STRUCTURAL BIOLOGY

Structure of coxsackievirus cloverleaf RNA and 3C^{pro} dimer establishes the RNA-binding mechanism of enterovirus protease 3C^{pro}

Dimagi Dias-Solange, My Tra Le, Keerthi Gottipati*, Kyung H. Choi*

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

In positive-strand RNA viruses, the genome serves as a template for both protein translation and negative-strand RNA synthesis. Enteroviruses use the cloverleaf RNA structure at the 5′ end of the genome to balance these two processes. Cloverleaf acts as a promoter for RNA synthesis and forms a complex with viral 3CD protein, the precursor to 3C^{pro} protease, and 3D^{pol} polymerase. The interaction between cloverleaf and 3CD is mediated by the 3C^{pro} domain, yet how 3C^{pro} promotes specific RNA-binding is not clear. We report the structure of coxsackievirus cloverleaf RNA-3C^{pro} complex, wherein two 3C^{pro} molecules interact with cloverleaf stem-loop D. 3C^{pro} dimer mainly recognizes the shape of the dsRNA helix through symmetric interactions, suggesting that 3C^{pro} is a previously undiscovered type of RNA binding protein. We show that 3CD protein also dimerizes on cloverleaf RNA and binds the RNA with higher affinity than 3C^{pro}. The structure provides insight into the RNA-binding mechanism of 3C^{pro} or 3CD with other cis-acting replication elements.

INTRODUCTION

RNA viruses have evolved several strategies to maximally use their small genomes for efficient replication. The viral RNA genomes are not simply protein-coding transcripts but often serve as functional macromolecules in their own right, playing central roles in diverse processes of viral pathogenesis. RNA structures within the genomes are involved in regulating genome replication, polyprotein translation, immune response, and even genome packaging and maturation (*1*–*6*). For instance, RNA structures present at the 5′ end of the genomes of enteroviruses and flaviviruses function as a switch to transition from host ribosome–mediated polyprotein translation of the genome to viral polymerase–mediated RNA synthesis (*1*–*3*).

Enteroviruses are small positive-sense RNA viruses in *Picornaviridae* that are associated with many human and animal diseases. Coxsackievirus B3 (CVB3), poliovirus (PV), and human rhinoviruses (HRV) are the causative agents of acute myocarditis, poliomyelitis, and respiratory infections, among others (*7*–*10*). No treatments are yet available to combat these viruses and vaccines are available only for PV. Enteroviruses encapsidate a ~7.5-kb, positive-sense RNA genome covalently linked to a peptide primer, VPg (virion protein genome linked), on the 5′ end. The genome consists of a 5′ untranslated region (5′UTR), a single open reading frame (ORF) and a 3′UTR terminating in a poly-A tail. The 5′UTR is ~750-nucleotide (nt) long and contains six conserved RNA domains (I to VI) that are critical for viral genome replication and polyprotein translation (*11*). Domain I, also called cloverleaf RNA, has a stem-loop structure with a four-way junction that functions as a promoter for negative-strand RNA synthesis (*12*–*14*). The cloverleaf structure also acts as a switch to direct the transition from viral translation to genome replication and serves as the assembly site for the viral replication complex (*15*, *16*). Domains II to VI constitute the internal ribosome entry site that recruits the ribosome for initiation of viral polyprotein translation. The single ORF is translated into a large polyprotein (VP4-VP2-VP3-VP1-2A-2B-2C-3A-3B-3C-3D), which is processed further into three smaller polyproteins: P1, P2, and P3 (*17*–*19*). P1 contains capsid proteins (VP1-4). P2 contains 2A protease, 2B ion channel protein, and 2C nucleoside-triphosphatase (NTPase), all of which are involved in host interaction functions required for robust virus replication (*20*). P3 proteins (3A-D), in both the mature and precursor (uncleaved) forms, are involved in viral replication. 3A and its precursor 3AB are small, membrane-binding proteins required for establishment of the viral replication complex. 3B is VPg, a peptide primer needed for initiating RNA synthesis (*21*). 3C^{pro} and its precursor 3CD (uncleaved 3C^{pro} and 3D^{pol}) function as a cysteine protease, responsible for polyprotein processing to generate mature viral proteins. 3D^{pol} is the RNA-dependent RNA polymerase that uses the peptide primer VPg to initiate RNA synthesis (*22*).

Picornaviruses use a primer-dependent initiation mechanism to synthesize both negative- and positive-strand RNAs (*23*, *24*). The viral polymerase 3D^{pol} first uses the cis-acting replication element (CRE) in the internal region of the RNA genome as a template and synthesizes a VPg-linked primer, VPg-pUpU (*13*, *25*). 3D^{pol} is recruited to CRE by a direct interaction with 3CD dimer bound to CRE (*26*). Following VPg-pUpU synthesis, the viral replication complex, composed of the viral 3CD and the host poly-rC binding protein 2 (PCBP2), assembles on the cloverleaf RNA at the 5′ end (*12*, *13*, *27*, *28*). The viral genome is then likely circularized via protein-protein interactions between the cloverleaf-3CD-PCBP2 complex at the 5′ end and the poly(A)-binding protein bound at the 3′ end of the genome (*29*, *30*). 3D^{pol} then uses the VPg-pUpU primer to initiate synthesis of negative-strand RNA at the 3′-end of the genome to produce a double-stranded RNA (dsRNA) intermediate (*23*). Genomic positive-strand RNA is next synthesized from the negative strand using the VPg-pUpU primer complementary to the 3′ end of the negative strand (*13*, *30*).

During viral replication, 3CD provides an essential RNA-binding activity, first, by interacting with the CRE RNA to recruit 3D^{pol} for initial VPg uridylylation and, subsequently, by binding the cloverleaf RNA to assemble a replication complex for negative-strand synthesis. The RNA binding domain of 3CD is 3C^{pro} protease, which

Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN 47405, USA.
*Corresponding author. Email: kegottip@iu.edu (K.G.); kaychoi@iu.edu (K.H.C.)

specifically recognizes CRE and cloverleaf, although a common RNA-binding motif has not been found in the protease domain. Thus, the molecular mechanism by which 3CD recognizes two distinct RNA elements, i.e., cloverleaf and CRE, is not known. Recently determined crystal structures show that enterovirus cloverleaf RNAs fold into "H"-shaped structures, stabilized by a unique adenosine-cytidine-uridine (A•C-U) base triple between stem-loops C and D (*3*, *31*, *32*). We previously determined the structures of the CVB3 and PV cloverleaf RNAs (*3*). Here, we report the structure of the CVB3 cloverleaf-3C^{pro} ribonucleoprotein complex. Unexpectedly, two 3C^{pro} proteins assemble on stem-loop D of cloverleaf via symmetric interactions with the dsRNA stem. 3C^{pro} primarily interacts with the phosphodiester backbones and sugar moieties, implying that the protein recognizes the shape of base-paired RNA stem as a dimer. Using mass photometry (MP) and biochemical binding assays, we show that two 3CD monomers assemble on RNA, rather than a preassembled 3CD dimer binding to the RNA, indicating that RNA structure directs 3C^{pro} dimer formation. We propose that 3C^{pro} dimer recognizes the shape of the dsRNA helix through symmetric interactions, and this mechanism is conserved in cloverleaf and CRE interaction.

RESULTS

Structure determination of the CVB3 cloverleaf-3C^{pro} dimer complex

We previously used the tRNA-scaffold approach to determine the structure of CVB3 cloverleaf RNA (CL-tRNA) (*3*). Crystals of the cloverleaf RNA-3C^{pro} complex were generated by co-crystallizing a CL-tRNA construct with a protease-inactive mutant of 3C^{pro} made by substituting Cys^{147} with Ala. Additionally, Gly^{55}, Asp^{58}, and Val^{63} of 3C^{pro} (corresponding to Glu^{55}, Asp^{58}, and Glu^{63} in PV 3C^{pro}) were mutated to Ala to prevent potential dimerization as described in the crystal structure determination of PV 3CD (*33*). The RNA-protein complex crystallized in space group $P2_12_12$, and diffraction data were collected to 2.87-Å resolution. Molecular replacement with the CL-tRNA [Protein Data Bank (PDB) code 8SP9] and 3C^{pro} crystal structures (PDB code 3ZYD) indicated that the asymmetric unit contained one molecule of cloverleaf RNA and two molecules of 3C^{pro}. Electron density was observed for the entirety of 3C^{pro} monomers except for the first amino acid at the N terminus and Arg^{143} for one molecule (M2) and three amino acids at the C terminus (residues 181 to 183) for both molecules. Electron density for the entire cloverleaf RNA was observed except for the disordered apical loop of stem-loop B (*14*–*25*), but density for the tRNA scaffold was missing except for C89 in the linker between the cloverleaf and tRNA scaffold (Fig. 1). In the crystal, each 3C^{pro} monomer interacts similarly with two copies of the other 3C^{pro} monomer related by crystallographic symmetry, i.e., M1 monomer interacts with two neighboring M2 molecules and vice versa (fig. S1). The tRNA portion of the CL-tRNA was not observed in the structure, and there is no space in the lattice to accommodate the tRNA, as tRNA would crash with the 3C^{pro} monomer (fig. S1). This indicates that the tRNA scaffold was cleaved during crystallization and excluded from the crystal. Urea-polyacrylamide gel analysis of the drop solution that the crystals grew in showed that a smaller RNA fragment was present in the solution, consistent with tRNA being cleaved before crystallization. Data collection, scaling, and refinement statistics are listed in Table 1.

Coxsackievirus 3C^{pro} binds the entire lateral surface of stem-loop D as a dimer

Enterovirus cloverleaf RNA consists of stem A and stem-loops B, C, and D and forms an H-type four-way junction where its four stems are stacked into two coaxial dsRNA helices (*3*, *31*). The stems A and D form one contiguous coaxial dsRNA helix, and stems B and C form the second coaxial dsRNA helix (Fig. 1). Stem-loops B and D are on diagonally opposite ends of the H-like arrangement of RNA helices and are antiparallel to each other. The structure of the cloverleaf and 3C^{pro} dimer complex shows that 3C^{pro} dimer interacts exclusively with stem-loop D, and, thus, the other stems of cloverleaf RNA are not involved in the interaction between cloverleaf and 3C^{pro} (Fig. 1A). Two 3C^{pro} molecules occupy the entire lateral surface of stem-loop D and wrap halfway around the dsRNA helix on one side of cloverleaf, opposite from stem C. Stem-loop D contains the noncanonical base pairs between ^{54}UCU^{56} and ^{71}UUU^{73} in the central pyrimidine mismatch region, and the twofold symmetry axis of 3C^{pro} dimer lies on the pyrimidine mismatch, roughly perpendicular to the helical axis of the stem (Fig. 1B). However, the shape of 3C^{pro} dimer is not complementary to the major groove of the dsRNA near the twofold axis, and there is no interaction with bases in the pyrimidine mismatch region (fig. S2). One monomer (M1) binds the dsRNA stem proximal to the apical loop, and the other (M2) binds toward the base of the stem and orients in the opposite direction to M1. Together, the 3C^{pro} dimer buries a large, positively charged surface area of ~1400 Å^2 between the 3C^{pro} dimer and the RNA (Fig. 2A).

The 3C^{pro} protein folds into a double–β barrel structure with the protease catalytic triad (Cys^{147}, His^{40}, and Glu^{71}) positioned at the interface of the two β barrels. The RNA-binding surface in 3C^{pro} dimer lies on the side opposite the protease active site, indicating that RNA binding should not interfere with the protease function of 3C^{pro} (Fig. 2A), consistent with previous reports (*34*–*36*). Three major areas in 3C^{pro} interact with RNA: the N-terminal α-helix, consisting of amino acids 2 to 13; the inter-domain interface region ^{82}KFRDI^{86}, also called the FG loop; and the hairpin loop ^{154}TGK^{156} near the C terminus of the protein (Fig. 2, B and C). The 3C^{pro} dimer has two symmetric RNA-binding interfaces, each of which is formed by residues from both monomers. Each interface is constituted by the N-terminal helix and nearby TGK hairpin of one 3C^{pro} molecule and the FG loop of the other molecule. Specifically, the N-terminal helix and TGK loop of M1 and the FG loop of M2 interact with the upper half of stem-loop D, proximal to the apical loop, while the opposite side of the 3C^{pro} dimer interacts with the bottom half of the stem (Fig. 2B). Both the FG loop and the TGK hairpin were previously identified as RNA binding regions, and single–amino acid substitutions in the FG loop (Arg^{84}, Asp^{85}, and Ile^{86}, CVB3 3C^{pro} numbering) and TGK hairpin (Thr^{154}, Gly^{155}, and Lys^{156}) have been shown to abolish RNA binding for PV, EV71, and HRV-14 3C^{pro} (*14*, *36*–*39*).

Two 3C^{pro} proteins provide symmetric interaction sites for stem-loop D

As noted above, the 3C^{pro} binding region on the cloverleaf RNA is localized to stem-loop D, a 34-nt-long dsRNA stem containing the pyrimidine mismatch (^{54}UCU^{56} and ^{71}UUU^{73}) and dinucleotide bulge (^{48}UA^{49}), capped by an apical tetraloop (Fig. 1B) (*3*). The 3C^{pro} dimer makes bipartite interactions with stem-loop D in five nucleotide stretches of ^{49}AGCAC^{53} and ^{65}GGUAC^{69} (Fig. 3). The conformations

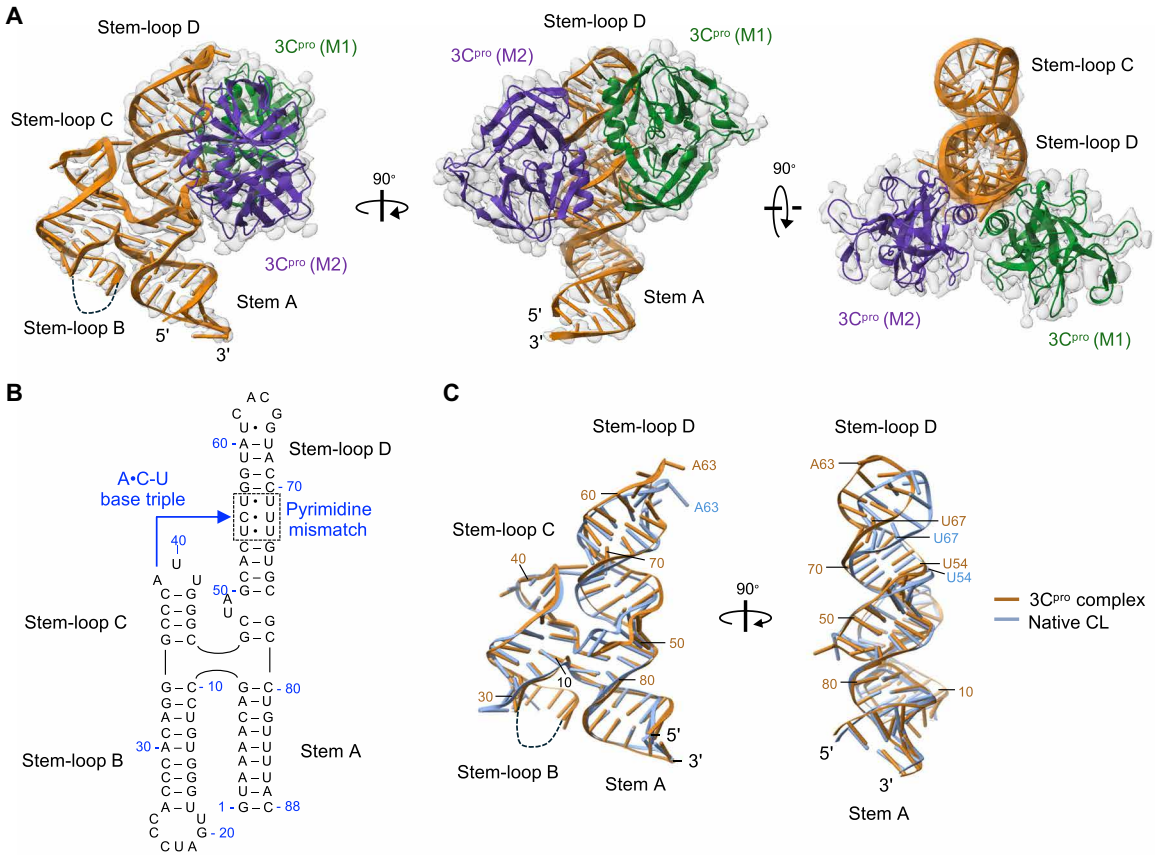

Fig. 1. Structure of the CVB3 cloverleaf RNA and 3C^{pro} dimer complex. (**A**) Overall structure of the CVB3 cloverleaf and 3C^{pro} dimer complex. The structure shows that two copies of 3C^{pro} interact along the lateral surface of the dsRNA of stem-loop D across from the stem-loop C of cloverleaf. Two 3C^{pro} monomers are depicted in green (M1) and violet (M2), and the cloverleaf RNA in gold. The 2Fo-Fc electron density map contoured at 1σ is shown as gray cage. (**B**) Secondary structure of CVB3 cloverleaf RNA. Four stems of cloverleaf RNA (A, B, C, and D) and the A•C-U base triple between stem-loops C and D are indicated. Pyrimidine mismatch region in stem-loop D is boxed. The nucleotide numbering corresponds to the cloverleaf RNA sequence in the genome. (**C**) Comparison of native and bound cloverleaf RNA structures. The cloverleaf RNA in the 3C^{pro} complex is colored as in (A), and the native cloverleaf (PDB code 8SP9) is colored in blue. The cloverleaf in the cloverleaf-3C^{pro} dimer complex can be superposed on the native structure with a root mean square deviation (RMSD) of 2.24 Å across all 70-nt pairs. Interaction with 3C^{pro} dimer increases the width of major groove between U54 and U67 from 8.8 Å in the native structure to 11 Å in the complex. Positions of cloverleaf U54 and U67 in the native and the complex structures are indicated.

of the two monomers within the 3C^{pro} dimer are similar and could be superimposed with a root mean square deviation (RMSD) of 0.52 Å over 179 Cα atoms. 3C^{pro} primarily interacts with the phosphate backbone and the hydroxyl groups of the sugar moieties in the dsRNA stem, indicating that 3C^{pro} recognizes the shape of the dsRNA, rather than specific nucleotides (Fig. 3). Symmetric interactions were observed across the major groove of stem-loop D near the twofold axis of 3C^{pro} dimer involving Arg^{13}, Arg^{84}, and Asp^{85} from each monomer (Fig. 3A). In M1, the side chain of Arg^{84} and backbone of Asp^{85} of FG loop interact with the phosphate backbone of U67 and U54, while the same atoms of M2 interact with phosphate backbones of C51 and C70 across the major groove (Fig. 3A). As a result of the symmetric interactions of 3C^{pro} dimer, the major groove edges between U54 and U67 and between C51 and C70 are maintained at a constant width of 11 Å. Further, Arg^{13} from the N-terminal helix of M1 interacts with the phosphate backbone of A68, while Arg^{13} of M2 interacts with phosphate backbone of A52 across the major groove (Fig. 3A). Arg^{13} of both monomers additionally interacts with the phosphate backbone of C53.

Symmetric interactions with purine bases are also observed with Glu^{5} and Asn^{80}. Glu^{5} and Asn^{80} of M1 interact with G50 and G76, while Glu^{5} and Asn^{80} of M2 interact with G66 and A60, respectively (Fig. 3A). Radiating from the twofold axis of 3C^{pro} dimer, additional residues from the N-terminal helix, FG loop, and TGK hairpin interact pseudo-symmetrically with the backbone phosphates and 3′ hydroxyl groups of dsRNA stem (Fig. 3, A and B). 3C^{pro} also interacts with single-stranded regions of A63 and G65 in the tetra-loop and A49 within the UA bulge. Details of RNA-protein interactions are shown in Fig. 3B.

The structure of the cloverleaf-3C^{pro} dimer complex is consistent with previous reports on the requirements of RNA structure for 3C^{pro} interactions. Mutations in the stem of stem-loop D of HRV14 cloverleaf were generally tolerated for 3C^{pro} interactions, suggesting that the nucleotide sequence in the stem is not important, provided the base pairing is maintained (*37*). Further, substitutions in the apical loop or base-pairing substitutions in the pyrimidine mismatch region did not affect interactions of stem-loop D with 3C^{pro} (*37*). Our structure of the cloverleaf-3C^{pro} dimer complex shows that 3C^{pro}

Table 1. X-ray crystallographic data collection and refinement statistics. Data collection statistics were obtained using HKL2000 (*66*). Model refinement statistics were obtained from Phenix (*67*). The final refined model and reflection files were deposited to PDB with accession number 9DCF. RMSDs, root mean square deviations.

	CL-tRNA + 3C^{pro}
Data collection	
Wavelength (Å)	1.00003
Resolution (Å)	48.93 to 2.87 (2.95 to 2.87)*
Space group	$P2_12_12$
Unit cell dimensions	a = 91.02 Å, b = 200.82 Å, c = 71.12 Å, α = β = γ = 90°
No. of reflections	29,950 (1384)
No. of molecules/AU	
RNA	1
Protein	2
Completeness (%)	98.8 (92.3)
Redundancy	11.0 (8.3)
$I/\sigma I$	5.0 (1.0)
$R_{meas}/R_{p.i.m}$	0.337/0.098 (2.504/0.833)
$CC_{1/2}$	0.947 (0.661)
Refinement	
Resolution range (Å)	48.93 to 2.87 (2.95 to 2.87)
Reflections used in refinement	22,756 (942)
R_{work}/R_{free}	23.69/27.33
No. of atoms	
Nucleotide	1627
Protein	2771
Solvent	12
RMSDs	
Bond angles (°)	1.28
Bond length (Å)	0.009
Ramachandran favored (%)	93.18
Ramachandran allowed (%)	6.82
Ramachandran outliers (%)	0.00
Average B factor	
Protein	56.78
RNA	130.59
Solvent	46.63

*Values in parentheses are for the highest resolution shell.

recognizes the overall structure of the dsRNA helix and that 3C^{pro} and cloverleaf binding does not involve many base-specific interactions, consistent with previous reports. Additionally, nuclear magnetic resonance (NMR) chemical shift perturbations using CVB3 stem-loop D and 3C^{pro} indicated that 3C^{pro} may interact with nucleotides adjacent to the triple pyrimidine mismatch (C53 and C70) and tetraloop including the closing UG base pair (^{61}UCACGG^{66}) (*40*, *41*). In the structure, 3C^{pro} dimer interacts with the face of stem-loop D where C53 and C70 present their major groove edge and with the tetraloop (A63, G65, and G66) (Fig. 3), consistent with the NMR studies. Two CVB3 cloverleaf mutations that abolished 3C^{pro} and 3CD interactions in a yeast three-hybrid system are individual deletions of C64 and U71 (*40*). 3C^{pro} does not interact with C64 in the tetraloop, as the base faces away from the 3C^{pro} dimer. However, deletion of C64 would render the tetra-loop a tri-loop, and, thus, the deletion likely affects the conformation of the loop, preventing optimal interactions with 3C^{pro} dimer. U71 is part of the pyrimidine mismatch, and 3C^{pro} dimer does not directly interact with this region. However, the twofold symmetry axis of the 3C^{pro} dimer is located in the pyrimidine mismatch, where four arginine residues (Arg^{13} and Arg^{84} from each 3C^{pro} monomer) bind in the major groove. Consequently, deletion of U71 in the pyrimidine mismatch would distort the dsRNA helix, and prevent symmetric 3C^{pro} dimer binding.

Cloverleaf RNA and 3C^{pro} dimer interaction does not require a large conformational change

We next determined whether cloverleaf and 3C^{pro} interactions induced any structural changes in the RNA and/or protein components by comparing the individual cloverleaf and 3C^{pro} structures with those in the complex (*3*). The cloverleaf RNA structures could be superposed with an RMSD of 2.24 Å over all 70-nt pairs (Fig. 1C). The largest differences are found in stem-loop D, including the apical loop, upper stem, and UA bulge, the areas that interact with 3C^{pro}. In the complex, because of the symmetric interaction with Arg^{13} and Arg^{84} with stem-loop D, the major groove between U54 and U67 was enlarged to 11.0 from 8.8 Å in the native structure (Figs. 1C and 2A). As a result, the apical loop is tilted toward stem-loop C in the complex compared to the native cloverleaf structure. The apical loop ^{62}CACG^{65} of stem-loop D forms the UNCG motif, wherein the sugar edge of U62 interacts with the Watson-and-Crick face of *syn*-G65. The UNCG tetraloop conformation is largely maintained in the complex with only a small adjustment in the positions of A63 and C64 bases (Fig. 1C). There was no gross structural change in 3C^{pro} upon cloverleaf interaction. Each of the two monomers in the 3C^{pro} dimer in the complex could be superposed with the native 3C^{pro} (PDB code 3ZYD) with an RMSD of 0.56 Å over 179 Cα atoms.

3CD dimerization is induced by RNA interaction

Although two 3C^{pro} molecules bind cloverleaf, the two 3C^{pro} monomers do not have extensive interactions with each other, with only 185 Å^2 of buried surface area. The 3C^{pro} dimer interface seen in our structure is distinct from those previously observed in the crystal structures of enterovirus 3C^{pro} dimers, which bury between 750 and 880 Å^2 of surface area (fig. S3) (*42*–*46*). The lack of extensive dimer interface in 3C^{pro} suggests that 3CD (or 3C^{pro}) dimerizes on RNA, rather than a preassembled 3CD (or 3C^{pro}) dimer binding RNA. Thus, we tested the oligomerization state of CVB3 3C^{pro} and 3CD in the absence and presence of cloverleaf RNA using MP. MP measures the interference between the light scattered by the molecule and the light reflected by a measurement surface, which is directly correlated with molecular mass (*47*). In vitro–transcribed cloverleaf RNA without the tRNA scaffold was used for MP. We tested protein and cloverleaf RNA interaction at two different salt concentrations, 40 and 100 mM KCl. When 3C^{pro} (50 nM) was incubated in the absence of RNA, a single peak at 39 ± 7 kDa was apparent at both salt concentrations, which corresponds to 3C^{pro} dimer (21.4 × 2 = 42.8 kDa) (Fig. 4A). Note that the size of 3C^{pro} monomer (21 kDa) is below the detection limit of MP and, thus, would not be detected. To determine whether 3C^{pro} exists in monomer and dimer equilibrium in solution, 3C^{pro} was analyzed by size exclusion chromatography. 3C^{pro} elutes as a single peak corresponding to the dimer size (fig. S4B). Thus, CVB3 3C^{pro} exists predominantly as a dimer in solution.

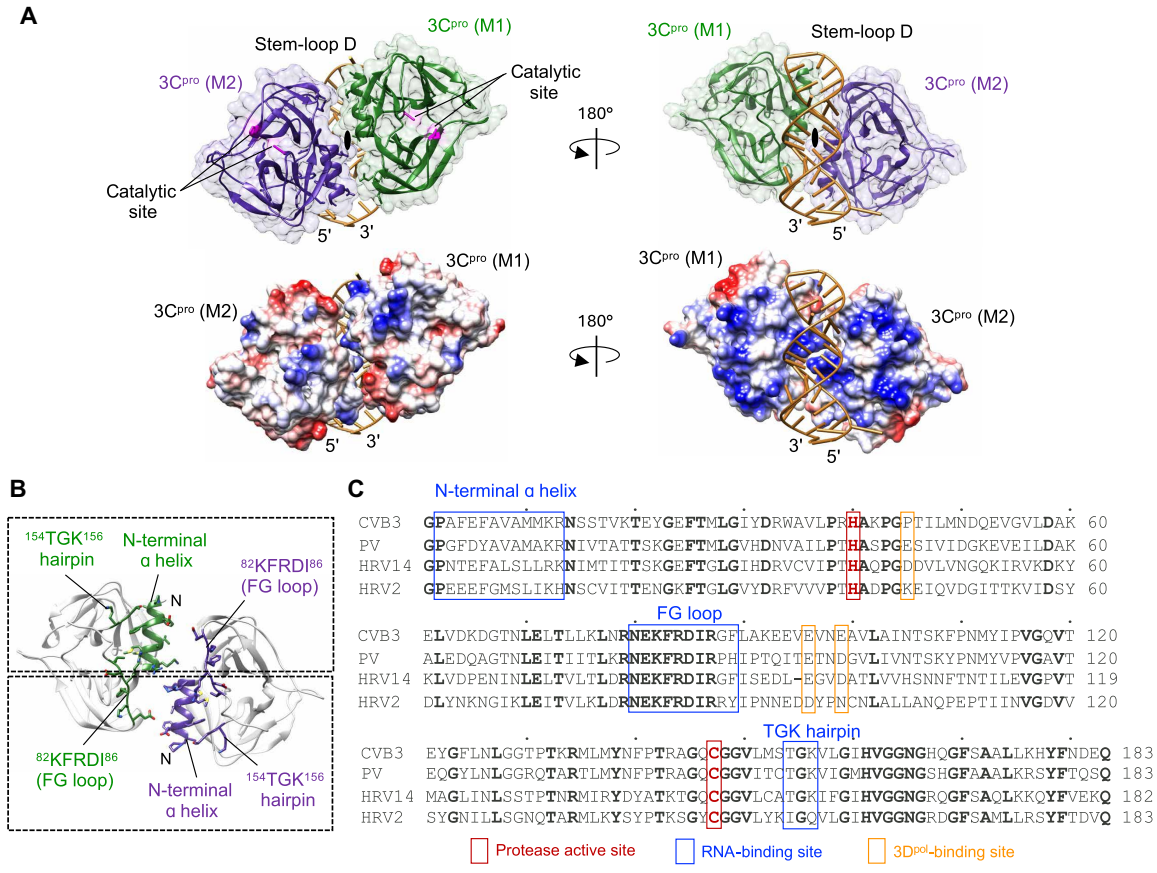

Fig. 2. Two 3C^{pro} proteins provide symmetric interaction sites for stem-loop D. (**A**) Arrangement of 3C^{pro} dimer on stem-loop D. Two 3C^{pro} monomers bind stem-loop D in opposite orientations. The 3C^{pro} residues are colored in green (M1) and violet (M2). The twofold axis of the dimer is indicated. 3C^{pro} folds into a double–β barrel structure, and the active site (His^{40}, Glu^{71}, and Cys^{147}, pink) is located between the two β barrels on the surface opposite the RNA-binding site. Electrostatic potential of the same surface of the dimer is shown below with positive charge shown in blue and negative charge in red. (**B**) Bipartite RNA-interaction surface formed by 3C^{pro} dimer. 3C^{pro} dimer forms two symmetric RNA-binding sites on stem-loop D. Each site is formed by the N-terminal helix and TGK hairpin of one 3C^{pro} monomer and FG loop of the other. (**C**) Sequence alignment of enterovirus 3C^{pro}. The 3C^{pro} sequences of CVB3 (AAC60718), PV (NP_740476), HRV14 (AAA45758), and HRV2 (WLV75740) were used for the alignment. The cloverleaf RNA-binding site, protease active site, and predicted 3D^{pol}-binding site are indicated. Conserved amino acids are shown in bold.

When 3C^{pro} (50 nM) was incubated with cloverleaf RNA (250 nM), an additional peak at 73 ± 12 kDa appeared, which corresponds to the molecular mass for 3C^{pro} dimer and one cloverleaf RNA (21.4 × 2 + 31.8 = 74.6 kDa). The amount of the 3C^{pro} dimer and cloverleaf RNA complex accounted for 17% of the total counts in 100 mM KCl and increased to 76% in 40 mM KCl (Fig. 4A). Thus, 3C^{pro} binds cloverleaf as a dimer, consistent with the structure.

When 3CD (50 nM) was incubated in the absence of RNA, a major peak appeared at 68 ± 12 kDa (40 mM KCl) or 73 ± 24 kDa (100 mM KCl), corresponding to a 3CD monomer (calculated mass of 73.8 kDa) (Fig. 4B). Additionally, a small peak at 147 ± 30 kDa (9% of the total counts), corresponding to the 3CD dimer, was observed at 40 mM KCl but not at 100 mM KCl (Fig. 4B). Thus, CVB3 3CD exists largely as a monomer with a small population of dimer. A ~45-kDa peak was also observed at 40 mM KCl, which was greatly reduced in 100 mM KCl. The identity of the peak is not clear, but it might be degradation products of 3CD, which aggregate at low-salt conditions but became soluble at high salt. Next, when 3CD (50 nM) was incubated with cloverleaf RNA (250 nM), one additional peak appeared at 174 ± 14 (40 mM KCl) or 170 ± 27 kDa (100 mM KCl), accounting for 19 to 23% of total counts. The mass corresponds to a 3CD dimer and one cloverleaf RNA (73.8 × 2 + 31.8 = 179.4 kDa). Thus, monomeric 3CD dimerizes on cloverleaf RNA, rather than binding as a preassembled dimer.

3D^{pol} domain of 3CD provides additional binding surface for interaction with cloverleaf RNA

The 3C^{pro} domain of 3CD is the major RNA binding domain, yet 3CD binds cloverleaf with ~10-fold higher affinity than the 3C^{pro} domain alone, suggesting that 3D^{pol} domain contributes to RNA interactions (*48*, *49*). Thus, we compared binding of CVB3 3C^{pro} and 3CD with stem-loop D (34 nt) and cloverleaf RNA using electrophoretic mobility shift assays (EMSAs). When stem-loop D was incubated with 3C^{pro} or 3CD, both proteins bind similarly (Fig. 5A). This confirms that the 3C^{pro} domain of 3CD is the major determinant for stem-loop D interaction, and the 3D^{pol} domain does not contribute to stem-loop D binding. This is also consistent with the structure, as 3C^{pro} interaction with cloverleaf RNA is localized on stem-loop D. In contrast, when full-length cloverleaf RNA was used for 3C^{pro} and 3CD interaction, 3CD binds to the cloverleaf with higher affinity than 3C^{pro}, indicating that the 3D^{pol} domain of 3CD provides an additional RNA-binding site to the cloverleaf RNA (Fig. 5B).

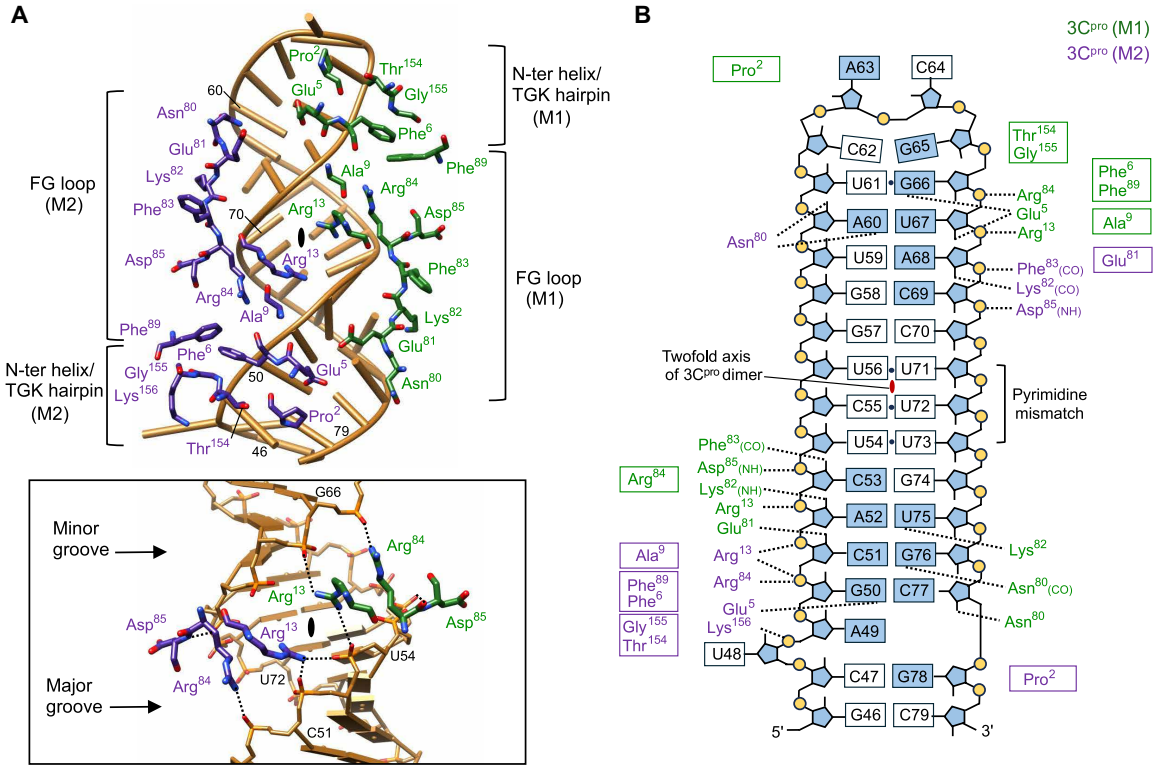

Fig. 3. 3C^{pro} dimer interacts primarily with dsRNA backbone across the major groove of stem-loop D. (**A**) RNA-binding site in 3C^{pro} dimer. 3C^{pro} residues that interact with stem-loop D are shown and colored as in Fig. 1. 3C^{pro} primarily binds the sugar-phosphate backbone with a few base-specific interactions. The symmetric interactions near the twofold axis of the 3C^{pro} dimer are shown below. Arg^{13}, Arg^{84}, and Asp^{85} from each monomer bind across the major groove and maintain the major groove width at 11 Å at both edges. (**B**) Schematic representation of 3C^{pro} and stem-loop D interactions. RNA regions that interact with the 3C^{pro} protein are shown in filled bases. Hydrogen bonds are shown as dotted lines, and van der Waals interactions are shown in boxes. Protein main-chain atoms that are involved in hydrogen bond interactions are indicated in parenthesis. The 3C^{pro} monomers are colored as in Fig. 1, and the twofold axis is shown.

To better understand how the 3D^{pol} domain enhances RNA binding affinity of 3CD, we modeled 3CD and cloverleaf interaction (Fig. 5C). The structure of PV 3CD protein shows that 3C^{pro} and 3D^{pol} domains, connected with a flexible seven-residue linker, do not interact with each other (PDB code 2IJD) (*33*). Two 3CD molecules were modeled onto the cloverleaf RNA by superposing the 3C^{pro} domain of 3CD with that of the complex. In the resulting model, the two 3D^{pol} domains of 3CD flank stem-loop C of cloverleaf in opposite orientations (Fig. 5C). The fingers subdomains of both 3D^{pol} monomers are located adjacent to stem-loop C, suggesting that the 3D^{pol} domains likely increase the binding affinity by providing potential surface of RNA interaction. Previously, Lys^{310} to Ala substitution in PV 3CD (corresponding to Lys^{127} in 3D^{pol} sequence) was shown to severely impair its ability to form the ternary complex with cloverleaf RNA and PCBP2 (*50*). In addition, transfections with PV cDNA constructs coding for 3D^{pol} mutations, Lys^{125}/Lys^{126}/Lys^{127} or Lys^{127}/Arg^{128}/Asp^{129} to Ala did not yield virus, suggesting that this area is important for viral replication (*51*). Lys^{127} is in the conserved basic region, ^{126}KKRDI^{130} of all enterovirus and rhinovirus 3D^{pol} proteins. In the cloverleaf-3CD model, the KKRDI region of 3D^{pol} (M1) is positioned near stem-loop C, suggesting that this region could play a role in RNA-protein interactions (fig. S5).

In the 3CD dimer-cloverleaf complex model, the cleavage site between 3C^{pro} (Gln^{183}) and 3D^{pol} (Gly^{184}) is positioned far away from the active site, indicating that the two monomers in the cloverleaf bound 3CD dimer would not cleave each other. The viral genome is likely circularized during replication, allowing the 3′ end of the genome to be positioned near the cloverleaf-3CD-PCBP2 complex at the 5′ end (*29*, *30*). Because 3CD also interacts with the 3′ end of the genome near the 3′ polyA tail (*49*), the 3CD in complex with cloverleaf RNA could cleave the 3CD bound at the 3′ end of the genome to generate 3D^{pol}. Alternatively, the 3CD dimer on cloverleaf RNA could provide an interaction site for 3D^{pol} and recruit 3D^{pol} to initiate RNA synthesis at the 3′ end of the genome.

DISCUSSION

What is the consensus RNA motif recognized by 3C^{pro}?

3C^{pro} specifically binds cloverleaf RNA, yet the cloverleaf-3C^{pro} dimer structure shows that 3C^{pro} primarily binds to the phosphate backbone of the dsRNA rather than interacting with specific bases. This suggests that 3C^{pro} recognizes the RNA sequence indirectly via directly interacting with the sugar-phosphate backbone. Thus, 3C^{pro} is a previously undiscovered type of RNA binding protein that recognizes the sequence-specific shape of base-paired dsRNA as a dimer. In this case, what provides the specificity of the 3C^{pro} dimer to bind cloverleaf or CRE RNAs? Crystal structures of enterovirus 3C^{pro} show an identical chymotrypsin fold. Thus, specific interactions between enteroviral 3C^{pro} proteins and their cognate cloverleaf RNAs may be required for virus-specific protein-RNA interactions.

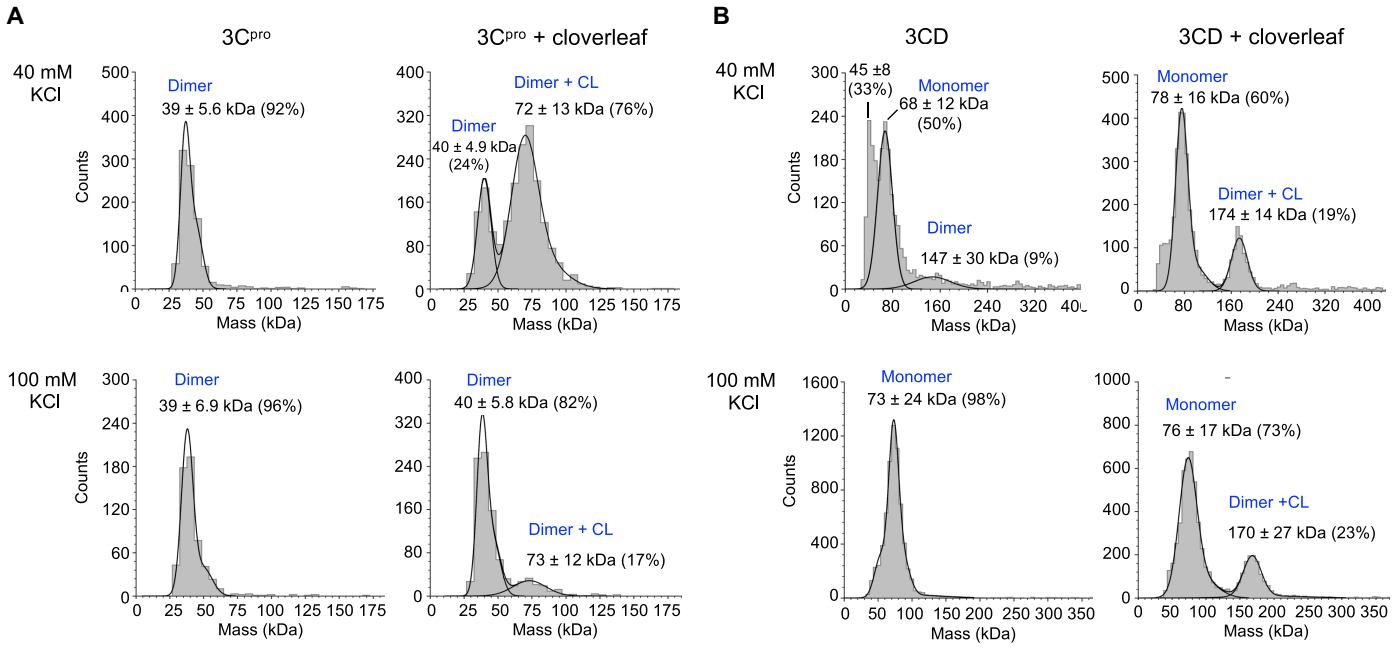

Fig. 4. 3C^{pro} and 3CD proteins bind cloverleaf as dimers as measured by MP. (**A**) 3C^{pro} interaction with cloverleaf RNA. Distribution of molecular mass of 3C^{pro} was determined in the absence and presence of cloverleaf RNA in a buffer containing either 40 mM (top) or 100 mM KCl (bottom). The oligomerization states are determined using the molecular masses of 3C^{pro} (21.4 kDa) and cloverleaf RNA (31.8 kDa). 3C^{pro} exists as a dimer in both 40 and 100 mM KCl conditions and binds cloverleaf RNA as a dimer. (**B**) 3CD interaction with cloverleaf RNA. Distribution of molecular mass of 3CD was determined as in (A). The oligomerization states are determined on the basis of the molecular mass of 3CD (73.8 kDa) and cloverleaf RNA (31.8 kDa). 3CD exists primarily as a monomer in 40 or 100 mM salt concentrations but binds cloverleaf as a dimer.

The N-terminal helix in the RNA-binding site of 3C^{pro} shows large sequence variation among enteroviruses (Fig. 2C). Thus, this area may play a role in distinguishing the cognate viral RNA from other RNAs. Additionally, although 3C^{pro} does not recognize specific bases, a characteristic recognition motif could be deduced from comparing the primary and secondary RNA structures of cloverleaf RNAs that bind CVB3 3C^{pro} (*52*). They include cloverleaf from PV, CVB3, HRV14, and bovine enterovirus (BEV), and cloverleaf mutants of CVB3 and HRV2 (Fig. 6A). To a first approximation, the 3C^{pro} recognition site is a stem-loop structure composed of an apical loop and a continuous dsRNA stem of at least 12 base pairs (BP 1-12), where 3C^{pro} interacts with BP 1-4 and BP 9-12. Second, similar base types are observed in symmetrically equivalent positions within the 3C^{pro}-binding site (BP 1-4 and BP 9-12), allowing 3C^{pro} to bind as a dimer. The two 3C^{pro} monomers of the dimer interact with ^{50}GCAC^{53}/^{66}GUAC^{69} in CVB3 cloverleaf RNA, which corresponds to GUAC/GUAC in PV and GUAC/GUAA in HRV2 cloverleaf RNAs. Third, the central four base pairs (BP 5-8) are U rich and may prefer to form noncanonical base pair interactions. The 3C^{pro} dimer interacts with the major groove of this central region and increases the width of the major groove to 11 Å (Fig. 1C). The width of the major groove in dsRNA is determined by base pair sequences and can vary by as much as 6 Å, ranging from 5 to 11 Å (*53*). Thus, 3C^{pro} may prefer binding to dsRNA sequences that form "breathable" stems that could easily expand upon binding. Last, a tetra-loop sequence at the apical stem-loop is preferred over tri-loop or penta-loop, which is seen in HRV 14 and BEV cloverleaf 2, respectively (Fig. 6A) (*52*). A purine (G or A) nucleotide is preferred at the 3′ terminal position within the tetra-loop (Fig. 6A).

Implications for 3CD dimer interaction with CRE

Picornaviruses use CRE, an internal stem-loop structure, to generate the uridylylated VPg primer. CRE has a large 14-nt loop, wherein the AAA sequence is present at 5- to 7-nt position of the loop (*54*). Viral 3CD and 3D^{pol} proteins assemble on CRE (Fig. 6B), and 3D^{pol} uses the AAA sequence within the apical loop of CRE as the template for the VPg uridylylation reaction (*55*, *56*). 3CD was proposed to bind CRE as a dimer, which then recruits 3D^{pol} to CRE (*26*, *57*–*60*). The discovery that 3CD (and 3C^{pro}) binds cloverleaf RNA as a dimer suggests that 3CD interaction with CRE is likely analogous to that observed with cloverleaf RNA. For 3CD and CRE interaction, it was shown that the 3C^{pro} domain of 3CD specifically recognizes the stem proximal to the apical loop of CRE and that its interaction with the stem is not sequence specific but does require a stabilized dsRNA helix (*25*, *57*, *61*). The same RNA-binding residues in 3C^{pro} (FG loop and TGK hairpin) were implicated for both cloverleaf and CRE interaction (*38*, *52*). In addition, the N-terminal residue Asp^5 was suggested to be involved in CRE interaction, because a substitution of Asp^5 to Ala in PV 3C^{pro} reduced CRE binding by twofold (*62*). Although the N-terminal region of 3C^{pro} has not been implicated in cloverleaf interaction, our structure shows that the N-terminal α-helix containing Glu^5 is part of the RNA-binding site (Fig. 3A). Thus, the 3C^{pro} dimer likely binds to the CRE stem-loop along the phosphate backbone, with a similar mode of interaction as observed in our structure of the cloverleaf-3C^{pro} dimer complex.

To provide further insight into the assembly of the VPg uridylylation complex, we modeled the CRE, 3C^{pro} dimer, and 3D^{pol} complex. CRE RNA from HRV14 (PDB code 1T28) was superimposed with stem-loop D of cloverleaf in the cloverleaf-3C^{pro} complex (Fig. 6C). In the resulting

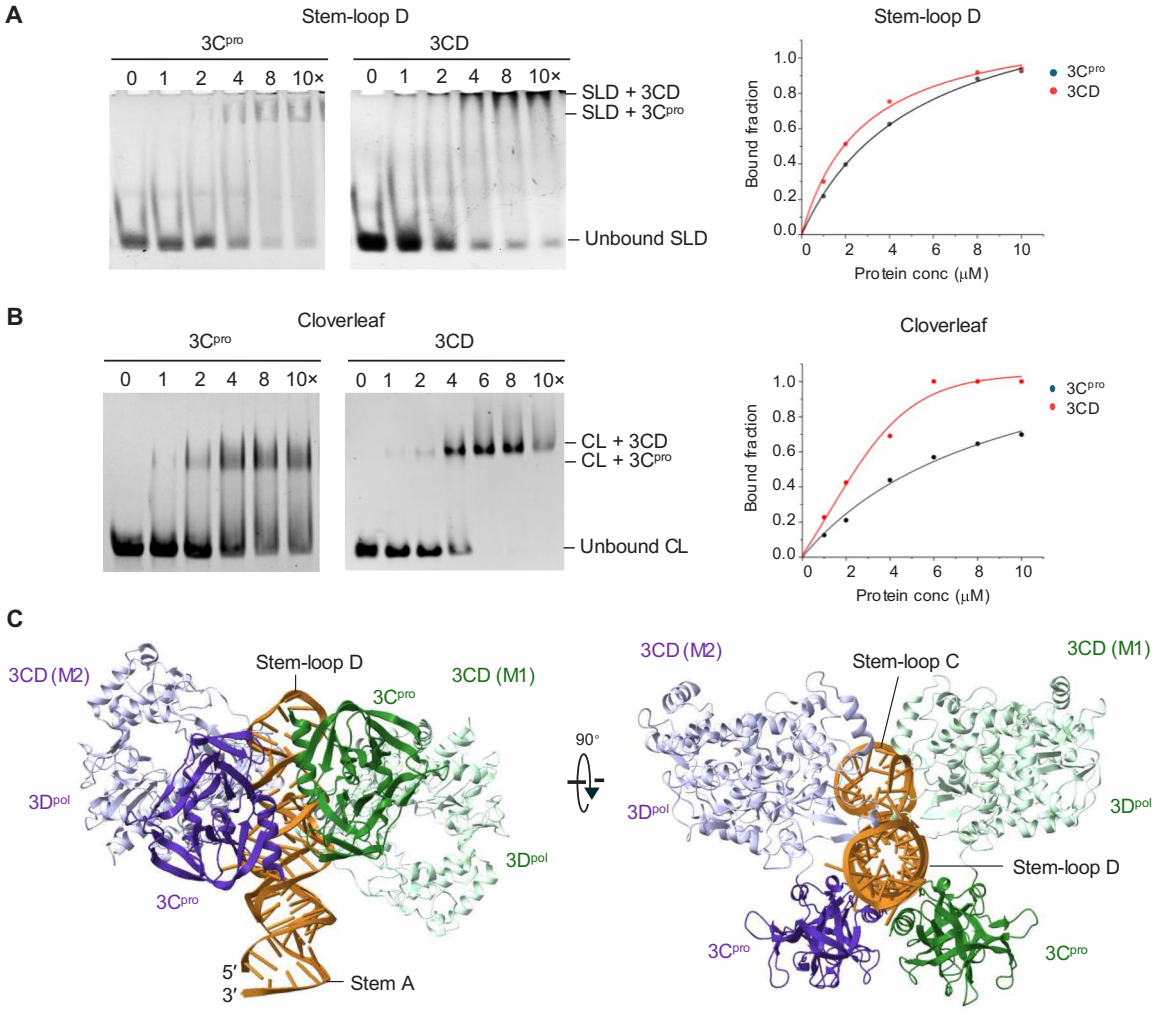

Fig. 5. 3CD increases the binding affinity to full-length cloverleaf RNA but not stem-loop D. (**A**) Gel shift assays of $3C^{pro}$ and 3CD with cloverleaf stem-loop D. Stem-loop D RNA (SLD; 1 μM) was incubated with increasing concentrations of protein (1 to 10 μM). Both proteins bind stem-loop D similarly. (**B**) Gel shift assays of $3C^{pro}$ and 3CD with cloverleaf RNA. Cloverleaf RNA (CL; 1 μM) was incubated with increasing molar excess of protein (1 to 10 μM). 3CD protein binds cloverleaf RNA with higher affinity than $3C^{pro}$. (**C**) Model of 3CD and cloverleaf RNA interaction. The 3CD and cloverleaf complex was modeled by superposition of $3C^{pro}$ domain of 3CD (PDB code 2IJD) onto each of the $3C^{pro}$ domains in the cloverleaf-$3C^{pro}$ dimer complex. $3D^{pol}$ domains (lighter colors) of both 3CD proteins extend toward and flank stem-loop C, likely providing additional interaction with cloverleaf RNA.

CRE-$3C^{pro}$ dimer model, the stem of CRE can interact with $3C^{pro}$ dimer similar to the way cloverleaf RNA and the $3C^{pro}$ dimer interact in their complex. We mapped the PV $3C^{pro}$ residues implicated in interaction with $3D^{pol}$, Glu^{45}, Glu^{96}, and Asp^{99} (corresponding to Pro^{45}, Glu^{96}, and Glu^{99} in CVB3 $3C^{pro}$) on the CRE-$3C^{pro}$ dimer complex (Figs. 2C and 6C) (*62*). All three residues are surface exposed. In particular, Glu^{96} and Glu^{99} of M1 monomer lie on the surface facing the apical loop of CRE, ideal for $3D^{pol}$ interaction (Fig. 6C). To model the $3C^{pro}$ dimer and $3D^{pol}$ interaction, we next used a structure of $3D^{pol}$ complexed with template and primer (PDB code 3KOA) and overlaid the RNA template on the AAA region of CRE (*63*). $3D^{pol}$ could be placed on $3C^{pro}$ without steric clashes (Fig. 6D). In the resulting model of the CRE-$3C^{pro}$ dimer-$3D^{pol}$ complex, the apical loop of CRE is inserted into $3D^{pol}$ from the front of the polymerase into the template-binding channel, and the active site of $3D^{pol}$ is placed near the AAA sequence in the CRE loop, ideally positioned for VPg uridylylation (Fig. 6D). It seems that the size of the CRE loop ensures that $3D^{pol}$ is within the proper distance from the $3C^{pro}$ surface. A longer or shorter CRE loop would not support stable engagement on the $3C^{pro}$ dimer, consistent with previous reports (*37*, *48*). Stabilizing interactions between $3C^{pro}$ and $3D^{pol}$ would also likely restrict the movement of $3D^{pol}$ on the CRE template, limiting the number of uridine 5′-monophosphates linked to VPg during the uridylylation reaction. In the current model of VPg uridylylation, the first uridine 5′-triphosphate (UTP) base pairs with the first A of the AAA sequence within the CRE loop. After addition of a single uridine to VPg, the nascent VPg-pU then slides back and the second UTP base pairs with the first A of the AAA sequence (*64*). It seems likely that catalysis in the confined space of the CRE-3CD dimer-$3D^{pol}$ complex may be a mechanism to restrict VPg uridylylation to VPg-pUpU.

METHODS

Construction, expression, and purification of CVB3 CL-tRNA

The CL-tRNA fusion construct for CVB3 was previously described (*3*). Briefly, the CVB3 cloverleaf RNA sequence (nucleotides 1 to 87 of AY752944.2) was inserted into the anticodon loop of human $tRNA^{Lys}$

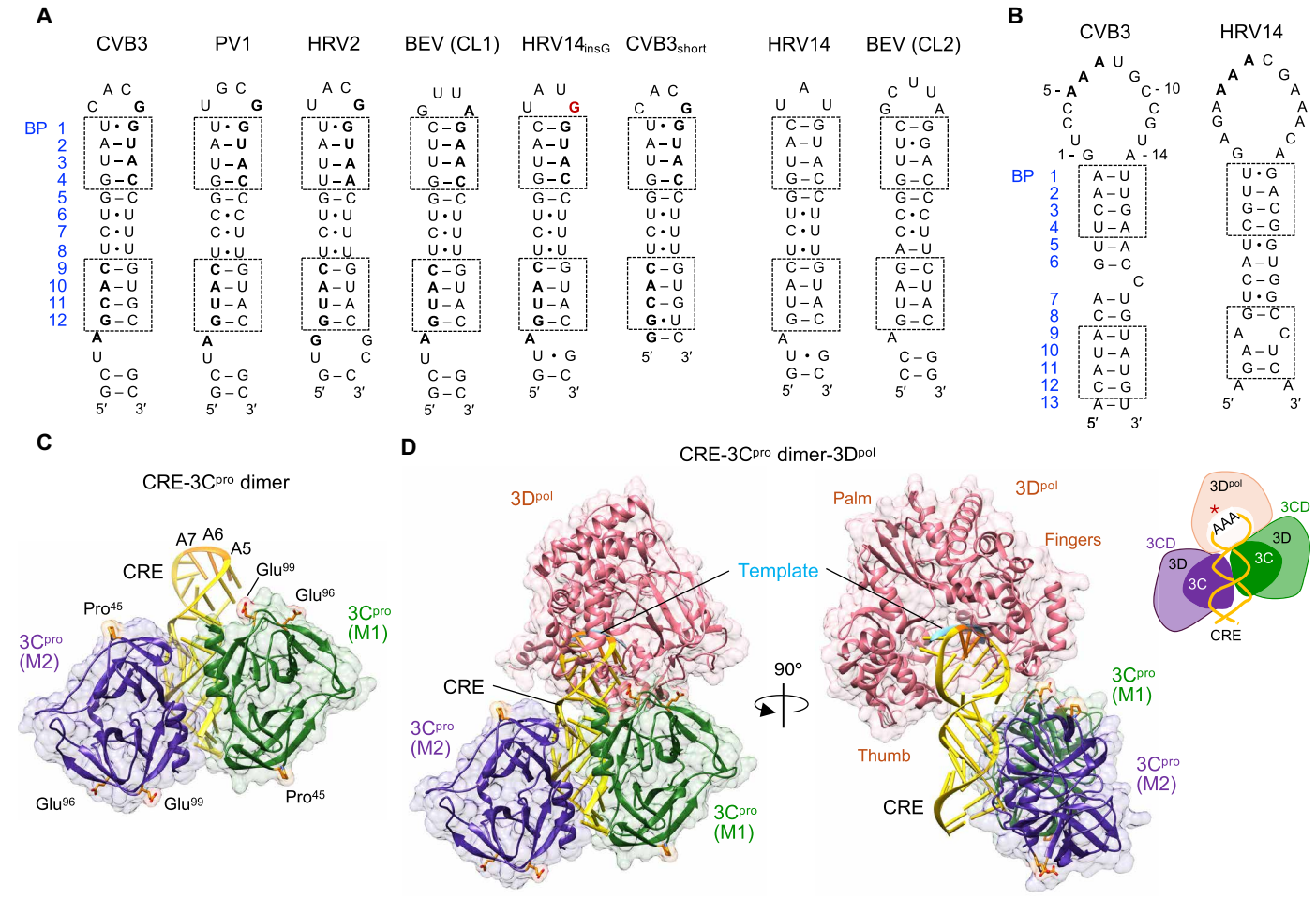

Fig. 6. 3C^{pro} interaction with cis-acting RNAs. (**A**) Sequence and secondary structure comparison of cloverleaf RNA. Cloverleaf RNA from CVB3, PV, HRV2, BEV, and mutants (HRV14$_{insG}$ and CVB3$_{short}$) bind CVB3 3C^{pro} (*52*). In contrast, the HRV14 cloverleaf and BEV cloverleaf 2 (CL2) RNAs do not bind CVB3 3C^{pro} (right). The base pairs (BP) that interact with 3C^{pro} dimer in the cloverleaf-3C^{pro} dimer complex are boxed, and symmetrically related nucleotides within the 3C^{pro} interaction sites are shown in bold. (**B**) The secondary structure of CVB3 and HRV14 CRE. CRE contains a 14-nt apical loop, wherein the AAA sequence positioned at 5 to 7 nt (bold) is used for VPg uridylylation. The predicted bipartite 3C^{pro} dimer interaction sites are indicated as in (A). (**C**) Model of the CRE-3C^{pro} dimer complex. The CRE-3C^{pro} dimer complex was modeled by superposing CRE (PDB code 1T28) onto stem-loop D of the clovlerleaf-3C^{pro} dimer complex. The AAA sequence in CRE was colored in orange and labeled. The predicted 3D^{pol} binding residues (Pro^{45}, Glu^{96}, and Glu^{99}) are depicted on 3C^{pro}. (**D**) Model of the CRE-3C^{pro} dimer-3D^{pol} complex. 3D^{pol} was modeled by superposing the template strand in the 3D^{pol}-template-primer complex (PDB code 3KOA) with the AAA sequence of CRE. A new 3D^{pol} recruited to the 3C^{pro} proximal to the apical loop positions the active site of the polymerase near the AAA sequence of CRE, conducive to VPg uridylylation. The 3D^{pol} subdomains are indicated. Schematic of the CRE, 3C^{pro} dimer, and 3D^{pol} complex is shown on the right.

with additional C88 to base pair with G1 to generate CL-tRNA (153 nt). DNA encoding the designed CL-tRNA was synthesized at Epoch Biosciences (Houston, TX) and inserted into pBluescript II SK(+) vector under the control of *Escherichia coli* (*E. coli*) *lpp* promoter and *rrnC* terminator. The plasmid containing CL-tRNA was transformed into *E. coli* BL21(DE3) cells, and the cells were grown overnight in 2× YT (yeast extract tryptone) medium at 37°C. The cells from 2 liters of culture were harvested and resuspended in 9 ml of 10 mM tris-HCl buffer (pH 7.4) containing 10 mM magnesium acetate. Saturated phenol (12 ml) was added to the cell suspension to extract RNA, and the solution was centrifuged at 20,000*g* for 30 min at 20°C to remove cell debris. The resulting aqueous phase was mixed with 0.1 volumes of 5 M NaCl and 2 volumes of absolute ethanol, and the RNA pelleted by centrifugation at 20,000*g* for 30 min at 4°C. The RNA pellet was dissolved in buffer A [40 mM sodium phosphate or tris-Cl (pH 7.0)] and loaded on HiLoad 16/20 Q HP anion-exchange column (Cytiva, formerly GE Life Sciences) equilibrated with buffer A. The CL-tRNA was eluted with 0.5 to 0.6 M NaCl gradient in buffer A. The RNA fractions were analyzed by electrophoresis on 8% urea-polyacrylamide gel. The pooled sample was buffer exchanged to water and concentrated to 15 to 20 mg/ml.

In vitro transcription of cloverleaf and stem-loop D RNA

CVB3 cloverleaf and stem-loop D RNAs were generated by in vitro transcription using T7 RNA polymerase. The DNA templates for CVB3 cloverleaf and stem-loop D RNAs were individually generated using polymerase chain reaction from the CL-tRNA plasmid. RNA was in vitro transcribed using standard T7 RNA polymerase transcription methods (*65*). The transcription reaction was prepared by combining 10 mM ribonucleoside triphosphates (rNTPs), 200 nM

DNA template, and T7 RNA polymerase (200 μg/ml) in buffer containing 40 mM tris-HCl (pH 8.3), 15 mM $MgCl_2$, 2.5 mM spermidine, 10 mM dithiothreitol (DTT), and 0.01% Triton X-100. After 15 min of incubation of the reaction mixture at 37°C, 2 μl of thermostable inorganic pyrophosphatase (2000 U/ml, New England Biolabs) was added and the reaction continued for 4 hours. At the end of the reaction, deoxyribonuclease (DNase) was added to the transcription product in 1:100 ratio and incubated for 30 min at 37°C. Last, 50 mM EDTA was added, and the sample was spun down. The supernatant was diluted in buffer A [40 mM sodium phosphate (pH 7.0)] and loaded onto a Mono Q 5/50 GL ion-exchange column (Cytiva). The desired RNA eluted between 520 and 570 mM NaCl in buffer A. Fractions were analyzed on a urea-polyacrylamide gel, and RNA-containing fractions were pooled and buffer exchanged into water. Typical yield is ~28 μg of RNA from 1 ml of T7 RNA transcription reaction for stem-loop D and ~250 μg for cloverleaf RNA.

Expression and purification of $3C^{pro}$ and 3CD proteins

The full-length CVB3 3CD protein (AAC60718) was designed with the catalytic Cys^{147} to Ala substitution and a C-terminal hexahistidine tag to aid purification of the recombinant protein. In addition, based on (*33*), Gly^{55}, Asp^{58}, and Val^{63} (corresponding to Glu^{55}, Asp^{58}, and Glu^{63} in PV $3C^{pro}$) were mutated to Ala to reduce aggregation. The full-length 3CD DNA construct was synthesized at Azenta (previously Genewiz) and subcloned into a pET28b(+) expression vector for recombinant expression. DNA plasmid encoding CVB3 $3C^{pro}$ was generated by subcloning the $3C^{pro}$ sequence from the plasmid containing the 3CD protein.

3CD and $3C^{pro}$ proteins were expressed in *E. coli* BL21(DE3) cells and purified as previously described with some modifications (*33*). Cells were grown in Luria broth (LB) supplemented with kanamycin (30 mg/ml) at 37°C to an optical density at 600 nm of ~0.9, and protein expression was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside with growth continued overnight at 18°C. For protein purification, the cell pellet from a 2-liter culture was resuspended in 30 ml of lysis buffer [50 mM tris (pH 8.0), 1 M NaCl, 10% (v/v) glycerol, 60 μM $ZnCl_2$, 10 mM imidazole, 5 mM β-mercaptoethanol, 0.1% NP-40, DNase (100 μg/ml), and cOmplete EDTA-free protease inhibitor cocktail (Roche)] and lysed by sonication. Protein in the soluble fraction of the lysate was loaded onto His-Pur Ni–nitrilotriacetic acid (Thermo Fisher Scientific) metal-affinity chromatography resin pre-equilibrated in lysis buffer (without the protease inhibitor cocktail and NP-40). Bound proteins were eluted using a gradient of 25 to 500 mM imidazole in elution buffer [50 mM tris (pH 8.0), 300 M NaCl, 10% (v/v) glycerol, 60 μM $ZnCl_2$, and 5 mM β-mercaptoethanol]. Fractions containing $3C^{pro}$ or 3CD were pooled and further purified using size exclusion chromatography on HiLoad Superdex 16/600 S200 prep-grade column (Cytiva) equilibrated with buffer containing 50 mM tris (pH 8.0), 250 M NaCl, 10% (v/v) glycerol, 60 μM $ZnCl_2$, and 1 mM Tris(2-carboxyethyl)phosphine (TCEP). Purity was assessed using SDS–polyacrylamide gel electrophoresis (PAGE).

Crystallization, x-ray data collection, and structure determination

Purified $3C^{pro}$ was mixed with CL-tRNA in a 1:1 molar ratio, and crystallization trials were performed in sitting-drop format using commercial sparse matrix crystallization screens. A drop volume of 0.3 μl and well solution of 50 μl were used for the trials. Successful hits were further optimized by sitting-drop vapor diffusion method at 20°C. Diffraction quality crystals were obtained after 12 to 15 days in 8% (v/v) Tacsimate (pH 6.0) and 25% (w/v) of polyethylene glycol 3350. For data collection, crystals were passed through mother liquor with 20% (v/v) glycerol and flash frozen in liquid nitrogen. Diffraction data to 2.87-Å resolution were collected at the Advanced Light Source beamline 8.2.2 (Lawrence Berkley National Laboratory, CA) and processed using HKL 2000 (*66*). The crystal belonged to space group $P2_12_12$ with unit cell dimensions of $a = 91.02$ Å, $b = 200.82$ Å, and $c = 71.12$ Å. The initial solutions were found by molecular replacement with CVB3 cloverleaf (PDB code 8SP9) and $3C^{pro}$ (PDB code 3ZYD) as search models using the Phaser program within the Phenix application suite (*67*). Reflections with $I/\sigma I \geq 0.6$ were included during refinement. The final model contained one molecule of CVB3 cloverleaf and two molecules of $3C^{pro}$ in the asymmetric unit with a solvent content of 66% and was refined to an R and R_{free} of 0.24 and 0.27, respectively. The tRNA scaffold was cleaved from the CL-tRNA as was later verified using TBE/urea-PAGE of crystals. The final CVB3 cloverleaf and $3C^{pro}$ complex structure contains the entire cloverleaf RNA, except nucleotides 14 to 25, and two $3C^{pro}$ monomers, except Gly^{1} and Arg^{143} in molecule M2 and three amino acids at the C terminus (residues 181 to 183) in both M1 and M2 molecules. For structural analysis, Coot and ChimeraX were used to determine atomic distances, angles, and surface areas (*68*, *69*). Atomic coordinates for the CVB3 CL and $3C^{pro}$ complex are deposited to PDB with accession number 9DCF.

MP analysis

MP measurements of cloverleaf RNAs (full-length cloverleaf and stem-loop D) and proteins ($3C^{pro}$ and 3CD) were performed on TwoMP mass photometer (Refeyn Ltd.). Cloverleaf RNA without the tRNA scaffold was used in the measurement. Before each measurement, RNA samples were premixed in an EMSA binding buffer [5 mM Hepes (pH 7.4), 40 or 100 mM KCl, 2.5 mM $MgCl_2$, 2 mM DTT, and 5% glycerol]. For mass measurement, silicone gaskets with six wells were placed on microscope coverslips (sample kit from Refeyn Ltd.), and 16 μl of the buffer was placed in the well. The microscope was focused on the coverslip surface. To obtain the mass of proteins, 4 μl of CVB3 $3C^{pro}$ or 3CD protein was mixed with 16 μl of the buffer (final concentration of 50 nM). To obtain the mass of the protein and RNA complexes, the protein samples were incubated with fivefold excess of cloverleaf RNA (250 nM). Sample binding to the coverslip surface was monitored for 60 s using the AcquireMP software (Refeyn Ltd.). Data analysis was performed using DiscoverMP (Refeyn Ltd.). To convert the measured optical-reflection contrast into molecular mass, a standard curve was created on the basis of dengue virus NS3 and NS5 proteins on the same day (fig. S6).

RNA-protein binding gel shift assays

EMSAs were used to measure the interaction between CVB3 cloverleaf RNAs (full-length cloverleaf and stem-loop D) and proteins ($3C^{pro}$ and 3CD). For EMSA, RNA (1 μM) was incubated with increasing molar amounts of protein (1 to 10 μM) in binding buffer [5 mM Hepes (pH 7.4), 25 mM KCl, and 2.5 mM $MgCl_2$, 2 mM DTT, and 5% glycerol] for 15 min at room temperature and loaded on a 6 to 12% native polyacrylamide gel. The gel was stained with ethidium bromide to visualize RNA. EMSA was also used to test whether CL-tRNA forms a similar complex with $3C^{pro}$ as the unmodified cloverleaf RNA (fig. S7). CL-tRNA, in vitro–transcribed cloverleaf RNA and tRNA alone (negative control) were incubated with increasing

amount of $3C^{pro}$. $3C^{pro}$ binds CL-tRNA and in vitro–transcribed cloverleaf RNA similarly, while it does not bind tRNA. Thus, the tRNA scaffold does not contribute to nucleoprotein complex formation. All EMSA experiments were repeated at least twice.

Supplementary Materials

This PDF file includes:

Figs. S1 to S7

REFERENCES AND NOTES

1. K. H. Choi, The role of the stem-loop A RNA promoter in flavivirus replication. *Viruses* **13**, 1107 (2021).
2. E. Lee, P. J. Bujalowski, T. Teramoto, K. Gottipati, S. D. Scott, R. Padmanabhan, K. H. Choi, Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. *Nat. Commun.* **12**, 2530 (2021).
3. K. Gottipati, S. C. McNeme, J. Tipo, M. A. White, K. H. Choi, Structural basis for cloverleaf RNA-initiated viral genome replication. *Nucleic Acids Res.* **51**, 8850–8863 (2023).
4. A. Schuessler, A. Funk, H. M. Lazear, D. A. Cooper, S. Torres, S. Daffis, B. K. Jha, Y. Kumagai, O. Takeuchi, P. Hertzog, R. Silverman, S. Akira, D. J. Barton, M. S. Diamond, A. A. Khromykh, West Nile virus noncoding subgenomic RNA contributes to viral evasion of the type I interferon-mediated antiviral response. *J. Virol.* **86**, 5708–5718 (2012).
5. P. S. Masters, Coronavirus genomic RNA packaging. *Virology* **537**, 198–207 (2019).
6. J. A. Fosmire, K. Hwang, S. Makino, Identification and characterization of a coronavirus packaging signal. *J. Virol.* **66**, 3522–3530 (1992).
7. M. Pons-Salort, E. P. Parker, N. C. Grassly, The epidemiology of non-polio enteroviruses: Recent advances and outstanding questions. *Curr. Opin. Infect. Dis.* **28**, 479–487 (2015).
8. C. C. Holm-Hansen, S. E. Midgley, T. K. Fischer, Global emergence of enterovirus D68: A systematic review. *Lancet Infect. Dis.* **16**, e64–e75 (2016).
9. L. Kainulainen, T. Vuorinen, K. Rantakokko-Jalava, R. Osterback, O. Ruuskanen, Recurrent and persistent respiratory tract viral infections in patients with primary hypogammaglobulinemia. *J. Allergy Clin. Immunol.* **126**, 120–126 (2010).
10. N. M. Chapman, K. S. Kim, Persistent coxsackievirus infection: Enterovirus persistence in chronic myocarditis and dilated cardiomyopathy. *Curr. Top. Microbiol. Immunol.* **323**, 275–292 (2008).
11. A. Kloc, D. K. Rai, E. Rieder, The roles of picornavirus untranslated regions in infection and innate immunity. *Front. Microbiol.* **9**, 485 (2018).
12. D. J. Barton, B. J. O'Donnell, J. B. Flanegan, 5′ cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. *EMBO J.* **20**, 1439–1448 (2001).
13. A. V. Paul, E. Wimmer, Initiation of protein-primed picornavirus RNA synthesis. *Virus Res.* **206**, 12–26 (2015).
14. R. Andino, G. E. Rieckhof, D. Baltimore, A functional ribonucleoprotein complex forms around the 5′ end of poliovirus RNA. *Cell* **63**, 369–380 (1990).
15. A. V. Gamarnik, R. Andino, Switch from translation to RNA replication in a positive-stranded RNA virus. *Genes Dev.* **12**, 2293–2304 (1998).
16. H. Toyoda, D. Franco, K. Fujita, A. V. Paul, E. Wimmer, Replication of poliovirus requires binding of the poly(rC) binding protein to the cloverleaf as well as to the adjacent C-rich spacer sequence between the cloverleaf and the internal ribosomal entry site. *J. Virol.* **81**, 10017–10028 (2007).
17. J. O. Cifuente, G. Moratorio, Evolutionary and structural overview of human picornavirus capsid antibody evasion. *Front. Cell. Infect. Microbiol.* **9**, 283 (2019).
18. A. Palmenberg, D. Neubauer, T. Skern, "Genome organization and encoded proteins" in *The Picornaviruses* (ASM Press, 2010), pp. 1–17.
19. E. Wimmer, C. U. Hellen, X. Cao, Genetics of poliovirus. *Annu. Revi. Genet.* **27**, 353–436 (1993).
20. J. Yuan, L. Shen, J. Wu, X. Zou, J. Gu, J. Chen, L. Mao, Enterovirus A71 proteins: Structure and function. *Front. Microbiol.* **9**, 286 (2018).
21. Y. Liu, D. Franco, A. V. Paul, E. Wimmer, Tyrosine 3 of poliovirus terminal peptide VPg(3B) has an essential function in RNA replication in the context of its precursor protein, 3AB. *J. Virol.* **81**, 5669–5684 (2007).
22. A. V. Paul, J. H. van Boom, D. Filippov, E. Wimmer, Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature* **393**, 280–284 (1998).
23. B. P. Steil, D. J. Barton, Conversion of VPg into VPgpUpUOH before and during poliovirus negative-strand RNA synthesis. *J. Virol.* **83**, 12660–12670 (2009).
24. K. H. Choi, Viral polymerases. *Adv. Exp. Med. Biol.* **726**, 267–304 (2012).
25. K. L. McKnight, S. M. Lemon, The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA* **4**, 1569–1584 (1998).
26. H. B. Pathak, J. J. Arnold, P. N. Wiegand, M. R. Hargittai, C. E. Cameron, Picornavirus genome replication: Assembly and organization of the VPg uridylylation ribonucleoprotein (initiation) complex. *J. Biol. Chem.* **282**, 16202–16213 (2007).
27. D. Flather, B. L. Semler, Picornaviruses and nuclear functions: Targeting a cellular compartment distinct from the replication site of a positive-strand RNA virus. *Front. Microbiol.* **6**, 594 (2015).
28. T. Lyons, K. E. Murray, A. W. Roberts, D. J. Barton, Poliovirus 5′-terminal cloverleaf RNA is required in cis for VPg uridylylation and the initiation of negative-strand RNA synthesis. *J. Virol.* **75**, 10696–10708 (2001).
29. J. Herold, R. Andino, Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol. Cell* **7**, 581–591 (2001).
30. D. A. Vogt, R. Andino, An RNA element at the 5′-end of the poliovirus genome functions as a general promoter for RNA synthesis. *PLOS Pathog.* **6**, e1000936 (2010).
31. N. K. Das, N. M. Hollmann, J. Vogt, S. E. Sevdalis, H. A. Banna, M. Ojha, D. Koirala, Crystal structure of a highly conserved enteroviral 5′ cloverleaf RNA replication element. *Nat. Commun.* **14**, 1955 (2023).
32. N. K. Das, J. Vogt, A. Patel, H. A. Banna, D. Koirala, Structural basis for a highly conserved RNA-mediated enteroviral genome replication. *Nucleic Acids Res.* **52**, 11218–11233 (2024).
33. L. L. Marcotte, A. B. Wass, D. W. Gohara, H. B. Pathak, J. J. Arnold, D. J. Filman, C. E. Cameron, J. M. Hogle, Crystal structure of poliovirus 3CD protein: Virally encoded protease and precursor to the RNA-dependent RNA polymerase. *J. Virol.* **81**, 3583–3596 (2007).
34. D. Shengjuler, Y. M. Chan, S. Sun, I. M. Moustafa, Z. L. Li, D. W. Gohara, M. Buck, P. S. Cremer, D. D. Boehr, C. E. Cameron, The RNA-binding site of poliovirus 3C protein doubles as a phosphoinositide-binding domain. *Structure* **25**, 1875–1886.e7 (2017).
35. B. Meng, K. Lan, J. Xie, R. A. Lerner, I. A. Wilson, B. Yang, Inhibitory antibodies identify unique sites of therapeutic vulnerability in rhinovirus and other enteroviruses. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13499–13508 (2020).
36. L. E. Leong, P. A. Walker, A. G. Porter, Human rhinovirus-14 protease 3C (3Cpro) binds specifically to the 5′-noncoding region of the viral RNA. Evidence that 3Cpro has different domains for the RNA binding and proteolytic activities. *J. Biol. Chem.* **268**, 25735–25739 (1993).
37. P. A. Walker, L. E. Leong, A. G. Porter, Sequence and structural determinants of the interaction between the 5′-noncoding region of picornavirus RNA and rhinovirus protease 3C. *J. Biol. Chem.* **270**, 14510–14516 (1995).
38. C. D. Amero, J. J. Arnold, I. M. Moustafa, C. E. Cameron, M. P. Foster, Identification of the oril-binding site of poliovirus 3C protein by nuclear magnetic resonance spectroscopy. *J. Virol.* **82**, 4363–4370 (2008).
39. S. R. Shih, C. Chiang, T. C. Chen, C. N. Wu, J. T. Hsu, J. C. Lee, M. J. Hwang, M. L. Li, G. W. Chen, M. S. Ho, Mutations at KFRDI and VGK domains of enterovirus 71 3C protease affect its RNA binding and proteolytic activities. *J. Biomed. Sci.* **11**, 239–248 (2004).
40. O. Ohlenschlager, J. Wohnert, E. Bucci, S. Seitz, S. Hafner, R. Ramachandran, R. Zell, M. Gorlach, The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure* **12**, 237–248 (2004).
41. Y. Ihle, O. Ohlenschlager, S. Hafner, E. Duchardt, M. Zacharias, S. Seitz, R. Zell, R. Ramachandran, M. Gorlach, A novel cGUUAg tetraloop structure with a conserved yYNMGg-type backbone conformation from cloverleaf 1 of bovine enterovirus 1 RNA. *Nucleic Acids Res.* **33**, 2003–2011 (2005).
42. S. C. Mosimann, M. M. Cherney, S. Sia, S. Plotch, M. N. James, Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* **273**, 1032–1047 (1997).
43. H. Jiang, C. Lin, J. Chang, X. Zou, J. Zhang, J. Li, Crystal structures of the 3C proteases from Coxsackievirus B3 and B4. *Acta Crystallogr. F. Struct. Biol. Commun.* **80**, 183–190 (2024).
44. Y. Kim, S. Lovell, K. C. Tiew, S. R. Mandadapu, K. R. Alliston, K. P. Battaile, W. C. Groutas, K. O. Chang, Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J. Virol.* **86**, 11754–11762 (2012).
45. Y. Zhai, X. Zhao, Z. Cui, M. Wang, Y. Wang, L. Li, Q. Sun, X. Yang, D. Zeng, Y. Liu, Y. Sun, Z. Lou, L. Shang, Z. Yin, Cyanohydrin as an anchoring group for potent and selective inhibitors of enterovirus 71 3C protease. *J. Med. Chem.* **58**, 9414–9420 (2015).
46. L. Costenaro, Z. Kaczmarska, C. Arnan, R. Janowski, B. Coutard, M. Sola, A. E. Gorbalenya, H. Norder, B. Canard, M. Coll, Structural basis for antiviral inhibition of the main protease, 3C, from human enterovirus 93. *J. Virol.* **85**, 10764–10773 (2011).
47. G. Young, N. Hundt, D. Cole, A. Fineberg, J. Andrecka, A. Tyler, A. Olerinyova, A. Ansari, E. G. Marklund, M. P. Collier, S. A. Chandler, O. Tkachenko, J. Allen, M. Crispin, N. Billington, Y. Takagi, J. R. Sellers, C. Eichmann, P. Selenko, L. Frey, R. Riek, M. R. Galpin, W. B. Struwe, J. L. P. Benesch, P. Kukura, Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).
48. R. Andino, G. E. Rieckhof, P. L. Achacoso, D. Baltimore, Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA. *EMBO J.* **12**, 3587–3598 (1993).
49. K. S. Harris, W. Xiang, L. Alexander, W. S. Lane, A. V. Paul, E. Wimmer, Interaction of poliovirus polypeptide 3CDpro with the 5′ and 3′ termini of the poliovirus genome. Identification of viral and cellular cofactors needed for efficient binding. *J. Biol. Chem.* **269**, 27004–27014 (1994).

50. T. B. Parsley, C. T. Cornell, B. L. Semler, Modulation of the RNA binding and protein processing activities of poliovirus polypeptide 3CD by the viral RNA polymerase domain. *J. Biol. Chem.* **274**, 12867–12876 (1999).
51. S. E. Diamond, K. Kirkegaard, Clustered charged-to-alanine mutagenesis of poliovirus RNA-dependent RNA polymerase yields multiple temperature-sensitive mutants defective in RNA synthesis. *J. Virol.* **68**, 863–876 (1994).
52. R. Zell, K. Sidigi, E. Bucci, A. Stelzner, M. Gorlach, Determinants of the recognition of enteroviral cloverleaf RNA by coxsackievirus B3 proteinase 3C. *RNA* **8**, 188–201 (2002).
53. A. Marin-Gonzalez, C. Aicart-Ramos, M. Marin-Baquero, A. Martin-Gonzalez, M. Suomalainen, A. Kannan, J. G. Vilhena, U. F. Greber, F. Moreno-Herrero, R. Perez, Double-stranded RNA bending by AU-tract sequences. *Nucleic Acids Res.* **48**, 12917–12928 (2020).
54. V. Thiviyanathan, Y. Yang, K. Kaluarachchi, R. Rijnbrand, D. G. Gorenstein, S. M. Lemon, High-resolution structure of a picornaviral internal cis-acting RNA replication element (cre). *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12688–12693 (2004).
55. A. V. Paul, J. Yin, J. Mugavero, E. Rieder, Y. Liu, E. Wimmer, A "slide-back" mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. *J. Biol. Chem.* **278**, 43951–43960 (2003).
56. A. V. Paul, E. Rieder, D. W. Kim, J. H. van Boom, E. Wimmer, Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *J. Virol.* **74**, 10359–10370 (2000).
57. J. Yin, A. V. Paul, E. Wimmer, E. Rieder, Functional dissection of a poliovirus cis-acting replication element [PV-cre(2C)]: Analysis of single- and dual-cre viral genomes and proteins that bind specifically to PV-cre RNA. *J. Virol.* **77**, 5152–5166 (2003).
58. Y. Yang, R. Rijnbrand, S. Watowich, S. M. Lemon, Genetic evidence for an interaction between a picornaviral cis-acting RNA replication element and 3CD protein. *J. Biol. Chem.* **279**, 12659–12667 (2004).
59. H. B. Pathak, S. K. Ghosh, A. W. Roberts, S. D. Sharma, J. D. Yoder, J. J. Arnold, D. W. Gohara, D. J. Barton, A. V. Paul, C. E. Cameron, Structure-function relationships of the RNA-dependent RNA polymerase from poliovirus (3Dpol). A surface of the primary oligomerization domain functions in capsid precursor processing and VPg uridylylation. *J. Biol. Chem.* **277**, 31551–31562 (2002).
60. M. Shen, Q. Wang, Y. Yang, H. B. Pathak, J. J. Arnold, C. Castro, S. M. Lemon, C. E. Cameron, Human rhinovirus type 14 gain-of-function mutants for oril utilization define residues of 3C(D) and 3Dpol that contribute to assembly and stability of the picornavirus VPg uridylylation complex. *J. Virol.* **81**, 12485–12495 (2007).
61. Y. Yang, R. Rijnbrand, K. L. McKnight, E. Wimmer, A. Paul, A. Martin, S. M. Lemon, Sequence requirements for viral RNA replication and VPg uridylylation directed by the internal cis-acting replication element (cre) of human rhinovirus type 14. *J. Virol.* **76**, 7485–7494 (2002).
62. M. Shen, Z. J. Reitman, Y. Zhao, I. Moustafa, Q. Wang, J. J. Arnold, H. B. Pathak, C. E. Cameron, Picornavirus genome replication. Identification of the surface of the poliovirus (PV) 3C dimer that interacts with PV 3Dpol during VPg uridylylation and construction of a structural model for the PV 3C2-3Dpol complex. *J. Biol. Chem.* **283**, 875–888 (2008).
63. C. Ferrer-Orta, M. Sierra, R. Agudo, I. de la Higuera, A. Arias, R. Perez-Luque, C. Escarmis, E. Domingo, N. Verdaguer, Structure of foot-and-mouth disease virus mutant polymerases with reduced sensitivity to ribavirin. *J. Virol.* **84**, 6188–6199 (2010).
64. E. Rieder, A. V. Paul, D. W. Kim, J. H. van Boom, E. Wimmer, Genetic and biochemical studies of poliovirus cis-acting replication element cre in relation to VPg uridylylation. *J. Virol.* **74**, 10371–10380 (2000).
65. J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).
66. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
67. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).
68. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
69. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

Acknowledgments: We thank J. Nysschen and G. Gonzalez-Gutierrez of Macromolecular Crystallography Facility at Indiana University and J. Nix at Advanced Light Source for help with x-ray data collection. We thank R. H. Simmons III and M. Bochman for help with MP experiments. **Funding:** Beamline 8.2.2 of the Advanced Light Source, a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231, is supported, in part by, the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. MP TwoMP Refeyn is acquired under the NIH grant 3R35GM133437-05S2 (to M. Bochman). This work was supported by NIH grants R01AI087856 and R21AI157336 (to K.H.C.), Indiana University Seed Funding program award (to K.G.), and the Indiana University John R. and Wendy L. Kindig Fellowship (to D.D.-S.). **Author contributions:** Sample preparation: D.D.-S. Data collection: D.D.-S., M.T.L., and K.G. Analysis: D.D.-S., M.T.L., K.G., and K.H.C. Supervision: K.G. and K.H.C. Writing—first draft: D.D.-S. and K.G. Writing—review and editing: D.D.-S., K.G., M.T.L., and K.H.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Atomic coordinates are available in the Protein Data Bank (PDB) under accession code 9DCF.

Submitted 6 September 2024
Accepted 4 February 2025
Published 12 March 2025
10.1126/sciadv.ads6862