

# The crystal structure of the human smacovirus 1 Rep domain

Lidia K. Limón, Ke Shi, Amy Dao, Jacob Rugloski, Kassidy J. Tompkins, Hideki Aihara, Wendy R. Gordon\* and Robert L. Evans III\*

Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455, USA.

\*Correspondence e-mail: wrgordon@umn.edu, evans858@umn.edu

Received 23 August 2023

Accepted 31 October 2023

Edited by K. K. Kim, Sungkyunkwan University School of Medicine, Republic of Korea

**Keywords:** HUH endonucleases; Rep domains; HUH-tags; ssDNA; bioconjugation; smacoviruses; crystal structure.

**PDB reference:** human smacovirus 1 Rep domain, 8fr5

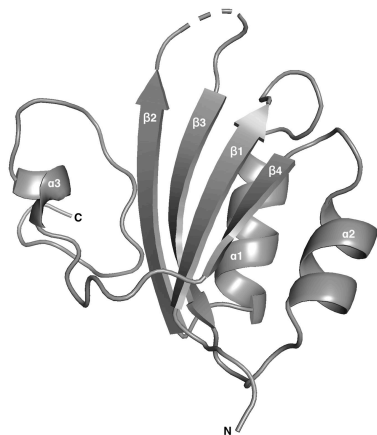
**Supporting information:** this article has supporting information at journals.iucr.org/f

Replication initiator proteins (Reps) from the HUH endonuclease family process specific single-stranded DNA sequences to initiate rolling-circle replication in viruses. Here, the first crystal structure of the apo state of a Rep domain from the smacovirus family is reported. The structure of the human smacovirus 1 Rep domain was obtained at 1.33 Å resolution and represents an expansion of the HUH endonuclease superfamily, allowing greater diversity in bioconjugation-tag applications.

## 1. Introduction

*Smacoviridae* is a family of small CRESS-DNA (circular Rep-encoding single-stranded DNA) viruses. These viruses have been found in the feces of multiple animals and are suspected to cause gastrointestinal disease in humans (Krupovic & Varsani, 2021; Li *et al.*, 2022). Indeed, CRESS-DNA viruses mainly infect eukaryotes. However, it was recently found that instead of direct infection of humans, smacoviruses may infect prokaryotes in the gut, making smacoviruses the smallest viruses to infect prokaryotes and functionally distinct from the majority of the family (Díez-Villaseñor & Rodríguez-Valera, 2019; Zhao *et al.*, 2019; Li *et al.*, 2022).

In addition to functional differences, there are putative structural differences in the replication initiator (Rep) domain in the HUH superfamily of enzymes responsible for processing single-stranded DNA (ssDNA) to replicate the genome during rolling-circle replication (Eisenberg *et al.*, 1977; Chandler *et al.*, 2013). Central to DNA processing of all HUH endonucleases is a structurally defined catalytic nickase domain that first recognizes a specific sequence/structure of DNA, nicks ssDNA at a 'nic site' to yield a sequestered 5'-end that remains covalently bound to the HUH endonuclease and a free 3'-OH that can be used as a primer for DNA replication, and finally facilitates a strand-transfer reaction to resolve the covalent intermediate (Fig. 1; Koonin, 1993; Ilyina & Koonin, 1992; Vega-Rocha *et al.*, 2007; Boer *et al.*, 2006; Chandler *et al.*, 2013; Lovendahl *et al.*, 2017). Named after a triad of residues, the HUH motif in the nickase domain is most often made up of two histidines separated by a bulky hydrophobic residue (U), but can also be histidine-U-glutamine. Several recent crystal structures have illustrated how viral Repls recognize and position ssDNA for cleavage (Luo *et al.*, 2018; Everett *et al.*, 2019; Tompkins *et al.*, 2021; Smiley *et al.*, 2023). Recent comparisons of CRESS-DNA Rep-domain protein sequences show that smacovirus Rep domains are both the smallest in size and the most divergent in sequence of the CRESS-DNA viral Repls (Tarasova & Khayat, 2022).



Finally, Rep domains from HUH endonucleases have been utilized as bioconjugation tags, termed HUH-tags, for applications that require covalent and specific protein–DNA bonds (Aird *et al.*, 2018; Sagredo *et al.*, 2016; Zdechlik *et al.*, 2020). Thus, structural information will guide their engineering to bind to desired DNA sequences (Tompkins *et al.*, 2021).

These interesting distinctions in function and domain composition suggest potential differences in structure and binding (*i.e.* bioconjugation) of the target DNA in smacoviruses. As a first step towards understanding the structural basis for the function of the smacovirus Rep domain in prokaryote infection, we solved a 1.33 Å resolution crystal structure of a smacovirus Rep domain and made structural comparisons with other CRESS-DNA viral Reps.

## 2. Materials and methods

### 2.1. Protein production and purification

**2.1.1. Cloning.** A codon-optimized gene block of the Rep-domain sequence from human smacovirus 1 (HSV1), accession No. AJE25845.1, was synthesized by Integrated DNA Technologies. An N-terminal His<sub>6</sub>-SUMO tag and 15 nucleotides homologous to the parent vector, pTD68, were included for cloning. The parent vector was linearized with the BamHI and XhoI restriction enzymes (New England Biolabs) and the gene block was ligated in using an In-Fusion HD Cloning Kit (Takara) as per the manufacturer’s protocol. The ligated plasmid was transformed into competent *Escherichia coli* Stellar cells and plated onto 100 µg ml<sup>-1</sup> ampicillin plates. After overnight incubation at 37°C, colonies were chosen and DNA was purified with a Qiagen Miniprep kit. Confirmation of the purified plasmid was performed by Sanger sequencing (Genewiz). Protein-production details are provided in Table 1.

**Table 1**

Macromolecule-production information.

Source organism	Human smacovirus
DNA source	Integrated DNA Technologies
Expression vector	pTD68
Expression host	<i>E. coli</i> BL21(DE3)
Complete amino-acid sequence	MTEPKWYDITVSKAKCPEEILRKWLDEN GERYAYGRERGEDGYEHFQVRVVLNRPT SWETMREIWGNSGHCSPITSIRNFDFVVLK EGDFVCSWIKVDP

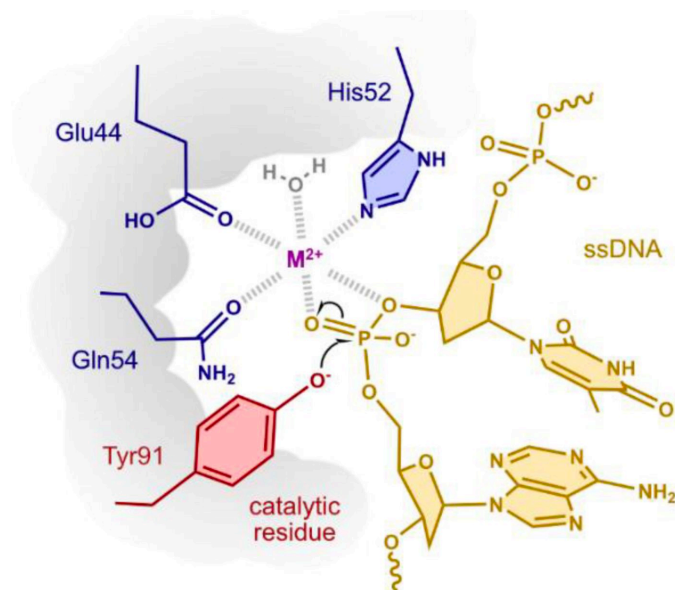
**2.1.2. Protein expression and purification.** Verified plasmids were transformed into *E. coli* BL21(DE3) cells and cultured in 1 l Luria–Bertani (LB) broth with 100 µg ml<sup>-1</sup> ampicillin at 37°C. The culture was induced at an OD<sub>600</sub> of between 0.6 and 0.9 using 0.5 mM isopropyl β-D-1-thiogalactopyranoside and the cells were grown for 20 h at 18°C. The cells were harvested by centrifugation and the pellet was resuspended in lysis buffer (50 mM Tris pH 7.5, 250 mM NaCl), 1 mM EDTA and a protease-inhibitor tablet (Pierce, Thermo Fisher) were added to prevent metal binding and degradation, respectively. Lysis was performed via sonication at 1 min intervals at 4°C. The homogenous suspension was centrifuged at 24 000g for 25 min at 4°C. The supernatant was incubated for 1 h on a rotator with 2 ml HisPure Ni-NTA agarose beads (ThermoFisher) and equilibrated with wash buffer (50 mM Tris pH 7.5, 250 mM NaCl, 1 mM EDTA, 30 mM imidazole). The supernatant was loaded onto a gravity column and allowed to flow through. Protein-bound beads were washed with 25 ml wash buffer and the protein was eluted with 5 ml elution buffer (50 mM Tris pH 7.5, 250 mM NaCl, 1 mM EDTA, 250 mM imidazole). The eluted protein was dialyzed in 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA. The His<sub>6</sub>-SUMO tag was cleaved with 5 µl ULP1 (1 U µl<sup>-1</sup>) overnight at 4°C and incubated with Ni-NTA agarose beads, and the flowthrough was collected. The protein-containing flowthrough was further purified using an Enrich SEC70 (Bio-Rad) size-exclusion chromatography column. Fractions containing the 16 kDa target protein were pooled and concentrated to 2.7 mg ml<sup>-1</sup> using a spin concentrator (Amicon Ultra-15 Centrifugal Filter Unit, 3 kDa molecular-weight cutoff).

### 2.2. Crystallization

A protein solution containing a 10 bp DNA oligonucleotide sequence of the smacovirus origin of replication (AGTAT-TACGC) and Mn<sup>2+</sup> was prepared in a 1:2:2 ratio. Drops consisting of 2 µl protein solution and 1 µl well solution were added to hanging-drop slides using the hanging-drop vapor-diffusion method. The well solution was composed of 0.1 M sodium acetate pH 5.0, 20% PEG 4000, 1 M guanidine-HCl. Upon crystal harvesting, 17% glycerol was added as a cryoprotectant. Crystallization details are listed in Table 2.

### 2.3. Data collection and processing

The data set was collected under cryoconditions on beamline 24-ID-C at the Advanced Photon Source (APS), Argonne National Laboratory using a Dectris EIGER2 16M pixel-array



**Figure 1**  
The catalytic activity of HUH endonucleases relies on the HUH/Q and tyrosine motifs to coordinate the nucleophilic attack on the DNA phosphate backbone (adapted from Tompkins *et al.*, 2021).

**Table 2**  
Crystallization.

Method	Hanging drop
Plate type	24-well plate, Hampton Research
Temperature (K)	298
Protein concentration (mg ml <sup>-1</sup> )	2.7
Protein buffer solution	50 mM Tris pH 7.5, 50 mM NaCl
Volume and ratio of drop	3 µl, 2:1 (protein:reservoir)
Volume of reservoir (µl)	500

**Table 3**  
Data collection and processing.

Values in parentheses are for the highest resolution shell.

X-ray source	Beamline 24-ID-C, APS
Wavelength (Å)	0.97910
Detector	EIGER2 16M
Exposure time (s)	0.20
Crystal-to-detector distance (mm)	170
Angle increment (°)	0.20
Resolution range (Å)	29.43–1.33 (1.38–1.33)
Space group	<i>P</i> 211
<i>a</i> , <i>b</i> , <i>c</i> (Å)	31.16, 49.37, 31.38
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 110.30, 90.00
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	1.96
Solvent content (%)	37.33
Total reflections	71420 (5923)
Unique reflections	19692 (1733)
Multiplicity	3.6 (3.4)
Mosaicity (°)	0.12
Completeness (%)	91.30 (86.80)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	14.08
Wilson <i>B</i> factor (Å <sup>2</sup> )	19.40
<i>R</i> <sub>merge</sub>	0.042
<i>R</i> <sub>meas</sub>	0.049
<i>R</i> <sub>p.i.m.</sub>	0.025
CC <sub>1/2</sub>	0.997

detector. The data set resulted in a 1.33 Å resolution model. Data-collection and processing details are provided in Table 3.

#### 2.4. Structure solution and structure refinement

Molecular replacement with other viral Reps did not provide sufficient phasing information; therefore, a molecular-replacement search model was first generated by *AlphaFold2* (Jumper *et al.*, 2021). The top generated model was then trimmed with *PyMOL* (version 2.0; Schrödinger) at the C-terminal end to remove short segments. The structure was solved with *Phaser* (McCoy *et al.*, 2007) using the trimmed *AlphaFold2*-predicted model and was refined with *Phenix* 1.17.1 (Liebschner *et al.*, 2019) and *Coot* (Emsley *et al.*, 2010). *MolProbity* (Chen *et al.*, 2010) was used for Ramachandran analysis. During refinement, it was determined that no ssDNA was bound to the structure. Structure solution and refinement statistics are listed in Table 4. The final model was deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank as PDB entry 8fr5.

### 3. Results and discussion

#### 3.1. Crystallization and structure determination

To uncover structural differences compared with other ssDNA-bound HUH-tags, attempts to co-crystallize the

**Table 4**  
Structure refinement.

Values in parentheses are for the highest resolution shell.

Reflections used in refinement	19612 (1679)
Reflections used for <i>R</i> <sub>free</sub>	1957 (168)
<i>R</i> <sub>work</sub>	0.187 (0.375)
<i>R</i> <sub>free</sub>	0.224 (0.414)
No. of non-H atoms	
Total	784
Macromolecules	736
Ligands	3
Solvent	45
No. of protein residues	92
R.m.s.d., bond lengths (Å)	0.005
R.m.s.d., angles (°)	0.77
Ramachandran favored (%)	100.0
Ramachandran allowed (%)	0
Ramachandran outliers (%)	0

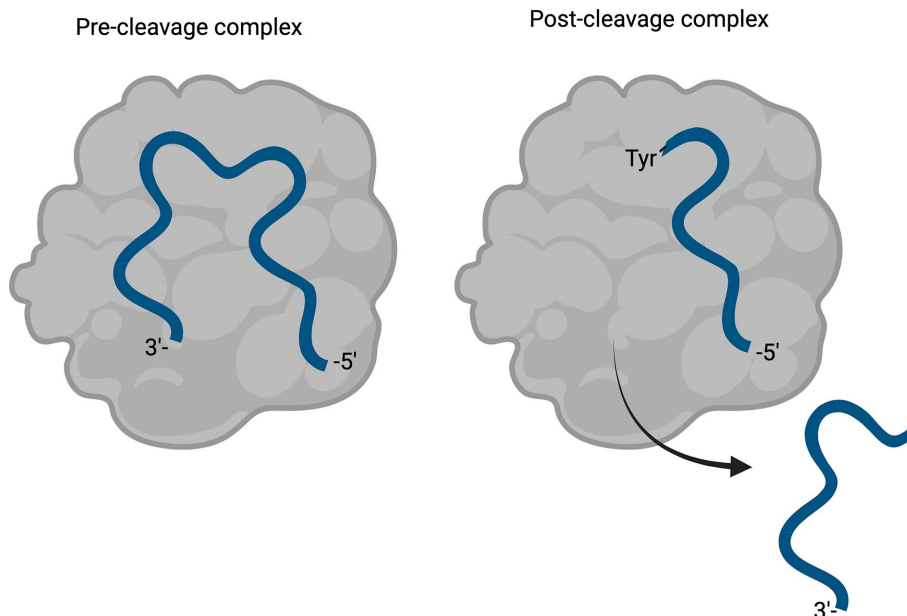
ssDNA-bound protein were performed by mutating the catalytic tyrosine (Tyr81) to a phenylalanine. This allows the coordination of the ssDNA but not covalent linkage to the ssDNA (Larkin *et al.*, 2005). This is because the covalently linked ssDNA is cleaved and the orientation of the ssDNA is changed (see Fig. 2), which does not inform us as to the pre-cleavage coordination orientation. While attempts to obtain the bound/coordinated structure were unsuccessful, we did obtain an unbound structure of HSV1 Rep at 1.33 Å resolution (Fig. 3). The lack of 2*F*<sub>o</sub> – *F*<sub>c</sub> electron density supporting the absence of ssDNA bound to HSV1 Rep is illustrated in Supplementary Fig. S1. Protein crystals formed within days in many of the well conditions screened. The well condition that resulted in the largest crystals was 0.1 M sodium acetate pH 5.0, 20% PEG 4000, 1 M guanidine–HCl. The crystal belonged to space group *P*211. The unit-cell parameters were *a* = 31.16, *b* = 49.37, *c* = 31.38 Å,  $\alpha$  = 90.00,  $\beta$  = 110.30,  $\gamma$  = 90.00°. There was one protein molecule in the asymmetric unit. The final values of *R*<sub>work</sub> and *R*<sub>free</sub> were 0.187 and 0.224, respectively.

#### 3.2. Structure analysis

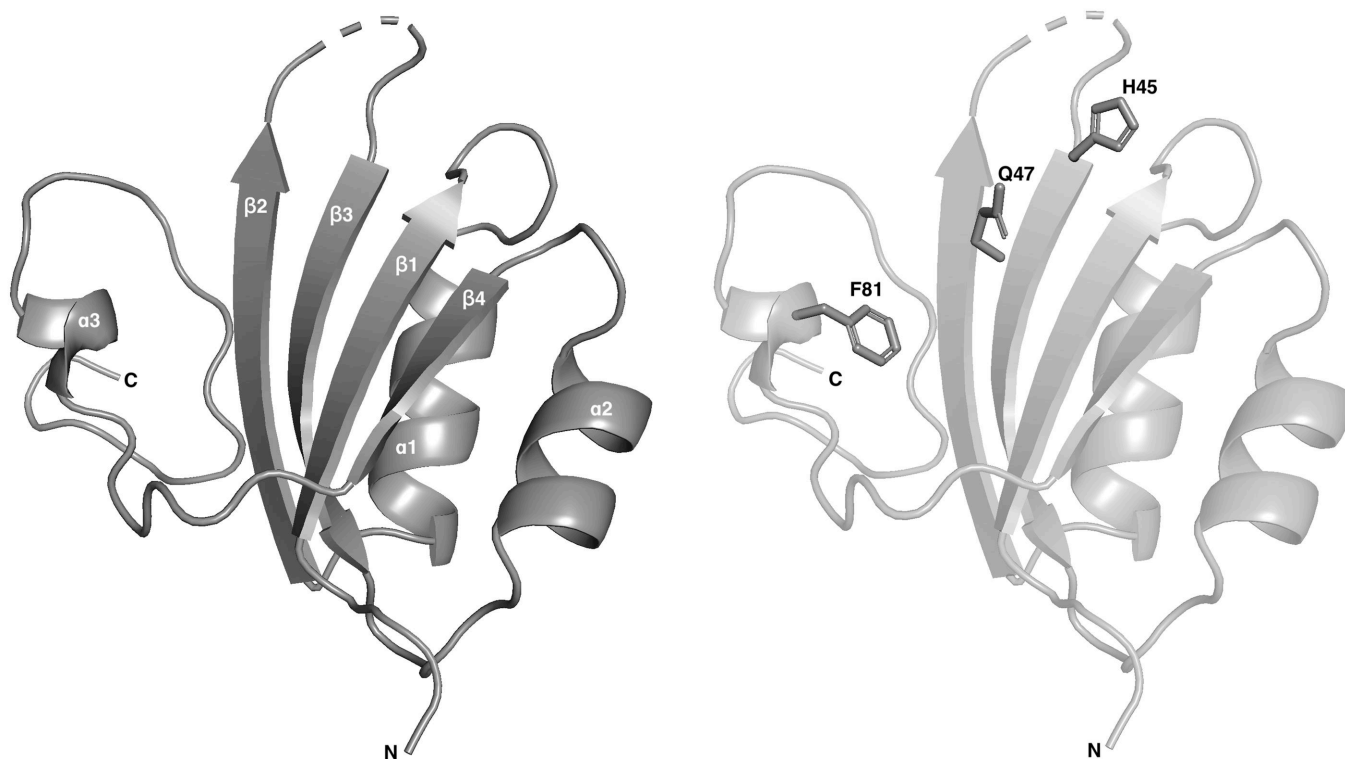
Attempts were made to model the GEDG residues in the electron density adjacent to the HUH/Q motif (Chandler *et al.*, 2013) but were unsuccessful, indicating that the flexibility of the loop in this region is unrestrained, thus resulting in poor electron density. Modeling of ssDNA in the electron density adjacent to the catalytic domains, HUQ and tyrosine motifs for the bound/coordinated structure was also unsuccessful. The resulting unbound structure consists of  $\beta$ 1,  $\alpha$ 1,  $\beta$ 2,  $\beta$ 3,  $\alpha$ 2,  $\beta$ 4 and  $\alpha$ 3 secondary structures, with the  $\beta$ -sheets in an anti-parallel layout (Fig. 3). The catalytically dead phenylalanine substituting for the reactive tyrosine residue resides within  $\alpha$ 3 and the coordinating histidine and glutamine residues reside within  $\beta$ 3. The overall fold of Rep is highly conserved among families of Reps (Fig. 4). When a sequence and structure alignment was performed using *PROMALS3D* (Pei *et al.*, 2008), we found that the Rep from porcine virus 2 (PCV2; PDB entry 5xor) from the circovirus family is structurally closer to that from wheat dwarf virus (WDV; PDB entry 6q1m) from the geminivirus family than that from HSV1

(Fig. 5). This is in agreement with the r.m.s.d. values of the superimposed structures. On superimposition of HSV1 Rep with WDV Rep (PDB entry 6q1m) the r.m.s.d. is 2.4 Å, while that with PCV2 Rep (PDB entry 5xor) is 3.3 Å. This can be compared with the r.m.s.d. value of 0.96 Å between WDV Rep

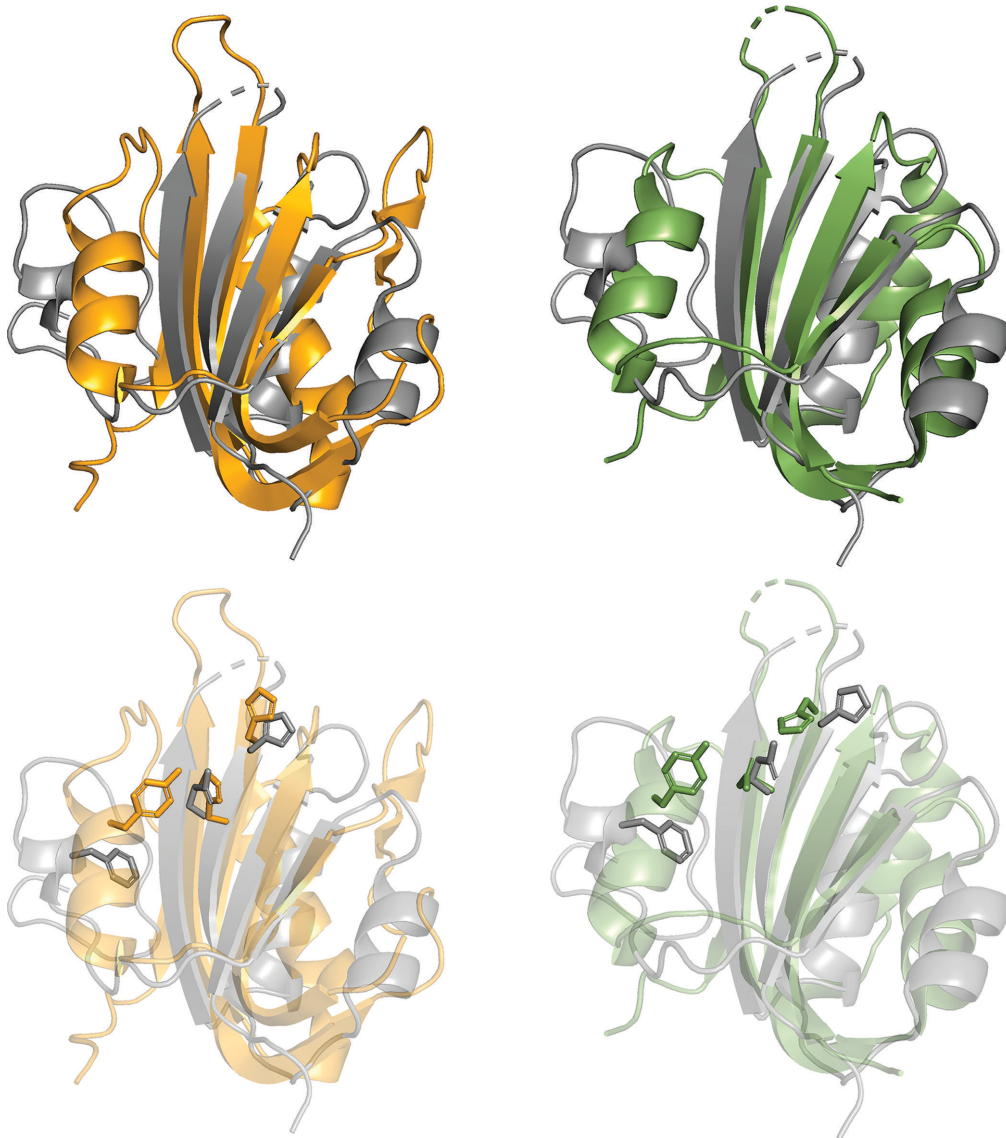
and PCV2 Rep. The difference may be due to the smaller protein size of HSV1 Rep, with fewer residues compared with WDV Rep and PCV2 Rep. HSV1 Rep is also structurally different from WDV Rep and PCV2 Rep in that  $\alpha 2$  and  $\alpha 3$  have shorter disordered loops connecting the  $\alpha$ -helices to the



**Figure 2**  
The coordination of the ssDNA by the Rep during the pre- and post-cleavage complexes. A mutation from tyrosine to phenylalanine does not allow the cleavage reaction to proceed but retains the ssDNA-binding ability of the Rep (Larkin *et al.*, 2005).



**Figure 3**  
Ribbon model of HSV1 Rep showing the secondary-structure organization (left) consisting of  $\beta 1$ ,  $\alpha 1$ ,  $\beta 2$ ,  $\beta 3$ ,  $\alpha 2$ ,  $\beta 4$  and  $\alpha 3$ , and the orientation of the catalytic HUQ and tyrosine (Y81F in our model) motif (right). In our noncoordinated structure, the noncoordinating amino acid 'U' is oriented away from the binding site and is therefore not shown here.



**Figure 4**  
Structural alignment of HSV1 Rep (gray) with WDV Rep (orange, PDB entry 6q1m) on the left and PCV2 Rep (green, PDB entry 5xor) on the right. Superimpositions illustrate the structural conservation among the Reps (top) and the orientation of the catalytic HUH/Q and tyrosine residues (bottom). The r.m.s.d. value on superimposition of HSV1 and WDV is 2.4 Å and that for HSV1 and PCV2 is 3.3 Å.

```

Conservation:          9   9                               9   9   9
PCV2                  1  GSPSKKNGRSGPQPHKRWVFTLNNPSEDER---KKIRDLPISLFDYFIVGEEGNEEGRTPHLQGFANFV  66
WDV                   1  --MASSTPRFRVYSKYLFLTYPQCTLEPQYALDSLRTLNLNKEYEPLYIAAVRELH-EDGSPHLHVLVQNK  67
HSV1                  1  -----MTEPKWYDITVSKAKC-PE---EILRKWLDENGERYAYGRERG-EDGYEHFQVRVVL  53
Consensus_aa:        .....K.h.hThspsp...p...pb.p.h.....hhhsCE...Es...Hhps.h...
Consensus_ss:        eeeeee   hh   hhhh   eeeehh   eeeeeeee
    
```

```

Conservation:          9   99
PCV2                  67  KKQTFNKVKWYLGAR-----CHIEKAKGTDQQNKEYCSKEGNL---LMECGAPRSQGQR-----  117
WDV                   68  LRASITPNPALNLRMDTSPFISFHPNIQAAK-DCNQVRDYITKEVSDVNTAEWGTFFVAVSTPGRKDRDA  136
HSV1                  54  NPTSWETMREIWGNS-----GHCSPTS---IRNFDYVLKEGDF---VCSWIKVPD-----  97
Consensus_aa:        ...ohpp.p.h.....sphp.hp...ps.-Yh.KEss...hph....s.....
Consensus_ss:        hhhhhhhh   hhhhhhhh   e   eeeee
    
```

**Figure 5**  
Structural and sequence comparison of HSV1 Rep with WDV Rep and PCV2 Rep using *PROMALS3D* (Pei *et al.*, 2008).  $\beta$ -Strands are shown in blue and  $\alpha$ -helices in red. Consensus amino-acid symbols: conserved amino acids are shown as bold uppercase letters; *h*, hydrophobic; *s*, small; *p*, polar; *c*, charged; -, negatively charged.

$\beta$ -sheets. Another difference among the families compared here is in the orientation of the HUH/Q and tyrosine residues in the catalytic motifs (Fig. 3), but this could also be explained by the absence of the divalent metal ion that is required to prime the active site for nucleophilic attack on the DNA substrate (Hickman *et al.*, 2002, 2004).

### Acknowledgements

X-ray crystallographic data were collected on beamline 24 (NE-CAT) at the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, which are funded by the National Institutes of Health (NIGMS P30 GM124165).

### Funding information

Funding for this research was provided by the National Institute of General Medical Sciences, National Institutes of Health (NIH) (R35 GM119483). LKL received funding and salary support from the Minnesota Muscle Training Grant (T32 AR007612). KS and HA received funding from the NIH (R35 GM118047).

### References

Aird, E. J., Lovendahl, K. N., St Martin, A., Harris, R. S. & Gordon, W. R. (2018). *Commun. Biol.* **1**, 54.  
 Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A. G., Pérez-Luque, R., Coll, M. & de la Cruz, F. (2006). *J. Mol. Biol.* **358**, 857–869.  
 Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G. & Ton-Hoang, B. (2013). *Nat. Rev. Microbiol.* **11**, 525–538.  
 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.  
 Díez-Villaseñor, C. & Rodríguez-Valera, F. (2019). *Nat. Commun.* **10**, 294.  
 Eisenberg, S., Griffith, J. & Kornberg, A. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 3198–3202.  
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.  
 Everett, B. A., Litzau, L. A., Tompkins, K., Shi, K., Nelson, A., Aihara, H., Evans, R. L. & Gordon, W. R. (2019). *Acta Cryst.* **F75**, 744–749.  
 Hickman, A. B., Ronning, D. R., Kotin, R. M. & Dyda, F. (2002). *Mol. Cell*, **10**, 327–337.

Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M. & Dyda, F. (2004). *Mol. Cell*, **13**, 403–414.  
 Ilyina, T. V. & Koonin, E. V. (1992). *Nucleic Acids Res.* **20**, 3279–3285.  
 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature*, **596**, 583–589.  
 Koonin, E. V. (1993). *Nucleic Acids Res.* **21**, 2541–2547.  
 Krupovic, M. & Varsani, A. (2021). *Arch. Virol.* **166**, 3245–3253.  
 Larkin, C., Datta, S., Harley, M. J., Anderson, B. J., Ebie, A., Hargreaves, V. & Schildbach, J. F. (2005). *Structure*, **13**, 1533–1544.  
 Li, R., Wang, Y., Hu, H., Tan, Y. & Ma, Y. (2022). *Nat. Commun.* **13**, 7978.  
 Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.  
 Lovendahl, K. N., Hayward, A. N. & Gordon, W. R. (2017). *J. Am. Chem. Soc.* **139**, 7030–7035.  
 Luo, G., Zhu, X., Lv, Y., Lv, B., Fang, J., Cao, S., Chen, H., Peng, G. & Song, Y. (2018). *J. Virol.* **92**, e00724-18.  
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.  
 Pei, J., Kim, B. H. & Grishin, N. V. (2008). *Nucleic Acids Res.* **36**, 2295–2300.  
 Sagredo, S., Pirzer, T., Aghebat Rafat, A., Goetzfried, M. A., Moncalian, G., Simmel, F. C. & de la Cruz, F. (2016). *Angew. Chem. Int. Ed.* **55**, 4348–4352.  
 Smiley, A. T., Tompkins, K. J., Pawlak, M. R., Krueger, A. J., Evans, R. L. III, Shi, K., Aihara, H. & Gordon, W. R. (2023). *mBio*, **14**, e02587-22.  
 Tarasova, E. & Khayat, R. (2022). *Viruses*, **14**, 37.  
 Tompkins, K. J., Houtti, M., Litzau, L. A., Aird, E. J., Everett, B. A., Nelson, A. T., Porschloegl, L., Limón-Swanson, L. K., Evans, R. L., Evans, K., Shi, K., Aihara, H. & Gordon, W. R. (2021). *Nucleic Acids Res.* **49**, 1046–1064.  
 Vega-Rocha, S., Byeon, I. L., Gronenborn, B., Gronenborn, A. M. & Campos-Olivas, R. (2007). *J. Mol. Biol.* **367**, 473–487.  
 Zdechlik, A. C., He, Y., Aird, E. J., Gordon, W. R. & Schmidt, D. (2020). *Bioconjug. Chem.* **31**, 1093–1106.  
 Zhao, L., Rosario, K., Breitbart, M. & Duffy, S. (2019). *Adv. Virus Res.* **103**, 71–133.