Virus Research 210 (2015) 178–187

Contents lists available at ScienceDirect

Virus Research

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

A RepA-like protein from bacteriophage BFK20 is a multifunctional protein with primase, polymerase, NTPase and helicase activities

Nora Halgasova[a], Barbora Solteszova[a], Vladimír Pevala[b], Július Košťan[c], Eva Kutejová[b], Gabriela Bukovska[a,*]

[a] Department of Genomics and Biotechnology, Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, 845 51 Bratislava, Slovakia
[b] Department of Biochemistry and Structural Biology, Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, 845 51 Bratislava, Slovakia
[c] Department for Structural and Computational Biology, Max F. Perutz Laboratories, Dr. Bohr-Gasse 9 (VBC 5), 1030 Vienna, Austria

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 6 August 2015
Accepted 10 August 2015
Available online 12 August 2015

Keywords:
Bacteriophage
Corynebacterium
DNA replication
Helicase
Protein expression
Replication protein

ABSTRACT

Phage BFK20 replication protein gp43 comprises an N-terminal prim-pol domain and a C-terminal domain similar to replicative helicases. We prepared and studied two recombinant proteins: gp43-1 containing both prim-pol and helicase domains, and gp43C with the helicase domain only. A SEC-MALS analysis indicated that gp43-1 forms a hexameric homooligomer. NTPase activity testing revealed that gp43-1 was able to hydrolyze a wide spectrum of NTPs, ATP the most efficiently. The ATPase activity of gp43-1 was strongly dependent on the presence of ssDNA in the reaction, but was low in the presence of dsDNA and in the absence of DNA. On the other hand, the ATPase activity of gp43C was very low regardless of the presence of DNA. The helicase activity of gp43-1 was detected using a fluorescence-based assay with a forked DNA substrate in the presence of ATP. However, no helicase activity could be detected for gp43C. We therefore conclude that the prim-pol domain is essential for the helicase and ssDNA-dependent ATPase activity of gp43-1.

1. Introduction

Bacteriophage replication is a complex process requiring the participation of both phage and host replication proteins. Phage genomes encode several families of replication proteins: initiator proteins, primases, helicase loaders, helicases and DNA polymerases (Weigel and Seitz, 2006a,b). The phage genomes typically have a modular structure, where genes encoding proteins with related functions are located close to each other. In most annotated phage genomes, the replication proteins are arranged in such modules. Moreover, several enzyme activities are often coupled into one large, multifunction protein. The most studied multifunctional phage replication proteins are gp4 from *Escherichia coli* phage T7, with both primase and helicase activities (Matson et al., 1983; Mendelman et al., 1992; Picha and Patel, 1998; Scherzinger et al., 1977; Toth et al., 2003) and alpha protein from *E. coli* phage P4, with origin recognition, primase and helicase activities (Ziegelin et al., 1993, 1995, 1997; Yeo et al., 2002).

Helicases are molecular motor proteins that use the energy from nucleoside triphosphate (NTP) hydrolysis to separate the complementary strands of double-stranded nucleic acids. RNA helicases have roles in all aspects of RNA metabolism while DNA helicases participate in many processes, including DNA replication, recombination and repair (Barry and Alberts 1994; Briggs et al., 2004; Carles-Kinch et al., 1997; LeBowitz and McMacken, 1986; Matson et al., 1983; McGlynn and Lloyd, 2002; Sharma et al., 2006), genome stability (Gabbai and Marians, 2010; Maher et al., 2011; Rezazadeh, 2011), chromatin remodeling (Ito et al., 1997; LeRoy et al., 1998; Saha et al., 2006) and the displacement of proteins from DNA (Byrd and Raney, 2006; Florés et al., 2005; Krejci et al., 2003). Based on the presence of conserved motifs and comparative structural and functional analysis (Gorbalenya and Koonin, 1993; Singleton et al., 2007), helicases have been divided into six superfamilies (SF). The two largest groups, SF1 and SF2, comprise enzymes that do not form ring-shaped oligomers. Families SF3 to SF6 contain helicases that function as toroidal, usually hexameric, structures (Singleton et al., 2007). SF3 comprises viral helicases, SF4 (F4 in Hall and Matson, 1999) contains helicases from the DnaB family, SF5, those from the Rho family and SF6 comprises the AAA+ATPases.

DNA helicases responsible for unwinding DNA at the replication fork are referred to as replicative helicases. They form ring-shaped structures that encircle and translocate along DNA and use the energy derived from NTP hydrolysis to separate duplex DNA

* Corresponding author. Fax: +43 421 2 59307416.
E-mail address: Gabriela.Bukovska@savba.sk (G. Bukovska).

http://dx.doi.org/10.1016/j.virusres.2015.08.011
0168-1702/

ahead of the replication fork. Bacterial replicative DNA helicases show homology to the *E. coli* DnaB protein. They are homohexamers that bind and translocate along single-stranded (ss) and double-stranded (ds) DNA and possess DNA-dependent ATPase and 5′ → 3′ helicase activities (Bujalowski et al., 1994; LeBowitz and McMacken, 1986; SanMartin et al., 1995). Examples of these include the RepA proteins, hexameric replicative helicases, which are encoded by plasmids found in most Gram negative bacteria (Dorrington and Rawlings, 1990; Scherzinger et al., 1997, 1984). Presently, there are two known types of bacteriophage replicative helicases. The most common type are also homologs of the *E. coli* bacterial DnaB protein and also have 5′ → 3′ unwinding activity. Examples include gp41 from T4 phage (Dong et al., 1995; Liu and Alberts, 1981) and G40P from SPP1 phage (Ayora et al., 2002; Bárcena et al., 1998). The second group comprises enzymes similar to the alpha protein from P4 phage which has origin recognition, primase and 3′ → 5′ helicase activities (Tocchetti et al., 2001; Ziegelin et al., 1993).

Bacteriophage BFK20 is a lytic phage of *Brevibacterium flavum* CCM 251 whose genome has been completely sequenced and annotated (EMBL AJ278322, Bukovska et al., 2006). Clusters of functionally related genes were defined and a replication module containing genes *orf29–orf46* was identified. Two putative helicases are encoded by *orf41* and *orf43*. Gp41 is a putative helicase from the SF2 superfamily while gp43 is a RepA-like protein with an N-terminal domain similar to the rare archaeo-eukaryotic primase-polymerase proteins and a C-terminal domain similar to SF4-type helicases. In our previous work we found that the N-terminal prim-pol domain, when expressed as a heterologous protein, possessed primase and polymerase activities (Halgasova et al., 2012). In this work we demonstrate that gp43 also possesses NTPase and helicase activities. Gp43 therefore appears to be a new member of the group of multifunctional phage replication proteins. We conclude that gp43 functions as a replicative helicase in the DNA replication process of BFK20.

2. Materials and methods

2.1. Bioinformatics analysis

The homology search on the gp43 sequence was done on the NCBI Blast Server (Altschul et al., 1997) using Blastp against the non-redundant database. A Conserved Domain Database (CDD, Marchler-Bauer et al., 2013) search was used to find conserved domains and conserved motifs in the protein. Protein sequence alignment was done using ClustalW (Higgins et al., 1996; Thompson et al., 1994).

2.2. Bacteriophage, bacteria and growth conditions

Bacteriophage BFK20 was propagated on *B. flavum* CCM 251 (hse^-, Aec^r) according to Koptides et al. (1992). Phage particles and phage DNA were isolated according to Sambrook and Russel (2001). *E. coli* XL1 blue (Agilent) was used for cloning and *E. coli* BL21 (DE3) (Novagen) for protein expression. They were grown in LB medium or Terrific Broth (TB) supplemented with 100 μg/ml kanamycin at 37 °C.

2.3. Cloning, expression and isolation of recombinant proteins

The forward (F) and reverse (R) primers used for PCR amplification, the restriction enzymes used for cloning, and the expression plasmids and recombinant proteins are summarized in Table S1. For *orf43* cloning, the sequence of the forward primer was designed so that the gp43-1 recombinant protein lacks the first five N-terminal amino acids of wild type gp43 (Bukovska et al., 2014). The gp43C recombinant protein was constructed using the region of *orf43* that contained the conserved helicase motifs. Amplification was performed using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific) in Phusion HF buffer with BFK20 DNA as a template on a T-Gradient thermal cycler (Whatman Biometra). The amplified PCR products were cloned into a pET28a+ expression vector and verified by sequence analysis using an eight-column capillary ABI 3100-Avant Genetic Analyser (Applied Biosystems). The resulting pET28-43-1 and pET28-43C plasmids were transformed into *E. coli* BL21 (DE3) cells. This cloning strategy resulted in the synthesis of gp43-1, a 1014 amino-acids protein with an N-terminal His-Tag sequence, and gp43C, a 334 residue protein with His-Tag sequences on both the N- and C-termini.

For protein expression, 1–2 ml of starter culture was inoculated into 200 ml LB (gp43-1) or TB (gp43C) medium supplemented with kanamycin (100 μg/ml). At an optical density of 0.5 at 600 nm, isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 0.5 mM. After additional growth for 3 h on an orbital shaker at 37 °C and 200 rpm, the cells were harvested by centrifugation at 2000 × *g* for 15 min at 4 °C, washed in physiological saline and again harvested. The pellets were suspended in 8 ml of 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole supplemented with 80 μl of Protease Inhibitor Cocktail (Sigma–Aldrich) and lysed using sonication 10 times for 15 s with amplitude 14 on a Soniprep 150 Plus (MSE). The cell debris were then removed by centrifugation at 20,000 × *g* for 30 min at 4 °C. The supernatant was loaded onto a chromatography column with 1 ml HIS-Select nickel affinity gel (Sigma–Aldrich) equilibrated with 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole. After protein loading, the column was washed with the same buffer and, in the case of gp43C, also with 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 60 mM imidazole. Affinity chromatography of both proteins was performed at room temperature. Proteins were eluted from the column with 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 0.4 M imidazole. Fractions containing the purified enzymes were pooled and dialyzed against 20 mM Tris–HCl pH 8.0, 40% (v/v) glycerol. Isolation of gp43C was also performed in buffers containing 50 mM HEPES pH 7.5, 0.5 M NaCl, 5 mM $MgCl_2$, 10% glycerol and imidazole at the concentrations given above. In this case gp43C was dialyzed against 25 mM HEPES pH 7.5, 0.5 M NaCl, 5 mM $MgCl_2$, 40% (v/v) glycerol. Purified gp43-1 and gp43C were quantified spectrophotometrically using calculated extinction coefficients. The control protein gp43N was expressed and isolated according to Halgasova et al. (2012).

2.4. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis

The purification of gp43-1 and gp43C was monitored by SDS-PAGE according to Laemmli (1970) using the Thermo Scientific Unstained Protein Molecular Weight Marker (#26610) as a molecular weight marker. The isolated recombinant proteins gp43-1 and gp43C were analyzed by Western blotting using a Panther semidry electroblotter (Owl). The His-Tag sequence was identified by immunoreaction with a His-Tag Monoclonal Antibody (Novagen) and a Goat anti-mouse IgG alkaline phosphatase conjugate (Novagen). A PageRuler Prestained Protein Ladder (#26616, Thermo Scientific) served as a molecular weight marker.

2.5. Gel filtration analysis

To probe the oligomeric states of gp43-1 and gp43C, analytical gel filtration was performed with a Superose 6 10/300 GL column (GE Healthcare) using the following buffers. For gp43-1: 20 mM Tris–HCl pH 8.0, 150 mM NaCl or 20 mM Tris–HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$ and for gp43C: 25 mM HEPES pH 7.5, 0.5 M NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol. Samples of purified proteins were