Virology 468-470 (2014) 322–329

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Review

Multiple roles of genome-attached bacteriophage terminal proteins

Modesto Redrejo-Rodríguez, Margarita Salas*

Instituto de Biología Molecular "Eladio Viñuela" (CSIC), Centro de Biología Molecular "Severo Ochoa" (Consejo Superior de Investigaciones Científicas—Universidad de Madrid), Universidad Autónoma, Nicolás Cabrera, 1, Cantoblanco, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 11 July 2014
Returned to author for revisions
31 July 2014
Accepted 4 August 2014

Keywords:
Bacteriophage
Protein-primed replication
Terminal protein
Nucleoid
Nuclear localization signal
Horizontal gene transfer

ABSTRACT

Protein-primed replication constitutes a generalized mechanism to initiate DNA or RNA synthesis in linear genomes, including viruses, gram-positive bacteria, linear plasmids and mobile elements. By this mechanism a specific amino acid primes replication and becomes covalently linked to the genome ends. Despite the fact that TPs lack sequence homology, they share a similar structural arrangement, with the priming residue in the C-terminal half of the protein and an accumulation of positively charged residues at the N-terminal end. In addition, various bacteriophage TPs have been shown to have DNA-binding capacity that targets TPs and their attached genomes to the host nucleoid. Furthermore, a number of bacteriophage TPs from different viral families and with diverse hosts also contain putative nuclear localization signals and localize in the eukaryotic nucleus, which could lead to the transport of the attached DNA. This suggests a possible role of bacteriophage TPs in prokaryote-to-eukaryote horizontal gene transfer.

Contents

Replication of linear genomes

DNA polymerases (DNAPs) are unable to start *de novo* DNA synthesis and require a free hydroxyl group to incorporate each new deoxynucleotide. In the case of circular genomes, a nick at the replication origin can generate an available 3′OH group to initiate the new strand synthesis. Alternatively, replication often starts by the synthesis of an RNA primer oligonucleotide that must be subsequently removed and refilled. However, at linear chromosome ends, removal of the primer fragment would generate an unreplicated single-stranded DNA (ssDNA) portion, which may lead to end shortening and loss of genetic information. This so-called end-replication problem can be overcome by different strategies. As far back as 1972 it was clear that previously observed head–tail concatemer intermediates during replication of bacteriophages T4 or T7 would allow them to evade the end-replication problem (Watson, 1972). Other phage linear genomes avoid end shortening by undergoing circularization prior to DNA replication, like λ phage (Campbell, 1994). Many eukaryotic viruses have similar strategies that also involve the presence of inverted terminal repetitions (ITR) that allow the formation of concatemeric replicative intermediates, e.g., in parvovirus or poxvirus replication, or genome circularization prior to replication, as in the case of herpesvirus (Boehmer and Lehman, 1997; Moss, 2013). In eukaryotic chromosomes telomerase extends directly the 3′

* Corresponding author. Tel.: +34 91 196 4675.
E-mail address: msalas@cbm.csic.es (M. Salas).

http://dx.doi.org/10.1016/j.virol.2014.08.003
0042-6822/

end, giving rise to a G-rich overhanging ssDNA that is protected from nucleolytic attacks and recombinational activities by various strategies, including folding into higher structures such as T-loops or G-quadruplexes (Blackburn, 1990). Alternatively, other organisms including bacteriophages, animal viruses like Adenovirus, mitochondrial and cytoplasmatic plasmids of yeast and filamentous fungi, linear chromosomes and plasmids of *Streptomyces*, as well as human Hepatitis B virus and some plant and animal RNA viruses (Salas, 1991), possess a terminal protein (TP), which is involved in priming the genome replication by different mechanisms (see below) and becomes covalently linked to the genome ends. More recently, the presence of a protein-primed replication strategy has been suggested in a wide variety of replicons, including archaeal viruses, eukaryotic transposable elements (TE) and virophages (Bamford et al., 2005; Kapitonov and Jurka, 2006; Fischer and Suttle, 2011; Peng et al., 2007). These predictions are based on the homology of the protein-primed DNA polymerases, a subgroup of B-family DNA polymerases that contain two specific subdomains, named TPR1 and TPR2, which are involved in the interaction with TP and in processivity and strand-displacement capacity, respectively (Dufour et al., 2000; Rodríguez et al., 2005).

The evolutionary origin of protein-primed DNA replication mechanisms remains uncertain, although the incidence of genes encoding protein-primed DNA polymerases, with an apparent monophyletic origin (Filee et al., 2002; Braithwaite and Ito, 1993), among taxa associated with all domains of life, has been highlighted as consistent with a replication mechanism aroused early in the evolution, the genes being acquired from the primordial gene pool (Peng et al., 2007; Koonin, 2006). In addition, terminal proteins of human Hepatitis B virus and RNA viruses, though they would have independent origins, most likely also appeared early in the evolution, most likely prior to or simultaneously with key events of eukaryogenesis (Koonin et al., 2008). Hence, the utilization of a terminal protein as a primer for genome replication seems to be an ancient solution for the end-replication problem, which might have arisen independently several times and, as we discuss below, provides a useful instrument that has often acquired several additional functions associated with genome maintenance and replication, like compartmentalization or gene transfer.

Alternative protein-primed genome replication mechanisms

Protein-primed based replication of Adenoviruses and several bacteriophages from different families like Φ29 or PRD1 can duplicate the full-length genome continuously from both origins, without requiring Okazaki fragments or other accessory proteins (Salas, 1991). The replication mechanism of *Bacillus subtilis* phage Φ29 DNA has been extensively studied, thanks to the development of an *in vitro* replication system with purified proteins and TP-DNA as template (reviewed in De Vega and Salas (2011)). Fig. 1 shows a scheme of Φ29 genome replication. Briefly, the TP-DNAP heterodimer specifically recognizes the TP-containing DNA ends, which are opened with the help of the double-stranded DNA binding protein (DBP). The DNAP then catalyzes the addition of the first dAMP to the TP present in the heterodimer complex, using the second base of the genome end as a template. The presence of repetitive sequences at the replication origins in the Φ29 genome allows maintenance of the terminal nucleotides by a backward movement of the initiation complex called sliding-back, described also in other bacteriophages, or jumping-back in the case of Adenovirus. At least in the case of Φ29-like phages, the specificity of the internal template nucleotide that directs the first nucleotide incorporated is determined by the amino acids sequence near to

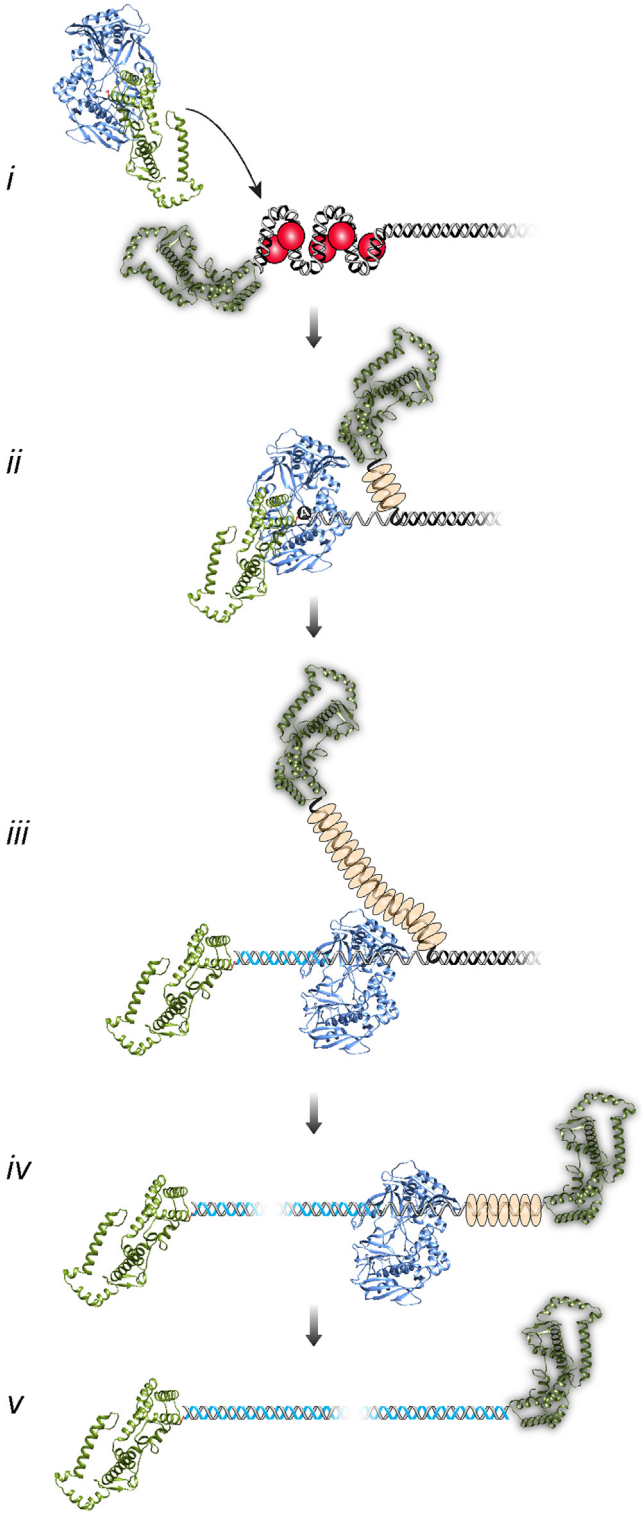

Fig. 1. Schematic representation of bacteriophage Φ29 TP-DNA replication. The DNAP (blue) and the TP (green) form an heterodimer that specifically binds to each of the genome ends (i). This binding is enhanced by the presence of the parental TP (dark green) and the end opening by the viral double-stranded DNA binding protein (DBP, red spheres). The synthesis of the new strand starts when the DNAP catalyzes the addition of the first nucleotide to its associated TP, using the second base of the genome end as a template (ii). Then, the DNAP dissociates, and the elongation progress asymmetrically from both ends (iii). The viral-encoded SSB protein (yellow ovals) binds to the displaced ssDNA and is removed by the DNAP during later stages of the replication process (iv). Eventually, continuous processive polymerization results in the generation of two fully replicated genomes (v). For simplicity only one of the TP-DNA ends is represented. See text for more details.