910–916 Nucleic Acids Research, 1995, Vol. 23, No. 6

Determination of the origin cleavage and joining domain of geminivirus Rep proteins

Françoise Heyraud-Nitschke*, Silke Schumacher[1], Jürgen Laufs[1], Sabine Schaefer, Jeff Schell and Bruno Gronenborn[1]

Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, D-50829 Köln, Germany and [1]Institut des Sciences Végétales, CNRS, Avenue de la Terrasse, F-91198 Gif-sur-Yvette, France

Received January 13, 1995; Revised and Accepted February 13, 1995

ABSTRACT

Replication of the single-stranded DNA genome of plant geminiviruses follows a rolling circle mechanism. It strictly depends on a 'rolling circle replication initiator protein', the M_r 41 kDa viral Rep protein, encoded by the C1 or AC1 genes. Using wheat dwarf virus (WDV) and tomato yellow leaf curl virus (TYLCV) as examples, we show that not only the full-size Rep proteins, but also a putative 30 kDa translation product of WDV open reading frame C1-N as well as an artificially shortened 24 kDa Rep of TYLCV, cleave and join single-stranded origin DNA *in vitro*. Thus the pivotal origin recognition and processing activities of geminivirus Rep proteins must be mediated by the amino-terminal domain of Rep.

INTRODUCTION

Geminiviruses are single-stranded DNA plant viruses with one or two circular genome components (for a review see ref. 1). They multiply in the nucleus of infected cells via a double-stranded DNA intermediate that is subsequently used as template for rolling circle replication of the viral-strand DNA (2,3). Geminiviruses represent a unique and simple system to study DNA replication and virus–host interactions in plants: they use host polymerase(s) to replicate their DNA, and only one viral protein is indispensable for replication (4,5). This M_r 41 kDa Rep protein is encoded by the AC1 (in bipartite geminiviruses) or C1 (in monopartite geminiviruses) gene and its *cis*-targets are located within the (large) intergenic region, a sequence including the ~200 almost identical bases in a given pair of bipartite geminivirus genomes, 'the common region' (6–8). The intergenic region of all geminiviruses contains a highly conserved element which consists of a GC-rich inverted repeat sequence able to form a hairpin-loop structure. The loop is a 10–13 nucleotide (nt) long sequence that invariably contains the nonamer TAATATTAC (1). Recently, genetic (9,10) and biochemical (11) analyses have demonstrated that this sequence element is the origin of viral-strand DNA synthesis. Geminivirus Rep protein initiates rolling circle replication by cleavage within the nonamer, 5′ of the penultimate adenosine ($TAATATT_{OH}$ p-AC), thereby generating a 5′ phosphoryl group to which the protein becomes covalently linked and a 3′ hydroxyl group which supposedly serves as a primer for viral-strand DNA synthesis. Based on what is known for rolling circle replication of phage ΦX174 (12), plasmids of Gram-positive (13) and Gram-negative bacteria (14,15) or parvoviruses (16), it is proposed (11) that, after the viral-strand DNA has been completely displaced, it becomes released in a Rep-catalyzed transfer reaction between the fixed 5′ end and the nonamer of the newly synthesized DNA strand by transesterification to the 3′ hydroxyl of thymidine 7 of the nonamer.

In addition to its key role in viral DNA replication, the geminivirus Rep protein specifically recognizes double-stranded sequences in the intergenic region (6,7) and represses transcription of its own gene (17,18). It also transactivates expression of the capsid gene promoter in geminiviruses of monocotyledonous plants (19). In the latter, of which wheat dwarf virus (WDV) and *Digitaria* streak virus (DSV) are representatives, the Rep protein is expressed from a spliced mRNA encompassing open reading frames (ORF) C1-N and C1-C (20,21). The M_r 40.5 kDa Rep protein of WDV (see Fig. 2) is a fusion between the amino-terminal 210 amino acids of the putative 30 kDa translation product of ORF C1-N (264 amino acids) and the 17 kDa domain (except the amino-terminal 9 amino acids) encoded by ORF C1-C. ORF C1-C has no AUG initiation codon and is probably not expressed individually *in vivo*. In contrast, the translation product of ORF C1-N alone could be derived from the unspliced mRNA (20,22); however, no direct evidence for the expression of this 30 kDa protein *in vivo* has been reported so far.

In order to attribute the various activities of the geminivirus Rep protein to defined protein domains, we expressed the 30 kDa translation product of ORF C1-N of WDV and a 24 kDa engineered version of the Rep protein of tomato yellow leaf curl virus (TYLCV, a monopartite geminivirus infecting dicotyledonous plants; 23,24) in *E.coli*. Their activities in *in vitro* cleavage assays (11) were analyzed and compared to those of the similarly expressed full-length Rep proteins. Here we show that the Rep amino-terminal domain catalyzes cleavage and joining at the viral-strand origin of replication.

* To whom correspondence should be addressed

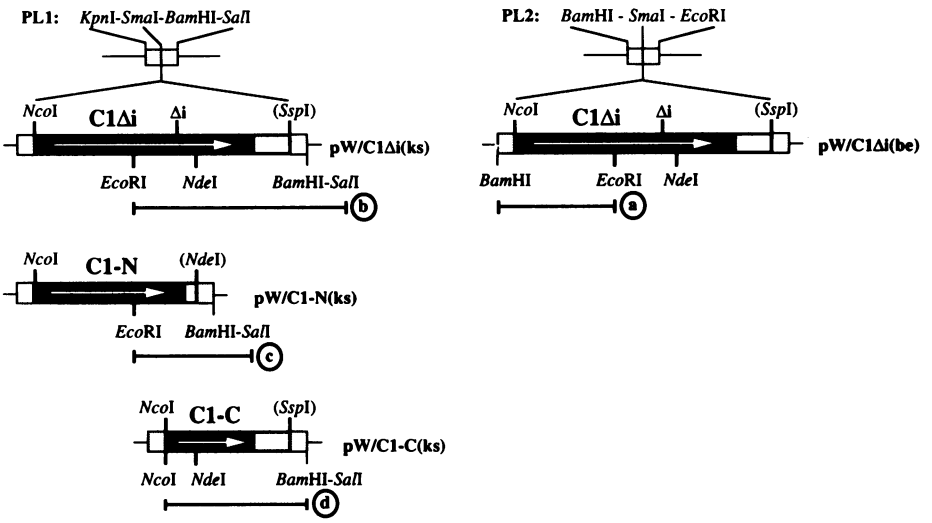

Figure 1. Construction of pMal-C1Δi, pMal-C1-N and pMal-C1-C from the intermediate plasmids pW/C1Δi(ks), pW/C1Δi(be), pW/C1-N(ks) and pW/C1-C(ks). C1Δi was cloned in the *Sma*I site of polylinkers PL1 (*Kpn*I–*Sma*I–*Bam*HI–*Sal*I; derived from pUC18) and PL2 (*Bam*HI–*Sma*I–*Eco*RI; derived from pGEX), whereas C1-N and C1-C were cloned in the *Sma*I site of PL1 only. Polylinkers and cloned fragments are not drawn to scale. Relevant restriction sites are indicated. ORFs are represented by the black boxes and the white arrows indicate the polarity of the translation products. Δi indicates the position of the splicing event that generates ORF C1Δi. ORFs C1Δi and C1-N were contained in *Nco*I–*Ssp*I and *Nco*I–*Nde*I fragments whose ends were filled-in prior to insertion into the *Sma*I site to recreate the *Nco*I sites; *Ssp*I and *Nde*I sites were destroyed (indicated by brackets). ORF C1-C was obtained by replacement of the 749 bp *Nco*I–*Nde*I fragment of ORF C1Δi in pW/C1Δi(ks) by a PCR generated 153 bp fragment. pMal-C1Δi was constructed by ligating fragments 'a' and 'b' to plasmid pMal-c2 cut by *Bam*HI+*Sal*I. Similarly, pMal-C1-N was constructed by ligating fragments 'a' and 'c' to the same plasmid. pMal-C1-C was constructed by inserting the filled-in fragment 'd' into pMal-c2, restricted by *Bam*HI and filled-in.

MATERIALS AND METHODS

Standard techniques in molecular biology were as described by Sambrook *et al.* (25).

Protein expression in *E.coli* and purification

Two different *E.coli* expression plasmids were used: for WDV-derived proteins, a pMal vector allowing fusion to a maltose binding protein (MBP; 26,27) and for TYLCV-derived proteins, a pGEX vector allowing fusion to a glutathione S-transferase (GST; 28).

The three WDV complementary-sense ORFs C1Δi, C1-N and C1-C were cloned in the vector pMal-c2 (New England Biolabs) and the cloning steps are depicted in Figure 1. ORF C1Δi was originally isolated from plasmid pWDVΔi (20) as a *Nco*I–*Ssp*I fragment (the *Nco*I restriction site overlaps the C1Δi AUG start codon and the *Ssp*I restriction site is located 3′ of ORF C1-C). The *Nco*I restriction site was filled-in and the resulting fragment was introduced in the *Sma*I cloning site of two polylinkers generating pW/C1Δi(ks) and pW/C1Δi(be). For the subsequent transfer of ORF C1Δi to pMal-c2, fragments 'a' (*Bam*HI–*Eco*RI) and 'b' (*Eco*RI–*Sal*I) were isolated and ligated to the *Bam*HI and *Sal*I cloning sites of pMal-c2. ORF C1-N of WDV was cloned in pMal-c2 in a similar way. ORF C1-N was isolated from plasmid pWS6 (29) as a 835 bp *Nco*I–*Nde*I fragment (the *Nco*I restriction site overlaps the C1-N AUG start codon and the *Nde*I restriction site is 3′ of ORF C1-N). This fragment was blunt-ended and introduced in the *Sma*I cloning site of the pUC18-derived polylinker PL1, generating pW/C1-N(ks). Fragments 'a' (*Bam*HI–*Eco*RI) and 'c' (*Eco*RI–*Sal*I) were isolated from pW/C1Δi(be) and pW/C1-N(ks) respectively, and ligated to the *Bam*HI and *Sal*I cloning sites of pMal-c2, generating pMal-C1-N.

To express the C1-C ORF individually, it was necessary to remove the stop codon preceding the first amino acid of C1-C. This was done by PCR using pWS6 as template and the following oligonucleotides as primers: p1 (5′-TGCGGATCA*GGATCC-ATGG*GA**CTTAGAGTG**-3′) and p2 (5′-GGCGTTGACGT*CA-TATG*TTGTG-3′). p1 overlaps the first three codons of ORF C1-C (shown in bold) and contains a mutant sequence (underlined) introducing *Bam*HI and *Nco*I restriction sites as well as an AUG codon in front of C1-C; p2 is an internal C1-C oligonucleotide overlapping the *Nde*I restriction site. The amplification reaction (50 µl) was in 10 mM Tris–HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, and contained 100 pg pWS6, 10 nmol each dNTP (final concentration: 200 µM), 100 pmol each p1 and p2 primers (final concentration: 2 µM) and 2 U *Taq* DNA polymerase (Perkin Elmer-Cetus). Thirty cycles (1 min 93°C, 45 s 42°C, 1 min 72°C) were run and the 182 bp product was gel-purified. It was subsequently digested with *Nco*I and *Nde*I (italicized restriction sites in the p1 and p2 sequences) and used to replace the 749 bp *Nco*I–*Nde*I fragment of ORF C1Δi in pW/C1Δi (ks), creating pW/C1-C(ks). ORF C1-C was released from pW/C1-C(ks) as a *Nco*I–*Bam*HI fragment (fragment 'd', Fig. 1) and, after treatment with the Klenow fragment of *E.coli* DNA polymerase I, was cloned in the filled-in *Bam*HI site of pMal-c2.

MBP-fusion proteins (MBP-Rep: 83.9 kDa, MBP-$RepN_{30}$: 73.6 kDa and MBP-$RepC_{17}$: 60 kDa) were expressed in *E.coli* from pMal-C1Δi, pMal-C1-N and pMal-C1-C respectively, and purification was from the soluble protein fractions as described by Riggs (27). The presence of a cleavage site for factor Xa between the MBP- and various carboxy-domains of the fusion proteins allowed purification of the corresponding Rep domains. Conditions for cleavage with factor Xa and separation of the products (MBP: 42.7 kDa, Rep: 41.2 kDa, $RepN_{30}$: 30.9 kDa; chromatography on

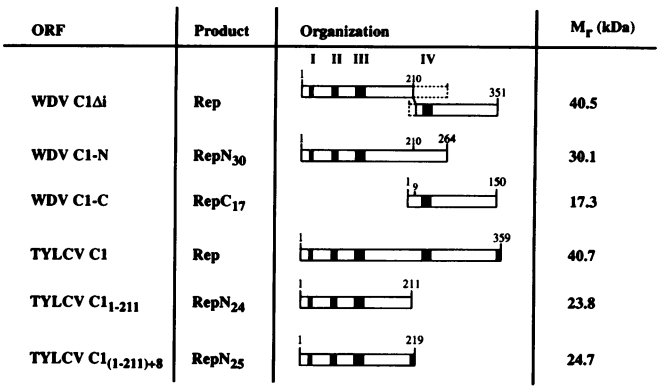

Figure 2. Organization of WDV and TYLCV Rep proteins. WDV Rep, $RepN_{30}$ and $RepC_{17}$ are the translation products of ORFs C1Δi, C1-N and C1-C respectively. C1Δi is expressed from a spliced transcript that joins ORFs C1-N and C1-C (the splice sites are represented by the diagonal line). The resulting Rep protein is a fusion between the amino-terminal 210 amino acids of the $RepN_{30}$ polypeptide and 141 amino acids encoded by ORF C1-C. Individual $RepN_{30}$ and $RepC_{17}$ domains comprise 264 and 150 amino acids respectively. TYLCV Rep protein and its two shortened versions, $RepN_{24}$ and $RepN_{25}$, are the translation products of ORFs C1, $C1_{1-211}$ and $C1_{(1-211)+8}$ respectively. $RepN_{24}$ consists of the amino-terminal 211 amino acids of Rep and $RepN_{25}$ consists of $RepN_{24}$ to which the last carboxy-terminal 8 amino acids of Rep (represented by the shaded box, not drawn to scale) have been added. All Rep domains were expressed as protein fusions to either MBP (WDV) or GST (TYLCV). Black boxes (not drawn to scale) represent conserved amino acid motifs as described by Koonin and co-workers (30,31); explanations are given in the text.

Q-sepharose) were according to the supplier (New England Biolabs).

For TYLCV, a GST-fusion system was chosen for expression in *E.coli*. Shortened versions of the M_r 40.7 kDa TYLCV Rep protein were engineered from pGEXEC1, a modified pGEX vector expressing the entire TYLCV Rep protein (11). DNA of pGEXEC1 was used as template to amplify the sequence coding for the 211 amino-terminal residues of Rep and introduce a termination codon after amino acid 211 (this ORF was referred to as $C1_{1-211}$). In a second version, the coding sequence of the last eight amino acids of the complete Rep protein were included prior to the termination codon in the carboxy-terminal primer (this ORF was referred to as $C1_{(1-211)+8}$). The following oligonucleotides were used:

amino-terminal' primer (pa): 5'-GTGGGATCCCGGGCTCC-GGCTCCGGCGACGACGACGACAAAATGCCAAGATCA-GGTCGTTT-3';

carboxy-terminal' primer 1 (pc1): 5'-ATACTCA*GTCGAC*TT-CAAGGCCGCGCGGCCGCATCCATGACG-3';

carboxy-terminal' primer 2 (pc2): 5'-ACTCA*GTCGAC*TT-CAGGCCTCACTTGCCTCTTCTTGGCGAGGCCGCGCG-GCCGCATCCATGACG-3'.

The sequences encoding the beginning and the respective ends of the ORFs are shown in bold, with start and stop codons underlined; the *SalI* restriction sites used for cloning are italicized. Standard amplification reactions [50 ng template, 250 μM each dNTP, 1 μM each primer, Promega buffer and 2.5 U of *Taq* polymerase (Promega)] were run for 30 cycles (1 min 94°C, 30 s 50°C, 30 s 72°C). A 691 bp fragment was amplified with primer combination pa/pc1 and a 713 bp fragment with primer combination pa/pc2. After restriction by endonucleases *SstI* and *SalI*, fragments of 347 and 369 bp respectively, bearing the new carboxyl termini of Rep were used to replace the 492 bp wild-type *SstI–SalI* fragment in plasmid pGEXEC1.

Protein expression and purification were as previously described (11).

In vitro cleavage-joining reactions

In vitro cleavage and joining assays were essentially as described by Laufs *et al.* (11). 5' end-labelled synthetic oligonucleotides encompassing the replication origin sequences of TYLCV or WDV were used as substrates. Reaction products were fractionated by electrophoresis on 12 or 15% denaturing polyacrylamide gels that were dried and autoradiographed.

RESULTS

Figure 2 schematically displays the WDV- and TYLCV-derived proteins that were used for *in vitro* analyses. Conserved amino acid motifs, as reported by Koonin and Ilyina, and Ilyina and Koonin (30,31), are indicated. Motif I, the function of which remains to be determined, has the sequence FLTY[PS]xC (residues in brackets: alternatives; x: no consensus at this position). Motif II, the 'two His motif' [HlHUlUQ (lower case letters: amino acids conserved in almost all sequences; U: hydrophobic amino acids)], is proposed to function as a ligand to divalent cations such as Mg^{2+}. Motif III (vxxYUxK) contains the conserved tyrosine residue (Y) which is assumed to mediate the link to the DNA after cleavage and prior to joining. Motif IV is the P-loop motif of the NTP-binding site with the consensus sequence G[PD][ST]RTGK[ST]. The proteins were produced in *E.coli*, either as fusions to a maltose binding protein (MBP) for WDV-derived proteins, or as fusions to a glutathione S-transferase (GST) for TYLCV-derived proteins. Besides the wild-type Rep proteins, translation products of WDV ORFs C1-N and C1-C (referred to as $RepN_{30}$ and $RepC_{17}$ respectively), as well as engineered short versions of the TYLCV Rep protein (referred to as $RepN_{24}$ and $RepN_{25}$) were expressed. The protein preparations were analyzed in standard *in vitro* cleavage-joining assays, as previously described (11). DNA substrates were 5' end-labelled synthetic oligonucleotides encompassing the origin of viral-strand DNA synthesis of WDV or TYLCV.

The 30 kDa translation product of WDV ORF C1-N ($RepN_{30}$) cleaves the viral-strand DNA origin

Substrates for WDV-derived proteins (Fig. 3A) were oligonucleotides WD27 (33mer) and WD28 (34mer). WD27 essentially consists of the loop sequence (nucleotides 2–13) and the 3' part of the inverted repeat sequence, whereas WD28 consists of the 5' part of the inverted repeat sequence and the loop sequence (nucleotides 20–31). Upon incubation of 5' end-labelled oligonucleotides WD27 (Fig. 3B, lanes 1–7) and WD28 (Fig. 3B, lanes 8–14), with either MBP-Rep (lanes 3 and 10) or MBP-$RepN_{30}$ (lanes 4 and 11), labelled cleavage products of 11 nt (lanes 3 and 4) and 29 nt (lanes 10 and 11) were generated respectively (the product sizes were determined by running the cleavage reactions along with oligonucleotides of defined sizes, data not shown). No cleavage was observed when MBP-$RepC_{17}$ was used instead of

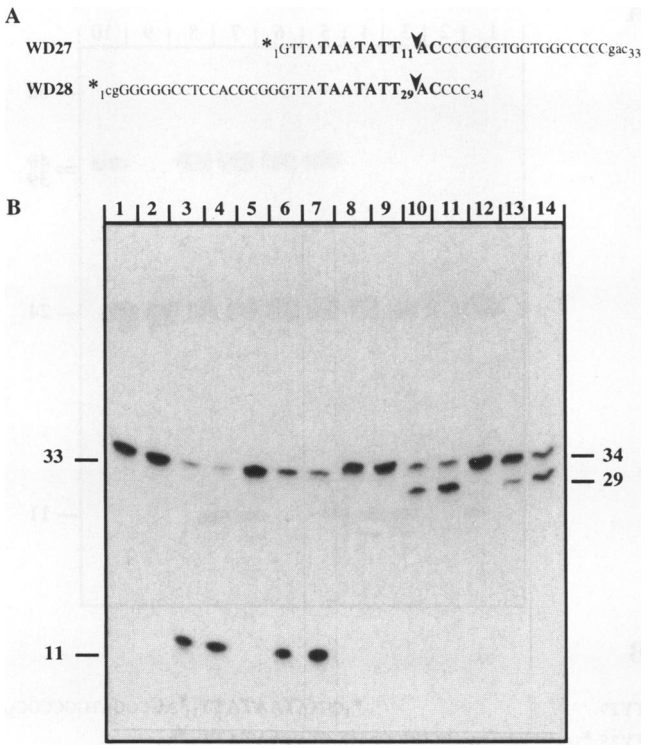

Figure 3. Site-specific cleavage of oligonucleotides WD27 and WD28 by WDV Rep and $RepN_{30}$. Oligonucleotides encompassing the viral-strand origin sequence of WDV were 5′ end-labelled and incubated with WDV Rep proteins. Products of the cleavage reactions were separated on a 15% denaturing polyacrylamide gel and visualized by autoradiography. **(A)** Nucleotide sequence of WD27 and WD28. Position 1 corresponds to the 5′ end of the oligonucleotide and the asterisk symbolizes the radiolabel (^{32}P). The arrowhead marks the cleavage site. Nucleotides contained in the hairpin-loop sequence are in upper case letters, with the conserved nonamer motif in bold. Sequences outside the inverted-repeat are indicated by lower case letters. **(B)** *In vitro* cleavage reactions were performed on WD27 (lanes 1–7) or WD28 (lanes 8–14) using the following protein preparations: MBP (lanes 2 and 9), MBP-Rep (lanes 3 and 10), MBP-$RepN_{30}$ (lanes 4 and 11), MBP-$RepC_{17}$ (lanes 5 and 12), Rep (lanes 6 and 13) and $RepN_{30}$ (lanes 7 and 14). No protein was added in assays run in lanes 1 and 8. Sizes of the substrate oligonucleotides (33 and 34 nt) and of the labelled products (11 and 29 nt) are indicated.

MBP-Rep or MBP-$RepN_{30}$ (lanes 5 and 12). In order to ascertain that the observed cleavage activity was due to the Rep domains of the fusion proteins, MBP-Rep and MBP-$RepN_{30}$ were proteolytically cleaved with factor Xa and the purified MBP-, Rep- and $RepN_{30}$-domains were individually used in cleavage assays. Rep (lanes 6 and 13) as well as $RepN_{30}$ (lanes 7 and 14) showed the same specific cleavage as MBP-Rep and MBP-$RepN_{30}$; in constrast, the MBP domain alone was inactive (lanes 2 and 9).

These data demonstrate that, *in vitro*, the entire WDV Rep protein as well as the 30 kDa domain of ORF C1-N specifically cleave the WDV viral-strand origin of replication between positions 7 and 8 of the nonamer sequence (TAATATT↓AC).

Both WDV Rep and $RepN_{30}$ cleave and join homologous and heterologous viral origins of replication *in vitro*

For TYLCV, we have shown that the Rep protein not only catalyzed an *in vitro* cleavage reaction but also a joining reaction at the viral-strand origin of replication (11). In order to find out whether the same was true for WDV, cleavage and joining assays were performed using an equimolar mixture of two 5′ end-labelled oligonucleotides as substrate.

Four products were detected upon incubation of WD27 and WD28 with MBP-Rep [Fig. 4A, lanes 3 and 7 (longer exposure of lane 3)] or MBP-$RepN_{30}$ [Fig. 4A, lanes 4 and 8 (longer exposure of lane 4)]: the 29 and 11 nt products are the labelled cleavage products of WD28 and WD27 respectively (see also lanes 5 and 6); the 51 and 16 nt products (the latter being clearly visible after long exposure) are those expected from the heterologous joining of the WD28 and WD27 cleavage products. They result, as schematically shown in Figure 4C, from the joining between the 5′ phosphoryl end of the unlabelled 22mer cleavage product of WD27 and the 3′ hydroxyl end of the labelled 29mer cleavage product of WD28, and between the 5′ phosphoryl end of the unlabelled 5mer cleavage product of WD28 and the 3′ hydroxyl end of the labelled 11mer cleavage product of WD27 respectively. Neither cleavage nor joining were detected if the two oligonucleotides were incubated with MBP-$RepC_{17}$ (lane 2) or MBP-βgalα (lane 1; this protein is the translation product of the *mal*E–lacZα gene fusion and is produced upon IPTG induction of *E.coli* cultures transformed with the wild-type pMal-c2 vector).

Interestingly, both activities were also observed when MBP-Rep or MBP-$RepN_{30}$ were incubated with a WDV/TYLCV oligonucleotide mixture. An example is shown in Figure 4B (lanes 6 and 7) for 5′ end-labelled oligonucleotides WD28 (34mer) and TY19 (24mer). The 29 and 11 nt products are the labelled cleavage products of WD28 and TY19 respectively; the 42 nt product results from a joining reaction between the 5′ phosphoryl end of the unlabelled 13 nt cleavage product of TY19 and the 3′ hydroxyl end of the labelled 29 nt cleavage product of WD28 (see also Fig. 4D for a schematic representation). The second recombinant joining product (16mer) was visible only after a longer exposure of the gel. It has to be noted that the joining of WDV-derived cleavage products with TYLCV-derived cleavage products consistently was less efficient than that observed between WDV-derived cleavage products only, as illustrated in lanes 2–3 and 6–7 where oligonucleotide combinations WD28/WD27 and WD28/TY19 were treated in parallel. Hence, WDV Rep and $RepN_{30}$ not only cleave and join their cognate viral-strand origin of replication but also perform these activities at a heterologous origin, albeit at a lower efficiency.

The 24 kDa amino-terminal domain of TYLCV Rep has cleaving and joining activities

Is there a basic difference in the domain organization of the Rep proteins that are expressed from a spliced mRNA and those that are expressed from a single ORF? To answer this question we designed a short version of the TYLCV Rep protein ($RepN_{24}$), consisting of the first 211 amino acids of the N-terminal domain. Since the stability in bacteria of such an artificially shortened Rep domain was not predictable, we also constructed a version ($RepN_{25}$) that contained, after amino acid 211, the carboxy-terminal eight hydrophilic amino acids of the full-size 40.7 kDa Rep protein to potentially recreate the natural C-terminal end. Both shortened domains (23.8 and 24.7 kDa in size respectively) were fused to glutathione S-transferase, expressed in *E.coli* and

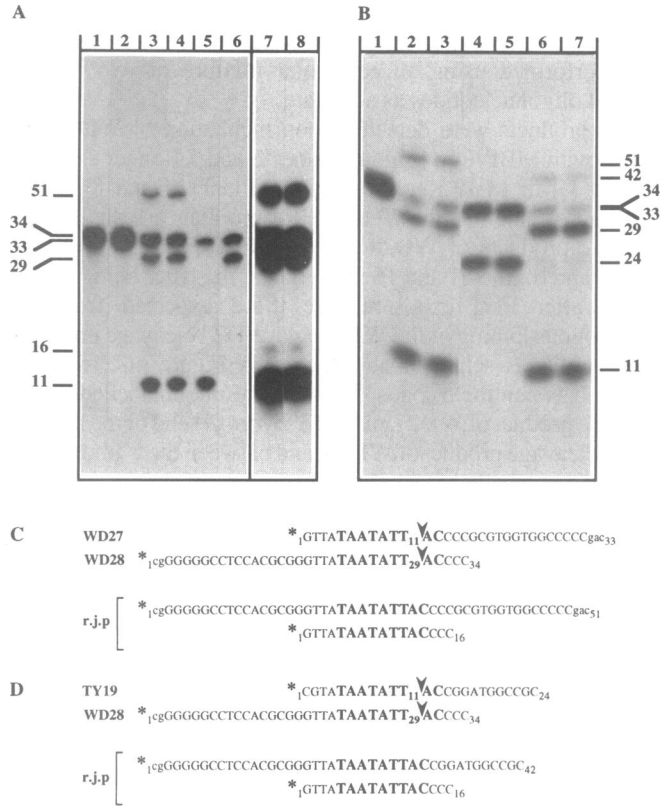

Figure 4. *In vitro* cleavage and joining reactions catalyzed by WDV Rep and $RepN_{30}$. (**A**) An equimolar mixture of 5′ end-labelled oligonucleotides WD27 and WD28 (lanes 1–4, 7 and 8) was used as substrate for MBP-βgalα (lane 1), MBP-$RepC_{17}$ (lane 2), MBP-Rep (lanes 3 and 7) or MBP-$RepN_{30}$ (lanes 4 and 8). Lanes 7 and 8 are overexposures of lanes 3 and 4 respectively. Lanes 5 and 6 contain cleavage reactions performed with MBP-Rep on WD27 and WD28 respectively. The lengths indicated on the left are in nt; 51 and 16: recombinant joining products [see (C) for a schematic representation]; 34 and 33: substrates; 29 and 11: cleavage products. (**B**) An equimolar mixture of 5′ end-labelled oligonucleotides WD27 and WD28 (lanes 1–3) was incubated with MBP-Rep (lane 2), MBP-$RepN_{30}$ (lane 3) or without protein (lane 1). In lanes 4–7, an equimolar mixture of 5′ end-labelled oligonucleotides TY19 and WD28 was incubated with MBP-βgalα (lane 5), MBP-Rep (lane 6), MBP-$RepN_{30}$ (lane 7) or without protein (lane 4). The lengths indicated on the right are in nt; 51 and 42: recombinant joining products [see (C) and (D) for a schematic representation]; 34, 33 and 24: substrates; 29 and 11: cleavage products. (**C**) Schematic representation of a cleavage/joining reaction using oligonucleotides WD27 and WD28 as substrates. Symbols are as described in Figure 3. r.j.p: recombinant joining products. (**D**) Schematic representation of a cleavage/joining reaction using oligonucleotides TY19 and WD28 as substrates. Symbols and abbreviations are the same as in (C).

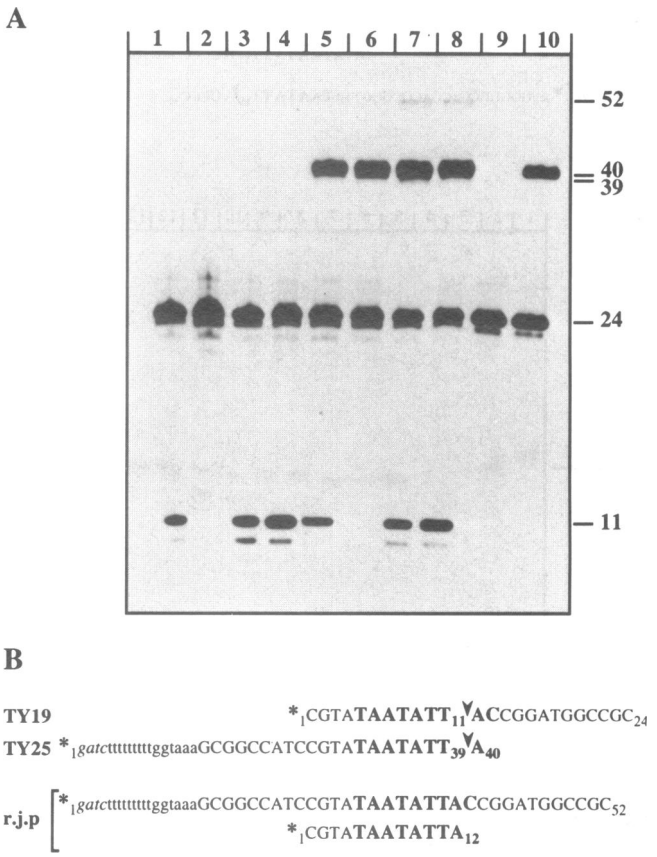

Figure 5. *In vitro* cleavage and joining reactions catalyzed by TYLCV Rep, $RepN_{24}$ and $RepN_{25}$. (**A**) For *in vitro* cleavage, 5′ end-labelled oligonucleotide TY19 (lanes 1–4 and lane 9) was incubated with GST-Rep (lane 1), GST (lane 2), GST-$RepN_{25}$ (lane 3), GST-$RepN_{24}$ (lane 4) or without protein (lane 9). For *in vitro* cleavage and joining, an equimolar mixture of 5′ end-labelled oligonucleotides TY19 and TY25 (lanes 5–8 and lane 10) was incubated with GST-Rep (lane 5), GST (lane 6), GST-$RepN_{25}$ (lane 7), GST-$RepN_{24}$ (lane 8) or without protein (lane 10). The lengths indicated on the right are in nt; 52: recombinant joining product [see (B) for a schematic representation]; 40 and 24: substrates; 39 and 11: cleavage products. (**B**) Schematic representation of a cleavage/joining reaction using oligonucleotides TY19 and TY25 as substrates. Symbols and abbreviations are the same as in Figure 4C. Italicized nucleotides are non-TYLCV sequences.

their cleavage-joining properties were compared to those of the wild-type TYLCV Rep.

The results are shown in Figure 5 where 5′ end-labelled oligonucleotide TY19 alone (24mer, lanes 1–4) and in combination with similarly labelled TY25 (40mer, lanes 5–8) were used as substrates. As previously demonstrated (11), GST-Rep cleaves TY19 (Fig. 5A, lane 1) generating products of 11 and 13 nt; TY19 being 5′ end-labelled, only the 11 nt long product is detected upon autoradiography (see also Fig. 5B for a schematic representation). Note that the 24mer used in this particular reaction is slightly contaminated by a 23mer (see also lane 9 where TY19 alone was loaded) which explains the additional appearance of a labelled 10mer. The same cleavage product was obtained when TY19 was incubated with GST-$RepN_{24}$ (lane 4) or GST-$RepN_{25}$ (lane 3), indicating that the 211 N-terminal amino acids of the TYLCV Rep protein retained the cleavage activity shown for the full-length protein.

To demonstrate the joining activity of GST-$RepN_{24}$ and GST-$RepN_{25}$, equimolar mixtures of TY19 and TY25 were used as substrates (lanes 7 and 8). When the cleavage-joining reaction was performed with GST-Rep (lane 5), a 52 nt joining product was generated (see also Fig. 5B for a schematic representation). The same 52 nt joining product was found when GST-$RepN_{24}$ (lane 8) or GST-$RepN_{25}$ (lane 7) were used in place of GST-Rep. The autoradiogram shown was overexposed to better visualize this 52mer joining product; as a consequence, the separation between 40 nt (TY25) and 39 nt (labelled cleavage product of TY25 of which only the 3′ terminal A is removed) is not very easily distinguished. In this particular experiment, the amount of

the 12 nt joining product (see Fig. 5B) was low and was only detected upon very long exposure (not shown). GST alone (derived from *E.coli* cultures transformed with the wild-type pGEX vector) was inactive in the cleavage-joining assays (lanes 2 and 6).

Taken together, the data demonstrate that the 211 N-terminal amino acids of the TYLCV Rep protein are sufficient for *in vitro* cleavage and joining at the TYLCV viral-strand origin of replication.

DISCUSSION

By expressing the different domains of the WDV Rep protein individually in *E.coli* and by shortening the 40.7 kDa wild-type TYLCV Rep protein to 24 kDa, we have shown that the 30 kDa amino-terminal domain of WDV Rep as well as the first 211 N-terminal amino acids of TYLCV Rep are capable to execute *in vitro* the cleavage and joining reactions that *in vivo* initiate and terminate a round of rolling circle replication. Hence, the cleavage and joining activities of the geminivirus Rep proteins must map to the amino-terminal domains of these proteins. Furthermore, some substrate requirements for the *in vitro* joining reaction have been outlined. Figures 4 and 5 show that the recombinant joining products corresponding to the reconstitution of the entire stem–loop sequence are readily produced (Fig. 4A: 51 nt product resulting from the transfer of the 5′ phosphoryl end of the 22mer cleavage product of WD27 to the 3′ hydroxyl end of the 29mer cleavage product of WD28; Fig. 5: 52 nt product resulting from the transfer of the 5′ phosphoryl end of the 13mer cleavage product of TY19 to the 3′ hydroxyl end of the 39mer cleavage product of TY25). In contrast, transfer of the 5′ phosphoryl end of the 5mer cleavage product of WD28 to the 3′ hydroxyl end of the 11mer cleavage product of WD27 (Fig. 4A) occurs at low efficiency. The transfer of the single adenosyl phosphate product of TY25 cleavage to the 3′ hydroxyl end of the 11mer cleavage product of TY19 (Fig. 5) could only be detected after prolonged exposure of the autoradiogram. Thus, an efficient transesterification reaction may require an interaction (possibly a base-pairing) between the acceptor sequence (3′ hydroxyl end of a cleavage site) and the donor sequence (5′ phosphoryl end of a cleavage site to which the Rep protein is bound). Results shown in Figure 4B, where an equimolar mixture of a WDV-derived oligonucleotide (WD28) and a TYLCV-derived oligonucleotide (TY19) was used as substrate for the WDV Rep or $RepN_{30}$ proteins, strengthen this hypothesis. Although cleavage of TY19 was as efficient as that of WD28, transfer of the 5′ phosphoryl end of the 13mer cleavage product of TY19 to the 3′ hydroxyl end of the 29mer cleavage product of WD28 was much less efficient than that observed when WD27 was used in place of TY19, probably because WD28 as hydroxyl end acceptor (29mer) does not easily base-pair with TY19 as phosphoryl end donor (13mer; Fig. 4D). For WDV, data obtained from protoplast transfection experiments had led to a similar conclusion (10,29). A 15 nt long sequence of the large intergenic region, highly homologous to the 3′ part of the stem–loop sequence only, served as a viral-strand origin of replication *in vivo*, but resolution of the polymeric replication intermediates into monomeric circular genomes was impaired. This result indicated that the formation of a particular secondary structure (hairpin-loop structure) appeared to be required for the release process.

The mapping of the cleavage/joining activity of the geminivirus Rep protein to the amino-terminal domain is a first step to further understand the multifunctionality and the modular organization of this protein. It experimentally substantiates predictions by sequence comparison analyses that have identified conserved motifs between various rolling circle replication initiator proteins and the 24 kDa amino-terminal domain of geminivirus Rep proteins (30,31). According to these comparisons, the conserved tyrosine in motif III is the DNA-linking residue of the active site involved in the cleavage/joining activity at the viral-strand origin of replication.

In addition to a replication initiator domain, all geminivirus Rep proteins contain a conserved nucleoside triphosphate (NTP) binding site that is located in the 17 kDa Rep domain of the monocot geminiviruses or in the equivalent carboxy-terminal part of the 41 kDa Rep of dicot geminiviruses (32,33). For TYLCV, it was demonstrated that the Rep protein expressed in *E.coli* possesses an ATPase activity *in vitro* (Desbiez *et al.*, submitted). Previous experiments indicated that ATP is not required for the cleavage/joining reaction (11). Here we unequivocally show that the physical separation of the ATPase domain from the cleavage/joining domain has no influence on the *in vitro* cleavage/joining activity. Studying the interaction of the AL1 protein of tomato golden mosaic virus (TGMV), the TGMV Rep, with its cognate intergenic region, Eagle and co-workers (18) have also demonstrated that mutations in the NTP-binding site do not affect the specific binding of Rep to its DNA target sequences. Taken together, the results of Eagle and co-workers (18) and the data presented here prove that the amino-terminal domain of the geminivirus Rep protein mediates DNA recognition as well as cleavage and joining at the viral-strand origin of replication. In the case of geminiviruses of monocotyledonous plants, as for instance WDV, this domain may be expressed as such *in vivo* from the unspliced complementary-sense transcript. This hypothesis, if verified, would raise the question of the biological role of an individually produced 30 kDa Rep domain. It is tempting to speculate that geminivirus Rep proteins act as dimers (or oligomers) since they probably possess only one active tyrosine for cleavage. A dimeric Rep protein complex, when covalently attached to the 5′ phosphoryl group of the displaced strand during DNA replication, would still have one reactive tyrosine available for a second cleavage and joining reaction, just as is the case for gpA protein of ΦX174 that has two reactive tyrosine residues per monomer (12). In addition, in the case of monocot geminiviruses, dimerization (or oligomerization) could also play a role in the regulation of Rep activity, for instance by heterodimerization between Rep and Rep_{30}. A control mechanism for the activity of a bacterial initiator protein that is based on dimer formation between active and inactive subunits, has been shown for the RepC protein of *S.aureus* plasmid pT181 (34).

For its full activity in the replication cycle of the virus however, Rep requires a functional ATPase domain (18, Desbiez *et al.*, submitted). The presence of the NTP-binding motif has prompted speculations that Rep may also act as a helicase (32,33) but this potential function still awaits proof.

We have begun to dissect the Rep protein into individual domains and to biochemically study their respective functions *in vitro*. The possibility to prepare individual functional domains of the protein will facilitate their structural analysis. Hence, the multifunctional geminivirus Rep protein provides a unique opportunity to unravel protein–DNA and protein–protein interac-

tions in plants at the molecular level. With regard to geminivirus replication, interesting features of these interactions have been highlighted in a recent work by Argüello-Astorga and co-workers (35,36) and an attractive model that involves the participation of host transcription factors has been proposed. Furthermore, some aspects of an intriguing functional similarity between the geminivirus Rep proteins or proteins involved in rolling circle replication of bacterial phages and plasmids, on the one hand, and proteins participating in bacterial conjugative DNA transfer or in the transfer and integration of the T-DNA from *A.tumefaciens* into the plant genome, on the other hand, are becoming apparent (37). Further studies on Rep proteins may yield some additional evolutionary surprises.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Bundesministerium für Forschung und Technologie to FH and in part by a grant from AFIRST to BG. SS gratefully acknowledges support by a predoctoral fellowship from the Deutscher Akademischer Austauschdienst. JL was supported by EMBO and AFIRST.

REFERENCES

1 Lazarowitz, S.G. (1992) *Crit. Rev. Plant Sci.* **11**, 327–349.
2 Saunders, K., Lucy, A. and Stanley, J. (1991) *Nucleic Acids Res.* **19**, 2325–2330.
3 Stenger, D.C., Revington, G.N., Stevenson, M.C. and Bisaro, D.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8029–8033.
4 Rogers, S.G., Bisaro, D.M., Fraley, R.T., Hoffmann, N.L., Brand, L., Elmer, J.S. and Lloyd, A.M. (1986) *Cell* **45**, 593–600.
5 Elmer, J.S., Brand, L., Sunter, G., Gardiner, W.E., Bisaro, D.M. and Rogers, S.G. (1988) *Nucleic Acids Res.* **16**, 7043–7060.
6 Fontes, E.P.B., Luckow, V.A. and Hanley-Bowdoin, L. (1992) *Plant Cell* **4**, 597–608.
7 Lazarowitz, S.G., Wu, L.C., Rogers, S.G. and Elmer, J.S. (1992) *Plant Cell* **4**, 799–809.
8 Thömmes, P., Osman, T.A.M., Hayes, R.J. and Buck, K.W. (1993) *FEBS Lett.* **319**, 95–99.
9 Heyraud, F., Matzeit, V., Schaefer, S., Schell, J. and Gronenborn, B. (1993) *Biochimie* **75**, 605–615.
10 Heyraud, F., Matzeit, V., Kammann, M., Schaefer, S., Schell, J. and Gronenborn, B. (1993) *EMBO J.* **12**, 4445–4452.
11 Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J. and Gronenborn, B. (1995) *Proc. Natl. Acad. Sci. USA*, in press.
12 van Mansfeld, A.D.M., van Teeffelen, H.A.A.M., Baas, P.D. and Jansz, H.S. (1986) *Nucleic Acids Res.* **14**, 4229–4238.
13 Noirot-Gros, M.-F., Bidnenko, V. and Ehrlich, S.D. (1994) *EMBO J.* **13**, 4412–4420.
14 Yasukawa, H., Hase, T., Sakai, A. and Masamune, Y. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10282–10286.
15 Pansegrau, W., Schröder, W. and Lanka, E. (1994) *J. Biol. Chem.* **269**, 2782–2789.
16 Im, D.-S. and Muzyczka, N. (1990) *Cell* **61**, 447–457.
17 Sunter, G., Hartitz, M.D. and Bisaro, D.M. (1993) *Virology* **195**, 275–280.
18 Eagle, P.A., Orozco, B.M. and Hanley-Bowdoin, L. (1994) *Plant Cell* **6**, 1157–1170.
19 Hofer, J.M.I., Dekker, E.L., Reynolds, H.V., Woolston, C.J., Cox, B.S. and Mullineaux, P.M. (1992) *Plant Cell* **4**, 213–223.
20 Schalk, H.-J., Matzeit, V., Schiller, B., Schell, J. and Gronenborn, B. (1989) *EMBO J.* **8**, 359–364.
21 Accotto, G.P., Donson, J. and Mullineaux, P.M. (1989) *EMBO J.* **8**, 1033–1039.
22 Dekker, E.L., Woolston, C.J., Xue, Y., Cox, B. and Mullineaux, P.M. (1991) *Nucleic Acids Res.* **19**, 4075–4081.
23 Kheyr-Pour, A., Bendahmane, M., Matzeit, V., Accotto, G.P., Crespi, S. and Gronenborn, B. (1991) *Nucleic Acids Res.* **19**, 6763–6769.
24 Navot, N., Pichersky, E., Zeidan, M., Zamir, D. and Czosnek, H. (1991) *Virology* **185**, 151–161.
25 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
26 Maina, C.V., Riggs, P.D., Grandea III, A.G., Slatko, B.E., Moran, L.S., Tagliamonte, J.A., McReynolds, L.A. and di Guan C. (1988) *Gene* **74**, 365–373.
27 Riggs, P.D. (1992) In Asubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (ed.), *Current Protocols in Molecular Biology.* Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., New York, Vol. 2, pp. 16.6.1–16.6.14.
28 Smith, D.B. and Johnson, K.S. (1988) *Gene* **67**, 31–40.
29 Kammann, M., Schalk, H.-J., Matzeit, V., Schaefer, S., Schell, J. and Gronenborn, B. (1991) *Virology* **184**, 786–790.
30 Koonin, E.V. and Ilyina, T.V. (1992) *J. Gen. Virol.* **73**, 2763–2766.
31 Ilyina, T.V. and Koonin, E.V. (1992) *Nucleic Acids Res.* **20**, 3279–3285.
32 Gorbalenya, A.E. and Koonin, E.V. (1989) *Nucleic Acids Res.* **17**, 8413–8440.
33 Gorbalenya, A.E., Koonin, E.V. and Wolf, Y.I. (1990) *FEBS Lett.* **262**, 145–148.
34 Rasooly, A., Wang, P-Z. and Novick, R.P. (1994) *EMBO J.* **13**, 5245–5251.
35 Argüello-Astorga, G.R., Guevara-González, R.G., Herrera-Estrella, L.R. and Rivera-Bustamante, R.F. (1994) *Virology* **203**, 90–100.
36 Argüello-Astorga, G.R., Herrera-Estrella, L.R. and Rivera-Bustamante, R.F. (1994) *Plant Mol. Biol.* **26**, 553–556.
37 Lessl, M. and Lanka, E. (1994) *Cell* **77**, 321–324.