Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017

Rebecca L Dineen, Christophe Penno, Philip Kelleher, Maxence J B Bourin, Mary O'connell-Motherway, Douwe van Sinderen

microbial biotechnology

Open Access

Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017

Rebecca L. Dineen,[1,2] Christophe Penno,[3] Philip Kelleher,[1,2] Maxence J. B. Bourin,[1,2] Mary O'Connell-Motherway[1] and Douwe van Sinderen[1,2]*

[1]*APC Microbiome Ireland, University College Cork, Western Road, Cork, Ireland.*

[2]*School of Microbiology, University College Cork, Western Road, Cork, Ireland.*

[3]*CNRS UMR 6553 EcoBio, Universite de Rennes 1, Campus de Beaulieu, Bat. 14A, Rennes cedex, 35042, France.*

Summary

pMP7017 is a conjugative megaplasmid isolated from the gut commensal *Bifidobacterium breve* JCM7017 and was shown to encode two putative replicases, designated here as RepA and RepB. In the current work, RepB was identified as the pMP7017 replicative initiator, as the *repB* gene, and its surrounding region was shown to be sufficient to allow autonomous replication in two bifidobacterial species, *B. breve* and *Bifidobacterium longum* subsp. *longum*. RepB was shown to bind to repeat sequence downstream of its coding sequence and this region was determined to be essential for efficient replication. Based on our results, we hypothesize that pMP7017 is an *i*teron-*r*egulated *p*lasmid (IRP) under strict auto-regulatory control. Recombinantly produced and purified RepB was determined to exist as a dimer in solution, differing from replicases of other IRPs, which exist as a mix of dimers and monomers. Furthermore, a stable low-copy *Bifidobacterium-E. coli* shuttle vector, pRD1.3, was created which can be employed for cloning and expression of large genes, as was demonstrated by the cloning and heterologous expression of the 5.1 kb *apuB* gene encoding the extracellular amylopullulanase from *B. breve* UCC2003 into *B. longum* subsp. *longum* NCIMB8809.

Received 16 January, 2021; accepted 22 March, 2021.
For correspondence. *E-mail d.vansinderen@ucc.ie; Tel. + 353 21 4901365; Fax + 353 21 4903101.
Microbial Biotechnology (2021) **14**(4), 1494–1511
doi:10.1111/1751-7915.13810
Funding information
No funding information provided.

Introduction

Bifidobacteria represent anaerobic, non-motile, high G + C% Gram-positive bacteria that belong to the phylum Actinobacteria. Bifidobacteria are considered key commensals in microbe-host interactions, purported to play a pivotal role in maintaining gastrointestinal health through various beneficial or probiotic activities (Leahy *et al.,* 2005; O'Callaghan and van Sinderen, 2016). They have been reported to reduce the risk of microbial infections by enforcing epithelial barrier function through acetate production and secretion of lipophilic antimicrobials (Lievin, 2000; Fukuda *et al.,* 2011). In addition, certain bifidobacterial strains have been shown to interact with their host *in vivo* via extracellular surface structures (Fanning *et al.,* 2012; Turroni *et al.,* 2013; Schiavi *et al.,* 2016; O'Connell Motherway *et al.,* 2019) and through the production of tryptophan catabolites, such as indole-3-lactic acid via aryl hydrocarbon receptor stimulation (Roager and Licht, 2018; Wong *et al.,* 2020).

These reported health benefits have been the driving force behind a wave of research activities to study the biology and genetics of bifidobacteria and the way in which they interact with their host. Currently, knowledge on the health benefits of bifidobacteria is for a considerable part deduced from *in silico* analyses due to the recent surge of available bifidobacterial genome sequences, a 16.5-fold increase in four years from 62 (O'Callaghan and van Sinderen, 2016) to 1021 (https://www.ncbi.nlm.nih.gov/taxonomy, accessed May 2020) and to a lesser degree based on findings from functional analyses (Cronin *et al.,* 2007; O'Connell Motherway *et al.,* 2009; Ruiz *et al.,* 2013; Sakanaka *et al.,* 2018; Koguchi *et al.,* 2020). The paucity of functional data is partly due to the inherent resistance of *Bifidobacterium* to the artificial uptake of DNA due to natural barriers such as restriction-modification systems (Bottacini *et al.,* 2018), but most significantly due to the scarcity of molecular techniques and tools suited to these bacteria (Ventura *et al.,* 2004; Cronin *et al.,* 2011; O'Callaghan and van Sinderen, 2016). Plasmids represent one of these crucial molecular tools, playing an integral role in the expansion of our understanding of gene function. This is in part due to their small size and relatively simple genetic organization, in combination with the ease with which they can be isolated and manipulated *in vitro* without adversely affecting the host (del Solar *et al.,*

1998). Plasmids are generally defined as autonomously replicating, extrachromosomal genetic elements, which are not essential to host survival, and which may carry a plethora of advantageous functions for rapid environmental adaptation (Sobecky, 1999; Heuer and Smalla, 2012; Dziewit *et al.,* 2015). Plasmids require one or more genes and/or *cis*-acting sequences, together termed the basic replicon, for replication and faithful transmission to each daughter cell during cell division. The genetic composition of such a basic replicon can vary widely and depends on the mode of replication.

Plasmids were originally thought to be relatively rare in bifidobacteria (Lee and O'Sullivan, 2010), yet a total of 49 plasmids of bifidobacterial origin have been identified to date, isolated from 12 species and ranging in size from 1.8 kb in the case of plasmid pMB1 isolated from *Bifidobacterium longum* subsp. *longum* (Rossi *et al.,* 1996) to 190 kb in the case of pMP7017 from *Bifidobacterium breve* (Bottacini *et al.,* 2015). Furthermore, recent evidence indicates a higher prevalence of megaplasmids in this genus due to the identification of pMP7017 homologues in fifteen *B. longum* strains (Odamaki *et al.,* 2018). The word 'megaplasmid' was initially coined to describe a 450 kb plasmid identified in the Gram-negative rhizosphere bacterium *Sinorhizobium meliloti* in 1981 (Rosenberg *et al.,* 1981), and in publications to date has been used to describe plasmids $>$ 100 kb (Schwartz, 2009), which appear in both phylogenetically and physiologically diverse groups of archaea and bacteria. While in-depth investigations have been conducted to characterize the replication functions of megaplasmids from *Rhizobia* (López-Guerrero *et al.,* 2012; Lagares *et al.,* 2014) and *Vibrio cholera* chromosome II (Stokke *et al.,* 2011; Baek and Chattoraj, 2014; de Lemos Martins *et al.,* 2018; Fournes *et al.,* 2018), relatively little is known about the replication mode of actinobacterial megaplasmids.

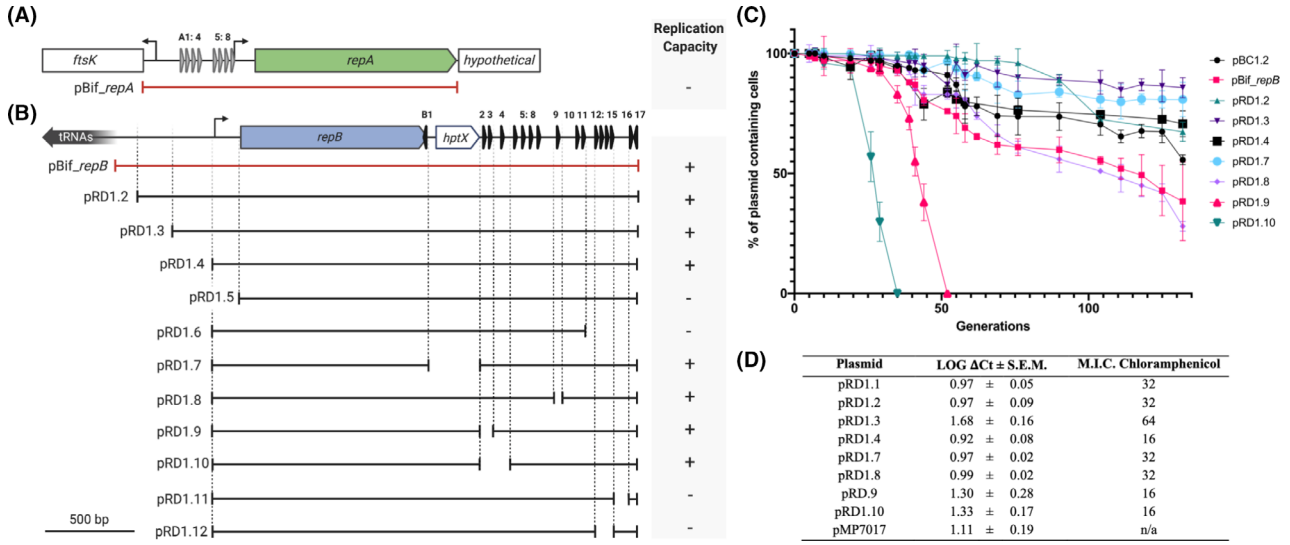
The 190 178 bp conjugative megaplasmid pMP7017 of *B. breve* JCM7017 represents the first and largest, conjugative megaplasmid identified among the actinobacterial gut commensals (Bottacini *et al.,* 2015). The report suggests that pMP7017 is the result of a co-integration of two plasmids, a notion based on the finding that there is a 4.8% deviation in G + C mol% from one arbitrary half to the other, while pMP7017 is also predicted to encode two replicative initiator proteins, denoted here as RepA and RepB. These putative replicases do not correspond or display homology to proteins from the RepABC family of plasmids; thus, this nomenclature is purely for the purpose of differentiation. In the current report, we investigate the replication functions of the conjugative bifidobacterial megaplasmid pMP7017 in an effort to further our understanding of the replicative strategies of this and, by inference, other large plasmids. While research efforts to date have primarily focused on plasmid replication in Gram-negative and pathogenic Gram-positive bacteria, the current work serves to address the knowledge gap of these systems in Gram-positive commensals.

Results

In silico *analyses of the regions flanking* repA *and* repB

As data pertaining to replication features of Gram-positive megaplasmids and bifidobacterial replicons are scarce, we focused our attention on two previously identified putative replicase genes (Bottacini *et al.,* 2015) to assess the intrinsic features of pMP7017 replication. Inspection of the *repA* gene (nucleotide coordinates 97 215…96 166; GenBank Accession No. KM406416) and surrounding sequences identified features characteristic of plasmid replicons within the 5' region of the *repA* coding sequence, such as an A + T rich region, a 35 bp palindrome and eight 15 bp direct repeats, designated here as A1 through to A8, with the consensus 5′-R**G**Y**GTCA**RW**ATTCTA**-3′ (Fig. 1A). An exploration of the non-coding region upstream of the *repB* gene (nucleotide coordinates 120 394…119 397) did not unveil any features typical of a plasmid origin. *In silico* scrutiny of the sequence located downstream of *repB* revealed a small 267 bp open reading frame (ORF) encoding a hypothetical, basic (pI 11.2) 10 kDa protein without any discerning motifs and referred to here as *hptX*. Furthermore, seventeen irregularly spaced 20 mer repeats, designated B1 through to B17 (Fig. 1B), were identified conforming to the consensus 5′-hvr**TGG**m**CTATC**rs**TG** hmdh-3′ with an additional sequence, B18, identified outside of the repeat region (nucleotide coordinates 115 314…115 334). Analysis of these repeats revealed that 20 mers B5 and B9 differ in their boundary sequences from the other repeats which appear to contain overlapping inversely orientated sequences, 5′- TTTTACACC-3′, resembling those of putative DnaA boxes observed in the chromosomal origin of *Bifidobacterium* species (Lee and O'Sullivan, 2010).

When replication occurs bi-directionally from a single origin such as with theta-type replication, there are two replication forks of opposite polarity, which have been shown to exhibit a GC-skew (Arakawa and Tomita, 2012; Zhao *et al.,* 2015; Zhang and Gao, 2017). Thus, in an effort to define the location of the replication origin and terminus the GC-skew properties across pMP7017 were plotted. A clear asymmetry was observed with two strong polarity switches at nucleotide positions 117 611 bp and 37 431 bp, presumably representing the approximate locations of the replication origin and terminus respectively. This trend in disparity is reflected in a Z-curve analysis which predicts the *oriV* to be located at position 120 395 bp (Luo *et al.*, 2019). These findings, therefore, suggest that the replication origin is located in the vicinity of *repB*.



(D)

Plasmid	LOG ΔCt ± S.E.M.	M.I.C. Chloramphenicol
pRD1.1	0.97 ± 0.05	32
pRD1.2	0.97 ± 0.09	32
pRD1.3	1.68 ± 0.16	64
pRD1.4	0.92 ± 0.08	16
pRD1.7	0.97 ± 0.02	32
pRD1.8	0.99 ± 0.02	32
pRD.9	1.30 ± 0.28	16
pRD1.10	1.33 ± 0.17	16
pMP7017	1.11 ± 0.19	n/a

Fig. 1. pRD derivative plasmids; stability and copy number.
A. Representation of the genomic architecture of *repA* and flanking regions. Grey arrowheads represent eight direct repeats preceding and overlapping the *repA* promoter, A1-A8. The region tested for replication is indicated by a red line.
B. Genetic organization of *repB* and surrounding regions. tRNAs are depicted as black arrows, while the seventeen 20 mer repeats, B1-B17, are represented as black arrowheads. The region tested for replication is indicated by a red line. A schematic representation of the pMP7017 replicon within each pRD1 derivative plasmid is depicted below.
C. Graph illustrating the segregational stability of the pRD1 derivatives over 132 generations, each data point is averaged from three independent replicates, and standard deviations are indicated as error bars.
D. Table of pRD1 plasmid copy numbers and MIC of chloramphenicol resistance.

Analysis of the protein sequence of RepA revealed a RepA_C domain (PF04796) at the N-terminus (interval: 57–215 aa, *E*-value: 1.6^{-16}). RepB also presents with this domain at its C-terminus (PF04796, interval: 193–361aa, *E*-value: 1.1^{-12}), which is part of the Rep_3 superfamily (CDD: cl19398, interval: 193–361aa, *E*-value: 5.09^{-08}). Members of this family of replicases are found in a number of IRPs from Gram-negative bacteria such as RK2 (Fang and Helinski, 1991), R6K (Filutowicz *et al.,* 1985), pSC101 (Kazuo and Mitsuyo, 1984), F (Murotsu *et al.,* 1981) and pPS10 (Kazuo and Mitsuyo, 1984), and low GC, non-conjugative plasmids of Gram-positive origin such as pSK639 from *Streptococcus epidermidis* (Apisiridej *et al.,* 1997). Homology modelling further supports this finding, as 57% of the total amino acid sequence of RepB (residues 102–291) was modelled heuristically at > 90% confidence, using PDB templates of the C-terminus of IRP initiators RepE from F plasmid and the R6K π monomer, represented by c2z9oB (Nakamura *et al.,* 2007) and c2nraC (Swan *et al.,* 2006) respectively. RepB is therefore structurally most similar to replicases of IRPs, which are known to replicate by means of the theta-type mechanism.

The repB *gene encodes the pMP7017 replication initiator protein*

To establish the potential involvement of either (or both) of the putative replication initiator-encoding genes in pMP7017 replication, the *repA* and *repB* genes, in addition to the regions flanking their coding sequences, were individually cloned into the bifidobacterial replication probe vector pBif (Alvarez-Martín *et al.,* 2007), generating pBif_*repA* and pBif_*repB* respectively. Recombinant plasmids for each of the replication genes were successfully obtained in *E. coli* and were then introduced into *B. breve* UCC2003 by electroporation to assess their ability to independently facilitate replication. Failure to obtain *B. breve* UCC2003 pBif_*repA* transformants suggests that the 1829 bp fragment encompassing *repA* and upstream repeat region was insufficient to allow autonomous replication (Fig. 1A). In contrast, recovery of *B. breve* UCC2003 transformants containing pBif_*repB* demonstrated that this 3258 bp region harbours all relevant features to facilitate autonomous replication in *Bifidobacterium* and thus is assumed to represent the basic pMP7017 replicon (Fig. 1B). To corroborate this finding, we adopted a strategy of site-directed insertional mutagenesis to investigate the dispensability of either *repA* or *repB* for replication using a non-replicating vector which integrates into the targeted gene via homologous recombination. Successful disruption of the *repA* coding sequence, as confirmed by Illumina sequencing, demonstrates that this gene is not essential for pMP7017 replication, while the inability to obtain a *repB* mutant despite multiple efforts indicates the indispensability of *repB*.

Iterons B12–B15 and flanking AT-rich region are essential cis-acting elements for pMP7017 replication

To determine the minimal boundaries of the functional *repB*-associated replicon, successively shorter fragments of the 3258 bp *repB*-containing replication region were cloned into the *Bifidobacterium* replication probe vector pBif and assayed for their capacity to support autonomous replication in *B. breve* UCC2003 (Fig. 1B). In this manner, the minimal replicon was determined to be the 2366 bp fragment of pRD1.4, which contains the *repB* promoter and coding sequence, *hptX* and seventeen 20 mer repeats. In order to determine the essentiality of individual iterons and the co-transcribed *hptX,* five additional constructs were made, pRD1.7 – pRD1.12, represented graphically in Fig. 1B. Plasmid derivatives pRD1.7, pRD1.8, pRD1.9 and pRD1.10 were replication-proficient in *B. breve* UCC2003, whereas derivatives pRD1.11 and pRD1.12 failed to replicate. It can be deduced from these data that the *hptX* coding sequence and iterons B2, B3, B4 and B9 are dispensable for replication and that iterons B12-B15 and flanking AT-rich region represent essential *cis*-acting elements.

Iterons B2, B3 and B4 confer stability of the pMP7017 replicon

While the *hptX* gene, iterons B2, B3 and B4, and iteron B9 individually were shown to be non-essential for pMP7017 replication, these features may nonetheless play a role in other functions, such as plasmid segregation, stability and/or copy number. To ascertain the role of these elements, if any, in the stability of the replicon, *B. breve* UCC2003 harbouring each of the replication-proficient pRD derivatives was grown in the absence of selection pressure for 132 generations and then assayed for the presence of the plasmid as a measure of stability, presented in Fig. 1C. These data show that plasmids containing all putative iterons, pBif_*repB*, pRD1.2, pRD1.3, pRD1.4 and pRD1.7, are quite stable as they are retained in the population for at least 132 generations without selective pressure. Instability was observed for plasmids pRD1.9, lacking two iterons (B2 and B3), and pRD 1.10 lacking three iterons (B2, B3 and B4) with a total loss of the plasmid occurring at ≤ 52 and ≤ 35 generations respectively. Furthermore, the absence of iteron B9 from plasmid pRD1.8 was shown to cause only a minimal loss of stability. The removal of the *hptX* gene appears to have no adverse effect on stability nor does the absence of the region located 5′ of the *repB* promoter. Plasmids pRD1.2 and pRD1.3 were shown to be the most stable among the assessed plasmids as they were maintained in $> 85\%$ of cells at 100 generations. Together these findings indicate that the iterons contribute to plasmid stability and highlight the 2542 bp fragment of pRD1.3 as the basic replicon due to the greater stability of this plasmid derivative, apparently containing all features necessary to support replication and ensure stable maintenance.

Iterons B2, B3, B4 and B9 do not participate in copy control

The plasmid copy number (PCN) of the eight replicating pRD derivatives and megaplasmid pMP7017 was determined by the relative quantification method using single-copy housekeeping genes on the *B. breve* UCC2003 chromosome as a reference. As data displayed in Fig. 1D illustrate, the PCN of all pRD1 plasmids approximate 1 copy per chromosome, with the exception of pRD1.3 with a somewhat higher value of 1.68 ± 0.16. These data illustrate the copy number of all pRD1 derivates is comparable to that of pMP7017 from which the replicon derives, existing at approximately 1 copy/chromosomal equivalent in *B. breve* UCC2003, irrespective of the absence of the 5′ region, *hptX* or iterons B2, B3, B4 and B9, thus suggesting these features do not play a major role in copy control. These non-essential iterons may nonetheless encode functions not explored in this work, considering their lack of participation in copy control they will subsequently be referred to as direct repeats.

repA *and* repB *are transcriptionally regulated in pMP7017*

To gain insight into *repA* and *repB* transcription and possible transcriptional regulation, the presumed promoter regions and 5′ UTRs of each *rep* gene were cloned upstream of the promoterless β-glucuronidase (β-GUS)-encoding gene in the reporter plasmid pNZ272 (Platteeuw *et al.,* 1994; Cronin *et al.,* 2007), generating plasmids pNZPP_*repA* and pNZPP_*repB* respectively. These plasmids were individually introduced into *B. breve* UCC2003 and *B. breve* UCC2003 pMP7017_199, facilitating transcriptional quantification of these promoter regions in both the absence and presence of the megaplasmid. *B. breve* UCC2003 pMP7017_199 contains a tetracycline-tagged pMP7017 derivative with an insertion in ORF199 (Bottacini *et al.,* 2015) encoding a surface-associated protein which displays no phenotypic deviation from the wild type and is used in the current work for the purpose of antibiotic-based plasmid selection. A bar chart of the data in Fig. S1A illustrates a moderate level of transcriptional activity from both promoters in pNZ272 when present in *B. breve* UCC2003 pMP7017_199, while the intracellular β-GUS activity of the pNZ272 transcriptional fusions in a

B. breve UCC2003 background was higher. *B. breve* UCC2003 pNZPP_*repA* revealed a marked increase in activity ~ 3.7-fold from its *B. breve* UCC2003 pMP7017_199 counterpart, while activity from the *repB* promoter increased ~1.6-fold in the absence of the megaplasmid. These findings indicate that the promoters of both *repA* and *repB* are repressed in strain *B. breve* UCC2003 pMP7017_199, suggesting that pMP7017 encodes trans-acting factors negatively regulating *repA* and *repB* transcription. To precisely define the location of the *repA* and *repB* promoters, the *repA* and *repB* transcription start sites (TSSs) were experimentally determined by primer extension (PE) analysis. A single extension product was obtained for *repA* corresponding to a position 71 nucleotides upstream of the predicted translational start site (Fig. S1B). Investigation of the preceding sequence revealed hexamers resembling the consensus −10 and −35 recognition sequences of bifidobacterial vegetative promoters (Bottacini *et al.,* 2017). PE analysis for *repB* resulted in an extension product 131 nucleotides 5′ of the predicted translational start site (Fig. S1CC) that is associated with a putative vegetative promoter sequence.

RepA binds to its corresponding promoter region

While it was unclear what role, if any, RepA plays in pMP7017 replication, mobility shift electrophoresis was performed to determine whether the RepA protein interacts with probe 124P, containing the *repA* promoter and eight direct repeats A1–A8. The electrophoresis mobility shift assay (EMSA) depicted in Fig. 2A clearly shows the reduced mobility of DNA probe 124P upon addition of RepA, highlighting the affinity of RepA for this region. Assuming that the eight direct repeats contained within the probe are the recognition sequences of RepA, the observation of one strong shift despite eight potential binding sites suggests that RepA binding involves cooperativity, hence site saturation appears instantaneous rather than incremental. To further elucidate this interpretation, the fraction of bound probe (Φ) was calculated and plotted as a function of protein concentration to generate a specific binding curve. From the curve obtained, we can see that RepA-124P complex formation is a non-linear function of RepA concentration consistent with cooperative interaction, and further analysis revealed a Hill coefficient of 6.27 ± 1.36 reflecting marked positive cooperativity. Together this substantiates the notion that binding of RepA to each of the eight repeats within probe 124P does not occur independently. Furthermore, an apparent K_d of ~ 121.8 nM $\pm$ 8.0 nM was determined, from this the K_a was calculated to be 5.26 pM^{-1} suggesting a high affinity of RepA for its DNA substrate.

RepA exists predominantly as a homodimer in solution

The pronounced cooperativity observed in RepA binding to its cognate recognition sequence signalled a propensity for RepA-RepA interaction. To establish its oligomeric state in solution, size exclusion chromatography was performed. From the chromatogram (Fig. S1D), one main peak was observed at 17.87 RV (ml) representing a molecular weight of 82.49 kDa and 67.31 % of the total protein, suggesting RepA exists predominantly as a homodimer. Two additional peaks can be seen at 164.8 kDa (17.23 RV; representing 16.98% of total protein) and 328 kDa (16.57 RV and representing 15.71% of the total protein) which we presume denote tetrameric and octameric forms of RepA, respectively, thus illustrating the multimerization ability of this protein.

RepB does not display any affinity for its promoter or coding region

In an effort to ascertain where within the pMP7017 replicon RepB binds and to assess whether the presumed RepB replication protein was able to physically interact with any of the identified DNA features flanking its coding sequence, nine IRD700-labelled probes were generated, designated here as A through to I, encompassing various segments of the replicon (Table S2). EMSAs with probes A and B, encompassing the *repB* promoter-containing region, and probes C and D, spanning the *repB* coding sequence, revealed an absence of binding activity (Fig. S2).

RepB binds specifically to iterons

In order to establish whether RepB binds directly and specifically to the putative iterons identified 3′ of its coding sequence, EMSAs were performed with DNA probes E through to I, each containing ≥ 1 20 mer repeat. EMSAs with each probe within this region resulted in clear mobility shifts, indicating that RepB is able to bind to the repeat-containing region downstream of the *repB* gene. EMSAs with short DNA fragments E and H (Fig. S3A and B), each containing an individual 20 mer site, resulted in a single mobility shift interpreted as a single species of protein–DNA complex, RepB-E and RepB-H respectively. To determine the quantitative parameters of these interactions, a titration curve was generated from the experimental data and fit to the Hill equation revealing a coefficient of ~ 1 in both cases with an apparent K_d of 254 $\pm$ 34 nM and 236.6 $\pm$ 66.5 nM for probe E and probe H respectively. This result suggests an approximate K_a of 4.1 pM^{-1} $\pm$ 0.3 pM, reflecting high-affinity binding of RepB to the 20 mer sites B1 and B9.

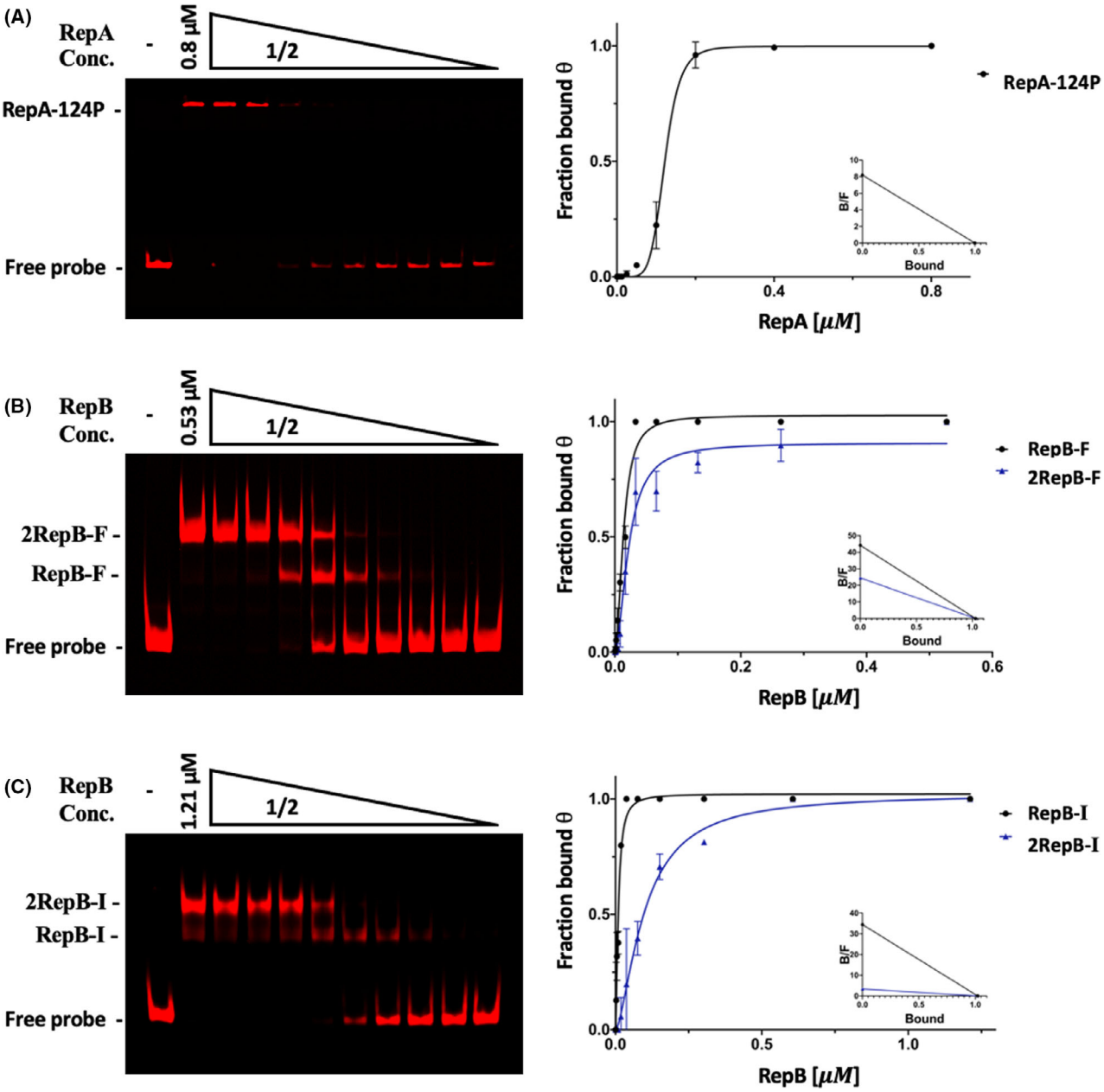


Fig. 2. The binding behaviour of RepA to its corresponding promoter region and of RepB to multi-site probes in EMSA.
A. EMSA image displaying the interaction of RepA with promoter probe 124P. The binding isotherm and Scatchard plot (inset) of this data are presented to the right.
B. A representative EMSA gel displaying two clear shifts representing the binding of RepB to the 2-site probe, F, containing two tandemly arranged iterons. The binding isotherm and Scatchard plot (inset) of this data are presented to the right.
C. EMSA image of RepB binding to 2-site probe, I, containing two invertedly orientated iterons. The binding isotherm and Scatchard plot (inset) of this data are presented to the right.

Additionally, the near equivalent estimations of dissociation suggest no predilection in terms of site orientation, while highlighting the interchangeability of residues 13 (A/G), 14 (C/G) and 18 (C/A) within the apparent recognition sequence, without any significant effect on affinity estimates.

RepB exists as a dimer in solution with no monomeric fraction

While quantitative evaluation of the binding isotherms allows for affinity approximation and estimation of the effective cooperativity behaviour, it was necessary to

determine the oligomeric state of the protein to support our interpretation of the graphical presentations of the ligand-binding data. To achieve this, size exclusion chromatography was performed, revealing that when RepB was assayed at a high protein concentration (0.8 mg ml^{-1}), it exists as a homodimer in solution illustrated by one main peak at 18.75 retention volume in the chromatogram (Fig. S1E) representing a molecular weight of 79.85 kDa and 97% of the total sample.

The binding behaviour of RepB to multi-site probes illustrates cooperative interactions

The EMSA (Fig. 2B) employing DNA probe F containing presumed binding sites B2 and B3 clearly depicts two distinct electrophoretic species, which can be explained by assuming that the initial shift is the result of a single binding event, while the double shift occurs due to subsequent RepB binding to the second site, representing nucleoprotein complexes RepB-F and 2RepB-F respectively. Analyses of the titration curve data generated from the binding reactions suggest positive cooperativity ($nH = \sim 2$), with an apparent dissociation constant of 13.7 nM $\pm$ 1.8 nM for RepB-F and 21.8 nM $\pm$ 4.9 nM for 2RepB-F. Correspondingly, the EMSA displayed in Fig. S3C demonstrates incremental binding of RepB to the 20 mer sites B2, B3 and B4 within probe G. Global sensitivity measures of the specific binding curves reveal positive cooperativity in the formation of complexes RepB-G ($nH = 1.5 \pm 0.34$) and 2RepB-G ($nH = 2.8 \pm 0.62$); conversely, a Hill coefficient of ~1 associated with 3RepB-G suggests independent binding at a third site. In consolidation with the results obtained from the EMSA employing DNA fragment F, it can be inferred that positive cooperativity occurs between sites B2 and B3 due to their relative proximity (3 bp intervening), while site B4 is bound independently. Although cooperativity has been indicated for contiguous tandem sites, it was interesting to establish what binding behaviour would be observed with convergent proximal sites on opposing strands, a spatial arrangement observed for sites B16 and B17. EMSAs with probe I containing these sites display the formation of two protein–DNA complexes, RepB-I and 2RepB-I (Fig. 2C). The persistence of nucleoprotein complex RepB-I, despite increasing RepB concentration, may be indicative of a faster K_{off} rate, potentially as a result of steric hindrance. Together these data illustrate an increased affinity of RepB for one recognition sequence in the presence of a second adjacent site, being reflected by an average ~ 20.7-fold reduction in the apparent K_d of the first binding event in addition to $nH \geq 2$ suggesting two interacting sites.

GGnCnAnnR are critical bases within iterons for binding of RepB

To verify that RepB specifically binds to the putative iterons downstream of its coding sequence, thirteen core nucleotides (4–16) within the 20 mer motif were individually substituted by PCR site-directed mutagenesis, and each probe (P1–P18) was then assessed for its ability to bind RepB by EMSA. The results of this assay, displayed in Fig. 3, revealed that substitutions of residues 5G>T (P2), 6G>T (P3), 8C>A (P5), 10A>C (P7) and 13G>T (P10) completely abolished the ability of RepB to bind, indicating that these residues are essential for RepB–iteron complex formation. Furthermore, the substitution of the remaining residues resulted in RepB–iteron binding events exhibiting a range of much-lowered affinities, which suggests a role in the stability of the complex. These data clearly show that alterations within the motif impact on RepB-mediated DNA binding affinity and indicate that RepB specifically binds to GGnCnAnnR.

The RepB C-terminal domain is required for dimerization and DNA binding

In silico structural homology modelling revealed a strong positive correlation of the C-terminus of the predicted 3D structure of RepB to the conserved iteron-binding domains of plasmid initiators RepE from F plasmid (Nakamura *et al.,* 2007) and π from R6K (Swan *et al.,* 2006) (Fig. S4A). Furthermore, structural analyses indicate RepB to be pseudo-symmetric with two putative DNA binding domains, a central winged helix-turn-helix 1 (wHTH1) and a C-terminal HTH2 flanked by a low complexity region rich in proline, serine and threonine residues (PST) (Fig. S4B). Oligopeptides enriched in these amino acids are known to form flexible structures which participate in ligand binding under specific conditions (Matsushima *et al.,* 2008); additionally, PST residues are known to serve as phosphorylation sites modulating protein activity (Yaffe and Smerdon, 2001; Johnson, 2009). To define the role of the C-terminal domain in both DNA binding and protein–protein interactions, a deletion mutant, RepB_ΔC, was created. RepB_ΔC was overexpressed and purified, followed by size exclusion chromatography to establish its oligomeric tendency in solution. From the chromatogram (Fig. S4C), one peaks at 18.7 RV (ml) representing 28.81 kDa and 100% of the sample, indicating RepB_ΔC exists as a monomer in solution. Purified RepB_ΔC was then tested for its ability to bind iterons by EMSA. The deletion variant showed no detectable binding to probe P17 as compared to

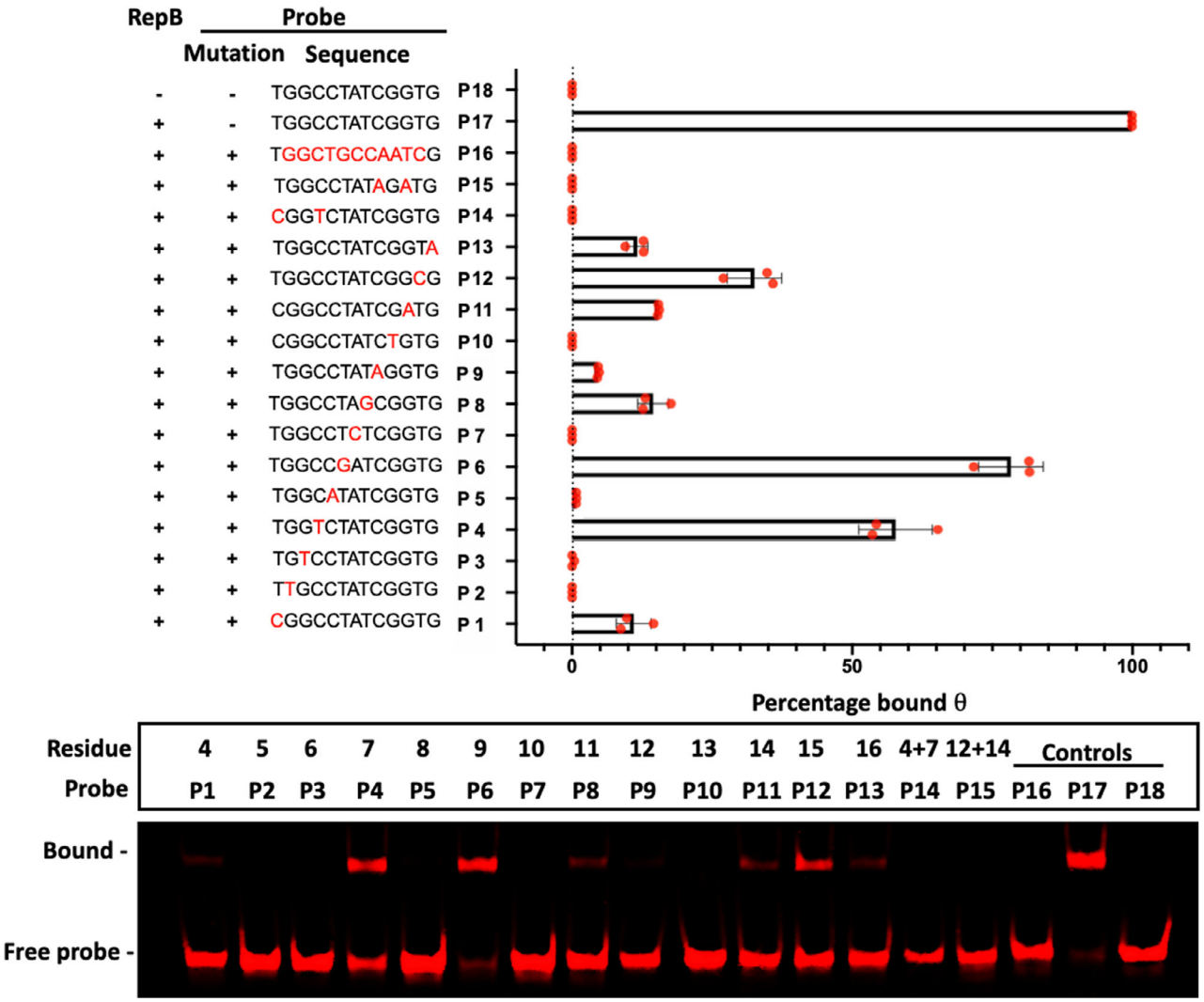


Fig. 3. Identification of the critical nucleotide residues for the binding of RepB to iterons. A bar chart representing the percentage of probe bound by RepB, the presence or absence of RepB in the reaction is denoted by a + or – symbol respectively. The sequence of each probe is given, and the incorporated mutation/s are indicated in red text. An EMSA gel image representing a replicate from which the bar chart is based is displayed at the bottom.

the RepB control, revealing a functional requirement of the C-terminal 98 residues of the RepB protein in dimerization and DNA binding (Fig. S4D).

Highly conserved RepA and RepB homologues are present in all bifidobacterial megaplasmids

A comparative analysis of the RepA and RepB protein sequences in twelve recently identified bifidobacterial megaplasmids confirmed the presence of both RepA and RepB homologues in all plasmids. RepB is conserved at a high level (protein per cent identity range: 88.75–89.43%), while, interestingly, RepA shows a higher level of conservation (protein per cent identity range: 89.97–96.56%). This indicates a potential functional importance of RepA despite its dispensability for pMP7017 replication.

The putative pMP7017 origin of replication is conserved between bifidobacterial replicons

A comparative analysis of the nucleotide sequence of the pMP7017 basic replicon between bifidobacterial megaplasmids, displayed in Fig. S5, illustrates the preservation of the *repB* coding sequence and iteron-containing region. Each replicon retains a minimum of eleven iterons with between one and three additional iterons located outside the replicon (data not shown).

Megaplasmids from *B. longum* subsp. *longum* strains, MCC10043, MCC10070 and MCC10096 contain two genes, *orfA* and *orfB*, constituting an IS200 transposable element that is not present in the other plasmids. The *hptX* gene identified in the pMP7017 replicon is absent in all *B. longum* megaplasmids, supporting the experimental data that this hypothetical gene is not required for replication. In consideration of these data, the *hptX*, *orfA* and *orfB* coding sequences were removed and the nucleotide sequences aligned to better assess the iteron-containing region in terms of its conservation (Fig. S6). The alignment displays the absence of sequences corresponding to pMP7017 20mers B2, B6, B7 and B9 in one or more *B. longum* megaplasmids, while a moderate level of variability is observed in the intervening sequences between the first ten repeats (B1–B10) which is not observed between the remaining iterons. Furthermore, no sequence resembling the DnaA consensus motif was identified in the *B. longum* megaplasmids. Analysis of the curvature, bendability and GC content of this highly conserved region reveals a conspicuous peak of curvature (> 7.5) from positions 118 659 bp to 118 686 bp, overlapping iterons B12, B13 and B14, followed by a region of high bendability in the AT-rich region with a GC content 14% less than the average from positions 118 772 bp to 118 264 bp.

pRD1.3: a novel bifidobacterial shuttle vector for cloning of large fragments

To verify the suitability of the pRD1 plasmids for cloning and expression purposes in bifidobacteria, the *apuB* gene from *B. breve* UCC2003, encoding an extracellular type II amylopullulanase (Motherway *et al.,* 2008), was directionally cloned into the multiple cloning site of plasmid pRD1.3, resulting in plasmid pRD1.3_*apuB*. Both pRD1.3 and pRD1.3_*apuB* were introduced into *B. longum* subsp. *longum* NCIMB8809, a strain unable to metabolize pullulan (Ryan *et al.,* 2006; Arboleya *et al.,* 2018). *B. longum* NCIMB8809 pRD1.3 and *B. longum* NCIMB8809 pRD1.3_*apuB* were then assessed for their ability to utilize pullulan by measuring optical density after 24 h of incubation in MRS supplemented with the sugar. From Figure 4A, it is obvious that *B. longum* NCIMB8809 pRD1.3 shows little to no growth (final OD_{600} 0.13 $\pm$ 0.013), whereas *B. longum* NCIMB8809 pRD1.3_*apuB* displays substantial growth (final OD_{600} 1.80 $\pm$ 0.16). The results show that pRD1.3_*apuB* confers pullulan-dependent growth in *B. longum* NCIMB8809, which validates the capacity of pRD1.3 as a novel bifidobacterial cloning and expression vector (Fig. 4B). Plasmid pRD1.3 is presumed to carry all replication and stability functions of the pMP7017 megaplasmid, and it may therefore be possible to employ pRD1.3 for cloning large (> 25 kb and up to 100 kb) DNA fragments.

Discussion

In the current work, *repA* was found to be non-essential for pMP7017 replication through successful site-directed mutagenesis, in addition to the inability of *repA* and surrounding regions to support autonomous replication in a replicase-deficient vector. As this gene is not critical for pMP7017 replication, *repA* may be a gene remnant encoding a dysfunctional replicase despite its high level of conservation among bifidobacterial megaplasmids. RepA was determined to exist predominantly as a dimer in solution with a tendency towards multimerization and was shown to bind its cognate promoter with high affinity.

We show that replication of the bifidobacterial megaplasmid pMP7017 depends on the plasmid-encoded initiator RepB, which exists as a dimer in solution at high concentrations with no observable monomeric fraction. RepB was demonstrated to be competent in iteron binding, displaying both iteron-dependent protein–protein interaction and iteron-independent interactions. The critical residues within the iteron sequence which facilitates the binding of RepB were established to be GGnCnAnnR, situated asymmetrically across the core of the 20 mer iteron. In all characterized IRPs, with the exception of the R6K π initiator, Rep proteins bind iterons as monomers, while Rep dimers are incompetent in iteron binding (Kunnimalaiyaan *et al.,* 2004). The study of Rep–iteron interactions is predominantly achieved with the use of dimerization-defective Rep mutants, through the addition of molecular chaperones in the binding reactions (DasGupta *et al.,* 1993; Ishiai *et al.,* 1994; Chattoraj *et al.,* 1996) or the use of chaotropic agents to induce dimer dissociation and facilitate the accumulation of iteron-binding active monomers. Interestingly, HIS-tagged WT RepB was shown to associate with a high affinity to iterons in EMSA in the absence of stimulation by host cell chaperones or the use of chaotropic agents. In addition, we did not observe a reduction in binding upon increasing RepB concentration nor did we observe a second shift representing a higher-order nucleoprotein complex in EMSA with single-site probes. Thus, it is tempting to speculate that the binding observed in this work is due to dimeric RepB. If this assumption is true, binding of iterons by dimeric RepB may inhibit replication as observed in the replication of plasmid R6K (Kunnimalaiyaan *et al.,* 2004). Dimeric RepB may then rely on the participation of temporally expressed cell-cycle chaperones, such as DnaK/DnaJ for monomerization as described for many initiators of IRPs (DasGupta *et al.,* 1993; Chattoraj, 2000; Zzaman *et al.,* 2004; Nakamura *et al.,* 2007).

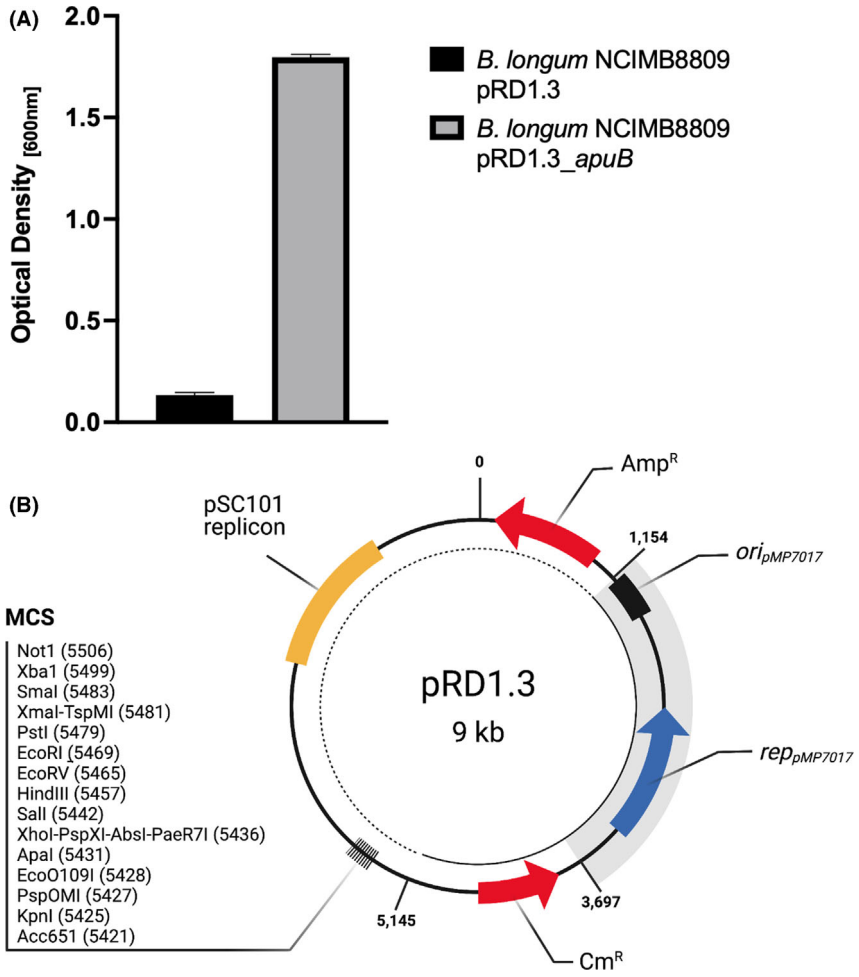


Fig. 4. Development of a novel bifidobacterial shuttle vector, pRD1.3.
A. Bar chart displaying the final optical density values of *B. longum* NCIMB8809 pRD1.3 and *B. longum* NCIMB8809 pRD1.3_*apuB* growth after 24 h of incubation with 10 % w/v pullulan as a sole carbon source.
B. Genetic map of plasmid pRD1.3 indicating the relevant features, the pMP7017 replicon is highlighted by a grey bar.

The C-terminal defective mutant, RepB_ΔC, was determined to be both monomeric and incompetent in iteron binding. While this suggests that the 98 aa C-terminal end of RepB is required for oligomerization and DNA binding, the observed phenotype may also be due to protein instability or indeed signify a lack of affinity of RepB monomers for the 20 mer sites in favour of alternative sites. It has previously been reported that even a single point mutation within the C-terminal domain of structurally similar replication initiators, i.e. π from R6K plasmid and RepE from F plasmid, is enough to severely affect or abolish the ability of the protein to bind iterons (Komori *et al.,* 1999; Swan *et al.,* 2006). Thus, a more conservative approach to mutagenesis is necessary to substantiate this finding.

While the pMP7017 minimal replicon was defined as a 2366 bp region (corresponding to pMP7017 nucleotide coordinates: 118 220. . .120 584) encompassing the *repB* coding sequence and seventeen repeats spanning 1159 bps, it was established that not all features contained within the replicon are essential. Deletion of the *hptX* coding sequence did not have any negative effect on replication. In addition, it was demonstrated that 20 mers B2, B3, B4 and B9 are not essential. While repeats B2-B4 were shown to participate in the stability of the replicon, the removal of these sequences displayed no perceived effect on plasmid copy number suggesting the pMP7017 copy control functions are encoded by alternative iterons. Contrastingly, the removal of iterons B12–B17 or the AT-rich region located between iterons B15 and B16 abolished the plasmid's capacity for replication, indicating this region contains essential *cis*-acting elements and thus likely constitutes the pMP7017 origin of replication. These data are supported by a comparative

analysis of the pMP7017 basic replicon between bifidobacterial megaplasmids which revealed a high degree of variably between and within repeats constituting the putative stability region, which is not observed in the highly conserved region containing the essential iterons B12–B17.

Although experimental validation is required to fully unravel the intricacies of pMP7017 replication, we here propose a mechanistic model for this bifidobacterial replicon (Fig. 5). We hypothesize that the pMP7017 iterons constitute a framework for the binding of RepB monomers that interact to form a structure that disrupts the DNA helix, facilitating DNA unwinding at the flanking AT-rich region. Reminiscent of the layout of DnaA boxes within the *oriC* of bacterial genomes, RepB iterons are arranged with varying affinities, distances and orientations suggesting a dynamic system regulating the formation of specific higher-order nucleoprotein complexes in the occlusion and stimulation of origin opening (Ekundayo and Bleichert, 2019). In addition to their importance for RepB assembly at the origin, the difference in sequence may constitute additional regulatory measures through weaker vs stronger binding sites. Furthermore, iteron-bound Rep protomers may inhibit origin melting as a result of coupling two origins in a process of replication inhibition, referred to as 'handcuffing'. This dual functionality of iterons is a copy control strategy commonly adopted by large, low copy number plasmids preventing plasmid over-replication through origin pairing (Chattoraj, 2000).

As the availability of genetic tools for the study of bifidobacteria is limited, the characterization of bifidobacterial replicons is necessary to develop new cloning and shuttle vectors. Considering the pMP7017 replicon is

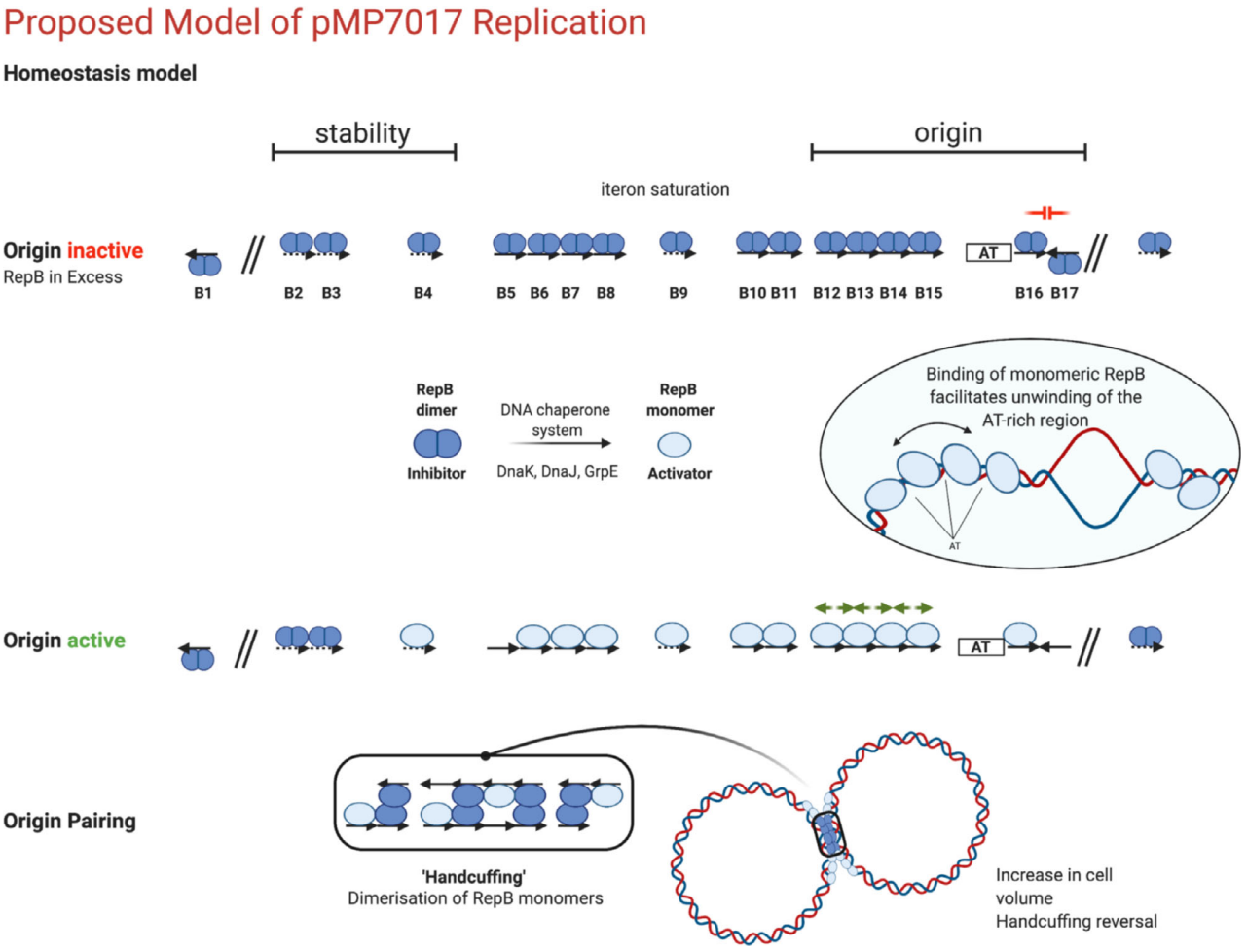


Fig. 5. Working model of pMP7017 replication initiation. Iteron saturation by RepB dimers prevents origin firing. Cell-cycle chaperones facilitate the conversion of RepB dimers to active monomers. Cooperative interaction (represented by green arrows) of RepB monomers bound to tandemly arranged iterons facilitates the unwinding of the flanking AT-rich region. While iteron-bound monomers on opposite template strands may participate in oligomerization to prevent over-replication in a process termed 'handcuffing', iterons are depicted as solid black lines and non-essential direct repeats are shown as broken arrows.

sufficient to drive replication of a plasmid of > 190 kb, it can be inferred that pRD1.3 represents a novel low-copy shuttle vector with the capacity to allow cloning of large fragments in bifidobacteria without instability or toxicity issues.

Experimental procedures

Bacterial strains, plasmids and cultivation conditions

Bacterial strains and plasmids are listed in Table S1. Bifidobacterial strains were routinely cultivated in reinforced clostridial medium (RCM [Oxoid, Basingstoke, Hampshire, United Kingdom]) or modified de Man, Rogosa and Sharpe (mMRS) medium prepared from first principles (Egan *et al*., 2018). The mMRS medium was supplemented with cysteine-HCl (0.05 % [wt/vol]) and lactose (1.0 % [wt/vol]) (Sigma) prior to inoculation. Bifidobacterial strains were cultivated in an anaerobic hood (Davidson and Hardy, Belfast, Ireland) for incubation at 37°C, maintaining an anaerobic environment. All *Escherichia coli* strains were cultivated exclusively in Luria–Bertani (LB) broth (Sambrook and Russell, 2001) at 37°C with agitation. Where applicable, growth media contained tetracycline (10 μg ml^{-1}), ampicillin (100 μg ml^{-1}), erythromycin (100 μg ml^{-1}), chloramphenicol (5 μg ml^{-1}) or kanamycin (50 μg ml^{-1}).

DNA manipulations and molecular cloning

Chromosomal DNA was isolated from *B. breve* JCM7017-199, a *B. breve* JCM7017-derivative strain containing a tetracycline resistance marker disrupting the open reading with locus tag pMP7017_199 of plasmid pMP7017, as previously described (Bottacini *et al.,* 2015). Plasmid DNA from *E. coli* was extracted using the High Pure Plasmid Isolation Kit (Sigma-Aldrich, Gillingham, UK). Standard oligonucleotides (Table S2) used in this study were synthesized by Eurofins (Ebersberg, Germany). Both standard and colony PCRs were performed using DreamTaq PCR 2× Master Mix (Thermo Fisher Scientific, Gloucester, United Kingdom), and colony PCRs were carried out as described previously (Mazé *et al.,* 2007). High-fidelity PCR was achieved using Q5 High-Fidelity 2X Master Mix (New England Biolabs, Hitchin, United Kingdom). PCR fragments were purified using the High Pure PCR Product Purification Kit (Sigma). Restriction endonucleases and T4 DNA ligase were used according to the manufacturer's instructions (Roche, Mannheim, Germany). Electroporation of plasmid DNA into *Bifidobacterium* spp. was conducted as previously described (Rossi *et al.,* 1997), while *E. coli* transformations were performed as described by Sambrook *et al*. (1989). All electroporation experiments were carried out using a GenePulser apparatus (Bio-Rad Laboratories, Richmond, CA, USA).

Insertional mutagenesis

Insertional mutagenesis was performed based on a previously described protocol (O'Connell Motherway *et al.,* 2009). For this purpose, a 525 bp internal fragment of ORF pMP7017_124 (*repA*) from codon positions 92 to 267 of the *repA* coding sequence and an internal 501 bp fragment of ORF pMP7017_146 (*repB*) from codon positions 101 to 268 of the *repB* coding sequence were amplified by high-fidelity PCR using B. *breve* JCM7017_199 total DNA as a template with primer combinations repAmut_F and repAmut_R, and repBmut_F and repBmut_R respectively (Table S2). The amplicons were ligated into the multiple cloning site of the integration vector pORI19 following HindIII and XbaI digestion. 5 μl of each ligation mixture was introduced into *E. coli* EC101 competent cells by electroporation. Recombinant *E. coli* derivatives were selected by blue-white screening on LB agar supplemented with 40 g ml^{-1} X-gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside) and 1 mM IPTG. Confirmation of recombination was achieved by restriction analysis prior to subcloning of the tetracycline resistance gene, *tetW*, from plasmid pAM5 into the unique SacI restriction site of each plasmid. The resulting constructs, designated pORI19_*repA* and pORI19_*repB*, were introduced into *B. breve* JCM7017 by electroporation, and transformants were selected on tetracycline-containing media. Site-specific recombination in such tetracycline-resistant transformants was confirmed by colony PCR with primers tetW_F and tetW_R to verify the presence of the *tetW* gene. Total DNA was isolated from such *tetW*-positive transformants and used as a template for PCR with primer pair *repA*conf and TetR_final to confirm integration within the expected genetic position of the megaplasmid. The *B. breve* JCM7017_*repA* mutant was then verified by genome sequencing performed employing an Illumina MiSeq Sequencing Platform.

Plasmid constructions

For the construction of plasmids pBif_*repA*, pBif_*repB*, pRD1.2, pRD1.3, pRD1.4, pRD1.5 and pRD1.6, various DNA fragments representing sections of the *repA* or *repB* gene and surrounding sequences were obtained by PCR amplification using total DNA isolated from *B. breve* JCM7017 as a template in all reactions and primer combinations as listed in Table S2. Recombinant DNA sequences for plasmids pRD1.7, pRD1.8, pRD1.9 and pRD1.10 were generated by overlap extension PCR containing two PCR segments with primer combinations

outlined in Table S2. Oligonucleotides were designed with an SphI restriction site incorporated at the 5′ end to allow direct cloning of amplicons. The generated amplicons and the bifidobacterial replication probe-*E. coli* vector pBif (Alvarez-Martín *et al.,* 2007) were digested with SphI and ligated. Ligation mixtures were dialysed and transferred into *E. coli* XL1blue competent cells by electroporation. pBif consists of a chloramphenicol resistance cassette derived from plasmid pC194 (Sangrador-Vegas *et al.,* 2007) in combination with the pSC101 replication functions of pWSK29. Colony PCR was performed to confirm the presence of the *repA* or *repB* gene using primer pairs pMP_0124F and pMP_0124R, or pMP_0146F and pMP_0146R, designated to generate a 237 bp or 244 bp amplification product respectively. Purified miniprep plasmid DNA of positive clones was used as a template for a second round of PCR with primer pRDconf_F in combination with the forward or reverse *repA* or *repB* cloning primers to determine the orientation of the gene, and plasmids containing *repA* or *repB* in the forward orientation were selected for further analysis.

Primer pair *apuB*F_Xba and *apuB*R_H3 was used to amplify a 5437 bp region containing the *apuB* coding sequence and promoter region from *B. breve* UCC2003 genomic DNA. Both the resulting *apuB*-encompassing amplicon and pRD1.3 plasmid were digested with restriction enzymes Xba1 and HindIII prior to ligation and dialysis. 5 μl of ligation mixture was electroporated into *E. coli* XL1blue competent cells, and the cells were resuspended in 950 μl of LB broth and incubated for 1 h incubation at 37°C shaking (160 r.p.m.) prior to spread-plating on LB agar containing 5 μg ml^{-1} of chloramphenicol and 10 μg ml^{-1} of tetracycline. Colony PCR was performed with primers *apuB*F_Xba and *apuB*R_H3 to select colonies containing the insert, prior to plasmid preparation and sequencing. Plasmid pRD1.3_*apuB* was isolated and transformed into *B. longum* NCIMB8809 by electroporation.

Overexpression constructs pQE30_*repA* and pQE30_*repB* were produced by cloning each coding sequence in frame with the plasmid-encoded N-terminal hexa-histidine using primer pairs pQE30-124N_Sph1 for *repA*, and pQE30-124N_HindIII or pQE30-146N_Sph1 and pQE30-146N _HindIII for *repB*.

To construct plasmids pNZPP_*repA* and pNZPP_*repB*, 200bp preceding the predicted translational start site of each gene was amplified by PCR with primer pairs 124PP_BgLII and 124PP_Pst1, or 146PP_BgLII and 146PP_Pst1 respectively. The resulting amplicons were cloned upstream of the promoterless β-glucuronidase gene in reporter plasmid pNZ272 and transformed into *L. lactis* NZ9000. The recombinant plasmids were then isolated and introduced into *B. breve* UCC2003 and *B. breve* UCC2003 pMP7017_199 by electroporation.

To create a C-terminal deletion in RepB, 669 bp of the *repB* gene excluding 297 bp of the 3′-end of this gene (representing the C-terminal amino acid positions 235 through to 332) was amplified by PCR using primers pQE146C-_SphIF and pQE146C-_H3R. The resulting amplicon was cloned into pQE30 in a manner to fuse a histidine tag to the N-terminus of the encoded, C-terminally truncated RepB protein (designated here as RepBΔC), resulting in the expression construct pQE30_RepBΔC.

The sequence integrity of inserts in all recombinant plasmids was checked by Sanger sequence analysis (performed by MWG, Ebersberg, Germany).

Plasmid stability

The stability of the pRD constructs was assayed by growing *B. breve* UCC2003 strains harbouring each of the individual constructs in non-selective media with successive subcultures at 12 h for approximately 145 generations. At regular time intervals, 100 μl of each culture was serially diluted to 10^{-9} and spot plated on both non-selective and selective agar plates. Colonies were counted, and the CFU/ml obtained on both non-selective and selective plates was determined. Total DNA was isolated from five resistant colonies for each construct and plasmid content assessed by PCR with primers pRDconf_F and pRDconf_R.

Carbohydrate utilization assay

To assess the pullulan-dependent growth potential of strains *B. longum* NCIMB8809 and *B. longum* NCIMB8809 pRD1.3_*apuB,* modified de Man Rogosa and Sharpe (mMRS) medium was freshly prepared from first principles (Man *et al.,* 1960) and supplemented with 1 % (v/v) of pullulan solution and 0.05 % (v/v) of L-cysteine. This medium was inoculated with 1 % (v/v) of stationary phase *B. longum* NCIMB8809 or *B. longum* NCIMB8809 pRD1.3_apuB cells, from cultures cultivated overnight in 1 % (v/v) glucose. The cultures were grown anaerobically at 37°C, and optical density was measured at OD_{600nm} after 24 h using a UV-1280 spectrophotometer (Shimadzu Corporation, Kyoto, Japan). Analyses were carried out in triplicate.

Antibiotic resistance of vectors

The minimal inhibitory concentration (MIC) of chloramphenicol supported by the constructs in *B. breve* UCC2003 as a host was measured by the E-test method according to the manufacturer's instructions (AB *Biodisk*, Solna, Sweden). MIC assays were performed using Reinforced Clostridial Agar (RCA [Oxoid, Basingstoke, Hampshire, United Kingdom]).

Determination of relative plasmid copy number

The relative copy number of pRD1-derived plasmids was assessed by quantitative real-time PCR using the culture and PCR conditions reported previously (Lee *et al.,* 2006). Amplification and detection were performed using a Fast real-time PCR system (Applied Biosystems, Foster City, CA) with QuantiNova SYBR Green PCR Kit (Qiagen, Manchester, UK). Primers 146A_F and 146A_R and, 146B_F and 146B_R (Table S2) were designed based on the pRD1 *repB* sequence, resulting in amplicons of 120 bp and 150 bp respectively. The 1-deoxy-D-xylulose 5-phosphate synthase gene (*dxsA*) (GenBank Accession No. ABE95261) and the housekeeping gene *dnaA* of *B. breve* UCC2003 were used as single-copy chromosomal reference genes, respectively, generating a 120 bp amplicon with primers *dxsA*_F and *dxsA*_R, and a 150 bp amplicon with primers UCC776B_F and UCC776B_R (Table S2). The relative copy number of the derivatives was calculated using the formula $N_{\text{relative}} = 1 + E^{-\Delta CT}$, where E is the amplification efficiency and ΔCT is the difference between the threshold cycle number of the *dxsA*- and *dnaA*-targeting reactions and that of the *repB* target. Experiments were performed in triplicate; mean results are indicated.

Heterologous protein overexpression and purification

RepA, RepB and RepBΔC were overexpressed and purified as N-terminal hexa-histidine fusion proteins using pQE30, an *E. coli* overexpression vector conferring an N-terminal six histidine tail to facilitate downstream purification. *E. coli* EC101 transformed with pQE30-*repA*, pQE30-*repB* or pQE30_RepBΔC was cultured in LB medium supplemented with 100 μg ml^{-1} ampicillin for growth at 37°C until the logarithmic phase (at OD_{600nm} of ~ 0.5–0.6) and induced by IPTG at a final concentration of 0.2 mM for 2 h at 37°C. Cells were harvested at 5000 × *g* for 15 min at 4°C and the pellet resuspended in 10 ml lysis buffer (50 mM Tris-HCl, 50 mM $CaCl_2$, 300 mM NaCl, 10 mM imidazole, pH 8.0). The cells were disrupted with 3 × 1 min rounds of bead-beating using the Mini-Bead-Beater-16 (BioSpec Products, Bartlesville, Oklahoma, USA) with 1 min intervals on ice. Cellular debris was removed by centrifugation at 12 000 × *g* for 30 min at 4°C. Proteins were purified from crude cell extracts via metal affinity chromatography.

Electrophoretic mobility shift assays (EMSAs)

EMSAs were performed essentially as described previously (Hamoen *et al.,* 1998). Binding reactions were carried out in a final volume of 20 μl in the presence of poly[d(I-C)] in binding buffer (20 mM Tris-HCl [pH 7.0], 5 mM $MgCl_2$, 0.5 mM DTT, 1 mM EDTA, 200 mM KCl, 10 % glycerol). Varying amounts of purified protein and the fixed concentration of probe were mixed on ice and subsequently incubated for 30 min at 37°C. Samples were loaded onto a 6 % non-denaturing polyacrylamide (PAA) gel prepared in TAE buffer (40 mM Tris-acetate [pH 8.0], 2 mM EDTA) and run in a 0.5–2.0 × gradient of TAE at 100 V for 90 min in an Atto Mini PAGE system (Atto Bioscience and Biotechnology, Tokyo, Japan). Fluorescent signals were detected using the Odyssey Infrared Imaging System (LI-COR Biosciences UK, Cambridge, UK) and captured using the supplied software (ODYSSEY V3.0). Band intensity of free probe and DNA–protein complex was calculated using the IMAGEJ software (Schindelin *et al.,* 2012; Schneider *et al.,* 2012). Graphical representations and statistical analysis were performed using Prism5 (GraphPad Software, San Diego, CA). Details pertaining to each probe used for EMSAs are outlined in Table S3.

Transcriptional start site determination by primer extension analysis

Bifidobacterium breve JCM7017 was grown in mMRS supplemented with 1% lactose to early exponential growth phase and total RNA isolated using a previously described method (Kuipers *et al.,* 1993). Primer extension analyses were performed by annealing 1 pmol of IRD700 labelled oligonucleotides to 15 μg of RNA as described previously (Ventura *et al.,* 2005). Sequence ladders of the presumed promoter regions immediately upstream of *repA* or *repB* were amplified from genomic DNA with the aforementioned primers using the Thermo Sequenase Primer Cycle Sequencing Kit (Thermo Scientific, Waltham, MA, USA). Separation was achieved on a 6.5% LI-COR Matrix KB Plus acrylamide gel. Signal detection and image capture were performed utilizing a LI-COR sequencing instrument (LI-COR Biosciences, Cambridge, United Kingdom).

GUS assay

Intracellular β-glucuronidase (β-GUS) activity was determined essentially as described previously (Sangrador-Vegas *et al.,* 2007). GusA activity was expressed in Miller Units and calculated using the equation: $1000*((OD_{420} - (1.75*OD_{550}))/(t \times v \times OD_{600}))$, where t is the reaction time [min], v is the cell volume [ml] and OD_{420}, OD_{550} and OD_{600} are absorbance values at 420, 550 and 600 nm respectively.

SEC RALS/LALS/RI analysis

Size exclusion chromatography was carried out on an AKTA Pure HPLC system (GE Healthcare, Cork, Ireland)

using a Superose 6 10/300 G/L column (GE Healthcare, Cork, Ireland) run in a buffer containing 50 mM Tris-HCl, 300 mM NaCl and 50 mM $CaCl_2$ at pH 7.5 with a flow rate of 0.5 ml min^{-1}. Proteins were injected at a final concentration of > 0.8 mg^{-ml}. Detection was performed using OmniSec REVEAL, a dual-angle light-scattering apparatus and refractometer (RALS/LALS/RI) (Malvern Instruments, Malvern, UK). Molecular weight calculations were performed with the OmniSec software (v10.4).

Nucleotide sequence analysis and comparative analysis

Sequence data were obtained from the Artemis-mediated (Rutherford *et al.,* 2000) genome annotations of the pMP7017 sequence (Bottacini *et al.,* 2015). Database searches were conducted using non-redundant sequences accessible at the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov) using BlastP or BlastN. Sequence verification and analysis were performed using the SeqMan and Seqbuilder programs of the DNASTAR software package (DNASTAR, Madison, WI, USA v10.1.2). Amino acid sequences obtained from NCBI© were aligned using ClustalOmega© prior to gap removal, and the finished alignment was generated using Jalview© in order to highlight conserved residues. Pfam (El-Gebali *et al.,* 2019) and NCBIs CDD search tool were used to identify functional domains. Structural analyses were performed using DALI (Holm, 2020) and Phyre2 (Kelley *et al.,* 2015), and structural comparisons and overlays were achieved with The PyMOL Molecular Graphics System, Version 2.3.5 Schrödinger, LLC. The Genomic Nucleotide Skew application (Developed by TU Munich; maintained by Department of Computational Systems Biology of the University of Vienna, Austria) was used to plot the GC-skew across the pMP7017 sequence using a window size of 200 bp and a step size of 100 bp. Nucleotide sequences of *repB* from pMP7017 and *B. longum* homologues were extracted from the NCBI database and aligned with Clustal Omega (Sievers *et al.,* 2011), and phylogeny was inferred using MEGAX (Stecher *et al.,* 2020) following gap removal. The Blast+ terminal (Camacho *et al.,* 2009) was used to determine percentage identity, and the analysis was conducted in R ('R: The R Project for Statistical Computing', https://www.r-project.org) and the comparative figures produced using the genoPlotR package (Guy *et al.,* 2010).

Author contributions

R.D. and D.v.S. designed and directed the project and experiments. R.D. performed all experiments outlined in this work. C.P. aided primer extension analysis. P.K. performed HPLC analysis of Rep proteins. M.J.B.B. assisted in the comparative analysis of the pMP7017 replicon. R.D. and D.v.S. analysed and interpreted the data. R.D drafted the manuscript and figures. R.D., D.v.S. and M.O.C.M. provided commentary and edits to the manuscript and figures in preparation of the final version of the article.

Conflict of interest

None declared.

References

Alvarez-Martín, P., O'Connell-Motherway, M., van Sinderen, D., and Mayo, B. (2007) Functional analysis of the pBC1 replicon from Bifidobacterium catenulatum L48. *Appl Microbiol Biotechnol* **76:** 1395–1402.

Apisiridej, S., Leelaporn, A., Scaramuzzi, C.D., Skurray, R.A., and Firth, N. (1997) Molecular analysis of a mobilizable theta-mode trimethoprim resistance plasmid from coagulase-negative staphylococci. *Plasmid* **38:** 13–24.

Arakawa, K., and Tomita, M. (2012) Measures of compositional strand bias related to replication machinery and its applications. *Curr Genomics* **13:** 4–15.

Arboleya, S., Bottacini, F., O'Connell-Motherway, M., Ryan, C.A., Ross, R.P., van Sinderen, D., and Stanton, C. (2018) Gene-trait matching across the Bifidobacterium longum pan-genome reveals considerable diversity in carbohydrate catabolism among human infant strains. *BMC Genom* **19:** 33.

Baek, J.H., and Chattoraj, D.K. (2014) Chromosome I Controls Chromosome II Replication in *Vibrio cholerae*. *PLoS Genet* **10:** e1004184.

Bottacini, F., Morrissey, R., Roberts, R.J., James, K., van Breen, J., Egan, M., *et al*. (2018) Comparative genome and methylome analysis reveals restriction/modification system diversity in the gut commensal *Bifidobacterium breve*. *Nucleic Acids Res* **46:** 1860–1877.

Bottacini, F., Motherway, M.O., Casey, E., McDonnell, B., Mahony, J., Ventura, M., and van Sinderen, D. (2015) Discovery of a conjugative megaplasmid in *Bifidobacterium breve*. *Appl Environ Microbiol* **81:** 166–176.

Bottacini, F., Zomer, A., Milani, C., Ferrario, C., Lugli, G.A., Egan, M., *et al*. (2017) Global transcriptional landscape and promoter mapping of the gut commensal *Bifidobacterium breve* UCC2003. *BMC Genom* **18:** 991.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Chattoraj, D.K. (2000) Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol Microbiol* **37:** 467–476.

Chattoraj, D.K., Ghirlando, R., Park, K., Dibbens, J.A., and Lewis, M.S. (1996) Dissociation kinetics of RepA dimers: implications for mechanisms of activation of DNA binding by chaperones. *Genes Cells* **1:** 189–199.

Cronin, M., Knobel, M., O'Connell-Motherway, M., Fitzgerald, G.F., and van Sinderen, D. (2007) Molecular

dissection of a bifidobacterial replicon. *Appl Environ Microbiol* **73:** 7858–7866.

Cronin, M., Ventura, M., Fitzgerald, G.F., and van Sinderen, D. (2011) Progress in genomics, metabolism and biotechnology of bifidobacteria. *Int J Food Microbiol* **149:** 4–18.

DasGupta, S., Mukhopadhyay, G., Papp, P.P., Lewis, M.S., and Chattoraj, D.K. (1993) Activation of DNA binding by the monomeric form of the P1 replication initiator RepA by heat shock proteins DnaJ and DnaK. *J Mol Biol* **232:** 23–34.

Dziewit, L., Pyzik, A., Szuplewska, M., Matlakowska, R., Mielnicki, S., Wibberg, D., *et al*. (2015) Diversity and role of plasmids in adaptation of bacteria inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals. *Front Microbiol* **6:** 152.

Egan, M., Bottacini, F., O'Connell Motherway, M., Casey, P.G., Morrissey, R., Melgar, S., *et al*. (2018) Staying alive: growth and survival of Bifidobacterium animalis subsp. animalis under in vitro and in vivo conditions. *Appl Microbiol Biotechnol* **102:** 10645–10663.

Ekundayo, B., and Bleichert, F. (2019) Origins of DNA replication. *PLoS Genet* **15:** e1008320.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., *et al*. (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* **47:** D427–D432.

Fang, F.C., and Helinski, D.R. (1991) Broad-host-range properties of plasmid RK2: importance of overlapping genes encoding the plasmid replication initiation protein TrfA.. *J Bacteriol* **173:** 5861–5868.

Fanning, S., Hall, L.j., Cronin, M., Zomer, A., MacSharry, J., Goulding, D., *et al*. (2012) Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc Natl Acad Sci USA* **109:** 2108–2113.

Filutowicz, M., McEachern, M., Greener, A., Mukhopadhyay, P., Uhlenhopp, E., Durland, R., and Helinski, D. (1985) Role of the π initiation protein and direct nucleotide sequence repeats in the regulation of plasmid R6K replication. In *Plasmids in Bacteria*. Helinski, D.R., Cohen, S.N., Clewell, D.B., Jackson, D.A., and Hollaender, A. (eds). New York: Springer, pp. 125–140.

Fournes, F., Val, M.-E., Skovgaard, O., and Mazel, D. (2018) Replicate once per cell cycle: replication control of secondary chromosomes. *Front Microbiol* **9:** 1833.

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., *et al*. (2011) Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469:** 543–547.

Guy, L., Roat Kultima, J., and Andersson, S.G.E. (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26:** 2334–2335.

Hamoen, L.W., Werkhoven, A.F.V., Bijlsma, J.J.E., Dubnau, D., and Venema, G. (1998) The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev* **12:** 1539–1550.

Heuer, H., and Smalla, K. (2012) Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev* **36:** 1083–1104.

Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci* **29:** 128–140.

Ishiai, M., Wada, C., Kawasaki, Y., and Yura, T. (1994) Replication initiator protein RepE of mini-F plasmid: Functional differentiation between monomers (initiator) and dimers (autogenous repressor). *Proc Natl Acad Sci USA* **91:** 3839–3843.

Johnson, L.N. (2009) The regulation of protein phosphorylation. *Biochem Soc Trans* **37:** 627–641.

Kazuo, Y., and Mitsuyo, Y. (1984) The replication origin of pSC101: the nucleotide sequence and replication functions of the ori region. *Gene* **29:** 211–219.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocols* **10,** 845–858.

Koguchi, H., Ishigami, N., Sakanaka, M., Yoshida, K., Hiratou, S., Shimada, M., *et al*. (2020) Application of recombinase-based in vivo expression technology to Bifidobacterium longum subsp. Longum for identification of genes induced in the gastrointestinal tract of mice. *Microorganisms* **8:** 410.

Komori, H., Matsunaga, F., Higuchi, Y., Ishiai, M., Wada, C., and Miki, K. (1999) Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution. *EMBO J* **18:** 4597–4607.

Kuipers, O.P., Beerthuyzen, M.M., Siezen, R.J., and De Vos, W.M. (1993) Characterization of the nisin gene cluster nisABTCIPR of Lactococcus lactis. Requirement of expression of the nisA and nisI genes for development of immunity. *Eur J Biochem* **216:** 281–291.

Kunnimalaiyaan, S., Krüger, R., Ross, W., Rakowski, S.A., and Filutowicz, M. (2004) Binding modes of the initiator and inhibitor forms of the replication protein pi to the gamma ori iteron of plasmid R6K. *J Biol Chem* **279:** 41058–41066.

Lagares, A., Sanjuán, J., and Pistorio, M. (2014) The plasmid mobilome of the model plant-symbiont *Sinorhizobium meliloti*: coming up with new questions and answers. *Microbiol Spectr* **2.** https://doi.org/10.1128/microbiolspec.PLAS-0005-2013.

Leahy, S.C., Higgins, D.G., Fitzgerald, G.F., and van Sinderen, D. (2005) Getting better with bifidobacteria. *J Appl Microbiol* **98:** 1303–1315.

Lee, C.L., Ow, D.S.W., and Oh, S.K.W. (2006) Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria. *J Microbiol Methods* **65:** 258–267.

Lee, J.-H., and O'Sullivan, D.J. (2010) Genomic insights into Bifidobacteria. *Microbiol Mol Biol Rev* **74:** 378–416.

de Lemos Martins, F., Fournes, F., Mazzuoli, M.-V., Mazel, D., and Val, M.-E. (2018) Vibrio cholerae chromosome 2 copy number is controlled by the methylation-independent binding of its monomeric initiator to the chromosome 1 crtS site. *Nucleic Acids Res* **46:** 10145–10156.

Lievin, V. (2000) Bifidobacterium strains from resident infant human gastrointestinal microflora exert antimicrobial activity. *Gut* **47:** 646–652.

López-Guerrero, M.G., Ormeño-Orrillo, E., Acosta, J.L., Mendoza-Vargas, A., Rogel, M.A., Ramírez, M.A., *et al*. (2012) Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. *Plasmid* **68:** 149–158.

Luo, H., Quan, C.-L., Peng, C., and Gao, F. (2019) Recent development of Ori-Finder system and DoriC database for microbial replication origins. *Brief Bioinform* **20:** 1114–1124.

Man, J.C.D., Rogosa, M., and Sharpe, M.E. (1960) A Medium for the cultivation of Lactobacilli. *J Appl Bacteriol* **23:** 130–135.

Matsushima, N., Yoshida, H., Kumaki, Y., Kamiya, M., Tanaka, T., Izumi, Y., and Kretsinger, R.H. (2008) Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. *Curr Protein Pept Sci* **9:** 591–610.

Mazé, A., O'Connell-Motherway, M., Fitzgerald, G.F., Deutscher, J., and van Sinderen, D. (2007) Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **73:** 545–553.

Motherway, M.O., Fitzgerald, G.F., Neirynck, S., Ryan, S., Steidler, L., and van Sinderen, D. (2008) Characterization of ApuB, an Extracellular Type II amylopullulanase from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **74:** 6271–6279.

Murotsu, T., Matsubara, K., Sugisaki, H., and Takanami, M. (1981) Nine unique repeating sequences in a region essential for replication and incompatibility of the mini-F plasmid. *Gene* **15:** 257–271.

Nakamura, A., Wada, C., and Miki, K. (2007) Structural basis for regulation of bifunctional roles in replication initiator protein. *Proc Natl Acad Sci USA* **104:** 18484–18489.

O'Callaghan, A., and van Sinderen, D. (2016) Bifidobacteria and their role as members of the human gut microbiota. *Front Microbiol* **7:** 925.

O'Connell Motherway, M., Houston, A., O'Callaghan, G., Reunanen, J., O'Brien, F., O'Driscoll, T., *et al*. (2019) A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation. *Mol Microbiol* **111:** 287–301.

O'Connell Motherway, M., O'Driscoll, J., Fitzgerald, G.F., and Van Sinderen, D. (2009) Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium breve UCC2003. *Microb Biotechnol* **2:** 321–332.

Odamaki, T., Bottacini, F., Kato, K., Mitsuyama, E., Yoshida, K., Horigome, A., *et al*. (2018) Genomic diversity and distribution of Bifidobacterium longum subsp. Longum across the human lifespan. *Sci Rep* **8:** 85.

Platteeuw, C., Simons, G., and de Vos, W.M. (1994) Use of the Escherichia coli beta-glucuronidase (gusA) gene as a reporter gene for analyzing promoters in lactic acid bacteria. *Appl Environ Microbiol* **60:** 587–593.

R: The R Project for Statistical Computing. (n.d.). Retrieved 24 November 2020, from https://www.r-project.org/

Roager, H.M., and Licht, T.R. (2018) Microbial tryptophan catabolites in health and disease. *Nat Commun* **9:** 1–10.

Rosenberg, C., Boistard, P., Dénarié, J., and Casse-Delbart, F. (1981) Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol Gen Genet* **184:** 326–333.

Rossi, M., Brigidi, P., Vara, G., Gonzalez Vara y Rodriguez, A., and Matteuzzi, D. (1996) Characterization of the plasmid pMB1 from *Bifidobacterium longum* and its use for shuttle vector construction. *Res Microbiol* **147:** 133–143.

Rossi, M., Brigidi, P., and Matteuzzi, D. (1997) An efficient transformation system for Bifidobacterium spp. *Lett Appl Microbiol* **24:** 33–36.

Ruiz, L., Motherway, M.O., Lanigan, N., and van Sinderen, D. (2013) Transposon mutagenesis in Bifidobacterium breve: construction and characterization of a Tn5 transposon mutant library for *Bifidobacterium breve* UCC2003. *PLoS One* **8:** e64699.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944–945.

Ryan, S.M., Fitzgerald, G.F., and van Sinderen, D. (2006) Screening for and identification of Starch-, Amylopectin-, and Pullulan-Degrading Activities in Bifidobacterial Strains. *Appl Environ Microbiol* **72:** 5289–5296.

Sakanaka, M., Nakakawaji, S., Nakajima, S., Fukiya, S., Abe, A., Saburi, W., *et al*. (2018) A transposon mutagenesis system for *Bifidobacterium longum* subsp. Longum based on an IS3 family insertion sequence, ISBlo11. *Appl Environ Microbiol* **84:** e00824-18.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press, pp. xxxviii–1546.

Sambrook, J., and Russell, D. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Sangrador-Vegas, A., Stanton, C., van Sinderen, D., Fitzgerald, G.F., and Ross, R.P. (2007) Characterization of plasmid pASV479 from Bifidobacterium pseudolongum subsp. Globosum and its use for expression vector construction. *Plasmid* **58:** 140–147.

Schiavi, E., Gleinser, M., Molloy, E., Groeger, D., Frei, R., Ferstl, R., *et al*. (2016) The surface-associated Exopolysaccharide of Bifidobacterium longum 35624 plays an essential role in dampening host proinflammatory responses and repressing local TH17 responses. *Appl Environ Microbiol* **82:** 7185–7196.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., *et al*. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* **9:** 676–682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675.

Schwartz, E. (Ed.). (2009) *Microbial Megaplasmids*. New York: Springer-Verlag.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al*. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7:** 539.

Sobecky, P.A. (1999) Plasmid ecology of marine sediment microbial communities. *Hydrobiologia* **401:** 9–18.

del Solar, G., Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., and Díaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62:** 434–464.

Stecher, G., Tamura, K., and Kumar, S. (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol* **37:** 1237–1239.

Stokke, C., Waldminghaus, T., and Skarstad, K. (2011) Replication patterns and organization of replication forks in *Vibrio cholerae*. *Microbiology* **157:** 695–708.

Swan, M.K., Bastia, D., and Davies, C. (2006) Crystal structure of pi initiator protein-iteron complex of plasmid R6K: Implications for initiation of plasmid DNA replication. *Proc Natl Acad Sci USA* **103:** 18481–18486.

Turroni, F., Serafini, F., Foroni, E., Duranti, S., O'Connell Motherway, M., Taverniti, V., *et al*. (2013) Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions. *Proc Natl Acad Sci USA* **110:** 11151–11156.

Ventura, M., van Sinderen, D., Fitzgerald, G.F., and Zink, R. (2004) Insights into the taxonomy, genetics and physiology of bifidobacteria. *Antonie Van Leeuwenhoek* **86:** 205–223.

Ventura, M., Zink, R., Fitzgerald, G.F., and van Sinderen, D. (2005) Gene structure and transcriptional organization of the dnaK Operon of Bifidobacterium breve UCC 2003 and application of the Operon in Bifidobacterial Tracing. *Appl Environ Microbiol* **71:** 487–500.

Wong, C.B., Tanaka, A., Kuhara, T., and Xiao, J.-Z. (2020) Potential effects of Indole-3-Lactic Acid, a metabolite of human Bifidobacteria, on NGF-induced neurite outgrowth in PC12 Cells. *Microorganisms* **8:** 398.

Yaffe, M.B., and Smerdon, S.J. (2001) PhosphoSerine/Threonine Binding domains: you can't pSERious? *Structure* **9:** R33–R38.

Zhang, G., and Gao, F. (2017) Quantitative analysis of correlation between AT and GC biases among bacterial genomes. *PLoS One* **12:** e0171408.

Zhao, H.-L., Xia, Z.-K., Zhang, F.-Z., Ye, Y.-N., and Guo, F.-B. (2015) Multiple factors drive replicating strand composition bias in bacterial genomes. *Int J Mol Sci* **16:** 23111–23126.

Zzaman, S., Reddy, J.M., and Bastia, D. (2004) The DnaK-DnaJ-GrpE chaperone system activates inert wild type pi initiator protein of R6K into a form active in replication initiation. *J Biol Chem* **279:** 50886–50894.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Promoter analysis of the *rep*A and *rep*B genes and HPLC chromatograms.

Fig. S2. EMSAs representing the lack of affinity of RepB for its promoter and coding regions.

Fig. S3. EMSA data illustrating the affinity of RepB for single-site probes, E and H, and probe G containing 3 putative iterons.

Fig. S4. Heuristic modelling of RepB to replicases of IRPs, HPLC chromatogram of RepB_▵C and EMSA displaying the lack of affinity of RepB_▵C for probe, P17.

Fig. S5. Comparative analysis of the pMP7017 basic replicon across bifidobacterial megaplasmids.

Table S1. Strains and plasmids used in this study.

Table S2. List of primers used in this study.

Table S3. Probe information for EMSAs.