High-Resolution Structure of the Nuclease Domain of the Human Parvovirus B19 Main Replication Protein NS1

Jonathan L. Sanchez[1,2,3], Niloofar Ghadirian[2] and Nancy C. Horton[4*]

Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721

Running title: DNA binding and cleavage by Human Parvovirus B19 NS1 nuclease domain

[1]BMCB Graduate Program, University of Arizona, Tucson, AZ 85721
[2]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721
[3]Current address: Department of Pharmacology, College of Medicine and Center for Innovation in Brain Science, University of Arizona
[4]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721

***To whom correspondence should be addressed:**
Dr. Nancy C. Horton
Department of Molecular and Cellular Biology
Telephone: (520) 626-3828.
E-mail: nhorton@u.arizona.edu.

Abbreviations

AAV Adeno Associated Virus

AAV2 Adeno Associated Virus serotype 2

AAV5 Adeno Associated Virus serotype 5

B19V, Human Parvovirus B19

C_α, alpha carbon atom

HBoV, Human Bocavirus

MVM, Minute Virus of Mice

MVM NS1, NS1 protein from MVM

MVM NS1-nuc, the N-terminal nuclease and origin binding domain of MVM NS1

NS1, B19V NS1 unless otherwise indicated

NS1-nuc, the N-terminal nuclease and origin binding domain of B19V NS1

NSBE, B19V NS1 Binding Elements in double-stranded DNA at the B19V origin of replication

PAGE, polyacrylamide gel electrophoresis

AAV2 Rep or AAV5 Rep, homologs of NS1 from AAV2 or AAV5

AAV2 Rep-nuc or AAV5 Rep-nuc, the N-terminal endonuclease domain of AAV2 Rep or AAV5 Rep

SDS, sodium dodecyl sulfate

TEV, Tobacco Etch Virus

Tris, Tris(hydroxymethyl)aminomethane

Tris-HCl, Tris titrated to a desired pH with HCl

trs, terminal resolution site or nicking site in DNA at the viral origin of replication

WDV, Wheat Dwarf Virus

WDV Rep-nuc, the N-terminal origin binding and nuclease domain of WDV Rep protein

ABSTRACT

Two new structures of the N-terminal domain of the main replication protein, NS1, of Human Parvovirus B19 (B19V) are presented. This domain (NS1-nuc) plays an important role in the "rolling hairpin" replication of the single-stranded B19V DNA genome, recognizing origin of replication sequences in double-stranded DNA, and cleaving (*i.e.* nicking) single-stranded DNA at a nearby site known as the trs. The three-dimensional structure of NS1-nuc is well conserved between the two forms, as well as with a previously solved structure of a sequence variant of the same domain, however shown here at significantly higher resolution (2.4Å). Using structures of NS1-nuc homologues bound to single- and double-stranded DNA, models for DNA recognition and nicking by B19V NS1-nuc are presented which predict residues important for DNA cleavage and for sequence-specific recognition at the viral origin of replication.

Importance:

The high-resolution structure of the DNA binding and cleavage domain of the main replicative protein, NS1, from the human pathogenic virus Human Parvovirus B19 is presented. Included also are predictions of how the protein recognizes important sequences in the viral DNA which are required viral replication. These predictions can be used to further investigate the function of this protein, as well as predict the effects on viral viability due to mutations in the viral protein and viral DNA sequences. Finally, the high-resolution structure facilitates structure-guided drug design efforts to develop anti-viral compounds against this important human pathogen.

Key words: viral origin of replication, DNA nicking, nuclease, single-stranded DNA binding, double-stranded DNA binding, parvovirus, protein structure, enzyme

Introduction

Human Parvovirus B19 (B19V) is a ubiquitous virus infecting the majority of the human population[1, 2]. B19V, a *Parvoviridae* family member of the genus *Erythrovirus* has been associated with a myriad of different illnesses; B19V was first discovered as the cause of aplastic crisis in patients with chronic hemolytic anemia[3], then as the causative agent of *erythema infectiosum* (Fifth disease) in 1983[4] which results in mild fever and a distinctive rash in children, and fever often with hepatitis and arthralgia in adults. B19V infection is also associated with pure red-cell aplasia from persistent infection in immunocompromised patients, and *hydrops fetalis* (a serious condition of the fetus) in pregnant women[1]. In addition, B19V infection has also been associated with other serious conditions such as inflammatory cardiomyopathy and the induction of autoimmune or autoimmune-like disease (short or long term)[5-8].

B19V is a single-stranded nonenveloped DNA virus of 5596 nucleotides, with an internal coding region flanked by palindromic sequences capable of forming terminal hairpin structures. B19V replicates in erythroid progenitors using a rolling hairpin mechanism[9]. The viral genome encodes six protein products: VP1 and VP2 which compose the viral capsid, NS1, the main replication protein, and three smaller nonstructural proteins[10-12]. Viral replication utilizes cellular factors, proposed to be coordinated by the viral NS1 protein[13-15], and is thought to make use of the terminal hairpins[13, 15] **(Fig. 1A**). Extension of the 3' end of one terminal hairpin by a cellular polymerase results in replication of the majority of the viral genome (Step 1, **Fig. 1A**), and NS1 is thought to bind to repeat sequences (NSBE1-4, **Fig. 1B**) and nick or cleave in one strand at the trs (the "terminal resolution site", Step 2, **Fig. 1A),** producing a new 3' end that can be used to prime synthesis of the remaining viral DNA (Step 3, **Fig. 1A**)[14]. Based on amino acid sequence and homology to other parvoviral replication proteins, B19V NS1 (NS1) is predicted to contain both nuclease and helicase domains involved in B19V replication. NS1 also contains a C-terminal domain involved in the protein's promoter transactivation activity[16] which acts upon its own viral promoter, p6, as well as several host promoters[17-21]. Furthermore, the genome of B19V is known to insert into host DNA, a reaction likely involving double-stranded DNA (dsDNA) recognition and strand nicking by NS1 followed by host DNA repair[22].

Prior studies show that the isolated N-terminal nuclease domain of B19V NS1 (NS1-nuc) binds sequence-specifically to the NSBE sequences in dsDNA, and is also responsible for sequence-specific DNA cleavage (or nicking) in single-, but not double-stranded DNA, at the trs site[23]. Since DNA encountered by NS1 is double-stranded, it is assumed that binding of NS1 to its target sequences in dsDNA induces strand separation nearby, allowing for the endonuclease activity of NS1 to cleave at the trs site. However, how NS1 recognizes its target sequences in dsDNA as well as at the nicking site are currently unknown, as is the mechanism by which NS1 binding to DNA at the NSBE induces strand separation. To begin to answer these questions, we determined the high-resolution crystal structure of NS1-nuc. Two crystal forms were solved, one at 2.4 Å resolution (Form I) and another at 3.5 Å (Form II) resolution which shows binding of a single Mg^{2+} in the DNA nicking active site. Using homology modeling, three-dimensional models for 1) trs sequence recognition in single-stranded DNA (ssDNA), 2) DNA cleavage in ssDNA, and 3) dsDNA binding at the NSBE sequences are presented.

Materials and Methods

Protein expression and purification

Purification of Form I NS1-nuc (residues 2-176 of B19V NS1) free of purification tags was performed as previously described using an N-terminal 6xhistidine and MBP-tagged fusion protein with a TEV protease cleavage site between the tags and the NS1-nuc sequences[23]. Following cell lysis using an Avestin Emulsiflex C3, cell debris were pelleted and the cell-free lysate incubated with pre-equilibrated Talon resin (Clonetech, Inc.). The partially pure eluted protein was then incubated with a 1:1 molar ratio of TEV protease[24] overnight at 4 °C. NS1-nuc free of MBP- and his-tags was then further purified with DEAE and Heparin FPLC (GE, Inc.). A longer construct of NS1-nuc (residues 2-209 of B19V NS1) was used in the Form II crystals and was prepared with an N-terminal 6xhistidine tag and TEV cleavage site and expressed in Tuner (DE3) cells overnight at 17°C following induction. Cells were lysed using an Avestin Emulisflex C3, then centrifuged to pellet cell debris. Purification proceeded with Talon resin (Clonetech, Inc.) chromatography followed by DEAE FPLC (GE). Purified protein was dialyzed into 0.1 M bis-tris-propane

pH 9.5, 150 mM NaCl, 1 mM 2-mercaptoethanol, and 50% glycerol, aliquoted, flash frozen in liquid nitrogen, then stored at -80°C until needed.

Crystallization, x-ray diffraction data collection, structure solution, and refinement

Crystallization proceeded using the hanging drop vapor diffusion method. NS1-nuc protein (Form I is free of the MBP tag, Form II retains the his-tag) was dialyzed extensively against 0.1 M bis-tris-propane pH 9.5, 150 mM NaCl, 1 mM 2-mercaptoethanol and concentrated to 5-10 mg/ml. Form I crystals appeared after several weeks at 17°C in the crystallization solution containing 15% PEG 3350, 0.1 M sodium citrate pH 4.5, 0.1 M NaCl, and 0.1 M LiCl and a 7:1 molar ratio of NS1-nuc to NSBE containing DNA (5'TCGCCGCCGGTAGGCGGGACTT, 5'AAGTCCCGCCTACCGGCGGCGA. For data collection, these crystals were exchanged into cryoprotectant (15% PEG 3350, 0.1 M sodium citrate pH 4.5, 0.1 M NaCl, 0.1 M LiCl, and 30% glycerol) prior to flash freezing in liquid nitrogen. Form II crystals appeared with the crystallization solution consisting of 2.5 M NaCl, 0.1 M Tris-HCl pH 7.0, and 200 mM $MgCl_2$ after several weeks at 4°C. For x-ray diffraction data collection, Form II crystals were harvested and exchanged into 2.5 M NaCl, 0.1 M Tris-HCl pH 7.0, and 30% glycerol then flash frozen and stored in liquid nitrogen. X-ray diffraction data collection was performed at SSRL BL 9-2 at 100K using Blu-Ice software[25]. Data processing, including integration, scaling, and merging, was performed with iMosflm[26] and SCALA[27, 28]. Structure solution of Form I NS1-nuc was performed using molecular replacement in PHASER[29] within the PHENIX software suite[30], and by searching for one copy of the nuclease domain using coordinates from PDB accession code 6USM[31]. Structure building and refinement proceeded through an iterative process using COOT[32, 33] and refinement using PHENIX[30, 34-36]. Solution of Form II proceeding using the same procedure, however with the Form I structural coordinates as the search model. Refinement of Form II made use of Form I as a reference structure, as well as secondary structure geometry restraints in the PHENIX software suite[30]. RMSD calculations were performed with UCSF Chimera[37], PyMol (Schrodinger), and DALI[38]. Images of structural models and electron density were prepared with UCSF

Chimera[37] and PyMol (Schrodinger). Electrostatic calculations performed with the software APBS[39] in PyMol (Schrodinger). Topology diagrams were made with PDBsum[40].

Sequence variation analysis of NS1-nuc

Sequences (195 total) of Human Parvovirus B19 NS1 (residues 1-176) were extracted from NCBI using BLASTp and the sequence of Form I NS1-nuc as the search sequence. WebLOGO[41] was used to create a figure to display amino acid sequence variations.

Results and Discussion

Structure Solution, Refinement, and Overall Analysis

Crystallographic data and structure refinement statistics are shown in **Table 1**. X-ray diffraction data were collected, scaled, and truncated to resolutions based on scaling, signal to noise, cross-correlation analysis, and final structure and map quality after structure refinement. These resolutions are 2.4 Å for Form I, where the highest resolution shell shows I/σ of 2.4, $CC_{1/2}$ of 38.5%, and final R_{work} and R_{free} of 19.9% and 25.7%, respectively. In the case of Form II, data were truncated at 3.5 Å where the highest resolution shell shows I/σ of 1.67, $CC_{1/2}$ of 89%, and final R_{work} and R_{free} of 26.1% and 29.6%, respectively. Residues corresponding to 2-173 of the NS1 sequence were included in the final refined model of Form I, however electron density for DNA was not evident despite its addition to the protein solution used in the Form I crystallization experiments. The additional residues on the C-terminus of the NS1-nuc construct used for Form II crystallization (residues 175-209) were also not evident in the electron density maps, possibly due to protein degradation during crystallization (**Fig. 1C**), and therefore the final model contains residues 2-174 of the NS1 sequence. **Figure 1D** shows two views of the NS1-nuc ribbon diagram of Form I, and **Figures 1E-F** show secondary structural elements and other notable structural features. The largest differences between Forms I and II NS1-nuc structures include a variation in the trace at residues 24-30 (**Fig. 2A-B**), as well as the absence of residues 127-128 in Form II (**Fig. 2A**), which are residues implicated in dsDNA binding (discussed further below). These structural differences likely originate from crystal

packing interactions since close packing with neighboring NS1-nuc copies occurs at these residues in all cases except residues 127-128 of Form II, and is distinct in each crystal form. These distinct interfaces likely stabilize different conformations of otherwise more mobile segments of NS1-nuc. The Form I map contains additional electron density between neighboring copies of NS1-nuc (**Fig. 2C**). Of the possible solvent molecules present in the crystallization solution citrate was found to fit best, based on refined R_{free}, refined temperature factors, and fit to map. (**Fig. 2C**). The citrate molecule is within hydrogen bonding (dashed lines, **Fig. 2C**) and/or salt bridging distance (note the crystallization conditions include pH 4.5) of appropriate groups in the two neighboring NS1-nuc copies, including the active site residues (His81, His83, and Tyr141) of one copy, and Lys119, Tyr130, and Thr97 in another. Both citrate locations overlap sites of nucleotide binding in the model of ssDNA binding created below. In the Form II crystal structure, significant (*i.e.* 3σ) positive difference density was also seen, and is located in the position occupied by Zn^{2+} in a previously determined structure of NS1-nuc (PDB accession code 6USM[31]) suggesting divalent cation binding (2Fo-Fc at 1σ map in blue, 1Fo-Fc map in magenta at 3σ when Mg^{2+} is omitted from the model, **Fig. 2D**). Mg^{2+} was modeled into this position (yellow sphere, **Fig. 2D**) since crystallization conditions contained 200 mM Mg^{2+}.

Comparison to other parvoviral Rep protein nuclease domains

Form I NS1-nuc was solved by molecular replacement using the NS1-nuc domain found in the PDB file with accession code 6USM[31]. After refinement these two structures of NS1-nuc were found to be very similar, with an RMSD of 0.8 Å over 164 residue C_α atoms (**Fig. 3A)**, and the largest differences in structure were found in loop segments at the exterior of the protein (thick red ribbons, **Fig. 3B**). The loop segment containing residues 147-148 is not present in the 6USM coordinates, but is well ordered in Form I NS1-nuc and contains a cis-proline (**Fig. 3C**). In addition, residues at the C-terminus of the domain are truncated to 171 in 6USM, but extend to 175 in Form I NS1-nuc. A single Zn^{2+} is located in the active site of 6USM in a nearly identical position as the Mg^{2+} ion modeled in Form II NS1-nuc (**Fig. 3D**). Differences in the amino acid sequences of NS1-nuc construct used in our studies and that of 6USM occur in thirteen positions (red

text, **Fig. 3E**). These originate from differences in biologically relevant sequences present in the NCBI data base (6USM follows the sequence of a laboratory isolate, Genbank entry AAG00943, Form I and II NS1-nuc follow the sequence of an isolate from a blood bank, Genbank entry ABN45789.1), and represent the two major variants present in NCBI (**Fig. 3F-G**). These substitutions are largely conservative in nature, and no large perturbations to the two structures are found at these positions.

Figure 4 compares the structure of NS1-nuc (Form I, blue, **Fig. 4A-C**) to homologous parvoviral structures present in the PDB originating from Adeno Associated Virus 2 (AAV2, PDB accession code 5DCX[42], magenta, **Fig. 4A**), Minute Virus of Mice (MVM, PDB accession code 3WRN[43], yellow, **Fig. 4B**), and Human Bocovirus (HBoV, PDB accession code 4KW3[44], green, **Fig. 4C**). Ribbon diagrams of pairwise superpositions (using C_α atoms) are shown in **Figures 4A-C**, with root mean square deviation values (RMSD using C_α atoms) mapped onto the structure of NS1-nuc (Form I) in **Figures 4D-F** (thicker, redder lines indicate greater RMSD, see legend in Å below each ribbon diagram). Most differences in structure occur to the positioning of loops and α-helices behind (as shown in the figure) the central β-sheet containing the active site histidine residues (marked in red in **Fig. 4A-C**). The structure of the AAV2 homolog contains an additional segment located between the N-terminal nuclease domain and the central helicase domain of AAV2 Rep (the homolog of B19V NS1), which are not present due to sequence truncation in the other domain structures (arrow, **Fig. 4A**). The overall RMSD values between NS1-nuc (Form I) and the other three structures are very similar, with 3.2 Å, 3.2 Å, and 3.0 Å over 163 C_α atoms of B19V NS1-nuc and the nuclease domains of AAV2 Rep, MVM NS1, and HBoV NS1, respectively.

Prediction of nick site ssDNA recognition and cleavage by NS1-nuc

A structure of a more distantly related viral replication enzyme (PDB accession code 6WE1[45], from Wheat Dwarf Virus, WDV)[46] provides the basis for a model of sequence-specific ssDNA recognition and nicking by NS1-nuc. The C_α atoms of the active site residues (His81, His83, Tyr141 in Form I NS1-nuc and His59, His61, Phe106 in WDV Rep-nuc, which contains the active site mutation Y106F) were used to superimpose the two structures, and the position of the ssDNA bound to WDV Rep-nuc identified (yellow,

Fig. 5A). A feature of the WDV Rep-nuc structure, the "ssDNA bridging motif"[46] (orange, **Fig. 5A**), is directly involved in ssDNA binding, making numerous interactions to the bases of several nucleotides and bridging the 5' and 3' ends of the bound ssDNA. In NS1-nuc, an insert is found in this position (magenta, **Fig. 5A**), which is predicted to interact with dsDNA (discussed further below), but also likely interacts with ssDNA. **Figure 5B** shows the electrostatic potential maps of NS1-nuc (mapped onto the protein surface in on the left, and as a field on the right), where blue indicates high positive charge, and red indicates high negative charge. The predicted ssDNA binding face (shown by the position of ssDNA taken from the alignment with the WDV Rep-nuc structure, magenta in left, yellow in right), shows a high degree of positive charge consistent with binding to negatively charged DNA.

The model of NS1-nuc bound to ssDNA created from the superposition with the WDV Rep-nuc structure allows for the prediction of residues likely to be involved in ssDNA recognition (light blue, red, and dark blue spheres, **Fig. 5C**). Only His9 of NS1-nuc (H9, red boxes, **Fig. 5C**) and His91 of WDV Rep-nuc (H91, **Fig. 5D**) appear to be conserved between the two structures. The model of NS1-nuc bound to nick site ssDNA also provides for the opportunity to model atoms in the active site **(Fig. 6A-B**). Both WDV Rep-nuc and NS1-nuc are members of the HUH nuclease superfamily, which require a divalent cation for DNA cleavage activity[47]. Prior work with NS1-nuc showed that Mg^{2+}, Co^{2+}, Ni^{2+}, and Mn^{2+} but not Zn^{2+}, Ca^{2+}, or Cu^{2+} confer DNA cleavage activity. Both the Mg^{2+} (from Form II NS1-nuc) and Zn^{2+} (from NS1-nuc in PDB accession code 6USM[31]) are bound in the same position in the active site, therefore the difference in activity with these two ions may derive from different chemical properties such as ligation geometries and ability to polarize ligated atoms[48-50]. In the current model of NS1-nuc bound to ssDNA, the Mg^{2+} (green sphere, **Fig. 6A**) is positioned near a non-esterified oxygen of the scissile phosphate (SP, the bond to be cleaved in the nicking/nuclease reaction). The distance between the modeled phosphate oxygen and Mg^{2+} is 1.4 Å, somewhat closer than the typical ligation distance for Mg^{2+} to oxygen ligands (1.9-2.1 Å)[51], but small adjustments in the position of the bound DNA could easily bring this distance to a more optimal value. The Mg^{2+} is also within ligation distance of the side chains of the active site residues His81 (2.5 Å) and His83 (2.5 Å)(**Fig. 6A**), as well as that of Glu72 (2.3 Å)(the relatively longer ligation distances may be

due to coordinate error in the relatively low resolution of the Form II structure). WDV Rep-nuc also possesses a nearby glutamic acid residue in the active site capable of ligation to the Mg^{2+} (Glu110, **Fig. 6B**). Divalent cation-dependent nucleases catalyze DNA cleavage by any or all of the following: 1) polarization of the nucleophile (in this case, the phenolic oxygen of Tyr141) to increase its nucleophilicity (this often occurs via ligation to the divalent cation and may result in deprotonation of the nucleophile), 2) stabilization of the transition state after nucleophilic attack (often via divalent cation ligation to a non-esterified oxygen of the scissile phosphate), and 3) stabilization of the leaving group (the O3') following bond breakage (often via direct ligation to the divalent cation or protonation from a divalent cation ligated water molecule)[48, 52-54]. In our NS1-nuc structure with modeled ssDNA, we find that the Mg^{2+} is positioned well to stabilize the transition state after nucleophilic attack via its predicted ligation to a non-esterified oxygen of the scissile phosphate (**Fig. 6A**). In the case of stabilization of the leaving group, the ssDNA bound model does not predict a direct ligation of the O3' leaving group to the Mg^{2+}, but protonation by a Mg^{2+} ligated water molecule could be possible. The model also does not predict direct ligation of the nucleophile (the oxygen of the side chain of Tyr141) to Mg^{2+}. However, divalent cation dependent nucleases also accelerate DNA cleavage by organizing reactive groups in the active site into a geometry favorable for nucleophilic attack and bond breakage[48, 52-54]. This geometry includes: 1) positioning the nucleophile within van der Waals radii of the phosphorus atom (the atom to be attacked by the nucleophile), and 2) arranging the three atoms of the bond making and breaking reaction (the attacking group, the phosphorus atom, and the leaving group) in an "in-line" configuration such that the angle between them is 180°[54]. We find in our model that the Tyr141 hydroxyl oxygen atom (the nucleophile of the DNA nicking reaction) is 3.6 Å from the phosphorus atom of the phosphodiester to be cleaved (the estimated van der Waals radii of the two atoms is 3.3 Å, 1.5 Å for oxygen, and 1.8 Å for phosphorus[55]), and the angle between the nucleophile, phosphorus atom, and leaving group (O3' of the 5' nucleotide) is 147°. Hence the active site moieties are poised in an appropriate position for the catalytic reaction to occur. Finally, the terminal amine of the Lys145 side chain is within hydrogen bonding distance to the Tyr141 OH nucleophile (2.9 Å), suggesting

a contribution to the catalytic reaction by acting as a general base to accept a proton from the nucleophile, and/or to stabilize a negative charge on the nucleophile following proton loss.

The ssDNA in the model also overlaps with the citrate molecules bound to NS1-nuc in the Form I structure. In this form, each asymmetric unit contains one NS1-nuc and one bound citrate molecule. However, because the citrate molecule binds between two copies of NS1-nuc, it has two distinct binding sites in a single copy of NS1-nuc. One location is very near the active site (**Fig. 6C**). Two carboxylate groups of citrate bind near the phosphate positions of nucleotides +1 and -1 in the modeled ssDNA, consistent with the affinity for negatively charged moieties in these locations. The second citrate binding site is found near the nucleotide at the -6 position of the modeled ssDNA (**Fig. 6D**). The citrate molecule is bound in a pocket formed on the surface of NS1-nuc, and ~4 Å closer to NS1-nuc than the -6 nucleotide, possibly predicting the true path of the bound ssDNA in this region, and thus implicating Tyr130 and Thr97 (**Fig. 2C**) in ssDNA binding as well.

The structure of WDV Rep-nuc bound to nick site DNA (**Fig. 7A**) suggests that sequence-specific recognition occurs through a combination of direct readout, consisting of hydrogen bonds and van der Waals interactions to the chemically distinct portions of the DNA bases, as well as indirect readout, derived from the sequence-specific energetics of DNA structure and base stacking[46]. Direct readout of the nick site DNA sequence is accomplished via hydrogen bonds between residues of WDV Rep-nuc and bases of nucleotides at the 2, -5, and -6 positions (see nucleotide numbering in **Fig. 7B**)[46]. Indirect readout of the nick site DNA sequence is suggested by the distorted U-shape of the bound ssDNA, as well as the base pairs between Ade1 and Thy-4 (a Watson-Crick bp) and Thy-1 and Ade-3 (a non-Watson-Crick bp). To understand how NS1-nuc recognizes its nick site in ssDNA, we substituted the DNA sequence of the ssDNA in the NS1-nuc/ssDNA model derived from the superposition with the WDV Rep-nuc/ssDNA structure (**Fig. 7C-D**). First, in terms of direct readout, the low amino acid sequence conservation in the DNA binding site residues makes direct readout contacts difficult to predict between NS1-nuc and the bound ssDNA, with the possible exception of the 2 position of the nick site DNA (which is Cyt in both viral nick sites, **Figs. 7B,D**). WDV Rep-nuc recognizes this base with hydrogen bonds from protein backbone atoms to the

base pairing atoms Cyt2 (**Fig. 7E**). In the model of NS1-nuc bound to nick site DNA, hydrogen bonds between the side chain of Asp133 and the O2 and N3 of Cyt2 would be possible with small adjustments (distances are 2.0 Å and 3.0 Å from the Asp133 carboxylate atoms to the O2 and N3 Cyt2 atoms, respectively, **Fig. 7F**), predicting a role for Asp133 in sequence-specific recognition at the 2 position of the nick site DNA. In addition, Arg5 approaches Cyt2 from behind the base, and Phe131 may form a π-hydrogen bond[56] to the NH_2 at the 4 of Cyt2 in this model, suggesting that these residues may also be important in DNA nick site sequence discrimination (**Fig. 7F**). In terms of indirect readout, the Watson-Crick bp between nucleotides at the 1 and -4 positions (Ade1 and Thy-4 in the WDV sequence, **Fig. 7G**) may be conserved as these nucleotides are Cyt1 and Gua-4 in the B19V nick site sequence (see model in **Fig. 7H**). The non-Watson-Crick base pair between Thy-1 and Ade-3 in the WDV structure occurs with a single hydrogen bond between the N3 of Thy-1 and N3 of Ade-3 (**Fig. 7I**). In the B19V nick site sequence, these bases are Ade-1 and Ade-3, and the structural model predicts that a single hydrogen bond is possible (after some adjustment due to the larger size of of the Ade base at the -1 position) between the N6 of Ade-1 and N3 of Ade-3 (green and dark blue sticks, **Fig. 7J**). Finally, in both models, a pyrimidine base is found at the turn of the bound DNA (Thy-2 in WDV, **Fig. 7A-B**, and Cyt-2 in B19V, **Fig. 7C-D**), which may be an important factor in indirect readout of the DNA sequence due to the formation of stacking interactions with the nucleotide at position -3 (an Ade in both cases).

A study of sequence preferences at the B19V nick site was performed with the NS1-nuc domain[23], which found the greatest preferences at nucleotides Cyt2, Cyt1, and Ade-1, with Cyt2 being most important, followed by Cyt1 and Ade-1. The nucleotides at these positions were also found to be the most important for nicking by WDV Rep-nuc using a sequence-specificity selection approach (HUH-seq)[46], and in the same order of importance. In the case of B19V, exhaustive substitutions of the nick site sequence were not tested, instead only mutation to the Watson-Crick base pairing partner was examined for cleavage by NS1-nuc. However, the large decrease in DNA cleavage activity of NS1-nuc found as a result of changing Cyt2 to a Gua[23] may be explained by the predicted tight binding pocket surrounding Cyt2 (**Fig. 7F**), since a Gua base would be too large to fit into this pocket and could not form the same interactions with Asp133 and Phe131.

The next most significant decrease in cleavage activity occurred with the substitution of Cyt1 to Gua. Since the model predicts a base pairing interaction with Gua-4 (**Fig. 7H**), it is clear that substitution of the Cyt with a Gua would disrupt this pairing. Strangely, substitution of Gua-4 with Cyt resulted in an increase in DNA cleavage by NS1-nuc. The Cyt-Cyt interaction (between Cyt1-Cyt-4), while not predicted to be favorable, may result in less disruption to the structure of the bound DNA due to the relatively small size of the Cyt bases (compared to two Gua bases as in Gua1-Gua-4). Finally, substitution of Ade-1 with Thy diminished cleavage by approximately 50%[23]. This substitution would disrupt the predicted hydrogen bonding with Ade-3, between the N6 of Ade-1 and the N3 of Ade-3 (green and blue sticks, **Fig. 7J**). A Thy at position -1 would not offer a hydrogen bond donor to take the place of the N6 of Ade-1, but interestingly, is the sequence seen in the WDV structure (Thy-1, yellow, **Fig. 7J).** The fact that substitution of Ade-1 with Thy is detrimental to DNA cleavage by NS1-nuc suggests a different configuration of these bases in the two structures, which would be necessary to accommodate the wild type Ade-1-Ade-3 interaction predicted in the bound B19V ssDNA. Conversely, substitution of Ade-3 to Thy had little effect, possibly due to the availability of the O2 of a Thy to take the place of the N3 of Ade-3 in the predicted hydrogen bonding interaction (**Fig. 7J**). It may also be that substitutions near the nicking site (in -1, 1, and 2) are more sensitive due to the requirement for precise positioning of atoms in the active site in order to achieve optimal cleavage activity.

Prediction of dsDNA binding by NS1-nuc

In addition to binding to ssDNA, B19V NS1 must also bind to target sequences in dsDNA known as NSBE, for NS1 Binding Elements. These sequences are located near the nick site (*i.e.* the trs) (**Fig. 1B**). It is the N-terminal nuclease domain of NS1 (*i.e.* NS1-nuc) which is also responsible for recognition of these sequences[23]. The structure of the nuclease domain from AAV5 Rep (AAV5 Rep-nuc) bound to dsDNA containing RBE repeats (PDB accession code 1RZ9[57]), which are sequences analogous to the NSBE of B19V, provides a framework for modeling the interactions of NS1-nuc with the NSBE sequences in dsDNA. **Figure 8A** shows a map of the AAV5 Rep-nuc domains bound to the RBE sites in dsDNA in this crystal

structure. Five copies of AAV5 Rep-nuc (colored ovals, **Fig. 8A**) bind the five four-base (imperfect) RBE repeats (boxed bp, **Fig. 8A**), and follow the helical twist of the DNA since each copy of AAV5 Rep-nuc interacts with bases in both the major and minor grooves in an equivalent manner (**Fig. 8B**). Interactions to the DNA by each copy of AAV5 Rep-nuc are not confined to a single quartet, but instead overlap. For example, one copy (yellow, **Fig. 8A**) interacts with base pairs of the first RBE quartet (boxed sequences, counting from right to left, **Fig. 8A**) in the major groove, as well as base pairs in the minor groove in the second RBE quartet. The next AAV5 Rep-nuc copy (green, **Fig. 8A**) interacts with the second RBE quartet bases via the major groove, and third quartet bases via the minor groove. Each AAV5 Rep-nuc copy therefore interacts with 2 quartets, and each quartet interacts with two copies of AAV5 Rep-nuc. Two segments of AAV5 Rep-nuc interact with dsDNA, and these same segments are largely conserved in the structure (but not sequence) of NS1-nuc (**Fig. 8C**). AAV5 Rep-nuc interacts with the minor groove of dsDNA via residues 101-111 (corresponding to 93-103 in NS1-nuc, **Fig. 8D**), and the major groove using residues 135-142 (corresponding to 124-130 in NS1-nuc, **Fig. 8D**).

Although identification of the likely dsDNA binding residues in NS1-nuc is relatively straightforward, prediction of exactly where NS1-nuc binds to the B19V DNA sequences is not. **Figure 8E** shows the NSBE and trs sequences of the B19V origin of replication. Rather than five quartet repeats as in AAV5, the sequence appears to have four octet repeats separated by two base pair spacings. The octet sequences are very GC-rich and were identified as NS1 binding sites in a prior binding study[58]. Additional DNA binding studies showed that as many as 5-7 copies of NS1-nuc can bind to dsDNA containing all four NSBE sequences[23]. But the differences in repeat size, spacing, and in DNA and protein sequence made the exact positioning and modeling protein-DNA interactions between NS1-nuc and DNA not possible. NS1-nuc may bind the NSBE sequences in a manner similar to that of AAV5 Rep-nuc, using overlapping quartets rather than four separate octets, but this has yet to be shown. Prior binding investigations indicated that the three octets closest to the trs are most important for NS1-nuc binding[23, 59]. These sequences total 28 bp, or 7 quartets, consistent with the 7 copies of NS1-nuc binding in the AAV5 Rep-nuc pattern. However, more information on NS1 interactions with dsDNA containing the NSBE sequences will be necessary to assign

the exact positioning and amino acid-nucleotide interactions between NS1-nuc and the DNA sequence.

Prior binding studies indicated that NS1-nuc binds DNA cooperatively, as evidenced by the shape of binding isotherms[23]. Cooperativity indicates that the binding of DNA by one copy of NS1-nuc increases the affinity of subsequent copies of NS1-nuc to the same DNA. This effect can occur when favorable protein-protein interactions occur between copies of NS1-nuc on the DNA, and/or when distortions made to the DNA upon binding of one copy facilitate the binding of subsequent copies by negating the requirement for the DNA distortion (which costs energy) by those subsequent protein copies. Cooperative DNA binding by AAV5 Rep-nuc is also suggested by the shape of the binding isotherm shown in Hickman, *et al.* (Figure 4B, lowest panel)[60]. However, no protein-protein interactions are found between copies of AAV5 Rep-nuc bound to dsDNA in the crystal structure, and similarly, no protein-protein interactions are predicted when the NS1-nuc structure is superimposed onto each of the five copies of AAV5 Rep-nuc bound to dsDNA (the closest approach between different copies of NS1-nuc is 8 Å). It is possible that residues beyond the C-terminus of AAV5 Rep-nuc or NS1-nuc extend enough to allow contacts between neighboring copies of AAV5 Rep-nuc or NS1-nuc when bound to DNA. Indeed, when a structure of AAV2 Rep-nuc containing an 13 additional residues at its C-terminus (PDB accession code 5DCX)[42, 61] is superimposed onto each of the five AAV5 Rep-nuc copies of the AAV5 Rep-nuc/dsDNA structure, the additional residues extend far enough to form interactions with neighboring AAV5 Rep-nuc copies. However, these residues were not present in the prior binding studies which showed strongly cooperative DNA binding by both AAV5 Rep and B19V NS1 nuclease domains[23, 60], suggesting the existence of some other mechanism of cooperativity beyond protein-protein interactions.

The other likely origin of cooperative DNA binding involves distortions in the bound DNA, where binding of each copy of a DNA binding protein distorts the DNA in a way which results in an increase in binding affinity of subsequent copies. The DNA conformation in the AAV5 Rep-nuc/dsDNA structure indeed shows distortions from ideal B-form DNA that follow the pattern of the quartet repeat[60]. Since binding sites of NS1-nuc to NSBE likely overlap as they do in RBE binding by AAV5 Rep-nuc, distortions induced by one copy of NS1-nuc could facilitate binding of neighboring copies by presenting DNA pre-

distorted in the manner which complements the protein-DNA binding interface, and eliminates the cost of DNA distortion to the binding of that subsequent copy to the DNA. Hence, we predict based on this analysis that DNA distortions play a large role in the cooperativity seen in NSBE binding by NS1-nuc.

Finally, we show the relative positions of the two types of DNA binding interfaces in **Figure 8F**. Regions of the red copy of NS1-nuc which are predicted to interact with dsDNA are shown in dark purple, and residues of the DNA nicking active site shown in dark blue. Though the ssDNA (green) and dsDNA (white) overlap minimally, it appears unlikely that NS1-nuc could bind to both simultaneously. The ssDNA interface on one copy of NS1-nuc is also located between adjacent copies of NS1-nuc on the dsDNA (modeled based on binding of AAV5 Rep-nuc binding to dsDNA). Note that the trs site is located distal to the dsDNA binding sites (**Fig. 8E**). In **Figure 8F**, with the NS1-nuc domains bound to the NSBE sequences, the trs would be ~1 turn of the DNA (if B form) to the right of the red copy of NS1-nuc. Since NS1-nuc binds and cleaves the trs only when single stranded, some mechanism of strand opening, presumably following NS1 binding to the NSBE, must occur prior to trs cleavage. The current best model for NS1 action at the B19V origin of replication derives from studies of AAV2 Rep[62]. A recent structural investigation showed that Rep68, which contains a SF3 helicase and ATPase domain C-terminal to the nuclease domain (corresponding to residues 210-481 in B19V NS1), forms rings which encircle bound ss and dsDNA. The authors propose a model[62], similar to one proposed previously[60], where the nuclease domains bring multiple copies of Rep to the origin via interaction with the RBE sequences, leading to formation of a heptameric ring structure which induces strand separation (possibly concurrent with loss of one Rep to form a hexameric ring). The strand separation then allows the trs to be cleaved by an available nuclease domain. Further studies will be required to confirm such a mechanistic model for B19V NS1.

Acknowledgments:

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health via Grant T32GM008659 (to J.L.S.), and the University of Arizona

Technology Research Initiative Fund (TRIF).Coordinates and structure factor amplitudes for Form I and II NS1-nuc structures have been deposited into the PDB under accession codes 7SZY and 7SZX.

Competing Interest: The authors declare that they have no conflicts of interest with the contents of this article.

References

[1] Young, N. S., and Brown, K. E. (2004) Parvovirus B19, *N Engl J Med 350*, 586-597.
[2] Cossart, Y. E., Field, A. M., Cant, B., and Widdows, D. (1975) Parvovirus-like particles in human sera, *Lancet 1*, 72-73.
[3] Pattison, J. R., Jones, S. E., Hodgson, J., Davis, L. R., White, J. M., Stroud, C. E., and Murtaza, L. (1981) Parvovirus infections and hypoplastic crisis in sickle-cell anaemia, *Lancet 1*, 664-665.
[4] Anderson, M. J., Jones, S. E., Fisher-Hoch, S. P., Lewis, E., Hall, S. M., Bartlett, C. L., Cohen, B. J., Mortimer, P. P., and Pereira, M. S. (1983) Human parvovirus, the cause of erythema infectiosum (fifth disease)?, *Lancet 1*, 1378.
[5] Tsay, G. J., and Zouali, M. (2006) Unscrambling the role of human parvovirus B19 signaling in systemic autoimmunity, *Biochem Pharmacol 72*, 1453-1459.
[6] Colmegna, I., and Alberts-Grill, N. (2009) Parvovirus B19: its role in chronic arthritis, *Rheum Dis Clin North Am 35*, 95-110.
[7] Franssila, R., and Hedman, K. (2006) Infection and musculoskeletal conditions: Viral causes of arthritis, *Best Pract Res Clin Rheumatol 20*, 1139-1157.
[8] Kerr, J. R. (2015) The role of parvovirus B19 in the pathogenesis of autoimmunity and autoimmune disease, *J Clin Pathol 69*, 279-291.
[9] Tattersall, P., and Ward, D. C. (1976) Rolling hairpin model for replication of parvovirus and linear chromosomal DNA, *Nature 263*, 106-109.
[10] Cotmore, S. F., McKie, V. C., Anderson, L. J., Astell, C. R., and Tattersall, P. (1986) Identification of the major structural and nonstructural proteins encoded by human parvovirus B19 and mapping of their genes by procaryotic expression of isolated genomic fragments, *J Virol 60*, 548-557.
[11] Ozawa, K., Ayub, J., Hao, Y. S., Kurtzman, G., Shimada, T., and Young, N. (1987) Novel transcription map for the B19 (human) pathogenic parvovirus, *J Virol 61*, 2395-2406.
[12] St Amand, J., Beard, C., Humphries, K., and Astell, C. R. (1991) Analysis of splice junctions and in vitro and in vivo translation potential of the small, abundant B19 parvovirus RNAs, *Virology 183*, 133-142.
[13] Zhi, N., Mills, I. P., Lu, J., Wong, S., Filippone, C., and Brown, K. E. (2006) Molecular and functional analyses of a human parvovirus B19 infectious clone demonstrates essential roles for NS1, VP1, and the 11-kilodalton protein in virus replication and infectivity, *J Virol 80*, 5941-5950.
[14] Ozawa, K., Kurtzman, G., and Young, N. (1986) Replication of the B19 parvovirus in human bone marrow cell cultures, *Science 233*, 883-886.
[15] Berns, K. I. (1990) Parvovirus replication, *Microbiol Rev 54*, 316-329.
[16] Xu, P., Zhou, Z., Xiong, M., Zou, W., Deng, X., Ganaie, S. S., Kleiboeker, S., Peng, J., Liu, K., Wang, S., Ye, S. Q., and Qiu, J. (2017) Parvovirus B19 NS1 protein induces cell cycle arrest at G2-phase by activating the ATR-CDC25C-CDK1 pathway, *PLoS Pathog 13*, e1006266.
[17] Nakashima, A., Morita, E., Saito, S., and Sugamura, K. (2004) Human Parvovirus B19 nonstructural protein transactivates the p21/WAF1 through Sp1, *Virology 329*, 493-504.
[18] Moffatt, S., Tanaka, N., Tada, K., Nose, M., Nakamura, M., Muraoka, O., Hirano, T., and Sugamura, K. (1996) A cytotoxic nonstructural protein, NS1, of human parvovirus B19 induces activation of interleukin-6 gene expression, *J Virol 70*, 8485-8491.
[19] Duechting, A., Tschope, C., Kaiser, H., Lamkemeyer, T., Tanaka, N., Aberle, S., Lang, F., Torresi, J., Kandolf, R., and Bock, C. T. (2008) Human parvovirus B19 NS1 protein modulates inflammatory signaling by activation of STAT3/PIAS3 in human endothelial cells, *J Virol 82*, 7942-7952.

[20] Wan, Z., Zhi, N., Wong, S., Keyvanfar, K., Liu, D., Raghavachari, N., Munson, P. J., Su, S., Malide, D., Kajigaya, S., and Young, N. S. (2010) Human parvovirus B19 causes cell cycle arrest of human erythroid progenitors via deregulation of the E2F family of transcription factors, *J Clin Invest 120*, 3530-3544.
[21] Fu, Y., Ishii, K. K., Munakata, Y., Saitoh, T., Kaku, M., and Sasaki, T. (2002) Regulation of tumor necrosis factor alpha promoter by human parvovirus B19 NS1 through activation of AP-1 and AP-2, *J Virol 76*, 5395-5403.
[22] Janovitz, T., Wong, S., Young, N. S., Oliveira, T., and Falck-Pedersen, E. (2017) Parvovirus B19 integration into human CD36+ erythroid progenitor cells, *Virology 511*, 40-48.
[23] Sanchez, J. L., Romero, Z., Quinones, A., Torgeson, K. R., and Horton, N. C. (2016) DNA Binding and Cleavage by the Human Parvovirus B19 NS1 Nuclease Domain, *Biochemistry 55*, 6577-6593.
[24] Tropea, J. E., Cherry, S., and Waugh, D. S. (2009) Expression and purification of soluble His(6)-tagged TEV protease, *Methods Mol Biol 498*, 297-307.
[25] McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M., and Kuhn, P. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines, *Journal of Synchrotron Radiation 9*, 401-406.
[26] Geoff, T., Battye, G., Kontogiannis, L., Johnson, O., Powella, H. R., and Leslie, A. G. W. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Cryst. D67*, 271-281.
[27] Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr D Biol Crystallogr 62*, 72-82.
[28] Evans, P. R. (1993) Data Reduction, *Proceedings of CCP4 Study Weekend*, 114-122.
[29] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.
[30] Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr D Biol Crystallogr 58*, 1948-1954.
[31] Tewary, S. K. (2020) Structure of nuclease domain of human parvovirus B19 non-structural protein 1 in complex with zinc.
[32] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.
[33] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.
[34] Grosse-Kunstleve, R. W., Terwilliger, T. C., Sauter, N. K., and Adams, P. D. (2012) Automatic Fortran to C++ conversion with FABLE, *Source Code Biol Med 7*, 5.
[35] Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., Richardson, D. C., Richardson, J. S., and Adams, P. D. (2012) Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution, *Acta Crystallogr D Biol Crystallogr 68*, 381-390.
[36] Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., and Urzhumtsev, A. (2013) Bulk-solvent and overall scaling revisited: faster calculations, improved results, *Acta Crystallogr D Biol Crystallogr 69*, 625-634.
[37] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.
[38] Holm, L. (2020) Using Dali for Protein Structure Comparison, *Methods Mol Biol 2112*, 29-42.
[39] Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G. W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2018) Improvements to the APBS biomolecular solvation software suite, *Protein Sci 27*, 112-128.
[40] Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S., and Thornton, J. M. (2018) PDBsum: Structural summaries of PDB entries, *Protein Sci 27*, 129-134.
[41] Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator, *Genome Res 14*, 1188-1190.
[42] Musayev, F. N., Zarate-Perez, F. (2015) Structural studies of AAV2 Rep68 reveal a partially structured linker and compact domain conformation.
[43] Tewary, S. K., Zhao, H., and Tang, L. (2014) Minute virus of mice non-structural protein-1N-terminal nuclease domain reveals a unique Zn2+ coordination in the active site pocket and shows a novel mode of DNA recognition at the origin of replication.
[44] Tewary, S. K., Zhao, H., and Tang, L. (2013) Crystal structure of the non-structural protein 1 N-terminal origin-recognition/nickase domain from the emerging human bocavirus.

[45] Tompkins, K., Litzau, L.A., Pornschloegl, L., Nelson, A.T., Evans III, R.L., Gordon, W.R. (2020) Wheat dwarf virus Rep domain complexed with a single-stranded DNA 8-mer comprising the cleavage site.
[46] Tompkins, K. J., Houtti, M., Litzau, L. A., Aird, E. J., Everett, B. A., Nelson, A. T., Pornschloegl, L., Limon-Swanson, L. K., Evans, R. L., Evans, K., Shi, K., Aihara, H., and Gordon, W. R. (2021) Molecular underpinnings of ssDNA specificity by Rep HUH-endonucleases and implications for HUH-tag multiplexing and engineering, *Nucleic Acids Res 49*, 1046-1064.
[47] Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., and Ton-Hoang, B. (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily, *Nat Rev Microbiol 11*, 525-538.
[48] Yang, W. (2011) Nucleases: diversity of structure, function and mechanism, *Q Rev Biophys 44*, 1-93.
[49] Kragten, J. (1978) *Atlas of metal-ligand equilibria in aqueous solution*, Halsted Press, Chichester, Eng.
[50] Pontius, B. W., Lott, W. B., and von Hippel, P. H. (1997) Observations on catalysis by hammerhead ribozymes are consistent with a two-divalent-metal-ion mechanism, *Proc Natl Acad Sci U S A 94*, 2290-2294.
[51] Harding, M. M. (1999) The geometry of metal-ligand interactions relevant to proteins, *Acta Crystallogr D Biol Crystallogr 55*, 1432-1443.
[52] Horton, N. C. (2008) DNA Nucleases, In *Protein-Nucleic Acid Interactions Structural Biology* (Rice, P. A., Correll, C.C., Ed.), pp 333-363, The Royal Society of Chemistry, Cambridge, UK.
[53] Horton, N. C., and Perona, J. J. (2002) Making the most of metal ions.[comment], *Nature Structural Biology 9*, 42-47.
[54] Gerlt, J. A. (1992) Mechanistic Principles of Enzyme-Catalyzed Cleavage of Phosphodiester Bonds, In *Nucleases, 2nd Ed.* (Lloyd, S., Linn, S., Roberts, R., Ed.), p 1, Cold Spring Harbor Press.
[55] Mantina, M., Chamberlin, A. C., Valero, R., Cramer, C. J., and Truhlar, D. G. (2009) Consistent van der Waals radii for the whole main group, *J Phys Chem A 113*, 5806-5812.
[56] Trachsel, M. A., Ottiger, P., Frey, H. M., Pfaffen, C., Bihlmeier, A., Klopper, W., and Leutwyler, S. (2015) Modeling the Histidine-Phenylalanine Interaction: The NH...pi Hydrogen Bond of Imidazole.Benzene, *J Phys Chem B 119*, 7778-7790.
[57] Hickman, A. B., Ronning, D.R., Perez, Z.N., Kotin, R.M., Dyda, F. (2003) Crystal Structure of AAV Rep complexed with the Rep-binding sequence.
[58] Raab, U., Beckenlehner, K., Lowin, T., Niller, H. H., Doyle, S., and Modrow, S. (2002) NS1 protein of parvovirus B19 interacts directly with DNA sequences of the p6 promoter and with the cellular transcription factors Sp1/Sp3, *Virology 293*, 86-93.
[59] Tewary, S. K., Zhao, H., Deng, X., Qiu, J., and Tang, L. (2014) The human parvovirus B19 non-structural protein 1 N-terminal domain specifically binds to the origin of replication in the viral DNA, *Virology 449*, 297-303.
[60] Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M., and Dyda, F. (2004) The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces, *Molecular cell 13*, 403-414.
[61] Musayev, F. N., Zarate-Perez, F., Bardelli, M., Bishop, C., Saniev, E. F., Linden, R. M., Henckaerts, E., and Escalante, C. R. (2015) Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker and Compact Domain Conformation, *Biochemistry 54*, 5907-5919.
[62] Santosh, V., Musayev, F. N., Jaiswal, R., Zarate-Perez, F., Vanderwinkel, B., Dierckx, C., Endicott, M., Sharifi, K., Dryden, K., Henkaerts, E., and Escalante, C. R. (2020) The CryoEM structure of AAV2 Rep68 in complex with ssDNA revewals a malleable AAA+ machine that can switch between oligomeric states, *BioRXiv*.

Tables

Table 1. Data collection and refinement statistics.

	Form I	Form II
PDB accession code	7SZY	7SZX
Site of data collection	SSRL BL9-2	SSRL BL9-2
Wavelength	0.97946 Å	0.97946 Å
Resolution range	45.06 Å - 2.4 Å (2.486 Å - 2.4 Å)	46.21 Å - 3.5 Å (3.625 Å - 3.5 Å)
Space group	P 2 21 21	P 32 2 1
Unit cell	45.0599 Å, 49.0999 Å, 74.9899 Å 90° 90° 90°	106.71 Å, 106.71 Å, 59.5904 Å 90° 90° 120°
Total reflections	13,532 (1322)	10,152 (976)
Unique reflections	6,799 (666)	5,099 (488)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Completeness (%)	98.42 (98.66)	98.70 (97.41)
Mean I/sigma(I)	7.21 (2.43)	5.37 (1.67)
Wilson B-factor	28.31	91.66
R-merge	0.1022 (0.5331)	0.07577 (0.4226)
R-meas	0.1446 (0.7539)	0.1072 (0.5977)
$CC_{1/2}$	0.981 (0.385)	0.993 (0.89)
CC*	0.995 (0.745)	0.998 (0.971)
Reflections used in refinement	6,796 (665)	5,074 (488)
Reflections used for R-free	333 (40)	508 (50)
R-work	0.1986(0.2629)	0.2605 (0.3421)
R-free	0.2566 (0.2891)	0.2956 (0.4067)
CC(work)	0.951 (0.778)	0.908 (0.759)
CC(free)	0.898 (0.601)	0.825 (0.685)
Number of non-hydrogen atoms	1439	1294
macromolecules	1337	1289
ligands	13	1
solvent	89	4
Protein residues	174	171
RMS(bonds)	0.003	0.004
RMS(angles)	0.61	0.87
Ramachandran favored (%)	95.93	94.01
Ramachandran allowed (%)	4.07	5.39
Ramachandran outliers (%)	0.00	0.60
Rotamer outliers (%)	1.43	0.00
Clashscore	2.28	6.09
Average B-factor	30.11	94.33
macromolecules	29.83	94.51
ligands	39.45	90.01
solvent	32.96	37.16

Statistics for the highest-resolution shell are shown in parentheses.

Figures

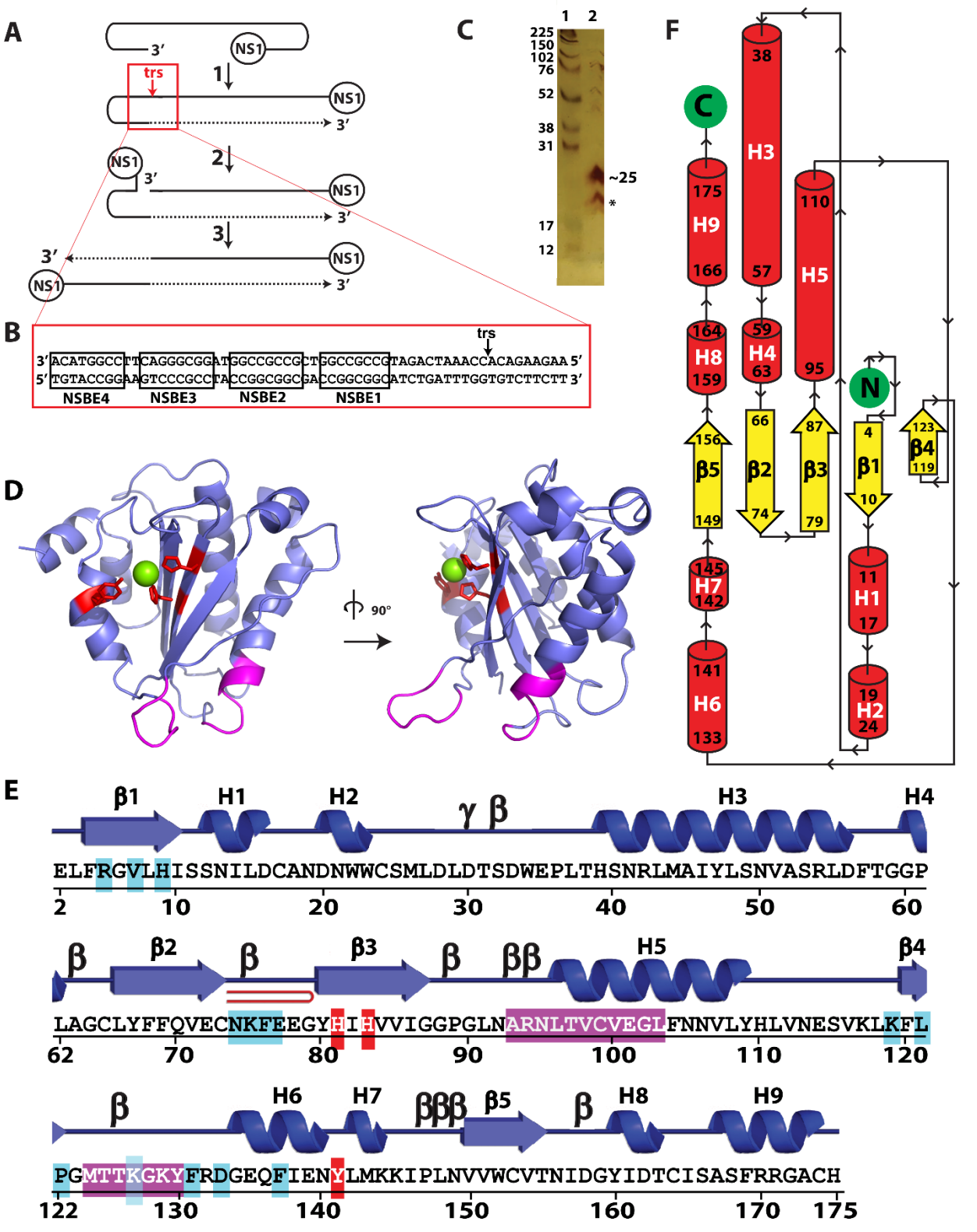

Figure 1. Overview of NS1-nuc function in replication and three-dimensional structure. A.

Replication using the cellular replication machinery is primed by the 3' folded over hairpin (or ITR). Upon double-stranded formation, transcription proceeds producing new copies of NS1. NS1 binds at the GC-rich sequences denoted NSBE1-4 and cleaves at the nearby terminal resolution site or trs, leaving NS1 covalently attached to the 5' end at the site of cleavage. Replication of the remaining segment of the genome is completed using the newly generated 3'OH and following unfolding of the terminal hairpin. **B.** DNA sequences at the viral origin of replication including the trs and NSBE sequences. **C.** Silver stained SDS-PAGE analysis of a drop containing Form II crystals (lane 2). The full-length construct is predicted to be 25 kDa, however a smaller species indicated by the asterisk (*) may be the form found in the crystals. **D.** Overview of Form I NS1-nuc with the positions of putative DNA cleavage active site residues in red and predicted dsDNA binding residues shown in magenta. Mg^{2+} identified in Form II shown as green sphere. **E.** Sequence, secondary structure, and notable features of Form I NS1-nuc. H1-H9: alpha helices, β1-β5: β-strands, β: β-turn, γ: γ-turn, red hairpin: β-hairpin, white letters in red boxes: putative DNA cleavage active site residues (*i.e.* active site residues), white in magenta boxes: predicted dsDNA binding residues, cyan boxes: predicted ssDNA binding residues. **F.** Topology diagram of Form I NS1-nuc with α helices as red cylinders, β strands as white arrows, N and C termini in green circles, and residue number of secondary structural elements in black.

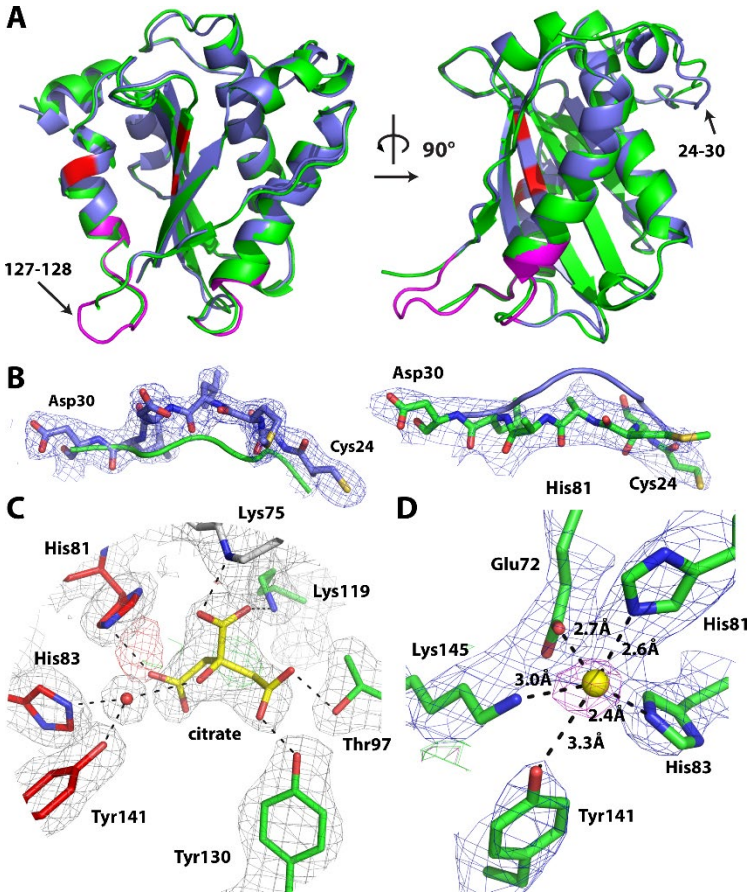

Figure 2. Comparison of the two new crystal forms of NS-nuc and binding sites of solvent molecule and Mg^{2+}. A. Two views of a cartoon representation of Form I (blue) and Form II (green). Putative active site residues are shown in red, predicted dsDNA binding sites in Form I shown in magenta. Arrows indicate regions disordered in Form II (left), and a region (residues 24-30) of deviation in the main chain position between the two forms (right). **B. Left:** Form I 2Fo-Fc map (at 1σ) with residues 24-30 of the Form I model (blue, red, yellow). The main chain trace of Form II is shown in green. **Right:** Form II 2Fo-Fc map (at 1σ) with residue 24-30 of the Form II model (green, blue, red, yellow). The main chain trace of Form I shown in blue. **C.** A citrate molecule modeled into the electron density map of Form I is found at the interface between two copies of NS1-nuc (red/white and green). The 2Fo-Fc map in the refined model is shown at 1σ in grey, the positive (3σ) 1Fo-Fc map shown in green, and the negative (-3σ) 1Fo-Fc map shown in red. Dashed lines indicate hydrogen bonding distances (2.7-3.2 Å) between appropriate hydrogen bonding

donating and accepting groups. **D.** A 3σ peak is found in the Form II 1Fo-Fc map (magenta) near five residues of the active site and modeled as Mg^{2+} (yellow sphere). 2Fo-Fc map shown at 1σ in blue.

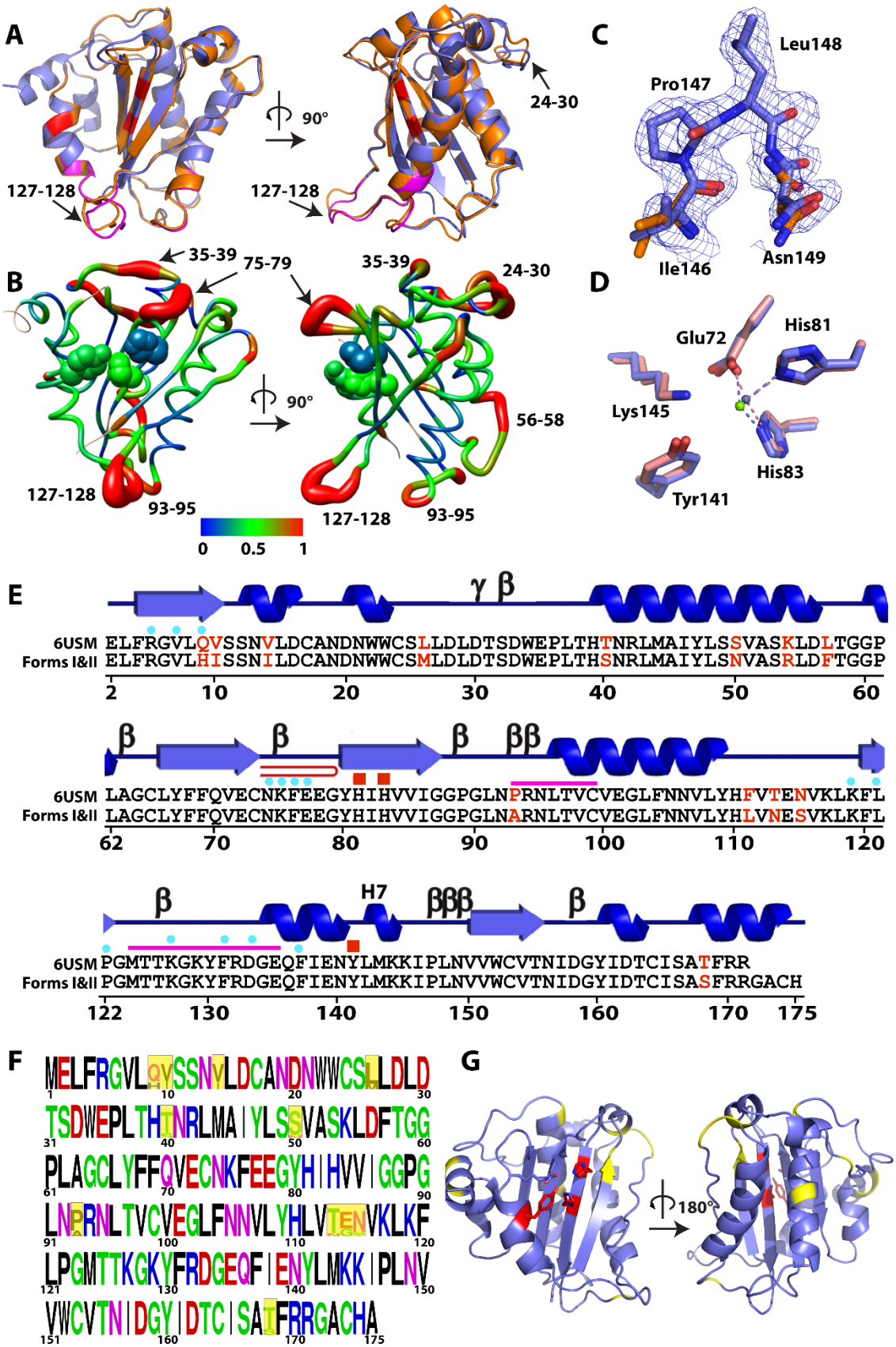
A
127-128
90°
127-128
24-30
B
35-39
75-79
35-39
24-30
127-128
93-95
90°
56-58
127-128
93-95
0
0.5
1
C
Leu148
Pro147
Ile146
Asn149
D
Glu72
His81
Lys145
His83
Tyr141
E
γ β
6USM ELFRGVLQVSSNVLDCANDNWWCSLLDLDTSDWEPLTHTNRLMAIYLSSVASKLDLTGGP
Forms I&II ELFRGVLHISSNILDCANDNWWCSMLDLDTSDWEPLTHSNRLMAIYLSNVASRLDFTGGP
2 10 20 30 40 50 60
β β β ββ
6USM LAGCLYFFQVECNKFEEGYHIHVVIGGPGLNPRNLTVCVEGLFNNVLYHFVTENVKLKFL
Forms I&II LAGCLYFFQVECNKFEEGYHIHVVIGGPGLNARNLTVCVEGLFNNVLYHLVNESVKLKFL
62 70 80 90 100 110 120
β H7 βββ β
6USM PGMTTKGKYFRDGEQFIENYLMKKIPLNVVWCVTNIDGYIDTCISATFRR
Forms I&II PGMTTKGKYFRDGEQFIENYLMKKIPLNVVWCVTNIDGYIDTCISASFRRGACH
122 130 140 150 160 170 175
F
G
180°

Figure 3. Comparison of NS1-nuc structures with sequence variations. A. Orthogonal views of a superposition of the two structures (Form I structure in slate blue, that from 6USM[31] in orange). Positions of active site residues are shown in red, and predicted dsDNA binding residues in magenta (Form I only). Regions of greater deviation identified with arrows. **B.** Orthogonal views of Form I NS1-nuc showing RMSD of C_α atoms between NS1-nuc and 6USM by color (see legend in Å). Residue numbers in segments with highest RMSD are also identified. The active site residues (His81, His83, Tyr141) are shown in spheres colored by C_α RMSD. **C.** 2Fo-Fc (1σ) map of Form I (in blue and red sticks) around the cis-proline at residue number 147. This region is not present in the model from 6USM (residues Ile146 and Asn149 shown in orange, blue, red sticks). **D.** Alignment of DNA nicking active sites from Form II (blue) and 6USM (light red). The modeled Mg^{2+} in Form II is shown in green, and the assigned Zn^{2+} of 6USM is shown in grey. **E.** Alignment of amino acid sequence of 6USM and Forms I and II crystal structures of NS1-nuc. Differences are highlighted in red text. Secondary structural elements are shown above the sequences, γ indicates γ-turn, β indicates β-turn, and red hairpin indicates β-hairpin. Active site residues are shown with red boxes above sequence, and predicted single- and double-stranded DNA binding residues are shown by blue dot or magenta line, respectively, above sequences. **F.** LOGO[41] representation of sequence variations of the 195 sequences of NS1-nuc (residues 1-176 of B19V NS1) from NCBI database. Significant variations boxed in yellow. **G.** Two views of Form I NS1-nuc with positions of significant sequence variants shown in yellow, and active site residues shown in red stick.

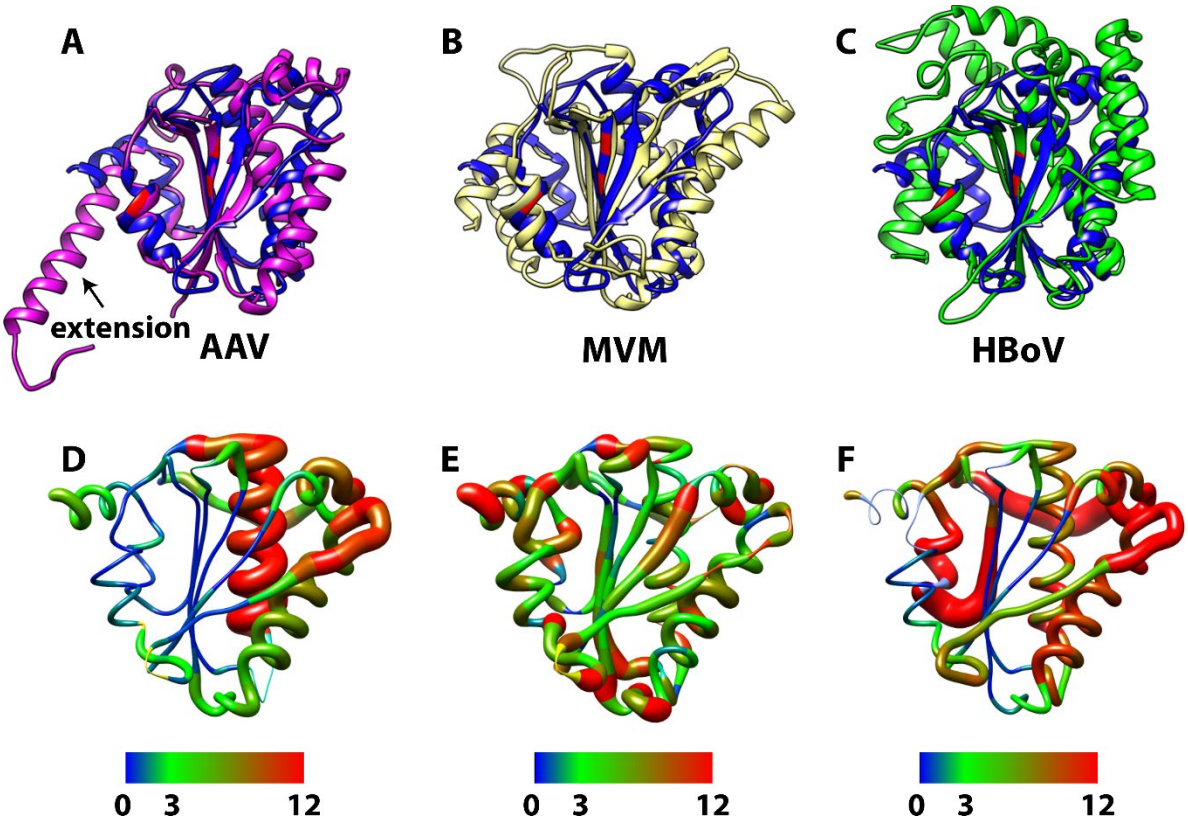

Figure 4. Comparison of NS1-nuc with parvoviral homologues. A. Comparison of Form I NS1-nuc (blue) with the nuclease domain of Adeno Associated Virus (AAV2) Rep (magenta, PDB accession code 5DCX[42]). Positions of active site residues in NS1-nuc (Tyr141, His81, His83) shown in red. **B.** As in A, but comparing NS1-nuc (blue) and the nuclease domain of Minute Virus of Mice (MVM) NS1 (yellow, PDB accession code 3WRN[43]). **C.** As in A, but comparing NS1-nuc (blue) and the nuclease domain of Human Bocavirus (HBoV) NS1 (green, PDB accession code 4KW3[44]). **D.** Ribbon diagram showing RMSD following superposition between C_α of NS1-nuc and AAV2 Rep-nuc. Ribbon color and thickness indicate local RMSD, scale show below, in Å. **E.** As in D but with NS1-nuc and MVM NS1-nuc. **F.** As in D but with NS1-nuc and HBoV NS1-nuc.

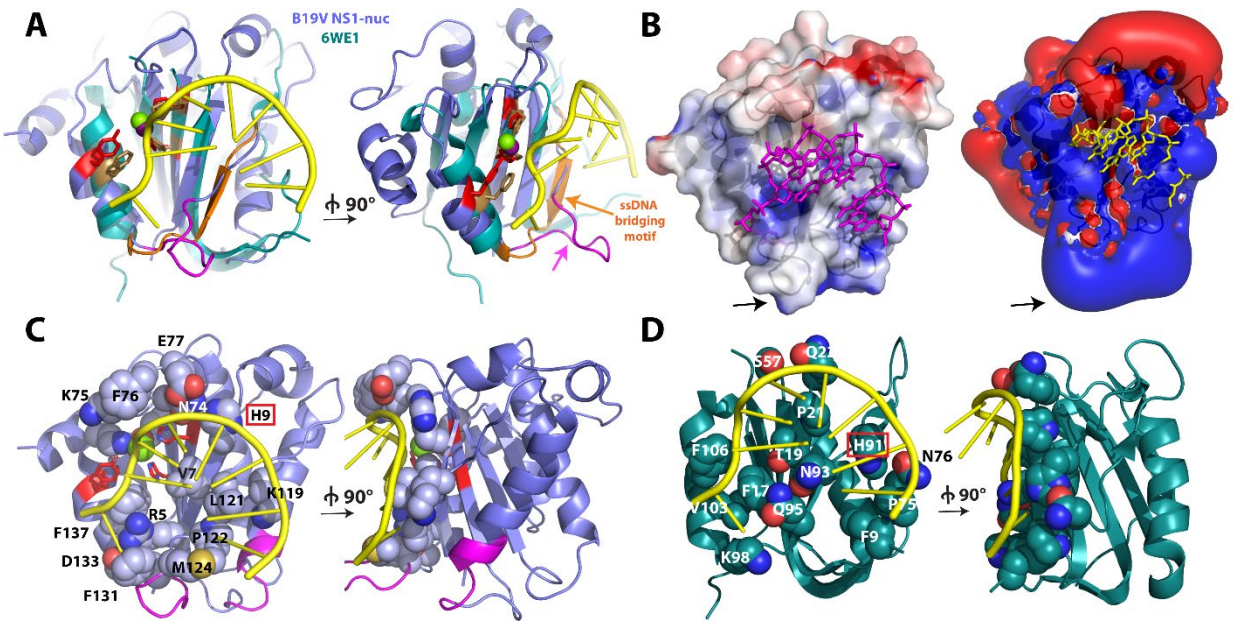

Figure 5. Nicking site recognition in ssDNA by WDV Rep-nuc and NS1-nuc. A. Orthogonal views of Form I NS1-nuc (slate blue) overlayed with a structure of Wheat Dwarf Virus (WDV) Rep-nuc (teal, PDB accession code 6WE1[45]) bound to ssDNA (yellow). Mg^{2+} from Form II NS1-nuc shown as green sphere, Mn^{2+} from WDV Rep-nuc shown as dark purple sphere. Active site residues of NS1-nuc and WDV Rep-nuc are shown in red and brown, respectively. The "ssDNA bridging" segment of WDV Rep-nuc is shown in orange, and residues 123-132 of NS1-nuc (a putative dsDNA binding element) shown in magenta. **B. Left:** Electrostatic potential of Form I NS1-nuc calculated using APBS[39] in PyMol mapped onto the surface of Form I NS1-nuc. ssDNA from PDB accession file 6WE1[45] after superposition of NS1-nuc and WDV Rep-nuc shown in magenta to mark the predicted ssDNA binding cleft. Red indicates negative charge, blue indicates positive charge. Orientation of NS1-nuc as in left side of panel A. **Right:** Electrostatic potential field calculated using APBS[39] in PyMol of Form I NS1-nuc, with ssDNA from WDV Rep-nuc/ssDNA of PDB accession code 6WE1[45] shown in yellow (charge ranges from -5 in red to +5 in blue). Orientation of NS1-nuc as in left side of panel A.**C.** As in A, but with residues implicated in ssDNA binding shown as spheres. Red box indicates the residue H9, which appears to be in a similar position as a H91 in WDV Rep-nuc. **D.** Orthogonal views of the nuclease domain of WDV Rep-nuc bound to nick site ssDNA (PDB

accession code 6WE1[45]) with residues within hydrogen bonding, van der Waals, or salt bridging distance shown as spheres. Red box indicates the residue H91, which appears to be in a similar position as a H9 in B19V NS1-nuc.

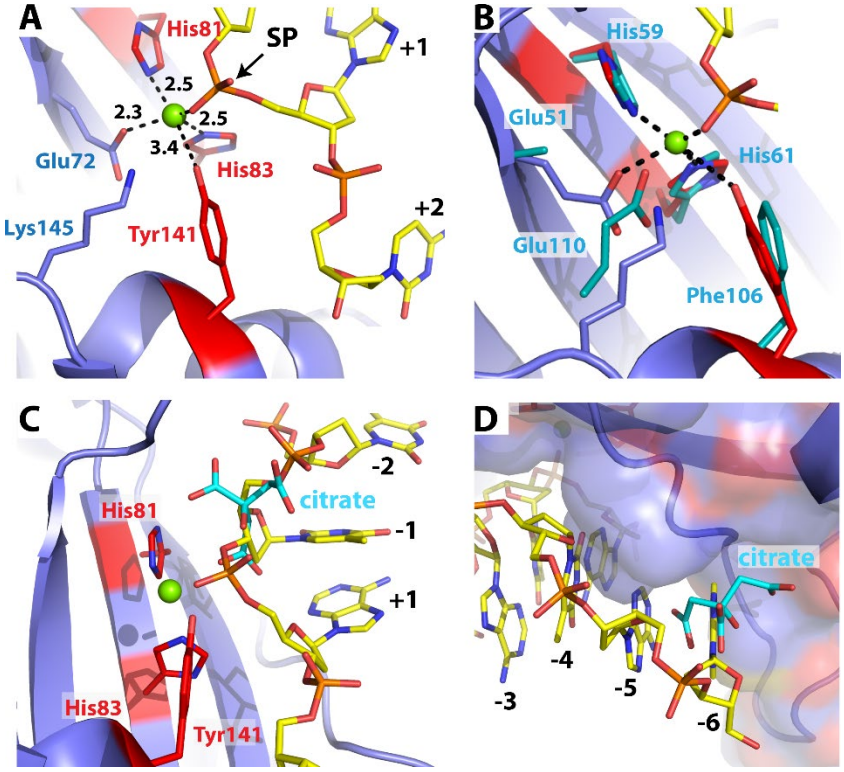

Figure 6. Interactions with bound Mg^{2+} and citrate, and comparison to WDV Rep-nuc. A. Form I NS1-nuc shown in cartoon, with active site residues in stick form. Mg^{2+} from Form II shown as a green sphere. ssDNA (shown in yellow, orange, red, and blue, numbering relative to nick site) from WDV Rep-nuc/ssDNA (PDB accession code 6WE1[45]) after superposition on to Form I NS1-nuc using the C_α of the two His and Tyr active site residues. Distances between Mg^{2+} and atoms of His81, His83, Tyr141 and Glu72, shown in Å. The distance between the Mg^{2+} ion and a non-esterified oxygen of the scissile phosphodiester of the modeled ssDNA is 1.4 Å. **B.** As in A, with selected side chains of WDV Rep-nuc (PDB accession code 6WE1[45]) after superposition onto NS1-nuc shown in teal shown in stick form. **C.** As in A, showing the location of a bound citrate molecule (cyan) near the active site and the backbone of modeled ssDNA between nucleotides +1 and -1. **D.** As in A, but with a surface rendering of NS1-nuc and showing a second position of citrate bound near nucleotide -6 of the modeled ssDNA.

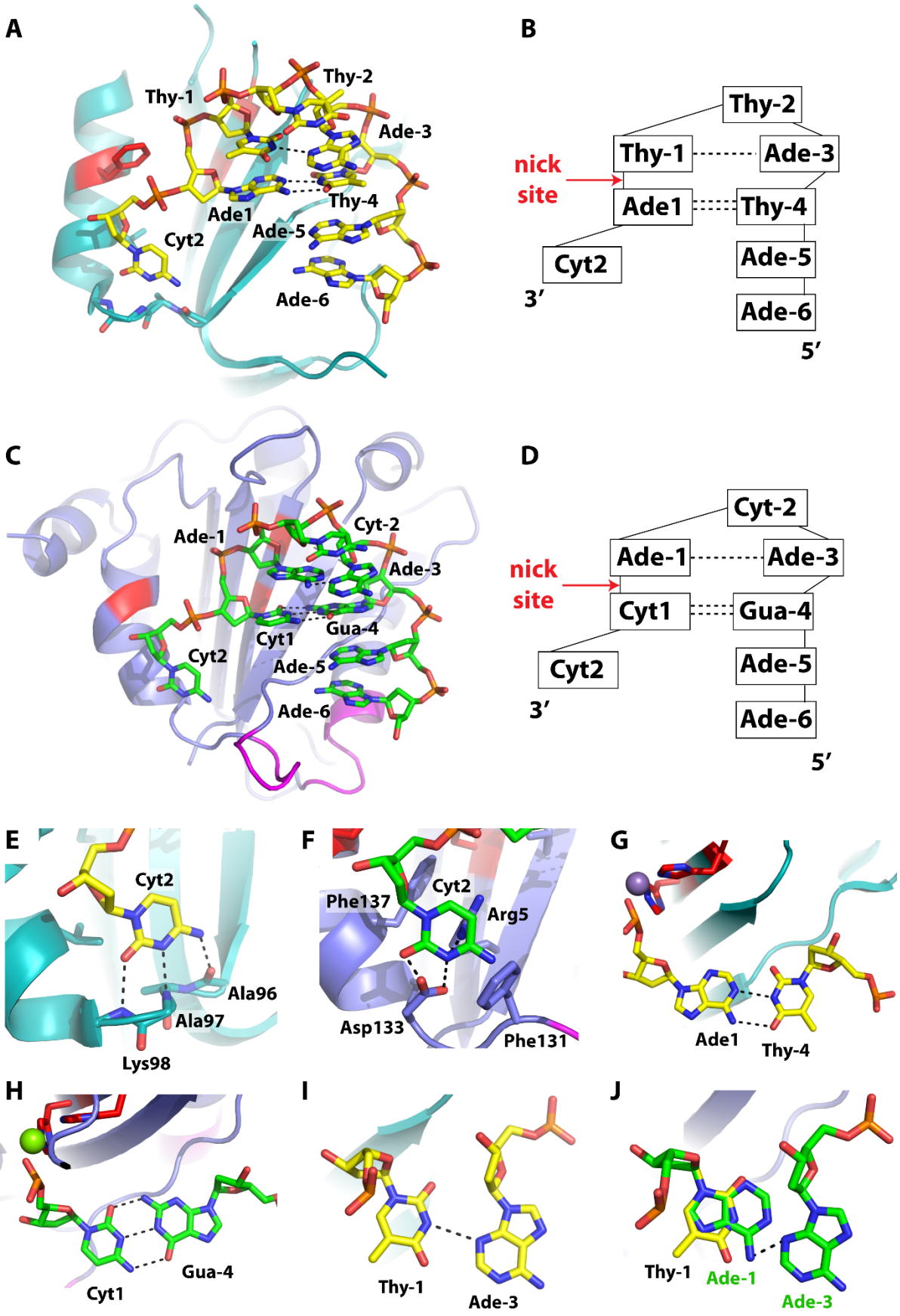
A
Thy-1
Thy-2
Ade-3
Thy-4
Ade1
Ade-5
Cyt2
Ade-6
B
Thy-2
Thy-1
Ade-3
nick site
Ade1
Thy-4
Cyt2
Ade-5
3′
Ade-6
5′
C
Ade-1
Cyt-2
Ade-3
Gua-4
Cyt1
Cyt2
Ade-5
Ade-6
D
Cyt-2
Ade-1
Ade-3
nick site
Cyt1
Gua-4
Cyt2
Ade-5
3′
Ade-6
5′
E
Cyt2
Ala96
Ala97
Lys98
F
Cyt2
Phe137
Arg5
Asp133
Phe131
G
Ade1
Thy-4
H
Gua-4
Cyt1
I
Thy-1
Ade-3
J
Thy-1
Ade-1
Ade-3

Figure 7. Nick site DNA binding by WDV Rep-nuc and NS1-nuc. A. Wheat Dwarf Virus (WDV) Rep-nuc bound to ssDNA containing the nick site (PDB accession code 6WE1[45]). **B.** Sequence of DNA in the WDV origin of replication around the nick site. Dashed lines indicate hydrogen bonding between bases. Stacking of bases is indicated by vertical alignment of bases. **C.** Model of NS1-nuc bound to nick (trs) site DNA based on the WDV Rep-nuc/DNA structure shown in A. **D.** B19V nick site DNA sequence and predicted interactions between bases. **E.** Interactions between Cyt2 and WDV Rep-nuc (PDB accession code 6WE1[45]). **F.** Model for interactions between Cyt2 and NS1-nuc. **G.** Watson-Crick base pair formed between Ade1 and Thy-4 formed in DNA bound by WDV Rep-nuc. Active site residues are shown in red and blue stick, and bound Mn^{2+} shown as purple sphere. **H.** Predicted Watson-Crick base pair formed by Cyt1 and Gua-4 in the model of B19V nick (trs) site DNA bound to NS1-nuc. **I**. Non-Watson-Crick base pair in WDV Rep-nuc bound DNA between Thy-1 and Ade-3. **J.** B19V NS1 nick (trs) site contains Ade at -1 and Ade at -3. Green shows interactions between -1 and -3 predicted in the model of NS1-nuc bound to nick site ssDNA, yellow shows the position of the Thy-1 in the structure with WDV Rep-nuc bound to DNA (coordinates for Ade-3 of the WDV structure overlap with the B19V model). The dashed line indicate close approach of N6 of Ade-1 and N3 of Ade-3 (2.3 Å). Adjustments in positioning of bases would be necessary to bring this to a reasonable distance, such as 2.8 Å.

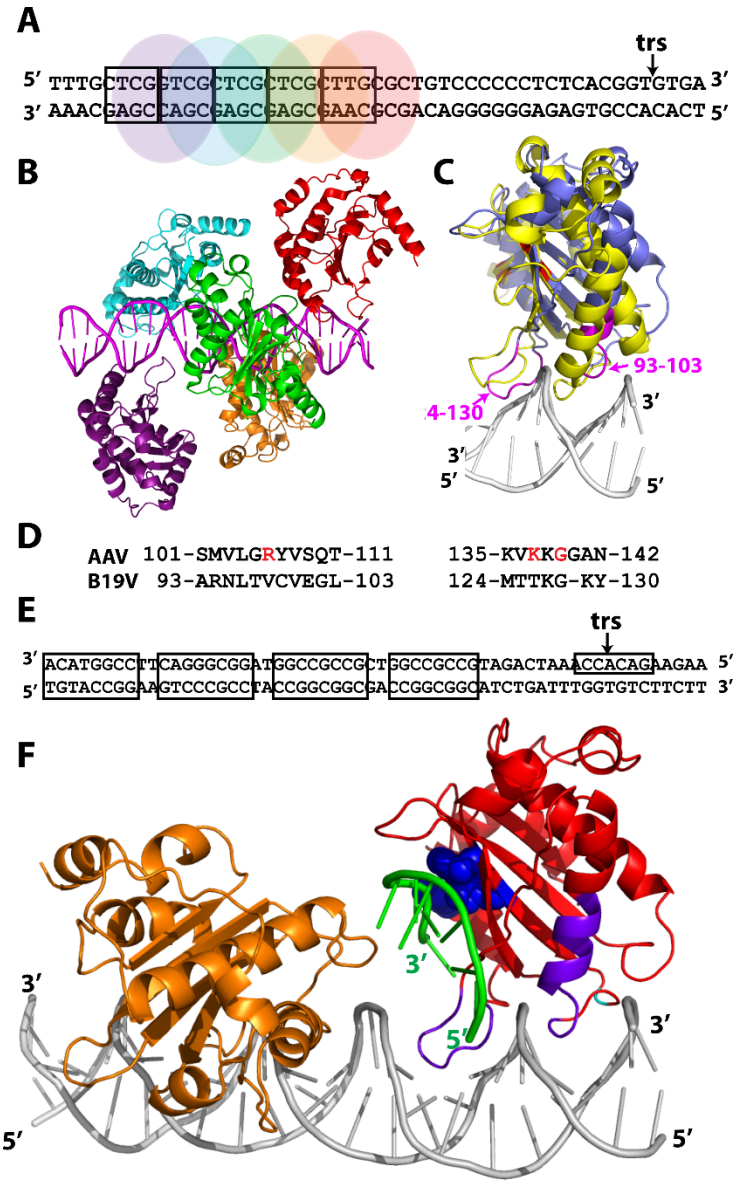

Figure 8. Predicted recognition of NSBE containing dsDNA by NS1-nuc. A. AAV5 viral origin of replication sequence with RBE (boxed bp), trs (arrow, *a.k.a.* nicking site), and positions of each AAV5 Rep-nuc (from PDB accession code 1RZ9[57]). The position of each copy of AAV5 Rep-nuc is colored differently and corresponds to AAV5 Rep-nuc domain colors from the crystal structure shown in panel B. **B.** Ribbon drawing of five copies of the AAV5 Rep-nuc bound to RBE sequences in the AAV5 origin of replication (from PDB accession code 1RZ9[57]). DNA is shown in cartoon form and colored in magenta. **C.** Overlay of NS1-nuc (blue) on AAV5 Rep-nuc (yellow) bound to RBE containing dsDNA (white). Putative DNA binding residues of NS1-nuc shown in magenta and active site residues highlighted in red. **D.**

Residues of the segments of AAV5 Rep-nuc closely approaching or interacting with the dsDNA, and corresponding residues of NS1-nuc after least squares alignment of the two protein structures. Residues shown in red indicate sequence-specific contacts from AAV5 Rep-nuc to the DNA (PDB accession code 1RZ9[57]). **E.** B19V origin of replication sequences with NSBE (four leftmost boxes) and nicking site/trs shown with arrow (boxed nucleotides around the trs indicate those in the model shown in panel F). **F.** Comparison of models of dsDNA (shown in white) and ssDNA (containing the trs sequence, shown in green) bound to NS1-nuc showing some overlap in binding sites. The active site residues His81, His83, and Tyr141 are shown as dark blue spheres, and the predicted dsDNA binding residues shown in dark purple. The location of the ssDNA binding site on NS1-nuc occurs at the interface between adjacent copies of NS1-nuc in the dsDNA binding model. Note that the nick site (trs) is located ~11 nucleotides to the right (as oriented in the figure) of the leftmost NSBE for NS1-nuc (the putative dsDNA binding sites, left-most four boxed sequences in panel E). The actual configuration of Rep and NS1 nuclease domains when bound to both dsDNA sites and the trs in ssDNA is currently unknown.