Virology 476 (2015) 61–71

Contents lists available at ScienceDirect

Virology

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

Structures of minute virus of mice replication initiator protein N-terminal domain: Insights into DNA nicking and origin binding

Sunil K. Tewary [a], Lingfei Liang [a], Zihan Lin [a], Annie Lynn [a], Susan F. Cotmore [b], Peter Tattersall [b,c], Haiyan Zhao [a,*], Liang Tang [a,*]

[a] Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045, USA
[b] Departments of Laboratory Medicine, Yale University Medical School, New Haven, CT 06510, USA
[c] Departments of Genetics, Yale University Medical School, New Haven, CT 06510, USA

ARTICLE INFO

Article history:
Received 23 September 2014
Returned to author for revisions
11 November 2014
Accepted 19 November 2014
Available online 18 December 2014

Keywords:
Parvovirus
DNA replication
Non-structural protein 1
Site-specific DNA binding
Nickase
Nuclease

ABSTRACT

Members of the *Parvoviridae* family all encode a non-structural protein 1 (NS1) that directs replication of single-stranded viral DNA, packages viral DNA into capsid, and serves as a potent transcriptional activator. Here we report the X-ray structure of the minute virus of mice (MVM) NS1 N-terminal domain at 1.45 Å resolution, showing that sites for dsDNA binding, ssDNA binding and cleavage, nuclear localization, and other functions are integrated on a canonical fold of the histidine-hydrophobic-histidine superfamily of nucleases, including elements specific for this *Protoparvovirus* but distinct from its *Bocaparvovirus* or *Dependoparvovirus* orthologs. High resolution structural analysis reveals a nickase active site with an architecture that allows highly versatile metal ligand binding. The structures support a unified mechanism of replication origin recognition for homotelomeric and heterotelomeric parvoviruses, mediated by a basic-residue-rich hairpin and an adjacent helix in the initiator proteins and by tandem tetranucleotide motifs in the replication origins.

Introduction

Minute virus of mice (MVM) belongs to the genus *Protoparvovirus* in the *Parvoviridae,* a family of small isometric viruses containing a linear single-stranded DNA (ssDNA) genome (Cotmore et al., 2014). It is broadly distributed in wild and laboratory mouse colonies, and exists as a range of genetically distinct allotropic strains that productively infect murine cells of specific differentiated lineages. The best characterized strains are MVMp, a "prototype" strain that infects cells of fibroblast origin, and MVMi, an "immunosuppressive" strain that productively infects T lymphocytes in culture, but also shows specificity for endothelium and hepatic erythropoietic precursors when infecting neonatal mice (Tattersall and Bratton, 1983). However, like many rodent protoparvoviruses, in human cells MVM is oncoselective and oncolytic, infecting tumor cell lines while being non-pathogenic for normal cells (Dupont, 2003), providing interesting therapeutic potential.

The MVM virion encapsidates a ssDNA genome of approximately 5 kb in a small protein capsid of ~280 Å in diameter with *T=1* icosahedral symmetry (Agbandje-McKenna et al., 1998; Cotmore and Tattersall, 2014). This genome contains two overlapping transcription units with P4 and P38 promoters positioned at 4 and 38 map units, respectively. Alternatively spliced mRNAs transcribed from the P4 promoter encode two major non-structural proteins, NS1 and NS2, that share a common 85 amino acid N-terminal domain, while the P38 promoter drives the synthesis of alternatively spliced transcripts encoding the capsid polypeptides (Pintel et al., 1983). The viral replication strategy, dubbed rolling hairpin replication (Cotmore and Tattersall, 2005a, 2013), is a linear adaptation of the more widely employed rolling-circle replication (RCR) mechanism (Kornberg and Baker, 1992). The NS1 protein of MVM is a multidomain, multi-functional, nuclear phosphoprotein that plays pivotal roles in initiating and directing viral DNA replication, as well as in viral DNA packaging and transcriptional activation of the viral promoters (Cotmore and Tattersall, 2014). Its N-terminal domain, the subject of the current study, contains overlapping site-specific double-stranded DNA (dsDNA) binding, ssDNA recognition, and origin-specific ssDNA nicking functions (Cotmore et al., 1995; Mouw and Pintel, 1998). It also contains a nuclear localization signal (NLS) that directs transport of the NS1 protein into the host cell nucleus during infection (Nuesch and Tattersall, 1993). The linear ssDNA genome of MVM is flanked by short palindromic sequences that can fold

* Corresponding authors. Tel.: +785 864 5838; fax: +785 864 5294.
E-mail addresses: zhaohy@ku.edu (H. Zhao), tangl@ku.edu (L. Tang).

http://dx.doi.org/10.1016/j.virol.2014.11.022
0042-6822/

into duplex hairpin telomeres, a feature common to all parvoviruses. For the members of some parvovirus genera, described as homotelomeric, these hairpin telomeres form part of a terminal repeat. On the other hand, the protoparvoviruses, such as MVM, are heterotelomeric, meaning that their left and right palindromes are different both in sequence and predicted structure (Cotmore and Tattersall, 2005b; Cotmore and Tattersall, 2014). In the MVM genome, these two hairpin sequences create disparate viral replication origins (Ori), called OriR (right) and OriL (left), when expressed in duplex replicative-form DNA. Within these origin sequences, NS1 binds site-specifically to 2–3 duplex reiterations of the tetranucleotide 5'–TGGT-3' (Cotmore et al., 1995; Cotmore et al., 2007; Mouw and Pintel, 1998). However, binding alone does not activate NS1's nicking function, which rather requires the cooperation of origin-specific cellular co-factors that use different mechanisms to further stabilize and orient NS1 in the nuclease complex (Christensen et al., 1999, 2001; Cotmore et al., 2000). These allow NS1 to unwind proximal dsDNA, likely in an ATP-dependent manner, generating a region of ssDNA that encompasses the resolution site, which is subsequently nicked by the NS1 nuclease activity (Christensen and Tattersall, 2002; Cotmore and Tattersall, 1989; Nuesch et al., 1995; Willwand et al., 1997; Wilson et al., 1991). Nicking occurs through a trans-esterification reaction that liberates a free 3′ hydroxyl group to prime unidirectional DNA synthesis, and leaves NS1 covalently attached to the new 5′ end of the DNA through a phosphotyrosine bond (Nuesch et al., 1995), through which it is thought to remain in the replication fork and serve as the 3'-to-5' replicative helicase (Christensen and Tattersall, 2002). Unlike other members of the *Parvoviridae*, viruses from genus *Protoparvovirus* have multiple additional cognate DNA binding sites for NS1 dispersed throughout their genomes, many of which bind NS1 with higher affinity than their Ori sequences, but without engaging its nickase activity (Cotmore et al., 2007). As a result, NS1 binds throughout duplex replicative-form viral DNA, potentially forming a unique type of chromatin and positioning NS1 to play additional roles in the viral life cycle. For example, NS1 is able to serve as a potent transcriptional activator of the otherwise silent viral P38 promoter, binding with high affinity to sites immediately upstream of this sequence and deploying its acidic carboxy-terminal transactivation domain (Legendre and Rommelaere, 1994; Lorson et al., 1996; Lorson et al., 1998).

In addition to its functions in the viral life cycle, MVM NS1 is also shown to inhibit the growth of transformed cells (Brandenburger et al., 1990; Caillet-Fauquet et al., 1990; Legendre and Rommelaere, 1992; Legrand et al., 1993) and to interfere negatively (Faisst et al., 1993; Legendre and Rommelaere, 1992) or positively (Vanacker et al., 1993) with the gene expression programmed by some heterologous promoters. However, genetic analysis has shown that both the transcription-regulating and cytotoxic activities of NS1 are largely confined to the amino and carboxy-terminal portions of the protein and can be dissociated from its replicative function (Legendre and Rommelaere, 1992). The nuclease activity of MVM NS1 was suggested to cause nicks in cellular chromatin (Op De Beeck and Caillet-Fauquet, 1997).

Four to five tandem tetranucleotide motifs 5′–GAGC-3′ are critical for initiator protein DNA binding in the origins of the homotelomeric parvovirus adeno-associated viruses (AAVs), which belong to the *Dependoparvovirus* genus (Hickman et al., 2004). In parvovirus B19 (B19V), a member of the homotelomeric *Erythroparvovirus* genus, the NS1-binding site in the DNA replication origin contains GC-rich 8-bp tandem repeats (Guan et al., 2009; Tewary et al., 2014), which can also be interpreted as tetranucleotide motifs (Tewary et al., 2014). However, the NS1-binding site (s) in the origins of the human bocavirus (HBoV), from the heterotelomeric *Bocaparvovirus* genus, have yet to be defined, and candidate sequences shared by both OriR and OriL are not obvious from simple visual inspection (Huang et al., 2012). Orthologs of the MVM NS1 nickase domain in AAV (Rep68/78) and HBoV were structurally characterized (Hickman et al., 2002; Tewary et al., 2013). The AAV Rep nickase domain was shown structurally to bind to the five tetranucleotide motifs as well as a stem-loop structure in the AAV Ori (Hickman et al., 2004). The MVM NS1 N-terminal domain (NS1N) shares low sequence identities of 17.8% and 18.7% with that of HBoV and AAV, respectively. Here, we report the crystal structure of the MVM NS1 N-terminal nickase domain (NS1N) at 1.45 Å resolution and structures in complex with metal ligands. These structures shed light on a nickase active site that is highly versatile in binding the metal ligands required for ssDNA binding and cleavage, and on a putative site for dsDNA binding, both built upon a canonical fold of the histidine-hydrophobic-histidine (HUH) superfamily of endonucleases. Comparative studies show structural elements that are unique to members of the *Protoparvovirus* genus, and distinct from those encoded by their *dependoparvovirus* and *bocaparvovirus* cousins. The structures support a unified mechanism of replication origin recognition for homotelomeric and heterotelomeric parvoviruses, which is mediated by a basic-residue-rich surface hairpin and an adjacent short helix in initiator proteins and by tandem tetranucleotide motifs in the viral DNA replication origin. Nevertheless, MVM NS1N doesn't seem to have a binding site for a DNA stem-loop as observed in the AAV Rep structure (Hickman et al., 2004), highlighting the variability in Ori binding mechanisms used by parvoviruses from different genera.

Results and discussion

Structure determination and the overall structure

The MVM NS1N encompassing residues 1 to 255 was overexpressed in *E. coli* and purified as a monomeric protein (see Materials and Methods). The structure was determined with the multiple isomorphous replacement method and refined at 1.45 Å resolution with excellent crystallographic and stereochemistry statistics (Table 1). The nickase domain consists of a centrally placed five-stranded antiparallel β-sheet (β6/β1/β5/β4/β8) flanked by four α-helices on one side (α2/α3/α4/α5) and three helices on the other side (α1/α6/α7) (Fig. 1 a). The central β-sheet forms a cleft that embraces the nickase active site (see below), which is surrounded by helix α6, a small loop between β4 and β5, and the loop L10.

Structural variations on a conserved fold among the Protoparvovirus, Dependoparvovirus and Bocaparvovirus genera

Superposition of the MVM NS1N structure with those of HBoV (Tewary et al., 2013) and AAV (Hickman et al., 2002) shows a core fold that is highly conserved among parvovirus NS1 nickases (Fig. 1 b, c and f). The core fold is comprised of the central β-sheet (β6/β1/β5/β4/β8), helices α6 and α7 on the left, and two long helices α3 and α4 on the right (Fig. 1 b and c). The central β-sheet and the two helices on the left are readily superimposable among the three structures, consistent with conserved functions of DNA cleavage by the nickase active site formed by central β-sheet and the two helices (α6/α7). Such a core fold bears apparent similarity to bacterial conjugative relaxases TrwC (Guasch et al., 2003) and TraI (Larkin et al., 2005) and transposase TnpA (Barabas et al., 2008; Ronning et al., 2005), which are all members of the HUH-superfamily nucleases (Fig. 1 d and e). Significant structural variations are present in the "right" portion of the molecules as oriented in the figure, which comprises β2, β3, α2 and α5 as well as