



The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication

Grace Campagnola,^a Vishnu Govindarajan,^a Annelise Pelletier,^a Bruno Canard,^b  Olve B. Peersen^a

^aDepartment of Biochemistry & Molecular Biology, Colorado State University, Fort Collins, Colorado, USA

^bCentre National de la Recherche Scientifique, Aix-Marseille Université CNRS UMR 7257, AFMB, Marseille, France

ABSTRACT Positive-strand RNA viruses replicate their genomes using virally encoded RNA-dependent RNA polymerases (RdRP) with a common active-site structure and closure mechanism upon which replication speed and fidelity can evolve to optimize virus fitness. Coronaviruses (CoV) form large multicomponent RNA replication-transcription complexes containing a core RNA synthesis machine made of the nsp12 RdRP protein with one nsp7 and two nsp8 proteins as essential subunits required for activity. We show that assembly of this complex can be accelerated 5-fold by preincubation of nsp12 with nsp8 and further optimized with the use of a novel nsp8L7 heterodimer fusion protein construct. Using rapid kinetics methods, we measure elongation rates of up to 260 nucleotides (nt)/s for the core replicase, a rate that is unusually fast for a viral polymerase. To address the origin of this fast rate, we examined the roles of two CoV-specific residues in the RdRP active site: Ala547, which replaces a conserved glutamate above the bound NTP, and Ser759, which mutates the palm domain GDD sequence to SDD. Our data show that Ala547 allows for a doubling of replication rate, but this comes at a fidelity cost that is mitigated by using a SDD sequence in the palm domain. Our biochemical data suggest that fixation of mutations in polymerase motifs F and C played a key role in nidovirus evolution by tuning replication rate and fidelity to accommodate their large genomes.

IMPORTANCE Replicating large genomes represents a challenge for RNA viruses because fast RNA synthesis is needed to escape innate immunity defenses, but faster polymerases are inherently low-fidelity enzymes. Nonetheless, the coronaviruses replicate their ≈ 30 -kb genomes using the core polymerase structure and mechanism common to all positive-strand RNA viruses. The classic explanation for their success is that the large-genome nidoviruses have acquired an exonuclease-based repair system that compensates for the high polymerase mutation rate. In this work, we establish that the nidoviral polymerases themselves also play a key role in maintaining genome integrity via mutations at two key active-site residues that enable very fast replication rates while maintaining typical mutation rates. Our findings further demonstrate the evolutionary plasticity of the core polymerase platform by showing how it has adapted during the expansion from short-genome picornaviruses to long-genome nidoviruses.

KEYWORDS RdRP, coronavirus, fidelity, polymerase

Positive-strand RNA viruses replicate their genomes using a virally encoded RNA-dependent RNA polymerase (RdRP) to carry out both minus and plus sense RNA synthesis. For the coronaviruses, this functionality is found in the nsp12 protein, and unlike most positive-strand RNA viruses that have single subunit RdRPs, the CoV polymerase is only active when complexed with the viral nsp7 and nsp8 proteins (1, 2). The structure of the polymerase from the original severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1) strain showed a surprising stoichiometry wherein one copy of nsp12 interacted with one nsp7/nsp8 heterodimer and a second nsp8 monomer (3),

Editor Anice C. Lowen, Emory University School of Medicine

Copyright © 2022 American Society for Microbiology. All Rights Reserved.

Address correspondence to Olve B. Peersen, Olve.Peersen@ColoState.edu.

The authors declare no conflict of interest.

Received 27 April 2022

Accepted 6 July 2022

Published 4 August 2022

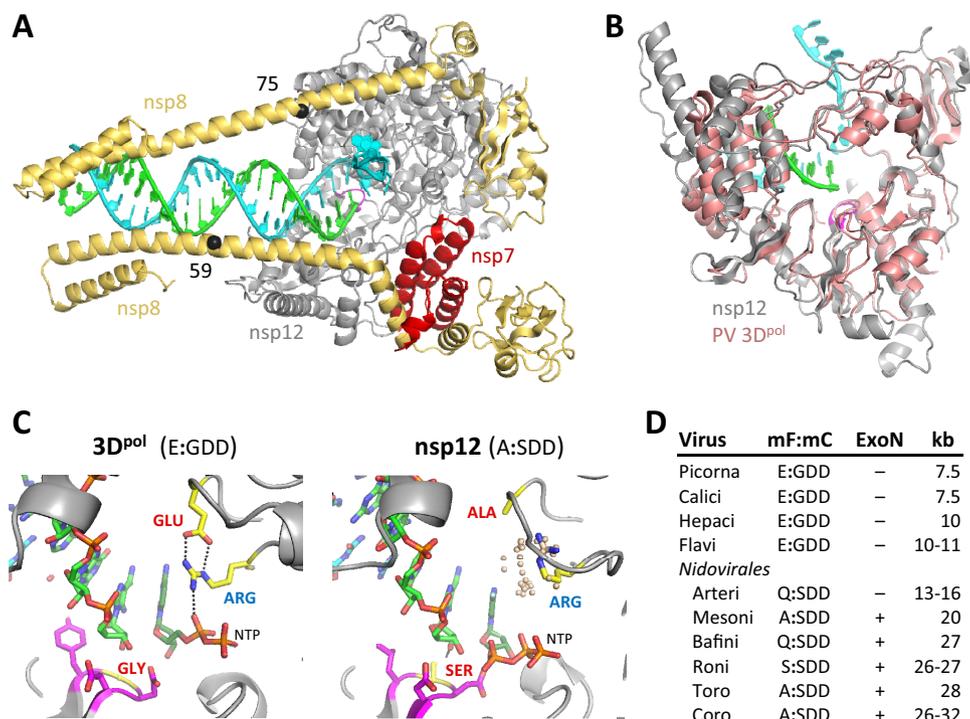


FIG 1 Structures of coronaviral and picornaviral polymerases. (A) Top view of the SARS-CoV-2 core replicase complexes of nsp12 RdRP (gray), nsp7 (red) and two nsp8 (gold) subunits, and template-product RNA duplex (cyan and green, respectively). (B) Structural superposition of nsp12 (gray) with the poliovirus 3D^{pol}-RNA (red) elongation complex showing the high structural similarity within the polymerase core of nsp12. (C) Detailed views of the 3D^{pol} and nsp12 active sites showing the locations of the key residues (red labels) targeted in this study; the Glu-versus-Ala in motif F (mF) and the Gly-versus-Ser in motif C (mC). The motif F arginine (blue labels) is conserved in all RdRPs. In 3D^{pol} and most other viral RdRPs, the Glu holds the Arg in place above the NTP-binding site via electrostatic interactions, but in the coronaviruses, this Glu is replaced by an Ala, and the Arg adopts a range of different conformations. This range is depicted by the tan spheres that show the positions of the arginine guanidinium group carbon atom in all 46 SARS-CoV nsp12 structures solved thus far. The panel shows the structures of the poliovirus 3D^{pol}-RNA-2'-dCTP complex (PDB ID [3OLA](#) [14] with Glu161, Arg174, and Gly327) and the nsp12-RNA-favipiravir complex (PDB ID [7CTT](#) [55] with Ala547, Arg555, and Ser759). (D) Comparison of residues found in these motif F and motif C positions across a range of positive-strand RNA viruses with different genome lengths.

and the more than 40 structures of SARS-CoV-2 replicase solved in the past 2 years have expanded on this core structure. Notably among these, the two nsp8 molecules contain long N-terminal helices that interact nonspecifically with a long product RNA duplex as “sliding pole” processivity factors (4) (Fig. 1A). These helices also provide binding sites for two nsp13 helicases (5), while additional surfaces of nsp12 provide binding sites for nsp9 and a nsp10/14 heterodimer that contains both exonuclease and methyltransferase domains (6, 7). A growing body of biochemistry and structural biology data point to two distinct experimental contexts for the nsp12 RdRP subunit. The first is in a nsp12-nsp7/8-nsp8 “core polymerase” complex that can carry out processive elongation *in vitro* and is often used for biochemistry and nucleoside analog incorporation studies. The second is in the context of a much larger viral replication-transcription complex (RTC) that adds the nsp13 helicase, nsp14 exonuclease, N7-methyltransferase, nsp15 endonuclease, and nsp16 2'-O-methyltransferase subunits alongside the small nsp9 and nsp10 accessory proteins that stabilize subunit conformations and may indirectly chaperone RTC assembly. The full RTC structure has not been solved yet, but the multitude of nsp12 and other CoV nsp structures are providing hints about its composition (8–10), and a credible model for a 2.5-MDa RTC with hexameric symmetry has been built by combining interactions identified in cryo-electron microscopy (cryo-EM) and crystal structures of proteins from both SARS-CoV viruses with computational docking approaches (11).

Despite having a very large multicomponent RTC, the nsp12 RdRP active site and core fold is very similar to that of the 52-kDa picornaviral 3D^{pol} enzymes that represent the smallest viral RdRPs (Fig. 1B). This is not altogether surprising as the nidoviruses evolved from the picornavirus-like supergroup of positive-strand RNA viruses (12, 13). Structural studies have shown that positive-strand RNA virus polymerases utilize an active-site closure mechanism that is based on a subtle movement of conserved motif A in the palm domain (14, 15), and biochemical and infectious virus studies show this core structure to be a platform upon which viruses can readily evolve their replication speed and fidelity to optimize fitness (16). Notably, replication rate and fidelity have been shown to be inversely correlated, with higher replication rates having a significant fidelity cost in both biochemical experiments and infectious virus studies (17–20). The CoV polymerase shares this structural core with a default *open* conformation active site, and mutations in motif A can alter replication fidelity based on the development of 5-fluoro-uracil resistance in mouse hepatitis virus CoV (21).

The ≈30-kB SARS-CoV genome is three to four times larger than those of more common RNA viruses such as flaviviruses and picornaviruses, and this size presents two key biochemical challenges for the CoV replication machinery. First, it must be very fast in order to replicate the entire genome before innate immune responses are activated in cells, and second, it must retain sufficiently high fidelity so as to maintain genetic integrity. Current data show that the CoV core replicase elongates at 150 to 200 nucleotides (nt)/s (22–25), which is more than twice as fast as poliovirus 3D^{pol}, and based on prior RdRP studies, this is likely to be accompanied by a fairly high mutation rate inherent to the polymerase active-site architecture (16). Several studies show nsp12 readily misincorporates native and modified nucleotides *in vitro*, including the nucleoside analogs remdesivir and favipiravir that show antiviral activity in cell culture and molnupiravir (N4-hydroxycytidine) that is approved as a CoV antiviral (24, 26, 27). Careful comparative measurements of nsp12 fidelity with other polymerases have not been made, and it is not known whether there are molecular features of nsp12 that allow it to maintain fidelity in light of its fast replication rate. The large-genome CoVs are also unique among RNA viruses in having an exonuclease-based repair mechanism via the ExoN domain of nsp14 that can offset the inherently reduced fidelity of a faster polymerase (28).

While the core of SARS-CoV nsp12 polymerase shares the canonical RdRP fold and mechanism, it does have two notable active-site differences from the typical positive-strand RNA virus polymerase (Fig. 1C); Ala547 within motif F reflects the loss of a key glutamic acid residue that normally sits above the active site and aids in positioning an arginine side chain and the NTP for catalysis and Ser759 within motif C results in a SDD sequence instead of a GDD sequence in the palm domain that interacts with the nascent RNA strand. Thus, positive-strand RNA virus polymerases generally use a glutamate-GDD configuration of these motifs F and C residues that we will denote as E:GDD, but CoVs have an A:SDD configuration (Fig. 1D). The SDD is present in all nidoviruses, but interestingly, the shorter genome arteriviruses have Q:SDD with a motif F glutamine that lacks the formal negative charge of the glutamate and would therefore have a weaker interaction with the active-site arginine.

In this study, we sought to determine how these two noncanonical features of the CoV nsp12 active site affect replication rate and fidelity. Detailed functional studies have been hampered by the difficulty of obtaining pure, highly active coronavirus replicase preparations. To overcome this complication, we first developed a new nsp8L7 fusion protein and used it to elucidate the pathway for core replicase assembly and then used assembled complexes in rapid kinetics and misincorporation studies to assess the biochemical effects associated with the Ala547 and Ser759 residues. Our data show that having Ala547 in motif F more than doubles the replication rate, and while this comes at a significant fidelity cost, the effect is mitigated by using SDD instead of GDD within motif C as a new fidelity control point. As a result, the overall nucleotide misincorporation level of nsp12 is on par with that of the much slower

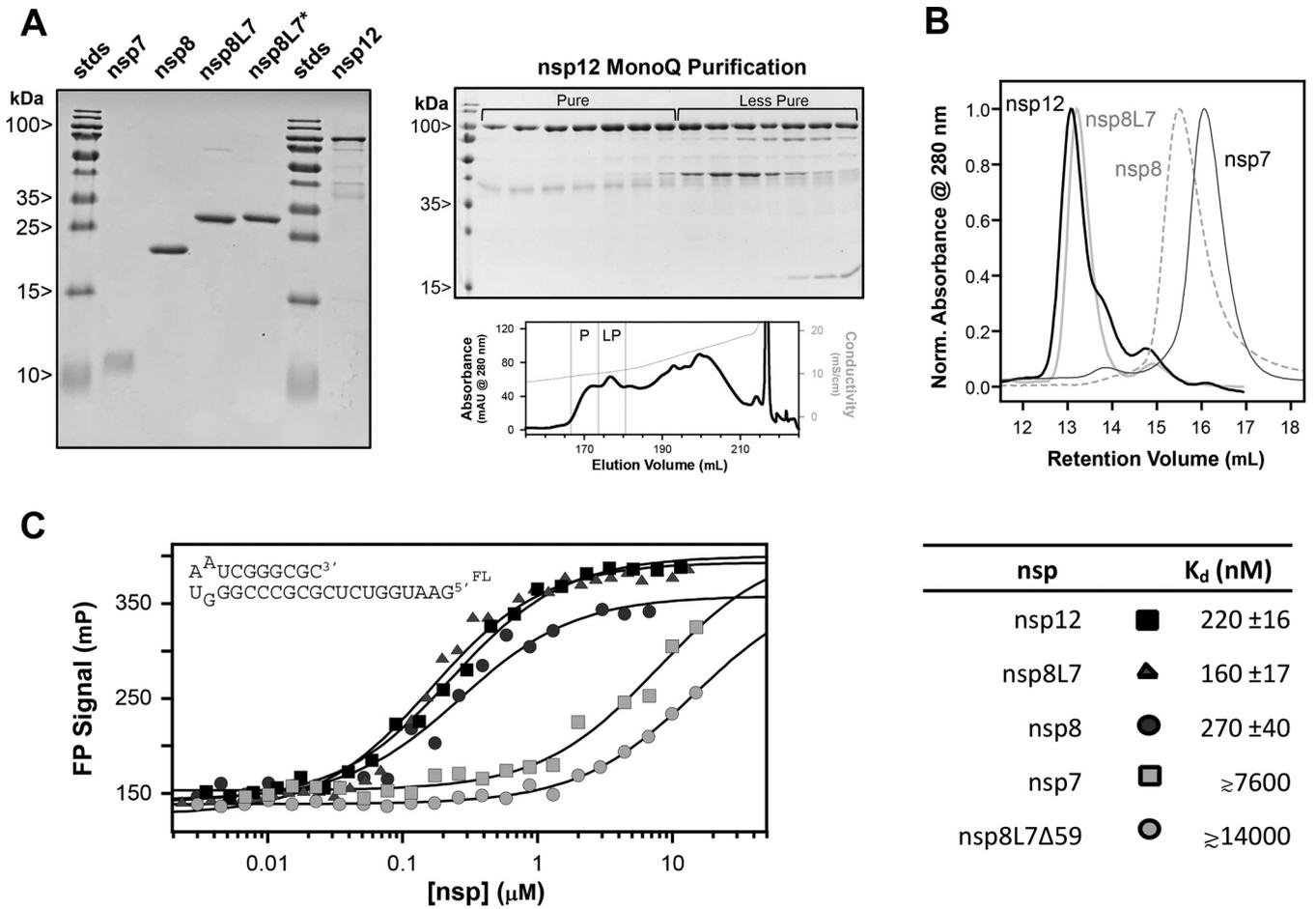


FIG 2 Protein purification and RNA binding. (A) SDS-PAGE of purified SARS-CoV-2 nonstructural proteins, including an engineered nsp8L7 fusion protein with a 6-residue linker (8L7* contains an 18-residue linker), MonoQ anion-exchange column chromatogram, and SDS-PAGE gel showing separation of nsp12 into “pure” (P) and “less pure” (LP) sample pools as indicated on the gel with brackets. (B) Superposed size-exclusion chromatograms showing that nsp12 (109 kDa), nsp8 (24 kDa), and nsp7 (11 kDa) elute at their expected sizes, while nsp8L7 (33 kDa) is well shifted from the expected elution volume. (C) Fluorescence polarization (FP) RNA-binding data for each of the nsp constructs obtained in 50 mM NaCl, pH 7.

poliovirus polymerase. These compensatory biochemical effects in CoV nsp12 show a new link between the core RdRP active-site mechanisms and maximizing virus fitness and provide function-based insights into nidoviral polymerase evolution.

RESULTS

Design of a highly active SARS-CoV-2 core replicase. Examination of the cryo-EM structure of the SARS-CoV-2 core replicase reveals that the nsp8 C terminus is in the vicinity of the nsp7 N terminus (3, 8). We designed nsp8L7, a single-chain version of the nsp7/8 heterodimer, with the goal of expediting the kinetics and specificity of complex formation. The nsp8L7 protein, in which nsp8 is connected to the N terminus of nsp7 via a 6-residue GSGSGS linker, allows us to delineate the biochemical roles of the two distinct nsp8 molecules in the replicase complex while retaining the long nsp8 N-terminal α -helix that interacts with the product RNA duplex. All proteins used in this study were expressed in *Escherichia coli* using T7 promoters and purified through metal-affinity, ion-exchange, and size-exclusion chromatography (Fig. 2A). Purification yield of nsp12 was fairly low at \approx 1 mg/liter of bacterial culture, and it copurified through metal-affinity chromatography with a number of bacterial proteins (29) that could be largely removed via a shallow salt gradient on a high-resolution MonoQ anion-exchange column (Fig. 2A). The size-exclusion data showed nsp12 (109 kDa), nsp8 (24 kDa), and nsp7 (11 kDa) eluting near their expected positions based on

standards, but the nsp8L7 (33 kDa) fusion protein eluted at an apparent molecular weight of ≈ 100 kDa (Fig. 2B). This indicates nsp8L7 forms oligomers in solution, which is consistent with several crystal structures showing that the nsp7/nsp8 heterodimer can itself form (nsp7/8)₂ dimers, and these dimers may further oligomerize into [(nsp7/nsp8)₂]₂ tetramers and larger hexadecamers via interactions involving the long N-terminal helix of nsp8 (30–33). RNA-binding experiments in the presence of 50 mM monovalent salt show that nsp8, nsp8L7, and nsp12 all bound a short hairpin primer-template RNA with ≈ 200 nM K_d values, while nsp7 binds weakly with a 7 μ M K_d (Fig. 2C). Deletion of the first 59 residues of the nsp8L7 fusion protein reduced binding affinity more than 80-fold to ≥ 14 μ M.

Core replicase assembly. Assembly of the core replicase complex, here defined as nsp12-nsp7/8-nsp8, from individual components and a hairpin primer-template RNA (Fig. 3A) was observed to be both slow and inefficient at micromolar concentrations, requiring ≈ 15 min and elongating only half of the RNA hairpin primer provided (Fig. 3B). Final concentrations were 1 μ M RNA and 2.5 μ M nsp12 either with 5 μ M nsp7 and 10 μ M nsp8 or with 5 μ M nsp8 and 5 μ M nsp8L7. This result is based on using short 15-s burst-phase elongation reactions to instantaneously assay the amount of competent complex present after various assembly times, and the data therefore do not reflect multiple turnover product formation over the course of the entire assembly window. Assembly with nsp8L7 was ≈ 2 -fold faster and $\approx 35\%$ more efficient than assembly with individual nsp7 and nsp8 proteins (Fig. 3B and D). Preincubation of nsp12 with nsp8 prior to the addition of nsp8L7 and RNA accelerated the assembly of functional replicase ≈ 3 -fold, but preincubation of nsp12 with nsp8L7 did not (Fig. 3C and D), indicating that the slow assembly from separate components is due to the lone nsp8 binding to the nsp12 pinky finger domain. Overall, the combination of nsp8 preincubation and using the nsp8L7 fusion instead of nsp7+nsp8 led to a 7.5-fold increase in core replicase assembly efficiency. Preassembly of nsp12+nsp8 followed by nsp8L7+RNA addition was used for all subsequent experiments unless otherwise stated.

To assess the lifetime of the nsp12-8 heterodimer, we assembled the complex, purified it by size-exclusion chromatography, and then made a series of reinjections of the purified sample on the same size-exclusion column over an 18-h period. These data showed that the heterodimer dissociated to a new equilibrium with a time constant of ≈ 6 h at room temperature in 100 mM NaCl, pH 7, 10% glycerol (Fig. 3E), and the final concentrations based on measured peak areas were ≈ 265 nM complex and ≈ 640 nM free nsp12 and nsp8, yielding an estimated K_d of ≈ 1.5 μ M for the nsp12-nsp8 interaction. We also note that the amount of actively elongating complex formed varied somewhat among different nsp12 preparations and sample pools (Fig. 2A), and we attribute this to conformational heterogeneity of nsp12 and the presence of copurifying bacterial proteins and chaperones that primarily interfere with formation of the nsp12-nsp8 heterodimer. The final assembly step after addition of nsp8L7 and RNA was consistently 2 to 3 min, and regardless of overall assembly efficiency, the active complexes formed exhibited consistent elongation rates and misincorporation levels in the assays described below.

Processive elongation rates. Processive elongation rates and nucleotide $K_{m,app}$ values were determined using a fluorescence based stopped-flow assay to measure the time needed to replicate a 26-nucleotide-long single-stranded RNA template (Fig. 4A). The assay is based on a fluorescence signal change triggered by the polymerase approaching the very 5' end of the template strand RNA and thus reports only on the population of elongation competent replicases (34). Notably, the assay result is insensitive to the replicase assembly efficiency as the signal from nonelongated RNAs remains constant; it only requires enough competent complexes to generate a fluorescence change with sufficient signal-to-noise ratio to be interpretable. Using titrations with an equimolar mixture of the four NTPs, we examined the concentration dependence of processive elongation rates as a function of temperature at pH 6.5 and as a function of

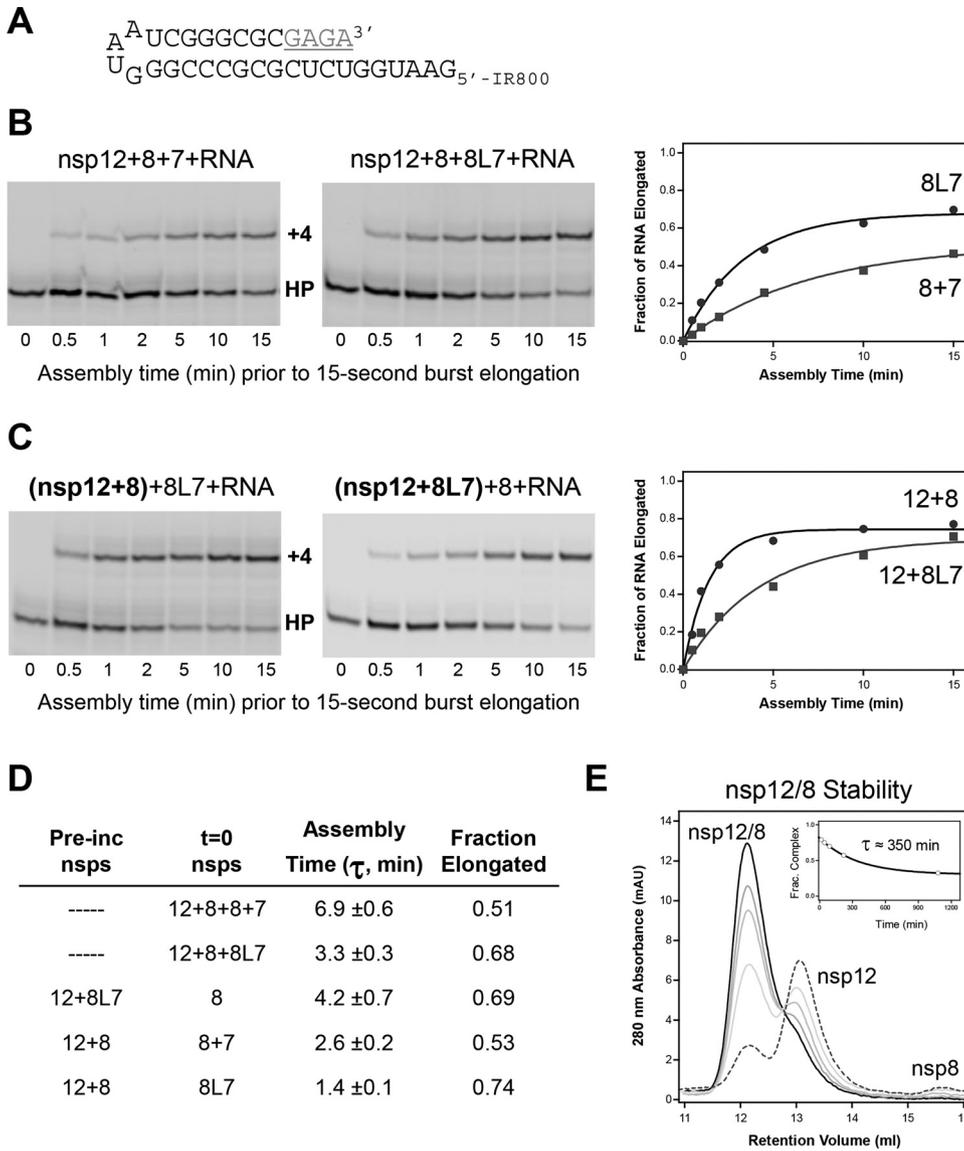


FIG 3 Core CoV replicase assembly. (A) Hairpin RNA (HP) used to assay replicase assembly via 15-s elongation reactions yielding a +GAGA (+4, gray) product imaged via a 5' IRDye800 label. (B) 7 M urea/Tris-borate-EDTA (TBE)-PAGE showing that the 8L7 fusion protein improves both the rate and efficiency of complex assembly compared to individual nsp7 and nsp8 subunits in reactions initiated by mixing all the components. (C) A 30-min preincubation of nsp12 with nsp8, but not nsp8L7, accelerates the core replicase assembly process. (D) Quantification of assembly times obtained by fitting the fraction of RNA elongated to a single exponential curve. The data in panels B to D reflect the final assembly step done with 1 μ M RNA, 2.5 μ M nsp12, and either 5 μ M nsp7 with 10 μ M nsp8 or 5 μ M nsp8 with 5 μ M nsp8L7. The 30-min preincubation steps were done at 2 \times final concentrations, i.e., 5 μ M nsp12 and 10 μ M cofactor proteins. (E) Size-exclusion chromatography was used to monitor the temporal stability of the nsp12/8 complex at room temperature. Area measurements of the nsp12/8 and nsp12 peaks showed the complex dissociates in \approx 6 h (inset) and has a \approx 1.5 μ M K_d based on estimated equilibrium concentrations.

pH at 30°C (Fig. 4B). The extrapolated maximal rates observed across these conditions ranged from 23 \pm 2 to 155 \pm 15 nt/s with NTP $K_{m,app}$ values in the 50 to 100 μ M range.

Based on these data, we carried out an expanded temperature range and pH study at a single 350 μ M NTP concentration, i.e., at near-maximal elongation rate conditions. The resulting data showed the CoV replicase to be an extremely fast RdRP, with processive elongation rates of 250 to 300 nt/s at 37°C and pH 7.5 (Fig. 4C). Individual deletions of the nsp8 N-terminal helix regions that do not directly contact nsp12, i.e., nsp8- Δ 75 and nsp8L7- Δ 59 (Fig. 1A), did not significantly affect the processive elongation rate. These deletions did reduce the amount of competent replicase formed, which led to lower signal-to-noise ratio in the stopped-flow data, but reliable elongation rates

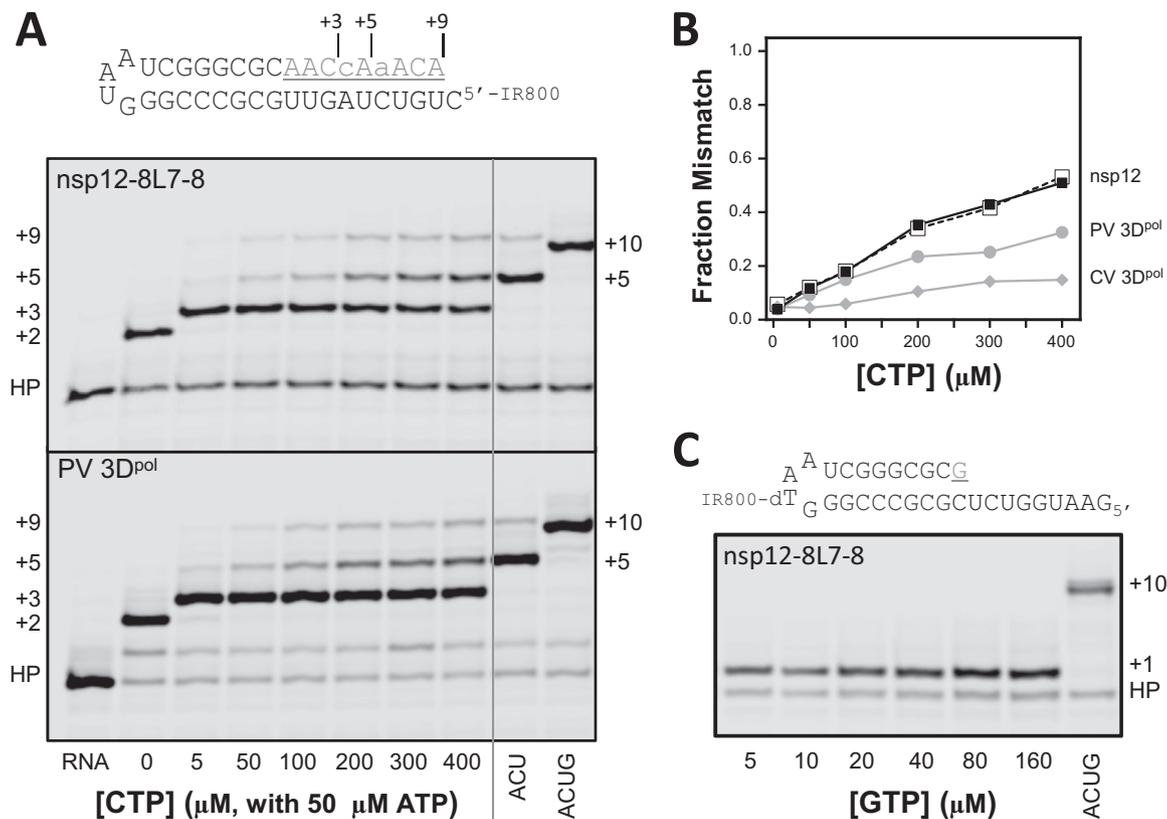


FIG 5 NTP misincorporation assay. (A) Hairpin RNA (HP) used to observe mismatch incorporation (underlined lowercase positions) by titrating the CTP concentration in 15-s burst-phase elongation assays. Elongation to the +5 and +9 bands required CTP:A and ATP:C misincorporation steps, respectively. (B) Comparison of misincorporation product formation by SARS-CoV-2 nsp12 and 3D^{pol} polymerases from poliovirus and coxsackievirus B3. nsp12 data were determined using both nsp8L7 + nsp8 (filled squares) and nsp7 + nsp8 (open squares) protein combinations. (C) Hairpin RNA (HP) used in a similar assay to measure GTP:U misincorporation and corresponding gel showing only the +1 cognate GTP incorporation product in 15-s burst assays with up to 160 μM GTP. The ACUG lanes reflect cognate elongation to +10 product and this forms a very stable hairpin that does not denature on the gels and therefore migrates faster than the misincorporation product RNAs.

individual nsp7 and nsp8 subunits (Fig. 5B). With the increasing amounts of +5 product generated at higher CTP concentrations, we also observe a second misincorporation that we attribute to an ATP:C mismatch, i.e., adenosine in place of guanosine, to produce a +6 species that is further elongated to a +9 AACcAaACA product (Fig. 5A). We did not observe the terminal ATP:C mismatch that would generate the full-length +10 product and attribute this to the instability of the elongation complex when there are no downstream nucleotides remaining in the nsp12 template entry channel. We quantitated the level of misincorporation using three gel bands: the +3 band reflecting initiation and cognate AAC addition, the +5 band that requires a CTP:A mismatch at the +4 position, and the +9 band that requires an ATP:C mismatch at the +6 position. At the molecular level, these misincorporation reactions occur in the context of processive elongation because each replicase complex begins with the original hairpin primer, and any intermediate +3 or +5 RNA products that dissociate prior to mismatch incorporation are unlikely to reform new elongation complexes within the 15-s reaction time. Using a similar self-priming RNA hairpin in which the cognate product was GAGACCAUUC, we observed only the cognate +1 product and did not detect any GTP:U mismatch incorporation (Fig. 5C).

We then compared these nsp12 misincorporation levels with those obtained from the picornaviral poliovirus and coxsackievirus B3 3D^{pol} enzymes. For a direct comparison with the nsp12 data, these experiments were done by preincubating 3D^{pol} and hairpin RNA and then adding the required nucleotides for processive elongation and

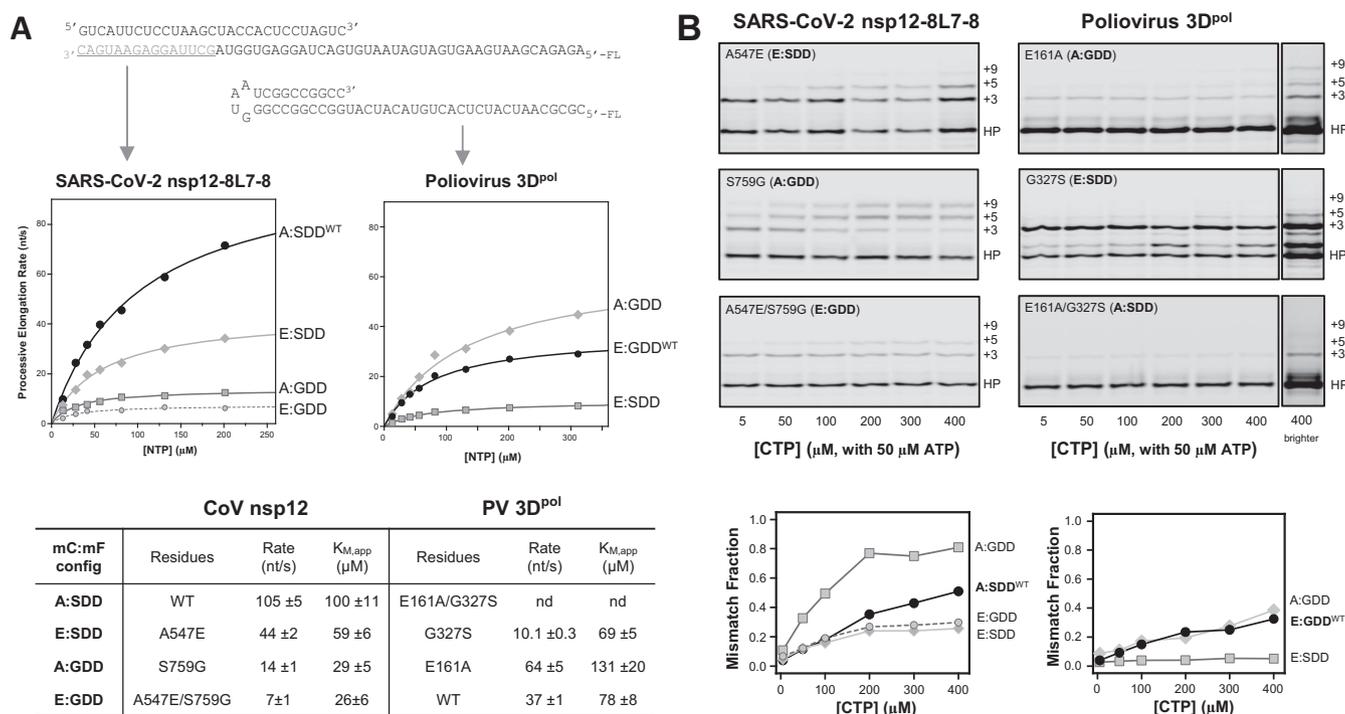


FIG 6 Characterization of polymerase active site motifs F and C mutants. (A) Primer-template RNAs and NTP concentration dependence of processive elongation by nsp12-8L7-8 complex and PV 3D^{pol} polymerases at 30°C, pH 7.0. (B) Misincorporation gels and corresponding mismatch incorporation plots for nsp12 and PV 3D^{pol} polymerases. For the 3D^{pol} data, the 400 μM CTP lane images were duplicated and brightened to better show weak misincorporation signals: The E161A mutant (A:GDD) did not initiate efficiently, which reduced overall product formation, but significant misincorporation was observed in the little product that was formed. In contrast, the G327S mutant (E:SDD) efficiently initiated to yield +3 product, but very little of that was further elongated by misincorporation. The double mutant (A:SDD) showed even less initiation, and no CTP concentration dependent misincorporation products were detected.

misincorporation; unlike in most of our prior studies, we did not preincubate with the first nucleotide(s) to generate stalled elongation complexes prior to elongation. Both 3D^{pol} enzymes also showed CTP concentration-dependent CTP:A misincorporation leading to the +5 product, and the subsequent ATP:C misincorporation leading to the +9 product (Fig. 5A). CV 3D^{pol} showed higher fidelity than PV 3D^{pol} (Fig. 5B), which is consistent with prior comparisons of GTP:U mismatch incorporation by the picornaviral enzymes (16). One particularly interesting observation from these data is that the amount of mismatch products made by nsp12 and poliovirus 3D^{pol} are roughly comparable, especially at lower CTP concentrations (Fig. 5B), suggesting that there are features of the nsp12 active site that maintain fidelity despite its significantly faster elongation rate.

Comparison with a canonical RdRP active site. To address the origin of this fidelity compensation, we generated individual A547E and S759G mutations in nsp12, and conversely E161A and G327S mutations in poliovirus 3D^{pol} (Fig. 1C and D), and analyzed their effects on processive elongation with rapid stopped-flow kinetics and on mismatch incorporation with burst-phase assays (Fig. 6). Introducing the motif F glutamic acid into the nsp12 active site reduced the maximal elongation rate 2.4-fold from 105 to 44 nt/s, the SDD-to-GDD mutation reduced rate by a greater 7.5-fold factor to only 14 nt/s, and the double mutation reduced it 15-fold to 7 nt/s (Fig. 6A). These mutations also decreased the NTP K_{m,app} values 2- to 3-fold, from 100 μM to 59, 29, and 26 μM. Fidelity increased similarly with both the A547E mutation and double mutations based on reductions in both CTP:A and ATP:C mismatches and decreased with S759G based on significantly higher levels of CTP:A mismatch incorporation (Fig. 6B). The plot in Fig. 6B reflects CTP:A mismatch events, and an analysis of only the ATP:C mismatch, i.e., +9 band divided by sum of +5 and +9 bands, showed similar effects

with $26\% \pm 2\%$ ATP:C product for wild-type nsp12 that decreased to $11 \pm 1\%$ with the higher fidelity A547E mutation, increased to $40 \pm 3\%$ with the lower fidelity S759G mutation, and decreased to $3 \pm 1\%$ with the double mutation. Thus, the two non-canonical residues in the native nsp12 active site have complementary functions that both benefit large-genome synthesis, with Ala547 accelerating replication rate and Ser759 improving fidelity.

In the context of poliovirus 3D^{pol}, an E161A mutation in motif F increased the processive elongation rate 1.6-fold and increased $K_{m,app}$ (Fig. 6A), effects that were opposite of those observed with the reverse mutation in nsp12, and again showing that the smaller alanine in motif F yields a faster polymerase. Interestingly, the E161A mutation also significantly destabilized the 3D^{pol}-RNA complex, resulting in lower signal amplitudes in the raw stopped-flow data (not shown) and inefficient initiation in the burst-phase elongation experiments to assess fidelity (Fig. 6B). The observed misincorporation level of the E161A mutant was comparable to that of wild-type 3D^{pol}, but note that this may be an underestimate of the true misincorporation efficiency because the unstable 3D^{pol} complexes may dissociate from the +3 product RNA faster than they misincorporate to generate the +4 and longer species. Last, the introduction of SDD into 3D^{pol} motif C via the G327S mutation slowed processive elongation 3.7-fold and significantly reduced misincorporation levels, with CTP:A being reduced 8-fold and ATP:C mismatch products being barely detectable. This is a clear increase in fidelity that is consistent with the higher fidelity associated with SDD over GDD in nsp12. We also attempted assays with the double mutant 3D^{pol}, but it was even less stable than the E161A single mutant. We could not obtain any elongation signals in the stopped-flow assay and barely detected the cognate +3 product in the burst assays (Fig. 6B), precluding any misincorporation assessment.

Overall, our data show that the nsp12-nsp7/8-nsp8 core CoV replicase complex could be efficiently assembled by using a nsp8L7 fusion protein in place of the nsp7/8 heterodimer and preincubating nsp12 and nsp8. This complex is a very fast viral polymerase with elongation rates approaching 300 nt/s at 37°C, and active-site mutants show this fast rate is enabled by nsp12 having an atypical alanine in motif F alongside a SDD sequence within the palm domain motif C that helps maintain fidelity.

DISCUSSION

Virus mutation rates as generated by the molecular replication machinery and the related mutation frequencies as observed in the resulting virus populations are inherently difficult values to determine (35). They reflect a complicated interplay of polymerase fidelity, co- and postsynthesis repair mechanisms, fitness, selection and adaptation effects, RNA modifications by host factors, the biochemical environment of the host cell, and even the molecular methodologies used to detect mutations. Nonetheless, careful comparative analyses indicate that positive-strand RNA virus mutation rates are 200- to 2,000-fold higher than what is seen for DNA-based genome replication (36–38). This largely reflects the lack of proofreading and repair systems in most RNA viruses, and the inherent error rates of DNA polymerase and viral RdRPs are comparable (39). However, the coronaviruses exhibit ≈ 10 -fold higher replication fidelity than most RNA viruses, and this is essential for the survival of these nonsegmented large-genome viruses. While much of this mutation rate reduction can be attributed to the ExoN proofreading enzyme that is unique to the large-genome nidoviruses, it is noteworthy that all nidoviral polymerases have two key active-site mutations that differentiate them from the RdRPs of smaller genome positive-strand RNA viruses. This raises the question of how the nidoviral RdRPs have adapted to enable higher replication rates and what effect this has on inherent polymerase fidelity. In this study, we address this issue by first elucidating a pathway for efficient *in vitro* assembly of the SARS-CoV-2 core replicase complex composed of nsp12-nsp7/8-nsp8 and then using these complexes to assay elongation rates and nucleotide mismatch incorporation levels.

Core replicase assembly is rate limited by the nsp12-nsp8 interaction. The assembly of functional CoV replicase complexes is inefficient, leading to seemingly low

elongation activity on a time scale of tens of minutes in assays in which all the protein, RNA, and NTP components are mixed simultaneously. High molar excess concentrations of nsp7 and nsp8 are often used in studies to increase efficiency, but even then, the assembly is not complete based on large amounts of unused primer. However, the subpopulation of RNA that is assembled into complexes is rapidly elongated in quench-flow and single-molecule experiments (22–24). These data underscore that CoV initiation is a multistep process that is at a minimum composed of slow assembly of the (nsp12-nsp7/8-nsp8-RNA) core replicase complex followed by rapid elongation. To separate these two confounding biochemical activities, we first sought to identify the slow step in the core replicase assembly process.

Building on the prior nsp7L8 fusion version of the nsp7/nsp8 heterodimer (1) and considering the now available structures of the CoV core replicase, we reversed the order of the two proteins to generate a single-chain nsp8L7 protein mimicking the nsp8/nsp7 dimer bound to nsp12. This greatly enhanced assembly of a highly active core replicase, and based on the order of addition data, we propose a model in which initial formation of a nsp12-nsp8 complex is the rate-limiting factor in replicase assembly that is followed by the rapid addition of nsp8L7 and RNA. Our finding that a nsp12-nsp8 complex forms first is consistent with the observation that cotranslational cleavage of a nsp7-nsp8-nsp12 fusion protein in insect cells yielded a purifiable nsp12-nsp8 complex that lacked nsp7 (2, 27). Importantly for *in vitro* experiments, using our two-step protocol with nsp12-nsp8 preassembly allows us to generate active replicase complexes at relatively low and close-to-stoichiometric concentrations of the nsp8 and nsp8L7 cofactor proteins. This minimizes formation of alternative complexes as these cofactors are themselves prone to oligomerization, as seen in our data (Fig. 2B) and characterized previously to yield complexes that may compete for RNA binding and elongate RNA independently of nsp12 (32, 33, 40–42).

nsp12 rate and fidelity. Stopped-flow rapid kinetics showed nsp12 elongation rates of almost 300 nt/s, making the CoV replicase the fastest viral RdRP characterized to date (Fig. 4C). These rates are consistent with the ≈ 170 nt/s observed at 37°C in magnetic tweezers studies (22, 25) and quench-flow experiments using both SARS-1 and SARS-2 replicases (23, 24). Notably, they are 3-fold faster than the ≈ 90 nt/s rates observed for poliovirus 3D^{pol} and 15-fold faster than the ≈ 20 nt/s coxsackievirus 3D^{pol} that share the same core structure and active-site closure mechanism (16, 34). The two nsp8 sliding pole helices do not significantly affect processive elongation rates (Fig. 4D), but their deletion reduced RNA binding (Fig. 2C) and significantly lowered amplitudes in the stopped-flow studies (Fig. 4D), consistent with the sliding poles acting as processivity factors that stabilize the elongation complex.

To determine whether there are coronavirus-specific features responsible for the remarkably high nsp12 RNA synthesis rate, we examined how the noncanonical Ala547 instead of a glutamate in motif F and Ser759 instead of a glycine in motif C affected rate and fidelity. Stopped-flow elongation data showed that both residues make significant contributions to the high replication rate of nsp12, with individual A547E and S759G mutations reducing elongation rates 2- and 7-fold, respectively. The two mutations had opposite effects on fidelity, which increased with the A547E mutation and decreased significantly with the S759G mutation (Fig. 6B). The S759G mutation provided a very clear indication of reduced fidelity in that it increased misincorporation levels almost 2-fold while reducing the processive elongation rate 7-fold, implying a ≈ 14 -fold decrease in nucleotide specificity. These data reflect misincorporation during processive elongation because given the slow core replicase assembly rate, the 15-s reaction time in our burst assays is too short to allow for multiple turnover reinitiation via RNA release and rebinding.

We then converted a canonical RdRP active site into a CoV type active site by making the opposite mutations in poliovirus 3D^{pol}. Mutating the motif F glutamate to alanine (E161A) increased 3D^{pol} elongation rate 1.6-fold, which is consistent with rate reduction observed with the opposite A547E mutation in nsp12. The observed fidelity effect arising from E161A was minor (Fig. 6B), but note that this may reflect an

underestimate of the true fidelity because the mutation itself significantly destabilized the 3D^{pol}-RNA complex. The RNA dissociation rate may therefore be comparable to the misincorporation rate, which would reduce the amount of misincorporated product observed in these burst-phase assays. The GDD to SDD mutation in 3D^{pol} showed a ≈ 3.7 -fold reduction in processive elongation rate and an ≈ 8 -fold reduction in misincorporation levels, suggesting it is a higher fidelity polymerase. However, as both effects are rate reductions and the inferred increase in nucleotide specificity is only ≈ 2 -fold, we cannot clearly establish from the burst-phase data if the reduced misincorporation level is a true fidelity effect or simply a reflection of slower kinetics.

Polymerase adaptations for large-genome replication. RNA virus genomes are limited by the tradeoffs inherent in maintaining genetic integrity in the context of replication by low-fidelity RNA-dependent RNA polymerases, and consequently the majority of RNA viruses have genomes in the 4- to 12-kb range. Larger genomes require more time for replication, but using a faster RdRP can be counterproductive because it tends to also increase mutation rates. Thus, the expansion of RNA genomes past the ≈ 12 -kb threshold typified by most flaviviruses requires both an increase in replication rate and a decrease in mutation rate, the latter of which could be achieved by either improving RdRP fidelity or by acquiring error-correcting functionality (12, 13, 37). The nidoviruses have undergone such an expansion, resulting in genomes of up to 32 kb, the existence of which is largely attributed to the acquisition of an error-correcting exonuclease (ExoN) enzyme. However, ExoN is present only in large-genome (>20 kb) nidoviruses, i.e., the coronaviruses and toroviruses, while smaller genome arteriviruses and some flavivirus-like viruses have managed to expand their genomes to 13 to 16 kb and 27 kb, respectively, without an ExoN (37). This suggests that the inherent speed and fidelity of the core viral RdRP platform can be modulated to allow for some expansion of the viral genome without adding proofreading error correction.

At the structural level, the positive-strand RNA virus polymerases share a core fold and palm-domain active-site closure mechanism that is based on a subtle movement of conserved motif A (10, 14, 43, 44). The active site itself includes three aspartates that are essential for magnesium coordination and catalysis and therefore present in all replicative polymerases (45). Beyond this, the exact amino acid sequence determines elongation rate and fidelity, with particularly strong effects arising from substitutions within motifs A and D (16) and more subtle effects coming from extended protein dynamics networks within the fingers domain (46–48). This led to a model in which the palm domain-based active-site closure mechanism provides a structural framework upon which individual viruses can evolutionarily fine tune RdRP function to optimize fitness (16, 18, 49). Outside of the immediate catalytic center, viral RdRPs commonly retain important residues that trigger the closure mechanism or interact directly with the RNA and NTP substrates. Among the latter is the glycine in the GDD sequence of the motif C β -hairpin structure that is located below the priming nucleotide ribose and a glutamate-arginine salt bridge in motif F that is positioned directly above the bound NTP α/β -phosphates (Fig. 1C). Interestingly, the nidoviral polymerases have different residues at both sites, leading us to propose a two-step process whereby the core RdRP structure and mechanism has adapted to support large-genome replication.

The molecular mechanism underpinning the effects of these mutations is best explained by first considering the catalytic geometry of picornaviral polymerases, in which high-resolution crystal structures have mapped out the entire RdRP nucleotide addition cycle (15). RNA-templated polymerases, i.e., RdRPs and reverse transcriptases, have a conserved arginine within motif F and the ring finger structure that is the structural homolog of an O-helix lysine found in DNA templated single subunit polymerases (14). Multiple structures show the arginine positioned directly above the NTP α/β -phosphate linkage in precatalytic RdRP-RNA-NTP complexes (14, 50–52), where it imposes an electrostatic positioning constraint on the NTP α -phosphate that will favor catalysis when a correctly base-paired NTP is bound in the active site (Fig. 1C). In the picornaviral RdRPs, this arginine (Arg174) is held in place by a salt bridge to a glutamate (Glu161) on the other β -strand of the ring finger, as shown in Fig. 1C

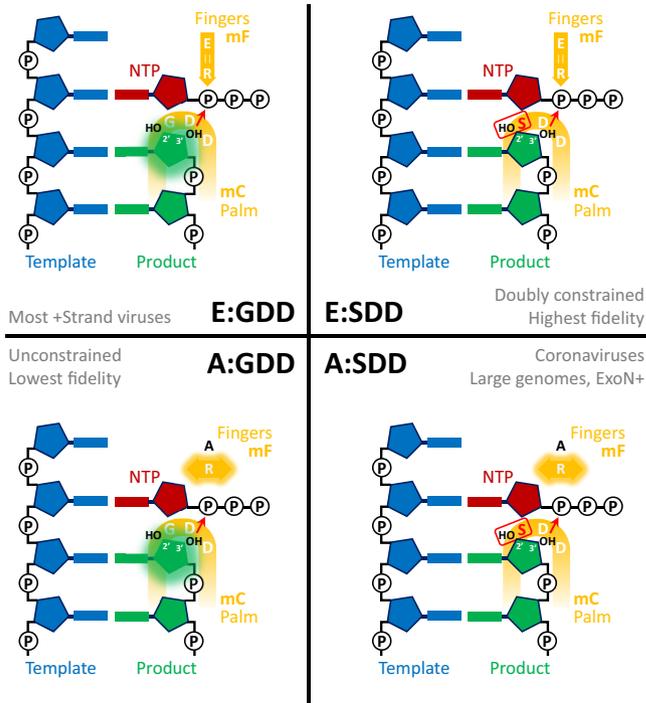


FIG 7 Schematic illustrations of the different RdRP active site configurations. Most positive-strand RNA virus RdRPs have an E:GDD configuration in which the glutamate-arginine interaction within motif F (E = R in yellow arrow) positions the arginine above the NTP α -phosphate for catalysis, while the GDD motif in motif C makes minimal interactions with the priming ribose group. This allows for some flexibility in primer positioning (indicated by the green halo) that can reduce nucleotide discrimination and fidelity. In contrast, coronaviruses have an A:SDD configuration, in which the motif F arginine is now more flexible (yellow halo), while the primer is held more rigid by the motif F serine interactions with the 2' hydroxyl of the priming ribose (red rectangle). The E:SDD configuration constrains both the arginine and the ribose, resulting in a high-fidelity enzyme, while the A:GDD configuration does not constrain either and results in a low-fidelity enzyme.

and schematically illustrated in the E:GDD panel of Fig. 7. Notably, this glutamate-arginine guanidinium group interaction is well ordered in PV 3D^{Pol} even in the absence of bound RNA and/or NTP (53). Removal of the salt bridge via the E161A mutation in poliovirus polymerase increases the elongation rate (Fig. 6A), and this likely arises from a combination of increased arginine flexibility, schematized by the halo around the R in the A:GDD panel of Fig. 7, and faster loading of the NTP into the active site; the rigid Glu-Arg salt bridge is also a steric barrier that the incoming NTP needs to physically bypass in order to base-pair with the templating nucleotide.

In this context, it is notable that all the nidoviral polymerases have two key active-site mutations: loss of the motif F glutamate and conversion of the palm domain motif C GDD sequence into an SDD. Comparing the 46 currently available nsp12 structures via the *coro* set of polymerase superpositions (9, 10) shows the motif F arginine (Arg555) adopting a wide range of positions in both the absence and presence of RNA and NTP (Fig. 1C). We propose that it is this inherent flexibility and lack of steric constraints associated with the Glu-Arg salt bridge that enables rapid NTP loading into the active site and fast elongation rates. The inherent molecular tradeoff of this fast elongation rate is lower fidelity due to fewer constraints on the NTP α -phosphate positioning during catalysis. To overcome this obstacle, the nidoviruses have established a new fidelity checkpoint by using a SDD instead of GDD sequence within motif C in the palm domain. The cryo-EM structures of several nsp12-RNA complexes show the SDD serine side-chain hydrogen bonding with the 2' hydroxyl of the priming nucleotide (Fig. 1C) (54–59), leading us to propose that the mechanistic explanation for the improved fidelity of SDD is a reinforcement of correct priming ribose positioning in the

active site. In effect, using SDD in motif C shifts a fidelity checkpoint from NTP positioning to primer 3' end positioning, and this allows for dramatic increases in elongation rate by releasing structural constraints on the motif F arginine. Conversely, having both fidelity control points present at the same time, as in the PV G327S mutant (Fig. 6B), results in a very high-fidelity enzyme, but also a slow enzyme, and consequently any virus carrying this combination would have significantly reduced growth rate. We speculate that this E:SDD configuration enzyme is slow because of conflicting interactions that independently constrain the positions of the NTP and the priming ribose into different sites, leading to a combined geometry that is less than ideal for catalysis (Fig. 7).

When considering the polymerase adaptations in going from short- to long-genome RNA viruses, it is also insightful to consider the intermediate length 13- to 16-kb arterivirus RdRPs. They have a Q:SDD configuration with a motif F glutamine that is a subtle mutation from the canonical glutamate and would likely interact similarly with the arginine, but the interaction would be weaker by virtue of being nonionic. The prediction of our model is that replication rate would increase, and this would then enable replication of slightly longer genomes in a time frame that can escape innate immune pressure, but fidelity would decrease because of the less restricted active-site geometry. We propose that it is this fidelity constraint that led to the subsequent emergence of the SDD sequence in motif C and shifting the fidelity checkpoint away from NTP α -phosphate positioning and toward priming ribose positioning. The presence of SDD in conjunction with a slight relaxation of the arginine interaction via a glutamine, e.g., Q:SDD, allowed for stable 13- and 16-kb arterivirus genomes. The jump to even larger genomes would require even faster polymerases, which could now be easily achieved by further loosening the constraints on the motif F arginine, resulting in the minimal alanine residue and A:SDD configuration found in the 26- and 32-kb coronaviruses and toroviruses or S:SDD found in roniviruses. However, at these very fast polymerase rates, the SDD fidelity checkpoint would no longer be sufficient for maintaining genome integrity, and therefore, these viruses also acquired ExoN domains to carry out postcatalysis error correction as a separate step. From the studies presented here, we cannot infer whether the arteriviral polymerases evolved from the picornaviral-like supergroup in parallel with the large-genome nidoviruses or were an intermediate on the pathway to large-genome evolution. However, the functional data led us to favor a model in which the relaxation of the interactions with the motif F arginine preceded the acquisition of the SDD in motif C; in the context of an ancestral 3D^{pol}-like enzyme, the E:SDD configuration is a slow and very high-fidelity polymerase that would likely reduce virus fitness and limit its emergence.

Summary. In this study, we addressed the molecular and mechanistic basis for the high speed and fidelity of the SARS-CoV-2 core replicase. Using a nsp8L7 fusion protein, we show that assembly of the (nsp12-nsp7/8-nsp8) core replicase complex is rate limited by the binding of the individual nsp8 molecule to the nsp12 polymerase fingers domain. The resulting complex is the fastest viral RdRP we have observed to date, with a processive elongation rate of 260 nt/s at 37° that is 3-fold faster than that of poliovirus polymerase. We have identified two noncanonical residues in the nsp12 active site that tailor the CoV enzyme to carry out large-genome replication: First, Ala547 in motif F replaces a canonical glutamate and allows for a doubling of replication rate by untying Arg555 in the active site, but this comes at a fidelity cost due to less restrictive NTP positioning during catalysis. Second, Ser759 in motif C restores fidelity by positioning the priming ribose for catalysis, in effect creating a new fidelity control point to compensate for the loss of NTP positioning. These structural and biochemical interactions suggest there were three major steps in the evolution of the nidovirus polymerase machinery (Fig. 7): (i) mutation away from the motif F glutamate to slightly increase elongation rates, (ii) the adoption of a SDD instead of GDD sequence in motif C as new fidelity control point via primer positioning, and (iii) further evolution of even faster elongation rates by adopting a motif F alanine that fully untethers the active-site arginine. The first two of

these events allow for stable maintenance of 13- to 16-kb arterivirus genomes, while the third enabled the rapid replication of larger 26- to 32-kb genomes but also required the acquisition of more powerful error-correcting ExoN functionality to maintain genomic stability. These findings further demonstrate ways in which the core structure and active-site closure mechanism shared by all positive-strand RNA virus polymerases can be evolutionarily tuned to optimize virus fitness.

MATERIALS AND METHODS

Protein expression and purification. A T7-based pET-26 plasmid with a N-terminal His₈ tag followed by a TEV cleavage site was used to express codon optimized versions of the Wuhan-1 strain SARS-CoV-2 proteins. The nsp8L7 fusion proteins were made by adding a GSGSGS linker and nsp7 to the C terminus of nsp8. All proteins were expressed in *E. coli* BL21(DE3) CodonPlus cells (Agilent Technologies) and grown at 37°C to an optical density at 600 nm (OD₆₀₀) of 0.8 prior to induction with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG). Nsp8, nsp7, and nsp8L7 and all N-terminal deletion constructs were expressed at room temperature for 18 h in NZCYM broth. Harvested cells were lysed at 18,000 lb/in² using a M-110L microfluidizer (Microfluidics Corp.) in a buffer containing 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol. The lysate was centrifuged at 17,000 rpm in a Beckman JA-20 rotor, and the supernatant was passed through 0.45- and 0.20- μ m filters prior to being loaded onto a 5-mL HisTrap FF column (Cytiva) followed by step elution with 50 mM Tris, pH 8.0, 300 mM NaCl, 350 mM imidazole. The protein-containing fractions were pooled, Tris(2-carboxyethyl) phosphine hydrochloride (TCEP) reducing agent was added to 1 mM, the sample was concentrated and then injected onto a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in 200 mM NaCl, 50 mM Tris, pH 7.0, and 20% glycerol. The eluted proteins were concentrated to 200 μ M and stored at -80°C.

Nsp12 and mutant polymerases were expressed in LB broth at 16°C for 18 h. The harvested cells were resuspended in 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol supplemented with 1% CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate) and lysed using the same method as above. The lysate was also loaded onto a 5-mL HisTrap FF column and step eluted in 50 mM Tris, pH 8.0, 300 mM NaCl, 350 mM imidazole. The protein-containing fractions were pooled, TCEP was added to 1 mM, and the sample was diluted 4-fold in a low-salt buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 20% glycerol) to reduce the ionic strength for ion-exchange chromatography on a MonoQ 10/100 GL column (Cytiva). Two sets of fractions containing nsp12 were pooled (Fig. 2A), concentrated to 30 μ M, and stored at -80°C in single-use 10- μ L aliquots. Poliovirus and coxsackievirus polymerases were expressed and purified as previously described (60, 61).

RNA selection criteria. For the experiments described below, we have used different RNA designs to accommodate conflicting experimental constraints. For fluorescence polarization (FP)-based RNA-binding studies, we chose a short primer-template hairpin that provides both duplex and single-stranded RNA regions while being small enough to elicit a significant FP change upon protein binding. For the assembly and fidelity studies, we chose hairpin RNAs that would provide a stable primer-template duplex for initiation, yet were small enough to fully denature on gels so as to provide clear single-base resolution that was not complicated by internal base pairing that can lead to smearing. For the stopped-flow kinetic assays, we needed long-lived RdRP-RNA complexes that could be diluted to \approx 50 nM before being loaded into the instrument for 10 to 12 sequential sample injections (Fig. 4) over a 3- to 5-min time period without significant dissociation. For the picornaviral polymerases, we can preincorporate a few nucleotides to form stable stalled elongation complexes with very slow k_{off} rates, but this does not work for CoV nsp12 (24), and we instead used a much longer RNA duplex section to gain stability via making full contact with the nsp8 sliding pole helices.

RNA binding. Fluorescence polarization (FP) RNA-binding assays were performed using a synthetic RNA hairpin with a 5'-GAAUGGUCUC-3' single-stranded section labeled at its 5' end with a fluorescein dye via a 6-carbon linker. Protein titrations were done with 20 nM RNA in binding buffer consisting of 50 mM HEPES (pH 7.0), 50 mM NaCl, 1 mM MgCl₂, 4 mM TCEP, and 0.1% (vol/vol) NP-40 detergent. FP was measured in black 384-well polystyrene plates using a Perkin-Elmer Victor³ multimode microplate reader. RNA dissociation constants were obtained by curve fitting the FP data to the quadratic form of the single-site binding isotherm.

Core replicase assembly. Assembly and initiation experiments were conducted by mixing final concentrations of 2.5 μ M nsp12, 5 μ M each nsp8 and nsp8L7, and 1 μ M RNA in 50 mM HEPES, pH 7.0, 62 mM NaCl, 1 mM MgCl₂, and 1 mM TCEP with various orders of addition and incubation times. The RNA for this assay is a synthetic (Integrated DNA Technologies, Coralville, IA) 8-base pair priming hairpin followed by various single-stranded template sequences as indicated in the figures. The RNA hairpin was labeled at its 5' end with an IRdye 800RS NHS ester (Li-Cor Biosciences). The proteins were preincubated as indicated at 2 \times final concentrations and mixed with remaining components at $t = 0$, and time point samples were taken thereafter. Each time point is a 15-s elongation reaction with 50 μ M NTPs (further diluting the replicase sample 2-fold) to assay for the amount of competent complex present at the specified assembly time. The reactions were manually quenched with 50 mM EDTA, 95% formamide, 0.02% bromophenol blue, and 0.02% xylene cyanol and analyzed by denaturing polyacrylamide gel electrophoresis using 7 M urea, 17% 19:1 acrylamide, 1 \times Tris-borate-EDTA (TBE) gels. Resolved RNA bands were imaged and quantified with a Li-Cor Odyssey CLx IR imager system and plotted as a ratio of elongated product species to the total amount of RNA.

Core replicase stability. The temporal stability of the nsp12-nsp8 complex was assessed with a large scale 300 μ L assembly reaction of 20 μ M nsp12 with 60 μ M nsp8 and using gel filtration chromatography in 100 mM NaCl, pH 7, 10% glycerol to purify the complex as a 0.5-mL sample from the leading edge of the peak at \approx 12-min retention volume. This purified complex was incubated at room temperature, and 50- μ L aliquots were periodically reinjected onto the column over an 18-h time course (Fig. 3E). The areas of the overlapping nsp12-8 and nsp12 peaks in the resulting chromatograms were determined by mathematical peak fitting (PeakFit, Systat Software) and corrected for their different extinction coefficients (<http://web.expasy.org/protparam>) to calculate the relative concentrations present in the loaded sample aliquots.

Elongation kinetics. Stopped-flow fluorescence was used to measure kinetics of processive elongation on a three-piece annealed RNA with a 29-bp duplex primer and a 26-nt single-stranded template with a fluorescein placed at the 5' end of the template. RNAs were synthesized by Integrated DNA Technologies (Coralville, IA) and annealed by mixing all three strands and heating to 70°C followed by slow cooling at 1°/min for 60 min in 66 mM KCl, 1 mM Tris, pH 8.0, and 0.1 mM EDTA. Replicase elongation complexes were generated by incubating 8 μ M nsp12 with 10 μ M nsp8 for 45 min at room temperature. 10 μ M nsp8L7 was incubated with 6 μ M RNA for 15 min prior to mixing 1:1 with the nsp12/8 complex and further incubated for 15 min. The replicase mix was then diluted 60-fold into 75 mM NaCl, 4 mM MgCl₂, 25 mM HEPES, pH 7.0, 1 mM TCEP and immediately loaded into the Bio-Logic SFM-4000 titrating stopped-flow instrument with a MOS-500 spectrometer, and kinetics data were collected using equimolar mixtures of all four NTPs at various concentrations and at various temperatures. Fluorescence data from the elongation reactions were fit to an increasing exponential function with a lag phase using Kaleidagraph (Synergy Software) as described (34). Elongation rates were calculated from the length of the observed lag phase as representing elongation over 24 of the 26 nucleotides in the ssRNA template, i.e., fluorescence change occurs after translocation past the third nucleotide from the 5' end, and plotted against NTP concentration to determine processive elongation rates and NTP $K_{m,app}$ values. PV 3D^{pol} stopped-flow elongation kinetics were performed using stalled +1 elongation complexes as previously described (34). All NTPs were purchased as 100 mM solutions from Jena Biosciences and diluted into single-use 1 and 10 mM aliquots stored at -80°C .

Misincorporation assays. 5 μ M nsp12 was incubated with 10 μ M nsp8 in 50 mM HEPES, pH 7.0, 75 mM NaCl, 1 mM MgCl₂, and 1 mM TCEP for 30 min and then mixed 1:1 with 10 μ M nsp8L7 with 2 μ M RNA in 50 mM HEPES, pH 7.0, 50 mM NaCl, 1 mM MgCl₂, and 1 mM TCEP that had been preincubated for 15 min. Replicase was assembled for an additional 15 min prior to the addition of 50 μ M ATP and various concentrations of CTP for 15 s before manual quenching with 50 mM EDTA in 95% formamide, 0.02% bromophenol blue. For the experiments with poliovirus and coxsackievirus polymerases, 5 μ M 3D^{pol} enzymes were preincubated with 1 μ M RNA for 15 min prior to initiating burst elongation using the same buffers, nucleotide mixtures, and reaction times as for the assembled nsp12 complex. Note that we did not preinitiate to generate stalled 3D^{pol} elongation complexes for these experiments, i.e., the nsp12 and 3D^{pol} assays are done under identical conditions requiring first nucleotide addition followed by processive elongation and misincorporation. Monovalent salts can significantly affect RNA binding, and the salt concentrations listed above include accounting for any salt introduced with the proteins whose purification buffers contained 200 to 300 mM NaCl. Quenched elongation samples were resolved by denaturing polyacrylamide gel electrophoresis using 7 M urea, 17% 19:1 acrylamide, and 1 \times TBE and imaged with a Li-Cor Odyssey CLx IR imaging system. Misincorporation was quantified by integrating the RNA band intensities and calculating the ratio of misincorporation products to the total amount of elongated products, i.e., all cognate and noncognate incorporations, and plotting them against the CTP concentrations. The quantitation for CTP:A mismatch is based on the sum of the +5 and +9 products divided by the total RNA elongated, i.e., sum of the +3, +5, and +9 products, and for ATP: C, it is based on the +9 product divided by the sum of the +5 and +9 products, i.e., what fraction of the +5 is further elongated to yield +9. Both quantitation schemes analyze only the material that was elongated in the burst-phase assay.

ACKNOWLEDGMENTS

This work was funded by National Institutes of Health grant R01 AI059130 to OBP and grant agreement 101005077 from the Innovative Medicines Initiative 2 Joint Undertaking (JU) (IMI-CARE) (B.C.).

REFERENCES

- Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ, Canad B, Imbert I. 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci U S A* 111:E3900-9. <https://doi.org/10.1073/pnas.1323705111>.
- Gordon CJ, Tchesnokov EP, Feng JY, Porter DP, Gotte M. 2020. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J Biol Chem* 295:4773-4779. <https://doi.org/10.1074/jbc.AC120.013056>.
- Kirchdoerfer RN, Ward AB. 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* 10:2342. <https://doi.org/10.1038/s41467-019-10280-3>.
- Hillen HS, Kocik G, Farnung L, Dienemann C, Tegunov D, Cramer P. 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature* 584:154-156. <https://doi.org/10.1038/s41586-020-2368-8>.
- Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET, Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-

- transcription complex. *Cell* 182:1560–1573.e13. <https://doi.org/10.1016/j.cell.2020.07.033>.
6. Yan L, Ge J, Zheng L, Zhang Y, Gao Y, Wang T, Huang Y, Yang Y, Gao S, Li M, Liu Z, Wang H, Li Y, Chen Y, Guddat LW, Wang Q, Rao Z, Lou Z. 2021. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell* 184:184–193.e10. <https://doi.org/10.1016/j.cell.2020.11.016>.
 7. Yan L, Yang Y, Li M, Zhang Y, Zheng L, Ge J, Huang YC, Liu Z, Wang T, Gao S, Zhang R, Huang YY, Guddat LW, Gao Y, Rao Z, Lou Z. 2021. Coupling of N7-methyltransferase and 3'-5' exonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell* 184:3474–3485.e11. <https://doi.org/10.1016/j.cell.2021.05.033>.
 8. Hillen HS. 2021. Structure and function of SARS-CoV-2 polymerase. *Curr Opin Virol* 48:82–90. <https://doi.org/10.1016/j.coviro.2021.03.010>.
 9. Peersen O. 2019. Polymerase structure alignments. <https://doi.org/10.5281/zenodo.3361874>.
 10. Peersen OB. 2019. A comprehensive superposition of viral polymerase structures. *Viruses* 11:745. <https://doi.org/10.3390/v11080745>.
 11. Perry JK, Appleby TC, Bilello JP, Feng JY, Schmitz U, Campbell EA. 2021. An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15. *J Biol Chem* 297:101218. <https://doi.org/10.1016/j.jbc.2021.101218>.
 12. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res* 117:17–37. <https://doi.org/10.1016/j.virusres.2006.01.017>.
 13. Lauber C, Goeman JJ, Parquet MdC, Nga PT, Snijder EJ, Morita K, Gorbalenya AE. 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathog* 9:e1003500. <https://doi.org/10.1371/journal.ppat.1003500>.
 14. Gong P, Peersen OB. 2010. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc Natl Acad Sci U S A* 107:22505–22510. <https://doi.org/10.1073/pnas.1007626107>.
 15. Gong P. 2021. Structural basis of viral RNA-dependent RNA polymerase nucleotide addition cycle in picornaviruses. *Enzymes* 49:215–233. <https://doi.org/10.1016/bs.enz.2021.06.002>.
 16. Campagnola G, McDonald S, Beaucourt S, Vignuzzi M, Peersen OB. 2015. Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *J Virol* 89:275–286. <https://doi.org/10.1128/JVI.01574-14>.
 17. Vignuzzi M, Wendt E, Andino R. 2008. Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* 14:154–161. <https://doi.org/10.1038/nm1726>.
 18. Gnadig NF, Beaucourt S, Campagnola G, Borderia AV, Sanz-Ramos M, Gong P, Blanc H, Peersen OB, Vignuzzi M. 2012. Coxsackievirus B3 mutator strains are attenuated *in vivo*. *Proc Natl Acad Sci U S A* 109:E2294–303. <https://doi.org/10.1073/pnas.1204022109>.
 19. Rozen-Gagnon K, Stapleford KA, Mongelli V, Blanc H, Failloux AB, Saleh MC, Vignuzzi M. 2014. Alphavirus mutator variants present host-specific defects and attenuation in mammalian and insect models. *PLoS Pathog* 10:e1003877. <https://doi.org/10.1371/journal.ppat.1003877>.
 20. Fitzsimmons WJ, Woods RJ, McCrone JT, Woodman A, Arnold JJ, Yennawar M, Evans R, Cameron CE, Lauring AS. 2018. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol* 16:e2006459. <https://doi.org/10.1371/journal.pbio.2006459>.
 21. Sexton NR, Smith EC, Blanc H, Vignuzzi M, Peersen OB, Denison MR. 2016. Homology-based identification of a mutation in the coronavirus RNA-dependent RNA polymerase that confers resistance to multiple mutagens. *J Virol* 90:7415–7428. <https://doi.org/10.1128/JVI.00080-16>.
 22. Seifert M, Bera SC, van Nies P, Kirchoerfer RN, Shannon A, Le TT, Meng X, Xia H, Wood JM, Harris LD, Papini FS, Arnold JJ, Almo S, Grove TL, Shi PY, Xiang Y, Canard B, Depken M, Cameron CE, Dulin D. 2021. Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife* 10:e70968. <https://doi.org/10.7554/eLife.70968>.
 23. Dangerfield TL, Huang NZ, Johnson KA. 2020. Remdesivir is effective in combating COVID-19 because it is a better substrate than ATP for the viral RNA-dependent RNA polymerase. *iScience* 23:101849. <https://doi.org/10.1016/j.isci.2020.101849>.
 24. Shannon A, Selisko B, Le NT, Huchting J, Touret F, Piorkowski G, Fattorini V, Ferron F, Decroly E, Meier C, Coutard B, Peersen O, Canard B. 2020. Rapid incorporation of favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat Commun* 11:4682. <https://doi.org/10.1038/s41467-020-18463-z>.
 25. Bera SC, Seifert M, Kirchoerfer RN, van Nies P, Wubulikasimu Y, Quack S, Papini FS, Arnold JJ, Canard B, Cameron CE, Depken M, Dulin D. 2021. The nucleotide addition cycle of the SARS-CoV-2 polymerase. *Cell Rep* 36:109650. <https://doi.org/10.1016/j.celrep.2021.109650>.
 26. Gordon CJ, Tchesnokov EP, Schinazi RF, Gotte M. 2021. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J Biol Chem* 297:100770. <https://doi.org/10.1016/j.jbc.2021.100770>.
 27. Gordon CJ, Tchesnokov EP, Woolner E, Perry JK, Feng JY, Porter DP, Gotte M. 2020. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem* 295:6785–6797. <https://doi.org/10.1074/jbc.RA120.013679>.
 28. Snijder EJ, Decroly E, Ziebuhr J. 2016. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv Virus Res* 96:59–126. <https://doi.org/10.1016/bs.aivir.2016.08.008>.
 29. Ahn DG, Choi JK, Taylor DR, Oh JW. 2012. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch Virol* 157:2095–2104. <https://doi.org/10.1007/s00705-012-1404-x>.
 30. Wilamowski M, Hammel M, Leite W, Zhang Q, Kim Y, Weiss KL, Jedrzejczak R, Rosenberg DJ, Fan Y, Wower J, Bierma JC, Sarker AH, Tsutakawa SE, Pingali SV, O'Neill HM, Joachimiak A, Hura GL. 2021. Transient and stabilized complexes of Nsp7, Nsp8, and Nsp12 in SARS-CoV-2 replication. *Biophys J* 120:3152–3165. <https://doi.org/10.1016/j.bpj.2021.06.006>.
 31. Biswal M, Diggs S, Xu D, Khudaverdyan N, Lu J, Fang J, Blaha G, Hai R, Song J. 2021. Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Res* 49:5956–5966. <https://doi.org/10.1093/nar/gkab370>.
 32. Konkolova E, Klima M, Nencka R, Boura E. 2020. Structural analysis of the putative SARS-CoV-2 primase complex. *J Struct Biol* 211:107548. <https://doi.org/10.1016/j.jsb.2020.107548>.
 33. Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M, Rao Z. 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol* 12:980–986. <https://doi.org/10.1038/nsmb999>.
 34. Gong P, Campagnola G, Peersen OB. 2009. A quantitative stopped-flow fluorescence assay for measuring polymerase elongation rates. *Anal Biochem* 391:45–55. <https://doi.org/10.1016/j.ab.2009.04.035>.
 35. Smith EC. 2017. The not-so-infinite malleability of RNA viruses: viral and cellular determinants of RNA virus mutation rates. *PLoS Pathog* 13:e1006254. <https://doi.org/10.1371/journal.ppat.1006254>.
 36. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J Virol* 84:9733–9748. <https://doi.org/10.1128/JVI.00694-10>.
 37. Ferron F, Sama B, Decroly E, Canard B. 2021. The enzymes for genome size increase and maintenance of large (+)RNA viruses. *Trends Biochem Sci* 46:866–877. <https://doi.org/10.1016/j.tibs.2021.05.006>.
 38. Peck KM, Lauring AS. 2018. Complexities of viral mutation rates. *J Virol* 92:e01031-17. <https://doi.org/10.1128/JVI.01031-17>.
 39. Arnold JJ, Cameron CE. 2004. Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg²⁺. *Biochemistry* 43:5126–5137. <https://doi.org/10.1021/bi035212y>.
 40. Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, Ferron F, Gorbalenya AE, Canard B. 2006. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J* 25:4933–4942. <https://doi.org/10.1038/sj.emboj.7601368>.
 41. Te Velthuis AJ, van den Worm SH, Snijder EJ. 2012. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both *de novo* initiation and primer extension. *Nucleic Acids Res* 40:1737–1747. <https://doi.org/10.1093/nar/gkr893>.
 42. Xiao Y, Ma Q, Restle T, Shang W, Svergun DI, Ponnusamy R, Sczakiel G, Hilgenfeld R. 2012. Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J Virol* 86:4444–4454. <https://doi.org/10.1128/JVI.06635-11>.
 43. Ferrero D, Ferrer-Orta C, Verdaguer N. 2018. Viral RNA-dependent RNA polymerases: a structural overview. *Subcell Biochem* 88:39–71. https://doi.org/10.1007/978-981-10-8456-0_3.
 44. Jia H, Gong P. 2019. A structure-function diversity survey of the RNA-dependent RNA polymerases from the positive-strand RNA viruses. *Front Microbiol* 10:1945. <https://doi.org/10.3389/fmicb.2019.01945>.
 45. Steitz TA. 1998. A mechanism for all polymerases. *Nature* 391:231–232. <https://doi.org/10.1038/34542>.
 46. Moustafa IM, Korboukh VK, Arnold JJ, Smidansky ED, Marcotte LL, Gohara DW, Yang X, Sanchez-Farran MA, Filman D, Maranas JK, Boehr DD, Hogle JM, Colina CM, Cameron CE. 2014. Structural dynamics as a contributor to

- error-prone replication by an RNA-dependent RNA polymerase. *J Biol Chem* 289:36229–36248. <https://doi.org/10.1074/jbc.M114.616193>.
47. Moustafa IM, Shen H, Morton B, Colina CM, Cameron CE. 2011. Molecular dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity. *J Mol Biol* 410:159–181. <https://doi.org/10.1016/j.jmb.2011.04.078>.
 48. Shen H, Moustafa IM, Cameron CE, Colina CM. 2012. Exploring the dynamics of four RNA-dependent RNA polymerases by a coarse-grained model. *J Phys Chem B* 116:14515–14524. <https://doi.org/10.1021/jp302709v>.
 49. McDonald S, Block A, Beaucourt S, Moratorio G, Vignuzzi M, Peersen OB. 2016. Design of a genetically stable high fidelity coxsackievirus B3 polymerase that attenuates virus growth *in vivo*. *J Biol Chem* 291:13999–14011. <https://doi.org/10.1074/jbc.M116.726596>.
 50. Shu B, Gong P. 2016. Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proc Natl Acad Sci U S A* 113:E4005–14. <https://doi.org/10.1073/pnas.1602591113>.
 51. Zamyatkin DF, Parra F, Alonso JM, Harki DA, Peterson BR, Grochulski P, Ng KK. 2008. Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *J Biol Chem* 283:7705–7712. <https://doi.org/10.1074/jbc.M709563200>.
 52. Zamyatkin DF, Parra F, Machin A, Grochulski P, Ng KK. 2009. Binding of 2'-amino-2'-deoxycytidine-5'-triphosphate to norovirus polymerase induces rearrangement of the active site. *J Mol Biol* 390:10–16. <https://doi.org/10.1016/j.jmb.2009.04.069>.
 53. Thompson AA, Peersen OB. 2004. Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J* 23:3462–3471. <https://doi.org/10.1038/sj.emboj.7600357>.
 54. Shannon A, Fattorini V, Sama B, Selisko B, Feracci M, Falcou C, Gauffre P, El Kazzi P, Delpal A, Decroly E, Alvarez K, Eydoux C, Guillemot JC, Moussa A, Good SS, La Colla P, Lin K, Sommadossi JP, Zhu Y, Yan X, Shi H, Ferron F, Canard B. 2022. A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. *Nat Commun* 13:621. <https://doi.org/10.1038/s41467-022-28113-1>.
 55. Peng Q, Peng R, Yuan B, Wang M, Zhao J, Fu L, Qi J, Shi Y. 2021. Structural basis of SARS-CoV-2 polymerase inhibition by favipiravir. *Innovation* 2:100080. <https://doi.org/10.1016/j.xinn.2021.100080>.
 56. Naydenova K, Muir KW, Wu LF, Zhang Z, Coscia F, Peet MJ, Castro-Hartmann P, Qian P, Sader K, Dent K, Kimanius D, Sutherland JD, Lowe J, Barford D, Russo CJ. 2021. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc Natl Acad Sci U S A* 118:e2021946118. <https://doi.org/10.1073/pnas.2021946118>.
 57. Bravo JPK, Dangerfield TL, Taylor DW, Johnson KA. 2021. Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. *Mol Cell* 81:1548–1552.e4. <https://doi.org/10.1016/j.molcel.2021.01.035>.
 58. Yin W, Mao C, Luan X, Shen DD, Shen Q, Su H, Wang X, Zhou F, Zhao W, Gao M, Chang S, Xie YC, Tian G, Jiang HW, Tao SC, Shen J, Jiang Y, Jiang H, Xu Y, Zhang S, Zhang Y, Xu HE. 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368:1499–1504. <https://doi.org/10.1126/science.abc1560>.
 59. Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, Ji W, Yan L, Zhu Y, Zhu C, Fang X, Yang X, Huang Y, Gao H, Liu F, Ge J, Sun Q, Yang X, Xu W, Liu Z, Yang H, Lou Z, Jiang B, Guddat LW, Gong P, Rao Z. 2020. Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* 182:417–428.e13. <https://doi.org/10.1016/j.cell.2020.05.034>.
 60. Hobdley SE, Kempf BJ, Steil BP, Barton DJ, Peersen OB. 2010. Poliovirus polymerase residue 5 plays a critical role in elongation complex stability. *J Virol* 84:8072–8084. <https://doi.org/10.1128/JVI.02147-09>.
 61. Campagnola G, Weygandt M, Scoggin K, Peersen O. 2008. Crystal structure of coxsackievirus B3 3Dpol highlights the functional importance of residue 5 in picornavirus polymerases. *J Virol* 82:9458–9464. <https://doi.org/10.1128/JVI.00647-08>.