Identification of a core sequence for the binding of BosR to the rpoS promoter region in *Borrelia burgdorferi*

Zhiming Ouyang, Jianli Zhou, Chad A. Brautigam, Ranjit Deka and Michael V. Norgard

1Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

2Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

The alternative sigma factor RpoS in *Borrelia burgdorferi* plays a central role in modulating host adaptive responses when spirochaetes cycle between ticks and mammals. The transcriptional activation of $\sigma^{54}$-dependent rpoS requires a Fur homologue designated BosR. Previously, BosR was shown to directly activate rpoS transcription by binding to the rpoS promoter. However, many other DNA binding features of BosR have remained obscure. In particular, the precise DNA sequence targeted by BosR has not yet been completely elucidated. The prediction of a putative Per box within the rpoS promoter region has further confounded the identification of the BosR binding sequence. Herein, by using electrophoretic mobility shift assays, we demonstrate that the putative Per box predicted in the rpoS promoter region is not involved in the binding of BosR. Rather, a 13 bp palindromic sequence (ATTTAANTTAAAT) with dyad symmetry, which we denote as the 'BosR box', functions as the core sequence recognized by BosR in the rpoS promoter region of *Borrelia burgdorferi*. Similar to a Fur box and a Per box, the BosR box probably comprises a 6–1–6 inverted repeat composed of two hexamers (ATTTAA) in a head-to-tail orientation. Selected mutations in the BosR box prevented recombinant BosR from binding to rpoS. In addition, we found that sequences neighbouring the BosR box also are required for the formation of BosR–DNA complexes. Identification of the BosR box advances our understanding of how BosR recognizes its DNA target(s), and provides new insight into the mechanistic details behind the unique regulatory function of BosR.

INTRODUCTION

*Borrelia burgdorferi*, the causative agent of Lyme disease, is transmitted to mammalian hosts through the bite of a tick vector (*Ixodes scapularis*) (Burgdorfer et al., 1982; Steere et al., 1983). To establish infection, *Borrelia burgdorferi* must adapt to and interact with these two markedly different mammalian and tick milieus. Host adaptation of *Borrelia burgdorferi* is achieved by dramatic changes in gene expression mediated by transcriptional regulators in response to various tick or mammalian stimuli (Antonara et al., 2011; Brooks et al., 2003; Coburn, 2001; He et al., 2011; Kung et al., 2013; Liang et al., 2002, 2004; Norris, 2006a, b, 2012; Pal & Fikrig, 2003; Parveen et al., 2003; Radolf et al., 2012; Roberts et al., 2002; Rosa et al., 2005; Samuels, 2011; Samuels & Radolf, 2010; Scheckelhoff et al., 2007; Troy et al., 2013). Among potential regulators encoded by *Borrelia burgdorferi*, the alternative sigma factor RpoS ($\sigma^5$) acts as a ‘gatekeeper’ during the transmission of spirochaetes from tick to mammals (Dunham-Emms et al., 2012; Hübner et al., 2001; Radolf et al., 2012; Samuels, 2011; Samuels & Radolf, 2010). When *Borrelia burgdorferi* resides in an unfed tick midgut, the expression of rpoS is repressed (Dunham-Emms et al., 2012; Ouyang et al., 2012; Radolf et al., 2012; Samuels, 2011; Samuels & Radolf, 2010). Upon sensing appropriate environmental signals during blood feeding by the tick, *Borrelia burgdorferi* produces RpoS, which, in turn, promotes the expression of virulence-associated outer surface lipoproteins such as outer surface protein C (OspC) (Grimm et al., 2004; Pal et al., 2004; Schwan et al., 1995; Seemanapalli et al., 2010). Spirochaetes expressing OspC migrate from the tick midgut to the salivary glands, and are transmitted to mammalian tissues, thereby establishing infection. The transcription of
rpoS is dependent on another alternative sigma factor, $\sigma^{24}$ (RpoN), which binds to a canonical $-24l/-12$ promoter upstream of the rpoS coding region (Burtnick et al., 2007; Fisher et al., 2005; Hübner et al., 2001; Lybecker & Samuels, 2007; Smith et al., 2007). Activation of $\sigma^{24}$-dependent rpoS transcription requires two additional key regulators: a putative bacterial enhancer-binding protein Rrp2 (BB0763) (Blevins et al., 2009; Boardman et al., 2008; Burtnick et al., 2007; Groschong et al., 2012; Ouyang et al., 2008; Yang et al., 2003), and a putative global regulator (BB0647) that has been named BosR (Boylan et al., 2003; Hyde et al., 2009, 2010; Katona et al., 2004; Ouyang et al., 2009, 2011).

BosR was originally annotated as a homologue of the bacterial ferric uptake regulator (Fur), or the peroxide stress response regulator PerR (Fraser et al., 1997). In a wide variety of bacterial pathogens, Fur acts as a global regulator to control iron homeostasis and many other essential functions (Carpenter et al., 2009; Lee & Helmann, 2007). As a member of the Fur family, PerR functions as a peroxide-sensing transcriptional regulator to modulate the bacterial oxidative stress response (Duarte & Latour, 2010; Lee & Helmann, 2007). Fur and PerR regulate gene transcription through binding to a consensus 19 bp Fur box (GATAATGATAATATTAC) or a 15 bp Per box (TTATAATTTATAA), respectively. In general, the Fur box is annotated as two $7-1-7$ heptamer (TGATAAT) inverted repeats, a $9-1-9$ inverted repeat (GATAATGAT) or three hexamer (GATAA) repeats (Baichoo & Helmann, 2002; Escolar et al., 1999; Fuangthong & Helmann, 2003; Lavrrar & McIntosh, 2003). The Per box, occupied by one PerR dimer, has been recognized as a $7-1-7$ inverted repeat composed of heptamers of TTATAAT in a head-to-tail orientation (Baichoo & Helmann, 2002; Fuangthong & Helmann, 2003).

Because it has been proposed that *Borrelia burgdorferi* does not require iron to support its growth (Posey & Gherardini, 2000), it remains unlikely that BosR is involved in modulating iron uptake in this pathogen. Rather, data from several recent studies (Hyde et al., 2009, 2010; Ouyang et al., 2009, 2011; Rudolf et al., 2012; Samuels, 2011; Samuels & Radolf, 2009) have convincingly demonstrated that BosR functions as a key activator to induce the central RpoN–RpoS regulatory pathway. By examining gene expression in isogenic bosR mutants, we and others (Hyde et al., 2009; Ouyang et al., 2009) found that bosR encodes a transcriptional regulator that modulates the expression of genes required for *Borrelia burgdorferi* mammalian infection and pathogenesis. In particular, the expression of rpoS and RpoS-dependent virulence factors, such as OspC (Grimm et al., 2004; Pal et al., 2004; Schwan et al., 1995; Seemanapalli et al., 2010) and DhpA (Blevins et al., 2008; Fischer et al., 2003; Ouyang et al., 2010; Shi et al., 2008; Weening et al., 2008), is abolished in bosR mutants. When a bosR mutation was complemented in cis or in trans, the expression of rpoS, ospC and dhpA is restored (Ouyang et al., 2009, 2011). Surprisingly, by using an electrophoretic mobility shift assay (EMSA), we found that BosR bound to a region(s) near the putative rpoS promoter that contains a motif of two pentamer direct repeats (TAAATTAAT), suggesting that BosR directly activates $\sigma^{24}$-dependent rpoS transcription (Ouyang et al., 2011). This represents the first example in any bacterium of $\sigma^{24}$-mediated gene transcription that requires a second accessory molecule (BosR) [in addition to the proposed enhancer-binding protein (Rrp2)] involved in DNA binding. The combined findings thus have revealed a previously unappreciated mechanism for $\sigma^{24}$-mediated gene control in bacteria.

Whereas our finding that BosR directly interacts with DNA to activate rpoS transcription in *Borrelia burgdorferi* is quite compelling, many of the DNA binding features of BosR remain unresolved. Our previous DNase I footprinting assays revealed three unusually large regions (BS1, BS2 and BS3) in rpoS that were protected by BosR (Ouyang et al., 2011). Specifically, BS1, BS2 and BS3 span 57, 75 and 73 bp of DNA, respectively. For the 16.8 kDa Fur or the 16.4 kDa PerR, one protein dimer usually protects an ~30 bp DNA sequence in DNase I footprinting assays (Baichoo & Helmann, 2002; Fuangthong & Helmann, 2003). Thus, it has been difficult to explain why the 20.2 kDa BosR, a protein with similar size to Fur or PerR, was able to interact with such a large DNA fragment. Moreover, although our previous data revealed that the motif of TAAATTAAT is necessary for BosR binding, the precise sequence targeted by BosR remained unknown. This, in turn, impeded an accurate interpretation of the three large BosR binding regions in rpoS, and has hindered the identification of additional BosR target genes. These information gaps also leave open the question of whether BosR binds to the rpoS promoter region via DNA motifs similar to the Fur box or Per box, especially given that (1) BosR was previously reported to interact with these two motifs (Katona et al., 2004) and (2) our recent sequence analyses predicted two putative Per boxes within the rpoS promoter region. To address these questions, we conducted more extensive analyses of the *Borrelia burgdorferi* rpoS promoter region by using EMSAs. Our combined data support that the putative Per box in the rpoS promoter region is not involved in BosR binding. Rather, BosR recognizes the rpoS promoter region via a ‘BosR box’, a 13 bp palindromic sequence consisting of a putative $6-1-6$ inverted repeat.

**METHODS**

**Strains and culture conditions.** Infectious *Borrelia burgdorferi* strain B31-M1 (Johnson et al., 1984) was routinely cultured at 37 °C and 5% CO2 in either BSK-II medium (Pollack et al., 1993) or BSK-H medium (Sigma) supplemented with 6% rabbit serum (Pel-Freeze). Spirochaetes were enumerated by dark-field microscopy. *Escherichia coli* strains were cultured in lysozyme broth (LB) supplemented with appropriate antibiotics at the following concentrations: kanamycin, 50 μg ml$^{-1}$; or streptomycin, 100 μg ml$^{-1}$.

**Recombinant BosR (rBosR) purification and analyses.** rBosR was purified from *E. coli* harbouring the BosR overexpression plasmid pOY73 as previously described (Coleman et al., 2013; Ouyang et al., 2011; Boardman et al., 2008; Yang et al., 2003; Katona et al., 2004; Smith et al., 2007).
2011). Purified rBosR was analysed using reversed-phase liquid chromatography directly coupled to tandem mass spectrometry (LC/MS/MS) (Protein Chemistry Technology Core, University of Texas Southwestern Medical Center), to verify the molecular mass and the sequence of rBosR. Sedimentation velocity experiments were performed in a Beckman-Coulter model XL-I analytical ultracentrifuge. Three samples were prepared by diluting rBosR to 5, 19 or 52 μM by using buffer A (20 mM Tris, 20 mM NaCl, 100 mM L-arginine, pH 7.5). For rBosR sedimentation, 390 μl of rBosR sample or buffer A was introduced into a dual-sectored, charcoal-filled Epon centerpiece, and centrifuged at 50 000 r.p.m. in an An-50Ti rotor. Absorbance data were collected at a wavelength of 280 nm. The partial specific volume of the proteins was estimated by using the program SEATERP (Lave et al., 1992). The solution density (1.00725 g ml−1) and viscosity (10.001565 Pa s) were measured using a DMA5000 density metre (Anton Paar) and an AMVn viscometer (Anton Paar), respectively. The c(s) distribution for the sedimenting species was generated using the program SEDFIT (Schuck, 2000), and the final c(s) analysis was performed in SEDPHAT. The molar masses of the rBosR monomer and dimer were fixed to their theoretical values (20 206 and 40 412 g mol−1, respectively) in this analysis. The global analysis of three datasets was also accomplished using SEDPHAT (Schuck, 2003). The s-value of the monomer (1.88S), the s-value of the dimer (2.91S), the log of the equilibrium association constant and the log of the kinetic off-rate were fitted using solutions of the Lamm equation that takes the chemical equilibrium and off-rate into account (Schuck, 2003). The χ2-value of the monomer (1.88S), the χ2-value of the dimer (2.91S), the log of the kinetic off-rate were fitted using solutions of the Lamm equation that takes the chemical equilibrium and off-rate into account (Dam et al., 2005; Schuck, 2000). It was also necessary to model approximately 2 % of the signal in each experiment as a large, incompetent species. For both the c(s) analyses and the Lamm equation analysis, the time-invariant noise in the data was accounted for as described by Schuck & Demeler (1999). To obtain the 68.3 % error intervals for the global parameters, the ‘error-surface projection’ method was implemented (Bevington & Robinson, 1992; Houtman et al., 2007), utilizing the F-statistics calculator in SEDPHAT as described by Brautigam (2011).

SDS-PAGE, immunoblotting and EMSAs were carried out as previously described (Ouyang et al., 2011).

RESULTS

rBosR in free solution exists in a monomer–dimer equilibrium

The molecular mass of rBosR, as determined by LC/MS/MS analysis, was 20 205 ± 2 Da (see Fig. S1b, available in the online Supplementary Material), in good agreement with the theoretical value of 20 206 Da. To determine the oligomerization state of rBosR in free solution, the protein was subjected to the sedimentation velocity mode of analytical ultracentrifugation. Our analyses revealed that, in 19 μM solution, rBosR displayed two sedimenting species (Fig. S1c). The more populated species had a sedimentation coefficient (c20,ω) of about 1.9S, whereas the secondary species had an s-value of 2.8S. Given the overall frictional ratio refined for this distribution (~1.3), these two sedimentation coefficients correspond to those of the BosR monomer (1.88S) and dimer (2.91S), respectively. Upon sedimenting at a lower concentration of rBosR (5 μM), the monomer species became more prevalent, whereas a higher concentration (52 μM) favoured the dimer (Fig. S1c). Therefore, rBosR probably exists in an equilibrium between its monomeric and dimeric forms under the conditions studied. Further global analyses of the sedimentation velocity data using SEDPHAT revealed that the dissociation constant of the BosR dimer was 28 μM.

BosR binds the Bacillus subtilis Per box but not the E. coli Fur box

Given that BosR was originally predicted to be a Fur or PerR homologue (Boylan et al., 2003; Fraser et al., 1997; Katona et al., 2004), it remains possible that BosR may regulate gene transcription by binding to consensus Fur or Per boxes. To test this possibility, EMSAs were performed. A 25 bp duplex oligonucleotide probe ZM356 containing the consensus E. coli Fur box, and a 31 bp probe ZM357 containing the consensus Bacillus subtilis Per box (Fig. 1a), were end-labelled with digoxigenin. Of note, these two sequences had been used previously to study the DNA binding activity of rBosR (Katona et al., 2004). Duplex nucleotides representing the BS1 or BS2 sites in the rpoS promoter region were also used in these assays. One duplex nucleotide probe (ZM126) that contains sequences of the Borrelia burgdorferi rpoS promoter region not bound by BosR (Ouyang et al., 2011) was used as a negative control. rBosR was incubated with each probe and protein–DNA complexes were separated on a 5 % native polyacrylamide gel. Digoxigenin-labelled DNA was detected by immunoblot. As expected, BosR bound the probe representing BS1 or BS2, but not the negative control probe ZM126 (Fig. 1b). Moreover, BosR bound the probe ZM357 (Bacillus subtilis Per box), but not ZM356 (E. coli Fur box) (Fig. 1b).

The putative Per box within the rpoS promoter is not recognized by BosR

Our data that BosR bound the Bacillus subtilis Per box but not the E. coli Fur box suggested that BosR may activate rpoS transcription through binding to a Per box in the rpoS promoter region. Consistent with this hypothesis, examination of the rpoS promoter region revealed two putative Per boxes (Fig. 2). The roles of these two sites in the binding of BosR were thus investigated.

The first putative Per box (Per1) is located 11 bp upstream of the rpoS translational start site (i.e. the ATG start codon) (Fig. 2). This 15 bp sequence (ATATAATACTATGA; underlined letters represent bases that are different from the conserved bases in the Per consensus sequence) has 12 out of 15 matches to the consensus Per box consensus, and Per1 comprises a putative 7–1–7 heptameric inverted repeat in a head-to-tail orientation (Fig. S2a). As shown in Fig. S2(b), BosR did not bind the 50 bp probe ZM125 or the 24 bp probe ZM124.

The second putative Per box (Per2) is located within the BosR BS1 binding site (Fig. 2). This 16 bp sequence (TTCAATCAATTACAA; underlined letters represent bases that are different from the conserved bases in the Per consensus sequence) has 12 out of 15 matches to the Per box consensus sequence. Notably, different from the classical 7–1–7 motif of the Per consensus, a 7–2–7 inverted
repeat is suggested by the Per2 sequence (Fig. 3a). Rather than only one base between the two heptameric repeats in the Per consensus, there are two bases (C and A) located between the two putative heptameric repeats in Per2. As shown in Fig. 3(b), although BosR bound the 60 bp ZM132, the recombinant protein did not bind the 32 bp probe ZM352 harbouring Per2.

EMSAs using other duplex nucleotide probes with variations in the Per2 sequence were also performed. As shown in Fig. 3(a), probe ZM353 potentially comprises a 7–2–7 inverted repeat (TTATAATCAATTATAA) with two perfect heptameric motifs (TTATAAT) that are conserved in the Per box. As such, the only difference between ZM353 and the Per box is that ZM353 contains two bases, C and A, EMSAs using other duplex nucleotide probes with variations in the Per2 sequence were also performed. As shown in Fig. 3(a), probe ZM353 potentially comprises a 7–2–7 inverted repeat (TTATAATCAATTATAA) with two perfect heptameric motifs (TTATAAT) that are conserved in the Per box. As such, the only difference between ZM353 and the Per box is that ZM353 contains two bases, C and A,
between the two perfect heptameric repeats, whereas the Per box contains only one base between the two heptamers (Fig. 3a). BosR did not bind to ZM353 (Fig. 3b), despite the fact that the protein bound to the Per box (ZM357) (Fig. 1b). Finally, the C between the two heptameric repeats in ZM353 was deleted, which resulted in the Per2 in the rpoS promoter region being replaced by a perfect Per box containing a 7–1–7 inverted repeat (ZM354). BosR bound this probe (ZM354) strongly (Fig. 3b). The combined data suggest that BosR does not interact with the Per2 sequence, and the Per2 site is not implicated in the binding of BosR to the rpoS promoter region in Borrelia burgdorferi.

The minimal sequence required for BosR binding

To identify the core sequence or DNA element(s) that BosR recognizes in the rpoS promoter, we determined the minimal sequences required for recognition by BosR. To this end, EMSAs were performed to delimit the BosR binding site in both BS1 and BS2. Of note, both regions showed high affinity for rBosR in our previous analyses (the dissociation binding constants (Kd) = 210.2 nM for BS1 and 36.6 nM for BS2) (Ouyang et al., 2011). We initially synthesized a series of duplex oligonucleotides to represent different versions of BS1 with varying lengths. Each of these sequences was labelled with digoxigenin and incubated with purified rBosR, and the bound complexes were separated by native PAGE in an EMSA. As shown in Fig. S3(b), BosR strongly bound the 60 bp probe ZM132 that comprises the full-length 57 bp BS1. When the 50 bp probe ZM351 was used in EMSAs, BosR (100 or 200 nM) still bound to this probe, but the shifted band was much weaker than that observed for the BosR–ZM132 complex, indicating that the affinity of ZM351 for BosR was substantially less than that of ZM132. Moreover, we tested the binding of BosR to four other versions of BS1, including the 32 bp probe ZM352, the 30 bp probe ZM404, the 35 bp probe ZM171 and the 25 bp probe ZM172. BosR failed to bind to any of these probes (Fig. S3b).

We also synthesized duplex oligonucleotides representing various versions of BS2. As shown in Fig. 4(b), BosR strongly bound the 60 bp probe ZM212, the 50 bp probe ZM355 and the 40 bp probe ZM174. Binding of BosR to each of these three probes yielded two complexes including a high-mobility complex (HMC) and a low-mobility complex (LMC). When additional bases were deleted from ZM174, the binding of BosR to the 35 bp probe ZM403 or the 32 bp probe ZM175 was greatly inhibited, because (1) the shifted band was much weaker than those seen in the BosR–ZM174 reaction, and (2) only the HMC was observed in the binding of BosR to ZM403 or ZM175. Additionally, when the 20 bp probe ZM402 was used in the EMSA, BosR failed to bind the probe. Taken together, these data suggested that BosR binds to BS1 or BS2 through a minimal recognition unit.

Identification of a potential BosR binding consensus sequence

Bacterial transcriptional regulatory proteins, such as PerR and Fur, typically bind to DNA with dyad symmetry. Therefore, BosR, as a Fur or PerR homologue, probably also binds to its DNA targets through a palindromic sequence. Along these lines, by scanning the sequence of the minimal unit representing BS2 (i.e. ZM174), two putative palindromic sequences with dyad symmetry were revealed, with one (ATT-TAGTTAAA) near the left end and the other (TTTAA-ATTAAAT) near the right end (Fig. 4a). Moreover, when the minimal unit representing BS1 (i.e. ZM351) was analysed, one apparent putative palindromic sequence with dyad symmetry (ATTTAATTTACAAAT) was noticed at the left end of this DNA (Fig. S3a). When these three sequences were compared, a 13 bp putative palindromic sequence, ATTTAAN-TTAAAT (N can be A or T), was predicted. This motif...
potentially constitutes a 6–1–6 hexameric (ATTTAA) inverted repeat in a head-to-tail orientation, suggesting that this palindromic sequence is the core sequence recognized by BosR. Herein, we have named this sequence the ‘BosR box’. BS2 thus contains two putative BosR boxes; the left or the right BosR box has 11 or 12, respectively, out of 13 matches to the theoretical BosR consensus (Fig. 4a). BS1 contains a nearly perfect BosR box at its left end; compared with the theoretical BosR box, this putative BosR box has an additional C (Fig. S3a). In addition, when aligning the BosR box with the right end of BS1, another imperfect putative BosR box (AAATAATTCAAAA) was also predicted; this motif has nine of 13 matches to the consensus BosR box sequence (Fig. S3a).

The BosR box is essential for BosR binding

We hypothesized that if the BosR box is essential for BosR binding, mutational changes in the sequence should partially inhibit or completely abolish BosR binding. Here, this mutagenesis approach was employed to study the roles of the BosR box in protein binding to ZM174 (the minimal recognition unit representing BS2). In particular, to keep the overall length of the probe constant, bases contained in the BosR box were substituted with random nucleotides. As shown in Fig. 5(a), the 40 bp probe ZM174 comprises two putative BosR boxes, with one located from bases 6 to 18, and the other one located from bases 23 to 35. Binding of 100 or 200 nM BosR to ZM174 yielded two complexes, including the HMC and the LMC species (Fig. 5b), suggesting the presence of two putative BosR boxes. The LMC is probably produced when both BosR boxes are occupied by protein, whereas the HMC is probably produced when only one BosR box is occupied. When mutations were introduced into the first putative BosR box, for example, when the A at position 6 was substituted with C (i.e. A6C mutation), the binding of 100 or 200 nM rBosR to the mutated probe yielded only one binding species (HMC) in EMSAs (Fig. 5b). Similar results were obtained when BosR was reacted with DNA harbouring other mutations (T7G or T8G) in the first putative BosR box. When a mutation was introduced into the second putative BosR box (T26C mutation), protein binding also yielded only the HMC species (Fig. 5b). When mutations were introduced into both putative BosR boxes (T8G/T26C, T8G/T23G/T24C), BosR failed to bind to these mutated probes under the conditions studied (Fig. 5b). These data suggest that the putative BosR boxes are required for rBosR binding to \( rpoS \) in \( Borrelia burgdorferi \).

Sequences flanking the BosR box are also required for BosR binding

As shown in Fig. S3 and Fig. 4, BosR failed to bind probes ZM352, ZM404, ZM171 or ZM402, despite the fact that...
each of these latter four probes contains one putative BosR box. These data suggest that, in addition to the BosR box, other sequences flanking the BosR box may also be required for protein binding. To test this hypothesis, mutations were introduced into the nucleotides neighbouring the putative BosR box in the probe ZM174 (Fig. 6a), and the ability of these mutated probes to bind to BosR was examined. Our results indicate that, similar to the probe ZM174, probes containing mutations such as A1G, G37T/G38T or A40G still bound to BosR and yielded two complexes including the HMC and LMC species (Fig. 6b), suggesting that these nucleotides are not required for recognition by BosR. However, the binding of BosR to the probe containing the C19T mutation produced only one binding species (HMC) (Fig. 6b). In addition, BosR did not bind to DNA with a C20T mutation. These data suggest that these two nucleotides (C19 and C20) contribute in some manner to the binding of BosR to rpoS.

To corroborate these results, we further synthesized artificial dsDNA containing the putative BosR box, and tested the ability of these probes to bind BosR. BosR did not bind the 34 bp probe ZM395 containing one putative BosR box, the 60 bp probe ZM162 containing one copy of TAAATTAAAT or the 60 bp probe ZM178 containing two copies of TAAATTAAAT (Fig. S4).

### Genome-wide distribution of the BosR box in Borrelia burgdorferi

To identify Borrelia burgdorferi genes potentially influenced by BosR, we searched the Borrelia burgdorferi B31 genome (Fraser et al., 1997) by using the Regulatory Sequence Analysis Tools (http://rsat.ulb.ac.be/rsat), and queried for genes containing a putative BosR box. Specifically, one mismatch of the BosR box sequence was allowed in the search, because the putative BosR box identified in the rpoS promoter region has one or two mismatches from the ‘perfect’ BosR box consensus. Gene promoter regions were defined as sequences from −400 to +50 bp (relative to the putative translation start codon ATG). The search identified 47 Borrelia burgdorferi genes carrying one or multiple BosR box sequences in their promoter regions (Table S1). Thirty genes were located on the main chromosome, and 17 genes were on linear or circular plasmids. Of the 30 chromosomal genes, 18 encode proteins with assigned functions and 12 encode hypothetical proteins. Of note, eight of these genes, bb0565 (cheW-2), rpoS, bba43, bba64, bbi36, bbi38, bbi39 and bbj25, were found to be regulated by BosR in our recent microarray analysis (Ouyang et al., 2009), further supporting the notion that the BosR box is the BosR binding site.

## DISCUSSION

We previously reported that BosR directly activates \( \sigma^{54} \)-dependent rpoS transcription in Borrelia burgdorferi via DNA binding (Ouyang et al., 2011). A novel DNA motif (TAAATTAAAT) containing two putative pentameric direct repeats is required for BosR binding. Despite these important initial findings, the precise sequence targeted by BosR remained unclear. Herein, our results provide new evidence that a 13 bp ‘BosR box’ is the core sequence recognized by BosR. The BosR box encompasses the TAAATTAAAT sequence identified in our previous study (Ouyang et al., 2011), but results from this study have
allowed important refinements and an expansion of our initial understanding of the DNA sequence(s) recognized by BosR.

With respect to DNA recognition, the 13 bp BosR box palindromic sequence (ATTTAANTTAAAT) probably comprises a 6–1–6 inverted repeat composed of two ATTTAA hexamers (or TAAATT in the complementary sequence) in a head-to-tail orientation (Fig. 7). Each hexamer probably interacts with one BosR monomer molecule, allowing the 6–1–6 inverted repeat to interact with one BosR dimer. This model is proposed because: (1) most characterized bacterial transcriptional regulators bind inverted repeats (Huffman & Brennan, 2002); (2) the 15 bp Per box is annotated as a 7–1–7 inverted repeat accommodating one PerR dimer (Baichoo & Helmann, 2002); (3) the 19 bp Fur box is primarily recognized as two overlapping heptamer inverted repeats [(7–1–7)], on opposite faces of the DNA helix, with each inverted repeat accommodating one dimeric Fur complex (Baichoo & Helmann, 2002); and (4) structure modelling revealed that the N-terminal helix–turn–helix DNA binding domain is similar among BosR and its homologues Fur and PerR (Ouyang et al., 2011). Alternatively, the BosR box may also be interpreted as a sequence containing two pentamer (TAAAT) direct repeats; each pentamer in this motif may interact with one BosR monomer molecule (Fig. 7). The existence of multiple interpretations for the recognition of one DNA sequence by a regulatory protein is not without precedent. For example, three different models have been proposed for the binding of Fur to the 19 bp Fur box (Baichoo & Helmann, 2002; Escolar et al., 1999; Fuangthong & Helmann, 2003; Lavrrar & McIntosh, 2003). Future structural analyses, such as co-crystallization of BosR with its DNA target(s), may shed light on additional mechanistic details of the BosR–DNA interaction(s).

In accordance with our new model, for the 57 bp BS1 or the 75 bp BS2 protected by BosR (Ouyang et al., 2011), either site contains two putative BosR boxes, suggesting that two BosR dimers may bind to BS1 or BS2. Each BosR dimer thus probably protects ~28–38 bp of DNA, which is in accordance with the fact that ~30 bp of DNA is protected by one Fur or PerR dimer (Baichoo & Helmann, 2002; Fuangthong & Helmann, 2003). In addition, our data indicate that other sequences neighbouring the core BosR box also play important roles in the formation of stable BosR–DNA complexes. These sequences flanking the BosR

---

**Fig. 6.** Sequences flanking the BosR box are required for BosR binding. (a) Sequences of the probes used in EMSAs. The BosR box is underlined. The number above the nucleotide denotes the position of the nucleotide in probe ZM174. Dashed lines represent identical nucleotides contained in probe ZM174. Each mutation is indicated in bold type. (b) EMSAs. Increasing amounts of rBosR (0, 100, 200 nM) were used in DNA binding assays. Probe names are indicated below the image. Arrowheads denote the HMC and LMC.

---

**Fig. 7.** Comparisons of the BosR box, the reference Fur box consensus and the Per box consensus. Arrows represent putative dimer binding sites. IR, inverted repeat; DR, direct repeat. The BosR box can be interpreted as a 6–1–6 IR or a 5–5 DR. The BosR box can be interpreted as a 6–1–6 IR or a 5–5 DR. The BosR box can be interpreted as a 6–1–6 IR or a 5–5 DR. The BosR box can be interpreted as a 6–1–6 IR or a 5–5 DR. The BosR box can be interpreted as a 6–1–6 IR or a 5–5 DR.
box may be required to recruit the supplementary BosR molecules to the initial binding site (i.e. the BosR box) to stabilize the BosR–DNA association. Thus, BosR may initially bind to the BosR box, then wrap around the DNA and oligomerize at the binding site. These DNA binding attributes of BosR together may account for the large DNA region protected by BosR observed in our previous DNase I footprinting assays (Ouyang et al., 2011).

The BosR box is unequivocally distinct from the Fur box or the Per box, despite the fact that the hexameric repeat found in the BosR box (TAAATT, or ATTTAA) and the heptamer repeat conserved in the Per box (TTATAAT) or the Fur box (TGATAAT) are all AT-rich sequences (Fig. 7). In fact, our data revealed that BosR did not bind DNA containing the Fur box. Moreover, despite BosR-bound DNA containing the Per consensus, our data clearly demonstrated that the BosR box, rather than the two putative Per boxes (Per1 and Per2) predicted in the rpoS promoter region, accounts for the binding of BosR. For the 15 bp Per1 (ATATAATACTATGA), although it putatively constitutes a 7–1–7 inverted repeat resembling the Per consensus, this sequence is not located in any of the three BosR-protected regions identified by DNase I footprinting assays (Ouyang et al., 2011), suggesting that BosR may not bind to Per1. This hypothesis was confirmed by our DNA binding assays, which showed that BosR did not bind the 50 bp DNA (ZM125) containing Per1. For the second 16 bp putative Per box Per2 (TTACAATCAATTACA), although it is located in the BosR-bound BS1, this putative Per box also is not involved in recognition by BosR. This conclusion was drawn based on multiple lines of evidence. First, different from the 7–1–7 inverted repeat of the Per box consensus, Per2 comprises a putative 7–2–7 inverted repeat. Second, although BosR bound the 31 bp probe ZM357 harbouring the Per consensus, our EMSAs indicated that BosR failed to bind the 32 bp probe ZM352 harbouring Per2 (Fig. 3b). Third, when the two imperfect putative heptamers (in Per2) were substituted with the heptamers conserved in the Per consensus, which yields a 7–2–7 inverted repeat containing two perfect heptamers (TTATAAT), the probe ZM353 still did not bind to BosR (Fig. 3). Finally, when the C between the two heptameric repeats in ZM353 was deleted, yielding a probe (ZM354) containing a perfect 7–1–7 inverted repeat, the probe strongly bound to BosR. Taken together, these data not only suggest that those two putative Per boxes predicted in the rpoS promoter region are not implicated in the binding of BosR, but the spacing of the two heptamers conserved in the Per box also plays important roles in protein binding.

Our results also point to the identification of novel candidate genes potentially regulated by BosR. To define the BosR regulon in Borrelia burgdorferi, Katona et al. (2004) performed a blastn analysis based on the Per box consensus sequence. However, the BosR box identified in our current study is disparate from the Per box. By scanning the Borrelia burgdorferi B31 genome for the BosR box, 47 Borrelia burgdorferi genes were found to contain one or multiple BosR boxes, suggesting that these genes may be regulated or influenced by BosR. Among these genes, potential BosR control over 41 genes was not previously appreciated. More importantly, in addition to rpoS, several other BosR-regulated genes are potentially essential for the Borrelia burgdorferi tick–mammal enzootic life cycle. For example, cheW-2 encodes a putative purine binding chemotaxis protein that is probably involved in the chemotactic and adaptive responses when spirochaetes reside in ticks (Charon & Goldstein, 2002). BB0578 (MCP-1) plays a critical role in Borrelia burgdorferi chemotaxis, probably by functioning as a methyl-accepting chemotaxis protein to regulate autophosphorylation of the sensor histidine kinase CheA (Charon & Goldstein, 2002). P66, a β3-chain integrin ligand acting as an adhesion molecule, is required for tissue colonization and the initial establishment of infection in animal hosts (Coburn & Cugini, 2003). Despite its unknown function, BBA64 was found to be required for the transmission or delivery of Borrelia burgdorferi from tick to mammalian hosts (Gilmore et al., 2010). Other BosR-regulated genes may also be potentially involved in Borrelia burgdorferi pathogenesis. Unravelling these novel genes potentially targeted by BosR will provide new insights into the molecular mechanism(s) by which BosR contributes to Lyme disease. Future efforts should reveal to what extent BosR directly modulates the expression of these other candidate virulence-associated genes during the tick–mammal transition of Borrelia burgdorferi, and decipher the precise details of how these genes are controlled by BosR.

Although BosR has been known for nearly two decades (Fraser et al., 1997), how this regulator exerts its regulatory function still remains obscure. In fact, whether BosR functions like Fur or PerR has been controversial. When Fur binds to the Fur box, Fur usually functions to repress gene transcription (Carpenter et al., 2009; Lee & Helmann, 2007). Similarly, the binding of PerR to the Per box also results in gene repression; gene transcription is allowed when PerR dissociates from the gene promoter (Duarte & Latour, 2010; Makthal et al., 2013; Traoré et al., 2006). Our previous results (Ouyang et al., 2009, 2011) revealed that the binding of BosR to the rpoS promoter region directly activates σ54-dependent rpoS transcription in Borrelia burgdorferi, providing the first clue that BosR may modulate Borrelia burgdorferi gene expression in a manner opposite to that of Fur or PerR. In this regard, our study herein provides new data to support that BosR may have unique features to set it apart from its homologues. In particular, our new data clearly demonstrate that BosR has a novel DNA binding activity. BosR binds not only to DNA sequences containing the Per box, but also DNA targets harbouring the BosR box. The molecular basis for this discrimination between BosR and its other homologues presumably lies in potential subtle structural differences among the three Fur family members. Future studies thus are warranted to elucidate those salient structural features of BosR that confer its novel DNA binding activity.
ACKNOWLEDGEMENTS

We thank Andrea Balbo (National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health) for acquiring solution density and viscosity data.

REFERENCES


