Genomic analysis reveals epistatic silencing of "expensive" genes in Escherichia coli K-12

Rajalakshmi Srinivasan, Deepti Chandraprakash, Revathy Krishnamurthi, Parul Singh, Vittore F. Scolari, Sandeep Krishna, Aswin Sai Narain Seshasayee

Genomic analysis reveals epistatic silencing of ''expensive'' genes in *Escherichia coli* K-12†

Rajalakshmi Srinivasan,[ab] Deepti Chandraprakash,[a] Revathy Krishnamurthi,[ac] Parul Singh,[ac] Vittore F. Scolari,[abd] Sandeep Krishna[a] and Aswin Sai Narain Seshasayee*[a]

A barrier for horizontal gene transfer is high gene expression, which is metabolically expensive. Silencing of horizontally-acquired genes in the bacterium *Escherichia coli* is caused by the global transcriptional repressor H-NS. The activity of H-NS is enhanced or diminished by other proteins including its homologue StpA, and Hha and YdgT. The interconnections of H-NS with these regulators and their role in silencing gene expression in *E. coli* are not well understood on a genomic scale. In this study, we use transcriptome sequencing to show that there is a bi-layered gene silencing system – involving the homologous H-NS and StpA – operating on horizontally-acquired genes among others. We show that H-NS-repressed genes belong to two types, termed ''*epistatic*'' and ''*unilateral*''. In the absence of H-NS, the expression of ''*epistatically* controlled genes'' is repressed by StpA, whereas that of ''*unilaterally* controlled genes'' is not. Epistatic genes show a higher tendency to be non-essential and recently acquired, when compared to unilateral genes. Epistatic genes reach much higher expression levels than unilateral genes in the absence of the silencing system. Finally, epistatic genes contain more high affinity H-NS binding motifs than unilateral genes. Therefore, both the DNA binding sites of H-NS as well as the function of StpA as a backup system might be selected for silencing highly transcribable genes.

1 Introduction

Horizontal gene transfer is a major force in bacterial evolution. It introduces drastic changes in an organism's gene content by transferring entire functional pathways. This allows the host organism to rapidly explore new niches and phenotypes, including pathogenesis. The effects of a horizontal gene transfer event extend beyond mere addition of new genes; it introduces new selective forces that might lead to a chain reaction of further modifications to the core or the conserved host genome. These might include, for example, changes in the metabolic or the regulatory circuitry of the cell,[1–3] deletions of core genes,[4] and fundamental changes in genomic base composition as imposed by acquired restriction-modification systems.[5,6] Further, there might be costs associated with expressing such acquired genes.[7] Therefore, it is not surprising that horizontally acquired genes are carefully regulated, not least at the level of gene expression.[8,9]

Gene expression in versatile bacteria such as the model organism *Escherichia coli* is regulated at the transcriptional level by a complex network of interactions mediated by sigma factors, which are components of the initiating RNA polymerase holoenzyme, and transcription factors. Transcription factors may be global or local in scope depending on various parameters including the numbers of genes that they regulate, their propensity to be involved in combinatorial control, and the numbers of environmental conditions in which they are active.[10] One of the global regulators of transcription in *E. coli* is the protein called H-NS, whose regulatory scope extends to 20% of the genes encoded by the genome.[11–13]

H-NS is a small protein of 137 amino acids, which recognises DNA that is A+T-rich in sequence[12,14,15] and/or specific DNA geometries such as intrinsic bending[16–18] and Holliday junctions.[19] It is also involved in transcription termination,[20] and can bind and stabilise mRNA.[21] As a transcriptional regulator, H-NS silences gene expression. It oligomerises on the DNA and forms DNA–protein–DNA bridges or rod-like structures, which might be structural motifs for strong transcriptional suppression.[22–24] Recent evidence has suggested

[a] *National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK, Bellary Road, Bangalore 560065, India. E-mail: aswin@ncbs.res.in*
[b] *Manipal University, Manipal 576104, India*
[c] *SASTRA University, Thanjavur, India*
[d] *Genomic Physics Group, UMR 7238 CNRS Microorganism Genomics, UPMC, Paris, France*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3mb70035f

that the 500 binding regions of H-NS collapse together spatially into two foci,[25] which might act as the bacterial heterochromatin. In particular, H-NS is a silencer of horizontally acquired genes,[13,15,26] which in enterobacteria display a tendency to be A+T-rich.[12,27] Despite being a regulator of horizontally acquired genes, H-NS itself is well conserved in enterobacteria. Functional homologs of H-NS have been detected even in distant bacterial genera such as *Bacillus* and *Mycobacteria.*[14] Taken together, H-NS is a powerful transcriptional repressor and a prominent link between the conserved, or the core genome of an organism and the accessory, or the recently acquired genome.

H-NS does not act alone. Its activity is enhanced or diminished by other proteins including StpA, Hha and YdgT (Cnu).[28] (a) StpA is a homologue of H-NS, the two proteins sharing nearly 70% sequence similarity.[29,30] A recent *in vitro* study has shown that StpA can also form DNA filaments similar to those formed by H-NS and thus silence gene expression.[31] Recent super-resolution microscopy showed different patterns of localisation for H-NS and StpA.[25] This is in curious contrast to ChIP-chip studies, which demonstrated that the genome-wide binding profile of StpA reflects that of H-NS.[32] H-NS and StpA repress each other transcriptionally.[29] The absence of H-NS not only abolishes StpA binding to two-thirds of its sites,[32] but also makes StpA subject to rapid degradation by the Lon protease.[33] Thus, the expression level of StpA – which contributes to its activity – is controlled by a balance between its synthesis and degradation. Prior studies, encompassing different strains of *E. coli* have shown that the Δ*stpA-hns* double mutant has a severe growth defect.[34,35] However, a mutation in *hns* has only a mild growth defect, except under certain conditions in stationary phase.[36] But a Δ*stpA* deletion does not appear to adversely affect growth as measured by optical density. (b) Hha is a small protein that can functionally replace the N-terminal oligomerisation domain of H-NS.[37] It also has a role in H-NS distinguishing between horizontally-acquired and core regions of the genome.[38] (c) YdgT is homologous to Hha and might share some of its functions.[39]

Taken together, the global gene regulatory system around H-NS appears to be a multi-layered silencing system that operates in regions of the genome with high A+T content, many of which are likely products of recent horizontal gene transfer. What is (are) the selective force(s) associated with the ability of H-NS to silence gene expression either on its own or in collaboration with other proteins? Here we present a study that attempts to answer this question using *E. coli* as a model.

2 Results

2.1 Growth characteristics

We first assessed the contributions of co-regulators acting around H-NS – namely, StpA, Hha and YdgT – to the growth of *E. coli*. We made single deletions of *hns*, *stpA*, *hha* and *ydgT* (Δ*hns*, Δ*stpA*, Δ*hha* and Δ*ydgT* respectively) in *E. coli* K12 MG1655. In addition, we made double deletions of *hns* with each of the other three co-regulators (Δ*stpA-hns*, Δ*hha-hns* and Δ*ydgT-hns*). Among the four single mutants, only Δ*hns* has a slight growth defect in LB medium (Fig. 1a). Among the double mutants (Fig. 1b), the Δ*stpA-hns* mutant appears to be deficient in growth, an observation not explained by the growth curves of Δ*hns* and Δ*stpA* taken in isolation. This strain shows a somewhat pronounced lag phase and a considerably smaller maximal growth rate than any of the other mutants. This is in agreement with earlier reports investigating the Δ*stpA-hns* double mutant.[34,35] But, the other double mutants – Δ*hha-hns* and Δ*ydgT-hns* – grow similarly to Δ*hns*.

2.2 "Epistatic" control of gene expression by H-NS and StpA

We studied the control of gene expression by H-NS, either alone or in combination with one of StpA, Hha and YdgT. For each of the strains used here (wildtype, Δ*hns*, Δ*stpA*, Δ*hha*, Δ*ydgT*, Δ*stpA-hns*, Δ*hha-hns* and Δ*ydgT-hns*), we performed transcriptome experiments – of cells grown to the mid-exponential phase in rich LB medium – using RNA-seq.

We made a gene expression matrix in which each mRNA-encoding gene was represented by a vector of processed read counts (see Methods). Hierarchical clustering of this matrix using uncentered correlation revealed the presence of two major clusters of samples (Fig. 1c). The first cluster contained the Δ*hns* single mutant and all the three double mutants. The second cluster comprised the wildtype and the single mutants of the three co-regulators. Thus, among the single mutants, Δ*hns* has the strongest effect on gene expression under the condition tested here, with most (>75%) responding genes being up-regulated in the mutant. Δ*stpA*, Δ*hha* and Δ*ydgT* have little to no effect. Among the double mutants, Δ*hha-hns* and Δ*ydgT-hns* are similar to Δ*hns*. However, Δ*stpA-hns* – forming a separate branch within the first cluster – has many genes that are upregulated relative to Δ*hns* as well as the wildtype, despite Δ*stpA* being similar to the wildtype.

Based on the above results, we restricted further analysis to H-NS and the homologous StpA. The deletion of *stpA* in a Δ*hns* background has a global impact on gene expression, whereas it does not in the wildtype. This suggests that certain genes are regulated by both the proteins in an epistatic manner. Herein, the word epistasis means that the transcriptional effect of Δ*stpA-hns* is not explained by the sum of the effects of the component single deletions. We classified genes into (a) those that are regulated *epistatically* by H-NS and StpA (for convenience referred to as "epistatic genes"); (b) those that are regulated *unilaterally* by H-NS ("unilateral genes"); (c) control genes, which are not regulated by H-NS or StpA, on the basis of our transcriptome data. Epistatically regulated genes were defined as those that are up-regulated in Δ*stpA-hns* relative to both the wildtype and Δ*hns*. About half of these epistatic genes are first up-regulated in Δ*hns* relative to the wildtype, and further up-regulated in Δ*stpA-hns*; the remaining do not respond to Δ*hns* but are up-regulated only in Δ*stpA-hns*. Unilateral genes are those which are up-regulated in Δ*hns* relative to the wildtype, but show no response to a further deletion of *stpA*. In total, we assembled a list of ∼360 epistatic and ∼610 unilateral genes (Fig. 2).

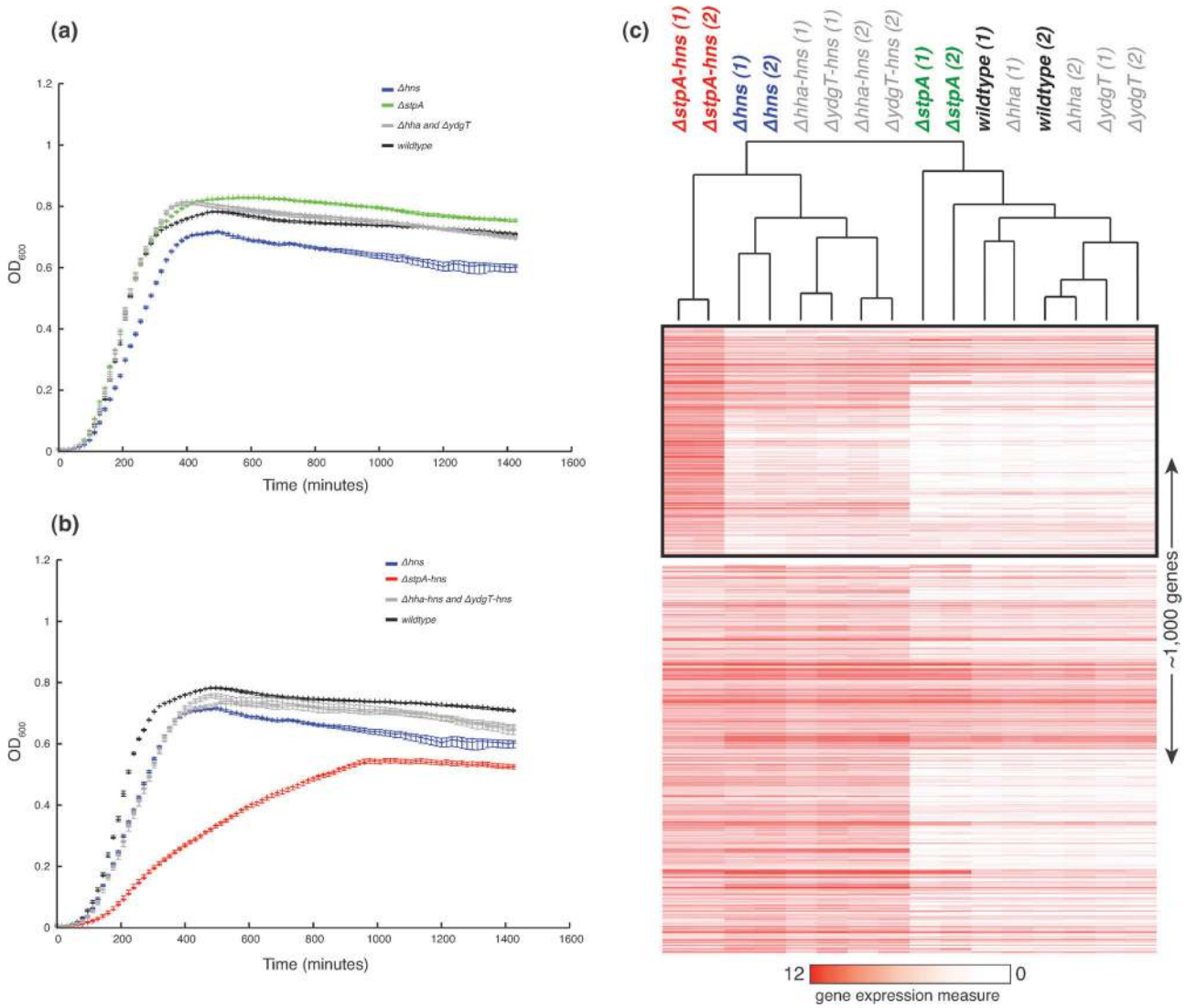

Fig. 1 (a) Growth curves of the wildtype and the various single mutant strains constructed in this study; (b) growth curves of the wildtype, Δ*hns*, and the various double mutant strains involving *hns*. The curves track the median across 6 replicates (2 biological × 3 technical), with the error bars marking the standard error. (c) Heatmap showing gene expression profiles of a selected set of genes in the various strains subjected to RNA-seq experiments. The cladogram was drawn using treeview and the heatmap using the matrix2png web server. Each column represents a strain, with the number in parenthesis identifying the two replicates. The wildtype, Δ*hns*, Δ*stpA*, and Δ*stpA-hns* strains are marked in bold text. Each row represents a gene. The colour in each cell shows the gene expression level of the corresponding gene in a strain. Red corresponds to high expression and white to low expression, with the range in-between shown by the intensity of the red colour. The region in the top, marked by the thick-bordered rectangle, shows genes which are specifically up-regulated in Δ*stpA-hns* when compared to Δ*hns*. The expression measure was computed as follows. The number of reads mapping to a gene was divided by the length of the gene. Then, this value was divided by the mode of the distribution of expression values across all genes in our dataset. This is a robust method of normalising for the variation in the total number of reads obtained for different samples. The expression measure is represented on a $\log_2$ scale.

2.3 H-NS and StpA binding to epistatically and unilaterally controlled genes

Using data from a previous ChIP-seq study of H-NS,[12] we find that much of the transcriptional up-regulation (in mutant compared with the wildtype) observed in our study can be explained by the direct effect of H-NS binding proximally to the gene (>70% of up-regulated genes are bound by H-NS). There is little difference between unilateral and epistatic genes in this respect. Kahramanoglou *et al.*[12] reported that the length of an H-NS binding region on the chromosome is correlated with the degree of repression of the target gene. We observe that epistatic genes are bound by longer H-NS binding tracts than unilateral genes (Fig. 3a; $P = 10^{-3}$; Wilcoxon test). This, tentatively, may be consistent with previously reported *in vitro* data that the heteromeric association of H-NS and StpA is more stable than the self-association of either protein.[40]

ChIP-chip experiments for StpA in wildtype and Δ*hns E. coli* K12[32] found that the binding profile of StpA is virtually identical to that of H-NS in the wildtype. However in Δ*hns*, StpA binding is abolished from two-thirds of its wildtype binding sites. By integrating these results with our transcriptome data,

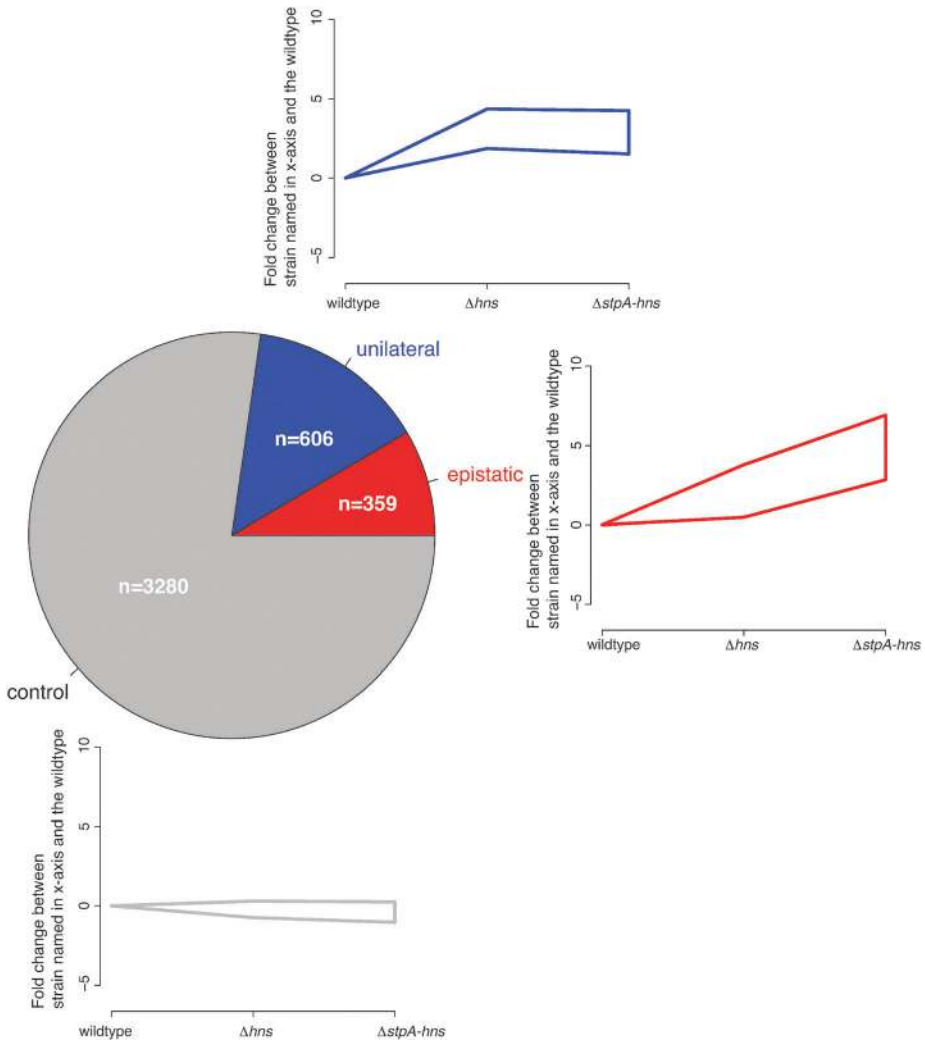

Fig. 2 The pie chart shows the number of genes belonging to each class of genes (epistatic, unilateral and the negative control) studied here. The three graphs show the inter-quartile range of the fold changes (on a $\log_2$ scale) of epistatic (red), unilateral (blue) and control (grey) genes in Δ*hns* and Δ*stpA-hns* relative to the wildtype. The lower bound is the first quartile and the upper bound the third quartile. The value for the wildtype is, by definition, zero.

we show that epistatic genes tend to be bound by StpA irrespective of whether H-NS is present or not (Fig. 3b). However, for the unilateral genes, StpA binding is lost in the absence of H-NS ($P < 10^{-10}$; Fisher's exact test).

What properties make the StpA target a specific subset of its binding sites in the absence of H-NS? Towards answering this, we assembled a set of DNA sequence motifs with high affinity for H-NS, based on publicly available *in vitro* affinity data for H-NS to 8-mer oligonucleotides.[14] For each epistatic and unilateral gene, we calculated the density of these high affinity binding motifs within the corresponding ChIP-seq-based H-NS binding region. We find that epistatic genes have a higher density of high-affinity H-NS binding motifs than unilateral genes (Fig. 3c; $P < 10^{-10}$; Wilcoxon test). This suggests that StpA, in the absence of H-NS, binds only to those target sites to which it has very high binding affinities.

To summarise, epistatically silenced genes tend to lie at sites where (a) H-NS binds in particularly long tracts and; (b) StpA binding is independent of H-NS availability. Thus epistatic control of gene expression by H-NS and StpA is explained by the genome-wide binding profiles of the two regulators.

2.4 Enrichment of horizontally-acquired and non-essential genes among epistatic genes

Is there a selective force that determines which genes are controlled unilaterally and which ones are regulated epistatically? Towards answering this question, we introduced various published genome-scale experimental data and computational predictions, and analysed our transcriptome data further.

H-NS is a well-known silencer of horizontally-acquired genes, a finding that is reflected in our data as well. We predicted ∼600 horizontally-acquired genes using the program Alien Hunter,[41] which defines regions of abnormal higher-order oligonucleotide usage as horizontally-acquired. For a nearly 50% A+T genome like that of *E. coli*, one would expect these predictions to pick genes with very high, or very low A+T contents. However, the data show that most horizontally-acquired genes thus predicted tend to be A+T-rich (see for example, ref. 12),

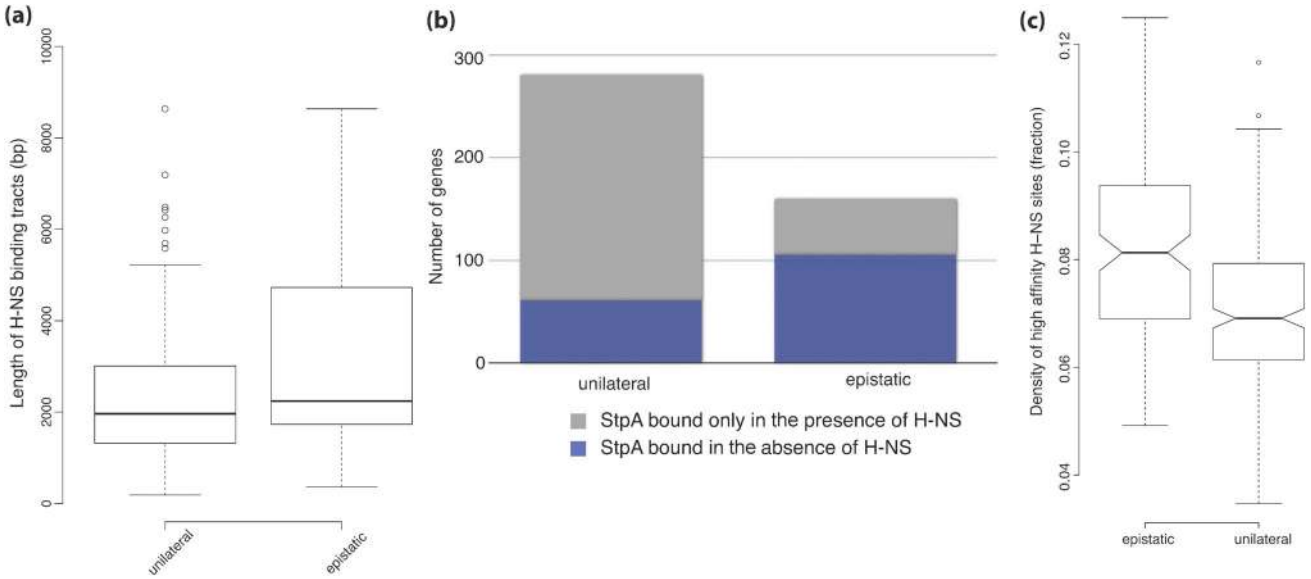

Fig. 3 (a) Distributions of the length of H-NS binding tracts for epistatic (right) and unilateral (left) genes. These numbers were obtained from ChIP-seq data published by Kahramanoglou and colleagues.[12] (b) The number of epistatic (right) and unilateral (left) genes which are bound by StpA in the wildtype and in Δ*hns*. These data were obtained from ChIP-chip experiments reported by Uyar *et al*.[32] (c) Distribution of the density of high affinity H-NS binding oligonucleotides[14] for epistatic and unilateral genes. In all the boxplots used in this paper, the whiskers are drawn at 1.5× IQR above and below the third and the first quartile respectively.

i.e. are asymmetric around the median A+T-content for all genes across the genome. Using this dataset, we find that a significantly higher proportion (30%) of epistatic genes are horizontally-acquired when compared to unilateral genes (21%; $P = 9.9 \times 10^{-4}$; Fisher's exact test).

Previous studies have made large-scale deletions of tracts of genes in the *E. coli* genome. These differ from single-gene deletion libraries such as the KEIO collection in the sense that hundreds of genes had been deleted in a single strain. Of the three such studies that we investigated, two had made deletions (removing 7–15% of the genes encoded) that did not affect the growth of *E. coli*,[42,43] whereas the third had resulted in cells with growth defects and aberrant morphology and nucleoid structure.[44] Thus, the groups of genes identified in each of the first two studies are non-essential to the growth of *E. coli* under standard laboratory conditions. The first study (Posfai dataset) used a comparative genomic survey to identify tracts of the genome to be deleted,[43] whereas the second study (Yu dataset) used a random gene disruption strategy.[42] The chromosomal deletions in the Posfai dataset covered ∼750 genes, whereas those in the Yu dataset covered ∼470 genes in total, with a maximum of ∼285 deleted in a single strain (cumulative Yu dataset). By overlaying these gene lists on our data, we find that epistatic genes are statistically enriched for non-essential genes, when compared to unilateral genes (Posfai dataset: 30% of epistatic *vs.* 20% of unilateral genes, $P = 7.2 \times 10^{-4}$; Yu dataset: 26% of epistatic *vs.* 16% of unilateral genes, $P = 2.8 \times 10^{-4}$; cumulative Yu dataset: 18% of epistatic *vs.* 11% of unilateral genes, $P = 10^{-3}$; Fisher's exact test). We note that the list of horizontally acquired genes, identified by the Alien Hunter program, overlaps with those presented by Posfai and Yu. However, the statistical enrichments stated above are independent of these overlaps. Finally, among the genes unique to the third study referred here,[44] where the deletions produced *E. coli* with growth defects, there is little enrichment for unilateral or epistatic genes (9% of each of unilateral, epistatic and control genes).

2.5 From silenced to de-repressed expression of epistatically and unilaterally controlled genes

Next, we used our transcriptome data to obtain the expression levels of epistatic and unilateral genes in the fully repressed wildtype cells. The expression levels of both sets of genes are low, as expected from the role of H-NS as a transcriptional silencer. Nevertheless, genes under epistatic control are expressed at significantly lower levels than those under unilateral regulation (Fig. 4a; $P < 10^{-10}$, for all relevant comparisons, Wilcoxon test). In a previous study, Vora and colleagues had used microarrays to define genes which fall within transcriptionally silent protein-bound regions of the chromosome.[45] In agreement with our transcriptome, epistatic genes are more likely to fall in such regions than unilateral genes, which in turn show a much higher tendency to be silent than control genes (27% for epistatic, compared to 17% for unilateral; $P = 1.1 \times 10^{-4}$, Fisher's exact test; 3% for control). These are consistent with the previously stated observation that epistatic genes are bound by longer H-NS binding tracts than unilateral genes.

Next, we compared the expression levels of epistatic and unilateral genes in the de-repressed state represented by Δ*stpA-hns*. We find a trend that is opposite to that in the wildtype cells. Both epistatic and unilateral genes are expressed at very high levels compared to the negative reference (Fig. 4b; $P < 10^{-10}$, for both comparisons, Wilcoxon test). In particular, epistatic genes are expressed at significantly higher levels than unilateral genes ($P = 2.1 \times 10^{-7}$, Wilcoxon test). Consequently, the fold

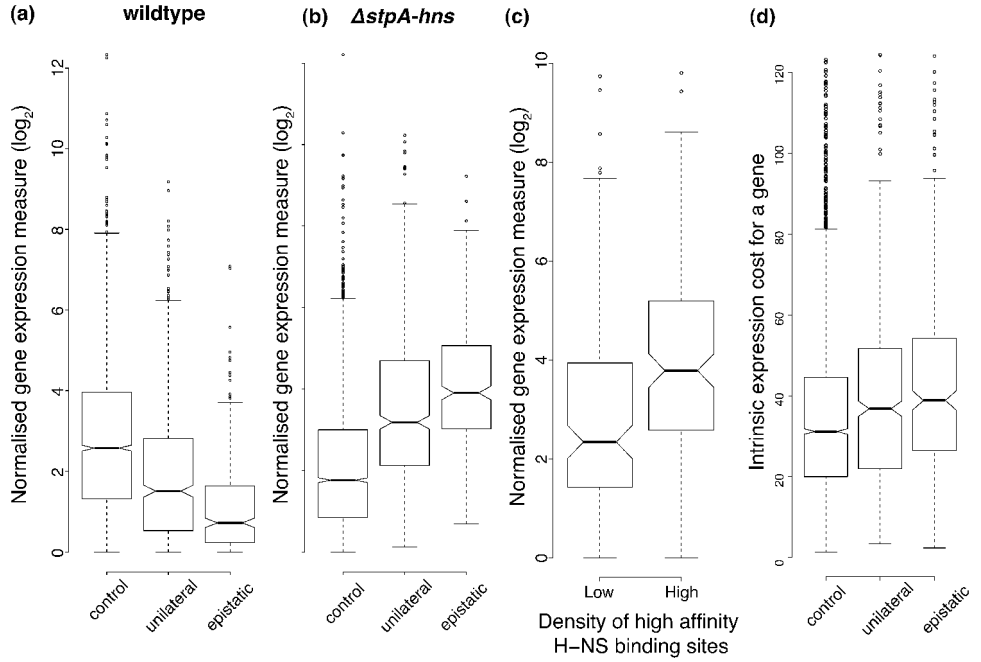

Fig. 4 (a) Distributions of expression levels, calculated as described in the legend to Fig. 1, for epistatic (right), unilateral (middle) and control (left) genes in the wildtype transcriptome data. (b) As in (a), except that these data are from the transcriptome of Δ*stpA-hns*. The *y*-axis is on the same scale as in (a). (c) Distributions of gene expression levels, in the double mutant, for genes ranked in the top 25% (labelled ''high'', right), in terms of the density of high-affinity H-NS binding sites, and those in the bottom 25% (labelled ''low'', left). (d) Distributions of intrinsic gene expression cost for epistatic (right), unilateral (middle) and control (left) genes.

change in expression between the de-repressed double mutant and the fully repressed wildtype is higher for epistatic than unilateral genes.

In summary, epistatic genes display properties that are at the extreme end of H-NS regulated genes, in terms of (a) high binding affinity to H-NS; (b) tendency to have been recently acquired and/or non-essential; (c) the degree of silencing imposed by H-NS (and StpA); (d) high expression levels in the absence of gene silencing.

2.6 De-repressed gene expression and H-NS target site selection

Does the de-repressed expression level select DNA sites bound by H-NS? To address this question, we tested for association between the de-repressed expression level of a H-NS-regulated gene (both unilateral and epistatic) and the density of high affinity H-NS binding sites. We find that genes with the top 25% of H-NS binding affinities have a significantly higher de-repressed expression levels than those in the bottom 25% ($P = 2.1 \times 10^{-7}$, Wilcoxon test; Fig. 4c). In fact, there is a weak, but significant, correlation between the de-repressed expression level of a gene and the density of high affinity H-NS binding sites targeting them (Spearman rank correlation coefficient = 0.23; $P = 4.7 \times 10^{-8}$). This is not entirely explained by the separation of H-NS-regulated genes into unilateral and epistatic, which differ from each other in their binding affinities to H-NS as well as their de-repressed expression levels. Even within the set of unilateral genes, the above correlation is maintained (Spearman rank correlation coefficient = 0.21; $P = 8.3 \times 10^{-4}$). This is not true for epistatic genes for which the binding affinities to H-NS and the expression levels are already so high that the trend might be saturated. Thus we speculate that selection might have favoured H-NS binding affinities to be more directed towards genes that are highly expressed in the absence of a silencing system.

2.7 A+T-content and gene expression cost: a hypothesis

We investigated whether de-repressed expression of epistatic genes would be expensive to the cell in terms of the metabolic burden it imposes. In the present context, we are interested in the cost of transcription, as our study does not access protein levels. In general, gene expression cost is measured by the cost of breaking activated P bonds during both transcription and translation, in addition to the metabolic processes that synthesise the nucleotide and amino acid building blocks.[46] Here we suggest that transcription of an A-rich gene at high levels may impose a higher metabolic cost than that of a G-rich gene. This is based on the following arguments: (a) the incorporation of an 'A' in the mRNA abstracts ATP – both phosphates and the adenosine – from metabolic reactions. An inspection of 2000 enzymatic reactions listed in the Ecocyc database[47] shows that 15% of these utilise ATP. Though the synthesis of GTP from GMP requires the conversion of ATP to ADP, the adenosine itself is conserved. The loss of a GTP itself by incorporation in an RNA molecule is less expensive as it is utilised in only 1% of enzymatic reactions in Ecocyc. (b) A recent study by Raghavan and colleagues showed that high-level expression of GFP has a greater adverse effect on fitness, when the gene is A+T-rich.[48] This is despite the fact that GFP itself is not toxic to the cell, and that the expression level of GFP is not affected by the A+T-content of the gene encoding it. This work noted however that this effect was dependent on translation. Nevertheless, some studies have shown that translation has a

positive effect on mRNA stability (reviewed by ref. 49). Therefore, in the absence of translation, the mRNA could be degraded rapidly, allowing recycling of nucleotides. The Raghavan study in fact reports that high-A+T mRNA has smaller half-life than high G+C-mRNA, consistent with the sequence preferences of RNase E.[50]

Based on (a) above, we costed each copy of a gene as the number of 'A' multiplied by 0.15 plus the number of 'G' multiplied by 0.01. This is speculatively representative of the intrinsic cost of transcribing a gene, based only on its sequence characteristics. As expected from the high A+T content of H-NS-bound genes, both unilateral and epistatic genes have higher intrinsic cost than control genes, with epistatic genes only slightly costlier than unilateral genes (Fig. 4d; $P < 10^{-10}$ for epistatic and 4.2×10^{-7} for unilateral genes, compared to the negative control; $P = 1.6 \times 10^{-3}$, comparing epistatic with unilateral genes; Wilcoxon test). The slightly higher intrinsic cost for epistatic genes may be because of the tendency of these genes to be a bit longer than unilateral genes (mean gene length of 1072 nt for epistatic and 975 nt for unilateral genes), though there is little difference between the two sets of genes in their A+T-contents (mean A+T-content of 52.2% and 52.4% across the ORF for epistatic and unilateral genes respectively). The cost of expressing a gene also depends on the level to which it is expressed. Higher gene expression implies higher cost. Thus, under any given condition, the product of the intrinsic cost of a gene and its expression level can be thought to represent its transcriptional cost. The fact that epistatic genes transition from being silent in the wildtype to being extremely highly expressed in Δ*stpA-hns* mean that their transcriptional cost will be very high in the double mutant. This will be more muted for unilateral genes. Given that epistatic and unilateral genes differ only slightly in their intrinsic cost, any difference in the total cost is likely because of the higher de-repressed expression of epistatic genes. Whether this difference between the two sets of H-NS-regulated genes makes any contribution to the growth defect of the double mutant is difficult to comment on.

2.8 Summary of results

Our study suggests that gene silencing by H-NS in *E. coli* is directed more towards genes that are expressed at extremely high levels in the absence of this repressive system (Fig. 5). This operates at two, partially overlapping, levels. Highly transcribable (*i.e.* in the absence of gene silencing) genes are first enriched for high-affinity H-NS binding sites. Moreover, these genes also recruit StpA as a molecular backup for H-NS, thus building a global epistatic gene silencing network.

3 Discussion

3.1 Gene expression cost and targeted gene silencing

While horizontal gene transfer is a major force in bacterial evolution, there are many barriers to it. First, it interferes with a well-evolved cellular machinery. This interference might be because of the biochemical activities of the horizontally-acquired gene products. Or it might stem from high costs of gene expression.[7] Thus, the establishment of a horizontally acquired gene represents a balance between fitness benefit or powerful parasitism (*e.g.* toxin–antitoxin systems) and the costs it imposes.

The genome of *E. coli* K12 MG1655 has ~600 horizontally-acquired genes, predicted based on their oligonucleotide content.[41] Most of these have higher A+T-content than other genes. The gene silencing system around H-NS is directed towards repressing many highly transcribable genes, which have high A+T content. Here, the word *transcribable* is used to refer to the expression of a gene in the de-repressed Δ*stpA-hns*. The selection for silencing might operate at multiple levels. First, highly transcribable genes contain a higher density of

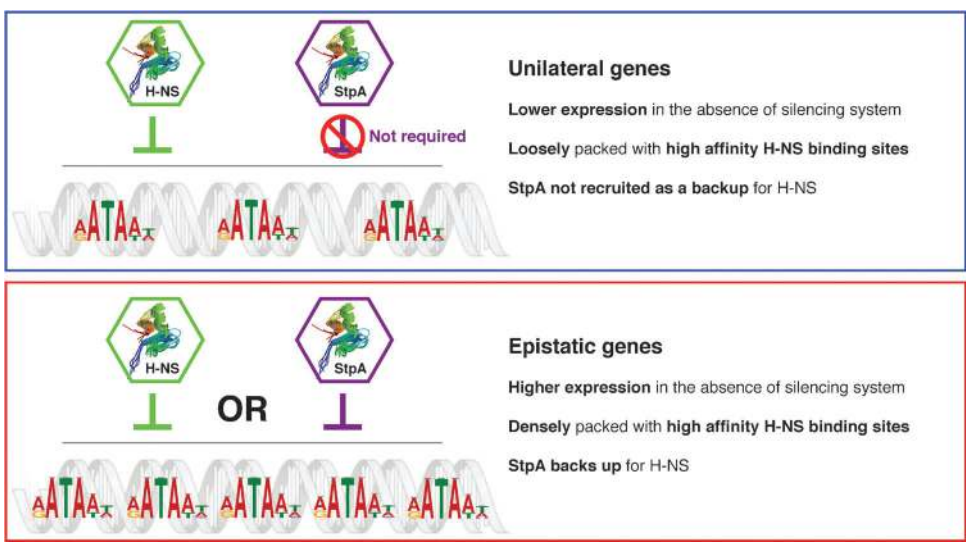

Fig. 5 A graphical summary of our results, comparing epistatic (red box) with unilateral (blue box) genes. The DNA sequence motif for H-NS is from Kahramanoglou *et al.*'s ChIP-seq study.[12] The structure of H-NS (within the green hexagon) is from the PDB ID 1HNS; this is also used to represent StpA (within the magenta hexagon) for which we could not find any structure.

high affinity H-NS binding motifs. This might imply that H-NS binding specificity has evolved to target highly transcribable DNA. A previous ChIP-chip study of H-NS has shown that higher expression of a gene might attract H-NS to it, at least for selected promoters.[11] This might be a conserved feature in many bacterial genomes, based on previous studies reporting highly similar sequence specificity for the silencing system in *Mycobacteria,*[14] which has only a low sequence similarity with H-NS. Second, StpA is recruited to highly transcribable genes in the absence of H-NS, thus presenting a second, backup layer of transcriptional repression. This selectivity of StpA to a subset of all its possible targets is likely to emerge from its relatively low expression levels.[30] Additionally, the formation of filaments of strong H-NS–StpA heterodimers at loci with high densities of high affinity H-NS binding motifs may also impose powerful transcriptional silencing. We also note here that certain gene functions might also select for epistatic gene silencing. For example, many genes involved in the biosynthesis of the O-antigen and capsular polysaccharides are epistatically regulated, as are several genes also under FNR control. We note that gene silencing can also be influenced by additional and unrelated players, such as the transcription terminator Rho,[51,52] and a study of their functional and/or physical interactions with H-NS is likely to emerge as an important area for future work.

Assuming that highly transcribable horizontally-acquired genes do provide their host with a selective advantage under certain conditions, their establishment in the host genome is facilitated by their silencing by H-NS (and StpA). An important question that emerges is the following. Is there a selective need for *E. coli* to maintain these sequences in a highly transcribable state? If not, one might surmise that mutations that decrease their transcription could have occurred. There might be several answers to this. One is that the acquisition of these genes occurred so recently that they have not had enough time to mutate sufficiently and get fixed in the population. Another possibility is that their expression is determined by the local geometry or topology of the DNA, which makes them highly permissible to transcription; this is in contrast to a few bases in the promoter region that permit high recruitment of RNA polymerase. Such a situation is likely to be more difficult to correct with point mutations. Alternatively, transcriptional silencing by H-NS might have ensured that there is no selection against their current sequence composition. A final hypothesis is that such a sequence composition is essential for these sequences to be functional. For example, it is believed that these sequences have important functional roles under certain stress conditions, including the stationary phase.[36,53] Under these circumstances, overall transcription is low, at least in part due to less negative supercoiling. Here, transcription may be favoured from highly transcribable DNA; in fact, a microarray study has shown that many genes that are up-regulated in response to a loss of negative supercoiling are A+T-rich.[54] Finally, the larger chromosomal context might play a role in gene expression patterns. Zarei, Scolari and colleagues[55,56] have observed that there is a statistical enrichment for H-NS regulated genes to be around the terminus of replication. Recent studies by Sobetzko *et al.*[57,58] observed that during exponential phase of growth, genes around the origin are more transcribed than those around the terminus. This is consistent with an inferred decrease in superhelical density from the origin to the terminus. However, under the stationary phase, the trend is reversed with higher expression observed around the terminus. We adopted the methods of Zarei, Scolari and colleagues[55,56,59] to compare the genomic coordinates of epistatic and unilateral genes in the context of chromosomal macrodomains, within which intrachromosomal recombination events occur preferentially.[60] Epistatic genes are preferentially located at the boundaries of these macrodomains – especially between the TER and the LEFT macrodomains and an edge of the RIGHT macrodomain (Fig. S1, ESI†). On the other hand, many unilateral genes are spread around the Ter macrodomain. The exact roles of global regulatory proteins including H-NS in chromosomal location and topology dependent gene expression programs remain to be understood in detail, although a recent study has studied the impact of local supercoiling on H-NS control of transcription from a specific bacterial promoter.[61]

3.2 An epistatic relationship between *hns* and *stpA*

Various studies, together straddling several strains of *E. coli,* have shown that a double mutant lacking both StpA and H-NS displays a growth and viability defect.[34,35] However, this effect cannot be predicted from the growth characteristics of the two single mutants alone. Though this appears to suggest a form of molecular backup between two homologous proteins, the backup is unequal: whereas the Δ*stpA* mutant hardly affects gene expression, a Δ*hns* mutant leads to upregulation of at least 500 genes as agreed upon by several studies.[12,13,15,34] Whereas the absence of *stpA* does not visibly affect *in vivo* DNA binding of H-NS on a genomic scale, that of *hns* abolishes binding of StpA to about two-thirds of its wildtype binding regions.[32] Though both H-NS and StpA repress each other's transcription, leading to the up-regulation of one in the absence of the other, the degree of up-regulation conferred on StpA by the absence of H-NS may not be sufficient to compensate for the lack of H-NS.[30] This in part might be due to the higher susceptibility of StpA to degradation by Lon in the absence of H-NS.[33] A complete backup of H-NS activity by StpA might not be desirable as there are conditions, such as a late stationary phase, in which an attenuated H-NS is favoured.[36]

The close relative of *E. coli*, *Salmonella enterica Typhimurium*, encodes three H-NS-like proteins. In this organism, the single Δ*stpA* mutant affects the expression of 5% of genes,[62] in contrast to what we, in this paper, and others, have reported for *E. coli*. The relationship of H-NS with StpA is in contrast to that with Sfh, another homolog of H-NS studied in *Salmonella*. Sfh binds to a subset of H-NS binding sites in wildtype cells, but expands its reach in Δ*hns.*[63] Sfh is different from StpA however, in the sense that it is a plasmid encoded protein, which serves the role of a stealth agent, as opposed to StpA which is well-ensconced in the chromosome. At the other end of the scale is *Yersinia*, also an enterobacteriaceae, which does not encode a H-NS homolog, and in which Δ*hns* is lethal.[64] Therefore, the

function of H-NS in concert with homologous regulators is taxa-dependent, offering scope for more detailed evolutionary analysis.

3.3 A molecular backup between two global transcriptional repressors

The use of homologous proteins (H-NS and StpA) acting together to silence expensive genes might confer robustness in the event of a mutation adversely affecting the function of either. Alternatively, during rapid exponential growth, the presence of multiple chromosomal copies might select for a second protein (read StpA) to target loci which may not be sufficiently bound by H-NS.[32,65] Under this model, one might expect selection to favour the positioning of epistatic genes in chromosomal regions of higher ploidy during exponential growth, *i.e.* closer to the origin than the terminus. We tested this hypothesis, using previously generated ChIP-seq mock-IP data for mid-exponential phase *E. coli,*[12,66] which by reflecting a quantitative genome sequencing experiment, identifies loci present in higher copies (Fig. S2, ESI†). We do not find any enrichment for epistatic genes to be present near the origin, or in any other way, in higher copy numbers. In fact, previous studies and our results presented in Fig. S1 (ESI†) report a tendency for H-NS-regulated genes to be localised around the edges of the chromosomal macrodomain comprising the terminus.[56] Therefore, the more parsimonious interpretation of our data might be in favour of a molecular backup in anticipation of a mutational event. Whether this might confer an evolutionary advantage especially in the face of large population sizes is debatable.

It has been previously suggested that there may be selection against target overlap between homologous regulators, on the basis of limited data available in the RegulonDB database.[67] However, this may not be valid for conditions and systems which are benefited by an epistatic circuit as described here.

Savageau's demand theory of gene regulation posits that the repressive mode of gene expression control is favoured over activation under situations where the target gene is rarely required,[68] which may well be true of the horizontally-acquired targets of H-NS and StpA. This is because the mutational loss of a repressor is expected to be selected against when a target gene is expressed constitutively in a low-demand regime. The fact that H-NS is a *global* transcriptional repressor, and that its loss leads to extremely *high expression* of *many* of its targets, might lead to an amplified selection against its loss or inactivation; this could select for backup by StpA.

Examples of molecular backups involving transcriptional regulators – not necessarily repressors – have been described on a genomic scale in the yeast *Saccharomyces cerevisiae*, where a whole genome duplication event generated a paralog of each gene, many of which have been retained.[69,70]

3.4 The growth defect of the Δ*stpA-hns* mutant: multiple causes?

The fact that a Δ*stpA* mutant in a Δ*hns* background has an effect on growth suggests that there are unique molecular features associated with the Δ*stpA-hns* double mutant, presumably at the transcriptional level. A study by the Uhlin group had demonstrated that the global regulator CRP is down-regulated by as much as five-fold in a Δ*stpA-hns* strain.[71] We generated a Δ*crp* mutant and observed that it shows only a small growth defect in LB medium; however a Δ*crp-hns* double mutant is very slow in growth (data not shown), suggesting a genetic interaction between the two. The above-mentioned work from the Uhlin group also showed that the double mutant might be experiencing a stringent response-like state, which is induced by the alarmone ppGpp. The down-regulation of CRP could be reversed in a mutant that was unable to synthesise ppGpp. This mutant, as well as overexpression of CRP, were reported to be partial suppressors of the growth defect seen in the Δ*stpA-hns* double mutant. The above study concluded that changes in the supercoiling state of the DNA in the Δ*stpA-hns* double mutant was the ultimate cause of these effects.

Our transcriptome data revealed a 3.5-fold down-regulation of the *crp* gene in the Δ*stpA-hns* strain. We performed an RNA-seq experiment with our Δ*crp* mutant and found that many genes that are down-regulated in Δ*stpA-hns* (when compared to Δ*hns*) are also differentially expressed in Δ*crp* (62% of genes down-regulated in Δ*stpA-hns* compared to Δ*hns* are differentially expressed in Δ*crp*). Most (92%) of these genes are also down-regulated in the Δ*crp* mutant, emphasising that the transcriptional effects are consistent between CRP and the H-NS–StpA pair. Half of these are not bound by H-NS, in contrast to the up-regulated genes where nearly three-fourths are bound by H-NS. This demonstrates that the down-regulation of a significant proportion of these genes is due to secondary effects, distal from H-NS and StpA. Note however that 50% of these genes are in fact bound by H-NS and it is not clear if H-NS – in some combination with CRP – can act as an activator at these loci. However, unlike in the study by the Uhlin group, we do not observe signatures of a stringent response transcriptome among the down-regulated genes, on the basis of microarray data generated by Traxler and colleagues.[72] The Uhlin study had also shown that changes in supercoiling effected by H-NS and StpA could be the primary cause of these effects. But, an overlap analysis performed using the NuST web server,[59] does not show any enrichment of supercoiling-sensitive genes[73] among those down-regulated in Δ*stpA-hns*.

Thus, a component of the growth defect observed in the double mutant might be explained by the down-regulation of CRP. Though this could indirectly emerge from the effect of H-NS (and StpA) on supercoiling, which in turn affects the stringent response,[71] we are unable to see signatures of these possible causes in our transcriptome. It may however be worth noting that the strain of *E. coli* used in the Uhlin study did not encode an active RelA (ppGpp synthase), whereas our strain has both RelA and SpoT, with the former accounting for much of the transcriptional effects of a Δ*relA-spoT* double mutant.[72]

Our analysis of transcriptome data generated here suggests that de-repressed expression of metabolically expensive genes might play an additional – admittedly minor – role in causing the growth defect of the Δ*stpA-hns* double mutant. And that this might be an additive effect spread across hundreds of genes. This may be supported, albeit tenuously at this point, by a

three-fold upregulation of the expression of genes encoding Tsx, a transporter of (deoxy)nucleosides,[74] and the transketolase TktB, which is involved in nucleoside biosynthesis not only by being a minor player in the production of ribose-5-phosphate,[75] but also by being part of the RNA degradosome.[76] In this context, it is also worth noting that TktA, a homolog of TktB, influences chromosome topology;[77] its role in the Δ*stpA-hns* or the Δ*hns* backgrounds is not known.

The extent to which gene expression cost might affect growth in the present context can only be speculated upon. More so since a rigorous mathematical definition of sequence-dependent gene expression cost that can be applied to functional genomic data is not available at present. However, cost of gene expression, including transcription and translation, is expressed in terms of growth of bacterial strains expressing specific non-selective genes such as GFP at high levels. Using such a definition of gene expression cost, a previous study had proposed that the cost of expressing unneeded proteins is at its highest during early stages of growth.[78] This, in the Δ*stpA-hns* context, may at best be reflected in the long lag phase and a slow early exponential growth. In fact, this may contribute to the small growth defect of the Δ*hns* single mutant as well.

Two additional gene expression-based hypotheses for the growth defect of Δ*stpA-hns*, namely the downregulation of an essential prophage-encoded gene *racR* and upregulation of a locus for capsular biosynthesis, were evaluated and tentatively invalidated. These are described in Fig. S3 (ESI†).

The slow growth phenotype of the double mutant may also derive contributions from non-transcriptional sources. For example, H-NS and StpA may be involved in repair of double strand DNA breaks,[79] with the former shown to bind to Holliday junctions *in vitro*.[19] The possible occurrence of such events especially as part of homologous recombination occurring during rapid exponential growth, might also contribute to the growth defect.

4 Methods

4.1 Strains and general growth conditions

The *Escherichia coli* variants used in the work are following: *E. coli* K-12 MG1655 wildtype (CGSC #6300); MG1655 Δ*hns* (Δ*hns*::*kan*^{r}); MG1655 Δ*stpA* (Δ*stpA*::*kan*^{r}); MG1655 Δ*hha* (Δ*hha*::*kan*^{r}); MG1655 Δ*ydgT* (Δ*ydgT*::*kan*^{r}); MG1655 Δ*stpA-hns* (Δ*hns-stpA*::*kan*^{r}); Δ*hha-hns* (Δ*hns-hha*::*kan*^{r}); Δ*ydgT-hns* (Δ*hns-ydgT*::*kan*^{r}). Luria broth was used for normal growth. 50 μg mL^{-1} of Ampicillin or Kanamycin was used as per requirement, primarily during the gene deletion process.

4.2 Construction of MG1655 knock-outs

All gene deletions were achieved by the λ Red recombination system, described by Datsenko and Wanner[80] using plasmids pKD46 and pKD4 or pKD13. Knockout strains generated by this method were selected on LB Kanamycin (50 μg mL^{-1}) plates and deletion was confirmed by PCR using specific primers. Primers used for the deletion and their detection are given in Table S1 (ESI†). Double knockouts were also generated by the same method (Datsenko and Wanner) after removing the kanamycin cassette using pCP20.

4.3 Growth curves

Overnight grown culture was inoculated in fresh LB to 1 : 100 ratio and the growth of cells monitored by measuring the optical density at 600 nm. All these growth experiments were performed in 96 well plates, incubated at 37 °C in a plate reader (Tecan, infinite® F200 PRO) with constant shaking at 87 rpm. OD_{600} was measured every ∼16 minutes. We note that this could cause issues with aeration; nevertheless, similar growth curves were observed in flasks rotating at 200 rpm in shakers.

4.4 RNA extraction and mRNA enrichment

For RNA extraction the overnight cultures were inoculated in 100 μL of fresh LB to bring the initial OD of the fresh culture to 0.03 and the flasks were incubated at 37 °C with shaking at 200 rpm. Two biological replicates were performed for each sample. Samples were collected at the mid-exponential phase (OD_{600} ∼ 0.5 for Δ*stpA-hns* and ∼0.9 for all other strains) of growth. Slight modifications of previously published protocols,[12] based on those recommended by the manufacturer for the TRIzol (Invitrogen) bacterial RNA isolation kit, were used. Briefly, cells were pelleted down at 8000 rpm for 10 minutes at 4 °C. Pellets were washed with autoclaved DEPC-treated RNAse-free water and cells were snap-frozen in liquid nitrogen and stored at −80 °C until required. Cells were homogenized with sterile pestle followed by RNA extraction using TRIzol reagent according to manufacturer's instruction until the 70% ethanol wash step. Total RNA was treated with DNAse I (Invitrogen Cat No. 18068-015) according to the manufacturer's instruction. Further precipitation of RNA and ribosomal RNA cleanup was achieved by MICROBExpress bacterial mRNA purification Kit. Total RNA was suspended in 10 μL of RNAse free water (Ambion®, Cat No. AM9932). Concentration and the quality of the RNA were determined on a Nanodrop and by visualization under agarose gel respectively.

Sequencing libraries were prepared using TruSeq RNA sample preparation kit v2 (Illumina, Catalog No. RS-122-2001) according to the manufacturer's guidelines. Prepared libraries were checked for quality on an Agilent Bioanalyzer, and sequenced for 50 cycles – from one end – on an Illumina HiSeq1000 platform. Raw data have been deposited with GEO under the accession number GSE40313.

4.5 Data analysis

The 50-mer single-end sequence reads were mapped to the *E. coli* K12 MG1655 genome using BWA.[81] Gene annotations were obtained from the Ecocyc database.[47] The number of reads falling within each gene was calculated based on the base position to which the first nucleotide of the read was mapped. A matrix of read counts was generated with 16 columns, one per strain (n = 8) each in duplicate, and 4245 rows, one for each mRNA gene. This matrix was fed into the Bioconductor (http://www.bioconductor.org) package EdgeR[82] for analysis of differential expressions. To calculate relative expression levels

of a gene within a sample, the read count for each gene was first divided by the length of the gene, and then by the mode of the distribution of expression measures thus obtained across all genes for that sample. These values are expressed on a $\log_2$ scale. The fold changes computed by EdgeR correlates well with those obtained by the above method (Pearson correlation coefficient = 0.94). The normalised gene expression matrix was subjected to single-linkage hierarchical clustering using the program Cluster, and the result viewed using Treeview and the Matrix2png web server (http://www.chibi.ubc.ca/matrix2png/). Third party data were obtained from various sources, all listed in the Results. Statistical analyses were performed in R (http://r-project.org).

4.6 Analysis of H-NS binding affinity to DNA sites

In vitro affinity data for H-NS to 8-mer oligonucleotides were generated by Gordon and colleagues[14] using a protein binding microarray in which various double stranded oligonucleotides were spotted and their affinities to a protein (H-NS in this case) applied to the array measured. The binding affinity towards each oligo was expressed in the form of a *z*-score (Z). H-NS and StpA bind to similar target sites; therefore the above-described semi-quantitative and relative measure of affinity of different DNA sites to H-NS can be readily applied to StpA. We first defined high affinity H-NS/StpA binding sites as those with $Z \geq 2$, which includes $\sim$5% of all the oligonucleotides tested. These sites include a majority of sequences containing the sites ATAAA (56%) and AATAA (68%), which form part of the high-affinity H-NS binding motifs identified by computational analysis by Lang *et al.*[83] and Kahramanoglou *et al.*[12] respectively. These sites should be clearly contrasted with the much larger binding regions identified by ChIP-seq or ChIP-chip analysis; in fact most ChIP-seq/ChIP-chip binding regions would contain one or more binding motifs. We scored each ChIP-seq-defined H-NS binding region by the density of high affinity sites contained in them, *i.e.* the proportion of all 8-mers within the binding region (using a sliding window of 1 nt) that were classified as having high affinity for H-NS. This value was then transferred to the gene associated with the binding region.

Author contributions

ASNS conceived and designed the study. RS performed the experiments and was helped by DC, RK and PS. ASNS performed the bioinformatics and analysed the data with help from VS. SK provided advice. ASNS wrote the manuscript, with input from RS.

Competing financial interests

ASNS serves as the 'DBT nominee' on the Institutional Biosafety Committee of Novozymes, Bangalore, India.

Accession

All transcriptome data are available at GEO at GSE40313.

Acknowledgements

Funding: RS and RK are funded by the INSPIRE scholarship from the Department of Science and Technology, Government of India. PS is funded by a Junior Research Fellowship from the University Grants Commission, Government of India. ASNS and SK are funded by the National Centre for Biological Sciences core funding, and ASNS by the Ramanujan Fellowship from the Department of Science and Technology, Government of India. VFS is founded by the PDI MSC scholarship of the Institut de recherche pour le développement, Government of France. High-throughput sequencing was performed at the Next Generation Genomics Laboratory, Centre for Cellular and Molecular Platforms. We thank Marco Cosentino-Lagomarsino, Mahadevan S, Babu Ponnusamy and Dasaradhi Palakodeti for critical reading of the manuscript. We thank the anonymous referees for their vital contributions that helped improve the manuscript. We thank CGSC for providing strains and plasmids.

References

1 W. Davids and Z. Zhang, *BMC Evol. Biol.*, 2008, **8**, 23, DOI: 10.1186/1471-2148-8-23.
2 M. J. Lercher and C. Pal, *Mol. Biol. Evol.*, 2008, **25**, 559–567.
3 C. Pal, B. Papp and M. J. Lercher, *Nat. Genet.*, 2005, **37**, 1372–1375.
4 A. T. Maurelli, R. E. Fernández, C. A. Bloch, C. K. Rode and A. Fasano, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 3943–3948.
5 E. P. Rocha, A. Danchin and A. Viari, *Genome Res.*, 2001, **11**, 946–958.
6 A. S. N. Seshasayee, P. Singh and S. Krishna, *Nucleic Acids Res.*, 2012, **40**, 7066–7073.
7 C. Park and J. Zhang, *Genome Biol. Evol.*, 2012, **4**, 523–532.
8 W. W. Navarre, M. McClelland, S. J. Libby and F. C. Fang, *Genes Dev.*, 2007, **21**, 1456–1471.
9 C. J. Dorman, *Nat. Rev. Microbiol.*, 2007, **5**, 157–161.
10 A. Martinez-Antonio and J. Collado-Vides, *Curr. Opin. Microbiol.*, 2003, **6**, 482–489.
11 D. C. Grainger, D. Hurd, M. D. Goldberg and S. J. W. Busby, *Nucleic Acids Res.*, 2006, **34**, 4642–4652.
12 C. Kahramanoglou, A. S. N. Seshasayee, A. I. Prieto, D. Ibberson, S. Schmidt, J. Zimmermann, V. Benes, G. M. Fraser and N. M. Luscombe, *Nucleic Acids Res.*, 2011, **39**, 2073–2091.
13 T. Oshima, S. Ishikawa, K. Kurokawa, H. Aiba and N. Ogasawara, *DNA Res.*, 2006, **13**, 141–153.
14 B. R. G. Gordon, Y. Li, A. Cote, M. T. Weirauch, P. Ding, T. R. Hughes, W. W. Navarre, B. Xia and J. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10690–10695.
15 S. Lucchini, G. Rowley, M. D. Goldberg, D. Hurd, M. Harrison and J. C. D. Hinton, *PLoS Pathog.*, 2006, **2**, e81, DOI: 10.1371/journal.ppat.0020081.
16 R. T. Dame, C. Wyman and N. Goosen, *Biochimie*, 2001, **83**, 231–234.
17 T. A. Owen-Hughes, G. D. Pavitt, D. S. Santos, J. M. Sidebotham, C. S. Hulton, J. C. Hinton and C. F. Higgins, *Cell*, 1992, **71**, 255–265.

18 R. Spurio, M. Falconi, A. Brandi, C. L. Pon and C. O. Gualerzi, *EMBO J.*, 1997, **16**, 1795–1805.
19 N. Sharadamma, Y. Harshavardhana, P. Singh and K. Muniyappa, *Nucleic Acids Res.*, 2010, **38**, 3555–3569.
20 S. Saxena and J. Gowrishankar, *J. Bacteriol.*, 2011, **193**, 3832–3841.
21 C. C. Brescia, M. K. Kaw and D. D. Sledjeski, *J. Mol. Biol.*, 2004, **339**, 505–514.
22 R. T. Dame, M. C. Noom and G. J. L. Wuite, *Nature*, 2006, **444**, 387–390.
23 Y. Liu, H. Chen, L. J. Kenney and J. Yan, *Genes Dev.*, 2010, **24**, 339–344.
24 S. Maurer, J. Fritz and G. Muskhelishvili, *J. Mol. Biol.*, 2009, **387**, 1261–1276.
25 W. Wang, G.-W. Li, C. Chen, X. S. Xie and X. Zhuang, *Science*, 2011, **333**, 1445–1449.
26 W. W. Navarre, S. Porwollik, Y. Wang, M. McClelland, H. Rosen, S. J. Libby and F. C. Fang, *Science*, 2006, **313**, 236–238.
27 C. J. Cardinale, R. S. Washburn, V. R. Tadigotla, L. M. Brown, M. E. Gottesman and E. Nudler, *Science*, 2008, **320**, 935–938.
28 D. M. Stoebel, A. Free and C. J. Dorman, *Microbiology*, 2008, **154**, 2533–2545.
29 A. Zhang, S. Rimsky, M. E. Reaban, H. Buc and M. Belfort, *EMBO J.*, 1996, **15**, 1340–1349.
30 J. M. Sonnenfield, C. M. Burns, C. F. Higgins and J. C. Hinton, *Biochimie*, 2001, **83**, 243–249.
31 C. J. Lim, Y. R. Whang, L. J. Kenney and J. Yan, *Nucleic Acids Res.*, 2012, **40**, 3316–3328.
32 E. Uyar, K. Kurokawa, M. Yoshimura, S. Ishikawa, N. Ogasawara and T. Oshima, *J. Bacteriol.*, 2009, **191**, 2388–2391.
33 J. Johansson and B. E. Uhlin, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 10776–10781.
34 C. M. Muller, U. Dobrindt, G. Nagy, L. Emödy, B. E. Uhlin and J. Hacker, *J. Bacteriol.*, 2006, **188**, 5428–5438.
35 B. Sonden and B. E. Uhlin, *EMBO J.*, 1996, **15**, 4970–4980.
36 S. Chib and S. Mahadevan, *J. Bacteriol.*, 2012, **194**, 5285–5293.
37 S. Rodriguez, J. M. Nieto, C. Madrid and A. Juárez, *J. Bacteriol.*, 2005, **187**, 5452–5459.
38 R. C. Banos, A. Vivero, S. Aznar, J. García, M. Pons, C. Madrid and A. Juarez, *PLoS Genet.*, 2009, **5**, e1000513.
39 A. Vivero, R. C. Banos, J. F. Mariscotti, J. C. Oliveros, F. García-del Portillo, A. Juárez and C. Madrid, *J. Bacteriol.*, 2008, **190**, 1152–1156.
40 P. G. Leonard, S. Ono, J. Gor, S. J. Perkins and J. E. Ladbury, *Mol. Microbiol.*, 2009, **73**, 165–179.
41 G. S. Vernikos and J. Parkhill, *Bioinformatics*, 2006, **22**, 2196–2203.
42 B. J. Yu, B. H. Sung, M. D. Koob, C. H. Lee, J. H. Lee, W. S. Lee, M. S. Kim and S. C. Kim, *Nat. Biotechnol.*, 2002, **20**, 1018–1023.
43 G. Posfai, G. Plunkett, 3rd, T. Fehér, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum and F. R. Blattner, *Science*, 2006, **312**, 1044–1046.
44 M. Hashimoto, T. Ichimura, H. Mizoguchi, K. Tanaka, K. Fujimitsu, K. Keyamura, T. Ote, T. Yamakawa, Y. Yamazaki, H. Mori, T. Katayama and J.-i. Kato, *Mol. Microbiol.*, 2005, **55**, 137–149.
45 T. Vora, A. K. Hottes and S. Tavazoie, *Mol. Cell*, 2009, **35**, 247–253.
46 A. Wagner, *Mol. Biol. Evol.*, 2005, **22**, 1365–1374.
47 I. M. Keseler, A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martínez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, M. Latendresse, L. Muniz-Rascado, Q. Ong, S. Paley, I. Schröder, A. G. Shearer, P. Subhraveti, M. Travers, D. Weerasinghe, V. Weiss, J. Collado-Vides, R. P. Gunsalus, I. Paulsen and P. D. Karp, *Nucleic Acids Res.*, 2013, **41**, D605–D612.
48 R. Raghavan, Y. D. Kelkar and H. Ochman, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 14504–14507.
49 A. Deana and J. G. Belasco, *Genes Dev.*, 2005, **19**, 2526–2533.
50 K. J. McDowall, S. Lin-Chao and S. N. Cohen, *J. Biol. Chem.*, 1994, **269**, 10790–10796.
51 J. M. Peters, R. A. Mooney, J. A. Grass, E. D. Jessen, F. Tran and R. Landick, *Genes Dev.*, 2012, **26**, 2621–2633.
52 S. Saxena and J. Gowrishankar, *J. Bacteriol.*, 2011, **193**, 3832–3841.
53 X. Wang, Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino and T. K. Wood, *Nat. Commun.*, 2010, **1**, 147, DOI: 10.1038/ncomms1146.
54 B. J. Peter, J. Arsuaga, A. M. Breier, A. B. Khodursky, P. O. Brown and N. R. Cozzarelli, *Genome Biol.*, 2004, **5**, R87, DOI: 10.1186/gb-2004-5-11-r87.
55 M. Zarei, B. Sclavi and M. C. Lagomarsino, *Mol. BioSyst.*, 2013, **9**, 758–767.
56 V. F. Scolari, B. Bassetti, B. Sclavi and M. C. Lagomarsino, *Mol. BioSyst.*, 2011, **7**, 878–888.
57 P. Sobetzko, M. Glinkowska, A. Travers and G. Muskhelishvili, *Mol. BioSyst.*, 2013, DOI: 10.1039/C3MB25515H.
58 P. Sobetzko, A. Travers and G. Muskhelishvili, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E42–E50.
59 V. F. Scolari, M. Zarei, M. Osella and M. C. Lagomarsino, *Bioinformatics*, 2012, **28**, 1643–1644.
60 M. Valens, S. Penaud, M. Rossignol, F. Cornet and F. Boccard, *EMBO J.*, 2004, **23**, 4330–4341.
61 Z.-A. Ouafa, S. Reverchon, T. Lautier, G. Muskhelishvili and W. Nasser, *Nucleic Acids Res.*, 2012, **40**, 4306–4319.
62 S. Lucchini, P. McDermott, A. Thompson and J. C. D. Hinton, *Mol. Microbiol.*, 2009, **74**, 1169–1186.
63 S. C. Dillon, A. D. S. Cameron, K. Hokamp, S. Lucchini, J. C. D. Hinton and C. J. Dorman, *Mol. Microbiol.*, 2010, **76**, 1250–1265.
64 R. C. Banos, J. I. Pons, C. Madrid and A. Juárez, *Microbiology*, 2008, **154**, 1281–1289.
65 M. C. Noom, W. W. Navarre, T. Oshima, G. J. L. Wuite and R. T. Dame, *Curr. Biol.*, 2007, **17**, R913–R914.
66 A. I. Prieto, C. Kahramanoglou, R. M. Ali, G. M. Fraser, A. S. N. Seshasayee and N. M. Luscombe, *Nucleic Acids Res.*, 2012, **40**, 3524–3537.
67 M. C. Lagomarsino, P. Jona, B. Bassetti and H. Isambert, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 5516–5520.

68 M. A. Savageau, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 2453–2455.
69 A. Gitter, Z. Siegfried, M. Klutstein, O. Fornes, B. Oliva, I. Simon and Z. Bar-Joseph, *Mol. Syst. Biol.*, 2009, **5**, 276, DOI: 10.1038/msb.2009.33.
70 J. Ihmels, S. R. Collins, M. Schuldiner, N. J. Krogan and J. S. Weissman, *Mol. Syst. Biol.*, 2007, **3**, 86, DOI: 10.1038/msb4100127.
71 J. Johansson, C. Balsalobre, S. Y. Wang, J. Urbonaviciene, D. J. Jin, B. Sondén and B. E. Uhlin, *Cell*, 2000, **102**, 475–485.
72 M. F. Traxler, S. M. Summers, H.-T. Nguyen, V. M. Zacharia, G. A. Hightower, J. T. Smith and T. Conway, *Mol. Microbiol.*, 2008, **68**, 1128–1148.
73 N. Blot, R. Mavathur, M. Geertz, A. Travers and G. Muskhelishvili, *EMBO Rep.*, 2006, **7**, 710–715.
74 R. Benz, A. Schmid, C. Maier and E. Bremer, *Eur. J. Biochem.*, 1988, **176**, 699–705.
75 A. Iida, S. Teshiba and K. Mizobuchi, *J. Bacteriol.*, 1993, **175**, 5375–5383.
76 M. E. Regonesi, M. Del Favero, F. Basilico, F. Briani, L. Benazzi, P. Tortora, P. Mauri and G. Deho, *Biochimie*, 2006, **88**, 151–161.
77 C. D. Hardy and N. R. Cozzarelli, *Mol. Microbiol.*, 2005, **57**, 1636–1652.
78 I. Shachrai, A. Zaslaver, U. Alon and E. Dekel, *Mol. Cell*, 2010, **38**, 758–767.
79 K. Shiraishi, Y. Ogata, K. Hanada, Y. Kano and H. Ikeda, *Genes Genet. Syst.*, 2007, **82**, 433–439.
80 K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6640–6645.
81 H. Li and R. Durbin, *Bioinformatics*, 2009, **25**, 1754–1760.
82 M. D. Robinson, D. J. McCarthy and G. K. Smyth, *Bioinformatics*, 2010, **26**, 139–140.
83 B. Lang, N. Blot, E. Bouffartigues, M. Buckle, M. Geertz, C. O. Gualerzi, R. Mavathur, G. Muskhelishvili, C. L. Pon, S. Rimsky, S. Stella, M. M. Babu and A. Travers, *Nucleic Acids Res.*, 2007, **35**, 6330–6337.