SARS-CoV-2 RdRp is a versatile enzyme with proofreading activity and ability to incorporate NHC into RNA by using diphosphate form molnupiravir as a substrate

Maofeng Wang, Cancan Wu, Nan Liu, Fengyu Zhang, Hongjie Dong, Shuai Wang, Min Chen, Xiaoqiong Jiang & Lichuan Gu*

State Key Laboratory of Microbial Technology, Shandong University, 72 Binhai Road, Qingdao 266237, P.R. China

*To whom correspondence should be addressed: L.G. (e-mail: lcgu@sdu.edu.cn)

Abstract

The coronavirus disease 2019 (COVID-19) has been ravaging throughout the world for almost two years and has severely impaired both human health and the economy. The causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) employs the viral RNA-dependent RNA polymerase (RdRp) complex for genome replication and transcription, making RdRp an appealing target for antiviral drug development. Although the structure of the RdRp complex has been determined, the function of RdRp has not been fully characterized. Here we reveal that in addition to RNA dependent RNA polymerase activity, RdRp also shows exoribonuclease activity and consequently proofreading activity. We observed that RdRp and nsp14-ExoN, when combined, exhibit higher proofreading activity compared to RdRp alone. Moreover, RdRp can recognize and utilize nucleoside diphosphate (NDP) as substrate to synthesize RNA and can also incorporate β-d-N4-hydroxycytidine (NHC) into RNA while using diphosphate form molnupiravir as substrate.

Introduction

The pandemic of coronavirus disease 2019 (COVID-19) has severely impacted human health and global economy since the outbreak in late 2019[1,2]. The causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is highly homologous to 2002 SARS-CoV with a similarity of 79.5%, and considered as new human infection along with β-coronavirus[3,4]. As of the date of writing, more than 254.02 million infections and more than 5.11 million deaths have been confirmed (https://www.worldometers.info/coronavirus/?utm_campaign). Although many kinds of vaccines have been developed to contain the pandemic, the rapidly spreading mutant strains arise serious concern that the vaccines would not yield the expected outcome[5-8]. As result, there is an urgent need to find effective medication or treatment to counter the virus[9-11].

SARS-CoV-2 is a positive-sense single-stranded RNA virus with a large genome of approximately 30 kb organized in 14 open reading frames[9,12]. Replication of the genome and transcription of genes are all dependent on a protein complex known as RNA-dependent RNA polymerase (RdRp). The RdRp of SARS-CoV-2 contains a catalytic subunit (non-structural protein 12, nsp12) and two accessory subunits (nsp7 and nsp8)[13,14]. Due to the significant role of RdRp and the lack of homologues in humans, RdRp has become the most appealing target for anti-coronavirus drug development[15,16].

The structure of the nsp12-nsp7-nsp8 complex has been determined in various forms, which includes active and inhibited forms[17-21]. The structure of SARS-CoV-2 RdRp

complex indicate that nsp12 contains an N-terminal β hairpin (residues 31 to 50) and a nucleotide transferase domain (NiRAN; residues 51 to 250)[20-22]. After the NiRAN domain is an interface domain (residues 251 to 365), connected to the RdRp catalytic domain (residues 366 to 920). The catalytic domain of RdRp is described as a cup-shaped right-hand conformation, composed of fingers, palm and thumb subdomains, forming a closed circle[23]. The combination of nsp7 and nsp8 stabilizes the closed conformation. The active site of the catalytic domain of SARS-CoV-2 RdRp is formed by seven conserved catalytic motifs (motifs A to G) which are necessary for catalysis[24-28].

These structures are crucial to understand the mechanism of RdRp activity and inhibition by remdesivir or other inhibitors[23]. The understanding of these structures can play an influential role in the development of more effective anti-coronavirus drugs. However, an analysis of the structure does not mean we have complete knowledge of RdRp. During an experiment to screen an array of FDA-approved drugs for RdRp inhibitors, we unexpectedly discovered that RdRp exhibits exoribonuclease activity. Additional research has indicated that RdRp displays proofreading activity as well. Further observation suggests that RdRp and nsp14-ExoN, when combined, exhibit higher proofreading activity compared to RdRp alone. Moreover, RdRp can also use nucleoside diphosphate (NDP) as substrate to synthesize RNA, which may give SARS-CoV-2 a great advantage for survival when ATP level inside host cells is low. In addition, we proved that molnupiravir is fully functionable in its diphosphate form in addition to triphosphate form. This can imply a new strategy for designing

89 COVID-19 treatment nucleoside analogue drugs with better efficacy and fewer side
90 effects.

91 **Results**

92 **Exoribonuclease activity was observed during characterization of RdRp**
93 **polymerase activity**

94 Since RdRp is by far the most attractive target for anti-coronavirus drug development,
95 thorough characterization of RdRp polymerase activity would greatly without doubt
96 contribute to discovering the effective strategy for drug screening[15,16]. Based on
97 previous research, SARS-CoV-2 RdRp polymerase activity was measured by
98 conducting a conventional RNA elongation assay[23]. When nucleoside triphosphate
99 (NTP) substrate is added, elongation catalyzed by RdRp complex occurred on the
100 RNA duplex, resulting in an intact double-stranded RNA product. Nsp12 alone could
101 not catalyze the extension reaction (Fig.1a). Surprisingly, when NTP was absent from
102 the system for control, the RNA duplex was markedly degraded. Degradation of RNA
103 duplex was inhibited when EDTA was added (Fig.1b). Our results indicated that in
104 addition to polymerase activity, RdRp also displays exoribonuclease activity despite
105 that neither sequence nor structure gives any clue of the existence of a functional
106 domain.

107 **SARS-CoV-2 RdRp also has proofreading activity**

108 Our finding that SARS-CoV-2 RdRp exhibits exoribonuclease activity prompted us to
109 speculate that RdRp may also possess proofreading activity. To test our speculation,
110 an RNA duplex (M1 RNA) consisting of a 13 nt and a 23 nt RNA with a U: G

111 mismatch at the 3'-end of the extending strand of RNA was prepared (Fig.2b). An
112 extension assay was then performed by using M1 RNA as the template. Result
113 showed that RdRp successfully extended the shorter RNA strand, just as it did with
114 the RNA duplex without mismatch, although the productivity was slightly lower
115 (Fig.3a). We hypothesized two possibilities: first, RdRp simply ignored the mismatch
116 and processed to 3' end leaving the U: G mismatch unchanged; second, RdRp
117 corrected the mismatch and then finished the extension. The nucleotide mismatch
118 always slowed down the extension reaction, no matter the case.
119 To figure out this process, nascent RNA strands were reverse transcribed into cDNAs.
120 The cDNAs were then amplified by 10 cycles of PCR and sequenced by
121 high-throughput sequencing. For each product, at least 20,000 sequences were
122 obtained (SI). The data showed three kinds of nascent RNAs were produced: (1)
123 mutation repaired GCUAUGUGAGAU**G**AAAGUUAACU, 39.29%; (2) mutation
124 ignored GCUAUGUGAGAU**U**AAAGUUAACU, 13.82%; (3) mutation leading to
125 frame shift GCUAUGUGAGAU**UG**AAAGUUAACU, 36.89%.
126 We also challenged RdRp with the RNA template (M2 RNA) consisting of a 13 nt and
127 a 23 nt RNA with a UU: GC double mismatch at the 3'-end of the extending strand of
128 RNA (Fig.2c). In this experiment, RNA extension was observed as well, but with a
129 much lower productivity (Fig.3a). The products were then processed and sequenced
130 as mentioned above. The results included four kinds of RNAs: (1) mutations half
131 repaired GCUAUGUGAGA**UG**AAAGUUAACU, 54.88%; (2) mutations completely
132 repaired GCUAUGUGAGA**CG**AAAGUUAACU, 20.69%; (3) mutation ignored

133 GCUAUGUGAGA**UU**AAAGUUAACU, 3.01%; (4) frame shift,
134 GCUAUGUGAGA**UUG**AAAGUUAACU, 11.42%.
135 The data strongly suggested that SARS-CoV-2 RdRp itself possesses proofreading
136 activity.

137 **Combination of RdRp and nsp14-ExoN exhibited higher proofreading activity**

138 Previous studies indicated that nsp14 is highly conserved within the Coronaviridae
139 family and critical for the SARS-CoV replication fidelity[7,29,30]. Since we discovered
140 that SARS-CoV-2 RdRp itself has proofreading activity, we supposed that the
141 combination of the two proteins may exhibit higher proofreading activity. To confirm
142 this hypothesis, we first tested the exoribonuclease activity of nsp14-nsp10 complex
143 in vitro and chose 300 nM as the proper concentration of nsp14 for the next
144 experiments[31] (Fig.3b).
145 We then conducted a proofreading activity assay for RdRp in the presence of 300 nM
146 of nsp14 with a RNA duplex containing a single mismatch (Fig.3c). High-throughput
147 sequencing indicated three kinds of nascent RNAs were produced: (1) mutation
148 repaired GCUAUGUGAGAU**G**AAAGUUAACU, 56.91%; (2) mutation ignored
149 GCUAUGUGAGAU**U**AAAGUUAACU, 22.85%; (3) mutation leading to frame shift
150 GCUAUGUGAGAU**UG**AAAGUUAACU, 10.13%. In the case of double mismatch,
151 nsp14 did not show significant role in improving the proofreading activity of RdRp:
152 (1) mutations half repaired 61.17%; (2) mutations completely repaired 17.21%; (3)
153 mutation ignored 5.72%; (4) frame shift 5.90% (Fig.3c). These data unambiguously

154 indicated that combination of RdRp and nsp14 is conducive to maintaining higher
155 replication fidelity especially for single mutation correction.

156 **RdRp synthesizes RNA by using NDP as substrate**

157 It is common sense that RNA polymerases synthesize RNA by using four types of
158 nucleoside triphosphate (ATP, UTP, CTP and GTP) as substrate. However, the
159 unexpected discovery of the exoribonuclease activity of RdRp gave us the unusual
160 idea that RdRp may also be able to use nucleoside diphosphate (NDP) as substrate. To
161 verify our hypothesis, we performed a RNA extension assay by using NTP and NDP
162 as substrate respectively. The result was fascinating, for RdRp synthesized the same
163 amount of RNA regardless NTP or NDP as substrate (Fig.4a). This result also raised a
164 question of whether the RNA made from NDP has the right sequence. To answer this
165 question, cDNA was synthesized through reverse transcription of the RNA product,
166 and then amplified by 10 cycles of PCR with primers containing the restriction sites
167 of Xho1 and BamH1. PCR products were then ligated into the vector pET-15b and
168 used to transform *E. coli* DH5α strain. Plasmids extracted from the positive clones
169 were sequenced by Sanger sequencing. The results confirmed that RNA synthesized
170 from NDP has the right sequence (Fig.4b). This means that unlike most other RNA
171 polymerase RdRp uses both NTP and NDP as substrate.

172 **SARS-CoV-2 RdRp incorporates NHC into RNA by using diphosphate form**
173 **molnupiravir as a substrate**

174 Since RdRp has become the most important target for the development of antiviral
175 drugs, many nucleoside analogue drugs have been constructed to target RdRp for

176 COVID-19 treatment[32-35]. Among all these chemicals reported by far, molnupiravir
177 has been viewed as a promising drug candidate. The molecular mechanism of
178 molnupiravir has been determined recently (Fig.5a). Once ingested by patients,
179 molnupiravir undergoes stepwise phosphorylation to yield the active nucleoside
180 triphosphate analogue (β-d-N4-hydroxycytidine (NHC) triphosphate, or called MTP).
181 The RdRp then uses MTP as substrate and incorporates NHC monophosphate into
182 RNA at the position where the nucleotide in the template is A or G, resulting in "error
183 catastrophe" and virus death[36-38].
184 Although the mechanism of molnupiravir has been well established, our finding that
185 RdRp can use NDP as a substrate still raises a question: is NHC diphosphate (MDP)
186 also an active form and recognized by RdRp as a substrate? To address this question,
187 the 13+23 nt RNA duplex was used to perform the RNA extension assay with NDPs
188 and MDP as the substrates in comparison with the extension using NTPs and MTP as
189 the substrates. 13+23 nt RNA duplex allows ten nucleotides (nt) of extension (Fig.5b).
190 In this case, when uridine diphosphate (UDP) was replaced by MDP, the incorporation
191 of M did not obstruct the incorporation of next subsequent nucleotides, RdRp still
192 finished the extension reaction. Similar results were obtained, in the case of using
193 MTP as a substrate (Fig.5c). Thus, it is reasonable to speculate that both MTP and
194 MDP are the active form of monulpiravir inside the human body.

195 **Discussion**

196 Since its outbreak in late 2019, the COVID-19 pandemic has been plaguing people for
197 two years with no sign of abating[3,12]. Although many kinds of vaccines have been

198 developed to suppress the pandemic, they had not produced the anticipated result to
199 withhold the spread of SARS-CoV-2. Development of potent antivirals against
200 SARS-CoV-2 is still an urgent need[10,11].

201 Since RNA-dependent RNA polymerase (RdRp) is critical for viral genome
202 replication and transcription and conserved in RNA virus species, RdRp has become
203 the most appealing target for antivirals development[11,15,23,26]. Thoroughly
204 characterization of RdRp both structurally and biochemically would no doubt be
205 beneficial to drug development. The discovery that RdRp has exoribonuclease activity
206 and proofreading activity is a complete surprise. Although it is general knowledge that
207 SARS-CoV-2 has proofreading activity, this activity was thought to be attributed to
208 nsp14-ExoN[29-31,39]. Sequence analysis of RdRp does not give any sign of the existence
209 of an exoribonuclease domain or a domain of unknown function[21,23]. We speculate the
210 exoribonuclease activity to be the "reverse reaction" of the polymerization. More
211 research is needed to confirm and explain the mechanism.

212 The substrate usage of RdRp surprised us greatly. As far as we know, all RNA
213 polymerases, including DNA dependent and RNA dependent RNA polymerases, use
214 nucleoside triphosphate (NTP) as substrate to synthesize RNA under the guidance of
215 the template chain[40]. Our finding that SARS-CoV-2 RdRp uses both NTP and NDP as
216 substrate to synthesize RNA with equal efficiency makes it an extremely unique RNA
217 polymerase. By far, we don't know why SARS-CoV-2 RdRp is so distinctive or if
218 there are any other enzyme-possessing similar features. We speculate that the ability
219 to use NDP as substrate could be a great advantage for SARS-CoV-2. During the late

220 stage of infection, the host cell may fail to produce enough ATP and thus other NTPs.
221 By using NDP as substrate, SARS-CoV-2 can continue the process of genome
222 replication and assembly of progeny virus while most of the metabolic activities stall.
223 The newly discovered feature of RdRp may profoundly change the strategy of
224 antiviral drug design. Many antiviral drugs are nucleoside analogues that target viral
225 polymerases as inhibitors to terminate RNA chain elongation[41,42]. However, such
226 chain-terminating antivirals often do not work as effectively as expected in the
227 treatment of COVID-19 because SARS-CoV-2 has exonucleolytic proofreading
228 activity that can remove the incorporated inhibitors[23,32,33,42-44]. It was thought that the
229 proofreading activity was carried out by nsp14-ExoN and incapacitating nsp14-ExoN
230 may be a way to deal with this issue[29,31,43]. The proofreading activity of RdRp is
231 unfavorable to this kind of RNA chain terminators. This may also explain why
232 molnupiravir, a mutagenizing nucleoside analog, is much more effective against
233 SARS-CoV-2 than remdesivir since molnupiravir can evade the proofreading
234 mechanism by both RdRp and nsp14-ExoN[36,37].
235 It has long been general knowledge that to become active against virus nucleoside
236 analogue drugs must undergo stepwise addition of phosphate groups to become the
237 triphosphate form. For this outcome, this type of antivirals must be recognized by
238 three kinds of kinases[45]. Which raises a concern that the triphosphate form of
239 antivirals may be incorporated into host mRNA[38]. Even worse, mutagenic
240 ribonucleoside analogs could be reduced into the 2'-deoxyribonucleotide form by host
241 ribonucleotide reductase and then incorporated into DNA[36,38]. Our finding that RdRp

242 also uses MDP as a substrate could largely resolve these concerns. There'd be no
243 issue about whether the prodrugs can become triphosphate form inside host cells since
244 the diphosphate form is also active. Furthermore, nucleoside diphosphate analogues
245 which can evade nucleoside diphosphate kinase would cease to have possibility of
246 being incorporated into mRNA or DNA. The best strategy is to use the
247 membrane-permeable nucleotide diphosphate analog for COVID-19 treatment[45,46].
248 The advantage of this strategy is that the drug is not only able to bypass all the
249 phosphorylation steps but also have no risk of incorporation into host mRNA or DNA
250 (Fig.6).

251 **Materials and Methods**

252 **Protein expression and purification**

253 Preparation of SARS-CoV-2 RdRp, composed of nsp12, nsp7 and two copied of the
254 nsp8 subunits, was carried out as described in ref.[23]. The SARS-CoV-2 *nsp12*, *nsp10*
255 and *nsp7* gene were cloned into a modified pET-21b vector with the C-terminus
256 possessing a 6×His-tag. The *nsp8* gene was cloned into the modified pET-32a vector
257 with the N-terminus possessing a trx-His_6-tag and PreScission Protease site. The
258 plasmids were transformed into *E. coli* BL21 (DE3) and the transformed cells were
259 cultured at 37 °C in LB with a final concentration of 100 μg/ml ampicillin. The
260 *nsp7*-pET-21b plasmid was transformed into *E. coli* Rosetta-gami2 (DE3) and the
261 transformed cells were cultured at 37 °C in LB with a final concentration of 100
262 μg/ml ampicillin and 25 μg/ml chloramphenicol. The *nsp14* gene (nsp14-ExoN
263 (residues 1–289)) was cloned into the modified pET-28b vector with the N-terminus

264 possessing a MBP-tag and PreScission Protease site. The plasmid was then
265 transformed into *E. coli* BL21 (DE3) and the transformed cells were cultured at 37 °C
266 in LB with a final concentration of 100 μg/ml kanamycin.
267 Bacterial cultures were incubated by shaking at 200 rpm, 37 °C to an OD600 of 0.8,
268 then the temperature was lowered to 16 °C and a final concentration of 0.2 mM of
269 isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein
270 expression for 20 h. Subsequently, the cells were harvested by centrifugation at 5000
271 ×*g* for 15 min at 4 °C. The pellet was resuspended in lysis buffer containing 25 mM
272 TrisHCl pH 8.0, 250 mM NaCl, 4 mM $MgCl_2$, 10% Glycerol and homogenized with a
273 high-pressure cell disrupter at 4 °C. The lysate was centrifuged at 25,200 ×*g* for 50
274 min at 4 °C, and the supernatant was then loaded onto Ni-NTA affinity
275 chromatography column for purification. Nsp12, nsp10, nsp7 were eluted with elution
276 buffer (25 mM TrisHCl pH 8.0, 200 mM NaCl, 4 mM $MgCl_2$, 250 mM Imidazole).
277 Nsp8 was subject to on-column tag cleavage by PPase (PreScission Protease) and then
278 eluted by the lysis buffer. The supernatant of nsp14-ExoN was loaded onto a
279 Dextrin-Sepharose gravity column (GE Healthcare), washed and then eluted with
280 elution buffer containing 25 mM TrisHCl pH 8.0, 100 mM NaCl and 20 mM maltose.
281 The MBP tag was then removed by PPase digestion. Next, all proteins were purified
282 by ion exchange chromatography (Source 15Q HR 16/10, GE Healthcare) and size
283 exclusion chromatography (Superdex 200 10/300GL, GE Healthcare) with 25 mM
284 TrisHCl pH 8.0, 200 mM NaCl, 4 mM $MgCl_2$.

285 For nsp12-nsp7-nsp8 complex assembly, protein samples were mixed with the molar
286 ratio of nsp7:nsp8:nsp12=2:2:1 at 4 °C overnight. The incubated RdRp complex was
287 then concentrated with a 100 kDa molecular weight cut-off centrifugal filter unit
288 (Millipore Corporation) and then further purified by size exclusion chromatography
289 using a SuperdexS200 10/300 GL column in 25 mM TrisHCl pH 8.0, 200 mM NaCl,
290 4 mM $MgCl_2$.

291 **Preparation of fluorescently labeled RNA for polymerase activity and**
292 **proofreading assays**

293 RNA template-product duplex was designed according to the published SARS-CoV-2
294 RNA extension assays[23]. FAM labeled 13 nt oligonucleotide with the sequence of
295 FAM - GCUAUGUGAGAUU and the 23 nt complementary oligonucleotide with the
296 sequence of CGAUACACUCUAAUUUCAAUUGA were synthesized for
297 polymerase activity assay. Two 23 nt oligonucleotides with U: G mutation (M1 RNA,
298 CGAUACACUCUACUUUCAAUUGA) or UU: CG mutation (M2 RNA,
299 CGAUACACUCUGCUUUCAAUUGA) were synthesized for proofreading assays.
300 All unmodified and 5' FAM-labeled RNA oligonucleotides were purchased from
301 Tsingke Biotechnology Co.,Ltd.. The oligonucleotides were mixed in equal molar
302 ratio in DEPC water, annealed by heating it to 95 °C for 10 min and gradually cooling
303 to room temperature to make the RNA duplex (Fig.2a, 2b, 2c).

304 **RdRp polymerase activity assays**

305 DEPC-treated water was used for all buffer preparation. RdRp at final concentration
306 of 2 μM was incubated with 200 nM 13+23 nt RNA duplex and 25 μM NTP (or

307 NDP/NMP) in a 20 μl reaction buffer containing 20 mM TrisHCl pH8.0, 10 mM KCl,
308 10 mM $MgCl_2$, 0.01% Triton-X100, 1 mM DTT for 40 min at 37 °C[23]. Eventually,
309 glycerol was added to a concentration of 6.5%. 10 μl RNA product for each reaction
310 was resolved on 20% denaturing polyacrylamide-urea gels and imaged with a
311 Tanon-5200 Multi Fluorescence Imager.

312 **RdRp exoribonuclease activity assay**

313 RdRp at final concentration of 2 μM was incubated with 200 nM 13+23nt RNA
314 duplex in 20 μl reaction buffer for 40 min at 37 °C. Reaction systems containing 50
315 mM EDTA with or without RdRp were prepared as the control group. Eventually,
316 glycerol was added to a concentration of 6.5%. 10 μl RNA product for each reaction
317 was resolved on 20% denaturing polyacrylamide-urea gels and imaged with a Clinx
318 Science Instruments Co., Ltd. (CSI).

319 **Nsp14-ExoN exoribonuclease activity assay**

320 Nsp10 and nsp14 were incubated at a molar ratio of 3:1 for nsp14-nsp10 complex
321 assembly[31]. Reaction buffer containing 300 nM of nsp14-nsp10 complex, 20 mM
322 TrisHCl pH 8.0, 10 mM KCl, 10 mM $MgCl_2$, 0.01% Triton-X100, 1 mM DTT and
323 250 nM M1 RNA was incubated at 37 °C for 1 h. Reaction system without
324 nsp14-nsp10 complex was used as a negative control. After incubation, glycerol was
325 added to a concentration of 6.5%. The product was resolved on 20% denaturing
326 polyacrylamide-urea gels, and imaged with a Tanon-5200 Multi Fluorescence Imager.

327 **The proofreading activities assays for RdRp and nsp14**

328 Double stranded RNAs were prepared by using the extension reaction catalyzed by 2
329 μM RdRp containing 250 nM M1 RNA / M2 RNA, 25 μM NTP, with or without 300
330 nM nsp14-nsp10 in a 20 μl reaction system at 37 °C for 1 h. 10 μl product for each
331 reaction was resolved on 20% denaturing polyacrylamide-urea gels to confirm the
332 dsRNAs production. The remaining dsRNA products were mixed with primer1
333 (acgatgcaaagtctcgacaaatggtcgataccaattcaCCGCTCGAGCGGAGTTAACTTT) and
334 dNTPs at 65 °C for 10 min and then cooled down to 24 °C for primer annealing. 10 ×
335 MMulV buffer, RNase inhibitor, and MMulV reverse transcriptase were added to the
336 above system at 24 °C for 2 h to obtain cDNA, and then the enzymes were inactivated
337 at 70 °C for 15 min. Subsequently, primer2
338 (cagataaactataattcctaatcgcgaggtggcactgcaaCGCGGATCCGCGGCTATGTGAGA(T))
339 was added, and the high-fidelity DNA polymerase pfu-Phusion was used to perform
340 PCR amplification. Individual PCR products were then purified by agarose gel
341 electrophoresis for high-throughput sequencing (Sangon Biotech).

342 **RNA extension assays with NDP as substrate**.

343 RdRp at final concentration of 2 μM was incubated with 200 nM 13+23 nt RNA
344 duplex and 25 μM NTP (or NDP/NMP) in a 20 μl reaction buffer containing 20 mM
345 TrisHCl pH8.0, 10 mM KCl, 10 mM $MgCl_2$, 0.01% Triton-X100, 1 mM DTT for 40
346 min at 37 °C.

347 RNA extension was performed by incubating 2 μM RdRp, 250 nM 13+23 nt RNA
348 duplex, 25 μM NDP in a 20μl reaction system at 37 °C for 1 h. 10 μl RNA product
349 was resolved on 20% denaturing polyacrylamide-urea gels and imaged with a

350 Tanon-5200 Multi Fluorescence Imager to confirm the extension. The remaining RNA
351 product, primer1 (CCGCTCGAGCGGAGTTAACTTT) and dNTPs were gently
352 mixed at 65 °C for 10 min. The mixture was cooled down to 24 °C then 10× MMulV
353 buffer and MMulV reverse transcriptase were added and incubated for 2 h to obtain
354 cDNA. Thereafter, the reverse transcriptase was inactivated at 70 °C for 15 min.
355 Subsequently, primer2 (CGCGGATCCGCG GCTATGTGAGAT) was added, and the
356 product was amplified by PCR with catalyzed by the high-fidelity DNA polymerase
357 pfu-Phusion. The PCR product was digested with Xho1 and BamH1 and ligated into
358 pET-15b vector, and transformed into *E-coli* DH5α to obtain clones. Plasmids from
359 positive clones were then sequenced with T7 universal primers.

360 **RNA extension assays with MDP as substrate**.

361 A RNA extension assay was performed as mentioned previously[37]. In this assay the
362 13+23 nt RNA duplex was used (Fig.5e). The 5'end of the RNA product strand
363 contains a 6-carboxyfluorescein (FAM) fluorescent label, which allows us to monitor
364 RNA elongation. RdRp can incorporate NHC monophosphate into RNA at the
365 position where the nucleotide in the template is A. MDP and MTP were purchased
366 from MedChemExpress.

367 For RNA extension RdRp at final concentration of 2 μM was incubated with 200 nM
368 13+23 nt RNA duplex and 50 μM NTP/NDP with or without MTP/MDP in 20 μl
369 reaction buffer for 40 min at 37 °C. 10 μl RNA product for each reaction was resolved
370 on 20% denaturing polyacrylamide-urea gels and imaged with a Tanon-5200 Multi
371 Fluorescence Imager to confirm the extension.

372 **Data availability**

373 All data generated or analyzed during this study are included in this published article

374 (and its Supplementary Information files).

375 **Acknowledgements**

376 The authors thank Yan Zhang at Zhejiang University for offering SARS-CoV-2 RdRp

377 gene, Carina Muyao Gu at CBe-Learn School for helping with the English expression

378 of the manuscript. This work was supported by Shandong Provincial Key Research

379 and Development Program (2020CXGC011305).

380 **Author contributions**

381 M.W. and L.G. conceived, designed the experiments and analyzed the data. M.W.

382 performed functional assays of RdRp. M.W., C.W., and X.J. participated in expression,

383 purification of the SARS-CoV-2 RdRp complex and other proteins. N.L., F.Z., H.D.,

384 S.W. and M.C. helped with the experiments and participated in figure and manuscript

385 preparation. M.W. and L.G. wrote the manuscript.

386 **Competing interests**

387 The authors declare no competing interests.

388 **References**

389 1 Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure,

390 replication, and pathogenesis. *J Med Virol* **92**, 418-423,

391 doi:10.1002/jmv.25681 (2020).

392 2 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China,

393 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).

394 3 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel
395 coronavirus: implications for virus origins and receptor binding. *The Lancet*
396 **395**, 565-574, doi:10.1016/s0140-6736(20)30251-8 (2020).
397 4 Chan, J. F. *et al.* Genomic characterization of the 2019 novel
398 human-pathogenic coronavirus isolated from a patient with atypical
399 pneumonia after visiting Wuhan. *Emerg Microbes Infect* **9**, 221-236,
400 doi:10.1080/22221751.2020.1719902 (2020).
401 5 Hoffmann, M. *et al.* SARS-CoV-2 variant B.1.617 is resistant to
402 bamlanivimab and evades antibodies induced by infection and vaccination.
403 *Cell Rep* **36**, 109415, doi:10.1016/j.celrep.2021.109415 (2021).
404 6 Liu, C. *et al.* Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and
405 convalescent serum. *Cell* **184**, 4220-4236 e4213,
406 doi:10.1016/j.cell.2021.06.020 (2021).
407 7 Mlcochova, P. *et al.*, doi:10.1101/2021.05.08.443253 (2021).
408 8 Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody
409 neutralization. *Nature* **596**, 276-280, doi:10.1038/s41586-021-03777-9 (2021).
410 9 Rudd, C. E. GSK-3 Inhibition as a Therapeutic Approach Against SARs CoV2:
411 Dual Benefit of Inhibiting Viral Replication While Potentiating the Immune
412 Response. *Front Immunol* **11**, 1638, doi:10.3389/fimmu.2020.01638 (2020).
413 10 Liu, C. *et al.* Research and Development on Therapeutic Agents and Vaccines
414 for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent Sci* **6**,
415 315-331, doi:10.1021/acscentsci.0c00272 (2020).

416 11 Quimque, M. T. J. *et al.* Virtual screening-driven drug discovery of
417 SARS-CoV2 enzyme inhibitors targeting viral attachment, replication,
418 post-translational modification and host immunity evasion infection
419 mechanisms. *J Biomol Struct Dyn* **39**, 4316-4333,
420 doi:10.1080/07391102.2020.1776639 (2021).
421 12 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in
422 China. *Nature* **579**, 265-269, doi:10.1038/s41586-020-2008-3 (2020).
423 13 Wu, C. *et al.* Analysis of therapeutic targets for SARS-CoV-2 and discovery of
424 potential drugs by computational methods. *Acta Pharm Sin B* **10**, 766-788,
425 doi:10.1016/j.apsb.2020.02.008 (2020).
426 14 Subissi, L. *et al.* One severe acute respiratory syndrome coronavirus protein
427 complex integrates processive RNA polymerase and exonuclease activities.
428 *Proc Natl Acad Sci U S A* **111**, E3900-3909, doi:10.1073/pnas.1323705111
429 (2014).
430 15 McKee, D. L., Sternberg, A., Stange, U., Laufer, S. & Naujokat, C. Candidate
431 drugs against SARS-CoV-2 and COVID-19. *Pharmacol Res* **157**, 104859,
432 doi:10.1016/j.phrs.2020.104859 (2020).
433 16 Chu, C. K. *et al.* Antiviral Activity of Nucleoside Analogues against
434 SARS-coronavirus (SARS-CoV). **17**, 285-289,
435 doi:10.1177/095632020601700506 (2006).
436 17 Zhai, Y. *et al.* Insights into SARS-CoV transcription and replication from the
437 structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol* **12**, 980-986,

438 doi:10.1038/nsmb999 (2005).

439 18 Johnson, M. A., Jaudzems, K. & Wuthrich, K. NMR Structure of the
440 SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5. *J Mol Biol* **402**,
441 619-628, doi:10.1016/j.jmb.2010.07.043 (2010).

442 19 Peti, W. *et al.* Structural genomics of the severe acute respiratory syndrome
443 coronavirus: nuclear magnetic resonance structure of the protein nsP7. *J Virol*
444 **79**, 12905-12913, doi:10.1128/JVI.79.20.12905-12913.2005 (2005).

445 20 Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from
446 COVID-19 virus. *Science* **368**, 779-782, doi:10.1126/science.abb7498 (2020).

447 21 Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12
448 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, 2342,
449 doi:10.1038/s41467-019-10280-3 (2019).

450 22 Lehmann, K. C. *et al.* Discovery of an essential nucleotidylating activity
451 associated with a newly delineated conserved domain in the RNA
452 polymerase-containing protein of all nidoviruses. *Nucleic Acids Res* **43**,
453 8416-8434, doi:10.1093/nar/gkv838 (2015).

454 23 Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA
455 polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499-1504,
456 doi:10.1126/science.abc1560 (2020).

457 24 Vázquez, A. L., Alonso, J. M. & Parra, F. Mutation analysis of the GDD
458 sequence motif of a calicivirus RNA-dependent RNA polymerase. *Journal of*
459 *virology* **74**, 3888-3891, doi:10.1128/jvi.74.8.3888-3891.2000 (2000).

460 25 Wang, Y. *et al.* Mutational analysis of the GDD sequence motif of classical
461 swine fever virus RNA-dependent RNA polymerases. *Virus Genes* **34**, 63-65,
462 doi:10.1007/s11262-006-0001-z (2007).
463 26 Jiang, Y., Yin, W. & Xu, H. E. RNA-dependent RNA polymerase: Structure,
464 mechanism, and drug discovery for COVID-19. *Biochem Biophys Res*
465 *Commun* **538**, 47-53, doi:10.1016/j.bbrc.2020.08.116 (2021).
466 27 Garriga, D., Ferrer-Orta, C., Querol-Audi, J., Oliva, B. & Verdaguer, N. Role
467 of motif B loop in allosteric regulation of RNA-dependent RNA
468 polymerization activity. *J Mol Biol* **425**, 2279-2287,
469 doi:10.1016/j.jmb.2013.03.034 (2013).
470 28 Gong, P. & Peersen, O. B. Structural basis for active site closure by the
471 poliovirus RNA-dependent RNA polymerase. *Proc Natl Acad Sci U S A* **107**,
472 22505-22510, doi:10.1073/pnas.1007626107 (2010).
473 29 Becares, M. *et al.* Mutagenesis of Coronavirus nsp14 Reveals Its Potential
474 Role in Modulation of the Innate Immune Response. *J Virol* **90**, 5399-5414,
475 doi:10.1128/JVI.03259-15 (2016).
476 30 Ma, Y. *et al.* Structural basis and functional analysis of the SARS coronavirus
477 nsp14-nsp10 complex. *Proc Natl Acad Sci U S A* **112**, 9436-9441,
478 doi:10.1073/pnas.1508686112 (2015).
479 31 Bouvet, M. *et al.* RNA 3'-end mismatch excision by the severe acute
480 respiratory syndrome coronavirus nonstructural protein nsp10/nsp14
481 exoribonuclease complex. *Proc Natl Acad Sci U S A* **109**, 9372-9377,

482 doi:10.1073/pnas.1201130109 (2012).

483 32 Choy, K. T. *et al.* Remdesivir, lopinavir, emetine, and homoharringtonine
484 inhibit SARS-CoV-2 replication in vitro. *Antiviral Res* **178**, 104786,
485 doi:10.1016/j.antiviral.2020.104786 (2020).

486 33 Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently
487 emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res* **30**, 269-271,
488 doi:10.1038/s41422-020-0282-0 (2020).

489 34 Williamson, B. N. *et al.* Clinical benefit of remdesivir in rhesus macaques
490 infected with SARS-CoV-2. *Nature* **585**, 273-276,
491 doi:10.1038/s41586-020-2423-5 (2020).

492 35 Xie, X. *et al.* A nanoluciferase SARS-CoV-2 for rapid neutralization testing
493 and screening of anti-infective drugs for COVID-19. *Nat Commun* **11**, 5214,
494 doi:10.1038/s41467-020-19055-7 (2020).

495 36 Zhou, S. *et al.* beta-d-N4-hydroxycytidine Inhibits SARS-CoV-2 Through
496 Lethal Mutagenesis But Is Also Mutagenic To Mammalian Cells. *J Infect Dis*
497 **224**, 415-419, doi:10.1093/infdis/jiab247 (2021).

498 37 Kabinger, F. *et al.* Mechanism of molnupiravir-induced SARS-CoV-2
499 mutagenesis. *Nat Struct Mol Biol* **28**, 740-746,
500 doi:10.1038/s41594-021-00651-0 (2021).

501 38 Malone, B. & Campbell, E. A. Molnupiravir: coding for catastrophe. *Nat*
502 *Struct Mol Biol* **28**, 706-708, doi:10.1038/s41594-021-00657-8 (2021).

503 39 Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L. & Denison, M. R. High fidelity

504 of murine hepatitis virus replication is decreased in nsp14 exoribonuclease
505 mutants. *J Virol* **81**, 12135-12144, doi:10.1128/JVI.01296-07 (2007).
506 40 Malone, B. *et al.* Structural basis for backtracking by the SARS-CoV-2
507 replication-transcription complex. *bioRxiv*, doi:10.1101/2021.03.13.435256
508 (2021).
509 41 Sheahan, T. P. *et al.* An orally bioavailable broad-spectrum antiviral inhibits
510 SARS-CoV-2 in human airway epithelial cell cultures and multiple
511 coronaviruses in mice. *Sci Transl Med* **12**, eabb5883,
512 doi:10.1126/scitranslmed.abb5883 (2020).
513 42 Zandi, K. *et al.* Repurposing Nucleoside Analogs for Human Coronaviruses.
514 *Antimicrob Agents Chemother* **65**, e01652-01620, doi:10.1128/AAC.01652-20
515 (2020).
516 43 Robson, F. *et al.* Coronavirus RNA Proofreading: Molecular Basis and
517 Therapeutic Targeting. *Mol Cell* **79**, 710-727,
518 doi:10.1016/j.molcel.2020.07.027 (2020).
519 44 Liu, C. *et al.* Structural basis of mismatch recognition by a SARS-CoV-2
520 proofreading enzyme. *Science* **373**, 1142-1146, doi:10.1126/science.abi9310
521 (2021).
522 45 Meier, C. Nucleoside diphosphate and triphosphate prodrugs - An unsolvable
523 task? *Antivir Chem Chemother* **25**, 69-82, doi:10.1177/2040206617738656
524 (2017).
525 46 Gollnest, T. *et al.* Membrane-permeable Triphosphate Prodrugs of Nucleoside

Analogues. *Angew Chem Int Ed Engl* **55**, 5255-5258, doi:10.1002/anie.201511808 (2016).

529 **Figure legends**

530 **Fig.1** RdRp both synthesizes and degrades RNA. **a** RdRp complex shows polymerase
531 activity in vitro. The primer strand is labeled with fluorescence at the 5′ end. The
532 positions of the primer and the full-length extension product are indicated. **b** RNA
533 degradation catalyzed by RdRp (RNA substrate is corresponding to 13+23nt RNA
534 duplex without fluorescence) when NTP is absent. EDTA prevents RNA degradation
535 by RdRp.

536 **Fig.2** The RNA duplexes for primer extension and RdRp proofreading. **a** The 13+23
537 nt RNA template-product duplex. The direction of RNA extension is shown. The color
538 of the depicted circles indicates the experimental design: blue, RNA template strand;
539 red, RNA product strand. The 5'end of the RNA product contains a FAM fluorescent
540 label. **b** The primer/template bearing a U: C mismatch (M1 RNA). **c** The
541 primer/template bearing a UU: GC double mismatches (M2 RNA).

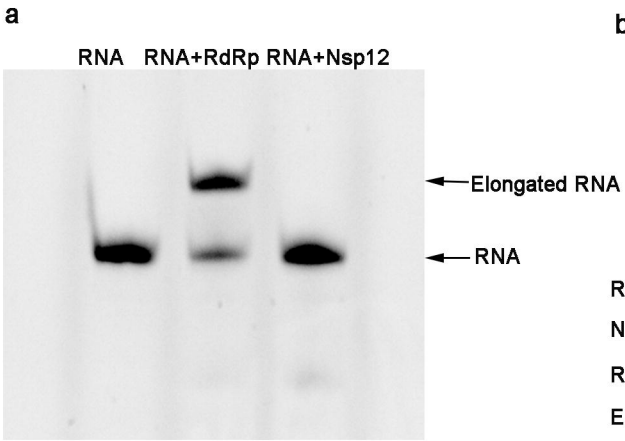
542 **Fig.3** Proofreading activity of RdRp and its combination with nsp14-ExoN. **a** The
543 13+23 nt RNA, M1 RNA, M2 RNA were used as template respectively. The positions
544 of the template RNA and the full-length extension product are indicated. **b**
545 Fluorescence display of M1 RNA cleavage products. **c** The 13+23 nt RNA, M1 RNA,
546 M2 RNA were used as template respectively. Primer strand extension by RdRp alone
547 or with nsp10 + nsp14. The positions of the template RNA and the full-length
548 extension product and degraded product are indicated.

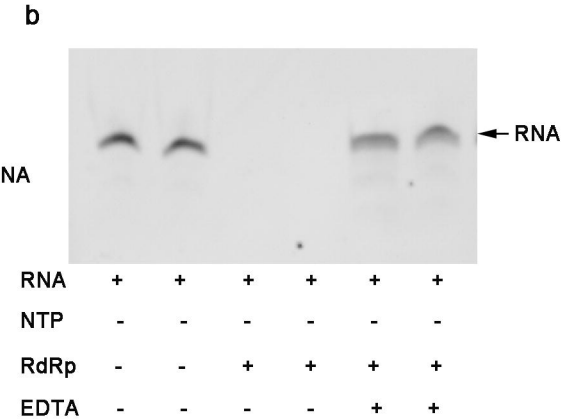
549 **Fig.4** RdRp uses both NTP and NDP as substrate. **a** RdRp was incubated with the
550 13+23 nt primer/template in the presence of NMP, NDP or NTP respectively. The

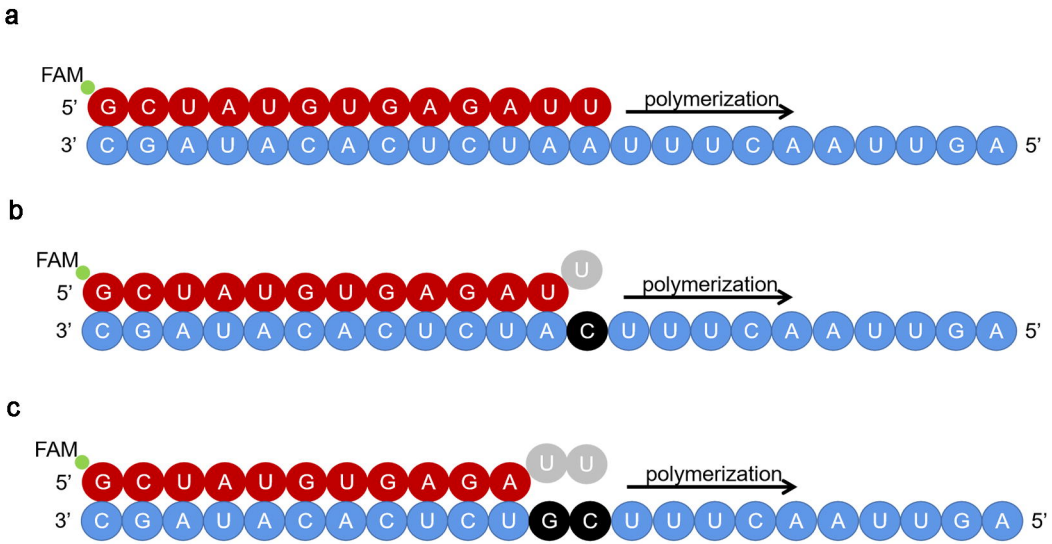
551 positions of the template RNA and the full-length extension product are indicated. **b**
552 Part of the Sanger sequencing chromatogram of the RT-PCR products corresponding
553 to the extended 10 nt RNA on the 13+23 nt RNA duplex with NDP as substrate. The
554 dash line circle represents the 10 nt extension.

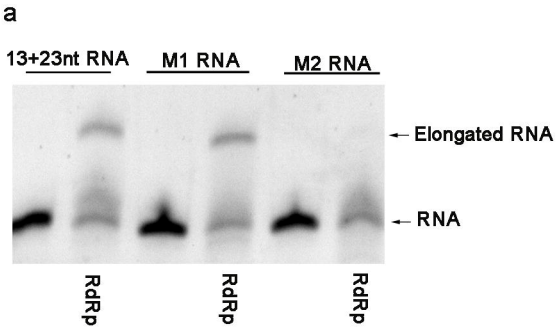
555 **Fig.5** RdRp incorporates NHC into RNA with MDP as substrate. **a** Metabolism of
556 nucleoside analogues molnupiravir in host cell. **b** The 13+23 nt RNA allows for RNA
557 extension by 10 nucleotides. NHC monophosphate can be incorporated into growing
558 RNA instead of U or C. **c** RNA elongation stalls at the expected positions when the
559 cognate NDP is withheld from the reaction. Extension to the end of the template is
560 possible when MDP replaces UDP in the presence of other substrate NDPs, showing
561 that using MDP as a substrate RdRp incorporates M into RNA and incorporation of M
562 does not prevent RNA extension.

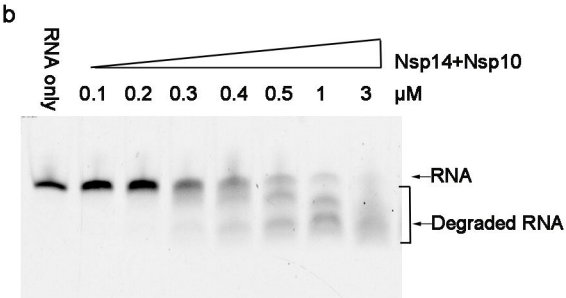
563 **Fig.6** Nucleotide diphosphate analog prodrugs designed by our strategy would have
564 less potential to cause side effects. The terminal phosphate group (β-phosphate) is
565 modified by lipophilic masks to give the prodrug more membrane permeability. Inside
566 cell the masks are removed by enzymatic or chemical reactions. The ribose moiety
567 and the base group should be designed in such a way that the diphosphate form drug
568 can evade NDPK thus eliminating the risk of being incorporated into host mRNA or
569 DNA.

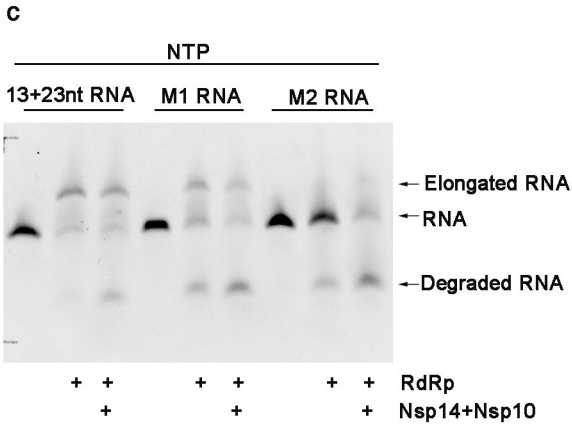

a
RNA
RNA+RdRp
RNA+Nsp12
Elongated RNA
RNA


b
RNA
RNA + + + + + +
NTP - - - - - -
RdRp - - + + + +
EDTA - - - - + +

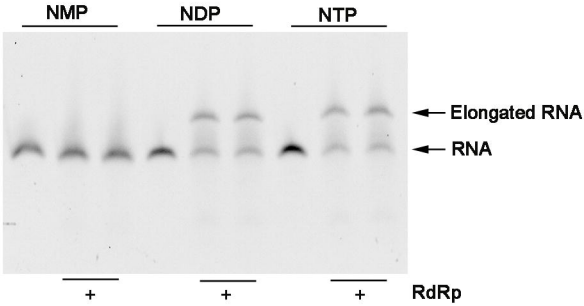

a
FAM
5'
G C U A U G U G A G A U U
polymerization
3'
C G A U A C A C U C U A A U U U C A A U U G A
5'
b
FAM
5'
G C U A U G U G A G A U
U
polymerization
3'
C G A U A C A C U C U A C U U U C A A U U G A
5'
c
FAM
5'
G C U A U G U G A G A
U U
polymerization
3'
C G A U A C A C U C U G C U U U C A A U U G A
5'

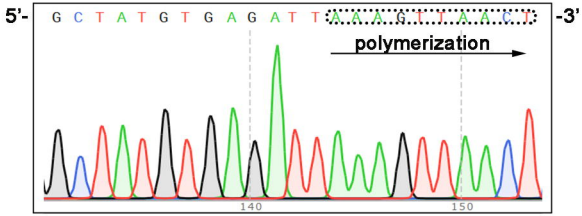

a
13+23nt RNA
M1 RNA
M2 RNA
Elongated RNA
RNA
RdRp
RdRp
RdRp

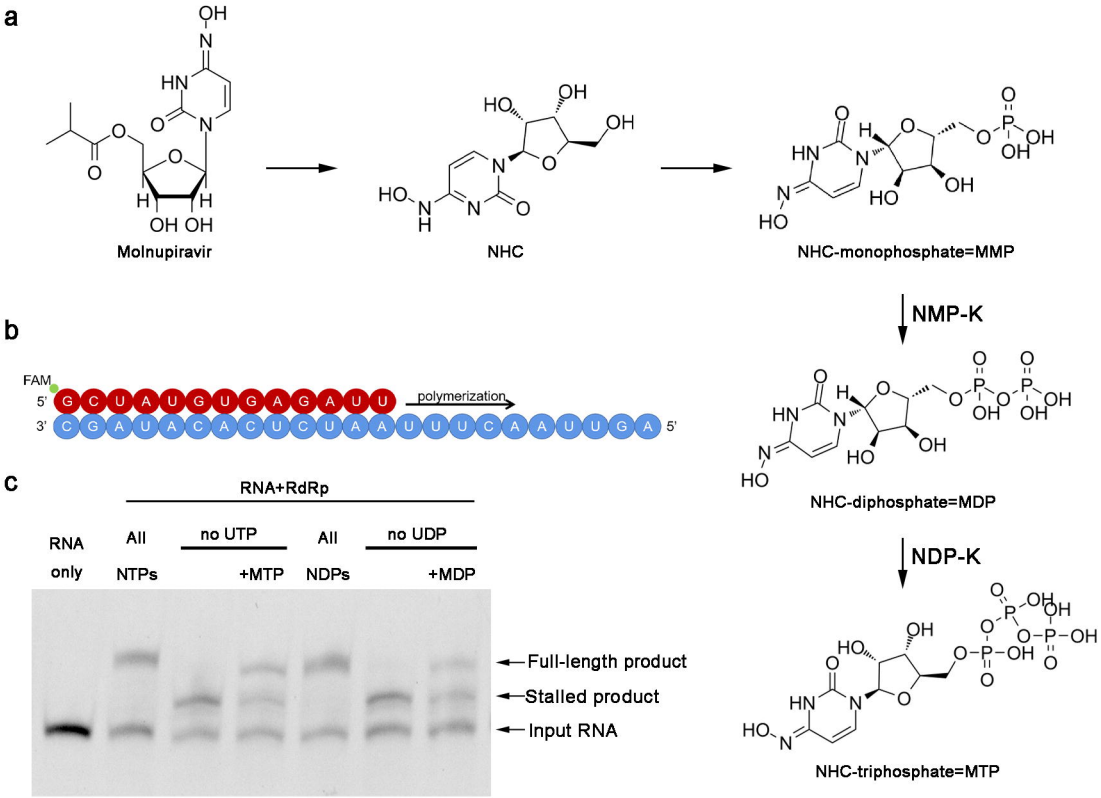

b
RNA only
Nsp14+Nsp10
0.1 0.2 0.3 0.4 0.5 1 3 μM
RNA
Degraded RNA

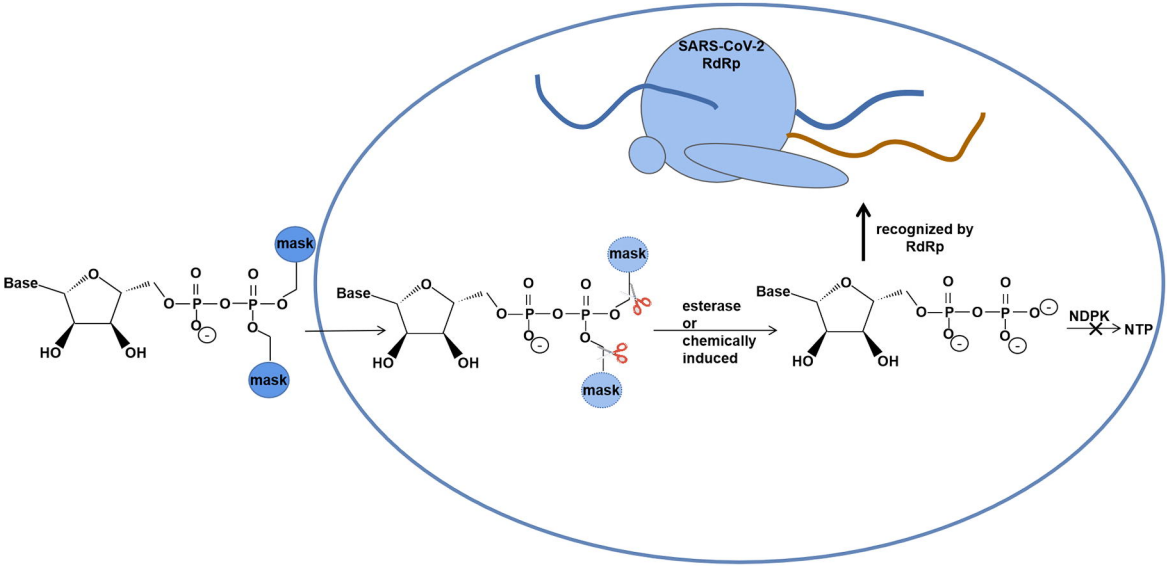

c
NTP
13+23nt RNA
M1 RNA
M2 RNA
Elongated RNA
RNA
Degraded RNA
+ + + + + + RdRp
+ + + Nsp14+Nsp10

a


NMP
NDP
NTP
Elongated RNA
RNA
+
+
+
RdRp


5'-
G C T A T G T G A G A T T A A A G T T A A C T
-3'
polymerization
140
150


a
Molnupiravir
NHC
NHC-monophosphate=MMP
NMP-K
NHC-diphosphate=MDP
NDP-K
NHC-triphosphate=MTP
b
FAM
5'
3'
G C U A U G U G A G A U U
C G A U A C A C U C U A A U U U C A A U U G A
polymerization
5'
c
RNA+RdRp
RNA only
All NTPs
no UTP +MTP
All NDPs
no UDP +MDP
Full-length product
Stalled product
Input RNA


SARS-CoV-2
RdRp
recognized by
RdRp
Base
HO
OH
mask
mask
mask
mask
esterase
or
chemically
induced
NDPK
NTP