Supplementary Online Material for:

Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication

José A. Ruiz-Masó[a], Rudi Lurz[b], Manuel Espinosa[a] and Gloria del Solar[a]*

[a]Centro de Investigaciones Biológicas, CSIC.
Ramiro de Maeztu, 9.
E-28040-Madrid,
Spain

[b]Max Planck Institute for Molecular Genetics.
Ihnestrasse, 73.
D-14195 Berlin,
Germany

To whom correspondence should be addressed. Tel: +34 918373112;
Fax: +34 915360432; Email: gdelsolar@cib.csic.es

SUPPLEMENTARY METHODS

To assess potential cooperative effects in the successive events of binding of RepB to DNA fragments containing the *nic* locus, the *bind* locus, or the entire *dso*, either a two-site or a three-site system model was employed, depending on whether two or three different protein-DNA complexes were observed. The following equations were used (1):

For a two-site system, cooperativity can be inferred whenever $K_2 > K_1^2 / 4$ using these three equations (where Z is the binding polynomial equal to $1 + K_1 \cdot L + K_2 \cdot L^2$):

$$\theta_0 = 1/Z$$
$$\theta_1 = K_1 \cdot L/Z$$
$$\theta_2 = K_2 \cdot L^2/Z$$

For a three-site system, the second and third ligand binding events are cooperative if $K_2 > K_1^2 / 3$, and if $K_3 > K_2 \cdot K_1 / 3$, respectively, by using these four equations (where Z is equal to $1 + K_1 \cdot L + K_2 \cdot L^2 + K_3 \cdot L^3$):

$$\theta_0 = 1/Z$$
$$\theta_1 = K_1 \cdot L/Z$$
$$\theta_2 = K_2 \cdot L^2/Z$$
$$\theta_3 = K_3 \cdot L^3/Z$$

where K_1, K_2 and K_3 are macroscopic equilibrium constants; L is the concentration of free protein; θ_0, θ_1, θ_2 and θ_3 are the fractions of free DNA and of the complexes C1, C2 and C3, respectively. Nonlinear least squares methods of parameter estimation were used.

SUPPLEMENTARY RESULTS

Since the number of RCR plasmids of the pMV158 family has increased up to the twenty four replicons reported so far, it was significant to align their *dso* regions and to know whether all of them shared the organization present in pMV158. Supplementary Figure S2 shows that these replicons exhibit a similar organization: a highly conserved nick sequence and, downstream of it, two DR clusters, the PDR and the DDR. The exceptions were the two plasmids from *Helicobacter pylori* (pHPK255 and pHP489), and plasmids pCI411, pLA106 and pCL2.1 from lactic acid bacteria, for which we have not found DDR (Supplementary Figure S2). These results were unexpected, since the existence of these two types of DR remained undisclosed in a previous inspection of the sequence of eleven of these replicons (2). In fact, some of the DR previously predicted to constitute the *bind* locus have now been reassigned as PDR. The newly defined PDR consist of two or three 7-bp or 8-bp repeats that are located a short distance 3' from the nick site and share some homology among the plasmids of the family. In contrast, the DDR are not conserved, except in very closely related plasmids. While the PDR are 10-18 nucleotides away from the nick site, the DDR are 45-91 nucleotides from it. In most plasmids, the DDR consist of either two or three 11-bp perfect or imperfect

repeats arranged in tandem, although in some of the plasmids containing three repeats, the most distal one has a 1-bp deletion or insertion (Supplementary Figure S2). In a few plasmids, DDR constituted of repeats of a length other than 11 bp, but still spanning about either 1 or 2 turns of the DNA double helix, were found (two 10-bp DR in the *Mycoplasma mycoides* plasmid pKMK1, two 12-bp DR in the *Lactobacillus plantarum* plasmid pA1, and three 21-bp DR in the *M. mycoides* plasmid pADB201). In the staphylococcal plasmid pE194, we found 20 bp-phased DDR consisting of two non-tandem 8-bp DR separated by a 12-bp spacer (Supplementary Figure S2). Similarly, in the lactococcal plasmid pBM02, two non-tandem imperfect 8-bp repeats also separated by a 12-bp spacer were observed. Thus, the DDR of the members of the pMV158 family, when present, are nearly in phase with the DNA helical repeat. As the DDR are proposed to be the primary binding sites of the cognate Rep proteins, their in-phase relative arrangement suggests that oligomeric forms of the initiators interact with the repeats of the *bind* locus on the same face of the DNA double helix. It is also worth noting that among plasmids of the pPSC22/ pFX2/ pWV01 sub-family, in which identical PDR and DDR are present, the spacer between the nick site and either the PDR or the DDR varies, though the distance between both sets of DR is maintained. Although further experimentation is required to draw any conclusion, it might be envisaged from this that phasing between PDR and DDR may be required for plasmid replication *in vivo* and, in fact, we have been unable to construct pMV158-derivative plasmids in which the *nic* locus and the DDR were placed out of phase (our unpublished observations). Interestingly, in pMV158 the sequence of the repeat unit of the PDR, although shorter than that of the DDR, is reminiscent to it (compare the sequences 5'-**GT-GCCGA**-3' in the PDR top strand, and 5'-AAA**GT**C**GCCGA**-3' in the DDR bottom strand, with identical bases in boldface letters). However, as a repeated pattern of either HO• or DMS footprints is not observed in the PDR (Figure 2B), it is clear that RepB, at least in its hexameric form, does not interact in the same way with the two direct repeats constituting this DNA region.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Stability of the RepB-*bind* complex. (**A**) Dissociation of C1 was initiated by addition of a 300-fold molar excess of the specific unlabelled DNA fragment. Samples were analyzed at the indicated times (lanes 3-5). Free DNA was loaded at lane 0. As controls, binding reactions without competitor DNA (lane 1), or with the labelled and unlabelled DNAs mixed prior to the addition of RepB (lane 2) were assayed. (**B**) Time course of the dissociation of the RepB-*bind* complex. The solid line is least-square fit of representative data to equation 1, whereas the dots depict the experimental values.

Figure S2. Conservation of the *dso* in plasmids of the pMV158 family. Members of the family are grouped based on the *dso* sequence identity distances, represented as a phylogenetic tree. Homology is found at the *nic* locus, whose nick sequence is conserved in all members. Divergences exist in the DDR, which does not even seem to exist in some cases. The putative DDR in the

Lactobacillus curvatus plasmid pLC2 correspond roughly to the repeats that have recently been proposed as the binding site of the pLC2 Rep protein (3), and differ from those previously proposed (2). The proposed DDR have only been proved to constitute the specific binding sites for their cognate Rep proteins in pMV158 (4) and in the *Enterococcus faecium* plasmid pJB01 (3). Features in pMV158 are highlighted.

SUPPLEMENTARY MATERIAL REFERENCES

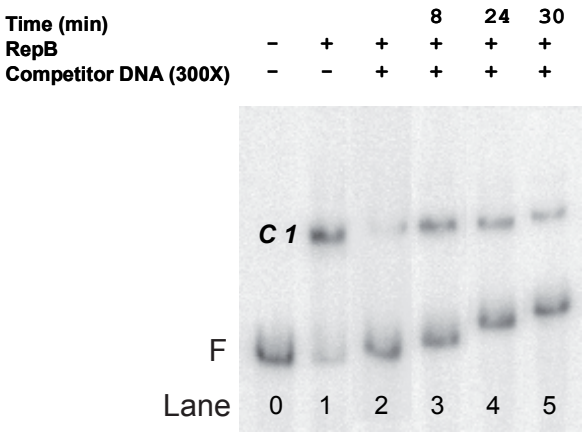
1. Senear, D.F. and Brenowitz, M. (1991) Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J. Biol. Chem.*, **266**, 13661-13671.
2. Moscoso, M., del Solar, G. and Espinosa, M. (1995) In vitro recognition of the replication origin of pLS1 and of plasmids of the pLS1 family by the RepB initiator protein. *J. Bacteriol.*, **177**, 7041-7049.
3. Kim, S.W., Jeong, E.J., Kang, H.S., Tak, J.I., Bang, W.Y., Heo, J.B., Jeong, J.Y., Yoon, G.M., Kang, H.Y. and Bahk, J.D. (2006) Role of RepB in the replication of plasmid pJB01 isolated from Enterococcus faecium JC1. *Plasmid*, **55**, 99-113.
4. de la Campa, A.G., del Solar, G. and Espinosa, M. (1990) Initiation of replication of plsmid pLS1. The initiator protein RepB acts on two distant regions. *J. Mol. Biol.*, **213**, 247-262.

SUPPLEMENTARY TABLE

Table S1**. DNA contour length statistics**

	1825 bp	***1825 bp + RepB***	***700 bp***	***700 bp+RepB***
Mean (nm)	596.63	578.23	234.45	220.06
Std. dev.	9.84	12.27	2.97	6.00
Std. err.	1.16	1.26	0.32	0.59

A


Time (min)
RepB
Competitor DNA (300X)
8 24 30
- + + + + +
- - + + + +
C 1
F
Lane 0 1 2 3 4 5

B

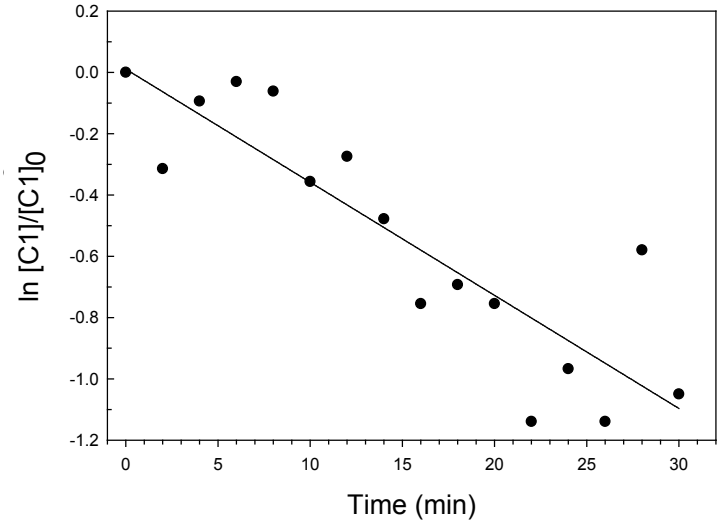
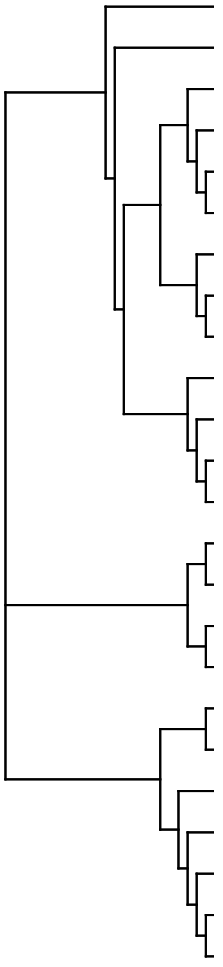

ln [C1]/[C1]0
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
0
5
10
15
20
25
30
Time (min)

Figure S1



	Nick sequence		*PROXIMAL DIRECT REPEATS*		*DISTAL DIRECT REPEATS*
pHPK255	TACTACGAC	**14**	TGTGTAA GCTACAA GCTACGA		
pHP489	TACTACGAC	**15**	GAGACCG GAGACCG GAGACAA		
pSSU1	TACTACGAC	**13**	GTGCCGA GTGCCAA	**85**	TCGGCGACATT TTGGCGACATT *T* TCGGCGACAAA
pSMQ172	TACTACGAC	**13**	GTGCCGA GTGCCAA	**85**	TCGGCGACATT TTGGCGAGTTT -TGGCGAGTTT
pMV158	TACTACGAC	**12**	GTGCCGA GTGCCAA	**84**	TCGGCGACTTT TCGGCGACTTT TCGGCGACTTT
pSB02	TACTACGAC	**12**	GTGCCGA GTGCCAA	**84**	TTGGCGACTTT TCGGCGACTTT TCGGCTACTTT
pPSC22	TACTACGAC	**18**	GTGCCGA GTGCCAA *TTTTT* GTGCCAA	**90**	CGCCAACGAAT CGCCAACGTTT *T* CGCCAACGTTT
pFX2	TACTACGAC	**12**	GTGCCGA GTGCCAA *TTTTT* GTGCCAA	**84**	CGCCAACGAAT CGCCAACGTTT *T* CGCCAACGTTT
pWV01	TACTACGAC	**12**	GTGCCGA GTGCCAA *TTTTT* GTGCCAA	**84**	CGCCAACGAAT CGCCAACGTTT *T* CGCCAACGTTT
pLH2	TACTACGAC	**13**	GTGCCGA GTGCCAA	**85**	TCAACAAATCG CCAACAAAATG ACAACATTTTT
pLC2	TACTACGAC	**13**	GTGCCGA GTGCCAA	**76**	TCAACAAAAAG TCAACAACCCG CCAACAAAATT
pJB01	TACTACGAC	**13**	GTGCCGA GTGCCAA	**76**	TCAACAAAAAG TCAACAAACCG CCAACAAATTT
pPF107-3	TACTACGAC	**13**	GTGCCGA GTGCCAA	**76**	TCAACAAAAAG TCAACAAACCG CCAACAAATTT
pLA106	TACTACCAC	**12**	TCGGTCA TTGGTCA		
pKMK1	TACTACCGA	**12**	GTGATGT GTGACAA	**91**	GTTTTGTTTT GTTTTATTTT
pCl411	TACTACAAC	**10**	TGTGGTCA TTTGGTCA TTTGGTCA		
pCL2.1	TACTACGAC	**10**	TGGTAATT TGGTAATT TGGTCAAA		
pA1	CACTACGAC	**14**	TGCAATG TGCAATG TGAAAAA	**45**	TGTTATATCAAT GTTTATAGCTAT
pADB201	TACTACGAT	**13**	AGTGATT TGTGACA TGTGACG	**55**	TGTTTTGCTAGCATTTGTTAA X3
pE194	TACTACGAC	**14**	TGTCCAT TGTCCAT TGTCCAA	**61**	ATGTTAAA *AGGTTGTTTTTT* ATGTTAAAG
pBM02	CACTACGAC	**12**	TGTCCAC TGTCCGT TATCCAT	**65**	TGACATTT *TTGGTGTGACAG* TGATTTTT
pLB4	TACTACGAC	**13**	TGTCCAT TGTCCAT TAAACAG	**64**	GGGTTTTCGGG GGGGTTTGAGA
pLF1311	CACTACGAC	**12**	TGTCCAT TGTCCAT TGAACAG	**64**	GGGTTTTCGGG AGGTTTTGAGA
pLF14	CACTACGAC	**12**	TGTCCAT TGTCCAT TGAACAG	**64**	GGGTTTTCGGG AGGTTTTGAGA

Figure S2