Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance

Aurore Sanchez1,2, Jérôme Rech1,2, Cyrielle Gasc1,2 and Jean-Yves Bouet1,2,*

1Laboratoire de Microbiologie et Génétique Moléculaires, Centre National de la Recherche Scientifique, F-31000 Toulouse, France and 2Laboratoire de Microbiologie et Génétique Moléculaires, Université de Toulouse, UPS, F-31000 Toulouse, France

Received October 1, 2012; Revised December 13, 2012; Accepted December 31, 2012

ABSTRACT
ParB proteins are one of the three essential components of partition systems that actively segregate bacterial chromosomes and plasmids. In binding to centromere sequences, ParB assembles as nucleoprotein structures called partition complexes. These assemblies are the substrates for the partitioning process that ensures DNA molecules are segregated to both sides of the cell. We recently identified the sopC centromere nucleotides required for binding to the ParB homologue of plasmid F, SopB. This analysis also suggested a role in sopC binding for an arginine residue, R219, located outside the helix-turn-helix (HTH) DNA-binding motif previously shown to be the only determinant for sopC-specific binding. Here, we demonstrated that the R219 residue is critical for SopB binding to sopC during partition. Mutating R219 to alanine or lysine abolished partition by preventing partition complex assembly. Thus, specificity of SopB binding relies on two distinct motifs, an HTH and an arginine residue, which define a split DNA-binding domain larger than previously thought. Bioinformatic analysis over a broad range of chromosomal ParBs generalized our findings with the identification of a non-HTH positively charged residue essential for partition and centromere binding, present in a newly identified highly conserved motif. We propose that ParB proteins possess two DNA-binding motifs that form an extended centromere-binding domain, providing high specificity.

INTRODUCTION
Low-copy number plasmids and most bacterial chromosomes have acquired partitioning (Par) mechanisms to ensure faithful segregation in dividing cells. This stable inheritance depends on active partition systems [reviewed in (1)]. Virtually, all Par systems are composed of three essential elements: a centromere, a centromere-binding protein (CBP) and an NTPase. Interactions between these partners are important for the partition process, and of these, the specific recognition of centromeres by CBPs to assemble partition complexes is crucial. Partition complexes are substrates for partition machineries, as they are responsible for the proper intracellular localization of the replicon onto which these are assembled. CBPs are functional equivalents of eukaryotic kinetochores, in which they act as the link between centromeres and NTPases.

Bacterial partition systems have been classified in three types according to the nature of the NTPase involved (2). The partition system we study here is of type I, representative of the most widespread system, present on most plasmids and on all chromosomes carrying a Par system. Although, the molecular mechanisms underlying partitioning of type I systems are still to be unraveled, evidence is now accumulating that in vivo type I adenosine triphosphatases (ATPases) generally show dynamic behaviors, either oscillating over the nucleoid (3–5) or forming metastable filaments (6,7). Most of the ParA proteins have also been found to polymerize in vitro in response to adenosine triphosphate (ATP) binding (7–10). These dynamic behaviors depend on the presence of the three Par elements. Notably, ParB stimulates ParA activities, such as polymerization (8,9) and ATP hydrolysis (11,12). This latter activity is further stimulated with centromere-bound ParB (13). Partition complexes assemblies on centromere DNA are thus important for the regulation of the partition process.

Bacterial centromeres, generically called parS sites, are diverse in composition and organization. Chromosomal parS sites typically consist of 16-bp inverted repeats clustered around the replication origin or spread throughout the Ori domain (14). Plasmid centromeres consist of...
inverted and/or direct repeats [reviewed in (15)]. The repeats can be extensive, for example, sopC of the F plasmid consists of 12 tandem repeats of a 43-bp motif (16), and parS of pTAR contains 13 heptameric repeats that are separated by integral helical turns (17). By contrast, the partition site O_{b3} of plasmid RK2/RP4 is a simpler palindromic core sequence of 13 bp, although related sites dispersed throughout this plasmid might also contribute to segregation (18). Centromeres of more complex organization also exist, such as that of plasmid P1 (19). This ~80-bp parS consists of two arms that harbor non-symmetrical heptamer (A-box) and hexamer (B-box) DNA motifs, separated by a central binding site for the host-encoded DNA-bending protein, integration host factor. Despite this high diversity, ParBs generally bind to centromere sites through helix-turn-helix (HTH) or in some cases ribbon-helix-helix (RHH) motifs [for review (15)].

In F plasmid, the partitioning locus sop specifies the ATPase, SopA, the CBB, SopB and the centromere, sopC. Each of the 12 tandem 43-bp repeat contains a 16-bp inverted repeat that constitutes the SopB-binding site (20). SopB is composed of several functional domains (Figure 1A). Owing to its high flexibility, the structure of the N-terminal part of SopB is unknown, which is also the case with others ParBs (21,22). This part contains the SopA-interaction domain (23), the arginine-finger and the putative arginine-loop motifs that stimulate SopA ATPase activity (13) and possibly a SopB dimer–dimer interaction region (24). SopB not only forms a SopB–sopC partition complex that recruits SopA but also extends the complex by spreading along the DNA on either side of sopC (25), and also probably in trans to different DNA molecules (26). By coating DNA, SopB forms extended partition complexes whose role in partitioning is not understood. In all cases, specific assembly of SopB on sopC to nucleate the formation of extended partition complexes is essential for F partitioning.

Our present knowledge of the protein–centromere contacts that define binding specificity is based on structural studies of only three ParB–parS complexes: KorB–O{b3} (RP4), ParB–parS (P1) and SopB–sopC (F). KorB binds O_{b3} through a classical HTH motif, and the recognition specificity is provided by two amino acid side chains external to the HTH motif (22). The P1 partition complex is more complicated owing to the unusual organization of its centromere site (see earlier in the text). Each HTH motif of ParB dimer contacts an A-box on each parS arm, and an additional DNA-binding module contacting the B-box is provided by the dimerization domain (21). The strong specificity of P1 ParB–parS interaction is shared between these two distinct DNA-binding motifs. Specificity determinants of SopB–sopC partition complex have been recently investigated at the nucleotide level by high-throughput analysis (27). This revealed that all 16 bp of sopC are required, to varying degrees, for maximal SopB binding strength. In addition, the co-crystal structure of a SopB–sopC complex was solved at ~3.5 A (26), enabling prediction of the likely contacts between sopC and residues of the HTH motif (Figure 1C). Because previous results had suggested binding specificity was entirely contained within the HTH (23), it was interesting that an arginine residue (SopB-R219) outside the HTH was close to the sopC guanine at position 7 (sopC-G7) in the co-crystal.

To better understand the underlying mechanisms involved in plasmid partitioning and centromere binding, we have now investigated the predicted interaction between SopB–R219 and sopC-G7. Strikingly, we found that this interaction is essential for SopB–sopC binding in vitro and is necessary for partition in vivo. These findings reveal that the specific contacts between SopB and sopC occur both inside and outside the HTH motif of SopB. Our results refine the basis of the high-binding specificity of F partition complex assembly. Based on sequence alignments and the known structure of a chromosomal ParB, we identified and showed that an arginine residue provides a specific contact with the chromosomal centromere. Extensive sequence analysis revealed that this positively charged residue is part of a newly defined and highly conserved motif present in chromosomal ParBs over the whole bacterial domain of life.

MATERIALS AND METHODS

Bacterial strains and plasmids

Strains are derivatives of Escherichia coli K-12 or E. coli B (C2833) and are listed, together with plasmids, in Supplementary Tables S1 and S2, respectively. Construction of plasmids is detailed in Supplementary Figure S4.

Media and growth conditions

Cultures were essentially grown at 37°C with aeration in LB (Miller, 1972) containing thymine (10 μg ml⁻¹) and antibiotics (Sigma) as appropriate (μg ml⁻¹): kanamycin (Km, 50), chloramphenicol (Cm, 10), streptomycin (Sm, 200) and spectinomycin (Sp, 20 in liquid medium, 30 in solid). Luria Broth (LB) was supplemented with 1.5% agar (Difco) for solid medium. For stability assay, cultures were grown at the indicated temperature with aeration in LB or in M9 minimal medium supplemented with glucose (0.4%), casamino acids (0.2%), MgSO₄ (1 mM), CaCl₂ (0.1 mM), thiamine (1 μg ml⁻¹), leucine (20 μg ml⁻¹) and thymine (40 μg ml⁻¹).

DNA manipulation and related procedures

Enzyme reactions, DNA preparation, agarose gel electrophoresis and transformation with plasmid DNA were carried out using standard procedures, according to suppliers’ recommendations where applicable (restriction endonuclease, DNA polymerase andligase, New England BioLabs; polymerase chain reaction with Phusion, Fermentas; plasmid and gel-fractionated fragment DNAs, Qiagen or Promega). SopB mutations were introduced in pYAS6 and pYAS47 plasmids by quick-change mutagenesis and verified by DNA sequencing (Supplementary Figure S4).
Protein purification

Intein-tagged SopB-G324, SopB-R219 mutants, ParBpt-G291 and ParBpt-G291-R194A proteins were purified from strain C2833 essentially as described previously (13), except that fractions were directly collected and frozen at −80°C after elution from the chitin column. ParBbcen-G306 does not bind to the chitin column. ParBbcen-G306 and ParBbcen-G306-R192A were prepared as described earlier in the text, except that FrI was incubated directly with dithiothreitol to induce full intein cleavage (see Supplementary Figure S3) and was stored in aliquots at −80°C. ParB proteins from Bcen were assayed directly with crude extracts.

Plasmids stability assay

Experiments were started from colonies of cells freshly transformed with the plasmids under test. SopB alleles, expressed from a pSC101 derivative under plac, were assayed in strain MC1061. Overnight cultures in M9–glucose–casamino acid medium containing selective antibiotics were diluted 400-fold into the same medium and grown to A600 = 0.25. Samples were then diluted serially into fresh M9 medium without the antibiotic selective for the plasmid under test and were processed as described previously (28). To determine the fraction of cells that retained the plasmid, samples were taken at the beginning and after 20 and 40 generations. The calculated loss rate per generation is given by the slope A of the equation (y = B.e^-Ax) from the fitted curves (using Graph Pad Prism 4 software) on graph plotting the per cent of cell mini-F+ as a function of the generation time, which provides a good approximation of the loss frequency (f) when <10% [f = 1 – (cell mini-F+/total cell)]^n, where n is the generation number. ParB alleles of Bcen or Pput were assayed for the stabilization of pDAG504 in same conditions as previously published (29,30). Briefly, for Bcen alleles, plasmids under test were introduced in strain DH10B, and growth cultures were performed in M9–glucose–casamino acid medium at 37°C. For Pput alleles, plasmids under test were introduced into strain DH5α, and growth cultures were performed in LB at 37°C.

Electromobility shift assay

Experiments were performed as previously described (9), using 1 nM radiolabeled DNA probes in the presence of sonicated salmon sperm DNA as competitor (100 μg ml⁻¹). Binding reaction products were resolved at 4°C by electrophoresis on 6% polyacrylamide gels in TGE (25 mM Tris base, 190 mM Glycine, 1 mM Ethylenediaminetetraacetic acid) for 2–4 h at 180 V. Polyacrylamide gel electrophoresis-purified complementary oligonucleotides (Sigma-Aldrich) were used to assemble duplex DNA probes by annealing (only top strands are shown; 16-bp SopB/ParB binding sites are indicated by upper case letters; bases corresponding to positions 7 and 10 are indicated in bold): sopC2: 5'attagtcTGGGACanccctcgtat-3'; sopC63: 5'attagtcTGGGACATGGGCcctcgtat-3'; sopC64: 5'attagtcTGGGACATGGGCcctcgtat-3'; sopC66: 5'attagtcTGGGACATGGGCcctcgtat-3'; and sopC67: 5'attagtcTGGGACATGGGCcctcgtat-3'. (C) Predicted interactions between amino acids of SopB with sopC nucleotides, adapted from the study by Pillet et al. (27). R190, R219, R195 and K191 correspond to arginines 190, 195, 219 and lysine 191. Straight and dotted lines represent suggested amino acid–nucleotide interactions. R and Y stand for purine and pyrimidine nucleotides, respectively. Green zones represent the subunits of an SopB dimer.
RESULTS

We have previously investigated the bases involved in the specific recognition of the F plasmid centromere by SopB [(27), summarized in Figure 1C]. In particular, the guanine at position 7 (sopC-G7) in the left arm (and its counterpart at position 10 in the right arm) was found to play a role in SopB–sopC interaction. In addition, the co-crystal structure of SopB–sopC complexes, determined at a medium resolution (~3 Å; (26)), suggests that G7 is close to an arginine residue (R219), in the major groove (Figure 1B). We estimated the distance between the O group of G7 and the NH group of R219 at 2.6 Å (Supplementary Figure S1), compatible with hydrogen-bonding between them (31). This R219 residue does not belong to the HTH motif responsible for the primary binding to DNA (Figure 1A), which contains three positively charged residues (R190, K191 and R195; Figure 1B) involved in specificity of binding to sopC (13,26). SopB-R219 is part of an α-helix separated from the HTH by an other α-helix (Figure 1A). Its potential role in centromere binding and thus in the partition process was further investigated.

SopB-R219 mutants are unable to stabilize mini-F

To determine whether the R219 residue plays a direct role in the partition of the mini-F plasmid, we constructed two SopB-R219 mutants by converting R219 to alanine (R219A) or lysine (R219K). These mutations were introduced into a pSC101 derivative, pYAS47, which carries sopAB under the control of the lac promoter, yielding pJYB161 (sopB-R219A) and pJYB162 (sopB-R219K). These plasmids were introduced into strain MC1061 carrying pDAG209, an unstable mini-F sopC+ but ΔsopAB, and loss of the mini-F during growth without selection was measured (Supplementary Figure S2A and Table 1). When wild-type SopB was provided in trans from pYAS47, no cell without pDAG209 was observed after 40 generations, indicating that the mini-F sopC+ was completely stable (loss rate < 0.01%). Without SopB (pAM238), the mini-F was lost at a frequency of 4.0% per generation as expected from the copy number (~5–6 at division) for segregation at random. The mutant alleles SopB-R219A (pJYB161) and -R219K (pJYB162) failed to stabilize pDAG209, which was lost at 3.4 and 3.1% per generation, respectively. Western blot analysis showed that in the stability assays, the mutant SopBs were produced at the same level as wild-type (wt) (Supplementary Figure S2B). These results indicate that the R219 residue of SopB is essential for the maintenance of mini-F during generation.

SopB-R219 variants do not bind sopC centromere

To test whether the inability of sopB-R219 alleles in stabilizing the mini-F is primarily because of a default in centromere-binding activity, we performed electromobility shift assays (EMSA). Wild-type SopB and R219 variants were expressed as intein-tagged proteins and purified as native proteins after cleavage of the intein (see ‘Materials and Methods’ section and Supplementary Figure S3).
Guanines 7 and 10 in sopC site are critical for SopB binding. The SopB–sopC co-crystal structure (26) predicts that the R219 residue of an SopB monomer is near the guanine in position 7 of the left arm of sopC, and thus that R219 of the second SopB monomer should be close to the guanine 10 on the right arm (Figure 1B and C). We previously observed by SPRi that mutations of these bases affect SopB binding to sopC (27). However, because of the low sensitivity of this technique, quantification of the loss in DNA-binding activity was not possible. Because SopB-R219 variants are deficient in sopC binding, we expected that if R219 directly interacts with guanines 7 and 10, we should observe a comparable loss in SopB binding if these bases are mutated.

We, therefore, performed EMSA in same conditions as described earlier in the text with wild-type SopB but with 30-bp sopC DNA probes carrying either transversion (C to A; probe sopC63) or transition (C to T; probe sopC64) at positions 7 and 10 (Figure 4). Both probes were barely shifted even at the highest SopB concentration (1 μM), indicating that these bases are strictly required for the formation of SopB-sopC complexes. This result is consistent with a direct interaction between R219 residues and guanine at positions 7 and 10.

Table 1. Stability of mini-F plasmids carrying F sopC or chromosomal parS centromeres

<table>
<thead>
<tr>
<th>F plasmids</th>
<th>Multicopy plasmids</th>
<th>Relevant characteristics</th>
<th>Loss/generation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDAG209</td>
<td>pAM238</td>
<td>sopA⁺B⁺</td>
<td>4.0 (±0.5)</td>
</tr>
<tr>
<td>(sopC)</td>
<td>pYAS47</td>
<td>sopA⁺B-R219A</td>
<td>3.4 (±0.3)</td>
</tr>
<tr>
<td></td>
<td>pJYB161</td>
<td>sopA⁺B-R219K</td>
<td>3.1 (±0.1)</td>
</tr>
<tr>
<td>pDAG551</td>
<td>pBBR1MCS5</td>
<td>Been parA⁺B⁺</td>
<td>2.6 (±0.1)</td>
</tr>
<tr>
<td>(parS)</td>
<td>pDAG562</td>
<td>Been parA⁺B-R192A</td>
<td>0.15 (±0.1)</td>
</tr>
<tr>
<td></td>
<td>pYJB167</td>
<td></td>
<td>2.7 (±0.4)</td>
</tr>
<tr>
<td>pDAG504</td>
<td>pAM238</td>
<td>Pput parA⁺B⁺</td>
<td>3.6 (±0.7)</td>
</tr>
<tr>
<td>(parSput)</td>
<td>pDAG503</td>
<td></td>
<td>0.25 (±0.1)</td>
</tr>
<tr>
<td></td>
<td>pJYB168</td>
<td>Pput parA⁺B-R194A</td>
<td>2.1 (±0.5)</td>
</tr>
</tbody>
</table>

Derivatives of strain MC1061 carrying the mini-F pDAG209 and the indicated plasmids were grown at 30 °C. Strains carrying the mini-Fs pDAG551 (consensus parS) or pDAG504 (parSput) were grown at 37 °C. The sop and par operons are under the control of the lac promoter on pAM238 or pBBR1MCS5. Samples of the cultures were plated at intervals for estimation of the proportion of mini-F-positive cells. Loss rates per generation, given by the slopes (A) of the fitted curves as presented in Supplementary Figure S2A, are average of at least three experiments, and standard deviations are indicated in brackets.

Figure 2. Interaction of SopB wild-type and R219 variants with DNA. The ability of wild-type SopB, SopB-R219A and -R219K to interact with an sopC DNA duplex or with non-specific DNA was tested in a gel mobility shift assay. Reaction mixtures, assembled on ice and incubated at 30 °C, were analyzed by electrophoresis on polyacrylamide gels. (A) SopBs binding to sopC centromere site. 32P-labeled 30-bp sopC2 probe was incubated alone (–) or with increasing concentrations (from left to right; 10, 30, 100, 300 and 1000 nM) of wild-type SopB (WT), SopB-R219A or SopB-R219K as indicated on top, in the presence of 100 nM competitor DNA. Positions of DNA probes and protein complexes are indicated on the left: B1 denotes the specific complex, whereas asterisk corresponds to a specific shift from a secondary form of the duplex DNA. (B) SopBs binding to non-specific DNA probe. 32P-labeled 114-bp repE probe was incubated alone (–) or with increasing concentrations (from left to right; 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 and 2000 nM) of wild-type SopB (WT) or SopB-R219A as indicated on top, in the absence of competitor DNA. NS1 and NS2 denote discrete complexes, and HMW corresponds to high molecular weight complexes.

Figure 3. SopB-R129A allele does not form foci cells. Left; localization of SopB–Venus fusion. Fluorescent foci from the mini-F expressing SopB–Venus display the expected pattern for partition complexes faithfully partitioned. Right; localization of SopB-R219A–Venus fusion. All cells displayed diffuse fluorescence throughout the cytoplasm. Cells without fluorescence are presumed to lack the mF carrying SopB-R219A–Venus as expected from the loss rate (see Supplementary Figure S2A and main text).

Guanines 7 and 10 in sopC site are critical for SopB binding

The SopB–sopC co-crystal structure (26) predicts that the R219 residue of an SopB monomer is near the guanine in position 7 of the left arm of sopC, and thus that R219 of the second SopB monomer should be close to the guanine 10 on the right arm (Figure 1B and C). We previously observed by SPRi that mutations of these bases affect SopB binding to sopC (27). However, because of the low sensitivity of this technique, quantification of the loss in DNA-binding activity was not possible. Because SopB-R219 variants are deficient in sopC binding, we expected that if R219 directly interacts with guanines 7 and 10, we should observe a comparable loss in SopB binding if these bases are mutated.

We, therefore, performed EMSA in same conditions as described earlier in the text with wild-type SopB but with 30-bp sopC DNA probes carrying either transversion (C to A; probe sopC63) or transition (C to T; probe sopC64) at positions 7 and 10 (Figure 4). Both probes were barely shifted even at the highest SopB concentration (1 μM), indicating that these bases are strictly required for the formation of SopB-sopC complexes. This result is consistent with a direct interaction between R219 residues and guanine at positions 7 and 10.
Chromosomal CBPs harbors extended DNA-binding domains essential for partition activity

The structure of a chromosomal homologue of SopB, Spo0J of *Thermus thermophilus*, solved in the absence of the parS centromeric DNA, also displays a typical HTH proposed to be the parS-binding motif (34). Interestingly, a helix (H9) located two helices away from the HTH recognition helix (H7) was shown to correspond to helix H8 of KorB (35), and it also aligns with helix H5 of SopB (Figure 6A). H5_{sopB} and H8_{korB} carry the R219 and the R240 residues, respectively, involved in specific recognition of their cognate centromere [this study, (22)]. Helix H9_{spo0J} also harbors an arginine residue (R179) in the corresponding helix (Figure 6A). In the 3D structure of the Spo0J dimer, the orientation of R179 relative to the HTH supports the possibility that it could provide a direct interaction with a specific base of parS (Figure 6B). Indeed, each arginine residue on both monomers has its side chain directed towards where the parS DNA would be bound in the HTH motif alignment.

Spo0/J_{the} binds the pseudopalindromic parS sequence 5'-TGTTCAATGGAACA-3'. This sequence corresponds to the highly conserved parS, characteristic of centromeres present on bacterial chromosomes, whose consensus is 5'-TGTTCAATGGAACA-3'. If the Spo0J-{R179} is involved in specific binding, we expect that it should be conserved at least in a subset of chromosomal ParBs. To test this hypothesis, we aligned ParBs from a large range of bacteria (Figure 6C). We found that R179 is highly conserved and present in most phyla. In the ε-Proteobacteria, the Chlamydiales and the Spirochaete groups, the arginine is replaced by a lysine, suggesting that the positively charged nature of the residue is the important feature, compatible with a direct interaction with a base of the parS site. If the H9 helices of ParB are functional equivalents to the H5 helix of SopB, these R/K residues should also be involved in chromosome segregation.

ParABS partition activity on bacterial chromosome is difficult to be assessed directly because of pleiotropic effects. However, partition function could be examined by the ability of parAB expressed from plac on a multicopy vector to stabilize a mini-F (Δsop) plasmid carrying a single parS in *E. coli* *in trans*. Indeed, this assay has been used previously to demonstrate the partition capacity of the *Bacillus subtilis* sop0J-parS and the *Pseudomonas putida* and *Burkholderia cenocepacia* parABS systems (29,30,36). To determine whether these positively charged residues are involved in partition, we tested the ParABS_{put} and ParABS_{cen} systems in stability assays by introducing mutations at positions equivalent to Spo0J-R179, parB-R194A and parB-R192A, respectively. We observed, as previously reported using the same experimental conditions (29,30), that both ParAB systems were able to efficiently stabilize their cognate parS carrying mF, pDAG504 or pDAG551 (Table 1; loss rates = 0.15–0.25%). We found, however, that both parB-R194A and parB-R192A alleles were impaired in their ability to support stabilization of pDAG504; *Bcen* ParB-R192A was completely deficient (loss rate = 2.7%), whereas *Put* ParB-R194A was highly but not completely deficient (loss rate = 2.1%). These results thus clearly demonstrate that these positively charged residues in H9 helices are essential, as is SopB-R219, for the partition function of chromosomal ParBs.

By performing EMSA, we then tested whether these arginine residues are, as for their SopB counterparts, directly involved in parS-specific binding (Figure 5). ParB proteins from *P. putida* were purified as for SopBs, whereas ParB proteins from *B. cenocepacia* were directly assayed from crude extracts expressing intein-tagged ParBs (see ‘Material and Methods’ section and Supplementary Figure S3). In contrast to wild-type, mutant ParBs from *Put* and *Bcen* were not able to bind to their respective parS sites. In both cases, the DNA binding deficiency is so high that no specific retardation was detected. These results thus indicate that R194 and R192 residues from *Put* and *Bcen*, respectively, are directly involved in centromere binding.

Finally, we found that the R/K residues of H9 helices are part of a highly conserved motif, GHA^{R/K}_{A/V}LL, that covered most of the H9 helix (Figure 6C). Secondary structure analysis on these motifs predicted that they all harbor an α-helical conformation (data not shown). We named this conserved motif box ‘CBM2’ for centromere-binding motif 2; the CBM1 being the typical HTH motif.

**DISCUSSION**

SopB is known to bind DNA specifically on the sopC centromere to function in the active partitioning of the...
F plasmid. We have shown here that the specificity of SopB–sopC interaction is more complex than anticipated, by identifying a new specific DNA-binding determinant, an arginine at position 219, located outside of the HTH motif. We demonstrated that SopB-R219A/K variants are completely deficient in sopC binding in vitro, whereas both mutants are still able to bind DNA non-specifically (Figure 2). Fluorescence microscopy revealed that, in contrast to wt SopB, neither mutant allele was able to form intracellular foci (Figure 4), supporting the idea that in vivo they are also deficient in sopC binding. In addition, these single mutations totally abolished active segregation (Table 1), demonstrating that the R219 residue is a crucial DNA-binding determinant of the SopB–sopC interaction. Extension of this result to chromosomal partition systems reveals that most ParBs also harbor a secondary CBM that was shown to be essential for partitioning functions (Figure 6 and Table 1).

Direct interaction between R219 and guanine 7 defines an extended DNA-binding domain in F SopB

HTH motifs are composed of two helices that contact the DNA in the major groove. The second helix, called the recognition helix, usually provides specificity by contacting specific bases of the target binding sequence. In the case of SopB, the recognition helix makes three specific base contacts through positively charged residues (K190, K191 and K195) and possibly two others through residues S189 and I192 (26). These numerous contacts were thought to be sufficient for providing a high specificity for centromere binding. The proximity evident in the SopB–sopC structure of the positively charged residue, R219, located two helices away from the HTH, to guanine at position 7 (G7) was thus intriguing, especially because G7 was found to be involved in SopB binding (27). We found the same strong deficiency of binding, as judged by gel retardation, between SopB-R219A binding to sopC and wild-type SopB binding to sopC carrying either transitions or transversions at positions 7 and 10 (compare Figures 2A and 4). In addition, replacing the arginine by a lysine at position 219 does not restore binding to sopC (Figure 2A). A positive charge at this position is thus not sufficient for DNA binding, arguing for specific contact between SopB-R219 and sopC-G7. In agreement with this finding, bioinformatic analysis (27) showed that (i) the arginine at position 219 is conserved among all closely related SopB homologues; and (ii) all sopC centromere sequences present in the databases harbor a guanine at position 7. Along with these data, our results thus clearly support a direct interaction between R219 residues of each SopB monomer with each G7 base of the palindromic sopC sequence.

If the R219-G7 interaction was anticipated from the structure, the finding that it is a critical one was not. Single mutations at position R219 completely abolish sopC recognition, revealing that the HTH is not sufficient to provide the specificity of the SopB–sopC interaction. However, specific contacts made by the recognition helix are also crucial for binding specificity, as mutations of two of the positively charged residues, K191 and K195, also abolish specific, but not non-specific, sopC binding (13). The centromere-binding specificity of plasmid F is thus shared by two motifs, the HTH and the R219 containing helix, which together define an extended DNA-binding domain.

Centromere-binding proteins harbor extended DNA-binding domains for highly specific interactions

Is this extended centromere-binding domain present in SopB a common characteristic among CBPs? Some co-crystal structures of ParB–parS are available to address such a question. Similarly to SopB, KorB of RK2/RP4 harbors two DNA-binding motifs (22). However, in sharp contrast to SopB, residues from the helices forming the HTH motif of KorB do not make specific contacts with bases in the O8 site (13-bp parS centromere). The KorB–O8 recognition specificity is mainly based on two side chain interactions outside the HTH (T211 and R240) that form direct hydrogen bonds with two bases next to the central base pair binding sequence. Thus, although the specificity determinants are not distributed as for F SopB, KorB also has an extended centromere-binding domain.

In the case of the P1 ParB, the specificity recognition is shared by two separate DNA-binding motifs, each contacting its specific binding sequence, the A- and B-boxes; the P1 centromere site being composed of a complex arrangement of these two DNA-binding boxes (37,38). The A-boxes are specifically bound by the HTH motifs, and the B-boxes are bound by the dimerized dimer motifs (21). These two motifs participate to the efficient binding of ParB to parS; therefore, they could be regarded as an extended DNA-binding domain. Although P1 ParB and F SopB are closely related to each other, SopB does not carry this second DNA-binding motif because of the lack of two small extra loops residues providing ‘wings’
that are specifically present in the ParB dimerization motif (26). This difference is in relation with the centromere structures where, in the particular case of P1, ParB recognizes two different binding sequences. Thus, as in KorB and SopB, P1 ParB has an extended DNA-binding domain composed of two centromere-binding motifs.

In the course of this study, we have identified the SopB-R219 counterpart present on chromosomal ParBs in most

Figure 6. ParB conserved motifs and residues involved in centromere binding. (A) Amino acid alignment of ParB homologues. Sequence alignments of SopB (F) with KorB (RP4) and Spo0J (T. thermophilus) are based on the study by Leonard et al. (34). Identical residues and highly conserved residues are indicated by asterisks and dots, respectively. Helices H1 to H5 of SopB, based on the crystal structure analysis (26), are schematically represented on top; Helices H2 and H3, corresponding to the HTH motif, are drawn in yellow. Arginine residues R219 (SopB), R240 (KorB) and R179 (Spo0J) are colored in red. (B) Ribbon diagrams of a ParB dimer. Structure is from T. thermophilus Spo0J (34). One monomer is colored in cyan and the other in green. The right diagram is a 90° horizontal rotation of the left diagram. HTH motifs and R179 residues of each monomer are colored in yellow and red, respectively. Each arm of the 16-bp parS DNA is expected to bind one HTH motifs of the Spo0J dimer. Orientations of R179 residues are compatible for providing additional specificity contacts in parS DNA. (C) Amino acid alignment of chromosomal ParBs. ParBs from a large spectrum of bacterial species were aligned with ClustalW; in the case of multichromosomal bacteria, ParBs are from the main chromosome. These ParBs are expected to bind the parS consensus sequence 5'-GTTNCACGTGNAAC-3' (30,48). Sequences highlighted in yellow and orange correspond to the HTH motif and the CBM2 (centromere-binding motif 2) box, respectively. Helices are represented as in (A) based on the crystal structure of Spo0J of T. thermophilus (34); amino acid sequence are numbered according to the Spo0J. The red plus sign (+) schematically represents the conserved R or K positively charged residues. ParBs are from Borrelia burgdorferi (Bbur), B. cenocepacia (Bcen), B. subtilis (Bsub), Campylobacter jejuni (Cjej), Campylobacter crescentus (Ccre), Chlamydia trachomatis (Ctra), Helicobacter pylori (Hpyl), Mycobacterium tuberculosis (Mtub), Myxococcus xanthus (Mxan), Neisseria meningitidis (Nme), P. putida (Pput), Porphyromonas uenonis (Puen), Rickettsia typhi (Rtyp), Staphylococcus aureus (Saur), Streptomyces coelicolor (Scoe), Streptococcus pneumoniae (Spne), T. thermophilus (Tthe) and Vibrio cholera (Vcho).
bacterial phyla (Figure 6C). As for SopB, this residue present in the two chromosomal Par systems tested here is essential for the partition function (Table 1) and is absolutely required for binding to its cognate centromere sequence (Figure 5). In addition, it belongs to a newly identified and well-conserved α-helical motif, GHAR/A/L. This motif aligns with helices H5 and H8 of SopB and KorB, respectively, which are also essential for centromere binding. We, therefore, named this new motif CBM2 for centromere-binding motif 2. We proposed that chromosomal ParB also harbor an extended centromere-binding domain composed of two DNA-binding motifs, the classical HTH (CBM1) and the newly identified CBM2.

We were unable to identify the CBM2 box and the positively charged residue on annotated chromosomal ParBs in only a few phyla, e.g. in cyanobacteria. The CBM2 motif, GHAR/[A]/L, is also not conserved on ParBs from secondary chromosomes in the case of bacteria carrying two or three chromosomes (data not shown). This could be because of divergent parS sites present on these chromosomes. Indeed, in the case of secondary chromosomes, parS sequences are different from the consensus one present on the main chromosome (14,39). It will be of interest to investigate whether a second DNA-binding motif is present on these Par systems.

In the case of type Ib ParBs that carry RHH DNA-binding motifs in place of an HTH, no co-crystal structure is available to determine whether specificity determinants are also present outside the primary binding motif. However, further DNA contacts providing additional binding specificity have been suggested by comparison of ParG, the ParB analogue of TP228, with the Arc and Mnt repressors harboring RHH motifs (40), supporting the possibility that type Ib ParBs could also carry extended centromere-binding domains.

Initially, the HTH motif of SopB was identified as the only determinant for centromere-specific binding (23). Our results along with these examples indicate that ParB–parS interactions are not as simple as initially proposed. Rather they suggest that bacterial CBPs have evolved, using different strategies, to possess extended DNA-binding domains. We thus propose that the requirement for at least two independent DNA-binding motifs is imposed by the need for a high centromere-binding specificity. Mutational analysis and co-crystal structure determination of other ParB–parS complexes, especially from chromosomal Par systems but also from ParBs with RHH, will be of particular interest to determine the extent of the different strategies developed for providing such high specificity for centromere binding.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–4 and Supplementary References [13,29,30,41–47].

ACKNOWLEDGEMENTS

The authors thank S. Depaul for technical assistance with stability assays and media preparation, N. Seddiki for help with PyMol software, Y. Ah-Seng for providing pYAS47 before publication, V. Morales for her expertise in protein purification and all members of the ‘segregation’ team for discussions. They are grateful to C. Johnson and reviewers for improvements in the manuscript.

FUNDING

Agence National pour la Recherche [2010 BLAN 1316 01]; Fonds European de Developpement Regional (for microscope equipments); Association pour le Recherche sur le Cancer (for microscope equipments). Funding for open access charge: Agence National pour la Recherche [2010 BLAN 1316 01].

Conflict of interest statement. None declared.

REFERENCES

18. Williams,D.R., Macarney,D.P. and Thomas,C.M. (1998) The partitioning activity of the RK2 central control region requires only incC, korB and KorB-binding site O(B)3 but other KorB-binding sites form destabilizing complexes in the absence of O(B)3. Microbiology, 144, 3369–3378.