP, which binds non-specifically to duplex DNA with a particular preference for targeting nicked and bent DNA. Despite their importance, the *in vivo* interactions of the two proteins to the DNA remain to be described at a high resolution and on a genome-wide scale. Further, the effects of these proteins on gene expression on a global scale remain contentious. Finally, the contrast between the functions of the homo- and heterodimeric forms of proteins deserves the attention of further study. Here we present a genome-scale study of HU- and IHF binding to the *Escherichia coli* K12 chromosome using ChIP-seq. We also perform microarray analysis of gene expression in single- and double-deletion mutants of each protein to identify their regulons. The sequence-specific binding profile of IHF encompasses ~30% of all operons, though the expression of <10% of these is affected by its deletion suggesting combinatorial control or a molecular backup. The binding profile for HU is reflective of relatively non-specific binding to the chromosome, however, with a preference for A/T-rich DNA. The

HU regulon comprises highly conserved genes including those that are essential and possibly supercoiling sensitive. Finally, by performing ChIP-seq experiments, where possible, of each subunit of IHF and HU in the absence of the other subunit, we define genome-wide maps of DNA binding of the proteins in their hetero- and homodimeric forms.

## INTRODUCTION

Nucleoid-associated proteins (NAPs) are considered to be global regulators of gene expression in bacteria. They alter the topology of bound DNA by bending, bridging or wrapping it, leading to multiple effects on the bacterial cell including transcriptional regulation (1). Studies of 12 types of NAPs in *Escherichia coli* showed that they are generally expressed at high levels, and differ from each other in their expression across the growth phase and the degree of sequence specificity (2,3). The global nature of the effects of NAPs on bacterial physiology has prompted several genome-scale studies of their binding and transcriptional effects in *E. coli* and *Salmonella enterica*; these have sometimes led to intriguingly conflicting conclusions on the functions of NAPs, thus underscoring their complexity (4–10).

Two NAPs, IHF and HU, are composed of two homologous subunits each (IhfA and IhfB; HupA and HupB). They are both members of the DNABII family of DNA-binding proteins and are strikingly similar to each other in sequence and in their unique structural fold (11).

\*To whom correspondence should be addressed. Tel: +91 98453 77507; Fax: +91 80 23636662; Email: aswin@ncbs.res.in  
Correspondence may also be addressed to Gillian M. Fraser. Tel: +44 1223 492 572; Fax: +44 1223 494 468; Email: gmf25@mole.bio.cam.ac.uk  
Correspondence may also be addressed to Nicholas M. Luscombe. Tel: +44 1223 330 245; Fax: +44 1223 333 346; Email: luscombe@ebi.ac.uk

However, the similarities end there: they differ in their sequence specificity, with IHF being sequence-specific and HU binding at low affinity along the chromosome (2,12) with some specificity toward gapped or nicked DNA (13–16). Whereas the ability of each subunit of HU to form homodimers and bind to the DNA in such a form is relatively well-established (17,18), such evidence is less clear for IHF (19,20). Moreover, the two proteins differ in the degree of conservation across bacteria: whereas at least one subunit of HU is found across most bacterial genomes making it the most conserved NAP, IHF has a more restricted occurrence. Their functions have been described to include regulation of transcription, replication and recombination via DNA binding (1) and extend to the control of translation initiation by HU via protein–RNA interactions (21,22).

Several molecular and genome-scale studies have investigated the role of IHF in transcriptional control. Notable among such studies are the description of its effects on the *nir* (23) and the *fim* (24,25) operons, wherein IHF represses the *nir* and activates the *fim* operon. Also remarkable is the role of IHF in helping the formation of activation loops at enhancer-dependent promoters (26). Compilation of results from molecular studies—performed under diverse conditions—by the curators of the RegulonDB database (27) identified over 150 genes as being regulated at the transcriptional level by IHF, with over two-thirds activated by IHF. A very early microarray study (28), primarily emphasizing technical aspects of data analysis, identified genes that are differentially regulated in an *ΔihfA* strain grown in MOPS minimal medium; however, it must be noted that the strain on which the experiment was performed (a derivative of K12 CP79) was different from that for which the microarray was designed (K12 MG1655). In *Salmonella enterica* Typhimurium, deletions of *ΔihfA*, *ΔihfB* and both *ΔihfA* and *ΔihfB* each led to different effects on transcription during growth in rich LB medium, thus suggesting distinct binding tendencies of the IhfA2 and IhfB2 homodimers and the IhfAB heterodimer (29). The number of genes responding transcriptionally to *Δihf* is substantially higher in *Salmonella* than reported in *E. coli*; these genes include virulence determinants in *Salmonella*. Finally, a genome-scale study of IHF binding to the *E. coli* genome using low-resolution microarrays showed a preference for the binding regions to be located in non-coding DNA (5).

Despite the near universal conservation of HU in the bacterial kingdom, only recently have genome-scale studies been performed to investigate its effect on gene expression. This is in spite of molecular studies investigating its role in controlling gene expression at specific loci, most notably the stabilization of the repression loop at the *gal* promoter (30). One study performed clustering analysis of microarray data obtained for *ΔhupA*, *ΔhupB* and *ΔhupA/ΔhupB* (*ΔhupAB*) strains during exponential, transition and stationary phases of growth, thus identifying distinct HupA2, HupB2 and HupAB regulons comprising genes used in energy metabolism, SOS response and osmolarity and acidic stress response (31). In spite of the established effect of HU on the supercoiled

state of the DNA, these authors found little association between genes comprising the HU regulon and those that respond to DNA supercoiling. Again however, this experiment was performed on a strain of *E. coli* (C600) which was not the same as that based on which (MG1655) the microarray was designed. A more recent microarray study of the double-deletion strain showed that genomic loci encoding HU-responsive genes tend to display high gyrase binding and therefore supercoiling sensitivity (32). Finally, in *Salmonella*, distinct regulons were identified for the three dimeric forms of HU, such that dissimilar sets of genes were differentially expressed during different phases of growth (33). To our knowledge, though HU is a major NAP, no study has investigated its *in vivo* binding to the chromosome on a genomic scale.

Despite the above studies, the binding characteristics of the two proteins have not been described at a high resolution and on a genome-wide scale. Further, as evident from the conflicting results of previous studies, the effects of these proteins on gene expression on a global scale remain a contentious issue. Finally, the contrast between the functions of the homo- and heterodimeric conformations of these proteins remains poorly understood and deserves the attention of further study. Here, we present a genome-scale study of the binding characteristics of HU and IHF to the *E. coli* K12 chromosome at four different time-points during batch growth in LB, using chromatin-immunoprecipitation coupled to high-throughput sequencing (ChIP-seq). We also perform microarray analysis of gene expression in single- and double-deletion mutants of each protein, to identify their regulons. Finally, by performing ChIP-seq experiments where possible, of each subunit of IHF and HU in the absence of the other subunit, we define genome-wide maps of DNA binding of the proteins in their hetero- and homodimeric forms.

## METHODS

### Strains and general growths conditions

The *E. coli* K-12 MG1655 bacterial strains used in this work are listed in Supplementary Table 1. Luria–Bertani (0.5% NaCl) broth and agar (15 g l<sup>-1</sup>) were used for routine growth. Where needed, ampicillin, kanamycin and chloramphenicol were used at final concentrations of 100, 30 and 30 μg ml<sup>-1</sup>, respectively.

### Construction of *E. coli* MG1655 knock-outs and FLAG-tagged strains

Disruption of *ihf* and *hup* genes in the *E. coli* chromosome was achieved by the λ Red recombination system (34), as previously described by Baba *et al.* (35). Primers designed for this purpose are shown in Supplementary Table 2. Sets of additional external primers were used to verify the correct integration of the PCR fragment by homologous recombination (Supplementary Table 3). The cassette was then removed by FLP-mediated site-specific recombination. Double-deletion strains were made by P1 transduction (36).

The 3xFLAG epitope was added at the C terminus of the IhfA, IhfB, HupA and HupB proteins by a PCR-based method with plasmid pSUB11 as template (37). Primers used for introducing the 3xFLAG tag are shown in Supplementary Table 2. The tagged construct was then introduced onto the chromosome of *E. coli* MG1655 using the  $\lambda$  Red recombinase system. At each stage, DNA and strain constructions were confirmed by PCR and/or sequencing. This approach resulted in the introduction of a kanamycin resistance cassette in the chromosome downstream of the tagged gene. The cassette was then removed by FLP-mediated site-specific recombination.

### RNA extraction and microarrays

To prepare cells for RNA extraction, 100 ml of fresh LB was inoculated 1:200 from an overnight culture in a 250 ml flask and incubated with shaking at 180 rpm in a New Brunswick C76 waterbath at 37°C. Two biological replicates were performed for each strain and samples were taken at exponential, late exponential, early stationary and stationary phase. The cells were pelleted by centrifugation (10 000g, 10 min, 4°C), washed in 1xPBS and pellets were snap-frozen and stored at -80°C until required. RNA was extracted using Trizol Reagent (Invitrogen) according to the manufacturer's protocol until the chloroform extraction step. The aqueous phase was then loaded onto mirVanaTM miRNA Isolation kit (Ambion) columns and washed according to the manufacturer's protocol. Total RNA was eluted in 50  $\mu$ l of RNAase free water. The concentration was then determined using a Nanodrop ND-1000 machine (NanoDrop Technologies), and RNA quality was tested by visualization on agarose gels and by Agilent 2100 Bioanalyser (Agilent Technologies, Palo Alto, CA, USA).

For the generation of fluorescence-labeled cDNA, we used the FairPlay III Microarray Labelling Kit (Stratagene). Briefly, 1  $\mu$ g of total RNA was annealed to random primers, and cDNA was synthesized in a reverse transcription reaction with an amino allyl modified dUTP. The amino allyl labeled cDNA was then coupled to a Cy3 dye (GE Healthcare) containing a NHS-ester leaving group. The labeled cDNA was hybridized to the probe DNA on microarrays by incubating at 65°C for 16 h. The unhybridized labeled cDNA was removed and the hybridized labeled cDNA was visualized using an Agilent Microarray Scanner.

### Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as previously described (4,38).

### Real-time qPCR

To measure the enrichment of the IhfA, IhfB, HupA, HupB or RNAP-binding targets in the immunoprecipitated DNA samples, real-time qPCR was performed using a MJ Mini thermal cycler (Bio-Rad). About 1  $\mu$ l of IP or mock-IP DNA was used with specific primers to the promoter regions (primer sequences

in Supplementary Table 3; results in Supplementary Table 4) and Quantitect SYBR Green (QIAGEN).

### RT-PCR for validation

To validate the results of the microarray analysis, quantitative reverse-transcriptase PCR (qRT-PCR) was carried out using specific primers to the mRNA targets showing up- or down-regulation, and control targets not showing differential expression (primer sequences in Supplementary Table 5; results in Supplementary Tables 6 and 7). RNA was extracted as described above from wild type,  $\Delta$ ihfA,  $\Delta$ ihfB,  $\Delta$ ihfAB,  $\Delta$ hupA,  $\Delta$ hupB and  $\Delta$ hupAB cells and 30 ng total RNA was used with the Express One-Step SYBR GreenER kit (Invitrogen) according to the manufacturer's guidelines, using a MJ Mini thermal cycler (Bio-Rad).

### Library construction and Solexa sequencing

Prior and post library construction, the concentration of the immunoprecipitated DNA samples was measured using the Qubit HS DNA kit (Invitrogen). Library construction and sequencing was done using the ChIP-Seq Sample Prep kit, Reagent Preparation kit and Cluster Station kit (Illumina). Samples were loaded at a concentration of 10 pM.

### Public data sources

The *E. coli* K12 MG1655 genome was downloaded from the KEGG database and gene coordinate annotations from the Ecocyc 11.5 database (39). Literature-derived transcriptional regulatory network and a list of operons were sourced from the RegulonDB 6.2 database (27). List of genes bound by IHF was obtained from Grainger *et al.* (5). ChIP-chip signals for DNA gyrase were obtained from Jeong *et al.* (40). Functional category annotation data for *E. coli* K12 MG1655 was obtained from the COG database. RNA-seq data was obtained from our previous publication (4).

### Analysis of genomic data

Reads obtained from the Illumina Genome Analyzer were mapped to both strands of *E. coli* K12 MG1655 genome using BLAT (41), as described previously. Binding regions for IHF were calculated using the per-base read count distribution as performed earlier (4); in addition to a statistical enrichment (binomial test) in the ChIP signal over the mock-IP (as proposed by PeakSeq) (42), we imposed a further 1.5-fold increase in the absolute signal. We also used the Bioconductor package BayesPeak (43,44) to identify binding regions for IHF and HU.

For HU, we calculated two gene-level measures of binding signal: (i) the highest read count obtained between -150 and +20 of the ORF and (ii) the median read count across the ORF body; the two measures provide equivalent results. In addition, we adapted a method used previously to analyse data from ChIP-chip experiments for nucleoporins in *Drosophila melanogaster*, to identify regions of enriched signal for HU (45). This adapted method calculates differences in log

(base 2)-transformed read count signals over 400 nt windows between the ChIP sample and the mock-IP sample, following normalization by DESeq (46). The left hand side of this distribution, plus its mirror image around the mode gives the null distribution. All data points over the 95th percentile of the null distribution were considered as representing significant binding in the sample.

Binding motifs were identified using the MEME software and subsequently, binding regions scanned for the occurrence of the motif using MAST (47). An operon was defined as bound by the protein of interest if at least 50 bp of the intergenic region upstream of the operon overlapped with a binding region. For long intergenic regions, only the first 400 bp immediately upstream of the operon were used.

Gene expression analyses were performed on a previously described custom-designed isothermal Agilent microarray platform, and analyzed as described earlier (4). Briefly, array data were background corrected using normexp (48) and normalization performed using VSN (49). Differential expression in the deletion strains compared with the wild-type was called at FDR-adjusted  $P$ -value of 0.05, and a fold change of at least two.

All statistical tests were carried out using  $R$ .

## RESULTS

### DNA-binding properties of IHF and HU subunits

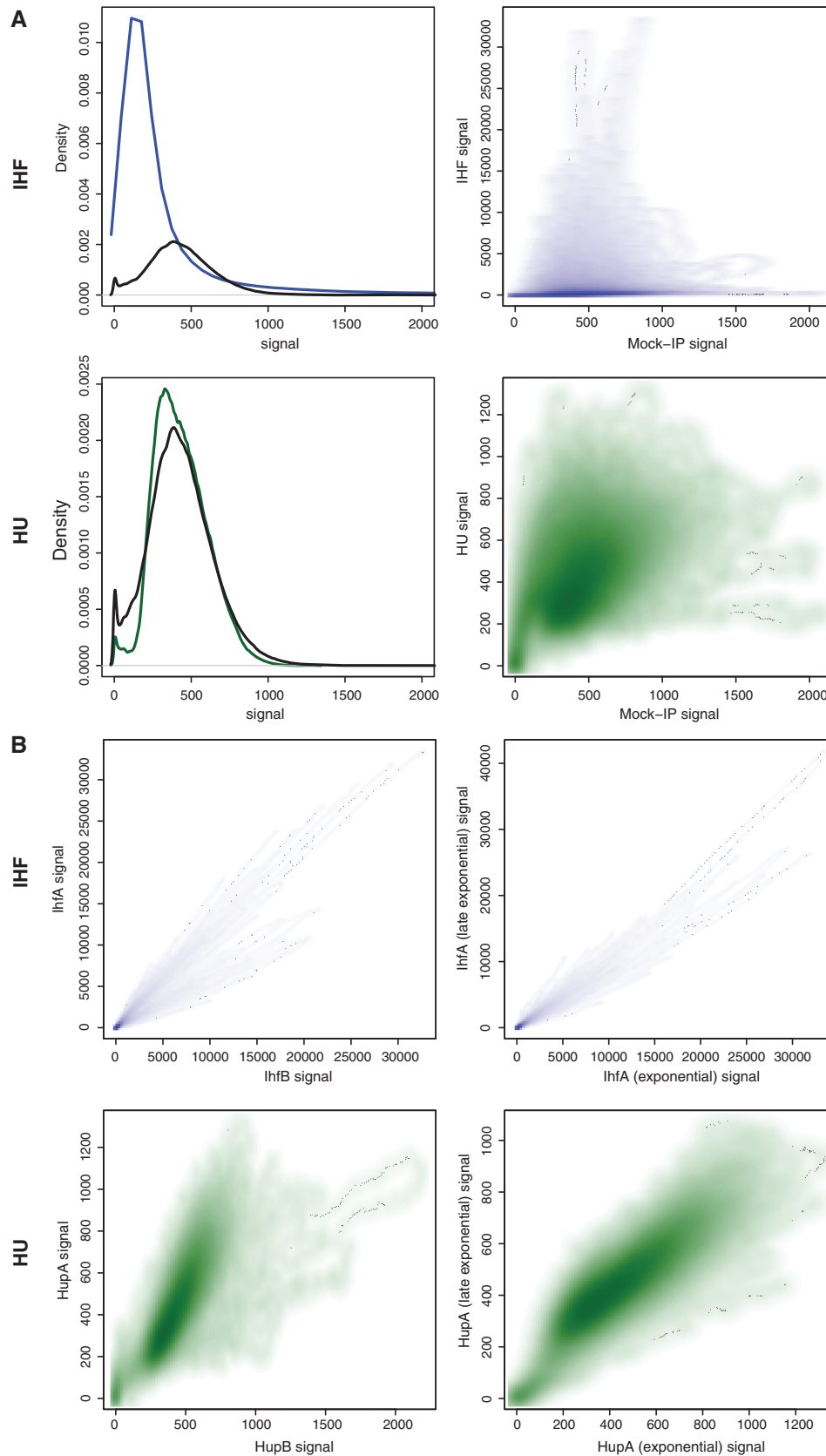
To study the binding characteristics of IHF and HU to the *E. coli* chromosome, we performed immunoprecipitation of each protein subunit—during mid-exponential, late-exponential, transition-to-stationary and stationary phases of growth—and sequenced the cross-linked DNA using an Illumina Genome Analyzer system (ChIP-seq). We also used control data from a mock-IP experiment (for mid-exponential phase) described in our previous study (4). We mapped the short sequence reads obtained from each sequencing experiment to the *E. coli* K12 MG1655 genome (KEGG ID: eco). For each sample, we then obtained a read count distribution, quantified by the number of reads that map to each base position on the chromosome.

We inspected the nature of the read count distributions by plotting their densities (Figure 1A). The distributions for the various IHF samples each had a heavy right tail corresponding to regions of specific binding. On the other hand, the distributions for HU were only slightly skewed to the right, and in this respect similar to that from the mock-IP experiment. Whereas the read counts obtained for IHF were only weakly correlated to the mock-IP control ( $\rho = 0.12$  for IhfA, mid-exponential phase sample; the weak correlation presumably arising from a systemic background), those for HU showed a more significant correlation with the mock-IP ( $\rho = 0.47$  for HupA, mid-exponential phase; Figure 1A). Similarly, plots of the distribution of mock-IP subtracted signal for HU (following division of read counts by the total number of reads obtained in that sample, and log transformation) was centered around zero with a relatively weak right-sided tail (Figure 2A). On the other hand, this distribution for

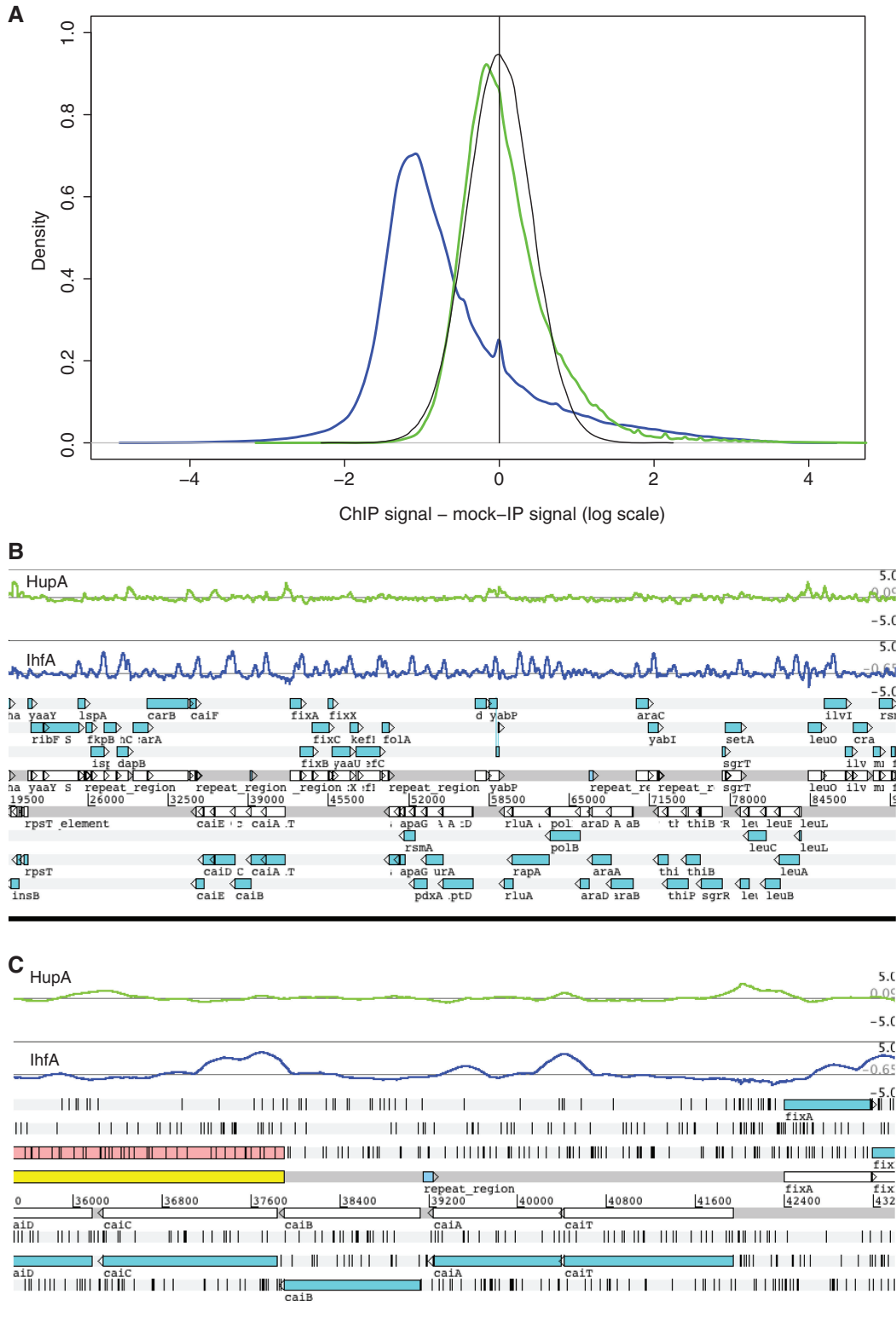
IHF was offset from zero with a peak well-below zero representing most of the genome with little or no binding, and those to the right corresponding to regions of enriched signal. Despite the strong resemblance of the HU data to the mock-IP, our HU experiment is representative of the protein's DNA binding profiles for the following reasons: (i) there is a considerable right-sided tail to the mock-IP-subtracted HU ChIP-seq signal; (ii) the read count profile for each HU subunit is more correlated with that for the other subunit ( $\rho = 0.83$ ) than with that for the mock-IP ( $\rho = 0.47$  and  $0.59$  for HupA and HupB, respectively; Figure 1B; Supplementary Figure S1); (iii) the profile at any given time-point is also more strongly correlated with that from the adjacent time-point than with the mock-IP profile ( $\rho = 0.88$  for HupA between exponential and late exponential phases; Figure 1B and Supplementary Figure S2); (iv) ChIP experiments for HU are reproducibly successful, unlike that for the mock-IP which typically provides very low concentrations of DNA not always sufficient for a sequencing reaction. Taken together, these provide a genome-wide, high-resolution, *in vivo* validation of prior molecular data suggesting that IHF binds DNA in a sequence-specific manner whereas HU binds *more* uniformly.

The strongly right-tailed distribution for IHF allowed us to identify regions of enriched signal—or binding regions—using a stringent version (Methods) of a procedure described earlier (4). Over 85% of the 1042 (1022) binding regions thus obtained for IhfA (and IhfB) overlap with those obtained using another published method BayesPeak (43,44). We noted a similar agreement between our method and another previously used in our lab to detect binding regions from eukaryotic ChIP-chip/seq experiments (45). In general, the signal enrichment in these IHF-bound regions is significantly higher than that for another sequence-specific, yet promiscuous, NAP: FIS (Figure 3A). During exponential phase, IHF-bound regions (either subunit) cover 13% of the genome, including upstream regions of 443 operons (17%). Genes identified as bound by either subunit of IHF during the two exponential-phase time-points in our study cover 68% of those identified in an earlier publication using mid-resolution ChIP-chip microarrays (5). We also recovered the known binding motif for IHF from these data (Figure 3B). We detected 2999 and 3162 occurrences of this motif within the binding regions of IhfA and IhfB, respectively. Of these motifs, <10% is localized to regions upstream of predicted operons. This proportion is small compared to our previous data for Fis for which over 20% of the binding regions fell upstream of operons. This is in line with the smaller number of bound operons (based on binding to upstream regions) per mega base pair of bound DNA for IHF when compared to Fis (approximately 720 operons per mega base pair of binding region for IHF, compared to approximately 1250 operons per mega base pair for Fis).

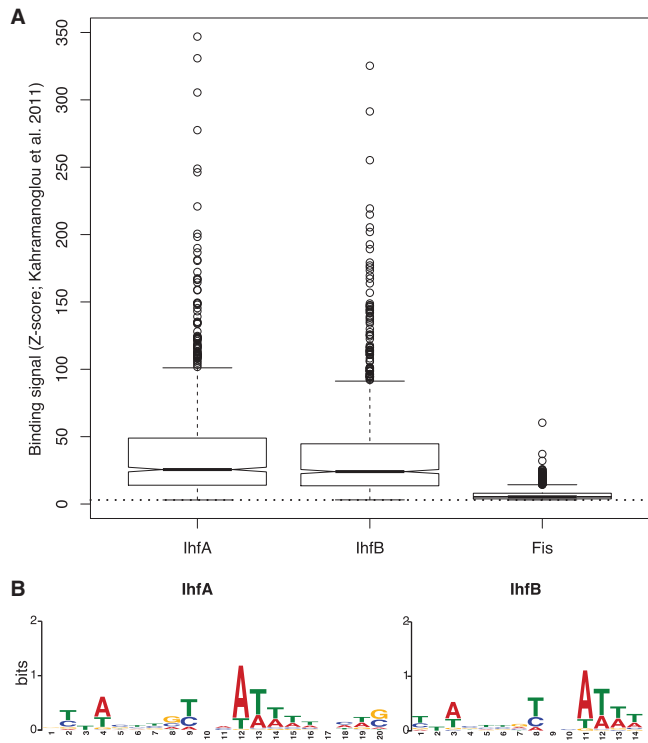
Because the binding profile from our HU ChIP-seq experiment shows a strong resemblance to that from the mock-IP with relatively weak signals (Figure 2A–C), we used two methods to characterize its binding. First, we obtained a HU occupancy measure for each gene in the



**Figure 1.** (A) The left panels show the distribution of read counts ( $x$ -axis truncated at 20) for IHF (blue), HU (green) and the mock-IP (black). The right panels show the correlation (at single-base resolution) in read counts between the mock-IP ( $x$ -axis) and HupA (green) or IhfA (blue) ChIP-seq ( $y$ -axis). (B) The left panels show the base-level correlation in read counts between IhfA and IhfB (blue), and HupA and HupB (green). The above are all for exponential phase data. The right panels show similar correlations for the same protein, but between the two exponential phase time-points. The signal is the number of reads mapping to a given base position divided by the total number of reads in the sample (multiplied by a factor  $10^7$ ).

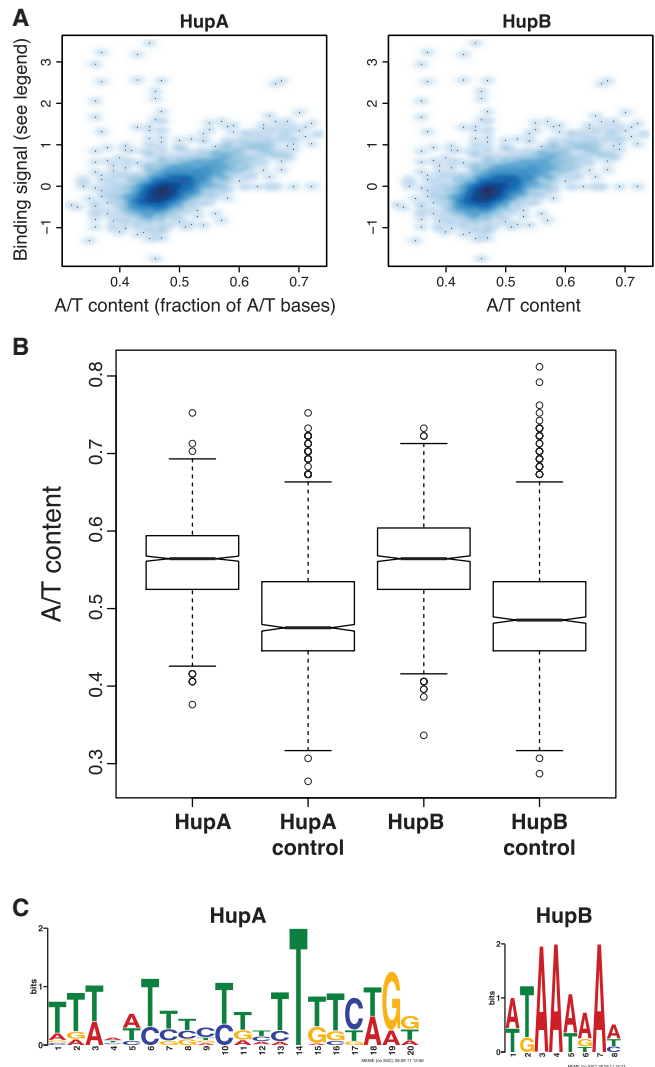


**Figure 2.** (A) Comparison of mock-IP-subtracted binding signals (log-scale of the number of reads mapping to a given position normalized by the total number of reads from that sequencing experiment multiplied by a factor  $10^7$ ) for IhfA (blue) and HupA (green). For HupA, the distribution is centered around zero with a slight right tail; this is similar to what might be expected when a simulated replicate of the mock-IP signal is subtracted from the reference mock-IP (black line; where each data point for the *in silico* replicate is derived from a normal distribution with mean equal to the signal on the mock-IP and standard deviation equal to that across the mock-IP dataset). On the other hand, for IhfA, the distribution has a strong offset from zero, with many points being below zero indicating lack of binding and a considerable number above zero indicating strong signal. (B) Tracks—rendered in Artemis—showing binding signals for IHF (blue) and HU (green) across an ~70 kb region of the genome. (C) A zoomed-in image of a portion of B, showing an ~8 kb region of the genome.



**Figure 3.** (A) Comparison of binding signals, represented by Z-scores as described in Kahramanoglu *et al.* (4), for IhfA, IhfB and Fis. This shows that binding signals for IHF are considerably higher than those for Fis (B) Weblogo representing the binding motifs for IHF.

genome, which was defined by the median of the read count distribution across the gene body. This value was then normalized by the corresponding value in the mock-IP data. This method is similar to that used to quantify nucleosome occupancies in eukaryotic studies (50). This normalized HU occupancy correlates positively with the A/T content of the bound DNA (Figure 4A). Second, for the exponential phase sample, we adapted a procedure used previously to investigate ChIP-chip data for nucleoporins in *Drosophila* (45)—which also showed wide-spread but low levels of binding, to identify 1104 and 1179 regions of enriched signal for HupA and HupB, respectively, with excellent agreements between the binding regions for the two subunits (>90% of peaks in the smaller list overlap with those in the second list). In agreement with our observations using gene-based occupancy profiles, these binding regions have significantly higher A/T content than the genomic average (Figure 4B). However, motif identification was not reliable, as different motifs were identified as significant for HupA and HupB despite their binding regions overlapping strongly (Figure 4C). This suggests that slight variations in the exact positioning of the binding regions might affect motif identification. Nevertheless, the one common feature of the identified motifs is A/T richness, which is in agreement with the findings described above and with results from an earlier report of *in vitro* specificity of HU–DNA interactions (51). This partiality towards A/T-rich genomic regions may be in line with previous



**Figure 4.** (A) Correlation between HU binding signal (as the median of read counts across the gene body, where read counts are divided by the total number of reads obtained for that sample and multiplied by a factor of  $10^7$ ; this number was then subtracted by the corresponding value from the mock-IP experiment) and A/T content (as a fraction of the total number of bases). These are for exponential phase data. (B) A/T content of regions of enriched mock-IP-subtracted ChIP-seq signal for HupA and HupB [modified from Kind *et al.* (45)], compared to the A/T content of randomly picked regions of the same length as regions of enriched signal (marked as controls). (C) Best motif identified by MEME for HupA and HupB.

reports suggesting a preference for HU to bind to bent DNA (52,53). In summary, HU binds largely in a non-specific fashion to the chromosome, with a particular preference toward targeting A/T-rich regions.

Finally, comparison of binding signals obtained from each subunit of the same protein indicates a high degree of correspondence between the two (Figure 1B). Notably for IHF, the proportions of the genome covered by the binding regions for IhfB (identified as described below) were considerably more than that for IhfA in three of the four time-points (excepting mid-exponential phase). Binding regions for IhfB are generally longer (by 5–10% median;  $P < 10^{-6}$ , Paired Wilcoxon test for all the above

three time-points; Supplementary Figure S3) than the corresponding region for IhfA suggesting that many IhfB binding regions are extensions of IhfA binding regions. This might be in concordance with a previous report showing that IhfB homodimers are more likely to form than IhfA dimers (19), but that such dimers may not exist freely in solution (20). For both proteins, there is high correlation among the binding profiles across time-points during our batch culture (Figure 1B).

### Effects of IHF and HU on global gene expression in *E. coli*

To investigate the effects of IHF and HU on gene expression in *E. coli*, we created single ( $\Delta ihfA$ ,  $\Delta ihfB$ ,  $\Delta hupA$  and  $\Delta hupB$ ) and double-deletion ( $\Delta ihfAB$ ,  $\Delta hupAB$ ) strains for the genes comprising the subunits of the two proteins (Supplementary Figures S4 and S5). We then performed microarray experiments measuring transcript abundance in these strains during exponential and late-exponential, transition to stationary and stationary phases of growth and compared them to that in the time-matched wild-type cells.

#### Effect of IHF on *E. coli* gene expression

We observe differential expression of only a small number of genes in the *ihf* single deletions (97 for IhfA and 56 for IhfB across all four conditions). Though a significantly larger number of genes are differentially expressed in the *ihfAB* double deletion, the number is much smaller (477 across all four conditions) than what we previously observed (4) for other sequence-specific nucleoid proteins such as Fis (1104 genes adopting the same criteria for calling differential expression as for the IHF data) and H-NS (1987 genes). Most of these effects are seen during the two exponential phases with only approximately 50 genes being differentially expressed—compared with the wild-type—during the stationary phase. Across the conditions, almost equal numbers of genes are up- or downregulated in  $\Delta ihfAB$ ; however, over two-thirds of the genes that are differentially expressed during late-exponential phase are upregulated (70%). Among the genes upregulated in  $\Delta ihfAB$ , there is a statistical enrichment for genes involved in ‘energy production and conversion’, a property that is seen particularly in late-exponential phase; however, these do not show any strong representations of individual metabolic pathways.

There is very little overlap among the sets of genes differentially expressed across different time-points (Supplementary Figure S4), despite the fact that the binding profile of IHF does not change significantly with growth phase. Further, similar to observations made earlier for Fis (4,6), there is very little correspondence between IHF binding and differential expression. Specific examples of IHF-bound genes that are differentially expressed in the double deletion includes the *fim* operon, which is strongly downregulated in the deletion strain in exponential phase. This is in agreement with prior molecular studies which have implicated IHF in both phase-switching and gene expression control at the *fim* operon (25).

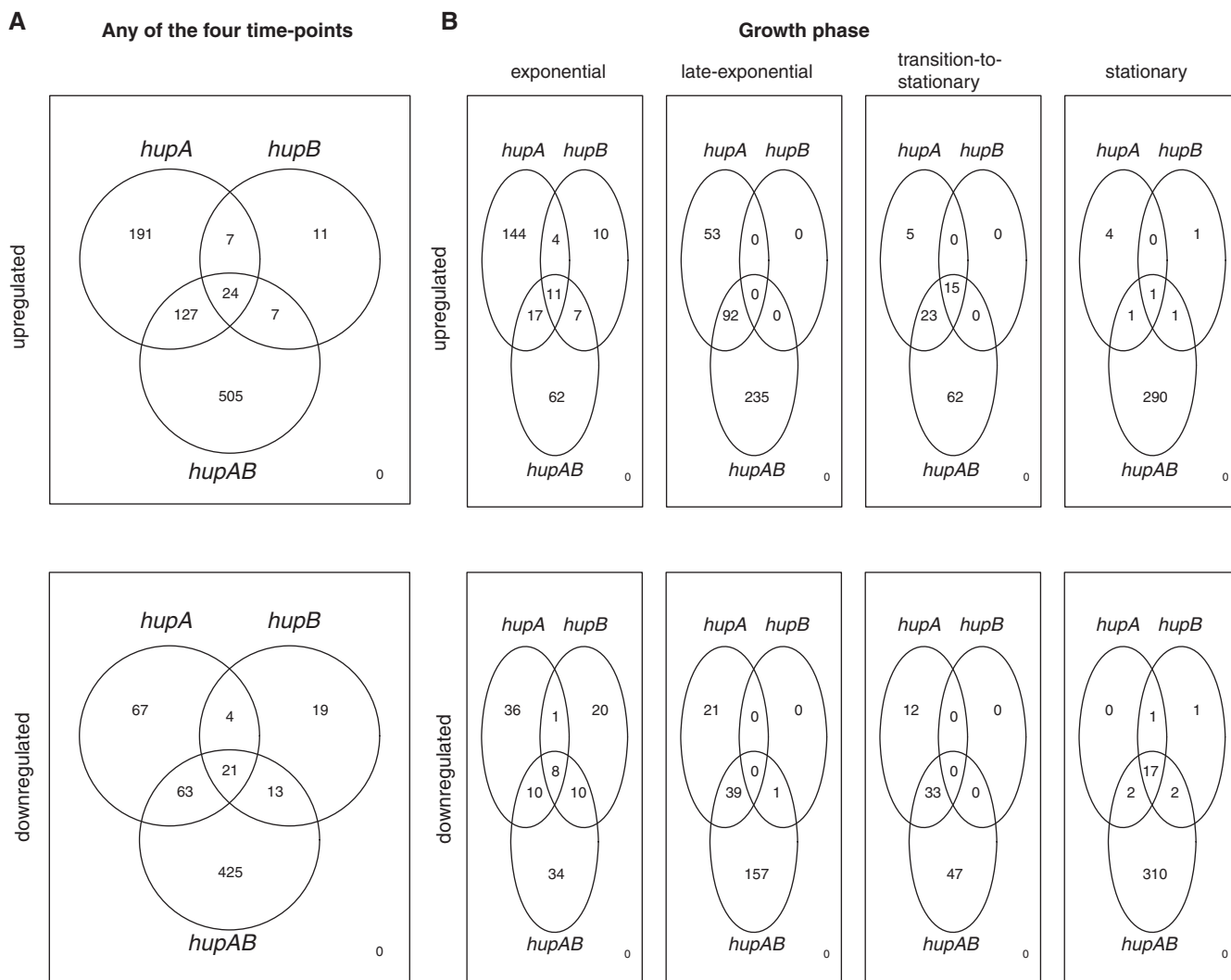
The lack of an observable global effect of IHF on the expression of genes bound by it might be explained by combinatorial regulation, i.e. the possible role of IHF as a facilitator of binding of other transcription factors to gene-upstream regions. For example, using ChIP-seq data previously generated in our lab, we find that there is a significant overlap between the genes bound by IHF and those by Fis (35% of genes bound in all conditions by IHF are also bound by Fis;  $P = 2 \times 10^{-5}$ , Fisher’s exact test). A previous study has shown that IHF is the second-most prolific transcription factor in terms of the number of other transcription factors with which it shares target genes (54). A striking example of this is the observed binding of IHF to a significant proportion (40%;  $P = 4 \times 10^{-5}$ , Fisher exact test) of genes regulated by  $\sigma^{54}$ , whose activation by AAA<sup>+</sup> ATPase transcription factors requires IHF-dependent DNA bending (55). The effect of such binding on gene expression might be highly specific to conditions, such as nitrogen limitation, not used in this study.

#### Effect of HU on *E. coli* gene expression

In contrast to IHF, mutants deficient in HU show large changes in gene expression; across the four conditions tested here, 1490 genes are up, or downregulated in either the single or the double mutants (Supplementary Figure S5). The greatest effect is seen in the double mutant in which 1266 genes are differentially expressed when compared to the wild-type; 512 genes change in expression in  $\Delta hupA$  whereas only 107 genes do so in  $\Delta hupB$ . Overall, a majority of differentially expressed genes are upregulated in  $\Delta hupAB$  (56%;  $P < 0.001$ , compared against random assignments of up and down regulation of genes) and  $\Delta hupA$  (69%;  $P < 0.001$ )—the two mutants that display global changes in gene expression. A statistically significant proportion of genes differentially expressed in  $\Delta hupA$  also change in expression in  $\Delta hupAB$  (43 and 54% of genes up- and downregulated in  $\Delta hupA$ ;  $P < 10^{-6}$ , Fisher’s exact test; Figure 5A); despite this, it must be noted that a significant component of each regulon is distinct from the other.

Genes that are upregulated in  $\Delta hupAB$  show an enrichment for essentiality for growth in rich media ( $P < 10^{-6}$  for sets, Fisher’s exact test; Figure 6); this is not true of genes differentially expressed in  $\Delta hupA$ . We then analyzed the COG functional categories of genes that are differentially expressed in these mutants, and find that genes involved in translation and ribosome biogenesis are upregulated in the double mutant but not in  $\Delta hupA$  (or in  $\Delta hupB$ ). We also find that genes involved in motility are upregulated in both the mutants. Finally, since HU is the most conserved NAP in bacteria, we analyzed the degree to which its target genes in *E. coli* are conserved across prokaryotes. Genes that are upregulated in  $\Delta hupAB$  tend to be highly conserved, whereas the same is not true of genes that are downregulated by  $\Delta hupAB$  or those that change in expression in  $\Delta hupA$ . The high degree of conservation observed for  $\Delta hupAB$  targets is not merely due to the aforementioned enrichment of genes involved in translation. It had previously been observed that genes





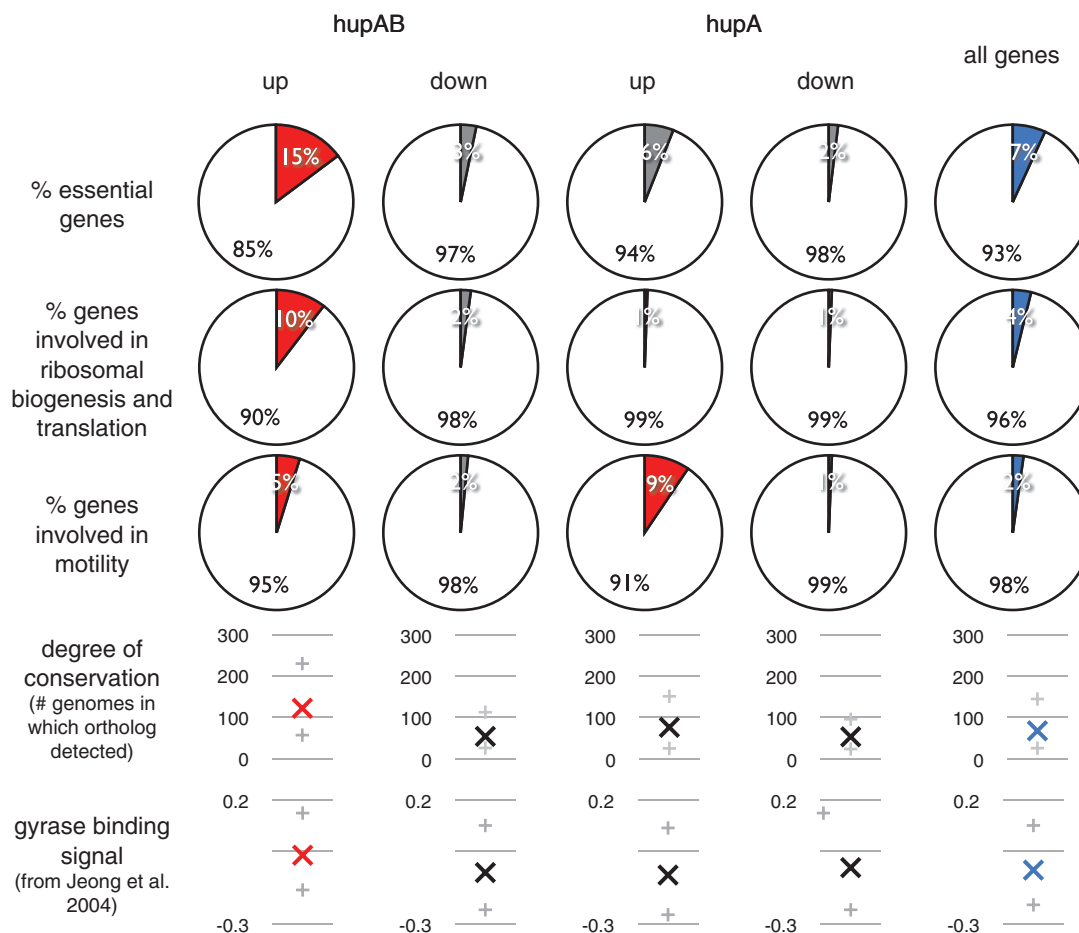
**Figure 5.** Venn diagrams showing the degree of overlap in numbers of genes that are differentially expressed in the various *hup* deletions at different time-points.

that are differentially expressed in  $\Delta hupAB$  tend to be bound by DNA gyrase (32), and are supercoiling-sensitive; in our data, this trend is relatively weak in  $\Delta hupAB$ , though statistically significant ( $P < 10^{-6}$ ; Mann-Whitney test), but absent in  $\Delta hupA$ .

It has been shown previously that deletion of HU leads to an increase in the accessibility of DNA to the DNA relaxing activity of topoisomerase I (32,56). This is, at first glance, at odds with the observation that genes, which are bound by the opposing DNA gyrase tend to be upregulated in the  $\Delta hupAB$  mutant in the present work and in an earlier work by Muskhelishvili's group (32). The authors of the above paper showed that there is little change in the unconstrained supercoiling levels in the double mutant (32). They further hypothesized that the increased accessibility of topoisomerase I to the DNA in the HU double mutant might be compensated by higher local negative supercoiling introduced at the upregulated loci by greater DNA gyrase binding and higher levels of transcription. To test this hypothesis we classified all

genes into four groups based on gyrase binding (40) and mid-exponential phase gene expression levels as measured using RNA-seq experiments in wild-type cells (4): (i)  $H_E H_G$ : high expression, high gyrase binding (high defined by the top third of the distribution); (ii)  $H_E L_G$ : high expression, low gyrase binding (low defined by the bottom third of the distribution); (iii)  $L_E H_G$ : low expression, high gyrase binding; (iv)  $L_E L_G$ : low expression, low gyrase binding. Though only  $\sim 27\%$  of all classifiable genes belong to  $H_E H_G$ ,  $\sim 54\%$  of genes upregulated in  $\Delta hupAB$  have high gyrase binding and high gene expression in wild-type cells ( $P < 10^{-6}$ ; Fisher's exact test; Supplementary Figure S6). This might indicate a possible role for increased local negative supercoiling, introduced by a combination of high transcription and DNA gyrase binding in determining upregulation in the  $\Delta hupAB$  mutant.

Analysis of expression patterns across different phases of growth reveals complex trends (Figure 5B). There is a progressive decrease in the number of genes that



**Figure 6.** Pie charts showing the proportions of genes in various sets (up or downregulated in *hupAB*, *hupA* or *hupB* across any of the four time-points) that are essential, involved in ribosome biogenesis and translation, and motility. Also shown are plots displaying the degree of conservation (as determined by the presence of bi-directional best-hit FASTA orthologs across 380 prokaryotic genomes) and DNA gyrase binding signal [from Jeong *et al.* (40)]; in these plots 'X' is the median of the distribution and '+' shows the first and the third quartiles. Red shows statistical enrichment (pie charts) or statistically higher medians (distributions) when compared to the reference set of all genes (blue).

are differentially expressed in  $\Delta hupA$  as the culture progresses through batch growth; however, there is only a slight overlap between the lists of differentially expressed genes in different time points. On the other hand, in  $\Delta hupAB$ , many genes change in expression during late exponential and stationary phases, though significant effects could be seen during the other two time-points; again each phase of growth sees a largely distinct set of genes being differentially regulated. These are described below.

During exponential phase, similar numbers of genes are differentially expressed in  $\Delta hupA$  and  $\Delta hupAB$ , with a small though statistically significant overlap between them. Essential, conserved genes and those involved in translation and ribosome biogenesis are over-represented among genes upregulated in  $\Delta hupAB$  but not in  $\Delta hupA$ . Similar enrichments are seen in the larger set of genes that is upregulated in  $\Delta hupAB$  during late-exponential phase; however, in contrast to the earlier time-point, almost all genes that are up-regulated in  $\Delta hupA$  also do so in  $\Delta hupAB$ . During the two exponential phase time-points, the number of genes upregulated in  $\Delta hupAB$  (67%) overwhelms those that are downregulated. Only a few genes

change in expression in  $\Delta hupB$  during this period. Though relatively few genes are differentially expressed during the transition to stationary phase, we note that there is a striking upregulation of various flagellar genes, involved in motility, in all three mutants at this time; we have validated several of these using RT-PCR (Supplementary Table S6). This is at odds with previous observations of a HU mutant that is non-motile in *E. coli* K12 W3110 (57) because of reduced transcription of the flagellin gene. Swimming motility assays performed by us resulted in smaller swarm diameters for the various *hup* mutants, the double mutant in particular; however, all mutants were motile (Supplementary Figure S7). Finally, during stationary phase, only  $\Delta hupAB$  shows global changes in gene expression. Unlike in the earlier time-points, only a slight majority (55%) of genes are upregulated, in which there is a statistical over-representation of translation-associated genes ( $P < 10^{-6}$ , Fisher's exact test followed by multiple correction by FDR); there is no functional enrichment detectable among downregulated genes.

In summary, deletion of both *hupA* and *hupB* has significantly greater impact on gene expression than that of either gene alone, with  $\Delta hupA$  displaying greater gene

expression changes than  $\Delta hupB$ . Further, genes that are upregulated in  $\Delta hupAB$  but not in  $\Delta hupA$  tend to be statistically enriched in essential cell processes such as translation, and are more conserved in prokaryotes than expected by random chance. These results may be consistent with our observation that during growth, cell densities are lower ( $\sim 25\%$  less than the wild-type) in the double deletion than in the wild-type or the single deletions; this possibly arises from a longer lag phase observed in the double deletion, although the growth rate of  $\Delta hupAB$  during exponential phase does not seem to be different from that of the wild-type [Supplementary Figure S8; also reported by (32)].

### Investigations of potential IHF and HU homodimers binding to the chromosome

Following from our observations of largely incongruent effects of single and double mutants of HU and IHF on *E. coli* gene expression, we performed ChIP-Seq experiments on each subunit of the two proteins, in strains carrying deletions of the second subunit. For IHF, our experiments did not yield enough DNA to perform sequencing reactions; this is suggestive of very weak or no homodimer binding to the DNA, at least in the absence of the second subunit. For HU, we were able to obtain binding profiles for each subunit in the absence of the other, which were strikingly similar to those in the wild-type (Supplementary Figure S9). This indicates that the homodimers bind to the chromosome in similar patterns as the heterodimer. This may be reflected in the fact that most bacterial genomes encode only one HU subunit, and is supportive of the fact that more genes change in expression in the double deletion than in the single mutants.

It has previously been suggested that HupA2 is the predominant form of HU in exponential phase, whereas the heterodimer takes over during later stages of growth (17). Western blots presented here show higher expression of HupA than HupB during exponential phase (Supplementary Figure S10). Gene expression data described above show greater gene expression changes in  $\Delta hupA$  than in  $\Delta hupB$ , especially during exponential phase. Our ChIP-seq data, however, show similar binding profiles for both subunits of HU across all stages of growth; results reported in this section further suggest that the binding profile of HupB is similar between the wild-type and  $\Delta hupA$  mutant. Though it is possible that there is a uniform reduction in binding signals for HupB in  $\Delta hupA$ , which might account for gene expression changes observed in  $\Delta hupA$  during exponential and late-exponential phases, there is little reorganization of HupB's binding profile.

## DISCUSSION

IHF and HU are two nucleoid-associated proteins that belong to the same DNA binding protein family, but show distinct levels of sequence specificities. IHF, a sequence-specific DNA binding protein, has extreme effects on the topology of bound DNA, which it bends by  $\sim 160^\circ$  (58). HU, the most conserved NAP, binds

more uniformly to the *E. coli* chromosome, with a preference for distorted DNA structures (13–16,52,53). Both proteins exist as heterodimers in *E. coli*. In this article, we report results from our genome-scale studies of the binding of IHF and HU to the chromosome of *E. coli* K12 MG1655, and its effects on gene expression at various time-points of batch culture, from growth to stasis.

IHF displays sequence-specific binding to the *E. coli* chromosome, with signal intensities significantly stronger than those observed for Fis, another sequence-specific NAP. The two subunits of IHF show similar binding profiles, indicative of preferential heterodimer formation. In the wild-type strain, IhfB binding regions cover more of the chromosome than those of IhfA (in three of the four conditions); this might be in line with a prior observation that IhfB homodimers form more readily than IhfA homodimers (19). However, our inability to recover enough DNA from ChIP experiments for IhfB in  $\Delta ihfA$  suggests that such homodimer formation may occur on the DNA (20), only in the presence of a nucleating heterodimer complex.

Across the four conditions tested, IHF binding regions target the upstream regions of over 30% of all predicted operons, indicative of a global role for the protein in regulating gene expression. However, only  $\sim 10\%$  of these change in expression when the genes coding for the two subunits of IHF are deleted. Additionally, compared to the effects of other sequence-specific NAPs such as Fis and H-NS (4), the overall effect of IHF on gene expression under the present conditions is less in terms of the number of genes that are differentially expressed in the deletion mutant(s). This might be linked to our observation that, in contrast to Fis where over 20% of binding motifs lie upstream of operons (4), only  $\sim 10\%$  of predicted IHF binding motifs are so positioned. We suggest that the minimal proximal effect of IHF on gene expression might be due to combinatorial regulation, i.e. the tendency of IHF to regulate genes jointly with other factors. Another possibility is that HU might compensate for the absence of IHF; this has been demonstrated for excisive recombination at specific sites (59), but remains to be investigated on a genomic scale in the context of transcription. In this context, it must be noted that IHF has important functions outside of transcriptional regulation such as recombination (60), which are not apparent in our transcriptome experiment; in fact the large majority of binding sites which are located in non-intergenic regions might have such functions.

In agreement with current knowledge, the binding profile for HU is reflective of relatively non-specific binding to the chromosome, however with a notable preference for A/T-rich DNA in concordance with previous *in vitro* studies (51). It has been shown previously that the composition of the HU dimer varies across the various phases of growth of *E. coli* with HupA2 being the dominant form during exponential phase and the heterodimeric form dominating during later stages of growth (17,61). Our western blots do indicate higher levels of expression of HupA than HupB during exponential and late-exponential growth phases. But, the binding

profile of each subunit of HU strongly correlates with that of the other subunit across the growth phases, including exponential growth; this might indicate that homo- and heterodimeric forms of HU could bind to the DNA interchangeably. Further, the binding profile of each subunit in the wild-type is similar to that in an otherwise isogenic strain that is lacking the other subunit. It is possible that the binding of HupB to the chromosome is uniformly less (across the genome) in  $\Delta hupA$  than in the wild-type, thus accounting for gene expression changes seen in  $\Delta hupA$  during exponential and late-exponential phases of growth. This interpretation may be in line with previous reports showing *in vitro* that HupB2 binds poorly to duplex DNA (18); however there is enough binding for us to recover in our ChIP experiments. However our data do not suggest any large-scale reorganization of the HupB binding profile following *hupA* deletion.

In apparent conflict with the consistency that the two subunits show in their binding, they have substantially different effects on gene expression. Briefly, we observe that  $\Delta hupAB$  has the greatest effects on gene expression, distantly followed by  $\Delta hupA$ , with minimal effects seen in  $\Delta hupB$ . Previous studies in *E. coli* C600 and *Salmonella enterica* have also observed discordance between the sets of genes differentially expressed in single and double deletions of HU subunits (11,33). The authors of the paper on *E. coli* C600 identified few genes as members of HupB2, interpreting this as a possible consequence of previously observed instability and low expression level of this form of the protein at 37°C, and the inability of HupB2 to introduce negative supercoiling on relaxed DNA in the presence of topoisomerase I (11). A similar observation— $\Delta hupB$  showing significantly smaller changes in gene expression than  $\Delta hupA$  and  $\Delta hupAB$ —was made in two of the three growth phase time-points tested in *S. enterica*; however, the extent of differential expression was similar in  $\Delta hupAB$  and  $\Delta hupA$  across all time-points (33).

The aforementioned study on *E. coli* C600 showed that the HU regulon is composed of genes involved in energy metabolism, SOS response, and osmolarity and acid stress responses (31). In contrast to the conclusions of a later study (32), which investigated the transcriptome of only the double mutant, these authors did not find any supercoiling dependence in the expression of the members of the HU regulon. Here we observe that genes that are upregulated in  $\Delta hupAB$  are statistically enriched for essential cellular functions such as translation and show statistically higher binding to DNA gyrase than other genes. Our analysis also agrees with a previous hypothesis that local negative supercoiling introduced by DNA gyrase and high transcription might compensate for the increased accessibility of DNA to topoisomerase I in the HU double mutant (32). Despite the fact that a significant proportion of genes that are upregulated in  $\Delta hupA$  also change in expression in  $\Delta hupAB$ , the above functional enrichments are not observed in  $\Delta hupA$ . Similarly, and in line with the fact that HU is the most conserved NAP in bacteria, genes upregulated in  $\Delta hupAB$  are more conserved across bacteria than other genes. This may also be reflected in the fact that the growth curve of  $\Delta hupAB$ , but not that of

$\Delta hupA$  or  $\Delta hupB$ , differs from that of the wild-type. Moreover, many bacterial genomes encode only one subunit of HU; the fact that a second subunit is encoded in *E. coli* might in part build in some redundancy to this conserved regulatory system. However, it is remarkable that the subunit of HU that is more conserved across bacteria is HupB, which appears to be the minor player in gene expression control at least in *E. coli* and *S. enterica* both of which encode both subunits of this protein.

## ACCESSION NUMBER

All sequence data have been deposited at NCBI SRA (Study accession SRP008538). All microarray data have been deposited at ArrayExpress (accession number E-MEXP 3461).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–7, Supplementary Figures 1–10, Supplementary Data set.

## ACKNOWLEDGEMENTS

The authors thank Vladimir Benes, David Ibberson and Sabine Schmidt at the Genomics Core Facility, EMBL-Heidelberg for the sequencing and microarray experiments. They thank the three anonymous referees for their critical comments and suggestions.

## FUNDING

Girton College, University of Cambridge; Ramanujan Fellowship, Department of Science and Technology, Government of India SR/S2/RJN-49/2010 (to A.S.N.S.); Spanish Ministry of Science and Innovation (to A.I.P.); Biotechnology and Biological Sciences Research Council (BBSRC) grant ‘Genomic Analysis of Regulatory Networks for Bacterial Differentiation and Multicellular Behaviour’ BB/E011489/1 (to G.M.F.) and BB/E01075X/1 (to N.M.L.); Isaac Newton Trust (to G.M.F.); European Molecular Biology Laboratory (EMBL) (to N.M.L.). Funding for open access charge: European Molecular Biology Laboratory.

*Conflict of interest statement.* None declared.

## REFERENCES

- Dillon, S.C. and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature reviews. Microbiology*, **8**, 185–195.
- Azam, T.A. and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.*, **274**, 33105–33113.
- Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, **181**, 6361–6370.
- Kahramanoglou, C., Seshasayee, A.S.N., 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.
5. 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.
  6. Cho, B.-K., Knight, E.M., Barrett, C.L. and Palsson, B.Ø. (2008) Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res.*, **18**, 900–910.
  7. Lucchini, S., Rowley, G., Goldberg, M.D., Hurd, D., Harrison, M. and Hinton, J.C.D. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathogens*, **2**, e81.
  8. 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.
  9. Noom, M.C., Navarre, W.W., Oshima, T., Wuite, G.J.L. and Dame, R.T. (2007) H-NS promotes looped domain formation in the bacterial chromosome. *Curr. Biol. CB*, **17**, R913–R914.
  10. Oshima, T., Ishikawa, S., Kurokawa, K., Aiba, H. and Ogasawara, N. (2006) *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res. Int. J. Rapid Publication Rep. Genes Genomes*, **13**, 141–153.
  11. Oberto, J., Drlica, K. and Rouvière-Yaniv, J. (1994) Histones, HMG, HU, IHF: Mème combat. *Biochimie*, **76**, 901–908.
  12. Benevides, J.M., Danahy, J., Kawakami, J. and Thomas, G.J. (2008) Mechanisms of specific and nonspecific binding of architectural proteins in prokaryotic gene regulation. *Biochemistry*, **47**, 3855–3862.
  13. Castaing, B., Zelwer, C., Laval, J. and Boiteux, S. (1995) HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J. Biol. Chem.*, **270**, 10291–10296.
  14. Kamashev, D., Balandina, A. and Rouvière-Yaniv, J. (1999) The binding motif recognized by HU on both nicked and cruciform DNA. *EMBO J.*, **18**, 5434–5444.
  15. Kamashev, D. and Rouvière-Yaniv, J. (2000) The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J.*, **19**, 6527–6535.
  16. Kobryn, K., Lavoie, B.D. and Chaconas, G. (1999) Supercoiling-dependent site-specific binding of HU to naked Mu DNA. *J. Mol. Biol.*, **289**, 777–784.
  17. Claret, L. and Rouvière-Yaniv, J. (1997) Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival. *J. Mol. Biol.*, **273**, 93–104.
  18. Pinson, V., Takahashi, M. and Rouvière-Yaniv, J. (1999) Differential binding of the *Escherichia coli* HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. *J. Mol. Biol.*, **287**, 485–497.
  19. Zuilianello, L., de la Gorgue de Rosny, E., van Ulsen, P., van de Putte, P. and Goosen, N. (1994) The HimA and HimD subunits of integration host factor can specifically bind to DNA as homodimers. *EMBO J.*, **13**, 1534–1540.
  20. Werner, M., Clore, G., Gronenborn, A. and Nash, H. (1994) Symmetry and asymmetry in the function of *Escherichia coli* integration host factor: implications for target identification by DNA-binding proteins. *Curr. Biol.*, **4**, 477–487.
  21. Balandina, A., Claret, L., Hengge-Aronis, R. and Rouvière-Yaniv, J. (2001) The *Escherichia coli* histone-like protein HU regulates rpoS translation. *Mol. Microbiol.*, **39**, 1069–1079.
  22. Balandina, A., Kamashev, D. and Rouvière-Yaniv, J. (2002) The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J. Biol. Chem.*, **277**, 27622–27628.
  23. Browning, D.F., Cole, J.A. and Busby, S.J.W. (2008) Regulation by nucleoid-associated proteins at the *Escherichia coli* nir operon promoter. *J. Bacteriol.*, **190**, 7258–7267.
  24. Corcoran, C.P. and Dorman, C.J. (2009) DNA relaxation-dependent phase biasing of the fim genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP. *Mol. Microbiol.*, **74**, 1071–1082.
  25. Dorman, C.J. and Higgins, C.F. (1987) Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. *J. Bacteriol.*, **169**, 3840–3843.
  26. Parekh, B.S. and Hatfield, G.W. (1996) Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. *Proc. Natl Acad. Sci. USA*, **93**, 1173–1177.
  27. Gama-Castro, S., Jiménez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muñoz-Rascado, L., Martínez-Flores, I., Salgado, H. et al. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.*, **36**, D120–D124.
  28. Arfin, S.M., Long, A.D., Ito, E.T., Toller, L., Riehle, M.M., Paegle, E.S. and Hatfield, G.W. (2000) Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J. Biol. Chem.*, **275**, 29672–29684.
  29. Mangan, M.W., Lucchini, S., Danino, V., Cróinin, T.O., Hinton, J.C.D. 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.
  30. Semsey, S., Virnik, K. and Adhya, S. (2006) Three-stage regulation of the amphibolic gal operon: from repressosome to GalR-free DNA. *J. Mol. Biol.*, **358**, 355–363.
  31. Oberto, J., Nabti, S., Jooste, V., Mignot, H. and Rouvière-Yaniv, J. (2009) The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS One*, **4**, e4367.
  32. Berger, M., Farcas, A., Geertz, M., Zhelyazkova, P., Brix, K., Travers, A. and Muskhelishvili, G. (2010) Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Rep.*, **11**, 59–64.
  33. Mangan, M.W., Lucchini, S.O., Cróinin, T., Fitzgerald, S., Hinton, J.C.D. and Dorman, C.J. (2011) Nucleoid-associated protein HU controls three regulons that coordinate virulence, response to stress and general physiology in *Salmonella enterica* serovar Typhimurium. *Microbiology*, **157**, 1075–1087.
  34. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
  35. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Systems Biol.*, **2**, 2006.0008.
  36. Thomason, L., Costantino, N. and Court, D. (2007) *E. coli* genome manipulation by P1 transduction. *Curr. Protocols Mol. Biol.*, (Suppl. 79), 1.17.
  37. 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.
  38. Grainger, D.C., Overton, T.W., Reppas, N., Wade, J.T., Tamai, E., Hobman, J.L., Constantinidou, C., Struhl, K., Church, G. and Busby, S.J.W. (2004) Genomic studies with *Escherichia coli* MelR protein: applications of chromatin immunoprecipitation and microarrays. *J. Bacteriol.*, **186**, 6938–6943.
  39. Keseler, I.M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T. et al. (2009) EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.*, **37**, D464–D470.
  40. Jeong, K.S., Ahn, J. and Khodursky, A.B. (2004) Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol.*, **5**, R86.
  41. Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
  42. Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carrier, N., Snyder, M. and Gerstein, M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat. Biotechnol.*, **27**, 66–75.
  43. Cairns, J., Spyrou, C., Stark, R., Smith, M.L., Lynch, A.G. and Tavaré, S. (2011) BayesPeak—an R package for analysing ChIP-seq data. *Bioinformatics*, **27**, 713–714.

44. Spyrou, C., Stark, R., Lynch, A.G. and Tavaré, S. (2009) BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinform.*, **10**, 299.
45. Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P. and Akhtar, A. (2008) Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell*, **133**, 813–828.
46. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
47. 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.
48. Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A. and Smyth, G.K. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, **23**, 2700–2707.
49. Huber, W., Heydebreck, A. von, Sülthmann, H., Poustka, A. and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**(Suppl. 1), S96–S104.
50. Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J. and Segal, E. (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.*, **4**, e1000216.
51. Krylov, A.S., Zasedateleva, O.A., Prokopenko, D.V., Rouviere-Yaniv, J. and Mirzabekov, A.D. (2001) Massive parallel analysis of the binding specificity of histone-like protein HU to single- and double-stranded DNA with generic oligodeoxyribonucleotide microchips. *Nucleic Acids Res.*, **29**, 2654–2660.
52. Pontiggia, A., Negri, A., Beltrame, M. and Bianchi, M.E. (1993) Protein HU binds specifically to kinked DNA. *Mol. Microbiol.*, **7**, 343–350.
53. Wojtuszewski, K. and Mukerji, I. (2003) HU binding to bent DNA: a fluorescence resonance energy transfer and anisotropy study. *Biochemistry*, **42**, 3096–3104.
54. Martínez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, **6**, 482–489.
55. Jovanovic, G. and Model, P. (1997) PspF and IHF bind co-operatively in the *psp* promoter-regulatory region of *Escherichia coli*. *Mol. Microbiol.*, **25**, 473–481.
56. Bensaid, A., Almeida, A., Drlica, K. and Rouviere-Yaniv, J. (1996) Cross-talk between topoisomerase I and HU in *Escherichia coli*. *J. Mol. Biol.*, **256**, 292–300.
57. Nishida, S., Mizushima, T., Miki, T. and Sekimizu, K. (1997) Immobile phenotype of an *Escherichia coli* mutant lacking the histone-like protein HU. *FEMS Microbiol. Lett.*, **150**, 297–301.
58. 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.
59. Goodman, S.D., Nicholson, S.C. and Nash, H.A. (1992) Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. *Proc. Natl Acad. Sci. USA*, **89**, 11910–11914.
60. Goosen, N. and van de Putte, P. (1995) The regulation of transcription initiation by integration host factor. *Mol. Microbiol.*, **16**, 1–7.
61. Claret, L. and Rouviere-Yaniv, J. (1996) Regulation of HU alpha and HU beta by CRP and FIS in *Escherichia coli*. *J. Mol. Biol.*, **263**, 126–139.