SARS-CoV-2 RdRp uses NDPs as a substrate and is able to incorporate NHC into RNA from diphosphate form molnupiravir

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 more than 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. Here, we reveal that RdRp can recognize and utilize nucleoside diphosphates (NDPs) as a substrate to synthesize RNAs with an efficiency of about two thirds of using nucleoside triphosphates (NTPs) as a substrate. NDPs incorporation is also template-specific and has high fidelity. Moreover, RdRp can incorporate β-d-N4-hydroxycytidine (NHC) into RNA while using diphosphate form molnupiravir (MDP) as a substrate. We also observed that MDP is a better substrate for RdRp than the triphosphate form molnupiravir (MTP).

Introduction

The pandemic of coronavirus disease 2019 (COVID-19) has severely impacted human health and the global economy since the outbreak in late 2019[1,2]. The causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is homologous to 2002 SARS-CoV with a genome sequence similarity of 79%[3,4], forming a sister clade to SARS-CoV and considered a newly β-coronavirus[5]. As of the date of writing, more than 491.75 million infections and more than 6.17 million deaths have been confirmed (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports). Although a variety of vaccines have been developed to contain the pandemic, the rapidly spreading mutant strains presses need to develop effective medication or treatment to counter the virus[6-9].

SARS-CoV-2 is a positive-sense single-stranded RNA virus[5] with a large genome of approximately 30 kb organized in 14 open reading frames[4,10-12]. 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]. Processivity is enhanced by the presence of nsp 7/8 and bound to nsp 12 in a 1:2:1 stoichiometry[15,16]. 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[17,18].

To date, the most promising RdRp inhibitors are nucleotide/nucleoside analogs (NA). NA prodrugs are designed to be metabolized to the active 5'-triphosphate form (5'-TP)

67 once inside cells[19], thus they can compete with endogenous nucleotides for
68 incorporation into newly synthesized viral RNAs, resulting in viral genomes lethal
69 mutation or inactivation[20]. Remdesivir is a broad antiviral drug approved for the
70 treatment of SARS-CoV-2, which can be incorporated into RNA to sterically hinder
71 RdRp, thereby blocking viral RNA synthesis[21]. Other NAs (favipiravir, ribavirin) are
72 integrated into nascent viral RNA at high rates but are not recognized as endogenous
73 nucleotides in subsequent rounds of replication, increasing the mutation rate and
74 resulting in an inviable genome, a process known as "lethal mutagenesis"[22-24]. The
75 recently approved NA antiviral drug molnupiravir significantly increases the
76 frequency of viral RNA mutations and impairs SARS-CoV-2 in both animal models
77 and humans[25,26]
78 By far none of these RdRp targeting drugs is effective enough to fulfill the need for
79 treatment of COVID-19. To develop a better RdRp targeting drug we need to better
80 understanding of both the properties of RdRp and how these inhibitors function.
81 Previous studies showed that some DNA polymerases[27,28] and HIV reverse
82 transcriptase[29] can use deoxynucleoside diphosphates (dNDPs) as a substrate to
83 synthesize DNA, and *Escherichia coli* RNA polymerase[30] use nucleoside
84 diphosphates (NDPs) to synthesize RNA. Although both dNDPs and NDPs showed
85 very weak activity in these experiments, the data prompted us to test if NDPs are an
86 active substrate for RdRp. In this study, we found that SARS-CoV-2 RdRp can
87 recognize and utilize NDPs as a substrate to synthesize RNAs with an efficiency of
88 about two thirds of using nucleoside triphosphates (NTPs) as a substrate. In addition,

89 we proved that molnupiravir is not only fully functionable in its diphosphate form but
90 also more active than in triphosphate form. This could imply a new strategy for
91 designing COVID-19 treatment nucleoside analogue drugs with more efficacy and
92 fewer side effects.

93 **Results**

94 **RdRp synthesizes RNA by using both NTPs and NDPs as a substrate**

95 Since RdRp is by far one of the most attractive targets for anti-coronavirus drug
96 development, a thorough characterization of RdRp polymerase activity will, without
97 doubt, greatly contribute to discovering the effective strategy for drug screening[17,18].
98 The RdRp of SARS-CoV-2 was purified by assembling nsp12, nsp7 and nsp8 (Fig.1a).
99 Based on previous research, SARS-CoV-2 RdRp polymerase activity was measured
100 by conducting a conventional RNA elongation assay[31]. When NTPs was added,
101 elongation catalyzed by RdRp complex occurred on the RNA duplex, resulting in an
102 intact double-stranded RNA product (Fig.1c).

103 To test if RdRp uses NDPs or NMPs as a substrate, we performed an RNA extension
104 assay by using NTPs, NDPs or NMPs as substrate, respectively. The result revealed
105 that RdRp synthesized RNA products of the same length regardless of NTPs or NDPs
106 as substrate (Fig.1d). To examine the template specificity of this RNA synthesis, we
107 challenged 13+23 nt RNA with various NDPs (Fig.1b). Only the cognate ADP
108 promoted RNA extension, indicating that NDPs incorporation is template-specific
109 (Fig.1e).

110 This result also raised the question of whether the RNA made from NDPs has the
111 right sequence. Hence, two experiments were performed: single-clonal sequencing of
112 ligated plasmids, and high-throughput sequencing of PCR products. First, the cDNA
113 was synthesized through reverse transcription of the RNA products, and then
114 amplified by 10 cycles of PCR with primers containing the restriction sites of XhoI
115 and BamHI. PCR products were then ligated into the vector pET-15b and used to
116 transform *E. coli* DH5α strain. Plasmids extracted from the positive clones were
117 sequenced by Sanger sequencing. The sequencing results indicated that the RNA
118 synthesized from NDPs has the right sequence (Fig.1f). Second, the cDNA amplified
119 by 10 cycles of PCR was sequenced by high-throughput sequencing (Fig.1g). At least
120 50,000 sequences were obtained for each product (SI). The ADP, GDP, CDP and UDP
121 (purchased from Sigma) used in these experiments were characterized by mass
122 spectrometry respectively to make sure these NDPs were not contaminated by NTPs
123 (S1).

124 **NDPs are nearly two thirds as efficient as NTPs as a substrate of RdRp**

125 Previous studies indicated that although some DNA polymerases and RNA
126 polymerases are able to use diphosphate form substrate. The activity of the
127 diphosphate form substrate, however, is much lower than that of the triphosphate form
128 substrate[28-30]. To determine if this is also the case for RdRp we performed a kinetics
129 study by using NTPs and NDPs as a substrate respectively. The 13+23 nt RNA duplex
130 was used as a template and the elongation of product over time was observed (Fig.2a).
131 Kinetics curves indicated that the relative incorporation efficiency of NDPs is about

132 two thirds (67.94% ± 2.66%) of that of NTPs (Fig.2b). These data suggested that the
133 incorporation rate for NDPs is slower than that of NTPs. Subsequently, we decided to
134 test if all four NDPs equally contribute to the slower incorporation rate. A 13+26 nt
135 RNA duplex was synthesized for the assay testing the incorporation efficiency for
136 each NDP (Fig.2c). Each reaction system contains three NTPs and the NDP of interest
137 (Fig.2d). Unexpectedly, our data showed that substitution of any single NTP by NDP
138 has no observable effect on the efficiency of RNA synthesis (Fig.2e). This may imply
139 that the speed of RNA synthesis is only affected by continuous NDP incorporation but
140 not the intermittent insertion.

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

143 Since RdRp has become one of the most important targets for the development of
144 antiviral drugs, many nucleoside analogue drugs have been constructed to target
145 RdRp for COVID-19 treatment[24,32-34]. Among all these chemicals reported by far,
146 molnupiravir has been regarded as a promising drug candidate. The molecular
147 mechanism of molnupiravir was determined recently. Once ingested by patients,
148 molnupiravir undergoes stepwise phosphorylation to yield the active nucleoside
149 triphosphate analogue (β-d-N4-hydroxycytidine (NHC) triphosphate, or MTP). The
150 RdRp then uses MTP as substrate and incorporates NHC monophosphate into RNA at
151 the position where the nucleotide in the template is G or A, resulting in "error
152 catastrophe" and virus death[26,35,36] (Fig.3a).

153 Although the mechanism of molnupiravir has been well established, our finding of-
154 RdRp's ability to utilize NDPs as a substrate still raises a question: is NHC
155 diphosphate (MDP) also an active form and recognized by RdRp as a substrate? To
156 address this question, the 13+26 nt RNA duplex was used to perform the RNA
157 extension assay with NDPs and MDP as the substrates in comparison with the
158 extension using NTPs and MTP as the substrates. The 13+26 nt RNA duplex allows
159 ten nucleotides (nt) of extension. The 13+26 nt RNA duplex contanis two adjacent G
160 and two adjacent A, allowing NHC incorporation into the RNA (Fig.3b). In this case,
161 when cytidine triphosphate (CTP) or uridine triphosphate (UTP) is replaced by MDP,
162 the incorporation of M does not hinder the incorporation of the next subsequent
163 nucleotide, and RdRp still completes the extension reaction. Similar results were
164 obtained using MTP as the substrate (Fig.3c). Therefore, it is reasonable to speculate
165 that both MTP and MDP are active forms of monulpiravir in humans.

166 **MDP is a better substrate than MTP**

167 Since both MTP and MDP are active substrates of RdRp we performed an assay to
168 figure out which one is the better substrate. By observing their efficiency in reaction,
169 we found that when the template is G, the relative incorporation efficiency of MTP is
170 0.78, while MDP is 0.87 in comparison to CTP. When the template is A, the relative
171 incorporation efficiency of MTP is 0.58, while MDP is 0.77 in comparison to UTP
172 (Fig.3d). In order to obtain a reliable result, the experiment was repeated six times
173 with two batches of RdRp. The results indicated MDP is the better substrate for RdRp

174 in comparison to MTP. MTP and MDP were characterized by HPLC and mass
175 spectrometry to make sure they are the right pure compounds (S2).
176 Next, we tried to determine if MDP is also more efficient in introducing mutations
177 than MTP. The products of RdRp-catalyzed extension reactions containing MTP or
178 MDP using 13+29 nt RNA as a template (Fig. 3e) were reverse-transcribed and
179 PCR-amplified, and the number of mutations was counted by high-throughput
180 sequencing. The statistical results were shown in Table 1. Similar mutation ratios
181 were found when the same concentrations of MTP and MDP were added. Higher ratio
182 of MTP or MDP to CTP associates with higher mutation rate, with MDP resulting in
183 slightly more mutations.

184 Table.1

substrate	NTP	50 μM CTP 50 μM MTP	50 μM CTP 50 μM MDP	5 μM CTP 50 μM MTP	5 μM CTP 50 μM MDP	0.8 μM CTP 50 μM MTP	0.8 μM CTP 50 μM MDP
all sequences	67883	64316	61673	59588	70906	59129	56214
no mutation （TCCCTTT）	67105	63626	61054	58386	69443	57831	54945
% of no mutation	98.85%	98.93%	99.00%	97.98%	97.94%	97.80%	97.74%
mutations*	778	690	619	1202	1463	1298	1269
% of mutation	1.15%	1.07%	1.00%	2.02%	2.06%	2.20%	2.26%

185 *The mutation resulting from the incorporation of M is shown as: TCCCTTC,
186 TTCCTTT, TACCTTT, TCCCCTT, TCACTTT, TCCCTCT, TCCCTAT etc.

187 **NDPs are a good substrate for SARS-CoV-2 RdRp but not other virus RNA**
188 **polymerases**

189 To test if other virus RNA polymerases also use NDPs as a substrate, we purified
190 SARS-CoV RdRp, NS5 protein of ZIKA virus and T7 RNA polymerase (RNAP) and
191 measured their activities when NDPs were used as s substrate. Results show that

192 amount these RNA polymerases tested SARS-CoV-2 RdRp had the highest activity
193 when using NDPs as a substrate. SARS-CoV RdRp could also use NDPs as a
194 substrate, but the product band seam a double RNA strand shorter than expectation
195 implying stalled extension. NS5 of ZIKA virus did not produce a well-defined product
196 band when using NDPs as a substrate (Fig.4a). The transcription of T7 RNAP was
197 measured based on fluorescence technology. We found that except for the case in
198 which ATP was replaced by ADP and weak transcription occurred, any other NTP
199 replaced by NDP would inhibit the transcription completely. (Fig.4b).

200 **Discussion**

201 Since its outbreak in late 2019, the COVID-19 pandemic has been plaguing people for
202 two years with no sign of abating[3,11]. Although many kinds of vaccines have been
203 developed to suppress the pandemic, they are not sufficient to withhold the spread of
204 SARS-CoV-2[37]. The development of potent antivirals against SARS-CoV-2 is still an
205 urgent need[6,38].

206 Since RNA-dependent RNA polymerase (RdRp) is critical for viral genome
207 replication and transcription and conserved in RNA virus species, RdRp has become
208 one of the most appealing targets for antivirals development[17,31,38,39]. Thoroughly
209 characterization of RdRp both structurally and biochemically would no doubt be
210 beneficial to drug development.

211 The substrate usage of RdRp is astonishing. To the best of our knowledge, most RNA
212 polymerases, including DNA-dependent and RNA-dependent RNA polymerases, use
213 nucleoside triphosphates (NTPs) as a substrate to synthesize RNA under the guidance

214 of a template strand. We found that SARS-CoV-2 RdRp synthesizes RNA with
215 comparable efficiency using NTPs and NDPs as a substrate, and other RNA viruses
216 that can utilize NDPs as a substrate have also been found. We speculate that the
217 ability to use NDPs as a substrate may be an advantage of RNA virus during infection.
218 During later stages of infection, host cells may not be able to produce enough ATP
219 and other NTPs. By using NDPs as a substrate, SARS-CoV-2 can continue the process
220 of genome replication and assembly of progeny viruses, while most metabolic
221 activities stop.
222 It has long been a general knowledge that to become active against virus nucleoside
223 analogue drugs must undergo stepwise addition of phosphate groups to become the
224 triphosphate form. In order to produce this outcome, this type of antivirals must be
225 recognized by three kinds of kinases[40]. This raises a concern that the triphosphate
226 form of antivirals may be incorporated into host mRNA[36]. Even worse, mutagenic
227 ribonucleoside analogs could be reduced into the 2'-deoxyribonucleotide form by host
228 ribonucleotide reductase and then incorporated into DNA[35,36]. Our finding that RdRp
229 also uses MDP as a substrate could largely resolve these concerns. Since the
230 diphosphate form is also active, there would be no issue regarding whether the
231 prodrugs can become triphosphate form inside host cells. Furthermore, nucleoside
232 diphosphate analogues, which can evade nucleoside diphosphate kinase, would cease
233 to have the possibility of being incorporated into mRNA or DNA. The best strategy is
234 to use the membrane-permeable nucleotide diphosphate analog for COVID-19
235 treatment[40,41]. The advantage of this strategy is that the drug is not only able to bypass

236 all the phosphorylation steps and has no risk of incorporation into host mRNA or
237 DNA (Fig.5).

238 **Materials and Methods**

239 **Protein expression and purification**

240 Preparation of SARS-CoV-2 RdRp, composed of nsp12, nsp7 and two copied of the
241 nsp8 subunits, was carried out as described in ref.[31]. The SARS-CoV-2 *nsp12* and
242 *nsp7* gene were cloned into a modified pET-21b vector with the C-terminus
243 possessing a 6×His-tag. The *nsp8* gene was cloned into the modified pET-32a vector
244 with the N-terminus possessing a trx-His_6-tag and PreScission Protease site. The
245 *nsp12*-pET-21b and *nsp8*-pET-32a plasmids were transformed into *E. coli* BL21
246 (DE3) and the transformed cells were cultured at 37 °C in LB with a final
247 concentration of 100 μg/ml ampicillin. The *nsp7*-pET-21b plasmid was transformed
248 into *E. coli* Rosetta-gami2 (DE3) and the transformed cells were cultured at 37 °C in
249 LB contanning a final concentration of 100 μg/ml ampicillin and 25 μg/ml
250 chloramphenicol.

251 Bacterial cultures were incubated by shaking at 200 rpm, 37 °C to an OD600 of 0.8,
252 then the temperature was lowered to 16 °C and a final concentration of 0.2 mM of
253 isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein
254 expression for 20 h. Subsequently, the cells were harvested by centrifugation at 5000
255 ×*g* for 15 min at 4 °C. The pellet was resuspended in lysis buffer containing 25 mM
256 TrisHCl pH 8.0, 250 mM NaCl, 4 mM $MgCl_2$, 10% Glycerol and homogenized with a
257 high-pressure cell disrupter at 4 °C. The lysate was centrifuged at 25,200 ×*g* for 50

258 min at 4 °C, and the supernatant was then loaded onto Ni-NTA affinity
259 chromatography column for purification. Nsp12, nsp7 were eluted with elution buffer
260 (25 mM TrisHCl pH 8.0, 200 mM NaCl, 4 mM $MgCl_2$, 250 mM Imidazole). Nsp8
261 was subject to on-column tag cleavage by PPase (PreScission Protease) and then
262 eluted by the lysis buffer. Next, all proteins were purified by ion exchange
263 chromatography (Source 15Q HR 16/10, GE Healthcare) and size exclusion
264 chromatography (Superdex 200 10/300GL, GE Healthcare) with 25 mM TrisHCl pH
265 8.0, 200 mM NaCl, 4 mM $MgCl_2$.
266 For nsp12-nsp7-nsp8 complex assembly, protein samples were mixed with the molar
267 ratio of nsp7:nsp8:nsp12=2:2:1 at 4 °C overnight. The incubated RdRp complex was
268 then concentrated with a 100 kDa molecular weight cut-off centrifugal filter unit
269 (Millipore Corporation) and then further purified by size exclusion chromatography
270 using a Superdex 200 10/300 GL column in 25 mM TrisHCl pH 8.0, 200 mM NaCl, 4
271 mM $MgCl_2$. SARS-CoV RdRp was purified by the same procedure. The NS5 RdRp
272 domain of ZIKV was purified as previously described[42].
273 **Preparation of fluorescently labeled RNA for polymerase activity**
274 RNA template-product duplex was designed according to the published SARS-CoV-2
275 RNA extension assays[31]. FAM labeled 13 nt oligonucleotide with the sequence of
276 FAM - GCUAUGUGAGAUU and the 23 nt complementary RNA strand with the
277 sequence of AGUUAACUUUAAUCUCACAUAGC were synthesized for
278 polymerase activity assay. A 26 nt RNA strand with the sequence
279 UAGCUGGCCUUAAAAUCUCACAUAGC was synthesized for substrate efficiency

280 assay. And a 29 nt RNA strand with the sequence
281 AGUUAACUUUAAAGGGAAUCUCACAUAGC were used for the detection of
282 mutation rates due to the incorporation of MTP and MDP into RNA. All unmodified
283 and 5' FAM-labeled RNA oligonucleotides were purchased from Tsingke
284 Biotechnology Co.,Ltd. The RNA strands were mixed in equal molar ratio in DEPC
285 water, annealed by heating it to 95 °C for 10 min and gradually cooling to room
286 temperature to make the RNA duplexes.

287 **RdRp polymerase activity assays**

288 DEPC-treated water was used in the preparation of all solutions. RdRp at final
289 concentration of 2 μM was incubated with 200 nM 13+23 nt RNA duplex and 50 μM
290 NTPs (or NDPs / NMPs) in a 20 μl reaction buffer containing 20 mM TrisHCl pH8.0,
291 10 mM KCl, 10 mM $MgCl_2$, 0.01% Triton-X100, 1 mM DTT for 1 min at 37 ℃[31], the
292 reactions were stopped with 2 × stop buffer (10 M urea, 50 mM EDTA). Eventually,
293 glycerol was added to a concentration of 6.5%. 10 μl RNA product for each reaction
294 was resolved on 20% denaturing polyacrylamide-urea gels and imaged with a
295 Tanon-5200 Multi Fluorescence Imager. The experiments described were performed
296 in triplet, unless specified otherwise.

297 **RNA extension assays with NDPs as substrate**

298 RdRp of different viruses at final concentration of 2 μM was incubated with 200 nM
299 13+23 nt RNA duplex and 50 μM NTPs (or NDPs/NMPs/ADP/GDP/CDP/UDP) in a
300 20 μl reaction buffer containing 20 mM TrisHCl pH8.0, 10 mM KCl, 10 mM $MgCl_2$,
301 0.01% Triton-X100, 1 mM DTT for 1 min at 37 ℃.

302 **Sanger sequencing**

303 The RNA product, primer1 (CCGCTCGAGCGGAGTTAACTTT) and dNTPs were
304 gently mixed at 65 °C for 10 min. The mixture was cooled down to 24 °C then 10×
305 MMulV buffer and MMulV reverse transcriptase were added and incubated for 2 h to
306 obtain cDNA. Thereafter, the reverse transcriptase was inactivated at 70 °C for 15 min.
307 Primer2 (CGCGGATCCGCGGCTATGTGAGAT) was then added, and the product
308 was amplified by PCR catalyzed by the high-fidelity DNA polymerase pfu-Phusion.
309 The PCR product was digested with XhoI and BamHI and ligated into pET-15b vector,
310 and transformed into *E-coli* DH5α to obtain clones. Plasmids from positive clones
311 were then sequenced with T7 universal primers.

312 **High-throughput sequencing**

313 The dsRNA products were mixed with primer1
314 (acgatgcaaagtctcgacaaatggtcgataccaattcaCCGCTCGAGCGGAGTTAACTTT) and
315 dNTPs at 65 °C for 10 min and then cooled down to 24 °C for primer annealing. 10 ×
316 MMulV buffer, RNase inhibitor, and MMulV reverse transcriptase were added to the
317 above system at 24 °C for 2 h to obtain cDNA, and then the enzymes were inactivated
318 at 70 °C for 15 min. Subsequently, primer2
319 (cagataaactataattcctaatcgcgaggtggcactgcaaCGCGGATCCGCGGCTATGTGAGA) and
320 high-fidelity DNA polymerase pfu-Phusion were added to perform PCR amplification.
321 PCR products were then purified by agarose gel electrophoresis for high-throughput
322 sequencing (Sangon Biotech).

323 **Substrate efficiency assay**

324 RNA extension was performed by incubating 2 μM RdRp, 250 nM 13+23 nt RNA
325 duplex, 50 μM substrate (which are NTPs, NDPs) in a 20 μl reaction system at 37 °C
326 for 0, 10, 20, 30, 40, 50, 60 s. RdRp catalyzed reactions each containing three NTPs
327 and one NDP (ADP GTP CTP UTP, ATP GDP CTP UTP, ATP GTP CDP UTP and
328 ATP GTP CTP UDP) and 13+26 nt RNA duplex was performed for 1 min. The
329 reactions were stopped with 2 × stop buffer. Grayscale analysis was then carried out
330 using ImageJ and ilustrated by using OriginPro 8 (OriginLab, USA).

331 **RNA extension assays with MDP as substrate**

332 A 13+26 nt RNA duplex was used in this assay (Fig.5e). The 5' end of the RNA
333 product strand was labeled with a 6-carboxyfluorescein (FAM) group, which allows
334 us to monitor RNA elongation. RdRp can incorporate NHC monophosphate into RNA
335 at the position where the nucleotide in the template is G or A. MDP and MTP were
336 purchased from MedChemExpress.

337 For RNA extension RdRp at final concentration of 2 μM was incubated with 200 nM
338 13+26 nt RNA duplex and 50 μM ATP GTP UTP or CTP GTP UTP with or without
339 50 μM MTP/MDP in 20 μl reaction buffer for 1min at 37 °C. The reactions were
340 stopped with 2 × stop buffer. 10 μl RNA product for each reaction was resolved on
341 20% denaturing polyacrylamide-urea gels and imaged with a Tanon-5200 Multi
342 Fluorescence Imager to confirm the extension. Grayscale analysis was then carried
343 out using ImageJ and ilustrated by using OriginPro 8 (OriginLab, USA).

344 **Incorporation mutation rate detection**

345 RdRp at final concentration of 2 μM was incubated with 200 nM 13+29 nt RNA
346 duplex and 50 μM ATP GTP UTP, 800 nM or 5 μM or 50 μM CTP and 50 μM
347 MTP/MDP in 20 μl reaction buffer for 1min at 37 °C. Mutation rates were determined
348 by High-throughput sequencing as mentioned above.

349 **Fluorescence assay for T7 RNA polymerase transcription**

350 Transcription of T7 RNA polymerase was measured by a fluorescence assay which
351 employs a pair of nucleic acid probes with fluorescent and quenching groups
352 respectively to detect mRNA production. When the probes pair with the mRNA FAM
353 and quenching group gets closer and the fluorescence produced by FAM will be
354 quenched. The change of the fluorescent signal is proportional to the mRNA
355 production.

356 **Data availability**

357 All data generated or analyzed during this study are included in this published article
358 (and its Supplementary Information files).

359 **Acknowledgements** The authors thank Yan Zhang at Zhejiang University for offering
360 SARS-CoV-2 RdRp gene, Peihui Wang at Cheeloo College of Medicine, Shandong
361 University for offering SARS-CoV RdRp gene, Haitao Yang at ShanghaiTech
362 University for offering ZIKV NS5 gene, Wuxi Biortus Biosciences Co. Ltd for
363 providing SARS-CoV-2 RdRp expressed in insect cells, Carina Muyao Gu at
364 CBe-Learn School for helping with the English expression of the manuscript. This
365 work was supported by Shandong Provincial Key Research and Development
366 Program (2020CXGC011305).

367 **Author contributions** M.W. and L.G. conceived, designed the experiments and
368 analyzed the data. M.W. performed functional assays of RdRp. M.W., C.W., and X.J.
369 participated in expression, purification of the SARS-CoV-2 RdRp complex and other
370 proteins. N.L., F.Z., H.D., S.W., M.C. helped with the experiments and participated in
371 figure and manuscript preparation. M.W. and L.G. wrote the manuscript.

372 **Competing interests** The authors declare no competing interests.

373 **References**

374 1 Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure,
375 replication, and pathogenesis. *J Med Virol* **92**, 418-423,
376 doi:10.1002/jmv.25681 (2020).

377 2 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China,
378 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).

379 3 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel
380 coronavirus: implications for virus origins and receptor binding. *The Lancet*
381 **395**, 565-574, doi:10.1016/s0140-6736(20)30251-8 (2020).

382 4 Chan, J. F. *et al.* Genomic characterization of the 2019 novel
383 human-pathogenic coronavirus isolated from a patient with atypical
384 pneumonia after visiting Wuhan. *Emerg Microbes Infect* **9**, 221-236,
385 doi:10.1080/22221751.2020.1719902 (2020).

386 5 Coronaviridae Study Group of the International Committee on Taxonomy of,
387 V. The species Severe acute respiratory syndrome-related coronavirus:
388 classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536-544,

389 doi:10.1038/s41564-020-0695-z (2020).

390 6 Liu, C. *et al.* Research and Development on Therapeutic Agents and Vaccines
391 for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent Sci* **6**,
392 315-331, doi:10.1021/acscentsci.0c00272 (2020).

393 7 Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune
394 escape. *Nat Rev Microbiol* **19**, 409-424, doi:10.1038/s41579-021-00573-0
395 (2021).

396 8 Chen, J., Gao, K., Wang, R. & Wei, G. W. Prediction and mitigation of
397 mutation threats to COVID-19 vaccines and antibody therapies. *Chem Sci* **12**,
398 6929-6948, doi:10.1039/d1sc01203g (2021).

399 9 Chen, J., Gao, K., Wang, R. & Wei, G. W. Revealing the Threat of Emerging
400 SARS-CoV-2 Mutations to Antibody Therapies. *J Mol Biol* **433**, 167155,
401 doi:10.1016/j.jmb.2021.167155 (2021).

402 10 Rudd, C. E. GSK-3 Inhibition as a Therapeutic Approach Against SARs CoV2:
403 Dual Benefit of Inhibiting Viral Replication While Potentiating the Immune
404 Response. *Front Immunol* **11**, 1638, doi:10.3389/fimmu.2020.01638 (2020).

405 11 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in
406 China. *Nature* **579**, 265-269, doi:10.1038/s41586-020-2008-3 (2020).

407 12 Zhang, Y. Z. & Holmes, E. C. A Genomic Perspective on the Origin and
408 Emergence of SARS-CoV-2. *Cell* **181**, 223-227,
409 doi:10.1016/j.cell.2020.03.035 (2020).

410 13 Wu, C. *et al.* Analysis of therapeutic targets for SARS-CoV-2 and discovery of

411 potential drugs by computational methods. *Acta Pharm Sin B* **10**, 766-788,
412 doi:10.1016/j.apsb.2020.02.008 (2020).
413 14 Subissi, L. *et al.* One severe acute respiratory syndrome coronavirus protein
414 complex integrates processive RNA polymerase and exonuclease activities.
415 *Proc Natl Acad Sci U S A* **111**, E3900-3909, doi:10.1073/pnas.1323705111
416 (2014).
417 15 Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12
418 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, 2342,
419 doi:10.1038/s41467-019-10280-3 (2019).
420 16 Hillen, H. S. *et al.* Structure of replicating SARS-CoV-2 polymerase. *Nature*
421 **584**, 154-156, doi:10.1038/s41586-020-2368-8 (2020).
422 17 McKee, D. L., Sternberg, A., Stange, U., Laufer, S. & Naujokat, C. Candidate
423 drugs against SARS-CoV-2 and COVID-19. *Pharmacol Res* **157**, 104859,
424 doi:10.1016/j.phrs.2020.104859 (2020).
425 18 Chu, C. K. *et al.* Antiviral Activity of Nucleoside Analogues against
426 SARS-coronavirus (SARS-CoV). **17**, 285-289,
427 doi:10.1177/095632020601700506 (2006).
428 19 Stein, D. S. & Moore, K. H. P. Phosphorylation of Nucleoside Analog
429 Antiretrovirals: A Review for Clinicians. *Pharmacotherapy: The Journal of*
430 *Human Pharmacology and Drug Therapy* **21**, 11-34,
431 doi:https://doi.org/10.1592/phco.21.1.11.34439 (2001).
432 20 Agostini, M. L. *et al.* Small-Molecule Antiviral β-d-N (4)-Hydroxycytidine

433 Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to
434 Resistance. *Journal of virology* **93**, e01348-01319, doi:10.1128/JVI.01348-19
435 (2019).
436 21 Picarazzi, F., Vicenti, I., Saladini, F., Zazzi, M. & Mori, M. Targeting the
437 RdRp of Emerging RNA Viruses: The Structure-Based Drug Design Challenge.
438 *Molecules* **25**, doi:10.3390/molecules25235695 (2020).
439 22 Crotty, S. *et al.* The broad-spectrum antiviral ribonucleoside ribavirin is an
440 RNA virus mutagen. *Nature Medicine* **6**, 1375-1379, doi:10.1038/82191
441 (2000).
442 23 Shannon, A. *et al.* Rapid incorporation of Favipiravir by the fast and
443 permissive viral RNA polymerase complex results in SARS-CoV-2 lethal
444 mutagenesis. *Nat Commun* **11**, 4682, doi:10.1038/s41467-020-18463-z (2020).
445 24 Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently
446 emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res* **30**, 269-271,
447 doi:10.1038/s41422-020-0282-0 (2020).
448 25 Sheahan, T. P. *et al.* An orally bioavailable broad-spectrum antiviral inhibits
449 SARS-CoV-2 in human airway epithelial cell cultures and multiple
450 coronaviruses in mice. *Sci Transl Med* **12**, eabb5883,
451 doi:10.1126/scitranslmed.abb5883 (2020).
452 26 Kabinger, F. *et al.* Mechanism of molnupiravir-induced SARS-CoV-2
453 mutagenesis. *Nat Struct Mol Biol* **28**, 740-746,
454 doi:10.1038/s41594-021-00651-0 (2021).

455 27 Yang, G., Franklin, M., Li, J., Lin, T. C. & Konigsberg, W. Correlation of the
456 Kinetics of Finger Domain Mutants in RB69 DNA Polymerase with Its
457 Structure. *Biochemistry* **41**, 2526-2534, doi:10.1021/bi0119924 (2002).
458 28 Burke, C. R. & Luptak, A. DNA synthesis from diphosphate substrates by
459 DNA polymerases. *Proc Natl Acad Sci U S A* **115**, 980-985,
460 doi:10.1073/pnas.1712193115 (2018).
461 29 Garforth, S. J., Parniak, M. A. & Prasad, V. R. Utilization of a
462 deoxynucleoside diphosphate substrate by HIV reverse transcriptase. *PLoS*
463 *One* **3**, e2074, doi:10.1371/journal.pone.0002074 (2008).
464 30 Gottesman, M. E. & Mustaev, A. Ribonucleoside-5'-diphosphates (NDPs)
465 support RNA polymerase transcription, suggesting NDPs may have been
466 substrates for primordial nucleic acid biosynthesis. *J Biol Chem* **294**,
467 11785-11792, doi:10.1074/jbc.RA119.009074 (2019).
468 31 Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA
469 polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499-1504,
470 doi:10.1126/science.abc1560 (2020).
471 32 Choy, K. T. *et al.* Remdesivir, lopinavir, emetine, and homoharringtonine
472 inhibit SARS-CoV-2 replication in vitro. *Antiviral Res* **178**, 104786,
473 doi:10.1016/j.antiviral.2020.104786 (2020).
474 33 Williamson, B. N. *et al.* Clinical benefit of remdesivir in rhesus macaques
475 infected with SARS-CoV-2. *Nature* **585**, 273-276,
476 doi:10.1038/s41586-020-2423-5 (2020).

477 34 Xie, X. *et al.* A nanoluciferase SARS-CoV-2 for rapid neutralization testing
478 and screening of anti-infective drugs for COVID-19. *Nat Commun* **11**, 5214,
479 doi:10.1038/s41467-020-19055-7 (2020).
480 35 Zhou, S. *et al.* beta-d-N4-hydroxycytidine Inhibits SARS-CoV-2 Through
481 Lethal Mutagenesis But Is Also Mutagenic To Mammalian Cells. *J Infect Dis*
482 **224**, 415-419, doi:10.1093/infdis/jiab247 (2021).
483 36 Malone, B. & Campbell, E. A. Molnupiravir: coding for catastrophe. *Nat*
484 *Struct Mol Biol* **28**, 706-708, doi:10.1038/s41594-021-00657-8 (2021).
485 37 Noreen, S., Maqbool, I. & Madni, A. Dexamethasone: Therapeutic potential,
486 risks, and future projection during COVID-19 pandemic. *Eur J Pharmacol*
487 **894**, 173854, doi:10.1016/j.ejphar.2021.173854 (2021).
488 38 Quimque, M. T. J. *et al.* Virtual screening-driven drug discovery of
489 SARS-CoV2 enzyme inhibitors targeting viral attachment, replication,
490 post-translational modification and host immunity evasion infection
491 mechanisms. *J Biomol Struct Dyn* **39**, 4316-4333,
492 doi:10.1080/07391102.2020.1776639 (2021).
493 39 Jiang, Y., Yin, W. & Xu, H. E. RNA-dependent RNA polymerase: Structure,
494 mechanism, and drug discovery for COVID-19. *Biochem Biophys Res*
495 *Commun* **538**, 47-53, doi:10.1016/j.bbrc.2020.08.116 (2021).
496 40 Meier, C. Nucleoside diphosphate and triphosphate prodrugs - An unsolvable
497 task? *Antivir Chem Chemother* **25**, 69-82, doi:10.1177/2040206617738656
498 (2017).

41 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).

42 Ferrero, D. S. *et al.* Supramolecular arrangement of the full-length Zika virus NS5. *PLoS Pathog* **15**, e1007656, doi:10.1371/journal.ppat.1007656 (2019).

521 **Figure legends**

522 **Fig.1** RdRp uses both NTPs and NDPs as a substrate. **a** Size-exclusion chromatogram
523 of the SARS-CoV-2 nsp12-nsp7-nsp8(RdRp) complex. RdRp was also characterized
524 by SDS-PAGE. **b** The 13+23 nt RNA template-product duplex. The direction of RNA
525 extension is shown. The color of the depicted circles indicates the experimental
526 design: blue, RNA template strand; red, RNA product strand. The 5'end of the RNA
527 product contains a FAM fluorescent label. **c** Incubation of the RdRp with RNA duplex
528 and NTPs leads to RNA extension. Nsp12 alone has almost no polymerase activity,
529 and requires nsp7 and nsp8 to form an RdRp complex to have polymerase activity. **d**
530 RdRp was incubated with the 13+23 nt duplex in the presence of NMPs, NDPs or
531 NTPs respectively. These are showing duplicate experiments performed with different
532 batches RdRp. The positions of the template RNA and the full-length extension
533 product are indicated on denaturing gel. **e** Electrophoretic separation of reaction
534 products of single nucleoside diphosphates as substrates. The positions of the original
535 RNA and extension products are indicated. **f** Part of the Sanger sequencing
536 chromatogram of the RT-PCR products corresponding to the extended 10 nt RNA on
537 the 13+23 nt RNA duplex with NDPs as substrate. The dash line circle represents the
538 10 nt extension. **g** Schematic of high-throughput sequencing of RdRp reaction
539 products.

540 **Fig.2** The activities of NDPs and NTPs are comparable. **a** Primer-extension reactions
541 on 13+23 nt RNA templates sampled at indicated time points. **b** Quantification of
542 elongation products for (**a**). The experiment was performed in triplet. **c** The 13+26 nt
543 RNA template-product duplex. The direction of RNA extension is shown. The 5'end

544 of the RNA product was labeled by FAM. **d** RdRp catalyzed reactions containing
545 different substrates. Lane 1: no substrate; lane 2: NTPs; lane 3: ATPGTPCTP; lane4:
546 ATPGTPCTP UDP; lane5: GTPCTPUTP; lane6: GTPCTPUTP ADP; lane7:
547 ATPCTPUTP; lane8: ATPCTPUTP GDP; lane9: ATPGTPUTP; lane10: ATPGTPUTP
548 CDP. The positions of the 13+26 nt RNA and extension products are indicated. **e**
549 Quantification of elongation products for (**d**).

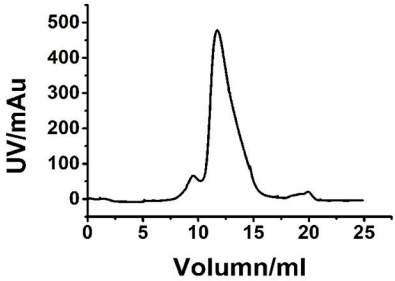
550 **Fig.3** RdRp incorporates NHC into RNA with MDP as a substrate. **a** Metabolism of
551 nucleoside analogues molnupiravir in host cell. A two-step model of
552 molnupiravir-induced RNA mutagenesis. **b** The 13+26 nt RNA allows for RNA
553 extension by 13 nucleotides. NHC monophosphate can be incorporated into growing
554 RNA instead of C or U. **c** RdRp catalyzed reactions containing different substrates.
555 Lane 1: no substrate; lane 2: NTPs; lane 3: ATPGTPUTP; lane4: ATPGTPUTP MTP;
556 lane5: ATPGTPUTP MDP; lane6: ATPGTPCTP; lane7: ATPGTPCTP MTP; lane8:
557 ATPGTPCTP MDP. The positions of the 13+26 nt RNA and extension products are
558 indicated. RNA elongation stalls at the expected positions when the cognate NTP is
559 absent from the reaction. **d** Quantification of elongation products for (**c**). **e** The 13+29
560 nt RNA allows for RNA extension by 16 nucleotides. NHC monophosphate can pair
561 with G or A for incorporation into growing RNA.

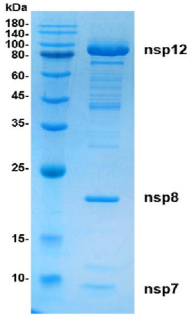
562 **Fig.4** Substrate properties of other RNA viruses **a** RdRp of SARS-CoV-2 and
563 SARS-CoV and NS5 of ZIKV were incubated with the 13+23 nt RNA duplex in the
564 presence of NTPs or NDPs respectively. The positions of the template RNA and the
565 full-length extension product are indicated. **b** The transcription of T7 RNA

566 polymerase was measured by monitoring the signal of a fluorescent probe.
567 Fluorescent signal continuously changed when transcription occurred, indicating the
568 presence of RNA product. However, when NDPs were used as a substrate, the
569 fluorescent signal remained unchanged, indicating the lack of RNA product.
570 Substitution of ATP by ADP led to weak transcription, any other NTP replaced by
571 NDP would inhibit the transcription completely.

572 **Fig.5** Nucleotide diphosphate analog prodrugs designed by our strategy would have
573 less potential to cause side effects. The terminal phosphate group (β-phosphate) is
574 modified by lipophilic masks to give the prodrug more membrane permeability. Inside
575 cell the masks are removed by enzymatic or chemical reactions. The ribose moiety
576 and the base group should be designed in such a way that the diphosphate form drug
577 can evade NDPK thus eliminating the risk of being incorporated into host mRNA or
578 DNA (Note: NDPs and dNDPs are phosphorylated by the same NDPK).

a

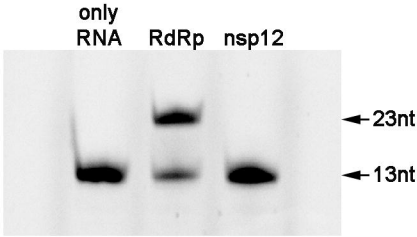

UV/mAu
500
400
300
200
100
0
0
5
10
15
20
25
Volumn/ml


kDa
180-
140-
100-
80-
60-
45-
35-
25-
15-
10-
nsp12
nsp8
nsp7

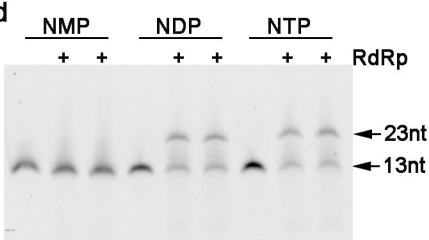
b

FAM
5'
3'
polymerization
5'

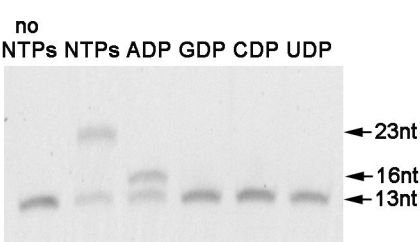
c


only RNA
RdRp
nsp12
23nt
13nt

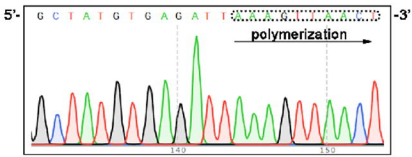
d


NMP
NDP
NTP
RdRp
23nt
13nt

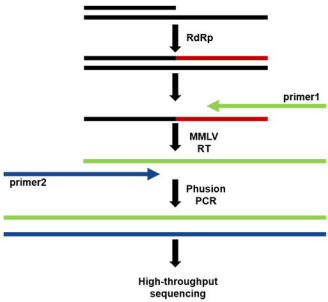
e

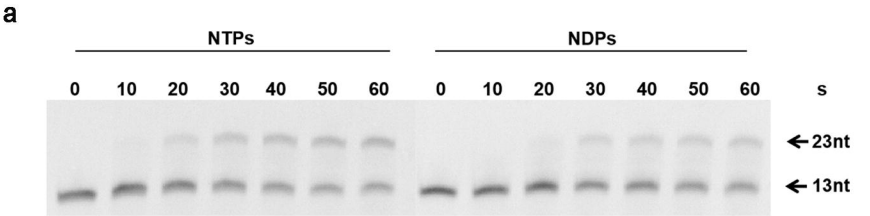

no NTPs
NTPs
ADP
GDP
CDP
UDP
23nt
16nt
13nt

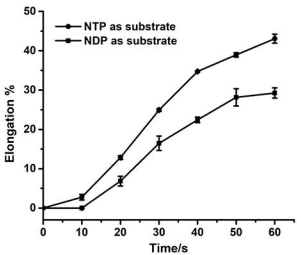
f


5'-
-3'
polymerization
140
150

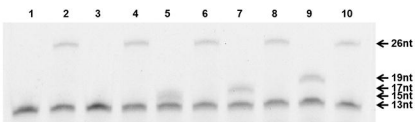
g

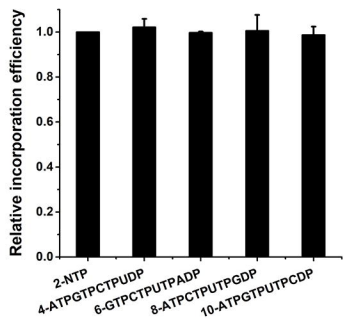

RdRp
primer1
MMLV RT
primer2
Phusion PCR
High-throughput sequencing

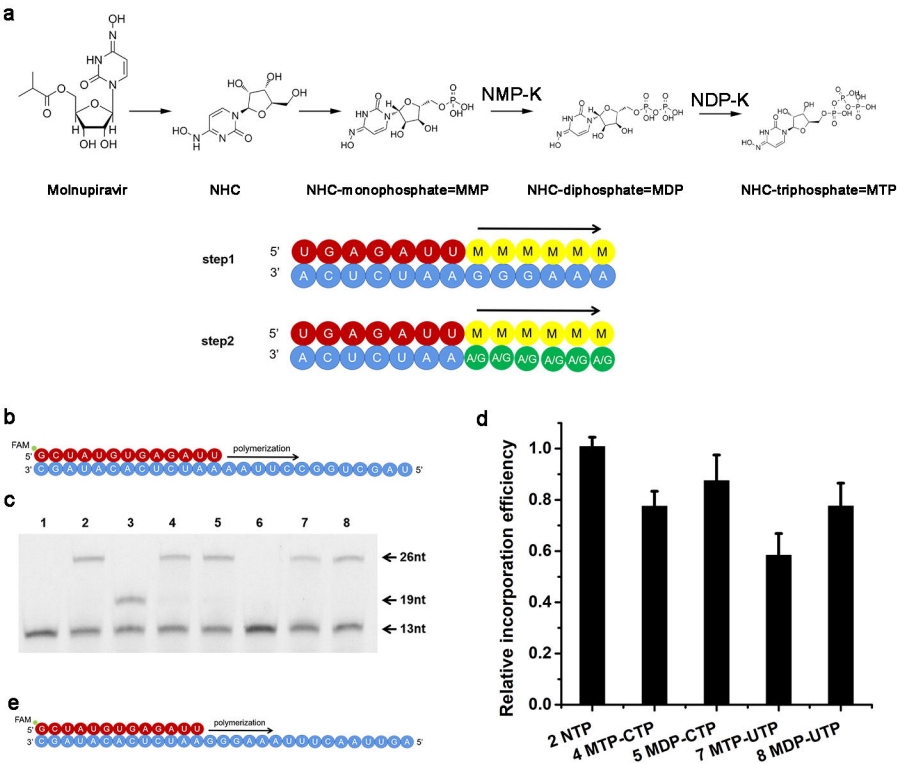

a
NTPs
NDPs
0 10 20 30 40 50 60
0 10 20 30 40 50 60
s
23nt
13nt


b
NTP as substrate
NDP as substrate
Elongation %
Time/s
0
10
20
30
40
50
0 10 20 30 40 50 60

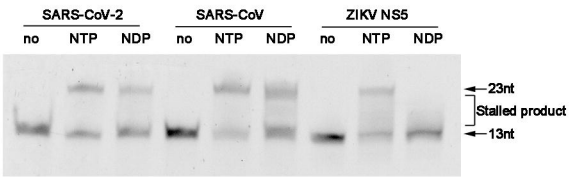
c
FAM
5'
3'
polymerization
5'


d
1 2 3 4 5 6 7 8 9 10
26nt
19nt
17nt
15nt
13nt

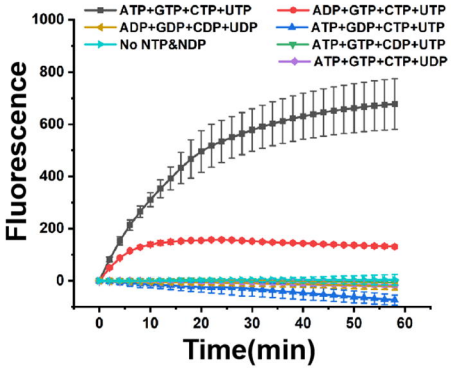

e
Relative incorporation efficiency
0.0
0.2
0.4
0.6
0.8
1.0
2-NTP
4-ATPGTPCTPUDP
6-GTPCTPUTPADP
8-ATPCTPUTPGDP
10-ATPGTPUTPCDP

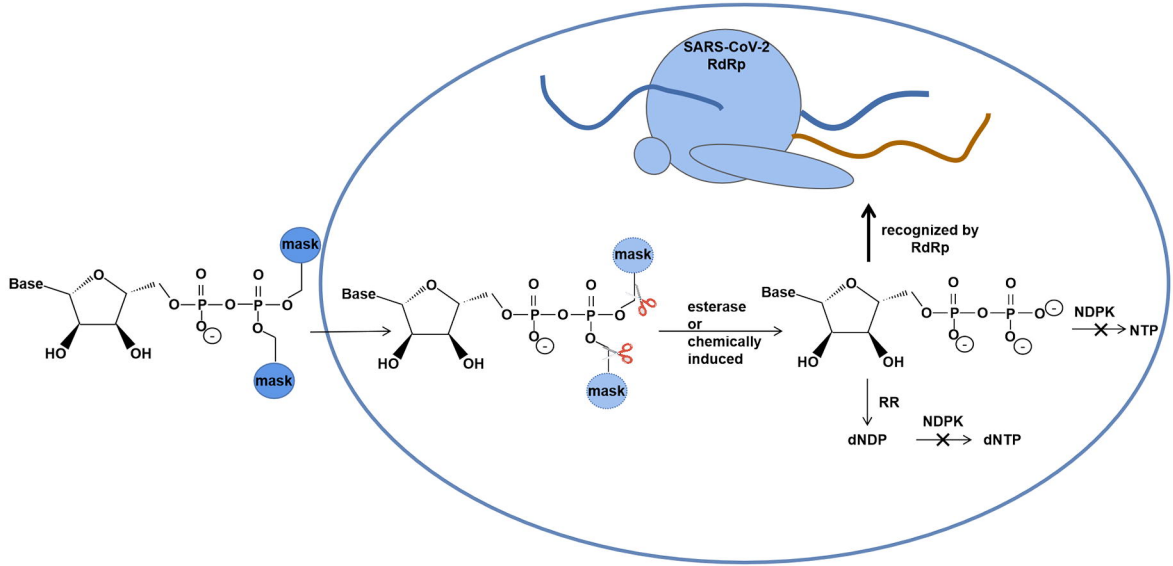

a
NMP-K
NDP-K
Molnupiravir
NHC
NHC-monophosphate=MMP
NHC-diphosphate=MDP
NHC-triphosphate=MTP
step1
step2
5'
3'
U G A G A U U M M M M M M
A C U C U A A G G G A A A
A/G A/G A/G A/G A/G A/G
b
FAM
polymerization
c
1 2 3 4 5 6 7 8
26nt
19nt
13nt
d
Relative incorporation efficiency
0.0
0.2
0.4
0.6
0.8
1.0
2 NTP
4 MTP-CTP
5 MDP-CTP
7 MTP-UTP
8 MDP-UTP
e

a


SARS-CoV-2
SARS-CoV
ZIKV NS5
no
NTP
NDP
no
NTP
NDP
no
NTP
NDP
23nt
Stalled product
13nt

b


ATP+GTP+CTP+UTP
ADP+GTP+CTP+UTP
ADP+GDP+CDP+UDP
ATP+GDP+CTP+UTP
No NTP&NDP
ATP+GTP+CDP+UTP
ATP+GTP+CTP+UDP
Fluorescence
1000
800
600
400
200
0
0
10
20
30
40
50
60
Time(min)


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