1   **Structural basis for backtracking by the SARS-CoV-2**

2   **replication-transcription complex**

3

4   **Brandon Malone[1,8], James Chen[1,8], Qi Wang[2], Eliza Llewellyn[1],**

5   **Young Joo Choi[1], Paul Dominic B. Olinares[3], Xinyun Cao[4],**

6   **Carolina Hernandez[5], Edward T. Eng[5], Brian T. Chait[3],**

7   **David E. Shaw[2,6], Robert Landick[4,7], Seth A. Darst[1,9],**

8   **Elizabeth A. Campbell[1,9,10]**

9

10   [1]Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY,

11   10065 USA.

12   [2]D. E. Shaw Research, New York, NY 10036 USA.

13   [3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller

14   University, New York, NY, 10065 USA.

15   [4]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI

16   53706 USA.

17   [5]The National Resource for Automated Molecular Microscopy, Simons Electron

18   Microscopy Center, New York Structural Biology Center, New York, NY, 10027

19   USA.

20   [6]Department of Biochemistry and Molecular Biophysics, Columbia University,

21   New York, NY 10032 USA.

22   [7]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI

23   53706 USA.

24   [8]These authors contributed equally: Brandon Malone, James Chen.

25   [9]Correspondence to: darst@rockefeller.edu, campbee@rockefeller.edu

26

27   **Author Contributions**: Conceptualization; B.M., J.C., Q.W., D.E.S., S.A.D.,

28   E.A.C. Protein purification, biochemistry; B. M., J.C., E.L., Y.J.C. Mass

29   spectrometry acquisition and analysis; P.D.B.O., B.C. Cryo-EM specimen

30   preparation; B.M., J.C. Cryo-EM data collection and processing: J.C., C.H.,

31   E.T.E. Model building, and structural analysis; J.C., B.M., S.A.D., E.A.C.

32   Molecular dynamics simulations and analysis; Q.W., D.E.S. Funding acquisition

33   and supervision; B.C., D.E.S., R.L., S.A.D, E.A.C. All authors contributed to the

34   writing of the manuscript.

35

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

37

38   **Classification:** Biological Sciences; Biophysics and Computational Biology.

39

40   **Keywords:** backtracking; coronavirus; cryo-electron microscopy; molecular

41   dynamics; RNA-dependent RNA polymerase

42

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

67 **backtracking by the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp)**

68 **may aid viral transcription and replication. Structures of SARS-CoV-2 RdRp**

69 **bound to the essential nsp13 helicase and RNA suggested the helicase**

70 **facilitates backtracking. We use cryo-electron microscopy, RNA-protein**

71 **crosslinking, and unbiased molecular dynamics simulations to characterize**

72 **SARS-CoV-2 RdRp backtracking. The results establish that the single-**

73 **stranded 3'-segment of the product-RNA generated by backtracking**

74 **extrudes through the RdRp NTP-entry tunnel, that a mismatched nucleotide**

75 **at the product-RNA 3'-end frays and enters the NTP-entry tunnel to initiate**

76 **backtracking, and that nsp13 stimulates RdRp backtracking. Backtracking**

77 **may aid proofreading, a crucial process for SARS-CoV-2 resistance against**

78 **antivirals.**

79

80 **Significance Statement**

81 The COVID-19 pandemic is caused by the severe acute respiratory syndrome

82 coronavirus 2 (SARS-CoV-2). The SARS-CoV-2 genome is replicated and

83 transcribed by its RNA-dependent RNA polymerase (RdRp), which is the target

84 for antivirals such as remdesivir. We use a combination of approaches to show

85 that backtracking (backwards motion of the RdRp on the template RNA) is a

86 feature of SARS-CoV-2 replication/transcription. Backtracking may play a critical

87 role in proofreading, a crucial process for SARS-CoV-2 resistance against many

88 antivirals.

89

4

Introduction

90

91 SARS-CoV-2 is the causative agent of the current COVID-19 pandemic (1, 2).

92 The SARS-CoV-2 genome is replicated and transcribed by its RNA-dependent

93 RNA polymerase holoenzyme [holo-RdRp, subunit composition

94 nsp7/nsp8$_2$/nsp12 (3, 4)] in a replication-transcription complex (RTC), which is

95 the target for antivirals such as remdesivir [Rdv; (5)]. The holo-RdRp is thought to

96 coordinate with many co-factors to carry out its function (6, 7). Some of these co-

97 factors, such as the nsp13 helicase (8) and the nsp10/nsp14 proofreading

98 assembly (9, 10), are also essential for viral replication and are antiviral targets

99 (11–13).

100 We recently reported the first views of the SARS-CoV-2 RTC in complex with

101 the nsp13 helicase [cryo-electron microscopy structures at a nominal resolution

102 of 3.5 Å; (14)]. The overall architecture of the nsp13-RTC places the nucleic acid

103 binding site of nsp13 directly in the path of the downstream template-strand RNA

104 (t-RNA), and cryo-electron microscopy (cryo-EM) difference maps revealed the

105 5'-single-stranded t-RNA overhang engaged with nsp13 before entering the

106 RdRp active site (14). The nsp13 helicase translocates on single-stranded

107 nucleic acid in the 5'->3' direction (15–22). Thus, this structural arrangement

108 presents a conundrum: The RdRp translocates in the 3'->5' direction on the t-

109 RNA strand, while nsp13 translocates on the same strand in the opposite

110 direction. Translocation of each enzyme opposes each other, and if the helicase

111 prevails it is expected to push the RdRp backward on the t-RNA (14). This

5

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

6

135

136 **Results**

137 **SARS-CoV-2 RdRp backtracked complexes for cryo-electron microscopy**.

138 Previously, DdRp backtracked complexes (BTCs) were generated for structural

139 studies by direct incubation of the DdRp with DNA-RNA scaffolds containing

140 mismatched nucleotides at the RNA 3'-end (27, 28, 30); these BTC-scaffolds

141 bind with the downstream Watson-Crick base pairs of the RNA-DNA hybrid

142 positioned in the DdRp active site and the single-stranded 3'-segment of

143 mismatched RNA extruding out the DdRp NTP-entry tunnel. To study RdRp

144 BTCs, we therefore designed and tested RNA scaffolds based on our original

145 SARS-CoV-2 RTC-scaffold but with three or five mismatched cytosine

146 nucleotides added to the product-RNA (p-RNA) 3'-end (BTC$_3$- and BTC$_5$-

147 scaffolds; Fig. 1A).

148 Native electrophoretic mobility shift assays revealed that although the holo-

149 RdRp (nsp7/nsp8$_2$/nsp12) bound the RTC-scaffold as observed previously [Fig.

150 1B, lane 1; SI Appendix, Fig. S1A; (14)], nsp13 was required for efficient binding

151 to the BTC-scaffolds (Fig. 1B). Stable nsp13-holo-RdRp complexes with BTC-

152 scaffolds were also observed by native mass-spectrometry (Fig. S1B and C).

153 To determine the structural organization of the SARS-CoV-2 BTC, we

154 assembled nsp13(ADP-AlF$_3$) and holo-RdRp with the BTC$_5$-scaffold (Fig. 1A;

155 hereafter called BTC$_5$) and analyzed the samples by single-particle cryo-EM. The

156 sample comprised two major classes: nsp13$_1$-BTC$_5$ (3.4 Å nominal resolution),

157 and nsp13$_2$-BTC$_5$ (3.6 Å; Fig. 1C; SI Appendix, Fig. S2 and S3). To eliminate

158 structural heterogeneity in the nsp13 subunits and obtain a higher-resolution view

8

159 of the BTC, the particles from both classes were combined and locally refined

160 inside a mask applied around the holo-RdRp and RNA (excluding the nsp13

161 subunits), leading to the BTC_5(local) combined map (3.2 Å; Fig. 1C; SI Appendix,

162 Fig. S2 and S3, Table S1).

163   The cryo-EM maps (Fig. 1C and 2) revealed two significant differences with

164 the nsp13-RTC structures (14): 1) The single-stranded downstream template-

165 RNA (t-RNA) engaged with nsp13.1 was resolved (Fig. 2A), and 2) a single-

166 stranded p-RNA 3'-segment was extruded into the RdRp NTP-entry tunnel (Fig.

167 2B).

168

169 **Nsp13 binds the downstream single-stranded t-RNA.** In the $nsp13_1$-BTC_5 and

170 $nsp13_2$-BTC_5 cryo-EM maps, the single-stranded 5'-segment of the t-RNA was

171 engaged with nsp13.1. This region of the cryo-EM density was well-resolved (Fig.

172 2A), allowing identification of the t-RNA segment engaged within the helicase as

173 +14 to +8 (numbering defined in Figure 1A), $^{5'}CCCAUGU^{3'}$. The five-nucleotide

174 segment connecting the t-RNA between the helicase and the RdRp (+7 to +3)

175 was disordered and not modeled.

176

177 **The SARS-CoV-2 RdRp NTP-entry tunnel accommodates the backtracked**

178 **RNA.** The cryo-EM maps also resolved a single-stranded p-RNA 3'-segment of

179 the BTC_5-scaffold extruding into the RdRp NTP-entry tunnel (Fig. 2B), confirming

180 the formation of a BTC (Fig. 3A). The overall architecture of the SARS-CoV-2

9

181   BTC is analogous to DdRp BTCs [Fig. 3; (14)]. The DdRp bridge helix [BH; (35)]

182   separates the DdRp active site cleft into a channel for the downstream template

183   DNA (over the top of the BH; Fig. 3B) and the NTP-entry tunnel (underneath the

184   BH; Fig. 3B). Similarly, the viral RdRp motif F [SI Appendix, Fig. S4A; (32)]

185   serves as the strand separating structural element for the backtracked RNA (Fig.

186   3A). The downstream t-RNA passes over the top of motif F, while the

187   backtracked RNA extrudes out the NTP-entry tunnel underneath motif F (Fig.

188   3A).

189   The RdRp NTP-entry tunnel provides a steric and electrostatic environment

190   conducive to channeling the backtracked RNA out of the active site without

191   specific polar protein-RNA interactions that could hinder the RNA movement (Fig.

192   3C and 4). Comparing the electrostatic surface potential of the NTP-entry tunnels

193   of the SARS-CoV-2 RdRp with eukaryotic and bacterial DdRps reveals a similar

194   overall electrostatic surface environment that may facilitate backtracked RNA

195   entry (Fig. 3C; SI Appendix, Fig. S4B), including a 'track' of conserved positively-

196   charged Arg and Lys residues of motif F (SARS-CoV-2 nsp12 K545, K551, R553,

197   and R555; Fig. 4; SI Appendix, Fig. S4A). Conserved residues of RdRp motifs C

198   and E complete the active-site/NTP-entry tunnel environment surrounding the

199   backtracked RNA (Fig. 4; SI Appendix, Fig. S4A).

200   In the nsp13-RTCs, the RTC-scaffold (Fig. 1A) is bound in a post-

201   translocated state (14); the 3' p-RNA A is base-paired to the t-RNA U at the -1

202   site near the catalytic nsp12-D760 (Fig. 5A). The next t-RNA base (A at +1) is

203   positioned to receive the incoming nucleoside-triphosphate (NTP) substrate, but

204   the site for the incoming NTP substrate is empty (Fig. 5A). By contrast, the BTC

205   structures were translocated by one base pair compared to the RTCs; the base

206   pair corresponding to the A-U Watson-Crick base pair at the 3'-end of the p-RNA

207   (located in the -1 site of the RTCs) was in the -2 position of the BTCs (Fig. 1A, 4,

208   and 5B). The -1 position of the BTC was occupied by the first C-A mismatch; the

209   p-RNA -1C made a non-Watson-Crick hydrogen bond with the opposing t-RNA A

210   (Fig. 4 and 5B). The next three mismatched p-RNA nucleotides (+1C, +2C, +3C)

211   trailed into the NTP entry tunnel (Fig. 4 and 5B). The 3'-nucleotide of the BTC$_5$-

212   scaffold p-RNA (+4C; Fig. 1A) was solvent-exposed at the outward-facing end of

213   the NTP-entry tunnel, lacked density and was therefore not modelled (Fig. 2B).

214   The trajectory of the backtracked nucleotides at positions +1/+2 was sharply bent

215   due to spatial constraints of motif F residues (Fig. 4A).

216

217   **Nsp13 stimulates backtracking.** The SARS-CoV-2 wild-type holo-RdRp

218   required the nsp13 helicase to bind the BTC-scaffolds efficiently (Fig. 1B).

219   However, we observed that the holo-RdRp containing nsp12 with a single amino

220   acid substitution (D760A) did not require nsp13 to bind the BTC-scaffolds (SI

221   Appendix; Fig. S1A, lane 4). Nsp12-D760 is a conserved residue of the RdRp

222   motif C that chelates a crucial Mg^{2+} ion in catalytic complexes [SI Appendix; Fig.

223   S4A; (32)], but in RdRp structures lacking substrate (including the BTC

224   structures), the Mg^{2+} ions are absent (14, 36, 37). The catalytic Asp residues of

225   the DdRps typically chelate the Mg^{2+} ion even in the absence of substrate (31,

226   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

12

250 under the conditions expected to favor backtracking the most (SI Appendix; Fig.

251 S5A, lane 7). These results affirm the view that nsp13 facilitates backtracking of

252 the SARS-CoV-2 RdRp.

253

254 **A mismatched nucleotide at the p-RNA 3'-end spontaneously frays and**

255 **enters into the RdRp NTP-entry tunnel.** The SARS-CoV-2 RTC is a highly

256 processive and rapid replicase/transciptase, capable of replicating an ~1 kb RNA

257 template at an average rate of ~170 nt/s (40). However, studies of other viral

258 RdRps suggest that misincorporation slows the overall elongation rate and may

259 induce backtracking (41–43). We used molecular dynamics simulations to

260 explore the fate of a mismatched nucleotide incorporated at the p-RNA 3'-end.

261 Starting with the $nsp13_2$-BTC_5 structure, the -1C was mutated to U, and the +2 to

262 +4 C's were removed. The resulting pre-translocated p-RNA had a matched -1U

263 and a mismatched +1C (-1U+1C; Fig. 5C). In three 5 μs simulations we observed

264 the 3'-mismatched +1C alternating between two positions, either remaining in the

265 vicinity of the active site (RMSD < 3.5 Å) or fraying away from the p-RNA:t-RNA

266 hybrid towards or into the NTP-entry tunnel (RMSD > 3.5 Å; Fig. 5C). Based on

267 analysis of the aggregated -1U+1C simulations, the mismatched +1C spent

268 about 40% of the time near the active site and about 60% of the time frayed

269 towards or in the NTP-entry tunnel. In control simulations with a fully matched p-

270 RNA 3'-end (-1U+1U), the matched +1U at the p-RNA 3'-end did not fray and

271 spent 100% of the time in the active site pocket (SI Appendix; Fig. S6).

272

Discussion

273

274   Our results establish that the SARS-CoV-2 RTC backtracks, that backtracking is

275   facilitated by the nsp13 helicase, and that the resulting single-stranded 3'-

276   segment of the p-RNA extrudes out the RdRp NTP-entry tunnel in a manner

277   analogous to the evolutionarily unrelated cellular DdRps (Fig. 3). Thus, a

278   secondary tunnel to accommodate backtracked RNA, facilitating fidelity and

279   possibly other functions (Fig. 6), appears to be a crucial feature of transcriptase

280   enzymes that evolved independently.

281   Backtracking of Φ6 and poliovirus RdRps has been reported based on

282   analysis of single-molecule observations (41–43). The nsp13 helicase facilitates

283   efficient backtracking of the SARS-CoV-2 RTC (SI Appendix; Fig. S5). We note

284   that in bacteria, the UvrD helicase has been shown to induce DdRp backtracking,

285   suggesting that a role for helicases in backtracking may be widespread (44).

286   Our results are consistent with the view that a matched nucleotide at the pre-

287   translocated p-RNA 3'-end remains base paired to the t-RNA (Fig. 5; SI

288   Appendix, Fig. S6), facilitating translocation and subsequent NTP addition and

289   thus rapid elongation (at a maximum elongation rate of ~170 nt/s, a translocation

290   event would occur approximately every 6 msec, on average, explaining why

291   translocation was not observed in our 5 μs simulations; Fig. 5; SI Appendix, Fig.

292   S6). However, upon misincorporation, the pre-translocated, mismatched

293   nucleotide at the p-RNA 3'-end spends more than half the time frayed from the t-

294   RNA and towards or in the NTP-entry tunnel (Fig. 5C), a configuration that is

295   likely recalcitrant to translocation and subsequent elongation. The favorable

14

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

15

319   which RdRp misincorporation or incorporation of nucleotide analogs can pause

320   the RdRp, allowing nsp13 to engage with the downstream single-stranded t-RNA

321   to induce backtracking (14). The resulting exposure of the p-RNA 3'-end out the

322   NTP-entry tunnel (Fig. 3A and 6) could provide access for the SARS-CoV-2

323   proofreading machinery [nsp10/14; (9, 12)] to degrade the p-RNA 3'-end, thus

324   removing the misincorporation or analog. This proofreading activity, which is

325   unique to the nidovirus order to which CoVs belong (10), is a major determinant

326   for the resistance of CoVs against many nucleotide analog inhibitors (13). Thus,

327   understanding RdRp backtracking and its potential role in CoV proofreading can

328   facilitate the development of therapeutics.

329

330

Materials and Methods

331
332   Detailed descriptions of SARS-CoV-2 nsp12, 7, 8, and 13 protein purification,

333   assembly of the RTC complexes, Native EMSAs, native mass-spectrometry,

334   cross-linking, specimen preparation for cryo-EM, cryo-EM data acquisition and

335   processing, model building and refinement, and molecular dynamics simulations

336   are provided in the SI Appendix.

337

352

References

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

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

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

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

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

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

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

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

370 8. K. C. Lehmann, E. J. Snijder, C. C. Posthuma, A. E. Gorbalenya, What we
371 know but do not understand about nidovirus helicases. *Virus Res* **202**, 12–32
372 (2015).

373 9. E. Minskaia, *et al.*, Discovery of an RNA virus 3'->5' exoribonuclease that is
374 critically involved in coronavirus RNA synthesis. *Proc National Acad Sci* **103**,
375 5108–5113 (2006).

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

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

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

383 13. E. C. Smith, H. Blanc, M. C. Surdel, M. Vignuzzi, M. R. Denison,
384 Coronaviruses lacking exoribonuclease activity are susceptible to lethal
385 mutagenesis: evidence for proofreading and potential therapeutics. *Plos Pathog*
386 **9**, e1003565 (2013).

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

389 15. A. O. Adedeji, *et al.*, Mechanism of Nucleic Acid Unwinding by SARS-CoV
390 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).

20

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

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

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

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

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

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

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

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

479   50. E. C. Smith, N. R. Sexton, M. R. Denison, Thinking Outside the Triangle:
480   Replication Fidelity of the Largest RNA Viruses. *Ann Rev Virol* **1**, 111–132
481   (2014).

482

483

484
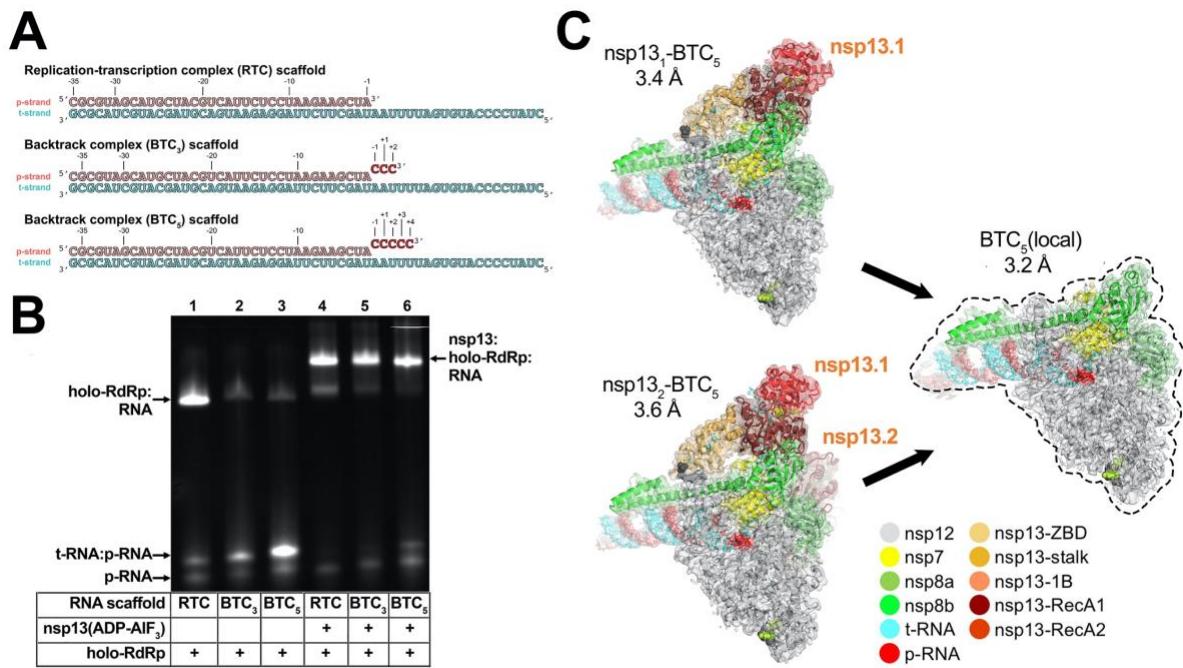
Figures

Figure 1

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

A. RNA scaffolds: (*top*) replication-transcription complex (RTC) scaffold (14);

(*bottom*) backtrack complex scaffolds (BTC3 and BTC5).

B. A native gel electrophoretic mobility shift assay reveals that holo-RdRp

requires nsp13(ADP-AlF3) 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.
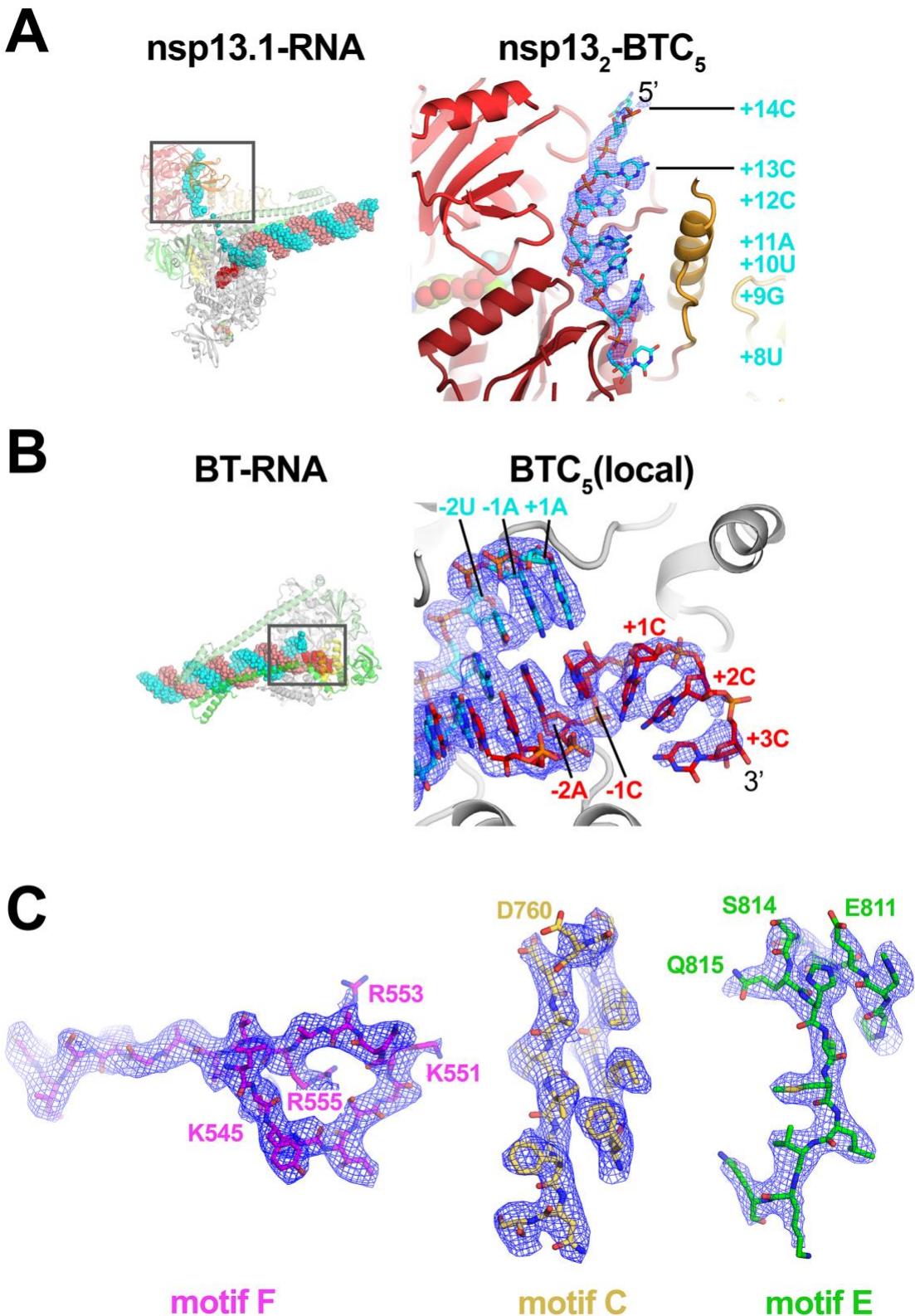
Figure 2

497 **Fig. 2. Cryo-EM density maps.**

498 **A.** (*left*) Overall view of $nsp13_1$-BTC_5. The boxed region is magnified on the right.

499 (*right*) Magnified view of the t-RNA segment (+14-5'-CCCAUGU-3'-+8) enclosed

500 in the nsp13.1 helicase subunit. The cryo-EM density map (from the $nsp13_2$-BTC

501 structure) for the RNA is shown (blue mesh).

502 **B.** (*left*) Overall view of the BTC structure. The boxed region is magnified on the

503 right.

504 (*right*) Magnified view of the region around the RdRp active site, showing the t-

505 RNA (cyan) and p-RNA (red) with the backtracked RNA segment. The cryo-EM

506 density map for the RNA [from BTC_5(local)] is shown (blue mesh).

507 **C.** BTC_5(local) cryo-EM density maps around nsp12 conserved motifs F, C, and

508 E. Selected residues are labeled.

509

510

24

511
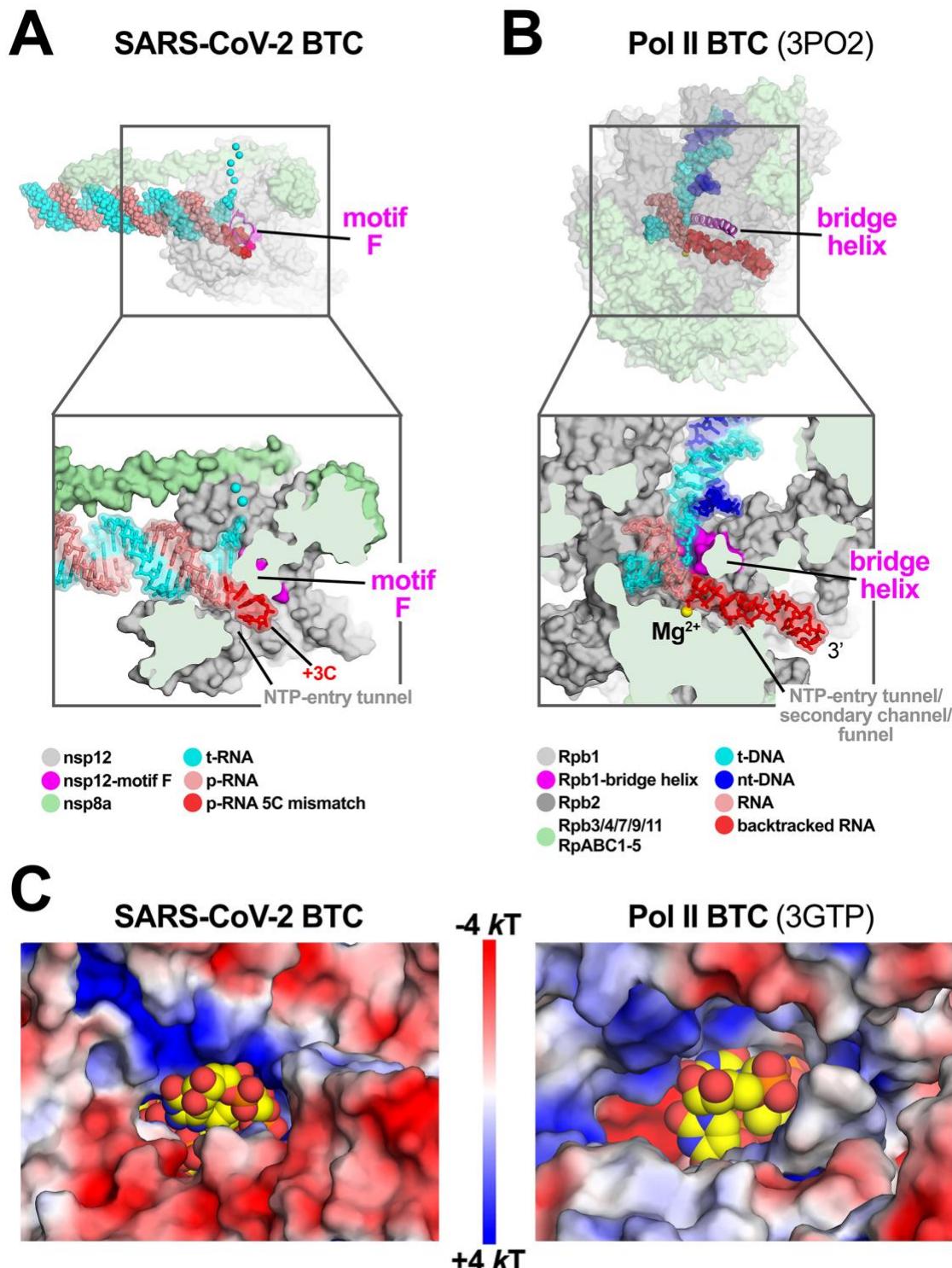
Figure 3

512

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

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

515 (*top*) Proteins are shown as transparent molecular surfaces, nucleic acids as

516 atomic spheres. The boxed regions are magnified on the bottom.

517 (*bottom*) Magnified, cross-sectional view. Proteins are shown as molecular

518 surfaces, nucleic acids in stick format with transparent molecular surface.

519 **A.** The SARS-CoV-2 BTC$_5$(local). Nsp8a and nsp12 are shown (nsp7 and nsp8b

520 are removed for clarity). Nsp12 motif F is shown as a magenta backbone ribbon

521 (*top*). Backtracked RNA (+1C to +3C of the BTC$_5$-scaffold; Figure 1A) extrude out

522 the NTP-entry tunnel.

523 **B.** A DdRp (*Saccharomyces cerevisiae* Pol II) BTC [PDB ID: 3PO2; (29)]. The

524 bridge helix is shown as a magenta backbone ribbon. The backtracked RNA

525 extrudes out the NTP-entry tunnel/secondary channel/funnel.

526 **C.** Views from the outside into the NTP-entry tunnels of the SARS-CoV-2 (*left*)

527 and an *S. cerevisiae* DdRp [PDB ID: 3GTP; (27) BTC. Protein surfaces are

528 colored by the electrostatic surface potential [calculated using APBS; (48)].

529 Backtracked RNA is shown as atomic spheres with yellow carbon atoms.

530

531

26
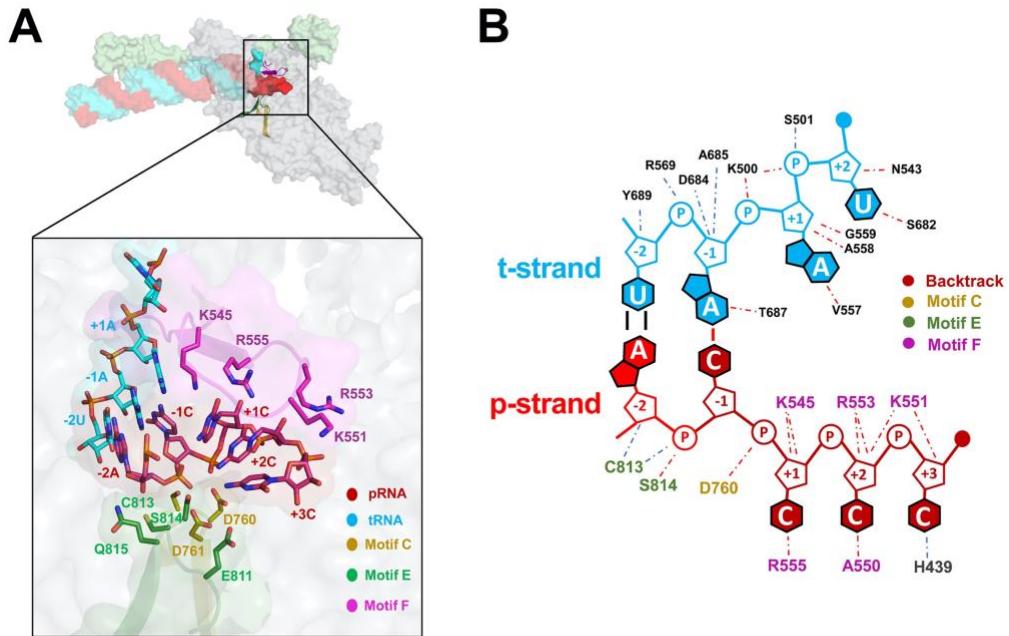
Figure 4

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

A. (*top*) Overall view of BTC$_5$(local). Proteins are shown as transparent molecular surfaces, nucleic acids as atomic spheres. Nsp8a and nsp12 are shown (nsp7 and nsp8b are removed for clarity). Nsp12 motifs C, E, and F are shown as backbone ribbons (colored according to the key on the bottom. The boxed region is magnified below.

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

B. Schematic illustrating the same protein-RNA interactions as (**A**). Drawn using Nucplot (49).
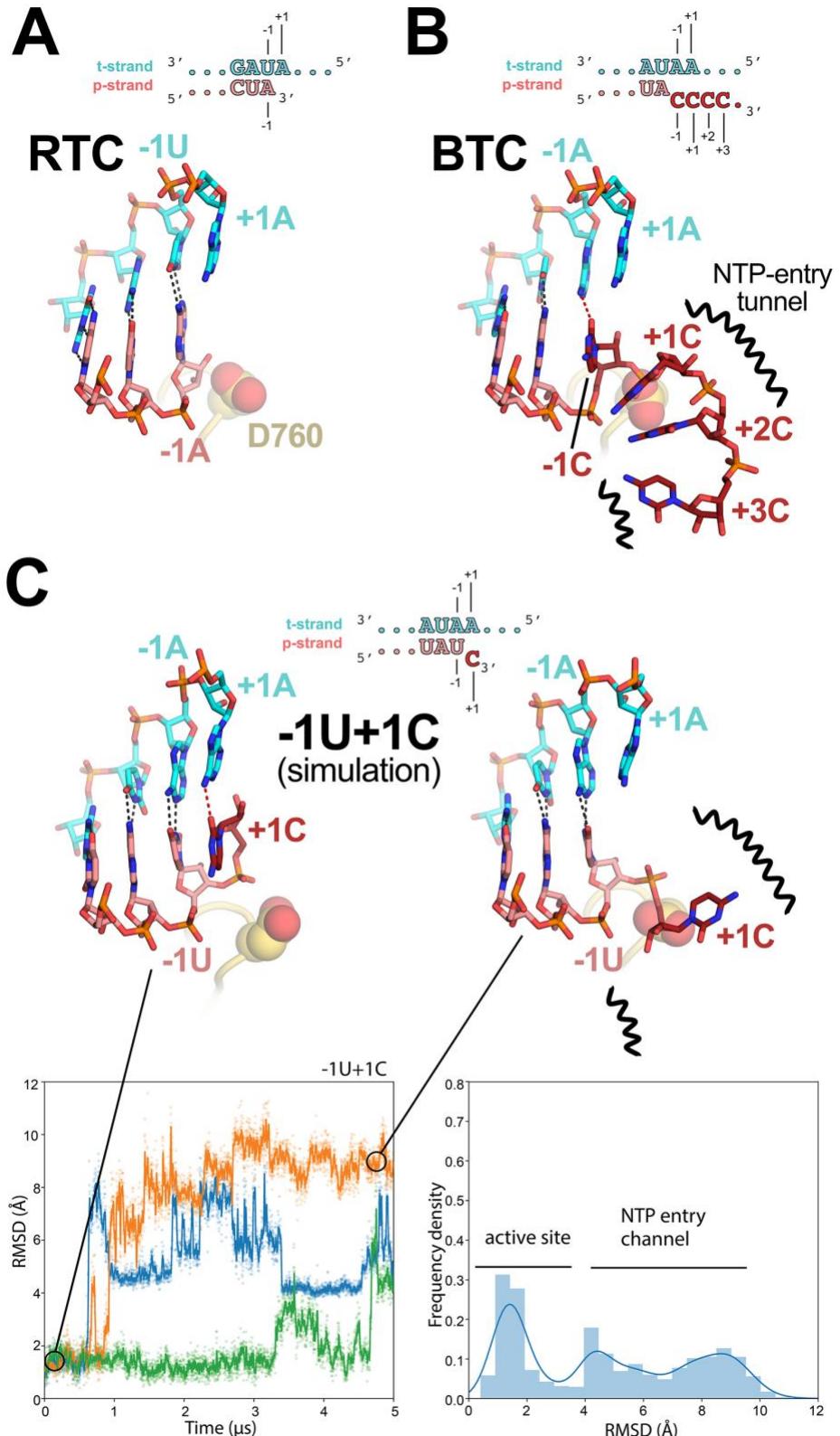
Figure 5

546 **Fig. 5. Comparison of active-site proximal RNA in the RTC and BTC**

547 **structures, and from simulations of a mismatched nucleotide at the p-RNA**

548 **3'-end.**

549 **A-B,** Comparison of the active-site proximal RNA in the RTC [**A**; PDB ID: 6XEZ;

550 (14)], BTC$_5$(local) (**B**), and from selected snapshots of molecular dynamics

551 simulations of a -1U+1C complex (**C**). The schematics denote the nucleotides

552 shown in the context of the RTC- (**A**) and BTC$_5$-scaffolds (**B**; full scaffold

553 sequences shown in Figure 1A) or generated from the BTC$_5$-scaffold for the

554 simulations (**C**). Carbon atoms of the t-RNA are colored cyan, p-RNA are colored

555 salmon except in the case of mismatched C's at the 3'-end, which are colored

556 dark red. Watson-Crick base pairing hydrogen-bonds are denoted as dark gray

557 dashed lines, other hydrogen-bonds as red dashed lines. Nsp12 motif C is shown

558 as a yellow-orange backbone ribbon, and the side-chain of D760 is shown as

559 atomic spheres.

560 **A.** The RTC is in a post-translocated state, with the A-U base pair at the p-RNA

561 3'-end in the -1 position (14).

562 **B.** The BTC$_5$(local) RNA is translocated compared to the RTC; the base pair

563 corresponding to A-U at the 3'-end of the RTC RNA in the -1 position is in the -2

564 position of the BTC RNA. A C-A mismatch occupies the BTC -1 site. The +1, +2

565 and +3 mismatched C's trail into the RdRp NTP-entry tunnel (denoted by black

566 squiggly lines). The +4C (present in the BTC$_5$-scaffold; Figure 1A) is exposed to

567 solvent, disordered and not modelled.

29

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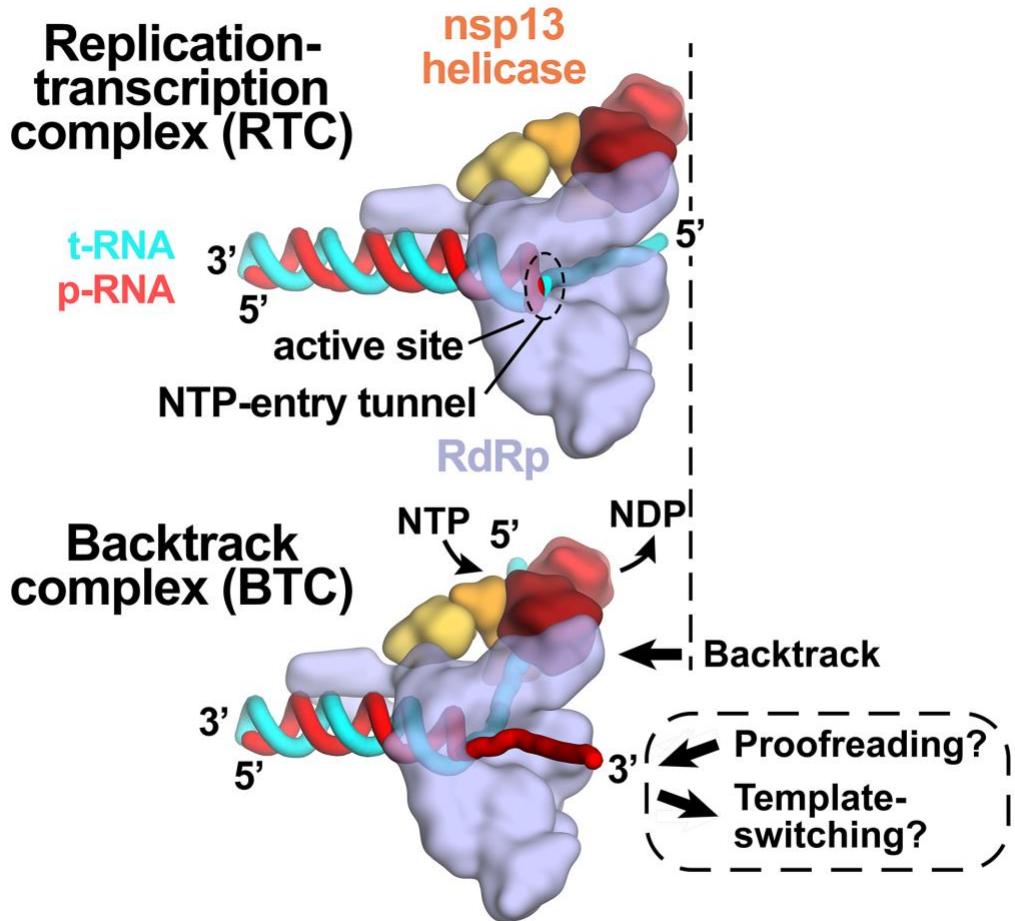
Fig. 6. Role of backtracking in proofreading and template-switching during

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).

32

601 **Protein expression and purification.**

602 *SARS-CoV-2 nsp12.* SARS-CoV-2 nsp12 was expressed and purified as

603 described (1). A pRSFDuet-1 plasmid expressing SARS-CoV-2 His$_6$-SUMO-

604 nsp12 (Addgene plasmid 159107) was transformed into *Escherichia coli* (*Eco*)

605 BL21-CodonPlus cells (Agilent). Cells were grown, followed by the addition of

606 isopropyl β-d-1-thiogalactopyranoside (IPTG) to induce protein expression

607 overnight. Cells were collected by centrifugation, resuspended and lysed in a

608 continuous-flow French press (Avestin). The lysate was cleared by centrifugation,

609 loaded onto a HiTrap Heparin HP column (Cytiva), and then eluted using a salt

610 gradient. The fractions containing nsp12 were pooled and loaded onto a HisTrap

611 HP column (Cytiva), washed, and eluted. Eluted nsp12 was dialyzed overnight in

612 the presence of His$_6$-Ulp1 SUMO protease. Cleaved nsp12 was passed through

613 a HisTrap HP column (Cytiva). Flow-through was collected, concentrated by

614 centrifugal filtration (Amicon), and loaded on a Superdex 200 Hiload 16/600

615 (Cytiva) for size-exclusion chromatography. Glycerol was added to the purified

616 nsp12, aliquoted, flash frozen with liquid N$_2$, and stored at -80°C.

617 *SARS-CoV-2 nsp7/8.* SARS-CoV-2 nsp7/8 was expressed and purified as

618 described (1). The pCDFDuet-1 plasmid expressing SARS-CoV-2 His$_6$-ppx-

619 nsp7/8 (ppx is a Prescission Protease cleavage site; Addgene plasmid 159092)

620 was transformed into *Eco* BL21(DE3). Cells were grown and protein expression

621 was induced overnight by the addition of IPTG. Cells were collected by

622 centrifugation, resuspended, and lysed in a continuous-flow French press

623 (Avestin). The lysate was cleared by centrifugation, then loaded onto a HisTrap

624 HP column (Cytiva), washed, and eluted. Eluted nsp7/8 was dialyzed overnight in

625 the presence of His$_6$-Prescission Protease to cleave the His$_6$-tag. Cleaved nsp7/8

626 was passed through a HisTrap HP column (Cytiva). Flow-through was collected,

627 concentrated by centrifugal filtration (Amicon), and loaded onto a Superdex 75

628 Hiload 16/600 (Cytiva). Glycerol was added to the purified nsp7/8, aliquoted,

629 flash frozen with liquid N$_2$, and stored at -80°C.

33

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

34

660   was loaded into a gold-coated quartz capillary tip that was prepared in-house and

661   then electrosprayed into an Exactive Plus with extended mass range (EMR)

662   instrument (Thermo Fisher Scientific) with a static direct infusion nanospray

663   source (2). The MS parameters used: spray voltage, 1.2 kV; capillary

664   temperature, 150 °C; in-source dissociation, 0 V; S-lens RF level, 200; resolving

665   power, 17,500 at m/z of 200; AGC target, 1×10^6; maximum injection time,

666   200 ms; number of microscans, 5; injection flatapole, 6 V; interflatapole, 4 V; bent

667   flatapole, 4 V; high energy collision dissociation (HCD), 200 V; ultrahigh vacuum

668   pressure, 7.2×10^{-10} mbar; total number of scans, at least 100. Mass calibration

669   in positive EMR mode was performed using cesium iodide. For data processing,

670   the acquired MS spectra were visualized using Thermo Xcalibur Qual Browser

671   (v. 4.2.47). MS spectra deconvolution was performed either manually or using

672   the software UniDec v. 4.2.0 (3, 4). The following parameters were used for the

673   UniDec processing: m/z range, 7,000 – 10,000 Th; background subtraction,

674   subtract curved at 100; smooth charge state distribution, enabled; peak shape

675   function, Gaussian; Beta Softmax distribution parameter, 20.

676   The expected masses for the component proteins based on previous nMS

677   experiments (1) include nsp7: 9,137 Da; nsp8 (N-terminal Met lost): 21,881 Da;

678   nsp13 (post-protease cleavage, has three Zn^{2+} ions coordinated with

679   9 deprotonated cysteine residues): 67,464 Da, and nsp12 (has two Zn^{2+} ions

680   coordinated with 6 deprotonated cysteine residues): 106,785 Da. The mass of

681   the assembled RNA duplex scaffold is 30,512 Da.

682   Experimental masses were reported as the average mass ± standard

683   deviation (S.D.) across all the calculated mass values within the observed charge

684   state series. Mass accuracies were calculated as the % difference between the

685   measured and expected masses relative to the expected mass. The observed

686   mass accuracies ranged from 0.016 – 0.035%.

687

688   **Preparation of SARS-CoV-2 nsp13-BTC₅ for Cryo-EM.** Purified nsp12 and

689   nsp7/8 were mixed in a 1:3 molar ratio and incubated at 22° C for 15 minutes.

35

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° 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-

36

720   normalized, drift-corrected, summed, and dose-weighted using MotionCor2 (7).

721   The contrast transfer function (CTF) was estimated for each summed image

722   using the Patch CTF module in cryoSPARC v2.15.0 (8) Particles were picked

723   and extracted from the dose-weighted images with box size of 256 px using

724   cryoSPARC Blob Picker and Particle Extraction. The entire dataset consisted of

725   10,685 motion-corrected images with 4,961,691 particles. Particles were sorted

726   using cryoSPARC 2D classification (N=100), resulting in 2,412,034 curated

727   particles. Initial models (Ref 1: decoy 1, Ref 2: complex, Ref 3: decoy 2;

728   SI Appendix; Fig. S2) were generated using cryoSPARC *ab initio* Reconstruction

729   on a subset of 85,398 particles. Particles were further curated using Ref 1-3 as

730   3D templates for cryoSPARC Heterogeneous Refinement (N=6), resulting in the

731   following: class1 (Ref 1), 258,097 particles; class2 (Ref 1), 263,966 particles;

732   class3 (Ref 2), 668,743 particles; class4 (Ref 2), 665,480 particles; class5 (Ref

733   3), 280,933 particles; class6 (Ref 3), 274,815 particles. Particles from class3 and

734   class4 were combined and further curated with another round of Heterogeneous

735   Refinement (N=6), resulting in the following: class1 (Ref 1), 67,639 particles;

736   class2 (Ref 1), 61,097 particles; class3 (Ref 2), 553,368 particles; class4 (Ref 2),

737   554,581 particles; class5 (Ref 3), 42,114 particles; class6 (Ref 3), 55,424

738   particles. Curated particles from class3 and class4 were combined, re-extracted

739   with a box size of 320 px, and further classified using Ref 2 as a 3D template for

740   cryoSPARC Heterogeneous Refinement (N=4). Classes from this round of

741   Heterogeneous Refinement (N=4) were as follows: class1 (Ref 2), 871,163

742   particles; class2 (Ref 2), 77,769 particles; class3 (Ref 2), 61,489 particles; class4

743   (Ref 2), 64,026 particles. Particles from class1 and class2 were combined and

744   further sorted using Heterogeneous Refinement (N=4) using class maps as

745   templates, resulting in the following: class1, 134,536 particles; class2, 270,170

746   particles; class3, 294,162 particles; class4, 172,295 particles. Classification

747   revealed two unique classes: $nsp13_1$-BTC (class1 and class2) and $nsp13_2$-BTC

748   (class3 and class4). Particles within each class were further processed using

749   RELION 3.1-beta Bayesian Polishing(9, 10). Polished particles were refined

750   using cryoSPARC Non-uniform Refinement, resulting in structures with the

37

751   following particle counts and nominal resolutions: $nsp13_1$-BTC (404,706 particles;

752   3.40 Å) and $nsp13_2$-BTC (466,457 particles; 3.45 Å).

753   To improve the resolution of the RNA in the BTC, particles from both classes

754   were combined in a cryoSPARC Non-uniform Refinement and density

755   corresponding to nsp13 was subtracted. Subtracted particles were further refined

756   with cryoSPARC Local Refinement using a mask encompassing the BTC and a

757   fulcrum point defined on the backtracked RNA. This map, BTC_5(local), contained

758   871,163 particles with a nominal resolution of 3.23 Å.

759   To improve the density of nsp13.2 in the $nsp13_2$-BTC map, particles were

760   subtracted using a mask defined around nsp13.2, leaving residual signal for only

761   nsp13.2. Subtracted particles were classified (N=4) in RELION 3.1 beta using a

762   mask around nsp13.2, resulting in the following classes: class1, 71,607 particles;

763   class2, 163,540 particles; class3, 176,461 particles; class4, 54,849 particles.

764   Subtracted particles in class1 and class2 were combined and reverted back to

765   the original particles, followed by refinement using cryoSPARC Non-uniform

766   Refinement. The resulting map of $nsp13_2$-BTC contains 235,147 particles with

767   nominal resolution of 3.59 Å. Local resolution calculations were generated using

768   blocres and blocfilt from the Bsoft package (11).

769

770   **Model building and refinement.** Initial models were derived from PDB: 6XEZ

771   (1). The models were manually fit into the cryo-EM density maps using Chimera

772   (12) and rigid-body and real-space refined using Phenix real_space_refine (13).

773   For real-space refinement, rigid body refinement was followed by all-atom and B-

774   factor refinement with Ramachandran and secondary structure restraints. Models

775   were inspected and modified in Coot (14).

776

777   **4-thiouridine crosslinking.** Nsp12 or nsp12-D760A were incubated with 3-fold

778   molar excess of nsp7/8 to assemble holo-RdRp (2 µM final) in transcription

779   buffer. The resulting holo-RdRp was added to a modified RNA scaffold

780 (SI Appendix; Fig. S5A) containing a photoactivable 4-thiouridine base (Horizon

781 Discovery) which was 5'-labelled by T4-polynucleotide kinase (New England

782 Biolabs) with γ-^{32}P-ATP (Perkin-Elmer). The holo-RdRp/RNA complex was left to

783 incubate for 5 minutes at 30°C in the dark. Nsp13 and ATP were added to a final

784 concentration of 2 uM and 2 mM, respectively, and incubated for five minutes at

785 30°C in the dark. The reaction mixture was transferred to a Parafilm covered

786 aluminum block at 4°C and irradiated with a 365-nm handheld UV lamp.

787 Reactions were quenched with LDS sample loading buffer (ThermoFisher

788 Scientific) and analyzed by gel electrophoresis on a NuPAGE 4-12% Bis-Tris gel

789 (ThermoFisher) at 150 Volts for 1 hour and visualized by autoradioagraphy.

790

Molecular dynamics simulations

792 *General simulation setup and parameterization.* Proteins, ADP, and ions were

793 parameterized with the DES-Amber SF1.0 force field (15). RNAs were

794 parameterized with the Amber ff14 RNA force field (16) with modified

795 electrostatic, van der Waals, and torsional parameters to more accurately

796 reproduce the energetics of nucleobase stacking (17). The systems were

797 solvated with water parameterized with the TIP4P-D water model (18) and

798 neutralized with 150 mM NaCl buffer. The systems each contained ~887,000

799 atoms in a 190×190×190 Å cubic box.

800 Systems were first equilibrated on GPU Desmond using a mixed NVT/NPT

801 schedule (19), followed by a 1 μs relaxation simulation on Anton, a special-

802 purpose machine for molecular dynamics simulations (20). All production

803 simulations were performed on Anton and initiated from the last frame of the

804 relaxation simulation. Production simulations were performed in the NPT

805 ensemble at 310 K using the Martyna-Tobias-Klein barostat (21). The simulation

806 time step was 2.5 fs, and a modified r-RESPA integrator (22, 23) was used in

807 which long-range electrostatic interactions were evaluated every three time

808 steps. Electrostatic forces were calculated using the *u*-series method (24). A 9-Å

809 cutoff was applied for the van der Waals calculations.

810 *System preparation.* The nsp13$_2$-BTC$_{-1U+1C}$ and the nsp13$_2$-BTC$_{-1U+1U}$ complexes

811 were prepared from the cryo-EM structure of the nsp13$_2$-BTC$_5$. AlF$_3$ and

812 CHAPSO were removed. Cytosines at the +2 and +3 positions of the p-RNA

813 were removed, and the cytosine at −1 was mutated to uracil. The resulting p-

814 RNA had a matched −1U and a mismatched +1C in nsp13$_2$-BTC$_{-1U+1C}$, and a

815 matched −1U and +1U in nsp13$_2$-BTC$_{-1U+1U}$. Missing loops and termini in

816 proteins were capped with ACE/NME capping groups. The two complexes were

817 prepared for simulation using the Protein Preparation Wizard in Schrödinger

818 Maestro. After a 1 µs relaxation simulation of the nsp13$_2$-BTC$_{-1U+1C}$ complex, the

819 −1U of the p-RNA formed a Watson-Crick base pair with the −1A in the t-RNA,

820 and the +1C of p-RNA formed a non-Watson-Crick C-A hydrogen bond with the

821 +1A of the t-RNA in the active site. After a 1 µs relaxation simulation of the

822 nsp13$_2$-BTC$_{-1U+1U}$ complex, the −1U and +1U of the p-RNA formed Watson-Crick

823 base pairs with the −1A and +1A of the t-RNA respectively.

824 *Simulation analysis.* All simulations were visually inspected using the in-house

825 visualization software Firefly. The average root-mean-square deviation (RMSD)

826 was calculated for +1C (or +1U) of the p-RNA between the last frame of the 1 µs

827 relaxation simulation and instantaneous structures from the trajectories, aligned

828 on the entire nps12 module.

829

830 **Quantification and statistical analysis.** The nMS spectra were visualized using

831 Thermo Xcalibur Qual Browser (versions 3.0.63 and 4.2.27), deconvolved using

832 UniDec versions 3.2 and 4.1 (3, 4) and plotted using the m/z software

833 (Proteometrics LLC, New York, NY). Experimental masses (SI Appendix;

834 Fig. S1B and C) were reported as the average mass ± standard deviation across

835 all the calculated mass values obtained within the observed charge state

836 distribution.

837 The local resolution of the cryo-EM maps (SI Appendix; Fig. S3B-D) was

838 estimated using blocres (11) with the following parameters: box size 15, sampling

839 1.1, and cutoff 0.5. Directional 3D FSC (SI Appendix; Fig. S3H-J) were calculated

840   by 3DFSC (25). The quantification and statistical analyses for model refinement

841   and validation were generated using MolProbity (26) and PHENIX (13).

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

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

2

865 **SUPPLEMENTAL FIGURES**

866

A

RNA scaffold	RTC	BTC$_5$	RTC	BTC$_5$	RTC	BTC$_5$	RTC	BTC$_5$
holo-RdRp	WT	WT	D760A	D760A	WT	WT	D760A	D760A
Nsp13 (ADP-AlF$_3$)					WT	WT	WT	WT

B

nsp13$_1$-BTC$_6$ + x MgADP

nsp13$_2$-BTC$_6$ + x MgADPAlF$_3$

C
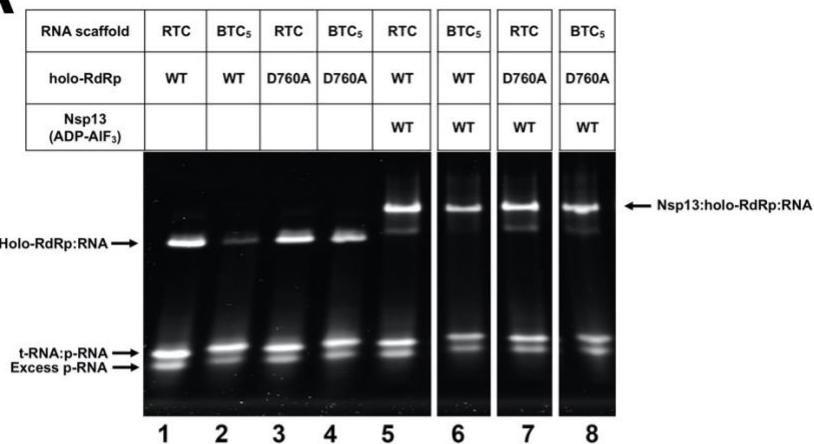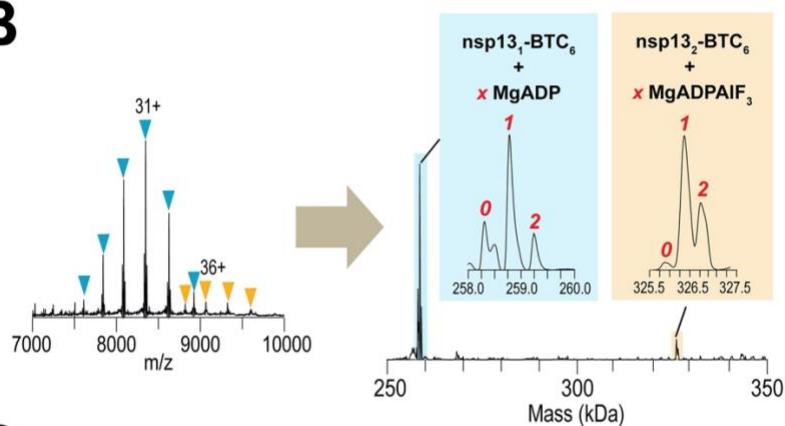
Mass assignments from nMS analysis of BTC$_6$ incubated with nsp13 and MgADPAlF$_3$

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

867

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

869 **BTC.**

870 **A.** A native gel electrophoretic mobility shift assay reveals that wt-holo-RdRp requires

871 nsp13(ADP-AlF$_3$) to bind the BTC$_5$-scaffold efficiently (compare lanes 1, 2, and 6) but

872 holo-RdRp with nsp12-D760A does not require nsp13 (lane 4).

873 **B.** The nMS spectrum and the deconvolved mass spectrum showing assembly of stable

874 nsp13-BTC$_6$ complexes. The peak for the nsp13$_2$-BTC$_6$ assembly is present at about

875 ~9% intensity relative to the predominant peak from nsp13$_1$-BTC$_6$.

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

877

878

879

880

881

882

883

884

885

886

4
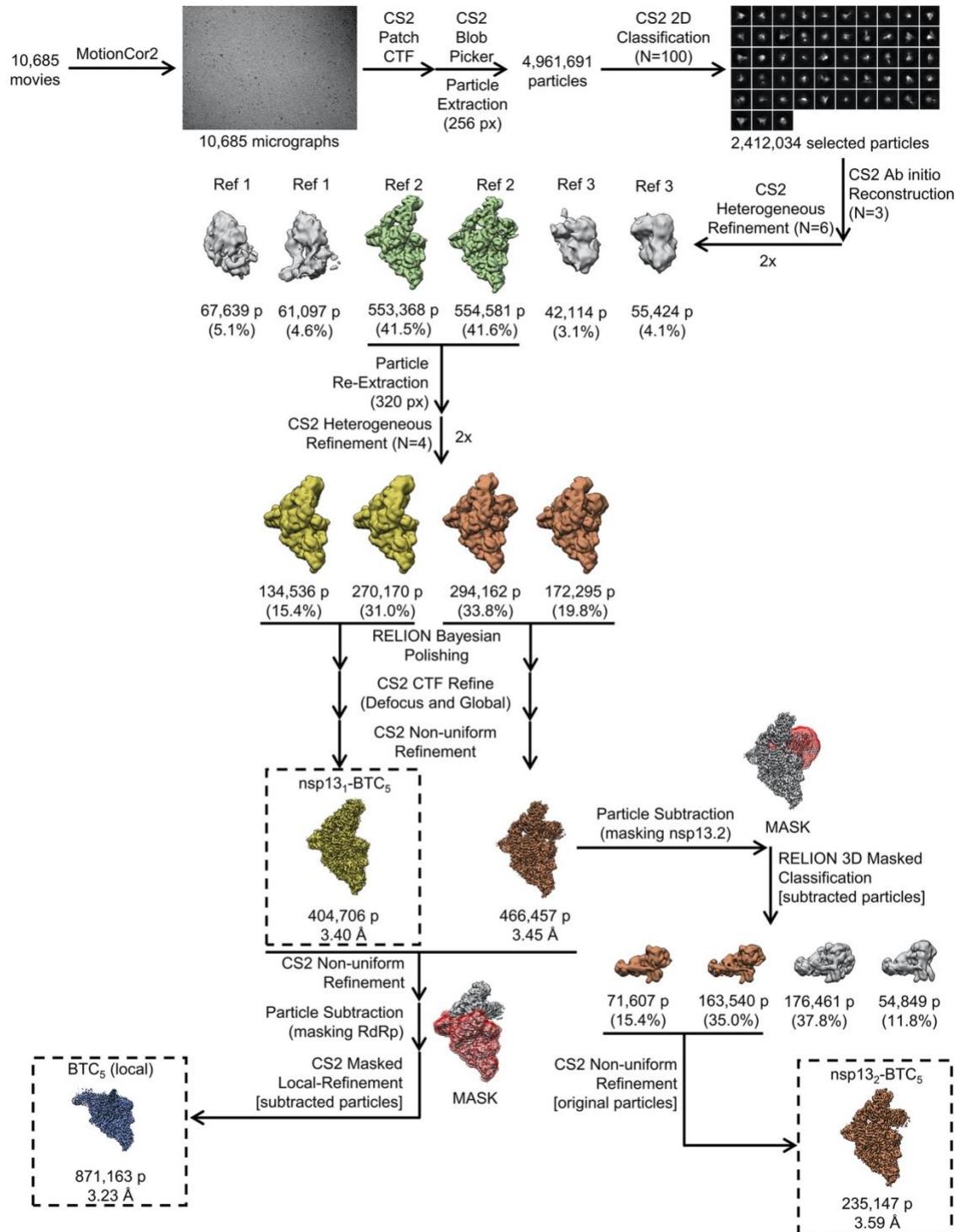
Figure S2

888 **Fig. S2. Cryo-EM processing pipeline.**

Figure S3

6
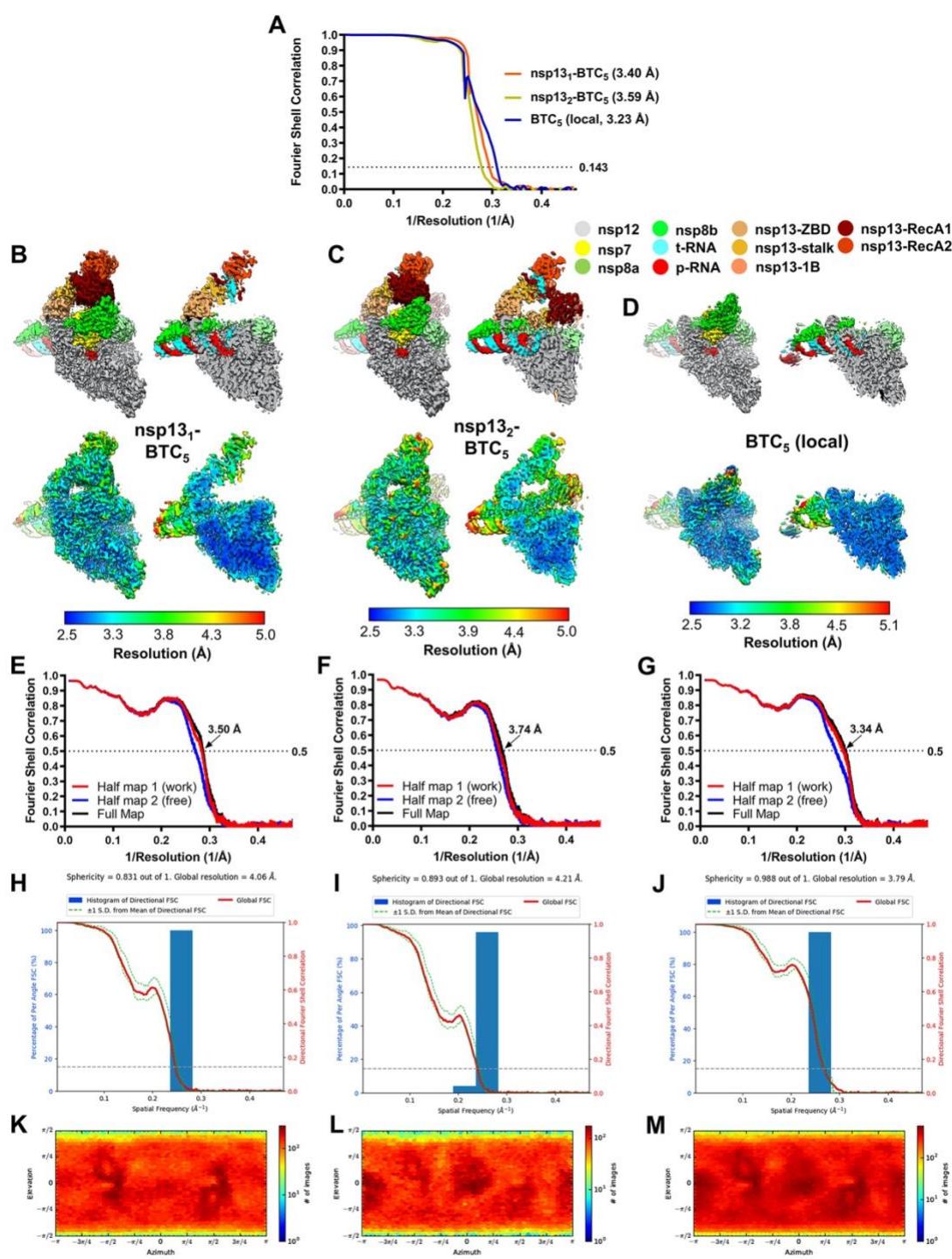
891 **Fig. S3. Cryo-EM analysis.**

892 **A.** Gold-standard FSC plots for $nsp13_1$-BTC_5, $nsp13_2$-BTC_5, and BTC_5(local), calculated
893 by comparing two independently determined half-maps from cryoSPARC (8). The
894 dotted line represents the 0.143 FSC cutoff.

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

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

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

899 **B.** $Nsp13_1$-BTC_5.

900 **C.** $Nsp13_2$-BTC_5.

901 **D.** BTC_5(local).

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

904 **E.** $Nsp13_1$-BTC_5.

905 **F.** $Nsp13_2$-BTC_5.

906 **G.** BTC_5(local).

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

908 **H.** $Nsp13_1$-BTC_5.

909 **I.** $Nsp13_2$-BTC_5.

910 **J.** BTC_5(local).

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

914 **K.** $Nsp13_1$-BTC_5.

915 **L.** $Nsp13_2$-BTC_5.

916 **M.** BTC_5(local).

917

918

919

920
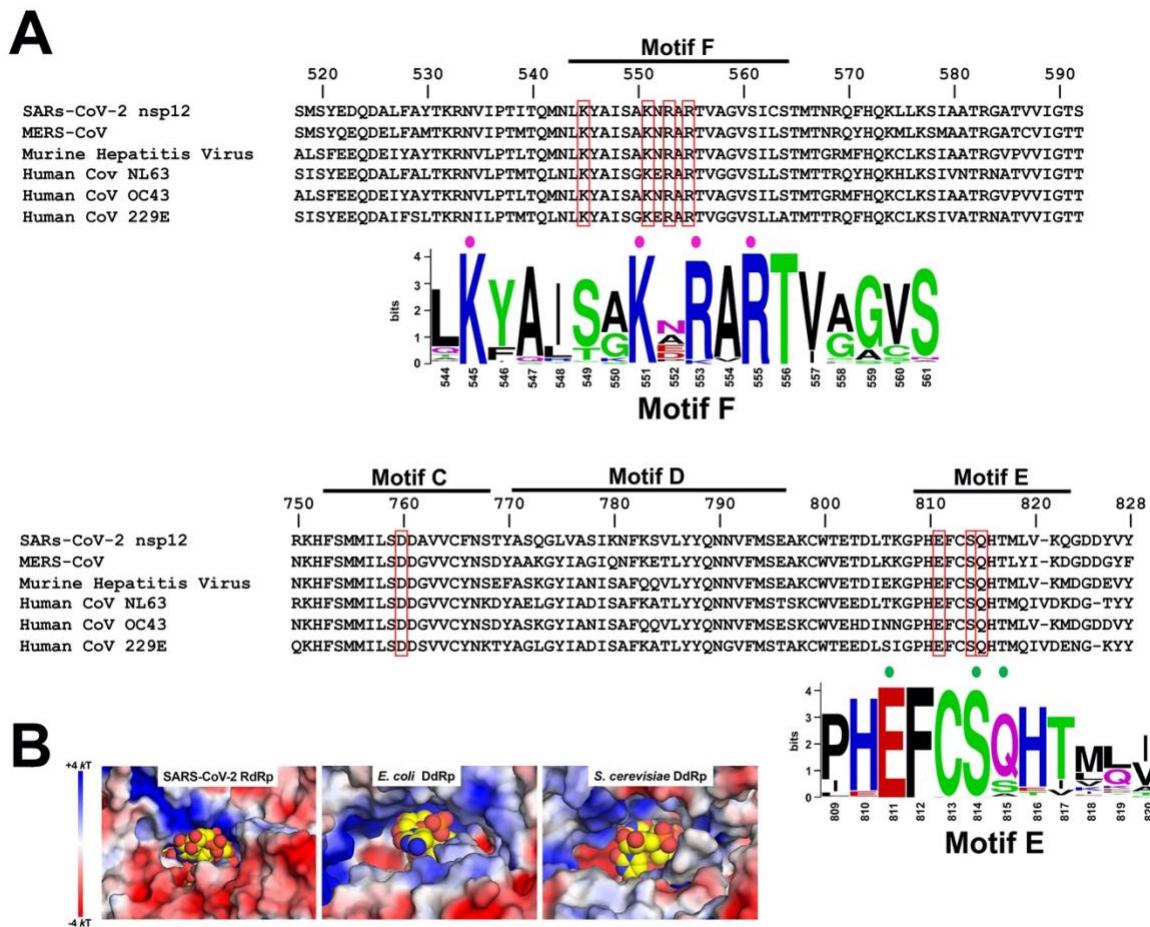
Figure S4

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.

8

Figure S5

937
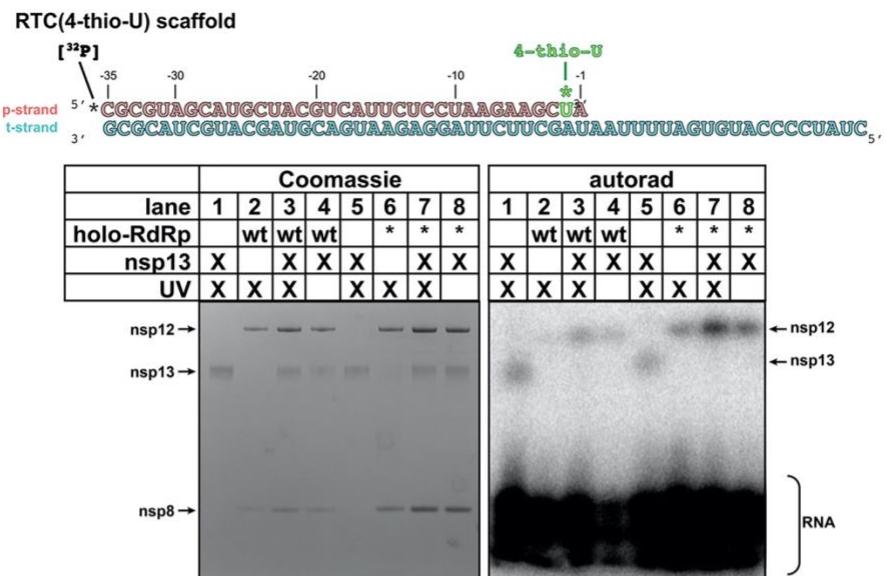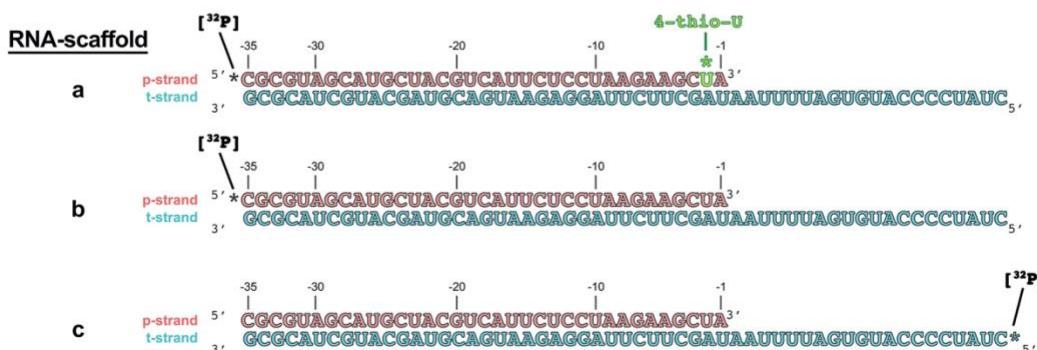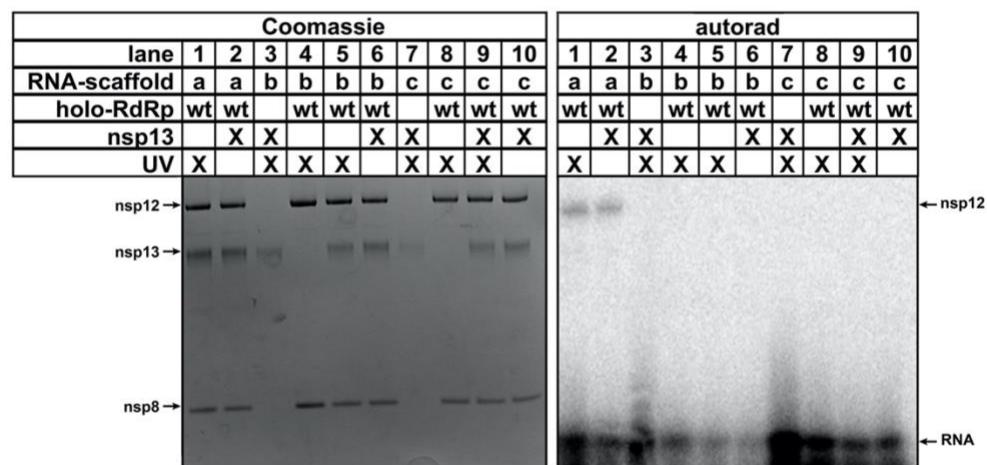
Fig. S5. 4-thio-U crosslinking analysis.

9

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

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

956

957

10
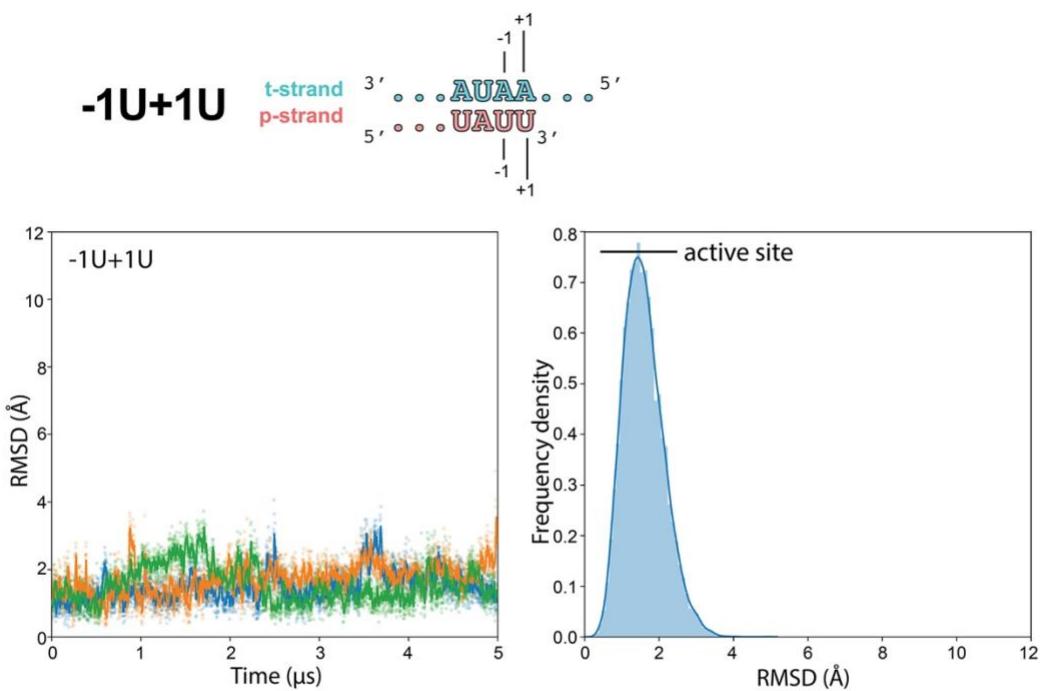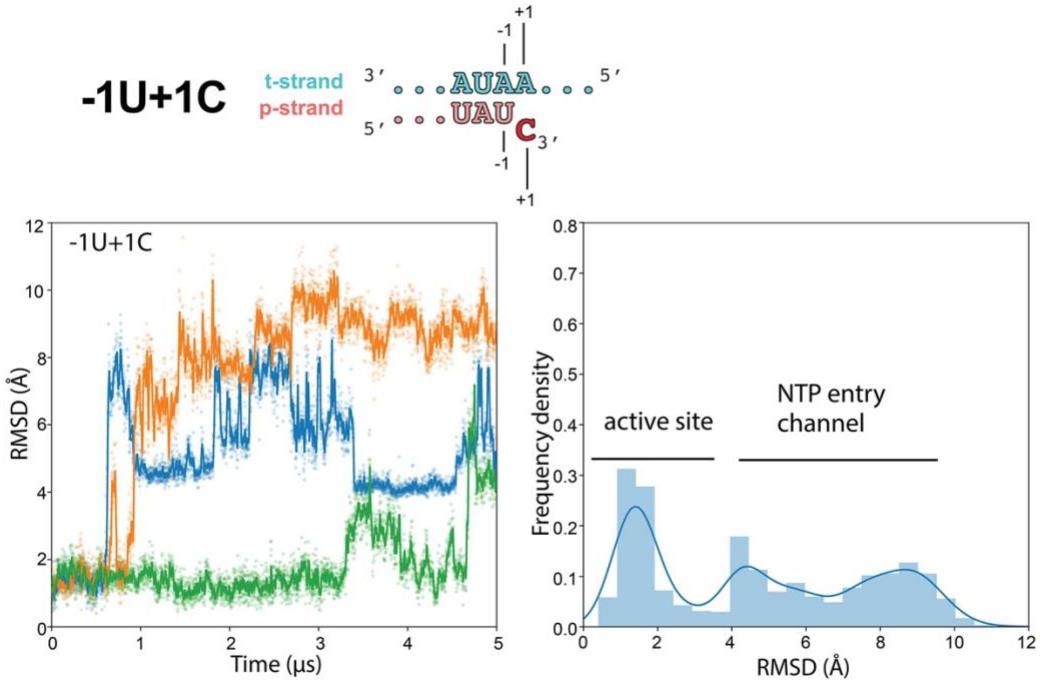
Figure S6

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

961 **nsp13$_2$-BTC$_{-1U+1U}$.**

962 Molecular dynamics simulations of the nsp13$_2$-BTC$_{-1U+1C}$ (*top*) and nsp13$_2$-BTC$_{-1U+1U}$

963 (*bottom*) complexes. The schematics illustrate the active-site proximal nucleotides in

964 each modeled complex. Each complex was simulated with 3 replicates. RMSD values

965 plotted as a function of time represent the heavy-atom RMSD of the +1 nucleotide of the

966 p-RNA (+1C for nsp13$_2$-BTC$_{-1U+1C}$ or matched +1U for nsp13$_2$-BTC$_{-1U+1U}$) compared with

967 the starting configuration (see Methods). The RMSD histograms (plotted on the right)

968 are aggregates of all 3 replicates.

969 (*top*) Nsp13$_2$-BTC$_{-1U+1C}$. As shown in Figure 5C, the mismatched p-RNA +1C spends

970 about 60% of the time frayed from the t-RNA +1A and near or in the NTP-entry tunnel

971 (RMSD ≥ ~3.5 Å).

972 (*bottom*) Nsp13$_2$-BTC$_{-1U+1U}$. With the p-RNA +1U matched with the t-RNA +1A for

973 Watson-Crick base pairing, the p-RNA +1U does not fray and spends all of its time in

974 the vicinity of the RdRp active site and base paired with the t-RNA.

975

976

977

978

979

980

981

982

983

984

985

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).

13

1014   10. J. Zivanov, *et al.*, New tools for automated high-resolution cryo-EM structure
1015   determination in RELION-3. *eLife* **7** (2018).

1016   11. G. Cardone, J. B. Heymann, A. C. Steven, One number does not fit all: mapping
1017   local variations in resolution in cryo-EM reconstructions. *Journal of structural biology*
1018   **184**, 226–236 (2013).

1019   12. E. F. Pettersen, *et al.*, UCSF Chimera--a visualization system for exploratory
1020   research and analysis. *Journal of computational chemistry* **25**, 1605–1612 (2004).

1021   13. P. D. Adams, *et al.*, PHENIX: a comprehensive Python-based system for
1022   macromolecular structure solution. *Acta Crystallographica Section D Biological*
1023   *Crystallography* **66**, 213–221 (2010).

1024   14. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta*
1025   *Crystallographica Section D Biological Crystallography* **60**, 2126–2132 (2004).

1026   15. S. Piana, P. Robustelli, D. Tan, S. Chen, D. E. Shaw, Development of a Force Field
1027   for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *J Chem*
1028   *Theory Comput* **16**, 2494–2507 (2020).

1029   16. J. A. Maier, *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and
1030   Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713.

1031   17. D. Tan, S. Piana, R. M. Dirks, D. E. Shaw, RNA force field with accuracy
1032   comparable to state-of-the-art protein force fields. *Proc National Acad Sci* **115**,
1033   201713027 (2018).

1034   18. S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, Water Dispersion Interactions
1035   Strongly Influence Simulated Structural Properties of Disordered Protein States. *J Phys*
1036   *Chem B* **119**, 5113–5123 (2015).

1037   19. K. J. Bowers, *et al.*, Scalable Algorithms for Molecular Dynamics Simulations on
1038   Commodity Clusters. *Acm Ieee Sc 2006 Conf Sc'06*, 43–43 (2006).

1039   20. D. E. Shaw, *et al.*, Anton 2: Raising the Bar for Performance and Programmability in
1040   a Special-Purpose Molecular Dynamics Supercomputer. *Sc14 Int Conf High Perform*
1041   *Comput Netw Storage Analysis*, 41–53 (2014).

1042   21. G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics
1043   algorithms. *J Chem Phys* **101**, 4177–4189 (1994).

1044   22. C. Predescu, *et al.*, Computationally efficient molecular dynamics integrators with
1045   improved sampling accuracy. *Mol Phys* **110**, 967–983 (2012).

1046   23. M. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular
1047   dynamics. *J Chem Phys* **97**, 1990–2001 (1992).

1048   24. C. Predescu, *et al.*, The u -series: A separable decomposition for electrostatics
1049   computation with improved accuracy. *J Chem Phys* **152**, 084113 (2020).

1050   25. Y. Z. Tan, *et al.*, Addressing preferred specimen orientation in single-particle cryo-
1051   EM through tilting. *Nature methods* **14**, 793–796 (2017).

1052   26. V. B. Chen, *et al.*, MolProbity: all-atom structure validation for macromolecular
1053   crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**, 12–21
1054   (2010).

1055   27. A. J. W. te Velthuis, Common and unique features of viral RNA-dependent
1056   polymerases. *Cell Mol Life Sci Cmls* **71**, 4403–20 (2014).

1057   28. T. D. Schneider, R. M. Stephens, Sequence logos: a new way to display consensus
1058   sequences. *Nucleic Acids Research* **18**, 6097–6100 (1990).

1059   29. M. Abdelkareem, *et al.*, Structural Basis of Transcription: RNA Polymerase
1060   Backtracking and Its Reactivation. *Mol Cell* **75**, 298-309.e4 (2019).

1061   30. D. Wang, *et al.*, Structural basis of transcription: backtracked RNA polymerase II at
1062   3.4 angstrom resolution. *Science* **324**, 1203–1206 (2009).

1063   31. E. Jurrus, *et al.*, Improvements to the APBS biomolecular solvation software suite.
1064   *Protein Sci* **27**, 112–128 (2018).

1065

1066

1067

1068