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

4

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

7

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

10

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

12

13

14

15

16

17

18

19

20

21

22

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 outcome⁵⁻⁸. As result, there is an urgent need to find effective medication or treatment to counter the virus⁹⁻¹¹.

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 forms¹⁷⁻²¹. 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 circle²³. 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 catalysis²⁴⁻²⁸.

These structures are crucial to understand the mechanism of RdRp activity and inhibition by remdesivir or other inhibitors²³. 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

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

Results

Exoribonuclease activity was observed during characterization of RdRp polymerase activity

Since RdRp is by far the most attractive target for anti-coronavirus drug development, thorough characterization of RdRp polymerase activity would greatly without doubt contribute to discovering the effective strategy for drug screening^{15,16}. Based on previous research, SARS-CoV-2 RdRp polymerase activity was measured by conducting a conventional RNA elongation assay²³. When nucleoside triphosphate (NTP) substrate is added, elongation catalyzed by RdRp complex occurred on the RNA duplex, resulting in an intact double-stranded RNA product. Nsp12 alone could not catalyze the extension reaction (Fig.1a). Surprisingly, when NTP was absent from the system for control, the RNA duplex was markedly degraded. Degradation of RNA duplex was inhibited when EDTA was added (Fig.1b). Our results indicated that in addition to polymerase activity, RdRp also displays exoribonuclease activity despite that neither sequence nor structure gives any clue of the existence of a functional domain.

SARS-CoV-2 RdRp also has proofreading activity

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

mismatch at the 3'-end of the extending strand of RNA was prepared (Fig.2b). An extension assay was then performed by using M1 RNA as the template. Result showed that RdRp successfully extended the shorter RNA strand, just as it did with the RNA duplex without mismatch, although the productivity was slightly lower (Fig.3a). We hypothesized two possibilities: first, RdRp simply ignored the mismatch and processed to 3' end leaving the U: G mismatch unchanged; second, RdRp corrected the mismatch and then finished the extension. The nucleotide mismatch always slowed down the extension reaction, no matter the case.

To figure out this process, nascent RNA strands were reverse transcribed into cDNAs. The cDNAs were then amplified by 10 cycles of PCR and sequenced by high-throughput sequencing. For each product, at least 20,000 sequences were obtained (SI). The data showed three kinds of nascent RNAs were produced: (1) mutation repaired GCUAUGUGAGAUGAAAGUUAACU, 39.29%; (2) mutation ignored GCUAUGUGAGAUUAAAGUUAACU, 13.82%; (3) mutation leading to frame shift GCUAUGUGAGAUUGAAAGUUAACU, 36.89%.

We also challenged RdRp with the RNA template (M2 RNA) consisting of a 13 nt and a 23 nt RNA with a UU: GC double mismatch at the 3'-end of the extending strand of RNA (Fig.2c). In this experiment, RNA extension was observed as well, but with a much lower productivity (Fig.3a). The products were then processed and sequenced as mentioned above. The results included four kinds of RNAs: (1) mutations half repaired GCUAUGUGAGAUGAAAGUUAACU, 54.88%; (2) mutations completely repaired GCUAUGUGAGACGAAAGUUAACU, 20.69%; (3) mutation ignored

GCUAUGUGAGAUUAAAGUUAACU, 3.01%; (4) frame shift,
GCUAUGUGAGAUUGAAAGUUAACU, 11.42%.

The data strongly suggested that SARS-CoV-2 RdRp itself possesses proofreading activity.

Combination of RdRp and nsp14-ExoN exhibited higher proofreading activity

Previous studies indicated that nsp14 is highly conserved within the Coronaviridae family and critical for the SARS-CoV replication fidelity^{7,29,30}. Since we discovered that SARS-CoV-2 RdRp itself has proofreading activity, we supposed that the combination of the two proteins may exhibit higher proofreading activity. To confirm this hypothesis, we first tested the exoribonuclease activity of nsp14-nsp10 complex in vitro and chose 300 nM as the proper concentration of nsp14 for the next experiments³¹ (Fig.3b).

We then conducted a proofreading activity assay for RdRp in the presence of 300 nM of nsp14 with a RNA duplex containing a single mismatch (Fig.3c). High-throughput sequencing indicated three kinds of nascent RNAs were produced: (1) mutation repaired GCUAUGUGAGAUGAAAGUUAACU, 56.91%; (2) mutation ignored GCUAUGUGAGAUUAAAGUUAACU, 22.85%; (3) mutation leading to frame shift GCUAUGUGAGAUUGAAAGUUAACU, 10.13%. In the case of double mismatch, nsp14 did not show significant role in improving the proofreading activity of RdRp: (1) mutations half repaired 61.17%; (2) mutations completely repaired 17.21%; (3) mutation ignored 5.72%; (4) frame shift 5.90% (Fig.3c). These data unambiguously

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

RdRp synthesizes RNA by using NDP as substrate

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

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

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

COVID-19 treatment³²⁻³⁵. Among all these chemicals reported by far, molnupiravir has been viewed as a promising drug candidate. The molecular mechanism of molnupiravir has been determined recently (Fig.5a). Once ingested by patients, molnupiravir undergoes stepwise phosphorylation to yield the active nucleoside triphosphate analogue (β -d-N4-hydroxycytidine (NHC) triphosphate, or called MTP). The RdRp then uses MTP as substrate and incorporates NHC monophosphate into RNA at the position where the nucleotide in the template is A or G, resulting in “error catastrophe” and virus death³⁶⁻³⁸.

Although the mechanism of molnupiravir has been well established, our finding that RdRp can use NDP as a substrate still raises a question: is NHC diphosphate (MDP) also an active form and recognized by RdRp as a substrate? To address this question, the 13+23 nt RNA duplex was used to perform the RNA extension assay with NDPs and MDP as the substrates in comparison with the extension using NTPs and MTP as the substrates. 13+23 nt RNA duplex allows ten nucleotides (nt) of extension (Fig.5b). In this case, when uridine diphosphate (UDP) was replaced by MDP, the incorporation of M did not obstruct the incorporation of next subsequent nucleotides, RdRp still finished the extension reaction. Similar results were obtained, in the case of using MTP as a substrate (Fig.5c). Thus, it is reasonable to speculate that both MTP and MDP are the active form of monulpiravir inside the human body.

Discussion

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

developed to suppress the pandemic, they had not produced the anticipated result to withhold the spread of SARS-CoV-2. Development of potent antivirals against SARS-CoV-2 is still an urgent need^{10,11}.

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

The substrate usage of RdRp surprised us greatly. As far as we know, all RNA polymerases, including DNA dependent and RNA dependent RNA polymerases, use nucleoside triphosphate (NTP) as substrate to synthesize RNA under the guidance of the template chain⁴⁰. Our finding that SARS-CoV-2 RdRp uses both NTP and NDP as substrate to synthesize RNA with equal efficiency makes it an extremely unique RNA polymerase. By far, we don’t know why SARS-CoV-2 RdRp is so distinctive or if there are any other enzyme-possessing similar features. We speculate that the ability to use NDP as substrate could be a great advantage for SARS-CoV-2. During the late

stage of infection, the host cell may fail to produce enough ATP and thus other NTPs.

By using NDP as substrate, SARS-CoV-2 can continue the process of genome replication and assembly of progeny virus while most of the metabolic activities stall.

The newly discovered feature of RdRp may profoundly change the strategy of antiviral drug design. Many antiviral drugs are nucleoside analogues that target viral polymerases as inhibitors to terminate RNA chain elongation^{41,42}. However, such chain-terminating antivirals often do not work as effectively as expected in the treatment of COVID-19 because SARS-CoV-2 has exonucleolytic proofreading activity that can remove the incorporated inhibitors^{23,32,33,42-44}. It was thought that the proofreading activity was carried out by nsp14-ExoN and incapacitating nsp14-ExoN may be a way to deal with this issue^{29,31,43}. The proofreading activity of RdRp is unfavorable to this kind of RNA chain terminators. This may also explain why molnupiravir, a mutagenizing nucleoside analog, is much more effective against SARS-CoV-2 than remdesivir since molnupiravir can evade the proofreading mechanism by both RdRp and nsp14-ExoN^{36,37}.

It has long been general knowledge that to become active against virus nucleoside analogue drugs must undergo stepwise addition of phosphate groups to become the triphosphate form. For this outcome, this type of antivirals must be recognized by three kinds of kinases⁴⁵. Which raises a concern that the triphosphate form of antivirals may be incorporated into host mRNA³⁸. Even worse, mutagenic ribonucleoside analogs could be reduced into the 2'-deoxyribonucleotide form by host ribonucleotide reductase and then incorporated into DNA^{36,38}. Our finding that RdRp

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

Materials and Methods

Protein expression and purification

Preparation of SARS-CoV-2 RdRp, composed of nsp12, nsp7 and two copied of the nsp8 subunits, was carried out as described in ref.²³. The SARS-CoV-2 *nsp12*, *nsp10* and *nsp7* gene were cloned into a modified pET-21b vector with the C-terminus possessing a 6×His-tag. The *nsp8* gene was cloned into the modified pET-32a vector with the N-terminus possessing a trx-His₆-tag and PreScission Protease site. The plasmids were transformed into *E. coli* BL21 (DE3) and the transformed cells were cultured at 37 °C in LB with a final concentration of 100 µg/ml ampicillin. The *nsp7*-pET-21b plasmid was transformed into *E. coli* Rosetta-gami2 (DE3) and the transformed cells were cultured at 37 °C in LB with a final concentration of 100 µg/ml ampicillin and 25 µg/ml chloramphenicol. The *nsp14* gene (*nsp14*-ExoN (residues 1–289)) was cloned into the modified pET-28b vector with the N-terminus

possessing a MBP-tag and PreScission Protease site. The plasmid was then transformed into *E. coli* BL21 (DE3) and the transformed cells were cultured at 37 °C in LB with a final concentration of 100 µg/ml kanamycin.

Bacterial cultures were incubated by shaking at 200 rpm, 37 °C to an OD600 of 0.8, then the temperature was lowered to 16 °C and a final concentration of 0.2 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression for 20 h. Subsequently, the cells were harvested by centrifugation at 5000 ×g for 15 min at 4 °C. The pellet was resuspended in lysis buffer containing 25 mM TrisHCl pH 8.0, 250 mM NaCl, 4 mM MgCl₂, 10% Glycerol and homogenized with a high-pressure cell disrupter at 4 °C. The lysate was centrifuged at 25,200 ×g for 50 min at 4 °C, and the supernatant was then loaded onto Ni-NTA affinity chromatography column for purification. Nsp12, nsp10, nsp7 were eluted with elution buffer (25 mM TrisHCl pH 8.0, 200 mM NaCl, 4 mM MgCl₂, 250 mM Imidazole).

Nsp8 was subject to on-column tag cleavage by PPase (PreScission Protease) and then eluted by the lysis buffer. The supernatant of nsp14-ExoN was loaded onto a Dextrin-Sepharose gravity column (GE Healthcare), washed and then eluted with elution buffer containing 25 mM TrisHCl pH 8.0, 100 mM NaCl and 20 mM maltose.

The MBP tag was then removed by PPase digestion. Next, all proteins were purified by ion exchange chromatography (Source 15Q HR 16/10, GE Healthcare) and size exclusion chromatography (Superdex 200 10/300GL, GE Healthcare) with 25 mM TrisHCl pH 8.0, 200 mM NaCl, 4 mM MgCl₂.

For nsp12-nsp7-nsp8 complex assembly, protein samples were mixed with the molar ratio of nsp7:nsp8:nsp12=2:2:1 at 4 °C overnight. The incubated RdRp complex was then concentrated with a 100 kDa molecular weight cut-off centrifugal filter unit (Millipore Corporation) and then further purified by size exclusion chromatography using a SuperdexS200 10/300 GL column in 25 mM TrisHCl pH 8.0, 200 mM NaCl, 4 mM MgCl₂.

Preparation of fluorescently labeled RNA for polymerase activity and proofreading assays

RNA template-product duplex was designed according to the published SARS-CoV-2 RNA extension assays²³. FAM labeled 13 nt oligonucleotide with the sequence of FAM - GCUAUGUGAGAUU and the 23 nt complementary oligonucleotide with the sequence of CGAUACACUCUAAUUCAAUUGA were synthesized for polymerase activity assay. Two 23 nt oligonucleotides with U: G mutation (M1 RNA, CGAUACACUCUACUUCAAUUGA) or UU: CG mutation (M2 RNA, CGAUACACUCUGCUUCAAUUGA) were synthesized for proofreading assays. All unmodified and 5' FAM-labeled RNA oligonucleotides were purchased from Tsingke Biotechnology Co.,Ltd.. The oligonucleotides were mixed in equal molar ratio in DEPC water, annealed by heating it to 95 °C for 10 min and gradually cooling to room temperature to make the RNA duplex (Fig.2a, 2b, 2c).

RdRp polymerase activity assays

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

NDP/NMP) in a 20 µl reaction buffer containing 20 mM TrisHCl pH8.0, 10 mM KCl, 10 mM MgCl₂, 0.01% Triton-X100, 1 mM DTT for 40 min at 37 °C²³. Eventually, glycerol was added to a concentration of 6.5%. 10 µl RNA product for each reaction was resolved on 20% denaturing polyacrylamide-urea gels and imaged with a Tanon-5200 Multi Fluorescence Imager.

RdRp exoribonuclease activity assay

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

Nsp14-ExoN exoribonuclease activity assay

Nsp10 and nsp14 were incubated at a molar ratio of 3:1 for nsp14-nsp10 complex assembly³¹. Reaction buffer containing 300 nM of nsp14-nsp10 complex, 20 mM TrisHCl pH 8.0, 10 mM KCl, 10 mM MgCl₂, 0.01% Triton-X100, 1 mM DTT and 250 nM M1 RNA was incubated at 37 °C for 1 h. Reaction system without nsp14-nsp10 complex was used as a negative control. After incubation, glycerol was added to a concentration of 6.5%. The product was resolved on 20% denaturing polyacrylamide-urea gels, and imaged with a Tanon-5200 Multi Fluorescence Imager.

The proofreading activities assays for RdRp and nsp14

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

RNA extension assays with NDP as substrate.

RdRp at final concentration of 2 μ M was incubated with 200 nM 13+23 nt RNA duplex and 25 μ M NTP (or NDP/NMP) in a 20 μ l reaction buffer containing 20 mM TrisHCl pH8.0, 10 mM KCl, 10 mM MgCl₂, 0.01% Triton-X100, 1 mM DTT for 40 min at 37 °C.

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

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

RNA extension assays with MDP as substrate.

A RNA extension assay was performed as mentioned previously³⁷. In this assay the 13+23 nt RNA duplex was used (Fig.5e). The 5'end of the RNA product strand contains a 6-carboxyfluorescein (FAM) fluorescent label, which allows us to monitor RNA elongation. RdRp can incorporate NHC monophosphate into RNA at the position where the nucleotide in the template is A. MDP and MTP were purchased from MedChemExpress.

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

Data availability

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

Acknowledgements

The authors thank Yan Zhang at Zhejiang University for offering SARS-CoV-2 RdRp gene, Carina Muyao Gu at CBe-Learn School for helping with the English expression of the manuscript. This work was supported by Shandong Provincial Key Research and Development Program (2020CXGC011305).

Author contributions

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

Competing interests

The authors declare no competing interests.

References

- 1 Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* **92**, 418-423, doi:10.1002/jmv.25681 (2020).
- 2 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 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,

- doi:10.1038/nsmb999 (2005).
- 18 Johnson, M. A., Jaudzems, K. & Wuthrich, K. NMR Structure of the SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5. *J Mol Biol* **402**, 619-628, doi:10.1016/j.jmb.2010.07.043 (2010).
- 19 Peti, W. *et al.* Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsP7. *J Virol* **79**, 12905-12913, doi:10.1128/JVI.79.20.12905-12913.2005 (2005).
- 20 Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779-782, doi:10.1126/science.abb7498 (2020).
- 21 Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, 2342, doi:10.1038/s41467-019-10280-3 (2019).
- 22 Lehmann, K. C. *et al.* Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic Acids Res* **43**, 8416-8434, doi:10.1093/nar/gkv838 (2015).
- 23 Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499-1504, doi:10.1126/science.abc1560 (2020).
- 24 Vázquez, A. L., Alonso, J. M. & Parra, F. Mutation analysis of the GDD sequence motif of a calicivirus RNA-dependent RNA polymerase. *Journal of 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

526 Analogues. *Angew Chem Int Ed Engl* **55**, 5255-5258,
 527 doi:10.1002/anie.201511808 (2016).
 528

Figure legends

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

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

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

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

positions of the template RNA and the full-length extension product are indicated. **b**

Part of the Sanger sequencing chromatogram of the RT-PCR products corresponding

to the extended 10 nt RNA on the 13+23 nt RNA duplex with NDP as substrate. The

dash line circle represents the 10 nt extension.

Fig.5 RdRp incorporates NHC into RNA with MDP as substrate. **a** Metabolism of

nucleoside analogues molnupiravir in host cell. **b** The 13+23 nt RNA allows for RNA

extension by 10 nucleotides. NHC monophosphate can be incorporated into growing

RNA instead of U or C. **c** RNA elongation stalls at the expected positions when the

cognate NDP is withheld from the reaction. Extension to the end of the template is

possible when MDP replaces UDP in the presence of other substrate NDPs, showing

that using MDP as a substrate RdRp incorporates M into RNA and incorporation of M

does not prevent RNA extension.

Fig.6 Nucleotide diphosphate analog prodrugs designed by our strategy would have

less potential to cause side effects. The terminal phosphate group (β -phosphate) is

modified by lipophilic masks to give the prodrug more membrane permeability. Inside

cell the masks are removed by enzymatic or chemical reactions. The ribose moiety

and the base group should be designed in such a way that the diphosphate form drug

can evade NDPK thus eliminating the risk of being incorporated into host mRNA or

DNA.

a**RNA RNA+RdRp RNA+Nsp12** **b**

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

a

b

c

a

b

c

a

a **b** **c**

	RNA+RdRp					
RNA	All	no UTP		All	no UDP	
only	NTPs	+MTP		NTPs	+MDP

NMP-K **NHC-diphosphate=MDP****NDP-K** **NHC-triphosphate=MTP**