



Allosteric Activation of SARS-CoV-2 RNA-Dependent RNA Polymerase by Remdesivir Triphosphate and Other Phosphorylated Nucleotides

Bing Wang,^{a,b} Vladimir Svetlov,^c Yuri I. Wolf,^d Eugene V. Koonin,^d Evgeny Nudler,^{c,e} Irina Artsimovitch^{a,b}

^aDepartment of Microbiology, The Ohio State University, Columbus, Ohio, USA

^bThe Center for RNA Biology, The Ohio State University, Columbus, Ohio, USA

^cDepartment of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York, USA

^dNational Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

^eHoward Hughes Medical Institute, New York University School of Medicine, New York, New York, USA

ABSTRACT The catalytic subunit of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA-dependent RNA polymerase (RdRp) Nsp12 has a unique nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain that transfers nucleoside monophosphates to the Nsp9 protein and the nascent RNA. The NiRAN and RdRp modules form a dynamic interface distant from their catalytic sites, and both activities are essential for viral replication. We report that codon-optimized (for the pause-free translation in bacterial cells) Nsp12 exists in an inactive state in which NiRAN-RdRp interactions are broken, whereas translation by slow ribosomes and incubation with accessory Nsp7/8 subunits or nucleoside triphosphates (NTPs) partially rescue RdRp activity. Our data show that adenosine and remdesivir triphosphates promote the synthesis of A-less RNAs, as does ppGpp, while amino acid substitutions at the NiRAN-RdRp interface augment activation, suggesting that ligand binding to the NiRAN catalytic site modulates RdRp activity. The existence of allosterically linked nucleotidyl transferase sites that utilize the same substrates has important implications for understanding the mechanism of SARS-CoV-2 replication and the design of its inhibitors.

IMPORTANCE *In vitro* interrogations of the central replicative complex of SARS-CoV-2, RNA-dependent RNA polymerase (RdRp), by structural, biochemical, and biophysical methods yielded an unprecedented windfall of information that, in turn, instructs drug development and administration, genomic surveillance, and other aspects of the evolving pandemic response. They also illuminated the vast disparity in the methods used to produce RdRp for experimental work and the hidden impact that this has on enzyme activity and research outcomes. In this report, we elucidate the positive and negative effects of codon optimization on the activity and folding of the recombinant RdRp and detail the design of a highly sensitive *in vitro* assay of RdRp-dependent RNA synthesis. Using this assay, we demonstrate that RdRp is allosterically activated by nontemplating phosphorylated nucleotides, including naturally occurring alarmone ppGpp and synthetic remdesivir triphosphate.

KEYWORDS allostery, NiRAN domain, RNA synthesis, remdesivir, ribosome pausing, SARS-CoV-2 RdRp, ppGpp, synonymous codons

Respiratory RNA viruses pose a major threat to humankind and have proved to be extremely refractory to modern disease control measures, which have limited the spread of water-, food-, and blood-borne epidemics, such as cholera, plague, and AIDS (1). In the 21st century, the H1N1pdm09, severe acute respiratory syndrome (SARS),

Citation Wang B, Svetlov V, Wolf YI, Koonin EV, Nudler E, Artsimovitch I. 2021. Allosteric activation of SARS-CoV-2 RNA-dependent RNA polymerase by remdesivir triphosphate and other phosphorylated nucleotides. *mBio* 12: e01423-21. <https://doi.org/10.1128/mBio.01423-21>.

Editor Richard Gerald Brennan, Duke University School of Medicine

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Irina Artsimovitch, artsimovitch.1@osu.edu.

This article is a direct contribution from Irina Artsimovitch, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Craig Cameron, University of North Carolina, and Katsuhiko Murakami, Pennsylvania State University.

Received 14 May 2021

Accepted 24 May 2021

Published 22 June 2021

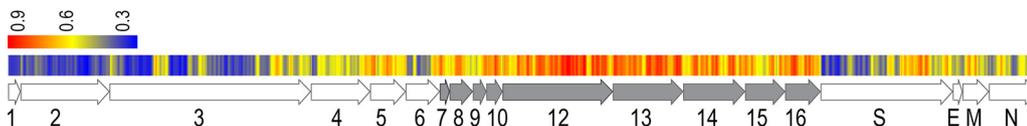


FIG 1 Conservation of amino acid residues in genomes of alpha-, beta-, gamma-, and deltacoronavirus genera; only those proteins that are present in all *Coronaviridae* are shown (see Data Set S1 in the supplemental material). The Nsps are indicated by numbers; Nsp7 to -16 (shown in gray), which comprise the replication-transcription complex (RTC), are more conserved than structural (E, M, N, S) proteins and other Nsps.

Middle East respiratory syndrome (MERS), and coronavirus (CoV) disease 2019 (COVID-19) epidemics have been caused by respiratory RNA viruses, the last three by betacoronaviruses (betaCoVs), which belong to the family *Coronaviridae* in the order *Nidovirales* (2). The ongoing COVID-19 pandemic led to a dramatic loss of human life and devastating economic and social disruptions around the world. The zoonotic origin of its causative agent, the SARS-CoV-2 clade of *Severe acute respiratory syndrome-related coronavirus* species, the rapid rise of mutant strains within the infected human population, and numerous instances of retransmission to zoonotic hosts speak to its resilience as a persistent human pathogen and the likelihood of the emergence of new betaCoV variants with pandemic potential (3–6). Adequate pandemic response measures, from the development of effective antivirals to genomic surveillance, require a detailed understanding of SARS-CoV-2's molecular and structural biology. However, CoVs have been studied less thoroughly than other viral pathogens, in part owing to their extraordinarily large genome size (by far the largest among the known RNA viruses) and complex biology (7).

Upon infecting human cells, the CoV plus-strand RNA genome is translated to produce a long polyprotein that is cleaved into several nonstructural proteins (Nsps), which are required for viral replication and gene expression by CoV-encoded protease (8). Among these, SARS-CoV-2 Nsp12 plays a central role as a catalytic subunit of RNA-dependent RNA polymerase (RdRp). The RdRp is the only protein that is universally conserved among RNA viruses (9) and therefore is an attractive target for broad-spectrum antivirals. Many nucleoside analogs identified as RNA synthesis inhibitors in other viruses have been actively pursued for retargeting against SARS-CoV-2 (10).

The transcription machinery of CoV is unique among RNA viruses in its complexity; the transcribing RdRp associates with the replicative helicase Nsp13, proofreading exonuclease Nsp14/10, and several other viral proteins in a large membrane-bound replication-transcription complex (RTC) (11). The RTC components are highly conserved among CoVs (Fig. 1). Furthermore, unlike many well-studied single-subunit viral RdRps (9), a minimally active SARS-CoV-2 RdRp consists of Nsp12 and three accessory subunits: Nsp7 and two copies of Nsp8 (7-8₂-12) (12–14) (Fig. 2A).

Nsp12 is a large (932-residue) multidomain protein. In addition to containing the RdRp module, composed of finger, palm, and thumb domains, Nsp12 contains a large *nidovirus* *RdRp-associated nucleotidyl transferase* (NiRAN) domain, which is connected to the finger domain through an interface domain (Fig. 2A). The NiRAN domain is unique to *Nidovirales* and has been suggested to perform a range of activities, from RNA capping to protein-primed initiation of RNA synthesis (15). A recent report identified the accessory RNA-binding protein Nsp9 as the physiological target of NiRAN NMPylase and showed that this activity is critical for viral replication (16). Consistent with these findings, the NiRAN domain active site was observed to bind Nsp9 in a single-particle cryogenic electron microscopy (cryoEM) study (17), apparently in a catalytically inactive arrangement. Thus, Nsp12 is a bifunctional enzyme with two active sites, one of which transfers an NMP moiety to the 3' end of the nascent RNA (active site 1 [AS1] in the RdRp domain) and the other one, to the N terminus of Nsp9 (AS2, in the NiRAN domain). AS1 and AS2 utilize standard nucleoside triphosphates (NTPs) as the substrates, but can also accommodate as ligands a variety of nucleotide derivatives. In particular, AS1 readily incorporates remdesivir (18) and favipiravir (19) monophosphates into

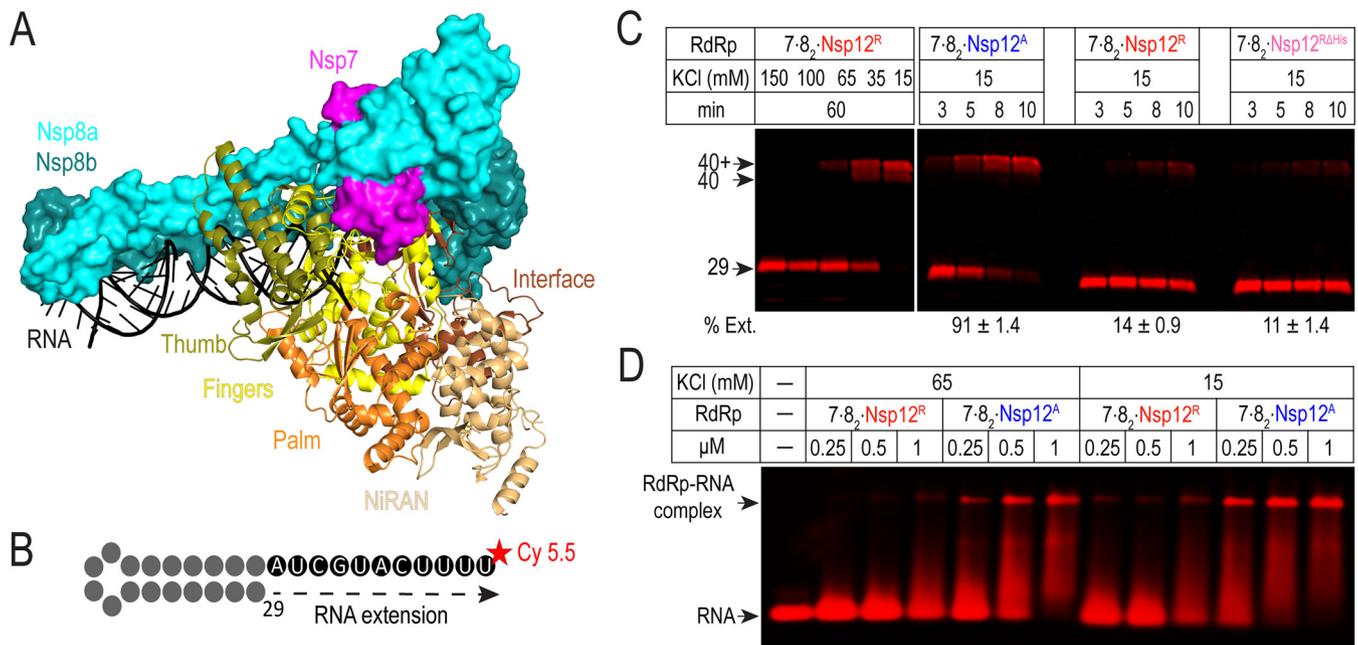


FIG 2 Activities of Nsp12^A and Nsp12^R. (A) Transcription of RdRp of SARS-CoV-2 Wuhan-Hu-1. Nsp7 and Nsp8 are shown at the surface, and Nsp12 is shown as a cartoon, with individual domains highlighted (PDB accession no. 6YYT). (B) The 29-nt RNA hairpin scaffold is extended by RdRp to produce a 40-nt product; additional extension is thought to be mediated by Nsp8 after the completion of RNA synthesis (40). Cy 5.5, cyanine 5.5. (C) RNA extension by RdRp at 37°C under the indicated conditions; 15 mM KCl is a permissive condition. Removal of the His tag (Δ His) does not increase Nsp12^R activity, but Nsp12^A expressed from an mRNA that retains rare codons is more active. Fractions of the extended RNA (% Ext.) at 10 min are shown (means \pm SEM; $n=3$). (D) Interactions with the RNA hairpin scaffold analyzed by electrophoretic mobility shift assays. RdRps at the indicated concentrations were incubated with 100 nM RNA.

RNA, and structural evidence suggests that AS2 is similarly promiscuous (14, 17, 20). Thus, the effects of NTPs and nucleoside analogs on both catalytic activities must be taken into account when interpreting experimental data and evaluating the antiviral potential of lead molecules.

SARS-CoV-2 RdRp contains intrinsically disordered regions (IDRs) that undergo large context-dependent conformational changes, e.g., upon interaction with the product RNA (13) or upon binding to ligands in the AS2 of NiRAN (14, 17, 20). These inherent dynamic properties suggest that RdRp's activity can be modulated, positively or negatively, by factors that control the folding of the enzyme. Here, we report that the RdRp used in several structural and functional studies is largely inactive because synonymous codon substitutions in Nsp12 designed to maximize its expression instead trigger its misfolding. We identify a region containing a cluster of rare codons that plays a critical role in the proper folding of Nsp12 and show that Nsp12 expression in a bacterial strain with slow ribosomes and/or incubation with the accessory Nsp7/8 subunits increases RdRp activity. We further show that nucleoside analogs that cannot be incorporated into RNA can nonetheless activate RNA chain extension, presumably through binding to AS2. Our findings have immediate implications for functional studies and identification of novel inhibitors of SARS-CoV-2 RdRp and highlight the need for improved mRNA-recoding algorithms during the rational design of other biotechnologically and medically important expression systems.

RESULTS

Nsp12s expressed from different coding sequences differ in activity and conformation. A rapidly growing collection of cryoEM structures of RdRp bound to different partners provides an excellent framework for understanding the mechanism of RNA synthesis and for the identification of novel RdRp inhibitors (12–14, 17, 21, 22). As is the case with other systems, structural models require validation by functional studies that critically depend on the availability of robust expression systems and

highly active RdRp preparations. Given that the structures obtained for RdRp produced in *Escherichia coli* (12, 14) and in insect cells (13) are closely similar, we used the *E. coli* expression platform (see Fig. S1 in the supplemental material) to initiate mechanistic studies of SARS-CoV-2 RdRp. For the sake of expediency, we used an Nsp12 expression vector described in reference 12 (we refer to Nsp12 produced from this vector as Nsp12^R, where R indicates the laboratory where this plasmid was constructed) and Nsp7- and Nsp8-producing vectors that we constructed for this study. Nsp12^R contains a noncleavable C-terminal His₁₀ tag, is soluble when produced in *E. coli*, and is easily purified under “native” (nondenaturing) conditions.

We found that the 7-8₂-12^R enzyme exhibited negligible activity on a number of different templates, including the optimal hairpin scaffold (Fig. 2B) used by Hillen et al. (13), which could be extended only at a very low concentration of salt. An extensive experimental survey of different combinations of purification schemes, RNA scaffolds, and reaction conditions failed to identify conditions that would support efficient primer extension, and the removal of the His tag, which has been proposed to interfere with RdRp activity (23), did not increase activity under permissive (15 mM KCl) conditions (Fig. 2C). In their follow-up study, Wang et al. reported similar results (24), prompting us to conclude that further attempts to boost the activity of the 7-8₂-12^R enzyme produced under these conditions would be futile.

Our survey of published reports failed to reveal an obvious reason for the observed low activity of the 7-8₂-12^R enzyme. Under similar reaction conditions, some RdRps were able to completely extend the RNA primer in minutes (13, 19, 23, 25), whereas others failed to do so in an hour (21, 24), regardless of the expression system. Idiosyncratic but reproducible variations in activity can arise from recombinant protein misfolding; indeed, coexpression of Nsp12 with cellular chaperones has been shown to enhance its activity (23, 26). A likely source of this variability may lie in the coding mRNA itself; whereas all recombinant Nsp12s have the same amino acid sequence (ignoring the tags), their coding sequences (CDSs) have been altered to match the codon usage of their respective hosts to maximize protein expression. Codon optimization is routinely used for protein expression in heterologous systems (27), yet protein function can be compromised even by a single synonymous codon substitution (28, 29). Furthermore, protein expression in a BL21 RIL strain, which alleviates codon imbalance by supplying a subset of rare tRNAs and is thus commonly used to express heterologous proteins in *E. coli*, can hinder proper folding (30). The abrogation of ribosome pausing at rare codons is thought to uncouple nascent peptide synthesis from its folding, giving rise to misfolded proteins (29, 31, 32).

Although robust viral gene expression may promote host takeover, SARS-CoV-2 mRNAs, including *nsp12*, are not efficiently translated in human cells (33). The viral *nsp12* mRNA contains clusters of rare codons (in humans) (Fig. S2), yet the resulting enzyme is active and able to sustain efficient infection. This suggests that pause-prone translation may facilitate the proper folding of Nsp12. In contrast, the *nsp12^R* codon usage matches that of highly expressed *E. coli* genes, raising the possibility that the *nsp12^R* CDS has been optimized for the maximum expression of soluble protein in a bacterial host but not for enzymatic activity and/or acquisition of a native structure. There is no *a priori* reason to believe that an overexpressed soluble protein retains all of its activity, and the abundance of Nsp12 in the heterologous host may be made possible by its diminished NTP binding and/or condensation activity, defects in RNA binding, or other functionalities. To evaluate this possibility, we designed an *nsp12^A* variant (where A indicates that it was constructed by I. Artsimovitch) that contains more rare codons (Fig. S2), including in the regions that bear rare codons in the viral mRNA. Interestingly, *nsp12^T* expression vector (constructed in T. Tuschl's lab) also gives rise to active RdRp (14); in this study, the viral *nsp12* mRNA was reverse transcribed and expressed in the *E. coli* BL21 RIL strain, which contains extra copies of the *argU*, *ileY*, and *leuW* rare tRNA genes. While comparable codon frequency measurements are not available for the RIL strain, it does not carry all rare tRNAs required for the efficient

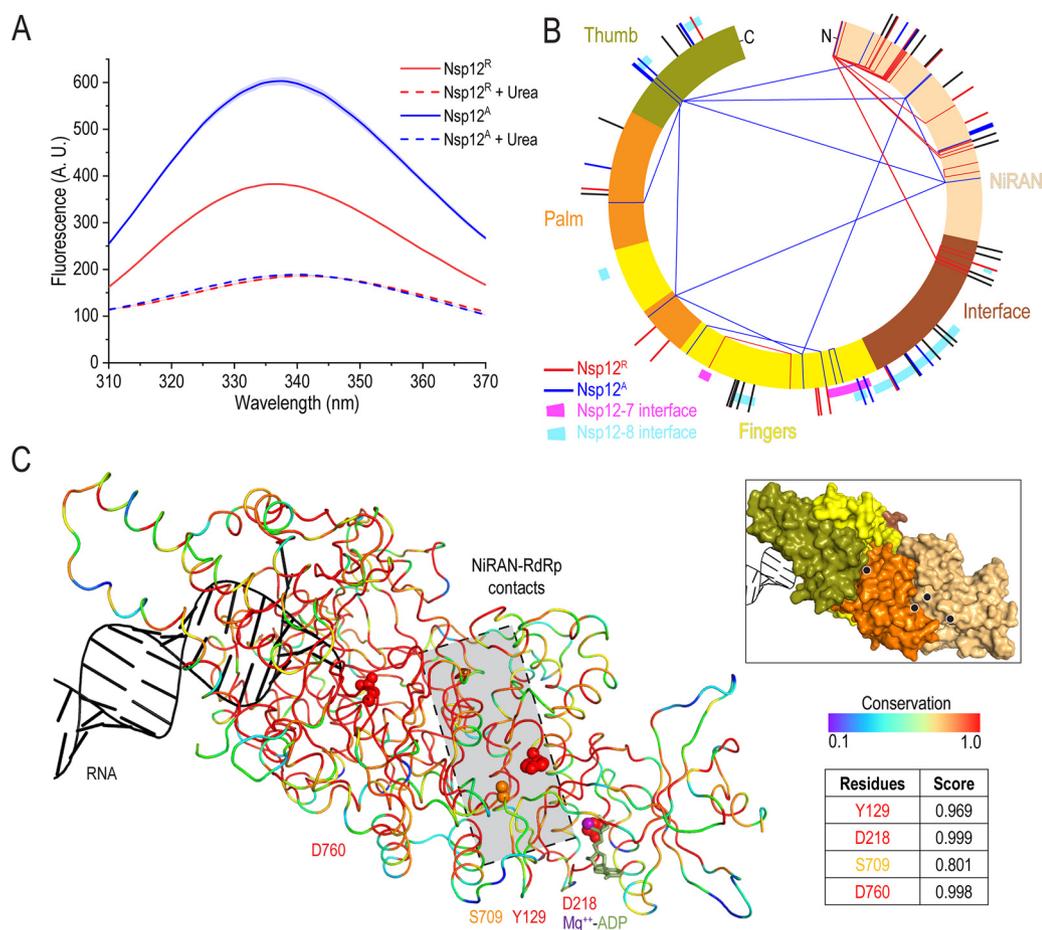


FIG 3 Differences between Nsp12^A and Nsp12^R. (A) Intrinsic tryptophan fluorescence of Nsp12 proteins. The spectra of denatured proteins confirm that their concentrations are identical. The means and SEM of triplicate measurements are shown as lines and shaded bands, respectively, in this and Fig. S4 and S6B. A.U., arbitrary units. (B) Map of the EDC modifications. Lines show the positions of monolinks (outside) and cross-links (inside) mapped onto the Nsp12 schematic, with the domains colored as in panel A. Colors indicate differences in reactivity; residues in red were reactive only in Nsp12^R, those in blue were reactive in Nsp12^A, and those in black were reactive in both proteins. Only high-confidence monolinks ($<10^{-5}$) and cross-links ($<10^{-3}$) are shown (see Data Set S2). (C) Conservation of the NiRAN-RdRp interaction surfaces mapped on the transcription complex structure (PDB accession no. 6XEZ). Amino acid residues are colored according to their conservation. Key residues in AS1 (D760), in AS2 (D218), and at the NiRAN-palm interface (Y129 and S709) are shown as spheres; ADP bound to AS2 is shown as sticks and the Mg²⁺ ion as a purple sphere.

translation of the viral *nsp12* mRNA, suggesting that *nsp12*^T codon usage is suboptimal. We found that RdRp assembled with Nsp12^A had a much higher activity on the hairpin scaffold (Fig. 2C). We also noted that Nsp12^A and Nsp12^T copurified with nucleic acids (Fig. S3A); subsequent gel shift assays revealed that 7·8₂·12^A readily bound the RNA hairpin, whereas 7·8₂·12^R did not (Fig. 2D). We found that the 7·8₂·12^T enzyme behaved similarly to 7·8₂·12^A (Fig. S3B), but since the *nsp12*^T expression vector lacks restriction sites required for protein engineering, we used *nsp12*^A in all subsequent experiments.

To test whether Nsp12^R was misfolded, we used several approaches. First, we assessed the Nsp12 thermal stability using differential scanning fluorimetry (34). We recorded melting temperatures (T_m) of 41.3°C for Nsp12^R and 47.3°C for Nsp12^A (Fig. S4A); for another *E. coli*-expressed Nsp12, a T_m of 43.6°C was reported (35). Second, we compared the intrinsic fluorescence spectra of Nsp12, which contains nine tryptophan residues that are expected to be sensitive to the microenvironment (36). Nsp12^A and Nsp12^R exhibited similar emission peaks, but the Nsp12^A intensity was 2-fold higher (Fig. 3A), suggesting that at least one Trp was more buried; the derivative spectra (Fig. S4B) did not reveal any additional differences. These results show that

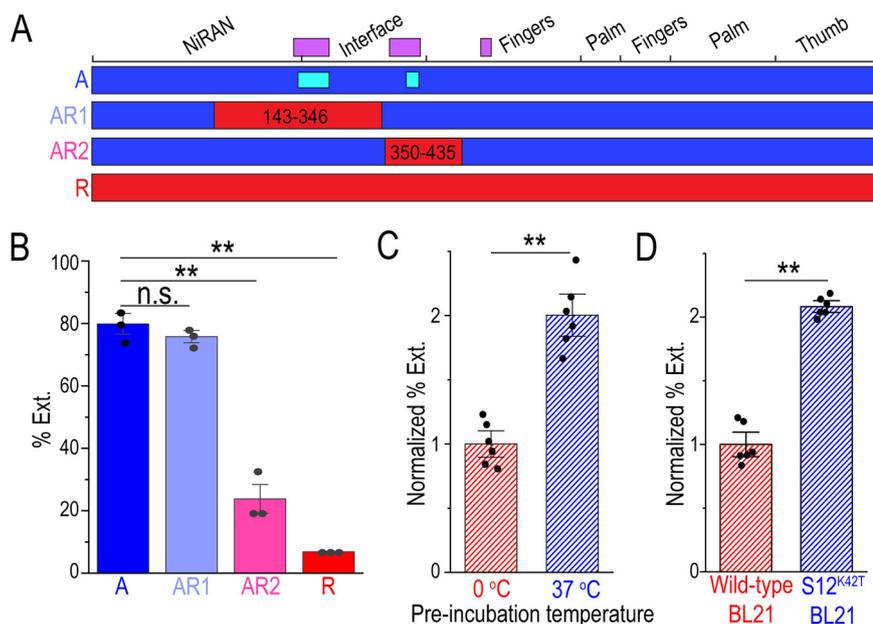


FIG 4 Determinants of Nsp12 activity. (A) The translational context around residue 400 is critical for the correct folding of Nsp12. The SARS-CoV-2 genomic *nsp12* RNA (with domain boundaries shown on the top) contains clusters of rare codons (purple bars); only the Nsp12^A CDS has rare codons (cyan bars) at the corresponding positions. (B) A chimeric Nsp12^{A-AR2} protein is defective in RNA synthesis. (C) Reactivation of Nsp12^R via 37°C preincubation with the accessory Nsp7 and Nsp8 subunits to form the RdRp holoenzyme. (D) Translation by slow ribosomes yields a more active Nsp12. RNA extension is shown as means \pm SEM, and the *P* value was calculated by an unpaired two-tailed *t* test. n.s., not significant; **, *P* < 0.01.

Nsp12^A and Nsp12^R are structurally distinct, but we cannot identify the regions of altered structure.

We next used a carboxyl- and amine-reactive reagent, EDC [1-ethyl-3-(3-dimethylaminopropyl)carbodiimide] to map solvent-accessible (surface) residues and intraprotein cross-links by mass spectrometry. We observed substantial differences in accessibility of several regions centered at residues 150 (NiRAN domain), 415 (fingers), 600 (palm), and 850 (thumb) and in cross-linking, particularly of the NiRAN domain (Fig. 3B).

Attenuated translation and accessory subunits promote an active Nsp12 conformation. Although an overall excess of underrepresented codons can slow down translation, in many cases, the ribosome has to pause at one or more specific rare codons to ensure proper protein folding at key junctures (37, 38). The differences between the codon frequencies between the 2.8-kb mRNAs encoding Nsp12^A and Nsp12^R are extensive (Fig. S2). The produced proteins also differ in their N and C termini (Fig. S1), but based on available structural data, an extra N-terminal glycine would not be expected to account for the observed dramatic differences in EDC reactivity (Fig. 3B and Data Set S2). Comparative analysis identified two regions that contained rare codon clusters in the native SARS-CoV-2 RNA and in Nsp12^A mRNAs, but not in Nsp12^R (Fig. 3A and Fig. S5). We constructed chimeric proteins in which these Nsp12^A segments were replaced with corresponding segments from Nsp12^R, generating proteins with identical amino acid sequences (Fig. 4A). We found that whereas swapping of codons (143 to 346) between the mRNAs producing active and inactive Nsp12 variants did not alter the RdRp activity, a chimeric protein containing codons 350 to 435 derived from the Nsp12^R CDS was defective (Fig. 4B).

Together with the EDC modification patterns (Fig. 3B), this result suggests that controlled translation of the codon 350–435 region is important for Nsp12 folding and that changes in contacts with Nsp7 (Fig. S6), which are critical for RdRp activity (13), may be partially responsible for the low activity of Nsp12^R. During expression of the viral genome, Nsp7, Nsp8, and Nsp12 are cotranslated with other Nsps as giant

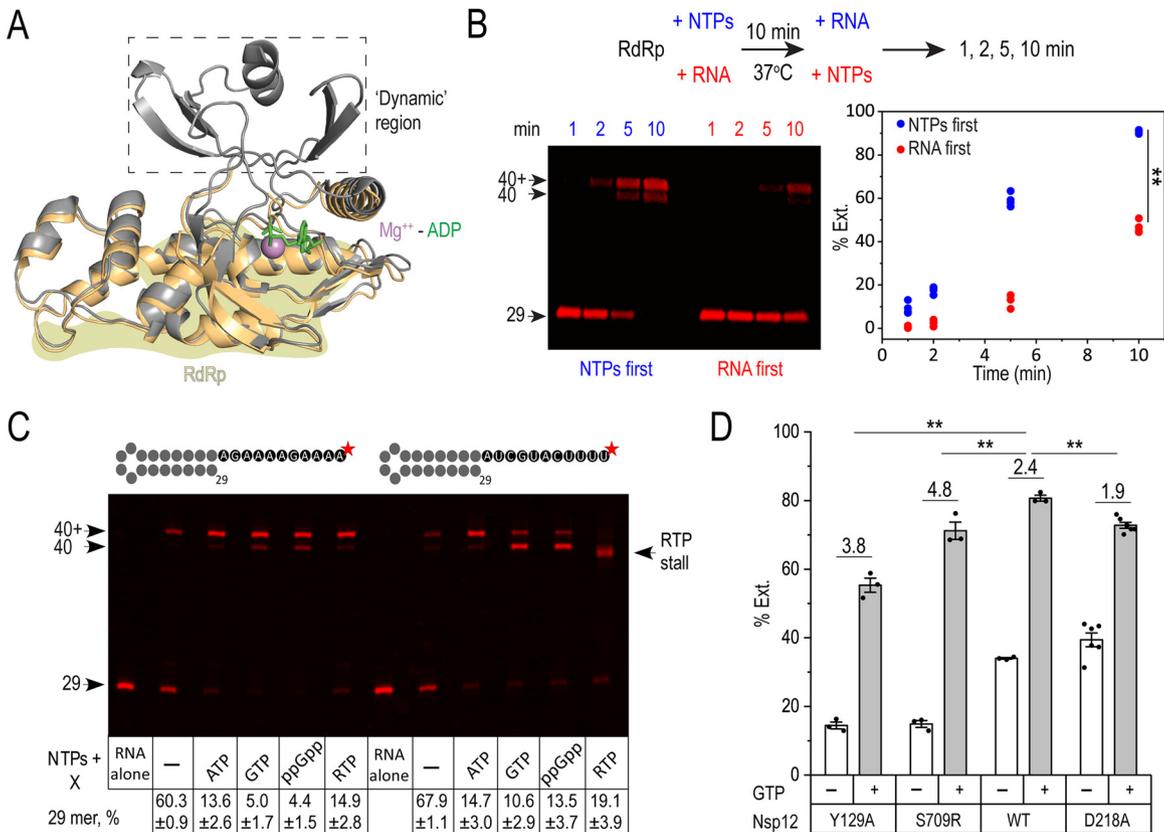


FIG 5 Allosteric activation of RdRp. (A) An overlay of NiRAN domain structures in the absence (wheat [PDB accession no. 6YYT]) and in the presence (gray [PDB accession no. 6XEZ]) of the bound ADP-Mg²⁺. The interface of the NiRAN domain and the RdRp domains is shown. (B) Activation of the Nsp12^R holoenzyme by preincubation with NTPs. (C) Activation of RNA synthesis (at 1 mM) on the CU (left) and 4N (right) templates. (D) Effects of Nsp12 substitutions on activation of RNA synthesis by 0.5 mM GTP; fold activation is shown above each set of bars. (B and D) RNA extension is shown as means ± SEM, and the *P* value was calculated by an unpaired two-tailed *t* test. n.s., not significant; **, *P* < 0.01.

precursors that are later processed into individual polypeptides, and the RdRp may be assembled concurrently with protein synthesis. Analysis of Nsp7/Nsp12 interactions by Trp fluorescence reveals that Nsp7 binds to both Nsp12 subunits and might favor a similar Nsp12^A-like state (Fig. S6). In support of the “scaffolding” function of the accessory subunits (35), we found that preincubation of Nsp12^R with Nsp7 and Nsp8 led to an increased activity (Fig. 4C).

We next tested if slowing translation during protein expression would promote Nsp12 folding. We constructed a BL21 strain with a K42T mutant of the ribosomal protein S12, which causes an approximately 2-fold reduction in the translation rate (39), and compared the Nsp12^R protein purified from this “slow” BL21 variant to the protein purified from wild-type BL21. We found that Nsp12^R purified from the mutant BL21 was approximately 2-fold more active (Fig. 4D), consistent with the favorable effect of attenuated translation.

Allosteric RdRp activation by nucleotides. Our results show that Nsp12^A and Nsp12^R differ dramatically in the conformations and interactions of their NiRAN domains (Fig. 3B). Although the NiRAN domain is not known to affect RNA chain extension directly, it interacts with the catalytic palm domain (13, 14, 21) and may modulate catalysis allosterically. The NiRAN domain is partially disordered in most unliganded structures of RdRp and transcription complexes but becomes ordered upon binding of ADP-Mg²⁺, GDP-Mg²⁺, and PP_i-Mg²⁺ to AS2 (Fig. 5A) (14, 17, 20). We hypothesized that, upon binding to nucleotides, the NiRAN domain would become more rigid, favoring an active RdRp conformation, thus leading to more efficient RNA elongation. First,

we compared rates of RNA synthesis under standard conditions in which RdRp is bound to the RNA scaffold prior to the addition of the NTP substrates to the “NTP-primed” reaction mixture, in which the order of reagent addition was reversed (Fig. 5B). The results show that preincubation with NTPs strongly potentiates Nsp12^R activity, an effect that may be mediated by the NiRAN domain.

Given that nucleotide binding to AS2 has been shown to remodel the NiRAN domain in the active Nsp12 (14), we surmised that NTP-mediated activation should also occur in Nsp12^A. To separate the direct and allosteric effects of NTPs, we used a CU template, which contains only purines in the transcribed region; as expected, the RNA was extended in the presence of CTP and UTP (Fig. 5C). In addition to detecting the runoff RNA (40 nucleotides [nt]), we detected a longer product that likely results from the terminal transferase activity of Nsp8, activity which prefers blunt over 3' recessed ends and ATP as a substrate (40). To assay the hypothetical allosteric activation of the RdRp by NTP bound to the NiRAN domain, we chose conditions under which less than 50% of the scaffold was extended.

Consistently with the allosteric effects of nontemplated nucleotides, transcription was activated in the presence of 1 mM ATP (>4-fold) or GTP (>10-fold) (Fig. 5C). An apparent promiscuity of AS2 suggests that other nucleotides might be able to substitute for ATP and GTP. To test this idea, we used a pause-promoting ATP analog, remdesivir triphosphate (RTP), and an allosteric effector of *E. coli* RNAP guanosine tetraphosphate (ppGpp). We found that the effects of RTP and ppGpp mimicked those of ATP and GTP, respectively, on the CU template (see Materials and Methods). Activation was also observed with the 4N template, on which ATP, GTP, and RTP but not ppGpp can be utilized as the substrates; as expected, RMP incorporation led to RdRp stalling before reaching the end of the template (18, 22). Consistently with the reported preference of the Nsp8 terminal transferase activity for ATP (40), the fraction of the extended RNA is reduced in the presence of GTP and ppGpp compared to that in the presence of ATP (Fig. 5C).

We hypothesized that RdRp-activating nucleotides act via binding to AS2 and stabilizing the RdRp-NiRAN interface. To test this hypothesis, we replaced two conserved residues at the interface. Tyr129 in the NiRAN domain is nearly invariant among all CoVs, whereas only small residues (Ser, Ala, Gly) are found at position 709 in the palm domain (Fig. 3C). Given this evolutionary conservation, we suspected that replacements of these amino acids might compromise the interdomain contacts, making RNA synthesis more dependent on the state of the NiRAN domain. Consistently, we found that Y129A and S709R substitutions reduced RNA synthesis activity while potentiating activation by 0.5 mM GTP (Fig. 5D).

The catalytic activity of the NiRAN domain has been shown to be independent of the RdRp function; Nsp9 modification occurs normally in an enzyme with substitutions in AS1 that inactivate the RdRp (16). To determine if the converse is true, we replaced Asp218, which coordinates the Mg²⁺ ion in the AS2 (14) and is critical for Nsp9 NMPylation and viral replication (16), with Ala. This substitution did not compromise RNA synthesis, confirming that AS1 and AS2 are functionally independent, but it modestly reduced GTP-dependent activation (Fig. 5D), suggesting that, if the allosteric GTP binds to the NiRAN AS2, Asp218 does not measurably contribute to nucleotide affinity. This observation is not entirely surprising because Asp residues are critical for substrate positioning but make lesser contributions to substrate binding in other viral polymerases (41).

DISCUSSION

Our results lead to two principal conclusions. First, SARS-CoV-2 Nsp12 depends on cotranslational folding, facilitated by ribosome pausing, and on interactions with the accessory subunits to attain the active conformation. Second, the two nucleotidyl transfer catalytic sites in Nsp12, a unique property of *Nidovirales*, appear to be connected allosterically,

with nucleotides including various analogs that bind to NiRAN AS2 and activate RNA chain extension in RdRp AS1.

Pause-free translation yields inactive RdRp. Our results demonstrate that over-optimized Nsp12^R mRNA produces a soluble but misfolded protein in which RNA binding and catalytic activity (Fig. 2C and D) are compromised. Notably, despite the dramatic differences in their activities, all structures of SARS-CoV-2 transcription complexes reported so far are closely similar (12–14), reflecting the bias introduced during cryoEM analysis, in which only a small fraction of “good” particles is selected based on image analysis (e.g., about 1% in a study of RdRp inhibition by remdesivir [42]). A preparation comprised of largely inactive enzymes remains amenable to the cryoEM analysis but would compromise biochemical experiments; to rephrase the fourth commandment of enzymology (43), thou shalt not waste clean thinking on dead enzymes. For example, a conclusion that SARS RdRp is more active than the SARS-CoV-2 enzyme (35) is predicated on the assumption that both RdRps are properly folded. Even more critically, inactive RdRps cannot be used to screen potential inhibitors.

While recoding is routinely used to optimize heterologous protein expression (27), the existence and frequent clustering of rare codons in mRNAs encoding many essential proteins, especially, large, multidomain ones, indicate their crucial role as regulators of protein folding. For example, native nonoptimal codons in intrinsically disordered regions (IDRs) are essential for the function of circadian clock oscillators (44, 45). IDRs often serve as platforms for protein-protein interactions (46) but can become trapped in unproductive states in the absence of their interaction partners. Our analysis supports this scenario by showing that an unstructured region that binds Nsp7 displays substantial differential sensitivity to EDC (Fig. 3B) and that interaction with Nsp7 locks Nsp12 in an active conformation (Fig. S6). When added to misfolded Nsp12, Nsp7/8 only modestly increases its activity (Fig. 4C). However, because all Nsps are produced as a giant precursor in coronavirus-infected cells (8), the accessory subunits may aid Nsp12 folding cotranslationally, as apparently happens during their coexpression in *E. coli* (23). Likewise, coexpression of *E. coli* RNA polymerase subunits suppresses assembly defects conferred by deletions in the catalytic subunits (47).

More broadly, our findings have implications for the heterologous expression of countless other proteins. Although examples of deleterious synonymous substitutions have been reported, these cases have been generally perceived as outliers. In retrospect, optimization-induced misfolding is likely to be far more prevalent than previously thought, with different recoding approaches impacting the structure and activity of the resulting protein in substantially different ways. The importance of cotranslational folding, particularly for large and dynamic proteins that contain essential mobile regions, emphasizes a need for the integration of diverse approaches, from ribosome profiling to machine learning, during a rational design of coding sequences to avoid misfolding traps.

Another important implication of the codon usage impact on SARS-CoV-2 protein folding, structure, and activity lies in the interpretation of genomic surveillance data. So far, the focus of the analysis of the genetic variability of SARS-CoV-2 has been on characterization of variants of concern, and designation of its evolutionary lineages has been in nonsynonymous changes, i.e., amino acid substitutions. Many of those amino acid substitutions show little-to-no impact on properties of proteins in which they appear (48, 49). We posit that many synonymic mutations, and even some nonsynonymic ones, may manifest their effects primarily at the level of cotranslational folding, rather than in the properties of the folded protein *in vitro*, or impact those through altering the ratio of folded to misfolded proteins during the infection.

Crosstalk between two catalytic sites of SARS-CoV-2 Nsp12. Decades of studies of viral RdRps focused on the mechanism of RNA synthesis and identification of nucleoside analogs that inhibit viral replication. During the COVID-19 pandemic, repurposing of the existing drugs targeting RdRp, justified by structural similarity among RdRp active sites (9), became an urgent priority. Among these drugs, remdesivir received the most attention, even though the estimates of its clinical effectiveness

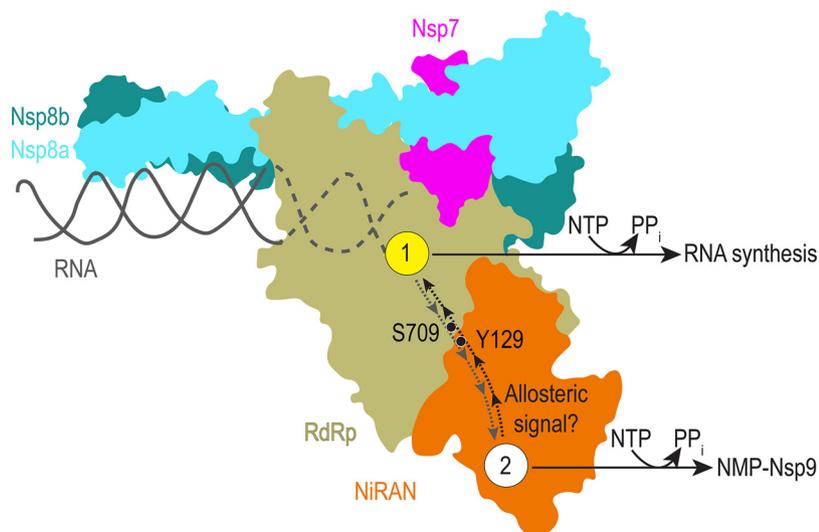


FIG 6 SARS-CoV-2 replication critically depends on two active sites in Nsp12 that mediate NMP transfer to RNA (AS1) and Nsp9 protein (AS2). Substrate (or inhibitor) binding to one site may be communicated to the other site through a highly conserved domain interface.

range from moderate to insignificant (50, 51). The CoV RdRp readily uses RTP as a substrate in place of ATP and temporarily stalls downstream at the site of RMP incorporation (18, 22). However, the proposed mechanisms of the inhibitory effect of RTP vary widely, from RdRp stalling to RNA chain termination to disassembly of the RdRp (52–54). It is presently unclear whether antiviral effects of remdesivir are due to delays in RNA synthesis or to errors in the product RNAs, as is the case with another purine analog, favipiravir (19).

Although efforts aimed at the identification of nucleoside analog inhibitors of RdRp are focused on AS1, it is clear that effects of nucleoside analogs on SARS-CoV-2 replication may be multifaceted. Nsp12 contains two active sites separated by more than 80 Å (Fig. 6), both of which can bind NTPs and nucleoside analogs. The functions of AS1 and AS2 are largely independent; Nsp12 containing double substitutions in AS1 that abolish elongation is fully competent for Nsp9 NMPylation (16), whereas the D218A substitution in Nsp12 that abolishes NMPylation blocks viral replication (16) but does not compromise RNA extension (Fig. 5D). In addition, each subunit of the Nsp8 dimer can also bind NTPs (40), and although there is no structural evidence of nucleotide binding to Nsp8 and its terminal transferase activity might be posttranscriptional, SARS-CoV-2 RdRp contains, all together, four nucleotide-binding sites.

Thus, one cannot assume that the observed effect of a nucleotide is mediated via the “primary” nucleotide binding to AS1; indeed, we show here that RTP promotes RNA synthesis when it cannot be incorporated into RNA, and this effect is even more pronounced with ppGpp (Fig. 4C). Competitive inhibitors binding in AS2 or transferring noncognate ligands to Nsp9 are likely to inhibit replication. In the latter case, misincorporation may have more lasting effects because errors in the nascent RNA can be corrected by the SARS-CoV-2 proofreading exonuclease Nsp14 (26).

We hypothesize that AS1 and AS2 are allosterically linked, enabling coordinated control of the RdRp activity. The NiRAN and palm domains form an extensive interface composed of highly conserved residues, including Tyr129 (Fig. 3C). Upon binding to AS2, nucleotides induce NiRAN folding and lead to subtle changes at the domain interface (14, 17, 20). We show that binding of nucleotides that cannot be incorporated into RNA potentiates RdRp activity (Fig. 5C). There is currently no direct evidence that this effect is triggered through their binding to AS2, but the effects of substitutions in the NiRAN domain (Fig. 5D) and structural data (14, 17, 20) support this model. Although rigorous computational, structural, and biochemical analyses will be required to test

this hypothesis, it is already clear that, when considering the effects of various nucleotide analogs on viral RNA synthesis, their binding to AS2 (and, perhaps, AS3 and AS4 as well) cannot be ignored.

The open active sites of viral RdRp can accommodate highly diverse substrates, some of which have been developed into therapeutics (10). Our findings that ppGpp activates RdRp similarly to GTP (Fig. 5C) suggests that other nucleotide-binding sites in Nsp12 are also promiscuous. Furthermore, the interplay between the binding of RTP to both catalytic and allosteric (relative to RNA synthesis) sites and competition therein with cellular NTPs call for a nuanced interpretation of the remdesivir inhibition mechanism, as does potential competition with ppGpp binding to the allosteric site (AS2). The biological activity of ppGpp has long been considered to be limited to bacteria and plastids, but its action as an alarmone has recently been demonstrated in human cells (55), raising the possibility that ppGpp and other nontemplating nucleotides impact SARS-CoV-2 replication in host cells.

Allosteric control of SARS-CoV-2 RdRp invites interesting parallels with *E. coli* Q β replicase, which also consists of four subunits, the phage-encoded RdRp (β -subunit) and three host RNA-binding proteins, the translation elongation GTPases EF-Tu and EF-Ts and a ribosomal protein, S1 (56). Similarly to Nsp7/8, EF-Tu and EF-Ts aid in the cotranslational assembly of Q β RdRp (57); EF-Tu also forms a part of the single-stranded RNA exit channel, assisting in RNA strand separation during elongation, whereas S1 acts as an initiation factor (56). EF-Tu and EF-Ts binding to ppGpp modulates host translation (58) and RNA synthesis by Q β (59), suggesting that RNA viruses from bacteria to humans may employ nucleotide analogs as sensors of cellular metabolism.

The SARS-CoV-2 RdRp subunit composition and dynamics resemble those of structurally unrelated bacterial RNA polymerases (RNAPs). Bacterial enzymes are composed of 4 to 7 subunits and are elaborately controlled by regulatory nucleic acid signals and proteins that induce conformational changes in the transcription complex, as revealed by many recent cryoEM studies (60–62). Notably, most natural and synthetic products that inhibit bacterial RNAPs alter protein interfaces or trap transient intermediates rather than block nucleotide addition and bind to many different sites (63). Compared to simpler RdRps, the SARS-CoV-2 enzyme, with several active sites and many conserved protein interfaces, may be an easier target for diverse small molecules that inhibit subunit or domain interactions or interrupt allosteric signals. Given the outsized importance of coronaviruses to human health, efforts to identify diverse inhibitors of RdRp, beyond nucleotide analogs, should be prioritized. Given the broad utilization of nucleotides by host enzymes, such as polymerases, kinases, lyases, etc., the chances of off-target side effects during therapeutic administration thereof are greatly elevated (64), whereas a druggable target unique to the pathogen, such as the SARS-CoV-2 NiRAN domain, bears inherently lower risks of nonspecific interactions. The potential for the allosteric regulation of NTP condensation by SARS-CoV-2 RdRp has recently been highlighted by a computational study that identified several motifs under allosteric control (65). Notably, the NiRAN domain makes extensive contacts with allosteric motif D and fewer contacts with motifs A and B (65), consistent with our findings that nontemplating phosphorylated nucleotide binding activates RdRp, solidifying its potential as the drug target.

MATERIALS AND METHODS

Construction of expression vectors. Plasmids used in this study are shown in Fig. S1 in the supplemental material. The SARS-CoV-2 *nsp7/8/12^A* genes were codon optimized for expression in *E. coli*, synthesized by GenScript, and subcloned into standard pET-derived expression vectors under the control of the T7 gene 10 promoter and *lac* repressor. The derivative plasmids were constructed by standard molecular biology approaches with restriction and modification enzymes from New England Biolabs, taking advantage of the existing or silent restriction sites engineered into the Nsp12 coding sequence. DNA oligonucleotides for vector construction and sequencing were obtained from Millipore Sigma. The sequences of all plasmids, including pET22a-Nsp12, were confirmed by Sanger sequencing at the Genomics Shared Resource Facility (The Ohio State University) and are available upon request.

Protein expression and purification. Nsp7/8 were overexpressed in *E. coli* XJB(DE3) cells (Zymo Research; catalog no. T5051). Nsp12 variants were overexpressed in *E. coli* BL21(DE3) cells (Novagen;

catalog no. 69450). Strains were grown in lysogenic broth (LB) with appropriate antibiotics: kanamycin (50 $\mu\text{g/ml}$), carbenicillin (100 $\mu\text{g/ml}$), and chloramphenicol (25 $\mu\text{g/ml}$). All protein purification steps were carried out at 4°C.

For Nsp7/8, cells were cultured at 37°C to an optical density at 600 nm (OD_{600}) of 0.6 to 0.8, and the temperature was lowered to 16°C. Expression was induced with 0.2 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG; GoldBio; catalog no. I2481C25) for 18 h. Induced cells were harvested by centrifugation (6,000 $\times g$), resuspended in lysis buffer A (100 mM HEPES, pH 7.5, 300 mM NaCl, 5% glycerol [vol/vol], 1 mM phenylmethylsulfonyl fluoride [PMSF; ACROS Organics; catalog no. 329-98-6], 5 mM β -mercaptoethanol [β -ME], 10 mM imidazole), and lysed by sonication. The lysate was cleared by centrifugation (10,000 $\times g$). The soluble protein was purified by absorption to Ni^{2+} -nitrilotriacetic acid (NTA) resin (Cytiva; catalog no. 17531801), washed with Ni-buffer A (20 mM HEPES, pH 7.5, 300 mM NaCl, 5% glycerol, 5 mM β -ME, 50 mM imidazole), and eluted with Ni-buffer B (20 mM HEPES, pH 7.5, 50 mM NaCl, 5% glycerol, 5 mM β -ME, 300 mM imidazole). The eluted protein was further loaded onto a Resource Q ion-exchange column (Cytiva; catalog no. 17117701) in Q buffer A (20 mM HEPES, pH 7.5, 5% glycerol, 5 mM β -ME) and eluted with a gradient of Q buffer B (20 mM HEPES, pH 7.5, 1 M NaCl, 5% glycerol, 5 mM β -ME). The fusion protein was treated with tobacco etch virus (TEV) protease at 4°C overnight supplemented with 20 mM imidazole and was passed through Ni^{2+} -NTA resin. The untagged, cleaved protein was loaded onto a Superdex 75 10/300 GL column (Cytiva; catalog no. 29148721) in Ni-buffer A. Peak fractions were assessed by SDS-PAGE and Coomassie blue staining. Purified protein was dialyzed into storage buffer A (20 mM HEPES, pH 7.5, 150 mM NaCl, 45% glycerol, 2.5 mM β -ME), aliquoted, and stored at -80°C .

For Nsp12, cells were cultured at 37°C to an OD_{600} of 0.6 to 0.8, and the temperature was lowered to 16°C. Expression was induced with 0.1 mM IPTG for 18 h. Induced cells were harvested by centrifugation, resuspended in lysis buffer B (100 mM HEPES, pH 7.5, 300 mM KCl, 5% glycerol, 2 mM MgCl_2 , protease inhibitor cocktail [Complete, EDTA-free; Roche Diagnostics; catalog no. 11836170001], 1 mM PMSF, 10 mM imidazole, 5 mM β -ME), and lysed by sonication. The cleared lysate was applied to Ni^{2+} -NTA resin (Cytiva), washed with Ni-buffer C (20 mM HEPES, pH 7.5, 300 mM KCl, 5% glycerol, 2 mM MgCl_2 , 5 mM β -ME, 0.1 mM PMSF) supplemented with 30 mM imidazole, and eluted with Ni-buffer D (20 mM HEPES, pH 7.5, 50 mM KCl, 5% glycerol, 2 mM MgCl_2 , 5 mM β -ME, 0.1 mM PMSF, 300 mM imidazole). The eluted protein was further purified by Resource Q (Cytiva) with Q buffer C (20 mM HEPES, pH 7.5, 5% glycerol, 2 mM MgCl_2 , 1 mM dithiothreitol [DTT]) and Q buffer D (20 mM HEPES, pH 7.5, 1 M KCl, 5% glycerol, 2 mM MgCl_2 , 1 mM DTT). Then the fusion protein was treated with an appropriate protease (TEV or SUMO protease) at 4°C. After an overnight treatment, protein was supplemented with 20 mM imidazole and passed through Ni^{2+} -NTA resin. The untagged protein was applied to the Superdex 200 increase 10/300 GL column (Cytiva; catalog no. 28990944) in SEC buffer (20 mM HEPES, pH 7.5, 300 mM KCl, 5% glycerol, 2 mM MgCl_2 , 1 mM DTT). Peak fractions were assessed by SDS-PAGE and Coomassie blue staining. Purified protein was dialyzed into storage buffer B (20 mM HEPES, pH 7.5, 150 mM KCl, 45% glycerol, 1 mM MgCl_2 , 1 mM DTT), aliquoted, and stored at -80°C .

Expression by slow ribosomes. To test the effect of slow translation on Nsp12 activity, a derivative of BL21 (IA659) containing a K42T substitution in the ribosomal protein S12 was constructed by P1 transduction from the DEV3 *E. coli* strain (KL16 *lac5 strA2*; obtained from Kurt Fredrick, The Ohio State University) and selection on streptomycin (50 mg/liter). This substitution reduces the translation rate ~ 2 -fold (39). Following sequencing of the *rpsL* gene to confirm the substitution, the slow BL21 strain was transformed with the plasmid encoding Nsp12^R. The protein was purified as described above.

RNA extension assays. An RNA oligonucleotide (5'-UUUUC AUGCUACGCGUAGUUUUUCUACGCG-3'; 4N) with cyanine 5.5 at the 5' end was obtained from Millipore Sigma (USA). Prior to the reaction, the RNA was annealed in 20 mM HEPES, pH 7.5, 50 mM KCl by heating the mixture to 75°C and then gradually cooling it to 4°C. Reactions were carried out at 37°C with 500 nM Nsp12 variant, 1 μM Nsp7, 1.5 μM Nsp8, 200 nM RNA, and 250 μM NTPs (Cytiva; catalog no. 27202501) in the transcription buffer (20 mM HEPES, pH 7.5, 15 mM KCl, 5% glycerol, 2 mM MgCl_2 , 1 mM DTT). RNA extension reactions were stopped at the desired times by adding 2 \times stop buffer (8 M urea, 20 mM EDTA, 1 \times Tris-borate-EDTA [TBE], 0.2% bromophenol blue). Samples were heated for 2 min at 95°C and separated by electrophoresis in denaturing 9% acrylamide (19:1) gels (7 M urea, 0.5 \times TBE). The RNA products were visualized and quantified using Typhoon FLA9000 (GE Healthcare) and ImageQuant software. RNA extension assays were carried out in triplicates. Means and standard errors of the means (SEM) were calculated by OriginPro 2021 (OriginLab), and an unpaired two-tailed *t* test was performed using Excel (Microsoft).

Electrophoretic mobility shift assays. RdRp (Nsp12:Nsp7:Nsp8 = 1:2:3; indicated concentrations in Fig. 2D represent the Nsp12 concentration) in 20 mM HEPES, pH 7.5, 65/15 mM KCl, 5% glycerol, 2 mM MgCl_2 , 1 mM DTT were incubated with 100 nM 4N RNA at 37°C for 5 min. Then reactions were mixed with 10 \times loading buffer (30% glycerol, 0.2% Orange G) and run on a 3% agarose gel in 1 \times TBE on ice. The gel was visualized by Typhoon FLA9000.

Activation of Nsp12^R. To test the effect of holo RdRp formation, 5 μM Nsp12^R mixed with Nsp7/8 (10 μM /15 μM) in storage buffer B was incubated at 0°C or 37°C for 15 min and then stored at -20°C . RNA extension was performed as described above, and the reaction was stopped at 8 min. To test the effect of NTPs, RdRp was first incubated with NTPs for 10 min at 37°C, and then 200 nM RNA was added to initiate the reaction; the final concentrations of RdRp (500 nM), RNA (250 nM), and NTPs (250 μM) were identical to those used in assays with the simultaneous addition of the RNA scaffold and substrates. The reaction was stopped by adding 2 \times stop buffer at the indicated times.

Allosteric activation by nucleotides. A CU RNA hairpin (5'-AAAAGAAAAGACGCGUAGUUUUUCUACGCG-3'; CU) labeled with cyanine 5.5 at the 5' end (Millipore Sigma) was annealed in 20 mM HEPES,

pH 7.5, 50 mM KCl by heating the mixture to 75°C and then gradually cooling it to 4°C. RdRp holoenzymes (500 nM wild-type or mutant Nsp12^Δ, 1 μM Nsp7, 1.5 μM Nsp8 [final concentrations]) were mixed with ATP, GTP, remdesivir triphosphate (RTP; MedChemExpress; catalog no. GS443902), or ppGpp (TriLink BioTechnologies; catalog no. N-6001) at concentrations indicated in Fig. 5 in 20 mM HEPES, pH 7.5, 15 mM KCl, 5% glycerol, 1 mM DTT and either 2 or 1 mM MgCl₂ (with 1 or 0.5 mM activating nucleotide, respectively). Reaction mixtures were incubated for 5 min at 37°C, and RNA chain extension was initiated by the addition of 200 nM RNA and 100 μM CTP and UTP. Following 15 min of incubation at 37°C, reactions were stopped at the desired times by adding 2× stop buffer (8 M urea, 20 mM EDTA, 1× TBE, 0.2% bromophenol blue).

Tryptophan fluorescence. Tryptophan fluorescence spectroscopy was performed using a model F-7000 fluorescence spectrophotometer (Hitachi). The excitation wavelength was set at 280 nm, and the emission spectra were recorded from 310 to 370 nm, with a 5-nm slit width of excitation and emission. The scan speed was 240 nm/min. The temperature was maintained at 37°C by a thermostatic water circulator (NESLAB RTE-7; Thermo Scientific). The samples were prepared in 20 mM HEPES, pH 7.5, 65 mM KCl, 5% glycerol, 2 mM MgCl₂, 1 mM DTT. One micromolar Nsp12 and 2 μM Nsp7 were used to record the spectra of Nsp12 and Nsp7, respectively. To record the spectra of Nsp7-12, 1 μM Nsp12 was incubated with 2 μM Nsp7 at 37°C for 15 min. To collect the spectra of denatured proteins, 1 μM Nsp12 was incubated in 8 M urea at room temperature for 1 h. Three independent measurements, each in three technical replicates, were performed. The same results were obtained with proteins purified 3 months apart. Means, SEM, and second derivatives of the emission spectra were calculated by OriginPro 2021.

Conservation analysis. To assess the relative conservation of coronavirus proteins, 3,309 diverse coronavirus genomes, representing alpha-, beta-, gamma-, and deltacoronavirus genera were downloaded from GenBank in May 2020. High-quality CDSs (containing no more than 32 contiguous codons with ambiguous bases) were translated into five (poly)proteins, conserved across all coronaviruses: orf1ab, S, E, M, and N. Alignments of five open reading frames (ORFs) were produced using the MUSCLE program (66). For each alignment, column homogeneity and the weighted fraction of nongap characters (both ranging from 0 to 1) were calculated as described previously (67). The product of these two values was used as the conservation index (ranging from 0 to 1). For the whole-genome pan-Coronaviridae conservation map, only consensus positions (those with a fraction of gaps below 0.5) were used.

EDC modification and mass spectrometry. Approximately 0.5 mg/ml of Nsp12 in 20 mM HEPES, pH 7.5, 50 mM KCl, 2 mM MgCl₂, 1 mM DTT was mixed with freshly prepared EDC [*N*-(3-dimethylamino-propyl)-*N*'-ethyl carbodiimide hydrochloride; Sigma, catalog no. 03449]. EDC was added to a final concentration of 2 mM, and the reaction was performed at room temperature for 30 min. The reaction was quenched with a 50× molar excess of Tris-HCl (pH 7.5) for 5 min. Cross-linked protein samples were separated using SDS-PAGE, the protein bands were stained with GelCode Blue, and tryptic peptides were generated using an in-gel tryptic digestion kit (Thermo Scientific; catalog no. 89871); peptides were purified using Pierce 10-μl C₁₈ tips (Thermo Fisher; catalog no. PI87782). Peptides were analyzed in the Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) coupled to an EASY-nLC (Thermo Scientific) liquid chromatography system, with a 2-μm, 500-mm EASY-Spray column. The peptides were eluted over a 180-min linear gradient from 96% buffer A (water) to 40% buffer B (acetonitrile) and then continued to 98% buffer B over 20 min with a flow rate of 200 nl/min. Each full mass spectrometry (MS) scan (*R*=60,000) was followed by 20 data-dependent MS2 (*R*=15,000) with high-energy collisional dissociation (HCD) and an isolation window of 2.0 *m/z*. The normalized collision energy was set to 35. Precursors of charge states 2 to 6 and 4 to 6 were collected for MS2 scans; monoisotopic precursor selection was enabled, and a dynamic exclusion window was set to 30.0 s. The resulting raw files were searched in enumerative mode with pFind3 (68) in open search mode against the Nsp12 sequence; the inferred modifications over a 1% cutoff were used as “variable” modifications in the subsequent pLink2 search. The same files were then searched in cross-link discovery mode using pLink2 (69) against the Nsp12 sequence, using [EDC] as the cross-linking reagent, trypsin as the enzyme generating the peptides, and variable modifications set as inferred by pFind3.

Data availability. Mass spectrometry data sets have been deposited into MassIVE (accession no. MSV000086827, available for download at <http://massive.ucsd.edu/MSV000086827/>), and processed data are presented in Data Set S2.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.1 MB.

FIG S2, PDF file, 0.3 MB.

FIG S3, PDF file, 0.3 MB.

FIG S4, PDF file, 0.2 MB.

FIG S5, PDF file, 0.1 MB.

FIG S6, PDF file, 0.1 MB.

DATA SET S1, XLSX file, 0.7 MB.

DATA SET S2, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We are grateful to Georgi Belogurov, Venkat Gopalan, and Markus Wahl for suggestions, to Elena Kudryashova for help with data acquisition, and to Dmitri Svetlov for discussions and expert editing.

This work was supported by a seed grant from The Ohio State University Office of Research to I.A., by the National Institutes of Health grants GM067153 (I.A.) and GM126891 (E.N.), and by the Blavatnik Family Foundation and the Howard Hughes Medical Institute (E.N.). Y.I.W. and E.V.K. are supported by the Intramural Research Program of the National Institutes of Health (National Library of Medicine).

Conceptualization, I.A.; cloning, protein purification, and biochemistry, B.W. and I.A.; mass spectrometry, V.S.; bioinformatics analysis, Y.I.W.; funding acquisition and supervision, I.A., E.N., and E.V.K.; manuscript first draft, B.W. and I.A. All authors contributed to finalizing the written manuscript.

We declare no competing interests.

REFERENCES

- Piret J, Boivin G. 2020. Pandemics throughout history. *Front Microbiol* 11:631736. <https://doi.org/10.3389/fmicb.2020.631736>.
- Zhang Y, Gargan S, Lu Y, Stevenson NJ. 2021. An overview of current knowledge of deadly CoVs and their interface with innate immunity. *Viruses* 13:560. <https://doi.org/10.3390/v13040560>.
- Hu B, Guo H, Zhou P, Shi ZL. 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19:141–154. <https://doi.org/10.1038/s41579-020-00459-7>.
- Case JB, Winkler ES, Errico JM, Diamond MS. 2021. On the road to ending the COVID-19 pandemic: are we there yet? *Virology* 557:70–85. <https://doi.org/10.1016/j.virol.2021.02.003>.
- Plante JA, Mitchell BM, Plante KS, Debbink K, Weaver SC, Menachery VD. 2021. The variant gambit: COVID-19's next move. *Cell Host Microbe* 29:508–515. <https://doi.org/10.1016/j.chom.2021.02.020>.
- Munir K, Ashraf S, Munir I, Khalid H, Muneer MA, Mukhtar N, Amin S, Ashraf S, Imran MA, Chaudhry U, Zaheer MU, Arshad M, Munir R, Ahmad A, Zhao X. 2020. Zoonotic and reverse zoonotic events of SARS-CoV-2 and their impact on global health. *Emerg Microbes Infect* 9:2222–2235. <https://doi.org/10.1080/22221751.2020.1827984>.
- Gulyaeva AA, Gorbalenya AE. 2021. A nidovirus perspective on SARS-CoV-2. *Biochem Biophys Res Commun* 538:24–34. <https://doi.org/10.1016/j.bbrc.2020.11.015>.
- Snijder EJ, Decroly E, Ziebuhr J. 2016. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv Virus Res* 96:59–126. <https://doi.org/10.1016/bs.aivir.2016.08.008>.
- Peersen OB. 2019. A comprehensive superposition of viral polymerase structures. *Viruses* 11:745. <https://doi.org/10.3390/v11080745>.
- Jiang Y, Yin W, Xu HE. 2021. RNA-dependent RNA polymerase: structure, mechanism, and drug discovery for COVID-19. *Biochem Biophys Res Commun* 538:47–53. <https://doi.org/10.1016/j.bbrc.2020.08.116>.
- Wolff G, Melia CE, Snijder EJ, Barcena M. 2020. Double-membrane vesicles as platforms for viral replication. *Trends Microbiol* 28:1022–1033. <https://doi.org/10.1016/j.tim.2020.05.009>.
- Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368:779–782. <https://doi.org/10.1126/science.abb7498>.
- Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature* 584:154–156. <https://doi.org/10.1038/s41586-020-2368-8>.
- Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET, Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* 182:1560–1573.e13. <https://doi.org/10.1016/j.cell.2020.07.033>.
- Lehmann KC, Gulyaeva A, Zevenhoven-Dobbe JC, Janssen GM, Ruben M, Overkleef HS, van Veelen PA, Samborskiy DV, Kravchenko AA, Leontovich AM, Sidorov IA, Snijder EJ, Posthuma CC, Gorbalenya AE. 2015. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic Acids Res* 43:8416–8434. <https://doi.org/10.1093/nar/gkv838>.
- Slanina H, Madhugiri R, Bylapudi G, Schultheiss K, Karl N, Gulyaeva A, Gorbalenya AE, Linne U, Ziebuhr J. 2021. Coronavirus replication-transcription complex: vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. *Proc Natl Acad Sci U S A* 118: e2022310118. <https://doi.org/10.1073/pnas.2022310118>.
- Yan L, Ge J, Zheng L, Zhang Y, Gao Y, Wang T, Huang Y, Yang Y, Gao S, Li M, Liu Z, Wang H, Li Y, Chen Y, Guddat LW, Wang Q, Rao Z, Lou Z. 2021. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell* 184:184–193.e10. <https://doi.org/10.1016/j.cell.2020.11.016>.
- Gordon CJ, Tchesnokov EP, Woolner E, Perry JK, Feng JY, Porter DP, Gotte M. 2020. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem* 295:6785–6797. <https://doi.org/10.1074/jbc.RA120.013679>.
- Shannon A, Selisko B, Le NT, Huchting J, Touret F, Piorkowski G, Fattorini V, Ferron F, Decroly E, Meier C, Coutard B, Peersen O, Canard B. 2020. Rapid incorporation of favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat Commun* 11:4682. <https://doi.org/10.1038/s41467-020-18463-z>.
- Naydenova K, Muir KW, Wu LF, Zhang Z, Coscia F, Peet MJ, Castro-Hartmann P, Qian P, Sader K, Dent K, Kimanius D, Sutherland JD, Lowe J, Barford D, Russo CJ. 2021. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc Natl Acad Sci U S A* 118: e2021946118. <https://doi.org/10.1073/pnas.2021946118>.
- Kirchdoerfer RN, Ward AB. 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* 10:2342. <https://doi.org/10.1038/s41467-019-10280-3>.
- Kokic G, Hillen HS, Tegunov D, Dienemann C, Seitz F, Schmitzova J, Farnung L, Siewert A, Hobartner C, Cramer P. 2021. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun* 12:279. <https://doi.org/10.1038/s41467-020-20542-0>.
- Dangerfield TL, Huang NZ, Johnson KA. 2020. Remdesivir is effective in combating COVID-19 because it is a better substrate than ATP for the viral RNA-dependent RNA polymerase. *iScience* 23:101849. <https://doi.org/10.1016/j.isci.2020.101849>.
- Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, Ji W, Yan L, Zhu Y, Zhu C, Fang X, Yang X, Huang Y, Gao H, Liu F, Ge J, Sun Q, Yang X, Xu W, Liu Z, Yang H, Lou Z, Jiang B, Guddat LW, Gong P, Rao Z. 2020. Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* 182:417–428.e13. <https://doi.org/10.1016/j.cell.2020.05.034>.
- Bera SC, Seifert M, Kirchdoerfer RN, van Nies P, Wubulikasimu Y, Quack S, Papini FS, Arnold JJ, Canard B, Cameron CE, Depken M, Dulin D. 2021. The nucleotide addition cycle of the SARS-CoV-2 polymerase. *bioRxiv* <https://doi.org/10.1101/2021.03.27.437309>.
- Shannon A, Le NT, Selisko B, Eydoux C, Alvarez K, Guillemot JC, Decroly E, Peersen O, Ferron F, Canard B. 2020. Remdesivir and SARS-CoV-2: structural requirements at both nsp12 RdRp and nsp14 exonuclease active-

- sites. *Antiviral Res* 178:104793. <https://doi.org/10.1016/j.antiviral.2020.104793>.
27. Gustafsson C, Govindarajan S, Minshull J. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol* 22:346–353. <https://doi.org/10.1016/j.tibtech.2004.04.006>.
 28. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. 2007. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525–528. <https://doi.org/10.1126/science.1135308>.
 29. Liu Y. 2020. A code within the genetic code: codon usage regulates cotranslational protein folding. *Cell Commun Signal* 18:145. <https://doi.org/10.1186/s12964-020-00642-6>.
 30. Zhang G, Hubalewska M, Ignatova Z. 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 16:274–280. <https://doi.org/10.1038/nsmb.1554>.
 31. Liutkute M, Samatova E, Rodnina MV. 2020. Cotranslational folding of proteins on the ribosome. *Biomolecules* 10:97. <https://doi.org/10.3390/biom10010097>.
 32. Pechmann S, Frydman J. 2013. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* 20:237–243. <https://doi.org/10.1038/nsmb.2466>.
 33. Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, Tamir H, Achdout H, Stein D, Israeli O, Beth-Din A, Melamed S, Weiss S, Israely T, Paran N, Schwartz M, Stern-Ginossar N. 2020. The coding capacity of SARS-CoV-2. *Nature* 589:125–130. <https://doi.org/10.1038/s41586-020-2739-1>.
 34. Abbott JA, Livingston NM, Egri SB, Guth E, Francklyn CS. 2017. Characterization of aminoacyl-tRNA synthetase stability and substrate interaction by differential scanning fluorimetry. *Methods* 113:64–71. <https://doi.org/10.1016/j.ymeth.2016.10.013>.
 35. Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, Wang Q, Sun Y, Fan Z, Qi J, Gao GF, Shi Y. 2020. Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. *Cell Rep* 31:107774. <https://doi.org/10.1016/j.celrep.2020.107774>.
 36. Chavez-Abiega S, Goedhart J, Bruggeman FJ. 2019. Physical biology of GPCR signalling dynamics inferred from fluorescence spectroscopy and imaging. *Curr Opin Struct Biol* 55:204–211. <https://doi.org/10.1016/j.sbi.2019.05.007>.
 37. Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, Liu Y. 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell* 59:744–754. <https://doi.org/10.1016/j.molcel.2015.07.018>.
 38. Zhou M, Wang T, Fu J, Xiao G, Liu Y. 2015. Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol Microbiol* 97:974–987. <https://doi.org/10.1111/mmi.13079>.
 39. Ruusala T, Andersson D, Ehrenberg M, Kurland CG. 1984. Hyper-accurate ribosomes inhibit growth. *EMBO J* 3:2575–2580. <https://doi.org/10.1002/j.1460-2075.1984.tb02176.x>.
 40. Tvarogová J, Madhugiri R, Bylapudi G, Ferguson LJ, Karl N, 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. <https://doi.org/10.1128/JVI.00291-19>.
 41. Gleghorn ML, Davydova EK, Basu R, Rothman-Denes LB, Murakami KS. 2011. X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc Natl Acad Sci U S A* 108:3566–3571. <https://doi.org/10.1073/pnas.1016691108>.
 42. Yin W, Mao C, Luan X, Shen DD, Shen Q, Su H, Wang X, Zhou F, Zhao W, Gao M, Chang S, Xie YC, Tian G, Jiang HW, Tao SC, Shen J, Jiang Y, Jiang H, Xu Y, Zhang S, Zhang Y, Xu HE. 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368:1499–1504. <https://doi.org/10.1126/science.abc1560>.
 43. Kornberg A. 2000. Ten commandments: lessons from the enzymology of DNA replication. *J Bacteriol* 182:3613–3618. <https://doi.org/10.1128/JB.182.13.3613-3618.2000>.
 44. Xu Y, Ma P, Shah P, Rokas A, Liu Y, Johnson CH. 2013. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* 495:116–120. <https://doi.org/10.1038/nature11942>.
 45. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y. 2013. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495:111–115. <https://doi.org/10.1038/nature11833>.
 46. Bugge K, Brakti I, Fernandes CB, Dreier JE, Lundsgaard JE, Olsen JG, Skriver K, Kragelund BB. 2020. Interactions by disorder—a matter of context. *Front Mol Biosci* 7:110. <https://doi.org/10.3389/fmolb.2020.00110>.
 47. Artsimovitch I, Svetlov V, Murakami KS, Landick R. 2003. Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* 278:12344–12355. <https://doi.org/10.1074/jbc.M211214200>.
 48. Cella E, Benedetti F, Fabris S, Borsetti A, Pezzuto A, Ciotti M, Pascarella S, Ceccarelli G, Zella D, Ciccozzi M, Giovanetti M. 18 March 2021. SARS-CoV-2 lineages and sub-lineages circulating worldwide: a dynamic overview. *Chemotherapy* <https://doi.org/10.1159/000515340>.
 49. Zeng L, Li D, Tong W, Shi T, Ning B. 19 January 2021. Biochemical features and mutations of key proteins in SARS-CoV-2 and their impacts on RNA therapeutics. *Biochem Pharmacol* <https://doi.org/10.1016/j.bcp.2021.114424>.
 50. Al-Abdoun A, Bizanti A, Barbarawi M, Jabri A, Kumar A, Fashanu OE, Khan SU, Zhao D, Antar AAR, Michos ED. 2021. Remdesivir for the treatment of COVID-19: a systematic review and meta-analysis of randomized controlled trials. *Contemp Clin Trials* 101:106272. <https://doi.org/10.1016/j.cct.2021.106272>.
 51. Kaka AS, MacDonald R, Greer N, Vela K, Duan-Porter W, Obley A, Wilt TJ. 2021. Major update: remdesivir for adults with COVID-19: a living systematic review and meta-analysis for the American College of Physicians practice points. *Ann Intern Med* 174:663–672. <https://doi.org/10.7326/M20-8148>.
 52. Olotu FA, Omolabi KF, Soliman MES. 2021. Piece of the puzzle: remdesivir disassembles the multimeric SARS-CoV-2 RNA-dependent RNA polymerase complex. *Cell Biochem Biophys* 79:175–187. <https://doi.org/10.1007/s12013-021-00977-y>.
 53. Tchesnokov EP, Gordon CJ, Woolner E, Kocinkova D, Perry JK, Feng JY, Porter DP, Gotte M. 2020. Template-dependent inhibition of coronavirus RNA-dependent RNA polymerase by remdesivir reveals a second mechanism of action. *J Biol Chem* 295:16156–16165. <https://doi.org/10.1074/jbc.AC120.015720>.
 54. Bravo JPK, Dangerfield TL, Taylor DW, Johnson KA. 2021. Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. *Mol Cell* 81:1548–1552.e4. <https://doi.org/10.1016/j.molcel.2021.01.035>.
 55. Ito D, Kawamura H, Oikawa A, Ihara Y, Shibata T, Nakamura N, Asano T, Kawabata SI, Suzuki T, Masuda S. 2020. ppGpp functions as an alarmone in metazoa. *Commun Biol* 3:671. <https://doi.org/10.1038/s42003-020-01368-4>.
 56. Takeshita D, Yamashita S, Tomita K. 2014. Molecular insights into replication initiation by Qbeta replicase using ribosomal protein S1. *Nucleic Acids Res* 42:10809–10822. <https://doi.org/10.1093/nar/gku745>.
 57. Takeshita D, Tomita K. 2010. Assembly of Q[β] viral RNA polymerase with host translational elongation factors EF-Tu and -Ts. *Proc Natl Acad Sci U S A* 107:15733–15738. <https://doi.org/10.1073/pnas.1006559107>.
 58. Rojas AM, Ehrenberg M, Andersson SG, Kurland CG. 1984. ppGpp inhibition of elongation factors Tu, G and Ts during polypeptide synthesis. *Mol Gen Genet* 197:36–45. <https://doi.org/10.1007/BF00327920>.
 59. Blumenthal T. 1977. Interaction of Q[β] RNA replicase with guanine nucleotides. Different modes of inhibition and inactivation. *Biochim Biophys Acta* 478:201–208. [https://doi.org/10.1016/0005-2787\(77\)90183-6](https://doi.org/10.1016/0005-2787(77)90183-6).
 60. Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina TV, Artsimovitch I, Landick R, Darst SA. 2018. Structural basis for transcript elongation control by NusG family universal regulators. *Cell* 173:1650–1662.e14. <https://doi.org/10.1016/j.cell.2018.05.017>.
 61. Said N, Hilal T, Sunday ND, Khatri A, Burger J, Mielke T, Belogurov GA, Loll B, Sen R, Artsimovitch I, Wahl MC. 2021. Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase rho. *Science* 371:eabd1673. <https://doi.org/10.1126/science.abd1673>.
 62. Guo X, Myasnikov AG, Chen J, Crucifix C, Papai G, Takacs M, Schultz P, Weixlbaumer A. 2018. Structural basis for NusA stabilized transcriptional pausing. *Mol Cell* 69:816–827.e4. <https://doi.org/10.1016/j.molcel.2018.02.008>.
 63. Chen J, Boyaci H, Campbell EA. 2021. Diverse and unified mechanisms of transcription initiation in bacteria. *Nat Rev Microbiol* 19:95–109. <https://doi.org/10.1038/s41579-020-00450-2>.
 64. Pich O, Muinos F, Lolkema MP, Steeghs N, Gonzalez-Perez A, Lopez-Bigas N. 2019. The mutational footprints of cancer therapies. *Nat Genet* 51:1732–1740. <https://doi.org/10.1038/s41588-019-0525-5>.
 65. Tee WV, Tan ZW, Lee K, Guarnera E, Berezovsky IN. 2021. Exploring the allosteric territory of protein function. *J Phys Chem B* 125:3763–3780. <https://doi.org/10.1021/acs.jpcc.1c00540>.

66. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>.
67. Esterman ES, Wolf YI, Kogay R, Koonin EV, Zhaxybayeva O. 2021. Evolution of DNA packaging in gene transfer agents. *Virus Evol* 7:veab015. <https://doi.org/10.1093/ve/veab015>.
68. Chi H, Liu C, Yang H, Zeng WF, Wu L, Zhou WJ, Wang RM, Niu XN, Ding YH, Zhang Y, Wang ZW, Chen ZL, Sun RX, Liu T, Tan GM, Dong MQ, Xu P, Zhang PH, He SM. 2018. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol* 36:1059–1061. <https://doi.org/10.1038/nbt.4236>.
69. Chen ZL, Meng JM, Cao Y, Yin JL, Fang RQ, Fan SB, Liu C, Zeng WF, Ding YH, Tan D, Wu L, Zhou WJ, Chi H, Sun RX, Dong MQ, He SM. 2019. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* 10:3404. <https://doi.org/10.1038/s41467-019-11337-z>.