OxyR-dependent formation of DNA methylation patterns in OpvAB^OFF and OpvAB^ON cell lineages of *Salmonella enterica*

Ignacio Cota[1], Boyke Bunk[2,3], Cathrin Spröer[2,3], Jörg Overmann[2,3], Christoph König[4] and Josep Casadesús[1,*]

[1]Departamento de Genética, Universidad de Sevilla, Facultad de Biología, Apartado 1095, 41080 Sevilla, Spain,
[2]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, 38124 Braunschweig, Germany,
[3]German Centre of Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany and [4]Pacific Biosciences, 1380 Willow Rd, Menlo Park, CA 94025, USA

ABSTRACT

Phase variation of the *Salmonella enterica opvAB* operon generates a bacterial lineage with standard lipopolysaccharide structure (OpvAB^OFF) and a lineage with shorter O-antigen chains (OpvAB^ON). Regulation of OpvAB lineage formation is transcriptional, and is controlled by the LysR-type factor OxyR and by DNA adenine methylation. The *opvAB* regulatory region contains four sites for OxyR binding (OBS_{A-D}), and four methylatable GATC motifs (GATC_{1–4}). OpvAB^OFF and OpvAB^ON cell lineages display opposite DNA methylation patterns in the *opvAB* regulatory region: (i) in the OpvAB^OFF state, GATC_1 and GATC_3 are non-methylated, whereas GATC_2 and GATC_4 are methylated; (ii) in the OpvAB^ON state, GATC_2 and GATC_4 are non-methylated, whereas GATC_1 and GATC_3 are methylated. We provide evidence that such DNA methylation patterns are generated by OxyR binding. The higher stability of the OpvAB^OFF lineage may be caused by binding of OxyR to sites that are identical to the consensus (OBS_A and OBS_c), while the sites bound by OxyR in OpvAB^ON cells (OBS_B and OBS_D) are not. In support of this view, amelioration of either OBS_B or OBS_D locks the system in the ON state. We also show that the GATC-binding protein SeqA and the nucleoid protein HU are ancillary factors in *opvAB* control.

INTRODUCTION

For decades, bacteriological research was based on the study of large populations of bacterial cells in batch cultures. This experimental approach assumed that the value of any parameter measured in the population would reflect a unimodal distribution around the average value in individual cells. This may be true for many cellular parameters. However, in the last two decades single cell analysis has shown that clonal populations of bacteria, even when growing in homogeneous environments, can exhibit phenotypic heterogeneity between individual cells (1–3). In certain cases, phenotypic heterogeneity reflects the occurrence of bistability, the formation of two subpopulations with distinct patterns of gene expression (4,5). Phenotypic diversity can be also generated by reversible ON-OFF switching of gene expression at high frequencies, a phenomenon known as phase variation (6,7). In bacterial pathogens, phase variation often occurs at loci that encode envelope structures and may be viewed as a strategy to generate programmed polymorphism (6,7). Indeed, lineage formation can help to evade the host immune system and to protect bacterial subpopulations against bacteriophage infection, among other potential adaptive advantages (7).

The molecular mechanisms of phase variation are diverse. Some are genetic, such as site-specific recombination (8) and slipped-strand mispairing in tracts of repetitive DNA sequences (9). In other cases, however, the formation of bacterial lineages has epigenetic origin, without alteration of the DNA sequence (10–12). Some of the best known examples of epigenetic phase variation involve the formation of heritable DNA adenine (Dam) methylation patterns (10–12). The list includes the *pap* operon of uropathogenic *Escherichia coli*, which encodes fimbrial adhesins for adherence to the urinary tract epithelium (10,13), the *agn43* aggregation gene of *E. coli* (14) and the glycosyltransferase operon *gtr* of *Salmonella enterica* (15). In all these cases, a transcriptional regulator binds a regulatory region that contains GATC sites, which become non-methylated because binding of the regulator hinders Dam methylase activity. The OFF and ON states of the phase variation locus thus differ in the methylation state of crit-

ical GATC sites (12). Because DNA base methylation often prevents or restrains binding of proteins to DNA, non-methylation can increase binding of the regulatory protein, thus generating a positive feedback loop that propagates the epigenetic state (12). However, in all phase variation systems both the OFF and ON states are metastable, which permits phenotypic switching after a number of generations (6,7). The switching frequencies are idiosyncratic for each phase variation locus, and may vary depending on culture conditions (12).

The *opvAB* locus of *S. enterica* serovar Typhimurium, previously annotated as *STM2209-STM2208*, is a *Salmonella*-specific locus that encodes cytoplasmic membrane proteins involved in control of O-antigen chain length (16). The *opvA* and *opvB* genes form a bicistronic transcriptional unit, which is transcribed from a canonical, σ^{70}-dependent promoter under the control of the LysR-type factor OxyR (16). Expression of *opvAB* is phase variable, and *S. enterica* batch cultures contain subpopulations of OpvAB[OFF] and OpvAB[ON] cells. Each subpopulation harbors a distinct type of O-antigen, and OpvAB-mediated modification renders OpvAB[ON] cells resistant to bacteriophages that use the O-antigen as receptor (17). However, the OpvAB[ON] subpopulation shows sensitivity to serum, reduced capacity to proliferate in macrophages and attenuation in the mouse model (16,17).

In this work we describe the epigenetic mechanism responsible for the formation of OpvAB[OFF] and OpvAB[ON] cell lineages in *S. enterica*. Each lineage shows a distinct pattern of GATC methylation at the *opvAB* regulatory region. We present evidence that such patterns are generated by differential OxyR binding at the *opvAB* regulatory region. We also show that the GATC-binding protein SeqA and the nucleoid protein HU are ancillary factors for *opvAB* lineage formation. Finally, we present a model of *opvAB* phase variation, partly based upon experimental evidence, partly inspired by literature data and containing some speculative elements as well.

MATERIALS AND METHODS

Bacterial strains, bacteriophage, media and culture conditions

The strains of *S. enterica* used in this study (Supplementary Table S1) belong to serovar Typhimurium, and originate from strain ATCC 14028. For simplicity, *S. enterica* serovar Typhimurium is routinely abbreviated as *S. enterica*. *E. coli* CC118 λ *pir* [*phoA20 thi-1 rspE rpoB argE(Am) recA1 (λ pir)*] and *E. coli* S17–1 λ *pir* [*recA pro hsdR* RP4–2-Tc::Mu-Km::Tn*7 (λ pir)*] were used for directed construction of point mutations. *E. coli* M15 [pREP4] (Qiagen, Valencia, CA, USA) was used for 6×His-OxyR[C199S] production. Plasmid pTP166 (18) was kindly provided by Martin G. Marinus, University of Massachusetts, Worcester, MA, USA.

Bertani's lysogeny broth (LB) was used as standard liquid medium. Solid LB contained agar at 1.5% final concentration. Green plates (19) contained methyl blue (Sigma-Aldrich, St Louis, MO, USA) instead of aniline blue. The indicator for monitoring β-galactosidase activity in plate tests was 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal; Sigma-Aldrich, 40 μg/ml). Antibiotics were used at the concentrations described previously (20). To grow *oxyR* strains on LB agar, 75 μl of a 10 mg/ml catalase solution (Sigma-Aldrich, St Louis, MO, USA) were spread on the surface of the plates.

Transductional crosses using phage P22 HT 105/1 *int201* (21) were used for construction of strains with altered chromosomal markers. The transduction protocol has been described elsewhere (22). To obtain phage-free isolates, transductants were purified by streaking on green plates. Phage sensitivity was tested by cross-streaking with the clear-plaque mutant P22 H5.

The oligonucleotides used in this study have either been described previously (16) or are listed in Supplementary Table S2. Gene disruption was achieved using plasmids pKD3, pKD4 and pKD13 (23) and oligonucleotides *PS1*, *PS2* or *PS4*. Verification of the constructs was achieved using oligonucleotides *E1* and *E2*. Antibiotic resistance cassettes introduced during strain construction were excised by recombination with plasmid pCP20 (23).

Directed construction of point mutations

Mutation of GATC sites within the *opvAB* regulatory region was achieved using previously described procedures and oligonucleotides (16). Additional primers are included in Supplementary Table S2. Antibiotic resistance cassettes from pKD3 and pKD4 were introduced in *opvAB*::*lac* and *opvAB*::*gfp* backgrounds using oligonucleotides *delGATC-PS1* and *delGATC-PS2*, respectively (16). The resulting strains were used as intermediates in the construction of point mutations. Mutation of OxyR binding sites was achieved in the same way using primers labeled *OxyRB* and *OxyRD* (Supplementary Table S2).

β-galactosidase assays

Bacterial cultures were grown in LB until stationary phase (O.D.$_{600}$ ~4). Levels of β-galactosidase activity were assayed using the CHCl$_3$-sodium dodecyl sulfate permeabilization procedure (24). All data are averages and standard deviations from more than three independent experiments.

Calculation of phase transition frequencies

Phase transition rates were estimated as described by Eisenstein (25). Briefly, a strain harboring an *opvAB*::*lac* fusion was plated on LB + X-gal. After 16 h growth at 37°C, colonies displaying ON and OFF phenotypes were chosen, resuspended in phosphate buffered saline (PBS) and respread on new plates. Phase transition frequencies were calculated using the formula $(M/N)/g$ where M is the number of cells that underwent a phase transition, N the total number of cells scored, and g the total number of generations that gave rise to the colony.

Flow cytometry

Bacterial cultures were grown in LB at 37°C until exponential phase (O.D.$_{600}$ ~0.3). Cells were then diluted in PBS to a final concentration of ~10^7/ml. Data acquisition and

analysis were performed using a Cytomics FC500-MPL cytometer (Beckman Coulter, Brea, CA, USA). Data were collected for 100 000 events per sample, and were analyzed with CXP and FlowJo8.7 software. Data are represented by a dot plot (forward scatter [cell size] versus fluorescence intensity [*opvAB::gfp* expression]).

Construction of plasmid pIZ1885 (pQE30::*oxyR*[C199S])

A DNA fragment containing *oxyR*[C199S] (16) was amplified using oligonucleotides *His-oxyR-BamHI-5* and *His-oxyR-SalI-3*, and cloned into pQE30 (Qiagen, Valencia, CA, USA) using the BamHI and SalI sites. The recombinant plasmid (pIZ1885) was verified by restriction analysis and DNA sequencing.

Purification of OxyR protein

For 6×His-OxyR[C199S] purification, plasmid pIZ1885 was transformed into *E. coli* M15 [pREP4] (Qiagen, Valencia, CA, USA). M15/pIZ1885 was grown in LB broth containing ampicillin, and expression of 6×His-OxyR[C199S] was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). After 3 h of induction, cells were centrifuged and resuspended in 10 ml of lysis buffer (20 mM Tris, 300 mM NaCl, 10 mM imidazole) per g of pelleted cells, and were lysed by sonication. The suspension was centrifuged at 10 000 rpm for 30 min and the supernatant containing the soluble fraction of 6×His-OxyR[C199S] was transferred to a HisTrap HP nickel affinity chromatography column (GE Healthcare, Wauwatosa, WI, USA). The column was washed with 4 ml of lysis buffer, 4 ml of washing buffer (20 mM Tris, 300 mM NaCl, 30 mM imidazole) and 4 ml of the same buffer with 50 mM imidazole. Protein elution was performed with 3 ml of elution buffer (20 mM Tris, 300 mM NaCl, 300 mM imidazole). Elution fractions enriched in 6×His-OxyR[C199S] were selected and combined. Imidazole was removed by transferring to an Amicon® ultra centrifugal filter (Merck Millipore, Darmstadt, Germany) and washing with storage buffer (20 mM Tris, 300 mM NaCl, 10% glycerol) or by dialyzing in cellulose membranes (Sigma-Aldrich, St Louis, MO, USA). 6×His-OxyR[C199S] was either used immediately or frozen in liquid nitrogen and stored at −80°C.

Gel mobility shift assay

A DNA fragment containing predicted OxyR binding sites in the *opvAB* regulatory region and labeled with 6-carboxyfluorescein (6-FAM) was prepared by polymerase chain reaction (PCR) amplification using primers *FAMGATClargo-5* and *FAMGATClargo-3* (Supplementary Table S2). The PCR product was purified with the Wizard® SV Clean-Up System (Promega). The *envR* control fragment was prepared using primers *envR-For-Dnase* and *envR-Rev-Dnase* (26), and was kindly provided by Elena Espinosa. Thirty five nanogram were used for each reaction. The FAM-labeled probe was incubated at room temperature for 30 min with increasing concentrations of purified 6×His-OxyR[C199S] in a final volume of 20 μl with 1× OxyR binding buffer [25 mM Tris–HCl pH 7.5, 50 mM KCl, 5 mM MgCl$_2$, 5% glycerol, 50 μg/ml bovine serum albumin (BSA), 1 mM DTT, 1 μg/ml poly(dI-dC)]. Protein–DNA complexes were subjected to electrophoresis at 4°C in a 5% non-denaturing polyacrylamide gel in Tris-glycine-ethylenediaminetetraacetic acid (EDTA) buffer (25 mM Tris–HCl pH 7.5, 380 mM glycine, 1.5 mM EDTA). The gel was then analyzed in a FLA-5100 Scanner (Fujifilm, Tokyo, Japan).

DNA methylation *in vitro*

PCR fragments were methylated *in vitro* using Dam methylase (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions and subsequently digested with MboI (New England Biolabs). The undigested product was purified using the Wizard® SV Clean-Up system (Promega, Madison, WI, USA).

DNase I footprinting

DNA probes containing the *opvAB* promoter and the upstream regulatory region, labeled with 6-carboxyfluorescein (6-FAM) at the opposite ends, were prepared by PCR amplification using the primer pairs *FAMGATClargo-5* + *FAMGATClargo-3* and *seqGATC-5* + *FAMGATClargoconFAM-3*. Dam-methylated versions of the probes were prepared as described above. DNase I footprinting was performed as described elsewhere (27) with minor modifications. DNase I footprinting reactions were performed in 15 μl reaction volumes containing 1× OxyR binding buffer and 2 μM 6×His-OxyR[C199S]. The binding reaction was allowed to equilibrate at room temperature for 30 min. A total of 1 μl (0.05 units) of DNase I (Roche Farma, Barcelona, Spain) was then added, mixed gently and incubated at 37°C for 5 min. The reaction was stopped by addition of 2 μl EDTA 100 mM followed by vigorous vortexing and thermal denaturation at 95°C for 10 min. Digestion products were desalted using MicroSpin G-25 columns (GE Healthcare, Wauwatosa, WI, USA) and analyzed on an ABI 3730 DNA Analyzer along with GeneScan 500-LIZ size standards (Applied Biosystems, Foster City, CA, USA).

SMRT® sequencing

Cultures of *S. enterica* were enriched for OpvAB[ON] cells if needed (17). SMRTbell[TM] template libraries were prepared according to the instructions from Pacific Biosciences (Menlo Park, CA, USA), following the procedure and checklist for 1 kb template preparation and sequencing. Briefly, for preparation of 600 bp libraries, 4 μg of genomic DNA were sheared in microTubes using adaptive focused acoustics (Covaris, Woburn, MA, USA). Size range was monitored on an Agilent 2100 Bioanalyzer from Agilent Technologies, Santa Clara, CA, USA. DNAs were end-repaired and ligated to hairpin adapters applying components from the DNA Template Prep, Pacific Biosciences, Menlo Park, CA, USA. SMRTbell[TM] templates were exonuclease-treated for removal of incomplete reaction products. Conditions for annealing of sequencing primers and binding of polymerase to purified SMRTbell[TM]

templates were assessed with the Pacific Biosciences' Binding Calculator. Six movies were taken for both states on the PacBio *RSII* (Pacific Biosciences, Menlo Park, CA, USA) using P4-C2 chemistry at 2 h collection time. Secondly, stationary phase cultures were enriched for OpvAB^{ON} cells and libraries were prepared as given above. In this case five movies were taken using P4-C2 chemistry at 3 h collection time.

Resulting data were mapped to the complete genome sequence (GenBank accession number CP001363.1) of *S. enterica* subsp. *enterica* serovar Typhimurium strain ATCC 14028, using the BLASR algorithm (28) as implemented in Pacific Biosciences' SMRT® Portal 2.1.0 within the 'RS_Modification_and_Motif_Analysis.1' protocol applying default parameter settings. According to the setup of the experiment the secondary analysis jobs were named 'OpvAB^{OFF}' and 'OpvAB^{ON}'. Besides the global methylation pattern, the methylation status of four GATC sites upstream of the *opvAB* operon was inferred using SMRT® View, investigating the chromosomal positions 2 361 489 and 2 361 490 (GATC$_1$), 2 361 439 and 2 361 440 (GATC$_2$), 2 361 416 and 2 361 417 (GATC$_3$) and 2 361 366 and 2 361 367 (GATC$_4$). Results are shown in supplementary .csv files S1 and S2 (OpvAB-OFF basemod summary and OpvAB-ON basemod summary, respectively).

Southern blot

Genomic DNA was isolated by phenol extraction and ethanol precipitation from stationary cultures in LB (O.D.$_{600}$ ∼4). A total of 16 μg of each DNA sample were digested with HaeIII and AccI (New England Biolabs, Ipswich, MA, USA), purified and divided into four fractions, three of which were subsequently digested with DpnI, MboI or Sau3AI (New England Biolabs). After digestion the samples were run in a 2% TAE-agarose gel at 100 V for 2 h. After electrophoresis, the DNA was denatured by treatment of the gel in acid conditions (0.25 M HCl, two washes 15 min each), followed by alkalinization (0.5 M NaOH, 1.5 M NaCl) and neutralization (0.5 M Tris, 1.5 M NaCl, pH 7.5; two washes, 30 min each). The gel was then washed in SSC 10× buffer (1.5 M NaCl, 150 mM trisodium citrate, pH 7) and the DNA was transferred by vacuum to an Amersham Hybond-N+ membrane (GE Healthcare, Wauwatosa, WI, USA) using a model 785 Vacuum Blotter (Bio-Rad, Hercules, CA, USA). The DNA in the membrane was then immobilized by UV crosslinking. A radioactive probe was prepared by PCR using dCTP [α-^{32}P] (Perkin Elmer, Waltham, MA, USA) and oligonucleotides *2208mut1DIRnuevo* and *2208mut4INVnuevo* (16). After the PCR reaction, nonincorporated nucleotides were removed by treatment in a Sephadex G-25 column (illustra MicroSpin G-25 columns, GE Healthcare, Wauwatosa, WI, USA) following the instructions of the manufacturer. Prior to hybridization the double-stranded DNA probe was denatured by heating at 95°C for 3 min, followed by incubation on ice. Hybridization with the probe was performed overnight at 42°C in hybridization buffer (0.5 M sodium phosphate pH 7.2, 10 mM EDTA, 7% sodium dodecylsulphate (SDS)). Excess probe was removed with washing buffer (40 mM sodium phosphate pH 7.2, 1% SDS) at 38°C (three washes, 30 min each).

The membrane was developed using a FLA-5100 Scanner (Fujifilm, Tokyo, Japan).

RESULTS

Both the absence and the overexpression of Dam methylase increase *opvAB* expression and abolish phase variation

Genes under Dam methylation control fall into two categories. One includes genes in which methylation and nonmethylation provide opposite signals (12). An example is the *traJ* gene of the *Salmonella* virulence plasmid, which is repressed by GATC methylation (29). In this class of genes, expression of the *dam* gene from a multicopy plasmid does not alter the wild-type phenotype (29). In other genes, however, a plasmid-borne *dam* gene does alter the gene expression pattern. This phenomenon is usually an indication that Dam dependent transcriptional control is more complex, and involves the formation of Dam methylation patterns (combinations of methylated and non-methylated GATC sites) (12). To ascertain whether *opvAB* belonged to the 'simple' or the 'complex' class of Dam methylation-dependent genes, the effect of introducing a *dam* gene carried on plasmid pTP166 was assayed. The results were as follows:

(i) In a wild-type background, an *opvAB*::*lac* translational fusion showed phase variation, and formed white (OpvAB^{OFF}) and blue (OpvAB^{ON}) colonies in the presence of X-gal. In a *dam* background, phase variation was abolished, and all colonies were Lac$^+$ (OpvAB^{ON}). Plasmid pTP166 yielded an intermediate phenotype (Figure 1A), suggesting that formation of the OpvAB^{OFF} and OpvAB^{ON} subpopulations might involve the establishment of a DNA methylation pattern in the GATC sites of the *opvAB* control region, rather than methylation or non-methylation of the full set of GATC sites. A similar phenomenon occurs in the *gtr* operon (15) which is repressed in a *dam* background while introduction of a cloned *dam* gene results in an intermediate phenotype.

(ii) Expression of *opvAB*::*lac* was also monitored by β-galactosidase assays (Figure 1B). Lack of Dam methylation increased expression of the *opvAB* operon as previously described (16). Introduction of the *dam* gene carried on the pTP166 plasmid yielded an intermediate *opvAB* expression level, as in the colonies described above.

(iii) Expression of an *opvAB*::*gfp* transcriptional fusion was monitored by fluorescence analysis (Figure 1C). A major OpvAB^{OFF} subpopulation and a minor OpvAB^{ON} subpopulation were detected in the wild-type. In a *dam* background, a single population in the ON state was observed, in accordance with the results obtained with a *opvAB*::*lac* fusion. In the presence of a cloned *dam* gene (pTP166), a single population with intermediate levels of expression was detected and a shift toward the ON state remained visible (Figure 1C).

Altogether, the above observations suggested that DNA methylation patterns might be formed at the *opvAB* control region. This region, located upstream of the *opvAB* pro-
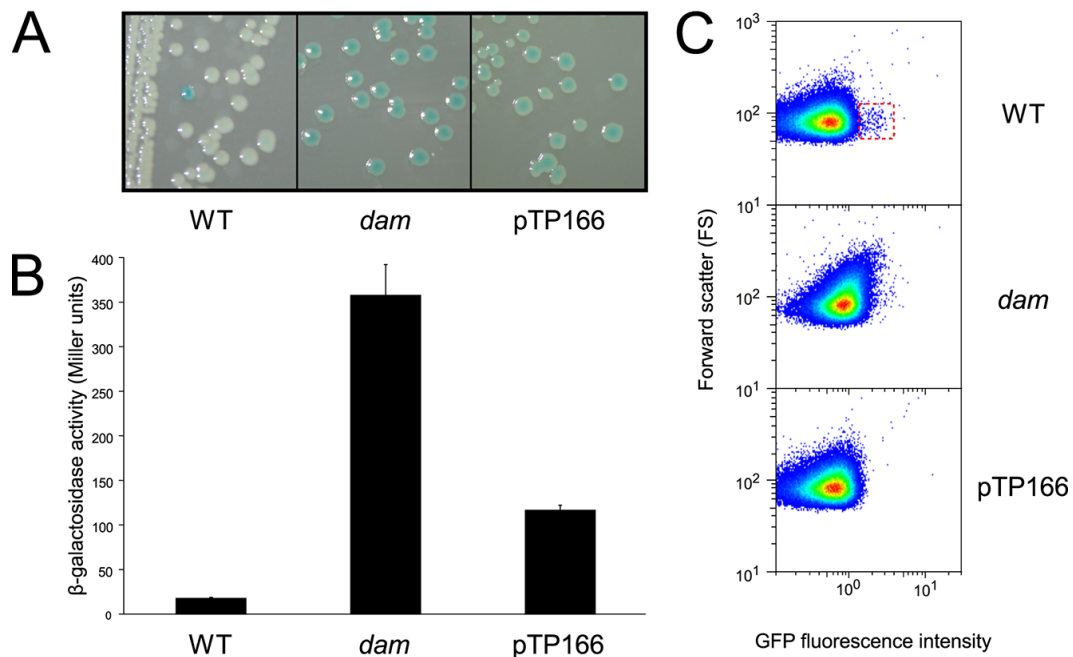
Figure 1. Regulation of *opvAB* expression by Dam methylation and formation of OpvAB subpopulations. (**A**) Visual observation of phase variation on LB + X-gal plates in *Salmonella enterica* strains carrying an *opvAB::lac* fusion in the wild-type, a *dam* mutant and a strain that overproduced Dam methylase (ATCC 14028/pTP166). (**B**) Averages and standard deviations of β-galactosidase activity of the same strains. (**C**) GFP fluorescence distribution in a strain carrying an *opvAB::gfp* fusion in the same backgrounds. Data are represented by a dot plot (forward scatter [cellular size] versus fluorescence intensity [*opvAB::gfp* expression]). All data were collected for 100 000 events per sample.

moter, contains four GATC sites separated by 46, 19 and 46 nt and centered at the -172.5, -122.5, -99.5 and -49.5 positions upstream of the transcription start site (Figure 2A and Supplementary Figure S1). From now on, these GATC sites will be referred to as $GATC_1$ to $GATC_4$, the latter being closest to the -35 module of the *opvAB* promoter (Supplementary Figure S1).

Roles of individual *opvAB* GATC sites in the formation of OpvAB^{OFF} and OpvAB^{ON} cell lineages

To study the contribution of each GATC site to *opvAB* regulation, mutations were introduced by site-directed mutagenesis. The mutations were designed to change GATC sites so that they would no longer be a substrate for Dam methylation. Because OxyR is essential for *opvAB* expression (16), alteration of consensus sequences was avoided inside putative OxyR binding sites. CATC sites were thus introduced in place of GATC sites, and every combination of mutated and non-mutated GATC sites was produced.

The effect of GATC mutations on *opvAB* expression was first analyzed by comparing the β-galactosidase activity of an *opvAB::lac* translational fusion in *dam*$^+$ and *dam* backgrounds (Figure 2B). Relevant observations were as follows:

(i) Mutation of $GATC_1$ and $GATC_3$ had a small effect on regulation by Dam methylation, although the absolute values of β-galactosidase activity were higher. Mutation of $GATC_2$ resulted in diminished regulation by Dam methylation. When $GATC_4$ was mutated, control

by Dam methylation showed an inverted pattern (expression was higher in a *dam*$^+$ background).

(ii) As a general rule, combinations of two or more mutations seemed to have an additive effect. A remarkable case was the combination of mutated $GATC_2$ and $GATC_4$ which exacerbated the inversion of regulation by Dam methylation caused by mutation of $GATC_4$ alone. It is noteworthy that mutations in $GATC_1$, $GATC_2$ and $GATC_3$ together did not abolish Dam-dependent regulation, whereas a single mutation in $GATC_4$ inverted the pattern of Dam-dependent regulation.

The overall conclusion from these experiments was that all four GATC sites are involved in Dam-dependent control of *opvAB* expression, and that the $GATC_4$ site may have an especially prominent role.

Even though disruption of OxyR binding sites had been avoided, GATC mutations affected *opvAB* expression irrespective of the presence or absence of DNA methylation, as observed in a *dam* background (Figure 2B). In the absence of Dam methylation, mutations in $GATC_1$ and $GATC_3$ increased *opvAB* expression whereas mutations in $GATC_2$ and $GATC_4$ resulted in lower *opvAB* expression. To separate such effects from those of Dam methylation itself, the β-galactosidase activity of *opvAB::lac* in a wild-type background was put in relation to the β-galactosidase activity in a *dam* background (Figure 2C). This representation leads to the interesting conclusion that mutations in $GATC_2$ and $GATC_4$ activate *opvAB* expression. Mutations in $GATC_1$ and $GATC_3$ show little effect on their own because *opvAB* expression is low in the wild-type, but they re-
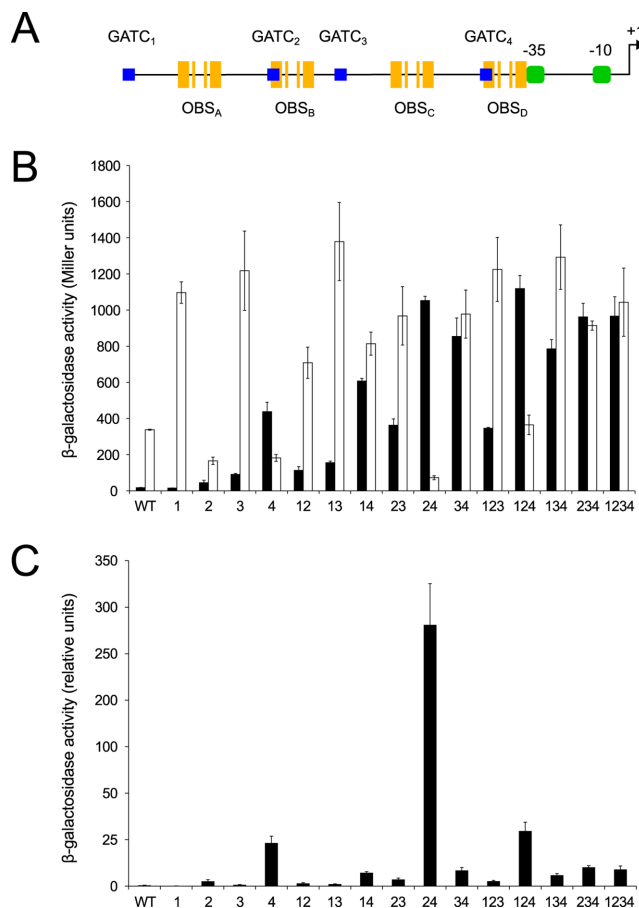
Figure 2. Effect of mutations in the *opvAB* GATC sites on *opvAB* expression. (**A**) Diagram of the *opvAB* regulatory region, with the GATC sites and the OxyR binding sites outlined. (**B**) Averages and standard deviations of β-galactosidase activity of strains carrying an *opvAB::lac* fusion in a wild-type background (black bars) and in a *dam* background (white bars). Mutated GATC sites are indicated by numbers 1–4. (**C**) Relative β-galactosidase activity of the *opvAB::lac* fusion in the same strains (activity in the wild-type divided by activity in a *dam* background).

press *opvAB* expression when combined with activating mutations in $GATC_2$ and/or $GATC_4$. Hence, the GATC sites in the *opvAB* regulatory region can be tentatively divided in two pairs: methylation of pair $GATC_1$ + $GATC_3$ seems to be associated with the $OpvAB^{ON}$ state while methylation of pair $GATC_2$ + $GATC_4$ seems to be associated with the $OpvAB^{OFF}$ state.

Analysis of fluorescence using an *opvAB::gfp* transcriptional fusion (Figure 3) allowed us to distinguish whether the differences in *opvAB* expression in GATC mutant backgrounds reflected differences in gene expression or differences in the sizes of the $OpvAB^{ON}$ and $OpvAB^{OFF}$ subpopulations. The main observations were as follows:

(i) In the wild-type, the $OpvAB^{ON}$ subpopulation comprised ∼0.18% cells.
(ii) Mutation of $GATC_4$ caused a drastic increase in the size of the $OpvAB^{ON}$ subpopulation. Mutations in $GATC_1$, $GATC_2$ and $GATC_3$ had a smaller effect, which was more clearly seen when they were combined with each other and/or with a mutation in $GATC_4$.

(iii) Two subpopulations were still distinguished when three GATC sites were mutated, provided that either $GATC_3$ or $GATC_4$ remained unaltered. The relative size of the $OpvAB^{OFF}$ and $OpvAB^{ON}$ subpopulations was however different in each case, with a predominant $OpvAB^{OFF}$ subpopulation when $GATC_4$ remained unaltered and a predominant $OpvAB^{ON}$ subpopulation when $GATC_3$ remained unaltered.
(iv) Mutation of both $GATC_3$ and $GATC_4$ eliminated subpopulation formation regardless of the presence of mutations in $GATC_1$ and $GATC_2$, and yielded an $OpvAB^{ON}$ population.

These observations are consistent with the gene expression analyses reported above, and permit to interpret the gene expression results in terms of subpopulation formation. Mutation of $GATC_4$ caused the most drastic increase in the proportion of $OpvAB^{ON}$ cells, thereby confirming that methylation of the $GATC_4$ site may have a relevant role in the formation of the $OpvAB^{OFF}$ subpopulation. Increase of $OpvAB^{ON}$ subpopulation was likewise observed when a mutated $GATC_4$ was combined with other mutated GATC sites (Figure 3).

OxyR binds the *opvAB* regulatory region

Four putative OxyR binding half-sites are found in the regulatory region of *opvAB* centered in the −148, −116, −75 and −43 positions (Supplementary Figure S1), and sharing 10, 8, 10 and 7 nt respectively with the 10-nt consensus sequence (16). The OxyR binding half-sites upstream of the *opvAB* promoter will be from now on referred to as OBS_A to OBS_D, the latter being immediately upstream of the *opvAB* −35 promoter module (Figure 2A). Assuming a helical periodicity of 10.5 bp (30), the OBS are predicted to be spaced by one, two and one helical turns, which means that all the OxyR binding sites may be on the same face of the DNA helix. The distance between OBS_A and OBS_B, and between OBS_C and OBS_D as well, is canonical for binding of the reduced form of OxyR (31). $GATC_2$ and $GATC_4$ overlap with OBS_B and OBS_D, respectively (Figure 2A).

To test whether OxyR binds the *opvAB* regulatory region, an electrophoretic mobility shift assay (EMSA) was carried out using purified OxyR protein (Figure 4A). To avoid uncontrolled oxidation of OxyR and because it was previously shown that the oxidation state of OxyR is not relevant for *opvAB* regulation (16), we used a mutant version of the OxyR protein, $OxyR^{C199S}$, which cannot be oxidized but retains the properties of the reduced form of OxyR (31,32). Purified 6×His-$OxyR^{C199S}$ protein (henceforth named OxyR for simplicity) was thus used. A DNA fragment containing the four regulatory GATC sites and the four OxyR binding half-sites was produced using a 6-FAM-labeled oligonucleotide and was incubated with increasing concentrations of OxyR. Binding was unambiguously detected. A DNA fragment from the regulatory region of an unrelated gene (*envR*) was used as a negative control, and binding was not detected (Figure 4A).
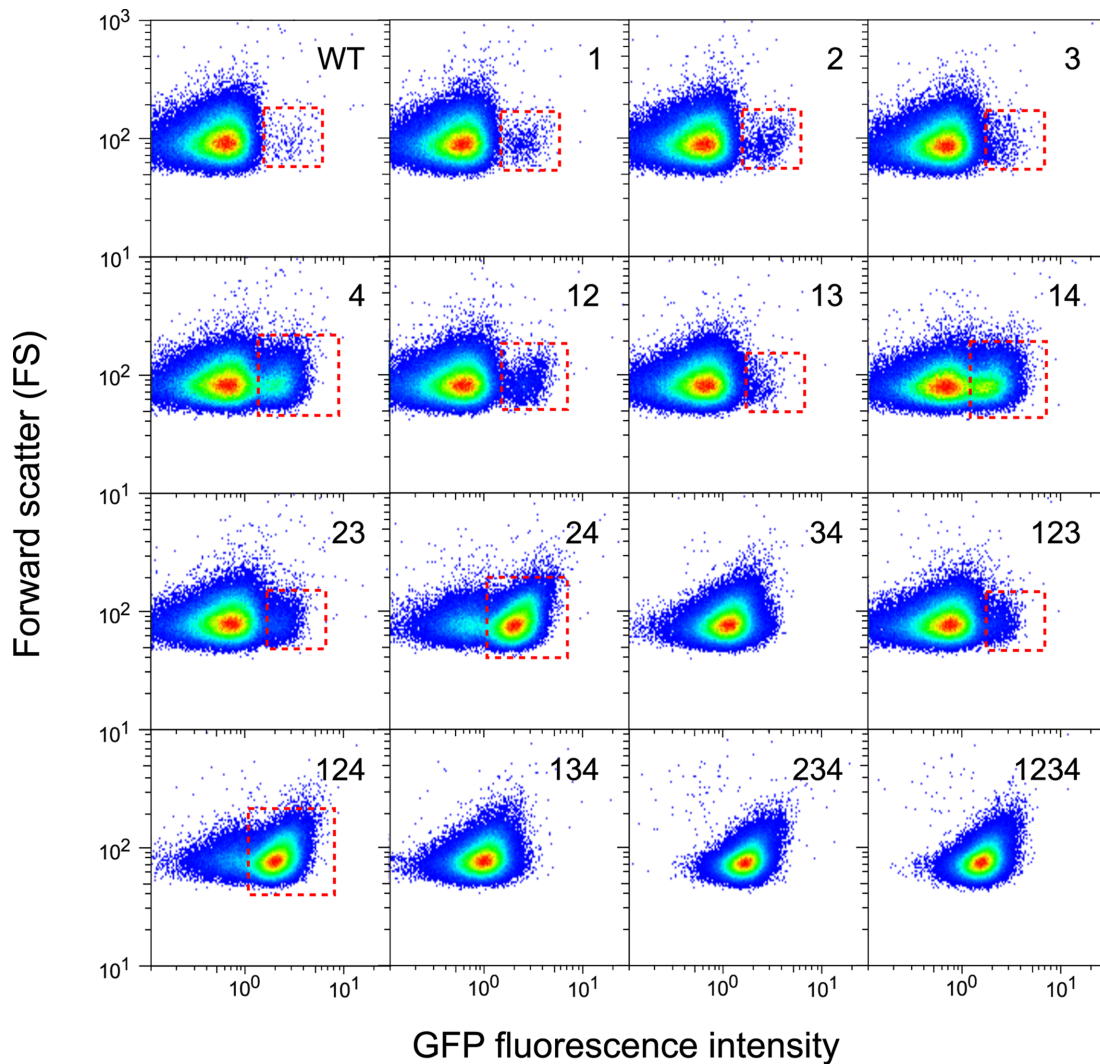
Figure 3. GFP fluorescence distribution in *Salmonella enterica* strains carrying an *opvAB::gfp* fusion and mutations in the *opvAB* GATC sites. Mutated GATC sites are indicated by numbers 1–4. Data are represented by a dot plot, and were collected for 100 000 events per sample.

OxyR protects the *opvAB* regulatory region

To define the binding pattern of OxyR to the *opvAB* regulatory region, purified OxyR was used in a footprinting assay performed using 6-FAM-labeled DNA fragments and DNase I (Figure 4B). The same DNA fragment used in the EMSA assays, containing both the GATC sites and predicted OxyR binding sites, was labeled at the alternate ends and used in parallel experiments. Methylated and non-methylated DNA probes were used, as well as a probe in which GATC sites 1–4 had been converted to CATC sites by site-directed mutagenesis. The analysis confirmed the ability of OxyR to bind the *opvAB* regulatory region *in vitro* (Figure 4 and Supplementary Figure S2). Relevant observations were as follows:

(i) Protection from DNase I digestion was detected in a 133 bp DNA span, albeit with regional differences. GATC$_1$, GATC$_2$, GATC$_3$ are located in the protected region. Fragment-specific binding patterns were de-

tected and the overall protection was less efficient when the DNA probe was either methylated or GATC-less.

(ii) The OBS$_A$ and OBS$_C$ sites were fully protected, while OBS$_B$ was partially protected.

(iii) OBS$_D$, which contains the GATC$_4$ site, was not protected.

The relevance of these observations may be limited as methylated and non-methylated DNA probes were used, and evidence presented above had suggested that *opvAB* regulation involved both methylated and non-methylated GATC sites (Figure 1). With this caveat, footprinting experiments confirmed the ability of OxyR to bind the *opvAB* regulatory region and defined the DNA region protected by OxyR binding. An additional, interesting observation was that OxyR protection extended outside the OxyR binding sites, as previously described for other LysR-type factors (33–36) (see below).
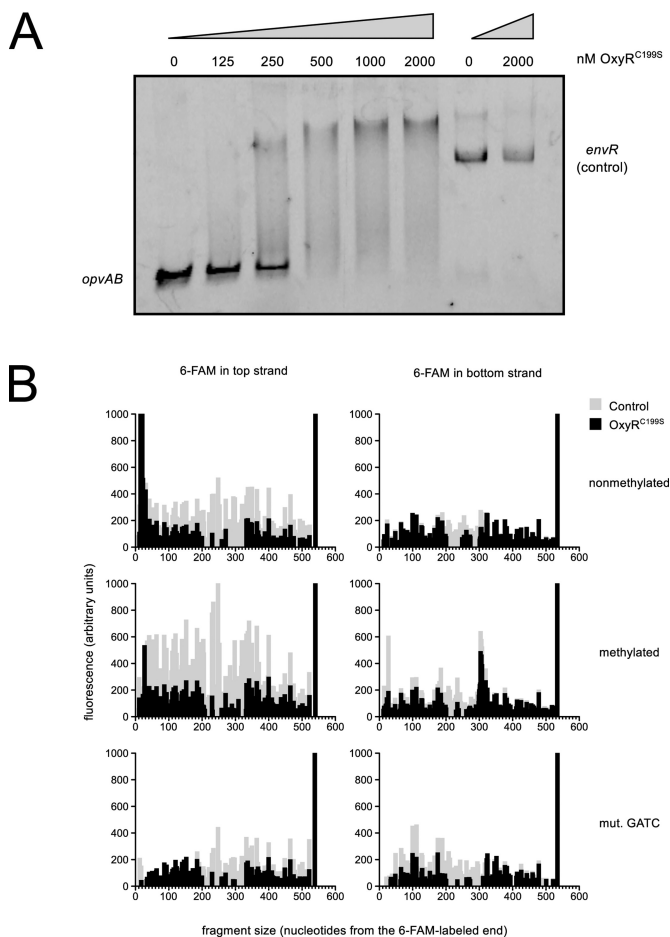
A

B

Figure 4. Binding of 6xHis-OxyR^{C199S} to the *opvAB* promoter region. (**A**) Electrophoretic mobility shift assay of 6xHis-OxyR^{C199S} binding to a DNA fragment containing the *opvAB* promoter and the upstream regulatory region. The regulatory region of *envR* was used as a negative control. (**B**) DNase I footprinting of 6xHis-OxyR^{C199S} binding to DNA fragments containing the *opvAB* promoter and regulatory region with a 6-FAM label in either the top or the bottom strand. Methylated, non-methylated and GATC-less versions of the fragment were used.

OpvAB^{OFF} and OpvAB^{ON} subpopulations are characterized by inverse patterns of Dam methylation

Single-molecule real-time (SMRT®) sequencing results showed that >97% of the total of 38 458 GATC sites present in the genome of *S. enterica* serovar Typhimurium are methylated, and that non-methylated sites are the exception. Within this set, several non-methylated GATC sites were detected upstream of the *opvAB* operon. In order to analyze them in more detail, position-specific base modification analyses were performed. Addition of the virulent P22 H5 phage to a culture of *S. enterica* results in selection of the OpvAB^{ON} subpopulation (17). Using this procedure, a culture was enriched in OpvAB^{ON} cells and the methylation state of the *opvAB* GATC sites was analyzed using SMRT® sequencing (37). An ordinary culture, which contains >99% OpvAB^{OFF} cells (16,17), was also subjected to SMRT® sequencing. A total of 246 373 (430 408) polymerase reads with a mean polymerase read length of 10 010 (8516) bp and mean sequence coverage of 178× (329×)

were obtained for the OpvAB^{ON} (OpvAB^{OFF}) SMRT sequencing. The results from position-specific base modification analysis are shown in supplementary .csv files S1 and S2, and can be summarized as follows:

(i) In an ordinary OpvAB^{OFF} culture, GATC$_1$ and GATC$_3$ were non-methylated, whereas GATC$_2$ and GATC$_4$ were methylated (Table 1).
(ii) In the OpvAB^{ON} culture, an inverse DNA methylation pattern was found: non-methylation of GATC$_2$ and GATC$_4$ and methylation of GATC$_1$ and GATC$_3$ (Table 1).

These observations confirm that establishment of the OFF and ON states of the *opvAB* locus involves the formation of DNA methylation patterns, as in other phase variation loci under Dam methylation control (10,13,15).

OxyR protects GATC sites from Dam methylation *in vivo*

OxyR has been previously described as a DNA methylation-blocking factor, able to induce the formation of non-methylated GATC sites (15,38). To test whether OxyR has a similar DNA methylation-blocking ability in the *opvAB* operon, the methylation state of the GATC sites in the *opvAB* regulatory region was tested *in vivo*. For this purpose, a Southern blot was performed using genomic DNA extracted from the wild-type strain and from an *oxyR* mutant. The methylation state of individual GATC sites was inferred from restriction analysis using enzymes that cut GATC sequences depending on their methylation state (MboI, DpnI and Sau3AI). GATC$_1$ and GATC$_3$ were found to be non-methylated while GATC$_2$ and GATC$_4$ were found to be methylated in the wild-type strain (Figure 5). In contrast, in an *oxyR* background, all four GATC sites were found to be methylated (Figure 5). These observations confirmed that OxyR has DNA methylation-blocking ability *in vivo* at the *opvAB* regulatory region.

Mutations in the OBS$_B$ and OBS$_D$ OxyR binding sites abolish phase variation

Of the four OxyR binding half-sites in the *opvAB* regulatory region, OBS$_A$ and OBS$_C$ are an absolute match (10 out of 10 nt) to the consensus sequences defined for OxyR binding (31). In contrast, OBS$_B$ and OBS$_D$ share only 8 and 7 out of 10 nt with the consensus sequence, respectively. The fact that *opvAB* phase variation is skewed toward the OFF state led us to hypothesize that the degree of OxyR binding site perfection played a role in such bias. To test our hypothesis, 1 nt change was introduced in OBS$_B$ and two nucleotide changes in OBS$_D$ so that their mutated versions would share 9 out of 10 nt with the consensus sequence. Construction of a perfect consensus sequence was avoided since it would inevitably destroy GATC$_2$ and GATC$_4$.

The consequences of OBS$_B$ and OBS$_D$ DNA sequence amelioration were analyzed using *opvAB::gfp* (Figure 6A) and *opvAB::lac* fusions (Figure 6B). Mutations in either OBS$_B$ or OBS$_D$ abolished *opvAB* phase variation, yielding a uniform OpvAB^{ON} population. In the case of OBS$_B$, a single nucleotide change led also to full expression of the

Table 1. DNA modification status according to SMRT® View for position specific base-modification analysis upstream of the *opvAB* operon[a]

Site	Genome position	OpvAB[OFF]	OpvAB[ON]
GATC$_1$	2 361 489+	unmodified (1.35, 31)	m6A (2.76, 59)
	2 361 490−	unmodified (1.22, 31)	m6A (3.99, 46)
GATC$_2$	2 361 439+	m6A (4.55, 55)	unmodified (0.93, 49)
	2 361 440−	m6A (2.85, 54)	unmodified (0.85, 37)
GATC$_3$	2 361 416+	unmodified (0.78, 55)	m6A (2.29, 45)
	2 361 417−	unmodified (0.43, 55)	m6A (2.15, 36)
GATC$_4$	2 361 366+	m6A (2.79, 45)	unmodified (1.02, 52)
	2 361 367−	m6A (3.14, 37)	unmodified (0.59, 47)

[a]Inter pulse duration ratios as well as strand-specific coverage values are given in parentheses.
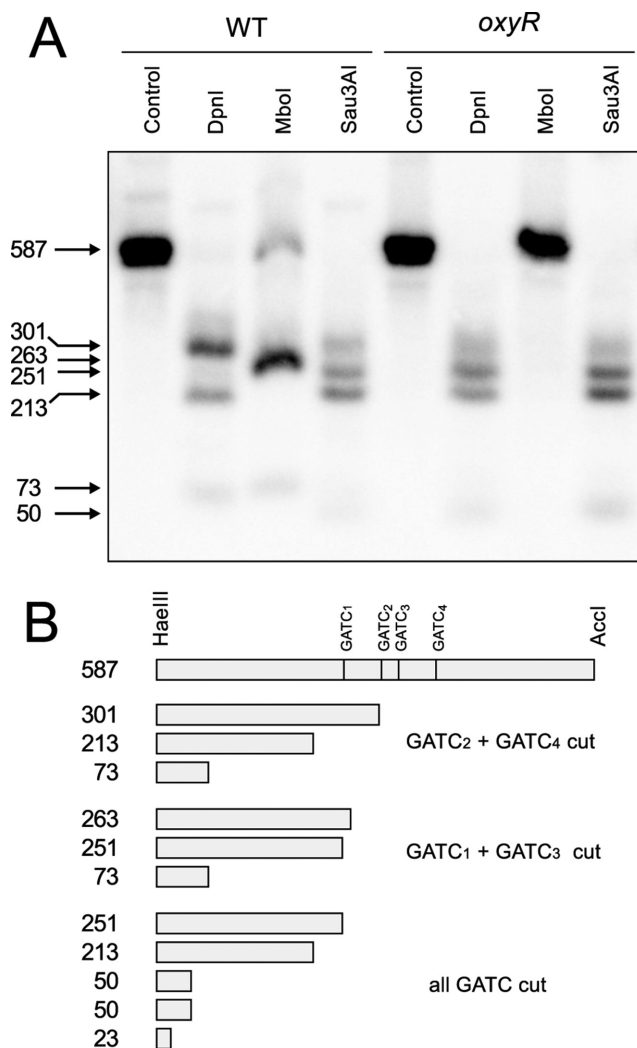
Figure 5. Methylation state of GATC sites in the *opvAB* regulatory region in wild-type and *oxyR* backgrounds. (**A**) Southern blot of genomic DNA obtained from wild-type and *oxyR* cultures and digested with HaeIII and with AccI (control) and DpnI, MboI or Sau3AI. Fragment sizes are indicated in base pairs. (**B**) Diagram of the HaeIII-AccI fragment and pattern of fragments obtained.

operon. The mutation in OBS$_D$ caused a smaller increase in expression and was epistatic to the mutation in OBS$_B$.

An interpretation of these observations is that OBS$_B$ and OBS$_D$ DNA sequence amelioration may 'trap' OxyR in the OpvAB[ON] configuration. In support of this view, absence of Dam methylation had no effect on *opvAB* expression in these mutant backgrounds (Figure 6). Hence, the preference of OxyR for certain OxyR-binding sites may be a key factor in regulation of *opvAB* phase variation, and alternative binding of OxyR upstream of the *opvAB* promoter may generate the OpvAB[OFF] and OpvAB[ON] subpopulations.

SeqA contributes to the small size of the OpvAB[ON] subpopulation

SeqA was considered a potential ancillary candidate for regulation of *opvAB* since it binds GATC sites (39) and is involved in regulation of other phase variation loci (40,41). Thus we analyzed the effect of a *seqA* mutation on *opvAB* expression and its influence on the formation of OpvAB subpopulations. A strain carrying a *seqA* null allele and an *opvAB::lac* fusion formed darker (Lac+) colonies on LB + X-gal than the wild-type, and displayed frequent sectoring. Nonetheless, two groups of differently colored colonies (light blue and dark blue) were still distinguishable (Figure 7A), which allowed calculation of phase transition frequencies. The OFF→ON transition rate was found to be 50-fold higher in a *seqA* background (3.0×10^{-3} compared with 6.1×10^{-5} in the wild-type), whereas the ON→OFF transition rates were similar (3.1×10^{-2} compared to 3.7×10^{-2} in the wild-type). Not surprisingly, the β-galactosidase activity of an *opvAB::lac* fusion was ∼10-fold higher in a *seqA* background (Figure 7B).

Fluorescence assays showed that mutation of *seqA* caused an increase in the size of the OpvAB[ON] subpopulation (Figure 7C). The effect was stronger in the presence of mutations in GATC$_1$ and/or GATC$_2$, and to a lesser extent in GATC$_3$ (Supplementary Figure S1). Interestingly, when GATC$_4$ was mutated, a mutation in *seqA* had an effect opposite to that observed in the wild-type: the OpvAB[ON] subpopulation was reduced (Supplementary Figure S1). When both GATC$_3$ and GATC$_4$ were mutated, the *seqA* mutation did not have a significant effect (Supplementary Figure S1). These results seem to indicate that the main role of SeqA in the regulation of *opvAB* is the maintenance of a
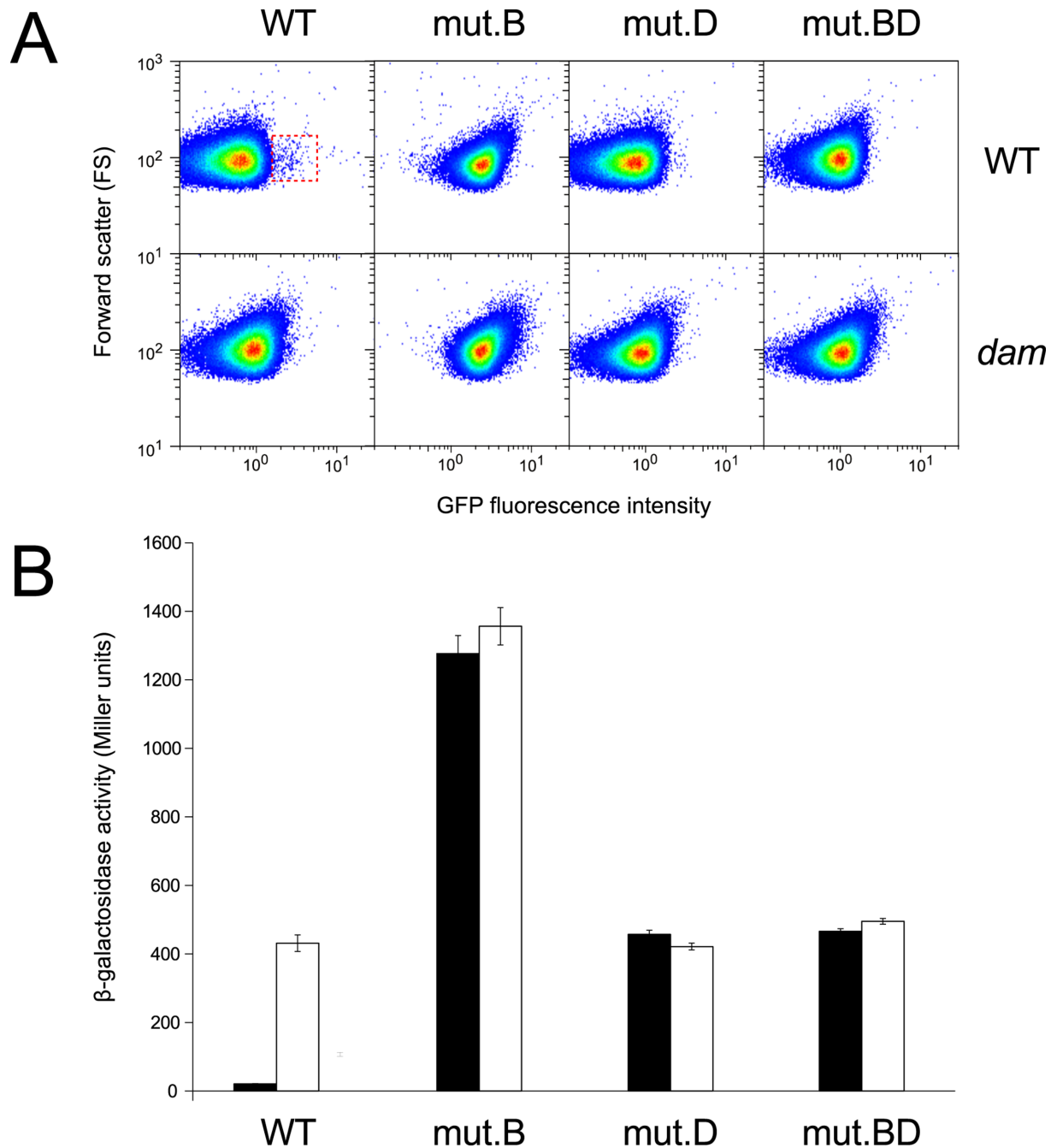
Figure 6. Effect of mutations in OBS_B and OBS_D on the expression of *opvAB*. (**A**) GFP fluorescence distribution in *Salmonella enterica* strains carrying an *opvAB::gfp* fusion and mutations in OBS_B (mut.B) and/or OBS_D (mut.D) in wild-type and *dam* backgrounds. Data are represented by a dot plot, and were collected for 100 000 events per sample. (**B**) Averages and standard deviations of β-galactosidase activity of *S. enterica* strains carrying an *opvAB::lac* fusion with mutations in OBS_B (mut.B) and/or OBS_D (mut.D) in wild-type (black bars) and *dam* (white bars) backgrounds.

low OFF→ON transition rate (in other words, repression of OpvAB^{ON} subpopulation formation).

HU is essential for the formation of the OpvAB^{ON} subpopulation

HU is a nucleoid-associated protein known to regulate a large number of genes in *E. coli* and *Salmonella* (42–44). The HU protein can exist in three forms: the HU αβ heterodimer and the corresponding homodimers. The het-

erodimer is the predominant form *in vivo* (45). We deleted *hupA* and/or *hupB*, the genes encoding the two proteins forming the HU heterodimer and tested the effect of the mutations on the expression of an *opvAB::gfp* fusion (Figure 8A). The OpvAB^{ON} subpopulation was found to be reduced from ~0.18% in the wild-type to 0.09% in single *hupA* and *hupB* mutants. Reduction of the OpvAB^{ON} subpopulation size was exacerbated in the double *hupA hupB* mutant:
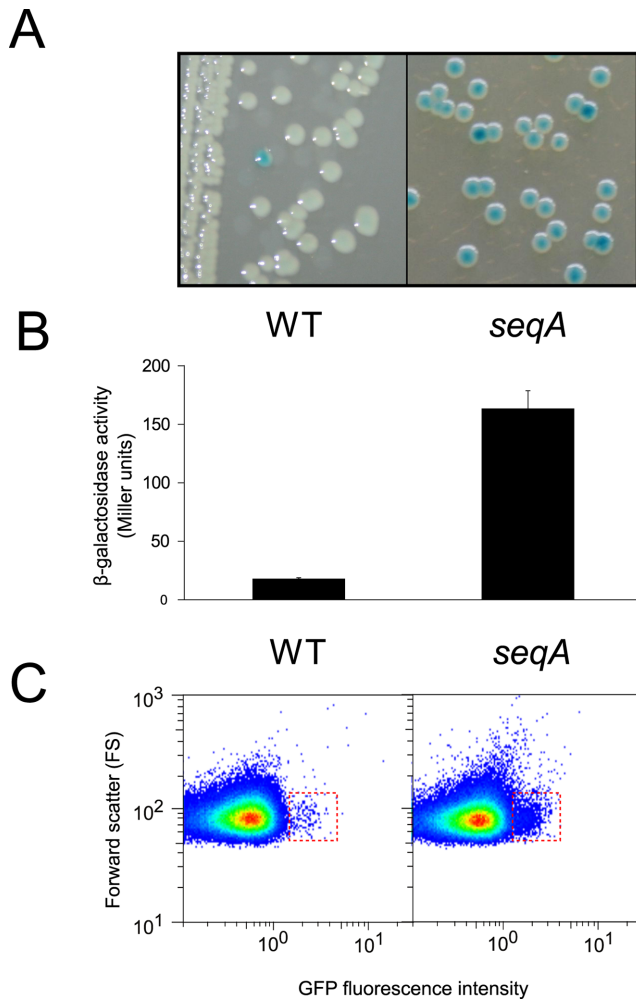
A

WT   *seqA*

B

WT   *seqA*

C

Figure 7. Role of SeqA in *opvAB* expression and in the formation of the OpvAB[OFF] and OpvAB[ON] subpopulations. (**A**) Colonies formed by *Salmonella enterica* strains carrying an *opvAB::lac* fusion in a wild-type background and in a *seqA* background. (**B**) Averages and standard deviations of β-galactosidase activity of *S. enterica* strains carrying an *opvAB::lac* fusion in a wild-type background and in a *seqA* background. (**C**) GFP fluorescence distribution in *S. enterica* strains carrying an *opvAB::gfp* fusion in a wild-type background and in a *seqA* background. Data are represented by a dot plot, and were collected for 100 000 events per sample.

the OpvAB[ON] subpopulation was virtually absent (Figure 8A).

When *hupA* and *hupB* mutations were introduced into an *opvAB::lac* background, a decrease in the β-galactosidase activity of the *opvAB::lac* fusion was likewise found (Figure 8B). In turn, when formation of Lac⁺ (OpvAB[ON]) colonies was scored on LB + X-gal plates, blue (Lac⁺) colonies were still visible in the *hupA* and *hupB* single mutants but not in the double mutant *hupA hupB* background (Figure 8C).

When the effect of the *hupA* and *hupB* mutations was tested in OpvAB[ON]-locked backgrounds, the population remained in the OpvAB[ON] state (Figure 8D), although *opvAB::lac* expression was slightly lower (Figure 8E). Hence, HU seems to be necessary for maintenance of the

OpvAB[ON] state in the wild-type but not in mutants locked in OpvAB[ON] state.

DISCUSSION

The *S. enterica opvAB* operon encodes membrane proteins that alter O-antigen chain length in the lipopolysaccharide (16). Expression of the *opvAB* operon is subject to phase variation, with switching frequencies of 6.1×10^{-5} (OFF→ON) and 3.7×10^{-2} (ON→OFF) per cell and generation in LB medium (16). As a consequence of these disparate switching levels, *S. enterica* populations (e.g. batch cultures) contain a major OpvAB[OFF] subpopulation (>99% cells) and a minor OpvAB[ON] subpopulation (<1% cells).

The regulatory region upstream of the *opvAB* promoter, depicted in Figures 2 and 9, and Supplementary Figure S1, contains four half-sites for binding of OxyR (OBS[A-D]), and 4 methylatable GATC motifs (GATC[1-4]). OxyR is a LysR-type transcriptional regulator that also acts as a sensor of oxidative stress. Although OxyR was first described as an activator of genes responsive to oxidative damage (46), its function in *opvAB* regulation is unrelated to oxidative damage and independent of its own oxidation state (16). The same is true for other OxyR-dependent phase variation systems such as *agn43* (47) and *gtr* (15). OxyR binds DNA as a tetramer (30).

SMRT® sequencing data show that *S. enterica* OpvAB[OFF] and OpvAB[ON] subpopulations differ in their pattern of Dam methylation at the *opvAB* regulatory region (Table 1). The patterns found are actually opposite: in the OpvAB[OFF] state, GATC[1] and GATC[3] are non-methylated, whereas GATC[2] and GATC[4] are methylated; in the OpvAB[ON] state, GATC[2] and GATC[4] are non-methylated, whereas GATC[1] and GATC[3] are methylated. Combinations of methylated and non-methylated GATC sites have been previously described in other phase variation loci including *pap* and *gtr* (10,15). In these loci, GATC non-methylation is the consequence of DNA methylation hindrance upon protein binding. In an analogous fashion, we provide evidence that DNA methylation patterns at the *opvAB* regulatory region are generated by OxyR binding (Figure 5).

OxyR has been shown to bind alternative pairs of half-sites in *gtr* (15), and *opvAB* may constitute another example of the same phenomenon albeit with a different genomic architecture. In *gtr*, the sites bound by OxyR in the OFF and ON lineages have identical number of nucleotides in common with the consensus sequence (15), which may explain why the *gtr* locus has similar ON→OFF and OFF→ON transition rates. In contast, the OBS[A] and OBS[C] sites of *opvAB* are identical to the consensus sequence for OxyR binding while OBS[B] and OBS[D] share only 8/10 and 7/10 nt with the consensus, respectively. This difference may explain the higher stability of the OpvAB[OFF] lineage, which results in a ~600-fold difference in the ON→OFF and OFF→ON transition rates. The relevance of the nucleotide sequence of OxyR binding sites for *opvAB* regulation is illustrated by the observation that single nucleotide changes in OBS[B] and OBS[D] lock the system in the ON state (Figure 6). The lower increase in *opvAB* expression caused by a mutation in OBS[D] (Figure 6) may be tentatively explained as a con-
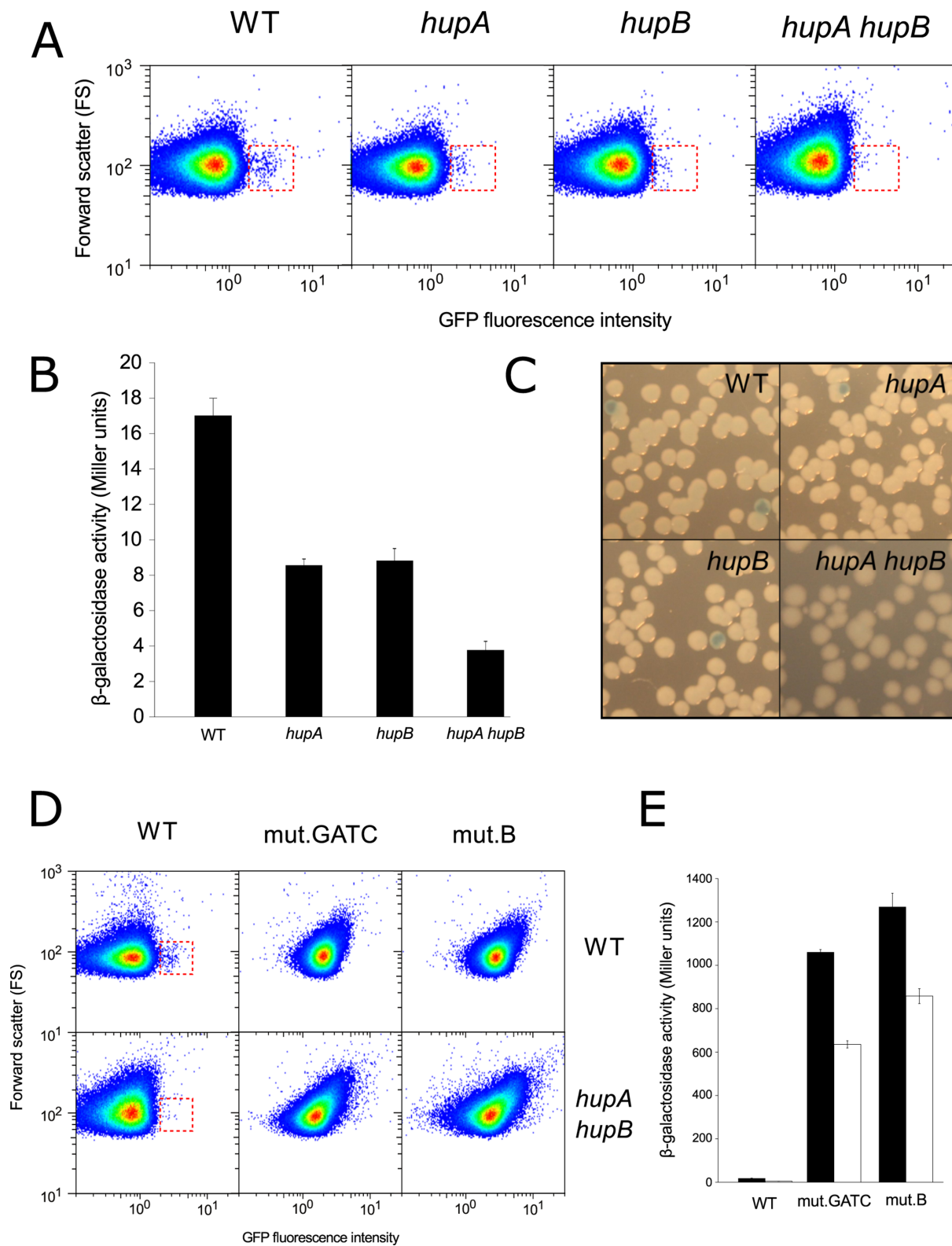
Figure 8. Role of HU in *opvAB* expression and in the formation of the OpvAB^OFF and OpvAB^ON subpopulations. (**A**) Dot plots of GFP fluorescence distribution in *Salmonella enterica* strains carrying an *opvAB::gfp* fusion in a wild-type background and in the absence of genes *hupA* and/or *hupB*. (**B**) Averages and standard deviations of β-galactosidase activity of *S. enterica* strains carrying an *opvAB::lac* fusion in a wild-type background and in the absence of genes *hupA* and/or *hupB*. (**C**) Visual observation of phase variation on LB + X-gal plates in strains carrying an *opvAB::lac* fusion a wild-type background and in the absence of genes *hupA* and/or *hupB*. (**D**) Dot plots of GFP fluorescence distribution in *S. enterica* strains carrying an *opvAB::gfp* fusion and a *hupA hupB* mutation in OpvAB^ON-locked backgrounds. (**E**) Averages and standard deviations of β-galactosidase activity of *S. enterica* strains carrying an *opvAB::lac* fusion in the wild-type (black bars) and in a *hupA hupB* background (white bars).
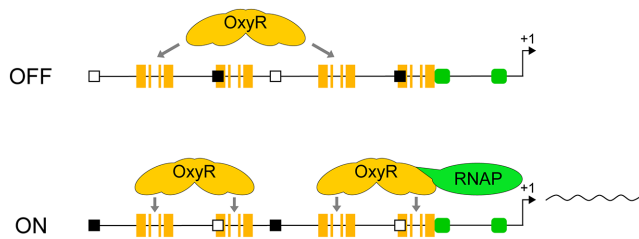
Figure 9. Model of *opvAB* phase variation. The diagram shows the Dam methylation states found in OpvAB^{OFF} and OpvAB^{ON} cell lineages and the hypothetical patterns of OxyR binding to cognate sites. Black and white squares represent methylated and nonmethylated GATC sites, respectively.

squence of the OBS$_D$ location immediately upstream of the −35 module: a mutation of OBS$_D$ may impair the interaction between OxyR and the RNA polymerase. In agreement with this view, it has been proposed that RNA polymerase may contact OxyR and other LysR-type transcription factors within the DNA region occupied by the regulator (48). The fact that the mutation in OBS$_D$ is epistatic over the mutation in OBS$_B$ (Figure 6) may support this interpretation.

Preferential methylation of GATC$_4$ may be an additional factor contributing to the stability of the OpvAB^{OFF} lineage. The DNA sequences that flank GATC$_1$, GATC$_2$ and GATC$_3$ are predicted to be relatively poor Dam methylation substrates compared with the flanking sequences of GATC$_4$ (49). Rapid methylation of GATC$_4$ may thus contribute to perpetuation of the OpvAB^{OFF} state.

Our tentative model, based on a combination of experimental data, information from the literature and some speculation as well, proposes that the predominant OFF state involves binding of OxyR to the OBS$_A$ and OBS$_C$ sites, which protects GATC$_1$ and GATC$_3$ from methylation (Table 1 and Figure 5). In this configuration, GATC$_2$ and GATC$_4$ are unprotected and therefore are methylated by Dam. A caveat to this model is that, to our knowledge, OxyR has not been described to bind non-consecutive half-sites. However, such binding pattern is consistent with two lines of evidence: (i) only OBS$_A$ and OBS$_C$ are fully protected in the footprinting assay (Figure 4B); (ii) OxyR has been hitherto described to bind DNA as a tetramer (31,50). An alternative hypothesis is that OxyR dimers may bind independently the OBS$_A$ and OBS$_C$ sites.

DNA bending, which is commonly induced by OxyR (31), specifically by the reduced tetramer structure (51) and by other LysR-type regulators (52–55), may contribute to methylation hindrance in GATC$_1$ and GATC$_3$. A DNA bend is induced by OxyR in *agn43* (47,56), another phase variation locus regulated by Dam methylation and OxyR. The occurrence of bending might thus help to understand why GATC$_1$ and GATC$_3$ are protected from methylation in the OpvAB^{OFF} configuration (Figure 5) despite their location outside OBS$_A$ and OBS$_c$ (Figure 9 and Supplementary Figure S1).

Another factor that might contribute to methylation hindrance in GATC$_1$ and GATC$_3$ in the OpvA^{OFF} lineage might be DNA wrapping, which has been proposed for other transcriptional regulators whose footprints extend outside the binding sites. Examples include the NtrC (33),

RcnR (34) and NorR (35,36) transcription factors from *E. coli*. CarP, also called PepA, an *E. coli* transcription factor, specifically prevents methylation of a GATC site which is not included in the binding footprint (57). GATC$_1$ and GATC$_3$, which lie in the extended OxyR-bound region, may be protected from Dam methylation in an analogous fashion.

In the ON state, OxyR binds to the OBS$_B$ and OBS$_D$ sites. As a consequence, GATC$_2$ and GATC$_4$ are protected from methylation and remain non-methylated, whereas GATC$_1$ and GATC$_3$ are unprotected and are methylated (Table 1). In this configuration, RNA polymerase is successfully recruited to the *opvAB* promoter and transcription of *opvAB* takes place. OxyR has been shown to recruit RNA polymerase by direct contact with the C-terminal domain of the α subunit (58,50), and the inverse is also true: RNA polymerase can recruit OxyR (50), which might contribute to maintenance of the OpvAB^{ON} state.

Additional factors involved in the formation of OpvAB cell lineages are the GATC-binding protein SeqA and the nucleoid protein HU. SeqA contributes to the stability of the OpvAB^{OFF} lineage, acting as a repressor of the OFF→ON transition (Figure 7). SeqA action seems to be exerted mostly on GATC$_3$ and GATC$_4$ (Supplementary Figure S1). Because SeqA binds hemimethylated GATC sites (59), a tentative speculation is that it might favor DNA methylation over OxyR binding during DNA replication, as previously suggested for *agn43* (40). In turn, HU contributes to formation of the OpvAB^{ON} lineage (Figure 8). Tentative interpretations may be that HU contributes to the establishment of the OpvAB^{ON} state either by inducing DNA bending or by stabilizing OxyR-mediated bending. The latter possibility may be more likely as HU often stabilizes bent DNA rather than bending DNA itself (60), and HU is not essential in OpvAB^{ON}-locked backgrounds (Figure 8). On the other hand, AT-rich DNA, such as that found in the *opvAB* regulatory region (which is 23% G + C only) is intrinsically prone to DNA bending (61,62).

In the model depicted in Figure 9, two OxyR tetramers are required to maintain the ON state but only one tetramer is necessary to maintain the OFF state (Figure 9). If true, this difference may be an additional factor to explain the high ON→OFF transition rate (together with the OxyR binding site differences and the preferential methylation of GATC$_4$). Upon passage of the DNA replication fork, the local concentration of OxyR will be halved, therefore facilitating the transition from ON (depending on two OxyR tetramers) to OFF (depending on one tetramer only).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Kussell,E. and Leibler,S. (2005) Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, **309**, 2075–2078.
2. Dhar,N. and McKinney,J.D. (2007) Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr. Opin. Microbiol.*, **10**, 30–38.
3. Zgur-Bertok,D. (2007) Phenotypic heterogeneity in bacterial populations. *Acta Agric. Slovenica*, **90**, 17–24.
4. Dubnau,D. and Losick,R. (2006) Bistability in bacteria. *Mol. Microbiol.*, **61**, 564–572.
5. Veening,J.W., Smits,W.K. and Kuipers,O.P. (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, **62**, 193–210.
6. van der Woude,M.W. and Baumler,A.J. (2004) Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.*, **17**, 581–611.
7. van der Woude,M.W. (2011) Phase variation: how to create and coordinate population diversity. *Curr. Opin. Microbiol.*, **14**, 205–211.
8. Silverman,M., Zieg,J., Hilmen,M. and Simon,M. (1979) Phase variation in *Salmonella*: genetic analysis of a recombinational switch. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 391–395.
9. Moxon,R., Bayliss,C. and Hood,D. (2006) Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.*, **40**, 307–333.
10. van der Woude,M., Braaten,B. and Low,D. (1996) Epigenetic phase variation of the *pap* operon in *Escherichia coli*. *Trends Microbiol.*, **4**, 5–9.
11. Low,D.A. and Casadesus,J. (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol.*, **11**, 106–112.
12. Casadesus,J. and Low,D.A. (2013) Programmed heterogeneity: epigenetic mechanisms in bacteria. *J. Biol. Chem.*, **288**, 13929–13935.
13. Hernday,A., Krabbe,M., Braaten,B. and Low,D. (2002) Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16470–16476.
14. Henderson,I.R. and Owen,P. (1999) The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and OxyR. *J. Bacteriol.*, **181**, 2132–2141.
15. Broadbent,S.E., Davies,M.R. and van der Woude,M.W. (2010) Phase variation controls expression of *Salmonella* lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. *Mol. Microbiol.*, **77**, 337–353.
16. Cota,I., Blanc-Potard,A.B. and Casadesus,J. (2012) *STM2209-STM2208 (opvAB)*: a phase variation locus of *Salmonella enterica* involved in control of O-antigen chain length. *PLoS One*, **7**, e36863.
17. Cota,I., Sánchez-Romero,M.A., Hernández,S.B., Pucciarelli,M.G., García-del Portillo,F. and Casadesus,J. (2015) Epigenetic control of *Salmonella enterica* O-antigen chain length: a tradeoff between virulence and bacteriophage resistance. *PLoS Genet.*, **11**, e1005667.

18. Marinus,M.G., Poteete,A. and Arraj,J.A. (1984) Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. *Gene*, **28**, 123–125.
19. Chan,R.K., Botstein,D., Watanabe,T. and Ogata,Y. (1972) Specialized transduction of tetracycline resistance by phage P22 in *Salmonella typhimurium*. II. Properties of a high-frequency-transducing lysate. *Virology*, **50**, 883–898.
20. Torreblanca,J., Marques,S. and Casadesus,J. (1999) Synthesis of FinP RNA by plasmids F and pSLT is regulated by DNA adenine methylation. *Genetics*, **152**, 31–45.
21. Schmieger,H. (1972) Phage P22-mutants with increased or decreased transduction abilities. *Mol. Gen. Genet.*, **119**, 75–88.
22. Garzon,A., Cano,D.A. and Casadesus,J. (1995) Role of Erf recombinase in P22-mediated plasmid transduction. *Genetics*, **140**, 427–434.
23. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.
24. Miller,J.H. (1972) *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, NY.
25. Eisenstein,B.I. (1981) Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. *Science*, **214**, 337–339.
26. Dillon,S.C., Espinosa,E., Hokamp,K., Ussery,D.W., Casadesus,J. and Dorman,C.J. (2012) LeuO is a global regulator of gene expression in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.*, **85**, 1072–1089.
27. Cameron,A.D. and Dorman,C.J. (2012) A fundamental regulatory mechanism operating through OmpR and DNA topology controls expression of *Salmonella* pathogenicity islands SPI-1 and SPI-2. *PLoS Genet.*, **8**, e1002615.
28. Chaisson,M.J. and Tesler,G. (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*, **13**, 238.
29. Camacho,E.M. and Casadesus,J. (2002) Conjugal transfer of the virulence plasmid of *Salmonella enterica* is regulated by the leucine-responsive regulatory protein and DNA adenine methylation. *Mol. Microbiol.*, **44**, 1589–1598.
30. Wang,J.C. (1979) Helical repeat of DNA in solution. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 200–203.
31. Toledano,M.B., Kullik,I., Trinh,F., Baird,P.T., Schneider,T.D. and Storz,G. (1994) Redox-dependent shift of OxyR-DNA contacts along an extended DNA-binding site: a mechanism for differential promoter selection. *Cell*, **78**, 897–909.
32. Kullik,I., Toledano,M.B., Tartaglia,L.A. and Storz,G. (1995) Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for oxidation and transcriptional activation. *J. Bacteriol.*, **177**, 1275–1284.
33. Lilja,A.E., Jenssen,J.R. and Kahn,J.D. (2004) Geometric and dynamic requirements for DNA looping, wrapping and unwrapping in the activation of *E.coli glnAp2* transcription by NtrC. *J. Mol. Biol.*, **342**, 467–478.
34. Iwig,J.S. and Chivers,P.T. (2009) DNA recognition and wrapping by *Escherichia coli* RcnR. *J. Mol. Biol.*, **393**, 514–526.
35. Tucker,N.P., Ghosh,T., Bush,M., Zhang,X. and Dixon,R. (2010) Essential roles of three enhancer sites in sigma54-dependent transcription by the nitric oxide sensing regulatory protein NorR. *Nucleic Acids Res.*, **38**, 1182–1194.
36. Bush,M., Ghosh,T., Tucker,N., Zhang,X. and Dixon,R. (2011) Transcriptional regulation by the dedicated nitric oxide sensor, NorR: a route towards NO detoxification. *Biochem. Soc. Trans.*, **39**, 289–293.
37. Flusberg,B.A., Webster,D.R., Lee,J.H., Travers,K.J., Olivares,E.C., Clark,T.A., Korlach,J. and Turner,S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods*, **7**, 461–465.
38. Haagmans,W. and van der Woude,M. (2000) Phase variation of Ag43 in *Escherichia coli*: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. *Mol. Microbiol.*, **35**, 877–887.
39. Waldminghaus,T. and Skarstad,K. (2009) The *Escherichia coli* SeqA protein. *Plasmid*, **61**, 141–150.
40. Correnti,J., Munster,V., Chan,T. and van der Woude,M. (2002) Dam-dependent phase variation of Ag43 in *Escherichia coli* is altered in a *seqA* mutant. *Mol. Microbiol.*, **44**, 521–532.

41. Jakomin,M., Chessa,D., Baumler,A.J. and Casadesus,J. (2008) Regulation of the *Salmonella enterica std* fimbrial operon by DNA adenine methylation, SeqA, and HdfR. *J. Bacteriol.*, **190**, 7406–7413.

42. Oberto,J., Nabti,S., Jooste,V., Mignot,H. and Rouviere-Yaniv,J. (2009) The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS One*, **4**, e4367.

43. Mangan,M.W., Lucchini,S., T,O.C., Fitzgerald,S., Hinton,J.C. and Dorman,C.J. (2011) Nucleoid-associated protein HU controls three regulons that coordinate virulence, response to stress and general physiology in *Salmonella enterica* serovar Typhimurium. *Microbiology*, **157**, 1075–1087.

44. Dorman,C.J. (2013) Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.*, **11**, 349–355.

45. Claret,L. and Rouviere-Yaniv,J. (1997) Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival. *J. Mol. Biol.*, **273**, 93–104.

46. Christman,M.F., Morgan,R.W., Jacobson,F.S. and Ames,B.N. (1985) Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in *Salmonella typhimurium*. *Cell*, **41**, 753–762.

47. Wallecha,A., Correnti,J., Munster,V. and van der Woude,M. (2003) Phase variation of Ag43 is independent of the oxidation state of OxyR. *J. Bacteriol.*, **185**, 2203–2209.

48. Zaim,J. and Kierzek,A.M. (2003) The structure of full-length LysR-type transcriptional regulators. Modeling of the full-length OxyR transcription factor dimer. *Nucleic Acids Res.*, **31**, 1444–1454.

49. Peterson,S.N. and Reich,N.O. (2006) GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence *pap* expression. *J. Mol. Biol.*, **355**, 459–472.

50. Kullik,I., Stevens,J., Toledano,M.B. and Storz,G. (1995) Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for DNA binding and multimerization. *J. Bacteriol.*, **177**, 1285–1291.

51. Choi,H., Kim,S., Mukhopadhyay,P., Cho,S., Woo,J., Storz,G. and Ryu,S.E. (2001) Structural basis of the redox switch in the OxyR transcription factor. *Cell*, **105**, 103–113.

52. Hryniewicz,M.M. and Kredich,N.M. (1991) The *cysP* promoter of *Salmonella typhimurium*: characterization of two binding sites for CysB protein, studies of in vivo transcription initiation, and demonstration of the anti-inducer effects of thiosulfate. *J. Bacteriol.*, **173**, 5876–5886.

53. Wang,L., Helmann,J.D. and Winans,S.C. (1992) The *A. tumefaciens* transcriptional activator OccR causes a bend at a target promoter, which is partially relaxed by a plant tumor metabolite. *Cell*, **69**, 659–667.

54. Fisher,R.F. and Long,S.R. (1993) Interactions of NodD at the *nod* box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. *J. Mol. Biol.*, **233**, 336–348.

55. van Keulen,G., Ridder,A.N., Dijkhuizen,L. and Meijer,W.G. (2003) Analysis of DNA binding and transcriptional activation by the LysR-type transcriptional regulator CbbR of *Xanthobacter flavus*. *J. Bacteriol.*, **185**, 1245–1252.

56. Lim,H.N. and van Oudenaarden,A. (2007) A multistep epigenetic switch enables the stable inheritance of DNA methylation states. *Nat. Genet.*, **39**, 269–275.

57. Charlier,D., Hassanzadeh,G., Kholti,A., Gigot,D., Pierard,A. and Glansdorff,N. (1995) *carP*, involved in pyrimidine regulation of the *Escherichia coli* carbamoylphosphate synthetase operon encodes a sequence-specific DNA-binding protein identical to XerB and PepA, also required for resolution of ColEI multimers. *J. Mol. Biol.*, **250**, 392–406.

58. Tao,K., Fujita,N. and Ishihama,A. (1993) Involvement of the RNA polymerase alpha subunit C-terminal region in co-operative interaction and transcriptional activation with OxyR protein. *Mol. Microbiol.*, **7**, 859–864.

59. Kang,S., Lee,H., Han,J.S. and Hwang,D.S. (1999) Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. *J. Biol. Chem.*, **274**, 11463–11468.

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

61. Carrera,P. and Azorin,F. (1994) Structural characterization of intrinsically curved AT-rich DNA sequences. *Nucleic Acids Res.*, **22**, 3671–3680.

62. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine-thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8490–8495.