Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex

Brandon Malone[1,8], James Chen[1,8], Qi Wang[2], Eliza Llewellyn[1], Young Joo Choi[1], Paul Dominic B. Olinares[3], Xinyun Cao[4], Carolina Hernandez[5], Edward T. Eng[5], Brian T. Chait[3], David E. Shaw[2,6], Robert Landick[4,7], Seth A. Darst[1,9], Elizabeth A. Campbell[1,9,10]

[1]Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY, 10065 USA.

[2]D. E. Shaw Research, New York, NY 10036 USA.

[3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, 10065 USA.

[4]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706 USA.

[5]The National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, 10027 USA.

[6]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032 USA.

[7]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706 USA.

[8]These authors contributed equally: Brandon Malone, James Chen.

[9]Correspondence to: darst@rockefeller.edu, campbee@rockefeller.edu

Author Contributions: Conceptualization; B.M., J.C., Q.W., D.E.S., S.A.D., E.A.C. Protein purification, biochemistry; B. M., J.C., E.L., Y.J.C. Mass spectrometry acquisition and analysis; P.D.B.O., B.C. Cryo-EM specimen preparation; B.M., J.C. Cryo-EM data collection and processing: J.C., C.H., E.T.E. Model building, and structural analysis; J.C., B.M., S.A.D., E.A.C. Molecular dynamics simulations and analysis; Q.W., D.E.S. Funding acquisition and supervision; B.C., D.E.S., R.L., S.A.D, E.A.C. All authors contributed to the writing of the manuscript.

Competing Interest Statement: The authors declare no conflict of interest.

Classification: Biological Sciences; Biophysics and Computational Biology.

Keywords: backtracking; coronavirus; cryo-electron microscopy; molecular dynamics; RNA-dependent RNA polymerase

Data deposition: The cryo-EM density maps have been deposited in the EMDataBank: EMD-23007 ($nsp13_1$-BTC_5), EMD-23008 ($nsp13_2$-BTC_5), EMD-23009 [BTC_5(local)]. Atomic coordinates have been deposited in the Protein Data Bank: 7KRN ($nsp13_1$-BTC_5), 7KRO ($nsp13_2$-BTC_5), 7KRP [BTC_5(local)] and will be released upon peer-review publication. Please contact Elizabeth Campbell, if you require the data before. The molecular dynamics trajectories described in this work are available at https://www.deshawresearch.com/downloads/download_trajectory_sarscov2.cgi/.

This PDF file include:

Main Text
Figures 1 to 6
Main Text References
Supplementary text and methods
Figures S1 to S6
Tables S1
SI References

Abstract

Backtracking, the reverse motion of the transcriptase enzyme on the nucleic acid template, is a universal regulatory feature of transcription in cellular organisms but its role in viruses is not established. Here we present evidence that backtracking extends into the viral realm, where

backtracking by the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) may aid viral transcription and replication. Structures of SARS-CoV-2 RdRp bound to the essential nsp13 helicase and RNA suggested the helicase facilitates backtracking. We use cryo-electron microscopy, RNA-protein crosslinking, and unbiased molecular dynamics simulations to characterize SARS-CoV-2 RdRp backtracking. The results establish that the single-stranded 3'-segment of the product-RNA generated by backtracking extrudes through the RdRp NTP-entry tunnel, that a mismatched nucleotide at the product-RNA 3'-end frays and enters the NTP-entry tunnel to initiate backtracking, and that nsp13 stimulates RdRp backtracking. Backtracking may aid proofreading, a crucial process for SARS-CoV-2 resistance against antivirals.

Significance Statement

The COVID-19 pandemic is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The SARS-CoV-2 genome is replicated and transcribed by its RNA-dependent RNA polymerase (RdRp), which is the target for antivirals such as remdesivir. We use a combination of approaches to show that backtracking (backwards motion of the RdRp on the template RNA) is a feature of SARS-CoV-2 replication/transcription. Backtracking may play a critical role in proofreading, a crucial process for SARS-CoV-2 resistance against many antivirals.

Introduction

SARS-CoV-2 is the causative agent of the current COVID-19 pandemic (1, 2). The SARS-CoV-2 genome is replicated and transcribed by its RNA-dependent RNA polymerase holoenzyme [holo-RdRp, subunit composition nsp7/nsp8$_2$/nsp12 (3, 4)] in a replication-transcription complex (RTC), which is the target for antivirals such as remdesivir [Rdv; (5)]. The holo-RdRp is thought to coordinate with many co-factors to carry out its function (6, 7). Some of these co-factors, such as the nsp13 helicase (8) and the nsp10/nsp14 proofreading assembly (9, 10), are also essential for viral replication and are antiviral targets (11–13).

We recently reported the first views of the SARS-CoV-2 RTC in complex with the nsp13 helicase [cryo-electron microscopy structures at a nominal resolution of 3.5 Å; (14)]. The overall architecture of the nsp13-RTC places the nucleic acid binding site of nsp13 directly in the path of the downstream template-strand RNA (t-RNA), and cryo-electron microscopy (cryo-EM) difference maps revealed the 5'-single-stranded t-RNA overhang engaged with nsp13 before entering the RdRp active site (14). The nsp13 helicase translocates on single-stranded nucleic acid in the 5'->3' direction (15–22). Thus, this structural arrangement presents a conundrum: The RdRp translocates in the 3'->5' direction on the t-RNA strand, while nsp13 translocates on the same strand in the opposite direction. Translocation of each enzyme opposes each other, and if the helicase prevails it is expected to push the RdRp backward on the t-RNA (14). This

112 reversible backward sliding, termed backtracking, is a well-studied feature of the
113 cellular DNA-dependent RNA polymerases [DdRps; (23–30)].

114 Backtracking by the cellular DdRps plays important roles in transcription
115 regulation, including the control of DdRp pausing during transcription elongation,
116 termination, DNA repair, and transcription fidelity (25). In backtracking, the DdRp
117 and associated transcription bubble move backwards on the DNA, while the RNA
118 transcript reverse threads through the complex to maintain the register of the
119 RNA-DNA hybrid (23–30). This movement generates a single-stranded 3'-
120 segment of the RNA transcript which is extruded out the secondary or NTP-entry
121 tunnel that branches off from the primary DdRp active-site cleft around the
122 conserved bridge helix (27–31).

123 Although evolutionarily unrelated to the DdRps, a secondary channel, formed
124 by the RdRp motif F β-hairpin loop and proposed to serve as an NTP-entry
125 tunnel, branches off from the main SARS-CoV-2 RdRp active site channel (32).
126 This NTP-entry tunnel is well positioned to receive the single-stranded 3'-
127 segment of backtracked RNA, a structural architecture analogous to the DdRps
128 (14). We envisaged that translocation by the helicase could mediate backtracking
129 of the RdRp, an otherwise energetically unfavorable process, enabling the key
130 viral functions such as proofreading (9, 10, 12, 33) and template switching during
131 subgenomic transcription (7, 34). Here we outline the structural basis for SARS-
132 CoV-2 RTC backtracking and describe the role of nsp13 in stimulating
133 backtracking.

134

135

Results

SARS-CoV-2 RdRp backtracked complexes for cryo-electron microscopy. Previously, DdRp backtracked complexes (BTCs) were generated for structural studies by direct incubation of the DdRp with DNA-RNA scaffolds containing mismatched nucleotides at the RNA 3'-end (27, 28, 30); these BTC-scaffolds bind with the downstream Watson-Crick base pairs of the RNA-DNA hybrid positioned in the DdRp active site and the single-stranded 3'-segment of mismatched RNA extruding out the DdRp NTP-entry tunnel. To study RdRp BTCs, we therefore designed and tested RNA scaffolds based on our original SARS-CoV-2 RTC-scaffold but with three or five mismatched cytosine nucleotides added to the product-RNA (p-RNA) 3'-end (BTC_3- and BTC_5-scaffolds; Fig. 1A).

Native electrophoretic mobility shift assays revealed that although the holo-RdRp (nsp7/$nsp8_2$/nsp12) bound the RTC-scaffold as observed previously [Fig. 1B, lane 1; SI Appendix, Fig. S1A; (14)], nsp13 was required for efficient binding to the BTC-scaffolds (Fig. 1B). Stable nsp13-holo-RdRp complexes with BTC-scaffolds were also observed by native mass-spectrometry (Fig. S1B and C).

To determine the structural organization of the SARS-CoV-2 BTC, we assembled nsp13(ADP-AlF_3) and holo-RdRp with the BTC_5-scaffold (Fig. 1A; hereafter called BTC_5) and analyzed the samples by single-particle cryo-EM. The sample comprised two major classes: $nsp13_1$-BTC_5 (3.4 Å nominal resolution), and $nsp13_2$-BTC_5 (3.6 Å; Fig. 1C; SI Appendix, Fig. S2 and S3). To eliminate structural heterogeneity in the nsp13 subunits and obtain a higher-resolution view

of the BTC, the particles from both classes were combined and locally refined inside a mask applied around the holo-RdRp and RNA (excluding the nsp13 subunits), leading to the BTC_5(local) combined map (3.2 Å; Fig. 1C; SI Appendix, Fig. S2 and S3, Table S1).

The cryo-EM maps (Fig. 1C and 2) revealed two significant differences with the nsp13-RTC structures (14): 1) The single-stranded downstream template-RNA (t-RNA) engaged with nsp13.1 was resolved (Fig. 2A), and 2) a single-stranded p-RNA 3'-segment was extruded into the RdRp NTP-entry tunnel (Fig. 2B).

Nsp13 binds the downstream single-stranded t-RNA. In the $nsp13_1$-BTC_5 and $nsp13_2$-BTC_5 cryo-EM maps, the single-stranded 5'-segment of the t-RNA was engaged with nsp13.1. This region of the cryo-EM density was well-resolved (Fig. 2A), allowing identification of the t-RNA segment engaged within the helicase as +14 to +8 (numbering defined in Figure 1A), $^{5'}$CCCAUGU$^{3'}$. The five-nucleotide segment connecting the t-RNA between the helicase and the RdRp (+7 to +3) was disordered and not modeled.

The SARS-CoV-2 RdRp NTP-entry tunnel accommodates the backtracked RNA. The cryo-EM maps also resolved a single-stranded p-RNA 3'-segment of the BTC_5-scaffold extruding into the RdRp NTP-entry tunnel (Fig. 2B), confirming the formation of a BTC (Fig. 3A). The overall architecture of the SARS-CoV-2

BTC is analogous to DdRp BTCs [Fig. 3; (14)]. The DdRp bridge helix [BH; (35)] separates the DdRp active site cleft into a channel for the downstream template DNA (over the top of the BH; Fig. 3B) and the NTP-entry tunnel (underneath the BH; Fig. 3B). Similarly, the viral RdRp motif F [SI Appendix, Fig. S4A; (32)] serves as the strand separating structural element for the backtracked RNA (Fig. 3A). The downstream t-RNA passes over the top of motif F, while the backtracked RNA extrudes out the NTP-entry tunnel underneath motif F (Fig. 3A).

The RdRp NTP-entry tunnel provides a steric and electrostatic environment conducive to channeling the backtracked RNA out of the active site without specific polar protein-RNA interactions that could hinder the RNA movement (Fig. 3C and 4). Comparing the electrostatic surface potential of the NTP-entry tunnels of the SARS-CoV-2 RdRp with eukaryotic and bacterial DdRps reveals a similar overall electrostatic surface environment that may facilitate backtracked RNA entry (Fig. 3C; SI Appendix, Fig. S4B), including a 'track' of conserved positively-charged Arg and Lys residues of motif F (SARS-CoV-2 nsp12 K545, K551, R553, and R555; Fig. 4; SI Appendix, Fig. S4A). Conserved residues of RdRp motifs C and E complete the active-site/NTP-entry tunnel environment surrounding the backtracked RNA (Fig. 4; SI Appendix, Fig. S4A).

In the nsp13-RTCs, the RTC-scaffold (Fig. 1A) is bound in a post-translocated state (14); the 3' p-RNA A is base-paired to the t-RNA U at the -1 site near the catalytic nsp12-D760 (Fig. 5A). The next t-RNA base (A at +1) is positioned to receive the incoming nucleoside-triphosphate (NTP) substrate, but

the site for the incoming NTP substrate is empty (Fig. 5A). By contrast, the BTC structures were translocated by one base pair compared to the RTCs; the base pair corresponding to the A-U Watson-Crick base pair at the 3'-end of the p-RNA (located in the -1 site of the RTCs) was in the -2 position of the BTCs (Fig. 1A, 4, and 5B). The -1 position of the BTC was occupied by the first C-A mismatch; the p-RNA -1C made a non-Watson-Crick hydrogen bond with the opposing t-RNA A (Fig. 4 and 5B). The next three mismatched p-RNA nucleotides (+1C, +2C, +3C) trailed into the NTP entry tunnel (Fig. 4 and 5B). The 3'-nucleotide of the BTC_5-scaffold p-RNA (+4C; Fig. 1A) was solvent-exposed at the outward-facing end of the NTP-entry tunnel, lacked density and was therefore not modelled (Fig. 2B). The trajectory of the backtracked nucleotides at positions +1/+2 was sharply bent due to spatial constraints of motif F residues (Fig. 4A).

Nsp13 stimulates backtracking. The SARS-CoV-2 wild-type holo-RdRp required the nsp13 helicase to bind the BTC-scaffolds efficiently (Fig. 1B). However, we observed that the holo-RdRp containing nsp12 with a single amino acid substitution (D760A) did not require nsp13 to bind the BTC-scaffolds (SI Appendix; Fig. S1A, lane 4). Nsp12-D760 is a conserved residue of the RdRp motif C that chelates a crucial Mg^{2+} ion in catalytic complexes [SI Appendix; Fig. S4A; (32)], but in RdRp structures lacking substrate (including the BTC structures), the Mg^{2+} ions are absent (14, 36, 37). The catalytic Asp residues of the DdRps typically chelate the Mg^{2+} ion even in the absence of substrate (31, 38), and this Mg^{2+} is retained in DdRp backtracked structures (27–30). Our RdRp

227 BTC structures suggest that in the absence of a Mg^{2+} ion, D760 presents an
228 electrostatic barrier to the phosphate backbone of the backtracked RNA (Fig.
229 5B), explaining the requirement for the helicase to surmount this barrier and why
230 removal of D760 stabilizes binding to the BTC-scaffolds.

231 To generate the SARS-CoV-2 BTCs for structural studies, we used the
232 BTC_5-scaffold with five mismatched C's at the p-RNA 3'-end (Fig. 1A). To study
233 the formation of SARS-CoV-2 BTCs from an RTC-scaffold (fully Watson-Crick
234 base paired p-RNA 3'-end), we analyzed UV-induced crosslinking from 4-thio-U
235 incorporated penultimate to the p-RNA 3'-end [RTC(4-thio-U)-scaffold; Si
236 Appendix, Fig. S5A; (39)]. Crosslinking was absolutely dependent on the
237 presence of 4-thio-U in the RNA, establishing specificity (SI Appendix; Fig. S5B).
238 RTCs assembled with wild-type nsp12 and the RTC(4-thio-U)-scaffold gave little
239 to no protein-RNA crosslinking upon UV exposure (SI Appendix; Fig. S5A, lane
240 2). These conditions favor a post-translocated RTC (14, 36, 37) with the 4-thio-U
241 sequestered in the RNA-RNA hybrid and thus not available for protein-RNA
242 crosslinking. Crosslinking of the p-RNA to nsp12 was substantially increased by
243 the addition of nsp13 (with 2 mM ATP, which is present in all the lanes;
244 SI Appendix; Fig. S5A, lane 3). Under these conditions, we propose that the
245 translocation activity of nsp13 backtracked a fraction of the complexes, freeing
246 the 4-thio-U from the RNA-RNA hybrid for crosslinking to nsp12. Replacing wild-
247 type nsp12 with nsp12-D760A (nsp12*; SI Appendix, Fig. S5A, lanes 6-8), which
248 is more prone to backtracking (SI Appendix; Fig. S1A), increased the protein-
249 RNA crosslinking under all conditions, with the maximal crosslinking occurring

under the conditions expected to favor backtracking the most (SI Appendix; Fig. S5A, lane 7). These results affirm the view that nsp13 facilitates backtracking of the SARS-CoV-2 RdRp.

A mismatched nucleotide at the p-RNA 3'-end spontaneously frays and enters into the RdRp NTP-entry tunnel. The SARS-CoV-2 RTC is a highly processive and rapid replicase/transciptase, capable of replicating an ~1 kb RNA template at an average rate of ~170 nt/s (40). However, studies of other viral RdRps suggest that misincorporation slows the overall elongation rate and may induce backtracking (41–43). We used molecular dynamics simulations to explore the fate of a mismatched nucleotide incorporated at the p-RNA 3'-end. Starting with the nsp13$_2$-BTC$_5$ structure, the -1C was mutated to U, and the +2 to +4 C's were removed. The resulting pre-translocated p-RNA had a matched -1U and a mismatched +1C (-1U+1C; Fig. 5C). In three 5 μs simulations we observed the 3'-mismatched +1C alternating between two positions, either remaining in the vicinity of the active site (RMSD < 3.5 Å) or fraying away from the p-RNA:t-RNA hybrid towards or into the NTP-entry tunnel (RMSD > 3.5 Å; Fig. 5C). Based on analysis of the aggregated -1U+1C simulations, the mismatched +1C spent about 40% of the time near the active site and about 60% of the time frayed towards or in the NTP-entry tunnel. In control simulations with a fully matched p-RNA 3'-end (-1U+1U), the matched +1U at the p-RNA 3'-end did not fray and spent 100% of the time in the active site pocket (SI Appendix; Fig. S6).

Discussion

Our results establish that the SARS-CoV-2 RTC backtracks, that backtracking is facilitated by the nsp13 helicase, and that the resulting single-stranded 3'-segment of the p-RNA extrudes out the RdRp NTP-entry tunnel in a manner analogous to the evolutionarily unrelated cellular DdRps (Fig. 3). Thus, a secondary tunnel to accommodate backtracked RNA, facilitating fidelity and possibly other functions (Fig. 6), appears to be a crucial feature of transcriptase enzymes that evolved independently.

Backtracking of Φ6 and poliovirus RdRps has been reported based on analysis of single-molecule observations (41–43). The nsp13 helicase facilitates efficient backtracking of the SARS-CoV-2 RTC (SI Appendix; Fig. S5). We note that in bacteria, the UvrD helicase has been shown to induce DdRp backtracking, suggesting that a role for helicases in backtracking may be widespread (44).

Our results are consistent with the view that a matched nucleotide at the pre-translocated p-RNA 3'-end remains base paired to the t-RNA (Fig. 5; SI Appendix, Fig. S6), facilitating translocation and subsequent NTP addition and thus rapid elongation (at a maximum elongation rate of ~170 nt/s, a translocation event would occur approximately every 6 msec, on average, explaining why translocation was not observed in our 5 μs simulations; Fig. 5; SI Appendix, Fig. S6). However, upon misincorporation, the pre-translocated, mismatched nucleotide at the p-RNA 3'-end spends more than half the time frayed from the t-RNA and towards or in the NTP-entry tunnel (Fig. 5C), a configuration that is likely recalcitrant to translocation and subsequent elongation. The favorable

296 environment of the NTP-entry tunnel (Fig. 3 and 4) may further encourage
297 backtracking. The resulting inhibition of translocation may enable the tight
298 engagement of the nsp13.1 helicase with the downstream single-stranded t-RNA
299 (Fig. 2A), allowing the 5'->3' translocation activity of the helicase to more robustly
300 backtrack the complex (SI Appendix; Fig. S5).

301 Our findings have implications for the processes of subgenomic transcription
302 and proofreading in SARS-CoV-2 [Fig. 6; (14)]. Generation of mRNAs for the
303 viral structural proteins begins with transcription initiation at the 3'-poly(A) tail of
304 the (+)-strand RNA genome. The process, called sub-genomic transcription,
305 ultimately generates a nested set of transcripts that are both 5'- and 3'-co-
306 terminal with the viral genome and involves a remarkable template-switch from
307 the 3'-portion of the genome to the 5'-leader (7, 34). The template-switching
308 event is thought to involve base-pairing between the 3'-end of the nascent
309 transcript and a complementary sequence (the Transcription Regulatory
310 Sequence, or TRS) near the (+)-strand 5'-leader (45). Backtracking could extrude
311 the 3'-end of the nascent transcript out the NTP-entry tunnel, making it available
312 for base pairing to the 5'-TRS (Fig. 6). Our results establishing that the SARS-
313 CoV-2 RTC can backtrack validates a key prediction of this model for the
314 mechanism of template-switching during sub-genomic transcription (14).

315 Nucleotide analogs that function by being incorporated into product RNA by
316 viral RdRps are important antiviral therapeutics (46). Notably, their incorporation
317 may induce backtracking by the RdRp (43). Rdv, a nucleotide analog, is the only
318 FDA-approved drug for COVID-19 treatment (5). Our results support a model in

which RdRp misincorporation or incorporation of nucleotide analogs can pause the RdRp, allowing nsp13 to engage with the downstream single-stranded t-RNA to induce backtracking (14). The resulting exposure of the p-RNA 3'-end out the NTP-entry tunnel (Fig. 3A and 6) could provide access for the SARS-CoV-2 proofreading machinery [nsp10/14; (9, 12)] to degrade the p-RNA 3'-end, thus removing the misincorporation or analog. This proofreading activity, which is unique to the nidovirus order to which CoVs belong (10), is a major determinant for the resistance of CoVs against many nucleotide analog inhibitors (13). Thus, understanding RdRp backtracking and its potential role in CoV proofreading can facilitate the development of therapeutics.

Materials and Methods

Detailed descriptions of SARS-CoV-2 nsp12, 7, 8, and 13 protein purification, assembly of the RTC complexes, Native EMSAs, native mass-spectrometry, cross-linking, specimen preparation for cryo-EM, cryo-EM data acquisition and processing, model building and refinement, and molecular dynamics simulations are provided in the SI Appendix.

Acknowledgments

We thank M. Ebrahim, J. Sotiris, and H. Ng at The Rockefeller University Evelyn Gruss Lipper Cryo-electron Microscopy Resource Center for help with cryo-EM. Some of the work reported here was conducted at the Simons Electron Microscopy Center (SEMC) and the National Resource for Automated Molecular Microscopy (NRAMM) and National Center for CryoEM Access and Training (NCCAT) located at the NYSBC, supported by grants from the NIH National Institute of General Medical Sciences (P41 GM103310), NYSTAR, the Simons Foundation (SF349247), the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM129539) and NY State Assembly Majority. This work was supported by the Pels Family Center for Biochemistry and Structural Biology (The Rockefeller University), and NIH grants P41 GM109824 and P41 GM103314 to B.T.C, R35 GM118130 to S.A.D, and R01 GM114450 to E.A.C.

References

1. F. Wu, *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

2. P. Zhou, *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

3. L. Subissi, *et al.*, One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities.

Proceedings of the National Academy of Sciences of the United States of America **111**, E3900-9 (2014).

4. R. N. Kirchdoerfer, A. B. Ward, Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature Communications* **10**, 2342–9 (2019).

5. U. S. F. and D. Administration, Remdesivir-EUA-Letter-Of-Authorization.pdf. *undefined* (2020).

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

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

8. K. C. Lehmann, E. J. Snijder, C. C. Posthuma, A. E. Gorbalenya, What we know but do not understand about nidovirus helicases. *Virus Res* **202**, 12–32 (2015).

9. E. Minskaia, *et al.*, Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc National Acad Sci* **103**, 5108–5113 (2006).

10. A. E. Gorbalenya, L. Enjuanes, J. Ziebuhr, E. J. Snijder, Nidovirales: Evolving the largest RNA virus genome. *Virus Res* **117**, 17–37 (2006).

11. R. Zhang, *et al.*, The nsp1, nsp13, and M proteins contribute to the hepatotropism of murine coronavirus JHM.WU. *J Virol* **89**, 3598–609 (2015).

12. M. R. Denison, R. L. Graham, E. F. Donaldson, L. D. Eckerle, R. S. Baric, Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity. *Rna Biol* **8**, 270–279 (2011).

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

14. J. Chen, *et al.*, Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **182**, 1560-1573.e13 (2020).

15. A. O. Adedeji, *et al.*, Mechanism of Nucleic Acid Unwinding by SARS-CoV Helicase. *Plos One* **7**, e36521 (2012).

16. E. M. Bautista, K. S. Faaberg, D. Mickelson, E. D. McGruder, Functional Properties of the Predicted Helicase of Porcine Reproductive and Respiratory Syndrome Virus. *Virology* **298**, 258–270 (2002).

17. K. A. Ivanov, J. Ziebuhr, Human Coronavirus 229E Nonstructural Protein 13: Characterization of Duplex-Unwinding, Nucleoside Triphosphatase, and RNA 5′-Triphosphatase Activities. *J Virol* **78**, 7833–7838 (2004).

18. N.-R. Lee, *et al.*, Cooperative translocation enhances the unwinding of duplex DNA by SARS coronavirus helicase nsP13. *Nucleic Acids Res* **38**, 7626–36 (2010).

19. K. J. Mickolajczyk, *et al.*, Force-dependent stimulation of RNA unwinding by SARS-CoV-2 nsp13 helicase. *Biophys J* (2020) https:/doi.org/10.1016/j.bpj.2020.11.2276.

20. A. Seybert, L. C. van Dinten, E. J. Snijder, J. Ziebuhr, Biochemical Characterization of the Equine Arteritis Virus Helicase Suggests a Close Functional Relationship between Arterivirus and Coronavirus Helicases. *J Virol* **74**, 9586–9593 (2000).

21. A. SEYBERT, A. HEGYI, S. G. SIDDELL, J. ZIEBUHR, The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5′-to-3′ polarity. *Rna* **6**, 1056–1068 (2000).

22. J. A. Tanner, *et al.*, The Severe Acute Respiratory Syndrome (SARS) Coronavirus NTPase/Helicase Belongs to a Distinct Class of 5′ to 3′ Viral Helicases. *J Biol Chem* **278**, 39578–39582 (2003).

23. N. Komissarova, M. Kashlev, RNA Polymerase Switches between Inactivated and Activated States By Translocating Back and Forth along the DNA and the RNA. *J Biol Chem* **272**, 15329–15338 (1997).

24. N. Komissarova, M. Kashlev, Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3’ end of the RNA intact and extruded. *Proc National Acad Sci* **94**, 1755–1760 (1997).

25. E. Nudler, RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**, 1438–1445 (2012).

26. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**, 33–41 (1997).

27. D. Wang, *et al.*, Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203–1206 (2009).

426 28. S. Sekine, Y. Murayama, V. Svetlov, E. Nudler, S. Yokoyama, The ratcheted
427 and ratchetable structural states of RNA polymerase underlie multiple
428 transcriptional functions. *Molecular Cell* **57**, 408–421 (2015).

429 29. A. C. M. Cheung, P. Cramer, Structural basis of RNA polymerase II
430 backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2012).

431 30. M. Abdelkareem, *et al.*, Structural Basis of Transcription: RNA Polymerase
432 Backtracking and Its Reactivation. *Mol Cell* **75**, 298-309.e4 (2019).

433 31. G. Zhang, *et al.*, Crystal structure of Thermus aquaticus core RNA
434 polymerase at 3.3 A resolution. *Cell* **98**, 811–824 (1999).

435 32. A. J. W. te Velthuis, Common and unique features of viral RNA-dependent
436 polymerases. *Cell Mol Life Sci Cmls* **71**, 4403–20 (2014).

437 33. M. L. Agostini, *et al.*, Coronavirus Susceptibility to the Antiviral Remdesivir
438 (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading
439 Exoribonuclease. *Mbio* **9**, e00221-18 (2018).

440 34. S. G. Sawicki, D. L. Sawicki, Advances in Experimental Medicine and
441 Biology. *Adv Exp Med Biol* **440**, 215–219 (1998).

442 35. W. J. Lane, S. A. Darst, Molecular Evolution of Multisubunit RNA
443 Polymerases: Structural Analysis. *Journal of Molecular Biology* **395**, 686–704
444 (2010).

445 36. Q. Wang, *et al.*, Structural Basis for RNA Replication by the SARS-CoV-2
446 Polymerase. *Cell* (2020) https:/doi.org/10.1016/j.cell.2020.05.034.

447 37. H. S. Hillen, *et al.*, Structure of replicating SARS-CoV-2 polymerase. *Nature*,
448 1–6 (2020).

449 38. P. Cramer, *et al.*, Architecture of RNA polymerase II and implications for the
450 transcription mechanism. *Science* **288**, 640–649 (2000).

451 39. E. J. Sontheimer, Site-specific RNA crosslinking with 4-thiouridine. *Mol Biol*
452 *Rep* **20**, 35–44 (1994).

453 40. M. Seifert, *et al.*, Signatures and mechanisms of efficacious therapeutic
454 ribonucleotides against SARS-CoV-2 revealed by analysis of its replicase using
455 magnetic tweezers. *Biorxiv*, 2020.08.06.240325 (2020).

456 41. D. Dulin, *et al.*, Backtracking behavior in viral RNA-dependent RNA
457 polymerase provides the basis for a second initiation site. *Nucleic Acids Res* **43**,
458 10421–9 (2015).

42. D. Dulin, *et al.*, Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. *Cell Reports* **10**, 983–992 (2015).

43. D. Dulin, *et al.*, Signatures of Nucleotide Analog Incorporation by an RNA-Dependent RNA Polymerase Revealed Using High-Throughput Magnetic Tweezers. *Cell Reports* **21**, 1063–1076 (2017).

44. V. Epshtein, *et al.*, UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372–377 (2014).

45. A. O. Pasternak, E. van den Born, W. J. M. Spaan, E. J. Snijder, Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *Embo J* **20**, 7220–7228 (2001).

46. E. D. Clercq, G. Li, Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* **29**, 695–747 (2016).

47. G. Cardone, J. B. Heymann, A. C. Steven, One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *Journal of structural biology* **184**, 226–236 (2013).

48. E. Jurrus, *et al.*, Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112–128 (2018).

49. N. M. Luscombe, R. A. Laskowski, J. M. Thornton, NUCPLOT: A Program to Generate Schematic Diagrams of Protein-Nucleic Acid Interactions. *Nucleic Acids Res* **25**, 4940–4945 (1997).

50. E. C. Smith, N. R. Sexton, M. R. Denison, Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. *Ann Rev Virol* **1**, 111–132 (2014).

Figures

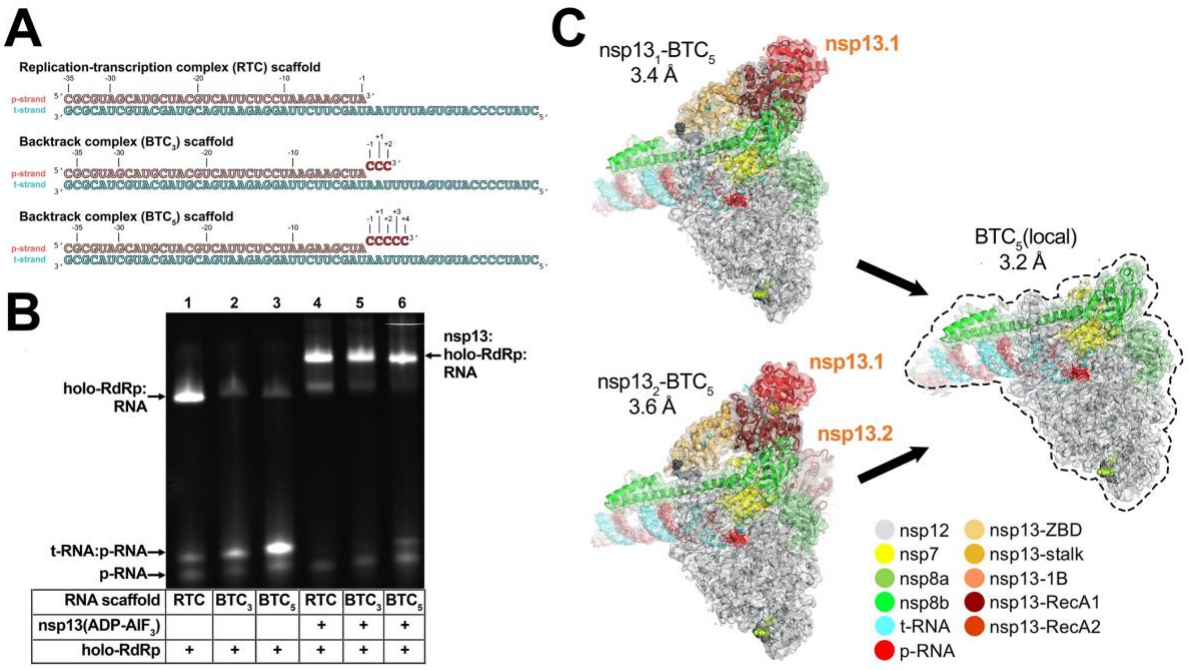


Figure 1

Fig. 1. SARS-CoV-2 backtrack complex.

A. RNA scaffolds: (*top*) replication-transcription complex (RTC) scaffold (14); (*bottom*) backtrack complex scaffolds (BTC_3 and BTC_5).

B. A native gel electrophoretic mobility shift assay reveals that holo-RdRp requires nsp13(ADP-AlF_3) to bind the BTC scaffolds efficiently.

C. Cryo-EM structures of SARS-CoV-2 BTCs. Shown is the transparent cryo-EM density [local-resolution filtered; (47)] with the refined models superimposed (Table S1). The models and density are colored according to the key.

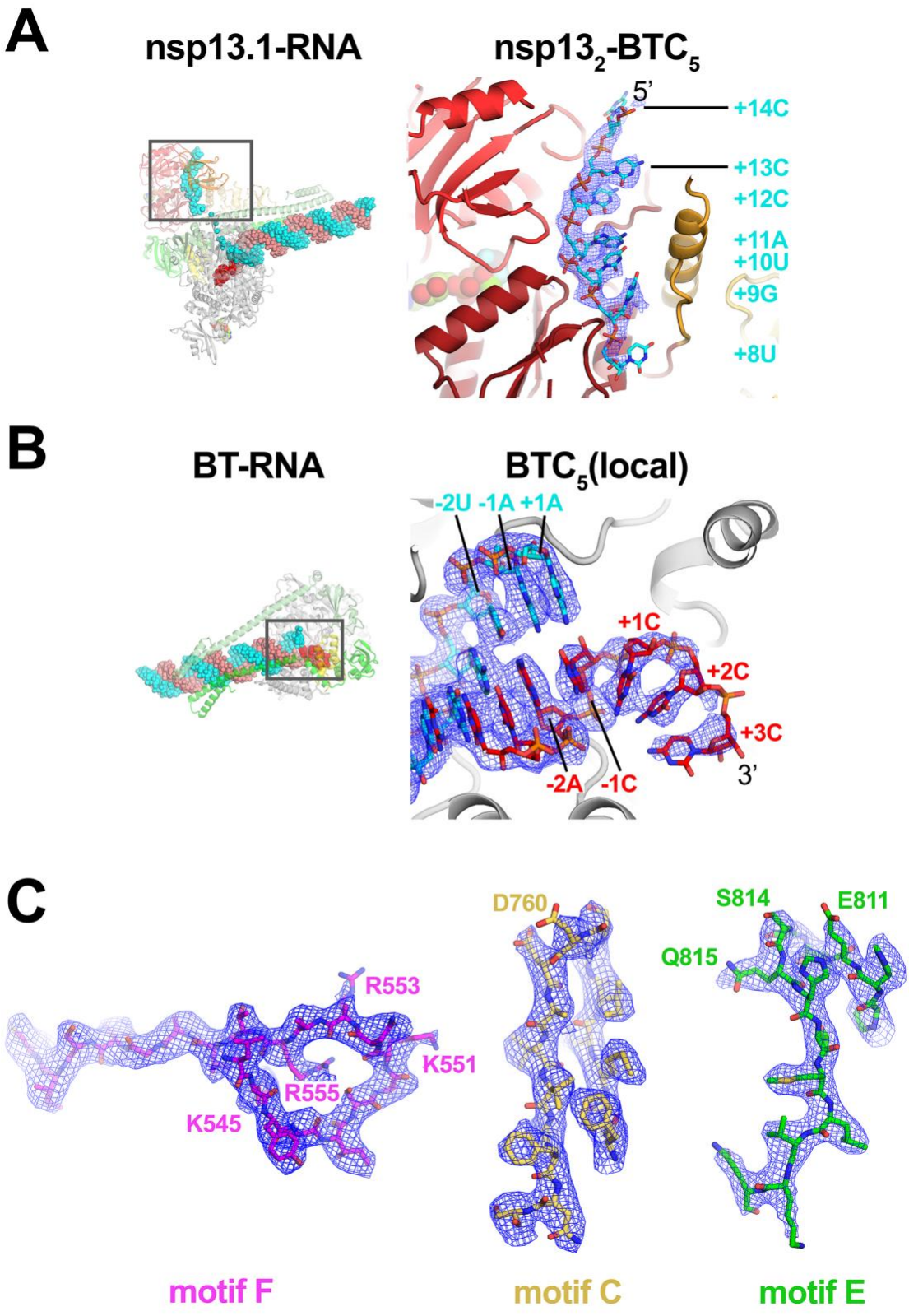


Figure 2

496

Fig. 2. Cryo-EM density maps.

A. (*left*) Overall view of nsp13$_1$-BTC$_5$. The boxed region is magnified on the right. (*right*) Magnified view of the t-RNA segment (+14-5'-CCCAUGU-3'-+8) enclosed in the nsp13.1 helicase subunit. The cryo-EM density map (from the nsp13$_2$-BTC structure) for the RNA is shown (blue mesh).

B. (*left*) Overall view of the BTC structure. The boxed region is magnified on the right. (*right*) Magnified view of the region around the RdRp active site, showing the t-RNA (cyan) and p-RNA (red) with the backtracked RNA segment. The cryo-EM density map for the RNA [from BTC$_5$(local)] is shown (blue mesh).

C. BTC$_5$(local) cryo-EM density maps around nsp12 conserved motifs F, C, and E. Selected residues are labeled.

511

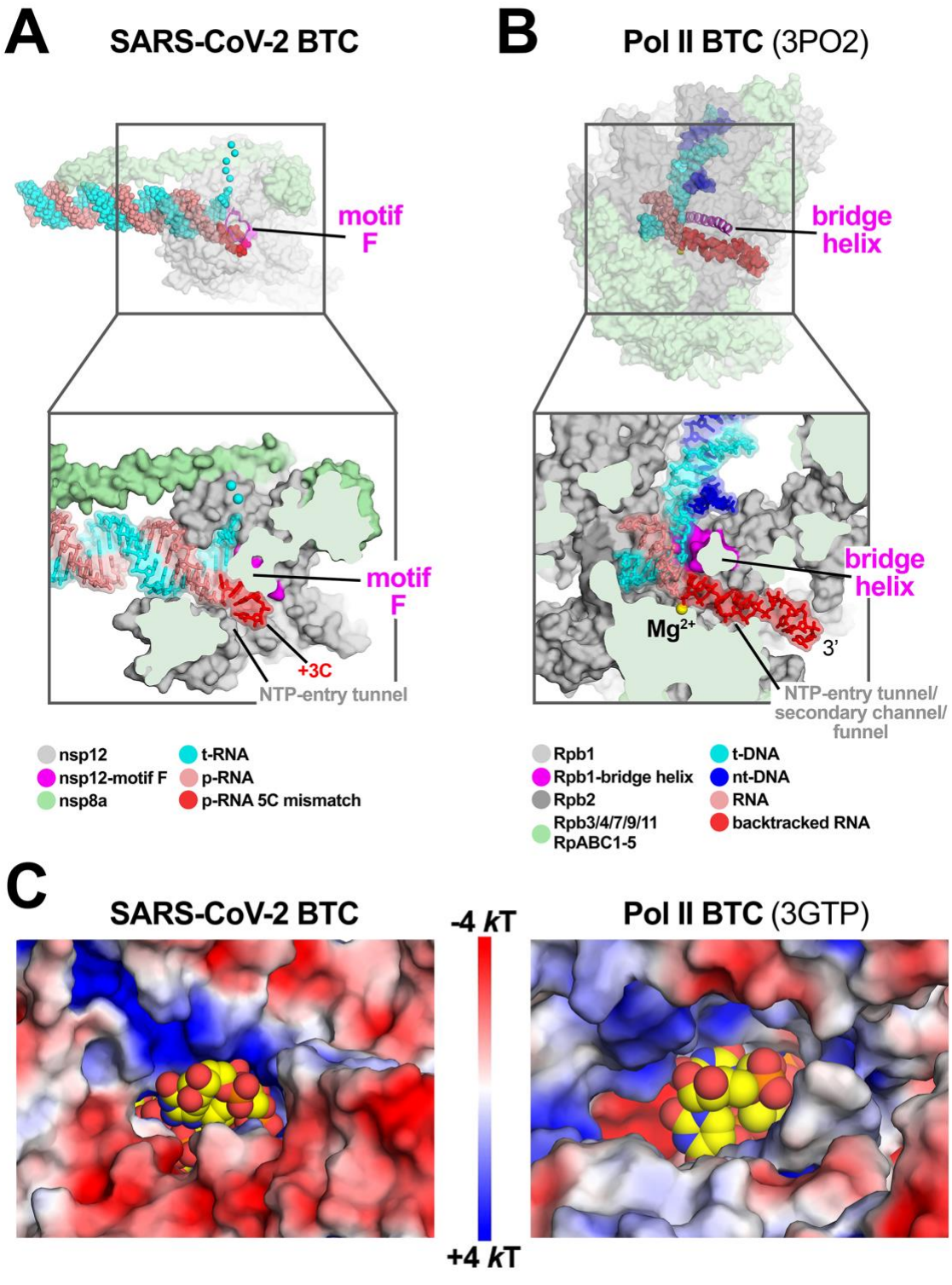


Figure 3

512

Fig. 3. SARS-CoV-2 RdRp and DdRp BTCs.

A, B. SARS-CoV-2 RdRp (**A**) and DdRp (**B**) BTCs.

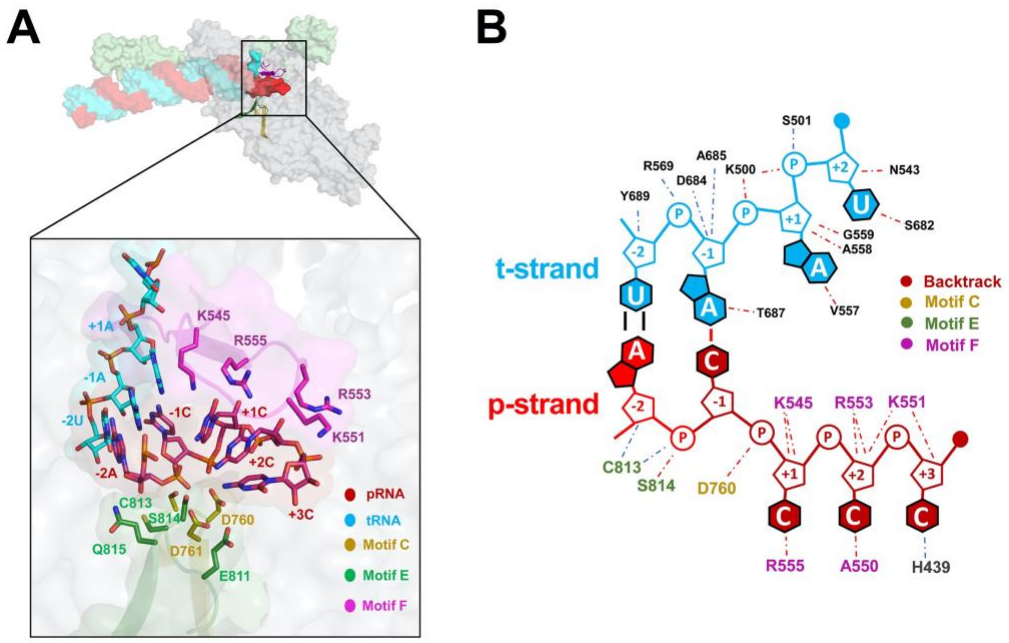
(*top*) Proteins are shown as transparent molecular surfaces, nucleic acids as atomic spheres. The boxed regions are magnified on the bottom.

(*bottom*) Magnified, cross-sectional view. Proteins are shown as molecular surfaces, nucleic acids in stick format with transparent molecular surface.

A. The SARS-CoV-2 BTC_5(local). Nsp8a and nsp12 are shown (nsp7 and nsp8b are removed for clarity). Nsp12 motif F is shown as a magenta backbone ribbon (*top*). Backtracked RNA (+1C to +3C of the BTC_5-scaffold; Figure 1A) extrude out the NTP-entry tunnel.

B. A DdRp (*Saccharomyces cerevisiae* Pol II) BTC [PDB ID: 3PO2; (29)]. The bridge helix is shown as a magenta backbone ribbon. The backtracked RNA extrudes out the NTP-entry tunnel/secondary channel/funnel.

C. Views from the outside into the NTP-entry tunnels of the SARS-CoV-2 (*left*) and an *S. cerevisiae* DdRp [PDB ID: 3GTP; (27) BTC. Protein surfaces are colored by the electrostatic surface potential [calculated using APBS; (48)]. Backtracked RNA is shown as atomic spheres with yellow carbon atoms.



532 **Figure 4**

533 **Fig. 4. Protein-RNA interactions in the BTC.**

534 **A.** (*top*) Overall view of BTC_5(local). Proteins are shown as transparent molecular
535 surfaces, nucleic acids as atomic spheres. Nsp8a and nsp12 are shown (nsp7
536 and nsp8b are removed for clarity). Nsp12 motifs C, E, and F are shown as
537 backbone ribbons (colored according to the key on the bottom. The boxed region
538 is magnified below.

539 (*bottom*) RNA is shown from -2 to +3. Proteins are shown as transparent
540 molecular surfaces. RdRp motifs C, E, and F are shown as transparent backbone
541 ribbons (colored according to the key) with side chains of residues that approach
542 the backtracked RNA (≤ 4.5 Å) shown.

543 **B.** Schematic illustrating the same protein-RNA interactions as (**A**). Drawn using
544 Nucplot (49).

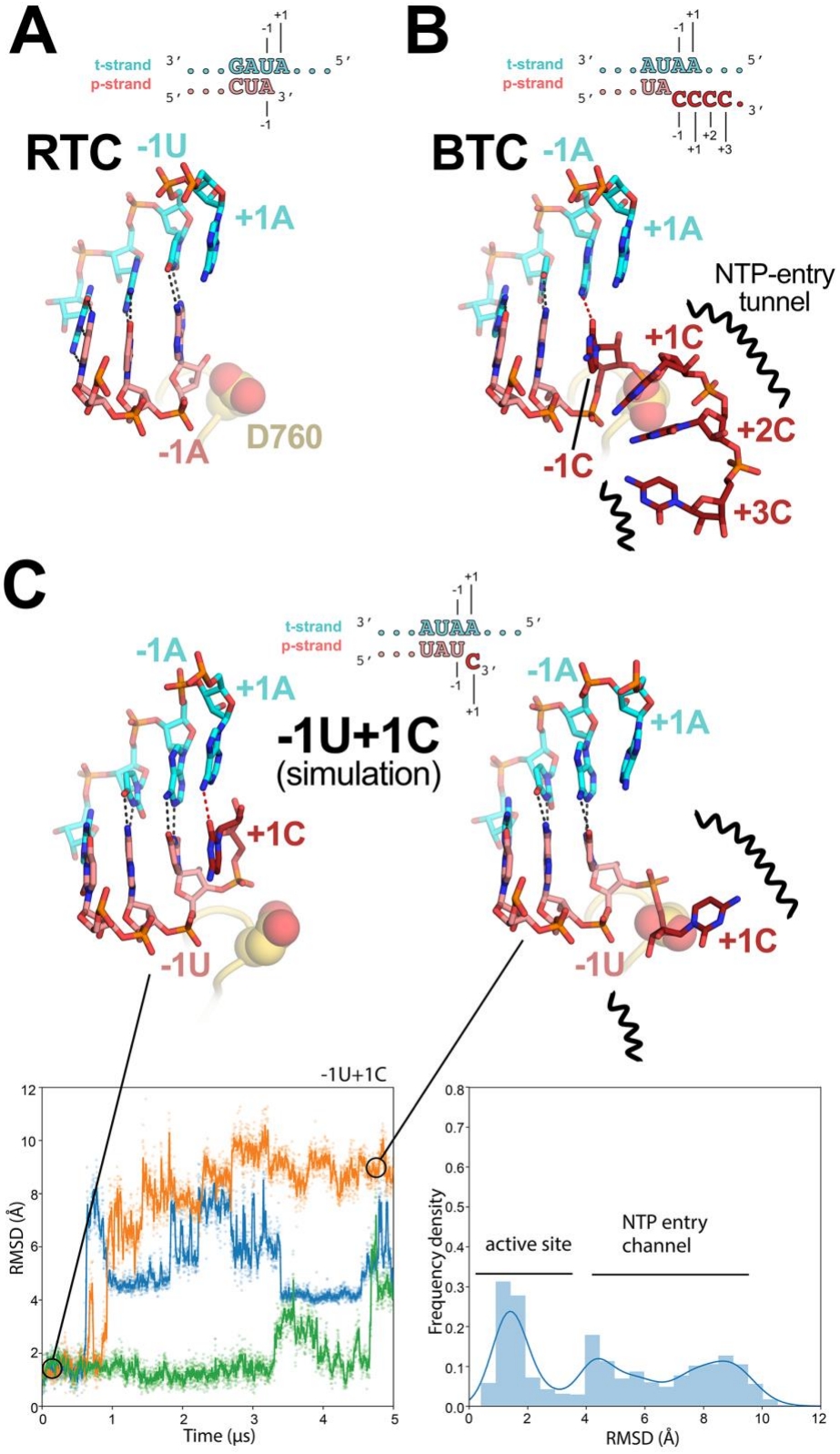


Figure 5

545

Fig. 5. Comparison of active-site proximal RNA in the RTC and BTC structures, and from simulations of a mismatched nucleotide at the p-RNA 3'-end.

A-B, Comparison of the active-site proximal RNA in the RTC [**A**; PDB ID: 6XEZ; (14)], BTC_5(local) (**B**), and from selected snapshots of molecular dynamics simulations of a -1U+1C complex (**C**). The schematics denote the nucleotides shown in the context of the RTC- (**A**) and BTC_5-scaffolds (**B**; full scaffold sequences shown in Figure 1A) or generated from the BTC_5-scaffold for the simulations (**C**). Carbon atoms of the t-RNA are colored cyan, p-RNA are colored salmon except in the case of mismatched C's at the 3'-end, which are colored dark red. Watson-Crick base pairing hydrogen-bonds are denoted as dark gray dashed lines, other hydrogen-bonds as red dashed lines. Nsp12 motif C is shown as a yellow-orange backbone ribbon, and the side-chain of D760 is shown as atomic spheres.

A. The RTC is in a post-translocated state, with the A-U base pair at the p-RNA 3'-end in the -1 position (14).

B. The BTC_5(local) RNA is translocated compared to the RTC; the base pair corresponding to A-U at the 3'-end of the RTC RNA in the -1 position is in the -2 position of the BTC RNA. A C-A mismatch occupies the BTC -1 site. The +1, +2 and +3 mismatched C's trail into the RdRp NTP-entry tunnel (denoted by black squiggly lines). The +4C (present in the BTC_5-scaffold; Figure 1A) is exposed to solvent, disordered and not modelled.

568 **C.** Molecular dynamics simulations of the nsp13$_2$-BTC$_{-1U+1C}$ complex. The
569 complex was simulated with 3 replicates. RMSD values plotted as a function of
570 time represent the heavy-atom RMSD of the +1C of the p-RNA compared with
571 the starting configuration (see Methods). The RMSD histograms (plotted on the
572 right) are an aggregate of all three replicates. Two structures taken from one of
573 the simulations are shown, one showing the +1C of the p-RNA in the active site
574 (t = 0 μs) and the other showing the +1C frayed into the NTP-entry tunnel
575 (t = 4.5 μs).
576

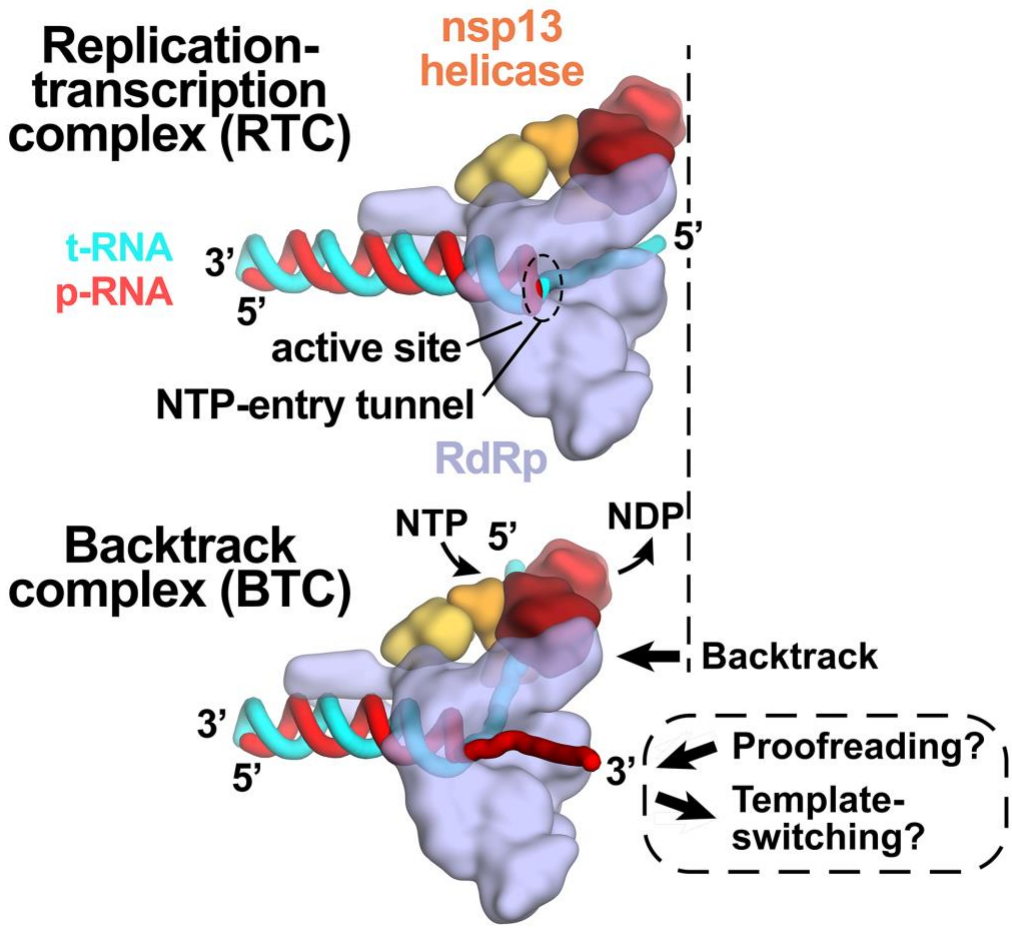


Figure 6

577

578

Fig. 6. Role of backtracking in proofreading and template-switching during sub-genomic transcription.

Schematic illustrating the proposed model for backtracking of the SARS-CoV-2 RTC and its potential role in proofreading and template-switching during sub-genomic transcription. The structural models are shown as cartoons (holo-RdRp, light blue; nsp13 helicase, orange shades; RNA strands, colored tubes as indicated).

(*top*) In the RTC, the elongating RdRp moves from left-to-right. The RdRp active site holds the p-RNA 3'-end. The NTP-entry tunnel provides access from solution to the RdRp active site. The downstream (5') single-stranded t-RNA is not engaged with nsp13.

(*bottom*) In the BTC, nsp13 translocates on the downstream (5') single-stranded t-RNA, pushing the RdRp backwards (right-to-left) on the RNA. This causes the p-RNA to reverse-thread through the complex, with the resulting single-stranded 3'-fragment extruding out the NTP-entry tunnel. The exposure of the p-RNA 3'-end could facilitate proofreading (9, 10, 12, 50) and also template-switching during sub-genomic transcription (7, 34).

Supplementary Information Text

METHODS

Structural biology software was accessed through the SBGrid consortium (1).

Protein expression and purification.

SARS-CoV-2 nsp12. SARS-CoV-2 nsp12 was expressed and purified as described (1). A pRSFDuet-1 plasmid expressing SARS-CoV-2 His_6-SUMO-nsp12 (Addgene plasmid 159107) was transformed into *Escherichia coli* (*Eco*) BL21-CodonPlus cells (Agilent). Cells were grown, followed by the addition of isopropyl β-d-1-thiogalactopyranoside (IPTG) to induce protein expression overnight. Cells were collected by centrifugation, resuspended and lysed in a continuous-flow French press (Avestin). The lysate was cleared by centrifugation, loaded onto a HiTrap Heparin HP column (Cytiva), and then eluted using a salt gradient. The fractions containing nsp12 were pooled and loaded onto a HisTrap HP column (Cytiva), washed, and eluted. Eluted nsp12 was dialyzed overnight in the presence of His_6-Ulp1 SUMO protease. Cleaved nsp12 was passed through a HisTrap HP column (Cytiva). Flow-through was collected, concentrated by centrifugal filtration (Amicon), and loaded on a Superdex 200 Hiload 16/600 (Cytiva) for size-exclusion chromatography. Glycerol was added to the purified nsp12, aliquoted, flash frozen with liquid N_2, and stored at -80°C.

SARS-CoV-2 nsp7/8. SARS-CoV-2 nsp7/8 was expressed and purified as described (1). The pCDFDuet-1 plasmid expressing SARS-CoV-2 His_6-ppx-nsp7/8 (ppx is a Prescission Protease cleavage site; Addgene plasmid 159092) was transformed into *Eco* BL21(DE3). Cells were grown and protein expression was induced overnight by the addition of IPTG. Cells were collected by centrifugation, resuspended, and lysed in a continuous-flow French press (Avestin). The lysate was cleared by centrifugation, then loaded onto a HisTrap HP column (Cytiva), washed, and eluted. Eluted nsp7/8 was dialyzed overnight in the presence of His_6-Prescission Protease to cleave the His_6-tag. Cleaved nsp7/8 was passed through a HisTrap HP column (Cytiva). Flow-through was collected, concentrated by centrifugal filtration (Amicon), and loaded onto a Superdex 75 Hiload 16/600 (Cytiva). Glycerol was added to the purified nsp7/8, aliquoted, flash frozen with liquid N_2, and stored at -80°C.

630 *SARS-CoV-2 nsp13*. SARS-CoV-2 nsp13 was expressed and purified as
631 described (1). The pet28 plasmid containing SARS-CoV-2 His_6-ppx-nsp13
632 (Addgene plasmid 159390) was transformed into *Eco* Rosetta(DE3) (Novagen).
633 Cells were grown, followed by the addition of IPTG to induce protein expression
634 overnight. Cells were collected by centrifugation, resuspended, and lysed in a
635 continuous-flow French press (Avestin). The lysate was cleared by centrifugation,
636 then loaded onto a HisTrap HP column (Cytiva), washed, and eluted. Eluted
637 nsp13 was dialyzed overnight in the presence of His_6-Prescission Protease to
638 cleave His_6-tag. Cleaved nsp13 was passed through a HisTrap HP column
639 (Cytiva). Flow-through was collected, concentrated by centrifugal filtration
640 (Amicon), and loaded onto a Superdex 200 Hiload 16/600 (Cytiva). Glycerol was
641 added to the purified nsp13, aliquoted, flash frozen with liquid N_2, and stored at -
642 80°C.

643

644 **Native electrophoretic mobility shift assays.** Nsp12 or nsp12-D760A were
645 incubated with 3-fold molar excess of nsp7/8 in transcription buffer (120 mM K-
646 acetate, 20 mM HEPES pH 8, 10 mM $MgCl_2$, 2 mM DTT) to assemble holo-RdRp
647 (2 μM final). The resulting complex was incubated with 1 μM of annealed RNA
648 scaffold (Horizon Discovery) for 5 minutes at 30°C. Nsp13 and pre-mixed ADP
649 and AlF_3 (Sigma-Aldrich) were added to a final concentration of 2 μM and 2 mM,
650 respectively, and incubated for an additional 5 minutes at 30°C. Reactions were
651 analyzed by native gel electrophoresis on a 4.5% polyacrylamide native gel
652 (37.5:1 acrylamide:bis-acrylamide) in 1X TBE (89 mM Tris, 89 mM boric acid,
653 1 mM EDTA) at 4°C. The gel was stained with Gel-Red (Biotium).

654

655 **Native mass spectrometry (nMS) analysis.** The reconstituted sample
656 containing 4 μM RTC and 8 μM nsp13 incubated with 2 mM ADP-AlF_3 was
657 buffer-exchanged into 150 mM ammonium acetate, 0.01% Tween-20, pH 7.5
658 using a Zeba microspin desalting column with a 40 kDa MWCO (ThermoFisher
659 Scientific). For nMS analysis, a 2–3 μL aliquot of the buffer-exchanged sample

was loaded into a gold-coated quartz capillary tip that was prepared in-house and then electrosprayed into an Exactive Plus with extended mass range (EMR) instrument (Thermo Fisher Scientific) with a static direct infusion nanospray source (2). The MS parameters used: spray voltage, 1.2 kV; capillary temperature, 150 °C; in-source dissociation, 0 V; S-lens RF level, 200; resolving power, 17,500 at *m/z* of 200; AGC target, 1 x 10^6; maximum injection time, 200 ms; number of microscans, 5; injection flatapole, 6 V; interflatapole, 4 V; bent flatapole, 4 V; high energy collision dissociation (HCD), 200 V; ultrahigh vacuum pressure, 7.2 × 10^{-10} mbar; total number of scans, at least 100. Mass calibration in positive EMR mode was performed using cesium iodide. For data processing, the acquired MS spectra were visualized using Thermo Xcalibur Qual Browser (v. 4.2.47). MS spectra deconvolution was performed either manually or using the software UniDec v. 4.2.0 (3, 4). The following parameters were used for the UniDec processing: m/z range, 7,000 – 10,000 Th; background subtraction, subtract curved at 100; smooth charge state distribution, enabled; peak shape function, Gaussian; Beta Softmax distribution parameter, 20.

The expected masses for the component proteins based on previous nMS experiments (1) include nsp7: 9,137 Da; nsp8 (N-terminal Met lost): 21,881 Da; nsp13 (post-protease cleavage, has three Zn^{2+} ions coordinated with 9 deprotonated cysteine residues): 67,464 Da, and nsp12 (has two Zn^{2+} ions coordinated with 6 deprotonated cysteine residues): 106,785 Da. The mass of the assembled RNA duplex scaffold is 30,512 Da.

Experimental masses were reported as the average mass ± standard deviation (S.D.) across all the calculated mass values within the observed charge state series. Mass accuracies were calculated as the % difference between the measured and expected masses relative to the expected mass. The observed mass accuracies ranged from 0.016 – 0.035%.

Preparation of SARS-CoV-2 nsp13-BTC_5 for Cryo-EM. Purified nsp12 and nsp7/8 were mixed in a 1:3 molar ratio and incubated at 22^o C for 15 minutes.

690 The mixture was buffer-exchanged into cryo-EM buffer (20 mM HEPES pH 8.0,
691 150 mM K-acetate, 10 mM $MgCl_2$, 1 mM DTT) using Zeba desalting columns
692 (ThermoFisher Scientific) and incubated with annealed BTC_5-scaffold (Fig. 1A) in
693 a 1:1.5 molar ratio. Purified nsp13 was concentrated by centrifugal filtration
694 (Amicon) and buffer exchanged into cryo-EM buffer using Zeba desalting
695 columns. Buffer exchanged nsp13 was mixed with ADP and AlF_3 and then added
696 to nsp7/8/12/RNA scaffold at a molar ratio of 1:1 with a final concentration of
697 2 mM ADP-AlF_3. Complex was incubated for 5 minutes at 30^o C and further
698 concentrated by centrifugal filtration (Amicon).

699

700 **Cryo-EM grid preparation.** Prior to grid freezing, 3-([3-
701 cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO,
702 Anatrace) was added to the sample (8 mM final), resulting in a final complex
703 concentration of 10 µM. The final buffer condition for the cryo-EM sample was
704 20 mM HEPES pH 8.0, 150 mM K-acetate, 10 mM $MgCl_2$, 1 mM DTT,
705 2 mM ADP-AlF_3, 8 mM CHAPSO. C-flat holey carbon grids (CF-1.2/1.3-4Au,
706 Electron Microscopy Sciences) were glow-discharged for 20 s prior to the
707 application of 3.5 µL of sample. Using a Vitrobot Mark IV (ThermoFisher
708 Scientific), grids were blotted and plunge-frozen into liquid ethane with 90%
709 chamber humidity at 4°C.

710

711 **Cryo-EM data acquisition and processing.** Structural biology software was
712 accessed through the SBGrid consortium (5). Grids were imaged using a 300 kV
713 Titan Krios (ThermoFisher Scientific) equipped with a K3 camera (Gatan) and a
714 BioQuantum imaging filter (Gatan). Images were recorded using Leginon (6) with
715 a pixel size of 1.065 Å/px (micrograph dimensions of 5,760 x 4,092 px) over a
716 nominal defocus range of -0.8 µm to -2.5 µm and 30 eV slit width. Movies were
717 recorded in "counting mode" (native K3 camera binning 2) with ~30 e-/px/s in
718 dose-fractionation mode with subframes of 50 ms over a 2.5 s exposure (50
719 frames) to give a total dose of ~66 e-/Å. Dose-fractionated movies were gain-

normalized, drift-corrected, summed, and dose-weighted using MotionCor2 (7). The contrast transfer function (CTF) was estimated for each summed image using the Patch CTF module in cryoSPARC v2.15.0 (8) Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC Blob Picker and Particle Extraction. The entire dataset consisted of 10,685 motion-corrected images with 4,961,691 particles. Particles were sorted using cryoSPARC 2D classification (N=100), resulting in 2,412,034 curated particles. Initial models (Ref 1: decoy 1, Ref 2: complex, Ref 3: decoy 2; SI Appendix; Fig. S2) were generated using cryoSPARC *ab initio* Reconstruction on a subset of 85,398 particles. Particles were further curated using Ref 1-3 as 3D templates for cryoSPARC Heterogeneous Refinement (N=6), resulting in the following: class1 (Ref 1), 258,097 particles; class2 (Ref 1), 263,966 particles; class3 (Ref 2), 668,743 particles; class4 (Ref 2), 665,480 particles; class5 (Ref 3), 280,933 particles; class6 (Ref 3), 274,815 particles. Particles from class3 and class4 were combined and further curated with another round of Heterogeneous Refinement (N=6), resulting in the following: class1 (Ref 1), 67,639 particles; class2 (Ref 1), 61,097 particles; class3 (Ref 2), 553,368 particles; class4 (Ref 2), 554,581 particles; class5 (Ref 3), 42,114 particles; class6 (Ref 3), 55,424 particles. Curated particles from class3 and class4 were combined, re-extracted with a box size of 320 px, and further classified using Ref 2 as a 3D template for cryoSPARC Heterogeneous Refinement (N=4). Classes from this round of Heterogeneous Refinement (N=4) were as follows: class1 (Ref 2), 871,163 particles; class2 (Ref 2), 77,769 particles; class3 (Ref 2), 61,489 particles; class4 (Ref 2), 64,026 particles. Particles from class1 and class2 were combined and further sorted using Heterogeneous Refinement (N=4) using class maps as templates, resulting in the following: class1, 134,536 particles; class2, 270,170 particles; class3, 294,162 particles; class4, 172,295 particles. Classification revealed two unique classes: nsp13$_1$-BTC (class1 and class2) and nsp13$_2$-BTC (class3 and class4). Particles within each class were further processed using RELION 3.1-beta Bayesian Polishing(9, 10). Polished particles were refined using cryoSPARC Non-uniform Refinement, resulting in structures with the

following particle counts and nominal resolutions: nsp13$_1$-BTC (404,706 particles; 3.40 Å) and nsp13$_2$-BTC (466,457 particles; 3.45 Å).

To improve the resolution of the RNA in the BTC, particles from both classes were combined in a cryoSPARC Non-uniform Refinement and density corresponding to nsp13 was subtracted. Subtracted particles were further refined with cryoSPARC Local Refinement using a mask encompassing the BTC and a fulcrum point defined on the backtracked RNA. This map, BTC$_5$(local), contained 871,163 particles with a nominal resolution of 3.23 Å.

To improve the density of nsp13.2 in the nsp13$_2$-BTC map, particles were subtracted using a mask defined around nsp13.2, leaving residual signal for only nsp13.2. Subtracted particles were classified (N=4) in RELION 3.1 beta using a mask around nsp13.2, resulting in the following classes: class1, 71,607 particles; class2, 163,540 particles; class3, 176,461 particles; class4, 54,849 particles. Subtracted particles in class1 and class2 were combined and reverted back to the original particles, followed by refinement using cryoSPARC Non-uniform Refinement. The resulting map of nsp13$_2$-BTC contains 235,147 particles with nominal resolution of 3.59 Å. Local resolution calculations were generated using blocres and blocfilt from the Bsoft package (11).

Model building and refinement. Initial models were derived from PDB: 6XEZ (1). The models were manually fit into the cryo-EM density maps using Chimera (12) and rigid-body and real-space refined using Phenix real_space_refine (13). For real-space refinement, rigid body refinement was followed by all-atom and B-factor refinement with Ramachandran and secondary structure restraints. Models were inspected and modified in Coot (14).

4-thiouridine crosslinking. Nsp12 or nsp12-D760A were incubated with 3-fold molar excess of nsp7/8 to assemble holo-RdRp (2 μM final) in transcription buffer. The resulting holo-RdRp was added to a modified RNA scaffold

(SI Appendix; Fig. S5A) containing a photoactivable 4-thiouridine base (Horizon Discovery) which was 5'-labelled by T4-polynucleotide kinase (New England Biolabs) with γ-^{32}P-ATP (Perkin-Elmer). The holo-RdRp/RNA complex was left to incubate for 5 minutes at 30°C in the dark. Nsp13 and ATP were added to a final concentration of 2 uM and 2 mM, respectively, and incubated for five minutes at 30°C in the dark. The reaction mixture was transferred to a Parafilm covered aluminum block at 4°C and irradiated with a 365-nm handheld UV lamp. Reactions were quenched with LDS sample loading buffer (ThermoFisher Scientific) and analyzed by gel electrophoresis on a NuPAGE 4-12% Bis-Tris gel (ThermoFisher) at 150 Volts for 1 hour and visualized by autoradioagraphy.

Molecular dynamics simulations

General simulation setup and parameterization. Proteins, ADP, and ions were parameterized with the DES-Amber SF1.0 force field (15). RNAs were parameterized with the Amber ff14 RNA force field (16) with modified electrostatic, van der Waals, and torsional parameters to more accurately reproduce the energetics of nucleobase stacking (17). The systems were solvated with water parameterized with the TIP4P-D water model (18) and neutralized with 150 mM NaCl buffer. The systems each contained ~887,000 atoms in a 190×190×190 Å cubic box.

Systems were first equilibrated on GPU Desmond using a mixed NVT/NPT schedule (19), followed by a 1 μs relaxation simulation on Anton, a special-purpose machine for molecular dynamics simulations (20). All production simulations were performed on Anton and initiated from the last frame of the relaxation simulation. Production simulations were performed in the NPT ensemble at 310 K using the Martyna-Tobias-Klein barostat (21). The simulation time step was 2.5 fs, and a modified r-RESPA integrator (22, 23) was used in which long-range electrostatic interactions were evaluated every three time steps. Electrostatic forces were calculated using the *u*-series method (24). A 9-Å cutoff was applied for the van der Waals calculations.

System preparation. The nsp13$_2$-BTC$_{-1U+1C}$ and the nsp13$_2$-BTC$_{-1U+1U}$ complexes were prepared from the cryo-EM structure of the nsp13$_2$-BTC$_5$. AlF_3 and CHAPSO were removed. Cytosines at the +2 and +3 positions of the p-RNA were removed, and the cytosine at −1 was mutated to uracil. The resulting p-RNA had a matched −1U and a mismatched +1C in nsp13$_2$-BTC$_{-1U+1C}$, and a matched −1U and +1U in nsp13$_2$-BTC$_{-1U+1U}$. Missing loops and termini in proteins were capped with ACE/NME capping groups. The two complexes were prepared for simulation using the Protein Preparation Wizard in Schrödinger Maestro. After a 1 μs relaxation simulation of the nsp13$_2$-BTC$_{-1U+1C}$ complex, the −1U of the p-RNA formed a Watson-Crick base pair with the −1A in the t-RNA, and the +1C of p-RNA formed a non-Watson-Crick C-A hydrogen bond with the +1A of the t-RNA in the active site. After a 1 μs relaxation simulation of the nsp13$_2$-BTC$_{-1U+1U}$ complex, the −1U and +1U of the p-RNA formed Watson-Crick base pairs with the −1A and +1A of the t-RNA respectively.

Simulation analysis. All simulations were visually inspected using the in-house visualization software Firefly. The average root-mean-square deviation (RMSD) was calculated for +1C (or +1U) of the p-RNA between the last frame of the 1 μs relaxation simulation and instantaneous structures from the trajectories, aligned on the entire nps12 module.

Quantification and statistical analysis. The nMS spectra were visualized using Thermo Xcalibur Qual Browser (versions 3.0.63 and 4.2.27), deconvolved using UniDec versions 3.2 and 4.1 (3, 4) and plotted using the m/z software (Proteometrics LLC, New York, NY). Experimental masses (SI Appendix; Fig. S1B and C) were reported as the average mass ± standard deviation across all the calculated mass values obtained within the observed charge state distribution.

The local resolution of the cryo-EM maps (SI Appendix; Fig. S3B-D) was estimated using blocres (11) with the following parameters: box size 15, sampling 1.1, and cutoff 0.5. Directional 3D FSC (SI Appendix; Fig. S3H-J) were calculated

by 3DFSC (25). The quantification and statistical analyses for model refinement and validation were generated using MolProbity (26) and PHENIX (13).

Table S1. Cryo-EM data collection, refinement, and validation statistics.

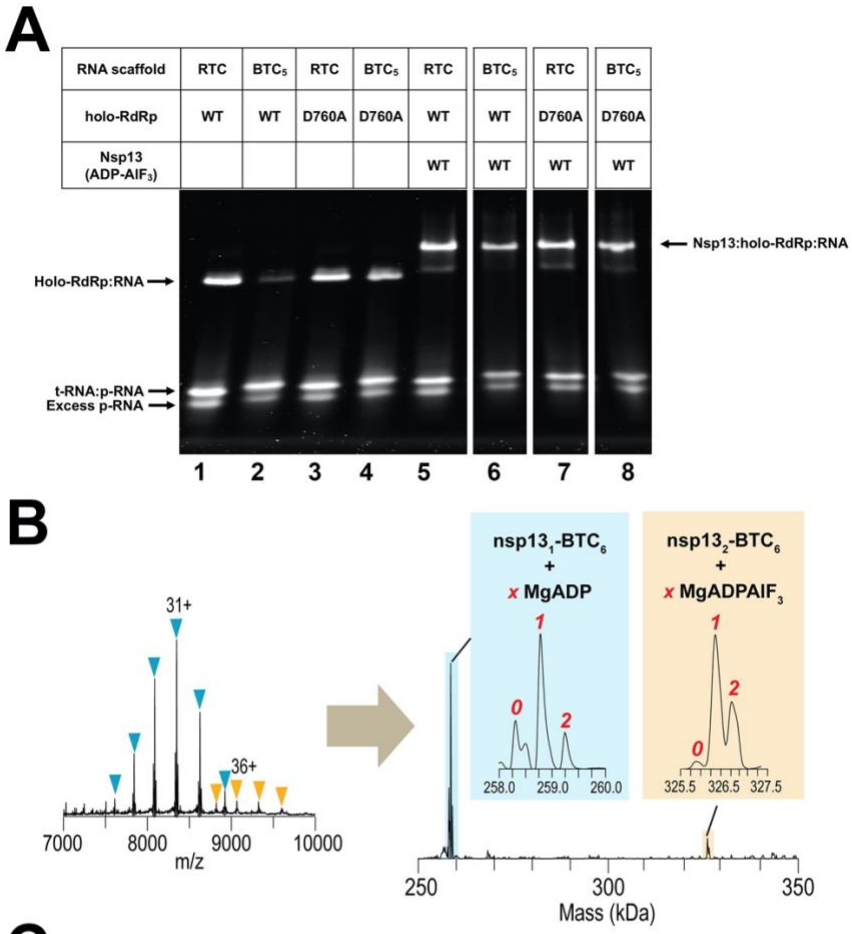
	nsp7/8/12/13/BTC_scaffold/ADP-AlF$_3$/CHAPSO		
Sample ID	nsp13$_1$-BTC	nsp13$_2$-BTC	BTC (local)
EMDB	EMD-23007	EMD-23008	EMD-23009
PDB	7KRN	7KRO	7KRP
Data collection and processing			
Microscope		TFS Titan Krios	
Voltage (kV)		300	
Detector		Gatan K3 Camera	
Electron exposure (e–/Å2)		66	
Defocus range (μm)		-0.8 to -2.5	
Data collection mode		Counting Mode	
Nominal Magnification		81,000x	
Pixel size (Å)		1.065	
Symmetry imposed		C1	
Initial particle images (no.)		4,961,691	
Final particle images (no.)	404,706	235,147	871,163
Map resolution (Å) - FSC threshold 0.143	3.40	3.59	3.23
Map resolution range (Å)	2.5-5.0	2.5-5.0	2.5-5.1
Refinement			
Initial model used (PDB code)	6XEZ	6XEZ	6XEZ
Map sharpening B factor (Å2)	-139.6	-127.6	-103.9

Model composition			
Non-hydrogen atoms	17351	21988	12561
Protein residues	1963	2553	1373
Nucleic acid residues	80	80	73
Ligands	5 Zn^{2+}, 2 Mg^{2+}, 3 CHAPSO, 2 ADP, 1 AlF_3	8 Zn^{2+}, 3 Mg^{2+}, 3 CHAPSO, 3 ADP, 2 AlF_3	2 Zn^{2+}, 1 Mg^{2+}, 3 CHAPSO, 1 ADP
B factors (Å2)			
Protein	45.65	74.56	38.31
Nucleic acid	128.79	163.6	140.77
Ligands	59.03	78.99	46.55
R.m.s. deviations			
Bond lengths (Å)	0.007	0.007	0.004
Bond angles (°)	0.711	0.735	0.609
Validation			
MolProbity score	2.66	2.68	1.97
Clashscore	9.18	8.96	6.12
Poor rotamers (%)	7.54	9.4	4.18
Ramachandran plot			
Favored (%)	91.4	92.76	97.07
Allowed (%)	8.6	6.93	2.93
Disallowed (%)	0	0.31	0

863

864

SUPPLEMENTAL FIGURES



C

Mass assignments from nMS analysis of BTC_6 incubated with nsp13 and $MgADPAlF_3$

Protein Assembly	Measured Mass ± SD (Da)*	Expected Mass (Da)	Δ Mass (Da)	% Mass Error
$nsp13_1$-BTC_6	258,307 ± 2	258,264	42	0.02
$nsp13_1$-BTC_6 + 1 MgADP	258,782 ± 3	258,716	66	0.03
$nsp13_1$-BTC_6 + 2 MgADP	259,246 ± 4	259,167	78	0.03
$nsp13_2$-BTC_6	325,844 ± 34	325,729	116	0.04
$nsp13_2$-BTC_6 + 1 $MgADPAlF_3$	326,342 ± 18	326,264	78	0.02
$nsp13_2$-BTC_6 + 2 $MgADPAlF_3$	326,776 ± 17	326,716	60	0.02

* Calculated from the average and S.D. of all the measured masses across the charge-state distribution ($n \geq 4$).

Figure S1

Fig. S1. Native gel electrophoresis mobility shift assay and nMS analysis of the BTC.

A. A native gel electrophoretic mobility shift assay reveals that wt-holo-RdRp requires nsp13(ADP-AlF_3) to bind the BTC_5-scaffold efficiently (compare lanes 1, 2, and 6) but holo-RdRp with nsp12-D760A does not require nsp13 (lane 4).

B. The nMS spectrum and the deconvolved mass spectrum showing assembly of stable nsp13-BTC_6 complexes. The peak for the $nsp13_2$-BTC_6 assembly is present at about ~9% intensity relative to the predominant peak from $nsp13_1$-BTC_6.

C. Mass assignments of the deconvolved peaks from the nMS analysis.

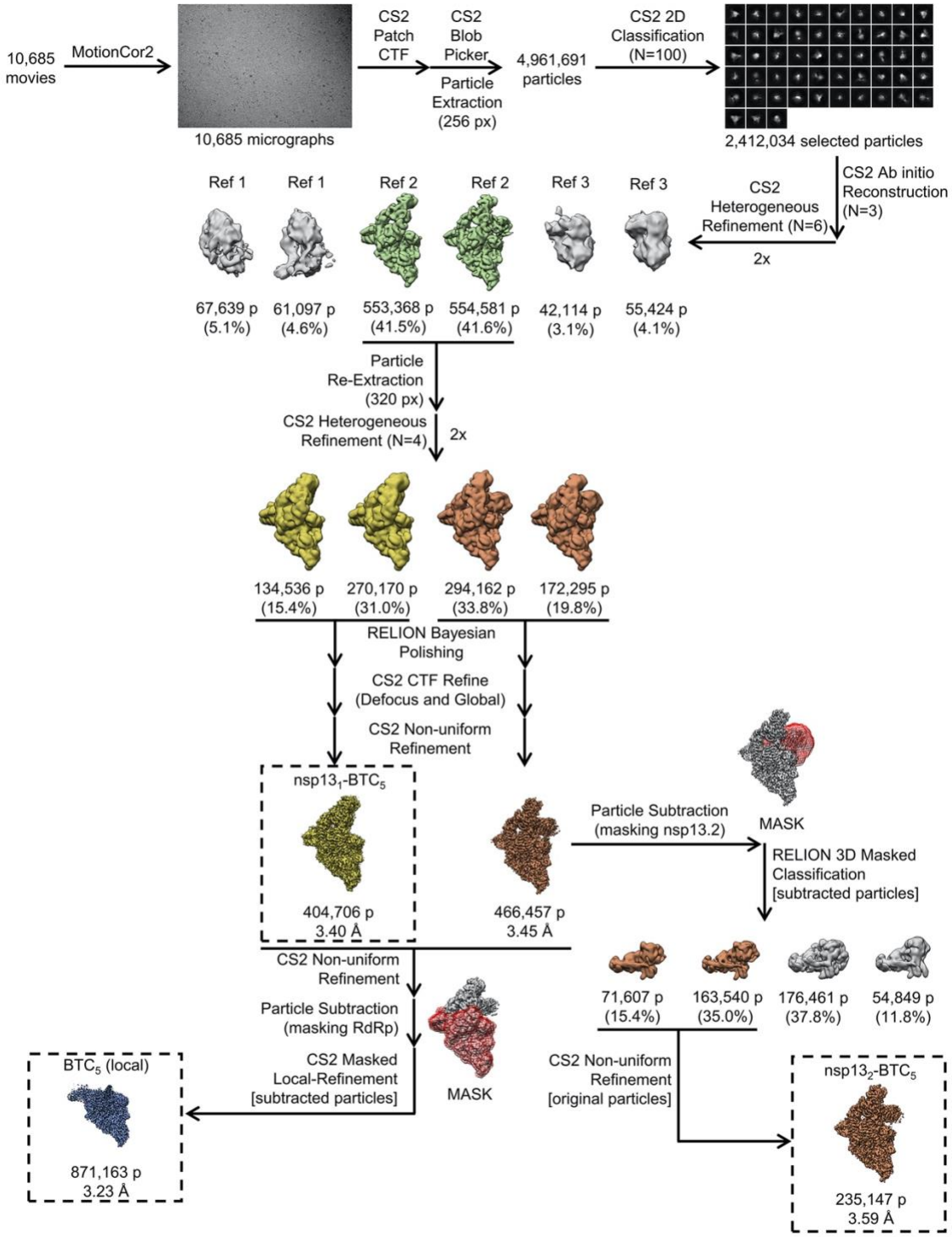


Figure S2

Fig. S2. Cryo-EM processing pipeline.

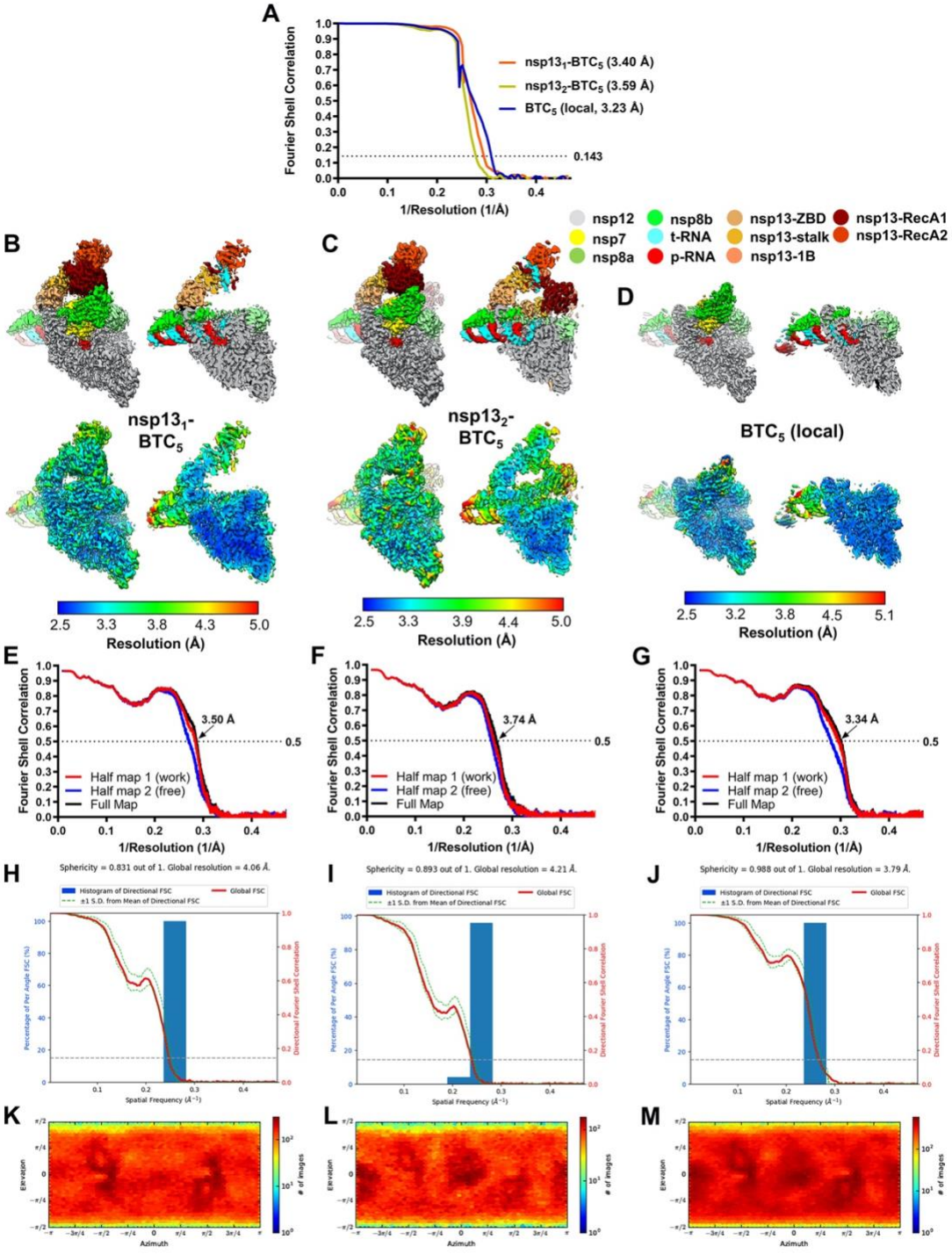


Figure S3

Fig. S3. Cryo-EM analysis.

A. Gold-standard FSC plots for nsp13$_1$-BTC$_5$, nsp13$_2$-BTC$_5$, and BTC$_5$(local), calculated by comparing two independently determined half-maps from cryoSPARC (8). The dotted line represents the 0.143 FSC cutoff.

B-D. Cryo-EM reconstructions filtered by local resolution(11). The view on the right is a cross-section.

(top) Colored by subunit according to the color key.

(bottom) Color by local resolution (key on the bottom).

B. Nsp13$_1$-BTC$_5$.

C. Nsp13$_2$-BTC$_5$.

D. BTC$_5$(local).

E – G. FSC calculated between the refined structures and the half map used for refinement (work, red), the other half map (free, blue), and the full map (black).

E. Nsp13$_1$-BTC$_5$.

F. Nsp13$_2$-BTC$_5$.

G. BTC$_5$(local).

H - J, Directional 3D Fourier shell correlation plots, calculated by 3DFSC(25).

H. Nsp13$_1$-BTC$_5$.

I. Nsp13$_2$-BTC$_5$.

J. BTC$_5$(local).

K – M. Particle angular distribution plots calculated in cryoSPARC. Scale shows the number of particles assigned to a particular angular bin. Blue, a low number of particles; red, a high number of particles.

K. Nsp13$_1$-BTC$_5$.

L. Nsp13$_2$-BTC$_5$.

M. BTC$_5$(local).

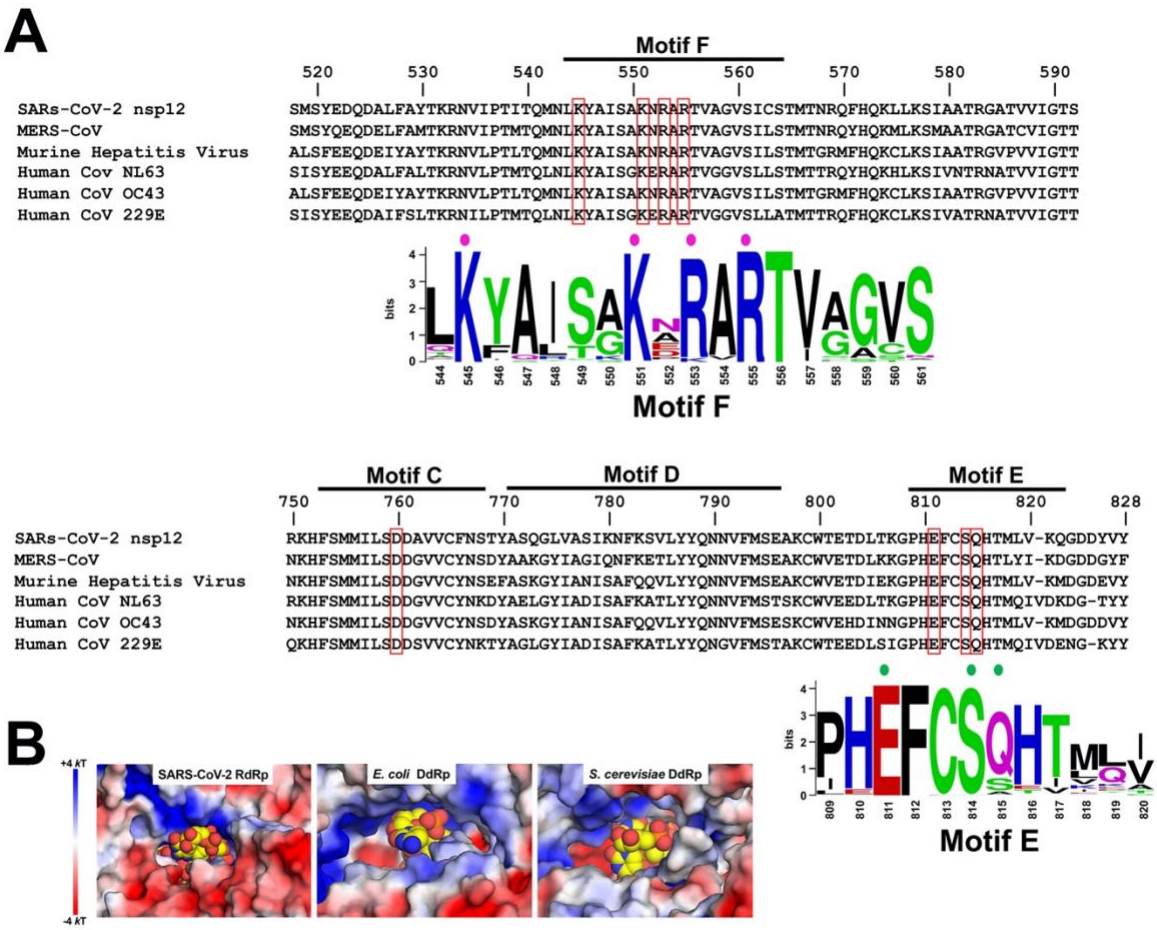


Figure S4

Fig. S4. Sequence conservation of nsp12 homologs and NTP-entry tunnel environment.

A. Sequence alignment of nsp12 homologs from six pathogenic and model CoV family members, covering RdRp motifs (27) (motifs F, C, D, and E denoted at the top of the sequence alignment) architecturally important for the NTP-entry tunnel. Selected residues discussed in the text are highlighted (red outlines). Sequence logos(28) for motif F and motif E are shown, with residues that interact with the backtracked RNA highlighted (colored dots above; see Figure 4). The sequence logos were generated from an alignment of 97 RdRp sequences from α-, β-, γ-, and δ-CoVs (Data S1).

B. Views from the outside into the NTP-entry tunnels of the SARS-CoV-2 BTC (*left*), an *E. coli* DdRp BTC [PDB ID: 6RI9; (29)] and an *S. cerevisiae* DdRp BTC [PDB ID: 3GTP; (30)]. Protein surfaces are colored by the electrostatic surface potential [calculated using APBS; (31)]. Backtracked RNA is shown as atomic spheres with yellow carbon atoms.

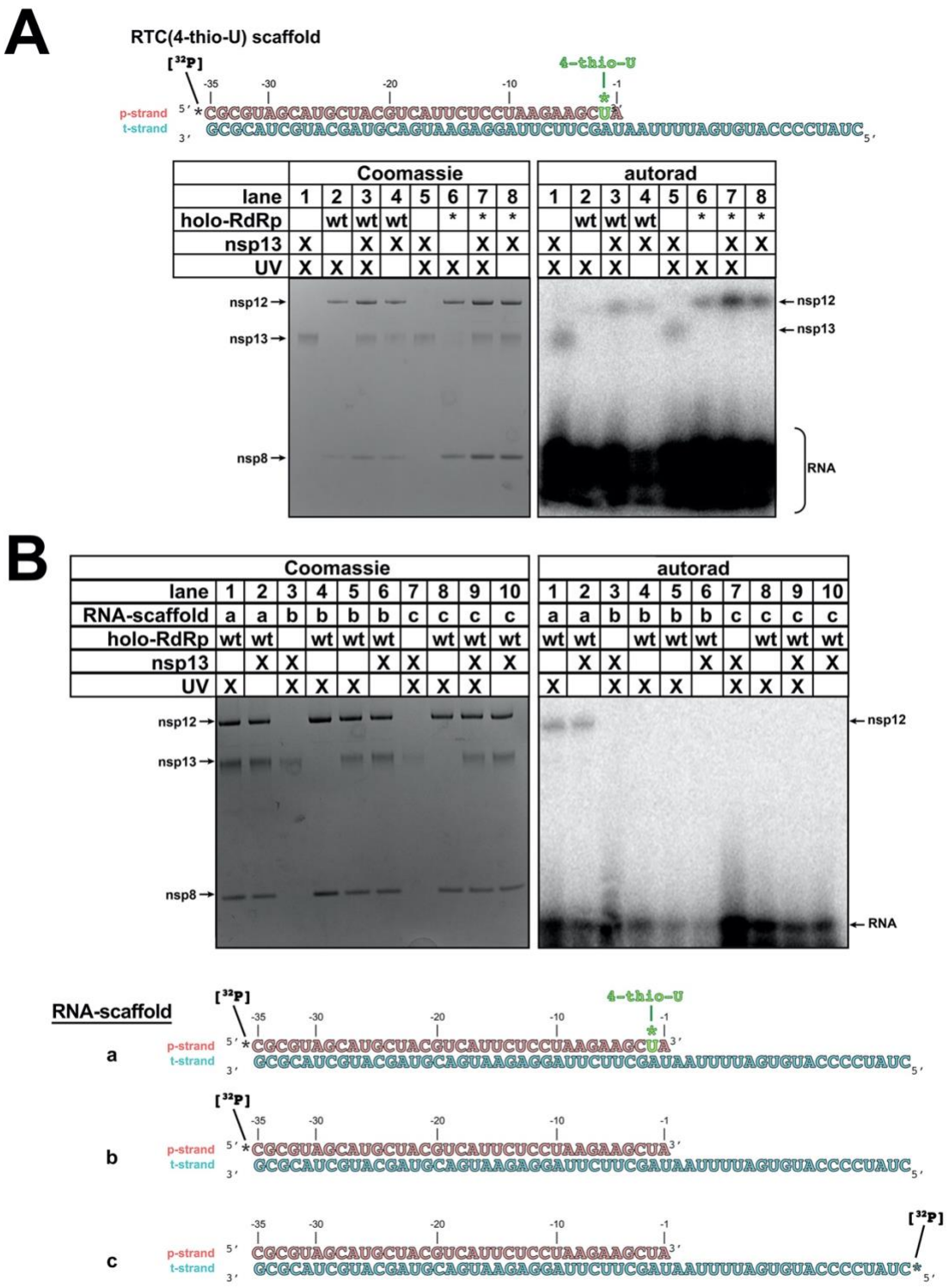


Figure S5

Fig. S5. 4-thio-U crosslinking analysis.

A. Protein-RNA crosslinking analysis: The 5'-[^{32}P]-labelled RTC(4-thio-U)-scaffold and the indicated proteins were incubated along with 2 mM ATP (present in every lane), exposed to UV as indicated, then analyzed by SDS polyacrylamide gel electrophoresis and autoradiography. The positions of nsp8, nsp12, and nsp13 bands are indicated. Lanes 1 and 5, containing nsp13 only, are identical controls indicating uniform UV exposure across the samples. Holo-RdRp(*) denotes the nsp12-D760A substitution that facilitates backtracking (see Figure S1A). The two panels show the same SDS polyacrylamide gel (left panel, Coomassie stained; right panel, visualized by autoradiography).

B. Protein-RNA crosslinks are specific. Lanes 1, 2; Analysis using the 5'-[^{32}P]-RTC(4-thio-U)-scaffold (RNA-scaffold 'a' shown on the bottom). Crosslinking to nsp12 serves as a positive control for the crosslinking reaction. Lanes 3-6; Analysis using RNA-scaffold 'b' (RTC-scaffold with 5'-[^{32}P]-labelled p-RNA). Lanes 7-10: Analysis using RNA-scaffold 'c' (RTC-scaffold with 5'-[^{32}P]-labelled t-RNA). The complete absence of protein-RNA crosslinks in lanes 3-10 indicates that the observed protein-RNA crosslinks arise from the 4-thio-U site-specifically incorporated in the p-RNA of the RTC(4-thio-U)-scaffold.

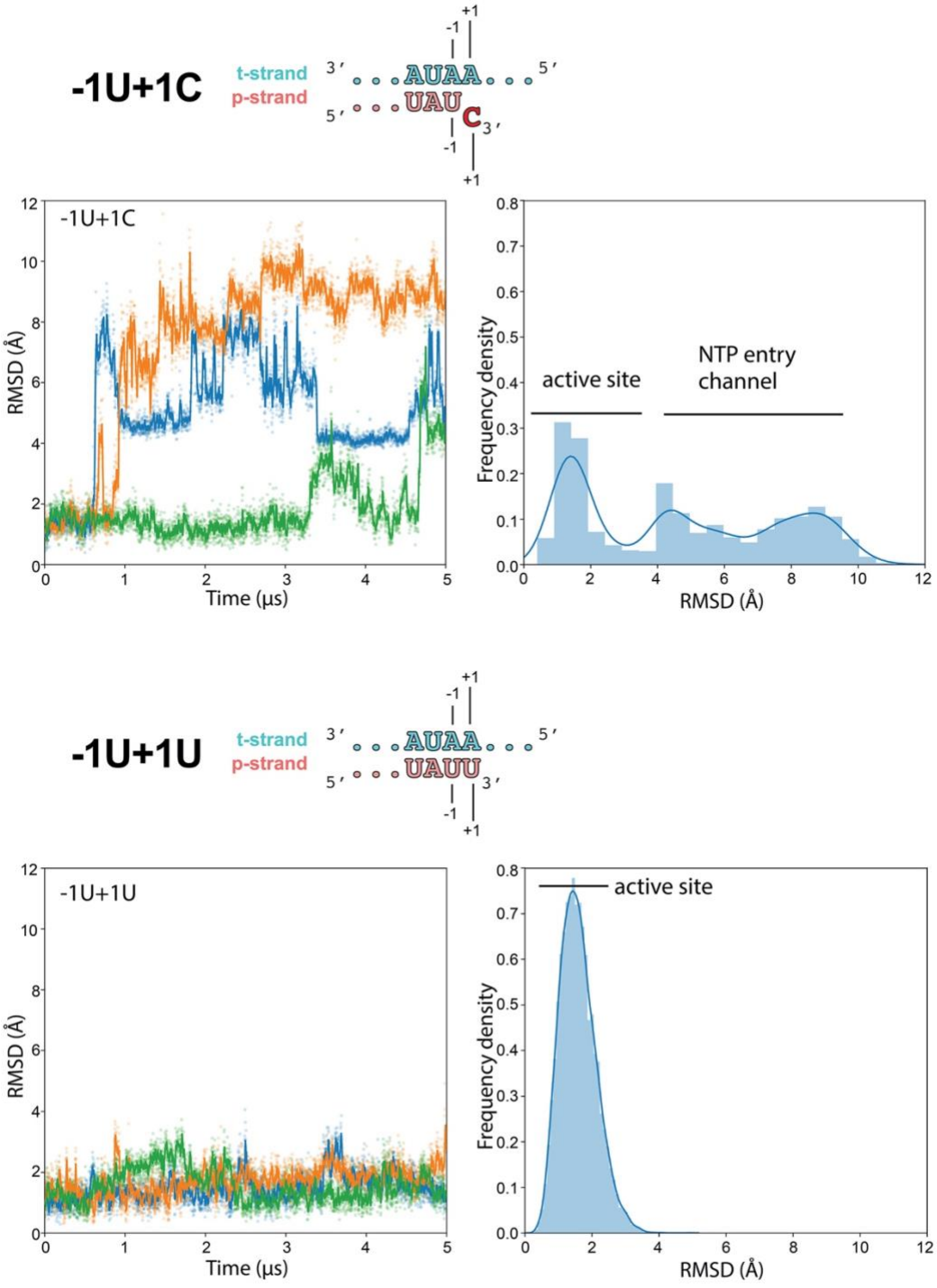


Figure S6

Fig. S6. Molecular dynamics simulations of nsp13$_2$-BTC$_{1U+1C}$ vs. nsp13$_2$-BTC$_{-1U+1U}$.

Molecular dynamics simulations of the nsp13$_2$-BTC$_{-1U+1C}$ (*top*) and nsp13$_2$-BTC$_{-1U+1U}$ (*bottom*) complexes. The schematics illustrate the active-site proximal nucleotides in each modeled complex. Each complex was simulated with 3 replicates. RMSD values plotted as a function of time represent the heavy-atom RMSD of the +1 nucleotide of the p-RNA (+1C for nsp13$_2$-BTC$_{-1U+1C}$ or matched +1U for nsp13$_2$-BTC$_{-1U+1U}$) compared with the starting configuration (see Methods). The RMSD histograms (plotted on the right) are aggregates of all 3 replicates.

(*top*) Nsp13$_2$-BTC$_{-1U+1C}$. As shown in Figure 5C, the mismatched p-RNA +1C spends about 60% of the time frayed from the t-RNA +1A and near or in the NTP-entry tunnel (RMSD ≥ ~3.5 Å).

(*bottom*) Nsp13$_2$-BTC$_{-1U+1U}$. With the p-RNA +1U matched with the t-RNA +1A for Watson-Crick base pairing, the p-RNA +1U does not fray and spends all of its time in the vicinity of the RdRp active site and base paired with the t-RNA.

SUPPLEMENTAL DATA FILES

Data File S1. Sequence alignment (Clustal format) of α- and β-CoV nsp12 sequences.

SI References

1. J. Chen, *et al.*, Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **182**, 1560-1573.e13 (2020).

2. P. D. B. Olinares, B. T. Chait, Methods in Molecular Biology. *Methods Mol Biology Clifton N J* **2062**, 357–382 (2019).

3. D. J. Reid, *et al.*, High-Throughput Deconvolution of Native Mass Spectra. *J Am Soc Mass Spectrom* **30**, 118–127 (2019).

4. M. T. Marty, *et al.*, Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Analytical chemistry* **87**, 4370–4376 (2015).

5. A. Morin, *et al.*, Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).

6. C. Suloway, *et al.*, Automated molecular microscopy: the new Leginon system. *Journal of structural biology* **151**, 41–60 (2005).

7. S. Q. Zheng, *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* **14**, 331–332 (2017).

8. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

9. S. H. W. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180**, 519–530 (2012).

10. J. Zivanov, *et al.*, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7** (2018).

11. G. Cardone, J. B. Heymann, A. C. Steven, One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *Journal of structural biology* **184**, 226–236 (2013).

12. E. F. Pettersen, *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605–1612 (2004).

13. P. D. Adams, *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213–221 (2010).

14. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **60**, 2126–2132 (2004).

15. S. Piana, P. Robustelli, D. Tan, S. Chen, D. E. Shaw, Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *J Chem Theory Comput* **16**, 2494–2507 (2020).

16. J. A. Maier, *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713.

17. D. Tan, S. Piana, R. M. Dirks, D. E. Shaw, RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc National Acad Sci* **115**, 201713027 (2018).

18. S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J Phys Chem B* **119**, 5113–5123 (2015).

19. K. J. Bowers, *et al.*, Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Acm Ieee Sc 2006 Conf Sc'06*, 43–43 (2006).

20. D. E. Shaw, *et al.*, Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *Sc14 Int Conf High Perform Comput Netw Storage Analysis*, 41–53 (2014).

21. G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms. *J Chem Phys* **101**, 4177–4189 (1994).

22. C. Predescu, *et al.*, Computationally efficient molecular dynamics integrators with improved sampling accuracy. *Mol Phys* **110**, 967–983 (2012).

23. M. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular dynamics. *J Chem Phys* **97**, 1990–2001 (1992).

24. C. Predescu, *et al.*, The u -series: A separable decomposition for electrostatics computation with improved accuracy. *J Chem Phys* **152**, 084113 (2020).

25. Y. Z. Tan, *et al.*, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature methods* **14**, 793–796 (2017).

26. V. B. Chen, *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**, 12–21 (2010).

27. A. J. W. te Velthuis, Common and unique features of viral RNA-dependent polymerases. *Cell Mol Life Sci Cmls* **71**, 4403–20 (2014).

28. T. D. Schneider, R. M. Stephens, Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18**, 6097–6100 (1990).

29. M. Abdelkareem, *et al.*, Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell* **75**, 298-309.e4 (2019).

30. D. Wang, *et al.*, Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203–1206 (2009).

31. E. Jurrus, *et al.*, Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112–128 (2018).