Journal of Microbiological Methods 90 (2012) 46–52

Contents lists available at SciVerse ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

DNA binding site analysis of *Burkholderia thailandensis* response regulators

Kristy L. Nowak-Lovato [a], Alexana J. Hickmott [a], Tuhin S. Maity [a], Martha L. Bulyk [b,c,d], John Dunbar [a], Elizabeth Hong-Geller [a,*]

[a] Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States
[b] Division of Genetics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115, United States
[c] Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115, United States
[d] Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, United States

ARTICLE INFO

Article history:
Received 19 March 2012
Received in revised form 27 March 2012
Accepted 27 March 2012
Available online 13 April 2012

Keywords:
Response regulator
Protein-binding microarray
Promoter binding site
Burkholderia
Two component systems
Gene regulation

ABSTRACT

Bacterial response regulators (RR) that function as transcription factors in two component signaling pathways are crucial for ensuring tight regulation and coordinated expression of the genome. Currently, consensus DNA binding sites in the promoter for very few bacterial RRs have been identified. A systematic method to characterize these DNA binding sites for RRs would enable prediction of specific gene expression patterns in response to extracellular stimuli. To identify RR DNA binding sites, we functionally activated RRs using beryllofluoride and applied them to a protein-binding microarray (PBM) to discover DNA binding motifs for RRs expressed in *Burkholderia*, a Gram-negative bacterial genus. We identified DNA binding motifs for conserved RRs in *Burkholderia thailandensis*, including KdpE, RisA, and NarL, as well as for a previously uncharacterized RR at locus BTH_II2335 and its ortholog in the human pathogen *Burkholderia pseudomallei* at locus BPSS2315. We further demonstrate RR binding of predicted genomic targets for the two orthologs using gel shift assays and reveal a pattern of RR regulation of expression of self and other two component systems. Our studies illustrate the use of PBMs to identify DNA binding specificities for bacterial RRs and enable prediction of gene regulatory networks in response to two component signaling.

Published by Elsevier B.V.

1. Introduction

Bacteria employ two-component signaling systems to couple the sensing of stress signals to adaptive changes in gene expression, thus ensuring tight regulation and coordinated expression of the genome in response to the environment (Beier and Gross, 2006; Cheung and Hendrickson, 2010; Laub and Goulian, 2007). Two-component systems represent the single largest paralogous family of signaling proteins in the bacterial kingdom and regulate diverse cellular processes, including chemotaxis, osmoregulation, metabolism, and transport. As the name implies, the prototypical two-component system is composed of two parts. First, a histidine kinase catalyzes autophosphorylation on a conserved histidine residue upon sensing changes in growth conditions and then transfers the phosphoryl group to the receiver domain of the second component, a response regulator (RR), which functions as a downstream effector protein, often as a transcription factor that regulates gene expression (Bourret et al., 1991; Pao and Saier, 1995). For example, under K^+ limitation, the histidine kinase KdpD activates the response regulator KdpE, which in its phosphorylated state, induces expression of the *kdpFABC* operon *via* increased affinity for a 23 base pair sequence in the *kdpFABC* promoter. The *kdpFABC* operon, which lies adjacent to the *kdpDE* operon, encodes an inducible high-affinity K^+ uptake system that scavenges K^+ to maintain ionic homeostasis in the cell (Gasell and Altendorf, 2001).

The rapid sequencing of bacterial genomes in the last several years has revealed a diversity of RRs with undefined regulatory functions. From 1123 distinct bacterial genomes, ~39,000 two-component proteins adjacent in the genome have been identified (Ulrich and Zhulin, 2010). The majority of RRs with DNA binding capability fall into three major families based on the structural similarity of their effector domains, (1) OmpR/PhoB family, winged helix-turn-helix motif (Kenney, 2002), (2) NarL family, helix-turn-helix motif (Baikalov et al., 1996), and (3) NtrC family, ATPase domain (Yang et al., 2004). Although the target genes of some RRs can be predicted based on genomic organization, such as KdpE control of *kdpFABC*, RRs can regulate multiple target genes scattered throughout a bacterial genome. The completion of sequenced bacterial genomes has enabled bioinformatics searches using consensus sequence motifs to predict DNA binding sites for specific RRs. Thus far, experimental confirmation of DNA binding sites for RRs has been limited. Aside from KdpE, DNA binding sites have been determined for the *Escherichia coli* RRs OmpR (Pratt and Silhavy, 1995), NarL (Baikalov et al., 1996; Maris et al., 2002), and PhoB (Makino et al., 1988), which regulate osmolarity, nitrate response, and phosphate availability, respectively. However, the great majority of bacterial RRs has been identified based only on sequence

* Corresponding author. Tel.: +1 505 665 2465; fax: +1 505 665 3024.
E-mail address: ehong@lanl.gov (E. Hong-Geller).

0167-7012/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.mimet.2012.03.019

homology, and their target DNA binding sites remain unknown or poorly characterized. Further experimental identification of target DNA binding sites and the cognate genes regulated by specific RRs can link extracellular inputs (*e.g.* nutrient deprivation, ion concentration, pH change) to a regulatory gene network and better define the molecular mechanisms activated in response to two-component signaling pathways.

We have chosen Gram-negative bacteria *Burkholderia* spp. as the model organism for discovery of RR DNA binding sites. The *Burkholderia* genus encompasses ~60 species, which exhibit a wide range of biological functions, including pathogenicity, bioremediation, and nitrogen fixation. The two best-characterized species, *Burkholderia pseudomallei* and *Burkholderia mallei*, the causative agents of human melioidosis and equine glanders, respectively, are categorized as Category B biothreat agents by the CDC. We have employed protein-binding microarray (PBM) technology to determine the DNA binding specificities of RRs expressed in *Burkholderia thailandensis*, a closely-related species to *B. pseudomallei* that is non-pathogenic in humans. The PBM is a rapid methodology to simultaneously screen all sequence variants of a defined length and obtain comprehensive binding site measurements of DNA–protein interactions *in vitro* (Berger and Bulyk, 2009; Berger et al., 2006; Mukherjee et al., 2004). PBMs have been successfully used to analyze transcription factor binding specificities in a wide variety of organisms, including *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, mice, and humans (Berger et al., 2008, 2006; Robasky and Bulyk, 2011). To date, use of PBMs in bacterial systems has been limited to a specific quorum sensing RR, LuxR, in the marine bacterium *Vibrio harveyi* (Pompeani et al., 2008), two nucleoid-associated proteins, H-NS and Lsr2, from *Salmonella enterica* and *Mycobacterium tuberculosis* (Gordon et al., 2011), and several TetR and MarR transcription factors in *Burkholderia xenovorans* (Maity et al., 2011). In this study, we demonstrate the successful application of PBMs to both known and previously-uncharacterized *Burkholderia* RRs, as a broadly applicable method to identify bacterial transcription factor binding sites for analysis of gene regulation in a wide range of bacterial species. We also perform comparative PBM analysis between a pair of RR orthologs in *B. thailandensis* and *B. pseudomallei* to investigate the overlap of DNA binding specificities in different *Burkholderia* species. We expect that identification of RR DNA binding sites in *Burkholderia* can provide molecular insights into how two-component systems monitor different environmental parameters and allow for prediction of cellular behavior across bacterial species.

2. Materials and methods

2.1. Cloning and expression of GST fusions to RRs

The RR genes were amplified from *Burkholderia* genomic DNA using sequence-specific primers by PCR in 50 μl reactions [1 μl 100 μM primer 1, 1 μl 100 μM primer 2, 50 ng genomic DNA isolated from *B. thailandensis* E264 or *B. pseudomallei* K96243, 5 μl 10× Pfu reaction buffer, 1 μl 100 mM dNTPs, 2.5 μl DMSO, 2.5 U of PfuUltra DNA polymerase (Agilent, Santa Clara, CA), and distilled H_2O for the remaining volume] using the following conditions, (1) 94 °C, 3 min, (2) 94 °C, 1 min; 50 °C, 1 min, 72 °C, 1 min for 30 cycles, and (3) 94 °C, 1 min; 50 °C, 1 min, 72 °C, 10 min. The primers introduced 5′ BamHI and 3′ HindIII restriction sites for cloning. The primer sets used were: (1) KdpE (BTH_I1025) (F) GATCGGATCCGAATGCCCATGAGTGAACCGACCGTCACC and (R) GATCAAGCTTTCAGCCCGCGCCGACGAGCCGGTAGCC, (2) PhoB (BTH_I1267) (F) GATCGGATCCATGCCCAGCAACATTCTCGTCATCGAA and (R) GATCAAGCTTTTACGCGTGTTTCGCGAGCCGGTA, (3) OmpR (BTH_I2094) (F) GATCGGATCCATGGAAACGAAAAACCCCTCCAAG, and (R) GATCAAGCTTTCAGGCCGCGCCGTCGGGGATGAA (4) NarL (BTH_I1849) (F) GATCGGATCCATGACCATACGCGTACTGTTGATCGAC, and (R) GATCAAGCTTTCAGGCCTCGGCCGGATGCGGCGC, (5) RisA (BTH_I2094) (F) GATCGGATCCATGGAAACGAAAAACCCCTCCAAG and (R) GATCAAGCTTTCAGGCCGCGCCGTCGGGGATGAA, (6) BTH_II2335 (F) GATCGGATCCATGACCACCGTTTCTTCCACGCCCCGC and (R) GATCAAGCTTCTACCGCCTGCGATGCTCCACCGCGAA, and (7) BPSS2315 (F) GATCGGATCCATGACTCCTGCCTCTTCCACGCCCCGC and (R) GATCAAGCTTCTACCGCCTGCGATGCTCGACCGCGAA.

The RR genes were cloned as N-terminal GST fusions into the pGEX-KG vector using T4 ligase (NEB, Ipswich, MA), transformed into BL21 *E. coli* competent cells, and induced for protein expression with 1 mM IPTG for 4 h. Cells were lysed with 1 mg ml^{-1} lysozyme on ice for 30 min, followed by treatment with 10 μg ml^{-1} DNase and 10 mM $MgCl_2$ for an additional 30 min, and centrifugation at 40,000 rpm for 1 h. GST fusion proteins were purified from the cleared supernatants by incubation with agarose beads cross-linked to glutathione for 1 h and eluted with 50 mM Tris–Cl (pH 8), 10 mM reduced glutathione. Protein samples were then dialyzed using a Slide-alyzer cassette (Thermo Scientific, Rockford, Il.) with a 10,000 MW cut-off to remove free glutathione, quantified using the BCA protein assay (Thermo Scientific), and stored at −80 °C in a final concentration of 30% glycerol.

2.2. Gel shift assays

To demonstrate BeFx-mediated enhancement of RR binding, the *pstS* and *nar* promoter regions were PCR-amplified for use as target DNA in gel shift assays. The following primers were used for PCR: (1) *pstS* promoter (F) ATCGGCCGGACAGGCCGG and (R) GAGACCTCCAGTGTGTGA and (2) *nar* promoter (F) GATCGGATCCCGACATCGTGAGACGAAGCCG and (R) GATCAAGCTTGACGATTCTCTCGAGACGAGG. For the *cstA* ($BTH\_II2252^{-156,-130}$) promoter and internal histidine kinase ($BTH\_II2334^{+447,+468}$) gel shift assays, each set of complementary oligonucleotides, (1) *cstA* (F) TGCTACGTAGCGGCCATACGTAGTTCC and (R) GGAACTACGTATGGCCGCTACGTAGCA, (2) BTH_II2334 (F) GGCTACGTGCGCTACGTCTGG and (R) CCAGACGTAGCGCACGTAGCC, and (3) non-specific oligos, (F) CGAGGGAGAATGATCGTTCTACCCTT and (R) AAGGGTAGAACGATCATTCTCCCTCG, was placed in a heat block at 95 °C for 5 min followed by removal of the heat block to the benchtop. The temperature of the heat block was allowed to decrease to room temperature to allow for oligonucleotide annealing.

Binding reactions (20 μl) containing indicated concentrations of GST–RR fusion proteins, 1 μM of target DNA sequences, 100 μM $BeCl_2$, 10 mM NaF, and 2 μl of 10× binding buffer (20 mM Tris–Cl pH 7.5, 0.5 mM EDTA, 5% glycerol, 1 mM DTT, 0.005% Triton X-100, 50 mM NaCl, 5 mM $MgCl_2$, and 2.5 mM $CaCl_2$), were incubated for 30 min at room temperature. Binding reactions were separated on 8% non-denaturing polyacrylamide gels run in 0.5× TBE buffer and visualized with Sybr Green DNA stain (Life Technologies, Grand Island, NY) using a ChemiDoc gel documentation system (Bio-Rad, Hercules, CA).

2.3. Protein-binding microarrays

A minimum of two PBMs were performed for each RR as previously described (Berger and Bulyk, 2009) with modifications. Briefly, microarrays were obtained from Agilent Technologies in a 4×44 K format, AMADID #015681 and #016060 (cat # G2514F). We performed primer extension from a universal 24-mer region to generate a double-stranded microarray platform. GST fusion proteins were diluted to a final concentration of 125 nM in a volume of 175 μl (PBS, 2% milk, 200 μg ml^{-1} BSA, 0.3 μg ml^{-1} salmon testes DNA) in individual chambers of a four chamber gasket coverslip. In addition, we included 2 μM $BeCl_2$, 200 μM NaF and 1× binding buffer in all incubation and buffer washing steps to maintain activation of the RRs and an optimized ionic environment during protein binding to the microarray. Microarrays were scanned (GenePix Pro 4200A, Axon Instruments, Sunnyvale, CA) to detect specific DNA–RR interactions at multiple

laser power settings and raw image files were quantified using GenePix Pro (Axon Instruments). Following data normalization, protein-binding signal intensities were converted to enrichment scores (E-scores). The Seed and Wobble algorithm (Berger and Bulyk, 2009; Berger et al., 2006) was used to construct position weight matrices (PWMs) by converting E-scores to probabilities using a Boltzmann distribution and displayed using the Web-based tool enoLOGOS (Workman et al., 2005) to represent the PBM-derived DNA binding specificities for each RR. In general, our criteria for determining that a RR exhibited specific binding were the observation that at least one 8-mer exhibited an E-score of ≥ 0.44 and that ≥ 5 of the top 20 E-score sequences resemble each other and are easily aligned. Given the degeneracy of the known DNA binding site for NarL, we loosened the first criterion for NarL (E-score 0.406), but NarL did meet the second criterion, in which ≥ 5 of the top 20 E-score sequences were easily aligned.

2.4. Genomic analysis to identify RR DNA binding sites

RR DNA binding sites obtained using PBM-generated PWMs were used as query motifs to search bacterial genomes to identify putative genes regulated by the RRs. The NCBI and Pathema databases were used to search for RR DNA binding sites and to map genomic localization. DNA binding sequence regions are labeled at nucleotide positions either upstream (−) or downstream (+) of the corresponding translational start sites.

3. Results

3.1. Activation of bacterial RRs using beryllofluoride

We cloned, expressed, and purified six N-terminal glutathione S-transferase (GST) fusion constructs to full-length RR orthologs in *B. thailandensis*, including KdpE, NarL, OmpR, and PhoB. Given that these RRs and their respective binding motifs are highly conserved among widely divergent bacterial species (Lozada-Chavez et al., 2006), we expected that *Burkholderia* orthologs of these RRs represent testable candidates to evaluate the use of PBMs for binding site motif analysis. RR activation to bind DNA is dependent on phosphorylation of a conserved aspartate residue by its cognate histidine kinase. In order to activate the RR to form a structure conducive to DNA binding, we incubated the GST–RR fusion proteins with beryllofluoride (BeFx), which forms a stable analog of the aspartyl phosphate moiety to simulate the phosphorylated form of the protein. BeFx has been previously demonstrated to enhance the downstream activities of some bacterial response regulators, including the ATPase activity of NtrC and binding of FliM peptides by CheY (Yan et al., 1999). To demonstrate that BeFx stimulated the DNA binding properties of RRs, we performed gel shift assays using GST–PhoB binding of the promoter region of *pstS*, which contains the *pho* box consensus sequence shared by regulatory regions in the phosphate regulon (Makino et al., 1988) (Fig. 1). Upon addition of $BeCl_2$ and NaF, the *pstS* promoter fragment bound to GST–PhoB and was shifted by increasing concentrations of GST–PhoB starting at 2 μM (Fig. 1, lanes 7–10). Notably, there is a concomitant disappearance of the unbound *pstS* promoter fragment, even upon incubation with 2 μM GST–PhoB, the lowest protein concentration analyzed. In contrast, we observed a markedly weaker shift of the *pstS* promoter fragment in the absence of BeFx, particularly detectable at 2 μM (Fig. 1, lanes 3–6), in which unbound *pstS* DNA was still present in the presence of 6 μM GST–PhoB. This observation is consistent with previous studies in which unphosphorylated RRs at high concentrations are capable of binding to target DNA (Nakashima et al., 1993). The non-specific DNA control from the *nar* promoter did not exhibit any shift at any GST–PhoB concentration. In addition to PhoB, we also observed BeFx-enhanced binding of KdpE and NarL response regulators to specific DNA targets (data not shown).

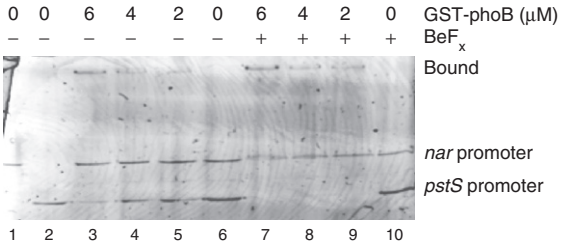

Fig. 1. Activation of bacterial response regulators using beryllofluoride. Specific binding between GST–PhoB and promoter sequence from *pstS* was analyzed by polyacrylamide gel shift analysis in the presence and absence of BeFx. A target sequence from *pstS*, which contains a pho box, and the non-specific *nar* sequence from the nar promoter were examined. GST–PhoB concentrations were increased from 0 to 6 μM with nucleic acid concentrations at 1 μM. Lane numbers are depicted at the bottom of the gel for reference in the Results section.

3.2. PBM analysis of Burkholderia RRs with known consensus binding site sequences

The PBM was designed based on a de Bruijn sequence of order 10, so that all possible 10-mer DNA sequence variants are represented once in an overlapping manner within ~44,000 probes (Berger et al., 2006). All contiguous and gapped 8-mers, including 4-gap-4, are each covered 32 times within the probe set (Berger and Bulyk, 2009), allowing comprehensive binding site measurements. To determine DNA binding specificities, we applied purified GST–RRs to PBMs in the presence of BeFx and subsequently detected DNA binding using a fluorescently-labeled anti-GST antibody. We used the Seed and Wobble algorithm to identify high-affinity binding site motifs and position weight matrices (PWMs) with the highest enrichment scores (E-score) based on normalized fluorescence intensity measurements (Berger and Bulyk, 2009; Berger et al., 2006) (Fig. 2A). For KdpE, we identified two nearly direct repeat 8-mer binding sites, TTTTTACA and TTTTTATA, with E-scores of 0.478 and 0.469, respectively. DNase I footprinting analysis of the *E. coli kdpFABC* promoter using *E. coli* KdpE had previously identified a 23-mer DNA sequence that contains both 8-mer sequences (Sugiura et al., 1992), thus validating the PBM results with *B. thailandensis* KdpE. In the absence of BeFx, we did not recover the two 8-mer motifs or any other conserved sequences with E scores >0.45, indicating that BeFx is required for KdpE DNA target binding. Although the 23-mer KdpE binding sequence in the *kdpFABC* operon is conserved between *E. coli* and *B. thailandensis* (Fig. 2A), the *B. thailandensis* sequence contains only the intact TTTTTACA sequence, not TTTTTATA, suggesting that there may be subtle differences in regulation of the *kdp* operon between *B. thailandensis* and *E. coli*. This PBM analysis further delimits the KdpE binding site from a 23-mer site from the *E. coli* studies to a minimal 8-mer sequence.

We next analyzed the RR NarL, a nitrate responsive activator and repressor of anaerobic respiratory gene expression, using PBMs. From DNAse I footprinting, phosphorylated full-length NarL and the NarL DNA binding domain were previously shown to bind to a 20-mer oligonucleotide sequence containing a tail-to-tail palindromic arrangement (**TACCCAT**TA**ATGGGTA**) in *E. coli* (Maris et al., 2002). The proposed degenerate consensus binding sequence for NarL is a heptameric sequence TACYYMT (where Y = C or T, M = A or C) (Baikalov et al., 1996; Buckler et al., 2002; Dickerson, 1998; Fraenkel et al., 1998). PBM analysis of NarL binding sites in the presence of BeFx revealed multiple sequences containing ACCCA or ACCCC (Fig. 2B), which overlap with the TACYYMT consensus heptamer. Despite the degeneracy of the NarL heptamer, we did not observe other binding site configurations in the top E-scores. Although the ACCCA and ACCCC sequences exhibited the highest binding affinity in the PBM, the top E-score was 0.406, a relatively low score compared to that

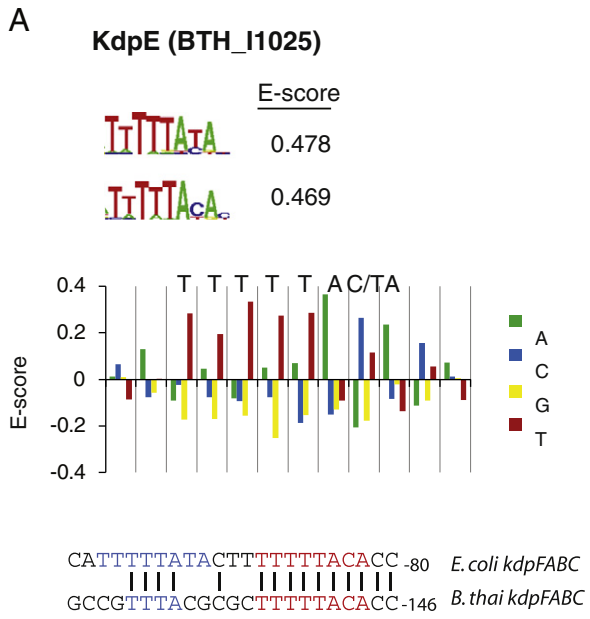

Fig. 2. PBM analysis of response regulators with known consensus binding sites. (A) The E-scores from PBM analysis of KdpE were averaged between two independent experiments and the resultant top two PWMs are shown using enoLOGOS. Use of the Seed-and-Wobble method to determine positional nucleotide preference based on the top 8-mer as a seed is displayed as a histogram. The KdpE promoter binding site is conserved between *E. coli* and *B. thailandensis* upstream of *kdpFABC*. Nucleotide positions are marked relative to the translational start sites of the adjacent gene. (B) The E-scores from PBM analysis of NarL were averaged between two independent experiments and the resultant top PWM is shown using enoLOGOS. The degenerate consensus sequence for NarL binding is shown for comparison.

of KdpE. Nevertheless, the ACCCA and ACCCC sequences clearly exhibited the highest binding affinity among the possible NarL binding site configurations and were preferred in the limited 10-mer sequence space of this PBM design.

We also used PBMs to analyze PhoB and OmpR binding in the presence of BeFx. However, we did not recover the expected canonical PhoB (CTGTAC/CTGTCA) (Okamura et al., 2000) or OmpR (GXXAC) binding motifs (Kenney, 2002). E-scores for these two RRs were less than 0.4, with very little consensus among the top 10-mer sequences (data not shown). GST–PhoB protein had exhibited specific DNA binding to its target sequence by gel shift (Fig. 1), indicating that PhoB was properly folded and functionally active. Response regulators in the OmpR and PhoB families have been shown to dimerize upon activation and bind to direct repeat nucleotide segments with four or five spacer length nucleotides (de Been et al., 2008). It is possible that dimerization or cooperativity in PhoB and OmpR is required for high-affinity DNA binding and thus, RR–DNA binding could not be easily detected in the 10-mer PBM format.

3.3. Identification of RisA DNA binding site motifs within Burkholderiales

Although PBM analysis of OmpR did not lead to identification of the GXXAC consensus sequence, we examined the binding site sequences of RisA (BTH_I2094), a structurally-related response regulator to OmpR with 70% nucleotide homology (Stenson et al., 2005). PBM analysis of RisA revealed TGTAACA and TGTTACA binding site motifs with a top E-score of 0.448 (Fig. 3A). Interestingly, the GXXAC motif occurs in both of these binding sites. To identify potential genes regulated by RisA, we examined intergenic sequences in the genomes of bacterial order *Burkholderiales* using the two high-affinity binding site motifs as the query sequences. We identified a direct repeat TGTAAC upstream of RisA itself in *B. thailandensis*, separated by four nucleotides, indicating the possibility of self-regulation by RisA (Fig. 3B). We also identified a degenerate direct TGTT/AA repeat upstream of an independent two-component histidine kinase and RR pair (BTH_I0210/BTH_I0209). An imperfect TGTA/GA repeat upstream of a RisA ortholog was also found in *Bordetella pertussis*, a member of the *Burkholderiales* order. The *B. pertussis* RisA had previously been shown to regulate gene expression of the *vrg* virulence response genes. Mutation and gel shift analyses of the *vrg6* and *vrg18* promoters had established a consensus sequence of AAATG/TTA for RisA binding (Croinin et al., 2005). Sequence analysis of the intergenic space upstream of vrg6 revealed the sequence AAATGTAAC, in which the last six nucleotides overlap with the high-affinity binding sequence identified using the PBM (Fig. 3B).

3.4. Comparison of DNA binding specificities between an orthologous pair of RRs in Burkholderia spp.

Given the success in identifying credible DNA binding motifs for *Burkholderia* orthologs of characterized RRs, we applied the PBM methodology to an unknown RR from the NarL/LuxR family, at gene locus BTH_II2335. PBM analysis yielded a high-affinity palindromic binding sequence CTACGTAG, E-score 0.484, in the presence of beryllium activation (Fig. 4A). To compare DNA binding specificities and provide insight into conservation of gene regulation among *Burkholderia* species, we examined BPSS2315 in *B. pseudomallei*, exhibiting ~96% nucleotide and amino acid identities to its *B. thailandensis* BTH_II2335 ortholog. Upon beryllium activation, the *B. pseudomallei* BPSS2315 ortholog also displayed high binding affinity to the CTACGTAG sequence, with an E-score 0.467 (Fig. 4A). Analysis of overlapping sequences with the primary CTACGTAG binding site revealed specific nucleotide and positional dependencies. For example, the internal 6-mer sequence, TACGTA, appeared in the top 10 binding sequences for BPSS2315, whereas the first 'C' nucleotide and last 'G' nucleotide of the full 8-mer motif were preferred in four of ten and seven of ten sequences, respectively. Comparison of binding sequences for BPSS2315 and BTH_II2335 demonstrated a preference by both RRs to bind the 6-mer internal TACGTA sequence, with markedly decreased affinity upon changing any nucleotide of the internal core, for example the last A (Fig. 4B).

To identify genes potentially regulated by the BTHII2335 and BPSS2315 RRs, we searched the *B. thailandensis* genome for the CTACGTAG motif. We found putative binding sites at nucleotide positions +449 and +458 within the cognate histidine kinase (BTH_II2334) adjacent to BTH_II2335, suggesting co-regulation of this histidine kinase–RR pair. The binding site within BTH_II2334 contains a direct repeat of CTACGT, with a spacer of three nucleotides. Putative sites were also found at −154 and −140 nucleotides upstream of the translational start site of *cstA*, a carbon starvation gene (BTH_II2252), which exhibits ~40% amino acid homology to NarL (Schmitt, 1999) (Fig. 4C). The site upstream of the *cstA* gene contains an imperfect repeat of the CTACGTAG sequence separated by 6 nucleotides. RRs have been shown to bind to direct or inverted nucleotide repeat segments with a spacer region of 2 to 11 nucleotides (de Been et al., 2008). Gel shift analysis demonstrated that the BPSS2315 RR can specifically bind to both target sequences, BTH_II2334$^{(+447,+468)}$ and *cstA*$^{(-156,-130)}$ (Fig. 4C). The BTH_II2335 RR also demonstrated binding to *cstA*$^{(-156,-130)}$ by gel

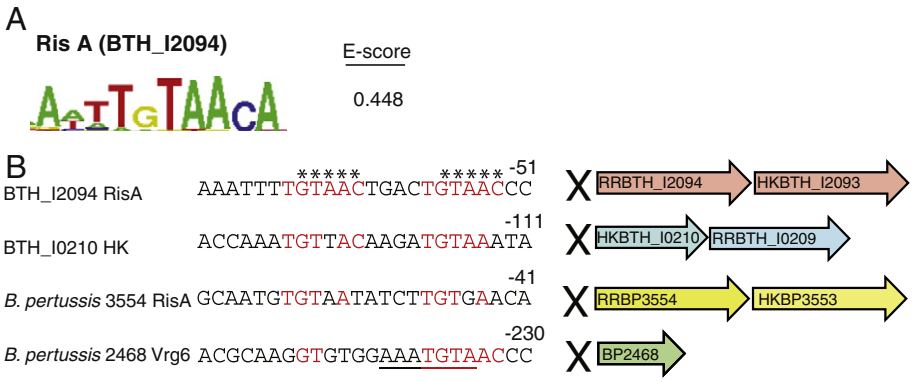

Fig. 3. PBM analysis of RisA and predicted DNA binding sequences identified in the bacterial order *Burkholderiales*. (A) E-scores from PBM analysis of RisA were averaged between two independent experiments and the resultant top PWM is shown using enoLOGOS. (B) Sequence alignments of predicted DNA binding sequences of RisA in *B. thailandensis* and *B. pertussis* are listed. Nucleotide positions are marked relative to the translational start sites of adjacent genes. Nucleotides highlighted in red overlap with target sequences identified from PBM analysis of RisA. The underlined sequence is a prediction of the RisA DNA binding site from mutational promoter studies for Vrg6 and Vrg18 (Croinin et al., 2005). Asterisks denote common GXXAC motif found in related OmpR response regulator target binding sites. Pictorial representations of gene organization are displayed with a black X to denote relative position of putative binding sites for the RR.

shift analysis (Fig. 4C), but did not exhibit detectable binding with BTH_II2334$^{(+447,+468)}$ (data not shown).

4. Discussion

Two-component signal transduction pathways in bacteria enable the rapid adaptation of cellular physiology in response to specific environmental cues. Bacteria with large genomes or those that exhibit broad metabolic versatility encode a larger number of two-component systems in order to couple a multitude of extracellular stimuli to complex intracellular responses. In this study, we evaluated the use of PBMs to couple identification of consensus binding motifs for *B. thailandensis* RRs to sequence analyses of bacterial genomes for prediction of candidate target genes regulated by specific RRs. Characterization of conserved RRs and their cognate binding sites across the *Burkholderia* genus or more distantly-related bacterial species, can provide molecular

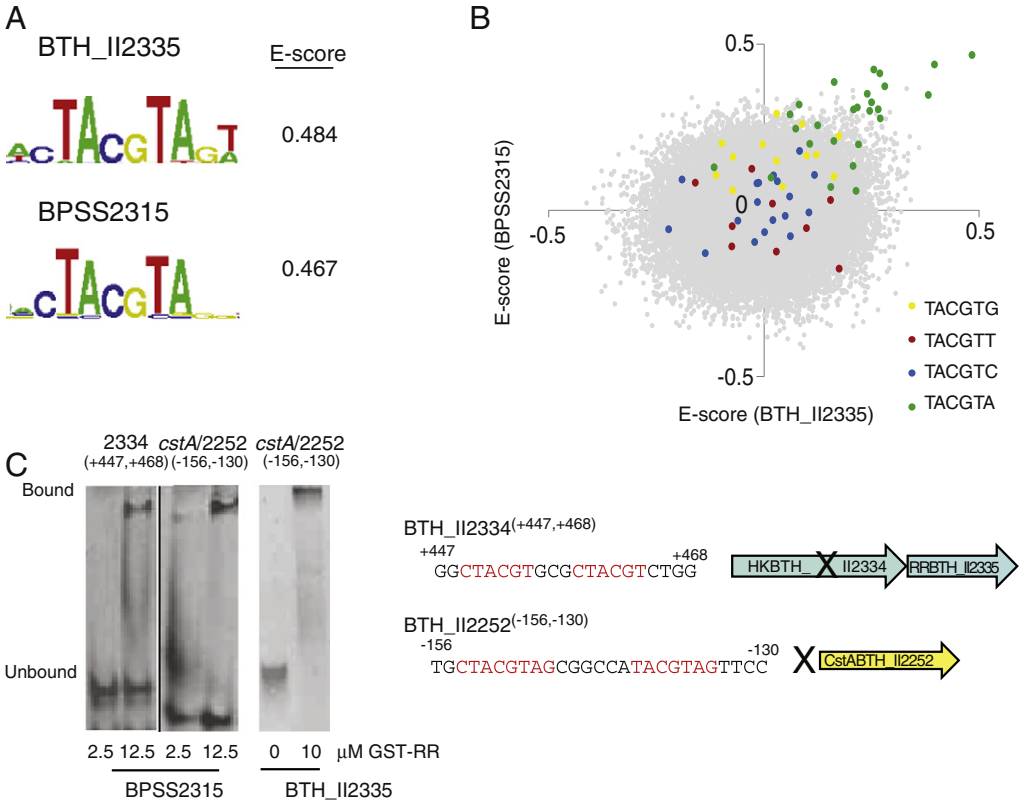

Fig. 4. Comparison of DNA binding specificities between orthologous response regulators. (A) E-scores from PBM analysis of BTH_II2335 (*B. thailandensis*) and BPSS2315 (*B. pseudomallei*) were averaged between two independent experiments and the resultant top PWMs are shown using enoLOGOS. (B) Scatter plot of ungapped 8-mers highlighting sequences containing the referenced 6-mers. The plot displays nucleotide preference at position six of the internal core 6-mer TACGTA consensus site for BPSS2315 and BTH_III2335. (C) Gel shift analysis was performed to detect binding between BTH_II2335 and BPSS2315 RRs and predicted binding sequences. The BTH_II2334$^{(+447,+468)}$ sequence is located within the histidine kinase coding region of BTH_II2334, upstream of its cognate response regulator BTH_II2335. *cstA*$^{(-156,-130)}$ is located in intergenic space upstream of the CstA carbon starvation response regulator. Oligonucleotides (1 µM) were incubated with 2.5 µM and 12.5 µM of GST–RR (BPSS2315) and 10 µM of GST–RR (BTH_II2335). Pictorial representations of gene organization are displayed with a black X to denote relative position of putative binding sites for the RR. Nucleotide position upstream (−) or downstream (+) of translational start sites of adjacent genes is noted.

insights into common signal transduction pathways and response networks to environmental stimuli in the bacterial kingdom.

We successfully discovered high-affinity DNA binding sites for five *Burkholderia* response regulators, including two from the OmpR RR family (KdpE and RisA) and three from the NarL/LuxR RR family (NarL and previously-uncharacterized orthologs BTH_II2335 and BPSS2335, from *B. thailandensis* and *B. pseudomallei*, respectively). However, we did not recover the known binding motifs for PhoB and OmpR using the PBM assays. There are potential limitations to the use of this 10-mer PBM platform for identification of RR binding sites. Many bacterial transcription factors dimerize upon activation and may require a longer target site than 10-mers for high-affinity binding. For example, the winged helix-turn-helix (wHTH) DNA-binding domain of OmpR/PhoB family members typically binds to DNA tandem repeats spanning more than 20 bp regions (Gao et al., 2007). In OmpR family members, the wing domains demonstrate great variability in protein–DNA contact, possibly accounting for degeneracy of DNA sequence binding, which introduces further complexity to the deconvolution of high-affinity binding sites using PBM analysis (Kenney, 2002). Furthermore, bacterial DNA-binding proteins may require as-yet uncharacterized metabolite or protein cofactors, normally available *in vivo*, for optimal binding affinity and specificity (Wall et al., 2004). It is also possible that some GST–RRs will not be expressed in an active form to efficiently bind target DNA or that the RRs have a fast off-rate of DNA binding. Thus, it is likely that this PBM design can be used to characterize DNA binding motifs for a subset of bacterial response regulators.

Comparison of DNA binding specificities between transcription factor orthologs can provide insights into the conservation of regulatory DNA–protein interactions between species. We found that the closely related RR orthologs, BTH_II2335 and BPSS2315, shared a conserved high-affinity binding site, CTACGTAG. This 8-mer sequence motif is a palindrome, which is characteristic of the DNA binding motifs of LuxR family response regulators (Pontes et al., 2008). Interestingly, identified gene targets for BTH_II2335 and BPSS2315 include the cognate histidine kinase (BTH_II2334) for BTH_II2335 and *cstA*, a carbon starvation gene that exhibits ~40% amino acid homology to the NarL family of RRs (Schmitt, 1999). These findings, in addition to potential RisA gene targets upstream of the RisA gene itself and an independent histidine kinase (BTH_I0210) indicate that a primary role of bacterial RRs may be to act as an autoregulator to modulate expression of self and other two-component proteins in response to external stimuli (Bijlsma and Groisman, 2003; Haydel et al., 2002; Tzeng et al., 2012).

Other studies have reported overlap of DNA binding specificity between orthologs that are less conserved than the *Burkholderia* orthologs we examined. The AP2 family of transcription factors from the distantly-related Apicomplexan species, *Cryptosporidium parvum* and *Plasmodium falciparum*, shares only 47% sequence identity (68% similarity), but still exhibits highly similar DNA-binding specificities (De Silva et al., 2008). Study of mouse transcription factors using PBMs indicates that orthologs with up to 67% amino acid sequence identity exhibited a common primary binding sequence, but distinct secondary DNA binding profiles. For example, the Irf4 and Irf5 transcription factors bind with highest affinity to 8-mers containing CGAAAC, but have different lower affinity sites (TGAAAG versus CGAGAC, respectively) (Badis et al., 2009). For the *Burkholderia* orthologs, we do note that BPSS2315 formed stable complexes with both predicted target sequences in BTH_II2334 and *cstA*, whereas BTH_II2335 bound to cstA, but not BTH_II2334. These differences in DNA binding specificity among orthologs suggest a model in which the evolution of gene regulation networks may lead to a gradual diversification of regulatory protein binding site preferences.

Other complementary technologies have been developed to perform high-throughput analysis of transcription factor binding sites. In particular, Chromatin Immunoprecipitation (ChIP) followed by sequence analysis, such as whole genome microarray detection (ChIP-chip) or whole genome sequencing (Chip-seq), has enabled the *in vivo* identification of transcription factor binding sites. For example, the atypical OmpR/PhoB response regulator Chx from *Chlamydia trachomatis* was found to act as an autoregulator and bind to its own promoter *in vivo*, in addition to other genomic sites (Hickey et al., 2011). Whereas the ChIP methodology provides transcription factor–DNA interactions under the specific cellular conditions used in the experiment, PBMs can provide data on the entire range of high-affinity to no affinity binding sites. A recent study employed a different *in vitro* genome-wide method, the DNA-affinity-purified-chip (DAP-chip) to map two-component RR binding sites in the sulfate-reducing bacteria, *Desulfovibrio vulgaris*. This method used His-tagged RRs activated by acetyl phosphate to affinity-purify sheared *D. vulgaris* genomic DNA and identify gene targets by whole genome amplification and hybridization to a *D. vulgaris* tiling array (Rajeev et al., 2011). Functional predictions were determined for 24 RRs in varied cell functions, such as energy metabolism, biofilm formation, and nutrient starvation. We expect that these different methodologies will serve to validate the expanding analysis of endogenous DNA binding sites for bacterial transcription factors.

Collectively, our studies support a model in which RRs autoregulate or regulate expression of other histidine kinase–RR pairs, suggesting that two component systems have evolved feedback loops to control signaling inputs from the environment. A systematic method to rapidly determine DNA binding sites of prokaryotic response regulators can greatly enhance our capability to predict target gene regulatory networks under different environmental and stress conditions across multiple bacteria species. Our demonstration that the PBM methodology can be applied to bacterial response regulators using beryllofluoride activation expands the potential applications of PBMs to analysis of bacterial two component systems and other transcription factors that require small molecule activation. Future research will be directed at analysis of differential DNA binding specificity in response to metabolite binding in bacterial pathogens.

Acknowledgments

This work was supported by a Los Alamos National Laboratory Directed Research and Development project (20080138DR). All authors state that there are no conflicts of interest.

References

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., et al., 2009. Diversity and complexity in DNA recognition by transcription factors. Science 324, 1720–1723.

Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P., Dickerson, R.E., 1996. Structure of the *Escherichia coli* response regulator narL. Biochemistry 35, 11053–11061.

Beier, D., Gross, R., 2006. Regulation of bacterial virulence by two-component systems. Curr. Opin. Microbiol. 9, 143–152.

Berger, M., Bulyk, M., 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. 4, 393–411.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep III, P.W., Bulyk, M.L., 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. 24, 1429–1435.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., et al., 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell 133, 1266–1276.

Bijlsma, J., Groisman, E., 2003. Making informed decisions: regulatory interactions between two-component systems. Trends Microbiol. 11, 359–366.

Bourret, R.B., Borkovich, K.A., Simon, M.I., 1991. Signal transduction pathways involving protein phosphorylation in prokaryotes. Annu. Rev. Biochem. 60, 401–441.

Buckler, D.R., Zhou, Y., Stock, A.M., 2002. Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from *Thermotoga maritima*. Structure 10, 153–164.

Cheung, J., Hendrickson, W.A., 2010. Sensor domains of two-component regulatory systems. Curr. Opin. Microbiol. 13, 116–123.

Croinin, T., Grippe, V., Merkel, T., 2005. Activation of the vrg6 promoter of *Bordetella pertussis* by RisA. J. Bacteriol. 187, 1648–1658.

de Been, M., Bart, M., Abee, J., Siezen, R., Francke, C., 2008. The identification of response regulator-specific binding sites reveals new roles of two-component systems in *Bacillus cereus* and closely related low-GC Gram-positives. Environ. Microbiol. 10, 2796–2809.
De Silva, E., Gehrke, A., Olszewski, K., Leon, I., Chahal, J., Bulyk, M., et al., 2008. Specific DNA-binding by Apicomplexan AP2 transcription factors. Proc. Natl. Acad. Sci. U. S. A. 105, 8393–8398.
Dickerson, R.E., 1998. DNA bending: the prevalence of kinkiness and the virtues of normality. Nucleic Acids Res. 26, 1906–1926.
Fraenkel, E., Rould, M.A., Chambers, K.A., Pabo, C.O., 1998. Engrailed homeodomain DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures. J. Mol. Biol. 284, 351–361.
Gao, R., Mack, T., Stock, A., 2007. Bacterial response regulators: versatile regulatory strategies from common domains. Trends Biochem. Sci. 32, 225–234.
Gasell, M., Altendorf, K., 2001. Analysis of KdpC of the K(+)-transporting KdpFABC complex of *Escherichia coli*. Eur. J. Biochem. 6, 1772–1781.
Gordon, B.R., Li, Y., Cote, A., Weirauch, M.T., Ding, P., Hughes, T.R., et al., 2011. Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. Proc. Natl. Acad. Sci. U. S. A. 108, 10690–10695.
Haydel, S., Benjamin Jr., W., Dunlap, N., Clark-Curtiss, J., 2002. Expression, autoregulation, and DNA binding properties of the *Mycobacterium tuberculosis* TrcR response regulator. J. Bacteriol. 184, 2192–2203.
Hickey, J.M., Weldon, L., Hefty, P.S., 2011. The atypical OmpR/PhoB response regulator ChxR from *Chlamydia trachomatis* forms homodimers *in vivo* and binds a direct repeat of nucleotide sequences. J. Bacteriol. 193, 389–398.
Kenney, L., 2002. Structure/function relationships in OmpR and other winged-helix transcription factors. Curr. Opin. Microbiol. 5, 135–141.
Laub, M.T., Goulian, M., 2007. Specificity in two-component signal transduction pathways. Annu. Rev. Genet. 41, 121–145.
Lozada-Chavez, I., Janga, S.C., Collado-Vides, J., 2006. Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Res. 34, 3434–3445.
Maity, T., Close, D., Valdez, Y., Nowak-Lovato, K., Marti-Arbona, R., Nguyen, T., et al., 2011. Discovery of DNA operators for TetR and MarR family transcription factors from *Burkholderia xenovorans*. Microbiology 158, 571–582.
Makino, K., Shinagawa, H., Amemura, M., Kimura, S., Nakata, A., Ishihama, A., 1988. Regulation of the phosphate regulon of *Escherichia coli*. Activation of pstS transcription by PhoB protein *in vitro*. J. Mol. Biol. 203, 85–95.
Maris, A.E., Sawaya, M.R., Kaczor-Grzeskowiak, M., Jarvis, M.R., Bearson, S.M.D., Kopka, M.L., et al., 2002. Dimerization allows DNA target site recognition by the NarL response regulator. Nat. Struct. Biol. 9, 771–778.
Mukherjee, S., Berger, M.F., Jona, G., Wang, X.S., Muzzey, D., Snyder, M., et al., 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat. Genet. 36, 1331–1339.
Nakashima, K., Sugiura, A., Kanamaru, K., Mizuno, T., 1993. Signal transduction between the two regulatory components involved in the regulation of the kdpABC operon in *Escherichia coli*: phosphorylation-dependent functioning of the positive regulator, KdpE. Mol. Microbiol. 7, 109–116.
Okamura, H., Hanaoka, S., Nagadoi, A., Makino, K., Nishimura, Y., 2000. Structural comparison of the PhoB and OmpR DNA-binding/transactivation domains and the arrangement of PhoB molecules on the phosphate box. J. Mol. Biol. 295, 1225–1236.
Pao, G.M., Saier Jr., M.H., 1995. Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. J. Mol. Evol. 40, 136–154.
Pompeani, A., Irgon, J., Berger, M., Bulyk, M., Wingreen, N., Bassler, B.L., 2008. The *Vibrio harveyi* master quorum-sensing regulator, LuxR, a TetR-type protein is both an activator and a repressor: DNA recognition and binding specificity at target promoters. Mol. Microbiol. 70, 76–88.
Pontes, M., Babst, M., Lochhead, R., Oakeson, K., Smith, K., Dale, C., 2008. Quorum sending primes the oxidative stress response in the insect endosymbiont, *Sodalis glossinidius*. PLoS One 3, e3541.
Pratt, L.A., Silhavy, T.J., 1995. Identification of base pairs important for OmpR–DNA interaction. Mol. Microbiol. 17, 565–573.
Rajeev, L., Luning, E., Dehal, P., Price, M., Arking, A., Mukhopadhyay, A., 2011. Systematic mapping of two component response regulators to gene targets in a model sulfate reducing bacterium. Genome Biol. 12.
Robasky, K., Bulyk, M.L., 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein–DNA interactions. Nucleic Acids Res. 39, D124–D128.
Schmitt, M., 1999. Identification of a two-component signal transduction system from *Corynebacterium diphtheriae* that activates gene expression in response to the presence of heme and hemoglobin. J. Bacteriol. 181, 5330–5340.
Stenson, T.H., Allen, A.G., al-Meer, J.A., Maskell, D., Peppler, M.S., 2005. *Bordetella pertussis* risA, but not risS, is required for maximal expression of bvg-repressed genes. Infect. Immun. 73, 5995–6004.
Sugiura, A., Nakashima, K., Tanaka, K., Mizuno, T., 1992. Clarification of the structural and functional features of the osmoregulated kdp operon of *Escherichia coli*. Mol. Microbiol. 6, 1769–1776.
Tzeng, Y., Zhou, X., Bao, S., Zhao, S., Noble, C., Stephens, D., 2012. Autoregulation of the MisR/MisS two-component signal transduction system in *Neisseria meningitidis*. J. Bacteriol. 188, 5055–5065.
Ulrich, L., Zhulin, I., 2010. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res. 38, D401–D407.
Wall, M.E., Hlavacek, W.S., Savageau, M.A., 2004. Design of gene circuits: lessons from bacteria. Nat. Rev. Genet. 5, 34–42.
Workman, C., Yin, Y., Corcoran, D., Ideker, T., Stormo, G., Benos, P.V., 2005. enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Res. 33, W389–W392.
Yan, D., Cho, H., Hastings, C., Igo, M., Lee, S., Pelton, J.G., et al., 1999. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. Proc Natl Acad Sci USA 96, 14789–14794.
Yang, X.F., Ji, Y., Schneider, B.L., Reitzer, L., 2004. Phosphorylation-independent dimer–dimer interactions by the enhancer-binding activator NtrC of *Escherichia coli*: a third function for the C-terminal domain. J. Biol. Chem. 279, 36708–36714.