Published in final edited form as:

Virology. 2015 February ; 476: 61–71. doi:10.1016/j.virol.2014.11.022.

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

Sunil Kumar Tewary[1], **Lingfei Liang**[1], **Zihan Lin**[1], **Annie Lynn**[1], **Susan F. Cotmore**[2], **Peter Tattersall**[2,3], **Haiyan Zhao**[1,*], and **Liang Tang**[1,*]

[1]Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045, USA

[2]Department of Laboratory Medicine, Yale University Medical School, New Haven, Connecticut, USA

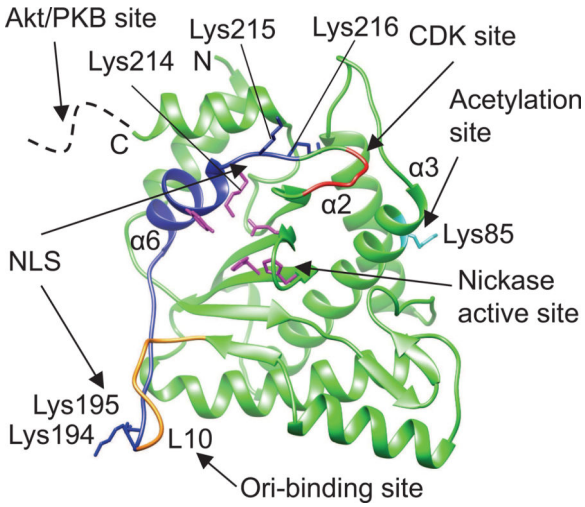
[3]Department of Genetics, Yale University Medical School, New Haven, Connecticut, USA

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.

Graphical Abstract

[*]Contact: Corresponding authors: L.T. or H.Z., Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045, USA, Tel: 785-864-5838 Fax: 785-864-5294, tangl@ku.edu or zhaohy@ku.edu.



Keywords

Parvovirus; DNA replication; non-structural protein 1; site-specific DNA binding; nickase; nuclease

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 ∼280Å 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, multifunctional, 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 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 via 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 via 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 1.45Å 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 inserted loops among them in MVM NS1N (Fig. 1 b and c). The AAV Rep nickase domain is the shortest, with 193 residues, while MVM and HBoV NS1N are markedly larger, consisting of 255 and 271 residues respectively and involving variable insertions among the five β-strands and the two helices. In MVM NS1N, there is a short β-hairpin (β2/β3) near the active site, which is not present in AAV and HBoV orthologs. Another noticeable structural variation lies in the positions of helices α6 and α7 of MVM NS1N, which adopt essentially the same conformation as in AAV Rep (Fig. 1c), while those in the HBoV ortholog apparently undergo a shift with respect to the central β-sheet (Fig. 1b).

These structural variations suggest that the right portions of the nickases are likely involved in virus-specific functions, for example, those related to assembly of NS1 proteins or their interactions with viral or cellular partners to enable virus replication. It is noticeable that AAV Rep nickase domain is the shortest among the three, which could be related to the fact that AAV is a member of the dependovirus genus whose replication requires helper viruses. In AAV Rep, it is likely that many elaborate insertions in the highly variable right portion of the nickase were lost during evolution, leaving over the structural core that is needed for only minimal functions, that is, Ori recognition and DNA nicking, which is consistent with the fact that AAV replication requires helper viruses such as herpesvirus and adenovirus and AAV alone is incapable of replicating as autonomous parvoviruses are. In MVM, it is known that cellular factors are engaged in Ori-binding and assembly of NS1. A heterodimeric cellular factor termed parvovirus initiation factor (PIF), also known as glucocorticoid modulating element binding protein, binds to a site immediately adjacent to the NS1-binding motifs in OriL to form a ternary nucleoprotein complex, which is required for NS1 nicking activity (Christensen et al., 1999, 2001). Cellular DNA-bending proteins of the HMG1/2 family are required for NS1-directed initiation of viral DNA replication, potentially through binding to the cruciform DNA structure formed by OriR (Cotmore et al., 2000; Cotmore and Tattersall, 1998). Thus, it would be interesting to test if the right portion of the MVM NS1N molecule is involved in physical interactions with any of those cellular factors.

The nickase active site

The nickase active site of MVM NS1N contains two histidine residues (His127/His129) separated by a spacer residue Cys128 (Fig. 2a and Fig. 1f). The two His residues are approached by Glu119, Lys 214, and the catalytic residue Tyr210 (Willwand et al., 1997). The structure-based sequence alignment of MVM NS1N with AAV Rep and HBoV NS1N shows strict conservation of these active site residues (Fig. 1f), reflecting the structural and functional unity of the active site architecture in the *Parvoviridae* family. Such an architecture of the active site, that is, the histidine-hydrophobic-histidine (HUH) motif and the catalytic tyrosine residue, comprises the characteristic structural elements that were originally identified in enzymes involved in RCR (Ilyina and Koonin, 1992; Koonin and Ilyina, 1993). These were later found to be ubiquitous among the HUH-superfamily endonucleases that cleave and/or ligate ssDNA and play roles in DNA replication in some bacterial and eukaryotic viruses, bacterial conjugation and transposition (Chandler et al., 2013).

Nevertheless, the MVM NS1N nickase active site exhibits differences from those of other HUH-superfamily endonucleases. A third His residue is frequently observed in conjugative relaxases TrwC (Guasch et al., 2003) and TraI (Larkin et al., 2005), while in MVM NS1N this position is occupied by Glu119, which is invariable among parvoviruses (Fig. 1f). In transposase TnpA, this position is a Gln residue (Barabas et al., 2008; Ronning et al., 2005). Such a difference may result in differential preferences and binding affinity for metal ligands. Several HUH endonucleases, such as gpA in bacteriophage phiX174 (Hanai and Wang, 1993), the A protein in bacteriophage P2 (Odegrip and Haggard-Ljungquist, 2001), conjugative relaxases TrwC (Guasch et al., 2003) and TraI (Larkin et al., 2005), contain two Tyr residues in the active site, which are functionally essential and are believed to alternate in the cleavage and ligation reactions. MVM NS1N contains only a single active Tyr residue (Willwand et al., 1997), which superposes well with Tyr18 in TrwC and Tyr16 in TraI. The lack of a second catalytic tyrosine residue in MVM NS1 supports the idea that no joining reaction is needed in NS1-directed DNA replication in MVM.

The MVM NS1 nickase active site is highly versatile in metal ligand binding

Divalent metal ions are required for DNA cleavage activities for the HUH-superfamily of endonucleases. It remains an open question what metal ion is physiologically used by a specific HUH-superfamily endonuclease. For example, it was reported that DNA cleavage activity of the relaxase TrwC, an HUH-superfamily endonuclease, can be activated by a variety of divalent metal ions including Mg^{2+}, Mn^{2+}, Ca^{2+}, Zn^{2+}, Cu^{2+} and Ni^{2+} in *in vitro* assays (Boer et al., 2006). It is often presumed that Mg^{2+} or Mn^{2+} are the physiological cofactors, given the abundance of Mg^{2+} in cells and similarity of the coordination configurations of Mg^{2+} or Mn^{2+}. DNA nicking of AAV Rep is metal-dependent (Davis et al., 2000). Mn^{2+} supported DNA nicking by AAV Rep (Hickman et al., 2002), while Mg^{2+} was unable to support ssDNA cleavage (Yoon et al., 2001). Mn^{2+} and Zn^{2+} but not Mg^{2+} bound to the AAV Rep nickase active site (Hickman et al., 2002), but Mg^{2+} was observed in the nickase active site of AAV Rep complexed with a DNA stem-loop harboring the secondary Rep-binding element (Hickman et al., 2004). TrwC nuclease activity can be activated by Mn^{2+} or Mg^{2+}, but none of the two metal ions were observed in the active site

in an X-ray crystallographic study of TrwC in complex with DNA oligonucleotide, potentially owing to the lack of the scissile phosphate in the bound DNA (Boer et al., 2006). Zn^{2+}, Cu^{2+} and Ni^{2+} supported TrwC nuclease activity, and were observed in the active site of TrwC soaked with those metal ions respectively, all showing a tetrahedral coordination configuration (Boer et al., 2006). In MVM NS1, Mg^{2+} supported DNA nicking (Nuesch et al., 1995; Willwand et al., 1997). We took the crystallographic approach to gain insights into metal ligand binding in MVM NS1N. The MVM NS1N native structure does not display any metal ion in the active site. The MVM NS1N crystals were soaked with a variety of divalent metal ions including Mg^{2+}, Mn^{2+}, Zn^{2+}, Cu^{2+}, Co^{2+} and Ni^{2+}, and high resolution X-ray structures were determined and refined (Table 2). These structures all show excellent electron density in the active sites for respective metal ions. The Mg^{2+}-, Mn^{2+}-, Co^{2+}- and Ni^{2+}-bound structures all show a nearly perfect six-member, octahedral configuration, in which the metal ion is coordinated with the epsilon nitrogen atoms of His127 and His129, a side chain carboxyl oxygen atom of Glu119, and three water molecules with an average coordinating distance of 2.2 Å (Fig. 2b). The Cu^{2+}-bound structure shows a tetrahedral configuration, in which the metal ion is coordinated with the epsilon nitrogen atoms of His127 and His129, a side chain carboxyl oxygen atom of Glu119, and one water molecules (Fig. 2c). The Zn^{2+}-bound structure shows a 5-member coordinating configuration, in which an additional water molecule participates (Fig. 2d). These data suggest that MVM NS1N nickase active site is capable of accommodating multiple types of metal ligands with different coordination configurations.

The architecture of the active site, that is, the conformations of His127, His129, Glu119, Lys214 and Tyr210 is essentially invariable in those metal-bound states as well as in the metal-free state. This suggests that the active site architecture of MVM NS1N is pre-configured for optimal binding of metal ions with various types of coordinating systems. Such a pre-configured architecture is enabled by a series of interactions (Fig. 2b). The imidazole ring of His129 is locked into the present rotamer conformation via an H-bond between its delta nitrogen with a water molecule, which in turn is H-bonded with the Ser27 side chain oxygen atom. The imidazole ring of His127 is immobilized through a stacking interaction with the Gln124 side chain. Side chains of Glu119 and Lys214 form a salt bridge. In comparison, metal binding leads to significant conformational changes in various regions in TrwC, including the nickase catalytic residue Tyr18, a loop adjacent to the active site harboring the second catalytic tyrosine residue, and the side chain conformations of the active site histidine residues (Boer et al., 2006), reflecting conformational flexibility in the nickase active site. The pre-configured architecture of the nickase active site observed in MVM NS1 might not exist in TrwC (Boer et al., 2006), thus providing a structural basis for capability of binding multiple types of metal ions in MVM NS1 but failure to bind Mg^{2+} and Mn^{2+} in TrwC as well as differences in the metal coordination systems observed in MVM NS1 and TrwC. The pre-configured, immobilized active site architecture in MVM NS1 and the conformational flexibility of the nickase active site in TrwC might be related to the functional differences of the two proteins, as TrwC must conduct a second cleavage reaction using the second catalytic tyrosine residue followed by DNA relegation while these do not occur in MVM NS1.

Taken together, these results indicate that MVM NS1N possesses a pre-configured nickase active site that is highly versatile for binding metal ions with various coordination systems.

Binding of the ssDNA substrate and the mechanism of DNA cleavage

To gain insight into the binding and cleavage of the ssDNA substrate, the MVM NS1N structure was superimposed with that of the Y16F mutant of the conjugative plasmid F factor TraI, a HUH endonuclease, in complex with a ssDNA molecule that extends beyond its nicking site (Larkin et al., 2005). Superimposition of the two structures resulted in a root mean square deviation (RMSD) of 2.64 Å for 104 Cα atoms. The MVM NS1N residues His127 and His129 align well with His157 and His159 of the TraI HUH motif, respectively, while Glu119 in MVM NS1N superposes with the third histidine residue (His146) in the TraI histidine triad (Fig. 3a). The MVM NS1N Tyr210 occupies a same position as the TraI catalytic residue Tyr16 (mutated into Phe in the structure). The metal ions in the MVM NS1N structures fit well with the Mg^{2+} ion in the TraI active site.

The superimposition shows that the ssDNA substrate docks onto the MVM NS1N active site via a curved path, circumventing the loop L10, crossing β6 and β1, and extending into the active site (Fig. 3a). Interestingly, two water molecules coordinating with Mg^{2+} in MVM NS1N occupy nearly the same positions as oxygen atoms O3' and OP1 in a backbone phosphate in the docked DNA from TraI (Fig. 3b), and the phosphate is positioned between the catalytic Tyr210 and the bound Mg^{2+}, with a distance of 2.69 and 2.38 Å from the oxygen atom OP1 of the DNA backbone phosphate to the side chain hydroxyl oxygen of Tyr210 and the Mg^{2+}, respectively (Fig. 3b). The oxygen atom O3' from the DNA backbone is at a distance of 2.51 Å from the Mg^{2+}. The phosphorous atom of the DNA backbone phosphate lies within 3.35 Å from the side chain hydroxyl oxygen of Tyr210. Such a conformation of the MVM NS1N active site together with the docked DNA may mimic the catalytic intermediate state of MVM NS1N poised for cleavage of the scissile phosphate bond of ssDNA. Based on the MVM NS1N structure and the docking analysis with the ssDNA substrate, we propose the following model for the nickase catalytic mechanism, which is enabled by a conserved Lys214:Glu119 pair unique to parvovirus NS1 nickases. The metal ion bound at the nickase active site is coordinated with two oxygen atoms, OP1 and O3', both bonded to the DNA backbone phosphorous, generating stress on the tetragonal structure of the phosphate and increasing the positive charge on the phosphorous atom so that it is susceptible to nucleophilic attack, which leads to formation of a penta-covalent intermediate with the fifth oxygen atom coming from the Tyr210 side chain hydroxyl group. Consequently, the scissile phosphate bond between the phosphorous atom and O3' is cleaved, a phospho-tyrosyl linkage is formed between the Tyr210 and the DNA 5'-phosphate, and a free hydroxyl group is generated at the 3' end of the DNA, which is to be used to prime virus DNA replication. The side chain NZ atom of Lys214 makes a salt bridge with Glu119 and is within 3.18 Å of the side chain hydroxyl oxygen of Tyr210. Such a Lys214:Glu119 pair may serve to generate the general base by abstracting a proton from NZ of Lys214, which then deprotonates the side chain hydroxyl group of Tyr210, turning it into a nucleophile to attack the DNA phosphorous. The Lys214:Glu119 pair is unique to parvovirus NS1 nickases and is absent in TrwC (Boer et al., 2006; Guasch et al., 2003) and

TraI (Larkin et al., 2005), in which acidic residues immediately adjacent to the catalytic tyrosine residues serve as the general base to activate the catalytic Tyr residue.

Before extending into the active site, the ssDNA spans an area where several acidic residues, namely, Glu33, Glu181 and Asp179 are lined up, together with a positively charged residue Lys31 (Fig. 3c). The distribution of charges in this area, that is, negatively charged area studded with a positively charged residue, may help guide the precise docking of ssDNA but preventing binding of non-specific DNA. Lys31 superimposes well with Arg10 in the Rep protein of AAV type 5 (Fig. 1f). The corresponding residue in AAV type 2 was reported to be required for DNA binding (Urabe et al., 1999). Interestingly, the corresponding position in HBoV NS1N is also occupied by a basic residue, Lys21 (Fig. 1f), confirming the role of this residue in ssDNA binding.

The mode of NS1 binding to the double-stranded Ori DNA

After entering the host cell nucleus, MVM genomic DNA is converted into a double-stranded replicative form of DNA through synthesis of the complementary strand DNA by a host polymerase which uses the OriL as the primer (Cotmore and Tattersall, 2005a, 2013). Such replicative form of DNA then allows transcription and synthesis of the NS1 protein, which in turn binds to the viral telomere to direct viral DNA replication and modulates host cellular processes in favor of virus replication (Cotmore and Tattersall, 2013). Thus, the initial binding site for NS1 in the Ori region in viral DNA is of double-stranded nature. Analysis of the electrostatic potential surface of MVM NS1N shows a highly basic region formed by the loop L10 (residues 186–202) and the N-terminal proximity of helix α4, which is rich in positively charged residues and can potentially serve to bind to DNA (Fig. 4 and 1f). To gain insight into the detailed protein:DNA interaction in Ori-recognition, the MVM NS1N structure is superimposed onto the structure of AAV Rep complexed with DNA, resulting in an RMSD of 2.42 Å for 159 Cα atoms (Fig. 1c). The loop L10 is considerably longer than the corresponding loops in AAV Rep and HBoV NS1 (Fig. 1b and c) (Hickman et al., 2004; Tewary et al., 2013), representing the longest in the *Parvoviridae* family. To avoid clashes between the lengthier loop L10 and DNA, the model was manually adjusted by moving the DNA slightly. This results in a plausible model for DNA binding of MVM NS1, in which the loop L10 inserts into the DNA major groove, and the N-terminal proximity of helix α4 interacts with the DNA minor groove, akin to those observed in AAV Rep (Hickman et al., 2004). The MVM NS1N loop L10 contains three basic residues Lys191, Lys194 and Lys195, corresponding to Lys135, Lys137 and Lys138 of AAV Rep, which may interact with DNA by inserting side chains into the DNA major groove as in AAV Rep (Fig. 4b) (Hickman et al., 2004). These residues in MVM NS1 form a KXXKK motif, showing consistence with the KXKK/RXRR motif in AAV Rep and HBoV NS1 respectively (Fig. 1f) but meanwhile indicating variations in details of interactions with DNA. The loop L10 contains two lysine residues (Lys189 and Lys199), which may also be involved in DNA interaction (Fig. 4b). In addition to the differences in lengths, the loop L10 in MVM NS1 adopts a conformation different from those in HBoV and AAV Rep (Fig. 1b and c). This suggests that these loops bear structural flexibility that allows conformational changes necessary for DNA binding. This is consistent with the weak electron density for the residues 191–193 located at the tip of the loop in MVM NS1N and the fact that the loops

of the two molecules in the crystallographic asymmetric unit in the HBoV NS1N X-ray structure adopt different conformations (Tewary et al., 2013). Several positively charged residues near the N-terminal end of helix α4, Lys143, Arg146 and Arg147, are proximal to the DNA phosphate backbones, providing additional charge-neutralizing interaction with DNA. The presence of positively charged residues around the N-terminal end of helix α4 is a conserved feature in AAV Rep and HBoV NS1, albeit the exact locations of those residues vary (Fig. 1f).

The conserved pattern of protein:DNA interactions as well as variations in details among MVM, AAV and HBoV are in agreement with similarities and differences in the sequences and structures of the NS1-bound DNA motifs in those parvoviruses. The Rep-binding site in the AAV Ori is five consecutive tetranucleotide motifs (Hickman et al., 2004). In MVM, the OriL and OriR also contain tetranucleotide motifs 5'-TGGT-3' (Cotmore and Tattersall, 2005a), although the repeating number differs from that in AAV. The GC-rich NS1-binding site in the Ori region of B19V can be viewed as five consecutive repeats of the tetranucleotide motif 5'-GCCG-3' (Guan et al., 2009; Tewary et al., 2014). While the exact DNA sequences of those motifs differ, the presence of consecutive tetranucleotide motifs appears to be a conserved feature among AAV, MVM and B19V, reflecting a conserved mode for NS1/Rep:DNA interactions.

AAV5 Rep contains an additional positively charged site opposite to the active site surface, which binds the Rep-binding element (RBE) stem-loop, strongly stimulating cleavage at the terminal resolution site (Hickman et al., 2004). The corresponding region on MVM NS1N, which is located at the back of the molecule when viewed as in Figure 1 a, is negatively charged (data not shown), and whether it is engaged in binding of additional DNA elements in the MVM Ori remains to be addressed. It is noticeable that such an RBE-binding site is not present in HBoV NS1N structure either (Tewary et al., 2013).

These results support a unified mechanism for recognition of replication origins by NS1/Rep for homotelomeric parvoviruses such as AAV and B19V and heterotelomeric parvoviruses such as MVM and HBoV, which is mediated by a basic-residue-rich surface loop or hairpin and an adjacent helix in NS1/Rep proteins and tandem tetranucleotide motifs in viral DNA replication origin.

Integration of multiple functions on MVM NS1N

In addition to the Ori-binding and ssDNA nicking, MVM NS1N encodes several functions essential for cellular localization and virus replication. Analysis of in-frame deletion and substitution mutants revealed that amino acid substitution in a triple lysine motif (residues 214–216) completely abrogated nuclear localization of the 672-residue NS1 protein. Substitution of a double lysine just upstream (residues 194–195) of this essential element also compromised nuclear localization, suggesting that the NLS is bipartite (Nuesch and Tattersall, 1993). The present structure shows that this bipartite NLS motif is located on the surface of NS1, spanning the putative DNA-binding loop L10, the helix α6 containing the catalytic residue Tyr210, and the nearby Lys214 (Fig. 5). While Lys214 is a part of the catalytic center, Lys215 and Lys216 are oriented away from the catalytic center. All these lysine residues are exposed on the molecular surface and would be very much available to

make contacts with the host factors such as importin to enable transportation into the host nuclei. Interestingly, the two elements (residues 194–195 and 214–216), while being located far apart as shown in the structure, appear to be adjacent in a crystallographic dimer (Fig. 5b), where they could possibly function together *in trans* across the dimer interface. This is consistent with the observation that, whereas mutations in either element disrupted the ability of the mutant NS1 to enter the nucleus, such mutant forms could be co-transported by wild-type NS1 expressed in the same cytosol (Nuesch and Tattersall, 1993).

Parvovirus H-1 (H-1PV), a close relative of MVM with 91% sequence identity for the NS1 proteins, has oncolytic and tumor-suppressive properties, which are potentially exploitable as novel cancer therapeutics (Li et al., 2013). In H-1PV, acetylation of the NS1 residue Lys85 modulates NS1-mediated transcription and cytotoxicity, and amino-acid substitution of the acetylation site strongly impairs NS1-mediated viral gene transcription, viral replication and cytotoxicity (Li et al., 2013). The present MVM NS1N structure shows that Lys85 is located on the helix α3 (Fig. 5), on the opposite side of the domain compared to the nickase active site. There is a predicted cyclin-dependent kinase (CDK) phosphorylation motif between residues Thr219 and Arg223. This motif is in close proximity to the second lysine cluster in the NLS motif, with a spacer of two residues (Fig. 5), and is conserved across the protoparvoviruses, but its function is currently unknown. Immediately after the C-terminus of MVM NS1N are two stretches of polypeptide (residues 261–267 and 276–278) that were shown to be required for NS1 oligomerization and helicase activity (Pujol et al., 1997), which harbors a potential Akt/PKB phosphorylation site at residue T278 that is also conserved across the protoparvoviruses. Mutation of Thr278 to either a phosphomimetic or a non-phosphorylable residue renders the virus non-infectious, and no complementation was observed in co-transfections with the two mutant genomes (P. Tattersall and S. Cotmore, unpublished results).

The nickase active site, the putative Ori-binding site and the NLS motif are located on the highly conserved core fold of the nickase domain, perhaps representing the minimal essential functions common to all parvoviruses. In comparison, the predicted CDK phosphorylation site, the acetylation site and the potential Akt/PKB phosphorylation site are located in regions that exhibit remarkable structural variations among NS proteins of parvoviruses from different genera (Fig. 1b and c), suggesting a great functional diversity for parvovirus NS1 proteins that are adapted to diverse host ranges and varied virus:host interactions. Integration of multiple functions onto the single domain of NS1N reflects a highly economic usage of protein coding capacity in parvoviruses, consistent with their small genome sizes. An interesting consequence of this genetic compactness is that the first 85 residues of the NS1N structure determined here also function as the N-terminal domain of NS2, whose C-terminal domain consists of 103 amino acid residues unrelated in sequence to NS1. NS2 is an ancillary viral protein with entirely separate functions from those of NS1, and it would be interesting to determine whether the structure of the shared 85 amino acid N-terminal sequence is maintained in the two viral proteins. Thus, the present MVM NS1N structure provides a framework for further studies to understand NS-guided virus replication and virus:host interactions.

MATERIALS AND METHODS

Protein expression and purification

The DNA fragment encoding MVM NS1N (residues 1–255; protein_id="AAA67109.1") was cloned into pET28a (Novagen) between *Nde1*/*BamH1* and a stop codon was introduced after Asp255. The N-terminally His-tagged protein was over-expressed by growth of *E. coli* strain B834(DE3) at 37°C in Luria-Bertani (LB) broth until OD_{600nm}=0.6, followed by induction at 16°C by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cells were harvested after 16 h of growth and lysed on a French press in the resuspension buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol and 5 mM β-mercaptoethanol). The protein was purified with a Ni-NTA column (Qiagen) followed by gel filtration chromatography on a Hiload 16/60 Superdex 75 column (GE Healthcare), which showed an elution volume of 64.8 ml corresponding to a monomer. The eluted fractions were collected in tubes containing the gel filtration buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 1 mM EDTA and 5% glycerol) supplemented with a final concentration of 100 mM L(+)Arginine (ACROS Organics) to prevent rapid protein precipitation and promote single crystal growth. The protein was concentrated to approximately 12 mg/ml using a Millipore centricon (molecular weight cutoff 10 kDa) prior to crystallization.

NS1N crystallization, X-ray data collection and structure determination

The purified MVM NS1N was crystallized with the hanging drop vapor diffusion method by mixing 1 μl protein solution with 1 μl of the well solution containing 2.8 M sodium formate and 100 mM sodium acetate trihydrate pH 5.4. Large crystals grew up to 0.6×0.4×0.08 mm^3 over 2–3 days at 20°C. Crystals were cryo-protected by transferring into a solution containing 3.5 M sodium formate and 100 mM sodium acetate trihydrate pH 5.4 for 1 min prior to flash freezing in liquid nitrogen. Heavy atom derivatives were obtained by soaking the native crystals in 3.5 M sodium acetate solution with 10 mM $HgCl_2$ or $Pb(C_2H_3O_2)_2$ for 5–7 min prior to flash freezing.

MVM NS1N native and heavy atom derivative X-ray data were collected at the Advanced Photon Source (APS) and Stanford Synchrotron Radiation Lightsource (SSRL), respectively. Additional data sets for the Zn^{2+}-containing crystal were collected at APS. Data were processed with xia2, which used the XDS algorithm (Kabsch, 2010; Winter et al., 2013). Structure determination by molecular replacement using the AAV Rep and HBoV1 NS1N structure as search models failed to give any correct solution. The structure was determined by the multiple isomorphous replacement method using the mercury and lead derivative data (Table 1). The experimental electron density map calculated with SOLVE in PHENIX was of excellent quality, which allowed automated building of most of the model using PHENIX (Adams et al., 2010). Small loops of regions of residues 72–81 and 191–195 were manually built using COOT (Emsley et al., 2010). Structure refinement coupled with manual model building was performed with the programs PHENIX (Adams et al., 2002) and COOT (Emsley et al., 2010) respectively. The crystallographic asymmetric unit contains one NS1N molecule. The refined model includes residues 6–253 with excellent refinement statistics and stereochemistry (Table 1). In the native structure, the Cys200 side chain thiol

group is covalently linked to β-mercaptoethanol, presumably resulted from chemical modification during protein purification. Crystals containing magnesium, manganese, zinc, nickel, cobalt and copper were obtained by soaking native crystals in a 10 mM metal ion solution for 5 min. These crystals were cryo-protected by transferring into a 3.5 M sodium formate, 100 mM sodium acetate trihydrate pH 5.4 and 15% xylitol for 1 min prior to flash freezing in liquid nitrogen. X-ray data were collected at SSRL or APS (Table 2). Structures were solved with the molecular replacement method using the native structure as the search model, and structure refinement and model building were performed with PHENIX and COOT, respectively (Table 2).

The coordinates and reflection data have been deposited with RCSB Protein Data Bank with the accession codes of 4PP4, 3WRN, 3WRO, 3WRQ, 3WRS, 3WRR and 4R94 for the native structure and the structures in complex with zinc, manganese, nickel, cobalt, copper and magnesium ions, respectively.

ACKNOWLEDGMENTS

We are grateful to the staff at the Advanced Photon Source beamlines GM-CA/CAT 23ID-B and 23ID-D and the Stanford Synchrotron Radiation Lightsource beamline 12-2 and 7-1 for help in X-ray data collection. The research in L.T.'s laboratory was supported by the NIH grant R01GM090010. P.T. and S.F.C. were supported by NIH grant R37AI026109.

REFERENCES

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002; 58:1948–1954. [PubMed: 12393927]

Agbandje-McKenna M, Llamas-Saiz AL, Wang F, Tattersall P, Rossmann MG. Functional implications of the structure of the murine parvovirus, minute virus of mice. Structure. 1998; 6:1369–1381. [PubMed: 9817841]

Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F. Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. Cell. 2008; 132:208–220. [PubMed: 18243097]

Boer R, Russi S, Guasch A, Lucas M, Blanco AG, Perez-Luque R, Coll M, de la Cruz F. Unveiling the molecular mechanism of a conjugative relaxase: The structure of TrwC complexed with a 27-mer DNA comprising the recognition hairpin and the cleavage site. Journal of molecular biology. 2006; 358:857–869. [PubMed: 16540117]

Brandenburger A, Legendre D, Avalosse B, Rommelaere J. NS-1 and NS-2 proteins may act synergistically in the cytopathogenicity of parvovirus MVMp. Virology. 1990; 174:576–584. [PubMed: 2137660]

Caillet-Fauquet P, Perros M, Brandenburger A, Spegelaere P, Rommelaere J. Programmed killing of human cells by means of an inducible clone of parvoviral genes encoding non-structural proteins. The EMBO journal. 1990; 9:2989–2995. [PubMed: 2167840]

Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nature reviews. Microbiology. 2013; 11:525–538. [PubMed: 23832240]

Christensen J, Cotmore SF, Tattersall P. Two new members of the emerging KDWK family of combinatorial transcription modulators bind as a heterodimer to flexibly spaced PuCGPy half-sites. Molecular and cellular biology. 1999; 19:7741–7750. [PubMed: 10523663]

Christensen J, Cotmore SF, Tattersall P. Minute virus of mice initiator protein NS1 and a host KDWK family transcription factor must form a precise ternary complex with origin DNA for nicking to occur. J Virol. 2001; 75:7009–7017. [PubMed: 11435581]

Christensen J, Tattersall P. Parvovirus initiator protein NS1 and RPA coordinate replication fork progression in a reconstituted DNA replication system. J Virol. 2002; 76:6518–6531. [PubMed: 12050365]

Cotmore SF, Agbandje-McKenna M, Chiorini JA, Mukha DV, Pintel DJ, Qiu J, Soderlund-Venermo M, Tattersall P, Tijssen P, Gatherer D, Davison AJ. The family Parvoviridae. Archives of virology. 2014; 159:1239–1247. [PubMed: 24212889]

Cotmore SF, Christensen J, Nuesch JP, Tattersall P. The NS1 polypeptide of the murine parvovirus minute virus of mice binds to DNA sequences containing the motif [ACCA]2-3. J Virol. 1995; 69:1652–1660. [PubMed: 7853501]

Cotmore SF, Christensen J, Tattersall P. Two widely spaced initiator binding sites create an HMG1-dependent parvovirus rolling-hairpin replication origin. J Virol. 2000; 74:1332–1341. [PubMed: 10627544]

Cotmore SF, Gottlieb RL, Tattersall P. Replication initiator protein NS1 of the parvovirus minute virus of mice binds to modular divergent sites distributed throughout duplex viral DNA. J Virol. 2007; 81:13015–13027. [PubMed: 17898054]

Cotmore SF, Tattersall P. A genome-linked copy of the NS-1 polypeptide is located on the outside of infectious parvovirus particles. J Virol. 1989; 63:3902–3911. [PubMed: 2527311]

Cotmore SF, Tattersall P. High-mobility group 1/2 proteins are essential for initiating rolling-circle-type DNA replication at a parvovirus hairpin origin. J Virol. 1998; 72:8477–8484. [PubMed: 9765384]

Cotmore, SF.; Tattersall, P. A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses. In: Kerr, J.; Cotmore, SF.; Bloom, ME.; Linden, RM.; Parrish, CR., editors. Parvoviruses. London, United kingdom: Hodder Arnold; 2005a. p. 171-181.

Cotmore, SF.; Tattersall, P. Structure and organization of the viral genome. In: Kerr, J.; Cotmore, SF.; Bloom, ME.; Linden, RM.; Parrish, CR., editors. Parvoviruses. London, United Kindom: Hodder Arnold; 2005b. p. 73-94.

Cotmore SF, Tattersall P. Parvovirus diversity and DNA damage responses. Cold Spring Harbor perspectives in biology. 2013:5.

Cotmore SF, Tattersall P. Parvoviruses: Small Does Not Mean Simple. Annual Review of Virology. 2014; 1 null.

Davis MD, Wu J, Owens RA. Mutational analysis of adeno-associated virus type 2 Rep68 protein endonuclease activity on partially single-stranded substrates. J Virol. 2000; 74:2936–2942. [PubMed: 10684315]

Dupont F. Risk assessment of the use of autonomous parvovirus-based vectors. Current gene therapy. 2003; 3:567–582. [PubMed: 14683452]

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica. Section D, Biological crystallography. 2010; 66:486–501.

Faisst SR, Faisst S, Grangette C, Schlehofer JR, Rommelaere J. NF kappa B upstream regulatory sequences of the HIV-1 LTR are involved in the inhibition of HIV-1 promoter activity by the NS proteins of autonomous parvoviruses H-1 and MVMp. Virology. 1993; 197:770–773. [PubMed: 8249299]

Guan W, Wong S, Zhi N, Qiu J. The genome of human parvovirus b19 can replicate in nonpermissive cells with the help of adenovirus genes and produces infectious virus. J Virol. 2009; 83:9541–9553. [PubMed: 19587029]

Guasch A, Lucas M, Moncalian G, Cabezas M, Perez-Luque R, Gomis-Ruth FX, de la Cruz F, Coll M. Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nature structural biology. 2003; 10:1002–1010. [PubMed: 14625590]

Hanai R, Wang JC. The mechanism of sequence-specific DNA cleavage and strand transfer by phi X174 gene A* protein. The Journal of biological chemistry. 1993; 268:23830–23836. [PubMed: 8226920]

Hickman AB, Ronning DR, Kotin RM, Dyda F. Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Mol Cell. 2002; 10:327–337. [PubMed: 12191478]

Hickman AB, Ronning DR, Perez ZN, Kotin RM, Dyda F. The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. Mol Cell. 2004; 13:403–414. [PubMed: 14967147]

Huang Q, Deng X, Yan Z, Cheng F, Luo Y, Shen W, Lei-Butters DC, Chen AY, Li Y, Tang L, Soderlund-Venermo M, Engelhardt JF, Qiu J. Establishment of a reverse genetics system for studying human bocavirus in human airway epithelia. PLoS Pathog. 2012; 8:e1002899. [PubMed: 22956907]

Ilyina TV, Koonin EV. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic acids research. 1992; 20:3279–3285. [PubMed: 1630899]

Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]

Koonin EV, Ilyina TV. Computer-assisted dissection of rolling circle DNA replication. Bio Systems. 1993; 30:241–268.

Kornberg, A.; Baker, TA. 2nd. New York: W.H. Freeman; 1992. DNA replication.

Larkin C, Datta S, Harley MJ, Anderson BJ, Ebie A, Hargreaves V, Schildbach JF. Inter-and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. Structure. 2005; 13:1533–1544. [PubMed: 16216584]

Legendre D, Rommelaere J. Terminal regions of the NS-1 protein of the parvovirus minute virus of mice are involved in cytotoxicity and promoter trans inhibition. Journal of virology. 1992; 66:5705–5713. [PubMed: 1388209]

Legendre D, Rommelaere J. Targeting of promoters for trans activation by a carboxy-terminal domain of the NS-1 protein of the parvovirus minute virus of mice. J Virol. 1994; 68:7974–7985. [PubMed: 7966588]

Legrand C, Rommelaere J, Caillet-Fauquet P. MVM(p) NS-2 protein expression is required with NS-1 for maximal cytotoxicity in human transformed cells. Virology. 1993; 195:149–155. [PubMed: 8317090]

Li J, Bonifati S, Hristov G, Marttila T, Valmary-Degano S, Stanzel S, Schnolzer M, Mougin C, Aprahamian M, Grekova SP, Raykov Z, Rommelaere J, Marchini A. Synergistic combination of valproic acid and oncolytic parvovirus H-1PV as a potential therapy against cervical and pancreatic carcinomas. EMBO molecular medicine. 2013; 5:1537–1555. [PubMed: 24092664]

Lorson C, Burger LR, Mouw M, Pintel DJ. Efficient transactivation of the minute virus of mice P38 promoter requires upstream binding of NS1. J Virol. 1996; 70:834–842. [PubMed: 8551622]

Lorson C, Pearson J, Burger L, Pintel DJ. An Sp1-binding site and TATA element are sufficient to support full transactivation by proximally bound NS1 protein of minute virus of mice. Virology. 1998; 240:326–337. [PubMed: 9454706]

Mouw M, Pintel DJ. Amino acids 16-275 of minute virus of mice NS1 include a domain that specifically binds (ACCA)2-3-containing DNA. Virology. 1998; 251:123–131. [PubMed: 9813208]

Nuesch JP, Cotmore SF, Tattersall P. Sequence motifs in the replicator protein of parvovirus MVM essential for nicking and covalent attachment to the viral origin: identification of the linking tyrosine. Virology. 1995; 209:122–135. [PubMed: 7747462]

Nuesch JP, Tattersall P. Nuclear targeting of the parvoviral replicator molecule NS1: evidence for self-association prior to nuclear transport. Virology. 1993; 196:637–651. [PubMed: 8372437]

Odegrip R, Haggard-Ljungquist E. The two active-site tyrosine residues of the a protein play non-equivalent roles during initiation of rolling circle replication of bacteriophage p2. Journal of molecular biology. 2001; 308:147–163. [PubMed: 11327759]

Op De Beeck A, Caillet-Fauquet P. The NS1 protein of the autonomous parvovirus minute virus of mice blocks cellular DNA replication: a consequence of lesions to the chromatin? J Virol. 1997; 71:5323–5329. [PubMed: 9188601]

Pintel D, Dadachanji D, Astell CR, Ward DC. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. Nucleic acids research. 1983; 11:1019–1038. [PubMed: 6828378]

Pujol A, Deleu L, Nuesch JP, Cziepluch C, Jauniaux JC, Rommelaere J. Inhibition of parvovirus minute virus of mice replication by a peptide involved in the oligomerization of nonstructural protein NS1. J Virol. 1997; 71:7393–7403. [PubMed: 9311818]

Ronning DR, Guynet C, Ton-Hoang B, Perez ZN, Ghirlando R, Chandler M, Dyda F. Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. Mol Cell. 2005; 20:143–154. [PubMed: 16209952]

Tattersall P, Bratton J. Reciprocal productive and restrictive virus-cell interactions of immunosuppressive and prototype strains of minute virus of mice. J Virol. 1983; 46:944–955. [PubMed: 6602222]

Tewary SK, Zhao H, Deng X, Qiu J, Tang L. The human parvovirus B19 non-structural protein 1 N-terminal domain specifically binds to the origin of replication in the viral DNA. Virology. 2014; 449:297–303. [PubMed: 24418564]

Tewary SK, Zhao H, Shen W, Qiu J, Tang L. Structure of the NS1 Protein N-Terminal Origin Recognition/Nickase Domain from the Emerging Human Bocavirus. J Virol. 2013; 87:11487–11493. [PubMed: 23966383]

Urabe M, Hasumi Y, Kume A, Surosky RT, Kurtzman GJ, Tobita K, Ozawa K. Charged-to-alanine scanning mutagenesis of the N-terminal half of adeno-associated virus type 2 Rep78 protein. Journal of virology. 1999; 73:2682–2693. [PubMed: 10074114]

Vanacker JM, Laudet V, Adelmant G, Stehelin D, Rommelaere J. Interconnection between thyroid hormone signalling pathways and parvovirus cytotoxic functions. Journal of virology. 1993; 67:7668–7672. [PubMed: 8230488]

Willwand K, Baldauf AQ, Deleu L, Mumtsidu E, Costello E, Beard P, Rommelaere J. The minute virus of mice (MVM) nonstructural protein NS1 induces nicking of MVM DNA at a unique site of the right-end telomere in both hairpin and duplex conformations in vitro. The Journal of general virology. 1997; 78(Pt 10):2647–2655. [PubMed: 9349487]

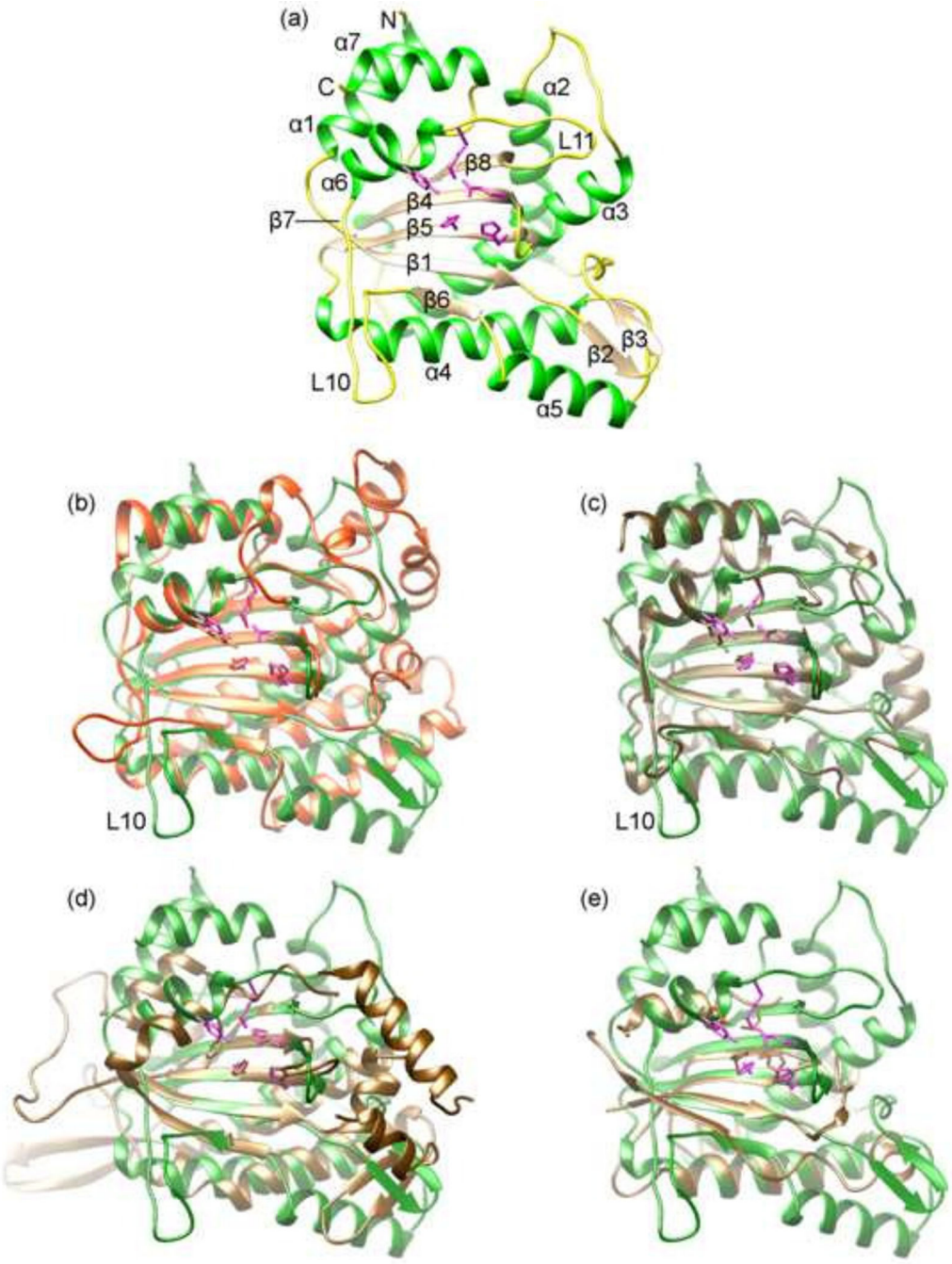
Wilson GM, Jindal HK, Yeung DE, Chen W, Astell CR. Expression of minute virus of mice major nonstructural protein in insect cells: purification and identification of ATPase and helicase activities. Virology. 1991; 185:90–98. [PubMed: 1833878]

Winter G, Lobley CM, Prince SM. Decision making in xia2. Acta crystallographica. Section D, Biological crystallography. 2013; 69:1260–1273.

Yoon M, Smith DH, Ward P, Medrano FJ, Aggarwal AK, Linden RM. Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus rep78 in vitro. J Virol. 2001; 75:3230–3239. [PubMed: 11238849]

HIGHLIGHTS

1. The structure of a parvovirus replication initiator protein has been determined;
2. The structure sheds light on mechanisms of ssDNA binding and cleavage;
3. The nickase active site is preconfigured for versatile metal ligand binding;
4. The binding site for the double-stranded replication origin DNA is identified;
5. A single domain integrates multiple functions in virus replication.



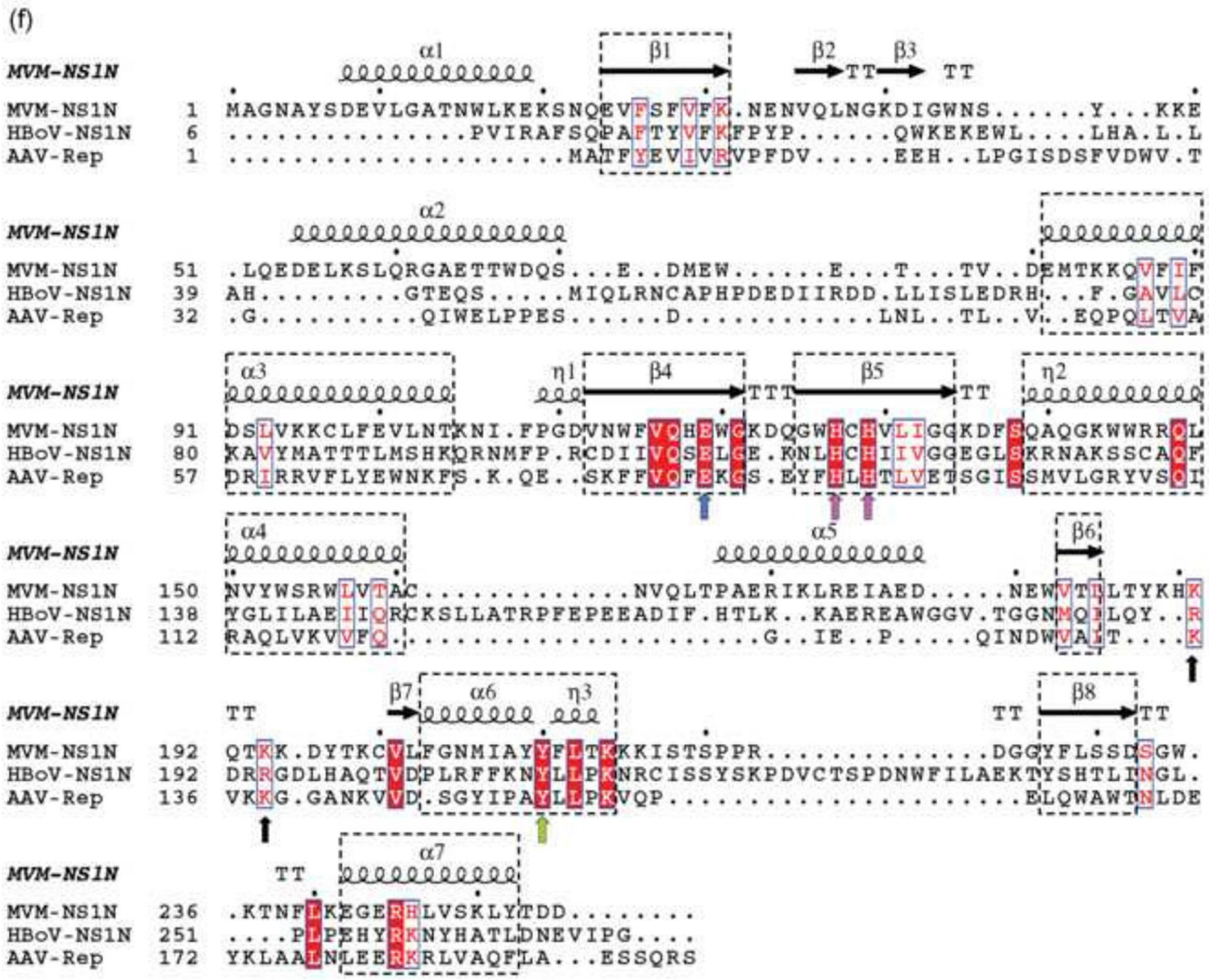


Fig 1. The overall structure of MVM NS1N

(a) The structure of MVM NS1N. The α helices, β strands and loops are in green, brown and yellow, respectively. Side chains of the nickase active site residues are shown as stick models in magenta. The amino- and carboxyl-termini are indicated with N and C respectively. The secondary structural elements are labeled. The position of the short β7 spanning two amino acid residues immediately N-terminal to helix α6 is indicated. (b–e) superimposition of MVM NS1N (green) with HBoV NS1N (orange; RCSB PDB accession code 4KW3), AAV Rep (brown; 1M55), TrwC (brown; 1QX0) and TnpA (brown; 2VJU) respectively. The nickase active site residues are shown as stick models. (**f**) Structure-based sequence alignment of the parvovirus N-terminal nuclease domains. The secondary structural elements based on the structure of MVM NS1N are shown above the sequences. α-helices, β-strands, turns (T) and 3_{10} helices (η) are indicated. The conserved core folds among MVM NS1N, AAV Rep and HBoV NS1N are highlighted in dashed rectangles. The conserved active site residues His127/His129, Tyr210 and Glu119 are indicated with arrows in magenta, green and blue, respectively. The Ori-binding loop of

MVM NS1N exhibits a KXXKK motif (black arrows), corresponding to KXKK in AAV Rep and RXRR in HBoV NS1, where X is any amino acid residue.

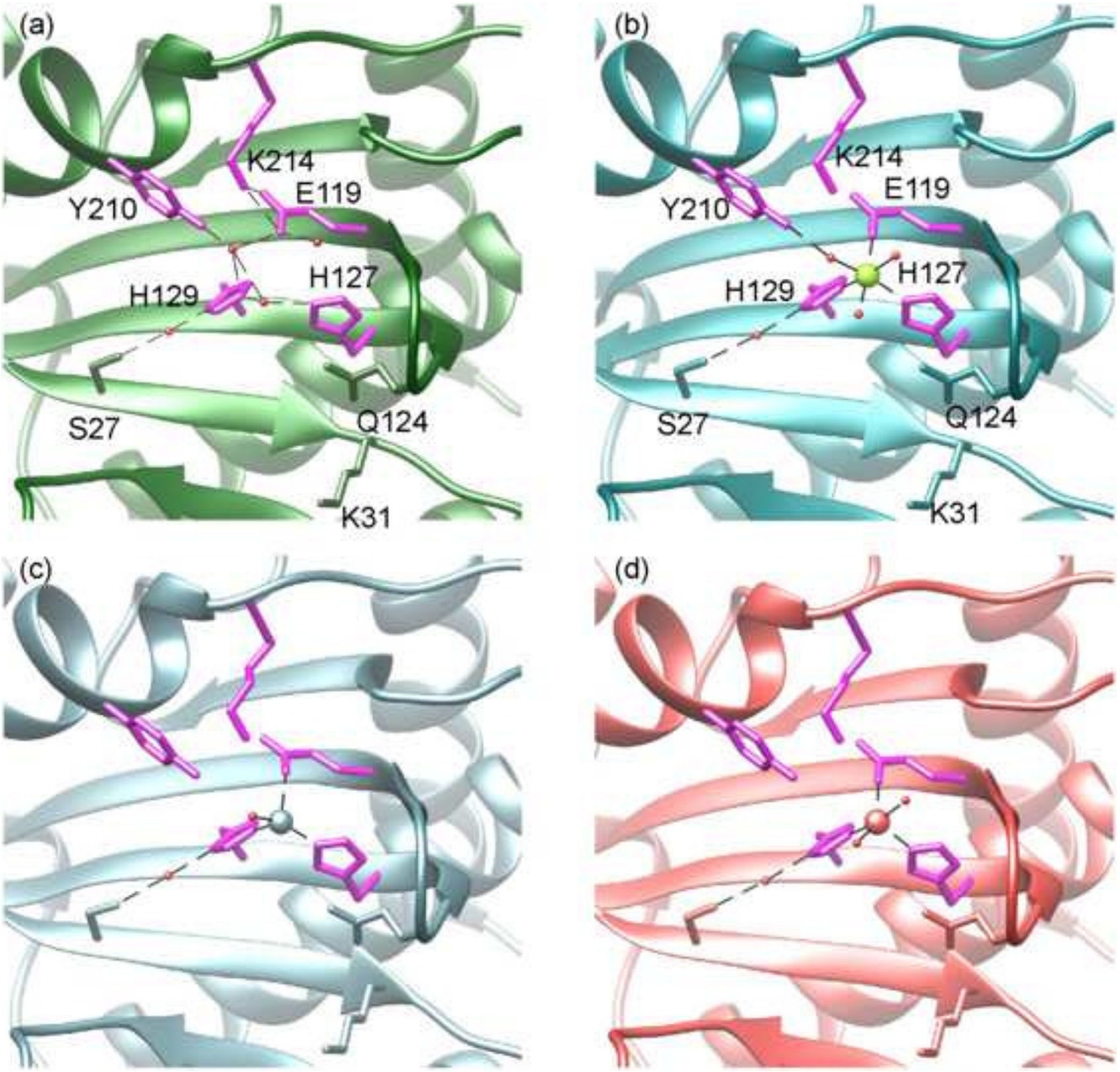


Fig 2. The MVM NS 1N nickase active site

(a) The H-bond network in the nickase active site of the native structure. The active site residues are shown as stick models in magenta and labeled. Residues Ser27, Gln124 and Lys31 are shown as stick models in green. Water molecules, red spheres. H-bond, dashed lines.

(b–d) The conformation of the nickase active site bound with the six-member, octahedrally coordinated Mg^{2+} (green sphere), the 4-member, tetrahedrally coordinated Cu^{2+} (light blue sphere) and the 5-member coordinated Zn^{2+} (red sphere), respectively.

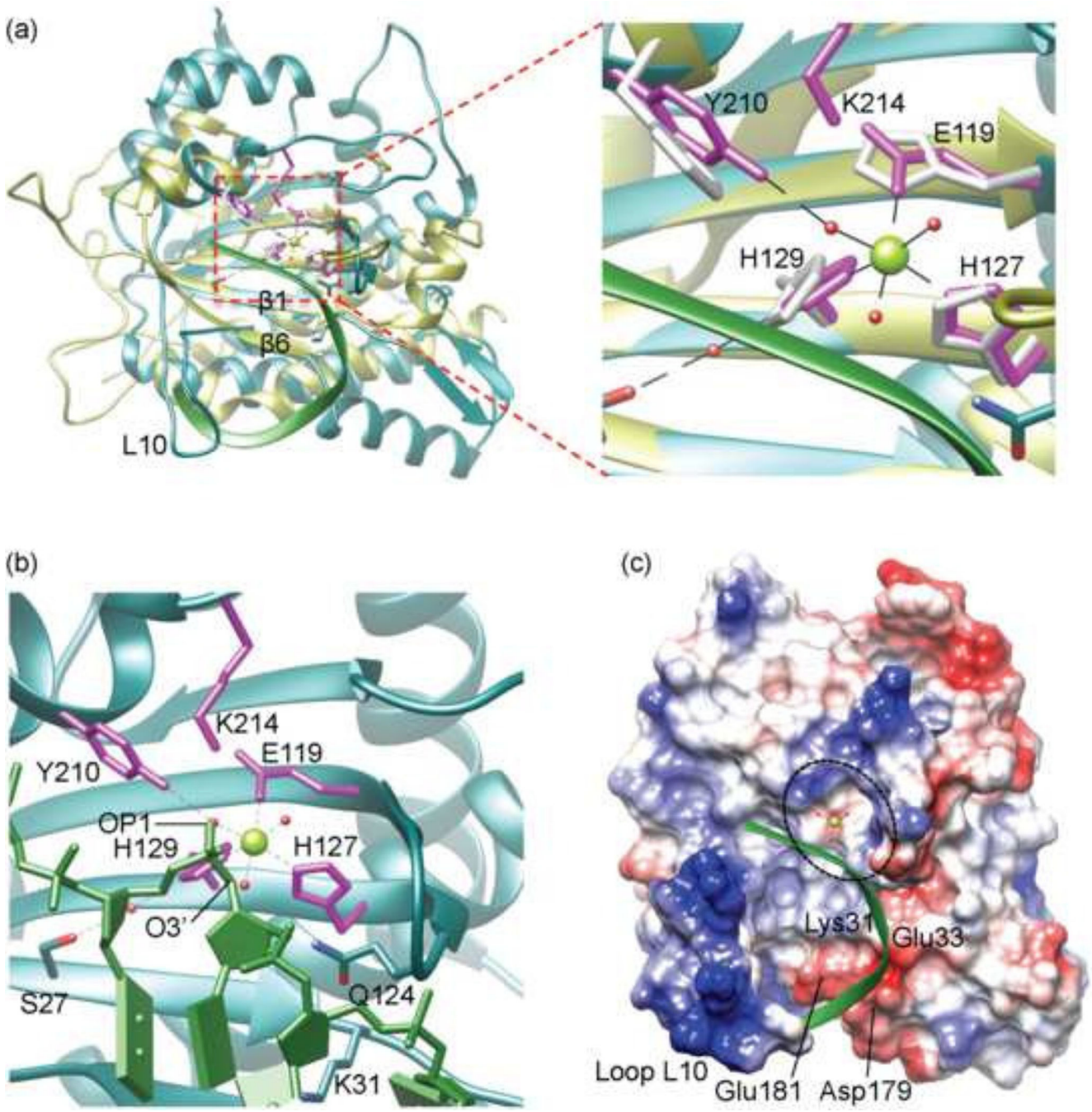


Fig 3. A model for binding of MVM NS 1N to the ssDNA substrate

(a) Superposition of MVM NS1N (cyan) with TraI (yellow; RCSB PDB entry code 2a0i). The nickase active site residues are shown as stick models in magenta for MVM NS1N and yellow for TraI. The bound Mg^{2+} in MVM NS1N is shown as a green sphere. The backbone of the bound ssDNA in TraI is shown as a ribbon diagram in green. Structural elements loop L10, β1 and β6 are labeled for clarity.

(b) Closeup of the MVM NS1N nickase active site docked with the ssDNA (green) from TraI. The three water molecules coordinated with the Mg^{2+} are shown as red spheres. Notice

the OP1 and O3’ atoms of ssDNA (indicated) nearly overlap with two Mg^{2+}-coordinating water molecules respectively.

(c) Electrostatic potential surface of MVM NS1N with docked ssDNA (green ribbon diagram) showing how the DNA aligns with the nickase active site (dashed ellipsoid) and approaches the catalytically essential metal ion (green sphere). Residues Lys31, Glu33, Asp179 and Glu181 as well as the loop L10 are indicated.

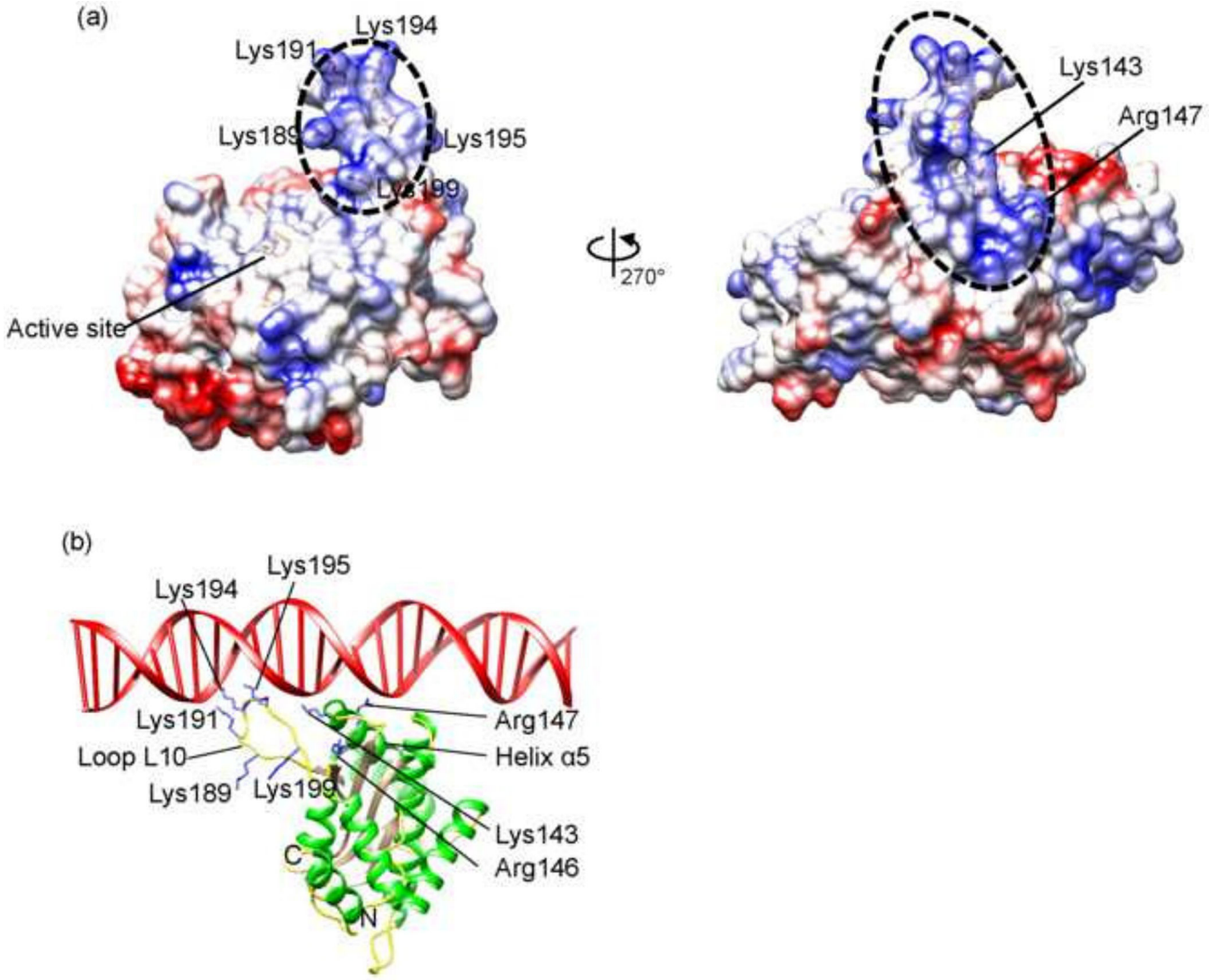


Fig 4. An Ori-binding model of MVM NS1N

(a) Electrostatic potential surface of MVM NS1N. Positively charged residues in the putative Ori-binding site (dashed ellipsoid) are indicated.

(b) The MVM NS1N structure with docked dsDNA.

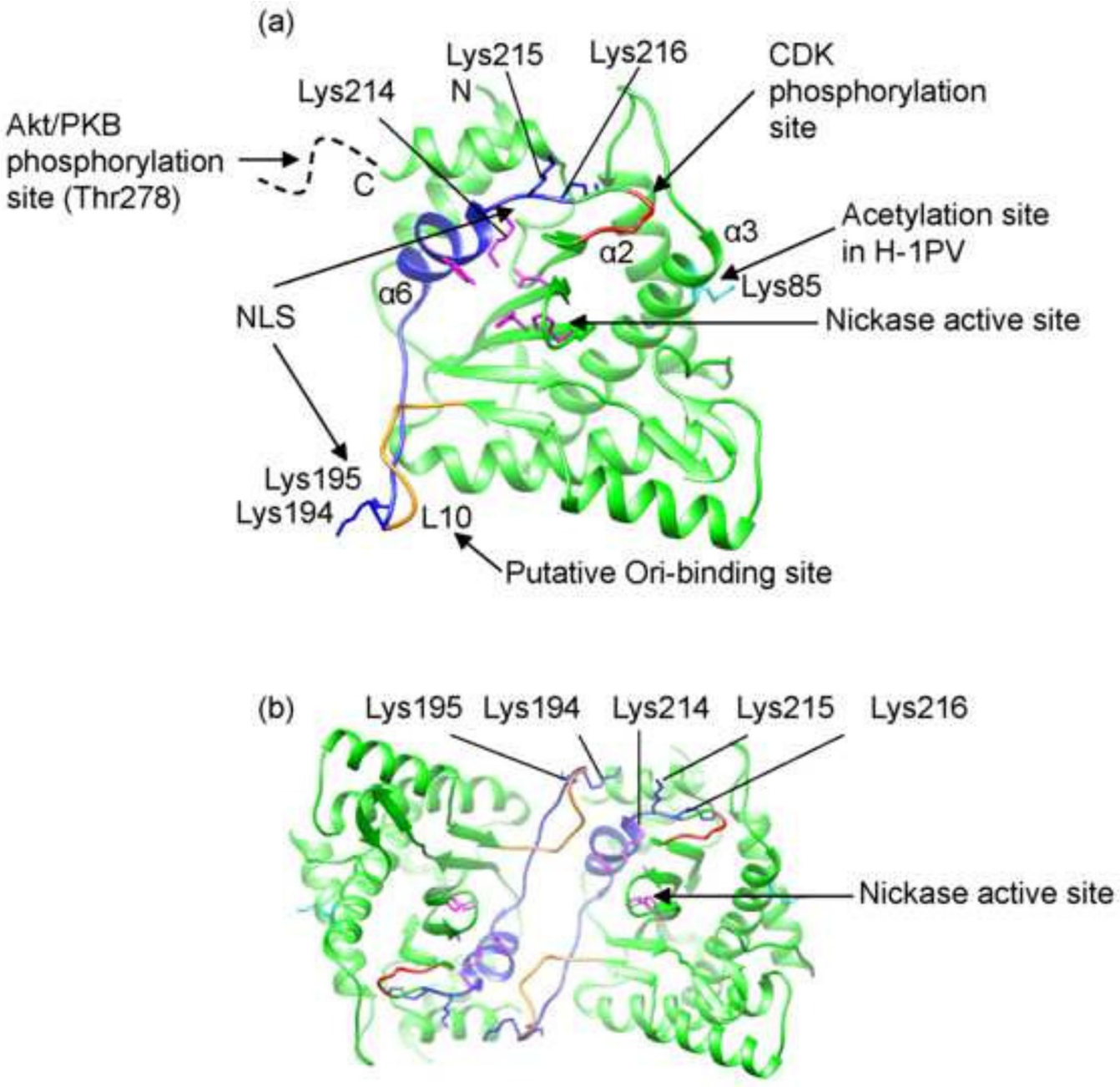


Fig 5. Functional motifs of MVM NS1N

(a) Functional motifs mapped onto the MVM NS1N structure. Nuclear localization signal (NLS) (residues 194–195 and 214–216), blue. Acetylation site (Lys85), cyan. Predicted cyclin-dependent kinase (CDK) phosphorylation site (residues 219–223), red. The potential Akt/PKB phosphorylation site at residue Thr278 is located in a region immediately after the C-terminus of the nickase domain (dashed curve). The nickase active site residues are shown in magenta. The putative Ori-binding site (loop L10) overlaps with the NLS and is partially in gold. The N- and C- termini are indicated.

(b) A crystallographic dimer of MVM NS1N view down the 2-fold axis showing that the first element in the NLS, Lys194-195, is adjacent to the second element (Lys214-216) in the NLS from a two-fold-related molecule.

Table 1

X-ray data collection and structure refinement statistics for the native crystal and heavy atom derivatives

Data collection	**Native**	**Pb**	**Hg**
Beamline	APS 23ID-D	SSRL 7-1	SSRL 7-1
Wavelength (Å)	0. 97949	0. 97946	0. 97946
Resolution (Å)	33-1.45 (1.49-1.45)*	34.5-1.56 (1.56-1.60)	34.4-1.56 (1.56-1.60)
No Measurements	142,612	183,404	148,706
Unique reflections	51,499 (3,649)	41,626 (2,901)	41,677 (2,794)
Completeness (%)	98.6 (95.8)	99.0 (95)	99.0 (91.3)
I/σ	20.4 (2.0)	16.3 (2.3)	16 (2.3)
R_{merge} (%)**	2.7 (35.2)	5.1 (38.2)	5.1 (51.3)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell (Å)	a=56.92, b=121.61, c=41.81	a=56.99, b=121.24, c=41.81	a=57.05, b=121.29, c=41.73
Structure refinement			
Resolution (Å)	33.71-1.45		
R_{work}/ R_{free} [a]	0.17/0.19		
Number of atoms			
Protein	2,074		
Water	242		
B-factors			
Protein	25.8		
Water	32.9		
R.m.s deviations			
Bond lengths (Å)	0.009		
Bond angles (°)	1.151		
Ramachandran plot(%)			
Most favored regions	97.56		
Allowed regions	2.44		
Disallowed regions	0.00		

*Values in the parentheses are for the outermost resolution shells.

**Rmerge = $\Sigma_{hkl}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity of reflection hkl and $\langle I(hkl)\rangle$ is the averaged intensity of symmetry-equivalent measurements

[a] $R_{work} = \Sigma_{hkl}||F_{obs}|-|F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are structure factors of the observed reflections and those calculated from the refined model, respectively. R_{free} has the same formula as R_{work}, except that it was calculated against a test set of the data that was not included in the refinement

Table 2

X-ray data collection and structure refinement statistics of metal-complexed NS1N

Data collection						
Dataset	Mn^{2+}	Ni^{2+}	Co^{2+}	Cu^{2+}	Zn^{2+}	Mg^{2+}
Beamline	SSRL 12-2	SSRL 12-2	SSRL 12-2	SSRL 12-2	SSRL 12-2	APS 23IDB
Wavelength (Å)	0. 97950	0. 97950	0. 97950	0. 97950	0. 97950	0.97934
Resolution (Å)	32.4-1.48 (1.48-1.52)*	30.3-1.54 (1.54-1.58)*	34.2-1.58 (1.58-1.62)*	33.6-1.62 (1.62-1.66)*	33.33-1.53 (1.53-1.57)*	50.00-1.67 (1.70-1.67)*
No. Measurements	205,515	195,161	118,205	195,667	223,108	151,693
Unique reflections	48,589 (3,538)	42,213 (3,139)	38,511 (2,884)	37,099 (2,707)	43,594 (3,148)	33,535 (1,610)
Completeness (%)	98.6 (98.8)	98.8 (98.7)	96.0 (98.7)	99.3 (99.4)	98.8 (98.7)	97.2 (95.9)
I/σ	17.8 (2.5)	13.7 (2.5)	17.1 (2.3)	13.1 (2.4)	20.3 (2.9)	37.1 (3.1)
Rmerge (%)**	3.8 (64.0)	5.5 (59.2)	3.5 (52.3)	7.0 (75.3)	3.9 (64.2)	3.6 (45.3)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell (Å)	*a*=57.24, *b*=121.35, *c*=41.75	*a*=57.28, *b*=121.26, *c*=41.65	*a*=57.11, *b*=120.95, *c*=41.50	*a*=57.06, *b*=120.84, *c*=41.59	*a*=56.78, *b*=121.43, *c*=41.56	*a*=57.02, *b*=121.34, *c*=41.58
Structure refinement						
Resolution (Å)	32.4-1.48	30.3-1.54	34.2-1.58	33.6-1.62	32.33-1.53	19.68-1.67
R_{work}/R_{free} **	0.18/0.20	0.18/0.20	0.19/0.22	0.17/0.20	0.17/0.20	0.16/0.19
Number of atoms						
Protein	2,074	2,074	2,074	2,074	2,074	2,074
Water	170	196	169	181	182	253
<B-factors>						
Protein	29.5	30.4	32.5	33.9	33.0	22.9
Water	35.1	35.8	36.9	37.6	36.9	44.5
Metal ion	Mn^{2+}	Ni^{2+}	Co^{2+}	Cu^{2+}	Zn^{2+}	Mg^{2+}
Occupancy	0.78	0.82	0.80	0.71	0.53	1.00
B-factor	15.5	18.6	18.2	20.6	16.7	20.9
R.m.s deviations						

Data collection						
Bond lengths (Å)	0.005	0.006	0.006	0.006	0.006	0.010
Bond angles (°)	0.938	0.927	0.910	0.912	0.921	1.190
Ramachandran plot (%)						
Most favored region	97.56	97.97	97.97	98.37	98.37	97.97
Allowed regions	2.44	2.03	2.03	1.63	1.63	2.03
Disallowed regions	0	0	0	0	0	0

*Values in the parentheses are for the outermost resolution shells.

**See Table 1 for definitions of R_{merge}, R_{work} and R_{free}