The EMBO Journal vol.12 no.1 pp.45-52, 1993

Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindromic DNA element

Pei-Zhi Wang, Steven J.Projan, Vitalia Henriquez and Richard P.Novick

Departments of Plasmid Biology and Applied Genetics, The Public Health Research Institute of the City of New York, Inc., 455 First Avenue, New York, NY 10016, USA

Communicated by C.Anagnostopoulos

The leading strand replication origin of pT181 plasmids consists of two adjacent inverted repeat elements (IR-II and IR-III), which are involved in origin recognition by the initiator (Rep) protein. The conserved core element, IR-II, which contains the initiation nick site, is induced by Rep to form a cruciform structure, probably the primary substrate for the initiation of rolling circle replication. The divergent repeat, IR-III, constitutes the determinant of origin recognition specificity. We show here that the distal arm of IR-III is not required for sequence-specific recognition, whereas the proximal arm and central region are required. Since the initiator is dimeric, we presume that it binds symmetrically to IR-III. A unique type of DNA-protein interaction is proposed, in which the lack of sequence requirement for the distal arm is a consequence of binding to the adjacent IR-II, which thereby polarizes the stringency of binding to the two arms of IR-III. In addition, genetic evidence indicates that both the spacing and the phasing of IR-II to IR-III are crucial for function and that the central segment of IR-III may serve to position the two flanking half-sites for optimal interaction of Rep with IR-III.

Key words: DNA-binding/DNA replication/plasmid/recognition/*Staphylococcus aureus*

Introduction

Two classes of prokaryotic replicons use rolling circle (RC) relication, namely single-stranded coliphages (reviewed by Baas, 1985; Baas and Jansz, 1988) and a collection of small high copy number plasmids from both Gram-positive (Figure 1) (reviewed by Gruss and Ehrlich, 1989; Novick, 1989) and Gram-negative bacteria (Gielow *et al.*, 1991; Kleanthous *et al.*, 1991; Yasukawa *et al.*, 1991), which replicate via a single-stranded DNA (ssDNA) intermediate. Of these replicons, the single-stranded isometric and filamentous phages, represented by the ϕX group and Ff group, respectively, and the small, multicopy antibiotic resistance staphylococcal plasmids, represented by the pT181 family (Figure 1), have been studied in most depth with respect to the mechanism of their viral or leading (+) strand synthesis. Each of these encodes a unique initiator protein, which is a site-specific endonuclease that recognizes the cognate (+) strand origin and introduces a site-specific nick that provides the 3′ OH primer for replication (Figure 2). Some initiators become covalently attached via a phosphotyrosine bond at the 5′ nick terminus, others evidently do not become stably attached (Baas and Jansz, 1988; Thomas *et al.*, 1990). Initiators of the pT181 plasmid family have site-specific topoisomerase I-like relaxing activity as well as endonuclease activity (Koepsel and Khan, 1987). The initiator protein fulfills a key role in leading strand synthesis, being involved in initiation and termination. The initiator proteins have somewhat different properties, resulting in variations in the mechanism of leading strand synthesis among these replicons. All of these replicons require specific host functions to accomplish leading strand synthesis. These have been characterized most fully for the single-stranded phages, for which at least three host-encoded proteins are involved, namely helicase, polymerase III holoenzyme and single-strand binding protein (for review, see Baas, 1985). Comparison of the properties of the various rolling circle initiator proteins is shown in Table I.

The six members of the pT181 plasmid family (Projan and Novick, 1988; Balson and Shaw, 1990) have a common functional organization and a high degree of sequence identity, but differ in the critical matching sequence elements in the replication initiator protein, Rep ('Rep' in this paper refers exclusively to initiator proteins and not to helicase, which is often referred to as 'Rep' in other papers) and (+)

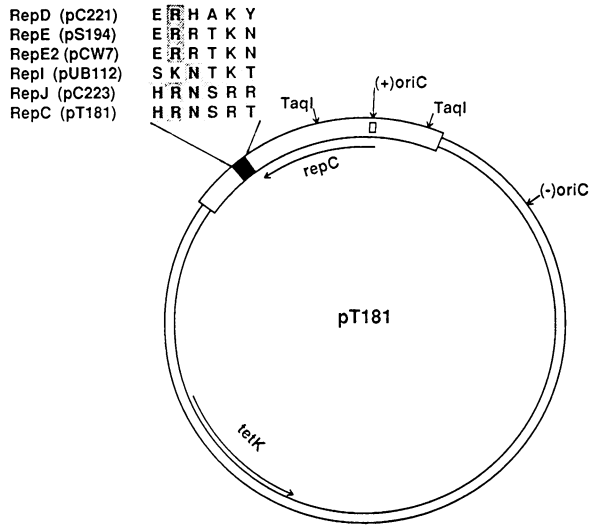

Fig. 1. Schematic diagram of pT181. The loci are described in detail in the text. *repC* is a gene encoding RepC, a *trans*-active, replicon specific, replication initiator protein. The six amino acid residue set containing the predicted determinant of origin specificity (Wang *et al.*, 1992) is indicated by black box and shown in the standard 1-letter code for the six Rep proteins from pT181 family. Abbreviations for the amino acid residues are A, Ala; E, Glu; H, His; K, Lys; N, Asn; R, Arg; S, Ser; and T, Thr. The amino acid residues in each protein that are identical with the corresponding residues in RepC are shaded. (+) *oriC* (internal to the *repC* coding sequence) and (−) *oriC* are the respective origins of leading and lagging strand replications. The *TaqI* sites used for (+) *oriC* cloning are indicated with arrows. *tetK* is a gene encoding tetracycline resistance.

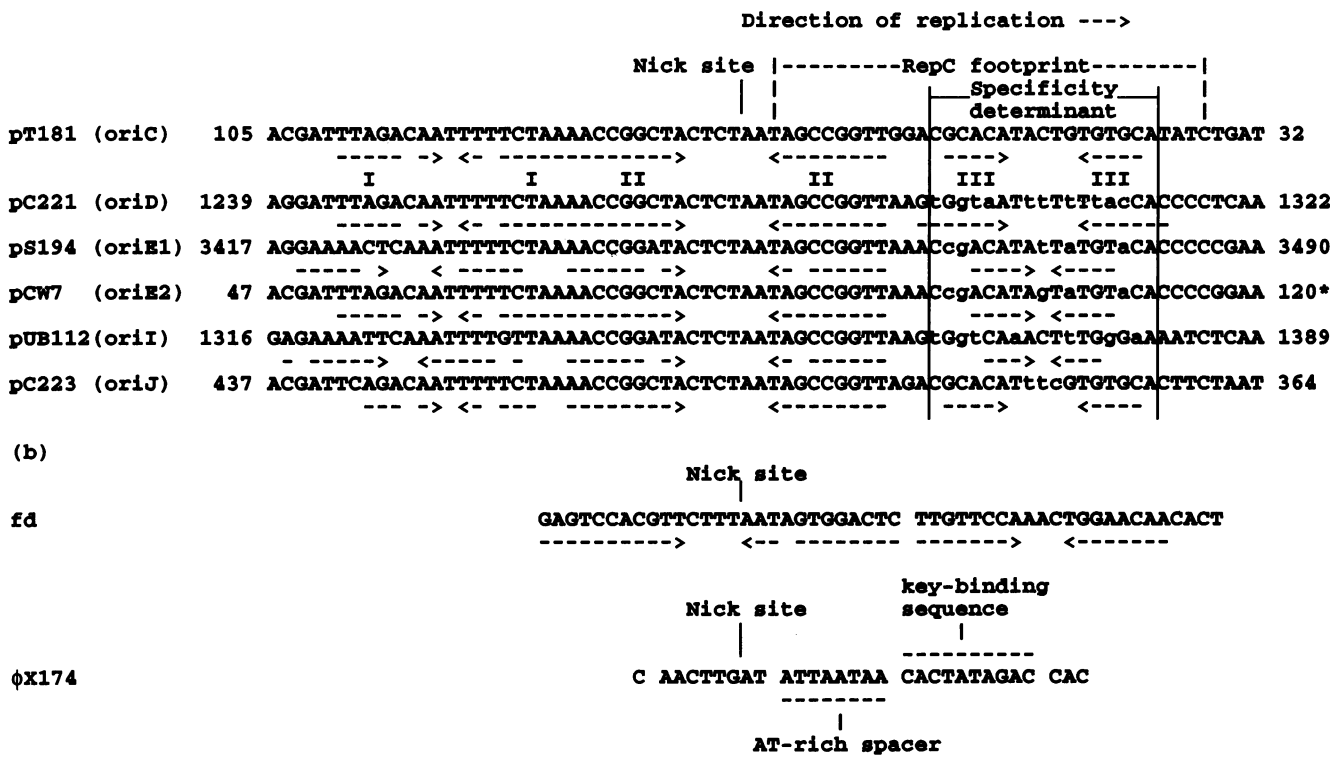

Fig. 2. The leading strand (+) origins of plasmids of the pT181 family and bacteriophages, fd and ϕX174. (**a**) Comparison of leading strand (+) origins of plasmids of the pT181 family. The direction of replication is indicated. The (+) strand nick site and RepC footprint are labeled as well (Koepsel *et al.*, 1986b). Inverted repeat (IR) sequences I, II and III are indicated by arrows (→ ←). The 17 nucleotides in IR-III region containing the specificity determinant are indicated by vertical lines. The bases in each of this region that are different from the homologously positioned bases in *oriC* are represented by lower case characters. Plasmid nucleotide coordinates (numbered at both ends of each DNA sequence) are from the published sequences (Projan and Novick, 1988; Balson and Shaw *et al.*, 1990). (**b**) The viral strand origins of fd (Dotto *et al.*, 1982) and ϕX174 (Baas *et al.*, 1981). The viral strand nick site is labeled and inverted repeat sequences are indicated by arrows (→ ←).

Table I. Comparison of the properties of initiator proteins

Description	RepC (pT181)	Gene A protein (ϕX174)	Gene II protein (fd)
1. Requirement of supercoiled RF DNA for cleavage	yes	yes	yes
2. Secondary structure at cleavage site	high	low	high
3. Nicking–closing reaction on RF DNA	yes	no	yes
4. Cleavage of single leading (+) strands	yes	yes	no
5. Complex formation with phage or plasmid DNA	attaches covalently to 5′-end of DNA after cleavage	attaches covalently to 5′-end of DNA after cleavage	weak
6. Energy for single-stranded DNA circularization	nicking at start	nicking at start	cleavage after
7. Linkage for energy transfer of the enzme to	5′-end	5′-end	3′-end
8. Energy conservation during replication cycle	yes	yes	no
9. Mode of action in RF DNA replication	distributive	processive	distributive
10. Complementation *in vivo*	*trans*	*cis*	*trans*
11. Rolling circle structures visualized by electron microscopy	not detected	looped	extended

ori, which are responsible for origin recognition (Figures 1 and 2). These six plasmids belong to five different incompatibility groups (Table II) and for the five compatible plasmids, pT181, pC221, pUB110, pC223 and pS194, each Rep protein recognizes only the (+) origin of the plasmid encoding it (Projan and Novick, 1988). Two of the plasmids, pS194 and pCW7, are weakly incompatible because of cross-reactivity between their Rep proteins and origins (Wang *et al.*, 1990).

The Rep proteins in the pT181 family share 75–85% amino acid sequence identity in pair-wise comparisons and are assumed to have the same structure. However, they show a high degree of divergence in two areas. One of these corresponds to the leading strand (2) replication origin, which is located in the N-terminal region of the Rep coding sequence; the other corresponds to the specificity determinant of the protein, which is located C-terminally (Projan and Novick, 1988; Wang *et al.*, 1992). Between these two divergent regions is a highly conserved domain containing the reactive tyrosine that becomes covalently attached to the

Table II. Plasmids

Plasmid	Phenotype[a]	Description	Reference
pT181	Tc^r	Inc3 *oriC* RepC	(Iordanescu, 1976a)
pC221	Cm^r	Inc4 *oriD* RepD	(Khan and Novick, 1983)
pS194	Sm^r	Inc5A *oriE1* RepE1	(Novick and Bouanchaud, 1971)
pCW7	Cm^r	Inc5B *oriE2* RepE2	(Balson and Shaw, 1990)
pUB112	Cm^r	Inc9 *oriI* RepI	(Iordanescu, 1976b)
pC223	Cm^r	Inc10 *oriJ* RepJ	(Wilson and Baldwin, 1978)
pSA7542	Tc^r	Cointegrate, pT181(cop620)::pE5, with *oriC* inactivated (RepC donor)	(Iordanescu, 1989)
pSA7461	Tc^r	Similar to pSA7542 except *repD* sequences substituted for *repC* (RepD donor)	(Iordanescu, 1989)
pRN5101	Em^r Tsr	Tsr mutant of pE194	(Novick *et al.*, 1984)
pRN6397	Em^r Tsr	pRN5101::pT181 *TaqI-C* (pT181 origin region cloned to pRN5101)	(Carleton *et al.*, 1984)
pRN6385	Em^r Tsr	pRN5101::pC221 *TaqI-F* (pC221 origin region cloned to pRN5101)	(Projan *et al.*, 1985)
pRN6428	Em^r Tsr	pRN5101::pS194 *TaqI-E* (pS194 origin region cloned to pRN5101)	(Projan and Novick, 1988)
pRN6647	Em^r Tsr	pRN5101::pC223 *TaqI-A* (pC223 orign region cloned to pRN5101)	(Projan and Novick, 1988)
pRN6648	Em^r Tsr	pRN5101::pUB112 *TaqI−D* pUB112 origin region cloned to pRN5101)	(Projan and Novick, 1988)
pWN8900	Ap^r	pBluescript::pT181 *ClaI−HpaI* (3677-811) (An *ori-C*-containing fragment cloned to pBluescript)	(Wang *et al.*, 1992)

[a]Abbreviations are: Tc^r, tetracycline resistance; Cm^r, chloramphenicol resistance; Sm^r, streptomycin resistance; Em^r, erythromycin resistance; Ap^r, ampicillin resistance; Tsr, thermosensitive replication.

5′ nick terminus during initiation (Thomas *et al.*, 1990). In addition, as these Rep proteins act as dimers (Thomas *et al.*, 1990), there would have to be a dimerization domain.

We have previously shown that the specificity of Rep−*ori* recognition is determined by a six amino acid residue element (aa 265−270) in the C-terminal region of the Rep protein (Figure 1). These six and flanking residues cannot form an amphipathic alpha helix or conform to the classical helix−turn−helix or other known DNA binding motifs. A novel type of interaction is suggested in which the binding domain has evolved to interact with a specific DNA structure characteristic of the cognate origin (Wang *et al.*, 1992).

There are three pairs of inverted repeats (IR-I, II and III) in the region containing the pT181 (+) origin (Gennaro *et al.*, 1989) (Figure 2). The central pair, IR-II, which is conserved among the six plasmids, contains the initiation nick site and forms a cruciform *in vivo* as well as *in vitro*. RepC binding enhances the extrusion of this cruciform (Noirot *et al.*, 1990). The adjacent divergent IR-III element has been predicted to contain the site of recognition specificity. The role of IR-I is unknown. Biochemical studies (Thomas *et al.*, 1990) indicated that the IR-II region is sufficient for the nicking-closing (topoisomerase I-like) activity associated with the Rep proteins and this, of course, is not plasmid-specific. The biologically functional initiation mechanism, however, requires IR-III as well as IR-II (Gennaro *et al.*, 1989). Thus the conserved IR-II core element distinguishes the origin regions from nonspecific DNA and the divergent IR-III region distinguishes the origin region of one member of the pT181 family from that of others. This type of two-tiered recognition has also been observed with other prokaryotic replication origins (reviewed by Bramhill and Kornberg, 1988) and with homeodomain recognition sites in eukaryotes (Hanes and Brent, 1991).

A number of DNA-binding proteins act as a homodimer and recognize symmetrical (or palindromic) DNA sites. In the case of repressor-operator interaction, if the binding *per se* represents the totality of the process, the DNA site could both be structurally and functionally symmetric. However, if functions other than DNA-binding are assigned to different parts of the protein, which act on a second DNA site directly or indirectly via other DNA-binding proteins, then the first DNA site might be functionally polarized by the second, namely functionally asymmetric although it is structurally symmetric. This question has not been raised and answered previously.

In this study, we have used site-directed mutagenesis to demonstrate that IR-III contains the determinant of origin recognition specificity for the pT181 plasmid family. Additionally, we have observed that the proximal arm of IR-III, adjacent to IR-II, is much more important than the distal arm, suggesting that attachment to IR-II may polarize the recognition of IR-III.

Results

Identification of the origin specificity determinant

To assay the relative functions of Rep proteins and origins, we utilized a two-plasmid *in vivo* system composed of a thermosensitive replication (Tsr) vector, pRN5101, into which is cloned the test origin, and a second plasmid encoding the appropriate Rep protein (Iordanescu, 1989; Wang *et al.*, 1992). The copy number of the origin-containing plasmid at the restrictive temperature is taken to represent the activity of the Rep−*ori* combination being tested. It should be indicated that the (+) origin is located within the N-terminal region of the *rep* coding sequence in all members of the pT181 family (Figure 1). As the (+) origin competes effectively for the cognate Rep protein, its presence could be a complicating factor in cross-complementation experiments and we have therefore used a derivative of *repC* in which the origin has been inactivated but does not change the amino acid sequence of the Rep protein (Iordanescu, 1989).

Inspection of the previously localized pT181 (+) origin (Gennaro *et al.*, 1989) in comparison with origin sequences for the other members of the pT181 family (Projan and Novick, 1988) suggested that the specificity determinant would correspond to a divergent region 3′ to the nick site and starting at the base of the conserved IR-II stem (Figure 2). This region contains the symmetry element IR-III that is always present and included in the RepC footprint, but that varies in configuration among the six origins (Figure 2). We first substituted the entire 17 nt IR-III element of

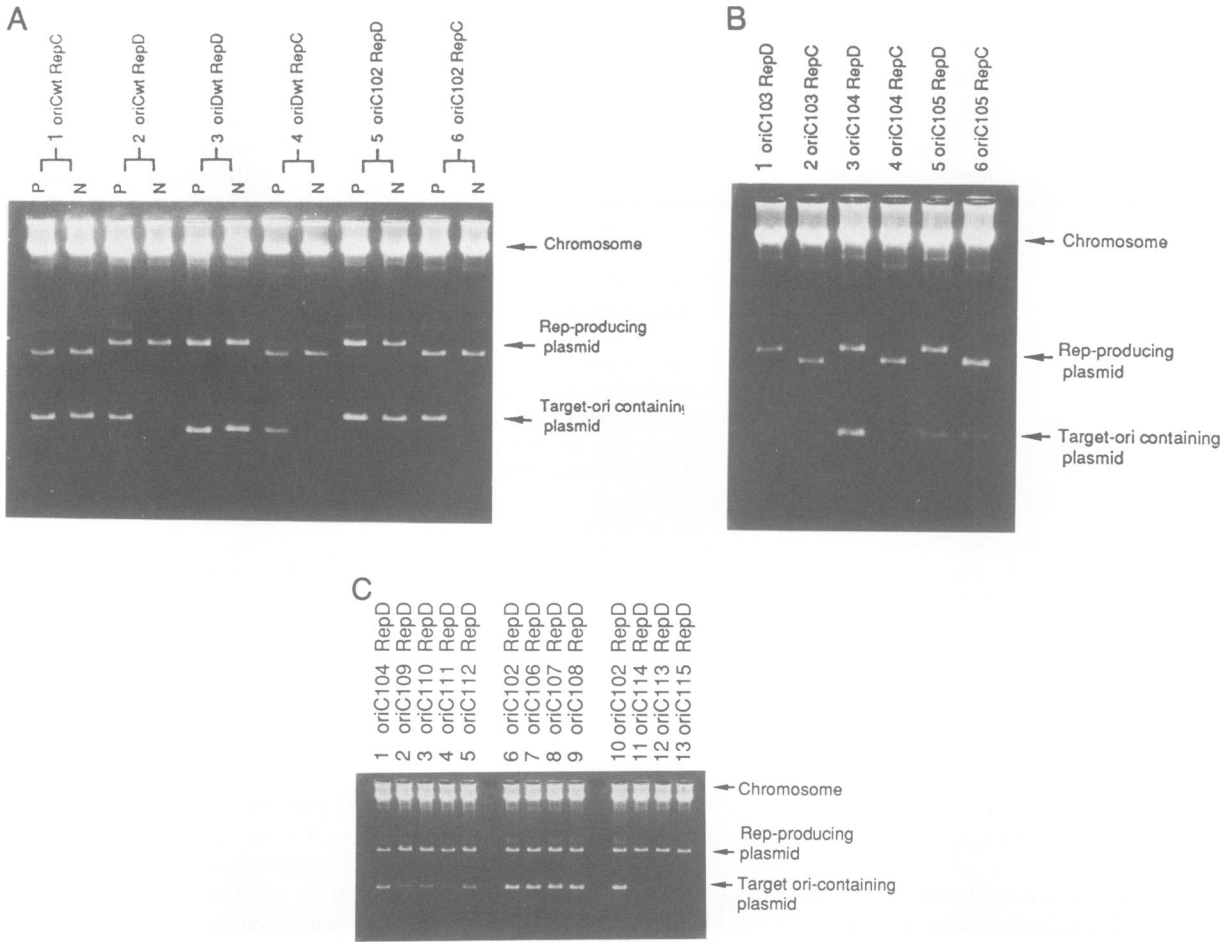

Fig. 3. Demonstration of the Rep protein specificity and activity for the corresponding wild type or mutant origins. Cultures were grown at permissive temperature (P) 32°C or non-permissive temperature (N) 43°C as indicated, respectively, and the cultures used to prepare sheared whole cell minilysates (Projan *et al.*, 1983) that were separated by agarose gel electrophoresis (1% agarose in TBE, 30 V for 16 h), stained and photographed. Chromosome, Rep protein-producing plasmid and target origin-containing plasmid are indicated, respectively. The plasmid vectors corresponding to the Rep proteins, and the wild-type and mutant origins are listed in Table III. **A.** Demonstration of the Rep protein specificity for the *oriC−oriD* hybrid, *oriC102* (see Table III, A). **B.** Demonstration of the Rep protein specificity for the origin mutations with recombined half-sites and central part (*oriC103−105*) (see Table III, B). **C.** Demonstration of RepD activity for the point mutations of *oriC104* and *oriC102*, and deletion and insertion mutations of *oriC102* (see Table III, C, D and E).

oriC with that of *oriD*, giving rise to *oriC102* as shown in Table III. This substitution has completely switched the origin specificity as the hybrid origin is recognized by RepD but not by RepC (Figure 3A). The overall activity of the hybrid origin, however, is ~2/3 that of the wild type *oriD* (Table III). Possibly the non-conserved nucleotides flanking IR-III would also be required for optimum activity. Further substitutions were made using *oriC102* and re-introducing one or more of the pT181-specific nucleotides in place of the corresponding pC221-specific nucleotides in this region. To begin with, *oriC104*, a derivative in which the 6 nt right (R) arm of IR-III has been restored to the pT181 sequence, has essentially the same properties as *oriC102*. Thus the specific sequence of the R arm (distal to IR-II), is not required for recognition of the origin, narrowing the specificity determinant to the nucleotides numbered 1−11 in Table III. Restoration of three divergent nucleotides (8, 9 and 11) in the central five nucleotide region eliminated nearly all activity with RepD and did not restore any activity with RepC (*oriC103*); substitution of the left (L) arm (nt 1−6) (*oriC105*) proximal to IR-II, eliminated specific recognition and most of the activity with either protein (Figure 3B, Table III). These results suggest that IR-III, the specificity determinant, does not function as a symmetry element.

Effect of base substitutions in IR-III region

Substitution of the individual nucleotides within the 11 nt segment (of which seven are divergent) confirmed the importance of the L arm for activity and showed that the central nucleotides were not individually very important. Two sets of base replacements were constructed (Table III) and tested with RepC and RepD as before (Figure 3C). Several conclusions were drawn from these experiments. (i) Single substitutions do not change the origin specificity, since all of these mutant origins are recognized by RepD but not by RepC. (ii) Single substitutions in the L arm of IR-III (*oriC109 to -112*) reduced the activity of the origin to <40%. This suggests that the base-specific contacts at these positions are critical for activity. In addition, our results also strongly suggest that the interaction between the recognition domain of the Rep protein and the origin DNA

Table III. Effect of site-directed mutations of oriC

Plasmid	oriC allele	[a]DNA sequence of IR-III region	[b]Diagram of IR-II and IR-III	Replication activity (copies/cell)	
				RepD	RepC
A. IR-III replacement in oriC with that corresponding to oriD					
		1 2 3 4 5 6 7 8 9 10 11	IR-II IR-III		
pRN6397	oriC wt	cgcaca-tactg-tgtgca		< 1	90
pRN6385	oriD wt	TGGTAA - TTTTT - TTACCA		100	< 1
pWN9002	oriC102	TGGTAA - TTTTT - TTACCA		66	< 1
B. Half sites or center part of IR-III replacements in oriC102 with those corresponding to oriC					
pWN9006	oriC103	TGGTAA - tactg - TTACCA		7	< 1
PWN9007	oriC104	TGGTAA - TTTTT - tgtgca		56	< 1
pWN9008	oriC105	cgcaca - TTTTT - TTACCA		23	15
C. Single replacements in oriC104 with residues corresponding to oriC					
pWN9007	oriC104	TGGTAA - TTTTT - tgtgca		56	< 1
pWN9114	oriC109	TGGT cA - TTTTT - tgtgca		11	< 1
pWN9115	oriC110	TGG aAA - TTTTT - tgtgca		17	< 1
pWN9116	oriC111	TG c TAA - TTTTT - tgtgca		8	< 1
pWN9117	oriC112	c GGTAA - TTTTT - tgtgca		24	< 1
D. Single replacements in oriC102 with residues corresponding to oriC					
pWN9002	oriC102	TGGTAA - TTTTT - TTACCA		66	< 1
PWN9012	oriC106	TGGTAA - TaTTT - TTACCA		66	< 1
PWN9013	oriC107	TGGTAA - TTcTT - TTACCA		53	< 1
pWN9014	oriC108	TGGTAA - TTTTg - TTACCA		45	< 1
E. Deletion and insertion mutations between IR-II and IR-III in oriC102					
pWN9002	oriC102	tt-$\overline{gga}$ - TGGTAA - TTTTT - TT		66	< 1
pWN9119	oriC114	tt-$\overline{a}$ - TGGTAA - TTTTT - TT		< 1	< 1
pWN9118	oriC113	tt-$\overline{gg}$ttacc$\overline{a}$ - TGGTAA - T		< 1	< 1
pWN9120	oriC115	tt-$\overline{gg}$ttacgttacc$\overline{a}$ - TGG		< 1	< 1

[a]Lower-case letters represent nucleotides corresponding to oriC, and capital letters represent those corresponding to oriD. Nucleotides between IR-II and IR-III are overlined, and inserted heterologous nucleotides are underlined.

[b]Thin lines (——) represent the nucleotide sequences corresponding to the wild-type oriC, and thick lines (▬) represent those corresponding to the wild-type oriD. Small bars (|) and white boxes (▭) indicate point and insertion mutations, respectively.

occurs in the major groove, because the Rep proteins distinguish C:G and G:C at base pair 3, and A:T and T:A at base pair 4. These base pairs display distinct functional groups in the major groove, but almost identical groups in the minor groove. (iii) Single substitutions in the central part of IR-III (*oriC106* to *-108*) have little or no effect on origin activity and no effect on specificity whereas the triple substitution in *oriC103* essentially inactivated the origin. Therefore, base-specific contacts in this region are probably not critical for recognition. Instead, its function may be to determine overall local structure which could be important for the specific Rep–*ori* interactions.

Fixed relative positions of the two Rep binding sites are required for an active origin

These Rep proteins are bivalent DNA-binding proteins, which have a C-terminal binding domain that contains the specificity determinant and a central binding domain that includes the active site tyrosine. The corresponding binding sites are represented by the IR-II and IR-III regions, respectively, and the question arises of what is the requisite structural relation between these two regions. For some bivalent DNA-binding proteins, such as lambda integrase, two independent DNA binding domains are used to bind two different DNA sequences. The distance between these

two DNA sequences can be changed without much effect on function so long as the correct helical phase of the DNA is preserved (Vargas *et al.*, 1989). The published RepC footprint (Koepsel *et al.*, 1986), obtained with linear DNA, showed that several bases located between the two sites are not bound by RepC. Therefore, the possibility was tested that a change in the relative positions of the two binding sites would not inactivate the origin if the correct helical phase of the DNA was maintained. A deletion of two base pairs (*oriC114*) and insertions of five (*oriC113*) and 10 (*oriC115*) base pairs in the intervening region were constructed to test this prediction. A sequence of 5′-ttacc-3′ was selected for insertion mutation because it was very different from the original sequence (5′-tggta-3′) in this position and introduced a *Bst*EII restriction site, which was used for further insertion of 5 bp to create a 10 bp insertion mutation. As shown in Table III, all three of these mutations inactivated the origin completely. This indicates that helical phase *per se* is not sufficient for functionality in this system; rather, the two sites must occupy fixed relative positions in order for the Rep protein to bind productively. In other words, we conclude that the protein must interact specifically with the IR-II and IR-III sequences and that it cannot do so if these are not in their native locations; the insertion of any foreign sequences will interfere with normal function regardless of what sequence has been inserted.

Discussion

In this report, we have localized the pT181 (+) origin specificity determinant to a 17 nucleotide symmetry element, IR-III, within the origin, which is divergent among the six closely related plasmids comprising the pT181 family (Figure 2). Although the entire IR-III element is footprinted by RepC (Koepsel *et al.*, 1986), only the first 11 nucleotides, comprising the L arm and center of the dyad, are required for Rep protein recognition; replacement of the six distal bases (R arm) by a non-complementary sequence reduced the activity of the origin somewhat (Gennaro *et al.*, 1989), but had no effect on the specificity (Table III, Figure 3A). The significance of this is considered below. The recognition element, which is presumed to represent the site of tight sequence-specific binding, lies 3′ to the nick site and is therefore present on the 5′ end of the displaced leading strand, to which RepC presumably remains attached during replication. These results lend further support to the following model for Rep protein binding and replication initiation by the plasmids of this family. Despite the presence of two adjacent symmetry elements, the strong and highly conserved IR-II element containing the nick site at its tip and the weaker, divergent IR-III, binding of the dimeric Rep proteins is asymmetric. The RepC footprint covers only the 3′ half of IR-II and its recognition is confined to the left arm and central region of IR-III. The protein binds to supercoiled DNA, with a dissociation constant in the nanomolar range (W.Shaw, personal communication), as well as to linear double and single stranded DNAs (Koepsel and Khan, 1987). Binding to supercoiled DNA promotes (or stabilizes) the formation of the IR-II cruciform which is believed to be the primary substrate for initiation (Noirot *et al.*, 1990). The specific six amino acid residue recognition site on the protein is located ~80 residues C-terminal to the active site tyrosine, so it is likely that the protein has two separate origin binding domains; with the cruciform extruded, the DNA conformation could accommodate the protein in the 'corner' formed by the cruciform and the flanking IR-III region, juxtaposing the two binding sites (see Figure 4), and the nicking reaction is believed to occur in that conformation. This configuration may explain the sensitivity of the pT181 (+) origin to changes in the distance between IR-II and IR-III. The deletion of 2 nt or the insertion of five abolished origin activity and this was not a simple consequence of changing of the helical phase since the insertion of 10 nt was equally lethal. We suggest that RepC must find recognizable sequences adjacent to IR-II; it cannot adjust to the insertion of new material by changing its conformation to accommodate the new spacing of the elements. The asymmetry of this binding is underscored by the fact that only one of the RepC subunits does the nicking and covalent attachment; the other may be carried along passively and may participate in termination (A.Rasooly and R.Novick, in preparation).

We note that two half-sites of IR-III are located in phase on adjacent helical turns so that symmetrical binding of the two subunits of the dimer would place one subunit about one turn further away from the nick site than the other subunit. It is difficult to envision a three-dimensional DNA–protein structure in which two subunits of a symmetrical dimer, each binding to one of the IR-III half-sites, would each bind to one stem of the IR-II cruciform. We suggest tentatively that only one of the subunits binds to the cruciform, the other facing symmetrically outwards as diagrammed schematically in Figure 4. If the second subunit is held in place through dimerization with the first subunit, it is not unreasonable that the sequence requirements for the right half-site are much weaker than for the left half-site. The binding of a protein dimer in a half-specific/half-non-specific model has been seen in the interaction of the glucocorticoid receptor (GR) with DNA (glucocorticoid response elements, GREs) (Luisi *et al.*, 1991). Upon binding DNA, the GR domains dimerize with the subunits spaced so as to interact with the hexameric pseudo-symmetric half-

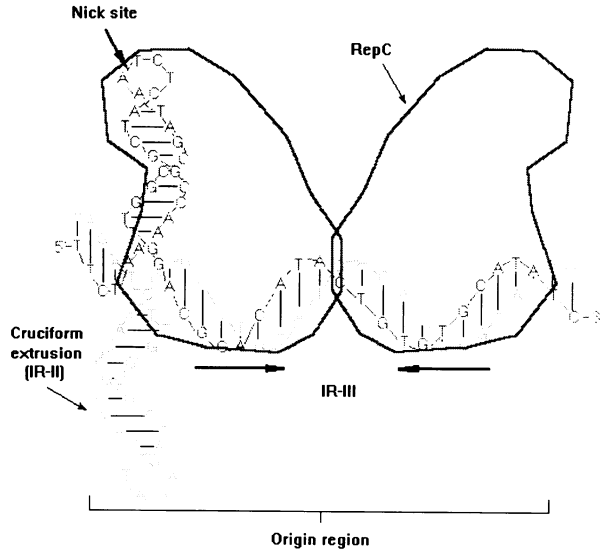

Fig. 4. Model for RepC–*oriC* interaction. RepCs are shown binding symmetrically to IR-III, which is indicated by arrows. One subunit of RepC is shown binding asymmetrically to the extruded cruciform containing the nick site as indicated. Dark letters represent nucleotides corresponding to the leading strand of pT181 and light letters represent nucleotides corresponding to the lagging strand.

sites of the response elements separated by three base pair spacers (GRE_{S3}). The insertion of an extra base pair between the hexameric half-sites (GRE_{S4}), does not affect either the binding or the crystal structure of the protein–DNA complex. The crystal structures of the GR DNA-binding domain complexed with GRE_{S4} ($GR-GRE_{S4}$) suggests that the protein retains its symmetrical dimeric structure for the $GR-GRE_{S4}$ as well as for the $GR-GRE_{S3}$ complexes; that is, it must bind to a different hexamer at one of the two half-sites. Indeed, the naturally occurring GREs that display only one clear half-site are bound in this mode (Beato, 1989).

The role of the central five bp of the IR-III element might then be to determine the spatial relation between the two half-sites. It is known for example that exchanging the two central A-T base pairs for G-C in the 434 operator causes an increase in the twist of the primary helix, thus shifting the relative positions of the flanking half-sites and reducing the affinity of the repressor by some 64-fold (Koudelka and Carlson, 1992) even though the repressor does not contact these bases. The central part of IR-III is predicted to show considerable variation in structure from plasmid to plasmid and this variation could contribute significantly to the spatial relation of the two half-sites and thus to the specificity of Rep–*ori* binding. Koepsel and Khan (1986) have reported RepC-induced DNA bending within the pT181 origin. The center of the induced bend corresponds to the region between IR-II and IR-III. However, as they used a linear DNA fragment in which the cruciform is unlikely to be present, the significance of this observation for our model of RepC binding is uncertain.

A number of other sequence-specific DNA–protein interactions share certain features of this system, but nothing quite matching it has been described. The most closely related system is that of the filamentous single stranded coliphages such as fd, whose (+) strand origin contains two IR elements, of which one contains the nick site and corresponds to the IR-II of pT181. The other IR, however, has not been shown to contain a determinant of specificity, nor has the binding interaction between initiator and origin been studied in detail (Baas, 1985). Both the pT181 nick site and the central element of IR-III show a remarkable degree of similarity with the corresponding elements of the fd origin (see Figure 2), which may well represent common ancestry.

Materials and methods

Bacterial strains and plasmids

The staphylococcal strains used are derivatives of strain RN450 (Novick, 1967). RN451 was derived by lysogenization with phage ϕ11 (Novick, 1967). RN4220 is a mutant that is an efficient acceptor of *Escherichia coli* DNA (Kreiswirth *et al.*, 1983). Plasmids are listed in Table II. The six members of the pT181 family are pT181, pC221, pS194, pCW7, pUB112 and pC223 (Table II) and their Rep proteins and (+) origins are referred to as RepC and *oriC*, RepD and *oriD*, Rep E1 and *oriE1*, RepE2 and *oriE2*, RepI and *oriI*, RepJ and *oriJ* (Table II, Figure 2), respectively, on the basis of the incompatibility groups of parental plasmids (note that *oriC* is unrelated to the *E.coli* chromosomal replication origin, which has the same abbreviation). pRN5101, used for cloning of the (+) origins, is a derivative of pE194, an unrelated plasmid with a mutation causing thermosensitive replication (Tsr) (Novick *et al.*, 1984). pSA7542 is a cointegrate formed between a pT181 copy mutant, *cop-620*, and an erythromycin resistance plasmid, pE5, whose replication system is unrelated to that of either pT181 or pE194 (Iordenescu, 1989). The pT181 replication origin on pSA7542 has been inactivated by a 4 nt replacement that eliminates the nick site, but does not change the amino acid sequence of the RepC protein. The erythromycin resistance gene, *ermC*, has been inactivated by filling in a restriction site (Iordanescu, 1989). pSA7461 is a derivative of pSA7542 in which the *repC* structural gene was replaced by the *repD* structural gene (Iordanescu, 1989).

Construction of origin variants

The 1571 nt *ClaI–HpaI* fragment (coordinates 3677 and 811, respectively) from pT181 (Khan and Novick, 1983) containing the leading (+) strand origin region was cloned into the phagemid, pBluescript KS(+) (Stratagene, La Jolla, CA), using *ClaI* and *Eco*RV sites, resulting in a derivative phagemid, pWN8900 (Wang *et al.*, 1992). Several new point mutations, a 2 bp deletion (*oriC114*) and a 5 bp insertion (*oriC113*) in the cloned origin (Table III) were created by oligonucleotide-directed mutagenesis (Kramer *et al.*, 1984; Kunkel *et al.*, 1987). A 10 bp insertion in the region between IR-II and IR-III (*oriC115*) (Table III) was created by filling in the *Bst*EII site within the previously inserted 5 nt linker. The 653 nt *Taq*I fragment containing each mutant origin region was cloned into pRN5101, a Tsr mutant of pE194. Each mutation was confirmed by DNA sequencing, using the dideoxynucleotide method (Sanger *et al.*, 1977) with sequenase (US Biochemical, Cleveland, OH). The mutagenic oligonucleotide primers were synthesized in the PHRI Microchemistry Laboratory by Dr Y.K.Yip. Most enzymes used were obtained commercially from Boehringer, Indianapolis, IN.

Transformation and transduction

Protoplast transformation of *Staphylococcus aureus* was carried out as described by Chang and Cohen (1979), as modified for *S.aureus* by Murphy (1983). Transduction was performed as described (Novick, 1967). Staphylococcal plasmids containing cloned fragments that had been manipulated in *E.coli* were re-introduced into *S.aureus* by protoplast transformation of a mutant staphylococcal strain, RN4220 (Kreiswirth *et al.*, 1983), which is an efficient acceptor of *E.coli* DNA.

Analysis of replication activity in vivo

The ability of a Rep protein to initiate the replication *in trans* of a plasmid containing a cloned target origin was tested by scoring for replication *in vivo* of the target origin-containing plasmid. These experiments made use of the Tsr vector, pRN5101, to which the various mutant origins had been cloned. Replication at the restrictive temperature was dependent on a functional pT181 origin and on the presence of a second plasmid producing RepC. For each of these clones, the origin-containing *Taq*I fragment was inserted at the *Cla*I site of pRN5101. These tests were evaluated qualitatively by visual inspection of agarose gel patterns or quantitatively by the measurement of plasmid copy numbers.

For electrophoretic analysis, cells were grown in CY broth (Novick and Brodsky, 1972), with plasmid-selective antibiotics, at 32°C to stationary phase. These cultures were used to inoculate CY broth without antibiotics (1:100) and duplicate cultures were grown at 32°C and 43°C respectively for 3 h (~4–5 generations). Whole cell lysates were prepared from these cultures, separated by agarose gel electrophoresis, stained and photographed.

Plasmid copy number determination

Copy numbers were determined by direct fluorescence densitometry of ethidium bromide-stained gels (Projan *et al.*, 1983) or by measuring the incorporation of [^{3}H]thymidine into plasmid compared with chromosomal DNA (Weisblum *et al.*, 1979). Fluorescence was measured with a Shimadzu CS-910 chromatogram scanner (Shimadzu Scientific Instruments, Columbia, MD) using the 254 nm Hg emission band for excitation and a 550 nm interference filter for measurements. Copy numbers were calculated from the ratio of the integrated plasmid signal to that of the chromosomal DNA, correcting for the lower ethidium bromide binding of supercoiled versus linear DNA as described and using previously determined values for chromosomal DNA/cell at different temperatures (Projan *et al.*, 1983).

For radioactive labeling, CY cultures prepared as above were incubated with [^{3}H]thymidine (1–5 μCi/ml) for 20 min prior to harvesting. Whole cell lysates were prepared from 1 ml samples, separated on agarose and stained with ethidium bromide. The plasmid and chromosomal bands were excised, autoclaved in 4 ml H_2O, then counted using Ultraflour (National Diagnostics, Manville, NJ) (4 ml/vial) in a Beckman LS200 scintillation counter, using non-radioactive gel slices to correct for quenching by agarose and by the ethidium bromide–DNA complex. Copies per cell, C_p, were calculated by the formula:

$$C_p = (D_p \times M_c)/(D_c \times M_p)$$

where D_p and D_c are the relative amounts of plasmid and chromosomal DNA (or radioactive counts) in the gel and M_c and M_p are total DNA/cell and plasmid molecular weight respectively.

Acknowledgements

We thank Drs L.A.Day, M.Gennaro and S.Iordanescu for discussions and a critical reading of the manuscript and S.Moghazeh for technical assistance. The PHRI computer facilities are funded by National Science Foundation grant PCM8313516. This work was supported by grant GM14372 to R.P.N. from the National Institutes of Health.

References

Baas,P.D. (1985) *Biochim Biophys Acta*, **825**, 111–139.
Baas,P.D. and Jansz,H.S. (1988) *Curr. Top. Microbiol. Immunol.*, **136**, 31–70.
Baas,P.D., Teertstra,W.R., Mansfeld,A.D.M.V., Jansz,H.S., Marel,G.A.V., Veenman,G.H. and Boom,J.H.V. (1981) *J. Mol. Biol.*, **152**, 615–639.
Balson,D.R. and Shaw,W.V. (1990) *Plasmid*, **24**, 74–80.
Beato,M. (1989) *Cell*, **56**, 335–344.
Bramhill,D. and Kornberg,A. (1988) *Cell*, **54**, 915–918.
Carleton,S., Projan,S.J., Highlander,S.K., Monghazeh,S. and Novick,R.P. (1984) *EMBO J.*, **3**, 2407–2414.
Chang,S. and Cohen,S.N. (1979) *Mol. Gen. Genet.*, **168**, 111–115.
Dotto,G.P., Horiuchi,K. and Zinder,N.D. (1982) *Proc. Natl. Acad. Sci. Sci.* USA, **79**, 7122–7126.
Genarro,M.L., Iordanescu,S., Novick,R.P., Murray,R.W., Steck.,T.R. and Khan,S.A. (1989) *J. Mol. Biol.*, **205**, 355–362.
Gielow,A., Diederich,L. and Messer,W. (1991) *J. Bacteriol.*, **173**, 73–79.
Gruss,A. and Ehrlich,S.D. (1989) *Microbiol. Rev.*, **53**, 231–241.
Hanes,S.D. and Brent,R. (1991) *Science*, **251**, 426–430.
Iordanescu,S. (1976a) *Arch. Roum. Pathol. Exp. Microbiol.*, **35**, 111–118.
Iordanescu,S. (1976b) *Arch. Roum. Pathol. Exp. Microbiol.*, **35**, 257–264.
Iordanescu,S. (1989) *Mol. Gen. Genet.*, **217**, 481–487.
Khan,S.A. and Novick,R.P. (1983) *Plasmid*, **10**, 251–259.
Kleanthous,H., Clayton,C.L. and Tabaqchali,S. (1991) *Mol. Mocrobiol.*, **5**, 2377–2389.
Koepsel.R.R. and Khan,S.A. (1986) *Science*, **233**, 1316–1318.
Koepsel,R.P. and Khan,S.A. (1987) *Nucl. Acids Res.*, **15**, 4085–4098.
Koepsel,R.R., Murray,R.W. and Khan,S.A. (1986) *Proc. Natl. Acad. Sci.* USA, **83**, 5484–5488.
Koudelka,G.B. and Carlson,P. (1992) *Nature*, **355**, 89–91.
Kramer,W., Drutsa,V., Jansen,H., Kramer,B., Pflugfelder,M. and Fritz,H. (1984) *Nucleic Acids Res.*, **12**, 9441–9456.
Kreiswirth,B., Lofdahl,S., Betley,M.J., O'Reilly,M., Schlivevert,P.M., Bergdoll,M.S. and Novick,R.P. (1983) *Nature*, **305**, 709–712.
Kunzel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
Luisi,B.F., Xu,W.X., Otwinowski,Z., Freedman,L.P., Yamamoto,K.R. and Sigler,P.B. (1991) *Nature*, **352**, 497–505.
Murphy,E. (1983) *Plasmid*, **10**, 260–269.
Noirot,P., Bargonetti,J. and Novick,R.P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8560–8564.
Novick,R.P. (1967) *Virology*, **33**, 155–166.
Novick,R.P. (1989) *Annu. Rev. Microbiol.*, **43**, 537–565.
Novick,R. and Bouanchaud,D. (1971) *Ann. N.Y. Acad. Sci.*, **182**, 279–294.
Novick,R.P. and Brodsky,R. (1972) *J. Mol. Biol.*, **68**, 285–302.
Novick,R.P., Adler,G.K., Projan,S.J., Carleton,S., Highlander,S., Gruss,A., Khan,S.A. and Iordanescu,S. (1984) *EMBO J.*, **3**, 2399–2405.
Projan,S.J., Carleton,S. and Novick,R.P. (1983) *Plasmid*, **9**, 182–190.
Projan,S.J. and Novick,R.P. (1988) *Plasmid*, **19**, 203–221.
Projan,S.J., Kornblum,J., Moghazeh,S., Edelman,I., Gennaro,M.L. and Novick,R.P. (1985) *Mol. Gen. Genet.*, **199**, 452–464.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Thomas,C.D., Balson,D.F. and Shaw,W.V. (1990) *J. Biol. Chem.*, **265**, 5519–5530.
Vargas,L.M., Kim,S. and Landy,A. (1989) *Science*, **244**, 1457–1461.
Wang,P.Z., Projan,S.J., Henriquez,V. and Novick,R.P. (1990) In Novick,R.P. (ed.), *Molecular Biology of the Staphylococci*. VHC Publishers, Inc., pp. 231–239.
Wang,P.Z., Projan,S.J., Henriques,V. and Novick,R.P. (1992) *J. Mol. Biol.*, **223**, 145–158.
Weisblum,B., Graham,M.Y., Gryczan,R. and Dubnau,D. (1979) *J. Bacteriol.*, **137**, 635–643.
Wilson,C.R. and Baldwin,J.N. (1987) *J. Bacteriol.*, **136**, 402–413.
Yasukawa,H., Hase,T., Sakai,A. and Masamune,Y. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10282–10286.

Received on July 22, 1992; revised on October 8, 1992