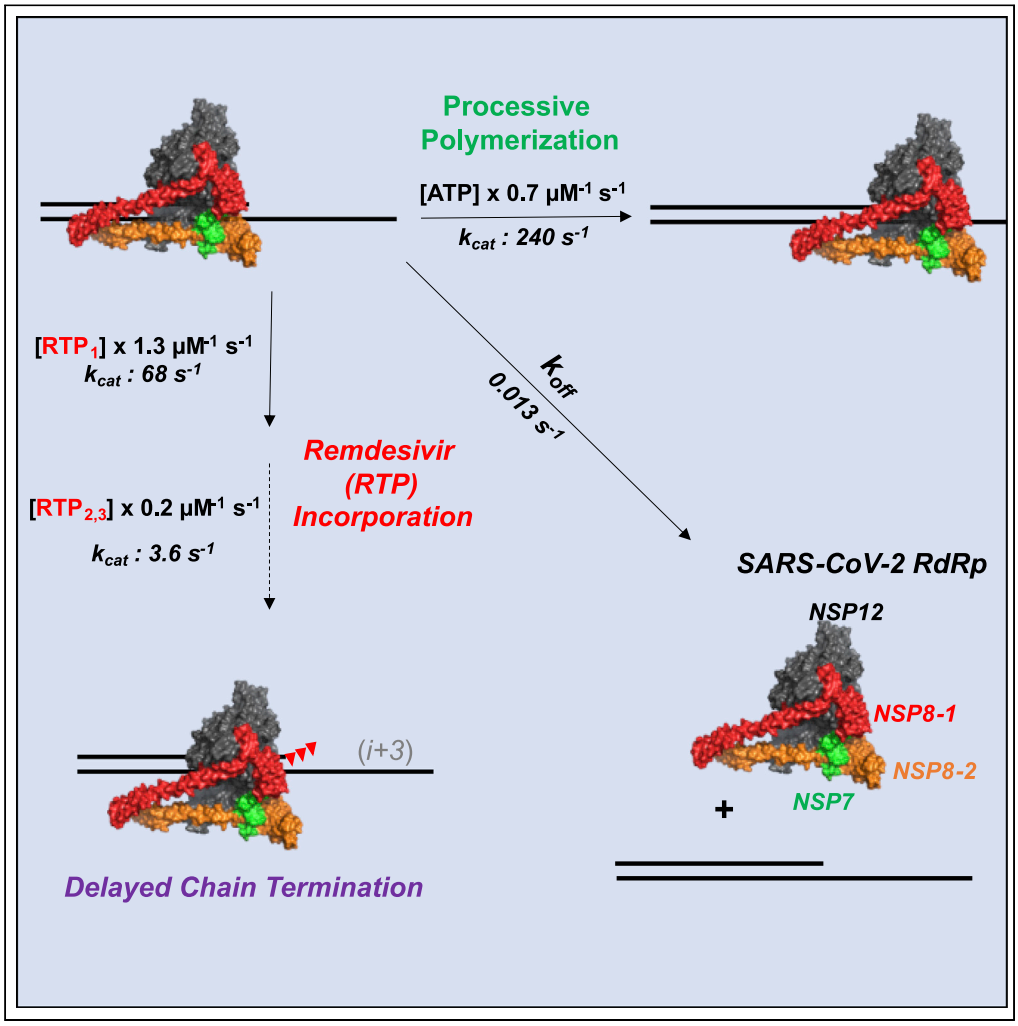
iScience

CellPress
OPEN ACCESS

Article

Remdesivir Is Effective in Combating COVID-19 because It Is a Better Substrate than ATP for the Viral RNA-Dependent RNA Polymerase



Tyler L. Dangerfield, Nathan Z. Huang, Kenneth A. Johnson

kajohnson@utexas.edu

HIGHLIGHTS

Co-expression of NSP12/7/8 with chaperones in *E. coli* gives soluble SARS CoV2 RdRp

Tag-free RdRp complex catalyzes fast and processive RNA polymerization

Polymerization rates are sufficient to replicate the 30 kb genome in 2 min

Remdesivir is incorporated with a specificity constant twice that observed for ATP

Dangerfield et al., iScience *23*, 101849
December 18, 2020
https://doi.org/10.1016/j.isci.2020.101849

iScience

CellPress
OPEN ACCESS

Article

Remdesivir Is Effective in Combating COVID-19 because It Is a Better Substrate than ATP for the Viral RNA-Dependent RNA Polymerase

Tyler L. Dangerfield,[1] Nathan Z. Huang,[1] and Kenneth A. Johnson[1,2,*]

SUMMARY

COVID-19 is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and is currently being treated using Remdesivir, a nucleoside analog that inhibits the RNA-dependent-RNA polymerase (RdRp). However, the enzymatic mechanism and efficiency of Remdesivir have not been determined, and reliable screens for new inhibitors are urgently needed. Here we present our work to optimize expression in *E. coli*, followed by purification and kinetic analysis of an untagged NSP12/7/8 RdRp complex. Pre-steady-state kinetic analysis shows that our reconstituted RdRp catalyzes fast (k_{cat} = 240–680 s^{-1}) and processive (k_{off} = 0.013 s^{-1}) RNA polymerization. The specificity constant (k_{cat}/K_m) for Remdesivir triphosphate (RTP) incorporation (1.29 $\mu M^{-1}s^{-1}$) is higher than that for the competing ATP (0.74 μM^{-1} s^{-1}). This work provides the first robust analysis of RNA polymerization and RTP incorporation by the SARS-CoV-2 RdRp and sets the standard for development of informative enzyme assays to screen for new inhibitors.

INTRODUCTION

A small outbreak of SARS-CoV-2 originated in Wuhan, China in December 2019 (Zhu et al., 2020) and swiftly led to a devastating global pandemic due to the highly contagious and virulent nature of this novel virus. Besides the push to develop a safe and widely available vaccine, work proceeds to develop direct-acting antiviral drugs. The viral RNA-dependent RNA polymerase (RdRp), consisting of non-structural proteins (NSP) 7, 8, and 12, is an important drug target as it catalyzes replication of the viral genome. Viral RNA and DNA polymerases are proven effective targets for the treatment of viral infections (Deval et al., 2014). For example, inhibitors of HIV reverse transcriptase are the cornerstone of multi-drug therapies to treat HIV infections (Hurwitz and Schinazi, 2012), and inhibitors of the hepatitis C virus (HCV) RdRp are central to treating HCV infections (Deval et al., 2014). Although there is a wealth of structural data on the SARS CoV-2 RdRp (Ferron et al., 2018; Hillen et al., 2020; Kirchdoerfer and Ward, 2019; Yin et al., 2020), there are no accurate kinetic studies to establish the mechanistic basis for effective inhibition.

The antiviral compound Remdesivir is an ATP analog, originally developed to treat Ebola infections (Pardo et al., 2020). On 22 October 2020 Remdesivir was approved by the FDA for treatment of COVID-19 based on promising clinical results and initial approval for compassionate use (Grein et al., 2020). It seems unlikely that an ATP analog could be effective because of the high concentrations of competing ATP *in vivo* but Remdesivir seems to show promise despite this hurdle. Therefore, accurate data are required to understand the mechanistic and kinetic basis for the effectiveness of Remdesivir and to use that knowledge to screen for and develop additional antiviral compounds. Although drug screening based on propagation of virus in cell culture has been effective in developing new therapies, this method is expensive and time consuming, making it difficult to screen millions of compounds in the search for new inhibitors. Assays based on analysis of well-defined kinetics of polymerization catalyzed by the viral RdRp *in vitro* could provide a critical component to find new drug candidates. However, preliminary studies attempting to address the kinetics and mechanism of Remdesivir incorporation were flawed because they were conducted using substandard enzyme preparations requiring 30 min of incubation to see a product band on a gel (Gordon et al., 2020b). Moreover, in lieu of accurate rate measurements, a surrogate assay was developed based on fractional extension of a short primer after it was labeled by incorporation of one nucleotide and then

[1]Department of Molecular Biosciences, The University of Texas at Austin, 100 W. 24th Street, Stop 5000, MBB 3.122, Austin, TX 78712, USA

[2]Lead Contact

*Correspondence: kajohnson@utexas.edu

https://doi.org/10.1016/j.isci.2020.101849

subsequently extended by incorporation of Remdesivir or ATP (Gordon et al., 2020b). Expression and purification of a viral polymerase for use in kinetic analysis must meet standards of purity and activity for the results to be valid. In particular, a polymerase must be capable of replicating the RNA at rates sufficient to account for the time required for viral replication *in vivo*, be fully active, and be free of contaminating nuclease activity. Most enzyme purification strategies rely on the use of tagged enzyme without regard to the effects of the tags on activity. For example, in prior work on HIV reverse transcriptase, purification without tags was essential for getting the most active enzyme preparations (Lu et al., 2010). Prior efforts to study the coronavirus RdRp *in vitro* have mostly relied on baculovirus-infected insect cell expression (Gordon et al., 2020b; Hillen et al., 2020; Yin et al., 2020) to overcome the low solubility of NSP12 when overexpressed in *E. coli*. However, baculovirus expression of protein is expensive, time consuming, and the low yield usually requires the use of tagged protein.

Prior coronavirus RdRp inhibition studies have relied on steady-state kinetic methods (Gordon et al., 2020b). However, steady-state kinetic assays do not yield meaningful results when applied to processive enzymes due to the slow, rate-limiting dissociation of the enzyme from the primer/template after the much faster rate of nucleotide incorporation (Kuchta et al., 1987). Rate-limiting dissociation masks inherent properties of nucleoside analogs, thereby misdirecting drug design studies. To circumvent problems with the steady-state approach and provide a robust system for studies on the RdRp complex, here we present pre-steady-state experiments to accurately monitor the kinetics of polymerization and quantify the kinetics of Remdesivir incorporation relative to ATP. This work is based on our optimization of the expression and purification of tag-free SARS-CoV-2 RdRp complex in *E. coli*, yielding protein that is capable of fast processive synthesis, as revealed by the pre-steady-state kinetic measurements. We reveal key features of Remdesivir and lay the foundation for robust assays to screen for and optimize new inhibitors for treating SARS-CoV-2 and future novel coronavirus infections.

RESULTS

Co-Expression of SARS-CoV-2 RdRp Subunits with Chaperones Gives Soluble Expression in *E. coli*

The SARS-CoV-2 RdRp complex, responsible for the replication of the positive-sense single-stranded RNA genome, consists of the NSP12 catalytic subunit with accessory proteins NSP7 and NSP8 (Hillen et al., 2020). Most mechanistic studies have been performed using protein purified from baculovirus-infected insect cells with histidine or other tags used in purification (Gordon et al., 2020b; Hillen et al., 2020; Yin et al., 2020). Because of the many advantages of *E. coli* expression, we set out to optimize expression of the RdRp complex in *E. coli*. Initial expression of untagged NSP12 alone yielded virtually all of the protein in inclusion bodies at all temperatures tested from 8–37°C (data not shown). A recent paper (Shannon et al., 2020) expressed a histidine-tagged NSP12 in *E. coli* by co-expression of chaperones from a separate plasmid, pG-Tf2 (Nishihara et al., 2000). This plasmid contains tetracycline inducible copies of Tf, GroEL, and GroES chaperones. When we tested expression of NSP12 alone with pG-Tf2, we observed a slight increase in solubility; however, the majority of the protein was still insoluble (data not shown).

Co-expression of proteins with their cellular protein partners has been an effective strategy to improve solubility in some difficult recombinant protein expression systems (Chiu et al., 2006). To test this strategy for the SARS-CoV-2 replication complex, an expression plasmid pQE-(NSP12)-pcl$^{ts,ind+}$-(NSP7-NSP8) was constructed so that it contains codon-optimized genes for NSP12, NSP7, and NSP8 in a single expression plasmid (Figure S1A). NSP12 was cloned under the IPTG inducible T5 promoter/lac operator from the pQE-30 parent plasmid, and NSP7 and NSP8 were cloned as a bicistronic operon under the temperature/nalidixic acid inducible bacteriophage λ promoter from the pcl$^{ts,ind+}$ backbone (Brandis and Johnson, 2009). This plasmid also has the kanamycin resistance gene, the high copy number pUC origin of replication, and on-board λcI$^{ts,ind+}$ and lacI repressors so protein can be expressed in any recA^{+} *E. coli* strain (Brandis and Johnson, 2009). For comparison of the untagged NSP12/7/8 complex to a histidine-tagged NSP12 and modified NSP7/NSP8 construct containing a GSGSGS linker used in a previous study (Shannon et al., 2020), the plasmid pQE-(NSP12-TEV-8xHis)-pcl$^{ts,ind+}$-(NSP7L8) was created (Figure S2). We prepared BL21 *E. coli*/pG-Tf2 harboring either pQE-(NSP12)-pcl$^{ts,ind+}$-(NSP7-NSP8) or pQE-(NSP12-TEV-8xHis)-pcl$^{ts,ind+}$-(NSP7L8) to test solubility of the complexes when co-expressed with chaperones from pG-Tf2. The chaperone/co-expression strategy yielded virtually all soluble protein for the untagged construct (Figures S3A and S3B) and mostly soluble protein for the his-tagged NSP12/NSP7L8 complex (not shown). We purified

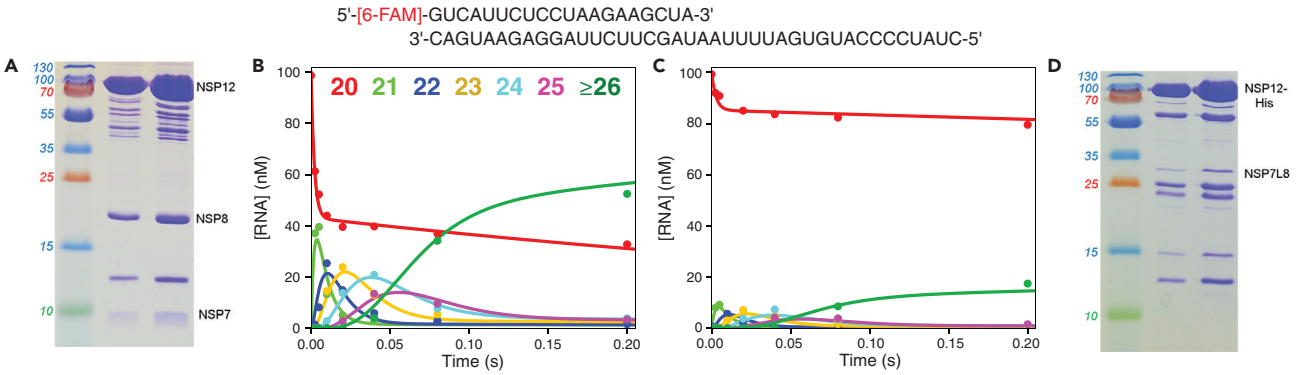


Figure 1. Tag-Free SARS-CoV-2 NSP12/7/8 Complex Has Higher Activity than the His-tagged NSP12/NSP7L8 Complex

Top: RNA substrate. The sequence was copied from the SARS-CoV-2 genome preceding the poly(A) tail. The primer strand was synthesized with a 5′-[6-FAM] label to monitor reaction kinetics by capillary electrophoresis.

(A) Gel showing purified NSP12/7/8 complex. Molecular weight markers are given to the left of the gel and 2 dilutions of the purified product are shown. Bands corresponding to NSP7, NSP8, and NSP12 are labeled to the right of the gel.

(B) NSP12/7/8 complex activity assay. A solution of 2 μM NSP12/7/8 complex, 5 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 250 μM each of ATP, CTP, and UTP to start the reaction. Approximately 60% of the starting RNA was turned over to product in the first turnover. Colors for each RNA length are given at the top of the panel.

(C) NSP12-His/NSP7L8 complex activity assay. A solution of 2 μM NSP12-His/NSP7L8 complex, 5 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 250 μM each of ATP, CTP, and UTP to start the reaction. Around 15% of the RNA was converted to product during a single turnover. RNA products of various lengths are colored as in (A). To a first approximation, note that the midpoint in forming the final product (12 nt incorporated) is ~0.06 s, giving an average polymerization rate of ~200 s^{-1}.

(D) Gel showing purified NSP12-His/NSP7L8 complex. Molecular weight markers are given to the left of the gel and 2 dilutions of the purified product are shown. Bands corresponding to NSP12-His and NSP7L8 are shown to the right of the gel.

the his-tagged complex on a nickel column and the untagged complex with a series of five different columns as described in the Transparent Methods section. The final step in the purification of the untagged complex was a size exclusion column, which separated NSP12 alone and complexes of NSP12/NSP7 and NSP12/NSP8 from the NSP12/NSP7/NSP8 complex. It is difficult to tell the exact stoichiometry from bands on a gel; however, the presence of smaller complexes lacking one or more subunits during size exclusion chromatography suggests the pooled complex of NSP12/NSP7/NSP8 has each subunit at least at a 1:1:1 ratio. For the his-tagged NSP12, the NSP7L8 polypeptide co-purified on the nickel column (Figure 1D) and for the untagged NSP12, NSP7, and NSP8 were all visible in the final purified product (Figure 1A). Because NSP8 has been shown to bind with two molecules per one molecule of NSP7 and NSP12 (Hillen et al., 2020), NSP8 was added in excess to the purified replication complex to saturate NSP8 binding. NSP8 was cloned into the $pcI^{ts,ind+}$ backbone (Figure S1B) and was overexpressed in the soluble fraction with heat induction in *E. coli* and purified with four different columns (Figure S4) as described in the Transparent Methods section.

Tag-free SARS-CoV-2 NSP12/7/8 Complex Has Higher Activity Than His-Tagged NSP12/NSP7L8 Complex

We initially tested the primer extension activity of the various purified SARS-CoV-2 proteins with a pre-steady-state rapid quench experiment. The RNA substrate used (Figure 1) in this paper consists of a 20 nt 5′-[6-FAM] labeled primer, annealed to a 40 nt template with the sequence from the SARS-CoV-2 genome (preceding the poly(A) tail) (Ferron et al., 2018; Ma et al., 2015). The experiment was performed with a large excess of enzyme over RNA. NSP8 was added to 5 μM for initial experiments, which is approximately 10 times the reported K_d for NSP8 binding to NSP12 (Peng et al., 2020). We tested both the untagged NSP12/7/8 and histidine-tagged NSP12-His/NSP7L8 protein preparations for primer extension activity. Reaction products were diluted into formamide and separated by capillary electrophoresis at 65°C in nanoPOP-6 polymer, which was sufficient to completely denature the RNA for separation (Figure S5).

For the untagged protein, a solution containing 2 μM NSP12/7/8 complex, 5 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was allowed to equilibrate for 10–60 min (the time to complete the experiment), then was mixed with 250 μM ATP, CTP, and UTP to start the reaction. Adding three of the four nucleotides afforded extension of the primer by 12 residues before a C in the template was encountered, terminating

synthesis. Following the addition of the three nucleotides, rapid extension of the RNA primer to form the 32 nt product was observed in 0.2 s (Figure 1B). Under the parallel conditions with the NSP12His/NSP7L8 complex, we also observed rapid formation of the 32 nt product, but only ~15% of the RNA primer was extended in a single turnover compared with 60% with the untagged NSP12/7/8 complex (Figure 1C), suggesting either a lower fraction of active protein, weaker RNA binding by the tagged polymerase and/or suboptimal concentration of the NSP7L8 accessory protein. No degradation of the primer by RNases was detected for either tagged or untagged RdRp complex during the roughly one-hour preincubation of the enzyme with RNA in the presence of Mg^{2+} while performing the experiment, indicating there was negligible RNase contamination in either enzyme preparation. It was previously reported for the SARS-CoV NSP12His/NSP7L8 complex that the activity was higher than the NSP12/7/8 complex (Subissi et al., 2014), presumably because the linked NSP7L8 overcame the weak binding of NSP7 and NSP8 to NSP12. However, that was not the case in our hands. The lower activity could be attributed to a number of factors including the histidine tag, the number of columns used for purification, lack of supplemented NSP7L8, etc. We have not repeated the purification and characterization of the tagged enzyme and we did not pursue further optimization of the histidine-tagged complex. Therefore, we cannot definitively conclude that the tag or the NSP7L8 variant interfered with activity; rather, our results serve as a cautionary tale. Because the untagged protein had higher activity and more closely mimics the natural protein, we opted to use the untagged RdRp complex for the remainder of our experiments.

As a control experiment, we also tested NSP8 alone for primer extension activity as a paper on the SARS-CoV (te Velthuis et al., 2012) suggested NSP8 has primer extension activity. A solution of 5 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 250 μM each of ATP, CTP, and UTP to start the reaction. Reactions were quenched with EDTA and time points were analyzed by capillary electrophoresis. No primer extension was observed on the timescale of the experiment, and no degradation of the primer strand occurred during the preincubation of the enzyme with the RNA (data not shown). This indicates that under our experimental conditions, NSP8 does not have measurable primer extension activity and that there is no detectable RNase activity in our tag-free preparation of NSP8.

Pre-steady-state Kinetics of UTP Incorporation

The next experiment was designed to measure the UTP concentration dependence on the rate of polymerization. A pre-equilibrated mixture containing 2 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with varying UTP concentrations (2.5–150 μM) to start the reaction. Reactions were quenched and analyzed as in Figure 1. Because there are two sequential templating A's, two UTPs were incorporated in rapid succession, as shown in Figure 2A. Data for the total amount of product at each nucleotide concentration are shown in Figure 2B—by summing the 21 and 22 nt products, we define the kinetics of the first UTP incorporation. Figure 2C shows the observed rate versus UTP concentration for the first UTP incorporation, obtained by fitting the data in Figure 2B to a single exponential function. Conventional equation-based fitting of the data in Figure 2C using a hyperbolic function provides estimates for the parameters k_{cat} = 220 ± 23 s^{-1} and K_m = 74 ± 11 μM for incorporation. To achieve more accurate results, the parameters reported in this paper were obtained by global data fitting based on numerical integration of the rate equations using KinTek Explorer to derive estimates for k_{cat} and K_m directly (Johnson, 2009, 2019; Johnson et al., 2009b). Data fitting by simulation also allows resolution of multiple sequential polymerization reactions during processive synthesis (Figure 2), and confidence contour analysis affords realistic error estimates based on the extent to which each parameter is constrained by the data (Johnson et al., 2009a). The scheme in Figure 2 shows the model used for fitting data in KinTek Explorer to extract k_{cat} (k_2 and k_4), and K_m (($(k_{-1}+k_2)/k_{-1}$), and (($(k_{-3}+k_4)/k_3$))) for the two UTP incorporations. The value of k_{cat}/K_m was then calculated from the estimates for k_{cat} and K_m. The parameters k_{cat}/K_m, k_{cat}, and K_m for each UTP incorporation derived from the simulation-based data fitting are summarized in Table 1. Of particular importance, note that the specificity constant is defined by k_{cat}/K_m.

SARS-CoV-2 RdRp Complex Weakly Binds RNA Substrate

Because only 60% of the RNA primer was extended in the previous experiments in spite of a large excess of enzyme over the RNA substrate, the following experiment was performed to differentiate low active enzyme concentration from weak binding to the RNA. A mixture of varying concentrations of the NSP12/7/8 complex (0.2–10 μM), 20 μM NSP8, 200 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 150 μM UTP to start the reaction. Reactions were quenched with EDTA, and the amount of total extended primer was determined by capillary electrophoresis. The concentration of product formed versus time at

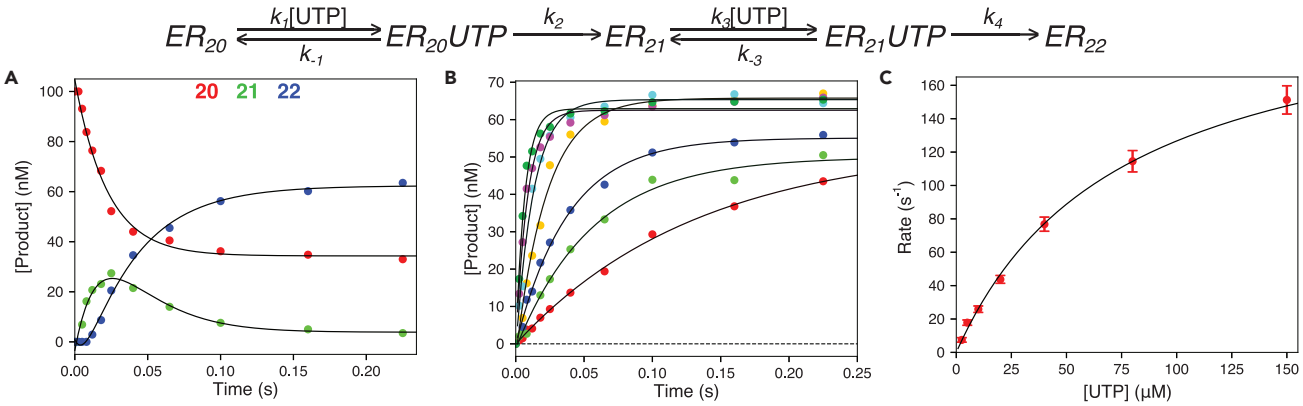


Figure 2. Pre-steady-state Kinetics of UTP Incorporation

Scheme: Kinetics of sequential UTP incorporations ER_n is the enzyme-RNA complex *n* nucleotides in length. k_1 and k_3 are k_{cat}/K_m for the first and second UTP incorporation, respectively, whereas k_2 and k_4 define k_{cat} for the first and second UTP incorporation, respectively. Experiment: a mixture containing 2 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with varying concentrations of UTP (2.5–150 μM) to start the reaction.

(A) UTP incorporation reaction at 20 μM UTP. The fits to double exponential functions are shown by the black lines going through the data points for each RNA species: 20 nt (red), 21 nt (green), and 22 nt (blue).

(B) Total product versus time at varying UTP concentrations. The sum of products 21 and 22 nt in length was plotted versus time for reactions performed at various UTP concentrations. The fits to single exponential functions are shown as the black lines going through the data points, with different colors for different concentrations (red to dark green: 2.5, 5, 10, 20, 40, 80, 150 μM).

(C) Rate of product formation versus UTP concentration. Rates and standard errors (bars) were derived from the singe exponential fit of the data in (B). This simple analysis illustrates the saturation of the rate versus nucleotide concentration with estimates of k_{cat} = 220 ± 23 s^{-1} and K_m = 74 ± 11 μM. Better estimates were derived from global data fitting as shown in Figure 3 and summarized in Table 1.

different enzyme concentrations is shown in Figure 3D. The amplitude versus nominal enzyme concentration graph (not shown) fits a hyperbola (K_d = 1.9 ± 0.2 μM), signifying that the binding of the enzyme to the RNA is weak compared with the fixed RNA concentration (200 nM) in the titration. This experiment suggests that the amplitude of 60 nM in Figure 2B is mostly due to weak RNA binding with some effect potentially from inactive enzyme. The K_d estimated from the global fit by simulation was 1.8 ± 0.2 μM for the enzyme binding to RNA, assuming 80% of the enzyme is active, as described below. This information aids in designing additional experiments in this paper, as the amount of product expected to form in a single turnover can be estimated based on the measured K_d for RNA binding.

SARS-CoV-2 RdRp Complex Dissociates Very Slowly from the RNA Substrate

Success of the experiment described in Figure 3D depends on a slow RNA dissociation rate. To measure the rate of RNA dissociation from the NSP12/7/8-RNA complex, a quench flow double-mixing experiment was performed. A solution of 1.25 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was allowed to equilibrate for 30 min, then mixed with 2 mg/mL heparin to start the reaction. The use of 2 mg/mL heparin has been shown to be an effective concentration (Donlin et al., 1991) as a trap for enzyme dissociated from the nucleic acid substrate for other polymerases, and the fact that the amount of product tends toward 0 indicates that heparin is an efficient trapping agent for this enzyme (Figure 3E). This solution was incubated for various amounts of time (mixing step 1) in the reaction loop before mixing with 150 μM UTP and 2 mg/mL heparin from the quench syringe and then allowed to react for 50 ms (mixing step 2). The reaction was then quenched with EDTA (0.3 M final concentration) by mixing in the collection tube. The decreasing amount of product formed in the rapid reaction with UTP provides a direct measurement of the decreasing amount enzyme-RNA complex as a function of time during mixing step 1 with the heparin trap (Figure 3E). The dissociation rate is slow, so the best fit obtained by simulation gives an RNA dissociation rate constant of 0.013 ± 0.001 s^{-1} (k_{off} in the Scheme in Figure 3). Confidence contours for the RNA off rate and UTP incorporation rate constants are shown in Figures 3G and 3H, respectively.

To obtain a minimal estimate of the active fraction of enzyme, we locked the RNA dissociation rate at 0.013 s^{-1} and fit the enzyme titration experiment in Figure 3D and UTP incorporation data in Figures 3A–3C using KinTek Explorer using the scheme shown at the top of Figure 3. The experiments were modeled with a pre-equilibration of the E and X states, where E is the active enzyme and X is inactive enzyme. This is not to say that enzyme equilibrates between active and inactive state; rather, this model provides a method to

Nucleotide	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)
UTP-1	2.3 ± 0.2	308 ± 17	130 ± 9
UTP-2	1.7 ± 0.5	680 ± 130	400 ± 80
ATP	0.74 ± 0.16	240 ± 30	320 ± 50
RTP-1	1.29 ± 0.06	68 ± 2	53 ± 2
RTP-2	0.19 ± 0.01	3.62 ± 0.04	19 ± 1
RTP-3	0.18 ± 0.01	3.76 ± 0.07	21 ± 2

Table 1. Kinetic Parameters for Nucleotide Incorporation[a]

[a]Rate constants were obtained by globally fitting data using KinTek Explorer. Standard errors were obtained from upper and lower limits derived by nonlinear regression analysis backed up by confidence contour analysis (Figures 3 and 5). In the nucleotide column, numbers next to the nucleotide represent each sequential incorporation where multiple incorporations were resolved by the data. Values for K_m and k_{cat}/K_m were calculated from the intrinsic rate constants derived in data fitting as described in the Methods. Values for k_{cat} were derived directly in fitting as the intrinsic rate constant for incorporation (Johnson, 2019; Kellinger and Johnson, 2010).

estimate the fraction of enzyme that is active. Equilibration between the E and X states was made very slow (10^{-5} s^{-1}) and allowed to reach equilibrium (computationally for ~6 days) before adding RNA in a separate step and then nucleotide in a final step, both of which occur on timescales fast enough to not perturb the E to X equilibrium.

Confidence contours for K_{active}, and the resulting k_{on} are shown in Figure 3F. This data fitting shows that the data define a lower limit of 60% active enzyme and that the resulting K_d (from varying k_{on}) changes based on the active fraction. At 60% active enzyme, the resulting K_d for RNA binding is 1.3 μM, whereas at 100% active enzyme the resulting K_d would be 2.2 μM, giving the range of possible K_d values from the data. The RNA dissociation rate does not vary significantly over this range because it is well defined by the data in Figure 3E and the interpretation is independent of the fraction of active enzyme over the limited range (60%–100%). The linear relationship between calculated binding rate constant and fraction of active enzyme reflects the fact that the net RNA binding rate with enzyme in excess is defined by the product of k_{on}[E]. The apparent RNA binding rate constants are far below diffusion limits (0.008 $\mu M^{-1}s^{-1}$ calculated for 80% active enzyme), which could be attributed to a two-step binding mechanism; however, the experiments presented here cannot differentiate multiple binding steps, so the simple one-step binding model was used.

Globally fitting the data shown in Figures 3A–3C was finalized assuming that 80% of the enzyme was active and with RNA binding kinetics defined from Figures 3D and 3E. The k_{cat} and K_m values for the first and second UTP incorporations are summarized in Table 1. These parameters do not significantly vary over the range of 60%-100% active enzyme (Table S1).

Remdesivir Is Incorporated More Efficiently Than ATP by the SARS-CoV-2 RdRp Complex

Because Remdesivir (Figure 4D) is an ATP (Figure 4A) analog, we first measured the kinetics of ATP incorporation by rapid quench methods so the specificity constant (k_{cat}/K_m) for ATP provides a benchmark for evaluating the relative efficiency of Remdesivir incorporation. A solution of 1.25 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 200 μM UTP and varying concentrations of ATP (5–400 μM) to start the reaction. The RNA substrate used in these studies encodes for two UTPs followed by four ATPs and then another UTP to form the 27 nt product. The data for each product band at 225 μM ATP are shown in Figure 4B showing the rise and fall of each intermediate.

Because the rates for each ATP incorporation appear similar and rapid after UTP incorporation, the kinetic parameters for each individual ATP incorporation were not well resolved in globally fitting the data (see scheme in Figure 4). Therefore, products ≥23 nt, corresponding to the products with ≥ one ATP incorporated, were summed to give the kinetics of incorporation of the first ATP (defined by K_5 and k_6). The data for the total product ≥23 nt at each ATP concentration are shown in Figure 4C. Fitting the data with known kinetics for incorporation of the two UTPs affords accurate estimates of the kinetic parameters governing the first ATP incorporation: $K_m = (k_{-5}+k_6)/k_5$ and $k_{cat} = k_6$ (Table 1). In particular, k_{cat}/K_m represents the most important parameter to evaluate

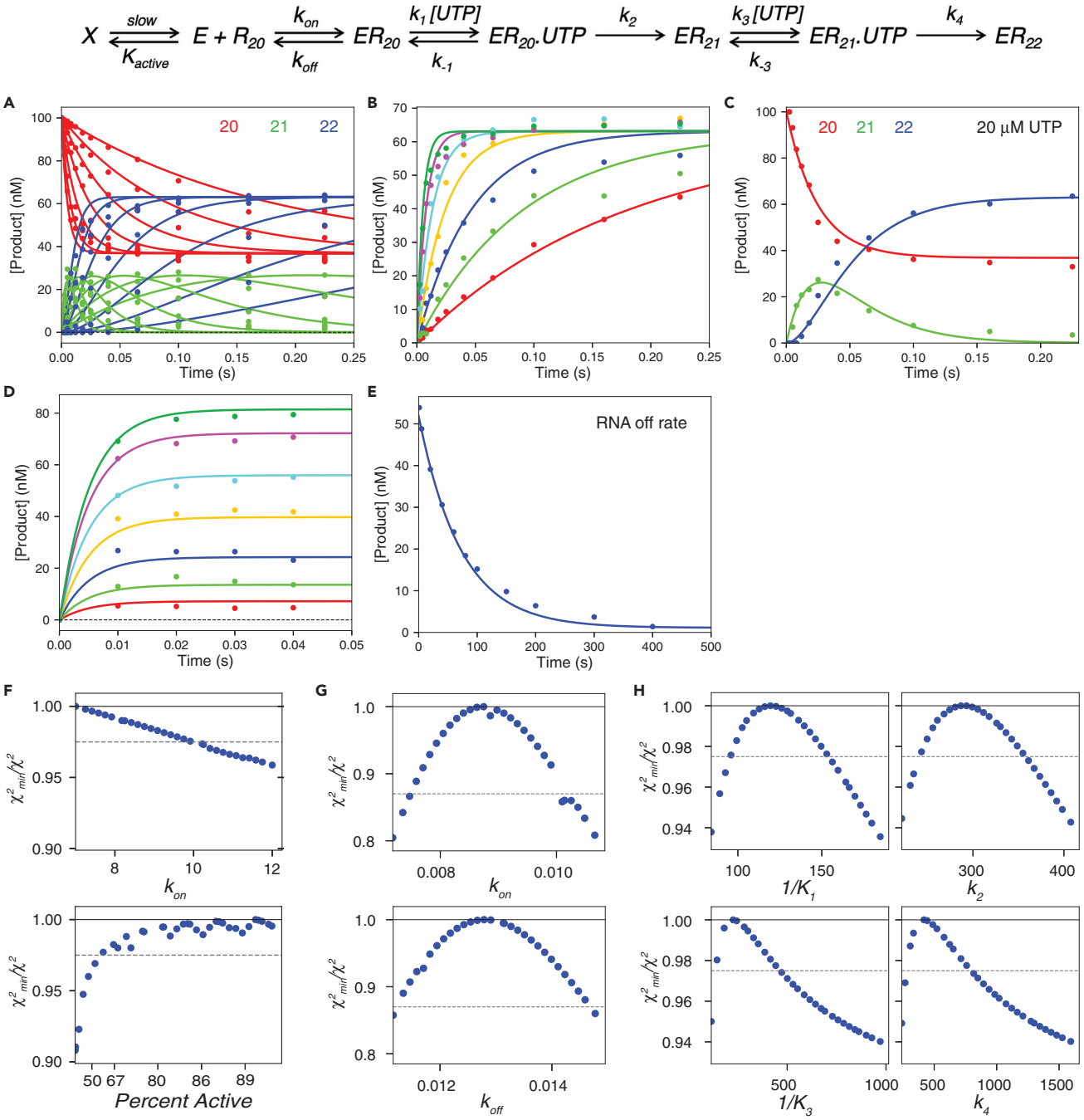


Figure 3. SARS-CoV-2 RdRp Complex Weakly Binds an RNA Substrate and Dissociates at a Very Slow Rate

Scheme: ER_n is the enzyme-RNA complex with RNA *n* nucleotides in length k_{off} is the RNA dissociation rate and k_{on} is the apparent RNA association rate. The ratio of k_{off}/k_{on} gives the K_d or the apparent equilibrium binding constant. Kinetic constants for UTP incorporation are labeled as in the scheme in Figure 2. K_{active} is the equilibrium constant for enzyme going from the active E state to the inactive X state, which occurs during the pre-equilibration step to estimate the active enzyme concentration, as described in the Transparent Methods. Experiments in panels A–C are the same as in Figure 2.

(A) UTP incorporation, concentration dependence. Data for the 20, 21, and 22 nt RNA are shown in red, green, and blue, respectively.

(B) UTP incorporation, concentration dependence. The concentration of total product versus time is plotted for each UTP concentration.

(C) UTP incorporation at 20 µM UTP. RNA of different lengths is colored as in (A).

(D) UTP incorporation, enzyme titration. A mixture of 0.2–10 µM NSP12/7/8 complex, 20 µM NSP8, 200 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 150 µM UTP to start the reaction. The amount of total product formed was plotted versus time for each enzyme concentration and fit by simulation in KinTek Explorer.

Figure 3. *Continued*

(E) RNA dissociation rate experiment. A solution containing 1.25 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was allowed to equilibrate for 30 min, then mixed with 2 mg/mL heparin in the first mixing step using the RQF-3 rapid-quench flow instrument. After the designated first mixing time (shown in the figure), the reaction was mixed with 125 μM UTP (concentration after dilution) from the quench syringe, held in the exit line for 50 milliseconds, then the reaction was quenched by mixing with EDTA to a final concentration of 0.3 M in a collection tube. The time axis in the figure is the time allowed for RNA dissociation after mixing with heparin trap, before the addition of nucleotide. The best fit by simulation is shown as the solid line yielding an RNA dissociation rate of 0.013 ± 0.001 s^{-1}

(F) Confidence contours used to estimate fraction active enzyme. Data in panels A–E were fit globally while allowing the equilibrium constant, K_{active} to vary, along with k_{on} (which compensates for variable active enzyme concentration). The data were fit to extract the percent of active enzyme and the corresponding k_{on} values, as described in the Transparent Methods. All other rate constants were also allowed to vary, except k_1 and k_3 that were locked at 100 $\mu M^{-1}s^{-1}$. Data in the panel show that the enzyme is at least 60% active.

(G) Confidence contours to define k_{on} and k_{off}. Data in panels D and E were fit assuming 80% active enzyme with all other rate constants locked at their best-fit values. This analysis give k_{on} = 0.008 ± 0.001 $\mu M^{-1}s^{-1}$ and k_{off} = 0.013 ± 0.001 s^{-1} and K_d for RNA binding of 1.8 ± 0.2 μM.

(H) Confidence contours for UTP binding and incorporation. Date in panels A–C were fit assuming 80% active enzyme and with k_{on} and k_{off} locked at their best fit values. The values for k_1 and k_3 were locked at 100 $\mu M^{-1}s^{-1}$. For each contour, the dashed line shows the χ^2 threshold. The smooth lines in each panel were derived from the global data fit. Rate constants derived from global data fitting, assuming 80% active enzyme, are given in Table 1.

Remdesivir incorporation because the ratio of k_{cat}/K_m values for the two nucleotides defines the ability of the enzyme to discriminate against Remdesivir in favor of ATP, as described below.

To investigate the kinetics of Remdesivir triphosphate (RTP) incorporation by the SARS-CoV-2 NSP12/7/8 complex, the concentration dependence of Remdesivir on the kinetics of its incorporation were measured by rapid quench methods, as described for ATP. A solution of 1.5 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was equilibrated then mixed with 150 μM UTP and various concentrations of RTP (3.5–315 μM) to start the reaction. Data at 315 μM RTP are shown in Figure 4E to illustrate the time dependence for formation and decay of each species. As with studies on ATP incorporation, because the kinetics of UTP incorporation are known (k_1 through k_4 in the scheme in Figure 4), the incorporation of Remdesivir can be easily modeled and fit by simulation. After the two templating A's, there are 4 templating U's followed by another templating A, giving 4 opportunities for RTP incorporation followed by the opportunity for one UTP incorporation. The rate constants for the first RTP were obtained by summing all of the products ≥23 nt in length as shown in Figure 4F. Fitting the data by simulation shows that the first RTP was efficiently incorporated: k_{cat}/K_m = 1.29 ± 0.06 $\mu M^{-1}s^{-1}$, $k_{cat} = k_6$ = 68 ± 2 s^{-1}, and K_m = 53 ± 2 μM (Table 1). Fitting the complete time course for each species (exemplified by one RTP concentration in Figure 4E) afforded resolution of all three incorporation reactions. Incorporations of the second and third RTP (k_7 through k_{10} in the scheme in Figure 4) occur with lower efficiencies with k_{cat}/K_m values (~0.19 $\mu M^{-1}s^{-1}$) and slower k_{cat} (3.6–3.8 s^{-1}) leading to lower K_m values (Table 1). The slower rates for the second and third incorporation allow the kinetics to be resolved in global data fitting. Even though the k_{cat} values are lower than for ATP, the data demonstrate that at least 3 RTPs are incorporated efficiently.

Confidence contours in Figure 5 show that each parameter is well constrained by the data. Error estimates in Table 1 were derived from the standard errors derived during nonlinear regression analysis. Such error estimates can be misleading if the parameters are not well constrained by the data. In the present case, the symmetrical confidence contour analysis establish that each parameter is well constrained and therefore, the standard error estimates are valid (Johnson, 2019). Multiplying the standard error by 1.96 gives the 95% confidence interval similar to values derived by confidence contour analysis. Incorporation of a fourth RTP occurs slowly, with only a small amount observed at the highest concentration. On the timescale of the experiment, no incorporation of UTP was observed after the four molecules of Remdesivir, but it may occur after longer times of incubation, especially due to the absence of the proofreading exonuclease in this experiment so each incorporation reaction is largely irreversible.

DISCUSSION

As of October 25, 2020, the World Health Organization reports over 43 million confirmed cases and over 1.15 million deaths due to SARS-CoV-2 globally with new infections increasing by 33% in a last week (WHO, 2019). Although development of a vaccine for SARS CoV-2 appears promising, it will still be important to develop an arsenal of direct-acting antiviral drugs for individuals who fail to develop neutralizing antibodies or neglect to get vaccinated. The viral RdRp is an attractive target for direct-acting antiviral drugs. In particular, nucleoside analogs may be effective in treating newly arising coronaviruses due to conservation of active site residues needed for activity. For example, nucleoside analogs and nonnucleoside inhibitors of the HIV reverse transcriptase are the cornerstone of treatments to control HIV infections

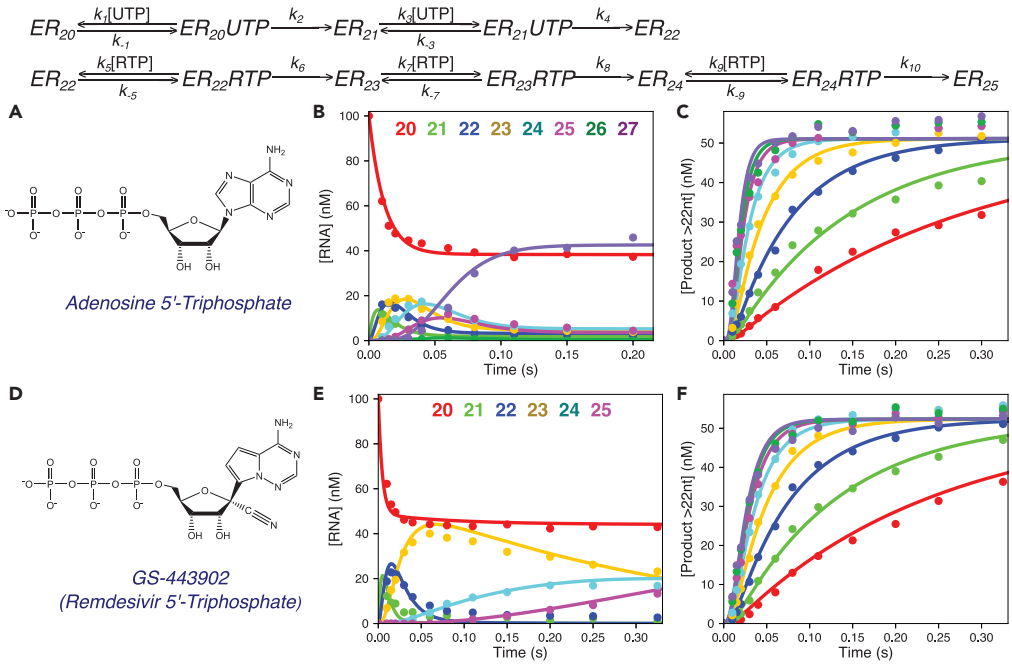


Figure 4. Remdesivir Is Incorporated more Efficiently than ATP by the SARS-CoV-2 RdRp Complex

Scheme: kinetic pathway for UTP and ATP incorporation. Species are labeled as in Figure 2, with added steps for sequential RTP incorporations; the same sequence applies to ATP incorporation. Net rate constants k_5, k_7, and k_9 define k_{cat}/K_m for each sequential RTP incorporation, whereas k_6, k_8, and k_{10} define k_{cat} for each sequential RTP incorporations. For ATP the pathway is identical, but we fit the data to only define k_{-5} and k_6 defining the kinetic parameters for the first incorporation, modeled as the sum of all products ≥23 nt in length.

(A) Chemical structure of adenosine 5′-triphosphate. Experiments: a solution containing 1.25 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 200 μM UTP and varying concentrations of ATP (5–400 μM) to start the reaction, and products were resolved and quantified using capillary electrophoresis.

(B) Reaction progress curve for individual RNA species at 225 μM ATP. Products of primer extension from 20 to 27 nt in length are shown as a function time after mixing. Note the reaction is over in less than 0.2 s. The solid lines through the data points show the best global fit by simulation for each RNA species with an average rate of ATP incorporation around 80 s^{-1} at this concentration. The color code for each RNA species is given at the top of the figure.

(C) Concentration of ATP incorporation product versus time at various ATP concentrations. We plot the sum of all products after the incorporation of the first ATP, i.e., products 23–27 nt in length to define the kinetics of the first ATP incorporation. The solid lines show the best global fit by simulation for the total RNA product ≥23 nt at each ATP concentration, shown as different colors (red to purple: 5, 10, 20, 40, 80, 150, 225, and 400 μM). These data define the parameters k_{cat} and K_m for ATP incorporation as shown in Table 1.

(D) Chemical structure of GS-443902 (Remdesivir 5′-triphosphate, RTP). RTP is an adenosine analog with a 1′ cyano group and modifications to the adenine base. Note that Remdesivir contains a 3′-OH, allowing continued polymerization after its incorporation. A mixture containing 1.5 μM NSP12/7/8 complex, 6 μM NSP8, 100 nM FAM-20/40 RNA, and 5 mM Mg^{2+} was mixed with 150 μM UTP and varying concentrations of RTP (3.5–315 μM) to start the reaction in the quench-flow instrument. Reactions were quenched with EDTA after various reaction times and the products quantified by capillary electrophoresis.

(E) Concentration of each RNA species versus time at 315 μM RTP. The concentrations of various species, 20 nt (red), 21 nt (green), 22 nt (blue), 23 nt (yellow), 24 nt (cyan), and 25 nt (purple) are shown with the solid lines from the best global fit of the data by simulation in KinTek Explorer, which included the time course for each species at each RTP concentration (not shown).

(F) Concentration of product containing RTP versus time for various concentrations of RTP. The time dependence of the first RTP incorporation is shown as the sum of species ≥23 nt in length. The best fit by simulation is shown as the solid colored lines through the data points, with different colors for each RTP concentration (red to purple: 3.5, 7, 14, 28, 56, 112, 210, and 315 μM). The parameters k_{cat} *and* K_m for each Remdesivir incorporation were derived by globally fitting all of the primary data defining the formation and decay of products 23, 24, and 25 nt in length and are summarized in Table 1.

(Skowron and Odgen, 2006). Similarly, HCV infections are effectively treated by the nucleoside analog, Sofosbuvir, which targets the viral RdRp. On October 22, 2020 Remdesivir became the first drug to be approved by the FDA for treating COVID-19 (FDA, 2020).

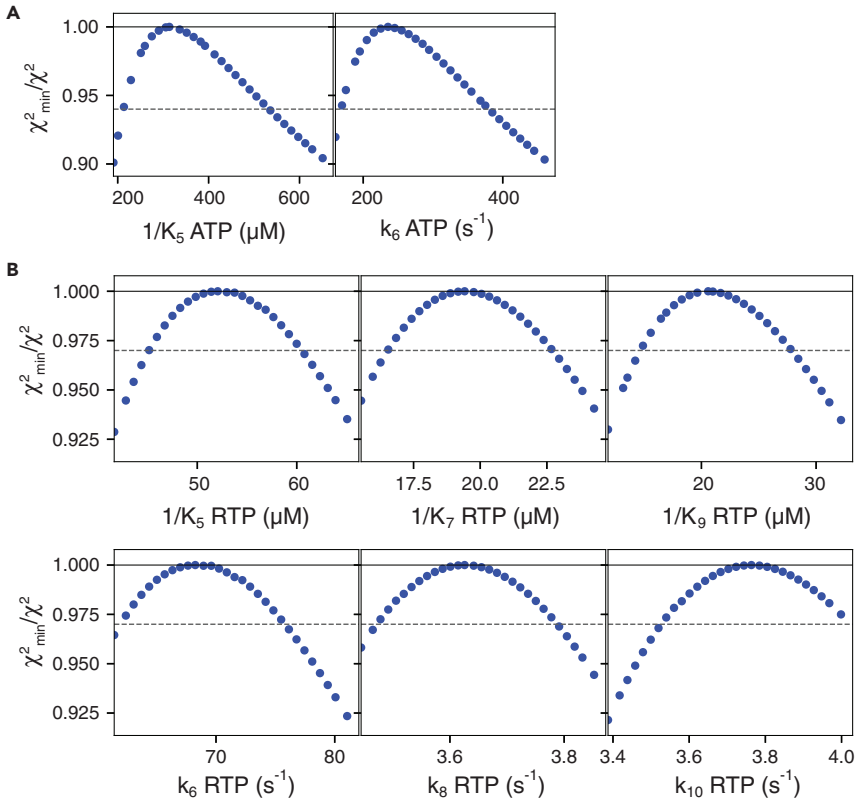


Figure 5. Confidence Contour Analysis

(A) Confidence contours for ATP incorporation. Confidence contours were derived from fitting the ATP incorporation data to determine $1/K_{5,ATP}$ and $k_{6,ATP}$. The data are slightly skewed, with a well-defined lower limit but a less well-defined upper limit due to the two sequential UTP incorporations that occur before ATP incorporation. The error estimates were based on a χ^2 threshold of 0.94 for the ratio of χ^2_{min}/χ^2 (dashed line) to define the 95% confidence limits (Johnson, 2019; Johnson et al., 2009a).

(B) Confidence contours for RTP incorporation. Confidence contours were derived from fitting the RTP incorporation data to determine $1/K_{5,RTP}$, $1/K_{7,RTP}$, $1/K_{9,RTP}$, $k_{6,RTP}$, $k_{8,RTP}$, and $k_{10,RTP}$. Based on the number of parameters and data points, the threshold was 0.97 in fitting the RTP incorporation data to define the 95% confidence limits. Values for UTP incorporation were locked at the values given in Table 1 during the fitting. Rate constants are defined according to the scheme in Figure 4. Error estimates listed in Table 1 are based on the standard errors derived during nonlinear regression analysis. The estimated errors are consistent with the confidence intervals derived from confidence contour analysis, noting that the confidence interval is 1.96 times the standard error range.

We found that Remdesivir is incorporated by the RdRp complex more efficiently than ATP, defined by the relative k_{cat}/K_m values for RTP versus ATP (1.29 versus 0.74 $\mu M^{-1}s^{-1}$). Although RTP is incorporated at least three-fold slower than ATP (68 s^{-1} versus 240 s^{-1}), it has a six-fold higher k_{cat}/K_m value due to a six-fold lower K_m relative to ATP (53 μM versus 320 μM). We define a *discrimination index* as the ratio of specificity constants for the normal nucleotide divided by that for the analog—this term quantifies the ratio by which the enzyme discriminates against the nucleotide analog (AnTP) relative to the normal nucleotide (NTP).

$$\text{Discrimination index} = D = (k_{cat}/K_m)_{NTP}/(k_{cat}/K_m)_{AnTP}$$

$$\text{Nucleotides incorporated per analog} = D \times [NTP]/[AnTP]$$

The value of the discrimination index for RTP versus ATP is less than one (0.57), which indicates that the analog is a better substrate than the cognate nucleotide. Of the nine nucleoside analogs approved by the FDA to treat HIV infections, only Stavudine (a thymidine analog) has a discrimination index less than one, but it has a higher toxicity index due to incorporation by the human mitochondrial DNA polymerase

(Johnson et al., 2001). Certainly, the discrimination index of 0.57 for RTP will help it to compete with ATP, but its effectiveness may still be limited by the high concentration ATP in the cell, typically around 3 mM (Traut, 1994). For example, an intracellular RTP concentration of 10 μM will lead to incorporation of RTP only one time out of 170 uridines in the template. However, this may be sufficient. For example, Sofosbuvir is effective in treating HCV infections even though it is incorporated inefficiently with a k_{cat}/K_m = 0.0007 $\mu M^{-1}s^{-1}$ compared with 0.1 $\mu M^{-1}s^{-1}$ for UTP, giving a discrimination index = 140 ± 45 (Villalba et al., 2020). The physiological concentration UTP ranges from 110 to 1000 μM UTP (Traut, 1994), so given a hypothetical concentration of 10 μM Sofosbuvir-triphosphate, Sofosbuvir will be incorporated once out of 1500–14,000 opportunities. In evaluating both Remdesivir and Sofosbuvir, we do not know the physiological concentrations of the activated, triphosphate form of the nucleoside analogs, so our calculations of relative effectiveness assume the same nominal concentration (10 μM) for comparison. A higher concentration of the triphosphate form would make the analogs more effective, so it would be important to obtain estimates of the physiological concentrations of the activated form of each analog. Nonetheless, this analysis explains why Remdesivir is effective in spite of having to compete against a high concentration of ATP.

We also discovered that further addition of RTP on top of an incorporated RTP proceeds with k_{cat}/K_m values lower than that for the first RTP mostly because the rate constants for incorporation (k_{cat}) were 18-fold slower (Table 1). Nonetheless, it is unlikely that the RdRp would incorporate multiple sequential RTPs *in vivo*. If the probability of incorporating one RTP is 1/170 = 0.6%, then the probability of incorporating two RTPs would be 0.004%. However, we do not know the rate of any NTP incorporation after RTP, and the unusual structure of Remdesivir could alter the kinetics of the incorporation of subsequent nucleotides. It has been suggested that Remdesivir may be a delayed chain terminator (Gordon et al., 2020b) but that hypothesis needs to be tested more rigorously in studies including the proofreading exonuclease to establish the balance between incorporation versus excision.

The major unanswered question is the extent to which incorporated Remdesivir is resistant to excision by the proofreading exonuclease, NSP14/10 (Ferron et al., 2018; Ma et al., 2015). For example, Sofosbuvir is more effective *in vivo* than a similar analog, Mericitabine, because it escapes removal by a highly active ATP-dependent excision reaction catalyzed by the HCV RdRp. Although Mericitabine is incorporated 40-fold more efficiently than Sofosbuvir, it has a half-life of less than one minute after incorporation, but Sofosbuvir persists with a half-life of at least one day (Villalba et al., 2020). Although inefficiently incorporated, Sofosbuvir resists excision. In contrast, RTP is more efficiently incorporated at physiological ATP concentrations than Sofosbuvir is at physiological UTP concentrations, but the question as to the resistance to exonuclease removal of RMP remains to be addressed quantitatively. It appears that resistance to exonuclease hydrolysis will be the predominant determinant of the effectiveness of nucleoside analogs in combating coronavirus infections (Robson et al., 2020). More work is needed to define the kinetic and structural features governing the exonuclease activity and selectivity.

Two criteria must be met when optimizing the expression and purification of a polymerase. The first essential requirement is that the enzyme must catalyze polymerization with rates commensurate with the time estimated for replication *in vivo*. The second criterion is based on measurement of the fraction of enzyme that is active, often addressed by an active site titration. In our studies, average rates of polymerization of 300 base pairs per second are sufficient to replicate the coronavirus genome in less than 2 min. There is sufficient information within the available data to show that our enzyme is at least 60% active relative to the nominal concentration determined by absorbance measurements. This estimate is based on confidence contour analysis demonstrating that a good fit to the data could not be obtained if the active enzyme was less than 60% of the nominal concentration. However, our equilibrium titration (Figure 3D) shows that the RNA binding affinity is weak (1.8 ± 0.2 μM). It is quite likely that the addition of the exonuclease subunits (NSP14/10) will lead to tighter RNA binding, which will afford a more accurate measurement of the active site concentration.

Although tagged protein expression in insect cells has many advantages and has proven to be an effective platform for expression of the SARS-CoV-2 RdRp for *in vitro* studies, the expense and time required for protein production are major limitations. The use of tags to aid in purification also have their own set of limitations and often alter the activity of enzymes. We initially tested expression of untagged NSP12 in *E. coli* and found the protein was almost totally insoluble under a wide range of conditions; however, this was almost completely alleviated by co-expression of NSP12 with NSP7 and NSP8 along with the chaperones Tf, GroEL/ES. The large quantity of soluble protein expressed in *E. coli* greatly facilitated our search

for conditions to optimize efficient purification of the tag-free enzyme. We also found the untagged enzyme to have higher activity than the 8xhistidine-tagged NSP12/NSP7L8 construct used in previous studies (Shannon et al., 2020). The lower activity could be due to either the histidine tag, the linked NSP7L8, or both. However, we have not repeated these results nor done an extensive characterization of the tagged enzyme. Therefore, our results only raise a caution in using tagged enzymes.

Previous studies on the coronavirus RdRp have mostly relied on steady-state methods that are unreliable except for comparison of two competing substrates when both are present in the reaction mixture. As might be expected, K_m estimates from steady state are around 1,000-fold lower when comparing our results to prior publications (Gordon et al., 2020b; Shannon et al., 2020). It has long been known from assays of DNA polymerase kinetics that steady-state measurements underestimate k_{cat} and K_m by factors of 100- to 1000-fold due to the rate-limiting dissociation of the enzyme-DNA complex in the steady state (Johnson, 2010). Prior steady-state kinetic studies on the coronavirus RdRp have reported k_{cat} values that rely on units of velocity as fraction of primer extended per minute that are meaningless (Gordon et al., 2020a, 2020b). These errors are attributable to a relatively inactive enzyme preparation and faulty experimental design. Robust assays must be employed so that the experimentalist knows which step in the reaction limits the observed output for studies evaluating the kinetics and mechanism of inhibition. This is especially important for high-throughput screens. For example, inhibitors that slow the steady-state rate of RNA dissociation may not be very effective *in vivo*. More importantly, the effects of inhibitors that slow the rates of incorporation will be masked by the slow RNA dissociation rate when measurements are made in the steady state.

Our yield of active, untagged protein allowed us to perform pre-steady-state experiments with the enzyme concentration in large excess over the RNA to observe polymerization during a single turnover. This is the key for obtaining reliable and quantitative definition of the kinetics of each single nucleotide incorporation event during processive synthesis. Measurements under pre-steady-state conditions provides accurate estimates of k_{cat} and K_m for each single nucleotide incorporation event during processive polymerization. Traditionally, the terms k_{pol} and $K_{d,app}$ have been used to define parameters derived from pre-steady-state DNA polymerase kinetics. The initial intention was to distinguish the results from faulty steady-state analysis and to indicate that if nucleotide binding occurs in one step, then $K_{d,app}$ is a true K_d (Johnson, 1993). However, in general, the $K_{d,app} = K_m$ for each incorporation during processive synthesis even when nucleotide binding occurs in more than one step (Kellinger and Johnson, 2010). Our analysis will allow for accurate comparison of binding and incorporation parameters for nucleoside analogs relative to their natural counterparts in the search for new inhibitors.

Limitations of the Study

Although nucleotide incorporation kinetics are important for developing analogs that efficiently compete with natural nucleotides for incorporation, this is only half of the story. The SARS-CoV-2 has a proofreading exonuclease complex consisting of NSP10/14 (Bouvet et al., 2012; Ma et al., 2015) that can likely excise nucleotide analogs once incorporated into the RNA, thus preventing chain termination and allowing the virus to continue replication. Moreover, studies performed in the absence of the proofreading exonuclease (NSP14/10) do not address the fundamental questions as to whether Remdesivir resists excision. In the absence of the exonuclease, continued polymerization will eventually bury Remdesivir. On the other hand, it is conceivable that after incorporation of several nucleotides on top of Remdesivir, the terminal nucleotide will be excised by the exonuclease and then reinserted by the polymerase in a repeated cycle of addition and excision to reach a steady state. Future studies on nucleotide analog excision by the NSP10/14 proofreading exonuclease complex, using pre-steady-state methods, will be crucial to analyze in conjunction with the polymerization data to discover the most useful antivirals for treating SARS-CoV-2 infection.

Resource Availability

Lead Contact

Questions about or requests for materials should be directed to and will be fulfilled by the Lead Contact, Kenneth A. Johnson (kajohnson@mail.utexas.edu).

Materials Availability

Plasmids used in this study have been deposited to Addgene: pcl$^{ts,ind+}$-(NSP8) – 160656 pQE-(NSP12)-pcl$^{ts,ind+}$-(NSP7-NSP8) – 160540.

Data and Code Availability

The published article includes all datasets generated or analyzed during this study.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101849.

ACKNOWLEDGMENTS

This work was supported by grants from NIAID (1R01AI110577 to KAJ) and the Welch Foundation (F-1604 to KAJ). Remdesivir triphosphate was a gift kindly provided by Brian Schultz and Joy Feng at Gilead Sciences.

AUTHOR CONTRIBUTIONS

Conceptualization, K.A.J., T.L.D., and N.Z.H.; Methodology, K.A.J., T.L.D., and N.Z.H.; Formal Analysis, K.A.J. and T.L.D.; Investigation, T.L.D. and N.Z.H.; Resources, K.A.J.; Writing—Original Draft, T.L.D.; Writing—Review and Editing, K.A.J., T.L.D., and N.Z.H.; Visualization, T.L.D.; Supervision, K.A.J.; Project Administration, K.A.J.; Funding Acquisition, K.A.J.

DECLARATION OF INTERESTS

K.A.J. is president of KinTek Corporation, which provided the quench flow instrument and KinTek Explorer software used in these studies.

Received: September 24, 2020
Revised: October 30, 2020
Accepted: November 18, 2020
Published: December 18, 2020

REFERENCES

Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., and Decroly, E. (2012). RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. Proc. Natl. Acad. Sci. U. S. A. *109*, 9372.

Brandis, J.W., and Johnson, K.A. (2009). High-cell density shake-flask expression and rapid purification of the large fragment of Thermus aquaticus DNA polymerase I using a new chemically and temperature inducible expression plasmid in Escherichia coli. Protein Expr. Purif. *63*, 120–127.

Chiu, J., Tillett, D., and March, P.E. (2006). Coexpression of the subunits of T7 DNA polymerase from an artificial operon allows one-step purification of active gp5/Trx complex. Protein Expr. Purif. *47*, 264–272.

Deval, J., Symons, J.A., and Beigelman, L. (2014). Inhibition of viral RNA polymerases by nucleoside and nucleotide analogs: therapeutic applications against positive-strand RNA viruses beyond hepatitis C virus. Curr. Opin. Virol. *9*, 1–7.

Donlin, M.J., Patel, S.S., and Johnson, K.A. (1991). Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. Biochemistry *30*, 538–546.

FDA. (2020). FDA approves first treatment for COVID-19. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19.

Ferron, F., Subissi, L., Silveira De Morais, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., et al. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. Proc. Natl. Acad. Sci. U. S. A. *115*, E162–E171.

Gordon, C.J., Tchesnokov, E.P., Feng, J.Y., Porter, D.P., and Gotte, M. (2020a). The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. J. Biol. Chem. *295*, 4773–4779.

Gordon, C.J., Tchesnokov, E.P., Woolner, E., Perry, J.K., Feng, J.Y., Porter, D.P., and Götte, M. (2020b). 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.

Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., Feldt, T., Green, G., Green, M.L., Lescure, F.-X., et al. (2020). Compassionate use of remdesivir for patients with severe covid-19. N. Engl. J. Med. *382*, 2327–2336.

Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. Nature *584*, 154–156.

Hurwitz, S.J., and Schinazi, R.F. (2012). Practical considerations for developing nucleoside reverse transcriptase inhibitors. Drug Discov. Today Technol. *9*, e183–e193.

Johnson, A.A., Ray, A.S., Hanes, J., Suo, Z.C., Colacino, J.M., Anderson, K.S., and Johnson, K.A. (2001). Toxicity of antiviral nucleoside analogs and the human mitochondrial DNA polymerase. J. Biol. Chem. *276*, 40847–40857.

Johnson, K.A. (1993). Conformational coupling in DNA polymerase fidelity. Annu. Rev. Biochem. *62*, 685–713.

Johnson, K.A. (2009). Fitting enzyme kinetic data with KinTek global kinetic explorer. Methods Enzymol. *467*, 601–626.

Johnson, K.A. (2010). The kinetic and chemical mechanism of high-fidelity DNA polymerases. Biochim. Biophys. Acta *1804*, 1041–1048.

Johnson, K.A. (2019). Kinetic Analysis for the New Enzymology: Using Computer Simulation to Learn Kinetics and Solve Mechanisms (KinTek Corporation).

Johnson, K.A., Simpson, Z.B., and Blom, T. (2009a). FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. Anal. Biochem. *387*, 30–41.

Johnson, K.A., Simpson, Z.B., and Blom, T. (2009b). Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. Anal. Biochem. *387*, 20–29.

Kellinger, M.W., and Johnson, K.A. (2010). Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. Proc. Natl. Acad. Sci. U. S. A. *107*, 7734–7739.

Kirchdoerfer, R.N., and Ward, A.B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat. Commun. *10*, 2342.

Kuchta, R.D., Mizrahi, V., Benkovic, P.A., Johnson, K.A., and Benkovic, S.J. (1987). Kinetic mechanism of DNA polymerase I (Klenow). Biochemistry *26*, 8410–8417.

Lu, M., Ngo, W., Mei, Y., Munshi, V., Burlein, C., Loughran, M.H., Williams, P.D., Hazuda, D.J., Miller, M.D., Grobler, J.A., et al. (2010). Purification of untagged HIV-1 reverse transcriptase by affinity chromatography. Protein Expr. Purif. *71*, 231–239.

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., and Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. Proc. Natl. Acad. Sci. U. S. A. *112*, 9436.

Nishihara, K., Kanemori, M., Yanagi, H., and Yura, T. (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. Appl. Environ. Microbiol. *66*, 884–889.

Pardo, J., Shukla, A.M., Chamarthi, G., and Gupte, A. (2020). The journey of remdesivir: from Ebola to COVID-19. Drugs Context *9*, 2020-4-14.

Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., et al. (2020). Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. Cell Rep. *31*, 107774.

Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., and Ng, W.L. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. Mol. Cell *79*, 710–727.

Shannon, A., Selisko, B., Le, N.-T.-T., Huchting, J., Touret, F., Piorkowski, G., Fattorini, V., Ferron, F., Decroly, E., Meier, C., et al. (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.

G. Skowron, and R. Odgen, eds. (2006). Reverse Transcriptase Inhibitors in HIV/AIDS Therapy, First edition (Humana Press), pp. 1–26.

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. Proc. Natl. Acad. Sci. U. S. A. *111*, E3900.

te Velthuis, A.J.W., van den Worm, S.H.E., and Snijder, E.J. (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.

Traut, T.W. (1994). Physiological concentrations of purines and pyrimidines. Mol. Cell. Biochem. *140*, 1–22.

Villalba, B., Li, J., and Johnson, K.A. (2020). Resistance to excision determines efficiency of hepatitis C virus RNA-dependent RNA polymerase inhibition by nucleotide analogs. J. Biol. Chem. *295*, 10112–10124.

WHO. (2019). WHO Coronavirus disease (COVID-19) pandemic. In COVID-19 Weekly Epidemiological Update (WHO). https://www.who.int/publications/m/item/weekly-epidemiological-update—27-october-2020.

Yin, W., Mao, C., Luan, X., Shen, D.-D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science *368*, 1499.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.

iScience, Volume *23*

Supplemental Information

Remdesivir Is Effective in Combating COVID-19 because It Is a Better Substrate than ATP for the Viral RNA-Dependent RNA Polymerase

Tyler L. Dangerfield, Nathan Z. Huang, and Kenneth A. Johnson

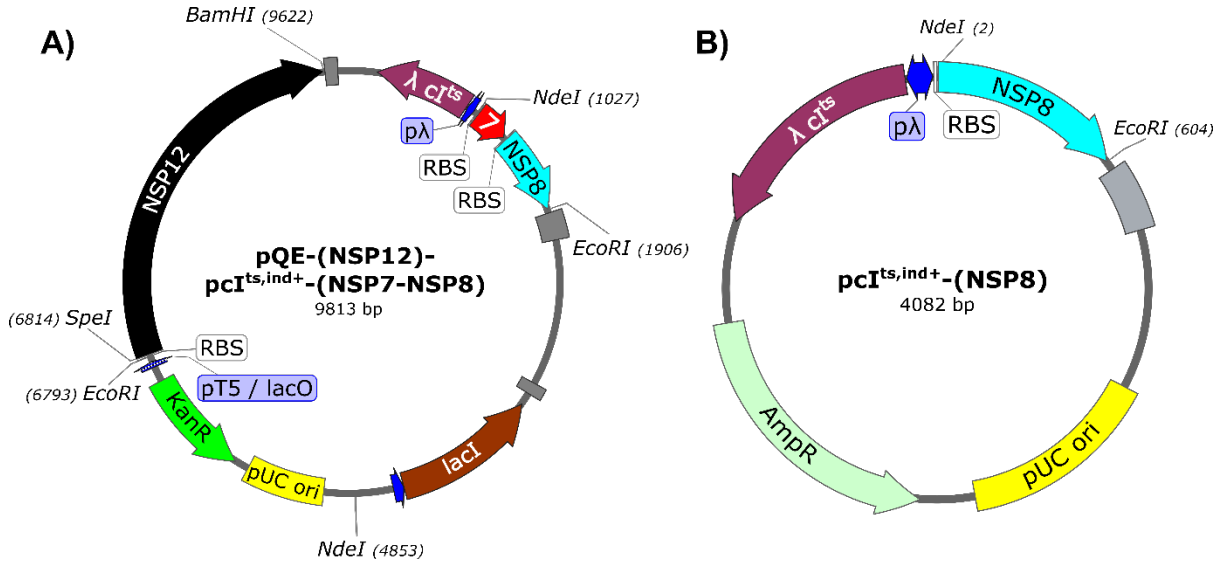


Figure S1: Expression plasmids, related to Figure 1. These maps show the plasmids that were used for *E. coli* expression of SARS-CoV-2 replication complex genes. Restriction sites are shown on the outside of the plasmid while unique features are shown on the plasmid or on the inside of the plasmid. A) Plasmid for NSP12/8/7 co-expression. This plasmid contains the NSP12 gene (black) under the T5 promoter and lac operator (blue, pT5/lacO), controlled by the on-board lacI gene (brown) under the placI promoter (blue) and induced with addition of IPTG. The NSP7 (7, red) and NSP8 (cyan) genes are arranged in a bicistronic operon, with ribosome binding sites (RBS, white) upstream of each gene—expression is under control of the lambda promoter system. The plasmid also features the kanamycin resistance gene (KanR, green) and a high-copy number pUC origin of replication (pUC ori, yellow). B) NSP8 expression plasmid. The plasmid shown was used for NSP8 (cyan) expression in *E. coli* under control of the lambda promoter (pλ, blue) and *E. coli* rrnBT_1T_2 terminator (grey) with an on-board temperature sensitive, recA cleavable $cI^{ts,ind+}$ repressor variant (λ $cI^{ts,ind+}$, purple). This plasmid also features a pUC origin of replication (pUC ori, yellow) and an ampicillin resistance gene (AmpR, green). Expression of NSP8 was induced by shifting the culture temperature from 30°C to 42°C for 30 minutes, followed by 3 hours of expression at 38°C.

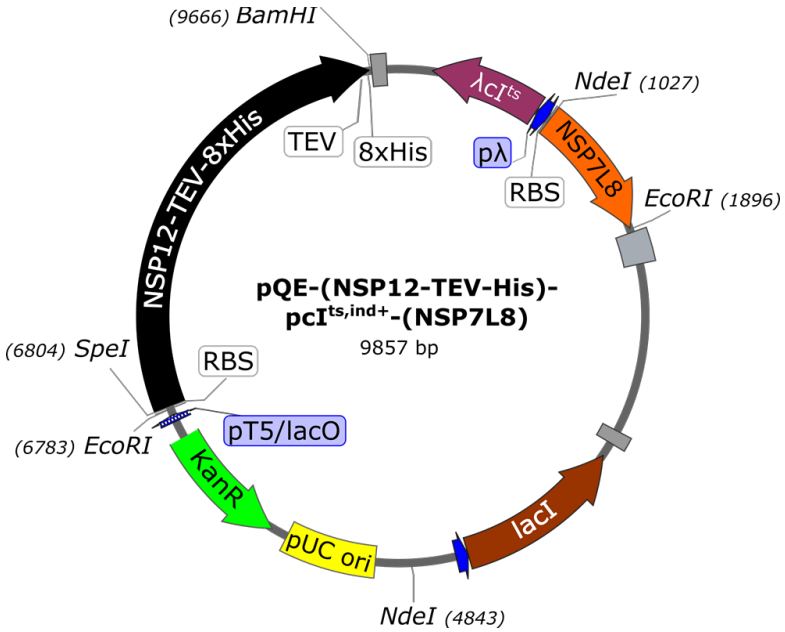


Figure S2: Histidine tagged NSP12/NSP7L8 expression plasmid, related to Figure 1. Contains the NSP12-TEV-8xHis gene (black) cloned under the T5 promoter/lac operator (pT5/lacO, blue), as well as the NSP7L8 gene (orange) containing a GSGSGS linker between NSP7 and NSP8 under control of the lambda promoter (pλ, blue). This plasmid also contains the lacI gene (brown) and the λcI$^{ts,ind+}$ repressor (purple) so expression can be performed in most *E. coli* strains. The kanamycin resistance gene (KanR, green) for selection and pUC origin of replication (yellow) are also on the plasmid.

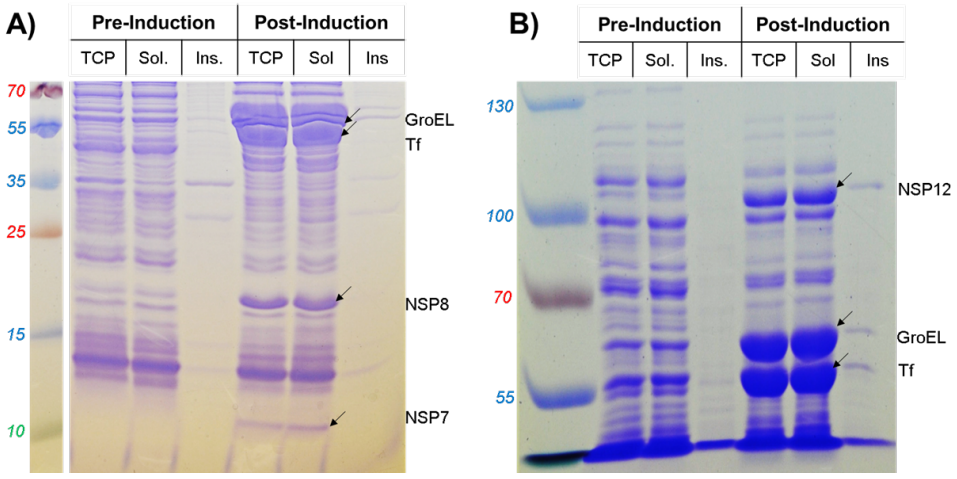


Figure S3: Soluble expression of NSP12/7/8 complex, related to Figure 1. Molecular weights of the protein ladder in kDa are given to the left of each gel in the color of its corresponding band. Samples corresponding to the total cell protein fraction (TCP), the soluble protein fraction (Sol.), and the insoluble protein fraction (Ins.) were from samples of *E. coli* cells before induction (Pre-Induction) and at the time of harvesting the pellet (Post-Induction). Molecular weights of proteins of interest are: NSP7, 9.4 kDa; NSP8, 22 kDa; NSP12, 107 kDa; Tf, 56 kDa; GroEL, 60 kDa. A) 15% SDS-PAGE of RdRp complex expression. The proteins of interest are labeled to the right of the gel with arrows pointing to the relevant bands B) 6% SDS-PAGE of RdRp complex expression. Proteins of interest are labeled as in (A).

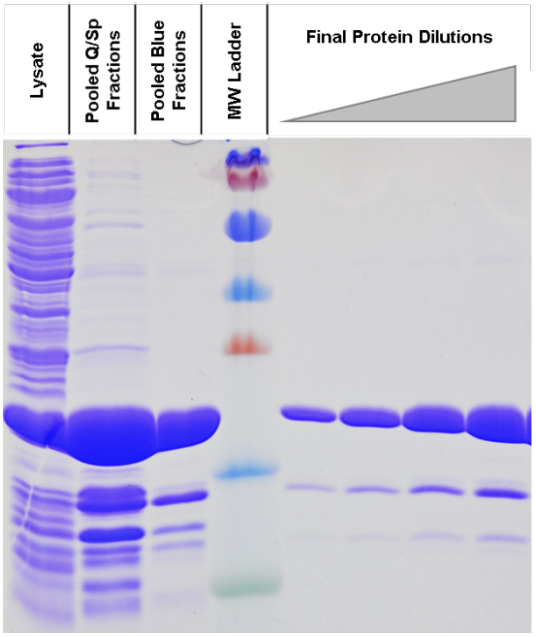


Figure S4: Expression and purification of NSP8, related to Figure 1. Bands in the molecular weight ladder are as follows in kDa in order from the bottom green band to the top blue band: 10, 15, 25, 35, 55, 70, 100. The leftmost lane contains a sample of the clarified lysate (soluble fraction) before loading on the Q/SP columns. The next two lanes to the right are samples collected after the Q/SP columns and after the HiTrap Blue-FF column. The lanes to the right of the molecular weight ladder are dilutions of the final protein, after the size exclusion column.

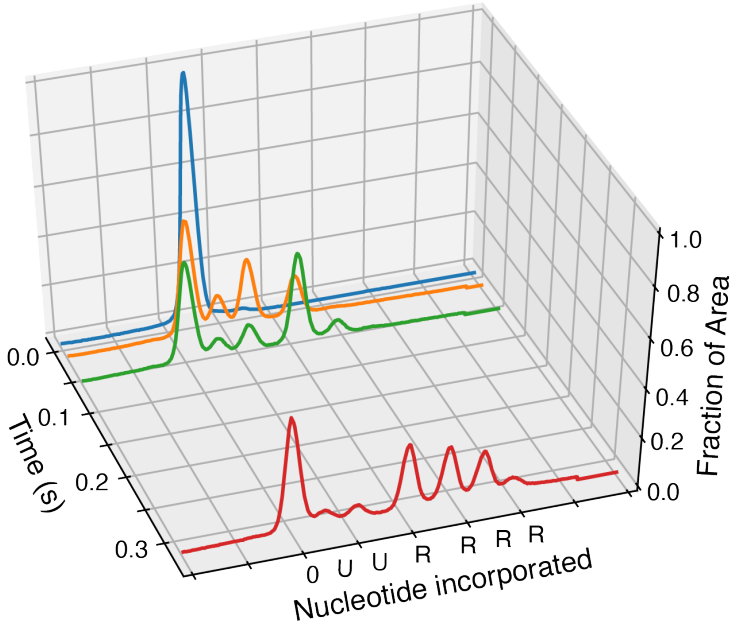


Figure S5: Sample chromatograms for Remdesivir incorporation experiment, related to Figure 4. A 3-D plot of chromatograms from different time points is shown with reaction time on the x axis, nucleotide position (retention

time relative to internal standard) on the y axis, and fractional intensity on the z axis. Peaks at each time point are sufficiently resolved and there is no indication of re-annealing of the RNA during electrophoresis.

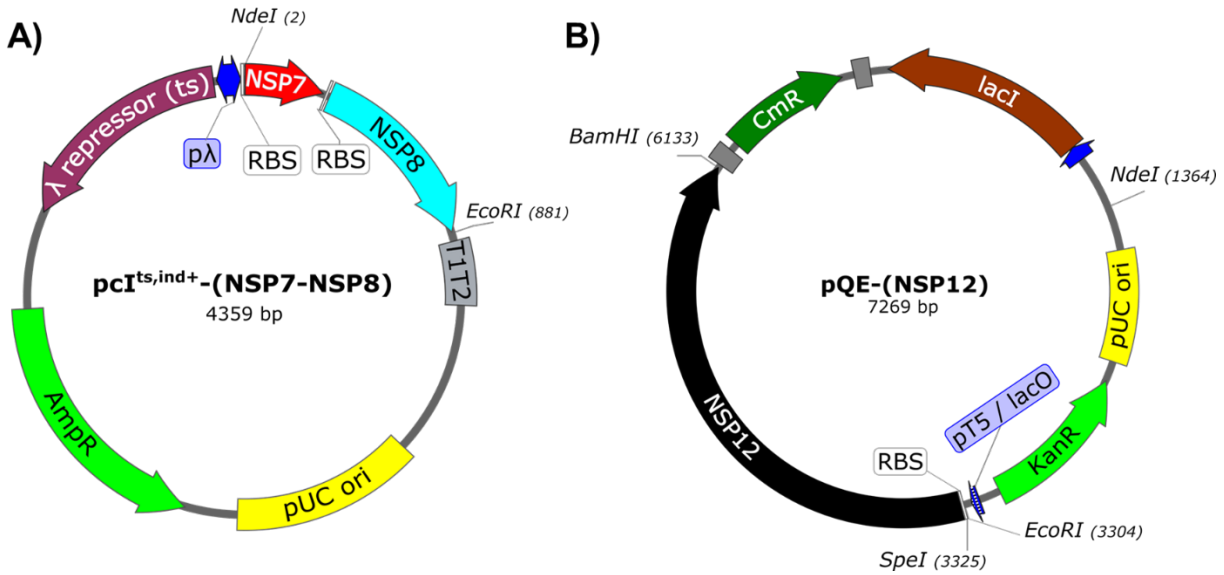


Figure S6: Precursor expression plasmids, related to Figure 1. A) Bicistronic NSP7/NSP8 expression plasmid. This plasmid contains the bicistronic NSP7-NSP8 operon cloned under the lambda promoter (pλ, blue) and *E. coli* rrnBT_1T_2 terminator (T1T2, grey), used with pQE-(NSP12) to make the final expression plasmid for the replication complex. The bicistronic operon contains ribosome binding sites (RBS, white) upstream of both NSP7 (red) and NSP8 (cyan) genes and a short linker between the stop codon for NSP7 and the RBS for NSP8. This plasmid also contains the temperature or nalidixic acid inducible λcI$^{ts,ind+}$ repressor (λ repressor, purple) cloned on the plasmid so expression can be performed in most *E. coli* strains. It also contains the ampicillin resistance gene (AmpR, green) for selection and the high copy number pUC origin of replication. Restriction sites of interest are given on the outside of the plasmid. B) NSP12 expression plasmid. This plasmid contains the codon optimized NSP12 gene (black) under control of the T5 promoter/lac operator (pT5/lacO, blue), used for constructing the final expression plasmid for co-expression with NSP7 and NSP8. This plasmid also features a pUC origin of replication (yellow), the chloramphenicol resistance gene (CmR, dark green), the kanamycin resistance gene (KanR, light green), and lacI repressor gene (brown). Restriction sites of interest are given on the outside of the plasmid map.

Table S1. Comparing 64% and 80% active enzyme, related to Figure 3

Parameter	64% active	80% active
k_{-1} (s^{-1})	10800 ± 800	10800 ± 800
k_2 (s^{-1})	370 ± 20	370 ± 20
k_{-3} (s^{-1})	17000 ± 3000	17000 ± 3000
k_4 (s^{-1})	380 ± 50	380 ± 50
k_{off} (s^{-1})	0.0132 ± 0.001	0.0142 ± 0.001

Rate constants and standard errors derived in fitting shown in Figure 3, where k_1 and k_3 were locked at 100 μM^{-1} s^{-1}.

Transparent Methods

METHOD DETAILS

Cloning:

All restriction enzymes, T4 ligase, T4 DNA polymerase, and Phusion DNA polymerase were purchased from New England Biolabs. Oligonucleotides used for cloning were synthesized by Integrated DNA Technologies with standard desalting and used without further purification. Plasmid figures were prepared with SnapGene® software (from GSL Biotech; available at snapgene.com). Genes were synthesized with codon optimization for *E. coli* by GenScript in pUC57 backbones. NSP12 was synthesized with an *NdeI* site integrated into the start codon and an *EcoRI* site at the 3' end of the gene, and the plasmid containing the gene was designated pUC57-(NSP12). Only the methionine from the start codon (ATG) was added to the wild-type NSP12 gene. NSP7 and NSP8 genes were synthesized as a bicistronic operon with an *NdeI* site integrated into the NSP7 start codon, a short linker region containing a ribosome binding site between the stop codon for NSP7 and the start codon for NSP8, and an *EcoRI* site at the 3' end of the NSP8 gene. The plasmid for cloning with the bicistronic operon was designated pUC57-(NSP7-NSP8). The bicistronic NSP7-NSP8 operon was cloned from pUC57-(NSP7-NSP8) into the $pcl^{ts,ind+}$ backbone (Brandis and Johnson, 2009) under the λ promoter, using *NdeI* and *EcoRI* restriction sites to yield $pcl^{ts,ind+}$-(NSP7-NSP8) shown in Figure S6A. The plasmid $pcl^{ts,ind+}$-(NSP8) in Figure S1B was made by PCR from $pcl^{ts,ind+}$-(NSP7-NSP8), using primers to delete the NSP7 gene. The NSP12 gene was cloned from pUC57-(NSP12) into the pQE-30 backbone with *SpeI* and *BamHI* sites, added by PCR, at the 5' and 3' ends of the NSP12 gene, respectively, to yield pQE-(NSP12) shown in Figure S6B. Ligation independent cloning (Aslanidis and de Jong, 1990) was used to join amplicons from $pcl^{ts,ind+}$-(NSP7-NSP8) and pQE-(NSP12), to form pQE-(NSP12)-$pcl^{ts,ind+}$-(NSP7-NSP8) shown in Figure S1A. The ampicillin resistance and chloramphenicol resistance genes were removed during this step, leaving only the kanamycin resistance gene on the final plasmid.

The following steps were performed to make the histidine-tagged NSP12 complex/NSP7L8 complex: The plasmid $pcl^{ts,ind+}$-(NSP7-NSP8) was amplified by PCR, using primers to replace the linker region between NSP7 and NSP8 with a GSGSGS linker, to yield $pcl^{ts,ind+}$-(NSP7L8). The plasmid pQE-(NSP12) was amplified by PCR using primers to add the following amino acid sequence to the C terminus of NSP12, as previously reported (Shannon et al., 2020): GGSENLYFQGHHHHHHHH, consisting of a 3 amino acid linker, a TEV protease site, and 8 histidine residues. The resulting plasmid was designated pQE-(NSP12-TEV-8xHis). A fragment of pQE-(NSP12-TEV-8xHis) and a fragment from $pcl^{ts,ind+}$-(NSP7L8) were amplified by PCR and joined with ligation independent cloning, as above, to yield pQE-(NSP12-TEV-8xHis)-$pcl^{ts,ind+}$-(NSP7L8) shown in Figure S2. pG-Tf2 (Nishihara et al., 2000) was purchased from Takara Biosciences. All constructs were verified by Sanger sequencing, which was performed by the University of Texas at Austin sequencing facility.

Protein expression:

Large scale expression protocols are given below for each construct. Samples for the total cell protein, soluble and insoluble fractions for gel electrophoresis analysis shown in the supplemental information were prepared as previously described (Brandis and Johnson, 2009).

NSP8 expression:
BL21 *E. coli* harboring pcI$^{ts,ind+}$-(NSP8) were inoculated into Terrific Broth (2.4% yeast extract, 2% tryptone, 0.4% glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4) with 100 μg/ml ampicillin and grown overnight at 30°C with shaking. The overnight starter culture (8 ml) was used to inoculate 1 liter of Terrific Broth with ampicillin in a 2.8 liter baffle-bottomed flask. Cells were grown at 30°C with shaking at 250 rpm until the OD_{600} reached 2. NSP8 expression was induced by incubation of the cells at 42°C with shaking for 20 minutes, then the cultures were incubated for an additional 3 hours with shaking at 38°C. Cells were harvested by centrifugation at 6,000 x g for 20 minutes at 4°C, flash frozen, and stored at -80°C.

NSP12/7/8 co-expression:
BL21 *E. coli* were transformed with pG-Tf2 with 20 ng/μl chloramphenicol for selection and made electrocompetent with water and glycerol washes (Au - Gonzales et al., 2013). The resulting cells are designated here as *E. coli* BL21/pG-Tf2. These cells were transformed with pQE-(NSP12)-pcI$^{ts,ind+}$-(NSP7-NSP8), plated on kanamycin/chloramphenicol plates at 30 μg/ml and 20 μg/ml, respectively, and incubated at 30°C. Colonies were inoculated into an 8 ml overnight starter culture which was later inoculated into 1 liter of Terrific Broth + kanamycin + chloramphenicol in a 2.8 liter baffle bottomed flask. The cells were incubated at 30°C with shaking at 250 rpm. When the OD_{600} reached 2, tetracycline was added to 10 ng/ml and the cells were incubated at 16°C with shaking for 20 minutes. IPTG and nalidixic acid were added to 0.5 mM and 50 μg/ml, respectively, and expression was carried out at 16°C with shaking for 16 hours. Cells were harvested and stored as above. Co-expression of histidine-tagged NSP12 and NSP7L8 was performed as for the NSP12/7/8 complex, but with *E. coli* BL21/pG-Tf2 harboring pQE-(NSP12-TEV-8xHis)-pcI$^{ts,ind+}$-(NSP7L8). A nalidixic acid (Sigma Aldrich) stock solution at 50 mg/ml was prepared fresh daily in 0.3 M NaOH.

Protein purification:

Unless otherwise noted, all steps were performed at 4°C. All columns/chromatography resins were purchased from GE healthcare and all dialysis tubing was from Spectra-Por.

NSP8 purification:
Cells were thawed and resuspended in Lysis Buffer (50 mM Tris-HCl pH 8, 2.5 mM EDTA, 150 mM NaCl, 1 mM DTT) at 5 ml buffer per gram cells. Phenylmethylsulfonyl fluoride (PMSF) and lysozyme were added to 10 mM and 0.3 mg/ml, respectively. The lysate was stirred at room temperature for 15 minutes, then sonicated on ice for 20 minutes. Sodium deoxycholate (Sigma) was added to 0.1%, NaCl was added to 0.5 M, and the lysate was clarified by centrifugation at 75,000 x g for 45 minutes. The supernatant was filtered using a 0.2 μm syringe filter (Whatman) and diluted with Dilution Buffer (50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 10% glycerol, 1 mM DTT) until the conductivity of the sample was equal to the conductivity of Buffer A (50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 50 mM NaCl, 10% glycerol, 1 mM DTT). The sample was then loaded on 2 x 5 ml Q Sepharose-FF columns with the outlet connected to 2 x 5 ml SP Sepharose-FF columns. The columns were washed with Buffer A and the Q columns were then removed. Proteins bound to the SP columns were eluted with a 200 ml gradient from 10 – 80% Buffer B (Buffer A with 1 M NaCl). Fractions containing NSP8 were pooled, diluted with Dilution Buffer to the conductivity of Buffer A, then loaded onto a 5 ml HiTrap Blue-FF column. The column was washed with 10% Buffer B, then bound proteins were eluted with a 50 ml gradient from 10 – 80% Buffer B. Fractions containing NSP8 were pooled and concentrated with Amicon Ultra centrifugal concentrators with a 10 kDa molecular weight cut-off (MWCO). The concentrated sample was injected onto a Superdex 200 10/300 GL column, equilibrated in 10% Buffer B, and eluted with the same buffer. Fractions containing NSP8 were pooled, concentrated by ultrafiltration as above, then dialyzed into Storage Buffer (50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 20 mM KCl, 1 mM DTT, 50% glycerol) with 1 kDa MWCO dialysis tubing. Protein concentration was determined by absorbance at 280 nm using the extinction coefficient 19,940 M^{-1} cm^{-1} calculated based on the amino acid sequence of NSP8. Aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

NSP12 8xHis/NSP7L8 complex purification:
Cells were thawed and re-suspended in Lysis Buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 30 mM imidazole, 10% glycerol, 2.5 mM DTT) at 5 ml per gram of cells. One EDTA-free protease inhibitor tablet (Roche) was added with 10 mM PMSF. Sodium deoxycholate was added to 0.1 % and lysozyme was added to 0.3 mg/ml. The lysate was stirred at room temperature for 15 minutes, sonicated on ice for 20 minutes, then clarified by centrifugation at 75,000 x g for 1 hour at 4°C. The sample was filtered through a 0.2 μm syringe filter, then loaded onto a 3 ml HisTrap-FF column, equilibrated in Buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 30 mM imidazole, 10% glycerol, 2.5 mM DTT). The column was washed with Buffer A until the UV reached baseline, then eluted with a 30 ml linear gradient from 0 – 100 % Buffer B (Buffer A with 500 mM imidazole). Fractions containing the NSP12-His/NSP7L8 complex were pooled, concentrated with Amicon Ultracel 15 centrifugal concentrators (30 kDa MWCO), then dialyzed into Storage Buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.25 mM EDTA, 2.5 mM DTT, 50% glycerol) at 4°C with 50 kDa MWCO dialysis tubing. Protein concentration was determined by absorbance at 280 nm using the extinction coefficient 162,850 M^{-1} cm^{-1}, calculated based on the amino acid sequence of each polypeptide assuming a stoichiometry of one NSP12-His to one NSP7L8. The protein was divided into aliquots, flash frozen in liquid nitrogen and stored at -80°C.

NSP12/7/8 complex purification:
Cells were thawed and re-suspended in Lysis Buffer (50 mM Tris-HCl pH 8, 2 mM EDTA, 250 mM NaCl, 10% glycerol, 5 mM DTT) at 5 ml/gram of cells. One protease inhibitor cocktail tablet (Roche) was added along with PMSF, sodium deoxycholate, and lysozyme to 10 mM, 0.1%, and 0.1 mg/ml respectively. The lysate was stirred at room temperature for 15 minutes then sonicated and clarified by centrifugation as above for the histidine-tagged complex. The sample was filtered through a 0.2 μm syringe filter then diluted with Dilution Buffer (50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 10% glycerol, 5 mM DTT) until the conductivity of the sample was the same as Buffer A (50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 50 mM NaCl, 10% glycerol, 2 mM DTT).The sample was loaded onto a 30 ml Q-Sepharose-FF column and unbound proteins were removed by washing with Buffer A. Bound proteins were eluted with a 300 ml gradient of 0 – 100% Buffer B (Buffer A with 1 M NaCl). Fractions containing the NSP12/7/8 complex were pooled, diluted with Dilution Buffer, and passed through a 5 ml HiTrap Blue-FF column. The flow-through was loaded onto a 5 ml Heparin Sepharose 6-FF column and unbound proteins were removed by washing with Buffer A. Bound proteins were eluted with a 60 ml gradient of 0 – 100% Buffer B. Fractions containing the NSP12/7/8 complex were pooled and diluted with Dilution Buffer. The sample was flowed through a 5 ml HiTrap SP-FF column equilibrated in Buffer A, and the flow-through was concentrated with Amicon Ultra centrifugal concentrators (30 kDa MWCO). The concentrated sample was injected onto a Superdex 200 10/300 GL column, equilibrated in Buffer A, and proteins were separated in the same buffer. Fractions containing the NSP12/7/8 complex were pooled and dialyzed into Storage Buffer (50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 50 mM NaCl, 50% glycerol, 1 mM DTT) with 1 kDa MWCO dialysis tubing. Protein concentration was determined by absorbance at 280 nm using the extinction coefficient 181,300 M^{-1} cm^{-1}, calculated from the amino acid sequence assuming one NSP12, one NSP7, and two NSP8 polypeptides per complex (Hillen et al., 2020; Yin et al., 2020). Aliquots were flash frozen in liquid nitrogen and stored at -80°C. Approximately 7 mg of NSP12/7/8 complex was obtained after purification from a one liter expression culture.

Preparation of RNA substrates:

RNA oligonucleotides shown in Figure 1A were synthesized by Integrated DNA Technologies with RNase-free HPLC purification. Oligos were resuspended in Annealing Buffer (10 mM Tris-HCl pH 7, 50 mM NaCl, 0.1 mM EDTA) and the concentration was determined by absorbance at 260 nm using the extinction coefficients 222,360 M^{-1} cm^{-1} and 403,100 M^{-1} cm^{-1} for the FAM-20 nt primer and 40 nt template, respectively. The two oligos were annealed by mixing at a 1:1 molar ratio, heating to 75°C, and cooling slowly to room temperature over approximately 2 hours. Oligonucleotides were stored at -20°C.

Kinetic methods:

Nucleotides were purchased from New England Biolabs. Heparin was purchased from Sigma Aldrich. Chemical structures for figures were prepared with the software ChemDraw and figures were prepared in Inkscape. Remdesivir triphosphate (GS-443902) was provided by Gilead Sciences. Concentration of RTP was determined by absorbance

at 245 nm using the extinction coefficient 24,100 M^{-1} cm^{-1}. The extinction coefficient was measured based on an accurate concentration determined by NMR with an internal formamide standard. Experiments were performed at 37 °C in Reaction Buffer (40 mM Tris-HCl pH 7, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT). Quench flow experiments were performed using a KinTek RQF-3 with reaction buffer in the drive syringes and unless otherwise noted, 0.6 M EDTA in the quench syringe for a final concentration of 0.2 M after quenching.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of reaction products

Kinetics time points were analyzed by capillary electrophoresis on an ABI 3130xl Genetic Analyzer with a 36 cm array and nanoPOP-6 polymer (Molecular Cloning Laboratories) at 65°C. Samples were prepared for analysis by mixing 1 μl of sample with 10 μl of HiDi formamide (ThermoFisher) containing a 28 nt Cy3-labeled DNA oligo internal standard for sizing. Samples were injected for 6 – 12 seconds, depending on the experiment, at 3.6 kV. Peaks area was determined with GeneMapper software and sizing and quantification was performed with a program written in-house.

Analysis of kinetic data

Data fitting/analysis was performed using KinTek Explorer simulation and data fitting software (Johnson, 2009; Johnson et al., 2009b) v10 (www.kintekexplorer.com). This software was also used in preparing figures for kinetic data. Conventional data fitting was performed in the software using built in functions. The equation for a single exponential is $y = A_0 + A_1(1 - \exp(-b_1 t))$, where A_0 is the y-value at time zero, A_1 is the amplitude, b_1 is the decay rate and t is time. The equation for a hyperbola is $y = A_0 + A_1[S]/(K_d+[S])$, where A_0 is the y value at time zero, A_1 is the amplitude and [S] is the x-axis variable. After global data fitting, confidence contour analysis was performed for each data set using the FitSpace function in KinTek Explorer (Johnson et al., 2009a). Parameter boundaries are reported in Table 1, using a χ^2 cut-off in the FitSpace calculation recommended by the software as a limit based on the number of parameters and number of data points used in the fitting of each data set (Johnson, 2019).

From intrinsic rate constants derived in data fitting, we calculated k_{cat}, K_m and k_{cat}/K_m. For example, for the first UTP, we use the following, with $k_1 = 100$ $\mu M^{-1}s^{-1}$ (for all fits).

$k_{cat} = k_2$
$K_m = (k_{-1} + k_2)/k_1$
$k_{cat}/K_m = k_1k_2/(k_{-1}+k_2)$

Estimating active enzyme concentration. Estimates of enzyme concentration based on absorbance or dye binding assays can have systematic errors and they do not establish the fraction of enzyme that is active. For these reasons, it is always important to estimate the concentration active sites relative the nominal enzyme concentration obtained from measurements of protein concentration. Because of COVID-19, we have been unable to obtain sufficient RNA templates to perform experiments with duplex RNA in excess of enzyme to provide a better estimate the concentration of enzyme active sites. Nonetheless, our existing data cannot be interpreted without a relatively high fraction of enzyme concentration. Therefore, to extract estimates of the minimal enzyme concentration necessary to account for the kinetic data, we added a step to the pathway to allow for a very slow equilibration between active (E) and (X) dead enzyme as shown in the pathway given in Figure 3, reproduced here.

$$X \underset{K_{active}}{\overset{slow}{\rightleftharpoons}} E + R_{20} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} ER_{20} \underset{k_{-1}}{\overset{k_1[UTP]}{\rightleftharpoons}} ER_{20}.UTP \xrightarrow{k_2} ER_{21} \underset{k_{-3}}{\overset{k_3[UTP]}{\rightleftharpoons}} ER_{21}.UTP \xrightarrow{k_4} ER_{22}$$

We do not imply that the reaction between E and X is reversible, only that we can model it as such to take advantage of the power of global data fitting and confidence contour analysis to estimate the fraction of enzyme in the active state. We set the rate constants for E – X equilibration to be on the order of 1e-05 s^{-1} and allowed the reaction to equilibrate (computationally) on for 500,000 s (~6 days). The equilibrated enzyme was then mixed with RNA for 250 s, which is long enough to allow RNA binding only to the E form without perturbing the E – X equilibration. In a third

mixing step in the software, we then added nucleotide to follow the fast reaction (< 1 s). In fitting the data, we locked the rate constant for E $\rightarrow$ X at 1e-05 s^{-1} and then allowed the reverse rate constant to float in globally fitting the all of the data shown in Figure 3. Confidence contour analysis showed that each of the rate constants governing binding and incorporation of the first and second UTP were well constrained by the data. Figure 3F shows the results from systematically varying the rate constant for the X $\rightarrow$ E reaction in order to compute a hypothetical equilibrium constant, K_{active} = [E]/[X] so we than calculate the fraction of active enzyme as [E] = $K_{active}/(1+K_{active})$, which we plot in the confidence contour shown in Figure 3F. These data show that there is a clear threshold at 60% active enzyme (relative to the nominal concentration). Below this value, the data fit gets significantly worse as judged by an increase in χ^2 (decrease in χ^2_{min}/χ^2). With increases in the fraction of enzyme that is active past 0.6, the data fit gradually gets better, but there are no significant changes in the value of χ^2 or in the values of parameters derived in fitting data. It should also be noted that the calculated second order rate constant for RNA binding shows an inverse correlation with the fraction of active enzyme over the narrow range from 60% to 100% active enzyme. This is consistent with an understanding that in experiments with enzyme in excess, the pseudo-first order rate constant for RNA binding will be given by the product of k_{on}[E]. This also leads to a modest change in the rate constant for RNA binding.

We then locked the fraction of active enzyme at 80% and then fit the titration (Figure 3D) and RNA dissociation (Figure 3E) globally to define the rate constants for RNA binding and dissociation, with confidence contours shown in Figure 3G. Similarly, we fit the UTP binding and incorporation results (Figures 3A, B, and C) with 80% active enzyme to get the confidence contours shown in Figure 3H. Data fitting for all of the figures in the paper, including Figure 3 was obtained using a nominal active enzyme concentration equal to 80% of the value estimated by absorbance measurements.

Supplemental References

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Research *18*, 6069-6074.

Au - Gonzales, M.F., Au - Brooks, T., Au - Pukatzki, S.U., and Au - Provenzano, D. (2013). Rapid Protocol for Preparation of Electrocompetent Escherichia coli and Vibrio cholerae. JoVE, e50684.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
BL21 *E. coli*	New England Biolabs	Cat#C2530H
Chemicals, Peptides, and Recombinant Proteins		
Isopropyl-beta-D-thiogalactoside (IPTG)	Gold Biotechnology	Cat# I2481C
Nalidixic Acid	Sigma Aldrich	Cat# N8878
Tetracycline	Fisher Scientific	Cat# BP912100
SARS-CoV-2 NSP8	This paper	N/A
SARS-CoV-2 NSP12/7/8 complex	This paper	N/A
Oligonucleotides		
RNA Primer: [6-FAM]-[6-FAM]-GUCAUUCUCCUAAGAAGCUA	Integrated DNA Technologies	N/A
RNA Template: CUAUCCCCAUGUGAUUUUAAUAGCUUCUUAGGAGAAUGAC	Integrated DNA Technologies	N/A
Cy3 Internal Standard DNA Oligo: [Cy3]-CCGTGAGTTGGTTGGACGGCTGCGAGGC	Integrated DNA Technologies	N/A
Recombinant DNA		
$pcl^{ts,ind+}$-(NSP8)	This paper	Deposited to Addgene --160656
pQE-(NSP12)-$pcl^{ts,ind+}$-(NSP7-NSP8)	This paper	Deposited to Addgene -- 160540
Software and Algorithms		
KinTek Explorer v10	KinTek Corp	https://kintekexplorer.com
Chemdraw Professional v16	Perkin Elmer	https://www.perkinelmer.com/product/chemdraw-professional-chemdrawpro
Inkscape v0.92	Inkscape	https://inkscape.org/
Gene Mapper v5	ThermoFisher Scientific	https://www.thermofisher.com/order/catalog/product/4370784#/4370784
SnapGene Viewer v5.1.5	GSL Biotech	https://www.snapgene.com/