Linear Chromosome-generating System of *Agrobacterium tumefaciens* C58

**PROTELOMERASE GENERATES AND PROTECTS HAIRPIN ENDS**

Received for publication, April 4, 2012. Published, JBC Papers in Press, May 10, 2012, DOI 10.1074/jbc.M112.369488

Wai Mun Huang, Jeanne DaGloria, Heather Fox, Qiurong Ruan, John Tillou, Ke Shi, Hideki Aihara, John Aron, and Sherwood Casjens

From the Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, Utah 84112 and the Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455

**Background:** Protelomerase is an enzyme that generates closed hairpin ends in bacterial linear chromosomes.

**Results:** *Atu2523* encodes the agrobacterial protelomerase that generates its telomeres.

**Conclusion:** Agrobacterial protelomerase is the most compact enzyme of its kind that can uniquely both form and bind hairpin telomeres.

**Significance:** The studies of the reaction mechanism is crucial in understanding why and how and the prevalence of the existence of linear chromosome in bacteria.

*Agrobacterium tumefaciens* C58, the pathogenic bacteria that causes crown gall disease in plants, harbors one circular and one linear chromosome and two circular plasmids. The telomeres of its unusual linear chromosome are covalently closed hairpins. The circular and linear chromosomes co-segregate and are stably maintained in the organism. We have determined the sequence of the two ends of the linear chromosome thus completing the previously published genome sequence of *A. tumefaciens* C58. We found that the telomeres carry nearly identical 25-bp sequences at the hairpin ends that are related by dyad symmetry. We further showed that its *Atu2523* gene encodes a protelomerase (resolvase) and that the purified enzyme can generate the linear chromosomal closed hairpin ends in a sequence-specific manner. *Agrobacterium* protelomerase, whose presence is apparently limited to biovar 1 strains, acts via a cleavage-and-religation mechanism by making a pair of transient staggered nicks invariably at 6-bp spacing as the reaction intermediate. The enzyme can be significantly shortened at both the N and C termini and still maintain its enzymatic activity. Although the full-length enzyme can uniquely bind to its product telomeres, the N-terminal truncations cannot. The target site can also be shortened from the native 50-bp inverted repeat to 26 bp; thus, the *Agrobacterium* hairpin-generating system represents the most compact activity of all hairpin linear chromosome- and plasmid-generating systems to date. The biochemical analyses of the protelomerase reactions further revealed that the tip of the hairpin telomere may be unusually polymorphically capable of accommodating any nucleotide.

---

* This work was supported, in whole or in part, by National Institutes of Health Grants GM095558 (to H. A.) and AI49003 (to S. R. C.). This work was also supported by National Science Foundation Grant MCB-0213124 (to W. M. H.).

* This work was selected as a Paper of the Week.

The atomic coordinates and structure factors (codes 4F41 and 4F43) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

To whom correspondence should be addressed: Dept. of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT 84112. Tel.: 801-581-8489; Fax: 801-581-3607; E-mail: waimun.huang@path.utah.edu.

Bacterial chromosomes and plasmids are usually circular. In some cases, linear chromosomes or linear plasmids with either covalent protein-bound or closed hairpin ends are stably maintained as an alternative genome configuration. To date, the bacteria known to harbor hairpin-ended linear chromosomes include only the large chromosome of all studied *Borrelia* species and one of the two chromosomes of *Agrobacterium tumefaciens* C58. *Borrelia* species cause diseases in mammals, including Lyme disease and relapsing fever in humans, and are unique among spirochetes in their chromosome architecture (1). Many strains of the genus *Agrobacterium* within the α-Proteobacteria cause the tumorous crown gall disease on a variety of plants and have been manipulated into an important tool for the genetic engineering of plants. Linear plasmids are somewhat more widely distributed. *Borrelia* spirochetes harbor, in addition to their linear chromosome, numerous linear and circular plasmids (2). Furthermore, in the γ-Proteobacteria, three linear hairpin-ended plasmid systems have been characterized as follows: the N15 of *Escherichia coli*, the δKO2 of *Klebsiella oxytoca*, and the PY54 of *Yersinia enterocolitica* (3–5). These linear plasmids are the nonintegrated prophages of the corresponding bacteriophages. The phage enzymes responsible for generating closed hairpin-ended plasmids are called protelomerases (6).

The protelomerases from phages δKO2 and N15 have been extensively characterized biochemically. They use a topoisomerase IB and tyrosine recombinase (Y-recombinase) type of cutting-and-rejoining mechanism to generate DNA hairpin ends at a specific DNA site called *tel* (7). Further examination of the mechanism showed that the protelomerases generate a pair of transient staggered cleavages 6 bp apart on the two strands of their specific target DNA sequence to form a 3’-covalent DNA-protein intermediate and a 5’-OH at each opening. After exchange of strand partners, the cleaved openings are re-ligated to generated closed hairpin ends as a result of two transesterification reactions. Although these enzymes use a tyrosine as the active site residue, they form a subfamily of Y-recombinases in that they act on a single duplex DNA site, rather than two target
Protelomerase of A. tumefaciens C58

sites that are typically utilized by the canonical Y-recombinases to effect DNA insertion or excision (i.e. typical Y-recombinases do not make double strand cuts on one target site, whereas protelomerases make double-stranded staggered cuts on the target DNA). Protelomerases are simple enzymes that utilize sequence-specific target sites without the need for cofactors (7). The protein responsible for generating the hairpin ends in the Borrelia burgdorferi linear plasmids has also been characterized. Like the phage enzymes, it generates a pair of 6-bp staggered cuts with 5′-protrusions and protein-linked intermediates at each of the openings that have 3′-phosphoryl-covalent attachment to the active site tyrosine and free 5′-OH (8, 9). Interestingly, the responsible protein, called ResT, is encoded on a circular plasmid of the bacterium (2). Whether the linear plasmids found in Borrelia had their origin in phages is unknown.

In addition to the above mentioned closed hairpin end generating systems of linear plasmids, protelomerase-like genes have also been identified based on amino acid sequence alignments in a number of marine Vibrio phages, including those from Vibrio harveyi phage VHML (10), Vibrio paraaemoltyicus phage V882 (11) and phage VP58.5 (12), and Halomonas aquamarina phage φHAP-1 (13). Furthermore, protelomerase-like proteins harboring the signature amino acids have also been identified in the two sequenced eukaryotic filamentous brown algal viruses Ectocarpus siliculosus EsV-1 and the Feldmannia irregularis FirrV (14, 15) and in a Cocolithovirus of the marine calcifying microalga Emiliania huxleyi (16). Whether these phages and viruses exist intracellularly as linear plasmids in the infected cells remains to be further explored. Nonetheless, this suggests that this new class of enzymes may have a wider distribution than previously recognized and may even be present in eukaryotes.

As a first step toward understanding the physiological significance of having closed linear chromosomes, we sought to examine the linear chromosome-generating system in A. tumefaciens C58. The type strain C58 (a representative of biovar 1 of the genus Agrobacterium (17)) is genetically tractable and has one circular and one linear chromosome but no linear plasmids (it does have the two circular T-plasmids that are responsible for the pathogenicity of the bacteria). The terminal fragments of the linear chromosome have rapid snapback properties upon heat denaturation (18). This observation strongly suggests that the linear chromosomal ends are closed hairpins, similar to those described above. A. tumefaciens C58 must therefore maintain both circular and linear chromosome resolving machineries. Amino acid similarity searches based on the phage-encoded protelomerase enzymes identified open reading frame Atu2523 on the circular chromosome from the two published sequences (18, 19); the nucleotide differences in the two published sequences have since been resolved and consolidated and are jointly listed under GenBank™ accession number AE008686. In the Borrelia system, which harbors a single linear chromosome and numerous smaller linear plasmids, the unique single copy gene for the protelomerase/resolving enzyme that is used to generate these different but related linear ends must possess somewhat relaxed nucleotide specificity (1, 2); however, the agrobacterial enzyme presumably has a more stringent specificity in its DNA target recognition because it acts exclusively on the two ends of its only linear chromosome, and no linear plasmids were found in the Agrobacterium genome. It should be noted that in systems where the linear plasmids exist as nonintegrated prophages, the target site for the phage protelomerase is invariably located upstream of the encoded gene as a full target of inverted repeat sequences. As such, the duplex target site is readily obtainable as the joined ends of the linear plasmid that reside in the interior of the phage genome. However, in linear replicons that are not phage-associated, such as the bacterial linear chromosome, the hairpin ends constitute only half of a target site. In this report, we complete the determination of the sequences of the linear chromosome to its telomeres and provide biochemical evidence that the protein encoded by Atu2523 is the agrobacterial protelomerase capable of generating closed hairpin ends identical to those of the linear chromosome of the organism. We also characterize the requirements of the enzyme and the minimal target site of the hairpin end generating system. Our results show that the Agrobacterium protelomerase also generates a pair of 6-bp staggered cleavages as an intermediate in the reaction confirming the universality of this reaction mechanism and that the Agrobacterium enzyme uniquely binds to and hence likely protects the hairpin ends.

**EXPERIMENTAL PROCEDURES**

**Plasmid Constructions**—Open reading frame Atu2523 was cloned into pET15b (Novagen) via the Ndel and BamHI sites in the T7 promoter-driven system to express the N-terminal His-tagged full-length TelA protein. The coding sequence was amplified by PCR using the following primer pair: AG-1 (5′-GAGCCGCC-ATATGCTGCGCCGAAAAACGAAAAACAA-3′) as upstream primer and AG-2 (5′-AGGGATCTTATCTTCTCGGGGC-ACGGGCGCCATCG-3′) as downstream primer, where the underlined nucleotides are the restriction sites used for cloning. The N-terminal deletion protein TelA-N11 was similarly constructed as the full-length protein, except primer AG-3 (5′-AACAAAAATATGCTCTGAAACGTCGATCGATA-3′) that starts the coding information beyond the Ndel site from residue 12 was used as the upstream primer. Subsequent systematic N-terminal deletions were constructed using a similar scheme of using as the first amino acid a residue beyond the intended truncation as the upstream primer, and again coupled with AG-2 as the downstream primer in the PCR amplification reaction for the construction. The C-terminal deletion TelA-C was constructed using primer AG-1 with primer AG-4 (5′-CGC- AACCTTTATATTTTGCGACATAGTCCCGTCTAG-3′), which ends at residue 432 of the coding sequence. In the latter case, the Ndel and HindIII cloning sites of pET15b were used. Mutant protein TelA/Y405F was constructed using a 34-mer complementary primer pair 5′-TGCAAAAGCTGTT-CGCTGATCGATCGATCG-3′ and 5′-AGCGTATAGGTCATGAAAGGCAGCGCATTTTCG-3′ carrying a Tyr to Phe substitution at residue 405 (the codon for the substituted residue is underlined), using the QuickChange II site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instruction. Similar site-directed mutagenesis was used to generate mutant proteins in the catalytic regions of the TelA.
proteins, namely R255A, K286N, Y363E, R366A, and H394A. The 50-bp synthetic target site and its truncation variants (described in Fig. 2) were cloned into a pSK plasmid (Stratagene) via HindIII and BamHI sites using synthetic oligonucleotides harboring the described sequences with the added restriction sites for cloning purposes. Similarly, substrate deletions (described in Fig. 4) as well as other substrate variants were constructed using symmetrically shorter oligonucleotide pairs or variant oligonucleotide pairs as target for the insertion into the pSK plasmid (Stratagene). All plasmid constructs were confirmed by DNA sequencing.

**Determination of Telomere Sequences**—Total genomic DNA from *A. tumefaciens* C58 was digested with either NdeI or MluI (both restriction enzymes generate sticky ends). Size fractions containing the 2-kbp NdeI fragment and a 1.8-kbp MluI fragment were isolated from agarose gels. The partially purified fragments were treated with S1 nuclease (room temperature, 5 min according to the manufacturer’s instruction) and ligated with T4 ligase to a duplex linker that was formed by a pair of complementary synthetic oligonucleotides (5'-phosphorylated LK1, 22-mer, 5'-P-CCCCATAGTGAGCTTGATAC-3' and LK-2, 20-mer, 5'-TAATACGACTCATATAGGG-3'). The duplex-linker thus formed has a blunt 5'-end and a 2-bp 3'-protrusion (underlined in the LK-1 sequence). This design ensured that the blunt end of the duplex linker was uniquely ligated to the blunt end generated by the S1 nuclease in the end fragments and not ligated to other linkers. The ligated products were then used as templates for PCRs. For the right-end NdeI fragment, an upstream primer (5'-TTTGTGCAGCCTTGGATAGG-3') designed from the known linear chromosome sequence and LK-2 oligonucleotide serving as 3' primer generated a PCR fragment of about 500 bp. Similarly, using the left-end MluI fragment, an upstream primer (5'-GGAAAGG-TGTGTCAATGG-3') and the LK-2 oligonucleotide generated a PCR fragment of 550 bp. These PCR fragments were directly inserted into the pCR-4-TOPO cloning system (Invitrogen) according to the manufacturer’s instructions. The resulting plasmid clones were sequenced using vector-derived primers to sequence across the entire inserts. The inserted sequences were compiled and compared with the published linear chromosome sequence.

**Purification and Assays of Protelomerase—Agrobacterium* Protelomerase (TelA) was purified as an N-terminal His<sub>6</sub> fusion protein from *E. coli* BL21 (DE3) cells (Novagen) harboring the expression plasmid pET/TelA essentially as described previously for the purification of protelomerase from *A. tumefaciens* (7). In brief, cultures were grown in LB medium (5 g of NaCl, 10 g of bacto-tryptone, 5 g of yeast extract/liter) supplemented with 130 μg ml<sup>-1</sup> ampicillin at 37 °C until the A<sub>600</sub> reached about 0.8. Induction with isopropyl 1-thio-β-D-galactopyranoside was done at 0.8 mM at room temperature for 15 h. Cells from the induced culture were harvested by centrifugation and stored frozen at −80 °C until use. Cells expressing TelA (or its derivatives) were lysed by lysozyme (Sigma, used at 0.6 mg ml<sup>-1</sup>) and briefly sonicated in the presence of 25% sucrose, 50 mM Tris-HCl, pH 7.5, 25 mM 2-mercaptoethanol, and protease inhibitors 0.13 mM benzamidine and 0.6 mM phenylmethylsulfonyl fluoride (PMSF). The crude extract was further treated with 150 μg ml<sup>-1</sup> RNase and 5 μg ml<sup>-1</sup> DNase, in 5 mM Mg<sub>2</sub>SO<sub>4</sub> at 0 °C for 1 h, and finally made into 2 mM NaCl and 0.6% Triton (Roche Applied Science) to complete the extraction. Soluble protein, as supernatant after centrifugation at 25,000 rpm at 4 °C in an SW40 rotor (Beckman Coulter), was loaded onto a 15-ml column of nickel-nitrolitriacetic acid-agarose (Qiagen) equilibrated with buffer N (10% glycerol, 50 mM Tris-Cl, pH 7.5, 10 mM 2-mercaptoethanol, 0.1 mM benzamidine). The column was washed extensively with buffer N containing 0.5 mM NaCl, and the protein was eluted with buffer N plus 0.5 mM NaCl and 0.8 mM imidazole. Fractions containing TelA were pooled and dialyzed against buffer N containing 0.5 mM NaCl to remove imidazole and eventually dialyzed against buffer S (50% glycerol, 50 mM Tris-Cl, pH 7.5, 0.5 mM NaCl, 15 mM 2-mercaptoethanol, and 0.1 mM benzamidine) and 0.5 mM PMSF for storage at −20 °C. Aliquots of about 10 μg were further purified by gel filtration on a Superdex 200 column (HiLoad 16/60, GE Healthcare) using FPLC at a flow rate of 1 ml min<sup>-1</sup> in buffer N containing 0.5 mM NaCl.

Protelomerase was assayed in 15-μl reactions containing 20 mM Tris-Cl, pH 7.5, 50 mM potassium glutamate, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5 mg of supercoiled or linearized plasmid substrate DNA, and 0.2–2 pmol of enzyme. Reactions were incubated at 30 °C for 30 min and stopped by the addition of SDS to 1% final concentration. Products were analyzed on 1% agarose electrophoresis gels and visualized by ethidium bromide staining. When oligonucleotides were used as substrates, the same reaction conditions were used except the substrate concentration was present at 15 pmol per reaction, and the enzyme concentration was increased to 40–100 pmol per reaction. Oligonucleotide reaction products were analyzed on 12% polyacrylamide gels using a vertical gel electrophoresis system.

**Oligonucleotide Substrates Used for the Determination of Cutting-Rejoining Sites**—These substrates were prepared by mixing complementary synthetic oligonucleotides at 10 pmol ml<sup>-1</sup> in 50 mM Tris-Cl, pH 7.5, and 5 mM MgCl<sub>2</sub>. The mixture was first heated to 90 °C for 3 min followed by slow cooling at 0.01 °C/s in a thermal cycler that allowed controlled annealing to form duplexes. When appropriate, oligonucleotides were labeled at the 5'-end with T4 polynucleotide kinase (New England Biolabs) and [γ-<sup>32</sup>P]ATP before annealing to provide a marker for the products. To prepare a duplex oligonucleotide substrate with an internal phosphate label, we used two halves of one strand of the full-length substrate, with the 5'-end of the designated internal break being first labeled with <sup>32</sup>P. The two halves were then mixed with their complementary full-length oligonucleotide to form a duplex with an internal break. The annealed products were then ligated with T4 DNA ligase to reconstitute the full-length duplex oligonucleotide substrate, and an internal <sup>32</sup>P was thus incorporated at the designated position. We define nucleotide numbering from left to right regardless of whether it is the top or bottom strand (as shown in Fig. 3), and because polynucleotide kinase adds a phosphate exclusively to the 5'-end of an oligonucleotide, the 3' or 5' orientation of the labeled internal phosphate at the top strand is different from that of the bottom strand. For example, to form substrate with label at 22T on the top strand, the two halves of
Proteolomerase of A. tumefaciens C58

the top strand consist of oligonucleotide 1–22 and oligonucleotide 23–60. The 5’-32P-phosphate was added to oligonucleotide 23–60. Thus, the 32P of 22T is added at the 3’-side of nucleotide number 22T. In contrast, when the bottom strand was internally labeled, the 32P was added at the 5’-side of the designated nucleotide. For example, in forming 28B substrate, the two halves of the bottom strand are formed by oligonucleotides 1–28 and 29–60 (numbering is again from left to right). The 5’-32P is added to the 5’-side of the left oligonucleotide 1–28 on the bottom strand before ligation to reconstitute the duplex substrate; hence, the 32P in this case was at the 5’-side of nucleotide 28B. The relevant phosphates are shown in Fig. 3A for clarity.

Other Methods—Electrospray ionization mass spectrometry analysis was done at the University of Utah Core Facilities. Neutral agarose gel (1%) was carried out in 50 mM NaOH and 1 mM EDTA running in TBE buffer pH 8.5 (0.135M Tris-base, 45 mM boric acid, and 2.5 mM EDTA-Na2). Alkaline agarose gel (1%) was carried out in 50 mM NaOH and 1 mM EDTA running at 40 V for 4 h at 4°C. Reaction products using oligonucleotide as substrates were analyzed in 12% polyacrylamide gel running in TBE buffer pH 8.5 (0.135 M Tris-base, 45 mM boric acid, and 2.5 mM EDTA-Na2).

RESULTS AND DISCUSSION

Telomeres of A. tumefaciens C58—The two published genomic sequences of A. tumefaciens C58 did not reach the telomeric ends of the linear chromosome, as the closed terminal fragments are expected to be absent from the collection of clones used in the sequencing projects. This is because the closed ends at the termini are not free to ligate to the vector during the library construction; hence, their sequences are excluded from the conventional genomic sequence. We set out to specifically determine the terminal sequences of the linear chromosome using a different strategy by first identifying the two terminal fragments of the linear chromosome and then opening the closed ends by single strand-specific nuclease, thereby making them accessible to ligation. This approach has been used successfully to determine the telomeres of the B. burgdorferi linear chromosome and some linear plasmids (1).

As was shown previously (18), the left terminal 2-kb Ndel DNA fragment and the right 1.8-kb MluI fragment both exhibit “snap-back” (rapid renaturation) properties when these fragments were heat-denatured followed by rapid cooling. These fragments were isolated from the total genomic DNA digests from agarose electrophoresis gels, followed by S1 nuclease treatment to nick open the hairpin ends, and these converted open ends were then blunt end-ligated to a double-stranded synthetic oligonucleotide linker. Using upstream genome sequence and the linker sequence as PCR primer targets, the end fragments were amplified and cloned into the pCR4-TOPO vector (Invitrogen) and sequenced (see under “Experimental Procedures” for details). We sequenced six clones from each of the Ndel and MluI fragment inserts, and the sequences were in perfect agreement with the published sequences where they overlapped and were followed by the extending new sequences, and the position of the linker sequence identified the tip of the linear chromosome.

When the targeted independently determined right and left telomeric sequences reported here were compared with the published genomic sequences, we found that the University of Washington sequence (accession number AE008686 (19)) was missing the last eight nucleotides at the left end and missing nine nucleotides at the right end. The Cereon sequence (accession number AE007870 (18)) ended about 700 bp from the right end. Interestingly, its left end has 19 nucleotides beyond the terminal sequence determined here. Eleven of these “extra” 19 bp were the complement of the last 11 bp of the left telomere determined here, and the last 8 bp were novel. This alignment supports the notion that the Cereon genomic end sequence was generated by the cloning of an extended or unfolded hairpin; for example, a nick might have occurred at the bottom strand of the left end (as depicted in Fig. 1A), and the hairpin was unfolded and inadvertently filled in by 3’-extension. The resulting fragment was then cloned into the library. We believe the last eight nucleotides of the published left end of the Cereon sequence (18) is not the Agrobacterium sequence. The fact that this inverted repeat sequence, which we determined to be at the left telomeric end, was found among the random clones constructed for the sequencing project further supports the contention that the left end was inadvertently extended and behaved like a “clonable” fragment where both ends of the fragment are ligatable to the vector.

When the left and right telomeres are incorporated into the context of the linear chromosome and aligned, as shown in Fig. 1, they form inverted repeats of each other for 25 bp from the ends (the boxed regions). Three exceptions are noted as follows: at the extreme right end, one nucleotide appears to be missing; we did not find the expected “A” at the extreme right end of the top strand (Fig. 1A, it is added here in gray type for discussion and clarification purposes, see below), and two nonsymmetric bps (Fig. 1A, not boxed) break up the long sequence into two shorter perfect inverted repeats. Since terminal inverted repeats are found at the ends of all other characterized linear plasmid hairpin ends (20). We reasoned that the right telomere should contain a missing A-T loop as described at the right tip, and its absence from the sequence determined here is most likely due to the S1 nuclease treatment inadvertently removing 1 bp during the process of opening the hairpin loop. As we will show below, by including the missing nucleotide at the right end, we successfully use the information either as the joined left and right telomere as well as the replicative ends of left or right termini in conjunction with the proteolomerase enzyme to generate these telomeres. Thus, we have included the A nucleotide at the right telomere as shown in Fig. 1A. As will be examined in further detail below, other options are also possible based on the biochemical analyses such that the missing nucleotide could be any one of the four bases. We also note that near the telomeres of the linear chromosome, the gene density is lower than that of the whole genome, and there are several transposable genes (and fragments thereof) near the telomeres. This could indicate recent mobile element activity in these regions. The most terminal genes with predicted functions are two putative acetylase genes (Atu4896 and Atu3009) 811 and 942 bp from the right and left ends, respectively. Functional genes have been found within several hundred base pairs of hairpin telom-
eres on other replicons with this type of telomere (3, 5, 21). By including the two telomeric sequences determined here, and resolving the differences of the two published sequences, the consolidated complete sequence of the <i>A. tumefaciens</i> C58 linear chromosome is 2,075,577 bp in length (GenBank™ accession number AE007870).

<i>Atu2523</i> is the Agrobacterium Protelomerase—BLAST searches (22) using the amino acid sequences of bacterial phage protelomerases as the query identified Atu2523 as the protelomerase-like gene encoding 442 residues at the physical map location of 2494 kbp on the circular chromosome of the Agrobacterium <i>C58</i> (see Fig. 1B for an abbreviated map). To confirm its identity, we cloned <i>Atu2523</i> into an expression vector and purified the encoded protein (see “Experimental Procedures” for detail). The resulting protein has an SDS-gel electrophoresis molecular mass of ~51 kDa, consistent with the mass calculated from the open reading frame. The purified protein was also subjected to a more rigorous equilibrium sedimentation analysis and was found to have a molecular mass of 49 ± 4 kDa consistent with a monomeric structure in solution (data not shown). Furthermore, when the preparation of the purified gene <i>Atu2523</i> protein was subjected to mass spectrometry analysis, it yielded a molecular mass of 52353.7 Da, in complete agreement with the calculated molecular mass of 52353 expected from its sequence with the initiating methionine.

The paradigm established in the phage <i>φKO2</i>, N15, and <i>Borrelia</i> linear replicon systems suggests that hairpin ends are generated from an inverted repeat formed by two telomeric sequences (4, 7, 9). Hence, a synthetic target of 50 bp was designed by joining the 25 bp from the left and right Agrobacterium C58 linear chromosome telomeres. (An A nucleotide was added to the right telomere to complete the inverted repeat, as described above.) The 50-bp inverted repeat sequence was cloned into a pSK vector (Stratagene), and the resultant 2.9-kbp plasmid was tested for its ability to serve as the substrate for the protein encoded by <i>Atu2523</i>. Fig. 2 shows that both the supercoiled and the linearized (by restriction enzyme AlwNI) forms of the 2.9-kbp plasmid carrying the 50-bp target do serve as substrate for the gene <i>Atu2523</i> protein, although the linearized duplex form is a much more proficient substrate. Fig. 2B shows that incubation of the gene <i>Atu2523</i> protein with the supercoiled circular substrate yielded a duplex sized DNA of 2.9 kb, which has the same migration if the plasmid was linearized by the restriction enzyme AlwNI (Fig. 2B, lanes 2 and 3 in the neutral agarose gel). When the 2.9-kbp product of the supercoiled circular substrate was analyzed on an alkaline gel, it migrated at a molecular size of about 5.8 kbp, twice the length of that analyzed under neutral conditions (Fig. 2B, lane 2). Similarly, the gene <i>Atu2523</i> protein converted the 2.9-kb linearized target (by AlwNI) into two fragments of 2.1 and 0.8 kb showing that the protein acted at the cloned target site; furthermore, both these product fragments were shown to be twice as long when analyzed under alkaline conditions (Fig. 2B, compare lane 4 in neutral and alkaline gels). The difference in product lengths when they are analyzed under neutral or alkaline conditions can only be explained if the new ends (not generated from the AlwNI restriction cut) were physically linked together. Hence, we conclude that the gene <i>Atu2523</i> protein generates closed hairpin ends within the cloned target sequence. Because this enzymatic activity describes the unique protelomerase that generates the closed hairpin telomeric ends.
of the _Agrobacterium_ C58 linear chromosome, we conclude that _Atu2523_ encodes the _Agrobacterium_ protelomerase TelA in accordance with the nomenclature of other protelomerases.

TelA Also Uses a Cutting-Religation Mechanism to Generate Hairpin Ends—Protelomerases can be thought of as specialized tyrosine recombinases that generate intra-molecular reaction products via a concerted cutting-and-religation type mechanism to allow each of the cleaved strands of the duplex to loop back to re-ligate with a new partner to form two closed hairpin ends (7). To elucidate the precise location within the 50-bp dyad symmetrical target sequence where TelA acts to generate the two hairpin ends, we used duplex oligonucleotide substrates because they can be easily modified. A 10-bp unrelated sequence (green color in Fig. 3A) was added as an extension to the right side of the 50-bp target sequence, such that it would be possible to easily recognize products that originated from the left or the right side of the substrate molecule. Fig. 3B shows that when the 60-bp oligonucleotide substrate was incubated with TelA, two products were generated as follows: a 25-bp product originated from the left side, and a 35-bp product originated from the right side, which includes the 10-bp extension.

To determine the precise nucleotide location where TelA cleavage-and-ligation occurs, we used a method of differential labeling of internal phosphates using 

\[ ^{32}P \] ATP to mark the various phosphates on the oligonucleotide in a panel of substrates in which a single internal phosphate at various locations along the 60-bp duplex was labeled with \[ ^{32}P \]. Such a panel was prepared by using duplex oligonucleotides that were formed initially by three synthetic oligonucleotides such that a nick (at a different position in each substrate) in the top or bottom strand was present. The 5′ openings at such nicks were labeled with poly-nucleotide kinase and [\( \gamma^{32}P \)]ATP, followed by sealing with T4 DNA ligase to reconstitute the 60-bp full-length duplex substrates; the backbone phosphates at various positions are thus labeled. Using such a panel of substrates, the exact cleavage location is the position where neighboring phosphates end up in different sized products. A phosphate label at the 5′-end of the top strand of the duplex substrate (Fig. 3C, lane T) and a 5′-phosphate label on the bottom strand (Fig. 3C, lane B) serve to mark positions of the 25-bp (left) and 35-bp (right) products, respectively. Fig. 3C also shows that on the top strand, the phosphate 3′ to nucleotide 22 (lane 22T) remains with the left side 25-bp product, whereas the phosphate 3′ to nucleotide 23 (lane 23T) goes with the right side 35-bp product. Similarly, for internally single labeled substrates on the bottom strand, phosphate labeled on the 5′-side of nucleotide 28 (numbered 3′ to 5′ in the bottom strand) remains with the 35-bp right side product, whereas phosphate labeled at the 5′ side of nucleotide 27 remains with the 25-bp left side product (Fig. 3D, lanes 27B and 28B). This set of experiments clearly demonstrates that cleavage occurs between nucleotides 22 and 23 on the top strand and between nucleotides 28 and 29 on the bottom strand, such that a staggered pair of nicks generates two halves, with 6-base 5′-overhangs as the intermediate in the reaction. The pattern of phosphate transfer further points out that the cleavages leave 5′-OH and 3′-phosphoryl openings. This confirms that the cleavage-ligation reaction of protelomerases, regardless of size, invariably makes a pair of 6-bp staggered cleavages at the center of the

---

**FIGURE 2. Enzymatic activity of protelomerase TelA.** A plasmid harboring a 50-bp target site consisting of 25 bp from each end of the telomeric sequence, drawn here as two pairs of inward pointing arrows (inner and outer arrows indicate perfect inverted repeats), was used as the substrate in the TelA assay. Lanes 1 and 2, supercoiled substrate without or with added TelA, respectively, were used; lanes 3 and 4, linearized substrate (plasmid DNA cut with AlwNI) without or with enzyme was used. The reaction products were analyzed on two 1% agarose electrophoresis gels running under native or alkaline conditions. M is the marker lane whose sizes in kbp are labeled on the side of the gels. Hairpin products of the reactions using circular or linearized substrates are labeled with black diamonds.

**FIGURE 3. Determination of the cutting-joining location of TelA.** A sequence of a 60-bp double-stranded oligonucleotide substrate consisting of the 50-bp target site plus a 10-bp extension (nucleotides 51–60 in green) added to the right end. The addition of the 10-bp extension distinguishes the left product (25 bp) from the right product (35 bp) when they are analyzed on a 12% polyacrylamide electrophoresis gel. The central 6 bp where hairpin formation occurs is in lowercase. Relevant phosphate moieties that were differentially labeled with \[ ^{32}P \] are depicted by a lowercase p, where each p was singly labeled in separate experiments. & ethidium bromide-stained 12% PAGE showing the activity of TelA on the 60-bp oligonucleotide substrate yielding two products of 25 and 35 bp. C and D, autoradiograms of a 12% polyacrylamide gel showing the products of the 60-bp substrates whose internal phosphates were differentially labeled by \[ ^{32}P \] on the top strand or the bottom strand respectively (see “Experimental Procedures” for the preparation of internally labeled oligonucleotide substrates). The positions of the internal label are given above the gel. T indicates that the 5′-end of the top strand is labeled to mark the position of the 25-bp product. B indicates that the 5′-end at the bottom strand is labeled to mark the position of the 35-bp product. Nucleotide numbering is from left to right on both strands. Substrates with internally labeled phosphate on the top strand are labeled 22T and 23T in C. Similarly, substrates with internally labeled phosphate on the bottom strand are labeled 28B and 27B in D; together they show that a staggered 6-bp cut (marked by filled triangles in A) form the intermediate of the reaction (see text for detail). + + + indicates that no protein was added, and + + indicates that TelA was present.
dyad symmetry of their respective target sites to generate closed hairpin ends (6, 7).

Minimal Target Size of TelA Protelomerase—Initially, we used the 50-bp synthetic target of the TelA protelomerase derived from the 25 bp at the right and left end of the A. tumefaciens C58 linear chromosome (called R-L target). Sequences beyond the 25-bp telomeric ends are divergent (Fig. 1A); the target sequence can be described as two short inverted repeats with a 3-bp spacer (Fig. 4A, lowercase type). For ease of cloning the target sites, we examined the “imperfect” inverted repeat as described here first. To determine whether the entire 50-bp conserved sequence is required for protelomerase action, we generated a set of systematically deleted target sites by removing bases symmetrically from both sides of the 50-bp target and cloning them into the pSK vector as was done for the parent 50-bp target. Fig. 4 shows that within the 50-bp target, substrate deletions 1–6 are all active. This shows that the central 26 bp covering only the internal inverted repeat (10 bp flanking each side of the central 6-bp cleavage locations) is sufficient for TelA protelomerase hairpin activity. Further removal of additional flanking bases from both sides (24-bp target or less) renders the target inactive for TelA hairpin reaction (Fig. 4, substrate 7 and 8). This 26-bp minimal target site is the smallest target size among the protelomerase proteins characterized to date. As noted, during replication of the linear chromosome, as the replication machinery traverses the left or the right telomeres around the hairpin ends, L-L’ or R-R’ (for the left end and its complement or the right end and its complement) duplex inverted repeat targets are formed. The three types of synthetic targets with the inverted repeats of the combined right and left telomeric sequence (R-L) or their replicated individual ends L-L’ and R-R’ worked equally well as substrates, see below.

Other Options for the Terminal Nucleotide at the Right Telomere—As described earlier, the terminal nucleotide at the right telomere was not recovered despite repeated attempts to both make different constructs and to sequence more clones from the targeted constructs. Yet there is no doubt that if this one nucleotide is not included at the right end, the target is not resolved by the protelomerase (see below). We therefore entertained the possibility that the terminal nucleotide is not “paired” with its “complement” (as in the cases of reaction type VII and VIII described in Fig. 5 called “flip-flop” ends after Baroudy et al. (23)). If this is true, our scheme for generating telomere sequence to the tip will not uncover the last nucleotide. This is because in the current schemes of telomere determination a nicking nuclease is used to open the hairpin loop of the telomere. If the last nucleotide in the loop was not paired, the nuclease would have removed it upon opening the loop, and the information for the last nucleotide will be lost regardless of whether the sequence of the targeted fragment was determined by first cloning it into special vectors or by direct sequencing of the genomic DNA from upstream unique primers (24). We therefore used a set of synthetic targets to test if any nucleotide as well as an unpaired configuration was capable of supporting the TelA protelomerase action to generate hairpin telomeres. Our approach was to design synthetic substrates based on telomeres whose last nucleotide may be any one of the four bases as listed in Fig. 5A.
### Protelomerase of A. tumefaciens C58

#### A

| Reaction type | Telomere | Replicated target | Reaction products |
|---------------|----------|-------------------|-------------------|
| I. AT tip     | 5'-TATCA -ATAGT | 5'-TATCA TGATA-\* ATAGT ACTAT- | 5'-TATCA -ATAGT ACTAT- | |
| II. 5 base loop | 5'-TATC -ATAGT | 5'-TATC -GATA- -ATAGT ACTAT- | No product | |
| III. 4 base loop | 5'-TATC -ATAGT | 5'-TATC -GATA- -CTAT- | No product | |
| IV. TA tip    | 5'-TACTCT -ATAGA \* | 5'-TACTCT -GATA- TCTAT- -ATAGA | |
| V. CG tip     | 5'-TATCC -ATAGG \* | 5'-TATCC -GATA- CCTAT- -ATAGG | |
| VI. GC tip    | 5'-TATCG -ATAGC \* | 5'-TATCG -GATA- GCTAT- -ATAGC | |
| VII. AA flip-flop | 5'-TATCA -ATAGA \* | 5'-TATCA -GATA- TCTAT- -ATAGA | |
| VIII. CC flip-flop | 5'-TATCC -ATAGG \* | 5'-TATCC -GATA- CCTAT- -ATAGG | |

#### B

![Image of gel electrophoresis](image)

- **Reaction type:**
  - I: - +
  - II: - +
  - III: - +
  - IV: + +
  - V: + +
  - VI: + +
  - VII: + +

#### C

![Image of gel electrophoresis](image)

- **Reaction type:**
  - I: +
  - II: +
  - III: +
  - IV: +
  - V: +

#### D

![Diagram of DNA structure](image)

- **TT-tip**
- **AA-tip**
protelomerase of A. tumefaciens C58

In A. tumefaciens, telomeres are the products of the reaction.

The N-terminal region of TelA (compare with the Borrelia protelomerase shown in Fig. 6A) was also investigated to deter-

at the tips of linear telomeres. In support of the stably maintained hairpin structure of this flip-flop-type telomere with the unpaired AA or TT at the tip, we have recently shown that such hairpins do bind TelA/R255A (data not shown), and in addition, we have obtained the x-ray structure of these complexes shown in Fig. 5D. These two types of flip-flop hairpin DNA adopt similar structures. Near the unpaired apex of the hairpin, the DNA strands of both structures have two alternate conformations. The dual positions of the nucleotides at the apex were also observed for the hairpin telomere from the native palindromic sequence bound to wild type TelA. The detailed telomeric hairpin DNA–TelA interactions by x-ray analyses will be described elsewhere. 2

N and C Termini of TelA Are Not Required for Enzymatic Activity—To date, numerous protelomerase-like proteins have been identified as being encoded by the genomes of β-Proteobacteria, spirochetes, cyanobacteria, bacterial phages, and eukaryotic algal viruses, although not all have been characterized biochemically for confirmation. These proteins range from 350 to 850 residues in length. The highest amino acid sequence conservation of these proteins resides in the central region where the active site tyrosine and the coordinating catalytic residues are located; a schematic alignment using the “catalytic core” region of about 200 residues as an anchor (shaded region) is shown in Fig. 6A. Based on protein alignment among protelomases and other tyrosine recombinases, it is predicted that in Agrobacterium TelA, Tyr-405 is the active site tyrosine, and the coordinating catalytic residues are Arg-255–Lys-286–Tyr-363–Arg-366–His-394 (the so-called pentad of tyrosine recombinases (25)). The alignment of the catalytic core region is supported by the genetic mutation studies in which amino acid substitutions of Y405F at the active site as well as those at R255A, K286N, Y363E, R366A, and H394A of the coordinating residues all render the protein inactive to form hairpins using the linearized substrate described above (data not shown). It is of interest to note that substitutions to Lys or His at residue Tyr-363, one of the coordinating residues, are totally permissible and that these are the residues found at the corresponding positions in TelK/Lys-380 from the phage fK02 enzyme (7) and His-315 from the enzyme of VHML phage (10). The Agrobacterium protein became inactive only when the residue is mutated to a negatively charged amino acid, Y363E.

The N-terminal region of TelA (compare with the Borrelia protelomerase shown in Fig. 6A) was also investigated to deter-

FIGURE 5. Effect of varying nucleotides at the tip of the right telomere of the linear chromosome. A, protelomerase reaction scheme. The columns describe the following: 1st column, the names of reaction types; 2nd column, varying nucleotides present at the right tip of the linear chromosome; 3rd column, the duplex substrates that would be generated upon replication, and 4th column, the hairpin products generated by TelA. The newly replicated strand is colored red, and the dot indicates the center of dyad symmetry. The cleavage sites are marked by triangles. Only the last five nucleotides of the right telomere are listed (see Fig. 4A for the context of the right telomere). Representatives of the actual experimental assays using either the cloned duplex replicated targets or the duplex oligonucleotides based on the replicated targets are given in the gels below. B, TelA activity on representative linearized cloned substrate (32 bp target) variants described in A. The products were analyzed on 1% agarose gel. C, TelA activity on representative target site variants using duplex oligonucleotide-based substrates. The 32-bp target plus 10 bp of unrelated sequence (similar to the scheme described in Fig. 3B) were used, except that the central 2 bp of target sequences were altered as described. Products of the oligonucleotide reactions were analyzed on 12% polyacrylamide gels. The reaction types (described in A) are listed above the gels. A — or — indicates the absence or presence of the enzyme TelA. S marks the position of the substrate, and P1 and P2 are the products of the reaction. D, crystal structure of the AA or TT tip hairpin DNA (of reaction type VII) bound to the TelA/R255A mutant. The DNA strands near the apex of the hairpin have two alternate conformations (depicted as two possible positions). The dual positions of the nucleotide at the apex were also observed for the hairpin telomere DNA from the natural palindromic sequence bound to wild type TelA (see footnote 2).

2 K. Shi, W. M. Huang, and H. Aihara, submitted for publication.
mine whether the entire length is required for activity. A systematic set of N-terminal truncations of TelA was constructed, each shorter than the previous one by about 15–20 residues. The truncations were partially purified and tested for their ability to generate hairpins using the linear substrate described in Fig. 2. Six N-terminal truncations were initially constructed which deleted from 11 to 72 residues without any ill effects in terms of protein stability or reaction kinetics (data not shown). Longer N-truncations became less stable with more breakdown and/or contaminants retained in similarly prepared protein samples as illustrated in the partially purified protein preparations shown in Fig. 6B where the native full-length protein A442, always appeared more homogeneous than the N-terminally truncated proteins. Nonetheless, the analysis showed that hairpin-generating activity is fully retained through protein N106, and N123 has about 50% of the activity, whereas N140 is inactive (Fig. 6C). This is in contrast to the other orthologous proteins as follows: TelK (from phage φKO2) (7) and the Borrelia protein ResT (9), where the entire N-terminal region is required for activity and any N-terminal truncations are not active.

Ten C-terminal residues of TelA can also be deleted without affecting the stability or activity of the protein in vitro (data not shown). The protein alignment based on the structural mod-
where the functioning TelA might effectively appear to behave like an N-terminal truncation (similar to N106). Efficient hairpin resolution then ensues as the complex acts in favor of duplex substrate resolution. After two newly formed hairpin telomeres are generated, TelA can again bind and protect the telomeres.

**Linear Chromosome and telA Gene Are Unique to the Biovar I Clade—Agrobacterium** strains fall into three biovars that differ in several physiological and biochemical characteristics as well as in their position within the 16 S rRNA gene tree for the greater Agrobacterium/Rhizobium clade (29). Prior to the complete sequencing of the *A. tumefaciens* C58 type strain, it had already been demonstrated that this strain in biovar I contains a circular chromosome I and a linear chromosome II (17). With the completion of the genomic sequences from a representative from each of the three biovars, namely that of *A. tumefaciens* C58 for biovar I, *Agrobacterium radiobacter* K84 for biovar II and *Agrobacterium vitis* S4 for biovar III, it is clear that multiple chromosomes and plasmids are present in different biovars, but a linear chromosome with the concomitant presence of a *telA* gene necessary for its resolution is found only in biovar I (30).

More recently, the genome sequence of a nonpathological *Rhizobium* strain formerly known as *Lupinus luteus* H13-3 was completed (31). It was found to harbor one circular chromosome, one linear chromosome, and one small circular plasmid. Sequence comparison showed that the circular chromosome of strain H13-3 bears 90% identity to the circular chromosome of *A. tumefaciens* C58. The linear chromosomes of the two strains show somewhat higher divergence and share only 75% similarity. Based on the similarity with strain C58, strain H13-3 is now renamed as *Agrobacterium* sp. H13-3 and becomes a member of the biovar I group (31). The presence of a linear chromosome can now be used as one of the criteria for *Agrobacterium* classification (30). *Agrobacterium* sp. H13-3 also encodes a TelA-like protein (AGROH133_08566) on its circular chromosome, and it shares 92% similarity with the protelomerase TelA (*Atu*2523) described here. Furthermore, it retains the gene synteny in the vicinity of the protelomerase gene. The complete sequence of *Agrobacterium* sp. H13-3 further shows that the two ends of its linear chromosome are inverted repeating sequences of each other (31). A more detailed DNA sequence alignment of the ends showed that the H13-3 right end and left end is 92.4 and 93.1% identical, respectively, to that of the corresponding region at the right end of the C58 chromosome described here (Fig. 8). Because the protelomerases from C58 and H13-3 are highly conserved (92% overall identity and are identical in the central catalytic core region, the shaded region depicted in Fig. 6A), we suspect that the target sites of H13-3 TelA are similar to those of C58 TelA. We had previously shown that TelK protelomerase (from *φKO2* of *Klebsiella*) and TelN protelomerase (from phage N15 of *E. coli*) shared 68% amino acid identity and their target sites are completely interchangeable (7). Hence, we believe that the target sites at the two telomeres of H13-3 linear chromosome are likely to be closely related to that of the C58 ends as the enzymes for their resolution are even more closely related. It is noted that the right telomere of C58 may be the more fluid of the two ends (as described here), and it is this end that may have been duplicated to form the similar telomeres of the H13-3 linear chromosome.
**Proteolomerase of A. tumefaciens C58**

Based on the detailed alignment, we concluded that the Agrobacterium sp. H13-3 sequencing project has not reached the end of the linear chromosome; and we predict that the linear chromosome may need to be extended 401 bp at the right end and 305 bp at the left end to reach the telomeres of its linear chromosome. More pertinent to the discussion here, it would be informative to identify the tip nucleotides of the telomeres of Agrobacterium sp. H13-3 linear chromosome.

In summary, A. tumefaciens C58 uses a simple and elegant way to protect and maintain its linear hairpin chromosome in coordination with its second and more conventional circular chromosome. Only two components are involved with the resolution of the telomeres as follows: a protelomerase protein and a specific telomeric sequence of at least 13 bp, which is replicated to form a target substrate of at least 26 bp of inverted repeating sequence. The same protelomerase that generates the hairpin also remains bound to and may be capable of protecting the hairpin. As is described in this report, the telomeric sequence of a bacterial linear chromosome is typically not revealed at the “end” of a sequencing project because of the hairpin nature of the telomere. Currently, there is no reliable method to predict the telomeric ends of a long linear chromosome. However, the responsible protelomerase especially in the catalytic domain is highly conserved. As a member of the tyrosine recombinase superfamily, specific protelomerase signature amino acids that are required to coordinate the active site tyrosine are readily recognized as unique residues with specific spacing among them (7, 25, 32). Hence, the presence of a protelomerase invariably predicts the presence of a linear chromosome or a linear plasmid in the organism. Based on the presence of protelomerase, we find that the occurrence of linear hairpin replicons is more widespread, especially in cyanobacteria (33), although none of the cyanobacterial systems have been examined in detail thus far. More recently, a protelomerase-like gene is found even in a bacterial endosymbiont Hamiltonella defensa (34), suggesting that linear hairpin chromosome may exist as an ancient chromosome configuration.

**Acknowledgments** — We thank Jennifer Ku for technical assistance in the early phase of this work and Chad Nelson (University of Utah Mass Spectrometry Core Facility) for mass analyses.

**REFERENCES**

1. Casjens, S., Murphy, M., Delange, M., Sampson, L., van Vught, R., and Huang, W. M. (1997) Telomeres of the linear chromosomes of Lyme disease spirochaetes. Nucleotide sequence and possible exchange with linear plasmid telomeres. *Mol. Microbiol.* **26**, 581–596

2. Casjens, S., Palmer, N., van Vught, R., Huang, W. M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R. J., Haft, D., Hickey, E., Gwinn, M., White, O., and Fraser, C. M. (2000) A bacterial genome in flux. The 12 linear and 9 circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.* **35**, 490–516

3. Casjens, S. R., Gilcreese, E. B., Huang, W. M., Bunny, K. L., Pedulla, M., Ford, M. E., Houtz, J. M., Hatfull, G. F., and Hendrix, R. W. (2004) The pKO2 linear plasmid prophage of Klebsiella oxytoca. *J. Bacteriol.* **186**, 1818–1832

4. Hertwig, S., Klein, L., Lurz, R., Langa, E., and Appel, B. (2003) PY54, a linear plasmid prophage of Versinia enteroxolitica with covalently closed ends. *Mol. Microbiol.* **48**, 989–1003

5. Ravin, V., Ravin, N., Casjens, S., Ford, M. E., Hatfull, G. F., and Hendrix, R. W. (2000) Genomic sequence and analysis of the atypical temperate bacteriophage N15. *J. Mol. Biol.* **299**, 53–73

6. Rybczynski, V. N., and Svarechovsky, A. N. (1999) The plasmid prophage N15. A linear DNA with covalently closed ends. *Mol. Microbiol.* **33**, 895–903

7. Huang, W. M., Joss, L., Hsieh, T., and Casjens, S. (2004) Protelomerase uses a topoisomerase IB/Y-recombinase type mechanism to generate DNA hairpin ends. *J. Mol. Biol.* **337**, 77–92

8. Chaconas, G., and Kobyrint, K. (2010) Structure, function, and evolution of linear replicons in *Borrelia*. *Annu. Rev. Microbiol.* **64**, 185–202

9. Kobyrint, K., and Chaconas, G. (2002) Rest, a telomere resolve encoded by the Lyme disease spirochete. *Mol. Cell.** 9**, 195–201

10. Oakey, H. J., Cullen, B. R., and Owens, L. (2002) The complete nucleotide sequence of the *Vibrio harveyi* bacteriophage VHML. *J. Appl. Microbiol.* **93**, 1089–1098

11. Lan, S. F., Huang, C. H., Chang, C. H., Liao, W. C., Lin, I. H., Jian, W. N., Wu, Y. G., Chen, S. Y., and Wong, H. C. (2009) Characterization of a new plasmid-like prophage in a pandemic *Vibrio parahaemolyticus* O3:K6 strain. *Appl. Environ. Microbiol.* **75**, 2659–2667

12. Zabala, B., Hammeri, J. A., Espejo, R. T., and Hertwig, S. (2009) The linear plasmid prophage Vp58.5 of *Vibrio parahaemolyticus* is closely related to the integrating phage VHML and constitutes a new incompatibility group of telomere phages. *J. Virol.* **83**, 9313–9320

13. Mobberley, I. M., Authement, R. N., Segall, A. M., and Paul, J. I. (2008) The temperate marine phage PhiHAP-1 of *Halomonas aquamarina* possesses a linear plasmid-like prophage genome. *J. Virol.* **82**, 6618–6630

14. Delaroque, N., Boland, W., Muller, D. G., and Knippers, R. (2003) Comparisons of two large phaeoviral genomes and evolutionary implications. *J. Mol. Evol.* **57**, 613–622

15. Delaroque, N., Muller, D. G., Bothe, G., Pohl, T., Knippers, R., and Boland, W. (2001) The complete DNA sequence of the *Ectocarpus siliculosus* Virus EsV-1 genome. *Virology* **287**, 112–132

16. Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T., Parkhill, J., Barrall, B. G., Chavchchie, C., Hamlin, N., Munagat, K., Norbertczak, H., Quail, M. A., Price, C., Rabbinowitch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. (2005) Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* **309**, 1090–1092

17. Allardet-Servent, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L., and Ramuz, M. (1993) Presence of one linear and one circular chromosome in the Agrobacterium tumefaciens C58 genome. *J. Bacteriol.* **175**, 7869–7874

18. Goodner, B., Hinkle, G., Guttung, S., Miller, N., Blanchard, M., Quorlo, B., Goldman, B. S., Cao, Y., Askemazi, M., Halling, C., Mullin, L., Houmield, K., Gordon, J., Vaudin, M., Iehorou, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Scarb, B., Grub, C., Elio, C., and Slater, S. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**, 2323–2328

19. Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G. E., Almeida, N. F. Jr., Woo, L., Chen, Y., Paulsen, I. T., Eisen, J. A., Karp, D. P., Bovee, D. S., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kuttyavin, T., Levy, R., Li, M. J., McClelland, E., Palmieri, A., Raymond, C., Rous, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z. Y., Dolan, M., Chunmey, F., Tingley, S. V., Tomb, J. F., Gordon, M. P., Olson, M. V., and Nester, E. W. (2001) The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**, 2317–2323

20. Casjens, S. (1999) Evolution of the linear DNA replicons of the *Borrelia* spirochetes. *Curr. Opin. Microbiol.* **2**, 529–534

21. Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J. F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vught, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Arthach, P., Bowman, C., Garland, S., Fuji, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., and
25. Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* 26, 391–406
26. Aihara, H., Huang, W. M., and Ellenberger, T. (2007) An interlocked dimer of the protelomerase TelK distorts DNA structure for the formation of hairpin telomeres. *Mol. Cell.* 27, 901–913
27. Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA. A view from high resolution crystal structures. *EMBO Rep.* 6, 39–45
28. Smogorzewska, A., and de Lange, T. (2004) Regulation of telomerase by telomeric proteins. *Annu. Rev. Biochem.* 73, 177–208
29. Young, J. M., Kuykendall, L. D., Martinez-Romero, E., Kerr, A., and Sawada, H. (2001) A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola*, and *R. vitis*. *Int. J. Syst. Evol. Microbiol.* 51, 89–103
30. Slater, S. C., Goldman, B. S., Goodner, B., Setubal, J. C., Farrand, S. K., Nester, E. W., Burr, T. J., Banta, L., Dickerman, A. W., Paulsen, I., Otten, L., Suen, G., Welch, R., Almeida, N. F., Arnold, F., Burton, O. T., Du, Z., Ewing, A., Godsby, E., Heisel, S., Houmier, K. L., Ihaveri, J., Lu, J., Miller, N. M., Norton, S., Chen, Q., Phoolcharoen, W., Ohlin, V., Ondrusk, D., Pride, N., Stricklin, S. L., Sun, J., Wheeler, C., Wilson, L., Zhu, H., and Wood, D. W. (2009) Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* 191, 2501–2511
31. Wibberg, D., Blom, J., Jänicke, S., Kollin, F., Rupp, O., Schneier-Bekel, S., Szczepanowski, R., Goesmann, A., Setubal, J. C., Schmitt, R., Pühler, A., and Schlüter, A. (2011) Complete genome sequencing of *Agrobacterium* sp. H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J. Biotechnol.* 155, 50–62
32. Deneke, J., Burgin, A. B., Wilson, S. L., and Chaconas, G. (2004) Catalytic residues of the telomere resolvase ResT. A pattern similar to, but distinct from, tyrosine recombinases and type IB topoisomerases. *J. Biol. Chem.* 279, 53699–53706
33. Welsh, E. A., Liberton, M., Stöckel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W., Jacobs, J. M., Aurora, R., Ghosh, B. K., Sherman, L. A., Smith, R. D., Wilson, R. K., and Pakrasi, H. B. (2008) The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15094–15099
34. Degnan, P. H., Yu, Y., Sisneros, N., Wing, R. A., and Moran, N. A. (2009) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9063–9068