The EMBO Journal Vol.17 No.15 pp.4511–4526, 1998

Protein domains and conformational changes in the activation of RepA, a DNA replication initiator

Rafael Giraldo[1,2,3], José M.Andreu[2] and Ramón Díaz-Orejas[1]

[1]Departamento de Microbiología Molecular and [2]Departamento de Estructura y Función de Proteínas, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), c/Velázquez 144, 28006 Madrid, Spain

[3]Corresponding author
e-mail: cibgs2q@pinar1.csic.es

RepA is the DNA replication initiator protein of the *Pseudomonas* plasmid pPS10. RepA has a dual function: as a dimer, it binds to an inversely-repeated sequence acting as a repressor of its own synthesis; as a monomer, RepA binds to four directly-repeated sequences to constitute a specialized nucleoprotein complex responsible for the initiation of DNA replication. We have previously shown that a Leucine Zipper-like motif (LZ) at the N-terminus of RepA is responsible for protein dimerization. In this paper we characterize the existence in RepA of two protein globular domains C-terminal to the LZ. We propose that dissociation of RepA dimers into monomers results in a conformational change from a compact arrangement of both domains, competent for binding to the operator, to an extended species that is suited for iteron binding. This model establishes the structural basis for the activation of DNA replication initiators in plasmids from Gram-negative bacteria.

Keywords: conformational changes/DNA replication/initiator activation/protein domains/RepA

Introduction

DNA replication is the key process for the transmission of genetic information in all organisms. According to the replicon model (Jacob *et al*., 1963), initiation of DNA replication requires a *cis*-acting DNA sequence (the replicator) and a *trans*-acting protein factor (the initiator) that binds to the replicator to form a specialized nucleoprotein complex. Initiator binding melts the DNA strands to form an open complex. Finally, the initiator must be able to guide the entrance into the open complex of other protein factors required for DNA replication (reviewed in Kornberg and Baker, 1992; Helinski *et al*., 1996; del Solar *et al*., 1998).

Both in prokaryotic and eukaryotic replicons, an activation step is required in order to enable the initiator protein to bind and melt the replicator DNA sequence (Kelman and O'Donnell, 1994). Therefore, the characterization at the molecular level of the determinants for the activation of initiators is a topic of great interest. Chromosomal initiators, such as the DnaA protein in *Escherichia coli* (reviewed in Messer and Weigel, 1996) and the Origin Recognition Complex (ORC) in eukaryotes (Bell and Stillman, 1992), require ATP binding and hydrolysis for the activation step. DnaA can also be activated by DnaK chaperone (Hwang *et al*., 1990). ORC function is regulated by other protein factors, including cyclin-dependent kinases (CDKs) (Newlon, 1997).

Plasmids are suitable models for dissecting the enzymology of DNA replication in Gram-negative bacteria. In most plasmid replicons, initiation of DNA replication requires a plasmid-encoded initiator protein, called Rep, of ~25–40 kDa. Rep-type initiators share clear sequence similarities (25–35% in pairwise alignments, pointing to a common structure and function) which we have used for building phylogenetic trees for plasmids in Gram-negative bacteria (del Solar *et al*., 1998).

With exceptions such as plasmid R1 (Giraldo and Díaz, 1992) and phage λ (Dodson *et al*., 1985), where the initiators bind to inversely-repeated sequences at the replicators, Rep-type proteins bind to tandem arrangements of directly-repeated sequences (iterons) to establish the initiation nucleoprotein complex. Iteron sequence length is typical of every replicon, but in all cases it extends to approximately two turns of a B-DNA double helix. The precise number and spacing between iteron repeats is also characteristic for each plasmid (reviewed in Helinski *et al*., 1996; del Solar *et al*., 1998). The Rep proteins of some replicons, such as *Pseudomonas* pPS10 (Nieto *et al*., 1992; García de Viedma *et al*., 1995a, 1996) and Enterobacterial pSC101 (Vocke and Bastia, 1983), F (Rokeach *et al*., 1985) and R6K (Filutowicz *et al*., 1985), have an important second function: they recognize inversely-repeated sequences (operators) which overlap the promoter of their own coding genes, acting as self-repressors. It has been shown for pPS10 (García de Viedma *et al*., 1995a, 1996), pSC101 (Manen *et al*., 1992) and F (Ishiai *et al*., 1994) replicons that dimers of their Rep proteins bind to the operator sequences, whereas monomers bind to the iterons.

Molecular chaperones, either the triad DnaK/DnaJ/GrpE or ClpA, have been implicated in the activation pathway of the Rep-type initiator proteins of plasmids P1 (Wickner *et al*., 1992, 1994; DasGupta *et al*., 1993; Sozhamannan and Chattoraj, 1993) and F (Kawasaki *et al*., 1990). Two alternative pathways for the action of chaperones in P1 replication have been proposed: Rep dimerization would be so tight that chaperones are required to dissociate dimers into monomers (Wickner *et al*., 1992, 1994); or chaperones would somehow 'activate' Rep to bind to the iterons (DasGupta *et al*., 1993; Sozhamannan and Chattoraj, 1993). A consensus proposal points out that chaperones would simultaneously induce both monomerization and an undetermined conformational change in Rep (Dibbens *et al*., 1997; Pak and Wickner, 1997). A role has also been proposed for the ClpX chaperone in the activation of the initiator of RK2 plasmid (Konieczny and

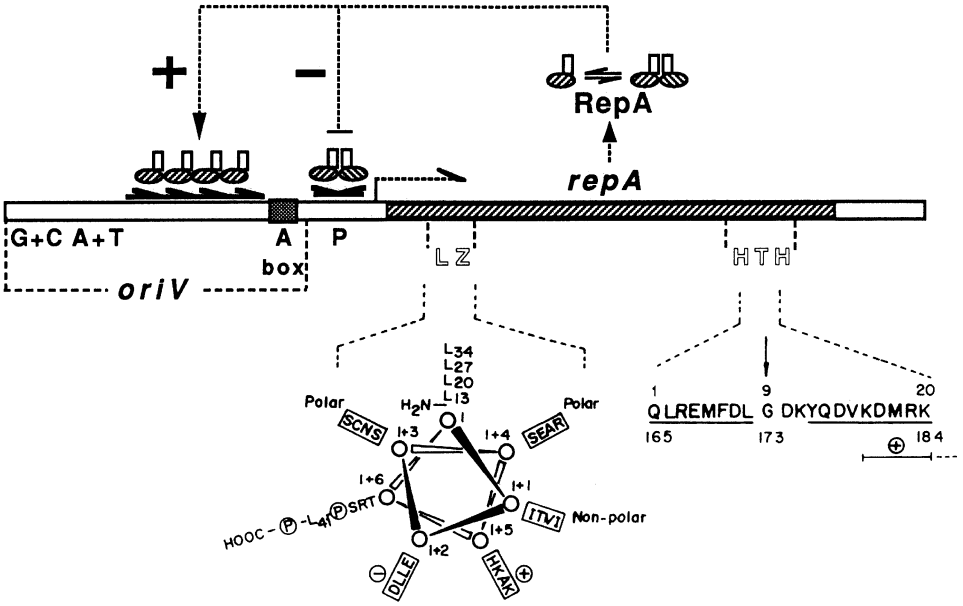


Fig. 1. Scheme of the basic replicon of the *Pseudomonas* plasmid pPS10. It consists of the gene (*repA*, striped box) for a *trans*-acting factor (RepA protein, striped ovals) and a *cis*-acting sequence (the replicator, *oriV*). RepA has an N-terminal LZ motif (white rectangles over the protein ovals, shown below as a helical-wheel projection), responsible for a monomer to dimer association equilibrium, and a C-terminal HTH motif (whose sequence is written below) involved in DNA binding. The sequence flanking the –35 box of the *repA* gene promoter (P) is the target for binding of RepA dimers, that recognize both arms of an 8 bp inverse repeat (operator, symmetric arrows), self-repressing (–) *repA* transcription. RepA monomers bind to four identical and contiguous 22 bp direct repeats (iterons, head-to-tail arrows) to initiate (+) DNA replication. There is a functional 9 bp DnaA box (shaded), the binding site for the chromosomal initiator DnaA, in between the most rightwards iteron and the operator. The replicator also includes A+T- and G+C-rich regions, common to other replicons.

Helinski, 1997). The ClpA chaperone is also a key factor in the way to proteolytic degradation of the P1 initiator by the ClpP protease (reviewed in Gottesman *et al*., 1997).

We have been studying RepA, the initiator protein of pPS10, a plasmid isolated from *Pseudomonas syringae* (Figure 1) (Nieto *et al*., 1992). RepA is one of the smallest (26.7 kDa) Rep-type proteins and so is an attractive model to dissect the structure and function of a DNA replication initiator. In a search for the determinants of the dual function of RepA, and so of the conversion of repressor protein dimers into initiator monomers, we had previously found that RepA dimerization is mediated by a Leucine Zipper-like (LZ) motif (reviewed in Lupas, 1996) at the N-terminus of the protein (García de Viedma *et al*., 1996). This was the first LZ motif found in a protein of prokaryotic origin (Giraldo *et al*., 1989). In addition, a helix–turn–helix (HTH) motif (reviewed in Brennan and Matthews, 1989) was identified at the protein C-terminus (García de Viedma *et al*., 1995b) and found to be responsible for RepA binding to both the operator and iteron DNA sequences.

By means of a combined *in vitro* approach, including biochemical and spectroscopic analyses and protein–DNA interaction functional studies, we describe in this paper the existence in RepA of two protein globular domains C-terminal to the LZ. Furthermore, we propose that dissociation of RepA dimers into monomers would result in a structural change from a compact arrangement of the protein domains, competent for binding to the operator, to an extended form that is suited for iteron binding. The emergent model establishes the structural basis for understanding the activation of Rep-type DNA replication initiators in plasmids from Gram-negative bacteria. We argue that conformational changes might be universal steps in the activation pathway for any initiator protein.

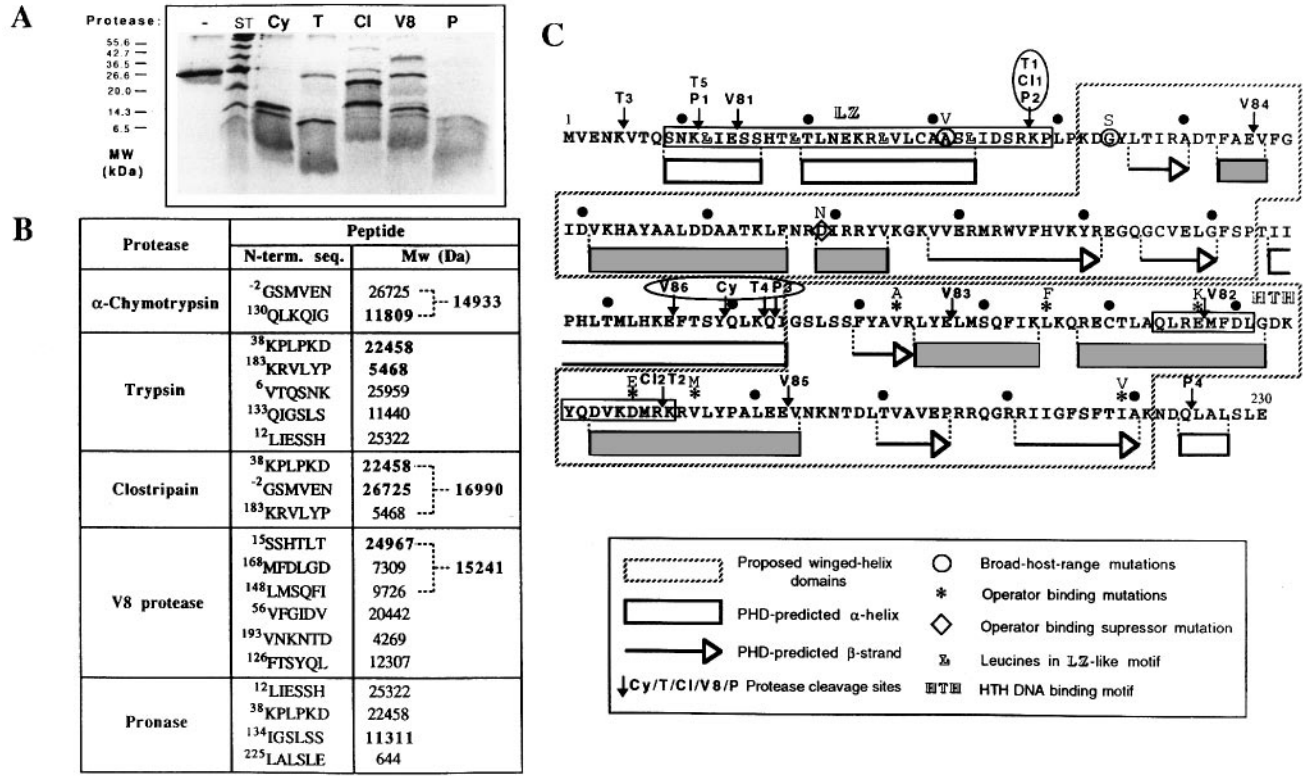
Results

Limited proteolysis and the proposal for protein domains in RepA

Limited proteolysis is a valuable tool for probing the accessibility of protein loops, which either protrude from the surface of a protein or connect several independently folded domains (reviewed in Hubbard, 1998). In order to analyse the structure of the RepA initiator from the *Pseudomonas* replicon pPS10, limited proteolysis of purified RepA was performed with several proteases (α-chymotrypsin, trypsin, clostripain, V8 protease and pronase). Digestion mixtures (Figure 2A) were N-terminal sequenced (Figure 2B) and the peptides were located in the RepA protein sequence (Figure 2C).

α-chymotrypsin breaks RepA into two halves of 129 and 101 amino acid residues, whereas the other proteases cut RepA into more pieces. Two cutting hotspots in the initiator, one C-terminal to the LZ motif (residue 37) and the other mapping around the α-chymotrypsin cleavage site (residues 125–133), establish three possible portions in RepA. A secondary structure prediction for RepA performed with the algorithm PHD (Rost and Sander, 1993) gives us a clue to the possible domain composition of the initiator: from the N-terminus, the mostly α-helical LZ motif (Giraldo *et al*., 1989; reviewed in Lupas, 1996) would be linked to a first potential domain, whose topology (β/α/α/α/β/β) would be repeated in a second putative C-terminal domain (Figure 2C).

However, the mere identification of proteolytic frag-



Protease	Peptide		
	N-term. seq.	Mw (Da)	
α-Chymotrypsin	$^{-2}$GSMVEN	26725	14933
	^{130}QLKQIG	**11809**	
Trypsin	^{38}KPLPKD	**22458**	
	^{183}KRVLYP	**5468**	
	^{6}VTQSNK	25959	
	^{133}QIGSLS	11440	
	^{12}LIESSH	25322	
Clostripain	^{38}KPLPKD	**22458**	16990
	$^{-2}$GSMVEN	**26725**	
	^{183}KRVLYP	5468	
V8 protease	^{15}SSHTLT	**24967**	15241
	^{168}MFDLGD	7309	
	^{148}LMSQFI	9726	
	^{56}VFGIDV	20442	
	^{193}VNKNTD	4269	
	^{126}FTSYQL	12307	
Pronase	^{12}LIESSH	25322	
	^{38}KPLPKD	22458	
	^{134}IGSLSS	**11311**	
	^{225}LALSLE	644	

Fig. 2. Limited proteolysis and the search for potential protein domains in RepA. (**A**) SDS–PAGE (20% polyacrylamide) corresponding to 1/3 volume of the digestion of 30 μg of purified RepA (–) with different proteases, either α-Chymotrypsin (Cy), Trypsin (T), Clostripain (Cl), Staphylococcal-V8 (V8) or Pronase (P), for 60 min. Protein molecular weight standards are shown (ST), together with their molecular masses (in kDa). (**B**) A list of the peptides that were identified after N-terminal sequencing of the same proteolysis reactions (2/3 volume) shown in (A). The N-terminal coordinate is indicated for each peptide [–2 corresponds to the GS sequence from the His_6-tag, previous to the first RepA residue, that remains after thrombin-digestion; see (C)]. The expected molecular masses for the fragments are also shown [those corresponding to peptides visualized in (A) are in bold]. (**C**) Mapping the proteolytic cleavage sites (B) in the sequence of RepA protein. Dots are located every 10 amino acid residues. Lower-case numbers after the protease codes [as in (A)] rank the frequency of cleavage. Ovals encircle the two proteolytic hot-spots that define the existence of domains (dashed boxes) in RepA. Below the sequence of the protein, a secondary structure prediction (PHD) is shown. The box at the bottom further explains the symbols used in this figure.

ments does not necessarily imply that these constitute independently folded domains. Thus, the application of complementary methods that provide information on protein structure is mandatory, and they require relatively large amounts of purified proteins.

Expression and purification of RepA fragments for structural and functional analyses

We have cloned four *repA* gene fragments, comprising combinations of the N-terminal LZ and the two globular domains proposed after proteolytic digestion. To summarize (Figure 3A), the constructs include: (i) an N-terminal deletion of the LZ, leaving intact both putative globular domains (ΔN37); (ii) a simultaneous deletion of both the LZ and the first domain, leaving the C-terminal domain with the HTH motif (ΔN132); (iii) a double deletion of the N-terminal LZ and the C-terminal domain, resulting in the isolated first domain (ΔN37C133); (iv) a C-terminal deletion, leaving the LZ and the contiguous first domain (ΔC133). Finally, a construct similar to ΔN37 but missing a few residues further into the N-terminus of the first putative domain (ΔN42) was included in the analysis. This fragment turned out to be a control for protein misfolding (see below).

RepA protein fragments, N-terminal-tagged with His_6 peptides, were overexpressed in *E.coli*, purified by means of nickel affinity chromatography and the fusion peptides were removed by thrombin cleavage before a final purification step (see Materials and methods; Figure 3B).

The chemical behaviour of RepA and its fragments gives us a clue to their stability. RepA–WT and the ΔC133 and ΔN132 fragments are expressed as mostly soluble to high yields. Once purified, RepA–WT and ΔN132 require a high ionic strength and a slightly acidic pH to keep them concentrated in solution, whereas ΔC133 is highly soluble even at low ionic strength. However, all RepA fragments having the LZ deleted (ΔN37, ΔN42 and ΔN37C133) are expressed as inclusion bodies and require solubilization with guanidinium hydrochloride (GuHCl) and refolding by dialysis (see Materials and methods). The solubility of the ΔN42 fragment is strictly dependent on high ionic strength, acidic pH and the maintenance of the His_6-tag so, in all the experiments presented in this paper, it was used as a fusion protein (His_6-ΔN42). ΔN37 and ΔN37C133 fragments, albeit purified from inclusion bodies, are correctly refolded since they acquire a stable three-dimensional structure and show sequence-specific DNA-binding activity (see below). Other RepA fragments that do not match with those defined by proteolysis were also cloned: ΔN92; ΔC58; ΔC103; and ΔC113. All four are expressed as inclusion bodies, but we have been unable to refold them as soluble proteins (data not shown), which

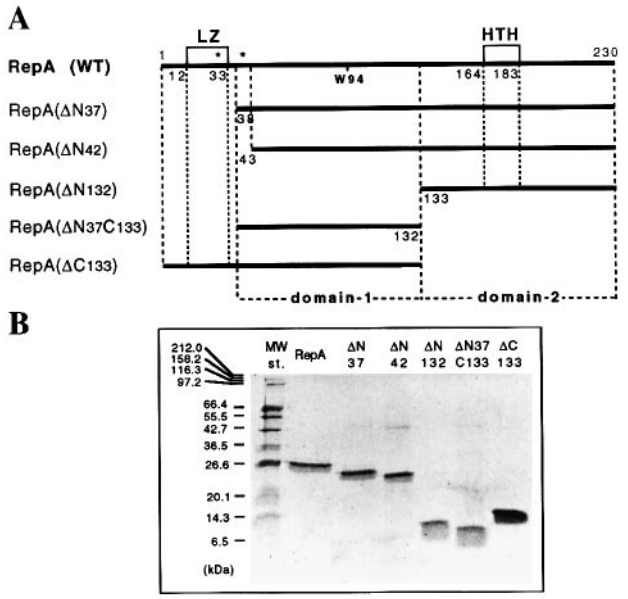


Fig. 3. Cloning and purification of RepA and its fragments. (**A**) Linear sketches of RepA and the fragments cloned, expressed and purified for the studies described in this paper. Dashed vertical lines mark the boundaries for the LZ and HTH motifs, as well as those for the putative domains defined by proteolysis. The location of point mutations that broaden the host range for pPS10 plasmid (*) (Giraldo *et al*., 1992; Fernández-Tresguerres *et al*., 1995) are indicated, as well as the unique tryptophane residue (W94) used as a probe in the fluorescence spectroscopy studies (Figure 5). (**B**) SDS–PAGE (15% polyacrylamide) of purified stocks of RepA–WT and its fragments. Molecular weight standards (MW st.) are the same as shown in Figure 2A. Masses for the RepA proteins are: 26.7 kDa (WT), 22.6 kDa (ΔN37), 22.0 kDa (ΔN42), 11.6 kDa (ΔN132), 11.1 kDa (ΔN37C133) and 15.3 kDa (ΔC133).

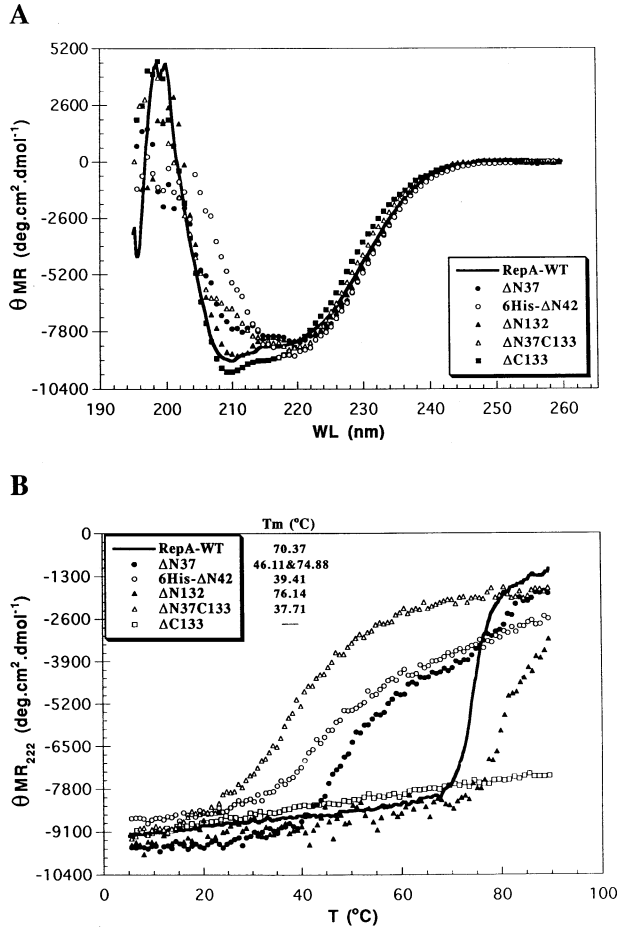


Fig. 4. CD spectroscopy on RepA–WT and its fragments. (**A**) CD spectra (260–195 nm range) were acquired at 5°C, for stocks of each protein (6–25 μM). Spectra are plotted to the same scale in mean molar ellipticity per residue (θMR) units (deg.cm^2.dmol^{-1}). The inset to the figure shows the symbols used to plot each protein spectrum. Structural changes associated with RepA monomerization are evident (see text). (**B**) Thermal denaturation profiles. The same samples studied in (A) were then heated from 5 to 90°C, measuring the evolution of the ellipticity at 222 nm (α-helix signature). Data were plotted as θMR_{222} (deg.cm^2.dmol^{-1}) versus temperature (°C). The inset shows the symbols used to mark the curve for each protein sample, together with its calculated melting temperature (T_m, 50% unfolding). These plots are compatible with the existence in RepA of two stable folding units (two globular domains).

suggests that they do not comprise independent and stable protein domains.

Using fusions between maltose binding protein and RepA, we had previously demonstrated that the LZ is involved in RepA dimerization (García de Viedma *et al*., 1996). Glutadialdehyde cross-linking (data not shown) and gel-filtration analyses (see below) of RepA and its fragments further support this conclusion, since only RepA–WT and ΔC133 can form dimers, although the appearance of heterogeneous aggregates at high concentrations of ΔN37 and ΔN37C133 suggests that the first putative domain could also be involved in additional protein–protein interactions.

The aim of this study is to gain understanding of the structural basis for the different functions of RepA dimers and monomers. To obtain homogeneous preparations of monomeric RepA is difficult: at the micromolar RepA–WT concentrations required for structural analysis, a significant proportion of protein dimers is present (see below), and our attempts to get soluble preparations of unfused LZ mutants showing reduced RepA dimerization (García de Viedma *et al*., 1996) have so far failed. Therefore, the ΔN37 fragment, having the dimerization determinant (the LZ motif) deleted, has been used to mimic RepA–WT monomers.

Spectroscopic analysis of cloned RepA fragments shows that they correspond to folded protein domains

Circular Dichroism (CD) spectroscopy. Far-UV-CD spectroscopy allows us to estimate the overall secondary structure of purified proteins. Spectra for mostly α-helical structures show double minima ellipticity (θ) values at 222 and 208 nm, and a maximum at 195 nm, whereas a high proportion of β-sheets leads to spectra with a broader minimum around 215 nm and a maximum at 200 nm (reviewed in Kelly and Price, 1997; Schmid, 1997). The variation of θ with temperature provides information on thermodynamic parameters related to the stability of protein domains (reviewed in Kelly and Price, 1997; Pace and Schultz, 1997).

CD spectra suggest that RepA monomerization results in a protein conformational change. We have performed CD spectroscopic measurements on RepA–WT and its purified fragments. The overall shape of the spectra (Figure 4A) is dominated by the α-helical component in RepA–WT and those fragments that comprise the LZ plus the first proposed domain (ΔC133) or just the second putative

domain (ΔN132). The LZ, a major α-helical element (a peptide including this RepA motif is ~40% α-helical; data not shown), is absent in ΔN37 and so its spectrum has shallower minima at 208 and 222 nm. The CD spectra for the two fragments in which the first proposed domain is present and the LZ is absent (ΔN37C133 and His_6-ΔN42) shift to shapes that correspond to a predominant β-sheet structure.

CD spectra were also acquired for equimolar mixtures of either ΔC133 and ΔN132 ('reconstituting' RepA–WT) or ΔN37C133 and ΔN132 ('mimicking' ΔN37). The experimentally obtained spectra are, in both cases, nearly identical to the algebraic addition of the spectrum for the individual components (data not shown). This additive character suggests that both proposed domains are independent structural units.

To get a deeper insight into the interpretation of the spectral changes mentioned above, we have analyzed them with the algorithm K2D, which de-convolutes a CD spectrum in its individual α-helical, β-sheet and random coil components (Andrade *et al*., 1993). K2D analysis confirms the conclusions already extracted from the visual inspection of the spectra: RepA monomerization would carry out a protein conformational change affecting the first putative domain that would slightly increase its β-sheet contents at the expense of the α-helical component, whereas the structure of the second proposed domain (including the HTH motif) would remain essentially unaltered (data not shown).

Thermal stability analysis indicates that each cloned fragment constitutes a folding unit. CD thermal denaturation profiles were acquired for RepA–WT and its fragments measuring the variation of ellipticity at 222 nm as a function of increasing temperature (Figure 4B). In the denaturation curve for a protein constituted by a single globular domain, each step or plateau may correspond to a defined folding state; the initial step of the curve at low temperatures corresponds to the folded (native) conformation, whereas the second plateau at high temperatures reflects the fully unfolded (denatured) state (Pace and Scholtz, 1997). The slope of the sigmoidal transition between states correlates with the cooperativity of the process (Kelly and Price, 1997).

For RepA–WT, a sharp transition between native and denatured states is observed, with a T_m value (temperature at which 50% of denaturation is achieved) of 70.4°C. Thus, RepA dimers behave as a single entity and the observed large slope indicates a very cooperative unfolding process. ΔN37 protein shows a transition through three states, possibly corresponding to the native and two distinct unfolded species. This is compatible with the presence of two independent protein domains, with T_m values of 46.1 and 74.9°C, in the monomeric RepA fragment that results from deleting the LZ. These domains would be tightly packed in the WT–protein dimer where they behave as a whole. Our proposal is confirmed by the curves obtained for ΔN37C133 and ΔN132 fragments: both of them show single transitions between two states with T_m values of 37.7 and 76.1°C, respectively, that closely resemble the values noted above for the first and second domains in ΔN37. The curve for His_6-ΔN42 loses the three step profile found in ΔN37, having instead a single transition between two states with a small slope and a low T_m (39.4°C), as expected for a loosely packed protein (Kelly and Price, 1997). Therefore, His_6-ΔN42 is really a misfolded control for ΔN37. Finally, dimeric ΔC133, in which the second RepA domain is absent, shows a linear profile in which the values of θ at 222 nm are essentially conserved up to 90°C. This thermostable behaviour suggests that the two copies of the first domain found in the ΔC133 dimers are even more tightly packed than they are in RepA–WT dimers. Thermal denaturation for all the other RepA proteins tested is irreversible.

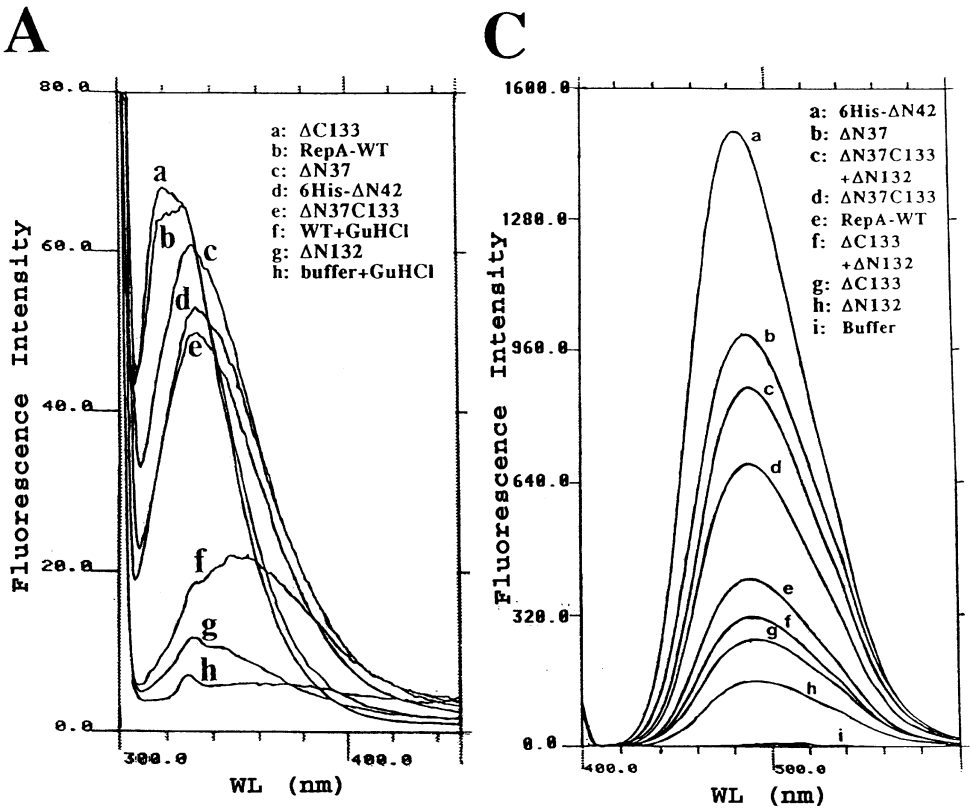
Thermal denaturation profiles were also performed with equimolar mixtures of either ΔC133 and ΔN132 (to mimic RepA–WT), or ΔN37C133 and ΔN132 (to resemble ΔN37) (data not shown). The three steps curve described above for ΔN37 was reproduced by mixing its N- and C-terminal domains, confirming the modular structure of this monomeric fragment. However, the highly cooperative transition observed for RepA–WT (see above) could never be reproduced by mixing its component fragments, suggesting that packing in RepA–WT dimers requires having both domains covalently linked.

Steady-state fluorescence spectroscopy

Intrinsic fluorescence indicates that Trp94 in RepA is buried in a hydrophobic core. Intrinsic fluorescence in proteins is dependent on the presence of Tyr and Trp residues, which have different spectroscopic properties. In order to selectively excite just Trp residues, a wavelength ⩾295 nm is used, so that it only tunes for the absorption tail of the indol but absorption by the Tyr phenol is negligible. Trp emission maximum varies from ~320 to 350 nm, depending on whether they are deeply buried in a hydrophobic cluster in a protein or totally exposed to the polar solvent, respectively (reviewed in Lakowicz, 1983; Schmid, 1997). This sensitivity of the emission wavelength of the fluorophore to its environment is useful in identifying hydrophobic cores in protein domains and protein–protein contacting interfaces. We have used the single Trp residue in RepA (W94; Figure 3A) as a spectroscopic probe for the presence of tertiary structure (that is, a defined three-dimensional fold) in the WT protein and its fragments.

W94 in each RepA protein (at 8.32 μM) was selectively excited at 295 nm and fluorescence emission spectra were acquired, either in ammonium sulfate buffer (native proteins) or in the same solution supplemented with GuHCl to 4.5 M (denatured proteins). Fluorescence emission maxima are between 320 and 334 nm for the different native RepA proteins. The exception is ΔN132, which has no Trp and so fluorescence emission is nearly null (Figure 5A and B). In GuHCl, emission maxima shift to values around 350 nm. This behaviour is expected if W94 is not exposed to the solvent in the native RepA proteins, but buried to different extents in a hydrophobic core.

Native ΔC133 shows the shortest wavelength for maximum emission and so it seems to have the most compact hydrophobic core, which we have tentatively related with its tight dimeric nature. The wider emission maximum for native RepA–WT expands from 320 nm (as ΔC133) to wavelengths closer to those for ΔN37, His_6-ΔN42 and ΔN37C133 (330–334 nm). This suggests the existence of at least two different states for W94 in RepA–WT, that



B

RepA protein	λemiss.max. (nm) (NH4)2SO4 buffer	+ GuHCl	Fbuff/FGuHCl
WT	320-325	347	3.07
ΔN37	330	350	2.61
6His-ΔN42	334	352	1.85
ΔN132	—	—	—
ΔN37C133	334	353	1.77
ΔC133	320	351	3.39

Fig. 5. Steady-state fluorescence spectroscopy on RepA–WT and its fragments. (**A**) Fluorescence emission spectra of W94 ($\lambda_{excitation}$ = 295 nm) in RepA proteins (8.32 μM) at 20°C. Spectra (identified as a–h in the inset and plotted to the same scale in arbitrary fluorescence units) were acquired in ammonium sulfate buffer (folded state), or in the same buffer supplemented with 4.5 M GuHCl (unfolded state; only that for RepA–WT is plotted since the others are similar). The spectrum for the buffer including GuHCl is also shown (the Raman band for water is the small peak at ~330 nm). (**B**) Maximum emission wavelength for each spectrum in (A). The ratios between the fluorescence intensities for the native and unfolded states (F_{buff}/F_{GuHCl}) are also shown. Data are compatible with all fragments having W94, to a different extent, buried in the protein hydrophobic core (an indication of being folded). (**C**) Extrinsic fluorescence emission of bis-ANS (16.64 μM) after incubation with 8.32 μM of each RepA protein (for 15 min at 20°C). The inset shows the identity of each sample, including equimolar mixtures of fragments that mimic full length and ΔN37 RepA. The emission intensity at 500 nm ranks the degree of exposure of hydrophobic regions and the compactness of RepA domains.

could be related with the presence of both protein dimers and monomers at the experimental conditions used, as confirmed by gel filtration analysis (see below; Figure 6, central panel). A compact dimeric conformation in RepA would result in a highly hydrophobic environment for W94, whereas in the monomers this residue would not be so deeply buried, pointing to a looser protein conformation. This can be quantified by the ratios between the emission intensities in ammonium sulfate buffer and in GuHCl (F_{buff}/F_{GuHCl}) for the different RepA proteins (Figure 5B), which distinguish between compact (WT and ΔC133, ratios >3.0), intermediate (ΔN37, ratio >2.5) and more relaxed (ΔN37C133 and His_6-ΔN42, ratios <2.0) species, compatible with our estimate for protein stability based on CD thermal profiles (Figure 4B).

The exposure of W94 to the solvent was also evaluated by quenching with acrylamide. W94 in all RepA fragments was significantly protected against quenching when compared with the behaviour of *N*-acetyl-tryptophaneamide, a control for fully exposed Trp (Lakowicz, 1983) (data not shown).

Finally, W94 and Tyr residues in each protein sample were simultaneously excited at 278 nm. A wide emission maximum around 320 nm in ammonium sulfate-containing

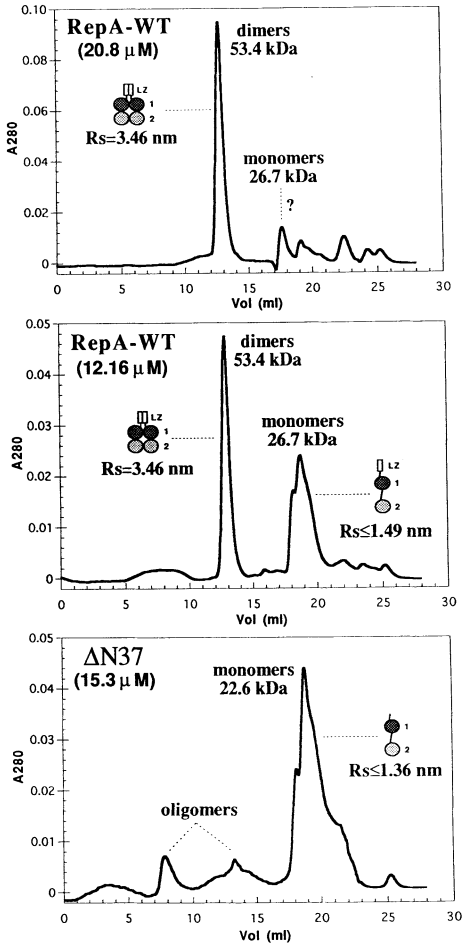


Fig. 6. FPLC gel-filtration analysis of RepA–WT and ΔN37. Protein samples in storage buffer were run at 0.4 ml/min through a Superose-12 HR-10/30 column at room temperature. Elution profiles were obtained monitoring A_{280} of the eluate. Insets to the panels indicate the sample identity, the concentration at injection and the species assigned to the observed peaks: monomers (wide and irregular), dimers (narrow and sharp) or larger aggregates. The estimated R_s and a cartoon with the proposed shapes (compact and spherical versus flexible and extended) for the proteins in the main peaks are also shown (colour coding for the domains is as in Figure 9).

buffer, due to fluorescence energy transfer from Tyr to Trp94, was resolved in GuHCl into two maxima at ~308 (Tyr) and ~350 nm (W94), due to uncoupling the emission of both types of aromatic residues (data not shown). This is an indication of the existence in all native RepA fragments of a network of aromatic residues within energy transfer distance (Lakowicz, 1983; Schmid, 1997).

An extrinsic fluorescent probe senses the degree of inter-domain compaction in RepA. A different way to study proteins by fluorescence spectroscopy relies on the use of extrinsic fluorescence probes such as 4,4′-dianilino-1,1′-binapthyl-5,5′-disulfonic acid (bis-ANS), which binds to hydrophobic patches in proteins when they are exposed to the solvent. On the contrary, bis-ANS hardly binds to hydrophobic residues when they are buried inside protein cores or hidden in protein–protein interfaces (Semisotnov *et al.*, 1991). If bis-ANS is bound, fluorescence emission is detected with a maximum at ~490–500 nm. There is a direct quantitative correlation between the amount of fluorophore bound and the intensity of the emission signal (Semisotnov *et al.*, 1991).

bis-ANS titrations were performed with RepA–WT or its fragments at 8.32 μM. The amount of fluorophore that gave maximum emission (16.64 μM) was selected to probe the exposure/compaction grade of the hydrophobic core in each protein (Figure 5C). His_6-ΔN42, which is thermally unstable (Figure 4B), has by far the most accessible hydrophobic core. Thereafter ranks ΔN37 and, closely, its 'reconstitution' by mixing equimolar amounts of ΔN37C133 and ΔN132. This nearly additive effect (within the experimental errors), as mentioned above for CD spectra, suggests that both RepA domains are quite independent in solution and, therefore, in RepA monomers they might be separated from each other in an extended arrangement. This proposal is confirmed by the emission spectrum for bis-ANS in RepA–WT (a mixture of dimers and monomers at this concentration; Figure 6, central panel). Its maximum is smaller than for monomeric ΔN37, suggesting that some hydrophobic patches exposed in RepA–WT monomers (most probably on the surface of the first domain, since ΔN37C133 binds more bis-ANS than ΔN132) could be sequestered from the solvent in the protein dimers. This could be the basis of the tendency to form heterogeneous poly-dispersed aggregates at high concentrations that exhibit both ΔN37 and ΔN37C133 fragments (see below; Figure 6). Finally, the most thermo-stable fragments (ΔC133 and ΔN132) have the least exposed hydrophobic residues and, therefore, the lowest fluorescence emission for bis-ANS.

The fluorescence experiments are proof, independent of CD spectroscopy, for the existence of two protein domains in RepA. They behave independently in the monomeric species (ΔN37), whereas they show a coupled behaviour in the dimeric one (RepA–WT). This suggests an extended versus a compact arrangement of both domains, respectively.

Gel filtration studies point to the existence of a compact conformation for RepA dimers versus an extended one for the monomers

Data presented above suggest the existence of two protein conformations for RepA in terms of secondary structure and interdomain compaction, which would also have a reflection in the overall form of both RepA dimers and monomers with further functional implications. To obtain a rough evaluation of RepA size and shape we have performed gel-filtration analysis (Ackers, 1970; Corbett and Roche, 1984) of the WT protein and its fragments.

The migration of a protein through a gel-filtration matrix is dependent on its mass, aggregation state and shape, related to its Stokes radius (R_s). Since the molecular masses and R_s values for the globular proteins used as calibration standards are tabulated (Potschka, 1987), it is possible to estimate these parameters for any protein. For a purified protein, a narrow and symmetrical elution peak reflects a single compact species, whereas a broader and more irregular peak shape is characteristic of a conformationally flexible molecule that is distributed in a heterogeneous population of species (Corbett and Roche, 1984).

Purified RepA–WT and its fragments were gel-filtered through a FPLC column that had been previously calibrated

with protein standards (data not shown; see Materials and methods). Dimeric RepA–WT (Figure 6, top panel) and ΔC133 (data not shown), injected at 20.8 μM, elute as sharp and symmetrical single peaks, suggesting that they have well-defined compact conformations. The estimated R_s values are 3.46 and 2.75 nm, respectively. When RepA–WT was injected at 12.16 μM, peaks corresponding to monomers and dimers are seen (Figure 6, central panel). At the same concentration, ΔC133 remains fully dimeric (data not shown), suggesting an even higher degree of compactness that could be related to the thermostability of this fragment (Figure 4B). The peak for RepA–WT dimers is coincident in position and sharpness with that seen when the more concentrated sample was injected (see above). However, the peak for the RepA–WT monomers is broader and has a more irregular shape (a shoulder is clearly observed). Very similar elution profiles are observed for the monomers of those RepA fragments comprising two domains without the LZ: ΔN37 (Figure 6, bottom panel) and His_6-ΔN42 (data not shown).

In general terms, monomeric RepA species elute with significant delay with respect to the expected values for globular proteins of their size (R_s values of 1.49, 1.36, 1.37, 1.23 and 1.10 nm for WT, ΔN37, His_6-ΔN42, ΔN37C133 and ΔN132, respectively). We have tentatively assigned an extended or elongated macromolecular shape to the peaks for WT, ΔN37 and His_6-ΔN42 monomers, since it has been shown that extended particles, due to a higher degree of penetration into the gel pores, tend to elute later than spherical ones with the same mass (Nozaki *et al*., 1976). We also propose that the linker between both domains could be flexible in the elongated species, generating a heterogeneous population of conformers responsible for the shoulders observed in the main peaks (Corbett and Roche, 1984). The elution profiles for RepA fragments that comprise the first domain with the LZ deleted (ΔN37, His_6-ΔN42 and ΔN37C133), in addition to monomers, show aggregated forms (e.g. species eluting with the exclusion volume) (Figure 6, bottom panel). This observation confirms our extrinsic fluorescence data on the existence of exposed hydrophobic surfaces, prone to aggregation, in those RepA fragments (see above).

DNA binding assays assign functions to RepA species: compact dimers bind to the operator whereas extended monomers bind to iterons

In previous work we had demonstrated that RepA–WT dimers bind to an inversely-repeated sequence (operator) at *repA* promoter (García de Viedma *et al*., 1995a), whereas RepA monomers bind to directly-repeated sequences (iterons) at the replicator (García de Viedma *et al*., 1996). We had also shown, by means of the isolation of defective mutants, that a C-terminal HTH motif in RepA is involved in protein binding to both operator and iteron sequences (García de Viedma *et al*., 1995b). In this paper we show that each RepA monomer is composed of an N-terminal LZ motif plus two globular domains (Figure 3A). We also propose that RepA dimers would have a compact arrangement of domains, whereas protein monomers would have a looser structure (see above). Thus, it is important to survey if, besides the monomer to dimer equilibrium (García de Viedma *et al*., 1996), having two protein conformational states in RepA provides the structural basis for the functional recognition of operator or iteron DNA sequences.

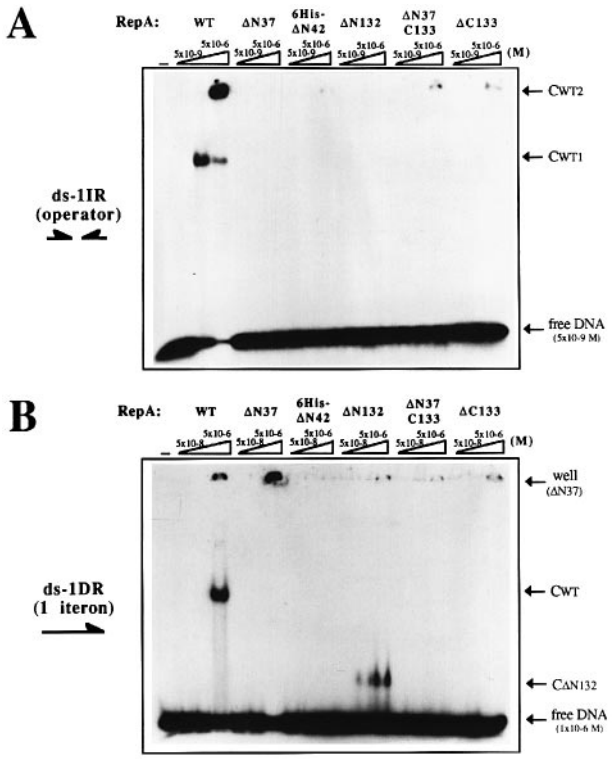


Fig. 7. EMSA of pPS10 ds-DNA sequences in complex with RepA (or its fragments). (**A**) A 39 bp (^{32}P-labelled) double-stranded oligonucleotide, including the *repA* operator sequence (ds-1IR), was incubated at 5×10^{-9} M with increasing amounts (ramp headers, from 5×10^{-9} to 5×10^{-6} M) of either purified RepA–WT or each of its fragments, and then run in a 6% polyacrylamide native gel. The autoradiograph shows two complexes for the WT protein (CWT1 and CWT2) (see text), and none for any of the RepA fragments. (**B**) The same type of experiment was performed with 1×10^{-6} M of a radiolabelled 45 bp double-stranded oligonucleotide including a single iteron plus the DnaA box (ds-1DR). DNA was incubated with purified RepA–WT or each of its fragments (concentration range from 5×10^{-8} to 5×10^{-6} M). Low-affinity single protein–DNA complexes appear only for RepA–WT (CWT), ΔN37 (well) and ΔN132 (CΔN132).

EMSA analysis of the DNA binding properties of RepA and its domains. Titrations of ^{32}P-labelled double-stranded oligonucleotides, containing either the operator or single iteron sequences, with RepA–WT and its fragments were performed by means of electrophoretic mobility shift assays (EMSA) (Figure 7).

RepA–WT binding to the operator sequence (Figure 7A) results in two retarded bands, corresponding to an initial RepA dimer–DNA complex and a second dimer bound to that complex (García de Viedma *et al*., 1995a). RepA–WT binding to the iteron sequence (Figure 7B) results in a single retarded band with higher mobility than the complex of RepA dimers at the operator, reflecting a unique protein monomer bound to DNA (García de Viedma *et al*., 1996). According to RepA–WT titrations, the values for the apparent dissociation equilibrium constant (K_{dapp}), estimated as the protein concentration that gives 50% of bound DNA (Rhodes, 1989), were ~5×10^{-8} M for RepA–WT binding to the operator and ~5×10^{-6} M for iteron binding. Higher affinity to the iterons requires the presence of at least two of the four 22 bp directly-repeated

sequences, probably reflecting cooperativity (data not shown). Leaving aside the large difference in affinity, the binding of RepA–WT to both kinds of DNA sequences is specific since it resists competition with polydI–dC and a mutated single iteron in excess (data not shown). There is a contribution of the DNA minor groove in the recognition by RepA of both types of DNA sequences; RepA–WT binding is displaced by micromolar concentrations of drugs that bind to the minor groove, such as Distamycin-A and Hoechst-33258 (reviewed in Wemmer and Dervan, 1997) (data not shown; see the footprinting data below). This was not evident from our previous description of the role of a major groove binding module, the HTH motif, in DNA recognition by RepA (García de Viedma *et al*., 1995b).

According to EMSA, apart from RepA–WT, no RepA fragments bind to the operator sequence (Figure 7A). This was expected, since either the fragments do not include the C-terminal protein domain that comprises the HTH motif (as in the cases of ΔN37C133, ΔC133) and/or the fragments are lacking the LZ motif required to dimerize, and so do not satisfy the symmetry of the operator (as in ΔN37, ΔN37C133, ΔN132 and His_6-ΔN42 that, as shown above, has its N-terminal domain misfolded). On the contrary, binding to the single iteron sequence is observed for at least two RepA fragments (ΔN37 and ΔN132) (Figure 7B). ΔN37 binds with a similar affinity as RepA–WT (K_{dapp} ~5×10^{-6} M), but the complex is less soluble and remains in the well of the gel. A complex of ΔN132 with the iteron appears with higher electrophoretic mobility than that observed for RepA–WT, as expected from its smaller protein molecular weight, and starts to be observed at a protein concentration 10-fold lower than that required for RepA–WT. However, the diffuse aspect of the band for ΔN132–iteron complex suggests a fast dissociation under the standard electrophoretic conditions. The same dissociation problem might well hide the existence in solution of short living complexes for those RepA fragments that would not bind to iteron DNA according to EMSA criteria.

Hydroxyl-radical footprinting maps the binding site for each RepA domain on the iteron sequence. We had previously obtained both DNase I and high resolution hydroxyl-radical footprinting data on the complexes formed by RepA–WT and the operator DNA sequence (García de Viedma *et al*., 1995a; for a summary see Figure 9A). We have now extended these studies to the complexes formed by RepA–WT and its fragments with the same double-stranded single iteron oligonucleotide used in EMSA (ds-1DR; Figure 7B), in order to: (i) analyze possible protein–DNA complexes of low stability in solution, previously undetected by EMSA; and (ii) investigate the structural basis for the functional preference of the monomeric (extended) over the dimeric (compact) RepA conformation for iteron recognition.

Under the experimental conditions in which RepA–iteron complexes were evident in EMSA (Figure 7B), we have been unable to detect any protection against DNase I. Since this nuclease and, at least in part, RepA (see above) recognize DNA through the minor groove, we believe that DNase I must displace bound RepA. Hydroxyl-radical footprinting reactions were then performed in order to test the protection by RepA of iteron DNA against a small chemical reagent that does not usually displace bound proteins. The degree to which hydroxyl-radicals cleave any deoxyribose link in DNA is decreased by the close binding of a protein. The oligonucleotide ds-1DR was incubated with RepA–WT (at 5×10^{-6} M, the K_{dapp} value according to the EMSA assays; Figure 7B), or with each RepA fragment at the same concentration, and the footprinting reactions were made. The results for the bottom strand are shown in Figure 8 (in Figure 9B, a summary of the footprints for the complementary top strand is also shown).

RepA–WT binding to a single iteron results in the protection of three sequence patches against hydroxyl radicals. These are marked as I–III in Figure 8, following the convention determined by the direction of the drawn arrow: 5′ to 3′ through the complementary top strand (Nieto *et al*., 1992). The protection is not complete, in agreement with the value of 50% DNA binding obtained by EMSA at that protein concentration (Figure 7B), and varies for each region (I–III). ΔN37 protects all three patches (I–III) evenly and strongly (by ~40%), compatible with its tight binding in EMSA. Although His_6-ΔN42 was expected not to bind to the iteron, due to misfolding problems (see above), it protects sequence patch II by just ~20%. Surprisingly, ΔN37C133, which did not bind to iteron DNA in EMSA, protects the three sequence patches almost as much as ΔN37 does (by ~35%). ΔN132 binds to iteron DNA in EMSA and strongly protects (by ~40%) sequence patch II. ΔC133 protects regions I and II in a negligible way, in accordance with its inability to bind DNA in EMSA. Finally, Distamycin-A, a drug that competes with RepA–WT binding to the iteron (see above), produces clear and uneven protection of the three sequence regions (I–III).

Hydroxyl-radical footprinting results allow us to map the binding site for each RepA domain along the iteron sequence (Figure 9B). ΔN132 strongly protects sequence patch II (that includes a 6 bp core sequence common to the operator and iteron repeats; Figure 9) and it also extends, though very weakly, further to region III (the boundary between the 3′ end of the iteron and the linker with the contiguous DnaA box). So, RepA C-terminal domain, including a HTH motif (García de Viedma *et al*., 1995a), binds to sequence patch II. His_6-ΔN42 binds weakly to the same place as ΔN132, suggesting a residually active conformation for its C-terminal domain.

The backbone contacts made by the proteins at the iteron sequence were then plotted on a model for a B-DNA double helix. The C-terminal domain in RepA binds to DNA mainly through the major groove, as expected for a HTH motif (Brennan and Matthews, 1989). ΔN37C133 (RepA first domain) binds not only to the 5′ end of the iteron (patch I), but also further to the 3′ edge (patches II and III). Although the protection found is reproducible and quite strong, the binding of ΔN37C133 to DNA must be of very low affinity (or fast dissociation), since we have been unable to detect it in EMSA (Figure 7B). It seems that the first domain in RepA, when it is found in the absence of the second domain and not in a dimeric and compact form (that is, ΔN37C133 rather than ΔC133), is also able to recognize the natural high-affinity site for the C-terminal domain (ΔN132). This unexpected result

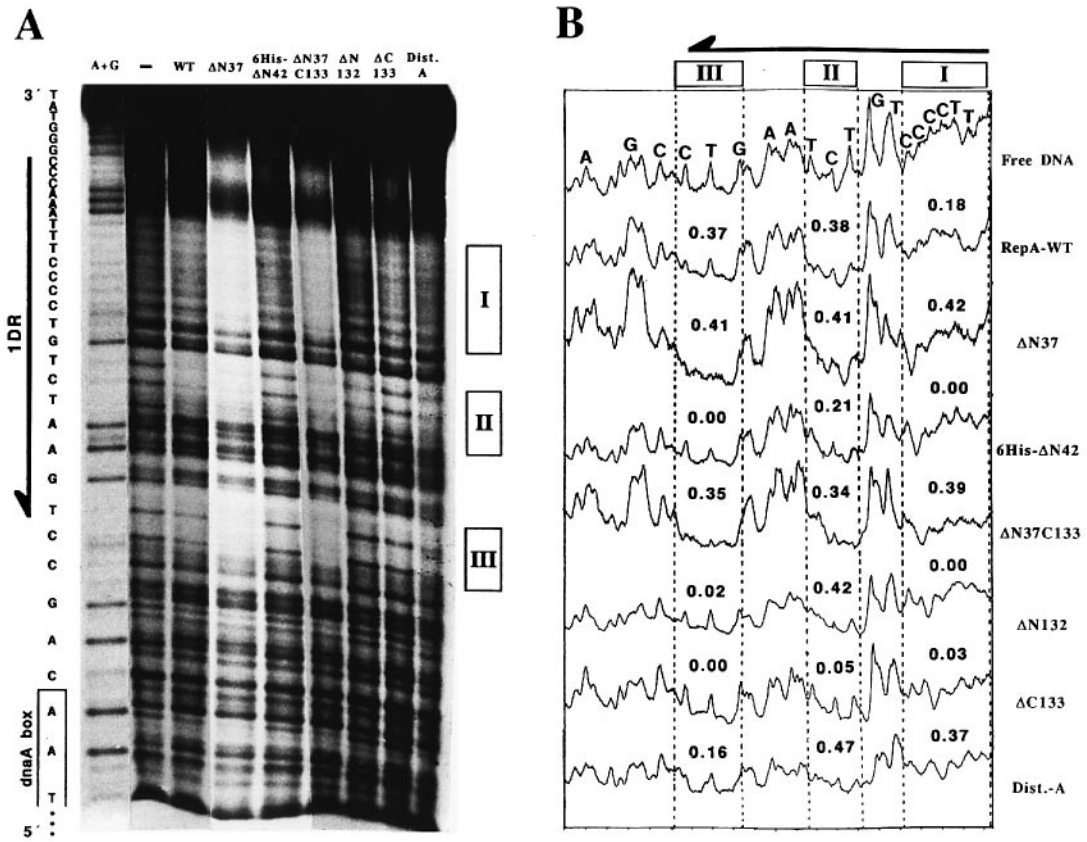


Fig. 8. Hydroxyl-radical footprinting of the complexes formed by RepA–WT (or its fragments) and the single iteron oligonucleotide. (**A**) Protection experiment performed with the indicated (headers on the gel tracks) RepA proteins (5×10^{-6} M) or Distamycin-A (25 μM) on the ds-1DR oligonucleotide (5×10^{-7} M), labelled at the bottom strand. The DNA sequence (A+G) is shown on the left, with the iteron (arrow) and the DnaA box (open rectangle) indicated. The three sequence patches (I–III) protected by the RepA proteins are drawn to the right of the gel.
(**B**) Densitometric PhosphorImager tracing of the gel tracks shown in (A). The sequence of the enlightened region is written over the plot for the free DNA. Drawing codes for the iteron and the protected sequences are as in (A). The numbers written over the plots for protein (or drug)–DNA complexes reflect the calculated protection fractions for each sequence patch.

suggests that both RepA domains, once the conformational change coupled to protein monomerization occurs, could share a similar three-dimensional structure (see Discussion and Figure 2C). Plotting the contacts made by RepA at the 5′ region of the iteron (patch I), it is evident that protein interaction in this place must involve the recognition of the DNA minor groove (Figure 9B), although we cannot exclude extra contacts through the major groove. This provides the basis to explain the displacement of RepA–WT binding by Distamycin-A (see above).

RepA–WT binds strongly to the 3′ end of the iteron sequence but weakly to the 5′ end, possibly due to a lower proportion of active extended monomers than compact dimers in the binding reactions (Figure 6, central panel). Monomeric ΔN37 is the protein species able to bind strongest and most evenly to the iteron DNA along its full sequence. Therefore, the extended conformation of RepA appears to be preferred for iteron binding, rather than the compact one, since it allows the first domain to recognize efficiently the 5′ end of the iteron. The protection of the sequence found in the boundary between the 3′ end of the iteron and the DnaA box (patch III) requires the simultaneous binding to the 5′ (I) and 3′ (II) iteron regions by either both RepA domains (as in RepA–WT and ΔN37) or two molecules of the first domain (as in ΔN37C133). This suggests that protection of region III could reflect a conformational change in the DNA, induced by RepA bound to an iteron in its full extension.

Discussion

In this paper we propose a structural base for the dual function of the RepA protein of the *Pseudomonas* plasmid pPS10, a DNA replication initiator that is also able to repress its own gene expression by means of binding to either directly or inversely-repeated DNA sequences, respectively. Using an *in vitro* approach, including RepA limited proteolysis and N-terminal sequencing, cloning and expression of RepA fragments, CD and fluorescence spectroscopies, gel filtration and protein–DNA interaction functional studies, we have found in RepA two protein globular domains, C-terminal to a LZ motif. Furthermore, we also propose that dissociation of RepA dimers into monomers would result in a conformational change from a compact arrangement of the protein domains, competent for binding to the operator, to an extended form that is suited for iteron binding.

A model for the relation between RepA protein structure and DNA-binding function

RepA dimers, stabilized by the N-terminal LZ motif, would adopt a compact conformation of domains to

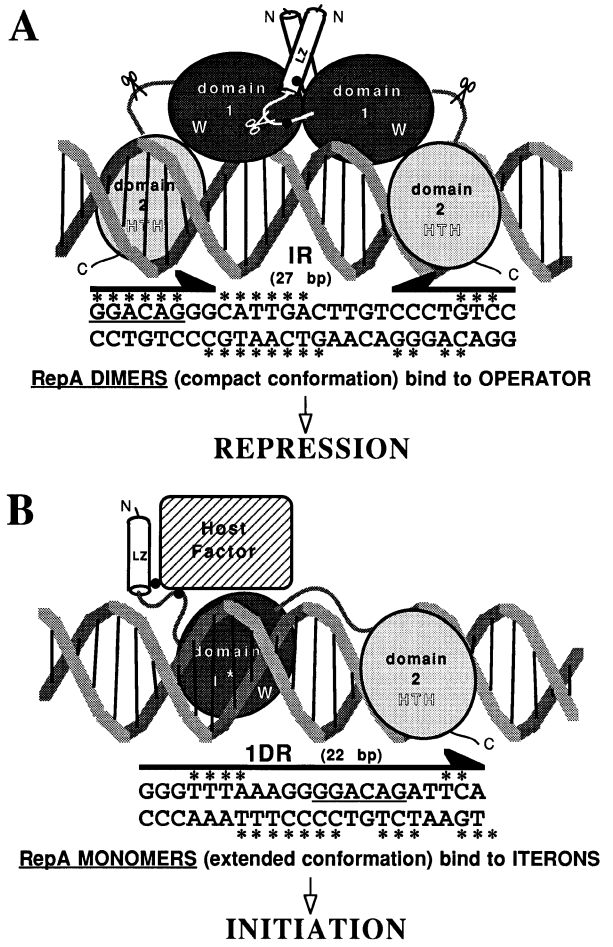


Fig. 9. A model on the structural basis for the dual DNA-binding function of RepA. (**A**) A compact dimeric conformation in RepA, established by a LZ coiled-coil and by contacts between the first globular domains, would be functional in the recognition of the operator DNA sequence (IR, inverted arrows). An HTH motif, found at the second globular domain, would bind through the major groove at opposite sides of a B-DNA double helix. Backbone contacts on both DNA strands (*) are according to García de Viedma *et al.* (1995a). This type of complex would result in the repression of *repA* promoter. Protein loops connecting the domains would be accessible to proteases (scissors), whereas residues (black dots) involved in the interaction with other replication factor/s would be hidden in the dimer interface. (**B**) An extended conformation of RepA monomers would be functional in the recognition of the iteron (1DR) sequence. The N-terminal domain recognizes the minor groove at the 5′ region of the iteron sequence, whereas the C-terminal domain (HTH motif) would recognize through the major groove the 3′ sequence core conserved in both IR and DR (underlined). Protein contacts at the DNA backbone in both strands are shown (*). Residues involved in binding to the host factor(s) would now be exposed and available for interactions. This type of complex would be competent for initiation. For further details, see text.

recognize the inversely-repeated operator target (Figure 9A). In this case, protein–DNA interactions would be mainly established between the HTH motif found in the second RepA domain and the DNA major groove. Some additional contacts of the first globular domain with the DNA backbone would fit with our previously published footprinting results on this kind of complex (García de Viedma *et al*., 1995a). Assuming a B-DNA structure for the operator sequence, the two copies of the second domain found in RepA dimers would bind to the double helix at opposite sides (~2.5 turns, although just 1.5 helical turns are schematically represented in Figure 9A). It is noteworthy that the footprinting of RepA–WT dimers at the operator sequence is asymmetric (García de Viedma *et al*., 1995a), with an extended protection on the left arm of the palindrome. Since up to two RepA dimers can be bound to the operator (García de Viedma *et al*., 1995a; Figure 7A), they could either be asymmetrically arranged or their binding could result in bending the left arm of the DNA palindrome towards the protein. Further experiments are required to address this question.

RepA monomerization would result in a protein species able to bind to the whole sequence of an iteron in a highly efficient manner, since the two protein domains would be arranged in an extended conformation (Figure 9B). The C-terminal domain in RepA would bind to the 3′ end of the iteron DNA (according to the orientation conventionally defined in Nieto *et al*., 1992) with a binding geometry similar to that found in the operator sequence (Figure 9A), since there is a common core sequence (GGACAG) in both types of DNAs that would be the target for the HTH motif (García de Viedma *et al*., 1995b). Coupled to RepA monomerization, a conformational change in the N-terminal domain would allow this to bind to the 5′ region of the iteron sequence. Plotting the contacts made by both RepA domains on the backbone of a B-DNA double helix, it turns out that the C-terminal domain (HTH) would be located at the major groove, whereas the N-terminal one would be laid on the minor groove, in such a way that they would be bound at nearly opposite faces of the DNA (~205° right-handed rotation and a coupled 12 bp translation relate the central base-pairs contacted by each domain). The protein–DNA contacts with the bottom strand backbone found at the middle of the iteron sequence suggest that the linker between RepA domains crosses over that strand. The extended conformation of RepA monomers, flexible in solution (Figure 6) and with low affinity for a single iteron sequence (Figure 7B), could be fixed after cooperative high-affinity binding to the multiple iteron sequences. A detailed structural and kinetic analysis is required to address this possibility.

The predicted secondary structure elements for both RepA domains (three α-helices flanked by three β-strands; see Figure 2C) resemble those found in a ‘winged-helix’ domain, a DNA-binding module that includes a HTH motif (Brennan, 1993). This observation fits with our results on the existence of two functional DNA binding domains in RepA and the way they bind: a winged-helix domain recognizes the DNA through both the major and the minor grooves, using a HTH motif in a three helix-bundle (shaded rectangles in Figure 2C) and a strand–loop–strand ‘wing’, respectively.

An analysis of the information contents of the iteron sequences in several plasmids has led to the proposal for the existence of two distinct conserved sequence patches (TGTG and CCC), spaced by 11 bp, that would be recognized by two hypothetical domains in a Rep-type protein (Chattoraj and Schneider, 1997). These sequences loosely match those we have found to be recognized at the pPS10 iterons by the second and first RepA domains, respectively: TcTG (patch II) and cCCCttt (patch I), although they are only ~6 bp apart in pPS10 iterons (Figures 8 and 9). Some additional information can be

extracted from comparative analyses with the iterons of pCEB2, a plasmid which Rep protein shares 82% sequence identity with pPS10 RepA (Charnock, 1997); the 3′ region of iterons (sequence patches II and III) is identical for both plasmids, whereas the 5′ region (patch I) changes in pCEB2 to aCC(t/C)(c/g)(a/T)T. This suggests a common binding mode for both initiators that has conserved the contacts for the C-terminal domain, but has developed alternative ones for the N-terminal domain.

Both the LZ and the first domain mediate protein–protein interactions in Rep-type initiators

Biochemical studies on several mutants in the Rep protein of F plasmid have identified, in addition to the N-terminal LZ (Giraldo *et al*., 1989), a central region involved in dimerization (Matsunaga *et al*., 1997). This region is coincident with the N-terminal domain of pPS10 initiator (del Solar *et al*., 1998) and we have observed the formation of protein oligomers by those RepA fragments (ΔN37, His_6-ΔN42 and ΔN37C133) in which the first domain is present in the absence of the LZ (Figure 6; our unpublished results). The LZ could then be a specific dimerization interface that modulates the tendency of the first domain to aggregate.

The Rep-coding gene in R6K plasmid has two alternative translational start codons, generating two protein products. The largest one includes an LZ (Giraldo *et al*., 1989) whereas the shortest does not, although both can form dimers pointing to the existence of a second dimerization interface (Filutowicz *et al*., 1994; Wu *et al*., 1997). We speculate that the full length dimer is mainly stabilized by the LZ motif, whereas the small dimer would be held together as we propose for the N-terminal domain in pPS10 RepA.

The aggregation through the first Rep domain of monomeric initiator molecules bound to iteron arrays could be the basis for the negative regulation mechanism called ‘hand-cuffing’, extensively characterized for R6K, RK2 and P1 plasmids (reviewed in Helinski *et al*., 1996; del Solar *et al*., 1998). It could also be related to the ability of the initiator of R6K to bring together, by protein–protein interactions, distant replicator DNA sequences where it binds (Miron *et al*., 1992). A mutation (P113S) in the proposed N-terminal domain of this Rep protein has been described as impairing DNA looping and hand-cuffing (Miron *et al*., 1994). We have previously shown that Rep-mediated DNA looping is also operational in the initiation complex of R1 plasmid (Giraldo and Díaz, 1992).

The structural and functional analogies between AraC, the dimeric transcriptional regulator of the arabinose catabolic genes in *E.coli*, and Rep-type proteins are noteworthy. AraC monomers are composed of an N-terminal dimerization domain linked to a C-terminal DNA binding-domain. The crystal structures of both the arabinose-ligated and unligated forms of the N-terminal domain (Soisson *et al*., 1997) show a β-barrel core and two antiparallel α-helices that establish a four-stranded coiled-coil in the dimers. Whereas the arabinose-bound species dimerizes exclusively through the coiled-coil, the unbound one establishes additional interactions between β-barrels that result in the formation of protein aggregates involved in DNA looping.

Structural basis for the phenotypes of some point mutations in Rep-type proteins

Many mutations have been isolated in Rep-type proteins affecting their function as initiators. Some of them map at the C-termini of the Rep proteins (e.g. those in the 154–208 region of F plasmid initiator; Matsunaga *et al*., 1997) clustering at the second protein domain, as defined here for pPS10 RepA (see García de Viedma *et al*., 1995b). They would affect either the HTH motif, and thus the ability of the Rep proteins to bind to DNA, or residues in the interface with the first domain, favouring or hampering the extended initiator conformation.

Other mutations that result in more efficient initiators have been isolated in the first domain of Rep proteins, between residues 138–184 in P1 (Sozhamannan and Chattoraj, 1993), 92–115 in pSC101 (Xia *et al*., 1993), 93–135 in F (Kawasaki *et al*., 1991) and 81–113 in R6K (York and Filutowicz, 1993). In the initiator protein of R6K, the putative RGD dimerization motif (residues 92–94) (Levchenko *et al*., 1994) and the S87N mutation, resulting in hyperactive initiation and cooperative iteron binding (York and Filutowicz, 1993; Levchenko *et al*., 1997), also map at the N-terminal domain. Therefore, we propose that these mutations would favour Rep monomerization and/or they could stabilize the extended conformation responsible for the initiation activity. Alternatively, we cannot discard that some of them could directly increase the iteron-binding activity that we have found in the N-terminal domain of pPS10 RepA.

Chaperone action could couple monomerization and conformational changes in Rep-type proteins

Molecular chaperones can promote the formation of Rep monomers *in vitro* (Wickner *et al*., 1992, 1994; DasGupta *et al*., 1993; Konieczny and Helinski, 1997; Pak and Wickner, 1997). However, at the concentrations used in the iteron binding assays, a significant amount of Rep must be monomeric (Figure 6, central panel; Ingmer *et al*., 1995; Chattoraj *et al*., 1996; García de Viedma *et al*., 1996). Some monomeric mutants in the P1 initiator have higher association rates to iteron DNA *in vitro* and show increased initiation frequency (Mukhopadhyay *et al*., 1994), but still require the action of chaperones *in vivo* (Sozhamannan and Chattoraj, 1993). This has led to the proposal that chaperones must also induce a conformational change in the Rep protein that ‘remodels’ its structure to become active in iteron binding (Dibbens *et al*., 1997).

We have shown previously that pPS10 replication *in vitro* requires DnaK chaperone (Giraldo *et al*., 1992), although dissociation of RepA dimers can be spontaneous (Figure 6, central panel; García de Viedma *et al*., 1996). It is tempting to propose that some of the RepA fragments that we have characterized here could mimic either substrates (ΔC133 dimers which remain tightly associated and compact at any concentration tested) or products (ΔN37 monomers, the extended species active in iteron binding) of chaperone action. We propose in this paper that RepA fragments containing the first domain, in the absence of the LZ, are prone to aggregation and that contacts between both copies of this domain could reinforce the hydrophobic packing between RepA molecules in the dimer. Therefore, the first domain in Rep-type

proteins could be the target for chaperone-mediated monomerization and stabilization of the initiator. Our results would provide a structural framework for understanding chaperone-assisted activation of Rep proteins. The high concentration of DNA required for RepA binding to a single iteron (1×10^{-6} M; Figure 7B) leaves the possibility of ligand-induced monomerization of RepA–WT dimers open. This mechanism for remodelling RepA structure could be a case for the allosteric effect of DNA substrates in the structure of DNA binding domains (Lefstin and Yamamoto, 1998).

Protein conformational changes: a universal feature in the activation of initiators?

The Epstein–Barr virus protein EBNA1 is a tight dimer that recognizes four copies of an inversely-repeated sequence in the replicator DNA. EBNA1 is the only initiator for which a three dimensional structure of its DNA binding-domain is available both in the absence and in the presence of a specific DNA fragment. A two-step mechanism for EBNA1 binding has been proposed (Dean and O'Donnell, 1996); an initial recognition of the DNA minor groove by flexible protein loops would be followed by a conformational change in EBNA1 that would allow an α-helix from each protomer to contact the major groove and bend the DNA. Overall, this model resembles our proposal for the existence of two major RepA conformations for DNA recognition.

Several of the six subunits of yeast ORC, the eukaryotic DNA replication initiator (Bell and Stillman, 1992), contact the conserved A and B1 sequences at the ARS replicators, where an extensive footprint on the DNA is found for each one of them (Lee and Bell, 1997). This suggests that, as the Rep-type initiators, they could adopt extended protein conformations when bound to the ARS.

The evidence that we present in this paper for the activation of a bacterial Rep-type initiator through the conversion between two protein structural states could well be the reflection of a universal mechanism for gaining a conformation efficient for DNA binding and distortion, the earliest steps in the initiation of DNA replication.

Materials and methods

Cloning and expression of pPS10 RepA protein and its fragments

In order to express large amounts of RepA proteins, a new expression vector was constructed (pRG-recA-NHis). Details will be published elsewhere. In summary, the *Nar*I–*Hin*dIII fragment of pUC18 (Yanisch-Perron *et al*., 1985) was substituted by tandem ligation of three linker fragments comprising: (i) the *recA* promoter (–42 to +13) with a *lexA* operator (Horii *et al*., 1980); (ii) the T7 ϕ10 gene leader sequence, including the ATG initiation codon embeded in a *Nco*I cloning site; and (iii) the coding sequence for the fusion peptide GSS**HHHHHH**SSGLVPRGS, that comprises a His_6 affinity tag (in bold) for immobilized metal-affinity chromatography (IMAC) and a thrombin cleavage-target (underlined), with a unique *Sac*II site in its DNA coding sequence. The resulting 2727 bp plasmid has a number of improvements over other *recA* promoter-based expression vectors (Olins *et al*., 1988): (i) it retains the complete multiple cloning sites from pUC18; (ii) it has a higher copy number; (iii) transcription occurs in the opposite direction to the Amp-resistance gene; and (iv) the N-terminal fusion peptide allows affinity purification and then thrombin cleavage.

The DNA pieces coding for RepA–WT and its fragments (Figure 3A) were obtained by 25 cycles of PCR with *Pfu* DNA polymerase (Stratagene) on pCN38 plasmid template (Nieto *et al*., 1992). Primer oligonucleotides (10 pmol/reaction) were made in a Beckman DNA synthesizer 1000M (solid-phase phosphoramidites chemistry, 0.1 μmol scale) and used without further purification. Coding strand oligonucleotides included (5′→3′): GCT tail, *Sac*II cloning site and a sequence annealing with the 5′ end of the relevant *repA* fragment, with estimated T_m values (Sambrook *et al*., 1989) between 60 and 62°C. Non-coding strand oligonucleotides comprised (5′→3′): GCT tail, *Hin*dIII site, TCA (complementary to TGA stop codon) and a sequence annealing with the 3′ end of each *repA* fragment, with T_m values selected as quoted above. PCR products were phenol/chloroform-extracted, ethanol-precipitated (Sambrook *et al*., 1989) and digested with *Sac*II and *Hin*dIII restriction endonucleases. DNA fragments were recovered from 1% agarose–TAE gels (Sambrook *et al*., 1989) using the Geneclean kit (Bio-101) and ligated with the pRG-recA-NHis vector. Recombinants were transformed into the *E.coli* JM109 strain (Yanisch-Perron *et al*., 1985) by the $CaCl_2$ method and selected at 37°C on LB-agar plates supplemented with ampicillin to 100 μg/ml (Sambrook *et al*., 1989). Plasmid DNA from selected clones was purified by the alkaline-lysis method (Sambrook *et al*., 1989) and verified by restriction analysis. Both DNA strands of positive recombinants were sequenced in a Perkin-Elmer ABI-Prism 377 DNA sequencer using fluorescent chain terminators.

The expression of each RepA construct was tested in C600 (Bachmann, 1996), CAG629 (*lon*) (New England BioLabs) and SG22099 (*clpA*), SG22080 (*clpX*) and SG22097 (*clpXP*) (Gottesman, 1990) *E.coli* strains. Preliminary tests were performed at 10 ml culture scale in 2× TY medium (Sambrook *et al*., 1989), supplemented with ampicillin to 100 μg/ml. Cells were grown at several alternative temperatures to an OD_{600} in the range of 0.35–0.50. The inducer (nalidixic acid, Sigma, from a fresh 10 mg/ml stock in 0.2 N NaOH) was then added to different final concentrations (10–100 μg/ml) and cells from 100 μl aliquots of the culture were harvested after different incubation times. Protein expression was checked by SDS–PAGE (15% polyacrylamide gels) (Laemmli, 1970) with New England BioLabs broad-range protein molecular weight markers. The best results were obtained after overnight induction with 25 μg/ml of nalidixic acid, in the SG22097 strain, at 33°C (RepA–WT, ΔC133) or 37°C (ΔN37, ΔN42 and ΔN37C133), and in the CAG629 strain at 28°C (ΔN132). This points to the involvement of ClpXP and Lon proteases in the degradation of RepA *in vivo*.

Purification of RepA protein and fragments

Large-scale protein expression was performed at the induction conditions determined in the small-scale pilot experiments, but in 5 l flasks with 3 l of medium. Cells were harvested by centrifugation (at 5000 r.p.m. for 15 min at 4°C, in a Sorvall GS3 rotor) and washed in 50 ml of cold 0.9% NaCl. Typical yields are between 6.5–8.5 g of wet cell mass, depending on the RepA construct. Cell pellets were resuspended in 50 ml of lysis solution (1.0 M KCl, 60 mM imidazole–HCl pH 8.0, 1% Brij-58, 1 mM p-NH_2-benzamidine, 10% glycerol) and then disrupted using a MSE sonicator (1 cm wide tip, 3–4 cycles, 45 s/cycle, at 60% amplitude) in an ice bath. Cell debris was separated from the soluble lysates by ultracentrifugation (15 000 r.p.m. for 60 min at 4°C, in a Beckman Ti45 rotor). Two main procedures were followed for protein purification, depending on having His_6-RepA soluble (WT, ΔN132 and ΔC133) or aggregated as inclusion bodies (ΔN37, ΔN42 and ΔN37C133).

WT, ΔN132 and ΔC133 purification. Chromatographic steps were performed at 4°C in a temperature-controlled cabin equipped with a Pharmacia standard chromatography setting. 10 ml of Chelating Sepharose Fast-Flow (Pharmacia) were packed into a XK-16/20 column (Pharmacia) and activated with $NiCl_2$ in excess. The column was equilibrated with at least 10 volumes of IMAC solution A (1.0 M KCl, 60 mM imidazole–HCl pH 8.0). Soluble lysate was then loaded at 2.0 ml/min flow rate into the column. This was extensively washed with IMAC solution A and then a linear 250 ml gradient was run from 0–100% of IMAC solution B (1.0 M KCl, 300 mM imidazole–HCl pH 8.0) at 1.0 ml/min. Five millilitre fractions were collected and analyzed by means of SDS–PAGE. Those containing His_6-RepA proteins were pooled, supplemented with $CaCl_2$ to 2 mM and with β-MeEtOH to 5 mM, and then digested with 50 units of Thrombin (Sigma) at room temperature for 4–6 h. The completeness of the digestion was verified by SDS–PAGE and then the protein buffer was changed to 0.1 M $(NH_4)_2SO_4$, 20 mM Mes–NH_4 pH 6.0, 5 mM β-MeEtOH, 0.1 mM EDTA, 10% glycerol by extensive dialysis at 4°C. In a second purification step, RepA proteins were loaded (at 1.0 ml/min) into 15 ml of SP-Sepharose Fast-Flow (Pharmacia) cationic exchanger, previously packed into a XK-26/20 column and equilibrated in dialysis solution. A 250 ml linear gradient was run, at 1.0 ml/min, from 0–100% of the same solution, but 0.5 M in $(NH_4)_2SO_4$. Fractions containing purified RepA

proteins were pooled, supplemented with $(NH_4)_2SO_4$ to 0.5 M and then concentrated to 0.75–1.0 mg/ml in an Amicon 50 ml filtration cell (kept in an ice bath) with a Diaflo PM-10 membrane. Protein concentration was determined by measuring A_{280} in 5 M GuHCl and calculating the molar extinction coefficients (18910, 5960 and 12950 $M^{-1}\cdot cm^{-1}$ for WT, ΔN132 and ΔC133 proteins, respectively) (Pace and Schmid, 1997). Purity of RepA preparations was determined to be ~95% by SDS–PAGE and Coomassie Blue staining. To achieve further purity assessment, 0.2 ml protein samples were dialyzed against 0.1 M NH_4-acetate pH 6.0 and then N-terminal sequencing (six cycles in an Applied Biosystems Procise-494 Protein Sequencer, running in pulsed-liquid mode) and amino acid analysis (Pharmacia Biochrom, standard hydrolysis in HCl with ninhydrine postcolumn detection) were performed. Protein stocks were stored at –80°C and are stable for at least a year.

ΔN37, ΔN42 and ΔN37C133 purification. Inclusion bodies were solubilized by means of 1 min sonication in 5 M guanidinium hydrochloride (GuHCl), 50 mM imidazole–HCl pH 8.0. Remaining particles were removed by ultracentrifugation (30 000 r.p.m. in a Beckman Ti45 rotor, for 60 min at 4°C) and the soluble fraction was loaded at 2.0 ml/min into the same Ni-IMAC column used for the soluble proteins (see above), but equilibrated in the GuHCl solution. After extensive column washing, a linear 250 ml gradient was run from 0–100% of column solution B (5 M GuHCl, 300 mM imidazole–HCl pH 8.0). Fractions containing His_6-RepA fragments were identified by SDS–PAGE and pooled. After adding β-MeEtOH to 10 mM, 10 ml aliquots of denatured proteins can be stored at –80°C for several months. These were diluted 10-fold to ~0.2 mg/ml (final volume 100 ml) with 5.6 M GuHCl, 0.56 M $(NH_4)_2SO_4$, 200 mM NH_4-acetate pH 6.0, 20 mM β-MeEtOH, 0.2 mM EDTA, 1.2% CHAPS (SIGMA). Fragment refolding was carried out by dialysis at 4°C, against 6× 1 l of 0.5 M $(NH_4)_2SO_4$, 50 mM NH_4-acetate pH 6.0, 10 mM β-MeEtOH, 0.1 mM EDTA, 10% glycerol for a total time of 36 h. Aggregated material was then removed by ultracentrifugation (30 000 r.p.m. in a Beckman Ti45 rotor, for 60 min at 4°C) and the soluble fraction was concentrated in an Amicon cell as described above. For removal of the fused peptide from His_6-ΔN37 and His_6-ΔN37C133, 1 ml of concentrated protein stock was diluted 10-fold in 1 M KCl, 25 mM HEPES–KOH pH 7.0, 2 mM $CaCl_2$, 5 mM β-MeEtOH, 10% glycerol and then digested at 4°C for 2–6 h with 25 units of thrombin. The final purification step was achieved after further diluting the digestion reaction 5-fold in 20 mM Mes–NH_4 pH 6.0 and loading the protein into a SP-Sepharose column. Chromatography, protein concentration (to 0.5 mg/ml, determined with molar extinction coefficients of 18910 $M^{-1}\cdot cm^{-1}$ for ΔN37 and His_6-ΔN42, and 12950 $M^{-1}\cdot cm^{-1}$ for ΔN37C133) and purity checking (~90%) were performed as stated above.

Limited proteolysis assays

RepA–WT stock (0.5 ml) at 1 mg/ml was dialyzed against 1 M KCl, 25 mM HEPES–KOH pH 7.5, 5 mM β-MeEtOH, 10% glycerol at 4°C. The protein was then diluted with water to 0.2 mg/ml. Proteases (SIGMA) were prepared as 1 mg/ml stocks (in 25 mM KCl, 20 mM HEPES–KOH pH 7.5, 5 mM β-MeEtOH, 50% glycerol) and stored at –80°C in small volume aliquots. The search for optimal proteolysis conditions was performed at room temperature with 160 μl aliquots of RepA–WT (32 μg), by adding 1 μl of each protease (diluted to 0.3 μg/μl). 20 μl samples were taken at t = 0, 1, 5, 10, 20, 30, 60 and 90 min, reactions were stopped with 1 μl of 20 mM p-NH_2-benzamidine, 50 mM EDTA and analyzed by SDS–PAGE (20% polyacrylamide gels) with Coomassie Blue staining. For the identification of the peptides generated by each protease, after 60 min digestion, samples were dialyzed against 0.1 M NH_4-acetate pH 6.0, 1 mM EDTA, 1 mM p-NH_2-benzamidine. Then one-third of the volume was checked by gel electrophoresis, whereas the other two-thirds were analyzed by N-terminal sequencing of the mixtures (see above). In every cycle the areas under the detected amino acid peaks were used to determine the abundance of the different peptides. Peptide sequences were identified with the program DNA Strider (version 1.0 for Macintosh). Up to four peptide sequences were identified without ambiguities in each mixture. Secondary structure predictions were performed with the PHD algorithm (Rost and Sander, 1993) in its home web site (http://www.embl-heidelberg.de/predict_protein/predict_protein.html#PP1SEC).

CD spectroscopy

CD studies (Schmid, 1997) were done with a Jasco-720 spectropolarimeter, using 0.1 cm path-length quartz cuvettes with 0.2 ml of each RepA protein (at several concentrations, in the range 6–25 μM). Proteins were studied in the storage buffer in which they are most stable (see above). CD spectra were acquired at 5°C between 260 and 195 nm (0.2 nm steps, 10 nm/min scan speed and 4 s time constant). Five spectra were averaged for each sample and the spectrum for the buffer was substracted as blank. The raw ellipticity data (in millidegrees) were transformed to mean molar ellipticity per residue (θMR, units: $deg\cdot cm^2\cdot dmol^{-1}$) and plotted using Kaleidagraph (Abelbeck Software, version 3.0.2 for Macintosh). CD spectra (range 240–200 nm) were analyzed into their secondary structure components with the K2D algorithm (Andrade *et al*., 1993), run online in a mirror web site (http://akilonia.cib.csic.es/k2d.html).

Thermal denaturation experiments (Pace and Schultz, 1997) were performed on the same samples from which spectra had just been acquired. Samples were overlayed with a couple of drops of mineral oil, to avoid evaporation when heating the cuvette. The temperature of the cell holder, controlled with a water bath and verified with a thermocouple device, was increased from 5 to 90°C at a 20°C/h rate (independent scans were performed at 50°C/h), recording variations in the ellipticity at 222 nm. Data were plotted as θMR_{222} ($deg\cdot cm^2\cdot dmol^{-1}$) versus temperature (°C). Melting temperature (T_m) values, corresponding to 50% unfolding, were calculated after non-linear curve fitting of the data points had been performed.

Fluorescence spectroscopy

Steady-state fluorescence spectra (Schmid, 1997) were acquired in a Shimadzu RF-540 spectrofluorimeter. Each RepA protein was studied in 0.25 M $(NH_4)_2SO_4$ buffer or in the same solution supplemented with GuHCl to 4.5 M. Protein stocks were diluted to 8.32 μM (final volume 0.35 ml, in 0.2×1.0 cm path length fluorescence cuvettes) and left to equilibrate at 20°C. Excitation spectra were acquired first, to monitor the emission of W94 (at 320 nm) and, in independent scans, of Y residues (at 308 nm). This confirmed that the single Trp residue in RepA is selectively excited at 295 nm. Emission spectra were acquired with the fluorimeter set to high sensitivity and medium scan-speed, with excitation and emission slits of 2 and 5 nm, respectively. Intrinsic fluorescence emission spectra for W94 were recorded from 300–450 nm at a constant ordinate scale (in arbitrary fluorescence units).

Extrinsic fluorescence studies were performed with bis-ANS (Sigma, prepared as a 1.46 mM stock in methanol) (Semisotnov *et al*., 1991). bis-ANS titrations (4.16–24.96 μM) were performed with each protein (8.32 μM) at 20°C, taking care not to induce protein aggregation since this would result in artefactual extra fluorescence emission. The fluorophore was left to equilibrate with the proteins for at least 15 min before exciting the samples at 395 nm (maximum absorption for bis-ANS). Emission spectra were acquired from 400–600 nm. Maximum relative emission at 500 nm was observed with 16.64 μM bis-ANS (2-fold molar excess over protein molecules).

FPLC analysis of RepA

Gel filtration chromatography (Corbett and Roche, 1984) was performed in a Pharmacia FPLC equipment with a Superose-12 HR-10/30 column, equilibrated with several volumes of 0.5 M $(NH_4)_2SO_4$, 20 mM Mes–KOH pH 6.0, 5 mM β-MeEtOH, 0.1 mM EDTA, 5% glycerol at room temperature. The column was calibrated with standards of known masses and hydrodynamic shapes (Potschka, 1987): cytochrome *c*, carbonic anhydrase, ovoalbumin, BSA, alcohol dehydrogenase and Dextran blue (V_o = 7.19 ml). 0.2 ml samples of either purified RepA–WT or its fragments were injected into the column and run at a flow rate of 0.4 ml/min. Detection was performed monitoring A_{280}. Stokes radius (R_s, in nm) associated with each RepA peak were calculated from a K_{av} versus R_s plot for the standards, fitted with a logarithmic curve (K_{av} = $0.7618–0.7972\cdot \log R_s$; correlation coefficient = 0.9790) (Ackers, 1970).

DNA binding assays

DNA oligonucleotides used as substrates for RepA in the DNA binding assays were:

1IR (operator): 5′-GAACAAGGACAGGGCATTGACTTGTCCCTGTCCCTTAAT-3′,

1DR (iteron+DnaA box): 5′-ATACCCGGGTTTAAAGGGGACAGATTCAGGCTG**TTATCCACA**CCC-3′,

and their complementary strands, where the inverted repeat and iteron sequences are shown underlined and the DnaA box is in bold. Oligonucleotides were synthesized as mentioned above, but at 1 μmol scale. They were purified after gel-electrophoresis in 10% polyacrylamide–urea denaturing gels, followed by reverse-phase chromatography through Waters Sep-Pack C-18 cartridges. The concentration of the purified DNA strands was determined by A_{260} in 0.2 N NaOH, with

molar extinction coefficients calculated from their nucleotide composition (Brown and Brown, 1991). DNA labelling of the individual single strands (50 pmol) was performed in 5′ with 20 μCi [γ-^{32}P]ATP (Amersham, >5000 Ci/mmol) and 2 units of T4-Polynucleotide kinase (at 37°C for 1 h). Labelling reactions were heat-inactivated and purified through Sephadex-G25 (in Costar Spin-X columns). Strand annealing was performed in TE buffer (Sambrook *et al*., 1989), mixing each one of the radiolabelled oligonucleotides with a 10% molar excess of the complementary cold strand, heating at 95°C for 10 min and then leaving the mixtures to cool slowly to room temperature. Annealed DNA stocks were stored at –20°C.

EMSA (reviewed in Rhodes, 1989) were performed in 20 μl volume. Concentrated RepA–WT protein (or its fragments) were added to radiolabelled double-stranded oligonucleotides in 20 mM Mes–NH_4 pH 6.0, 5 mM β-MeEtOH, 0.1 mM EDTA, 6% glycerol, 50 mg/ml BSA. The $(NH_4)_2SO_4$ concentration was kept at 0.25 M. Reactions were assembled in ice and then transferred to room temperature for 30 min (no difference was appreciated in the stability of complexes in the incubation interval 10 min–16 h). Then 5 μl of (5×) loading buffer (0.01% bromophenol blue, 20% glycerol) were added to the samples, which were loaded into 6% polyacrylamide (19:1)–0.5×TBE gels. Electrophoresis was run at 150 V, keeping temperature at ~10°C, until the marker dye had migrated 11 cm. Gels were dried on 3MM paper (Whatman) and autoradiographs were taken (Kodak X-OMAT films).

Hydroxyl-radical footprinting (reviewed in Rhodes, 1989) was performed on the double-stranded 1DR oligonucleotide with each strand alternatively labelled. Protein samples were previously dialyzed against 1 M KCl, 20 mM Mes–KOH pH 6.0, 0.5 mM β-MeEtOH, 0.1 mM EDTA. Binding reactions were performed in 20 μl final volume. Protein and DNA stocks were diluted to 5×10^{-6} M and 5×10^{-7} M, respectively. KCl concentration was kept at 0.3 M. Fenton reaction to generate hydroxyl-radicals from H_2O_2 was performed as described (Giraldo and Díaz, 1992). After thiourea quenching, 1 volume of formamide loading buffer was added to samples. These were boiled and loaded into 17.5% polyacrylamide (19:1)–urea sequencing-type gels, together with the Maxam and Gilbert reaction for purines (Sambrook *et al*., 1989). Gels were fixed in 10% acetic acid and dried on 3MM paper. Autoradiography was performed as before and, in addition, PhosphorImager (Molecular Dynamics) screens were exposed overnight for quantitative purposes. Screens were then scanned and data processed with the ImageQuant software. To correct for heterogeneities in cleavage and loading, the areas of rectangles including relevant sequence patches were normalized by a factor calculated as the ratio between the value of the whole volumetric integration for the free DNA track and that for the sample track. Protection fractions were then obtained as the ratios between the normalized area for each protected patch and that for the same sequence in the free DNA track.

Acknowledgements

We thank Dr Maciej Zylicz for the kind gift of the *Clp* mutant strains. We deeply acknowledge the excellent technical contribution of Mrs Ana Serrano. We are also grateful to Dr Germán Rivas for many discussions on RepA hydrodynamics and for the critical reading of the manuscript. R.G. is indebted to all other members of R.D.-O. and J.M.A. labs for their support through all this project. R.G. is a recipient of a contract from the Spanish MEC (Programa de Incorporación de Doctores a Grupos de Investigación en España). This work was supported by a grant of the Spanish DGICYT (PB94-0127) to R.D-O. and J.M.A.

References

Ackers,G.K. (1970) Analytical gel chromatography of proteins. *Adv. Protein Chem.*, **24**, 343–446.

Andrade,M.A., Chacón,P., Merelo,J.J. and Morán,F. (1993) Evaluation of secondary structure of proteins from U.V. circular dichroism using an unsupervised learning neural work. *Protein Eng.*, **6**, 383–390.

Bachmann,B.J. (1996) Derivations and genotypes of some mutant derivatives of *Escherichia coli* K- 12. In Neidhardt,F.C., Curtiss,R.,III, Ingraham,J., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology.* ASM Press, Washington DC, pp. 2460–2488.

Bell,S.P. and Stillman,B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, **357**, 128–134.

Brennan,R.G. (1993) The winged-helix DNA-binding motif: Another Helix–Turn–Helix takeoff. *Cell*, **74**, 773–776.

Brennan,R.G. and Matthews,B.W. (1989) The Helix–Turn–Helix DNA binding motif. *J. Biol. Chem.*, **264**, 1903–1906.

Brown,T. and Brown,D.J.S. (1991) Modern machine-aided methods of oligodeoxyribonucleotide synthesis. In Eckstein,F. (ed.), *Oligonucleotides and Analogs: A Practical Approach.* Oxford University Press, Oxford, UK, pp. 1–24.

Charnock,C. (1997) Characterization of the cryptic plasmids of the *Pseudomonas alcaligenes* type strain. *Plasmid*, **37**, 189–198.

Chattoraj,D.K. and Schneider,T.D. (1997) Replication control of plasmid P1 and its host chromosome: The common ground. *Prog. Nucleic Acid Res. Mol. Biol.*, **57**, 145–186.

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.

Corbett,R.J.T. and Roche,R.S. (1984) Use of high speed size-exclusion chromatography for the study of protein folding and stability. *Biochemistry*, **23**, 1888–1894.

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.

Dean,F.B. and O'Donnell,M. (1996) DNA–protein interactions: Two steps to binding replication origins? *Curr. Biol.*, **6**, 931–934.

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

Dibbens,J.A., Muraiso,K. and Chattoraj,D.K. (1997) Chaperone-mediated reduction of RepA dimerization is associated with RepA conformational change. *Mol. Microbiol.*, **26**, 185–195.

Dodson,M., Roberts,J., McMacken,R. and Echols,H. (1985) Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: Complexes with λO protein and with λO, λP and *Escherichia coli* DnaB proteins. *Proc. Natl Acad. Sci. USA*, **82**, 4678–4682.

Fernández-Tresguerres,M.E., Martín,M., García de Viedma,D., Giraldo,R. and Díaz-Orejas,R. (1995) Host growth temperature and a conservative amino acid substitution in the replication protein of pPS10 plasmid influence plasmid host range. *J. Bacteriol.*, **177**, 4377–4384.

Filutowicz,M., Davis,G., Greener,A. and Helinski,D.R. (1985) Autorepressor properties of the π-initiation protein encoded by plasmid R6K. *Nucleic Acids Res.*, **13**, 103–114.

Filutowicz,M., York,D. and Levchenko,I. (1994) Cooperative binding of initiator protein to replication origin conferred by single amino acid substitution. *Nucleic Acids Res.*, **22**, 4211–4215.

García de Viedma,D., Giraldo,R., Ruíz-Echevarría,M.J., Lurz,R. and Díaz-Orejas,R. (1995a) Transcription of *repA*, the gene of the initiation protein of the *Pseudomonas* plasmid pPS10, is autoregulated by interactions of the RepA protein at a symmetrical operator. *J. Mol. Biol.*, **247**, 211–223.

García de Viedma,D., Serrano-López,A. and Díaz-Orejas,R. (1995b) Specific binding of the replication protein of plasmid pPS10 to direct and inverted repeats is mediated by an HTH motif. *Nucleic Acids Res.*, **23**, 5048–5054.

García de Viedma,D., Giraldo,R., Rivas,G., Fernández-Tresguerres,E. and Díaz-Orejas,R. (1996) A leucine zipper motif determines different functions in a DNA replication protein. *EMBO J.*, **15**, 925–934.

Giraldo,R. and Díaz,R. (1992) Differential binding of wild-type and a mutant RepA protein to *oriR* sequence suggests a model for the initiation of plasmid R1 replication. *J. Mol. Biol.*, **228**, 787–802.

Giraldo,R., Nieto,C., Fernández-Treguerres,M.E. and Díaz,R. (1989) Bacterial zipper. *Nature*, **342**, 866.

Giraldo,R., Martín,M., Fernández-Tresguerres,M.E., Nieto,C. and Díaz,R. (1992) Mutations within the minimal replicon of plasmid pPS10 increase its host range. In Hughes,P., Fanning,E. and Kohiyama,M. (eds), *DNA Replication: The Regulatory Mechanisms*. Springer Verlag, Berlin, Germany, pp. 225–237.

Gottesman,S. (1990) Minimizing proteolysis in *Escherichia coli*: Genetic solutions. *Methods Enzymol.*, **185**, 119–129.

Gottesman,S., Wickner,S. and Maurizi,M.R. (1997) Protein quality control: Triage by chaperones and proteases. *Genes Dev.*, **11**, 815–823.

Helinski,D.R., Toukdarian,A.E. and Novick,R.P. (1996) Replication control and other stable maintenance mechanisms of plasmids. In Neidhardt,F.C., Curtiss,R.,III, Ingraham,J., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and

Umbarger,H.E. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology.* ASM Press, Washington DC, pp. 2295–2324.
Horii,T., Ogawa,T. and Ogawa,H. (1980) Organization of the *recA* gene of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **77**, 313–317.
Hubbard,S.J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta*, **1382**, 191–206.
Hwang,D.S., Crooke,E. and Kornberg,A. (1990) Aggregated DnaA protein is dissociated and activated for DNA replication by phospholipase or DnaK protein. *J. Biol. Chem.*, **265**, 19244–19248.
Ingmer,H., Fong,E.L. and Cohen,S. (1995) Monomer–dimer equilibrium of the pSC101 RepA protein. *J. Mol. Biol.*, **250**, 309–314.
Ishiai,M., Wada,C., Kawasaki,Y. and Yura,T. (1994) Replication initiator RepE of mini-F plasmid: Functional differentiation between monomers (initiator) and dimers (autogenous repressor). *Proc. Natl Acad. Sci. USA*, **91**, 3839–3843.
Jacob,F., Brenner,S. and Couzin,F. (1963) On the regulation of DNA replication in bacteria. *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 329–348.
Kawasaki,Y., Wada,C. and Yura,T. (1990) Roles of *Escherichia coli* heat shock proteins DnaK, DnaJ and GrpE in mini-F plasmid replication. *Mol. Gen. Genet.*, **220**, 277–282.
Kawasaki,Y., Wada,C. and Yura,T. (1991) Mini-F plasmid mutants able to replicate in the absence of σ^{32}: Mutations in the *repE* coding region producing hyperactive initiator protein. *J. Bacteriol.*, **173**, 1064–1072.
Kelly,S.M. and Price,N.C. (1997) The application of circular dichroism to studies of protein folding and unfolding. *Biochim. Biophys. Acta*, **1338**, 161–185.
Kelman,Z. and O'Donnell,M. (1994) DNA replication: Enzymology and mechanisms. *Curr. Opin. Genet. Dev.*, **4**, 185–195.
Konieczny,I. and Helinski,D.R. (1997) The replication initiator protein of the broad-host-range plasmid RK2 is activated by the ClpX chaperone. *Proc. Natl Acad. Sci. USA*, **94**, 14378–14382.
Kornberg,A. and Baker,T. (1992) *DNA Replication.* 2nd edn. W.H.Freeman, New York, NY.
Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
Lakowicz,J.R. (1983) *Principles of Fluorescence Spectroscopy.* Plenum Press, New York, NY.
Lee,D.G. and Bell,S.P. (1997) Architecture of the yeast Origin Recognition Complex bound to origins of DNA replication. *Mol. Cell. Biol.*, **17**, 7159–7168.
Lefstin,J.A. and Yamamoto,K.R. (1998) Allosteric effects of DNA on transcriptional regulators. *Nature*, **392**, 885–888.
Levchenko,I., York,D. and Filutowicz,M. (1994) The dimerization domain of R6K plasmid replication initiator protein π revealed by analysis of a truncated protein. *Gene*, **145**, 65–68.
Levchenko,I., Inman,R. and Filutowicz,M. (1997) Replication of the R6K γ origin *in vitro*: Dependence on wt π and hyperactive πS87N protein variant. *Gene*, **193**, 97–103.
Lupas,A. (1996) Coiled coils: New structures and new functions. *Trends Biochem. Sci.*, **21**, 375–382.
Manen,D., Upegui,G.L. and Caro,L. (1992) Monomers and dimers of the RepA protein in plasmid pSC101 replication: Domains in RepA. *Proc. Natl Acad. Sci. USA*, **89**, 8923–8927.
Matsunaga,F., Ishiai,M., Kobayashi,G., Uga,H., Yura,T. and Wada,C. (1997) The central region of RepE initiator protein of mini-F plasmid plays a crucial role in dimerization required for negative replication control. *J. Mol. Biol.*, **274**, 27–38.
Messer,W. and Weigel,C. (1996) Initiation of chromosome replication. In Neidhardt,F.C., Curtiss,R.,III, Ingraham,J., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology.* ASM Press, Washington DC, pp. 1579–1601.
Miron,A., Mukherjee,S. and Bastia,D. (1992) Activation of distant replication origins *in vivo* by DNA looping as revealed by a novel mutant form of an initiator protein defective in cooperativity at a distance. *EMBO J.*, **11**, 1205–1216.
Miron,A., Patel,I. and Bastia,D. (1994) Multiple pathways of copy control of γ replicon of R6K: Mechanisms both dependent on and independent of cooperativity of interaction of π protein with DNA affect the copy number. *Proc. Natl Acad. Sci. USA*, **91**, 6438–6442.
Mukhopadhyay,G., Sozhamannan,S. and Chattoraj,D.K. (1994) Relaxation of replication control in chaperone-independent initiator mutants of plasmid P1. *EMBO J.*, **13**, 2089–2096.
Newlon,C.S. (1997) Putting it all together: Building a prereplicative complex. *Cell*, **91**, 717–720.
Nieto,C., Giraldo,R., Fernández-Tresguerres,E. and Díaz,R. (1992) Genetic and functional analysis of the basic replicon of pPS10, a plasmid specific for *Pseudomonas* isolated from *Pseudomonas syringae patovar savastanoi*. *J. Mol. Biol.*, **223**, 415–426.
Nozaki,Y., Schechter,N.M., Reynolds,J.A. and Tanford,C. (1976). Use of gel chromatography for the determination of the Stokes radii of proteins in the presence and absence of detergents: A reexamination. *Biochemistry*, **15**, 3884–3889.
Olins,P.O., Devine,C.S., Rangwala,S.H. and Kavka,K.S. (1988) The T7 phage gene 10 leader *RNA*, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene*, **73**, 227–235.
Pace,C.N. and Schmid,F.X. (1997) How to determine the molar absorbance coefficient of a protein. In Creighton,T.E. (ed.), *Protein Structure: A Practical Approach*. Oxford University Press, Oxford, UK, pp. 253–259.
Pace,C.N. and Scholtz,J.M. (1997) Measuring the conformational stability of a protein. In Creighton,T.E. (ed.), *Protein Structure: A Practical Approach*. Oxford University Press, Oxford, UK, pp. 299–321.
Pak,M. and Wickner,S. (1997) Mechanism of protein remodeling by ClpA chaperone. *Proc. Natl Acad. Sci. USA*, **94**, 4901–4906.
Potschka,M. (1987) Universal calibration of gel permeation chromatography and determination of molecular shape in solution. *Anal. Biochem.*, **162**, 47–64.
Rhodes,D. (1989) Analysis of sequence-specific DNA-binding proteins. In Creighton,T.E. (ed.), *Protein Function: A Practical Approach*. Oxford University Press, Oxford, UK, pp. 177–198.
Rokeach,L.A., Søgaard-Andersen,L. and Molin,S. (1985) Two functions of the E protein are key elements in the plasmid F replication control system. *J. Bacteriol.*, **164**, 1262–1270.
Rost,B. and Sander,C. (1993) Prediction of protein structure at better than 70% accuracy. *J. Mol. Biol.*, **232**, 584–599.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schmid,F.X. (1997) Optical spectroscopy to characterize protein conformation and conformational changes. In Creighton,T.E. (ed.), *Protein Structure: A Practical Approach*. Oxford University Press, Oxford, UK, pp. 261–297.
Semisotnov,G.V., Rodionova,N.A., Razgulyaev,O.I., Uversky,V.N., Gripas,A.F. and Gilmanshin,R.I. (1991) Study of the molten globule intermediate state in protein folding by a hydrophobic fluorescence probe. *Biopolymers*, **31**, 119–128.
Soisson,S.M., MacDougall-Shackleton,B., Schleif,R. and Wolberger,C. (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science*, **276**, 421–425.
Sozhamannan,S. and Chattoraj,D.K. (1993) Heat shock proteins DnaJ, DnaK and GrpE stimulate P1 plasmid replication by promoting initiator binding to the origin. *J. Bacteriol.*, **175**, 3546–3555.
Vocke,C. and Bastia,D. (1983) The replication initiator protein of plasmid pSC101 is a transcriptional repressor of its own cistron. *Proc. Natl Acad. Sci. USA*, **82**, 2252–2256.
Wemmer,D.E. and Dervan,P.B. (1997) Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol.*, **7**, 355–361.
Wickner,S., Skowyra,D., Hoskins,J. and McKenney,K. (1992) DnaJ, DnaK and GrpE heat shock proteins are required in *oriP1* DNA replication solely at the RepA monomerization step. *Proc. Natl Acad. Sci. USA*, **89**, 10345–10349.
Wickner,S., Gottesman,S., Skowyra,DF., Hoskins,J., McKenney,K. and Maurizi,M.R. (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl Acad. Sci. USA*, **91**, 12218–12222.
Wu,J., Sektas,M., Chen,D. and Filutowicz,M. (1997) Two forms of replication initiator protein: positive and negative controls. *Proc. Natl Acad. Sci. USA*, **94**, 13967–13972.
Xia,G., Manen,D., Yu,Y. and Caro,L. (1993) *In vivo* and *in vitro* studies of a copy number mutation of the RepA replication protein of plasmid pSC101. *J. Bacteriol.*, **175**, 4165–4175.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp19 and pUC19 vectors. *Gene*, **33**, 103–119.
York,D. and Filutowicz,M. (1993) Autoregulation-deficient mutant of the plasmid R6K-encoded π protein distinguishes between palindromic and nonpalindromic binding sites. *J. Biol. Chem.*, **268**, 21854–21861.

Received March 10, 1998; revised May 26, 1998;
accepted June 12, 1998