

Convergence of DNA methylation and phosphorothioation epigenetics in bacterial genomes

Chao Chen^{a,b,1}, Lianrong Wang^{a,1}, Si Chen^a, Xiaolin Wu^{a,b}, Meijia Gu^a, Xi Chen^a, Susu Jiang^a, Yunfu Wang^b, Zixin Deng^a, Peter C. Dedon^c, and Shi Chen^{a,d,2}

^aKey Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, School of Pharmaceutical Sciences, Wuhan University, Wuhan 430071, China; ^bTaihe Hospital, Hubei University of Medicine, Shiyuan 442000, Hubei, China; ^cDepartment of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^dMedical Research Institute, Wuhan University, Wuhan 430071, China

Edited by Philip C. Hanawalt, Stanford University, Stanford, CA, and approved March 14, 2017 (received for review February 13, 2017)

Explosive growth in the study of microbial epigenetics has revealed a diversity of chemical structures and biological functions of DNA modifications in restriction–modification (R-M) and basic genetic processes. Here, we describe the discovery of shared consensus sequences for two seemingly unrelated DNA modification systems, ^{6m}A methylation and phosphorothioation (PT), in which sulfur replaces a nonbridging oxygen in the DNA backbone. Mass spectrometric analysis of DNA from *Escherichia coli* B7A and *Salmonella enterica* serovar Cerro 87, strains possessing PT-based R-M genes, revealed d(G_{PS}^{6m}A) dinucleotides in the G_{PS}^{6m}AAC consensus representing ~5% of the 1,100 to 1,300 PT-modified d(G_{PS}A) motifs per genome, with ^{6m}A arising from a yet-to-be-identified methyltransferase. To further explore PT and ^{6m}A in another consensus sequence, G_{PS}^{6m}ATC, we engineered a strain of *E. coli* HST04 to express Dnd genes from *Hahella chejuensis* KCTC2396 (PT in G_{PS}ATC) and Dam methyltransferase from *E. coli* DH10B (^{6m}A in G^{6m}ATC). Based on this model, in vitro studies revealed reduced Dam activity in G_{PS}ATC-containing oligonucleotides whereas single-molecule real-time sequencing of HST04 DNA revealed ^{6m}A in all 2,058 G_{PS}ATC sites (5% of 37,698 total GATC sites). This model system also revealed temperature-sensitive restriction by DndFGH in KCTC2396 and B7A, which was exploited to discover that ^{6m}A can substitute for PT to confer resistance to restriction by the DndFGH system. These results point to complex but unappreciated interactions between DNA modification systems and raise the possibility of coevolution of interacting systems to facilitate the function of each.

DNA phosphorothioation | DNA methylation | epigenetics | restriction–modification | bioanalytical chemistry

The emergence of convergent technologies has led to a growing appreciation for the diversity of DNA modifications in microbial epigenetics and restriction–modification (R-M) systems (1–3). DNA methylation, the most extensively studied genetic modification, was originally discovered in bacteria in the context of R-M systems involving a methyltransferase (MTase) that modifies “self” DNA at specific target sites and a cognate restriction endonuclease (REase) that discriminates and destroys unmodified invading DNA (3–5). R-M systems are ubiquitous in prokaryotes and are classified into four types (I, II, III, and IV) based on their molecular structure, sequence recognition, cleavage position, and cofactor requirements (3, 6, 7). However, some MTases exist alone, without an apparent cognate REase partner. These so-called “orphan” MTases include DNA adenine methylase (Dam), which modifies the adenine N-6 in the GATC motif, DNA cytosine methylase (Dcm), which methylates C-5 of the second cytosine in CC(AT)GG sequences, and cell cycle-regulated methylase (CcrM), which methylates the N-6 of adenine in GANTC (N = A, T, C, or G) (8). Despite the absence of cognate REases, orphan MTases still confer immunity against the virulence of a parasitic R-M complex with the same target sites (9). In addition to defense against bacteriophages and transposons, DNA methylation is involved in various cellular functions, including chromosome replication, DNA

mismatch repair, transcriptional regulation, transposition, and pathogenesis (1, 10, 11).

The chemical diversity of DNA modifications expanded with our discovery of phosphorothioates (PTs) as physiological modifications of the DNA sugar-phosphate backbone in which a nonbridging oxygen is replaced by sulfur (12). The PT system is more complicated than methylation and entails sequence-selective and R_p-stereo-specific incorporation of sulfur by five Dnd proteins (12, 13). Since the original report in *Streptomyces lividans*, the five-member *dndABCDE* gene cluster has been found in diverse, taxonomically unrelated bacteria and archaea (14, 15). *dndA* can be located either adjacent to clustered *dndBCDE* or elsewhere in the genome, consistent with the function of DndA as a cysteine desulfurase; DndA can also be functionally replaced by an iron–sulfur cluster assembly protein (IscS) homolog in some bacterial strains (16). DndB binds to the promoter of the *dnd* operon to regulate *dndBCDE* transcription and affect PT frequency (17, 18). DndC possesses ATP pyrophosphatase activity and shows significant homology to PAPS reductase family proteins (13, 16). By comparison, DndD has ATPase activity that is possibly related to DNA nicking during sulfur incorporation, and structural analysis suggests that DndE binds to nicked double-stranded DNA (13, 19, 20). In some strains, a divergently transcribed cluster, *dndFGH*, occurs in the vicinity of the *dndBCDE* operon. Without PT protection, unrestrained DndFGH restriction activity leads to double-stranded DNA damage that triggers the SOS response, cell filamentation, and prophage

Significance

New technologies have fostered renewed interest in bacterial epigenetics, with DNA modifications defending against other microbes and controlling gene expression. Here, we describe how two diverse modifications, adenine methylation (^{6m}A) and phosphorothioation (PT), have evolved to occupy the same genomic sites in bacteria. Using mass spectrometry, single-molecule real-time sequencing, and genetic engineering, we discovered co-incident PT and ^{6m}A in G_{PS}^{6m}AAC and G_{PS}^{6m}ATC, representing PT consensus sequences, a GA-containing consensus for an unknown methyltransferase, and the GATC consensus for Dam methyltransferase. ^{6m}A was found in all G_{PS}ATC sites in vivo, with ^{6m}A substituting for PT in preventing DNA cleavage by PT restriction enzymes. These findings show how different DNA modification systems have evolved to avoid conflict.

Author contributions: L.W., P.C.D., and Shi Chen designed research; C.C., X.W., M.G., X.C., and S.J. conducted experiments; C.C., L.W., Si Chen, X.W., Y.W., Z.D., P.C.D., and Shi Chen analyzed data; C.C., L.W., P.C.D., and Shi Chen wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹C.C. and L.W. contributed equally to this work.

²To whom correspondence should be addressed. Email: shichen@whu.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702450114/-DCSupplemental.

induction (21, 22). Dnd is thus regarded as an unusual bacterial R-M system.

The complexities of different DNA modification systems have begun to come to light with the application of new technologies. For example, the application of liquid chromatography-coupled tandem quadrupole mass spectrometry (LC-MS/MS) to quantify nuclease-resistant PT-modified dinucleotides has revealed significant diversity in the quantity, sequence setting, and bacterial distribution of PT modifications, such as the $d(G_{PS}A)$ and $d(G_{PS}T)$ motifs in 1:1 proportions in the enteric pathogen *Escherichia coli* B7A (15). Subsequent application of single-molecule, real-time (SMRT) sequencing technology to map genomic PT sites in *E. coli* B7A revealed that these $d(G_{PS}A)$ and $d(G_{PS}T)$ motifs are located within conserved 4-bp complementary $G_{PS}AAC/G_{PS}TTC$ contexts, with no apparent wider consensus sequence (23). However, only 12% of the 40,701 possible GAAC/GTTC sites on the chromosome were found to be PT-modified, even in the presence of active DndFGH (23). These unique features differentiate *dndBCDE-dndFGH* from typical methylation-based R-M systems and suggest that it represents a type of bacterial defense system with highly unusual target selection by Dnd proteins.

Despite having similar functional roles, DNA methylation and PT modification are considered to be two R-M defense systems that have evolved independently in prokaryotes. Here, we report the discovery of an intriguing collision of the two modification systems, which were found to share DNA substrates to generate a hybrid $d(G_{PS}^{6m}A)$ modification motif. This commonality is associated with significant rules of interaction and interference: sulfur replacement on the DNA backbone remarkably inhibits methylation, and methylation can substitute for PT to confer resistance to restriction by DndFGH. This interaction between two distinct endogenous DNA modification systems suggests that such “conflict” drives the evolution of different consensus sequences to avoid interference and ensure the protective function of each.

Results

Discovery of the Doubly Modified $d(G_{PS}^{6m}A)$ Consensus in Bacteria. PT modifications occur sequence-selectively in DNA, with all 16 possible dinucleotide contexts represented in archaeobacterial and eubacterial organisms (15). The PT context of $d(G_{PS}A)$, which occurs in *E. coli* B7A and *Salmonella enterica* serovar Cerro 87 (15), drew our attention because adenine is the substrate of N6-adenine-specific methyltransferases (MTases). This enzyme class catalyzes the $6m$ A modification widespread in R-M systems, such as M.TaqI, M.ClaI, M.EcoBI, and M.EcoRI, as well as solitary Dam and CcrM MTases. Adenine MTases modify a variety of sequence contexts such as TCGA, ATCGAT, TGA(N)8TGCT, GAATTC, GATC and GANTC (underlined bases indicate methylation). What was notable here is the coincidence of recognition sequences of some adenine MTases with those for PT modifications, such as TGA(N)8TGCT for M.EcoBI and others (24) and GAAC/GTTC for Dnd enzymes.

These coincidences prompted us to search for PT and $6m$ A in the same sequence motifs. Here, we searched first for $d(G_{PS}^{6m}A)$ dinucleotides in *E. coli* B7A and *S. enterica* serovar Cerro 87 possessing $d(G_{PS}A)$ motifs, using an LC-MS/MS method that exploits the nuclease resistance of PT-containing dinucleotides. Genomic DNA is hydrolyzed and persistent PT-containing dinucleotides are resolved from the bulk 2'-deoxyribonucleosides by HPLC, which allows sensitive MS identification and detection of the dinucleotides. Systematic MS scanning of the HPLC eluent of hydrolyzed genomic DNA from these bacteria revealed a molecule with m/z 611 (Fig. 1A). Collision-induced dissociation (CID) analysis yielded fragment ions with m/z values of 150, 362, and 460 (Fig. 1A), which are consistent with the structure $d(G_{PS}^{6m}A)$ shown in Fig. 1B. This structure was validated in several ways. First, the m/z 611 species disappeared in DNA from PT-deficient mutant strains CX3 and XTG103 of *E. coli* B7A and *S. enterica* serovar Cerro 87,

respectively, which points to the presence of a PT structure in the dinucleotide. Second, the CID fragmentation properties parallel those of $d(G_{PS}A)$, which is 14 atomic mass units (amu) lighter due to the absence of the N⁶-methyl group: a protonated molecular ion $[M+H]^+$ m/z of 597 (vs. 611), a protonated adenine base (m/z 136 vs. 150), and other fragments (m/z 348, 446) lighter by 14 amu (12). Finally, the structure in Fig. 1B was confirmed by LC-MS/MS analysis of a synthetic $d(G_{PS}^{6m}A)$ standard that possessed PT as a mixture of R_P and S_P stereochemistries (Fig. 1C), with R_P as the naturally occurring configuration (Fig. 1A). These results established the cooccurrence of a PT and a methyl group in the same $d(GA)$ context to yield a hybrid $d(G_{PS}^{6m}A)$ in R_P configuration.

We next performed quantitative analysis to establish the frequency of $d(G_{PS}^{6m}A)$ in the bacterial genomes. LC-MS/MS analysis using calibration curves based on chemical standards revealed that, compared with the frequency of $d(G_{PS}A)$ at 301 ± 2 and 258 ± 17 per 10^6 nt, the methylated $d(G_{PS}^{6m}A)$ occurred at much lower levels of 15 ± 1 and 2 ± 0 per 10^6 nt in *E. coli* B7A and *S. enterica* serovar Cerro 87, respectively. Interestingly, we are unaware of adenine MTases with GAAC/GTTC as the precise recognition sequence, so it is possible that the frequency of $d(G_{PS}^{6m}A)$, low as it is, resulted from the overlap of GA in the consensus sequences of a MTase and the PT-inserting Dnd proteins in a subset of GAAC/GTTC sites. To further analyze the intersecting DNA modifications, we engineered bacteria with PT and methylation systems, as discussed next.

Interaction Between PT Modification and Methylation in Bacterial Genomes. This coincidence of PT and $6m$ A modifications at the same sites in bacterial genomes raised questions about both the target selection mechanisms underlying the modification proximity and the implications of proximity for modification function. Here, we explored this issue with PTs and Dam-mediated methylation, both of which are considered to be postreplicative modifications, with hemimodified sites created by passage of the replication fork later fully modified (25, 26). We used the PT modification system from *Hahella chejuensis* KCTC2396 and *dam* from *E. coli* DH10B, both of which recognize GATC as a consensus sequence and generate $G_{PS}ATC$ and $G^{6m}ATC$, respectively. To assess how Dam reacts with PT-modified DNA, oligonucleotide substrates harboring GATC or $G_{PS}ATC$ (Table S1) were reacted with Dam and S-adenosylmethionine (SAM) in vitro. As shown in Fig. 2A, Dam is able to use $G_{PS}ATC$ with both R_P and S_P stereoisomers of PT as a substrate to yield $d(G_{PS}^{6m}A)$. However, a time course of the Dam methylation reaction measured by LC-MS/MS (Fig. 2B) revealed that both the rate and extent of Dam activity were significantly reduced in the presence of PTs, regardless of whether the substrate was single- or double-stranded. The sulfur substitution of $G_{PS}ATC$ thus interfered with either recognition or catalysis by Dam in an in vitro setting.

We next assessed the interaction of Dam and the PT-modifying enzymes DndBCDE in vivo. Here, we used *E. coli* HST04, which lacks *dam*, *dcm*, and *hds* MTases, as well as PT modification genes, as a host for vectors expressing *dnd* genes B, C, D, and E (*dndBCDE*) and *dam* genes (pWHU701 and pWHU702, respectively); *E. coli* HST04 possesses the *iscS* gene that substitutes for *dndA*. LC-MS/MS analysis revealed the formation of $d(G_{PS}^{6m}A)$ in exponentially growing *E. coli* HST04(*dndBCDE*, *dam*) at a frequency of 219 ± 2 per 10^6 nt, compared with 228 ± 4 $d(G_{PS}A)$ per 10^6 nt in *E. coli* HST04(*dndBCDE*) lacking *dam* (Table S2). This finding stands in contrast to the observation of PT inhibition of Dam activity in vitro and suggests either that Dam activity in vivo is not affected by the presence of a PT or that $6m$ A forms first followed by PT insertion, with Dnd proteins unaffected by $6m$ A. As described next, however, only a fraction of the 37,610 GATC sites containing $6m$ A also contained PT, which points to different target selection mechanisms by Dnd and Dam proteins.

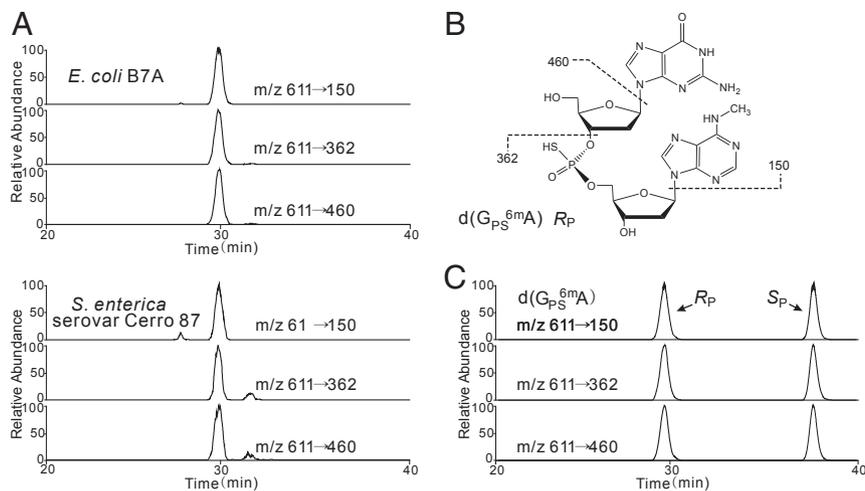


Fig. 1. LC-MS/MS analysis of isolated and synthetic $d(G_{PS}^{6mA})$. (A) Naturally occurring $d(G_{PS}^{6mA})$ R_P in *E. coli* B7A (Upper) and *S. enterica* serovar Cerro 87 (Lower) was eluted at 29.5 min, identical to the retention time of synthetic $d(G_{PS}^{6mA})$ (C) in the R_P configuration. (B) The major fragment ions of $d(G_{PS}^{6mA})$ observed in LC-MS/MS analysis are shown in the structural insert, with the protonated molecular ion $[M+H]$ appearing at m/z 611.

Comparative Genome Maps of PT and $6mA$ Sites by SMRT Sequencing.

The evidence for coincidence motivated an attempt to apply the SMRT sequencing platform to compare the locations of PT and $6mA$ across the genome of engineered *E. coli* HST04. SMRT sequencing of DNA modifications exploits the kinetic signatures, defined as pulse width and interpulse duration (IPD), arising when the DNA polymerase copies past the modification (27). Building on SMRT applications to nucleobase methylation (27), 5-hydroxymethylcytosine (28), and nucleobase damage (29), we recently demonstrated that SMRT sequencing is applicable to PT mapping, with identification of $G_{PS}AAC/G_{PS}TTC$ and $C_{PS}CA$ in *E. coli* B7A and *Vibrio cyclitrophicus* FF75, respectively (23). Both $6mA$ and $d(G_{PS}A)$ modifications of template DNA generate kinetic signatures with regard to the A base. The discovery of the hybrid $d(G_{PS}^{6mA})$ prompted us to compare the maps of PT and $6mA$ across the *E. coli* HST04 genome.

To this end, we used SMRT sequencing to map $G_{PS}ATC$ in *E. coli* HST04(*dndBCDE*) and $G^{6mA}ATC$ and hybrid $G_{PS}^{6mA}ATC$ in *E. coli* HST04(*dndBCDE, dam*). In the 4.57×10^6 bp in the *E. coli* HST04(*dndBCDE*) genome, 2,058 of the 37,698 GATC sites (5%) were found to be PT-modified (Fig. 3), which is consistent with the $d(G_{PS}A)$ quantitation by LC-MS/MS (228 ± 4 PT per 10^6 nt) (Table S2). Among the 2,058 $G_{PS}ATC$ sites, 1,019 are located on the (+) strand and 1,039 on the (-) strand, with PT occurring on both strands of the same GATC site 517 times. The remaining 1,024 PT modification sites are single-stranded $G_{PS}ATC$ sites.

The map of $G^{6mA}ATC$ and $G_{PS}^{6mA}ATC$ sites across the *E. coli* HST04(*dndBCDE, dam*) genome revealed modification of nearly 100% of all GATC sites (37,610 out of 37,698 GATC sites) (Fig. 3). However, the similarity of the SMRT sequencing IPD ratios for $G_{PS}^{6mA}ATC$ and $G^{6mA}ATC$ made it difficult to distinguish between them in the genome modification maps in *E. coli* HST04(*dndBCDE, dam*). Knowing that nearly all of the 2,058 PT-modified GATC sites detected in *E. coli* HST04(*dndBCDE, dam*) contained $6mA$ in the *dam*-expressing *E. coli* HST04(*dndBCDE, dam*), we can conclude that all $G_{PS}ATC$ sites in *E. coli* HST04(*dndBCDE, dam*) also contain $6mA$, which leads to the observation that IPD ratios for $G_{PS}^{6mA}ATC$ sites were 2.8-fold larger than those for $G_{PS}ATC$, with 89% of the $G_{PS}^{6mA}ATC$ sites having an IPD ratio $\geq 2\times$ larger than $G_{PS}ATC$. Using a Gaussian distribution as the smoothing kernel, we estimated the probability density functions of the IPD ratio values for 2,058 $G_{PS}ATC$, 2,058 $G_{PS}^{6mA}ATC$, and 35,552 $G^{6mA}ATC$ on chromosomes. As shown in Fig. S1, the

immediate observation was that the average IPD ratios of both $G_{PS}^{6mA}ATC$ and $G^{6mA}ATC$ were larger than that of $G_{PS}ATC$ and that the differences were significant (both P values < 0.0001). The IPD ratios of $G_{PS}ATC$ were largely concentrated within a small range of [1.3, 3.5] and formed a sharp peak; in contrast, the $G_{PS}^{6mA}ATC$ and $G^{6mA}ATC$ sites exhibited more widely ranging IPD

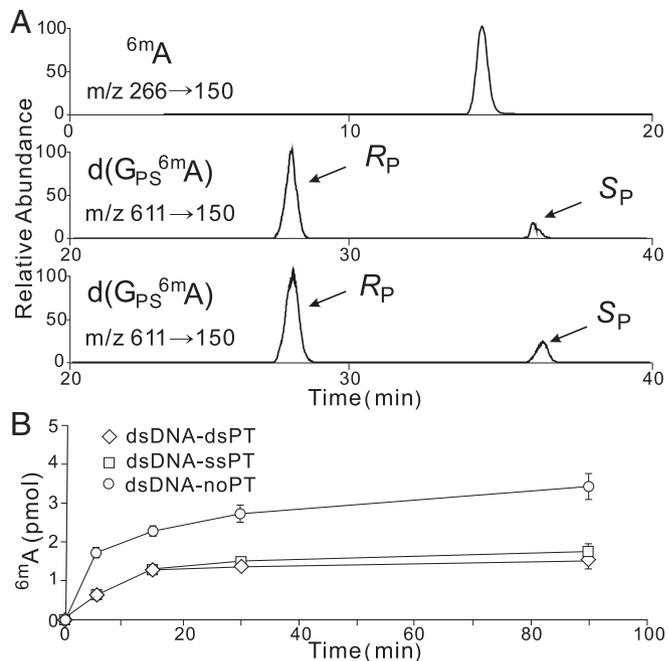


Fig. 2. Dam enzymatic activity with different oligonucleotide substrates. (A) LC-MS/MS analysis of the Dam-methylated products using duplex oligonucleotides possessing normal GATC (dsDNA-noPT), single-stranded $G_{PS}ATC$ (dsDNA-ssPT), and double-stranded $G_{PS}ATC$ (dsDNA-dsPT) as substrates. The oligonucleotides are listed in Table S1. Dam recognizes $G_{PS}ATC$ in both R_P and S_P configurations. (B) Time course of the Dam enzymatic reaction with dsDNA-noPT, dsDNA-ssPT, and dsDNA-dsPT; 4 units of Dam were incubated with 20 pmol of a DNA substrate in 10 μ L of $1\times$ *dam* methyltransferase buffer supplemented with 80 μ M SAM for 0 to 90 min at 37 $^{\circ}$ C, followed by inactivation at 65 $^{\circ}$ C for another 15 min. $6mA$ was measured by LC-MS/MS at the indicated times (SI Materials and Methods).

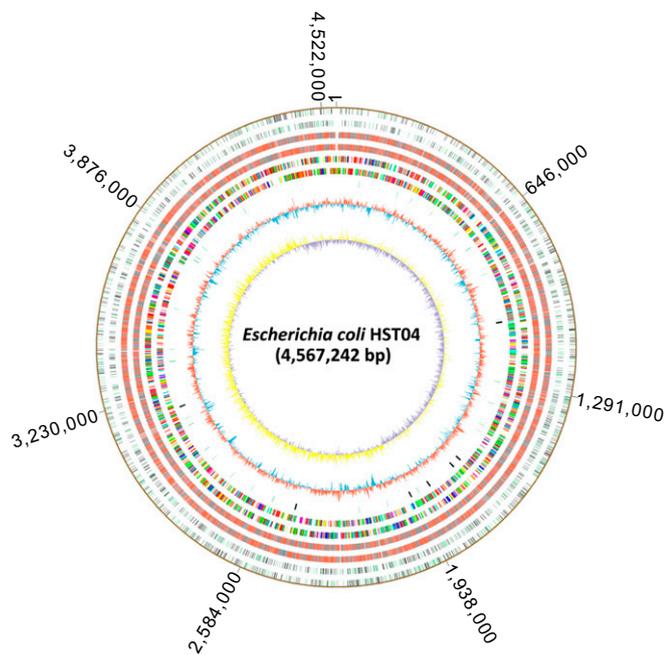


Fig. 3. PT and methylation site mapping across the *E. coli* HST04 genome. From outer to inner circles: Circles 1 and 2, (forward, reverse strands) 2,058 $G_{PS}^{6m}ATC$ sites in *E. coli* HST04 (*dndBCDE*) detected by SMRT in ORFs (black) and noncoding regions (green). No site is detected in noncoding RNA. These sites were almost completely converted to methylation-PT double modification sites in *E. coli* HST04 (*dndBCDE*, *dam*). Circles 3 and 4, methylation sites *E. coli* HST04 (*dndBCDE*, *dam*) (excluding sites in circles 1 and 2) in ORFs (gray), noncoding RNA (blue), and noncoding regions (red). Circles 5 and 6, predicted protein-coding sequences colored according to cluster of orthologous groups of proteins (COG) function categories. Circle 7, tRNA/rRNA operons. Circle 8, GC content. Circle 9, GC skew.

ratios, and the corresponding probability density curves were flatter (Fig. S1). In addition, the $G_{PS}^{6m}ATC$ and $G^{6m}ATC$ probability density curves were similar in shape: (i) both curves were symmetric about their means; (ii) the highest probability density of each did not differ much between the curves (0.405 for $G_{PS}^{6m}ATC$ and 0.437 for $G^{6m}ATC$); and (iii) the ranges covering the middle 90% IPD ratios for the curves did not differ much ([3.98, 7.32] for $G_{PS}^{6m}ATC$ and [3.58, 6.67] for $G^{6m}ATC$). Therefore, the similar IPD ratios of $G_{PS}^{6m}ATC$ and $G^{6m}ATC$ made it difficult to distinguish between them, consistent with the SMRT result that the 2,058 $G_{PS}^{6m}ATC$ sites were grouped with the remaining 35,552 methylation $G^{6m}ATC$ sites by the SMRT analysis portal (Fig. 3). We nonetheless conclude that PT and $6mA$ did not interfere with each other in terms of modification location or frequency in vivo, with normal target recognition behavior for proteins in both modification systems.

Development of a Temperature-Dependent PT-Associated Restriction System. With evidence for proximal PT and $6mA$ modifications, we next sought to assess the effects of $6mA$ on PT-dependent restriction. Here, we engineered *E. coli* HST04 to express the DndFGH restriction proteins from *H. chejuensis* KCTC2396, which, as noted for *H. chejuensis* Dnd modification proteins described earlier, lacks Dam MTase and possesses PTs as $d(G_{PS}^{6m}A)$ in the Dam methylation consensus GATC. To this end, we first cloned the *dndFGH* gene from *H. chejuensis* KCTC2396 into the low-copy plasmid pACYCDuet-1 to generate pWHU705. Without the protection of PT modifications, the DndFGH restriction endonuclease activity should be lethal or at least toxic to bacteria as a result of DNA damage and the SOS response (5, 21, 22). Surprisingly, we were able to prepare the pWHU705 plasmid,

and the restriction proteins DndFGH exerted no apparent toxicity or growth inhibition in *E. coli* HST04 at 37 °C (Fig. 4A). However, a DndFGH-mediated bactericidal effect emerged when the cells were grown at 15 °C (Fig. 4B). Similar thermoregulated restriction was also observed for WXY-1, a *dndBCDE*-deficient mutant of *E. coli* B7A. Without PT protection, WXY-1 demonstrated a slight growth defect in liquid culture at 37 °C, but restriction became more evident when the cells were cultured at 15 °C (Fig. S2). To characterize this effect of growth temperature, we examined *dndFGH* transcription in *E. coli* HST04(pWHU705) and WXY-1 at different temperatures using real-time quantitative RT-PCR (qRT-PCR) with GAPDH as an internal control. Levels of *dndFGH* transcript increased significantly when cells growing at 37 °C were shifted to 15 °C for 1 h (Table S3). These results reveal that DndFGH restriction systems in *E. coli* B7A and *H. chejuensis* KCTC2396 are thermoregulated, with DndBCDE-DndFGH possibly functioning as an unusual temperature-dependent defense system. The thermoregulated behavior of DndFGH proved useful for studying the effects of $6mA$ on PT restriction activity.

DNA Methylation Protects Against PT-Associated Restriction. The temperature-sensitive DndFGH system in *E. coli* HST04 provided an opportunity to test the effects of $6mA$ on PT-dependent restriction. Here, we compared the bactericidal activity of DndFGH in the presence of PT or $6mA$ by quantifying viability [colony-forming units (cfu)] after exponentially growing cells were plated and cultured at 37 °C or 15 °C. In the absence of either PT or $6mA$, the viability of *E. coli* HST04 cells at 15 °C was significantly decreased in the presence of DndFGH (Fig. 4C). Cotransformation of *E. coli* HST04 with pWHU705 and pWHU701, which express *H. chejuensis* KCTC2396 DndFGH and DndBCDE, respectively, restored viability at 15 °C to 66% of the level at 37 °C (Fig. 4C). This partial recovery of cell viability could result from the amount of DndBCDE expressed from pWHU701 under these growth conditions, which may not be fully sufficient to counteract the temperature-induced DndFGH. Remarkably, methylation of GATC sites conferred protection against DndFGH restriction activity to almost the same extent as PT modification (64%) (Fig. 4C), which suggests that the A methylation of GATC sites hampered DNA cleavage activity or interfered with the ability of DndFGH to bind to target sites. The experiment to assess the mechanism of A methylation on DndFGH is not yet possible due to the lack of a system for in vitro reconstitution of DndFGH activity.

Discussion

The advent of SMRT sequencing has accelerated the pace of discovery and characterization of prokaryotic epigenetics. In particular, methylation-based R-M systems and epigenetic marks have been detected in the vast majority of bacterial and archaeal genomes (3, 30). The chemical diversity of DNA modifications in bacteria has recently grown beyond methylation with the discovery of 7-deazaguanine modifications (2) and PT modifications of the DNA backbone (12, 15). We have shown that PTs function in R-M systems with DndFGH (22, 23, 31) but that PTs are also present in bacteria that lack restriction genes. Although it is well-established that many bacteria possess multiple R-M systems and MTases, each with unique methylation consensus sequences (3, 32), we have discovered a proximity of two diverse DNA modifications in the same core consensus sequence, with interactions that point to a functional cooperation of the two modification systems.

Although SMRT sequencing is not able to resolve closely spaced DNA modifications, including PT and $6mA$ in $G_{PS}^{6m}ATC$ consensus sequences as observed here, we were able to detect $d(G_{PS}^{6m}A)$ dinucleotides by LC-MS/MS in two unrelated bacteria, *E. coli* B7A and *S. enterica* serovar Cerro 87, by virtue of the PT-induced nuclease inhibition that produces a limit digest of PT-containing dinucleotides. The fact that $d(G_{PS}^{6m}A)$ occurred at much lower

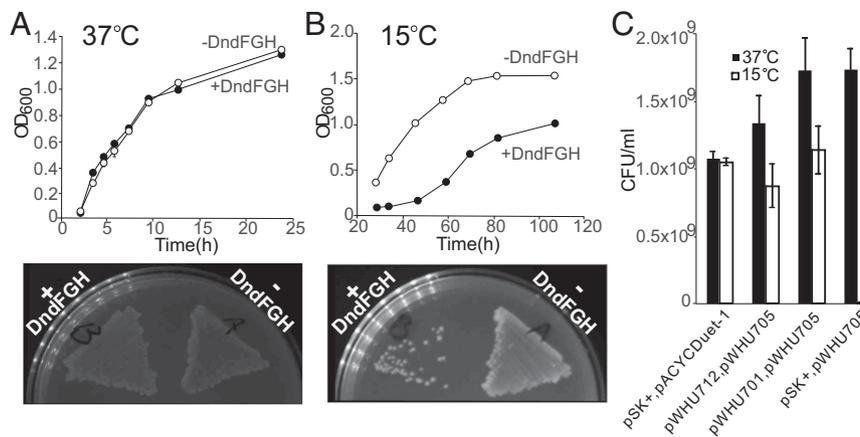


Fig. 4. Growth and cell viability analysis of bacterial strains. (A and B) Effect of pWHU705, expressing DndFGH of *H. chejuensis* KCTC2396, on the growth of *E. coli* HST04 at 37 °C (A) and 15 °C (B) in liquid LB broth (Upper) and on LB-agar plates (Lower). *E. coli* HST04 harboring empty pBluescript SK+ or pACYCDuet-1 plasmid was used as the control. (C) Cell viability assessment to compare protection mediated by DNA methylation and PT modification against DndFGH restriction. Exponentially growing cells at 37 °C were appropriately diluted and plated onto fresh LB-agar plates, which were incubated at 37 °C or 15 °C for cfu counting. Error bars indicate the SDs of three independent experiments.

frequencies than the $d(G_{PS}A)$ dinucleotide (2 to 15 per 10^6 nt versus 250 to 300 per 10^6 nt) supports the idea that $d(G_{PS}^{6mA})$ is occurring in the larger consensus sequence of a DNA methyltransferase. Examples of such sequences in other *E. coli* strains include the GAACC consensus of many *E. coli* R-M systems (33) and the larger recognition sequences for EcoBI/Eco1158I (TGAN₃TGCT), Eco377I (GGAN₈ATGC), and Eco777I (GGAN₆TATC) (24). However, *E. coli* B7A does not possess genes for these enzymes, so the source of $d(G_{PS}^{6mA})$ in B7A must be another MTase. Similarly, $d(G_{PS}^{6mA})$ dinucleotides must occur in the GAAC consensus sequence for Dnd proteins in *S. enterica* serovar Cerro 87, so DNA^N-adenine-methyltransferase from phage origin (M2.SenCer87ORF144P) and methyl-directed repair DNA adenine methylase (M.SenCer87DamP) (33), which target GATC, cannot account for the $d(G_{PS}^{6mA})$ modification. Interestingly, the $G_{PS}ATC$ sites may also contain ^{5m}C due to the activity of C5-cytosine-specific DNA methylase (M1.SenCer87ORF144P) (33), which raises the possibility of sites containing all three modifications, $G_{PS}^{6mA}AT^{5m}C$, further complicating the interplay among DNA modification systems.

The discovery of proximate PT and ^{6mA} motivated the development of a model system to further characterize coincident DNA modification consensus sequences. Here, we focused on the GATC recognition sequence for DndBCDE from *H. chejuensis* KCTC2396 and Dam MTase. Although the in vitro activity of Dam was partially inhibited by a preexisting PT modification, the SMRT sequencing map of ^{6mA} in GATC sites in *E. coli* HST04 (*dndBCDE*, *dam*) showed complete modification of all GATC sites (37,610 of 37,698 sites), which suggests either that the Dam inhibition is an artifact of the in vitro situation, with in vivo factors overcoming any inhibitory effect of PT, or that Dam methylation occurs in synchrony with or before PT modification. As a double-stranded DNA modification at GATC, ^{6mA} methylation by Dam does indeed occur quickly after DNA replication (34) so it is possible that PT and ^{6mA} are inserted in tandem at hemimodified sites postreplication.

The *E. coli* HST04 model system provided an opportunity to test these in vitro observations in vivo. The observation of PT modifications in 2,058 of the 37,698 GATC sites (5%) in *E. coli* HST04 (*dndBCDE*) is similar to previous SMRT analyses in *E. coli* B7A (12% of GAAC/GTTC) and *V. cyclitrophicus* FF75 (14% of CCA) (23), with the 228 ± 4 PT per 10^6 nt measured by LC-MS/MS analysis corroborating the quantitative nature of SMRT mapping. However, the map of the $G_{PS}ATC$ sites across the genome

(GenBank accession no. CP013952) showed PTs in 939 ORFs, 1,119 in noncoding regions, and none in tRNAs or rRNA genes (Fig. 3), which stands in contrast to the slight enrichment of PTs in tRNA genes in *E. coli* B7A (23), in which *dnd* genes are native.

We also made the surprising observation of the temperature dependence of the DndFGH restriction system in two bacteria. Thermoregulated R-M systems are relatively rare, with two recent studies describing this phenotype for the restriction endonucleases LlaJI and LmoH7 (35, 36). The first is a plasmid-associated R-M system that mediates temperature-dependent resistance to phages in *Lactococcus lactis*, with LlaJI activity most pronounced when plaque assays are performed at 19 °C and completely abolished at 37 °C (35). The other reported temperature-sensitive restriction phenomenon involves the endonuclease LmoH7, which shares no sequence similarity, genetic components, or recognition sites with LlaJI, but it can confer phage resistance in *Listeria monocytogenes* ECII strains at low temperature (36). Similar to LlaJI and LmoH7, the mechanism responsible for *dndFGH* induction may be due to the instability of mRNA at high temperature or the involvement of an unknown temperature-sensitive transcriptional regulator. At this point, we do not know whether the temperature sensitivity occurs at the level of biochemical activity or protein levels, but the phenomenon points to the potential for an intriguing fitness advantage for bacteria in terms of resisting phage infection, for example, in moderate-temperature environments outside the core temperature of homeothermic animals (e.g., soil, water).

Although the coincidence of PT and ^{6mA} in the same DNA sequence allows an organism to repurpose consensus sequences, the phenomenon has implications for the effectiveness of R-M systems and the real and potential epigenetic functions of both PT and ^{6mA}. The artificial $G_{PS}^{6mA}ATC$ system used here ignores the potential disruptive effects of PT on the binding of ^{6mA}-sensitive protein binding in Dam-regulated epigenetic functions so it will be important to explore the coincidence of PT and DNA methylation in bacteria with both native modification systems. There are also interesting implications for R-M activity against phage infection. Many phages possess orphan MTases that modify and protect the phage genome against host restriction enzymes. For example, several phages have been found to possess GATC-specific Dam MTase activities, which not only modulate phage propagation but also protect against hundreds of R-M systems (37). Based on our observation of the interchangeability of ^{6mA} and PT against PT-dependent restriction, this protection would also be extended against

GATC-specific PT R-M systems in diverse bacteria. Interestingly, phages such as *Staphylococcus* phage K have evolved to lack 5'-GATC-3' motifs in their DNA genome (38).

In summary, the discovery of proximate PT and ^{6m}A DNA modifications in several bacteria point to the potential for widespread coincidence of epigenetics systems (Fig. S3). The results reveal complex but unappreciated interactions between DNA modification systems and raise the possibility of coevolution of interacting systems to facilitate the function of each.

Materials and Methods

Detailed descriptions of materials and methods are provided in *SI Materials and Methods*.

Bacterial Strains, Plasmids, and Growth Conditions. All strains, plasmids and primers used in this study are listed in Table S1. In-frame deletion mutants of *E. coli* B7A (WXY-1, CX2 and CX3, *dndBCDE*, *dndFGH*, and *dndBCDEFGH*) were constructed by homologous recombination using the thermo- and sucrose-sensitive plasmid pKOV-Kan, as previously described (31). Cell viability was determined by colony counting on LB-agar plates, plates placed at 15 °C for

24 h and shifted to 37 °C for colony counting for assessing low-temperature viability. qRT-PCR analysis of gene transcription was performed as described in detail in *SI Materials and Methods*.

DNA Modification Analysis. For in vitro DNA methylation, duplex oligonucleotides containing PT modifications were reacted with Dam MTase, and modifications were analyzed by LC-MS/MS as described in *SI Materials and Methods*. PT modifications were detected by LC-MS/MS analysis of enzymatically hydrolyzed DNA as previously described (39); see *SI Materials and Methods* for details. SMRT sequencing to map PT and methylation modifications was performed as described previously (23) and in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Bochen Zhu and Yuanjie Ge for research assistance. This work was supported by grants from the National Science Foundation of China (31520103902, 31670086, and 31300038) and the 973 program of the Ministry of Science and Technology (2013CB734003), Hubei Province's Outstanding Medical Academic Leader Program, the Young One Thousand Talent program of China, the US National Science Foundation (CHE-1019990), the US National Institute of Allergy and Infectious Diseases (AI112711), and the Singapore-MIT Alliance for Research and Technology sponsored by the National Research Foundation of Singapore.

- Casadesús J, Low D (2006) Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* 70:830–856.
- Thiaville JJ, et al. (2016) Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc Natl Acad Sci USA* 113:E1452–E1459.
- Blow MJ, et al. (2016) The epigenomic landscape of prokaryotes. *PLoS Genet* 12: e1005854.
- Meselson M, Yuan R, Heywood J (1972) Restriction and modification of DNA. *Annu Rev Biochem* 41:447–466.
- Blumenthal RM, Cheng X (2002) Restriction-modification systems. *Modern Microbial Genetics*, eds Streips UN, Yasbin RE (John Wiley & Sons, New York), 2nd Ed, pp 177–225.
- Roberts RJ, et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31:1805–1812.
- Loenen WA, Raleigh EA (2014) The other face of restriction: Modification-dependent enzymes. *Nucleic Acids Res* 42:56–69.
- Low DA, Weyand NJ, Mahan MJ (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect Immun* 69:7197–7204.
- Takahashi N, Naito Y, Handa N, Kobayashi I (2002) A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. *J Bacteriol* 184:6100–6108.
- Heussipp G, Fälker S, Schmidt MA (2007) DNA adenine methylation and bacterial pathogenesis. *Int J Med Microbiol* 297:1–7.
- Wion D, Casadesús J (2006) N6-methyl-adenine: An epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol* 4:183–192.
- Wang L, et al. (2007) Phosphorothioation of DNA in bacteria by *dnd* genes. *Nat Chem Biol* 3:709–710.
- Zhou X, et al. (2005) A novel DNA modification by sulphur. *Mol Microbiol* 57: 1428–1438.
- He X, et al. (2007) Analysis of a genomic island housing genes for DNA S-modification system in *Streptomyces lividans* 66 and its counterparts in other distantly related bacteria. *Mol Microbiol* 65:1034–1048.
- Wang L, et al. (2011) DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proc Natl Acad Sci USA* 108:2963–2968.
- You D, Wang L, Yao F, Zhou X, Deng Z (2007) A novel DNA modification by sulfur: DndA is a Nifs-like cysteine desulfurase capable of assembling DndC as an iron-sulfur cluster protein in *Streptomyces lividans*. *Biochemistry* 46:6126–6133.
- Cheng Q, et al. (2015) Regulation of DNA phosphorothioate modifications by the transcriptional regulator DptB in *Salmonella*. *Mol Microbiol* 97:1186–1194.
- He W, et al. (2015) Regulation of DNA phosphorothioate modification in *Salmonella enterica* by DndB. *Sci Rep* 5:12368.
- Hu W, et al. (2012) Structural insights into DndE from *Escherichia coli* B7A involved in DNA phosphorothioation modification. *Cell Res* 22:1203–1206.
- Yao F, Xu T, Zhou X, Deng Z, You D (2009) Functional analysis of *spfD* gene involved in DNA phosphorothioation in *Pseudomonas fluorescens* Pf0-1. *FEBS Lett* 583:729–733.
- Cao B, et al. (2014) Pathological phenotypes and in vivo DNA cleavage by unrestrained activity of a phosphorothioate-based restriction system in *Salmonella*. *Mol Microbiol* 93:776–785.
- Gan R, et al. (2014) DNA phosphorothioate modifications influence the global transcriptional response and protect DNA from double-stranded breaks. *Sci Rep* 4:6642.
- Cao B, et al. (2014) Genomic mapping of phosphorothioates reveals partial modification of short consensus sequences. *Nat Commun* 5:3951.
- Kasarjian JK, Iida M, Ryu J (2003) New restriction enzymes discovered from *Escherichia coli* clinical strains using a plasmid transformation method. *Nucleic Acids Res* 31:e22.
- Dyson P, Evans M (1998) Novel post-replicative DNA modification in *Streptomyces*: Analysis of the preferred modification site of plasmid pJ101. *Nucleic Acids Res* 26:1248–1253.
- Barras F, Marinus MG (1989) The great GATC: DNA methylation in *E. coli*. *Trends Genet* 5:139–143.
- Flusberg BA, et al. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7:461–465.
- Song CX, et al. (2011) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods* 9:75–77.
- Clark TA, Spittle KE, Turner SW, Korlach J (2011) Direct detection and sequencing of damaged DNA bases. *Genome Integr* 2:10.
- Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE—A database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* 38:D234–D236.
- Xu T, Yao F, Zhou X, Deng Z, You D (2010) A novel host-specific restriction system associated with DNA backbone S-modification in *Salmonella*. *Nucleic Acids Res* 38:7133–7141.
- Krebs J, et al. (2014) The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res* 42:2415–2432.
- Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE—A database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* 43:D298–D299.
- Marinus MG, Løbner-Olesen A (2014) DNA methylation. *Ecosal Plus* 6(1).
- O'Driscoll J, et al. (2004) Lactococcal plasmid pNP40 encodes a novel, temperature-sensitive restriction-modification system. *Appl Environ Microbiol* 70:5546–5556.
- Kim JW, et al. (2012) A novel restriction-modification system is responsible for temperature-dependent phage resistance in *Listeria monocytogenes* ECI. *Appl Environ Microbiol* 78:1995–2004.
- Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D (2013) Bacteriophage orphan DNA methyltransferases: Insights from their bacterial origin, function, and occurrence. *Appl Environ Microbiol* 79:7547–7555.
- Krüger DH, Barcak GJ, Smith HO (1988) Abolition of DNA recognition site resistance to the restriction endonuclease EcoRII. *Biomed Biochim Acta* 47:K1–K5.
- Lai C, et al. (2014) In vivo mutational characterization of DndE involved in DNA phosphorothioate modification. *PLoS One* 9:e107981.
- DuPont HL, et al. (1971) Pathogenesis of *Escherichia coli* diarrhea. *N Engl J Med* 285: 1–9.
- Murase T, Nagato M, Shirota K, Katoh H, Otsuki K (2004) Pulsed-field gel electrophoresis-based subtyping of DNA degradation-sensitive *Salmonella enterica* subsp. *enterica* serovar Livingstone and serovar Cerro isolates obtained from a chicken layer farm. *Vet Microbiol* 99:139–143.
- Lee HK, et al. (2001) *Hahella chejuensis* gen. nov., sp. nov., an extracellular-polysaccharide-producing marine bacterium. *Int J Syst Evol Microbiol* 51:661–666.
- Alting-Mees MA, Short JM (1989) pBluescript II: Gene mapping vectors. *Nucleic Acids Res* 17:9494.
- Lalio M, Heath J (2001) A new method for generating point mutations in bacterial artificial chromosomes by homologous recombination in *Escherichia coli*. *Nucleic Acids Res* 29:E14.
- Chang AC, Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* 134:1141–1156.