

# A nucleobase-binding pocket in a viral RNA-dependent RNA polymerase contributes to elongation complex stability

Wei Shi<sup>1,2,†</sup>, Han-Qing Ye<sup>1,†</sup>, Cheng-Lin Deng<sup>1</sup>, Rui Li<sup>1,2</sup>, Bo Zhang<sup>1</sup> and Peng Gong<sup>1,3,\*</sup>

<sup>1</sup>Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, No.44 Xiao Hong Shan, Wuhan, Hubei 430071, China, <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China and <sup>3</sup>Drug Discovery Center for Infectious Diseases, Nankai University, Tianjin 300350, China

Received September 09, 2019; Revised November 30, 2019; Editorial Decision December 02, 2019; Accepted December 05, 2019

## ABSTRACT

The enterovirus 71 (EV71) 3D<sup>pol</sup> is an RNA-dependent RNA polymerase (RdRP) that plays the central role in the viral genome replication, and is an important target in antiviral studies. Here, we report a crystal structure of EV71 3D<sup>pol</sup> elongation complex (EC) at 1.8 Å resolution. The structure reveals that the 5'-end guanosine of the downstream RNA template interacts with a fingers domain pocket, with the base sandwiched by H44 and R277 side chains through hydrophobic stacking interactions, and these interactions are still maintained after one in-crystal translocation event induced by nucleotide incorporation, implying that the pocket could regulate the functional properties of the polymerase by interacting with RNA. When mutated, residue R277 showed an impact on virus proliferation in virological studies with residue H44 having a synergistic effect. *In vitro* biochemical data further suggest that mutations at these two sites affect RNA binding, EC stability, but not polymerase catalytic rate ( $k_{cat}$ ) and apparent NTP affinity ( $K_{M,NTP}$ ). We propose that, although rarely captured by crystallography, similar surface pocket interaction with nucleobase may commonly exist in nucleic acid motor enzymes to facilitate their processivity. Potential applications in antiviral drug and vaccine development are also discussed.

## INTRODUCTION

Human enterovirus 71 (EV71), a member of the *Enterovirus A* (EV\_A) species belonging to the *Enterovirus* genus of the *Picornaviridae* family, is one of the major causative agents of hand, foot and mouth disease (HFMD), and can some-

times cause severe neurological complications such as encephalitis, aseptic meningitis, and acute flaccid paralysis (1–3). The genome of EV71 is a single-stranded positive-sense RNA of approximately 7400 nucleotides (nt) encoding a 240-kD polyprotein. The polyprotein is co- and/or post-translationally processed by host or viral proteases into four structural proteins (VP1–VP4) and seven non-structural proteins (2A–2C and 3A–3D) (4). The 3D (also known as 3D<sup>pol</sup>) protein is an RNA-dependent RNA polymerase (RdRP) that plays the central role in viral RNA genome replication, and has become a major target for antiviral studies. To date, numerous crystal structures of picornavirus RdRP proteins have been solved. Along with RdRPs from other RNA viruses, the polymerase catalytic core adopts an encircled right hand architecture with palm, fingers and thumb domains surrounding the active site (5–7). The fingers domain can be further divided into index, middle, ring and pinky subdomains following a nomenclature first utilized in describing the poliovirus (PV, a member of the *Enterovirus C* species) RdRP (6). The index finger subdomain interacts with the thumb to make the hallmark encirclement distinguished from other classes of processive polymerases (5–6,8). A collection of viral RdRP structures in complex with RNA or with RNA/NTP further have illustrated the mechanism of the RdRP catalytic mechanisms, in particular, the mechanism of the polymerase nucleotide addition cycle (NAC) with unique features in the pre-catalysis active site closure and post-catalysis translocation events (9–13).

Compared to the relatively conserved catalytic mechanisms, the regulatory mechanisms of viral RdRPs are quite diverse. Other viral proteins, RNA elements, host factors or even regions as part of the RdRP molecule could affect the function of RdRP during the entire genome replication process. The key interactions with the RdRP catalytic core involved in these regulation processes can be classified into two major types: one from other molecules and the other

\*To whom correspondence should be addressed. Peng Gong. Tel: +86 27 87197578; Fax: +86 27 87197578; Email: gongpeng@wh.iov.cn

†These authors wish it to be known that, to their opinions, the first two authors should be regarded as Joint First Authors.

from region(s) that are part of the RdRP molecule but are beyond the RdRP catalytic core. Examples of the first type include the regulation of flavivirus NS5 by viral protein NS3 and viral RNA elements within the 5'-untranslated region (UTR) (14–16), picornavirus 3D<sup>pol</sup> by viral protein 3CD and host SUMOylation (SUMO stands for small ubiquitin-like modifier) machineries (17,18), and coronavirus nsp12 by nsp7 and nsp8 (19). Representative cases of the second type have been identified in the *Flaviviridae* RdRPs, with the methyltransferase (MTase, part of the flavivirus NS5) and the N-terminal domain (NTD, part of the pestivirus NS5B) regulating the corresponding polymerases, respectively, through intra-molecular interactions (20–25).

In this work, we identified and characterized a type of regulatory interaction that is quite unique if compared with the aforementioned interaction types. These interactions only involve a small surface pocket within the RdRP fingers domain and a nucleobase of the RNA that is part of RdRP–RNA catalytic complex. We unintentionally identified these interactions when solving an EV71 RdRP elongation complex (EC) crystal structure. Through a set of biochemical characterizations, we further demonstrated that this surface pocket facilitates EC stability, likely through its interactions with the nucleobase within the RdRP–RNA complex, while the elongation catalytic rates are not affected. Cell-based virological data comparing the wild type (WT) and mutant viruses further suggest that this pocket is important to virus proliferation. To the best of our knowledge, regulatory sites within the RdRP catalytic core that interact with the downstream RNA have not been reported with support of structural data. Our study provides an example of evolution and utilization of nucleobase anchoring sites by viral RdRPs to facilitate or maintain their essential catalytic properties, and the identified surface pocket may also serve as an intervention target for developing antiviral applications.

## MATERIALS AND METHODS

### Plasmid construction, protein expression and protein purification

The EV71 3D<sup>pol</sup> gene within the DNA clone of strains HeN09-17/HeN/CHN2009 (GenBank accession no. JX678881, genotype C) and BrCr-ts (GenBank accession no: AB204853.1, genotype A) was cloned into a pET26b-Ub vector (26,27). The resulting plasmids were transformed into *Escherichia coli* strain BL21(DE3)pCG1 (kindly provided by Dr Craig Cameron, Pennsylvania State University, State College, PA, USA) for the production of the full-length 3D<sup>pol</sup> protein with a native glycine at its N-terminus according to previously described methods and a C-terminal hexa-histidine tag for affinity purification (26,28). All 3D<sup>pol</sup> point mutations were introduced by using the QuickChange site-directed mutagenesis method and the corresponding WT plasmid as the template (29). Cell growth, isopropyl-β-D-thiogalactopyranoside (IPTG) induction, cell harvesting, cell lysis, protein purification and protein storage were performed as described previously (10,28), except that the temperature for overnight culture growth was 37°C. The final buffer condition for protein storage was 5 mM Tris (pH 7.5), 200 mM NaCl, 0.02% (wt./vol.) NaN<sub>3</sub>, and 5 mM Tris(2-carboxyethyl)phosphine

(TCEP). 3D<sup>pol</sup> concentrations were measured by absorbance at 280 nm using an extinction coefficient of 71 280 M<sup>-1</sup>cm<sup>-1</sup> calculated by the ExpASY ProtParam program (<http://web.expasy.org/protparam/>). The typical protein yield was 10–15 mg per liter of bacteria culture.

### RNA preparation and 3D<sup>pol</sup> EC assembly

The 31-mer template RNA (T31) was prepared by *in vitro* T7 RNA polymerase transcription and subsequent glmS ribozyme cleavage according to protocols described previously (30,31). T31 was subjected to a self-annealing process before being annealed to an 8-mer RNA (P8, Integrated DNA technologies) at a 1:1.1 molar ratio to yield the T31/P8 construct according to protocols described previously (31). The EV71 3D<sup>pol</sup> EC assembly, purification, and storage were carried out using protocols described previously (10), except that the assembly reaction was conducted in a buffer containing 50 mM HEPES (pH 7.0), 50 mM NaCl, 75 mM KCl, 8 mM MgCl<sub>2</sub> and 4 mM TCEP.

### EC crystallization, NTP soaking of the EC crystals and crystal harvesting

The EC crystals were grown by sitting-drop vapor diffusion at 16°C using a 12.5 mg/ml EC sample. Rhombohedron-shape crystals grew to their final size within 1–2 weeks in a precipitant solution containing 0.17 M sodium acetate, 0.085 M Tris (pH 8.5), 25.5% (wt./vol.) PEG4000 and 15% (vol./vol.) glycerol. Crystal soaking trials were done for 10 h using the precipitant solution supplemented with 5 mM ddCTP and 10 mM MgCl<sub>2</sub>. Crystals were directly cooled and stored in liquid nitrogen prior to data collection.

### Crystallographic data processing and structure determination

X-ray diffraction data was collected at Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U1 (wavelength: 0.9792 Å, temperature: 100 K). Data of at least 180° were typically collected in 1° oscillation steps. Reflections were integrated, merged and scaled using D\*Trek (32). The initial structure solution was obtained using the molecular replacement program PHASER (33) with coordinates derived from EV-B EC structure (PDB entry: 5F8G, chains A-C) as the search model (10). Manual model building and structure refinement were done using Coot and Phenix, respectively (34,35). The 3500-K composite simulated-annealing (SA) omit 2F<sub>o</sub>-F<sub>c</sub> electron density maps were generated using CNS (36). Unless otherwise indicated, protein structure superpositioning was done using the maximum likelihood-based structure superpositioning program THESEUS (37).

### Fluorescence polarization (FP)-based RNA binding assays

The binding affinity of the WT EV71 3D<sup>pol</sup> or its variant to the template/primer RNA was assessed using a fluorescence polarization (FP)-based assay. A T31/P8F (with a 1:1.1 molar ratio) construct was used as the RNA substrate. Except for bearing a 6-FAM (6-carboxyfluorescein)

at the 5'-end of P8, it is otherwise identical to the P31/P8 construct used in the EC assembly. For each 3D<sup>pol</sup> construct, a set of 14 mixtures were prepared, each with 25 nM T31/P8F construct (according to T31 concentration) and concentrations of 3D<sup>pol</sup> in the range of 3.2 nM to 5.6 μM in a FP buffer containing 50 mM potassium glutamate (pH 7.5), 50 mM arginine, 5 mM dithiothreitol (DTT), 5 mM MgCl<sub>2</sub>, 30 mM NaCl and 5% (vol./vol.) glycerol. The mixtures were loaded into the wells of a black 384-well plate (OptiPlate 384-F, PerkinElmer). After at least 60-min of incubation at room temperature (r.t.), the fluorescent signals were measured in the 2102 EnVision Multilabel reader (PerkinElmer) using an excitation wavelength of 485 nm (monochromator) and a 535-nm high band pass emission filter. The FP values were calculated based on the fluorescent signals parallel (F<sub>∥</sub>) and perpendicular (F<sub>⊥</sub>) to the excitation light and the equation  $FP = (F_{\parallel} - F_{\perp}) / (F_{\parallel} + F_{\perp})$ . The FP values at different 3D<sup>pol</sup> concentrations were fitted to the one-site binding model for determination of the dissociation constant ( $K_d$ ) values following the equation  $FP = \text{offset} + \text{amplitude} \cdot [3D^{\text{pol}}] / ([3D^{\text{pol}}] + K_d)$ . Five replicates were obtained for each 3D<sup>pol</sup> concentration. The FP values were decreased at high-concentration data points, possibly resulting from protein aggregation and therefore these data points were omitted in fitting routines.

### EC assembly assays

To characterize the process of EC formation for WT EV71 3D<sup>pol</sup> and its variants, the EC assembly reactions were carried out as described above with a total volume of 20 μl except that the 3D<sup>pol</sup> and RNA concentrations were 6 and 4 μM, respectively, the KCl concentration was 55 mM, and the NaCl concentration was 20 mM. At each indicated reaction time point, an aliquot of the reaction mixture was withdrawn and mixed with an equal volume of stop solution containing 95% (vol./vol.) formamide, 20 mM EDTA (pH 8.0), and 0.02% (wt./vol.) xylene cyano. The RNA species were analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) followed by Stains-All staining and band intensity quantification as previously described (31). For each 3D<sup>pol</sup> construct, pre-incubation of 3D<sup>pol</sup> and T31/P8 at 22.5°C prior to the addition of the NTP mixture for various time were performed in parallel to the experiment without pre-incubation.

### EC stability assays

The EC stability assay were modified from protocols described previously (38). EC assembly reactions were formed as described above for 120 min. NaCl was then added to reach a final concentration of 300 mM to prevent 3D<sup>pol</sup>-RNA rebinding and reinitiation, and the reaction mixture was incubated at 30°C. At various time points after the addition of NaCl, a 10 μl aliquot was withdrawn from the mixture and supplemented with CTP to reach a final CTP concentration of 300 μM. The CTP-triggered reaction proceeded for 3 min at 25°C. Reaction quenching, gel electrophoresis, staining and quantification were performed as described in the EC formation assays (31). Experiments were performed in triplicates for EV71-C 3D<sup>pol</sup> trials.

### Stopped-flow fluorescence assays for determining 3D<sup>pol</sup> elongation rate constants

Two types of stopped-flow fluorescence assays were established in measuring the EC elongation rates. The first assay was modified from protocols described previously (39,40). The 11-mer RNA primer (P11L) with a LI-COR label at the 5'-end and the 35-mer RNA template T35F<sub>term</sub> with a 5'-Fluorescein (Integrated DNA technologies) were annealed at a 1:1 molar ratio to yield the T35F<sub>term</sub>/P11L construct before being used in a stopped-flow assay for measuring the overall 3D<sup>pol</sup> elongation rates over a stretch of 17 nt. The LI-COR function was not used in this study due to the lack of an appropriate imaging instrument. By modifying the sequence at the 5' region of the RNA template from previously described methods (39), the fluorescent signal increase was only observed upon the incorporation of the last five nucleotides. Kinetics experiments were performed using a Chriscan SF3 instrument (Applied Photophysics) with equal volume mixing of the pre-assembled EC and elongation NTPs. Fluorescence excitation wavelength was 492 nm (monochromator, bandwidth 2 nm) and fluorescent signal was detected using a 515-nm high band pass filter. ECs were pre-assembled at 22.5°C for 90 min using 1 μM T35F<sub>term</sub>/P11L construct, 8 μM 3D<sup>pol</sup> and 100 μM each ATP and GTP in a SF buffer containing 50 mM HEPES (pH 7.0), 20 mM NaCl, 55 mM KCl, 5.5 mM MgCl<sub>2</sub> and 4 mM TCEP. The reaction mixture was then 20-fold diluted using the SF buffer to reach a final RNA concentration of 50 nM. The fluorescence signal change was measured at 22.5°C after rapid-mixing, producing a final solution with 25 nM RNA, 200 nM 3D<sup>pol</sup>, 150 μM ATP/GTP/CTP and 2.5–160 μM UTP in SF buffer.

Data analysis and fitting were done as described previously (39–41). The data of rapid signal increase observed at the end of the elongation reaction were fitted into a single exponential rise curve. The time of intersection point of this exponential curve with the lag phase signal was taken as a measure of time of elongating 17-nt single-stranded region, not including the five terminal nucleotides that do not contribute to the lag phase. The average elongation rate was calculated for all 17 elongation steps constituting the lag phase. This approximation treatment was demonstrated to reasonably estimate the average elongation rate when compared to more thorough data analysis to model the entire elongation reaction steps as a series of irreversible steps with an equal rate constant (40). The observed elongation rates were then plotted as a function of UTP concentration to determine the  $k_{\text{pol}}$  and apparent  $K_M$  values by fitting the data to the Michaelis–Menten type equation,  $\text{rate} = k_{\text{pol}} \cdot [\text{UTP}] / ([\text{UTP}] + K_M)$ .

The P8 primer and T31-F<sub>int</sub> template (Integrated DNA technologies) were annealed into duplex mimicking T31/P8 RNA duplex to determine the incorporation rates of a 2'-deoxy-CMP (2dCMP) directed by a templating guanosine 5 nt upstream of the fluorescein (Figure 7D). The stopped-flow routine was performed as described above. The data of rapid signal decrease observed for 2dCMP incorporation were fitted into a single exponential decay curve  $Y = \text{offset} + \text{amplitude} \cdot \exp(-\text{rate} \cdot t)$ . Due to deviation from the single exponential curvature, the latter portion of the signal



was omitted in the fitting routine (Figure 7D). The incorporation rates were then plotted as a function of the 2'-deoxy-CTP (2dCTP) concentration, and the data were fitted to the Michaelis-Menten type equation to determine the  $k_{\text{pol}}$  and apparent  $K_M$  values.

### Genome-length EV71 cDNA mutant construction and the immunofluorescence assay (IFA)

EV71 genome-length cDNA clones with different mutations were constructed using the infectious cDNA clone of pACYC-EV71-FL as the backbone. Fusion PCR was used to engineer all mutations into the backbone (27). All constructs were validated by DNA sequencing. RNA transcription, transfection and subsequent IFA were carried out as previously described (27,42)

## RESULTS

### The global structure of the EV71 RdRP EC resembles those of the previously reported enterovirus RdRP ECs except for the placement of the downstream RNA

Based on established methods in obtaining picornaviral RdRP EC crystals (43), we assembled a genotype-C EV71 RdRP EC using an RNA construct T31/P8 comprising a 31-nt template (T31) and an 8-nt primer (P8) to direct the incorporation of a GAGA tetra-nucleotide by providing GTP and ATP as the only NTP substrates (Figure 1A), purified the EC by anion-exchange chromatography, and crystallized the EC. The crystal structure was solved at 1.8 Å resolution in space group C2 by molecular replacement using previously reported genotype-B EV71 RdRP EC structure as the search model (Table 1) (10). The final model had  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 0.181 and 0.214, respectively, and the crystallographic asymmetric unit contained only one EC. The RdRP in this EC is structurally consistent with RdRPs in the previously reported enterovirus EC structures (10,43–45) with the root-mean-square deviation (RMSD) values in the range of 0.5–1.1 Å for all superimposable  $\alpha$ -carbon atoms (the new structure as the reference; 99–100% coverage). The majority of the RNA is well ordered including all downstream nucleotides that form a stem-loop structure (Figure 1A and B). Based on available structural information, the downstream RNA of the picornaviral RdRP EC is flexible and does not have a consistent conformation (Figure 1A, bottom right) (10,43–45). The crystal contacts could play important roles in stabilizing the RNA in certain conformations and resulting in the high-resolution feature of this EC crystal form. The high-resolution nature of this structure allows identification of a total of five ordered magnesium ions associated with this EC, interacting with the upstream (two ions) and downstream (two ions) RNA and residues of the catalytic motifs A and C in the palm domain (one ion).

### Structural identification of interactions between a fingers domain pocket and the 5'-end guanosine of the RNA template

Interestingly, the guanine moiety of the 5'-end guanosine of the downstream RNA was observed to interact with a fingers domain pocket, with its guanine plane sandwiched by

**Table 1.** X-ray diffraction data collection and structure refinement statistics

| EC form—PDB  | Native—6KWQ           | ddCTP—6KWR            |
|--|-----------------------|-----------------------|
| <b>Data collection<sup>1</sup></b>                 |                       |                       |
| Space group  | C2                    | C2                    |
| Cell dimensions                                    |                       |                       |
| $a, b, c$ (Å)                                      | 124.3, 75.5, 74.0     | 130.0, 76.5, 68.1     |
| $\alpha, \beta, \gamma$ (°)                        | 90, 91.5, 90          | 90, 94.9, 90          |
| Resolution (Å) <sup>2</sup>                        | 47.0–1.76 (1.82–1.76) | 38.2–2.50 (2.59–2.50) |
| $R_{\text{merge}}$                                 | 0.077 (0.43)          | 0.075 (0.46)          |
| $R_{\text{meas}}$                                  | 0.090 (0.50)          | 0.092 (0.57)          |
| $I/\sigma I$                                       | 7.8 (2.0)             | 7.6 (1.6)             |
| Completeness (%)                                   | 99.8 (99.9)           | 98.9 (99.6)           |
| Redundancy   | 3.8 (3.7)             | 2.9 (2.9)             |
| <b>Refinement</b>                                  |                       |                       |
| Resolution (Å)                                     | 1.76                  | 2.50                  |
| No. reflections                                    | 67,730                | 22,873                |
| $R_{\text{work}}/R_{\text{free}}$ <sup>3</sup> (%) | 18.1/21.4             | 21.6/25.7             |
| No. atoms  |                       |                       |
| Protein/RNA  | 3,695/906             | 3,600/755             |
| Ligand / Ion /                                     | 13/6/648              | 27/2/105              |
| Water  |                       |                       |
| B-factors (Å <sup>2</sup> )                        |                       |                       |
| Protein  | 27.7                  | 60.5                  |
| Ligand / Ion /                                     | 38.7/28.5/37.2        | 68.5/74.9/58.4        |
| Water  |                       |                       |
| R.m.s. deviations                                  |                       |                       |
| Bond lengths (Å)                                   | 0.006                 | 0.008                 |
| Bond angles (°)                                    | 0.83                  | 1.05                  |
| Ramachandran stat. <sup>4</sup>                    | 95.3/4.4/0.0/0.2      | 92.2/7.6/0.0/0.3      |

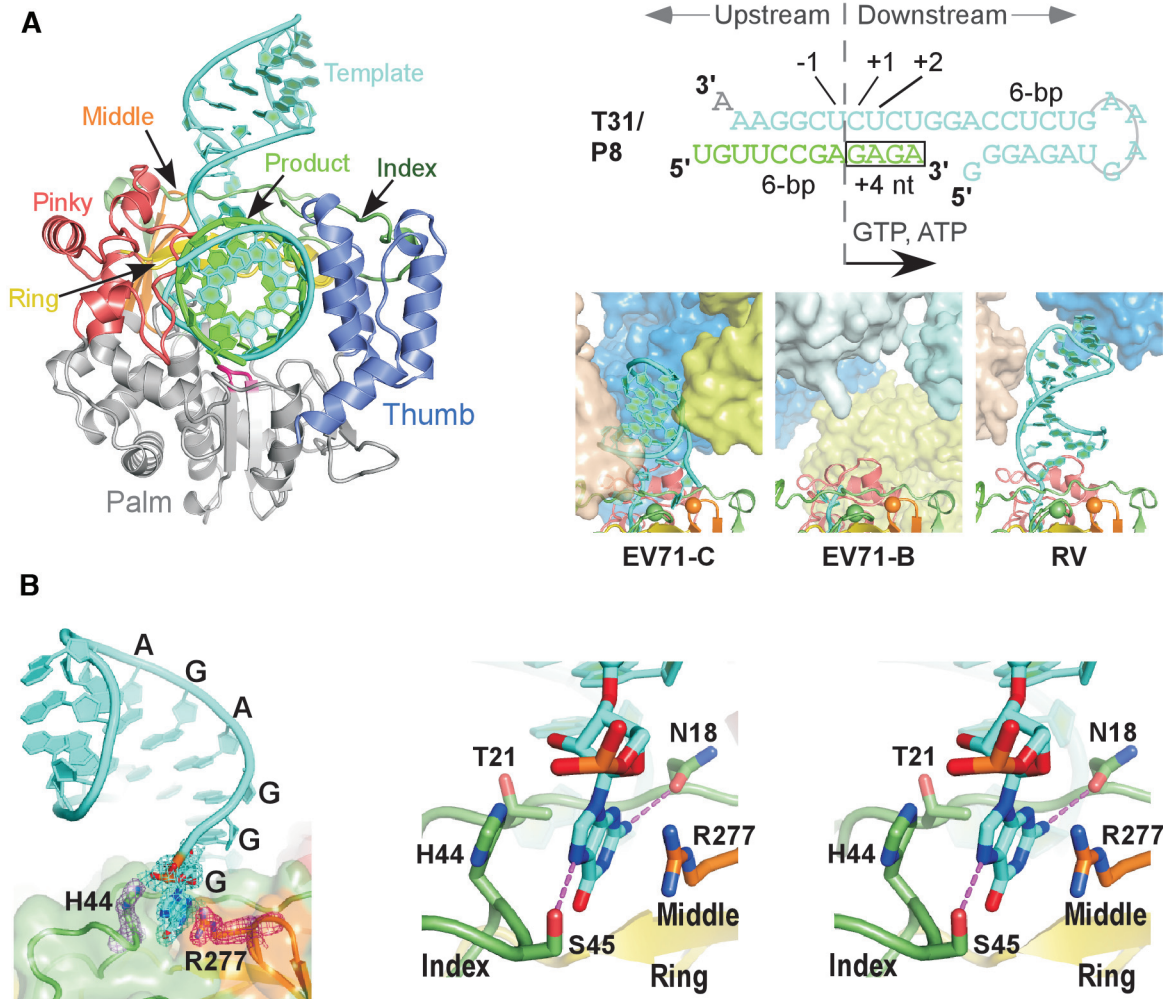
<sup>1</sup>One crystal was used for data collection for each structure.

<sup>2</sup>Values in parentheses are for highest-resolution shell.

<sup>3</sup>5% of data are taken for the  $R_{\text{free}}$  set.

<sup>4</sup>Values are in percentage and are for most favored, additionally allowed, generously allowed, and disallowed regions in Ramachandran plots, respectively.

the imidazole moiety of the index finger residue H44 and the guanidinium moiety of the middle finger residue R277 likely through stacking interactions (average distances to the guanine plane: 3.4 Å and 3.6 Å for guanidinium and imidazole moieties, respectively; angles between neighboring planes: 10–15°) (Figure 1B). Hydrogen bonding and other hydrophobic interactions involving index finger residues N18, T21 and S45 further stabilize the guanine base, with side chains of residue S45 interacting with the guanine-specific N2 position and residue N18 interacting with the purine-specific N7 position (Figure 1B, right). Although the crystal contacts between the downstream RNA and the neighboring ECs could contribute to the establishment and maintenance of these interactions, the delicate interaction network nevertheless indicates potential functional roles of this fingers domain pocket. This pocket comprises the tip of the middle finger, residues 18–21 and a unique 'kink' region (residues 44–55) of the index finger. Structurally, the kink region is highly conserved in RdRPs from the *Picornaviridae*, conserved to some extent in RdRPs from the *Caliciviridae*, but has not been found in RdRPs from other viral families in the positive-strand RNA viruses including the *Flaviviridae*, *Permutotetraviridae* and *Leviviridae* (Figure 2A). Sequence conservation analysis of representative picornavirus RdRPs indicates that residues corresponding to EV71 RdRP H44 and R277 are not well conserved, but the majority of residues at these two positions are able to



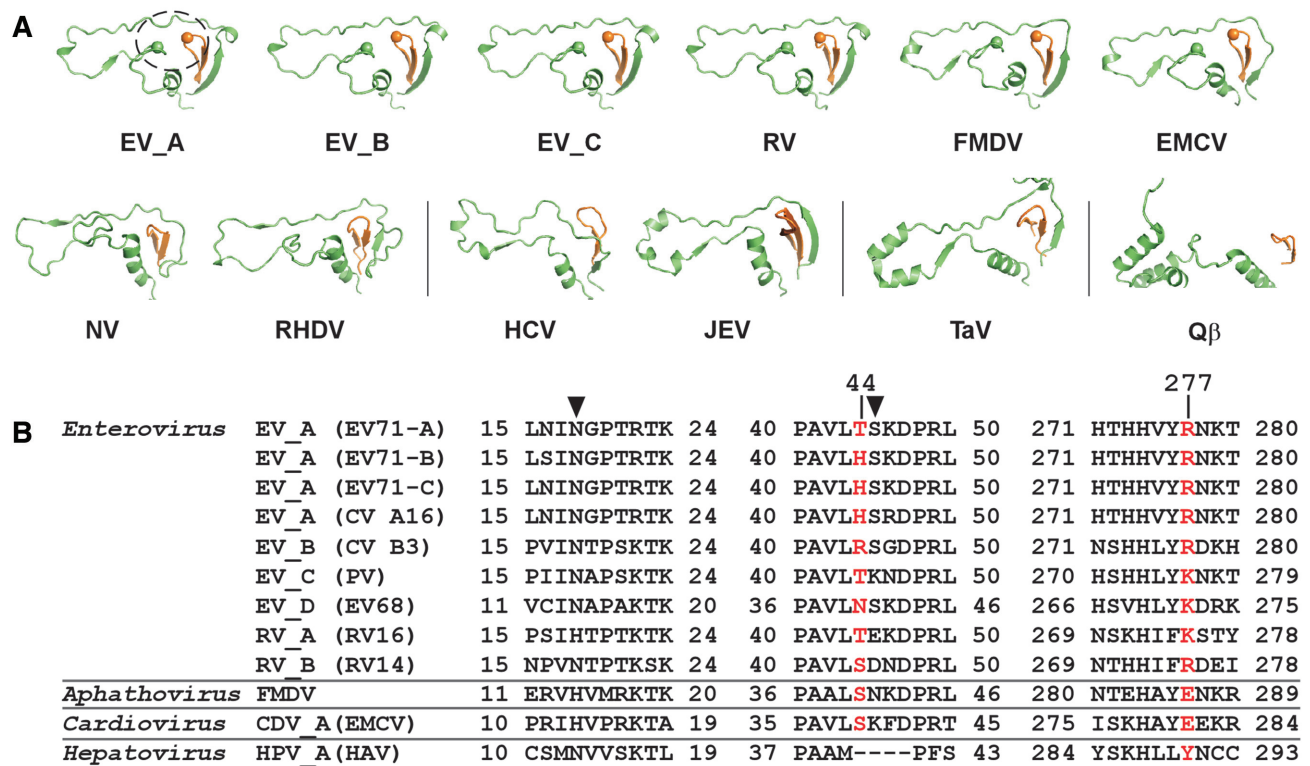
**Figure 1.** A nucleobase-binding pocket identified in an EV71 RdRP EC crystal structure. (A) EC structure (left, viewed into the front channel), RNA construct for EC assembly (top right) and a comparison of downstream RNA in enterovirus RdRP EC structures (bottom right). RNA strands and structural regions of the RdRP are color-coded as indicated. The signature sequence YGDD is in magenta. The boxed region in the RNA sequence indicates the tetra-nucleotide incorporated into the product strand during EC assembly. Neighboring complexes (in different colors) around the downstream RNA of the EC in each crystal lattice were shown in surface representation. The  $\alpha$ -carbon of the two key residues participating in the stacking interactions with the nucleobase are shown as spheres. PDB entries used: EV71-C–6KWQ; EV71-B–5F8G; RV (rhinovirus)–4K50. (B) The interactions between the pocket and the 5'-nucleobase. The coloring scheme is the same as in panel A. Left: a global view of the interactions with composite SA omit electron density maps (contoured at  $1.0 \sigma$ ) of the terminal guanosine and residues H44 and R277 overlaid. Right: stereo-pair images of the interaction details with key side chains and the 5'-guanosine shown in sticks. Dashed lines indicate hydrogen bonding interactions between the guanine and side chains of residues N18 and S45.

provide hydrophobic interactions through their side chains (Figure 2B). We therefore hypothesized that this fingers domain pocket involving the structurally unique kink region may have been utilized by some picornaviruses to optimize their RdRP function through its capability of interacting with RNA bases.

#### One-register in-crystal translocation does not disrupt the interactions between the 5'-end guanine and the fingers domain pocket

Some picornavirus RdRP EC crystal lattices allow in-crystal catalysis and translocation (10,43). As the first step to assess the interactions between the 5'-end guanine and the fingers domain pocket, we performed NTP soaking experiments to test whether progressive synthesis can 'pull' the

5'-end guanine out of the pocket. Based on the template sequence, CTP is the desired substrate in the next two NACs. However, the EC crystals soaked in CTP-containing solutions diffracted only poorly and did not yield useful structural data. With 2',3'-dideoxy CTP (ddCTP) used in the soaking trials, we obtained a 2.5-Å resolution dataset (Table 1). The resulting structure was solved still in the C2 space group, albeit with apparent changes in the unit cell parameters. Compared with the original C2 lattice, the unit cell parameters  $a$ ,  $c$ , and  $\beta$  are 5.7 Å (4.6%) longer, 5.9 Å (8.0%) shorter, and 3.4 Å wider, respectively. These observations indicate that the crystal lattice had adjusted itself to accommodate the ddCTP-induced polymerization reactions, while the CTP-soaking was likely too destructive to the lattice. In the ddCTP-derived structure, one ddCMP was incorporated into the product RNA chain and one translocation



**Figure 2.** Structure and sequence conservation analyses of the nucleobase-binding pocket. Virus abbreviations: FMDV–foot-and-mouth disease virus; EMCV–encephalomyocarditis virus; NV–norovirus; RHDV–rabbit hemorrhagic fever virus; HCV–hepatitis C virus; JEV–Japanese encephalitis virus; TaV–*Thosea asigna* virus; Qβ–bacteriophage Qβ; HAV–hepatitis A virus. (A) Side by side comparison of the index/middle (in green/orange) fingers subdomains of RdRPs from different positive-strand RNA viruses derived from a superposition (least square methods) of the highly conserved RdRP motif C residues. The dashed circle in the EV\_A structure indicates the location of the nucleobase-binding pocket. PDB entries used: EV\_A–6KWQ, EV\_B–3N6L, EV\_C–1RA6, RV–1TP7, FMDV–1U09 and EMCV–4NYZ for *Picornaviridae* (top row); NV–1SH0 and RHDV–1KHV for *Caliciviridae*; HCV–1C2P and JEV–4K6M for *Flaviviridae*; TaV–4XHI for *Permutotetraviridae*; Qβ–3MMP for *Leviviridae*. (B) Sequence alignment of three regions constituting the pocket (first two are part of the index finger and last one is part of the middle finger) for RdRPs from the *Enterovirus* genus and three representative RdRPs from other genera of the *Picornaviridae*. Residues equivalent to the nucleobase stacking H44 and R277 in EV\_A RdRP are shown in red, while those equivalent to the hydrogen bonding N18 and S45 are indicated by solid triangles.

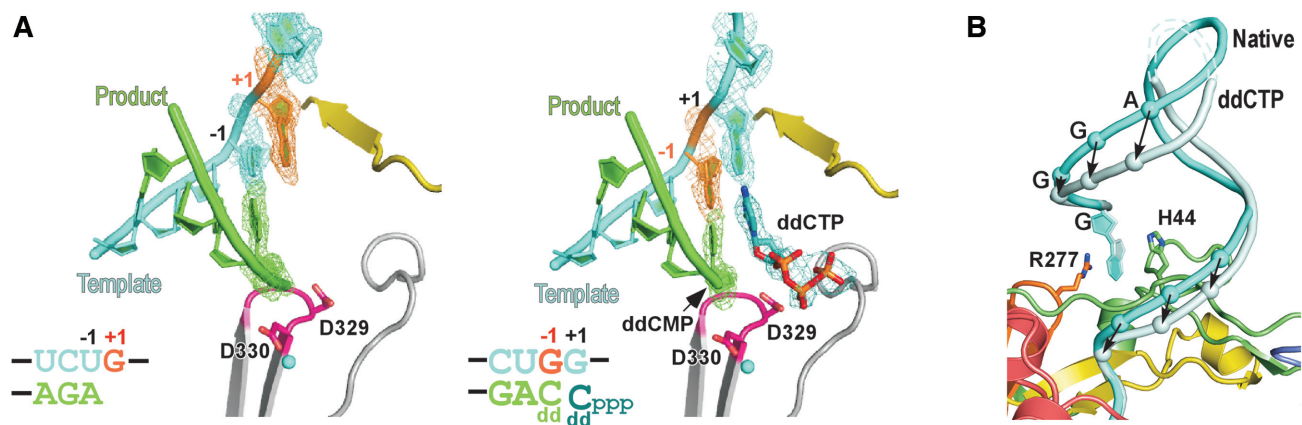
event had occurred, while the second ddCTP molecule was bound in the active site (Figure 3A, left). Consequently, the downstream RNA stem-loop underwent a small-scale twist as one nucleotide was pulled into the active site (Figure 3B and Supplementary Video S1). However, the 5'-end guanine was still buried in the pocket with most of the interactions maintained (Figure 3A, right). The incorporation of the ddCMP into the product chain prevented further catalysis and translocation due to the absence of a 3'-hydroxyl group, and therefore the disruption of the 5'-end guanine interactions with the fingers domain pocket may not be achieved through this approach. Nevertheless, the observation of the maintenance of these interactions and the twisting of the downstream RNA hairpin upon one translocation event together argue this type of interaction may be strong enough to serve as a regulatory factor of the polymerase.

#### RdRP residue R277 is important to EV71 proliferation, while H44 may play an auxiliary role

To assess the biological relevance of H44 and R277 in viral replication, we engineered a set of point mutations at

these two sites into an infectious cDNA clone of EV71. Besides alanine mutations, a threonine mutation at residue 44 was also tested because threonine is the other naturally occurring amino acid at this position in EV71 (Figure 2B). After *in vitro* transcription, equal amount of the WT and mutant viral RNAs were transfected into Vero cells. Viral protein expression was compared between the WT and mutant viruses. An immunofluorescence assay (IFA) was used to monitor virus replication by detecting expression of the viral VP1 protein. As shown in Figure 4A, all mutants except for the H44A-R277A showed IFA-positive signal at 24 h post-transfection (hpt), and the number of IFA-positive cells increased over time. There were comparable fractions of IFA-positive cells upon transfection with the RNAs of the WT, H44A and H44T (about 20, 50 and 80% positive cells at 24, 36, 48 hpt, respectively). In contrast, the R277A and H44T-R277A produced much fewer IFA-positive cells than the WT (about 1, 10 and 20% positive cells at 24, 36, 48 hpt, respectively), and no IFA-positive cells were observed at 72 hpt for H44A-R277A. These data collectively indicate that 3D<sup>pol</sup> residue R277 is critical for the EV71 viral replication and viability, while residue H44 may play a synergistic role with R277.





**Figure 3.** A comparison of the native and ddCTP-derived EC structures. (A) The EC structures are shown in cartoon representations with composite SA omit electron density map (contoured at  $1.2\sigma$ ) overlaid for the native (left) and ddCTP-derived ECs. For clarity, the majority of the polymerase and RNA up- and downstream of the active site are not shown. Except for the templating +1 nt in the native complex and the -1 nt in the ddCTP complex are shown in orange to correlate the nucleic acids translocation, the coloring scheme is as in Figure 1. Magnesium ions are shown as cyan spheres. (B) A comparison of the downstream RNA of the two EC structures. The arrows indicate the movement of the backbone phosphates of equivalent nucleotides from the native (cyan) conformation to the ddCTP (light cyan) conformation. Dashed lines indicate the likely path of the disordered RNA in the ddCTP complex. RdRP structure is only shown for the native complex for clarity.

#### RdRP residue R277 showed an apparent effect on RNA binding ability, and H44 also contributed

To further dissect the molecular mechanism of how residues R277 and H44 contributed to virus proliferation, in particular related to the polymerase properties, we performed a series of biochemical characterizations comparing the WT and corresponding mutant RdRP proteins. RNA binding affinity was measured using an FP assay derived from a previous study (46). Various concentrations of WT or mutant RdRPs were incubated with the T31/P8 RNA construct bearing a fluorescein label at the 5'-end of the P8. Upon RdRP binding, the tumbling freedom of the fluorescent probe was reduced, resulting in an increase of the FP signal. The data were fitted to the one-site binding model to estimate the dissociation constant ( $K_d$ ). The results showed that WT, H44A, R277A and H44A-R277A RdRP had  $K_d$  values of 0.18, 0.17, 0.40 and 0.60  $\mu\text{M}$ , respectively (Figure 4B). Alanine mutation of H44 alone does not weaken the RNA binding ability of RdRP, but could display a synergistic effect for the reduced binding ability of R277A. The variations in RNA binding affinities for WT and mutants are consistent with the data of virus proliferation. Note that the FP signals gradually decreased when polymerase concentrations were greater than 1  $\mu\text{M}$ , possibly due to protein aggregation under relatively high concentrations (Figure 4B). We therefore omitted the data affected by this effect (circled dots in Figure 4B) in the fitting routine.

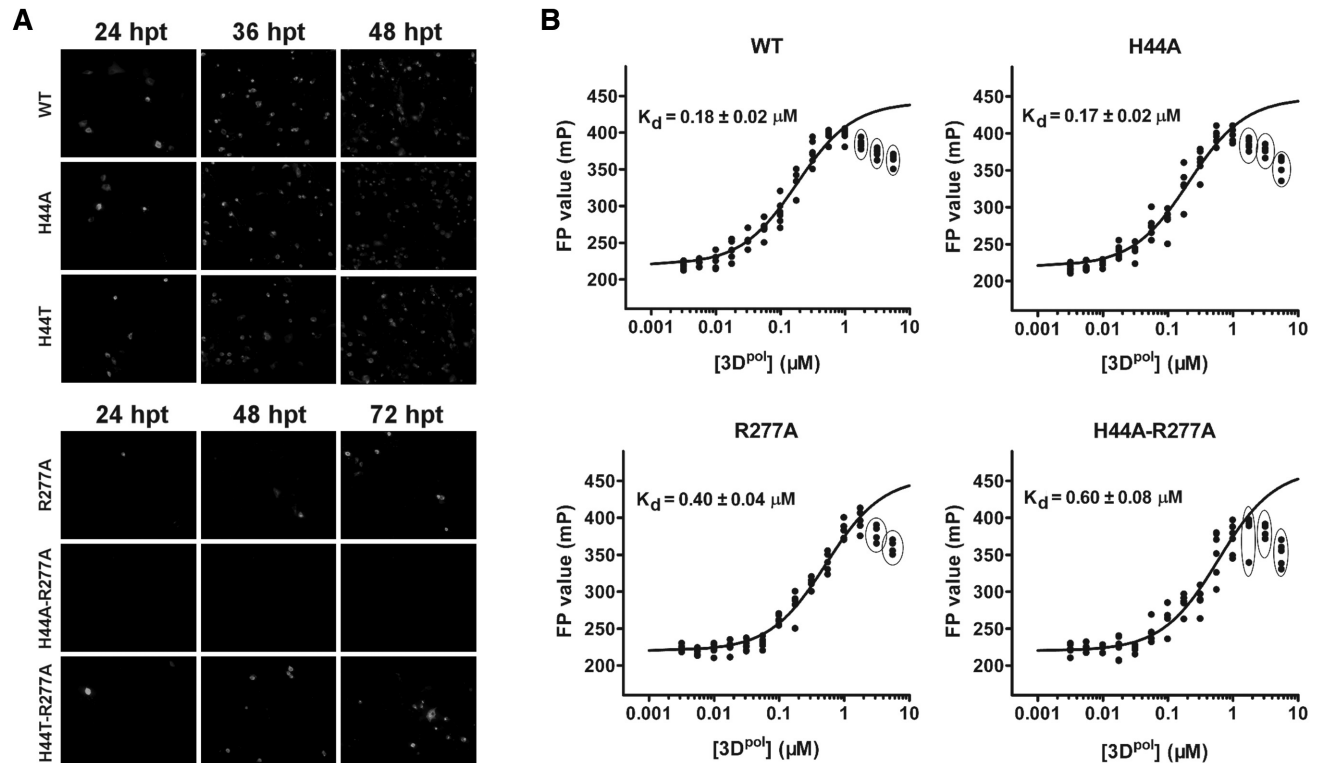
#### R277A-containing mutants required longer time to reach equilibrium of forming binary complex with RNA

The effects of RdRP residues H44 and R277 on EC formation were examined by monitoring the 12-mer production from T31/P8. In the presence of ATP and GTP, T31/P8 directed incorporation of a tetra-nucleotide GAGA to yield a 12-mer product. This is a multi-step process including RdRP-RNA binding and four rounds of nucleotide in-

corporation. If the process was initiated without a pre-incubation of RdRP and T31/P8, the 12-mer production exhibited a single-exponential behavior with the R277A and H44A-R277A mutants having slower production rates than the WT and H44A mutant (Figure 5A and B). We further conducted the RdRP and T31/P8 pre-incubation time test to determine the pre-incubation time required for each RdRP protein to reach an equilibrium while binding to the RNA construct, and applied these pre-incubation time periods in subsequent EC formation trials. After full pre-incubation, the 12-mer production was much faster than that without pre-incubation and the initial burst rates for all RdRP proteins were too fast to be detected in manual mixing experiments (Figure 5, compare panels B and C). These observations first indicate that the RdRP-RNA binding is the rate-limiting step in the conversion process. Secondly, the difference in the 12-mer production rate observed in the reactions without pre-incubation suggests that the R277A and H44A-R277A mutants likely have slower association rates with the RNA than the WT and H44A mutant have. The differences in the 12-mer amount after the initial burst of synthesis likely reflect different percentages of the RdRP-RNA binary complex formed after pre-incubation, and are consistent with the differences in  $K_d$  value determined in the FP assay (Figure 4B). Together with the FP data, these results suggest that the R277A-containing mutations affect both the RNA binding affinity and kinetics of the EV71 RdRP. We note that the usage of the Stains-All staining method in the quantitation analysis was shown as a valid approach at least for semi-quantitative analyses (21,31).

#### Both H44 and R277 residues contributed to RdRP EC stability

Since the interactions between the 5'-end guanine and the fingers domain pocket were directly observed in the EV71 RdRP EC structure, an immediate question to ask



**Figure 4.** Mutations at the H44 and R277 sites affect virus proliferation and RdRP–RNA binding. (A) IFA analysis of genome-length EV71 viral RNA replication for WT and mutant viruses in transfected Vero cells at the indicated time points. (B) Binding of the T31/P8F RNA to the EV71.C 3D<sup>pol</sup> constructs as detected by FP signal change. The dissociation constant ( $K_d$ ) derived from the data fitting following the one-site binding model was overlaid. Data points not included in the fitting routine were indicated by solid circles.

is whether important properties of the EC are affected by these interactions. We first carried out a test of EC stability for WT and mutant RdRPs. The results showed that all ECs but the WT EC had gradually dissociated over time, as reflected by increasing fractions of non-extendable 12-mer products when treated with high-salt condition (300 mM NaCl) at 30°C over time (Figure 6A and B). WT RdRP formed the most stable EC, having nearly undetectable dissociation even after a 96-h treatment. The H44A and R277A mutants formed EC with moderately reduced stability, with about 28 and 25% EC dissociated after 96 h, respectively, while the H44A-R277A forms EC with the lowest stability, with 84% of EC dissociated at 96 h. By fitting the H44A-R277A data to a single exponential decay model, the dissociation rate constant ( $k_{off}$ ) of EC was estimated as  $0.031 \text{ h}^{-1}$ , corresponding to an EC half life ( $t_{1/2}$ ) of about 22.4 h (Figure 6B). These data suggest that both H44 and R277 contributed to EC stability. Note that the stability of our EC is much higher than that of a PV RdRP EC obtained after incorporating two nucleotides (with a  $T_{1/2}$  of 3 h for WT PV RdRP EC) (38).

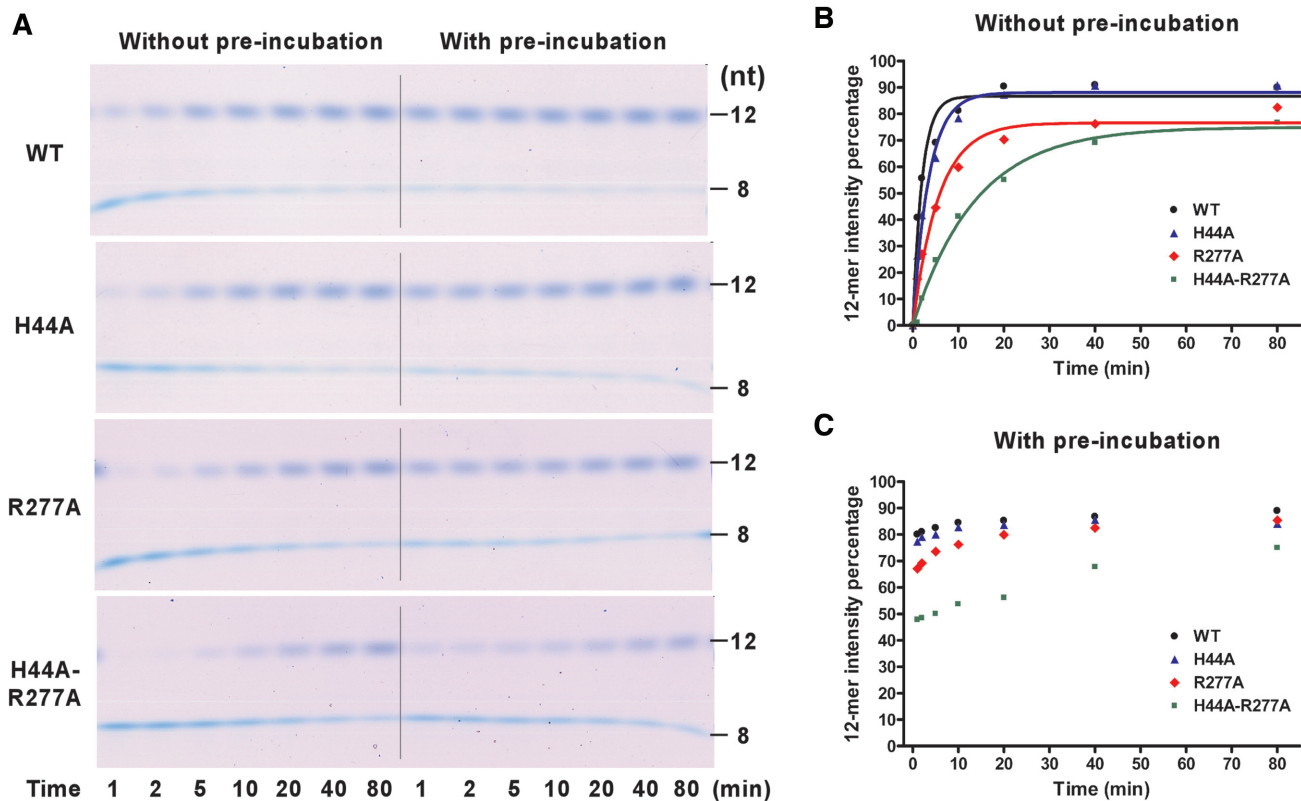
While the RdRP from the genotype-C EV71 (EV71-C) has a histidine at residue 44, those from genotype-A EV71 (EV71-A) and CV A16 have a threonine at this position (45). We introduced alanine or histidine mutation at residue 44 of RdRPs from these two viruses and assessed the EC stability along with corresponding WT proteins. The results showed that, for both RdRPs, both the WT and T44H

could form highly stable ECs, while the EC derived from the T44A mutant showed moderately reduced stabilities (Figure 6C and D). Together with the virological data, these results suggest that both histidine and threonine at residue 44 could fulfill the function in stabilizing EC through the structurally observed interactions, with no significant differences between these two residue types.

#### RdRP mutations at residues 44 and 277 do not apparently affect polymerase elongation rates and NTP substrate affinity

We next assessed the elongation rate, another important property of a polymerase EC, for the WT and mutant RdRPs from EV71-C. Using fluorescence-based stopped-flow methods derived from previous kinetics studies characterizing enterovirus RdRPs (39,40), a T33F<sub>term</sub>/P11L (see ‘Materials and Methods’ section) RNA construct was used to measure the average elongation rates on a 17-nt stretch of the template (Figure 7A). The template T33F<sub>term</sub> formed an eight base pair (bp) duplex with the P11L primer, and the EC was assembled after incorporating a GAGA tetranucleotide sequence by providing GTP and ATP as the only NTP substrates. The EC was then mixed in the stopped-flow instrument with all four NTPs, and the fluorescence signal only started to increase when the polymerase reached the fifth nucleotide to the 5'-end of the template (40) (Figure 7A and B). Therefore, the time of the fluorescence signal lag phase can be used to estimate the overall time for



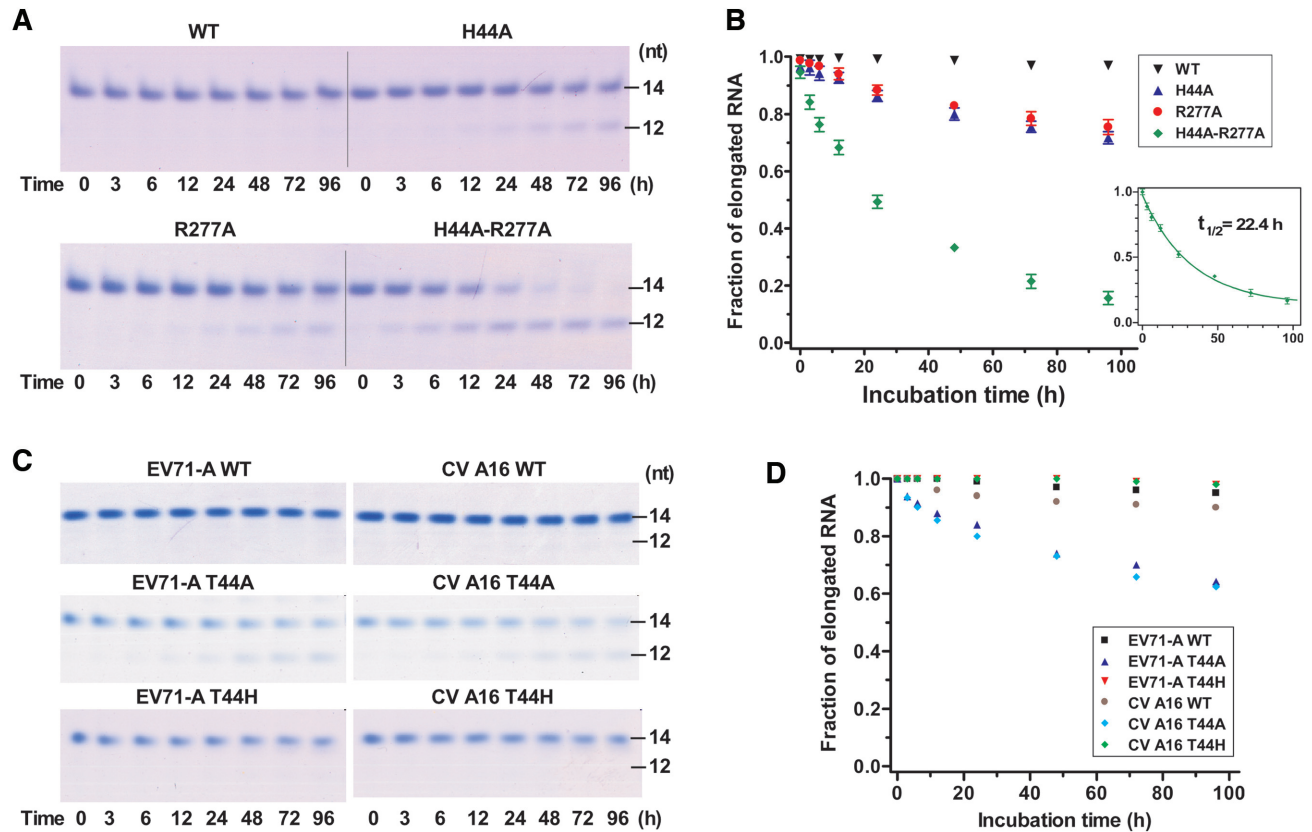


**Figure 5.** A comparison of EC formation of EV71\_C RdRP constructs with or without protein–RNA pre-incubation. (A) Monitoring the 12-mer (P12) containing EC formation in a time course manner for four RdRP constructs with (left) or without (right) pre-incubation. (B and C) Quantitation analyses of EC formation with (B) or without (C) pre-incubation indicated by the 12-mer intensity percentage values ( $P_{12_{int}}/(P_{8_{int}}+P_{12_{int}})$ ). For each RdRP construct, the data in panel B were fitted to a single exponential rise model to aid the comparison.

the polymerase EC to synthesize a total of 17 nt. By varying UTP concentrations in the NTP mixture, we were able to measure the average elongation rate constant ( $k_{pol}$ ) and the Michaelis constant ( $K_M$ ) for UTP (Figure 7B and C). The results showed that the WT and mutant RdRPs exhibited comparable  $k_{pol}$  (5.5–6.3 nt/s) and  $K_M$  (10.0–15.4  $\mu$ M) values (Figure 7C). While these data suggest that the H44 and R277 may not affect the overall elongation rate in multiple rounds of nucleotide addition, we further estimated the enzyme parameters in a single nucleotide addition assay starting with a situation mimicking the EC formed by the T31/P8 in our structural study and ending with incorporation of a 2dCMP using a T31- $F_{int}$ /P8 construct (Figure 7D). By placing an internal fluorescein-labeled thymidine at the fifth position downstream of the incorporation site, the incorporation induces a decrease of fluorescence signal that can be fitted to a single exponential decay model. The results showed that the WT and mutant RdRPs also have comparable  $k_{pol}$  (3.9–5.4  $s^{-1}$ ) and  $K_M$  (99.5–132.4  $\mu$ M) values for single nucleotide addition (Figure 7E and F). While the  $k_{pol}$  values are comparable to those obtained in the multi-round addition reactions, the  $K_M$  values for 2dCTP is about 10-fold of those for UTP. Taken together, the stopped-flow kinetics data demonstrate that mutations at EV71 RdRP residues 44 and 277 do not significantly influence the elongation rate and NTP substrate affinity.

## DISCUSSION

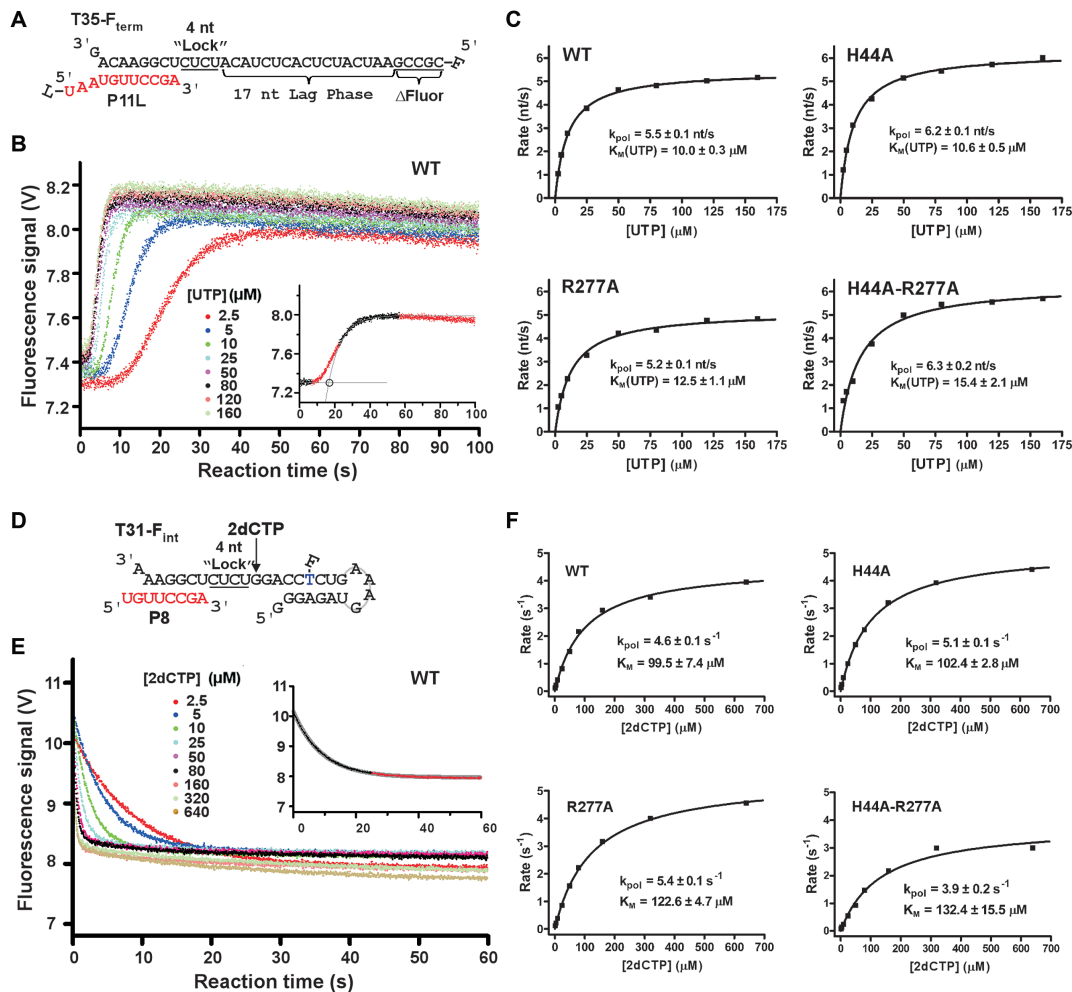
The nucleobase-binding pocket identified in this study is formed between the ‘kink’ region of the RdRP index finger and the middle finger tip. In majority of the RdRP structures from the positive-strand RNA viruses, the index finger begins with a  $\beta$ -strand, arises to the top of the right-hand architecture, goes across the top to interact with the thumb tip and then folds back and descends to the starting point (Figure 2A). Although not directly participating in catalysis, the  $\beta$ -strand plays a structural role in forming a five-stranded  $\beta$ -sheet with the middle finger and the bottom part of the ring finger. The interactions between the tips of the index finger and thumb makes the encirclement of the right-hand architecture and may restrict large-scale movement of the fingers domain typically observed in other classes of single subunit polymerases for NTP repositioning toward catalysis (reviewed in (8)). Indeed, viral RdRPs utilize a small-scale rearrangement, mainly involving motifs A and D in the palm domain to close the active site (9–10,13). Collectively, the tip and bottom parts of the index finger play relatively conserved structural roles, while the middle part including the kink region may have the potential to vary in structure and function. Indeed, the kink region is only highly conserved in structure for the *Picornaviridae* RdRPs (Figure 2A, top row), is structurally varied in the *Caliciviridae* RdRPs (Figure 2A, bottom left), and is not present in



**Figure 6.** A comparison of EC stability of EV RdRP constructs. (A) Monitoring the 14-mer (P14) product formation by EV71-C RdRP ECs after a high-salt challenge for various time points. (B) Quantitation analysis of EC stability indicated by the intensity fraction values of the elongated RNA ( $P14_{int}/(P12_{int}+P14_{int})$ ). Data obtained from the H44A-R277A construct were fitted to a single exponential decay model to estimate the EC half life ( $T_{1/2}$ ). Error bars indicate the SD values for measurements in triplicate. (C) Monitoring the P14 formation by EV71-A and CV A16 RdRP ECs after a high-salt challenge for various time points. (D) Quantitation analysis of EC stability same as in panel B for data presented in panel C.

RdRPs from the other three families with RdRP structures available (Figure 2A, bottom row, and excluding the *Caliciviridae*). Although structurally conserved, the residues responsible for nucleobase stacking and purine base recognition are only conserved in enteroviruses (stacking: T/H and R/K for EV\_A residues 44 and 277 equivalents, respectively; recognition: N and S for residues 18 and 45 equivalents, respectively) (Figure 2B). Here we have shown that histidine and threonine can be exchanged by each other at EV\_A residue 44 (Figure 6). In a previous study, a PV bearing a K276L mutation (equivalent to position 277 in EV71\_A RdRP) in the RdRP region was found to revert to WT phenotype with an arginine replacing the leucine (47), suggesting that arginine and lysine are interchangeable at this position. Interestingly, when the same site was mutated to alanine (K276A) in another study, the mutant PV was found to produce RNAs with shorter poly(A) tails (48). With the role of EC stability contribution considered, it is conceivable that an alanine mutation at this position could alter the reiterative polyadenylation process. In order to assess the contribution of the hydrogen bonding interactions between the pocket and the terminal nucleotide, we mutated S45 to a phenylalanine (S45F) or a leucine (S45L) and found that the EC stability was also impaired with these mutations, albeit to different extents (Supplementary Figure S1A). It is also

worth noting that the intra-complex interaction between a protein surface pocket and a base within an RNA tightly bound to that protein may be advantageous if compared to interactions between the same pocket and a base-containing small molecule. When a GG dinucleotide was provided at 10 or 100  $\mu$ M (2.5- or 25-fold of the T31 RNA concentration) in the EC stability assay as a challenging agent, no obvious inhibitory effect was observed (Supplementary Figure S1B). The regulatory function of the nucleobase-binding pocket identified in this study may only be utilized by a small group of picornaviruses. However, similar regulation mode through surface pocket-nucleic acid interactions may occur in nucleic acid motor enzymes as suggested below. Note that W5, a residue within the index finger  $\beta$ -strand of PV RdRP, was shown to modulate EC stability and was proposed to interact with the RNA template through stacking interactions (38). Structurally, PV RdRP W5 interacts with a metal ion mediated residue cluster comprising H270, H272 and C281 in the middle finger  $\beta$ -strands (9). Though a similar nucleobase-binding pocket was not structurally observed, W5 and its neighboring residues could offer hydrophobic interactions necessary for nucleobase binding. Although lacking direct structural evidence, it is possible that W5 may achieve its contribution in EC stability through an interaction mode analogous to what is observed



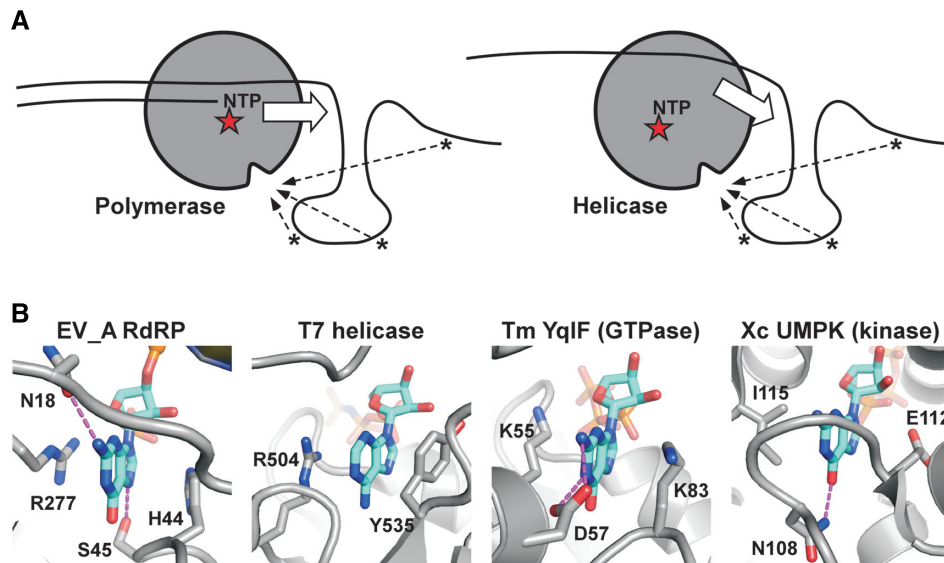
**Figure 7.** A comparison of elongation rate of EC formed by four represented EV71.C RdRP constructs. (A) The RNA constructs used to determine an average elongation rate constant. The 5'-fluorescein (F) labeled template (T35-F<sub>term</sub>) and the P11L primer were designed to have a 17-nt lag phase after a 4-nt incorporation (corresponding to the underlined CUCU 'Lock' template sequence) for EC assembly and before the fluorescence signal increase upon the last five nucleotides incorporation (ΔFluor). (B) Data from the WT construct showing a decrease in lag time with increasing concentration of UTP. Bottom right: a simple analysis was applied by fitting the latter part of the signal increase to a single exponential rise and by using the interception of this curve and the initial signal level to estimate the lag time for synthesizing the 17 nt. (C) The average elongation rates derived from data in panel B were fitted to the Michaelis–Menten equation to calculate the  $k_{pol}$  and  $K_M$  values for each construct. (D) The RNA constructs used to determine a single-nucleotide elongation rate constant. The internal fluorescein labeled template (T31-F<sub>int</sub>) and the P8 primer were designed to have an immediate fluorescence signal decrease upon a 2dCMP incorporation after a 4-nt incorporation (corresponding to the underlined CUCU 'Lock' template sequence) for EC assembly. (E) Data from the WT construct showing faster signal decrease with higher 2dCTP concentration. Top right: the data were fitted to a single exponential decay model to estimate the elongation rate with the latter part of the data omitted due to deviation of the single-exponential behavior. (F) The elongation rates derived from data in panel E were fitted to the Michaelis–Menten equation to generate the  $k_{pol}$  and  $K_M$  values for each construct.

in the current study. While both the W5 and nucleobase-binding pocket are located at the downstream side of the RNA, a previously characterized basic patch within the RNA template entry channel is at the upstream side interacting with the template backbone phosphates (49), not adjacent to neither downstream site mentioned above (Supplementary Figure S2).

NTP-driven nucleic acid motor enzymes, such as polymerase, helicase and translocases, utilize the energy provided by phosphoryl transfer or NTP hydrolysis reactions to move along nucleic acids. Typically in these proteins, the NTP binding pocket is the active site while the surface of the protein is accessible to the interactions with the nucleic acids up- or down-stream of the active site. Although not al-

ways essential to the enzymatic activities, the sequence-non-specific interactions between these proteins and the nucleic acids beyond the active site could contribute to the processivity of these enzymes, because they provide extra anchoring point(s) to the nucleic acids (Figure 8A). In this work, we identified, in EV\_A RdRPs, unique interactions of this category with a small pocket suitable for nucleobase binding that captured the downstream RNA and contributed to the stability of the RdRP EC. Introducing mutations at two residues responsible for the stacking interactions with the nucleobase resulted in reduced levels of virus proliferation, suggesting that this type of interactions likely occurred in the viral genome replication process. During viral genome replication, a replicating RdRP could encounter





**Figure 8.** Similar nucleobase-binding pocket may be generally present in nucleic acid motor enzymes. **(A)** Working models showing interaction opportunities (dashed arrows) between the nucleobases in the nucleic acids and the nucleobase-binding surface pocket in nucleic acid motors such as polymerases (left) and helicases (right). The red stars indicate the enzyme active sites. Block arrows indicate the moving direction of the enzymes. **(B)** A comparison of the nucleobase-binding pocket identified in this study and those observed in other nucleobase-binding proteins (not limited to nucleic acids motor enzymes) with similar sandwich-like stacking interactions. The nucleotides and the side chains of protein residues participating in nucleobase stacking and hydrogen bonding (dashed lines) interactions are shown in sticks. PDB entries used: EV\_A RdRP-6KWQ; T7 helicase-1E0J; Tm (*Thermotoga maritima*) YqIF-3CNN; Xc UMPK (*Xanthomonas campestris* uridylylate kinase)-3EK5).

a variety of RNA structural forms (e.g. double stranded, single stranded, etc.) with or without nucleobases accessible to such a binding pocket. Therefore, this type of interaction may only occur occasionally or randomly during the genome replication process. However, once occurring, it probably improves the stability of the RdRP–RNA complex and therefore makes positive contribution of the processive synthesis of genome-length RNA. When the unpaired 5'-guanosine was not present, as in the EC formed using a construct with a 30-mer template (T30), the EC stability was only slightly reduced comparing to that formed using the T31/P8 construct for the WT RdRP (Supplementary Figure S3). When the H44A-R277A mutant was combined with the T30/P8 construct, the EC stability was also apparently affected (Supplementary Figure S3). These data suggest that either a breathing nucleotide (i.e. being in an equilibrium between the paired and unpaired states) or other unpaired nucleotide can bind to the fingers domain pocket and contribute to EC stability. We further proposed that similar interactions may commonly occur in NTP-driven nucleic acid motors. The fact that similar interactions were not previously identified may be related to the randomness of the interaction sites along the nucleic acids. The inter-EC interactions in our crystal lattice that stabilize the downstream RNA played a key role for identifying this type of interactions, inspiring us to collect further evidence supporting its role in enzyme stability. In the DNA-dependent T7 RNA polymerase EC, residues E168, R647 and N671 form a platform and allow the +1 non-template nucleotide to bind, possibly contributing to the maintenance of the downstream edge of the transcription bubble (50) (Supplementary Figure S4A). If compared to the T7 RNA polymerase EC, the 5' nucleotide in our EV71 RdRP EC is equivalent

to the +3 non-template nucleotide, and therefore these two interaction modes are analogous but not identical (Supplementary Figure S4B). The sandwich-like nucleobase binding interactions observed in our study are very common in nucleotide binding enzymes (not to be confused with the aforementioned NTP-driven nucleic acid motors) including helicases (51), GTPases (52) and kinases (53), while the residue types providing the hydrophobic interactions to the nucleobase are quite diverse (Figure 8B). Hence, it is not difficult for nucleic acid motors to assemble such a nucleobase-binding pocket during evolution. Once functional contribution, such as processivity or stability enhancement, is established based on the interactions between the pocket and nucleobases, key residues constituting the pocket could become more conservative as positions 44 and 277 in enterovirus RdRPs.

Members of the *Enterovirus* genus include causative agents of HFMD, poliomyelitis and common cold. The identification of the functionally conserved nucleobase-binding pocket in this study may provide opportunities for developing inhibitors or live attenuated vaccines. As discussed above, the assembly of this surface pocket is analogous to nucleotide binding pocket observed in other enzymes. Therefore, it may have the potential to bind not only nucleobases within a nucleic acid strand, but also nucleotide analogs commonly studied in antiviral research for RNA viruses. With respect to the development of live-attenuated vaccine, this pocket has two advantages. First, the two key residues are far apart in the primary structure, and therefore it is less susceptible to reversion if both sites are chosen for mutation in the vaccine strain. Second, the residue type for both positions is not strictly conserved as long as it can provide suitable hydrophobic interactions to the nucle-

obase, and allows a variety of amino acid combinations to be tested for viable viruses with attenuated features.

In summary, we identified a nucleobase-binding pocket in enterovirus RdRP by crystallography and analyzed its relevance to polymerase function through *in vitro* biochemical characterization and cell-based virological studies. Taken together, this pocket contributes to RdRP EC stability and virus proliferation, likely through its nucleobase binding capability. We propose that similar regulatory mechanisms, although rarely reported, may commonly exist in nucleic acid motor enzymes. Our work could serve as a general reference for future studies to dissect the mechanism behind nucleobase anchoring sites identified in these enzymes through non-crystallographic approaches.

## DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6KWQ and 6KWR.

## SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr Olve Peersen for pointing out the ‘kink’ region as a potential interesting element in picornavirus RdRPs, Dr Hanzhong Wang for providing the cloning materials for the EV71-A (BrCr ts strain) RdRP gene, Dr Guoliang Lu and Dr Bo Shu for X-ray diffraction data collection, Meihua Wang for preparing the template RNA T30, Liu Deng and Qiaojie Liu for laboratory assistance, synchrotron SSRF (beamline BL17U Shanghai, China) for access to beamline and The Core Facility and Technical Support, Wuhan Institute of Virology, for access to instruments. *Author contributions*: P.G. and B.Z. conceived the research; W.S. performed the crystallographic experiments; C.-L.D. performed the virological experiments; W.S., H.-Q.Y., and R.L. performed the biochemical experiments. P.G. and W.S. analyzed the crystallographic data; C.-L.D. and B.Z. analyzed the virological data; W.S., H.-Q.Y., and P.G. analyzed the biochemical data; W.S. and P.G. assembled the initial manuscript and all authors contributing to the writing of the paper.

## FUNDING

National Key Research and Development Program of China [2018YFA0507200, 2018YFD0500100, 2016YFC1200400]; National Natural Science Foundation of China [31670154, 31802147]; Chinese Academy of Science Funds: Advanced Customer Cultivation Project of Wuhan National Biosafety Laboratory [2018ACCP-MS06, 2019ACCP-MS06]; ‘One-Three-Five’ Strategic Programs, Wuhan Institute of Virology [Y605191SA1]. Funding for open access charge: Ministry of Science and Technology of the People’s Republic of China [2018YFA0507200]; National Natural Science Foundation of China [31670154]. *Conflict of interest statement*. None declared.

## REFERENCES

- Ooi, M.H., Wong, S.C., Lewthwaite, P., Cardoso, M.J. and Solomon, T. (2010) Clinical features, diagnosis, and management of enterovirus 71. *Lancet Neurol.*, **9**, 1097–1105.
- Huang, P.N. and Shih, S.R. (2014) Update on enterovirus 71 infection. *Curr. Opin. Virol.*, **5**, 98–104.
- Wang, S.M. and Liu, C.C. (2009) Enterovirus 71: epidemiology, pathogenesis and management. *Expert Rev. Anti Infect. Ther.*, **7**, 735–742.
- Lin, J.Y., Chen, T.C., Weng, K.F., Chang, S.C., Chen, L.L. and Shih, S.R. (2009) Viral and host proteins involved in picornavirus life cycle. *J. Biomed. Sci.*, **16**, 103.
- Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F. and Weber, P.C. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat. Struct. Biol.*, **6**, 937–943.
- Thompson, A.A. and Peersen, O.B. (2004) Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J.*, **23**, 3462–3471.
- Ng, K.K., Pendas-Franco, N., Rojo, J., Boga, J.A., Machin, A., Alonso, J.M. and Parra, F. (2004) Crystal structure of norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. *J. Biol. Chem.*, **279**, 16638–16645.
- Shu, B. and Gong, P. (2017) The uncoupling of catalysis and translocation in the viral RNA-dependent RNA polymerase. *RNA Biol.*, **14**, 1314–1319.
- Gong, P. and Peersen, O.B. (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *PNAS*, **107**, 22505–22510.
- Shu, B. and Gong, P. (2016) Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *PNAS*, **113**, E4005–E4014.
- Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E. and Verdaguer, N. (2007) Sequential structures provide insights into the fidelity of RNA replication. *PNAS*, **104**, 9463–9468.
- Zamyatkin, D.F., Parra, F., Alonso, J.M., Harki, D.A., Peterson, B.R., Grochulski, P. and Ng, K.K. (2008) Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *J. Biol. Chem.*, **283**, 7705–7712.
- Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D. 3rd, Wetmore, D.R., McGrath, M.E., Ray, A.S. *et al.* (2015) Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science*, **347**, 771–775.
- Johansson, M., Brooks, A.J., Jans, D.A. and Vasudevan, S.G. (2001) A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. *J. Gen. Virol.*, **82**, 735–745.
- Tay, M.Y., Saw, W.G., Zhao, Y., Chan, K.W., Singh, D., Chong, Y., Forwood, J.K., Ooi, E.E., Gruber, G., Lescar, J. *et al.* (2015) The C-terminal 50 amino acid residues of dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *J. Biol. Chem.*, **290**, 2379–2394.
- Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L. and Gamarnik, A.V. (2006) A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.*, **20**, 2238–2249.
- Paul, A.V., Rieder, E., Kim, D.W., van Boom, J.H. and Wimmer, E. (2000) 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.
- Liu, Y., Zheng, Z., Shu, B., Meng, J., Zhang, Y., Zheng, C., Ke, X., Gong, P., Hu, Q. and Wang, H. (2016) SUMO modification stabilizes enterovirus 71 polymerase 3D To facilitate viral replication. *J. Virol.*, **90**, 10472–10485.
- Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and Imbert, I. (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *PNAS*, **111**, E3900–E3909.
- Lu, G. and Gong, P. (2013) Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog.*, **9**, e1003549.

21. Liu, W., Shi, X. and Gong, P. (2018) A unique intra-molecular fidelity-modulating mechanism identified in a viral RNA-dependent RNA polymerase. *Nucleic Acids Res.*, **46**, 10840–10854.
22. Wu, J., Lu, G., Zhang, B. and Gong, P. (2015) Perturbation in the conserved methyltransferase-polymerase interface of flavivirus NS5 differentially affects polymerase initiation and elongation. *J. Virol.*, **89**, 249–261.
23. Zhao, Y., Soh, T.S., Zheng, J., Chan, K.W., Phoo, W.W., Lee, C.C., Tay, M.Y., Swaminathan, K., Cornvik, T.C., Lim, S.P. *et al.* (2015) A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog.*, **11**, e1004682.
24. Li, W., Wu, B., Soca, W.A. and An, L. (2018) Crystal structure of classical swine fever virus NS5B reveals a novel N-terminal domain. *J. Virol.*, **92**, e00324-18.
25. Jia, H. and Gong, P. (2019) A structure-function diversity survey of the RNA-dependent RNA polymerases from the positive-strand RNA viruses. *Front. Microbiol.*, **10**, 1945.
26. Gohara, D.W., Ha, C.S., Kumar, S., Ghosh, B., Arnold, J.J., Wisniewski, T.J. and Cameron, C.E. (1999) Production of “authentic” poliovirus RNA-dependent RNA polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in *Escherichia coli*. *Protein Expr. Purif.*, **17**, 128–138.
27. Shang, B.D., Deng, C.L., Ye, H.G., Xu, W.B., Yuan, Z.M., Shi, P.Y. and Zhang, B. (2013) Development and characterization of a stable eGFP enterovirus 71 for antiviral screening. *Antivir. Res.*, **97**, 198–205.
28. Lu, G.L. and Gong, P. (2013) Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog.*, **9**, e1003549.
29. Papworth, C.B., Bauer, J.C., Braman, J. and Wright, D.A. (1996) Site-directed mutagenesis in one day with >80% efficiency. *Strategies*, **9**, 3–4.
30. Batey, R.T. and Kieft, J.S. (2007) Improved native affinity purification of RNA. *RNA*, **13**, 1384–1389.
31. Wu, J.Q., Lu, G.L., Zhang, B. and Gong, P. (2015) Perturbation in the conserved methyltransferase-polymerase interface of flavivirus NS5 differentially affects polymerase initiation and elongation. *J. Virol.*, **89**, 249–261.
32. Pflugrath, J.W. (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr. D*, **55**, 1718–1725.
33. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
34. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.
35. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D*, **66**, 213–221.
36. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
37. Theobald, D.L. and Wuttke, D.S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22**, 2171–2172.
38. Hobdley, S.E., Kempf, B.J., Steil, B.P., Barton, D.J. and Peersen, O.B. (2010) Poliovirus polymerase residue 5 plays a critical role in elongation complex stability. *J. Virol.*, **84**, 8072–8084.
39. Gnadig, N.F., Beaucourt, S., Campagnola, G., Borderia, A.V., Sanz-Ramos, M., Gong, P., Blanc, H., Peersen, O.B. and Vignuzzi, M. (2012) Cocksackievirus B3 mutator strains are attenuated in vivo. *PNAS*, **109**, E2294–E2303.
40. Gong, P., Campagnola, G. and Peersen, O.B. (2009) A quantitative stopped-flow fluorescence assay for measuring polymerase elongation rates. *Anal. Biochem.*, **391**, 45–55.
41. McDonald, S., Block, A., Beaucourt, S., Moratorio, G., Vignuzzi, M. and Peersen, O.B. (2016) Design of a genetically stable high fidelity coxsackievirus B3 polymerase that attenuates virus growth in vivo. *J. Biol. Chem.*, **291**, 13999–14011.
42. Deng, C.L., Yeo, H.M., Ye, H.Q., Liu, S.Q., Shang, B.D., Gong, P., Alonso, S., Shi, P.Y. and Zhang, B. (2014) Inhibition of enterovirus 71 by adenosine analog NITD008. *J. Virol.*, **88**, 11915–11923.
43. Gong, P., Kortus, M.G., Nix, J.C., Davis, R.E. and Peersen, O.B. (2013) Structures of coxsackievirus, rhinovirus, and poliovirus polymerase elongation complexes solved by engineering RNA mediated crystal contacts. *PLoS One*, **8**, e60272.
44. Gong, P. and Peersen, O.B. (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 22505–22510.
45. Bi, P., Shu, B. and Gong, P. (2017) Crystal structure of the coxsackievirus A16 RNA-dependent RNA polymerase elongation complex reveals novel features in motif A dynamics. *Virol. Sin.*, **32**, 548–552.
46. Mestas, S.P., Sholders, A.J. and Peersen, O.B. (2007) A fluorescence polarization-based screening assay for nucleic acid polymerase elongation activity. *Anal. Biochem.*, **365**, 194–200.
47. Richards, O.C. and Ehrenfeld, E. (1997) One of two NTP binding sites in poliovirus RNA polymerase required for RNA replication. *J. Biol. Chem.*, **272**, 23261–23264.
48. Kempf, B.J., Kelly, M.M., Springer, C.L., Peersen, O.B. and Barton, D.J. (2013) Structural features of a picornavirus polymerase involved in the polyadenylation of viral RNA. *J. Virol.*, **87**, 5629–5644.
49. Kortus, M.G., Kempf, B.J., Haworth, K.G., Barton, D.J. and Peersen, O.B. (2012) A template RNA entry channel in the fingers domain of the poliovirus polymerase. *J. Mol. Biol.*, **417**, 263–278.
50. Yin, Y.W. and Steitz, T.A. (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, **116**, 393–404.
51. Singleton, M.R., Sawaya, M.R., Ellenberger, T. and Wigley, D.B. (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, **101**, 589–600.
52. Kim, D.J., Jang, J.Y., Yoon, H.J. and Suh, S.W. (2008) Crystal structure of Y1qF, a circularly permuted GTPase: implications for its GTPase activation in 50 S ribosomal subunit assembly. *Proteins*, **72**, 1363–1370.
53. Tu, J.L., Chin, K.H., Wang, A.H. and Chou, S.H. (2009) Unique GTP-binding pocket and allostery of uridylylase kinase from a gram-negative phytopathogenic bacterium. *J. Mol. Biol.*, **385**, 1113–1126.