Analysis of phage Mu DNA transposition by whole-genome *Escherichia coli* tiling arrays reveals a complex relationship to distribution of target selection protein B, transcription and chromosome architectural elements

JUN GE, ZHENG LOU, HONG CUI, LEI SHANG and RASIKA M HARSHEY*

Section of Molecular Genetics and Microbiology & Institute of Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

**Corresponding author (Fax, 512-471-7088; Email, rasika@uts.cc.utexas.edu)*

Of all known transposable elements, phage Mu exhibits the highest transposition efficiency and the lowest target specificity. *In vitro*, MuB protein is responsible for target choice. In this work, we provide a comprehensive assessment of the genome-wide distribution of MuB and its relationship to Mu target selection using high-resolution *Escherichia coli* tiling DNA arrays. We have also assessed how MuB binding and Mu transposition are influenced by chromosome-organizing elements such as AT-rich DNA signatures, or the binding of the nucleoid-associated protein Fis, or processes such as transcription. The results confirm and extend previous biochemical and lower resolution *in vivo* data. Despite the generally random nature of Mu transposition and MuB binding, there were hot and cold insertion sites and MuB binding sites in the genome, and differences between the hottest and coldest sites were large. The new data also suggest that MuB distribution and subsequent Mu integration is responsive to DNA sequences that contribute to the structural organization of the chromosome.

[Ge J, Lou Z, Cui H, Shang L and Harshey RM 2011 Analysis of phage Mu DNA transposition by whole-genome *Escherichia coli* tiling arrays reveals a complex relationship to distribution of target-selection protein B, transcription and chromosome architectural elements. *J. Biosci.* **36** 587–601] **DOI** 10.1007/s12038-011-9108-z

1. Introduction

In their long and successful evolutionary history, transposable elements (TEs) have learnt to balance self-propagation with host survival. Regulated expression of TE proteins and their judicious use of target sites is one aspect of this balancing act, while DNA accessibility controlled by genome-wide events like replication, transcription and chromosome-organizing proteins is the other (Craig 1997; Wu and Burgess 2004; Berry *et al.* 2006). While molecular details of target selection have been studied *in vitro* for many individual TEs, *in vivo* studies are beginning to reveal the influence of host strategies in controlling transposition events (Lewinski *et al.* 2006; Beauregard *et al.* 2008; Brady *et al.* 2009; Parks *et al.* 2009; Guo and Levin 2010). An understanding of *in vivo* target site selection mechanisms is essential to understanding how TEs have shaped genome evolution.

Transposable element Mu is also a phage with a temperate lifestyle, which utilizes repeated rounds of replicative transposition to multiply its genome during the lytic phase of growth (Symonds *et al.* 1987). Efficient transposition involves the phage-encoded transposase MuA and the ATP-dependent DNA-binding protein MuB, which directs the transposition complex to selected DNA targets. Several insights into the target selection strategies of Mu are at hand, largely from *in vitro* studies. MuA is responsible for recognition of a 5 bp weak target consensus

Keywords. A-tracts; MuB; Mu DNA transposition; Fis; nucleoid-associated proteins; target site selection

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/Sep2011/pp587–601/suppl.pdf*

of 5′-NY(G/C)RN-3′ *in vitro* (Mizuuchi and Mizuuchi 1993; Haapa-Paananen *et al.* 2002), which is also observed *in vivo* (Manna *et al.* 2005). *In vitro*, MuB is responsible for delivering target DNA to MuA via MuA-MuB interactions (Chaconas and Harshey 2002). MuB binds non-specifically and co-operatively on DNA, exhibiting a tendency to form larger polymers or filaments on A/T-rich DNA (Adzuma and Mizuuchi 1991; Greene and Mizuuchi 2004). MuB-bound DNA is preferentially chosen for Mu integration (Adzuma and Mizuuchi 1988; Mizuuchi and Mizuuchi 1993). Analysis of DNA sequences surrounding Mu insertion peaks *in vitro* suggests that MuB likely binds in an interspersed manner to locally A/T-rich segments in natural DNA, preventing integration in the bound regions but directing integration to adjacent sites free of MuB (Ge and Harshey 2008).

A new role of MuB has recently emerged in providing immunity to integration within Mu ('Mu genome immunity'), where it binds strongly during active Mu replication (Ge *et al.* 2010). This property of MuB is consistent with its observed behaviour on synthetic A/T-only DNA, where strong MuB binding occluded the entire bound region from insertions, these being directed to the junction of MuB-bound/unbound regions (Ge and Harshey 2008). However, the Mu genome is not A/T-rich, suggesting that other cellular features are responsible for MuB binding patterns *in vivo*. The paradoxical properties of MuB in both promoting and preventing Mu integration suggest that an analysis of MuB distribution and Mu insertion patterns might provide useful insights into transposon-host target selection strategies. While a microarray analysis of Mu transposition targets in *E. coli* has been reported using an ORF array (Manna *et al.* 2004), MuB distribution has not been analysed.

We report here a detailed analysis of MuB binding and Mu integration patterns using chromatin immunoprecipitation (ChIP) coupled with DNA hybridization to *E. coli* genome tiling arrays (Ren *et al.* 2000). The data confirm *in vitro* results, showing a positive correlation between MuB binding and Mu transposition. However, no apparent relationship emerged between MuB binding, transposition and transcription. These and other results, including a similar analysis in a strain deleted for the nucleoid-organizing protein Fis, suggest that MuB binding is modulated by some aspect of chromosome architecture.

2. Materials and methods

2.1 *Bacterial strains and plasmids*

Mu lysogenic *E. coli* strain HM8305 (F′ *pro*$^+$ *lac*:Mu *cts*62/Δ*pro lac his met rpsL* Mu^r) (Bukhari and Taylor 1975) was used to construct CW28 (HM8305 ΔMu*B*) and its derivative SJG18 (CW28 Δ*fis*). Both of the mutants were generated by the λ Red recombination system (Datsenko and Wanner 2000). Plasmid pJG8 contains 9*c-myc* epitope tag fused to the N-terminus of Mu*B* in pIL164 (Lee and Harshey 2001) with the Mu*A* gene deleted.

2.2 *Phage purification and phage DNA extraction*

Procedures for Mu prophage induction, phage purification and DNA isolation were as previously described (Au *et al.* 2006; Ge *et al.* 2010).

2.3 *Chromatin immunoprecipitation*

CW28 and SJG18 carrying pJG8 were used to perform ChIP experiments. Cell growth, phage induction and ChIP procedures for MuB were as described; samples were prepared 40 min after prophage induction, at which time point Mu is actively replicating (Ge *et al.* 2010).

2.4 *DNA samples preparation for the analysis of Mu transposition targets*

For Mu insertion sites analysis, cells were allowed to lyse completely, and isolated phage DNA was treated as follows: 15 μg phage DNA was completely digested by PvuI at 37°C for 4 h; the digestion products separated by electrophoresis on 1% agarose gel, and DNA fragments between 1.5–3.5 kb, which contain MuR-end linked to host DNA, were purified using Qiaquick Gel Extraction Kit® (Qiagen). Purified DNA was suspended in 250 μl TE buffer (pH 7.5), and sonicated twice for 15 s each in an ice bath using SONICS Vibra cell® (SONICS, model VC 505). The sizes of sheared DNA ranged from 300 to 1000 bp with an average of 500 bp. This DNA was purified with Qiaquick PCR purification Kit® (Qiagen).

To prepare reference DNA for Mu insertion sites analysis, chromosomal DNA from CW28 and SJG18 strains was purified with Wizard® Genomic DNA Purification Kit (Promega). The purified DNA was diluted to 20 ng/μl with TE buffer to a final volume of 500 μl and subjected to sonication 4 times for 15 s each in an ice bath. Sheared DNA size and purification was as described above. The DNA was labelled directly with Cy5, i.e. without PCR amplification.

2.5 *DNA amplification*

A two-step protocol was used to amplify the DNA samples. In the first or round-A step, a 10 μl reaction mixture containing 40 ng DNA (either ChIP sample, Input DNA from ChIP sample, or reference DNA for Mu targets analysis), 2 μl 5× sequenase buffer and 1 μl round-A

primer (40 μM) (5′-GTTTCCCAGTCAC-GATCNNNNNNNNN; this is essentially a random primer and hence constitutes both forward and reverse primers; the specific sequence has no match in the *E. coli* genome, so it serves as an efficient tag during round-B PCR) was prepared. The mixture was heated at 94°C for 2 min, cooled to 8°C and held for 2 min in a PCR machine (PTC-2000; MJ Research). Next, 5 μl of round-A mixture (1 μl of 5× sequenase buffer, 1.5 μl of dNTP mix (5 mM each), 0.75 μl of DTT (0.1 M), 1.5 μl of BSA (0.5 mg/ml) and 0.3 μl of Sequenase (13 U/μl; USB) was added. The mixture was then ramped from 8°C to 37°C over 8 min, held for 8 min, heated for 2 min at 94°C and cooled to 8°C. After the addition of 1 μl diluted Sequenase (0.3 μl of Sequenase, 0.7 μl sequenase dilution buffer), the mixture was again ramped from 8°C to 37°C over 8 min, held for 8 min, heated for 2 min at 94°C and cooled to 4°C. 45 μl of TE was added to the mixture to bring the volume to about 60 μl. Round-B reaction mixture was composed of 15 μl of round-A reaction products, 20 μl of 5× Phusion HF buffer (New England Biolabs), 1 μl of round-B primer (500 μM) (5′-GTTTCCCAGTCACGATC; *see* round-A primer description above), 2 μl of dNTP (10 mM each), 1 μl of Phusion Hot Start DNA Polymerase (2 U/μl; NEB) and 61 μl of H_2O. The mixture was heated for 2 min at 98°C, cycled for 15 s at 98°C, 30 s at 40°C, 30 s at 50°C, and 15 s at 72°C (30 cycles total), incubated for 5 min at 72°C and then cooled to 4°C. Amplified DNA was purified using Qiaquick PCR purification Kit.

2.6 *DNA labelling and microarray hybridization*

A whole-genome tiling array for ChIP-on-chip assay from NimbleGen was used for microarray analysis. The microarray contains the whole *E. coli* MG1655 genome (NC_000913.1) arranged on one slide into 386486 contiguous 50 bp oligonucleotide sequences overlapping by 26 bp, every 24 bp on average. The procedures of DNA labelling are described in the NimbleGen Arrays User's Guide for ChIP-chip Analysis. Cy5 and Cy3-labelled random 9-mers (Trilink Biotechnologies) were employed. Sample DNA (ChIP or processed Mu DNA) was amplified with Cy5-9mer primer, and reference DNA (Input or whole genome DNA) with Cy3-9mer primer. The samples were loaded on microarray slides and subjected to standard hybridization procedures (NimbleGen Arrays User's Guide). Arrays were scanned using GenePix 4000B (Molecular Devices).

2.7 *Data processing*

The ratio value of each probe (fluorescence intensity of Cy5 over Cy3) is the relative enrichment of that probe sequence; for ChIP-chip it is termed the relative MuB binding preference, or BBP, and for transposition target analysis it is termed the transposition target preference, or TTP. The raw data from scanning were log-scaled and normalized with the Tukey biweight mean using NimbleScan software (NimbleGen). The average $\log_2$ BBP or $\log_2$ TTP from three independent biological experiments, each containing probes representing forward and reverse strands of the genome on the slides, were used to identify MuB binding peaks or Mu transposition target peaks using NimbleScan software (NimbleGen). Data were visualized using SignalMap software (NimbleGen). Other analyses were performed with Matlab r2007a software (The MathsWorks), Microsoft Excel 2007 (Microsoft) and algorithms written with Perl.

2.8 *Mu transposition targets and MuB binding loci affected by fis deletion*

To map the target loci affected by the *fis* deletion, a differential profile was generated by subtracting values in the *fis*$^-$ target profile from those in the wild-type target profile for each probe. The mean and standard deviation were then calculated for this differential profile, and those probes with values that fell out of the range of mean ± 3× standard deviation were considered as presenting significant differences between wild type and *fis*$^-$ backgrounds. These probes were then mapped to the *E. coli* genome. To reduce the array noise interference, only those loci with five continuous probes were identified as gene loci affected by the *fis* deletion. The MuB-binding loci were identified by a similar procedure.

3. Results and discussion

3.1 *Genome-wide distribution of MuB and Mu insertions*

c-myc-MuB, which is functionally indistinguishable from wild-type MuB (Ge and Harshey 2008; Ge *et al.* 2010), was used to measure genome-wide MuB binding. It was supplied from plasmid pJG8 to a Mu*B*$^-$ lysogen CW28 for ChIP-chip analysis (supplementary figure 1). Whole-genome tiling microarrays and ChIP sample preparation are described in Materials and methods; the array was hybridized with Cy5-labelled ChIP DNA and Cy3-labelled whole-genome DNA, both amplified by random primers. The 'ratio' value of each probe (fluorescence intensity of Cy5 over Cy3) is the relative enrichment of that probe sequence in the ChIP sample, referred to as the relative MuB binding preference, or BBP. A scatter plot of $\log_2$ BBP values from three biologically independent experiments showed that the observed variation between experiments for most of the probes is acceptable because the three data groups are in an apparent linear relationship with their average values (supplementary

figure 2A). The average $\log_2$ BBP value for each probe calculated from both forward and reverse strands on three slides was used for subsequent analysis. $\log_2$ BBP values have a median of −0.01, mean of 0.03 and standard deviation of 0.31 (supplementary figure 2B, left), and ranged from −1.13 to 5.35, representing an 89-fold difference in MuB binding on preferred (hot) and non-preferred (cold) 50 bp probes (figure 1). For 95% of the genome, $\log_2$ BBP values were within 2-fold deviation (above or below) from the median, whereas for 99% of the genome $\log_2$BBP values fell within 3-fold deviation apart from the median. $\log_2$ BBP values above and below 1-fold deviation apart from the median were considered strong and weak MuB binding, respectively.

Mu insertion sites were analysed by using phage DNA from the same strain used for ChIP, as detailed in Materials and methods. The method takes advantage of the fact that Mu DNA packaged in the phage particles is linked to its insertion sites (Ge *et al.* 2010). Briefly, *E. coli* DNA linked to the right end of the Mu genome were labelled with Cy5 and hybridized to the tiling array with the *E. coli* genome DNA labelled with Cy3. The ratio of Cy5 *vs* Cy3 was the relative efficiency of that probe sequence used as a transposition target, named as transposition target preference, or TTP. The average $\log_2$ TTP value for each probe was calculated similarly to BBP (supplementary figure 2C). $\log_2$ TTP values have a median of −0.02, mean of −0.12 and standard deviation of 0.87 (supplementary figure 2B, right), and ranged from −3.87 to 3.16, representing a 131-fold difference in Mu transposition on hot and cold probes (figure 1). For 94% of the genome, $\log_2$ TTP values were within 2-fold deviation (above or below) from the median, whereas for 99% of the genome $\log_2$ BBP values fell within 3-fold deviation apart from the median. As with BBP, $\log_2$ TTP values above and below 1-fold deviation apart from the median were considered hot and cold Mu insertion sites, respectively. Although both $\log_2$ BBP and $\log_2$ TTP data seem slightly skewed from the normal distribution, the assumption of normality still applies, since the data size is very large and most of the values are in the normal range.

Overall, the data in figure 1 show that both TTP and BBP are generally randomly distributed across the whole genome, consistent with previous observations that Mu transposes randomly, and that MuB is a non-specific DNA-binding protein. Despite the general random nature of Mu transposition and MuB binding, there are hot and cold insertion sites and MuB binding sites in the genome, and differences between the hottest and coldest sites are large.

3.2 *Complex relationship between MuB binding and Mu transposition*

To determine the relationship of MuB binding to Mu transposition, $\log_2$ BBP values were partitioned into 13 equal intervals of 0.5. The binned average $\log_2$ TTP value for each group was plotted against the average $\log_2$ BBP value of that group (figure 2A). The scatter plot showed a positive relationship between these two values, i.e. MuB binding positively modulates Mu transposition target choice in a narrow range of $\log_2$ TTP (around 3-fold). To see whether target choice coincides with MuB binding on a larger genome-wide scale, moving medians of $\log_2$ TTP and $\log_2$ BBP within windows of 150 kb and steps of 1 kb were plotted against genome location (figure 2B); a weak but significant negative correlation was observed (r=−0.2621). Both MuB binding and transposition values were seen to be grouped into domains with high and low activity. With a few exceptions, the two values were negatively correlated throughout the genome, with peaks and troughs of MuB binding generally conflicting with those of transposition. A close examination revealed that these peaks and troughs are not exactly matched, but are slightly offset with respect to each other. The appearance of discrepancy between the two sets of data is due to the fact that figure 2A plots all TTP and BBP values without consideration of whether a specific TTP value is physically matched to its corresponding BBP value, while figure 2B looks at specific locations within 150 kb moving windows. Thus, even though overall target preference is positively related to MuB binding, a wider lens reveals that Mu transposition and MuB binding are not co-incident.

To further examine the relationship between BBP and TTP, 25 highest MuB binding peaks and hot and cold transposition sites were compared. These data showed that while there was a general agreement between MuB binding and transposition, there were several deviations from it (tables 1 and 2). For example, 12 genes which had highest MuB binding peaks (*mmuM, tsx, yegT, rcsB, gatZ, yegV, ymfR, subB, deaD, rpsF, cspE* and *lpp*) correspond to hot transposition peaks, although these 12 are not among the 25 hottest transposition peaks in the genome (listed in table 2). Figure 3A shows a detailed distribution of Mu insertions in *mmuM*. Six of the 25 MuB peaks (*tyrS, cspA, fliA, tnaL, argZ* and *purH*) do not correspond to a hot target directly, but there are hot transposition peaks in their vicinity (within 0.5 and 2.5 kb). Figure 3B shows details within *cspA*. The remaining seven MuB peaks (*acpP, rplT, glyV, leuZ, rpmJ, infC* and *aceE*) all fall within long stretches of cold target areas, but correspond to the 'relatively hot' regions in these areas. Figure 3C shows *rplT* details.

In summary, the data presented in this section show that while there is a generally positive relationship between MuB binding and Mu transposition, Mu insertions are offset with respect to MuB binding. These data support *in vitro* results which showed that Mu transposition occurs on either side of a MuB-protected region on a plasmid (Mizuuchi and

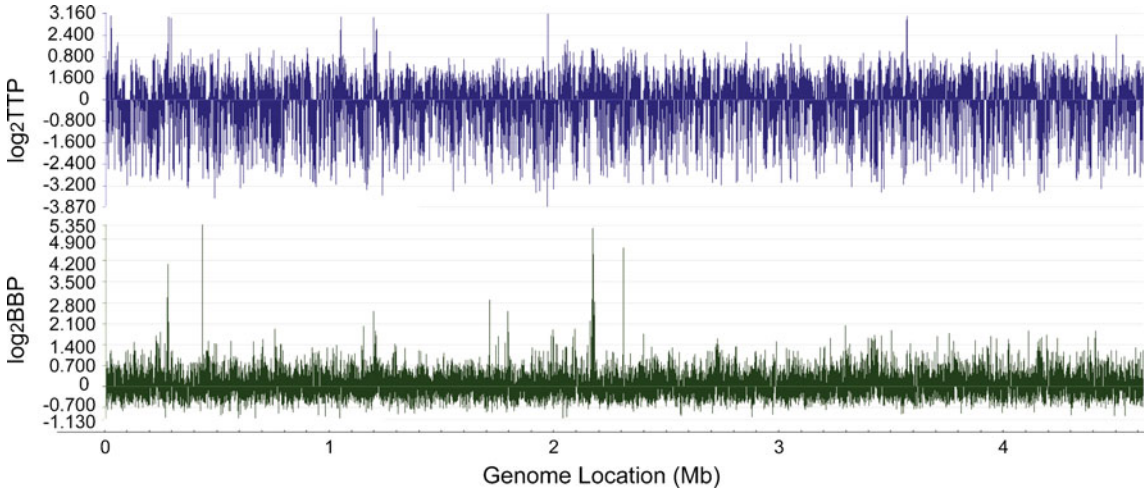

Figure 1. Genome wide mapping of Mu transposition targets and MuB binding on the *E. coli* chromosome. Top panel, an overview of Mu transposition target profile. The values of $\log_2$TTP were plotted against the location of each probe in the 4.64 Mb *E. coli* genome. These values range from −3.98 to 3.20. Bottom panel, an overview of MuB binding profile. The values of $\log_2$BBP were plotted as in the top panel, and range from −1.13 to 5.36.

Mizuuchi 1993), and that MuB bound to synthetic A/T-only DNA promotes integration adjacent to but not within the bound region (Ge and Harshey 2008). All of these results paint the coherent picture of Mu transposition occurring adjacent to and away from but not within MuB-bound DNA.

3.3 *MuB binding is positively related to A/T content and negatively related to CGG frequency*

MuB binding is influenced by the A/T content of DNA *in vitro* (Greene and Mizuuchi 2004; Tan *et al.* 2007; Ge and Harshey 2008). However, genes with a high frequency of

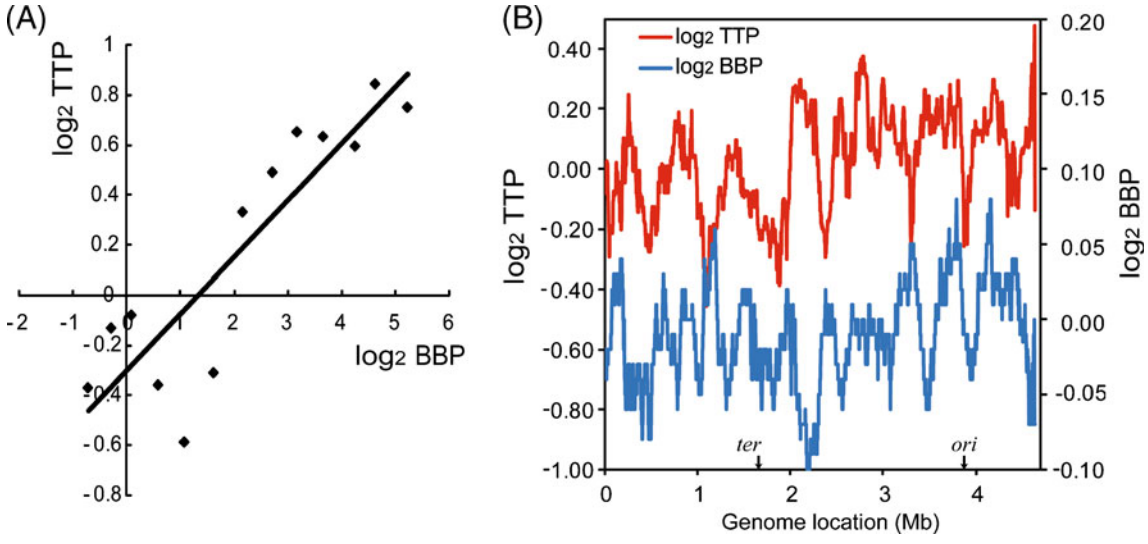

Figure 2. The relationship of TTP and BBP. (**A**) TTP is positively related to BBP values in the scale of short probes (50 bp). $\log_2$ BBP values ranging from −1.13 to 5.35 were grouped into 13 groups with an equal interval of 0.5 (i.e. −1.13~−0.63, −0.63~−0.13....). The averages of $\log_2$ BBP and $\log_2$ TTP in each group were calculated and plotted against each other. A linear trend line was added showing the positive correlation between $\log_2$ BBP and $\log_2$ TTP. (**B**) In the scale of 150 kb, $\log_2$ TTP is negatively related to $\log_2$ BBP. The moving median of $\log_2$ BBP and $\log_2$ TTP are both with a 150 kb window and 1 kb step, and were plotted against the location of the start of the first probe in each window. Position of the replication origin and termination are labelled *ori* and *ter*, respectively. The Pearson correlation coefficient is −0.2621.

Table 1. A list of 25 MuB binding peaks, arranged in accordance with their log_2BBP scores

ID	START	END	SCORE	Name	bnum
0	274480	276449	1.7	*mmuM*	b0261
1	430474	431171	1.7	*tsx*	b0411
2	1150750	1151231	1.7	*acpP*	b1094
3	1715057	1715202	1.7	tyrS	b1637
4	1797488	1798185	1.7	*rplT*	b1716
5	2175874	2177627	1.7	*yegT*	b2098
6	2312108	2315397	1.7	*rcsB*	b2217
7	3717935	3718224	1.55	*cspA*	b3556
8	2171890	2175731	1.51	*gatZ*	b2095
9	4390394	4390683	1.51	*glyV*	b4162
10	1999786	2000275	1.49	*fliA*	b1922
11	2177914	2180627	1.47	*yegV*	b2100
12	1989873	1990138	1.45	*leuZ*	b1908
13	3440478	3441055	1.45	*rpmJ*	b3299
14	3886442	3886611	1.45	*tnaL*	b3707
15	1204622	1204815	1.34	*ymfR*	b1150
16	2151536	2151777	1.3	*sibB*	b2074
17	3305802	3306283	1.3	*deaD*	b3162
18	1798352	1798881	1.28	*infC*	b1718
19	2816064	2816665	1.28	argZ	b2696
20	4423043	4423188	1.28	*rpsF*	b4200
21	124240	124961	1.26	*aceE*	bO115
22	656533	656726	1.26	*cspE*	b0623
23	4205797	4205918	1.26	*purH*	b4006
24	1755357	1755502	1.25	*lpp*	b1677

Start, start position of the peak; End, end position of the peak; Score, log_2 ratio of the BBP peak, generated by NimbleScan Software; Name, name of the gene within which the MuB binding peak falls; bnum, NCBI ID of the genes.

CGG are preferred Mu targets *in vivo* (Manna *et al.* 2005). N-CGG-N is a MuB-independent subset of the target consensus N-YSR-N [N(T/C)(G/C)(G/A)N] determined *in vitro* (Mizuuchi and Mizuuchi 1993; Haapa-Paananen *et al.* 2002). To determine the relationship between these two parameters and MuB binding *in vivo*, the average A/T percentage and additive CGG counts (the total number of CGG tri-nucleotides within the given DNA sequence) of 100 continuous array sequences (or probes) covering the genome were plotted against the additive log_2 BBP values of the same regions, respectively (figure 4). With a few exceptions, A/T percentage was positively correlated to log_2BBP (figure 4A), indicating that MuB prefers to bind DNA with high A/T content. In contrast, CGG count was negatively correlated to log_2 BBP (figure 4B), indicating that MuB avoids DNA with high CGG frequency. These data are consistent with a limited earlier analysis of Mu insertions in two hot target genes *in vivo* (Ge and Harshey 2008). CGG is the central region of the 5 bp Mu target site consensus whose selection is a function of the transposase MuA. By binding A/T and G/C sequences separately, the MuB and MuA proteins appear to function in a mutually synergistic manner in target selection.

3.4 *Two distinctive patterns of MuB binding are related to corresponding cold and hot regions of transposition*

Genome-wide, we observed two striking MuB binding patterns characterized by low but continuous binding, which we have named 'platform' binding. There are seven ribosomal RNA (*rrn*) operons in *E. coli* (Hillebrand *et al.* 2005), all seven of which displayed an identical such pattern (figure 5A, B). The two operons shown – *rrnA* and *rrnD* – are oppositely oriented with respect to each other on the genome (Condon *et al.* 1995; Nomura 1999). The binding regions are large (>5 kb), and this long platform pattern corresponds to a stretch or cluster of cold transposition sites. Within each *rrn* operon, however, there were spikes showing relatively higher MuB binding and transposition compared with the surrounding regions. These spikes correspond to the boxB-A region between 16s rRNA and 23s rRNA genes, which serves as the anti-terminator for 23s rRNA transcription. This profile of platform binding and cold insertion sites is not *rrn*-specific because three other regions showed a similar pattern. Of these, one is an area rich in pseudogenes and also contains the gene for Ile tRNA (figure 5C). The second is a complex region bordering an *IS*2 insertion *ins*CD-4, which contains a remnant of an ETT2 (type III secretion system) pathogenicity island (Ren *et al.* 2004), and putative members of the NarL family of response regulators (figure 5D). The third region encodes genes involved in ribosome function – *tufA* (chain elongation factor EF-Tu), *fusA* (chain elongation factor EF-G), *rpsG* (30S ribosomal subunit protein S7) and *rpsL* (30S ribosomal subunit protein S12) (figure 5E). A hallmark of *rrn* operons is that they are highly transcribed, although not all the operons have similar transcription profiles (Condon *et al.* 1992). However, compared with the high transcription activity of genes in the A, B and E panels in figure 5, those in panels C and D have average transcription profiles (Wei *et al.* 2001). Therefore, the patterns of MuB binding and Mu insertion cannot be directly correlated to transcriptional activity alone in these regions.

The second platform pattern of MuB binding is similar to the first except that the BBP values are higher, and it undulates, i.e. rises and falls, within a larger range of BBP (figure 6). Regions showing this pattern are transposition hot sites. All seven *IS*1 elements show this pattern (figure 6A, B; only 2/7 *IS*1 elements are shown). A common

Table 2. A list of 25 most preferred and 25 least preferred Mu targets, arranged according to their $\log_2$ TTP scores

Hot transposition target peaks						Cold transposition target peaks					
ID	START	END	SCORE	Name	bnum	ID	START	END	SCORE	Name	bnum
1	19802	20283	2.84	*insB_1*	b0021	1	1973761	1973930	−3.51	*motB*	b1889
2	3581732	3582189	2.84	*insB_6*	b3445	2	3881641	3881810	−3.46	*dnaA*	b3702
3	289864	290417	2.79	*insB_3*	b0274	3	3468398	3468543	−3.4	*tufA*	b3338
4	1976545	1976786	2.79	*insB_5*	b1893	4	4195185	4195354	−3.4	*nudC*	b3996
5	278392	278681	2.74	*insB-2*	b0263	5	1236288	1236409	−3.35	*ycg8*	b1188
6	1049482	1049747	2.74	insB_*4*	b0988	6	4174775	4174920	−3.35	*secE*	b3981
7	22514	22659	2.38	*ileS*	b0026	7	3480066	3480283	−3.24	*yheS*	b3352
8	4516775	4517256	2.08	*insA_7*	b4294	8	483662	483783	−3.19	*acrB*	b0462
9	2062706	2062875	2.03	*cobS*	b1991	9	2002228	2002589	−3.19	*fliD*	b1924
10	2061842	2062515	1.98	*cobT*	b1991	10	3333484	3334493	−3.19	*murA*	b3189
11	51200	51417	1.93	*apaG*	b0049	11	1552612	1552733	−3.14	*maeA*	b1478
12	3103916	3104517	1.83	*nupG*	b2964	12	4449949	4450142	−3.14	*ytfR*	b4230
13	48080	48753	1.77	*kefC*	b0047	13	94377	94642	−3.08	*murE*	b0086
14	500178	500563	1.77	*gsk*	b0477	14	2443970	2444451	−3.08	*mepA*	b2328
15	2052323	2052876	1.77	*amn*	b1982	15	2750846	2751039	−3.08	*smpA*	b2617
16	2347485	2347798	1.77	*yfaH*	b2238	16	3020246	3020511	−3.08	*recN*	b2881
17	3981464	3981609	1.77	*asIB*	b3800	17	3566418	3566539	−3.08	*glyC*	b3429
18	378001	378338	1.72	*frmB*	b0355	18	4175399	4175544	−3.08	*secE*	b3981
19	2862792	2863225	1.72	*ygbN*	b2740	19	366289	366434	−3.03	*lacI*	b0345
20	898140	898525	1.67	*ybjS*	b0859	20	1167862	1167983	−3.03	*ycfQ*	b1111
21	2050859	2051100	1.67	*shiA*	b1981	21	1958737	1959026	−3.03	*argS*	b1876
22	2054267	2054412	1.67	*amn*	b1983	22	3923065	3923498	−3.03	*mnmG*	b3741
23	2861904	2862505	1.67	*ygbM*	b2739	23	612364	612605	−2.98	*fes*	b0585
24	3697770	3698155	1.67	*ldrD*	b4453	24	2399955	2400124	−2.98	*nuoE*	b2285
25	3860195	3860532	1.67	*glvB*	b3682	25	2887306	2887595	−2.98	*cysI*	b2763

Score, $\log_2$ ratio of the TTP peaks. Other descriptions as in table 1.

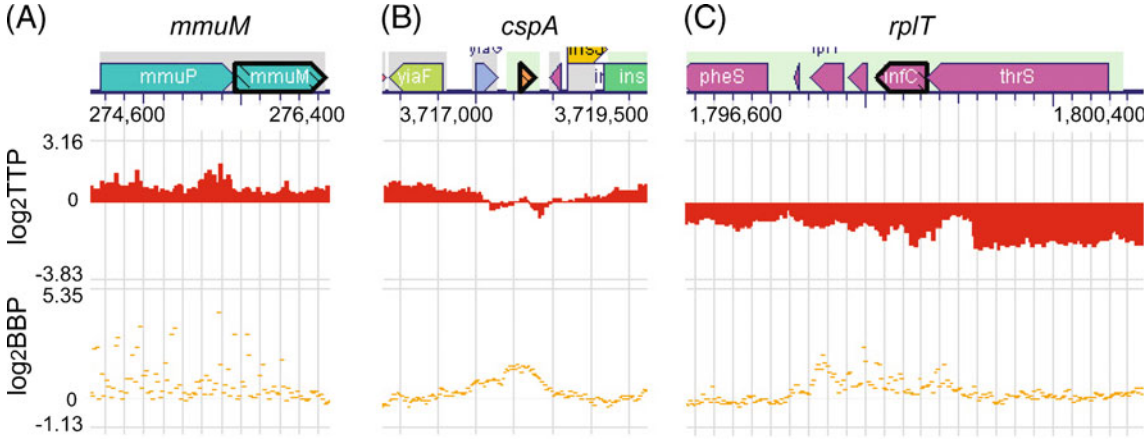

Figure 3. Mu transposition patterns in three genes listed in table 1 showing high levels of MuB binding. (**A**) *mmuM* shows uniformly high levels of both Mu transposition and MuB binding, but does not fall within the 25 hottest genes listed in table 2. (**B**) *cspA* shows an inhibition of transposition in the region of the MuB binding peak, but high transposition on either side of it. (**C**) *rplT* lies in a stretch of the genome which is 'cold' but this gene itself receives more insertions than its immediate neighbours.

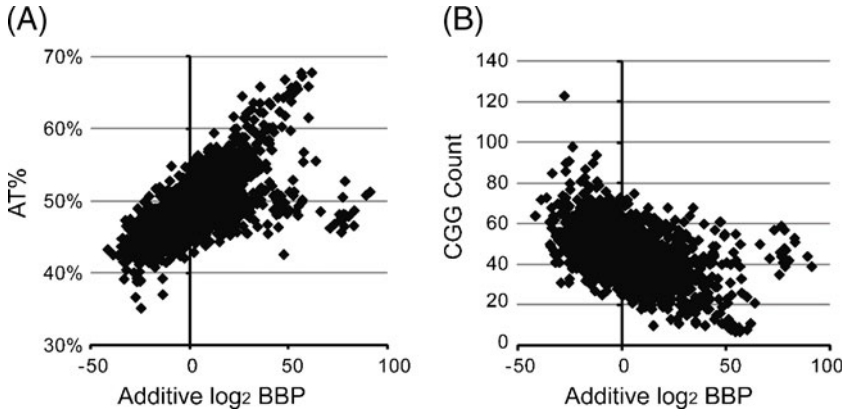

Figure 4. Effect of AT-content and CGG-frequency of target DNA on MuB binding. (**A**) The additives of $\log_2$BBP of 100 continuous tiling array sequences or 'probes' covering the whole genome were plotted against the average AT% of each region. (**B**) As in **A**, except they were plotted against the CGG count in each region.

attribute of all the *IS*1 elements is that they span a short region (~600 bp) and have low transcription profiles (Wei *et al.* 2001). However, a long (>10 kb) pseudogene region also shows this pattern (figure 6C). Here, several hot insertion spikes coincided with MuB binding peaks. Transcription is high throughout this region. Thus, again, the patterns of MuB binding and Mu insertion are not correlated to transcription. These data also show that transcription per se does not block Mu transposition.

A recent technology called IPOD (*in vivo* protein occupancy display), which reveals protein occupancy across the bacterial chromosome at the resolution of individual

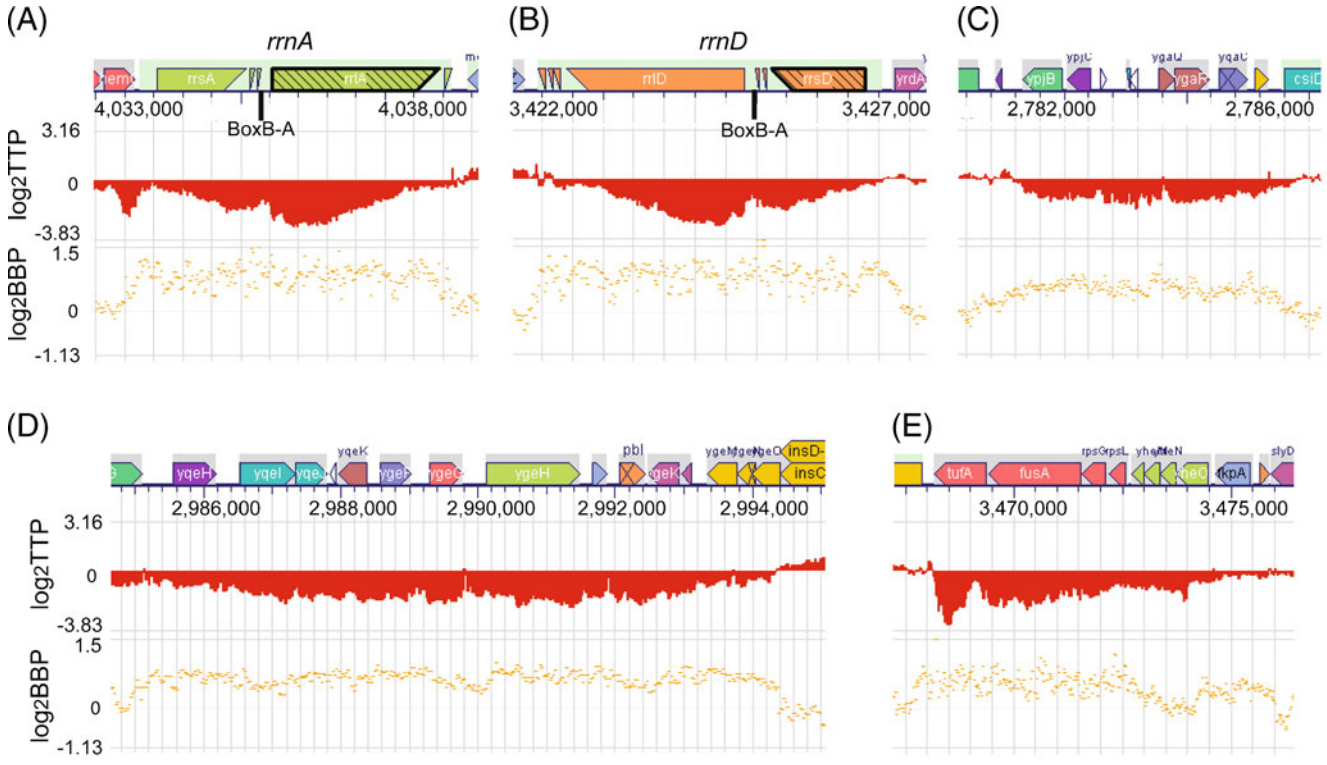

Figure 5. 'Platform' MuB binding patterns and cold transposition sites. (**A**) *rrnA* operon, containing *rrsA, ileT, alaT, rrlA*, and *rrfA*. (**B**) *rrnD* operon, containing *rrfF, thrV, rrfD, rrlD, alaU, ileU, rrsD*. (**C**) *ypjB, ypjC, b2651, IleY, ygaQ, ygaR, ygaC, ygaD*. (**D**) *yqeH, yqeI, yqeJ, yqeL, yqeK, ygeF, ygeG, ygeH, ygeI, pbl, ygeM, ygeN, ygeO*. (**E**) *tufA*, *fusA*, *rpsG*, *rpsL*. See text for description of platform binding.

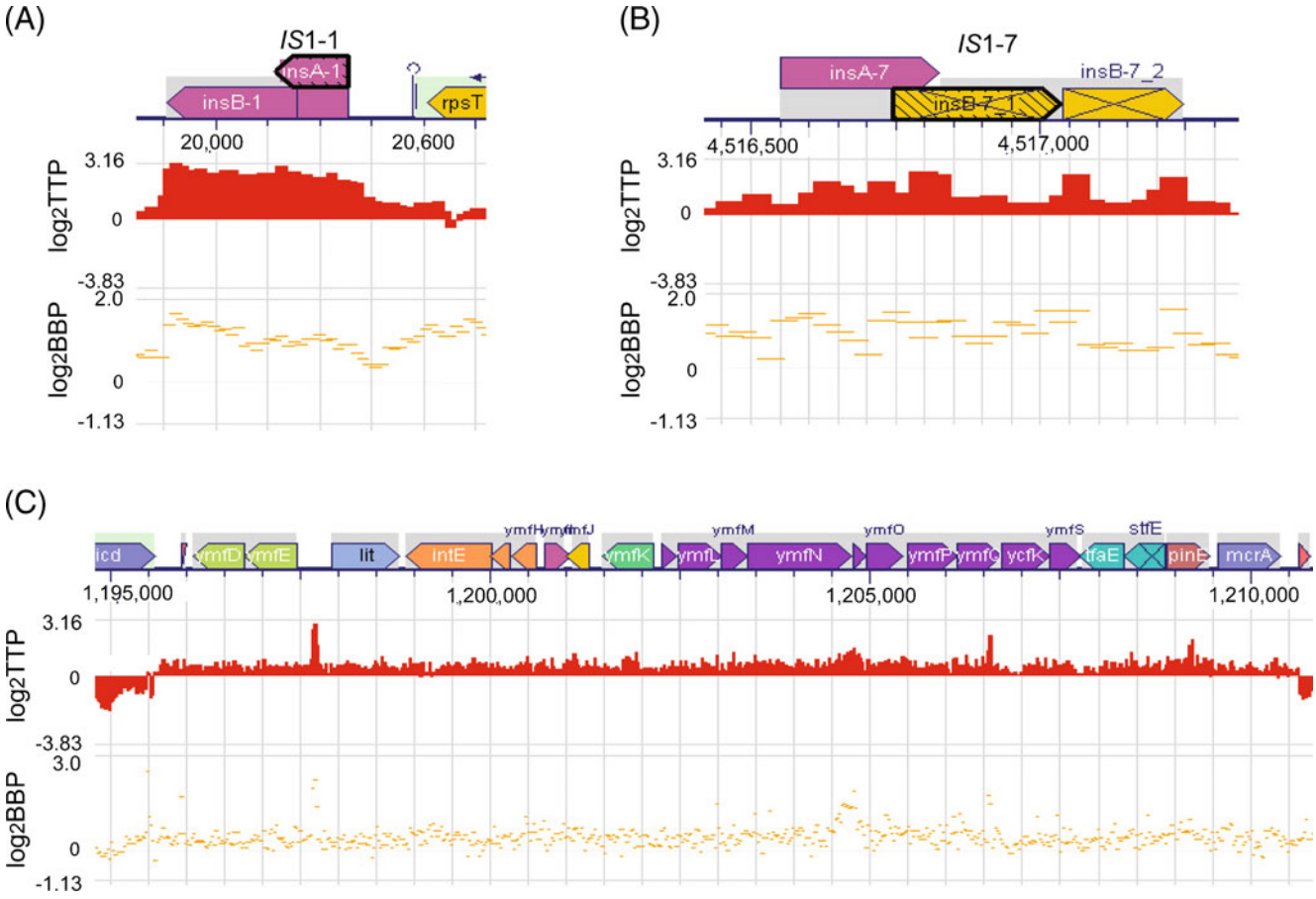

Figure 6. A second platform pattern of MuB binding and hot transposition sites. (**A**) *insB-1, insA-1, rpsT*. (**B**) *insA-7, insB-7*_1, *insB7*_2. (**C**) *ymfD, ymfE, lit, intE, xisE, ymfH, ymfI, ymfJ, ymfK, ymfT, ymfL, ymfM, ymfN, ymfR, ymfO, ymfP, ymfQ, ycfK, ymfS, tfaE, stfE, pinE, mcrA*.

binding sites, showed extensive protein occupancy domains (EPODs) in *E. coli*, some of which were localized to highly expressed genes and were enriched in RNA-polymerase (these included the *rrn* operons), but the majority of which were localized to transcriptionally silent loci (Vora *et al.* 2009). The latter were dominated by conserved hypothetical ORFs, had high intrinsic DNA curvature and were highly enriched in binding sites of nucleoid proteins. In the light of these data, one interpretation of the platform MuB binding patterns and their opposing consequences for TTP is that they are influenced by EPODs at both transcriptionally active and silent regions.

3.5 *MuB binding is positively related to transcription*

Previous studies have shown a negative correlation between Mu transposition and transcription (Manna *et al.* 2004). To determine if MuB binding shows a similar correlation, $\log_2$ BBP values were compared with transcript copy number (TSC) (Wei *et al.* 2001) (figure 7A). TSC is a measure of mRNA abundance in exponentially growing cells in LB, and higher TSC represents higher gene expression. An induced *lac* operon has a TSC value of 10, while an uninduced operon has a TSC<0.2. When moving medians of $\log_2$BBP over 150 kb and TSC over 101 genes (the two windows cover equivalent distances; *see* figure 7 legend) were plotted against genome location, the overall trend of MuB binding was seen to be positively related to that of transcription, with peaks and troughs of MuB binding generally matching with those of transcription.This result suggests that transcription *per se* does not exclude MuB binding.

As reported earlier (Manna *et al.* 2004), our TTP profiling also showed an inverse relationship between $\log_2$ TTP and TSC in general (figure 7B). However, close observation revealed some exceptions where gene loci with high transcription were hot transposition targets (figure 7B; indicated with a star). In a more detailed region shown in figure 6C (1.195–1.210 Mb), where TSC values range

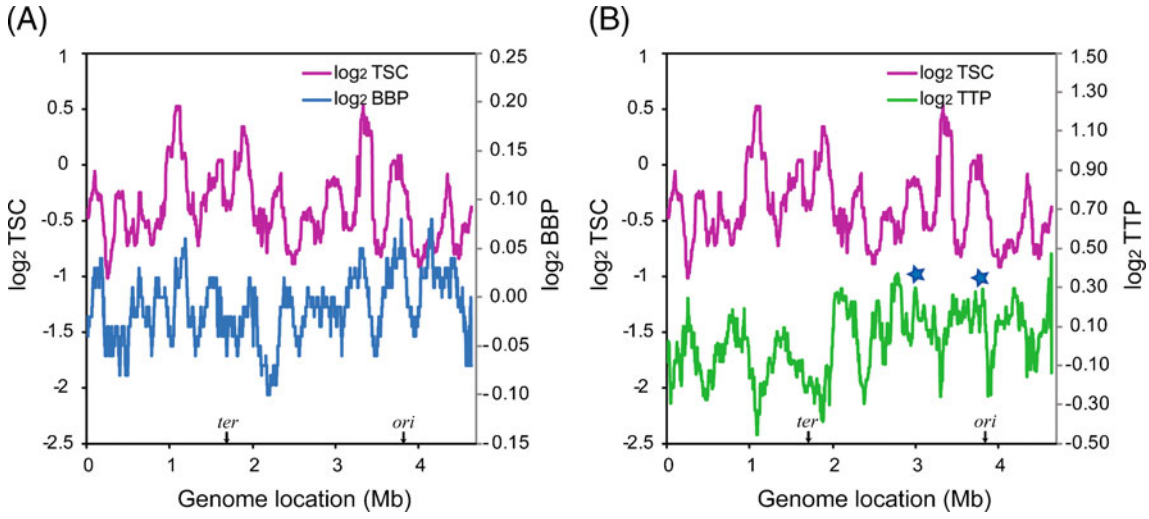

Figure 7. Relationship of MuB binding, Mu transposition and transcription profile of the *E. coli* genome. (**A**) A moving median of $\log_2$BBP with a 150 kb window and a 1 kb step, was plotted against the location of the start of the first probe in each window. A moving median of transcription copy number (mTSC) with a window of 101 genes and a step of one gene, was plotted against the location of the start site of the first gene in each window. (The average length of genes in *E. coli* is 1.1 kb, and that of the intergenic region ranges from100~1000 bp with an average of 116 bp, so the length of 101 genes is around 120 kb. Therefore, 101genes span a distance comparable to 150 kb.) TSC values are from Wei *et al.* 2001. Blue, BBP; purple, TSC. (**B**) As in **A**, except the moving median of TTP is compared with that of TSC. Green, TTP; Stars, TTP peaks coincident with TSC peaks.

between 1.66 and 6.22, not only is Mu transposition above average but displays several transposition peaks. Similarly, in the region of 4.423–4.424 Mb (supplementary figure 3A), where TSC of the four genes *rpsF*, *priB*, *rpsR* and *rpll* ranges from 3.564 to 7.159, both $\log_2$ TTP and $\log_2$ BBP show some high values. In supplementary figure 3B, *rmf* and *fabA* have TSC of 23.278 and 2.239, while their TTP and BBP values are above average.

In summary, transcription does not insulate DNA from MuB binding. It follows that transcription should also not inhibit Mu transposition. Indeed, several regions of high transcription supported Mu transposition. This suggests that transcription itself may not be the cause of the negative relationship between TTP and TSC, but rather that MuB binding and transcription may be responding to some common cellular feature, e.g. EPODs discussed above. Experimental data have also shown a good correlation between transcriptional activity and the number and stability of looped DNA domains in particular regions of the *E. coli* genome (Dillon and Dorman 2010).

3.6 *MuB binding responds to A-tract frequency of the Escherichia coli chromosome: Analysis in a fis⁻ strain*

The *E. coli* genome is highly organized and condensed through the concerted action of NAPs (nucleoid associated proteins), condensins (such as MukBEF), topoisomerases and DNA supercoiling. 400~450 independently segregated supercoiling domains are thought to be dynamically organized both in size and in numbers over the life time of a cell (Alexandrov *et al.* 1999; Deng *et al.* 2005; Sinden and Pettijohn 1981; Stein *et al.* 2005). The domains are thought to be stabilized through extensive binding by several NAPs, which include H-NS (or StpA), HU, IHF, Fis and the stationary phase-specific DNA-binding protein Dps (Murphy and Zimmerman 1997; Azam and Ishihama 1999; Schneider *et al.* 2001 Dame 2005; Dillon and Dorman 2010). The domains and their bound proteins are implicated in regulating transcription (Ussery *et al.* 2001; Dorman and Deighan 2003; Blot *et al.* 2006). The basic underlying factor facilitating DNA condensation is encoded in the DNA itself, where stretches of A/T sequences or A-tracts (defined as the sequence A_nT_m, where (n+m) ≥4) impart an intrinsic curvature to DNA (Hodges-Garcia *et al.* 1989; Haran *et al.* 1994). In the *E. coli* genome, A-tracts are over-represented and distributed 'quasi-regularly' with a 10–12 bp periodicity throughout the genome, organized in ~100-bp-long clusters (Tolstorukov *et al.* 2005). A-tracts introduce local bends of the DNA duplex and these bends accumulate when properly phased, suggesting that the A-tract clusters would facilitate DNA looping and superhelical branching, positioning promoters at the apices of superhelices (Laundon and Griffith 1988; Rippe *et al.* 1995). A-tracts have therefore been proposed to

constitute the 'structural code' for DNA compaction (Tolstorukov *et al.* 2005).

The Fis protein is one of the most abundant NAPs in *E. coli* (Ussery *et al.* 2001). A recent ChIP-chip analysis has shown that Fis is directly involved in structuring the supercoiling domains of the *E. coli* chromosome through stabilization of DNA crossovers, loops and bends (Cho *et al.* 2008). If most NAPs bind to stretches of A/T sequences, which are also preferred by MuB, we reasoned that these architectural elements would likely be off limits for MuB. Indeed, comparison of our MuB binding data with the available Fis binding data (Cho *et al.* 2008) showed a striking inverse relationship between the binding profiles of the two proteins. In the figure 8A, the binding profile of MuB was compared with that of Fis protein by plotting the moving median of both $\log_2$ BBP and $\log_2$ Fis enrichment ratio (defined similarly to BBP) with a 150 kb window and 1 kb step, against genome location. Both MuB and Fis binding values could be grouped into fluctuating high and low binding regions. Fis and MuB binding showed weak but significant negative relationship (r=−0.2621). With several exceptions, the peaks of MuB binding correlated with the troughs of Fis binding, and vice versa. Curiously, in figure 8A, the MuB and Fis binding patterns appear to be in or out of 'phase' along four large, but relatively specific and well-defined domains of the *E. coli* chromosome; the curves are roughly 'out of phase' from position 1–2 Mb and from 3.5–4.5 Mb; and are 'in phase' from position 4.4–0.5 Mb and

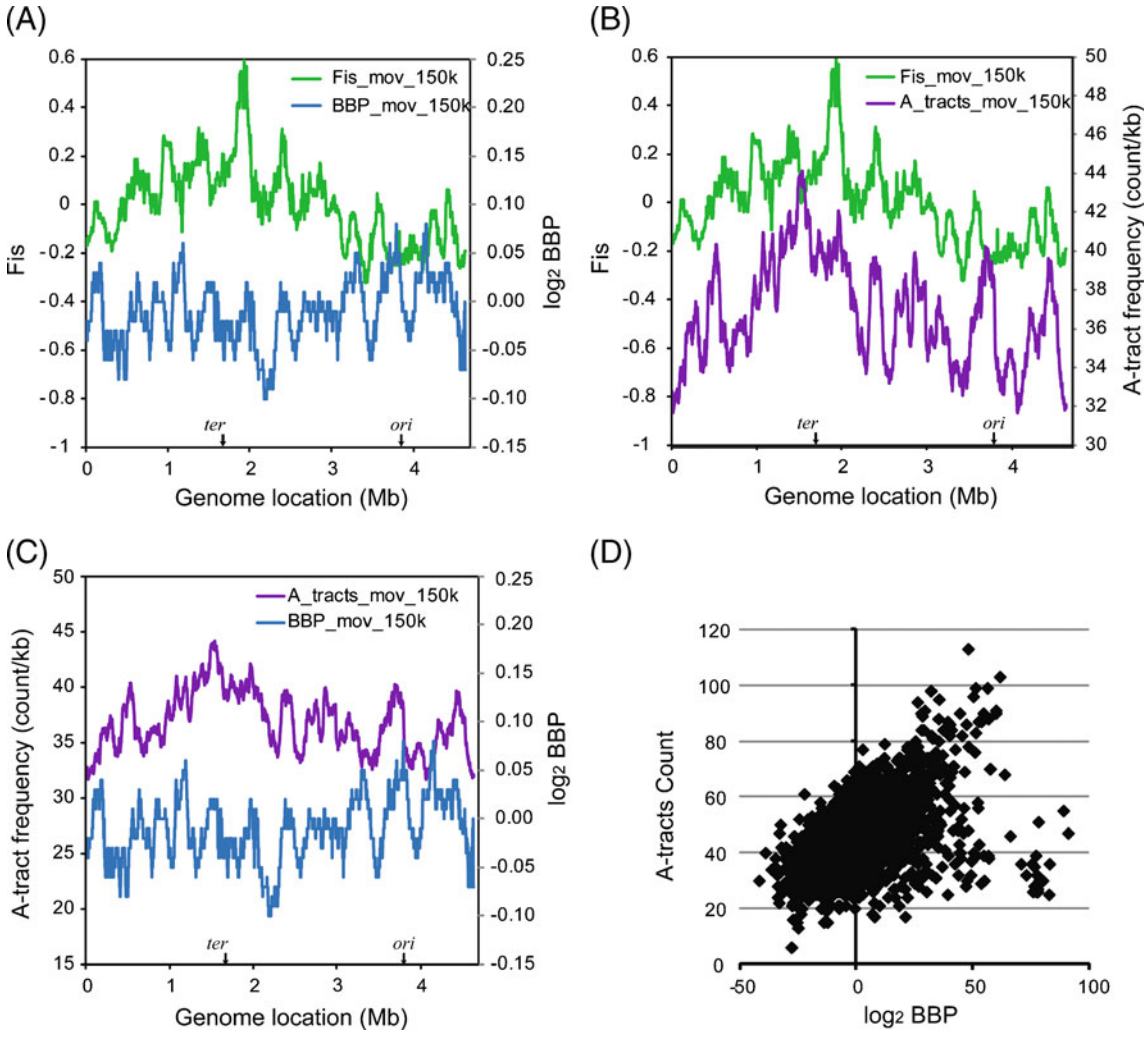

Figure 8. The relationship between MuB binding, Fis binding, and A-tracts. (**A**–**C**) Moving medians of $\log_2$ BBP, Fis enrichment and A-tract frequency with the moving window of 150 kb and step 1 kb were plotted against the location of the start of the first probe in each window. The Pearson correlation coefficients for A–C are −0.2661, 0.6329 and 0.2525, respectively. (**D**) The additive of $\log_2$BBP values from 100 continuous probes was plotted against the total count of A-tracts in that region. The 100 continuous probes cover 2426 bp of the genome.

from 2–3.5 Mb. Both genetic and cytological studies have suggested that the *E. coli* chromosome has a ring organization with four structured macrodomains (Ori, Ter, Left and Right) and two less structured regions (Niki *et al.* 2000; Boccard *et al.* 2005; Espeli *et al.* 2008). The macrodomains are between 3.76–0.04 Mb (Ori), 0.59–1.2 Mb (Right), 1.2–2.18 Mb (Ter) and 2.18–2.87 Mb (Left). Indeed, the in-phase regions of Fis and BBP approximately fall within the Ori and Left macro domains, and one of the two 'out of phase' regions falls in a less structured region (~0.25–0.59 Mb). This observation reinforces the notion that MuB binding is responding to chromosome domain structure.

Fis binds to intrinsic DNA curvatures introduced by clusters of A-tracts. A comparative analysis done by plotting the moving median of both A-tracts frequency and $\log_2$ Fis enrichment ratio showed that Fis binding regions exhibited a strong positive relationship with A-tract frequency (r=0.6320; figure 8B). A similar relationship between A-tracts frequency and H-NS binding on *E. coli* chromosome had been reported before (Oshima *et al.* 2006). These results suggest that A-tract frequency can be used to represent the binding profiles of various NAPs. The results in figure 8A and B would predict that BBP should be negatively correlated with A-tract frequency. However, although there were peaks and troughs of MuB binding and A-tract frequency which were mismatched (figure 8C), the overall trend of BBP was in accordance with that of A-tract frequency (r=0.2515). Close observation revealed that in each of the trend panels (figure 8A–C), there are exceptions and offsets to the overall relationships. These exceptions and offsets likely reflect the competition for A-tract clusters among MuB, Fis and other NAPs. When the additive A-tract frequency and $\log_2$ BBP values of 100 probes representing 2426 bp of the genome were plotted against each other (figure 8D), the plot showed that MuB does tend to bind to chromosome regions with high A-tract frequency, but that there are also low A-tract frequency regions that are preferably bound. The contradiction comes from those deviations in panel 8A and 8B, where Fis and MuB show concurrence in some loci, and Fis is absent from a few of the A-tract rich areas. The overall positive correlation between MuB and A-tract frequency indicates that MuB manages to occupy A-tract cluster rich regions, and therefore should wield an influence on the chromosome supercoiling structure.

To further examine if MuB binding is affected by Fis, a binding profile was generated in a *fis*$^-$ mutant strain (CW28Δ*fis*). The $\log_2$ BBP values from the parent CW28 strain were plotted against those from CW28Δ*fis* (figure 9A). Perhaps not surprisingly, given the redundancy of the NAPs, both BBP and TTP profiles were found to be similar at majority of sites in both wild-type and *fis* mutant genomes, changing dramatically only at some loci (supplementary table 1). These loci mapped mainly to ORF regions. The fact that only a small portion of the changes (9/106 loci in BBP and 16/213 in TTP) were within the intergenic regions, where the majority of Fis binding sites map, suggests that these regions otherwise occupied by Fis are taken over by other NAPs in the *fis* deletion strain. There were 106 loci where MuB binding was differentially affected in wild-type *vs* *fis*$^-$ strains; of these, 71 had increased BBP and 35 had decreased BBP. Similarly, Mu transposition profiles or $\log_2$ TTP values were largely similar in both genomes (figure 9B), changing only at some loci listed in supplementary table 2. There were 213 loci differentially affected in TTP; of these 138 had increased

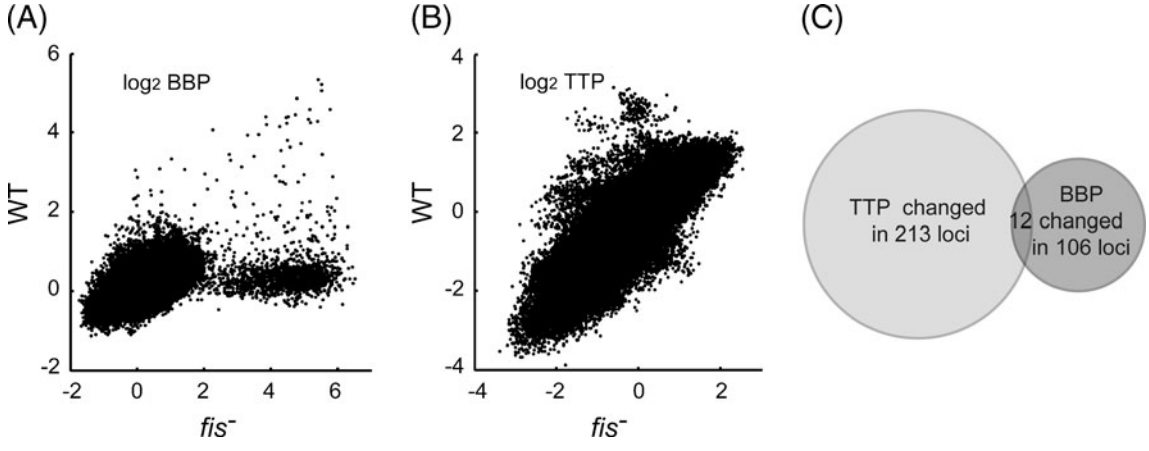

Figure 9. Comparison of BBP and TTP profiles in wild type and *fis*$^-$ strains. (**A**) $\log_2$ BBP values from the wild-type (WT) strain were plotted against those from a *fis*$^-$ strain. (**B**) $\log_2$ TTP values from the WT strain were plotted against those from *fis*$^-$ strain. All profiles are an average of three independent experiments. (**C**) Venn diagram representing BBP and TTP changes in the *fis*$^-$ strain.

TTP and 75 loci had decreased TTP. Surprisingly, only 12 of these loci were shared i.e. both BBP and TTP changed significantly at these loci (figure 9C). The rare coincidence of BBP and TTP indicates that the choice of transposition targets is influenced by factors besides MuB binding.

Fis binds 894 regions of the *E. coli* genome, 67% of which are in intergenic regions (Cho *et al.* 2008). We note, however, that only 8.5% of differential MuB binding sites and 7.5% of differential transposition targets fall into intergenic regions (supplementary tables 1 and 2). Fis is also reported to differentially regulate the expression of 923 genes (Cho *et al.* 2008). However, only a small proportion of the loci that showed changes in BBP and TTP were those reported to be trancriptionally regulated by Fis (*see* supplementary tables 1 or 2 legend). We can therefore surmise that changes in Mu transposition targets and MuB binding are not controlled by transcription events regulated by Fis. That target availability is only minimally perturbed in the absence of a major NAP such as Fis, shows that chromosome architecture is inherently robust. In light of this robustness, we consider even the small change in MuB binding and transposition observed in the absence of Fis to be significant.

In summary, the results in this section suggest that the MuB is excluded from occupying Fis sites, even though both proteins bind A-tract DNA. In the absence of Fis, the largely stable MuB binding patterns and Mu transposition profiles suggest that the presence of multiple NAPs shield the *E. coli* genome against significant perturbations of chromosome structure. This observation may also have implications for Mu transposition during the lytic growth, when the genome is being dramatically rearranged. For example, MuB could stand-in for NAPs if necessary, maintaining the structure of the chromosome for as long as possible in order to ensure efficient Mu transposition.

4. Summary

The several new insights obtained from this *in vivo* study are as follows: (1) MuB binds throughout the Mu genome, consistent with its non-specific DNA binding properties. Despite the general random nature of Mu transposition and MuB binding, there were hot and cold insertion sites and MuB binding sites in the genome, and differences between the hottest and coldest sites were large. (2) Mu transposition is positively correlated with MuB binding, but transposition peaks do not necessarily correspond to MuB binding peaks. Transposition appears to be in the vicinity of MuB-bound DNA, supporting *in vitro* studies where Mu transposition occurred next to but not within MuB-bound DNA. Further support for this conclusion comes from the finding that MuB shows preference for binding AT-rich regions but not CGG regions, which are favoured Mu insertion sites. (3) An overall positive relationship was observed between MuB binding and transcription, and yet a similar relationship expected between transposition and transcription was generally not seen. The data suggest that a direct relationship between transcription and transposition is unlikely. (4) MuB-preferred A/T DNA is also preferred by chromosome remodelling proteins such as Fis, and yet the binding profiles of the two proteins were distinct, suggesting that MuB is excluded from regions important for molding the chromosome architecture. In the absence of Fis, which also regulates expression of a large number of genes, MuB binding and Mu insertion profiles changed in a manner unrelated to changes in transcription profiles, again supporting the conclusion that Mu transposition in unrelated to transcription. Although the observed changes in binding and transposition profiles in the absence of Fis were not large, Fis is only one of several known NAPs in *E. coli*, a redundancy which highlights the robustness of the nucleoid structure. (5) The complex relationship between MuB binding and the various parameters examined in this study preclude any specific conclusion to be drawn regarding why MuB binds strongly within the Mu genome; immunity of the Mu genome to self-integration is proposed to stem from such binding. (6) Given that the architectural code, i.e. A-tracts, is also the substrate for MuB binding, and that NAPs apparently block MuB access to this code, we propose that the architectural code of the host genome provides a vital defense against invading transposable elements like Mu.

Acknowledgements

This work was supported by National Institutes of Health grant GM 33247 and in part by the Robert Welch Foundation Grant F-1351.

References

Adzuma K and Mizuuchi K 1988 Target immunity of Mu transposition reflects a differential distribution of Mu B protein. *Cell* **53** 257–266

Adzuma K and Mizuuchi K 1991 Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage Mu. Involvement of protein oligomerization in the ATPase cycle. *J. Biol. Chem.* **266** 6159–6167

Alexandrov AI, Cozzarelli NR, Holmes VF, Khodursky AB, Peter BJ, Postow L, Rybenkov V and Vologodskii AV 1999 Mechanisms of separation of the complementary strands of DNA during replication. *Genetica* **106** 131–140

Au TK, Agrawal P and Harshey RM 2006 Chromosomal integration mechanism of infecting mu virion DNA. *J. Bacteriol.* **188** 1829–1834

Azam TA 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

Beauregard A, Curcio MJ and Belfort M 2008 The take and give between retrotransposable elements and their hosts. *Annu. Rev. Genet.* **42** 587–617

Berry C, Hannenhalli S, Leipzig J, and Bushman FD 2006 Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* **2** e157

Blot N, Mavathur R, Geertz M, Travers A and Muskhelishvili G 2006 Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.* **7** 710–715

Boccard F, Esnault E and Valens M 2005 Spatial arrangement and macrodomain organization of bacterial chromosomes. *Mol. Microbiol.* **57** 9–16

Brady T, Lee YN, Ronen K, Malani N, Berry CC, Bieniasz PD and Bushman FD 2009 Integration target site selection by a resurrected human endogenous retrovirus. *Genes Dev.* **23** 633–642

Bukhari AI and Taylor AL 1975 Influence of insertions on packaging of host sequences covalently linked to bacteriophage Mu DNA. *Proc. Natl. Acad. Sci. USA* **72** 4399–4403

Chaconas G and Harshey RM 2002 Transposition of phage Mu DNA; in *Mobile DNA II* eds NL Craig, R Craigie, M Gellert and AM Lambowitz (Washington DC: SM Press) pp 384–402

Cho BK, Knight EM, Barrett CL and Palsson BO 2008 Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. *Genome Res.* **18** 900–910

Condon C, Philips J, Fu ZY, Squires C and Squires CL 1992 Comparison of the expression of the seven ribosomal RNA operons in Escherichia coli. *EMBO J.* **11** 4175–4185

Condon C, Squires C and Squires CL 1995 Control of rRNA transcription in Escherichia coli. *Microbiol. Rev.* **59** 623–645

Craig NL 1997 Target site selection in transposition. *Annu. Rev. Biochem.* **66** 437–474

Dame RT 2005 The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56** 858–870

Datsenko KA and Wanner BL 2000 One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97** 6640–6645

Deng S, Stein RA and Higgins NP 2005 Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.* **57** 1511–1521

Dillon SC and Dorman CJ 2010 Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8** 185–195

Dorman CJ and Deighan P 2003 Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Genet. Dev.* **13** 179–184

Espeli O, Mercier R and Boccard F 2008 DNA dynamics vary according to macrodomain topography in the E. coli chromosome. *Mol. Microbiol.* **68** 1418–1427

Ge J and Harshey RM 2008 Congruence of *in vivo* and *in vitro* insertion patterns in hot E. coli gene targets of transposable element Mu: opposing roles of MuB in target capture and integration. *J. Mol. Biol.* **380** 598–607

Ge J, Lou Z and Harshey RM 2010 Immunity of replicating Mu to self-integration: a novel mechanism employing MuB protein. *Mobile DNA* **1** doi: 10.1186/1759-8753-1-8

Greene EC and Mizuuchi K 2004 Visualizing the assembly and disassembly mechanisms of the MuB transposition targeting complex. *J. Biol. Chem.* **279** 16736–16743

Guo Y and Levin HL 2010 High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in Schizosaccharomyces pombe. *Genome Res.* **20** 239–248

Haapa-Paananen S, Rita H and Savilahti H 2002 DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection *in vitro*. *J. Biol. Chem.* **277** 2843–2851

Haran TE, Kahn JD and Crothers DM 1994 Sequence elements responsible for DNA curvature. *J. Mol. Biol.* **244** 135–143

Hillebrand A, Wurm R, Menzel A and Wagner R 2005 The seven E. coli ribosomal RNA operon upstream regulatory regions differ in structure and transcription factor binding efficiencies. *Biol. Chem.* **386** 523–534

Hodges-Garcia Y, Hagerman PJ and Pettijohn DE 1989 DNA ring closure mediated by protein HU. *J. Biol. Chem.* **264** 14621–14623

Laundon CH and Griffith JD 1988 Curved helix segments can uniquely orient the topology of supertwisted DNA. *Cell* **52** 545–549

Lee I and Harshey RM 2001 Importance of the conserved CA dinucleotide at Mu termini. *J. Mol. Biol.* **314** 433–444

Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, Shinn P, *et al.* 2006 Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* **2** e60

Manna D, Breier AM and Higgins NP 2004 Microarray analysis of transposition targets in Escherichia coli: the impact of transcription. *Proc. Natl. Acad. Sci. USA* **101** 9780–9785

Manna D, Deng S, Breier AM and Higgins NP 2005 Bacteriophage Mu targets the trinucleotide sequence CGG. *J. Bacteriol.* **187** 3586–3588

Mizuuchi M and Mizuuchi K 1993 Target site selection in transposition of phage Mu. *Cold Spring Harb. Symp. Quant. Biol.* **58** 515–523

Murphy LD and Zimmerman SB 1997 Isolation and characterization of spermidine nucleoids from Escherichia coli. *J. Struct. Biol.* **119** 321–335

Niki H, Yamaichi Y and Hiraga S 2000 Dynamic organization of chromosomal DNA in Escherichia coli. *Genes Dev.* **14** 212–223

Nomura M 1999 Engineering of bacterial ribosomes: replacement of all seven Escherichia coli rRNA operons by a single plasmid-encoded operon. *Proc. Natl. Acad. Sci. USA* **96** 1820–1822

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.* **13** 141–153

Parks AR, Li Z, Shi Q, Owens RM, Jin MM and Peters JE 2009 Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell* **138** 685–695

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, *et al.* 2000 Genome-wide location and function of DNA binding proteins. *Science* **290** 2306–2309

Ren CP, Chaudhuri RR, Fivian A, Bailey CM, Antonio M, Barnes WM and Pallen MJ 2004 The ETT2 gene cluster, encoding a

second type III secretion system from Escherichia coli, is present in the majority of strains but has undergone widespread mutational attrition. *J. Bacteriol.* **186** 3547–3560

Rippe K, von Hippel PH and Langowski J 1995 Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* **20** 500–506

Schneider R, Lurz R, Luder G, Tolksdorf C, Travers A and Muskhelishvili G 2001 An architectural role of the Escherichia coli chromatin protein FIS in organising DNA. *Nucleic Acids Res.* **29** 5107–5114

Sinden RR and Pettijohn DE 1981 Chromosomes in living Escherichia coli cells are segregated into domains of supercoiling. *Proc. Natl. Acad. Sci. USA* **78** 224–228

Stein RA, Deng S and Higgins NP 2005 Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Mol. Microbiol.* **56** 1049–1061

Symonds N, Toussaint A, Van de Putte P and Howe MM 1987 *Phage Mu* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory)

Tan X, Mizuuchi M and Mizuuchi K 2007 DNA transposition target immunity and the determinants of the MuB distribution patterns on DNA. *Proc. Natl. Acad. Sci. USA* **104** 13925–13929

Tolstorukov MY, Virnik KM, Adhya S and Zhurkin VB 2005 A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucleic Acids Res.* **33** 3907–3918

Ussery D, Larsen TS, Wilkes KT, Friis C, Worning P, Krogh A and Brunak S 2001 Genome organisation and chromatin structure in Escherichia coli. *Biochimie* **83** 201–212

Vora T, Hottes AK and Tavazoie S 2009 Protein occupancy landscape of a bacterial genome. *Mol. Cell* **35** 247–253

Wei Y, Lee JM, Richmond C, Blattner FR, Rafalski JA and LaRossa RA 2001 High-density microarray-mediated gene expression profiling of Escherichia coli. *J. Bacteriol.* **183** 545–556

Wu X and Burgess SM 2004 Integration target site selection for retroviruses and transposable elements. *Cell Mol. Life Sci.* **61** 2588–2596

MS received 30 May 2011; accepted 29 June 2011

*e*Publication: 16 August 2011

Corresponding editor: DURGADAS P KASBEKAR