Molecular Cell
Short Article

Protein Occupancy Landscape of a Bacterial Genome

Tiffany Vora,[1,2,3] Alison K. Hottes,[1,2] and Saeed Tavazoie[1,2,*]
[1]Department of Molecular Biology
[2]The Lewis-Sigler Institute for Integrative Genomics
Princeton University, Princeton, NJ 08544, USA
[3]Present address: School of Sciences and Engineering, The American University in Cairo, 11835 New Cairo, Egypt
*Correspondence: tavazoie@genomics.princeton.edu
DOI 10.1016/j.molcel.2009.06.035

SUMMARY

Protein-DNA interactions are fundamental to core biological processes, including transcription, DNA replication, and chromosomal organization. We have developed in vivo protein occupancy display (IPOD), a technology that reveals protein occupancy across an entire bacterial chromosome at the resolution of individual binding sites. Application to *Escherichia coli* reveals thousands of protein occupancy peaks, highly enriched within and in close proximity to noncoding regulatory regions. In addition, we discovered extensive (>1 kilobase) protein occupancy domains (EPODs), some of which are localized to highly expressed genes, enriched in RNA-polymerase occupancy. However, the majority are localized to transcriptionally silent loci dominated by conserved hypothetical ORFs. These regions are highly enriched in both predicted and experimentally determined binding sites of nucleoid proteins and exhibit extreme biophysical characteristics such as high intrinsic curvature. Our observations implicate these transcriptionally silent EPODs as the elusive organizing centers, long proposed to topologically isolate chromosomal domains.

INTRODUCTION

Replication, maintenance, and expression of genetic information are processes that are orchestrated through precise interactions of hundreds of proteins with chromosomal DNA. For decades, research has focused on the behavior and functional consequences of DNA-protein interactions at individual loci. However, understanding systems-level behaviors, such as chromosomal organization, genome replication, and transcriptional network dynamics, requires observations at the scale of the entire system. Microarray-based chromatin immunoprecipitation (ChIP-chip) allows global measurements of chromosomal occupancy for individual proteins (Ren et al., 2000). In another global approach, methylase protection, a fraction of all occupied sites are monitored in vivo, independently of the identity of the bound proteins (Tavazoie and Church, 1998). However, there currently exists no comprehensive approach for simultaneous, high-resolution monitoring of all in vivo protein-DNA interactions across the genome. We have developed such a technology and used it to profile protein occupancy of the *E. coli* chromosome at the resolution of individual binding sites.

RESULTS

In Vivo Protein Occupancy Display

In order to globally profile the occupancy of all proteins on chromosomal DNA, we first stabilize in vivo protein-DNA interactions through covalent crosslinking with formaldehyde (Figure 1A). After cell lysis and sonication, protein footprints are minimized to a mode of ~50 bp through DNase I digestion (Figure 1B). Phenol extraction is then used to trap amphipathic protein-DNA complexes at the interface between the organic and aqueous phases. Following interface isolation and crosslink reversal, short DNA fragments are end labeled and hybridized to a high-density tiling array containing 25-mer oligonucleotides at the resolution of one every four base pairs across the entire genome. After scanning and data normalization, a high-resolution global protein occupancy profile is achieved. For each probe on the chip, protein occupancy enrichment or depletion levels are quantified using a z-score that represents the probe-by-probe relative signal intensity with respect to the mean, and normalized to the standard deviation, of signals from replicate hybridizations of whole genomic DNA (Experimental Procedures).

Global Protein Occupancy Profile of the *E. coli* Chromosome

The vast fraction of characterized protein-DNA interactions occur via sequence-specific interactions of transcription factors with DNA within, and in close proximity to, noncoding regulatory regions (Gama-Castro et al., 2008). Consistent with this, we see highly significant occupancy enrichment in noncoding regions as compared to coding regions (Figure 1C). This difference in occupancy is clearly discernable in a local chromosomal view where high-amplitude peaks are largely confined to the regions between genes (Figure 2A). Independent biological replicates demonstrate that the position and relative amplitude of these occupancy peaks show a high level of reproducibility (Figure 2A). Although there is, overall, relative depletion of occupancy within open reading frames (ORFs), occasionally this is interrupted by a sharp occupancy peak (Figures 2A and S1 [available online]). The functional role of these intragenic interactions is not known but could represent a significant gap in our understanding of bacterial gene expression. At high resolution, occupancies of

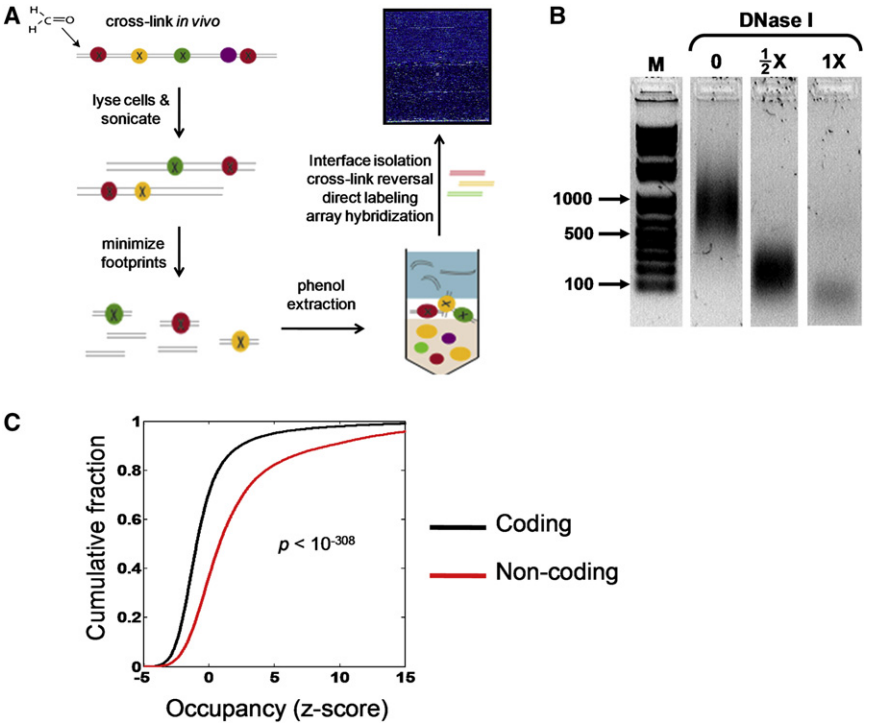


Figure 1. In Vivo Protein Occupancy Display

(A) Schematic for isolation and genome-wide display of protein-bound sites across a bacterial genome. Formaldehyde crosslinking preserves in vivo protein-DNA interactions. Following cell lysis and sonication, protein footprints are minimized through DNase I treatment. Phenol extraction enriches for protein-DNA complexes at the interface between the aqueous and organic phases. Following interface isolation, crosslinks are reversed, and the resulting DNA fragments are end labeled and hybridized to a tiling array.

(B) Gel fractionation shows that DNase I treatment leads to a drop in the mode of fragment length distribution from ~1000 bp (no DNase I) to ~200 bp (1/2× DNase I), to below 100 bp (1× DNase I). The samples were separated on the same gel, and extraneous lanes were removed for clarity.

(C) Cumulative probability distribution of occupancy (z-score: standard deviations from the mean) for both coding and noncoding regions determined during late exponential phase growth. The z-score values were smoothed by averaging within a moving window of 128 base pairs.

individual proteins can be readily discerned, displaying footprints on the scale of a typical transcription-factor-binding site (Figures 2B and S2). An automated peak detection algorithm identified ~2063 individual occupied sites in a population of *E. coli* cells growing in late exponential phase (Figure S3). The pattern of peaks is reproducible in biological replicates and shows condition-dependent variation (Figure S4).

Discovery of Extended Protein Occupancy Domains

Intriguingly, examination of the entire genome-wide occupancy profile revealed contiguous regions of protein binding, many of which extend beyond a kilobase in length (Figures 3A–3D and S5). We performed a systematic search for these extended protein occupancy domains (EPODs) under early exponential growth using an automated algorithm that identified regions 1024 bp or longer with contiguous median occupancy values above the 75th percentile of all genome-wide values (Supplemental Experimental Procedures). These domains had a median length of 1.6 kb and extended as long as 14 kb (Figure S6A). We wondered whether the extreme signal in these domains corresponded to the footprint of RNA polymerase within highly transcribed regions. To test this possibility, we performed transcriptional profiling under identical cellular growth conditions (Experimental Procedures). As can be seen (Figure 3A), we found clear cases where the boundaries of an EPOD coincided with those of highly transcribed regions such as those containing ribosomal protein genes (Figure 3A). However, we found many cases where EPODs existed in a transcriptionally silent state, across both genes and intergenic regions, and even long operons (Figures 3B–3D and S7). Due to their extreme and bimodal RNA expression behavior, we performed an automated classification of EPODs by clustering them into two populations using their median expression level across domains (Supplemental Experimental Procedures and Table S3). This resulted in 121 domains in the highly expressed class (*he*EPODs) and 151 in the transcriptionally silent class (*ts*EPODs). Previously published RNA polymerase ChIP-chip data (Grainger et al., 2005), from cells grown under identical conditions, allowed us to compare RNA polymerase occupancy of *ts*EPODs and *he*EPODs relative to a background set generated by randomly sampling genomic sequences from the overall EPOD length distribution (Figure 4A). As expected, *he*EPODs showed extremely high levels of RNA polymerase occupancy ($p < 10^{-246}$). In comparison, *ts*EPODs showed lower levels of RNA polymerase occupancy ($p < 0.02$) relative to control.

In order to gain further insight into the potential role of EPODs, we looked for enrichment of specific functional categories in genes that overlapped them (Table S1). As expected, *he*EPODs were highly enriched in processes and pathways that are highly expressed, including translation and tRNAs. The most significantly enriched classes within *ts*EPODs were predicted and hypothetical ORFs, with marginally significant enrichment in prophage and prophage-related genes. On the other hand, *ts*EPODs, by and large, avoid putatively essential genes (Table S2). The number of *ts*EPODs, their apparently random, yet widespread genome-wide distribution, and their enrichment within transcriptionally silent ORFs of unknown function, suggested that they may fulfill an architectural role. In fact, there exists compelling evidence that the *E. coli* chromosome is organized into domains, subserving both chromosomal compaction and topological domain isolation (Postow et al., 2004). Evidence for such in vivo organization comes from both genetic and biochemical studies (Garcia-Russell et al., 2007; Postow et al., 2004), including visualization of rosette-like structures by microscopy

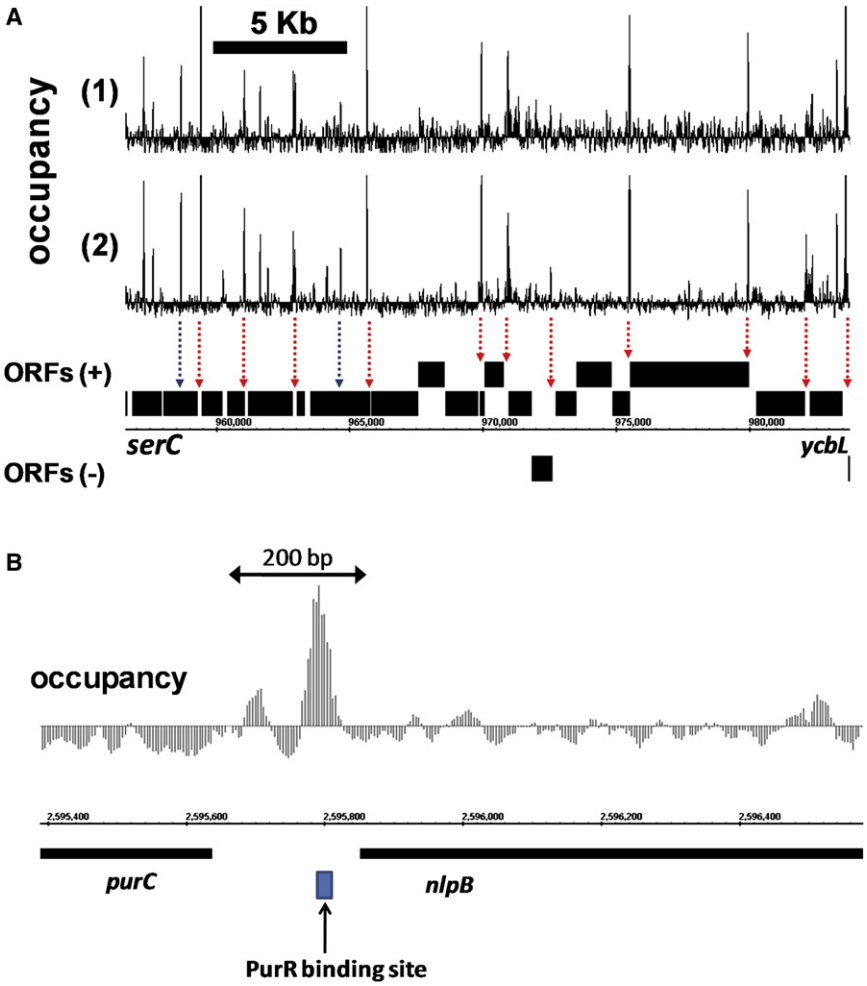


Figure 2. Protein Occupancy Profile of the *E. coli* Genome during Late Exponential Phase Growth

(A) At low spatial resolution, high-amplitude occupancy peaks are largely confined to intergenic (noncoding) regions of the genome (red arrows). However, similar peaks can less frequently be seen within coding regions as well (blue arrows). Two independent biological replicates show highly reproducible occupancy profiles across this region.

(B) At high spatial resolution, multiple occupancy peaks are discernable within a single intergenic region. Peaks are localized to the typical footprint of individual transcription factors and often overlap experimentally determined binding sites (PurR, RegulonDB).

(Delius and Worcel, 1974b; Hinnebusch and Bendich, 1997; Pettijohn, 1996; Postow et al., 2004). However, the formation, composition, maintenance, and dynamics of these domains remain open questions (Bendich, 2001; Postow et al., 2004; Travers and Muskhelishvili, 2007). Investigators have argued that such domains may be organized through the binding and cooperation of abundant proteins collectively referred to as nucleoid proteins (Azam and Ishihama, 1999). These proteins have characteristics that suit them well for this task. These include high abundance, low sequence specificity, tendency to cause DNA curvature, and propensity to bind curved DNA. In addition, some of these factors (e.g., H-NS) are known to form at least homodimeric interactions (Stella et al., 2005), a capacity that as argued previously (Dame et al., 2000; Skoko et al., 2006) may allow distant chromosomal sites to be brought together to form topologically isolated domains. Low-resolution ChIP-chip studies against known nucleoid proteins (Grainger et al., 2006) revealed both a bias toward interaction in noncoding regions and a correlation with Fis and H-NS binding, suggesting cooperative interaction of nucleoid proteins in maintaining genomic architecture.

We sought evidence for the involvement of nucleoid proteins in the formation of *ts*EPODs. The availability of probabilistic sequence specificity models, in the form of position weight matrices (PWM) (Gama-Castro et al., 2008), allowed us to determine the relative occupancy potential of these regions through computational analysis of a subset of these factors: H-NS, IHF, and Fis (Supplemental Experimental Procedures). We found that, indeed, as a population, *ts*EPODs have significantly higher PWM scores for all of these nucleoid proteins (e.g., for H-NS $p < 10^{-28}$). The same was not true for *he*EPODs, as their PWM score distribution did not deviate significantly from background (Figures 4B and S8). On the contrary, the PWM score distribution for LacI (a nonnucleoid transcription factor) showed the opposite trend, with significantly lower values ($p < 10^{-7}$) within *ts*EPODs (Figure 4C). Consistent with the preference of nucleoid proteins for A/T-rich DNA (Cho et al., 2008; Grainger et al., 2006), we also saw a highly skewed A:T frequency bias: 59% within *ts*EPODs, as compared to 49% for the background and 50% for *he*EPODs ($p < 10^{-30}$, Figure 4D). We also found *ts*EPODs to display extreme biophysical characteristics (Pedersen et al., 2000) such as high curvature ($p < 10^{-24}$) and stacking energy ($p < 10^{-34}$), again consistent with the hypothesis that these regions constitute chromosomal organizing centers (Figures 4E and S9). Consistent with our computational analyses above, we saw significant enrichment for the high-affinity binding of nucleoid proteins in our *ts*EPODs relative to background (Figure 4F) within individual ChIP-chip profiles for H-NS, IHF, and Fis (Grainger et al., 2006). Intriguingly, we also saw a highly significant enrichment for the binding of Fis within *he*EPODs (Figure 4F). This is consistent with the locus-specific role of Fis in the regulation of highly expressed genes, including ribosomal RNAs (Aiyar et al., 2002; Cho et al., 2008; Grainger et al., 2006).

DISCUSSION

In total, our observations argue in favor of a model in which the binding of *ts*EPODs by nucleoid proteins establishes them as

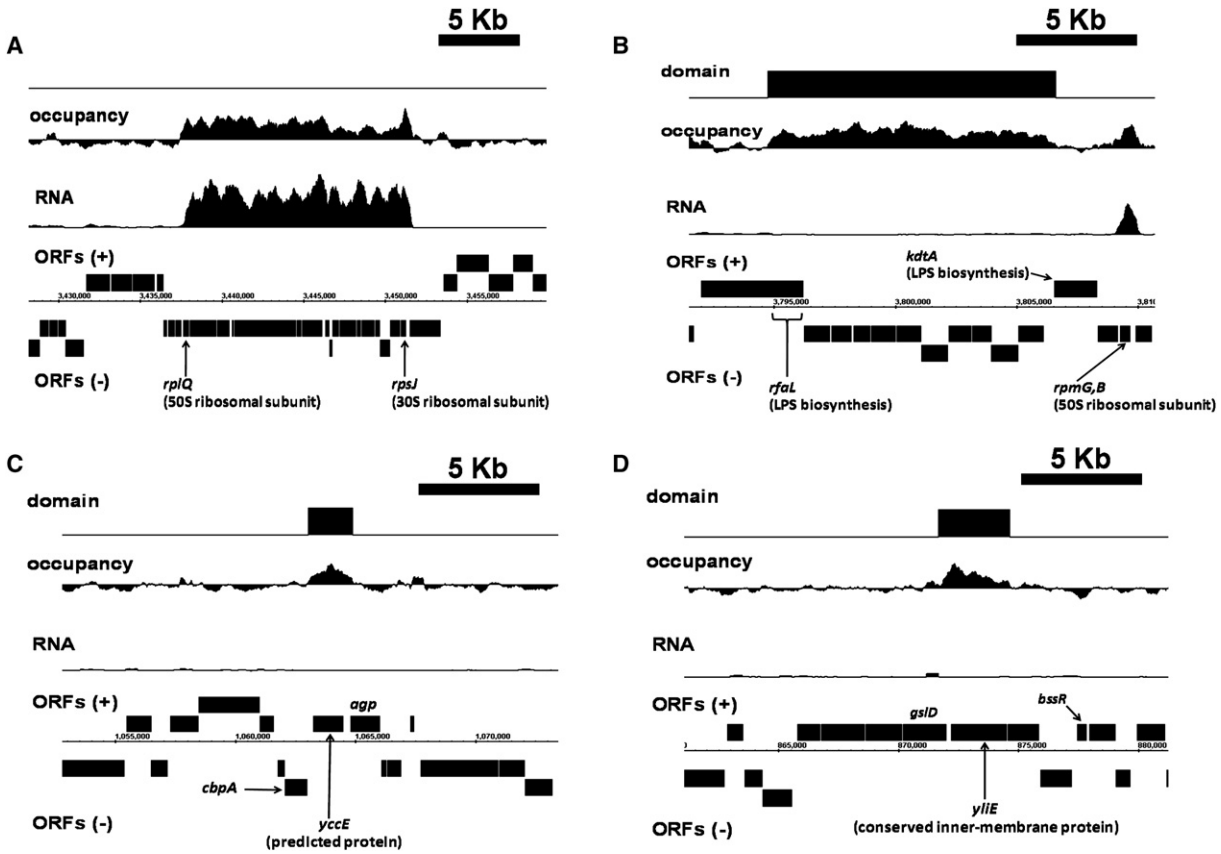


Figure 3. Extended Protein Occupancy Domains

Protein occupancy and RNA expression profiles are shown for early exponential phase growth, smoothed by averaging within a moving window of 512 base pairs. At the bottom, open reading frames (ORFs) are annotated on both strands. Automatically detected transcriptionally silent extended protein occupancy domains (EPODs) are shown at the top.

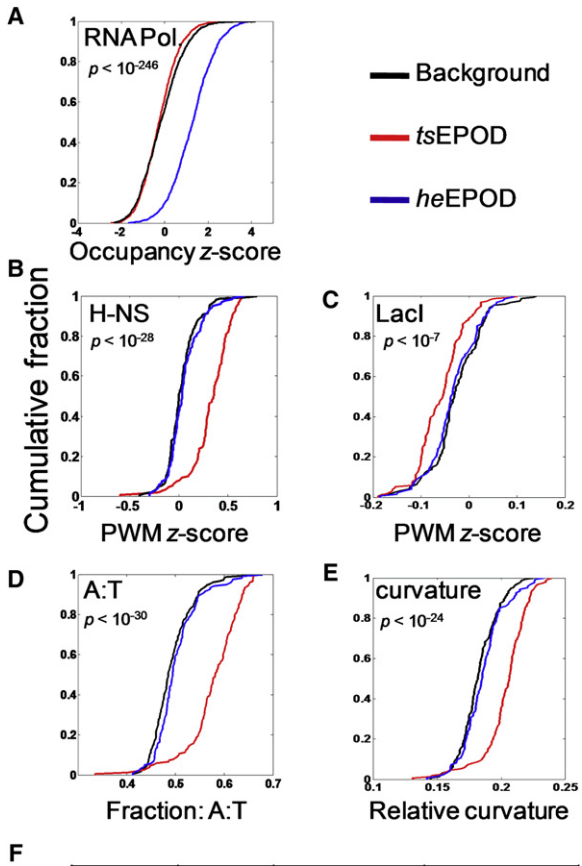
(A) A high-protein-occupancy and high-RNA-expression domain encompassing a region with genes encoding ribosomal protein subunits.

(B) An EPOD automatically detected within a transcriptionally silent region with genes encoding LPS biosynthesis products. A neighboring region encoding 50S ribosomal protein subunits (*rpmG* and *rpmB*) shows an equal level of protein occupancy but shows a high level of RNA expression.

(C and D) Transcriptionally silent EPODs detected within genes of unknown function encoding a predicted protein, *yccE* (C) and a conserved inner membrane protein, *yliE* (D).

chromosomal organizing centers. We argue that the underlying biophysical properties of these regions may largely dictate this role. IHF is known to have a preference for curved DNA, causing it to bend sharply upon binding; the nucleoid proteins HU and H-NS bind strongly to curved DNA as well (Swinger and Rice, 2004). Fis, H-NS, and IHF restrain supercoils (Pettijohn, 1996), and both H-NS (Dame et al., 2000) and Fis (Skoko et al., 2006) show oligomerization and DNA compaction in vitro. We propose that nucleation starts with nucleoid proteins preferentially binding these curved regions of DNA. Because several of the nucleoid proteins prefer to bind curved DNA, these initial protein-DNA interactions make the region more favorable for further binding events. In this way, a wave of nucleoid proteins may spread across these regions, reinforced through the maintenance of curvature and intradomain protein-protein interactions. Homo- and heterodimeric protein-protein interactions, for example, as shown for H-NS (Stella et al., 2005), can then bring these domains in contact with each other, forming the classic rosette structures visualized by EM (Delius and Worcel, 1974a; Postow et al., 2004).

Our observations do not suggest that every *ts*EPOD is essential to chromosomal organization at all times. Rather, a subset of *ts*EPODs could be involved in the formation of higher-order structure in any one cell, or across different environmental conditions. For relevant discussions see Deng et al. (2005), Postow et al. (2004), and Valens et al. (2004). The lack of any discernable fitness deficit for a reduced genome *E. coli* strain, MDS42 (Kolisnychenko et al., 2002), which is missing 24% of the ORFs contained in *ts*EPODs, supports this dynamic and redundant picture. In fact, in vivo protein occupancy display (IPOD) analysis of this reduced genome showed that the occupancy pattern of the remaining EPODs is largely preserved, with 44% of EPOD sequences in MDS42 exactly overlapping those defined in MG1655 (Figure S10). Although there are a minority of loci with substantially different occupancy patterns, most of the residual discrepancy is due to differences in the exact definition of EPOD boundaries and not their locations. These observations provide additional support for our proposed model, namely that specific chromosomal regions, by virtue of their sequence



F

Factor	Total	*ts*EPOD (*P*)	*he*EPOD (*P*)
H-NS	93	49 (2 x 10^{-5})	20 (NS)
IHF	125	30 (7.4 x 10^{-5})	12 (NS)
Fis	200	43 (5.0 x 10^{-2})	69 (2 x 10^{-5})
CRP	68	7 (NS)	5 (NS)

Figure 4. Distinct Occupancy Composition and Biophysical Properties within Extended Protein Occupancy Domains

The cumulative distribution of various measures are shown for transcriptionally silent EPODs (red), highly expressed EPODs (blue), and a matched background control (black). The Wilcoxon rank sum test is used to determine statistical significance of observed deviations relative to background.

(A) Experimentally determined relative RNA polymerase occupancy.

(B and C) Computationally scored PWM binding preference for a nucleoid protein (H-NS) and a nonnucleoid transcriptional repressor (LacI) using a genome-wide relative measure (z-score).

(D) Cumulative distributions of A:T frequencies within EPODs.

(E) Cumulative distribution of predicted relative curvature values within EPODs.

(F) Distribution of binding sites for various nucleoid proteins and CRP within *ts*EPODs and *he*EPODs.

composition, act as extended protein occupancy domains, which in turn may allow them to participate in organizing large-scale chromosomal topology. However, we also raise the possibility that the establishment of these transcriptionally silenced protein occupancy domains may subserve other functions. For example, others have argued for the role of nucleoid proteins such as H-NS in the silencing of horizontally transferred DNA (Dorman, 2007).

A closer inspection of some EPODs suggests that our automated classification of them into the two groups of highly expressed and transcriptionally silent may not capture the full range of their diversity. Indeed, one of the longest *ts*EPODs is defined over a cluster of genes encoding enzymes in the pathway of lipopolysaccharide (LPS) biosynthesis (Figure 3B). Analysis of strand-specific RNA abundance of this locus (Figure S11) clearly shows that, although this region is classified as transcriptionally "silent," there is low-level expression that is mostly confined to the first three genes in the operon (*rfaQ*, *rfaG*, and *rfaP*). These observations suggest that extended protein occupancy may be present at loci with low-level expression, and that it may be caused by processes that are distinct from those operating at absolutely silent loci.

We have developed IPOD, a global, in vivo approach for monitoring the protein occupancy of an entire bacterial genome at the resolution of individual binding sites. Aqueous/organic phase separation has been previously used to enrich on the basis of nucleosome density in *S. cerevisiase* (Nagy et al., 2003), and Grainger et al. demonstrated that crosslinked RNA-polymerase-bound sequences are preferentially partitioned to the organic phase in *E. coli* (Grainger et al., 2006). Here we have shown that localization of small nucleoprotein complexes at the aqueous/organic interface is a simple yet powerful strategy for profiling protein occupancy across an entire prokaryotic genome. Although the identity of the protein bound at each site is not known, increasingly accurate sequence-specificity models of protein-DNA interactions should allow probabilistic assignments to known DNA-binding proteins. In fact, since IPOD analysis allows measurements of correlated occupancy of many sites across different conditions, it should aid in the refinement of existing sequence-specificity models and the discovery of new ones.

The ability to simultaneously monitor both protein occupancy and transcriptional output, at high spatial and temporal resolution, promises to allow true systems-level modeling of transcriptional network dynamics and chromosomal organization. At large spatial scales, these data have revealed the existence of transcriptionally silent protein occupancy domains. Our diverse observations implicate these regions as the long-proposed domain-organizing centers of the *E. coli* chromosome.

EXPERIMENTAL PROCEDURES

Protein Occupancy Display

In Vivo Crosslinking and Footprint Minimization

In vivo formaldehyde crosslinking was performed as in Laub et al. (2002) with minor variations. Batch cultures of *E. coli* MG1655 were grown to early (2.4 × 10^7 CFU/ml) or late (2.4 × 10^8 CFU/ml) exponential phase, in Luria-Bertani medium (0.1% Bacto Tryptone, 0.05% yeast extract, 0.05% NaCl), at which point 30 ml of cells was mixed with 300 μl 1M sodium phosphates (pH 7.6) and 810 μl 37% formaldehyde. Batch cultures of *E. coli* MDS42 were grown to early exponential phase (OD_{600} 0.3). All cultures were grown at 37°C with shaking. Duration of exposure to formaldehyde, at room temperature with shaking, was varied from 5 to 20 min without noticeable differences in crosslinking efficiency; 20 min exposure was used in the experiments presented here. Crosslinking was quenched by addition of 2 ml 2 M glycine. The samples were shaken at room temperature for 10 min and then moved to ice for an

additional 10 min to complete quenching. The cells were pelleted by centrifugation at 5525 × g, 4°C for 4 min and then washed twice with ice-cold 1× phosphate-buffered saline. The remaining liquid was removed, and the cells were frozen in a dry-ice slurry and stored at −80°C for not more than 1 month.

Cell pellets were thawed on ice and resuspended in 500 μl Lysis Buffer A (10 mM Tris [pH 8.0], 20% sucrose, 50 mM NaCl, 10 mM EDTA) with 20 mg/ml of freshly added lysozyme. The samples were incubated at 37°C for 30 min and then mixed with 500 μl 2× IP buffer (100 mM Tris [pH 7.0], 300 mM NaCl, 2% Triton X-100) with 0.7 mg/ml of freshly added PMSF. The samples were incubated at 37°C for an additional 15 min. The cells were pelleted by centrifugation at 850 × g, 4°C, for 3 min, and the supernatants were gently removed by pipetting. The cell pellets were resuspended in 1 ml Lysis Buffer B (5×: 250 mM HEPES [pH 7.5], 2.5 M NaCl, 5 mM EDTA, two Complete Protease Inhibitor Mini tablets [Roche P/N 11393100], filter sterilized) and moved to ice. The cells were sonicated on ice using a Misonix Sonicator 3000 with a microtip at power level 2 for three 10 s pulses, with 10 s rests on ice between pulses. The lysates were clarified by centrifugation at 16,100 × g, 4°C, for 5 min. The supernatants were transferred to a separate tube and stored at −80°C for not more than 1 month.

Sonicated cellular lysates were thawed on ice. Aliquots of 350 μl cell lysate were treated with 5 μl RNaseA (10 mg/ml), 38 μl rDNaseI (38U, Ambion), and 37 μl 10× rDNaseI buffer at 37°C for 1 hr.

Protein-DNA Complex Isolation, Crosslink Reversal, and DNA Labeling

Protein-DNA complexes were isolated by phenol extraction. To achieve this, 150 μl 10 mM Tris and 500 μl 25:24:1 phenol:chloroform:isoamyl alcohol were added to the samples. The samples were vortexed for 10 s and centrifuged at top speed for 2 min at room temperature. A white disk was readily discernible at the aqueous/organic interface. To purify this interface, all aqueous and organic liquid was removed by pipetting. A second extraction was performed by adding 500 μl 10 mM Tris and 450 μl 24:1 chloroform:isoamyl alcohol, vortexing and centrifuging as in the previous step. Again, all liquid was removed from the interface by pipetting, and residual liquid was removed by wicking. For crosslink reversal, the interface was suspended in 500 μl 10 mM Tris and 50 μl 10% SDS and placed at ~100°C for 30 min. The tubes were placed on ice, then moved to 65°C for 3 hr following addition of 5 μl proteinase K (20 mg/ml). After heat treatment, the solutions were phenol/chloroform extracted and ethanol precipitated in the presence of glycogen to purify the DNA. The DNA pellets were resuspended in 50 μl water and quantified using a Nanodrop spectrophotometer. Two micrograms of the fragmented DNA, isolated from DNA-protein complexes, was used as the input in a labeling reaction with the Enzo Terminal Labeling kit (P/N 42630). The labeling reactions were assembled according to the manufacturer's instructions; the labeling reaction was incubated at 37°C for 30 min to 1 hr. No qualitative difference in labeling efficiency was observed between reactions labeled for 30 min or for 1 hr.

Tiling-Array Hybridization

We designed an Affymetrix tiling array for the MG1655 *E. coli* genome, containing probes that cover the entire genome at a resolution of 4 bp between steps; however, the steps alternate strand coverage, so there is an 8 bp step between probes on the same strand. There are a total of 2.47 million probes on the array, of which 2,300,160 directly enter analysis as *E. coli* probe pairs. The Affymetrix system pairs each 25-mer perfect match genomic sequence with a corresponding 25-mer that has a mismatch in the 13th position. The mismatch probe is intended as a crosshybridization control. In addition to the *E. coli* sequences, 33,996 probe pairs are sequence-specific controls against other genomes, including *B. subtilis*, lambda phage, and *A. thaliana*. The tiling array (Ecoli_Tab520346F) is a standard-sized 200 μl volume microarray with 5 μm features. The FS450_0002 washing protocol (Affymetrix) was used. Briefly, this includes two posthybridization washes, a streptavidin-phycoerythrin stain, a poststain wash, an antibody amplification stain, a second streptavidin-phycoerythrin stain, a final wash, and addition of holding buffer to the microarray. After completion of the wash cycle, the microarrays were scanned a single time at room temperature.

Generation of Reference Genomic Hybridizations

As the Affymetrix platform is a "single-color" hybridization system, it was necessary to choose an appropriate reference sample in order to determine relative enrichment/depletion of occupancy across the genome. To accomplish this, we used a whole-genome DNA preparation from the wild-type (MG1655) *E. coli* strain, grown to stationary phase in an overnight culture. The genomic DNA was exposed to a low concentration of DNase I in the presence of cobalt to fragment the DNA to a median size of approximately 300 base pairs. The DNA was labeled as above, and hybridizations were performed with six biological replicates.

RNA Expression Profiling

To obtain tiling-resolution RNA measures across the *E. coli* genome, wild-type cells were grown in biological duplicate in LB to OD_{600} = 0.3. The cultures were moved to ice, and the QIAGEN RNeasy kit (P/N 74104) was used to isolate total cellular RNA. Immediately following elution in RNase-free water, residual DNA was removed by treatment with DNaseI at 37°C for 15 min. The samples were treated again using the RNeasy kit, resuspended in 40 μl RNase-free water, and stored at −20°C.

The RNA samples were reverse transcribed to DNA using the SuperScript system from Invitrogen (P/N 18053017). RNA (10 μg) was incubated with 5 μg random hexamer primer at 70°C for 10 min. The samples were moved to ice and then mixed with 8 μl SuperScript buffer, 4 μl DTT, 2 μl 10 mM dNTP mix, 3 μl DEPC-treated water, and 1 μl RNasin. Then, 2 μl of SuperScript II Enzyme was added, and the samples were incubated at 25°C for 10 min, and 42°C for 2.5 hr, then moved to 95°C for 5 min to terminate the reaction. To fragment the RNA template, 2 μl 1 N NaOH was added, and the samples were placed at 65°C for 15 min. At room temperature, the pH was readjusted by adding 2 μl 1 N HCl. The cDNA was cleaned using the QIAGEN QiaQuick Nucleotide Removal Kit (P/N 28304) and resuspended in 40 μl water. Hybridizations were performed in biological duplicate as above, using 1.3 μg cDNA.

Data Processing and Normalization

Protein Occupancy Display

We developed in-house computational and statistical analysis tools for use with the *E. coli* tiling array. We used a previous study (Choe et al., 2005) as a model for overall design of normalization among arrays and perfect match adjustment on single arrays. Analysis scripts were written in Perl, MatLab, and R to standardize the statistical manipulations across all data sets. The output file from the scanning process is a CEL file; we used a proprietary Affymetrix utility, bpmap_bcel_join, to extract the perfect match and mismatch raw signal intensities from the CEL files. Choe et al. (2005) demonstrated that subtracting the mismatch value from the perfect match value was a simple yet effective method for correcting the perfect match signal for crosshybridization.

In order to quantify relative enrichment/depletion of sequences for protein occupancy, we utilized signals from six independent genomic DNA hybridizations to calculate a z-score. The z-score for each probe is defined as the experimental value for the probe minus the mean of the six references (for that probe), divided by the standard deviation of the six references (for that probe). In exponentially growing cells, multiple origin replication events give rise to inflated hybridization signal in and around the original of replication. This increased intensity around the origin was clearly visible during early exponential growth but decreased significantly for cells in late exponential phase. In order to correct for this, we used a local normalization protocol where the signal from each probe was normalized using the ratio between the mean signal intensity between the experimental signal and the mean of the six reference genomic replicates, calculated within an 80 kb window centered on the data point for that probe. Occupancy z-scores were calculated using these locally normalized values. Therefore, a positive z-score reflects the overrepresentation of a DNA sequence in the pool of DNA-protein complexes relative to genomic DNA, while a negative z-score indicates relative depletion of the DNA sequence in the set of DNA-protein complexes. For many of the analyses, the z-score values where spatially smoothed using a moving average window ranging from 32 to 512 base pairs.

RNA Expression

The *E. coli* Affymetrix tiling array was utilized for RNA expression analysis as described above. The resulting data consisted of PM-MM signal values at the resolution of 8 bp for each strand across the entire *E. coli* genome. The data for two replicate experiments were quantile normalized and averaged to yield mean expression values for each strand. We used linear interpolation

to generate RNA signal at 4 bp resolution for each strand. These strand-specific profiles were then smoothed using a 512 bp moving average window. In order to generate a strand-independent transcriptional output profile, we then took the larger of the two strand's signal at each genomic coordinate. This smoothed, strand-independent transcriptional profile was used for subsequent analysis.

SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, three tables, and eleven figures and can be found with this article online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00479-1.

ACKNOWLEDGMENTS

We thank the members of the Tavazoie laboratory for helpful comments on the manuscript. T.V. was supported by a NASA predoctoral fellowship. A.K.H. was assisted by fellowship #08-1090-CCR-EO from the New Jersey State Commission on Cancer Research. S.T. was supported by grants from the NSF (CAREER), DARPA, NHGRI, NIGMS (P50 GM071508), and the NIH Director's Pioneer Award (1DP10D003787-01). The oligonucleotide array data were deposited at NCBI Gene Expression Omnibus with accession number GSE16414.

Received: December 21, 2008
Revised: April 7, 2009
Accepted: June 24, 2009
Published: July 30, 2009

REFERENCES

Aiyar, S.E., McLeod, S.M., Ross, W., Hirvonen, C.A., Thomas, M.S., Johnson, R.C., and Gourse, R.L. (2002). Architecture of Fis-activated transcription complexes at the Escherichia coli rrnB P1 and rrnE P1 promoters. J. Mol. Biol. *316*, 501–516.

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.

Bendich, A.J. (2001). The form of chromosomal DNA molecules in bacterial cells. Biochimie *83*, 177–186.

Cho, B.K., Knight, E.M., Barrett, C.L., and Palsson, B.O. (2008). Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. Genome Res. *18*, 900–910.

Choe, S.E., Boutros, M., Michelson, A.M., Church, G.M., and Halfon, M.S. (2005). Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. Genome Biol. *6*, R16.

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.

Delius, H., and Worcel, A. (1974a). Electron microscopic studies on the folded chromosome of Escherichia coli. Cold Spring Harb. Symp. Quant. Biol. *38*, 53–58.

Delius, H., and Worcel, A. (1974b). Letter: Electron microscopic visualization of the folded chromosome of Escherichia coli. J. Mol. Biol. *82*, 107–109.

Deng, S., Stein, R.A., and Higgins, N.P. (2005). Organization of supercoil domains and their reorganization by transcription. Mol. Microbiol. *57*, 1511–1521.

Dorman, C.J. (2007). H-NS, the genome sentinel. Nat. Rev. Microbiol. *5*, 157–161.

Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muniz-Rascado, L., Martinez-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.

Garcia-Russell, N., Orchard, S.S., and Segall, A.M. (2007). Probing nucleoid structure in bacteria using phage lambda integrase-mediated chromosome rearrangements. Methods Enzymol. *421*, 209–226.

Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J., and Busby, S.J. (2005). Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. Proc. Natl. Acad. Sci. USA *102*, 17693–17698.

Grainger, D.C., Hurd, D., Goldberg, M.D., and Busby, S.J. (2006). Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. Nucleic Acids Res. *34*, 4642–4652.

Hinnebusch, B.J., and Bendich, A.J. (1997). The bacterial nucleoid visualized by fluorescence microscopy of cells lysed within agarose: comparison of Escherichia coli and spirochetes of the genus Borrelia. J. Bacteriol. *179*, 2228–2237.

Kolisnychenko, V., Plunkett, G., III, Herring, C.D., Feher, T., Posfai, J., Blattner, F.R., and Posfai, G. (2002). Engineering a reduced Escherichia coli genome. Genome Res. *12*, 640–647.

Laub, M.T., Chen, S.L., Shapiro, L., and McAdams, H.H. (2002). Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. Proc. Natl. Acad. Sci. USA *99*, 4632–4637.

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.

Pedersen, A.G., Jensen, L.J., Brunak, S., Staerfeldt, H.H., and Ussery, D.W. (2000). A DNA structural atlas for Escherichia coli. J. Mol. Biol. *299*, 907–930.

Pettijohn, D.E. (1996). The nucleoid. In *Escherichia coli* and *Salmonella*, F. Neidhardt, ed. (Washington, DC: ASM), pp. 158–166.

Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R. (2004). Topological domain structure of the Escherichia coli chromosome. Genes Dev. *18*, 1766–1779.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306–2309.

Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S.M., Marko, J.F., and Johnson, R.C. (2006). Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis. J. Mol. Biol. *364*, 777–798.

Stella, S., Spurio, R., Falconi, M., Pon, C.L., and Gualerzi, C.O. (2005). Nature and mechanism of the in vivo oligomerization of nucleoid protein H-NS. EMBO J. *24*, 2896–2905.

Swinger, K.K., and Rice, P.A. (2004). IHF and HU: flexible architects of bent DNA. Curr. Opin. Struct. Biol. *14*, 28–35.

Tavazoie, S., and Church, G.M. (1998). Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in E. coli. Nat. Biotechnol. *16*, 566–571.

Travers, A., and Muskhelishvili, G. (2007). A common topology for bacterial and eukaryotic transcription initiation? EMBO Rep. *8*, 147–151.

Valens, M., Penaud, S., Rossignol, M., Cornet, F., and Boccard, F. (2004). Macrodomain organization of the Escherichia coli chromosome. EMBO J. *23*, 4330–4341.