The nucleotide addition cycle of the SARS-CoV-2 polymerase

Bera, Subhas Chandra; Seifert, Mona; Kirchdoerfer, Robert N.; van Nies, Pauline; Wubulikasimu, Yibulayin; Quack, Salina; Papini, Flávia S.; Depken, Martin; Dulin, David; More Authors

DOI
10.1016/j.celrep.2021.109650

Publication date
2021

Document Version
Final published version

Published in
Cell Reports

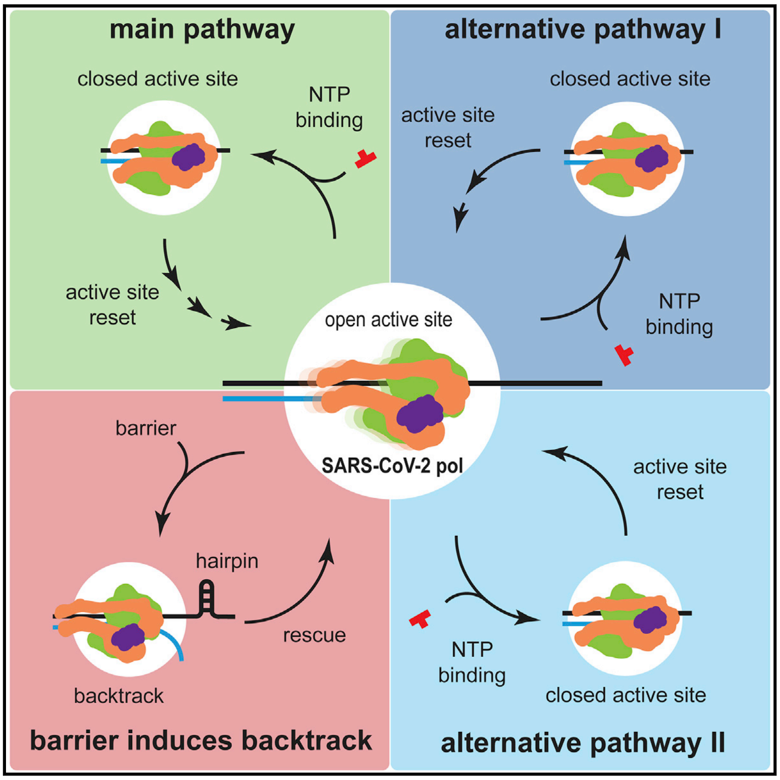
Citation (APA)
Bera, S. C., Seifert, M., Kirchdoerfer, R. N., van Nies, P., Wubulikasimu, Y., Quack, S., Papini, F. S., Depken, M., Dulin, D., & More Authors (2021). The nucleotide addition cycle of the SARS-CoV-2 polymerase. *Cell Reports*, *36*(9), Article 109650. https://doi.org/10.1016/j.celrep.2021.109650

Cell Reports

Article

The nucleotide addition cycle of the SARS-CoV-2 polymerase

Graphical abstract



Highlights

- The SARS-CoV-2 polymerase is a stable and processive RNA polymerase
- The SARS-CoV-2 polymerase uses one fast and two slow nucleotide addition pathways
- The nucleotide addition cycle starts with a thermally activated translocation
- RNA secondary structures induce deep SARS-CoV-2 polymerase backtrack

Authors

Subhas Chandra Bera, Mona Seifert,
Robert N. Kirchdoerfer, ...,
Craig E. Cameron, Martin Depken,
David Dulin

Correspondence

s.m.depken@tudelft.nl (M.D.),
d.dulin@vu.nl (D.D.)

In brief

Bera et al. reveal the complete nucleotide addition cycle of the SARS-CoV-2 polymerase and show that the SARS-CoV-2 polymerase is a processive RNA polymerase that backtracks when elongating through RNA template containing secondary structures.

Bera et al., 2021, Cell Reports *36*, 109650
August 31, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.celrep.2021.109650

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

The nucleotide addition cycle of the SARS-CoV-2 polymerase

Subhas Chandra Bera,[1,7] Mona Seifert,[1,7] Robert N. Kirchdoerfer,[2] Pauline van Nies,[1] Yibulayin Wubulikasimu,[1] Salina Quack,[1] Flávia S. Papini,[1] Jamie J. Arnold,[4] Bruno Canard,[3] Craig E. Cameron,[4] Martin Depken,[5,*] and David Dulin[1,6,8,*]

[1]Junior Research Group 2, Interdisciplinary Center for Clinical Research, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Cauerstr. 3, 91058 Erlangen, Germany

[2]Department of Biochemistry and Institute for Molecular Virology, University of Wisconsin-Madison, Madison, WI 53706, USA

[3]Architecture et Fonction des Macromolécules Biologiques, CNRS and Aix-Marseille Université, UMR 7257, Polytech Case 925, 13009 Marseille, France

[4]Department of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA

[5]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, the Netherlands

[6]Department of Physics and Astronomy, and LaserLaB Amsterdam, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV, Amsterdam, the Netherlands

[7]These authors contributed equally

[8]Lead contact

*Correspondence: s.m.depken@tudelft.nl (M.D.), d.dulin@vu.nl (D.D.)

https://doi.org/10.1016/j.celrep.2021.109650

SUMMARY

Coronaviruses have evolved elaborate multisubunit machines to replicate and transcribe their genomes. Central to these machines are the RNA-dependent RNA polymerase subunit (nsp12) and its intimately associated cofactors (nsp7 and nsp8). We use a high-throughput magnetic-tweezers approach to develop a mechanochemical description of this core polymerase. The core polymerase exists in at least three catalytically distinct conformations, one being kinetically consistent with incorporation of incorrect nucleotides. We provide evidence that the RNA-dependent RNA polymerase (RdRp) uses a thermal ratchet instead of a power stroke to transition from the pre- to post-translocated state. Ultra-stable magnetic tweezers enable the direct observation of coronavirus polymerase deep and long-lived backtracking that is strongly stimulated by secondary structures in the template. The framework we present here elucidates one of the most important structure-dynamics-function relationships in human health today and will form the grounds for understanding the regulation of this complex.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the third zoonotic coronavirus outbreak in less than 20 years, after SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV). To date, SARS-CoV-2 has infected worldwide near 200 million people, which has led to more than four million deaths, with numbers still on the rise. Although several vaccines are now available, we still lack easily administered antiviral drugs to protect non-vaccinated populations against the current or future coronavirus outbreaks (Pan et al., 2021). The ∼30 -kb-long positive single-stranded RNA ((+)ssRNA) genome of coronaviruses encodes many structural and nonstructural proteins (nsp). Among the latter are the viral proteins that constitute the multi-subunits RNA-dependent RNA polymerase (RdRp) responsible for replication and transcription of the viral genome (Snijder et al., 2016). This replication/transcription machinery may differ in composition of accessory factors, but likely has the same core (referred to as polymerase from here on): the RdRp subunit (nsp12) and accessory factors (nsp7 and nsp8) (Kirchdoerfer and Ward, 2019; Hillen et al., 2020). Because of its central role in the virus life cycle, the coronavirus polymerase represents a major drug target (Robson et al., 2020). Nucleotide analogs, such as remdesivir, are the only therapeutic option to treat coronavirus infection, and a precise understanding of the nucleotide addition cycle would tremendously help the development of antiviral drugs.

The nsp12 structure includes the typical features from (+) ssRNA virus RdRps (Gong and Peersen, 2010), with a cupped right hand shape that include palm, finger, and thumb subdomains (Hillen et al., 2020; Kirchdoerfer and Ward, 2019; Wang et al., 2020b; Gao et al., 2020). The coronavirus polymerase complex replicates and transcribes the viral genome at a high pace, having the highest nucleotide addition rate measured for any RNA polymerase to date (i.e., ∼170 nt/s at 37°C) (Seifert et al., 2020a; Shannon et al., 2020; Dangerfield et al., 2020). The fidelity of RdRps is on par with DNA polymerases lacking exonuclease activity (Arnold and Cameron, 2004). Although

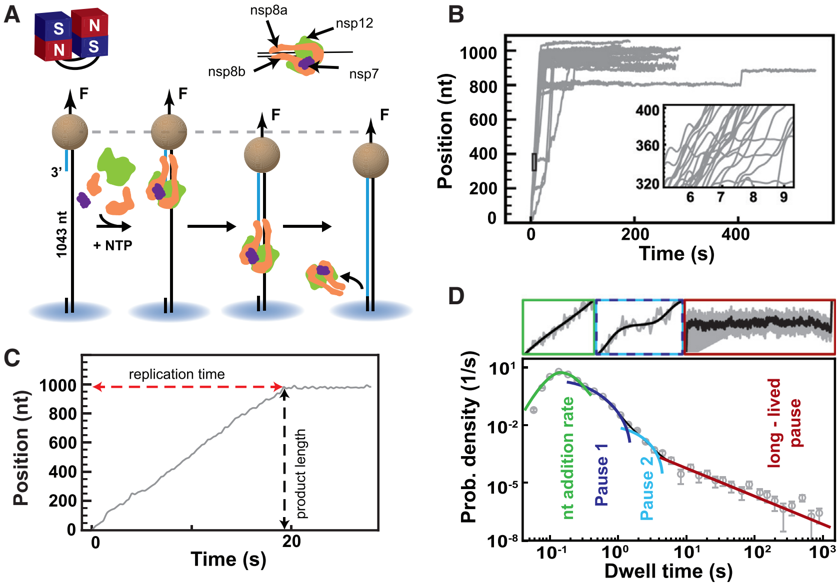


Figure 1. A magnetic tweezers assay to characterize SARS-CoV-2 polymerase elongation kinetics

(A) Schematic of the magnetic tweezers assay to monitor RNA synthesis by the SARS-CoV-2 polymerase. A magnetic bead is attached to a glass coverslip surface by an RNA construct (1,043-nt-long ssRNA template), which experiences a constant force (*F*). nsp7, nsp8, and nsp12 assemble at the 3′ end of the primer ssRNA strand to form an RNA-synthesis competent polymerase. The elongating polymerase converts the ssRNA template into dsRNA, reducing the end-to-end extension of the tether.

(B) Traces of SARS-CoV-2 RNA synthesis activity acquired at 25 pN, 25°C, and with 500 μM NTP (inset) showing bursts of nucleotide addition interrupted by pauses of various durations.

(C) The product length (black) and the total replication time (red) are extracted from the SARS-CoV-2 RNA synthesis activity traces.

(D) A log-binned histogram of the probability density distribution of the dwell times extracted from the SARS-CoV-2 polymerase activity traces. The above insets are examples of the kinetic events that dominate a given dwell time. The distribution has been fitted using a maximum likelihood estimation (MLE) approach using the stochastic-pausing model (STAR Methods, solid lines). The model includes four different probability distribution functions: uninterrupted ten nucleotide additions (green), exponentially distributed pause 1 and pause 2 (blue and cyan, respectively), and the power law distributed long-lived pause (red). The error bars denote one standard deviation of 1,000 bootstrap procedures.

RdRp infidelity enhances viral evolution (Lauring et al., 2013), the stability of the large coronavirus genome is not as tolerant to mutation as viruses with smaller genomes. Therefore, to further reduce mutational burden, the coronavirus genome also encodes a 3′-to-5′ exonuclease, nsp14, that proofreads the RNA product (Ferron et al., 2018; Smith et al., 2013; Eckerle et al., 2007, 2010). Importantly, nsp14 also confers coronaviruses with protection against many antiviral nucleotide analogs (Robson et al., 2020).

Despite the coronavirus polymerase being a validated therapeutic target, we have little understanding of either its structure-dynamics-function relationships or the mechanochemistry governing the thousands of reiterative cycles of nucleotide addition. Furthermore, we have no knowledge on the mechanisms that underlies the surveillance of the viral genome replication transcription by other viral co-factors. Processes such as nucleotide mismatch incorporation in physiological concentration of all nucleoside triphosphates (NTPs) are asynchronous and transient, making their observation difficult for classic ensemble approaches. Single-molecule assays enable the direct observation of single enzymatic complex and sheds light on rare—but essential—biochemical events, such as nucleotide mismatch and analog incorporation (Dulin et al., 2017), which are hidden to even the best-resolved bulk assays. Single molecule magnetic and optical tweezers have been successfully applied to investigate polymerases mechanochemistry, with near base-pair resolution on thousands of cycles and under any solution condition (Ostrofet et al., 2019). These techniques have shown a great success in providing the most complete characterization of the mechanochemical cycles in elongation of many model DNA polymerases (Manosas et al., 2012a, 2012b; Morin et al., 2012; Ibarra et al., 2009; Hoekstra et al., 2017; Wuite et al., 2000), cellular RNA polymerases (Abbondanzieri et al., 2005; Gabizon et al., 2018; Dangkulwanich et al., 2013), and viral RdRps (Seifert et al., 2020b; Dulin et al., 2015c, 2015d, 2017).

Using our recently developed single-molecule high-throughput magnetic tweezers assay (Seifert et al., 2020a), we reveal here the mechanochemical cycle of nucleotide addition by the SARS-CoV-2 core polymerase complex, which represents the first step toward building higher order assemblies and defining the function of each factor. Moreover, our discovery of multiple discrete catalytically competent forms of the polymerase complex, with only one potentially being surveilled by the proofreading exonuclease, suggests strategies to greatly enhance the potency of "next-generation" anti-coronavirus nucleotide analogs.

RESULTS

A high-throughput magnetic tweezers assay to investigate SARS-CoV-2 polymerase RNA synthesis kinetics

We investigated the mechanochemistry of the SARS-CoV-2 polymerase during RNA synthesis using a high-throughput magnetic tweezers assay (Seifert et al., 2020a). We designed an RNA hairpin construct (Figure S1A; STAR Methods), which provides a 1,043-nt ssRNA template when stretched with a force above ∼22 pN (Figure S1B). The hairpin construct is flanked at one end by a biotin-labeled handle to attach the magnetic bead and at the other end by a digoxigenin-labeled handle to anchor the RNA construct to the coverslip glass surface of a flow chamber (Figure 1A). The applied force is controlled by the distance between the magnetic beads and a pair of permanent magnets (Ostrofet et al., 2018). The SARS-CoV-2 polymerase, formed by nsp12, nsp8, and nsp7, assembles at the 3′ end of an ~800-nt-long primer and converts the 1,043-nt-long ssRNA into double-stranded RNA (dsRNA) while elongating, which proportionally decreases the construct end-to-end extension (Figure 1A). At each nucleotide addition cycle, the polymerase must translocate along the template, but also "reset" its active site to perform the next cycle. The mechanical tension applied to the primer-template probes the free energy, equilibrium, and rate of translocation and any other conformational changes occurring along the mechanical reaction coordinate (i.e., the direction of the applied force) (Bustamante et al., 2004). The observed nanometer-scale change in extension is subsequently converted in nucleotides (STAR Methods) (Dulin et al., 2015c). High-throughput magnetic tweezers enable the simultaneous observation of hundreds of RNA hairpin tethered magnetic beads (Figure S1C), providing the acquisition of dozens of SARS-CoV-2 polymerase elongation traces (Figure 1B), from which we characterized both the final product length and the total replication time (Figure 1C).

As observed with another RNA virus, RdRps (Seifert et al., 2020b; Dulin et al., 2015c, 2015d, 2017), the SARS-CoV-2 polymerase activity traces show bursts of nucleotide additions interrupted by pauses of durations ranging from subseconds to hundreds of seconds (Figure 1B). We scanned the activity traces with successive non-overlapping windows of 10 nt and measured the duration of ten successive nucleotide addition cycles, coined "dwell times," to extract the detailed kinetic information from SARS-CoV-2 polymerase RNA synthesis (Figure 1D).

We previously introduced a stochastic-pausing model to describe RdRp's dwell time distributions for Φ6, poliovirus, and human rhinovirus C RdRps (Seifert et al., 2020b; Dulin et al., 2015c, 2015d, 2017), and this model was applied to the SARS-CoV-2 dwell time distributions (STAR Methods). This model includes four probability distribution functions describing the events that kinetically dominate the dwell time: uninterrupted 10-nt additions (gamma distribution), short-lived pause 1 and pause 2 (exponential distributions), and the long-lived pauses (power law distribution) (Figure 1D). The long-lived pause durations are consistent with a $t^{-3/2}$ power-law tail, similar to what is expected from polymerase backtracking (Depken et al., 2009; Dulin et al., 2015c) (STAR Methods).

Unlike other methods used to analyze polymerase elongation kinetics, the pauses and the pause-free nucleotide addition bursts are characterized without any time binning, using a maximum likelihood estimation (MLE) algorithm applied directly to the dwell times (Dulin et al., 2015c) (STAR Methods). Our approach reduces both biases introduced by using arbitrary thresholds to discriminate between pauses and nucleotide addition bursts or by fitting binned data. The model contains seven free parameters (i.e., the nt addition rate, pause 1 and pause 2 exit rates), their respective probabilities, and the backtracking probability (all probabilities sum up to 1) to describe SARS-CoV-2 RdRp elongation kinetics over four orders of magnitude in time. This is on par with other models describing the elongation kinetics of the bacterial RNA polymerase (Abbondanzieri et al., 2005; Shaevitz et al., 2003; Gabizon et al., 2018) and replicative DNA polymerases (Ibarra et al., 2009; Hoekstra et al., 2017) at the single-molecule level.

SARS-CoV-2 polymerase is a processive RNA polymerase

We first investigated whether the observed pauses were related to the exchange of the polymerase itself or its factors. If one of the pauses originated from the exchange of polymerase factor(s), we expect the kinetics of this specific pause to be affected by varying the concentration of proteins in the reaction buffer. To test this, we performed two types of experiments. In the first set of experiments, we flushed the flow chamber with reaction buffer containing NTP and the polymerase factors at different concentrations of nsp12 (i.e., 0.1, 0.2, and 0.4 μM) while maintaining the nsp12:nsp8:nsp7 stoichiometry at 1:9:9 (i.e., with an excess of nsp7 and nsp8). In this case, elongation starts as soon as the polymerase has assembled on the primer-template, while the polymerase factors are present in the solution to enable protein exchange. In the second set of experiments, we pre-assembled (PA) the polymerase by incubating the polymerase factors in the flow chamber without NTP, subsequently rinsed the flow chamber to remove any free proteins, and started the polymerase RNA synthesis activity by injecting a reaction buffer solution containing 500 μM NTP (Figure 2A; STAR Methods).

In the presence of the polymerase factors in solution, the product length and replication time were not significantly changed for the different polymerase factors concentrations we evaluated, with mean values of (899 ± 28) nt and (30 ± 3) s, respectively (Figures 2B and 2C). A dwell time analysis of these traces similarly reported that the nucleotide addition, pause 1, pause 2, and the long-lived pause distributions were also unresponsive to changes in concentration of the polymerase factors in the flow cell (Figures 2D, 2E, and S2A). We then evaluated the elongation kinetics of the pre-assembled complex. Because the traces may have started at an undefined time during the injection of the reaction buffer containing NTPs, we did not evaluate the replication time for the pre-assembled polymerase activity traces. We performed a dwell time analysis of these traces (Figure S2B), which showed again that none of the parameters of the stochastic-pausing model were affected by the absence of viral proteins in the solution. Furthermore, pre-assembling the complex and performing the primer elongation reaction in either 150 mM potassium glutamate or 150 mM sodium chloride did not change the polymerase mean product length (Figure S2C). We conclude that neither the product length, the replication time, nor the dynamics observed in the traces resulted from viral protein disassembly or exchange during RNA synthesis. The coronavirus polymerase is a processive RNA polymerase.

The short pauses are the signature of slow nucleotide addition pathways distinct from the nucleotide addition burst pathway

The polymerase complex has potentially four NTP binding sites: the RdRp and the NiRAN domains of nsp12, and the two nsp8s (Wang et al., 2021). Concerning the latter, nsp8 has been shown to have 3′-terminal adenylyltransferase activity stimulated by short 5′-poly-U template (Tvarogová et al., 2019). We do not have such a template (STAR Methods), and we therefore do not expect any impact on the kinetics we are monitoring. The NiRAN nucleotide binding activity has been shown to allosterically assist replicase assembly but not the elongation kinetics

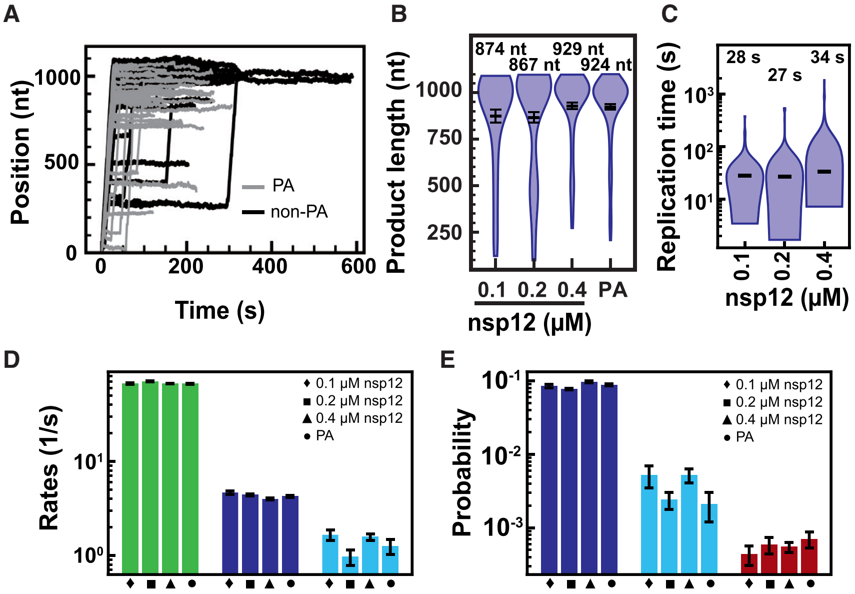


Figure 2. SARS-CoV-2 elongation dynamics does not result from polymerase co-factors disassembly or exchange

All experiments were conducted at 35 pN and 25°C with 500 µM NTP.

(A) Traces of SARS-CoV-2 RNA synthesis activity of pre-assembled (PA, gray) and non-pre-assembled (non-PA, black) polymerase. PA polymerase may start synthesizing RNA while flushing reaction buffer with NTP and therefore only the part of trace collected at the end of the infusion is represented.

(B and C) Product length (B) and replication time (C) of SARS-CoV-2 polymerase as a function of nsp12, nsp7, and nsp8 concentration at a constant stoichiometry of 1:9:9. The median replication times and the mean product lengths are indicated above the violin plots and represented as horizontal thick black lines, flanked by 1 SD error bar extracted from 1,000 bootstraps.

(D) Nucleotide addition rates (green bars), pause 1 (dark blue bars), and pause 2 (cyan bars) exit rates for PA polymerase and concentration of nsp12 as described in panels and nsp7 and nsp8 concentrations as described in (B) and (C).

(E) Probabilities to enter pause 1 (dark blue bars), pause 2 (cyan bars), and long-lived pauses (red bars) for the conditions described in (D). Error bars are 1 SD extracted from 100 bootstraps. All statistics, kinetics parameters, and error estimations are provided in Table S1.

(Wang et al., 2021). We therefore considered only the NTP binding sites of nsp12 RdRp to be affected by the nucleotide concentration. We subsequently investigated how nucleotide concentration affects the RNA synthesis kinetics to determine which state is catalytically competent, the saturating NTP concentration of such states, and their nucleotide addition kinetics parameters, such as the Michaelis-Menten (MM) constant K_M and the maximum nucleotide-addition rate k_{max}. We varied the NTP concentration between 10 µM and 1 mM at a constant force of 25 pN (Figure 3A). We noted that decreasing the NTP concentration significantly increases the average total replication time from (21 ± 1) s to (283 ± 9) s (Figure 3B), without affecting the product length (Figure S3A). The shape of the dwell time distribution is also affected by decreasing the concentration of NTP: pause 1 and pause 2 shoulders inflate dramatically, whereas the nucleotide addition peak fades away behind the pause 1 distribution (Figures 3C and S3B). Fitting the distributions using the stochastic-pausing model showed that SARS-CoV-2 polymerase nucleotide addition rate is constant at ~76 nt × s^{-1} over the whole range of NTP concentration (Figure 3D). This is a rather surprising result because one would expect polymerase nucleotide addition rate to increase with NTP concentration, up to some maximal rate k_{max} when NTP binding is saturated. The observed NTP concentration independence suggests that the chemistry of nucleotide addition is not rate-limiting and must be followed by a second (nearly) irreversible step that dominates the overall reaction timescale within the NTP concentration explored. Another surprising result is that pause 1 and pause 2 exit rates decreased by one order of magnitude when decreasing in NTP concentration (Figure 3D; Table S1). This indicates that pause 1 and pause 2 are the kinetic signatures of slow and very slow nucleotide additions (SNA and VSNA) pathways, in addition to the nucleotide addition bursts (NAB) pathway. Pause 1 and pause 2 exit rates as a function of NTP concentrations are well described by MM kinetics (Figure 3D), with the exit rate written as

$$k^{a}([\mathrm{NTP}]) = \frac{k^{a}_{\max}[\mathrm{NTP}]}{K^{a}_{M}+[\mathrm{NTP}]},\quad a=1,2\,, \qquad \text{(Equation 1)}$$

where K^{a}_{M} is a force- and translocation-dependent analog of the MM constant, and k^{a}_{max} is the maximal exit rate from pause a. We extracted $K^{1}_{M} = (86 \pm 34)\ \mu\mathrm{M}$ and $K^{2}_{M} = (250 \pm 112)\ \mu\mathrm{M}$and $k^{1}_{max} = (5.4 \pm 0.6)\ \mathrm{nt}{\times}\mathrm{s}^{-1}$and $k^{2}_{max} = (2.1 \pm 0.4)\ \mathrm{nt}{\times}\mathrm{s}^{-1}$ at a force of 25 pN (Figure 3D). Pause 1 probability was constant down to 500 µM NTP, and subsequently increased by almost 6-fold when further reducing NTP concentration down to 20 µM (Figure 3E; Table S1). We note that pause 1 probability did not increase further below 20 µM NTP, which is likely due to our analysis being unable to fit the gamma distributions at large pause 1 probability (above ~0.3) (Figure S3B). Pause 2 probability increased by ~2.3-fold (Figures 3E and S3C) (i.e., not proportionally to pause 1). The change in probability with NTP concentration indicates that SNA and VSNA pathways are distinct from the NAB pathway. To verify whether the kinetics response of the polymerase to the NTP titration was not just specific to the 25 pN force, we performed the same set of experiments at 35 pN (Figure S3D). The trend observed at 25 pN was conserved at 35 pN. The median replication time decreased when increasing NTP concentration (Figure S3E), whereas the average product length remained constant (Figure S3F). The nucleotide addition rate was constant for NTP concentrations down to 50 µM NTP and then decreased, which likely resulted from the poor MLE fit of the gamma-distribution due to the large pause 1 probability (Figures S3G–S3I). Pause 1 and pause 2 exit rates remained well described by the effective MM equation (Equation 1), i.e., $K^{1}_{M} = (109 \pm 16)\ \mu\mathrm{M}$ and $K^{2}_{M} = (156 \pm 60)\ \mu\mathrm{M}$ and $k^{1}_{max} = (5.6 \pm 0.2)\ \mathrm{nt}{\times}\mathrm{s}^{-1}$ and

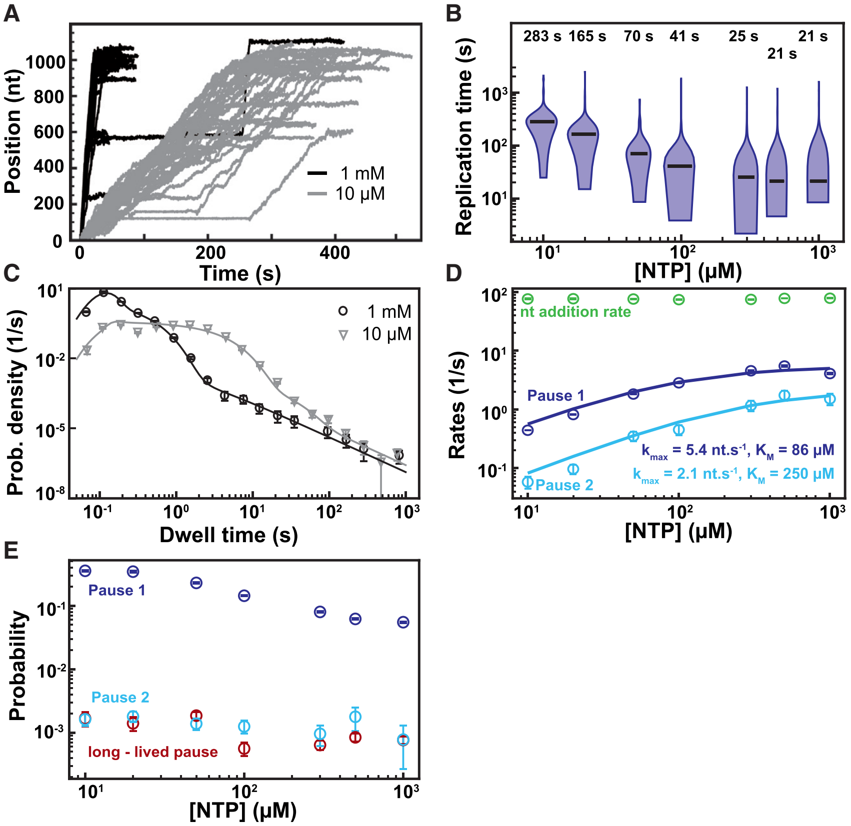


Figure 3. Pause 1 and pause 2 are the kinetic signatures of low-efficiency catalytic states

All experiments were conducted at 25 pN, 25°C, 0.6 µM nsp12, 1.8 µM nsp7, and 1.8 µM nsp8.

(A) SARS-CoV-2 polymerase activity traces at 1 mM NTP (black) and 10 µM NTP (gray).

(B) The replication time as a function of NTP concentration. The median replication times are indicated above the violin plots and represented as horizontal thick black lines, flanked by 1 SD error bars extracted from 1,000 bootstraps.

(C) The dwell time distributions and their corresponding MLE fits (solid lines) at 1 mM (black circles) and 10 µM NTP (gray triangles).

(D) The nucleotide addition rates (green circle), pause 1 (dark blue circle), and pause 2 (cyan circle) exit rates as a function of NTP concentration. The solid lines are Michaelis-Menten fit to pause 1 and pause 2 exit rates NTP concentration dependence. The extracted Michaelis-Menten parameters are indicated next to the fits using the color code described above.

(E) The probabilities for pause 1 (dark blue circle), pause 2 (cyan circle), and long-lived pauses (red circle). The error bars in (C) denote 1 SD extracted from 1,000 bootstraps. The error bars in (D) and (E) are 1 SD extracted from 100 bootstraps. All statistics, kinetics parameters and error estimations are provided in Table S1.

$k^2_{max} = (1.9 \pm 0.2)$ nt$\times$s^{-1} (Figure S3H). Pause 1 and pause 2 probabilities increased when decreasing NTP concentration, but faster than at 25 pN, suggesting a force dependence for these pauses (Figure S3I; Table S1). In conclusion, the coronavirus polymerase presents three catalytic pathways, where pause 1 and pause 2 are, respectively, the kinetic signatures of a slow and a very-slow-nucleotide addition pathways branching off from the nucleotide addition burst pathway (described by the nucleotide addition rate) and the long-lived pauses are catalytically incompetent.

A large conformational change dominates the coronavirus polymerase nucleotide addition cycle kinetics

By varying the tension on the template, we probed whether a rate-limiting conformational change occurs during the nucleotide addition cycle (Bustamante et al., 2004). Modeling the experimental data enables us to extract the magnitude of the conformational change along the RNA, as well as the conformer-conversion rate at zero force. Increasing the force from 20 to 60 pN increased the total duration of the activity traces and the number of pauses in these traces (Figure 4A). The median replication time increased by 4-fold, whereas the average product length only mildly decreased (Figures 4B and 4C). Surprisingly, even at forces as high as 60 pN, the SARS-CoV-2 polymerase demonstrated a strong RNA synthesis activity (Figure 4A).

We used the stochastic-pausing model (Figure 1D; STAR Methods) to fit the dwell-time distributions obtained from the activity traces collected at forces varying from 20 pN to 60 pN (Figures S4A–S4J). The nucleotide addition rate decreased exponentially between 20 pN and 30 pN and remained mostly constant for larger forces (Figure 4D; Table S1). A similar plateau at high force has been reported for T7 DNA polymerase (Hoekstra et al., 2017). To describe the force dependence of the nucleotide addition rate when using an ssRNA template, we used an Arrhenius equation with an offset rate A

$$k_{ss}(F) = k^0_{ss} \times \exp(-F\delta x / k_B T) + A, \qquad \text{(Equation 2)}$$

where $k^0_{ss} + A$ is the nucleotide addition rate at zero force, F is the applied force, δx is the distance to the transition state of the reaction, and $k_B T$ is the thermal energy. We extracted $A = (68 \pm 1)$ nt × s^{-1} and $k^0_{ss} = (697 \pm 634)$ nt$\times$s^{-1}, and consequently also a very high nucleotide addition rate on par with recent estimations (Shannon et al., 2020; Dangerfield et al., 2020). The distance to the transition state $\delta x = (0.75 \pm 0.18)$ nm is larger than a single base distance, which may indicate a large polymerase conformational change not due to translocation.

Pause 1 and pause 2 kinetics were also force-dependent. Their respective exit rates were constant at tensions up to ~35 pN and decreased at higher forces (Figure 4E; Table S1). Our data are consistent with pause 1 and pause 2 exiting through at least two irreversible kinetic steps: one NTP addition step that is rate limiting at low forces (up to ~35 pN), with a relatively weak force dependence and another one strongly force-dependent related to a conformational rearrangement that is rate limiting at high forces.

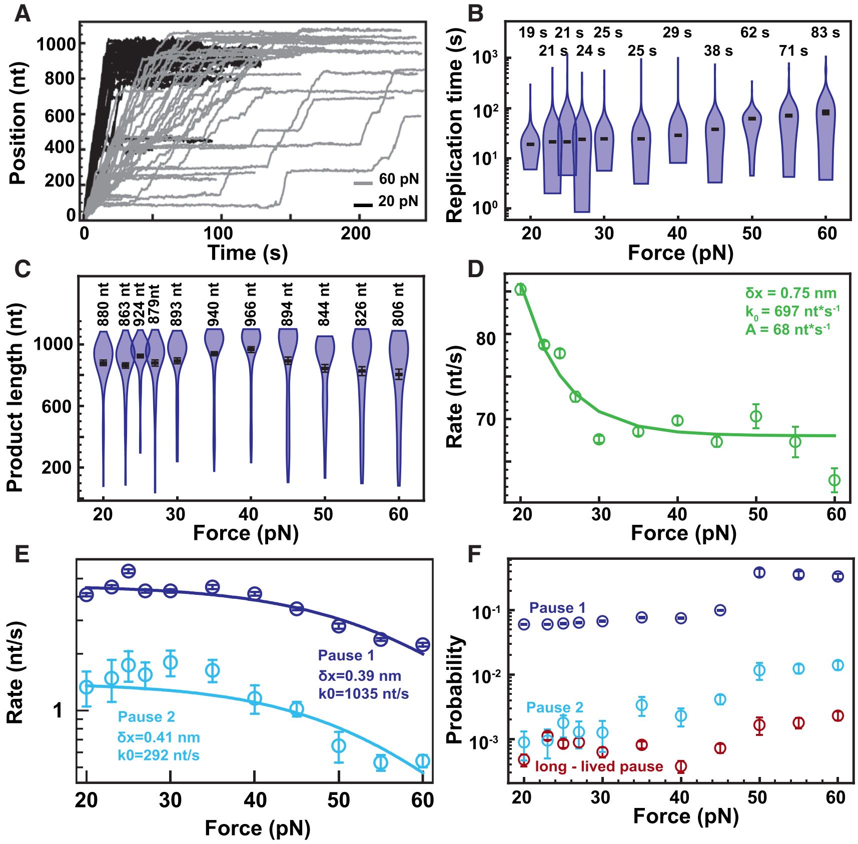


Figure 4. The force dependence of SARS-CoV-2 polymerase kinetics indicates a rate-limiting large conformational change

All experiments were conducted at 500 μM NTP, 25°C, 0.6 μM nsp12, 1.8 μM nsp7, and 1.8 μM nsp8.

(A) SARS-CoV-2 polymerase activity traces at 20 pN (black) and 60 pN (gray).

(B and C) Replication time (B) and product length (C) of SARS-CoV-2 polymerase as a function of force. The median replication times and the mean product lengths are indicated above the violin plots and represented as horizontal thick black lines, flanked by 1 SD error bars extracted from 1,000 bootstraps.

(D) The nucleotide addition rate of elongating SARS-CoV-2 polymerase as a function of force. The data were fitted (solid line) with the Arrhenius equation (Equation 2).

(E) Pause 1 (dark blue circle) and pause 2 (cyan circle) exit rates as a function of force. The dashed lines are the fits of the corresponding rates using Equation 3.

(F) The probabilities to enter pause 1 (dark blue circle), pause 2 (cyan circle), and long-lived pause (red circle) as a function of force. The error bars in (D)–(F) are 1 SD extracted from 100 bootstraps. All statistics, kinetics parameters, and error estimations are provided in Table S1.

We see no appreciable force dependence in the low force regime when the polymerase elongates on an ssRNA template and we therefore described the observed average pause-exit rate as

$$k_a(F)=\frac{1}{\dfrac{1}{k^a([\mathrm{NTP}]^*)}+\dfrac{1}{k_{\mathrm{ss},a}(F)}},\ k_{\mathrm{ss},a}(F)=k^0_{\mathrm{ss},a}\exp\left(-\frac{\delta x_a F}{k_\mathrm{B}T}\right),\ a=1,2, \qquad \text{(Equation 3)}$$

where $k^a([\mathrm{NTP}]^*)$ parameters are extracted from the fit of Equation 1 to pause 1 and pause 2 exit rates as a function of NTP concentration at 25 pN (Figure 3D), i.e., in the weakly force-dependent regime (Figure 4E), using $[\mathrm{NTP}]^* = 500\ \mu\mathrm{M}$. Equation 3 describes the observed pause 1 and pause 2 exit rates, and we extracted $k^0_{\mathrm{ss},1} = (1,035 \pm 209)\ \mathrm{s}^{-1}$ and $\delta x_1 = (0.39 \pm 0.02)$ nm for pause 1 and $k^0_{\mathrm{ss},2} = (292 \pm 275)\ \mathrm{s}^{-1}$ and $\delta x_2 = (0.41 \pm 0.07)$ nm for pause 2 (Figure 4E). The probability of pause 1 and pause 2 also increased when raising the force from 20 to 45 pN by almost 2- and 5-fold, respectively (Figures 4F and S4K; Table S1), which indicates that the polymerase branches off from the NAB to the SNA and the VSNA pathways from a pre-translocated state that is further populated with increasing tension. The long-lived pause probability showed no force dependence up to 45 pN (Figure 4F). At 50 pN, all pauses showed an ~3-fold jump in probability, which may originate from a change in the dsRNA helix conformation from A- to S-form when being overstretched (Lipfert et al., 2014; Melkonyan et al., 2019).

In conclusion, the second irreversible and rate-limiting step of the nucleotide addition cycle is force-dependent and related to a conformational change more than 3 times larger than a single base translocation. The two low-efficiency catalytic pathways (pause 1 and pause 2) also support a two successive irreversible steps model, where the first one is NTP concentration-dependent and the second one is force-dependent.

Duplex RNA barrier to the polymerase translocation increases pause 1 probability and induces polymerase backtracking

The coronavirus genome is heavily structured (Lan et al., 2020), and it is therefore particularly relevant to interrogate the ability of SARS-CoV-2 polymerase to elongate through a closed RNA hairpin (Figure 5A). The conversion of the hairpin into a dsRNA tether happens in two phases, where the polymerase first progresses through the hairpin stem, followed by an ssRNA to a dsRNA conversion. The second part of the trace leads to either a decrease or an increase in extension for forces above and below ~10 pN, respectively (traces at 20 pN and 9 pN in Figure 5B) (Dulin et al., 2015c). Here, we focused our investigation only on the initial ~0.5-kbp (kilobase pair) hairpin unwinding activity of the polymerase. At the end of each trace, we verified whether the polymerase fully converted the template into a linear dsRNA by increasing the force to 45 pN, as dsRNA extension is rather constant in this force range, unlike ssRNA (Figures S5A and S5B) (Dulin et al., 2015c). We discarded the traces that showed a large difference in extension (Figure S5B). A direct comparison of the polymerase activity traces at either 9 pN or

20 pN show that unwinding the hairpin stem at low assisting force induced an increase in the number of pauses and their duration (Figure 5B).

We observed an increase in the magnitude of the magnetic bead position fluctuation during the long-lived pauses at 9 pN, and we investigated these fluctuations. To this end, we improved the microscope resolution by increasing the objective magnification by 2-fold to 100× to reduce the tracking noise (Dulin et al., 2015b). Furthermore, we introduced a custom autofocus, which significantly increased the stability of the measurement (Figures S5D–S5H; STAR Methods) by annihilating the impact of the drift for a duration as long as ∼100 s (i.e., a ∼50-fold improvement compared to previous studies in similar conditions) (Ostrofet et al., 2020; Dulin et al., 2015b; Huhle et al., 2015) (Figures S5G and S5H; STAR Methods). Consequently, we were able to confirm that the fluctuations observed in Figure 5B originated from polymerase backtracks as deep as ∼30 nt upstream of the last incorporated nucleotide (Figures 5C, 5D, and S5I). Our results provide a direct observation of SARS-CoV-2 polymerase backtracks and support a recent structural study on a pre-assembled backtracked coronavirus polymerase (Malone et al., 2021).

Decreasing the force from 20 pN to 9 pN increased the median replication time by almost an order of magnitude (Figure 5E), whereas the mean product length remained largely constant (Figure 5F). Applying our stochastic-pausing model to the dwell time distributions (Figure S6C), we observed that the nucleotide addition rate decreased when RNA synthesis is performed against a dsRNA barrier, from (85 ± 1) nt × s^{-1} at 20 pN and ssRNA template to (52 ± 1) nt × s^{-1} at the same force but using a dsRNA template (Figures 4D and 5G). The nucleotide addition rate further decreased to (45 ± 3) nt × s^{-1} at 18 pN to remain largely constant at lower forces (Figure 5G; Table S1).

Although pause 1 and pause 2 exit rates measured at 20 pN were comparable when using either the closed or the open hairpin, lowering the force to 9 pN drastically decreased pause 1 and pause 2 exit rates by 6- and 7-fold, respectively (Figure 5G; Table S1). The decrease is well described by an Arrhenius equation (Equation 2; $A = 0$), providing $\delta x_{1,\mathrm{ds}} = (0.75 \pm 0.08)$ nm, $\delta x_{2,\mathrm{ds}} = (0.81 \pm 0.08)$ nm, $k^0_{1,\mathrm{ds}} = (0.13 \pm 0.05)\ \mathrm{s}^{-1}$, and $k^0_{2,\mathrm{ds}} = (0.02 \pm 0.01)\ \mathrm{s}^{-1}$. The distance to the transition state for both pauses are similar and consistent with the gain in extension from a single base pair melting (i.e., two single nucleotides) (Vieregg et al., 2007). We therefore suggest that $\delta x_{1/2}$ relates to single base pair polymerase forward translocation, and therefore translocation becomes rate-limiting for pause 1 and pause 2 when the polymerase is elongating through a dsRNA barrier (closed hairpin). The exit rates of the pauses at zero force, $k^0_{1,\mathrm{ds}}$ and $k^0_{2,\mathrm{ds}}$, highlight the difficulty for the polymerase alone to exit pause 1 and pause 2 when elongating through dsRNA. Finally, elongating through a dsRNA barrier significantly increases the probability to pause: pause 1, pause 2, and backtrack pause probabilities increased by 3-, 1.4-, and 8-fold while decreasing the force from 20 to 9 pN when using a dsRNA hairpin template (Figure 5H; Table S1).

In conclusion, the coronavirus polymerase is rather inefficient at elongating through a duplex RNA template, which suggests that the nsp13 helicase or other viral co-factors assist the polymerase during viral genome replication and transcription.

A nucleotide addition cycle model for the coronavirus polymerase

To describe the NTP and force dependence of the coronavirus polymerase on either the ss- or dsRNA template, we introduce a model for the nucleotide addition cycle of the coronavirus polymerase, which may be applicable to other related viral RdRps by extension (Figures 6A and S6). Our model supports a polymerase translocation mechanism at the beginning of the nucleotide addition cycle, where the step forward is thermally activated. This translocation step could be mistaken for a power-stroke because it does not demonstrate a force dependence in rate. The forward translocation is then stabilized by NTP binding, followed by the closure of the active site (Yang et al., 2012). Pause 1 probability increases when raising the applied tension on the ssRNA template (open hairpin) or decreasing the applied tension on the dsRNA template (closed hairpin). We therefore conclude that the SNA pathway (pause 1) branches out from the pre-translocated state of the NAB pathway (Figures 6A and S6). Furthermore, pause 1 probability never becomes null at saturating NTP, but rather plateaus above 500 μM. This indicates that translocation cannot be considered equilibrated with respect to NTP binding and pause 1 entry rate and is even rate-limiting for both pause 1 and pause 2 when using a dsRNA hairpin template. The NTP concentration independence of the SARS-CoV-2 polymerase nucleotide addition rate (NAB pathway) at two different forces (25 and 35 pN) indicates the existence of another irreversible and rate-limiting step following the nucleotide addition chemistry. The force dependence of this state suggests a large conformational change of the polymerase at the end of the nucleotide addition cycle. The SNA and VSNA pathways (pause 1 and pause 2) exit rates are also consistent with two successive irreversible steps, one related to the chemistry of nucleotide addition and the other related to a tension-dependent conformational change. An irreversible conformational change was also observed in a pre-steady state kinetic study of poliovirus RdRp (Arnold and Cameron, 2004) and was inferred to be the translocation of the polymerase. However, our data show the distance to the transition state of the reaction is three times as large as a single base translocation for both the NAB, SNA, and VSNA. We therefore conclude that this conformational change cannot be solely due to translocation and likely represents a large conformational change in the elongating polymerase.

Performing a global fit of the probability to enter pause 1 as a function of NTP concentration (both at 25 and 35 pN), and of the applied tension (at 500 μM NTP) (STAR Methods), we show that our model describes the kinetics of the elongating SARS-CoV-2 polymerase well (Figures 6B–6D; STAR Methods). This model extracts a step size from the pre- to the post-translocated states of 0.23 nm, which is in excellent agreement with a single nucleotide translocation distance (STAR Methods). The model also accounts for the increase in pause 1 probability as a function of the applied tension when using a dsRNA template (Figure 6E; STAR Methods). The long-lived pauses are the kinetic signature of polymerase backtracking, which dramatically increased when the polymerase elongates through a dsRNA hairpin. We conclude that the backtrack branches off from the NAB pathway pre-translocated state (Figures 6A and S5I).

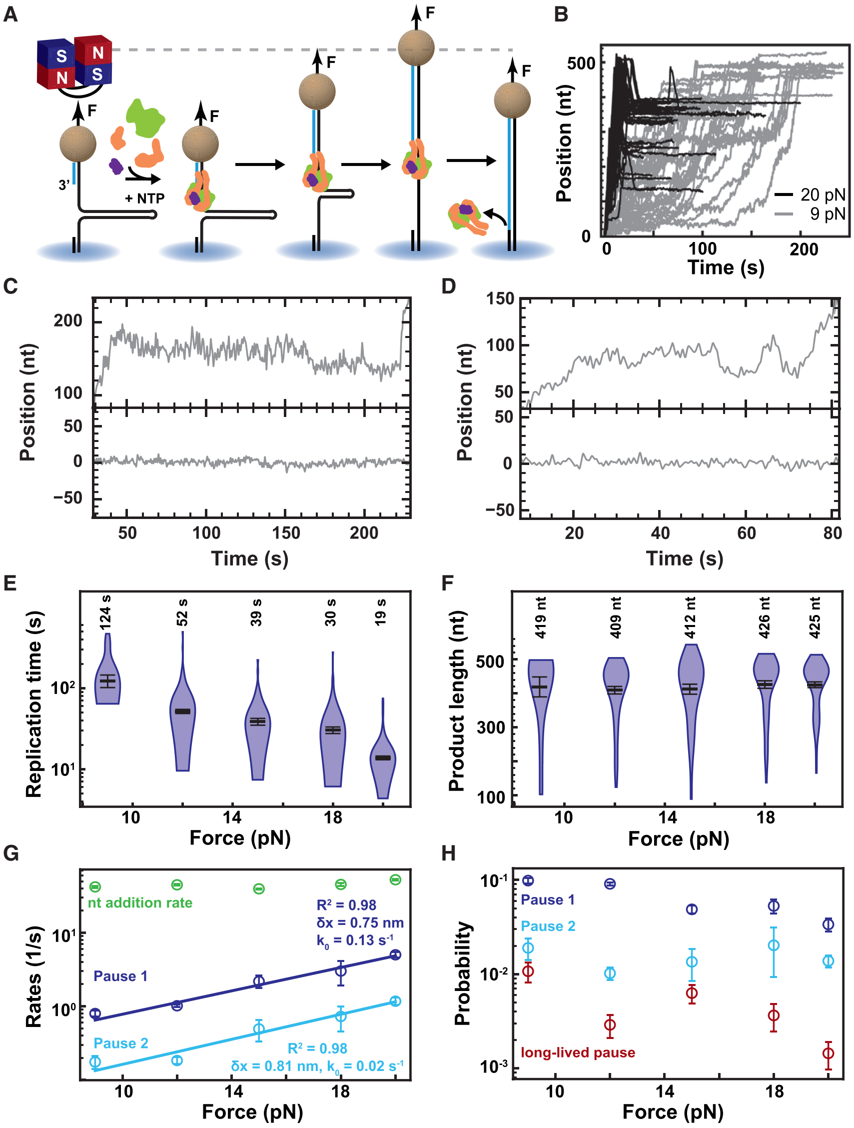


Figure 5. Template secondary structures represent a strong barrier to the elongating SARS-CoV-2 polymerase

All experiments were conducted at 500 μM NTPs, 25°C, 0.2 μM nsp12, 1.8 μM nsp7, and 1.8 μM nsp8.

(A) The applied force is lower than the hairpin opening force and therefore the polymerase elongates through the 499-nt hairpin stem, with a net increase in extension of 1 nt + 1 bp. Subsequently, the change in tether extension is described by the difference in extension of an ssRNA and dsRNA as a function of force (Figure 1A).

(B) SARS-CoV-2 polymerase activity traces on dsRNA at 20 pN (black) and 9 pN (gray). At 20 pN, an increase in extension appears in the first part of the trace when the polymerase unwinds the hairpin stem, and a subsequent decrease ends the activity trace in the second part (converted as the first part of the trace, not considered for analysis).

(C and D) Top: zoom-in of SARS-CoV-2 polymerase backtrack acquired at 9 pN force and 500 μM NTP using the ultra-stable magnetic tweezers at 58 Hz camera frequency acquisition and low-pass filtered at 1 Hz. Bottom: 1 Hz low-pass filtered trace of a tether showing no polymerase activity acquired simultaneously.

(E and F) Replication time (E) and product length (F) of SARS-CoV-2 polymerase on dsRNA as a function of force. The median replication times and the mean product lengths are indicated above the violin plots and represented as horizontal thick black lines, flanked by 1 SD error bars extracted from 1,000 bootstraps.

(G) The nucleotide addition (green circle), pause 1 (dark blue circle), and pause 2 (cyan circle) exit rates as a function of force. The solid lines represent Arrhenius fits to pause 1 and pause 2 exit rates (Equation 1).

(H) Probabilities to enter pause 1 (dark blue circle), pause 2 (cyan circle), and long-lived pauses (red circle) as a function of force. The error bars in (E) and (F) are 1 SD extracted from 100 bootstraps. All statistics, kinetics parameters, and error estimations are provided in Table S1.

DISCUSSION

Understanding the mechanism, regulation, and inhibition of the SARS-CoV-2 replication and transcription complex necessitates a precise and complete description of the nucleotide-addition cycle. Although structural information has been gathered at an incredible pace (Gao et al., 2020; Hillen et al., 2020; Wang et al., 2020b; Chen et al., 2020), a mechanistic framework to interpret the structural information is not available. Here, we have interrogated the nucleotide-addition cycle of the SARS-CoV-2 core polymerase complex on a 1-kb-long template using a high-throughput magnetic tweezers approach (Seifert et al., 2020a). Consistent with previous studies of other RdRps (Seifert et al., 2020b; Dulin et al., 2015c, 2015d, 2017), RNA synthesis by this complex is described best as a series of bursts of nucleotide addition interrupted by pauses of various durations (Figure 1). We propose that these pauses originate from distinct structural/conformational states of the complex and/or transactions performed by the complex, potentially resulting from nucleotide misincorporation. Following polymerase initiation, and over the 1,000 cycles of nucleotide addition monitored, there was no evidence for exchange of the protein components of the complex or dissociation of the complex (Figure 2). These results demonstrate that the SARS-CoV-2 polymerase complex is both stable and processive.

We observed that the nucleotide addition rate of the SARS-CoV-2 polymerase decreased when increasing the tension and eventually plateaued above 30 pN force at a minimum of ∼70 nt/s (Figure 4D). Why does the polymerase nucleotide addition rate not decrease to zero, as observed for ϕ29 DNA polymerase (Ibarra et al., 2009)? We suggest that ϕ29 is extremely sensitive to the remodeling of the primer-template-polymerase interactions on changes in tether tension, whereas SARS-CoV-2 polymerase active site is more resilient to an increased tension. To

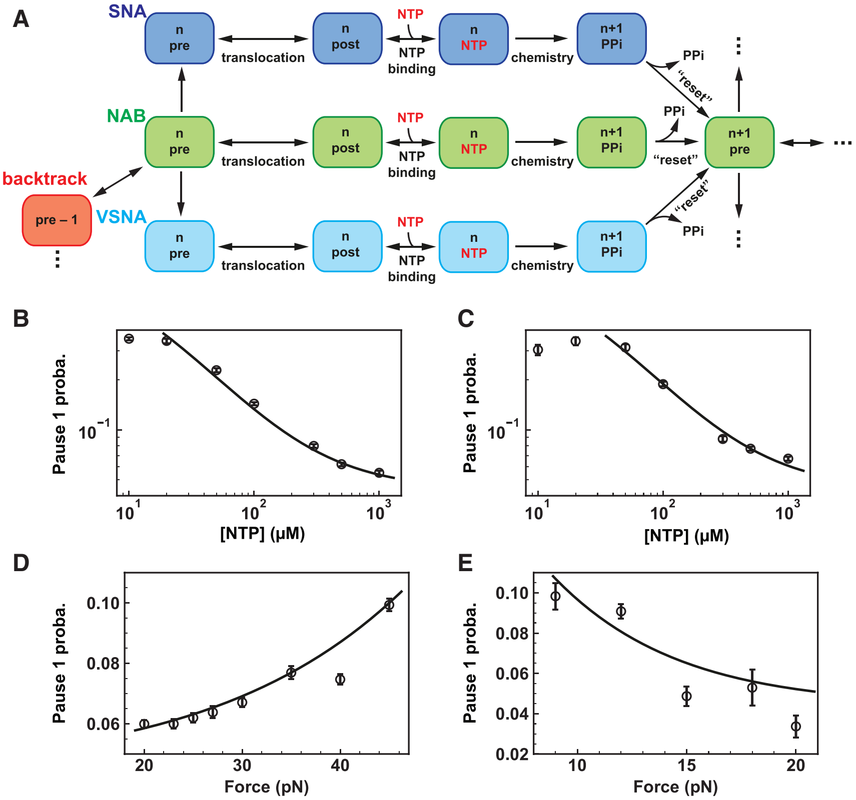


Figure 6. A mechanochemical model of the coronavirus polymerase nucleotide addition cycle

(A) The mechanochemical cycle starts from the pre-translocated state (pre) at position n on the product RNA of the nucleotide addition burst (NAB) pathway. From this position, the polymerase either use thermal energy to translocate forward onto the post-translocated state and carry on with nucleotide binding, chemistry (n + 1 RNA product length) and a large conformational change that "resets" the polymerase toward the next nucleotide addition cycle, or enters either the slow-nucleotide addition (pause 1) or the very-slow-nucleotide addition (pause 2) (SNA and VSNA, respectively) pathway. The polymerase may also enter a backtrack state, which appears (as in the trace) as polymerase position behind the maximum position reached on the template and/or long-lived pauses and is favored by elongating through a closed hairpin.

(B–E) Pause 1 probabilities, as a function of the NTP concentration at either 25 pN or at 35 pN (B and C) and as a function of force for either an ssRNA or dsRNA templates at 500 µM NTP (D and E), were fitted with Equation S4 of the kinetic model describing the entry in pause 1 (SNA) (STAR Methods).

verify this hypothesis, we modeled the downstream position of the template strand by aligning the recent cryoelectron microscopy (cryo-EM) structure of an elongating SARS-CoV-2 polymerase (Hillen et al., 2020) to the crystal structure of an elongating poliovirus RdRp (Gong and Peersen, 2010). The upstream duplex RNA product aligns well, and therefore we could infer the template strand position downstream nsp12. We found the angle between the downstream template strand and the duplex product can increase up to ~33° (i.e., from ~87° to ~121°) before sterically clashing with nsp12 residues (Figures 7A and S7). In other words, the SARS-CoV-2 template strand is loosely constrained downstream of the active site. Similarly, T7 DNA polymerase, which nucleotide addition rate also plateaus at high force to a ~80 nt/s lower bound (Hoekstra et al., 2017), shows a little constrained template strand. Indeed, the template strand angle with the duplex may increase from ~92° to ~149° on applied force, before clashing with polymerase residues (Figure 7B) (Doublié et al., 1998). Performing a similar alignment with ϕ29 DNA polymerase shows the template strand is more constrained downstream of the polymerase active site (i.e., the angle ranges from ~122° to ~136°) and is tightly constrained within the polymerase template channel (Figure 7C) (Berman et al., 2007). We therefore suggest that ϕ29 DNA polymerase sensitivity to force results from the deformation of the active site due to the aforementioned tight interactions.

Here, we show that the polymerase incorporates nucleotides through three distinct catalytic pathways that we named nucleotide-addition burst (NAB), slow-nucleotide-addition (short pauses ≤1 s duration) (SNA), and very-slow-nucleotide-addition (pauses ~1–5 s duration) (VSNA). The nucleotide addition cycle is conserved for all pathways, i.e., all starts with a thermally activated forward translocation, similarly to what has been observed with cellular DNA-dependent RNA polymerases (Gabizon et al., 2018; Dangkulwanich et al., 2013; Abbondanzieri et al., 2005). Following translocation, the nucleotide addition cycles carry on with nucleotide binding and two successive irreversible steps where the first is linked to the chemistry of nucleotide addition and the second to a large conformational change in the elongating polymerase-primer-template complex (Figures 6A and S6). The force dependence of the second irreversible step kinetics indicates a global conformational change that displaces the RNA template between its exit and entry points into the polymerase complex by a distance equivalent to three bases (Figure 4). What could this large conformational change be? Structural work on the related poliovirus and EV71 RdRps suggests that reopening of the active site at the end of the nucleotide addition cycle is concomitant with a large conformational change of the motif B loop and a displacement of the D-motif (Sholders and Peersen, 2014; Yang et al., 2012; Gong and Peersen, 2010), with a rearrangement of the interactions between the template strand and the G-motif (Wang et al., 2020a). Altogether, these rearrangements would release the "grip" of the polymerase on the primer-template to reset the polymerase and enable translocation at the next nucleotide addition cycle.

The structural origin of the SNA and VSNA pathways is difficult to ascertain because these pathways have not been previously identified. Mutating the G-motif to alter the interactions with the +1 position of the template strand significantly affects the

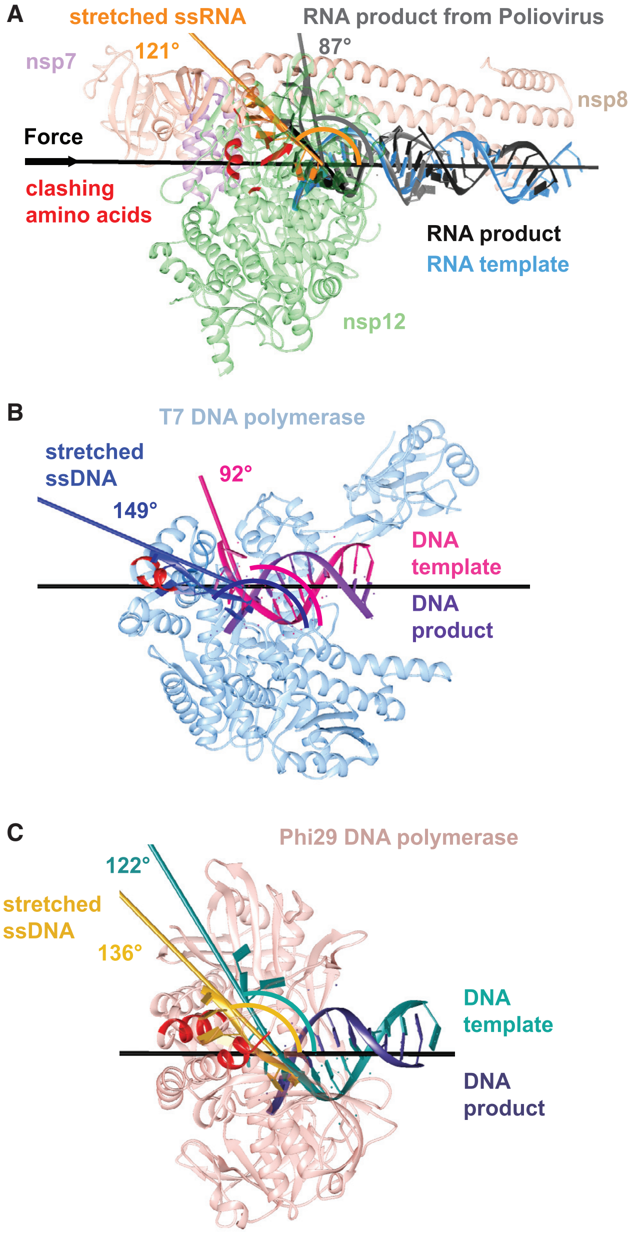


Figure 7. Comparison of structural models of elongating SARS-CoV-2 polymerase, T7 DNA polymerase, and Φ29 DNA polymerase with their respective downstream template strand positions

(A) SARS-CoV-2 polymerase-RNA complex (PDB: 6yyt) aligned and overlapped with the ssRNA template (gray) and poliovirus RdRp (not shown) structure (PDB: 3ol6). The angle measured between the force axis (black) and the ssRNA template is ~87°. This angle may increase up to ~121° (orange) on applied force, being limited by nsp12 residues (red).

(B) The angle measured between the force axis (black) and the ssDNA template (pink) of T7 DNA polymerase (PDB: 1t7p) is ~92° and may reach ~149° (purple) on applied force, being limited by the polymerase residues (red).

(C) The angle between the force axis (black) and the ssDNA template (teal) of Φ29 DNA polymerase (PDB: 2pyl) is ~122°, which may increase up to ~136° (gold) on applied force, being limited by the polymerase residues (red).

kinetics of nucleotide addition (Wang et al., 2020a). Because pause 1 and pause 2 probability increases with tension (Figures 4F and S4K), one could hypothesize that destabilizing the aforementioned interaction led the polymerase to the SNA/VSNA state. However, our previous study on poliovirus showed the error-prone mutant H273R of poliovirus RdRp induced a large increase in the VSNA pathway probability (Dulin et al., 2017), although the mutation is distant from the catalytic site (Korboukh et al., 2014). A network of interactions between the polymerase motifs, co-factors, and the primer-template likely lead the polymerase into the SNA and VSNA states in a stochastic manner. This state may be evolutionary conserved to regulate replication and transcription kinetics but also to enable the surveillance of the coronavirus polymerase by other viral co-factors.

How the proofreading exonuclease, nsp14, detects mismatch incorporation by nsp12 RdRp is unknown. Is there a kinetic competition between nsp12 and nsp14? Does a pause in the elongating nsp12 provide the time needed for the exonuclease to intervene, as for some high-fidelity DNA polymerases (Hoekstra et al., 2017; Manosas et al., 2012a)? We have identified three pauses in the SARS-CoV-2 RdRp elongation kinetics. Although the long-lived backtrack pauses are catalytically incompetent, pause 1 (SNA) and pause 2 (VSNA) are the signatures of slow nucleotide incorporations and therefore represent good candidates fitting the kinetic fingerprint of nucleotide misincorporation. Pause 1 probability is too high to be consistent with mismatch incorporation, but pause 2 kinetics are on par with pre-steady state kinetic analysis of SARS-CoV-1 polymerase, where the average synthesis rate of a 10-nt product including a UTP:G mismatch was measured at $(0.14 \pm 0.01)\ s^{-1}$ in the presence of 50 μM NTP (Shannon et al., 2020). At 25 pN and 50 μM NTP, we measured a similar exit rate for pause 2 at $(0.23 \pm 0.01)\ s^{-1}$ (Table S1). Pause 2 probability at the saturating concentration of NTP is similar to what we measured for poliovirus RdRp, where pause 2 was identified as the kinetic signature of nucleotide mismatch incorporation (Dulin et al., 2017). This would suggest that the fidelity of the coronavirus polymerase is on par with other viral RdRps. Furthermore, in a model where pause 2 is the kinetic signature of nucleotide mismatch incorporation, the lifetime of pause 2, $(k^2_{max})^{-1} \approx 0.5$s would indicate the lower bound time interval for nsp14 to intervene. Future studies investigating the elongation kinetics of a—yet to identify—low-fidelity mutant of the SARS-CoV-2 polymerase will identify the pathway to mismatch incorporation and whether it is associated with the VSNA pathway.

By significantly improving the stability of our magnetic tweezers assay, we directly monitored SARS-CoV-2 polymerase backtracking as deep as ~30 nt upstream of the last incorporated nucleotide (Figures 5C, 5D, and S5I). Our result supports a recent cryo-EM study on a pre-assembled backtracked coronavirus RdRp complex (Malone et al., 2021). The dwell time distribution of pauses resulting from polymerase backtrack is mathematically described by a power law with a −3/2 exponent (Depken et al., 2009), which is consistent with our findings for the long-lived third pause. Altogether, our direct observation and dwell time analysis support polymerase backtracking as the origin for the long-lived pauses, which further generalizes backtracking as a common property in the viral RdRp world (Seifert et al., 2020b; Dulin et al., 2015d, 2017). We show here that SARS-CoV-2 polymerase backtracking is strongly stimulated by the presence of secondary structure downstream the polymerase (Figure 5F). Because the coronavirus genome is heavily structured (Lan et al., 2020), polymerase backtracking would likely be ubiquitous during coronavirus genome replication and transcription without the assistance of other viral co-factors. The helicase nsp13 or the single-stranded RNA binding protein nsp9 are good candidates for this support role, and several cryo-EM studies have reported that nsp9 and nsp13 associate to the coronavirus polymerase (Chen et al., 2020; Yan et al., 2020, 2021). A very recent study even identifies nsp13 as an essential component to establish the polymerase into a stable backtrack state (Malone et al., 2021). Therefore, nsp13 potentially modulates coronavirus polymerase backtracking, which may be essential to promote strand switching during viral genome recombination and transcription (Sola et al., 2015; Jochheim et al., 2021).

Having characterized the three catalytic pathways of the coronavirus polymerase, it is now possible to understand by which pathway it incorporates nucleotide analogs. Most importantly, we show that the entry into any of the slow nucleotide addition pathways occurs before nucleotide (analog) binding, and therefore reaction conditions (as NTP concentration) are extremely important to ensure proper evaluation of how nucleotide analogs are incorporated and the yield of incorporation. In the companion study, we leverage the benchmark provided here to fingerprint the action of various nucleotide analogs, demonstrating why remdesivir is better incorporated than T-1106 (a chemically more stable version of favipiravir) (Seifert et al., 2020a). This result likely explains the higher efficacy of remdesivir against SARS-CoV-2 infected cells than favipiravir. Being able to tune the polymerase into a drug-incorporation-competent pathway using small-molecule binding may represent an avenue to further improve drug efficacy.

Replication and transcription is at the heart of coronavirus lifecycle and form an important target for drug development. However, a mechanistic understanding of these processes has been lacking, forming an obstacle to efficiently target the key aspect that regulates viral genome synthesis in infected cells. Here, we provide a complete mechanistic framework describing the coronavirus core polymerase nucleotide addition cycle. We will leverage this single molecular platform to elucidate the structure-function-dynamics relationship of the multi-subunit RdRp complexes processing the coronavirus genome, offering new targets for drug development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
    - ○ Lead contact
    - ○ Materials availability
    - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
    - ○ High-throughput magnetic tweezers apparatus
    - ○ Ultra-stable magnetic tweezers
    - ○ Recombinant Protein Expression of RdRp (nsp12) and cofactors (nsp7 and nsp8) from SARS-CoV-2
    - ○ Construct fabrication
    - ○ Flow-cell assembly
    - ○ Single molecule SARS-CoV-2 polymerase activity experiments
    - ○ Data processing
    - ○ SARS-CoV-2 activity trace product length analysis
    - ○ Model fitting
    - ○ Stochastic pausing model
    - ○ Maximum likelihood estimation fitting routine
    - ○ Modeling of Pause 1 propensity under force and concentration sweeps
    - ○ The reaction kinetic scheme
    - ○ The observables
    - ○ Catalytic rate
    - ○ Pause 1 probability
    - ○ The effect of a dsRNA structure downstream the polymerase
    - ○ Structure modeling
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109650.

ACKNOWLEDGMENTS

D.D. thanks OICE for hosting his lab. D.D. was supported by the Interdisciplinary Center for Clinical Research (IZKF) at the University Hospital of the University of Erlangen-Nuremberg, the German Research Foundation (DFG-DU-1872/3-1), the Netherlands Ministry of Education, Culture and Science (OCW) BaSyC (Building a Synthetic Cell) Gravitation grant (024.003.019), and the Netherlands Organisation for Scientific Research (NWO). R.N.K. was supported by NIAID NIH (AI123498). B.C. acknowledge grants by the Fondation pour la Recherche Médicale (Aide aux équipes), the SCORE project H2020 SC1-PHE-Coronavirus-2020 (101003627), and the REACTing initiative (REsearch and ACTion targeting emerging infectious diseases). J.J.A. and C.E.C. were supported by NIAID NIH (AI045818).

AUTHOR CONTRIBUTIONS

J.J.A., B.C., C.E.C., and D.D. conceived the research. S.C.B. and M.S. performed the single-molecule experiments. S.C.B., M.S., M.D., and D.D. analyzed the data. P.v.N. wrote the MLE analysis routine. F.S.P. made the RNA used for the study. Y.W. helped to develop the magnetic tweezers autofocus. R.N.K. provided SARS-CoV-2 polymerase proteins. M.D. performed theoretical analysis and modeling. S.Q. derived the structural model of the polymerases. All the authors have discussed and interpreted the results and participated in writing the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 10, 2021
Revised: July 10, 2021
Accepted: August 11, 2021
Published: August 31, 2021

REFERENCES

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460–465.

Arnold, J.J., and Cameron, C.E. (2004). Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg^{2+}. Biochemistry *43*, 5126–5137.

Berman, A.J., Kamtekar, S., Goodman, J.L., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M., and Steitz, T.A. (2007). Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. EMBO J. *26*, 3494–3505.

Bustamante, C., Chemla, Y.R., Forde, N.R., and Izhaky, D. (2004). Mechanical processes in biochemistry. Annu. Rev. Biochem. *73*, 705–748.

Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T., et al. (2020). Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. Cell *182*, 1560–1573.e13.

Chien, M., Anderson, T.K., Jockusch, S., Tao, C., Li, X., Kumar, S., Russo, J.J., Kirchdoerfer, R.N., and Ju, J. (2020). Nucleotide Analogues as Inhibitors of SARS-CoV-2 Polymerase, a Key Drug Target for COVID-19. J. Proteome Res. *19*, 4690–4697.

Cnossen, J.P., Dulin, D., and Dekker, N.H. (2014). An optimized software framework for real-time, high-throughput tracking of spherical beads. Rev. Sci. Instrum. *85*, 103712.

Cowan, G. (1998). Statistical Data Analysis (Oxford University Press).

Dangerfield, T.L., Huang, N.Z., and Johnson, K.A. (2020). Remdesivir Is Effective in Combating COVID-19 because It Is a Better Substrate than ATP for the Viral RNA-Dependent RNA Polymerase. iScience *23*, 101849.

Dangkulwanich, M., Ishibashi, T., Liu, S., Kireeva, M.L., Lubkowska, L., Kashlev, M., and Bustamante, C.J. (2013). Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. eLife *2*, e00971.

Depken, M., Galburt, E.A., and Grill, S.W. (2009). The origin of short transcriptional pauses. Biophys. J. *96*, 2189–2193.

Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J.S., Gough, G.R., Gilham, P.T., and Moras, D. (1989). Crystallographic structure of an RNA helix: [U(UA)6A]2. J. Mol. Biol. *209*, 459–474.

Doublié, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature *391*, 251–258.

Dulin, D., Berghuis, B.A., Depken, M., and Dekker, N.H. (2015a). Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. Curr. Opin. Struct. Biol. *34*, 116–122.

Dulin, D., Cui, T.J., Cnossen, J., Docter, M.W., Lipfert, J., and Dekker, N.H. (2015b). High Spatiotemporal-Resolution Magnetic Tweezers: Calibration and Applications for DNA Dynamics. Biophys. J. *109*, 2113–2125.

Dulin, D., Vilfan, I.D., Berghuis, B.A., Hage, S., Bamford, D.H., Poranen, M.M., Depken, M., and Dekker, N.H. (2015c). Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. Cell Rep. *10*, 983–992.

Dulin, D., Vilfan, I.D., Berghuis, B.A., Poranen, M.M., Depken, M., and Dekker, N.H. (2015d). Backtracking behavior in viral RNA-dependent RNA polymerase provides the basis for a second initiation site. Nucleic Acids Res. *43*, 10421–10429.

Dulin, D., Arnold, J.J., van Laar, T., Oh, H.S., Lee, C., Perkins, A.L., Harki, D.A., Depken, M., Cameron, C.E., and Dekker, N.H. (2017). Signatures of Nucleotide Analog Incorporation by an RNA-Dependent RNA Polymerase Revealed Using High-Throughput Magnetic Tweezers. Cell Rep. *21*, 1063–1076.

Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L., and Denison, M.R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J. Virol. *81*, 12135–12144.

Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B., et al. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog. *6*, e1000896.

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., and Imbert, I. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. Proc. Natl. Acad. Sci. USA *115*, E162–E171.

Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R.H., and Bustamante, C.J. (2018). Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. Nat. Commun. *9*, 2930.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science *368*, 779–782.

Gong, P., and Peersen, O.B. (2010). Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. USA *107*, 22505–22510.

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.

Hoekstra, T.P., Depken, M., Lin, S.N., Cabanas-Danés, J., Gross, P., Dame, R.T., Peterman, E.J.G., and Wuite, G.J.L. (2017). Switching between Exonucleolysis and Replication by T7 DNA Polymerase Ensures High Fidelity. Biophys. J. *112*, 575–583.

Huhle, A., Klaue, D., Brutzer, H., Daldrop, P., Joo, S., Otto, O., Keyser, U.F., and Seidel, R. (2015). Camera-based three-dimensional real-time particle tracking at kHz rates and Ångström accuracy. Nat. Commun. *6*, 5885.

Ibarra, B., Chemla, Y.R., Plyasunov, S., Smith, S.B., Lázaro, J.M., Salas, M., and Bustamante, C. (2009). Proofreading dynamics of a processive DNA polymerase. EMBO J. *28*, 2794–2802.

Jochheim, F.A., Tegunov, D., Hillen, H.S., Schmitzova, J., Kokic, G., Dienemann, C., and Cramer, P. (2021). Dimeric form of SARS-CoV-2 polymerase. bioRxiv. https://doi.org/10.1101/2021.03.23.436644.

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.

Korboukh, V.K., Lee, C.A., Acevedo, A., Vignuzzi, M., Xiao, Y., Arnold, J.J., Hemperly, S., Graci, J.D., August, A., Andino, R., and Cameron, C.E. (2014). RNA virus population diversity, an optimum for maximal fitness and virulence. J. Biol. Chem. *289*, 29531–29544.

Lan, T.C.T., Allan, M.F., Malsick, L.E., Khandwala, S., Nyeo, S.S.Y., Bathe, M., Griffiths, A., and Rouskin, S. (2020). Structure of the full SARS-CoV-2 RNA genome in infected cells. bioRxiv. https://doi.org/10.1101/2020.06.29.178343.

Lauring, A.S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. Nat. Rev. Microbiol. *11*, 327–336.

Lipfert, J., Skinner, G.M., Keegstra, J.M., Hensgens, T., Jager, T., Dulin, D., Köber, M., Yu, Z., Donkers, S.P., Chou, F.C., et al. (2014). Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA. Proc. Natl. Acad. Sci. USA *111*, 15408–15413.

Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T., et al. (2021). Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. Proc. Natl. Acad. Sci. USA *118*, e2102516118.

Manosas, M., Spiering, M.M., Ding, F., Bensimon, D., Allemand, J.F., Benkovic, S.J., and Croquette, V. (2012a). Mechanism of strand displacement synthesis by DNA replicative polymerases. Nucleic Acids Res. *40*, 6174–6186.

Manosas, M., Spiering, M.M., Ding, F., Croquette, V., and Benkovic, S.J. (2012b). Collaborative coupling between polymerase and helicase for leading-strand synthesis. Nucleic Acids Res. *40*, 6187–6198.

Melkonyan, L., Bercy, M., Bizebard, T., and Bockelmann, U. (2019). Overstretching Double-Stranded RNA, Double-Stranded DNA, and RNA-DNA Duplexes. Biophys. J. *117*, 509–519.

Morin, J.A., Cao, F.J., Lázaro, J.M., Arias-Gonzalez, J.R., Valpuesta, J.M., Carrascosa, J.L., Salas, M., and Ibarra, B. (2012). Active DNA unwinding dynamics during processive DNA replication. Proc. Natl. Acad. Sci. USA *109*, 8115–8120.

Ostrofet, E., Papini, F.S., and Dulin, D. (2018). Correction-free force calibration for magnetic tweezers experiments. Sci. Rep. *8*, 15920.

Ostrofet, E., Papini, F.S., Malinen, A.M., and Dulin, D. (2019). A single-molecule view on cellular and viral RNA synthesis. In Biophysics of RNA-Protein Interactions, C. Joo and D. Rueda, eds. (Springer, New York).

Ostrofet, E., Papini, F.S., and Dulin, D. (2020). High spatiotemporal resolution data from a custom magnetic tweezers instrument. Data Brief *30*, 105397.

Pan, H., Peto, R., Henao-Restrepo, A.M., Preziosi, M.P., Sathiyamoorthy, V., Abdool Karim, Q., Alejandria, M.M., Hernández García, C., Kieny, M.P., Malekzadeh, R., et al.; WHO Solidarity Trial Consortium (2021). Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. N. Engl. J. Med. *384*, 497–511.

Papini, F.S., Seifert, M., and Dulin, D. (2019). High-yield fabrication of DNA and RNA constructs for single molecule force and torque spectroscopy experiments. Nucleic Acids Res. *47*, e144.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1992). Numerical Recipes in C: The Art of Scientific Computing (Cambridge University Press).

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.

Seifert, M., Bera, S.C., Van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.-T.-N., Grove, T.L., Papini, F.S., Arnold, J.J., Almo, S.C., et al. (2020a). Signatures and mechanisms of efficacious therapeutic ribonucleotides against SARS-CoV-2 revealed by analysis of its replicase using magnetic tweezers. bioRxiv. https://doi.org/10.1101/2020.08.06.240325.

Seifert, M., van Nies, P., Papini, F.S., Arnold, J.J., Poranen, M.M., Cameron, C.E., Depken, M., and Dulin, D. (2020b). Temperature controlled high-throughput magnetic tweezers show striking difference in activation energies of replicating viral RNA-dependent RNA polymerases. Nucleic Acids Res. *48*, 5591–5602.

Shaevitz, J.W., Abbondanzieri, E.A., Landick, R., and Block, S.M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. Nature *426*, 684–687.

Shannon, A., Selisko, B., Le, N.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.

Sholders, A.J., and Peersen, O.B. (2014). Distinct conformations of a putative translocation element in poliovirus polymerase. J. Mol. Biol. *426*, 1407–1419.

Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M., and Denison, M.R. (2013). Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. PLoS Pathog. *9*, e1003565.

Snijder, E.J., Decroly, E., and Ziebuhr, J. (2016). The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. Adv. Virus Res. *96*, 59–126.

Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. Annu. Rev. Virol. *2*, 265–288.

Tvarogová, J., Madhugiri, R., Bylapudi, G., Ferguson, L.J., Karl, N., and Ziebuhr, J. (2019). Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3′-Terminal Adenylyltransferase Activity. J. Virol. *93*, e00291-19.

Vieregg, J., Cheng, W., Bustamante, C., and Tinoco, I., Jr. (2007). Measurement of the effect of monovalent cations on RNA hairpin stability. J. Am. Chem. Soc. *129*, 14966–14973.

Wang, M., Li, R., Shu, B., Jing, X., Ye, H.Q., and Gong, P. (2020a). Stringent control of the RNA-dependent RNA polymerase translocation revealed by multiple intermediate structures. Nat. Commun. *11*, 2605.

Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C., et al. (2020b). Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. Cell *182*, 417–428.e13.

Wang, B., Svetlov, V., Wolf, Y.I., Koonin, E.V., Nudler, E., and Artsimovitch, I. (2021). Allosteric Activation of SARS-CoV-2 RNA-Dependent RNA Polymerase by Remdesivir Triphosphate and Other Phosphorylated Nucleotides. MBio *12*, e0142321.

Wuite, G.J., Smith, S.B., Young, M., Keller, D., and Bustamante, C. (2000). Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. Nature *404*, 103–106.

Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M., et al. (2020). Architecture of a SARS-CoV-2 mini replication and transcription complex. Nat. Commun. *11*, 5874.

Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M., et al. (2021). Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. Cell *184*, 184–193.e10.

Yang, X., Smidansky, E.D., Maksimchuk, K.R., Lum, D., Welch, J.L., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2012). Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. Structure *20*, 1519–1527.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Digoxigenin	Roche	Cat # 11333089001, RRID:AB_514496
Chemicals, peptides, and recombinant proteins		
Dynabeads M-270 Streptavidin	Invitrogen	Cat # 65305
HEPES	Roth	CAS # 7365-45-9
NaCl	Roth	CAS # 7647-14-5
Potassium Glutamate	Sigma	CAS # 6382-01-0
Magnesium chloride	Roth	CAS # 7791-18-6
EDTA	Roth	CAS # 6381-92-6
BSA	Roth	CAS # 90604-29-8
Polystyrene beads	Sigma	SKU # LB3-1ML
NTP	Promega	Cat # P1132
Nsp12, nsp7 and nsp8	Kindly provided by Robert N. Kirchdoerfer	N/A
Biotin-16-UTP	Jena Bioscience	NU-821-BIO16
Digoxigenin-11-UTP	Jena Bioscience	NU-821-DIGX
Antarctic Phosphatase	New England Biolabs	M0289
T4 Polynucleotide Kinase	New England Biolabs	M0201
T4 RNA ligase 2	New England Biolabs	M0239
SyBrSafe Nucleic Acid Gel Stain	Invitrogen	S33102
Agarose	Sigma-Aldrich	A9539
TAE buffer 50x	Thermo Scientific	B49
TBE buffer 10x	Carl Roth	3061
RNA Gel Loading Dye (2X)	Thermo Scientific	R0641
Critical commercial assays		
Phusion High-Fidelity PCR kit	New England Biolabs	E0553
Monarch PCR & DNA Cleanup Kit	New England Biolabs	T1030
HiScribe T7 High Yield RNA Synthesis Kit	New England Biolabs	E2040
RNA Clean & Concentrator	Zymo Research	R1015, R1017
Oligonucleotides		
TAATACGACTCACTATAGGCCTCACTTCTGCTATTTCGC	Papini et al., 2019	P42
TAATACGACTCACTATAGGTCTTCGCCAGACGGCATTTA	Papini et al., 2019	P43
TAATACGACTCACTATAGGGCAGGCAAGTCCGATTTTTTG	Papini et al., 2019	P44
GGAACCAAAGGATATTCAGACG	Papini et al., 2019	P45
TAATACGACTCACTATAGGGGAACCAAAGGATATTCAGACG	Papini et al., 2019	P46
AACAAGAAACTTCCTTGGCTG	Papini et al., 2019	P47
TAATACGACTCACTATAGGTGCTGAGGAACCGGAGTG	Papini et al., 2019	P48
TAATACGACTCACTATAGGGTAGTGATTAACATTCGACAGCATGCGCAC	Papini et al., 2019	P49
CAGGATCACGTTACCGCC	Papini et al., 2019	P50
TAATACGACTCACTATAGGAACGGCGCCTATGACG	Papini et al., 2019	P51
TGGATCCGTGGGCGC	Papini et al., 2019	P52

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GCGACTTAGCTGAGGCC	Papini et al., 2019	P20
GGTGCCCACAGAACGTC	Papini et al., 2019	P24
Recombinant DNA		
pMTM2	Papini et al., 2019	https://doi.org/10.1093/nar/gkz851
Software and algorithms		
Python 3	Python Software Foundation	https://www.python.org/download/releases/3.0/
MLE fitting Python 3 routine of the pause stochastic model	Zenodo	https://doi.org/10.5281/zenodo.5179600
LabVIEW	National Instruments	https://www.ni.com/en-us.html
Other		
Objective Lens	Nikon	MRL01502
Camera	Teledyne DALSA Falcon2	FA-80-12M1H
Opto-mechanics	Thorlabs	https://www.thorlabs.de/
Tube lens	Qioptic	G322304000
Piezo	Physic Instrument (PI)	P-726.1CD
Motors	Physic Instrument (PI)	M-126-PD1, C-150
Pump	Cole-Parmer GmbH	Item # 77122-32
LED	OSRAM Opto	GTIN # 4062986000272
Permanent magnet	Supermagnet	Cat # W-05-G
Cover glass	OMNILAB	5161060
Foil heater	Thorlabs	HT10K
SARS-CoV-2 pol complex	PDB	6yyt
Poliovirus RNA pol complex	PDB	3ol6
T7 DNA pol complex	PDB	lt7p
Phi29 DNA pol complex	PDB	2pyl

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. David Dulin (d.dulin@vu.nl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All original data can be made availbale upon reasonable request to the lead author.

The original code to perform the MLE fitting of the stochastic-pausing model is available on Zenodo: https://doi.org/10.5281/zenodo.5179600

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No animal or human subjects or plants or microbe strains or cell lines are used in this study.

METHOD DETAILS

High-throughput magnetic tweezers apparatus

The high-throughput magnetic tweezers used in this study have been described in detail elsewhere (Ostrofet et al., 2018). Shortly, a pair of vertically aligned permanent magnets (5 mm cubes, SuperMagnete, Switzerland) separated by a 1 mm gap are positioned above a flow cell (see paragraph below) which is mounted on a custom-built inverted microscope. The vertical position and rotation of the magnets are controlled by two linear motors, M-126-PD1 and C-150 (Physik Instrumente PI, GmbH & Co. KG, Karlsruhe, Germany), respectively. The field of view is illuminated through the magnets gap by a collimated LED-light source, and is imaged onto a large chip CMOS camera (Dalsa Falcon2 FA-80-12M1H, Stemmer Imaging, Germany) using a 50 × oil immersion objective (CFI Plan Achro 50 XH, NA 0.9, Nikon, Germany) and an achromatic doublet tube lens of 200 mm focal length and 50 mm diameter (Qioptic, Germany). To control the temperature, we used a system described in details in Seifert et al. (2020b). A flexible resistive foil heater with an integrated 10 MΩ thermistor (HT10K, Thorlabs) is wrapped around the microscope objective and further insulated by several layers of kapton tape (KAP22-075, Thorlabs). The heating foil is connected to a PID temperature controller (TC200 PID controller, Thorlabs) to adjust the temperature within ~0.1°C.

Ultra-stable magnetic tweezers

To improve the stability of the instrument, we introduced an autofocus (i.e., along the z axis) to the magnetic tweezers assay. We hypothesized that reference bead (a surface melted polystyrene bead) subtraction alone could not fully correct for large drift (> 100 nm) during long timescale (> 20 minutes) measurements. Therefore, correction for the drift by establishing an autofocus with a high resolution piezo nanopositioner would correct for such detrimental effect. To achieve this, we selected a suitable surface-attached reference bead in the field of view and considered it for autofocus in the tracking algorithm. During the acquisition, the current z axis position (average of last 60 z-positions) of the reference bead was compared every 1 s interval against its target z axis position (average of 60 z-positions at the beginning/suitable part of measurement). If the difference between the current and the target z-positions was larger than 0.5 nm, the drift/difference was subsequently corrected by adjusting the position of the objective using the piezo nanopositioner. Of note, the drift in our assay is much smaller than 0.5 nm/s. To quantitatively evaluate the improvement introduced by the autofocus into the measurement, we extracted the Allan deviation (AD) of a reference bead either with or without autofocus correction and either drift-corrected or not, i.e., by further subtracting the position of another reference bead. We estimated the AD at two different acquisition frequency, i.e., 58 and 500 Hz, using 100x microscope objective magnification (60 nm pixel size in the image plane) (Figures S5D–S5H). The AD compares the noise in the bead position for partially overlapping, successive time windows of increasing duration τ. Therefore, the AD quantifies the magnitude of the noise when average over a given time interval, and informs on its origin (Dulin et al., 2015b; Ostrofet et al., 2020). For shot-noise limited measurements, the AD decreases as $1/\sqrt{\tau}$, to eventually increases when other sources of noise, e.g., mechanical drift, cumulates and dominates at larger τ. Therefore, the minimum of AD indicates the time interval over which the assay is stable and the best resolution achievable. This is clearly visible in Figure S5G, where the AD at 58 Hz acquisition frequency reaches a minimum of $AD \sim 0.1\,nm$ in absence of reference bead subtraction at $\tau \sim 1\,s$. This minimum decreases to $AD \sim 0.01\,nm$ at $\tau \sim 30\,s$ when subtracting the reference bead position. Applying the autofocus, the drift is further reduced, with a minimum of $AD \sim 0.008\,nm$ at $\tau \sim 100\,s$, indicating a significant improvement in stability. Similar results were obtained at 500 Hz (Figure S5H). We note that the resolution does not significantly improve below $AD \sim 0.008\,nm$, as we likely reached a physical limit in what is achievable in position correction for our assay.

Recombinant Protein Expression of RdRp (nsp12) and cofactors (nsp7 and nsp8) from SARS-CoV-2

This protocol was described Chien et al. (2020). SARS-CoV-2 nsp12: The SARS-CoV-2 nsp12 gene was codon optimized and cloned into pFastBac with C-terminal additions of a TEV site and strep tag (Genscript). The pFastBac plasmid and DH10Bac *E. coli* (Life Technologies) were used to create recombinant bacmids. The bacmid was transfected into Sf9 cells (Expression Systems) with Cellfectin II (Life Technologies) to generate recombinant baculovirus. The baculovirus was amplified through two passages in Sf9 cells, and then used to infect 1 L of Sf21 cells (Expression Systems) and incubated for 48 hr at 27°C. Cells were harvested by centrifugation, resuspended in wash buffer (25 mM HEPES pH 7.4, 300 mM NaCl, 1 mM $MgCl_2$, 5 mM DTT) with 143 µL of BioLock per liter of culture. Cells were lysed via microfluidization (Microfluidics). Lysates were cleared by centrifugation and filtration. The protein was purified using StrepTactin Superflow agarose (IBA). StrepTactin eluted protein was further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Life Sciences) in 25 mM HEPES, 300 mM NaCl, 100 µM $MgCl_2$, 2 mM TCEP, at pH 7.4. Pure protein was concentrated by ultrafiltration prior to flash freezing in liquid nitrogen. SARS-CoV-2 nsp7 and nsp8: The SARS-CoV-2 nsp7 and nsp8 genes were codon optimized and cloned into pET46 (Novagen) with an N-terminal 6x histidine tag, an enterokinase site, and a TEV protease site. Rosetta2 pLys *E. coli* cells (Novagen) were used for bacterial expression. Cultures were grown to an OD600 of 0.8 and induced with a final concentration of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and growth temperature was reduced to 16°C for 16 hr. Cells were harvested by centrifugation and pellets were resuspended in wash buffer (10 mM Tris pH 8.0, 300 mM NaCl, 30 mM imidazole, 2 mM DTT). Cells were lysed via microfluidization and lysates were cleared by centrifugation and filtration. Proteins were purified using Ni-NTA agarose beads (QIAGEN) and eluted with wash buffer containing 300 mM imidazole. Eluted proteins were digested with 1% w/w TEV protease during overnight room temperature dialysis (10 mM Tris pH 8.0, 300 mM NaCl, 2 mM DTT). Digested proteins were passed back over Ni-NTA to remove undigested protein before concentrating the proteins by ultrafiltration. Nsp7 and nsp8 proteins were further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Life Sciences). Purified proteins were concentrated by ultrafiltration prior to flash freezing with liquid nitrogen.

Construct fabrication

The fabrication of the RNA hairpin has been described in detail in Papini et al. (2019). The RNA hairpin is made of a 499 bp double-stranded RNA stem terminated by a 20 nt loop that is assembled from three ssRNA annealed together (Figure S1A), and two handles, one of 856 bp at the 5′ end and one 822 bp at the 3′ end. The handles include either a 343 nt digoxygenin-labeled ssRNA or a 443 nt biotin-labeled ssRNA (Figure S1A). Upon applied force above ∼22 pN, the hairpin opens and frees a 1043 nt ssRNA template for SARS-CoV-2 polymerase (Figure S1B). To obtain the different parts of the RNA construct, template DNA fragments were amplified via PCR, purified (Monarch PCR and DNA cleanup kit) and *in vitro* transcribed (NEB HiScribe T7 High Yield RNA Synthesis Kit). Transcripts were then treated with Antarctic Phosphatase and T4 Polynucleotide Kinase. RNAs were purified using the RNA Clean & Concentrator-25 kit (Zymo Research). Individual RNA fragments were annealed and ligated with T4 RNA ligase 2 (NEB) to assemble the RNA hairpin.

RNA template for SARS-CoV-2 magnetic tweezers experiments (highlighted in yellow and in gray are the loop and the ssRNA template preceding the hairpin stem where the polymerase loads, respectively)

GUUCUACAUAGCGUGCAGACGUGAAUUUAAUCUCGCUGACGUGUAGACACAGUGCGUCUGCUGUCGGGUCCCUCUGGUGACUGGGUAGUUGGACUUGCCCUUGGAAGACAUAGCAAGACCCUGCCUCUCUAUUGAUGUCACGGCGAAUGUCGGGGAGACAGCAGCGGCUGCAGACAUCAGAUCGGAGUAAUACUCUCCGUAACUGGCCUUCUCUGAAUUCCGACGUUGUUAAGAUGGCAGAGCCCGGUAAUCGCUACUUGACCAGAUAAGCUUUCCGUGGAUGGUUUAGAGGAAUCACAUCCAAGACUGGCUAAGCACGAAGCAACUCUUGAGUGUAAAAUUGUUGUCUCCUGUAUUCGGGAUGCGGGUACUAGAUGACUGCAGGGACUCCGACGUUAAGUACAUUACCCCGUCAUAGGCGCCGUUCAGGAUCACGUUACCGCCAUAAGAUGGGAGCAUGACUUCUUCUCCGCUGCGCCCACGGAUCCAGUAGUGAUUAACAUUCGACAGCAUGCGCACUAAUCACUACUGGAUCCGUGGGCGCAGCGGAGAAGAAGUCAUGCUCCCAUCUUAUGGCGGUAACGUGAUCCUGAACGGCGCCUAUGACGGGGUAAUGUACUUAACGUCGGAGUCCCUGCAGUCAUCUAGUACCCGCAUCCCGAAUACAGGAGACAACAAUUUUACACUCAAGAGUUGCUUCGUGCUUAGCCAGUCUUGGAUGUGAUUCCUCUAAACCAUCCACGGAAAGCUUAUCUGGUCAAGUAGCGAUUACCGGGCUCUGCCAUCUUAACAACGUCGGAAUUCAGAGAAGGCCAGUUACGGAGAGUAUUACUCCGAUCUGAUGUCUGCAGCCGCUGCUGUCUCCCCGACAUUCGCCGUGACAUCAAUAGAGAGGCAGGGUCUUGCUAUGUCUUCCAAGGGCAAGUCCAACUACCCAGUCACCAGAGGGACCCGACAGCAGACGCACUGUGUCUACACGUCAGCGAGAUUAAAUUCACGUCUGCACGCUAUGUAGAACCCUCAGCCAACUCGGUCGCGUCGGA

The template contains 250 U (24%), 253 A (24%), 273 C (26%) and 267 G (26%).

Flow-cell assembly

The fabrication procedure for flow cells has been described in details in Ostrofet et al. (2018). To summarize, we sandwiched a double layer of Parafilm by two #1 coverslips, the top one having one hole at each end serving as inlet and outlet, the bottom one being coated with a 0.01% m/V nitrocellulose in amyl acetate solution. The flow cell is mounted into a custom-built holder and rinsed with ∼1 mL of 1x phosphate buffered saline (PBS). 3 μm diameter polystyrene reference beads are attached to the bottom coverslip surface by incubating 100 μL of a 1:1000 dilution in PBS (LB30, Sigma Aldrich, stock conc.: 1.828*1011 particles per milliliter) for ∼3 minutes. The tethering of the magnetic beads by the RNA hairpin construct relies on a digoxygenin/anti-digoxygenin and biotin-streptavidin attachment at the coverslip surface and the magnetic bead, respectively. Therefore, following a thorough rinsing of the flow cell with PBS, 50 μL of anti-digoxigenin (50 μg/mL in PBS) is incubated for 30 minutes. The flow cell was flushed with 1 mL of 10 mM Tris, 1 mM EDTA pH 8.0, 750 mM NaCl, 2 mM sodium azide buffer to remove excess of anti-digoxigenin followed by rinsing with another 0.5 mL of 1x TE buffer (10 mM Tris, 1 mM EDTA pH 8.0 supplemented with 150 mM NaCl, 2 mM sodium azide). The surface is then passivated by incubating bovine serum albumin (BSA, New England Biolabs, 10 mg/mL in PBS and 50% glycerol) for 30 minutes, and rinsed with 1x TE buffer.

Single molecule SARS-CoV-2 polymerase activity experiments

20 μL of streptavidin coated Dynal Dynabeads M-270 streptavidin coated magnetic beads (ThermoFisher) was mixed with ∼0.1 ng of RNA hairpin (total volume 40 μL) and incubated for ∼5 minutes before rinsing with ∼2 mL of 1x TE buffer to remove any unbound RNA and the magnetic beads in excess. RNA tethers were sorted for functional hairpins by looking for the characteristic jump in extension of the correct length (∼0.5 μm at 30 pN) due to the sudden opening of the hairpin during a force ramp experiment (Papini et al., 2019). The flow cell was subsequently rinsed with 0.5 mL reaction buffer (50 mM HEPES pH 7.9, 10 mM DTT, 2 μM EDTA, 5 mM $MgCl_2$). After starting the data acquisition at a force that would keep the hairpin open, 100 μL of reaction buffer containing the indicated concentrations of nsp12, nsp7, nsp8 (1:9:9 stoichiometry for experiments presented in Figure 2, and 1:3:3 for all other experiments) and NTPs were flushed in the flow cell to start the reaction. For the pre-assembled polymerase experiments, 0.6 μM nsp12, 1.8 μM nsp7 and nsp8 were incubated for five minutes in the flow cell, while applying 35 pN force on the tether. The excess polymerase proteins were subsequently flushed away with 0.3 mL of reaction buffer (flow cell volume ∼40 μL), followed by the injection of 100 μL of reaction buffer with 500 μM NTP. The experiments were conducted at a constant force as indicated for a duration of 20 to 60 minutes. The camera frame rate was fixed at 58 Hz and the temperature set to 25°C. A custom written Labview routine controlled the data acquisition and the (x-, y-, z-) positions analysis/tracking of both the magnetic and reference beads in real-time (Cnossen et al., 2014). Mechanical drift correction was performed by subtracting the reference bead position from the magnetic bead positions.

Data processing

The activity traces of SARS-CoV-2 polymerase converts the tether from ssRNA to dsRNA, which concomitantly decreases the end-to-end extension of the tether. The change in extension measured in micron was subsequently converted into replicated nucleotides N_R, low-pass filtered at 2 Hz and the dwell times were extracted using a dwell time window of 10 nt as described in Seifert et al. (2020a). The choice of a dwell time window size of 10 nt is constrained by the resolution of the assay, i.e., position fluctuation due to the noise should not influence the dwell time distribution, and to minimize the probability of entering any pause states several times per dwell time (Dulin et al., 2015a, 2015c). For data acquired on the closed hairpin, the data was converted into nucleotides based on the average increase in extension for completed activity traces and low-pass filtered at 1 Hz. Dwell times were extracted as for the data on the ssRNA. The dwell times of all the traces for a given experimental condition were assembled and further analyzed using a maximum likelihood estimation (MLE) fitting routine to extract the parameters from the stochastic-pausing model.

SARS-CoV-2 activity trace product length analysis

To extract the product length of the polymerase complex, only the traces where the beginning and the end could clearly be distinguished and for which the tether did not rupture for ten minutes following the last observed elongation activity were considered. We represented the mean product length, as well as one standard deviation of the mean from 1000 bootstraps as error bars.

Model fitting

There are many kinetic models that are consistent with the empirical dwell-time distributions we observe. The first thing we need to account for is multiple entry into a pause state in one dwell-time window.

Stochastic pausing model

For concreteness, we first consider the situation where we have three characteristic times (corresponding to nucleotide addition, and two to two different pauses) in a single nucleotide addition step. Based on the data, we assume that we have separation of timescales $\tau_0 \ll \tau_1 \ll \tau_2$, and that each process (nucleotide addition, or pausing) dominates the single-nt dwell-time distribution for times around its characteristic timescale. This assumption washes out most details of the kinetic scheme that connects the pauses with nucleotide addition, but allows us to determine the general form of the dwell-time distribution without specifying how the pauses are connected to the nucleotide addition pathway. Under the above assumptions, we have the approximate single-nt dwell-time distribution

$$P_{1\mathrm{nt}}(t) = \frac{p_0}{\tau_0} e^{-\frac{t}{\tau_0}} + \frac{p_1}{\tau_1} e^{-\frac{t}{\tau_1}} + \frac{p_2}{\tau_2} e^{-\frac{t}{\tau_2}}, \; \tau_0 < \tau_1 < \tau_2, p_0 + p_1 + p_2 = 1.$$

Moving over to Laplace space this becomes

$$\psi_{1\mathrm{nt}}(s) = \int_0^\infty dt P_{1\mathrm{nt}}(t) e^{-st} = \frac{p_0}{1+\tau_0 s} + \frac{p_1}{1+\tau_1 s} + \frac{p_2}{1+\tau_2 s},$$

Using the fact that convolutions in real space are products in Laplace space, the first passage time distribution across a N nucleotide window can be written as

$$\psi_{N\mathrm{nt}} = \left(\frac{p_0}{1+\tau_0 s} + \frac{p_1}{1+\tau_1 s} + \frac{p_2}{1+\tau_2 s} \right)^N$$

$$= \sum_{n_0=0}^{N} \sum_{n_1=0}^{N-n_0} \binom{N}{n_0, n_1, N-n_0-n_1} p_0^{n_0} p_1^{n_1} p_2^{N-n_0-n_1} \left(\frac{1}{1+\tau_0 s} \right)^{n_0} \left(\frac{1}{1+\tau_1 s} \right)^{n_1} \left(\frac{1}{1+\tau_2 s} \right)^{N-n_0-n_1}$$

$$= \underbrace{p_0^N \left(\frac{1}{1+\tau_0 s} \right)^N}_{\text{only fast timescale}} + \underbrace{\sum_{n_0=0}^{N-1} \left(N^{n_0, N-n_0, 0} \right) p_0^{n_0} p_1^{N-n_0} \left(\frac{1}{1+\tau_0 s} \right)^{n_0} \left(\frac{1}{1+\tau_1 s} \right)^{N-n_0}}_{\text{at least one intermediate but no slow timescale}}$$

$$+ \underbrace{\sum_{n_0=0}^{N-1} \sum_{n_1=0}^{N-n_0-1} \left(N^{n_0, n_1, N-n_0-n_1} \right) p_0^{n_0} p_1^{n_1} p_2^{N-n_0-n_1} \left(\frac{1}{1+\tau_0 s} \right)^{n_0} \left(\frac{1}{1+\tau_1 s} \right)^{n_1} \left(\frac{1}{1+\tau_2 s} \right)^{N-n_0-n_1}}_{\text{at least one slow timescale}}$$

In the last step we have separated out the terms that contain only the fast process (first term), at least one of the intermediate process but none of the slow process (second term), and those that contain at least one slow process (third term).

This means that we do not have to keep track of the fast timescale when considering the second term, nor the fast and intermediate timescale when considering the third term. We approximate

$$\psi_{N\mathrm{nt}} \approx \underbrace{p_0^N\left(\frac{1}{1+\tau_0 s}\right)^N}_{\psi^0_{N\mathrm{nt}}} + \underbrace{\sum_{n_0=0}^{N-1}\left(N^{n_0,N-n_0}\right)p_0^{n_0}p_1^{N-n_0}\left(\tfrac{1}{1+\tau_1 s}\right)^{N-n_0}}_{\psi^1_{N\mathrm{nt}}}$$

$$+\underbrace{\sum_{n_0=0}^{N-1}\sum_{n_1=0}^{N-n_0-1}\left(N^{n_0,n_1,N-n_0-n_1}\right)p_0^{n_0}p_1^{n_1}p_2^{N-n_0-n_1}\left(\tfrac{1}{1+\tau_2 s}\right)^{N-n_0-n_1}}_{\psi^2_{N\mathrm{nt}}}$$

This approximation breaks down in the short time limit $(t \ll \tau_0)$ as the two last terms no not account for that nucleotide addition is needed to traverse the window even when there are pauses in the dwell-time window (see below).

The experimental dwell time distributions are all well fit with exponential shoulders for Pause 1 and Pause 2 using a maximum-likelihood approach. This corresponds to exchanging the second $(\psi^1_{N\mathrm{nt}})$ and third $(\psi^2_{N\mathrm{nt}})$ term in the above with simple exponential processes that capture the probability and the average time of each pause. The probability and average time can respectively be written as

$$q_i = \int_0^\infty \mathrm{d}t P_{N\mathrm{nt}}(t) = \psi^i_{N\mathrm{nt}}(0),\, T_i = \frac{\int_0^\infty \mathrm{d}t\, t\, P_{1\mathrm{nt}}(t)}{\int_0^\infty \mathrm{d}t\, P_{1\mathrm{nt}}(t)} = -\partial_s \ln \psi^i_{N\mathrm{nt}}(0), i=1,2.$$

The estimated weights and timescales over the Nnt can be calculated as

$$q_1 = (p_0+p_1)^N - p_0^N,\, T_1 = \frac{Np_1(p_0+p_1)^{N-1}}{(p_0+p_1)^N - p_0^N}\tau_1$$

$$q_2 = 1-(p_0+p_1)^N,\, T_2 = \frac{Np_2}{1-(p_0+p_1)^N}\tau_2.$$

Based on this, the fitting function would be

$$P_{N\mathrm{nt}}(t) \approx \frac{1-q_1-q_2}{\tau_0(N-1)!}\left({}^{t}\!/_{\tau_0}\right)^{N-1} e^{-t/\tau_0} + Q(t)\left(\frac{q_1}{T_1}e^{-t/T_1} + \frac{q_2}{T_2}e^{-t/T_2}\right)$$

where we have introduced the regularizing function

$$Q(t) = \frac{(t/(\tau_0 N_{\mathrm{dw}}))^{N_{\mathrm{dw}}-1}}{1+(t/(\tau_0 N_{\mathrm{dw}}))^{N_{\mathrm{dw}}-1}}$$

to account for the fact that even when there is a pause, the short timescales are still dominated by a rapid succession on nucleotide addition steps. The fit results dependence on these cut-offs is negligible as long as they are introduced in regions where the corresponding term is sub-dominant. Here the cut is placed under the center of the elongation peak, guaranteeing that it is placed where pausing is sub-dominant.

We can now translate between the probabilities and timescales over a Nnt window to the 1nt window through

$$p_0 = (1-(q_1+q_2))^{1/N},\, p_1 = (1-q_2)^{1/N} - (1-(q_1+q_2))^{1/N},\, p_2 = 1-(1-q_2)^{1/N}$$

$$\tau_1 = \frac{(p_0+p_1)^N - p_0^N}{Np_1(p_0+p_1)^{N-1}}T_1,\, \tau_2 = \frac{1-(p_0+p_1)^N}{Np_2}T_2$$

Here, we explicitly also allow for a third, very improbable (never entered twice in a dwell-time window), and fat-tailed pause

$$P_{N\mathrm{nt}}(t) \approx \frac{1-q_1-q_2}{\tau_0(N-1)!}\left({}^{t}\!/_{\tau_0}\right)^{N-1} e^{-t/\tau_0} + Q(t)\left(\frac{q_1}{T_1}e^{-t/T_1} + \frac{q_2}{T_2}e^{-t/T_2} + \frac{a_{\mathrm{bt}}}{2(1+t/1\mathrm{s})^{3/2}}\right). \tag{S1}$$

The additional third pausing term captures the asymptotic power-law decay (amplitude a_{bt}) of the probability of dwell-times dominated by a backtrack. The backtracked asymptotic term needs to be further regularized for times shorter than the diffusive backtrack step. We have introduced a regularization at 1 s, but the precise timescale does not matter, as long as it is set within the region where the exponential pauses dominate over the backtrack.

Maximum likelihood estimation fitting routine

The above stochastic-pausing model was fit to the dwell time distributions using a custom Python 3.7 routine (https://doi.org/10.5281/zenodo.5179600). The dwell-time distribution is fit to the experimentally collected dwell-times $\{t_i\}_i$ by minimizing the likelihood function (Cowan, 1998)

$$L = -\sum_i \ln P_{N\mathrm{nt}}(t_i) \tag{S2}$$

with respect to timescales and probabilistic weights. We calculated the statistical error on the parameters by applying the MLE fitting procedure on 100 bootstraps of the original dataset (Press et al., 1992), and reported the standard deviation for each fitting parameters.

Modeling of Pause 1 propensity under force and concentration sweeps

In the main text we show that Pause 1 and Pause 2 are described by Michaelis-Menten kinetics (Figure 4D), and their probabilities increase with force (Figure 3F). We therefore hypothesize that they originate from the pre-translocated state of the polymerase. Pause 1 (p1) dominates in probability, and though we do not know how Pause 2 or other the long-lived pauses are connected to Pause 1, we can ignore their effect on the analysis of Pause 1 and elongation.

The reaction kinetic scheme

We consider here the working-model reaction scheme illustrated in Figure S6 (marked with gray dash line). The nucleotide addition reaction starts with the polymerase fluctuating in position between the pre- and post-translocated states (rates $k_{\mathrm{post}}(F)$ and $k_{\mathrm{pre}}(F)$), and is stabilized forward upon nucleotide binding. Translocation is not assumed equilibrated, but binding is assumed equilibrated with dissociation constant K_{D}. Once bound, the nucleotide is irreversibly incorporated at rate $k_{\mathrm{irr},1}$.

We assume that all force-dependent rates can be written in terms of a shift of the transition state according to

$$k_{\mathrm{X}}(F) = k_{\mathrm{X}}(F_0)e^{-\delta_X(F-F_0)/k_{\mathrm{B}}T}$$

In the above, δ_X is the distance to the transition state in the direction of a growing transcript, F is the tether tension and F_0 is an (arbitrary) reference tension.

By summing probabilities over paths, the first passage-time distribution of completing the irreversible step when starting from the pre-translocated state can be written in Laplace space as

$$\Psi_{\mathrm{irr},1}(s) = \sum_{n=0}^{\infty}\left(\frac{k_{\mathrm{post}}(F)}{s+k_{\mathrm{in},1}+k_{\mathrm{post}}(F)}\,\frac{K_{\mathrm{pre}}(F,[\mathrm{NTP}])}{s+K_{\mathrm{pre}}(F,[\mathrm{NTP}])+K_{\mathrm{irr},1}([\mathrm{NTP}])}\right)^n$$

$$\times\frac{k_{\mathrm{post}}(F)}{s+k_{\mathrm{in},1}+k_{\mathrm{post}}(F)}\,\frac{K_{\mathrm{irr},1}([\mathrm{NTP}])}{s+K_{\mathrm{pre}}(F,[\mathrm{NTP}])+K_{\mathrm{irr},1}([\mathrm{NTP}])}$$

$$=\frac{K_{\mathrm{irr},1}([\mathrm{NTP}])k_{\mathrm{post}}(F)}{(s+k_{\mathrm{post}}(F)+k_{\mathrm{in},1})(s+K_{\mathrm{pre}}(F,\mathrm{NTP}])+K_{\mathrm{irr},1}([\mathrm{NTP}]))-k_{\mathrm{post}}(F)K_{\mathrm{pre}}(F,[\mathrm{NTP}])}.$$

In the equation above, we have defined the effective rates out of the binding equilibrated post-translocated state as

$$K_{\mathrm{pre}}(F,[\mathrm{NTP}]) = k_{\mathrm{pre}}(F)\frac{K_{\mathrm{D}}}{[\mathrm{NTP}]+K_{\mathrm{D}}},\quad K_{\mathrm{irr},1}([\mathrm{NTP}]) = k_{\mathrm{irr},1}\frac{[\mathrm{NTP}]}{[\mathrm{NTP}]+K_{\mathrm{D}}}.$$

For notational convenience we define the dimensionless rates

$$\gamma_{\mathrm{in},1} = \frac{k_{\mathrm{in},1}}{k_{\mathrm{in},1}+k_{\mathrm{post}}(F)},\quad \gamma_{\mathrm{pre}} = \frac{k_{\mathrm{pre}}(F)}{k_{\mathrm{in},1}+k_{\mathrm{post}}(F)},\quad \gamma_{\mathrm{irr},1} = \frac{k_{\mathrm{irr},1}}{k_{\mathrm{in},1}+k_{\mathrm{post}}(F)}.$$

The observables

From the first-passage time distribution we can deduce the probability to lock in the next base before entering the pause as

$$P_{\mathrm{irr},1} = \Psi_{\mathrm{irr},1}(0) = (1-\gamma_{\mathrm{in},1})\frac{[\mathrm{NTP}]}{[\mathrm{NTP}]+K_{\mathrm{P1}}},$$

where we have introduced the modified dissociation constant

$$K_{\mathrm{P1}} = K_{\mathrm{D}}\frac{\gamma_{\mathrm{pre}}\gamma_{\mathrm{in},1}}{\gamma_{\mathrm{irr},1}}$$

The inverse average time it takes to move from the pre-translocated state through the irreversible step can be written as

$$k_{\mathrm{irr},1} = \left(- \frac{\partial \ln \Psi_{\mathrm{irr},1}(s)}{\partial s}\bigg|_{s=0} \right)^{-1} = V^{\max}_{\mathrm{irr},1} \frac{[\mathrm{NTP}] + K_{\mathrm{P1}}}{[\mathrm{NTP}] + K_{\mathrm{D}}^{\mathrm{app}}} \tag{S3}$$

where

$$V^{\max}_{\mathrm{irr},1} = \frac{\gamma_{\mathrm{irr},1}}{1+\gamma_{\mathrm{irr},1}}, \; K_{\mathrm{D}}^{\mathrm{app}} = K_{\mathrm{D}} \frac{1+\gamma_{\mathrm{pre}}}{1+\gamma_{\mathrm{irr},1}}.$$

Catalytic rate

We see no appreciable concentration dependence in the total catalytic rate at $F = 25$ pN (nucleotide addition rate, Figure 3D), which is only consistent with Equation S3) if $[\mathrm{NTP}] \ll K_{\mathrm{P1}}, K_{\mathrm{D}}^{\mathrm{app}}$. This implies that the probability to enter Pause 1,

$$P_{\mathrm{P1}} = 1 - P_{\mathrm{irr},1} = \frac{\gamma_{\mathrm{in},1}[\mathrm{NTP}] + K_{\mathrm{P1}}}{[\mathrm{NTP}] + K_{\mathrm{P1}}}, \tag{S4}$$

is constant and close to unity, contradicting both the concentration dependence and quantitative values seen in Figure 3D. Thus, the irreversible step of our working-model reaction scheme cannot be rate-limiting, we conclude there is another irreversible, concentration independent, and rate-limiting step. This second, irreversible step (rate $k_{\mathrm{irr},2}$), is associated with a large conformational change of the polymerase-primer-template complex (Figure S6, red dashed line).

Pause 1 probability

Next, we turned to the pause probability P_{P1} (Equation S4), over the concentration sweeps at tether tension $F = 25$ pN (Figure 3E, using all but lowest concentration point) and 35 pN (Figure S4F, using all but two lowest concertation points), and one force sweep at $[\mathrm{NTP}] = 0.5$ mM (Figure 3F, using all but three highest force points). We minimized the error-weighted square deviation under the assumption that the distance to the transition state in the pre- to post-translocation $\delta > 0$, and that the distance from the other direction is $a_{\Delta\mathrm{bp}} - \delta > 0$. From this we estimated (Figures 6B–6D)

$$\delta = 0.0 \text{ nm}, a_{\Delta\mathrm{bp}} = 0.23 \text{ nm}, \; \gamma_{\mathrm{in},1} = 0.046, \; K_{\mathrm{P1}} = 9.0 \text{ µM}$$

at a tether tension of 25 pN. Allowing δ to be negative results in a very small negative value. In our data, we thus see no evidence that the pre- to post- translocation step is force dependent ($\delta \approx 0$ nm), while the post- to pre-translocation step acts over a distance $a_{\Delta\mathrm{bp}} = \delta x_{\mathrm{ss}} - \delta x_{\mathrm{ds}}$ corresponding well to the difference in average length of a single- (δx_{ss}) and double-stranded (δx_{ds}) base in the present force range (Dulin et al., 2015c). We further see that translocation is not fast compared to NTP binding and pausing, as there is a non-zero probability ($\gamma_{\mathrm{in},1} > 0$) to enter a pause even at saturating NTP concentrations; consequently, translocation is not equilibrated.

The effect of a dsRNA structure downstream the polymerase

In the presence of a hairpin, the energy of the post-translocated state is destabilized by the tension dependent melting energy

$$\varepsilon(F) = \varepsilon_0 - 2a_{\mathrm{ss}}F$$

of the outer hairpin base pair. Here ε_0 is the relaxed melting energy and a_{ss} is the typical extension of a single-stranded (ss) base in our force range. Assuming that the distance to the transition state from pre- to post-translocated state is negligible also in the presence of a hairpin, only the rate from post- to pre-translocated state changes

$$k_{\mathrm{pre}}^{\mathrm{hp}}(F) = k_{\mathrm{pre}}^{\mathrm{no-hp}}(F)\exp(\varepsilon(F)/k_{\mathrm{B}}T) \tag{S5}$$

in the presence of hairpin. We have already fitted out $a_{\Delta\mathrm{bp}}$, the difference between a ss base and a ds base-pair in the present force range. dsRNA is $\sim$ 95% stretched in our force range, with a crystallographic length of 0.28 nm/bp (Dock-Bregeon et al., 1989), giving $\delta x_{\mathrm{ds}} \approx 27$ nm, and thus $a_{\mathrm{ss}} = 0.27 \text{ nm} + a_{\Delta\mathrm{bp}} = 0.50$ nm; in accord with what is expected from the literature (Vieregg et al., 2007). Using this value, injecting Equation S5 into Equation S4, we fitted Pause 1 probability in the presence of a hairpin (Figure 6E), and extract a zero-tension melting energy $\varepsilon_0 = 18\, k_{\mathrm{B}}T$, i.e., a melting force of $\sim$ 18 pN, which agrees well with the hairpin opening fully at $\sim$ 22 pN (Figure S1B). We can also extract the probability to enter Pause 1 at zero force when the polymerase replicates through a dsRNA template, i.e., $P_{\mathrm{P1,hp}}(F = 0) \approx 0.31$; This should be compared to $P_{\mathrm{P1,no-hp}}(F = 0) \approx 0.049$ at the same conditions but without a hairpin.

Structure modeling

Comparison of the structural model of elongating SARS-CoV-2 polymerase (PDB: 6yyt) (Hillen et al., 2020) was overlapped with the ssRNA template of the poliovirus RdRp (PDB: 3ol6) (Gong and Peersen, 2010) using UCSF Chimera 1.15 (Figures 7A and S7). An axis representing the force was modeled through the center of the dsRNA product strand to represent the force applied in the magnetic tweezers setup. An additional 6 nucleotide ssRNA model was aligned to basepair position n-1. It was then tilted toward the force axis until a clash with the amino acids of the RdRp was observed. Angles were measured between the backbone of RNA template positions and the force axe.

Structures for T7 DNA polymerase (PDB: 1t7p) as published in Doublié et al. (1998) and Φ29 DNA polymerase (PDB: 2pyl) as published in Berman et al. (2007) were modeled accordingly (Figures 7B and 7C).

QUANTIFICATION AND STATISTICAL ANALYSIS

This information is provided in the figure legends, the STAR Methods, and the Table S1.