Check for updates

OPEN

Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis

Florian Kabinger [1,5], Carina Stiller [2,5], Jana Schmitzová[1,5], Christian Dienemann[1], Goran Kokic [1], Hauke S. Hillen [3,4], Claudia Höbartner [2✉] and Patrick Cramer [1✉]

Molnupiravir is an orally available antiviral drug candidate currently in phase III trials for the treatment of patients with COVID-19. Molnupiravir increases the frequency of viral RNA mutations and impairs SARS-CoV-2 replication in animal models and in humans. Here, we establish the molecular mechanisms underlying molnupiravir-induced RNA mutagenesis by the viral RNA-dependent RNA polymerase (RdRp). Biochemical assays show that the RdRp uses the active form of molnupiravir, β-D-N^4-hydroxycytidine (NHC) triphosphate, as a substrate instead of cytidine triphosphate or uridine triphosphate. When the RdRp uses the resulting RNA as a template, NHC directs incorporation of either G or A, leading to mutated RNA products. Structural analysis of RdRp–RNA complexes that contain mutagenesis products shows that NHC can form stable base pairs with either G or A in the RdRp active center, explaining how the polymerase escapes proofreading and synthesizes mutated RNA. This two-step mutagenesis mechanism probably applies to various viral polymerases and can explain the broad-spectrum antiviral activity of molnupiravir.

Coronaviruses use an RNA-dependent RNA polymerase (RdRp) for the replication and transcription of their RNA genome[1–6]. RdRp is an important target for the development of antiviral drugs against coronaviruses[1,7–10]. Structures of RdRp have been reported for SARS-CoV-1[11] and SARS-CoV-2[12–16] and provide insights into the mechanisms of RNA-dependent RNA synthesis[17]. The structures also enable mechanistic studies that can rationalize the molecular processes underlying the antiviral activity of compounds targeting RdRp.

Antiviral drugs often target viral polymerases and function as nucleoside analogs that terminate RNA chain elongation. However, such chain-terminating antivirals are generally not effective against SARS-CoV-2 because coronaviruses carry an exonucleolytic proofreading activity that can remove misincorporated nucleotides from the nascent RNA 3′ end[18–20]. The nucleoside analog remdesivir can circumvent proofreading because its incorporation does not terminate elongation but only stalls RdRp after the addition of three more nucleotides[14,21–24]. Remdesivir was the first FDA-approved drug for the treatment of patients with COVID-19[25–28], but its effectiveness is disputed[29], emphasizing the need to develop new antiviral drugs.

Another promising drug candidate for the treatment of patients with COVID-19 is molnupiravir (or EIDD-2801), which also targets the RdRp of SARS-CoV-2. Molnupiravir is an isopropylester prodrug of the nucleoside analog β-D-N^4-hydroxycytidine (NHC or EIDD-1931)[30,31]. Molnupiravir interferes with the replication of various viruses[30–36], including SARS-CoV-2[37,38]. It inhibits SARS-CoV-2 replication in human lung tissue[39], blocks SARS-CoV-2 transmission in ferrets[40] and reduces SARS-CoV-2 RNA in patients[41]. In contrast to approved drugs such as remdesivir that are administered by infusion, molnupiravir is orally available. Molnupiravir has been tested in phase I trials[42] for safety, tolerability and pharmacokinetics, and phase II/III studies are currently ongoing[41] (NCT04405739, NCT04405570 and NCT04575597). Available data suggest that

molnupiravir acts as a mutagenizing agent that causes an 'error catastrophe' during viral replication[30,37,43]. Indeed, NHC can introduce mutations into viral RNA, as shown for Venezuelan equine encephalitis virus[44]. Also, the sequencing of influenza virus populations has indicated that NHC causes G-to-A and C-to-U transitions in viral RNA[30], and the same transitions have been found for SARS-CoV-2[37].

Despite this progress, a systematic biochemical and structural analysis of molnupiravir- or NHC-induced RNA mutagenesis by viral RNA polymerases is lacking. In this Article, we quantify the effects of molnupiravir or NHC on the RNA synthesis activity of SARS-CoV-2 RdRp using a purified biochemical system and defined synthetic RNAs. Together with structural analysis, we establish the molecular mechanism of molnupiravir-induced RNA mutagenesis. Our results provide detailed insights into the mechanism of action of molnupiravir, which is entirely distinct from that of remdesivir or chain-terminating nucleoside analogs.

Results

SARS-CoV-2 RdRp readily incorporates NHC into RNA. We first tested whether purified SARS-CoV-2 RdRp can use the active form of molnupiravir, NHC triphosphate ('MTP') (Fig. 1a,b), as a substrate for RNA synthesis. We conducted RNA elongation assays in a defined biochemical system using recombinant RdRp and synthetic RNA template–product duplexes (Methods). We used four different RNA duplexes that differed at position +1 of the template strand (Supplementary Table 1), which directs binding of the incoming nucleoside triphosphate (NTP) substrate (Fig. 1c). The RNA product strand contained a fluorescent label at its 5′ end that allowed us to monitor and quantify RNA elongation.

When nucleotides G or A were present at the RNA template position +1, NHC monophosphate ('M') was readily incorporated instead of C or U, respectively (Fig. 1d,e). Time-dependent RNA

[1]Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany. [2]Universität Würzburg, Lehrstuhl für Organische Chemie I, Würzburg, Germany. [3]University Medical Center Göttingen, Department of Cellular Biochemistry, Göttingen, Germany. [4]Max Planck Institute for Biophysical Chemistry, Research Group Structure and Function of Molecular Machines, Göttingen, Germany. [5]These authors contributed equally: Florian Kabinger, Carina Stiller, Jana Schmitzová. ✉e-mail: claudia.hoebartner@uni-wuerzburg.de; patrick.cramer@mpibpc.mpg.de
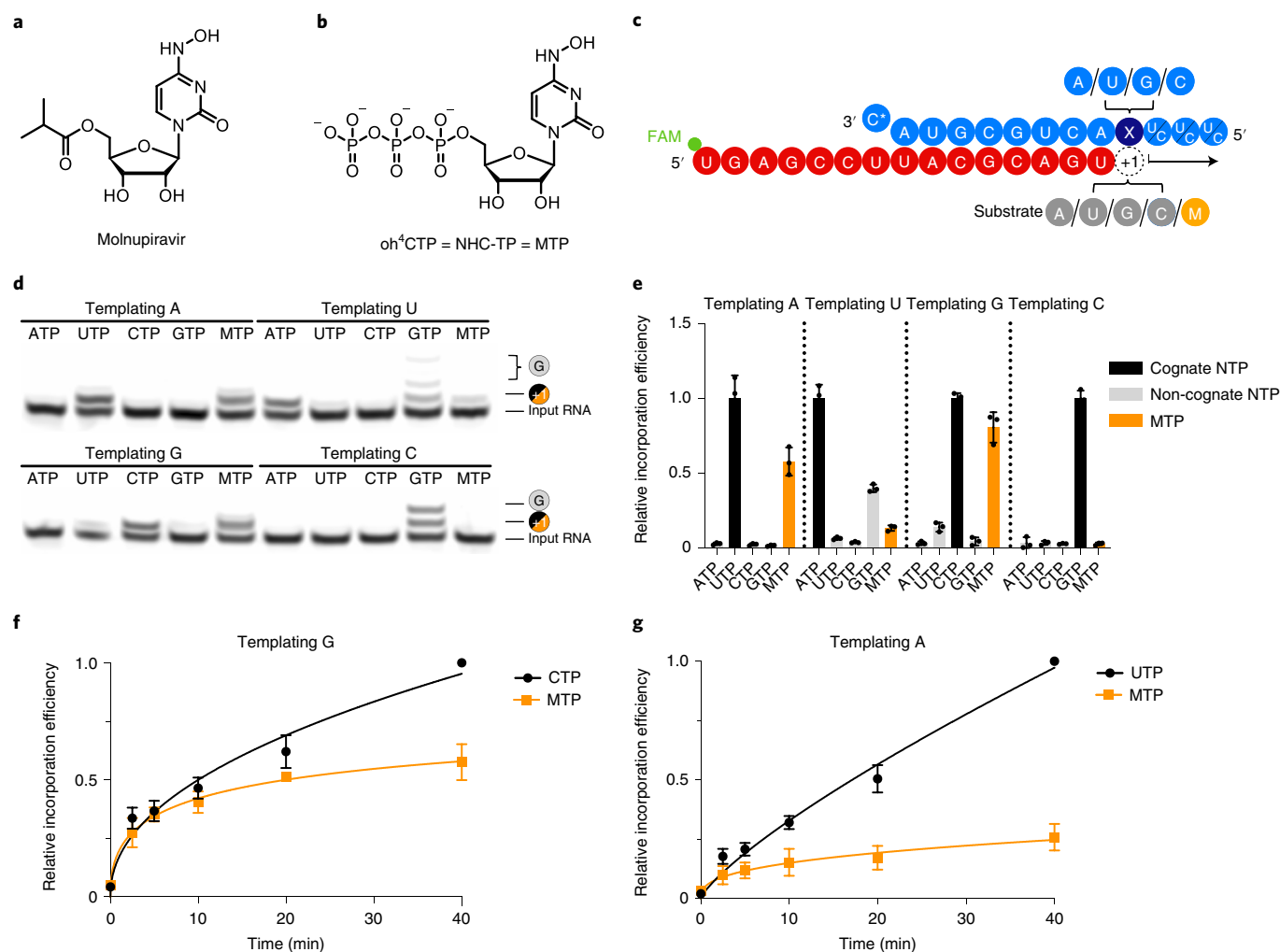
Fig. 1 | RdRp incorporates NHC opposite G and A in the template. a, Chemical structure of molnupiravir. **b**, Chemical structure of NHC triphosphate (MTP). **c**, The RNA template–product duplex. The direction of RNA extension is shown. The color of the depicted circles indicates the experimental design: blue, RNA template strand; dark blue, +1 templating nucleotide; red, RNA product strand; gray, NTP substrate; orange, MTP. The 5′ end of the RNA product contains a FAM fluorescent label. C* at the 3′ end of the template indicates dideoxy-C (ddC). **d**, NHC monophosphate is incorporated into growing RNA instead of C or U when G or A are present in the template +1 position. **e**, Quantification of nucleotide incorporation efficiency relative to the cognate NTP (dark gray) after triplicate measurements. Non-cognate NTPs and MTP are depicted in light gray and orange, respectively. Individual data points and boxes represent mean ± s.d. **f**, Quantification of time-dependent M incorporation opposite a templating G residue after triplicate measurements. Incorporation efficiency is calculated relative to cognate C incorporation. Data are mean ± s.d. **g**, Quantification of time-dependent M incorporation opposite a templating A residue after triplicate measurements. Incorporation efficiency is calculated relative to cognate U incorporation. Data are mean ± s.d. An uncropped gel image for **d** and data behind the graphs in **e**–**g** are available as source data.

elongation experiments showed that M was slightly less efficiently incorporated than the cognate nucleotide C (Fig. 1f). M incorporation opposite A was also observed, but was substantially reduced compared to incorporation of the cognate nucleotide U (Fig. 1g). These results can be explained by base pairing of an incoming MTP substrate with either G or A in the RNA template strand. Consistent with this model, NHC adopts different tautomeric forms[45] that have been predicted to allow for base pairing with either G or A[46].

RdRp does not stall after NHC incorporation. We next tested whether the incorporation of NHC monophosphate (M) into nascent RNA interferes with further RNA extension. We first conducted RNA elongation assays with a scaffold that allowed for RNA extension by four nucleotides (nt) (Fig. 2a). We observed that incorporation of M instead of the cognate C or U did not prevent incorporation of three subsequent nucleotides (Fig. 2b). Furthermore, we

tested RNA extension with a scaffold that allowed for the incorporation of 11 nt (Fig. 2c). In this case too, the RdRp reached the end of the template when uridine triphosphate (UTP) or cytidine triphosphate (CTP) was replaced by MTP, although, again, incorporation of M instead of U was less efficient than incorporation of M instead of C (Fig. 2d). These results demonstrate that M incorporation into nascent RNA does not prevent further RNA elongation. Thus, longer RNA products containing M nucleotides may be synthesized by the RdRp in the presence of MTP. This posed the question of what happens when M-containing RNA is used as a template in a second step of RNA synthesis.

RdRp uses NHC-containing templates to direct RNA mutagenesis. To investigate the templating properties of NHC, we prepared an M-containing RNA by solid-phase synthesis using the phosphoramidite building block M-PA, which we synthesized
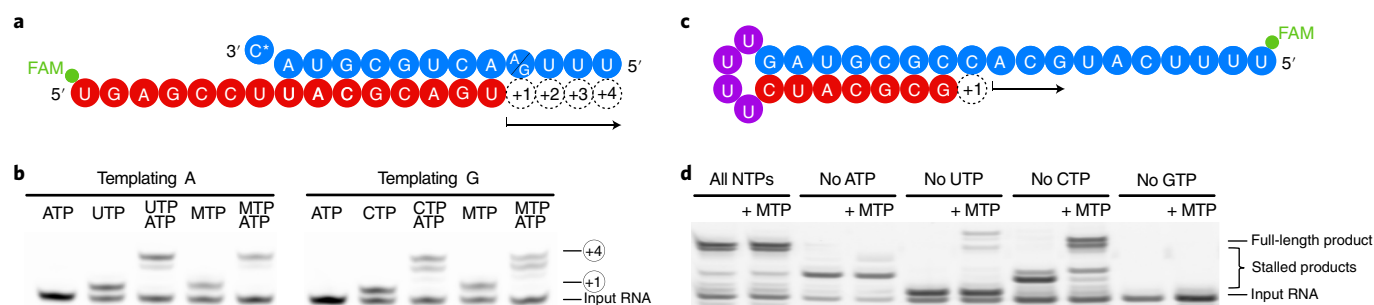
Fig. 2 | NHC incorporation does not stall SARS-CoV-2 RdRp. a, The RNA template–product duplex (as in Fig. 1c) allows for RNA extension by four nucleotides. The direction of RNA extension is indicated. The 5′ end of the RNA product contains a FAM fluorescent label. C* at the 3′ end of the template indicates dideoxy-C (ddC). **b**, RNA elongation to the end of the template in **a** is possible when MTP replaces either CTP or UTP in the presence of adenosine triphosphate (ATP). The experiment was performed once. **c**, The RNA template–product hairpin duplex allows for RNA extension by 11 nucleotides. **d**, RNA elongation stalls at the expected positions when the cognate NTP is withheld from the reaction. Extension to the end of the template is possible when MTP replaces either CTP or UTP in the presence of other substrate NTPs, showing that incorporation of M does not prevent RNA extension. Note that more efficient RNA extension is seen at higher NTP/MTP concentrations, and also for MTP replacing UTP (not shown). The experiment was performed once. Uncropped gel images for **b** and **d** are available as source data.
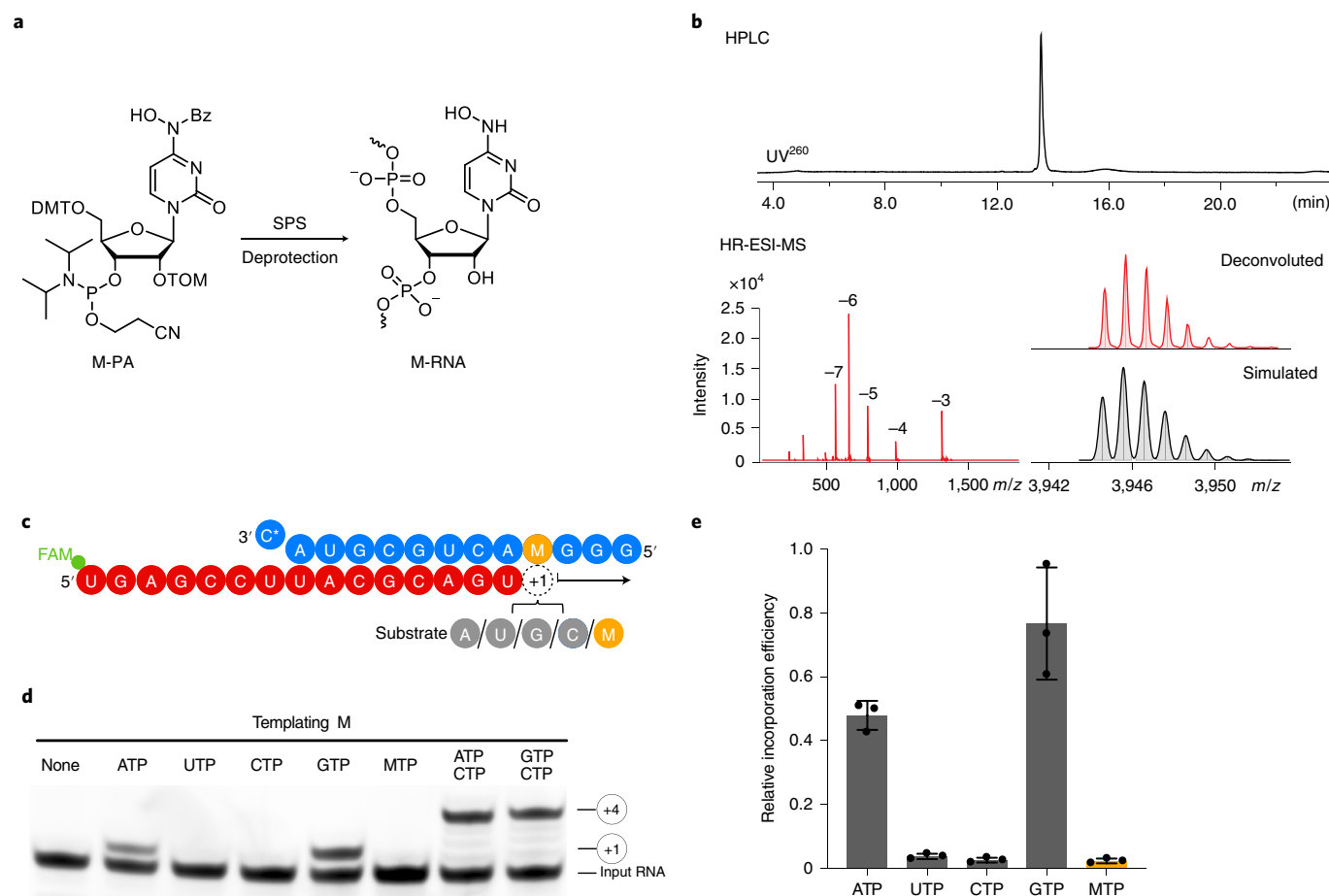

Fig. 3 | NHC can direct incorporation of G and A into RNA. a, Scheme of synthesis of RNA containing NHC monophosphate (M) at a defined position. 5′-O-DMT-2′-O-TOM-protected *N4*-hydroxycytidine phosphoramidite (M-PA) used for solid-phase synthesis of M-containing RNA (M-RNA). **b**, Analysis of M-containing RNA by denaturing HPLC confirms the homogeneity of the synthetic RNA (top). HR-ESI-MS analysis proves the presence of NHC and absence of unmodified RNA (bottom). **c**, The RNA template–product scaffold with M in template position +1, where it is used by the RdRp to direct binding of the incoming NTP substrate. The 5′ end of the RNA product contains a FAM fluorescent label. C* at the 3′ end of the template indicates dideoxy-C (ddC). **d**, When present at position +1 of the template strand, M can direct the incorporation of G or A into nascent RNA, but not C or U. **e**, Quantification of the experiment in **d** after triplicate measurements. Incorporation efficiencies are calculated relative to C incorporation opposite templating G. Individual data points and error bars represent mean ± s.d. An uncropped gel image for **d** and data behind the graph in **e** are available as source data.
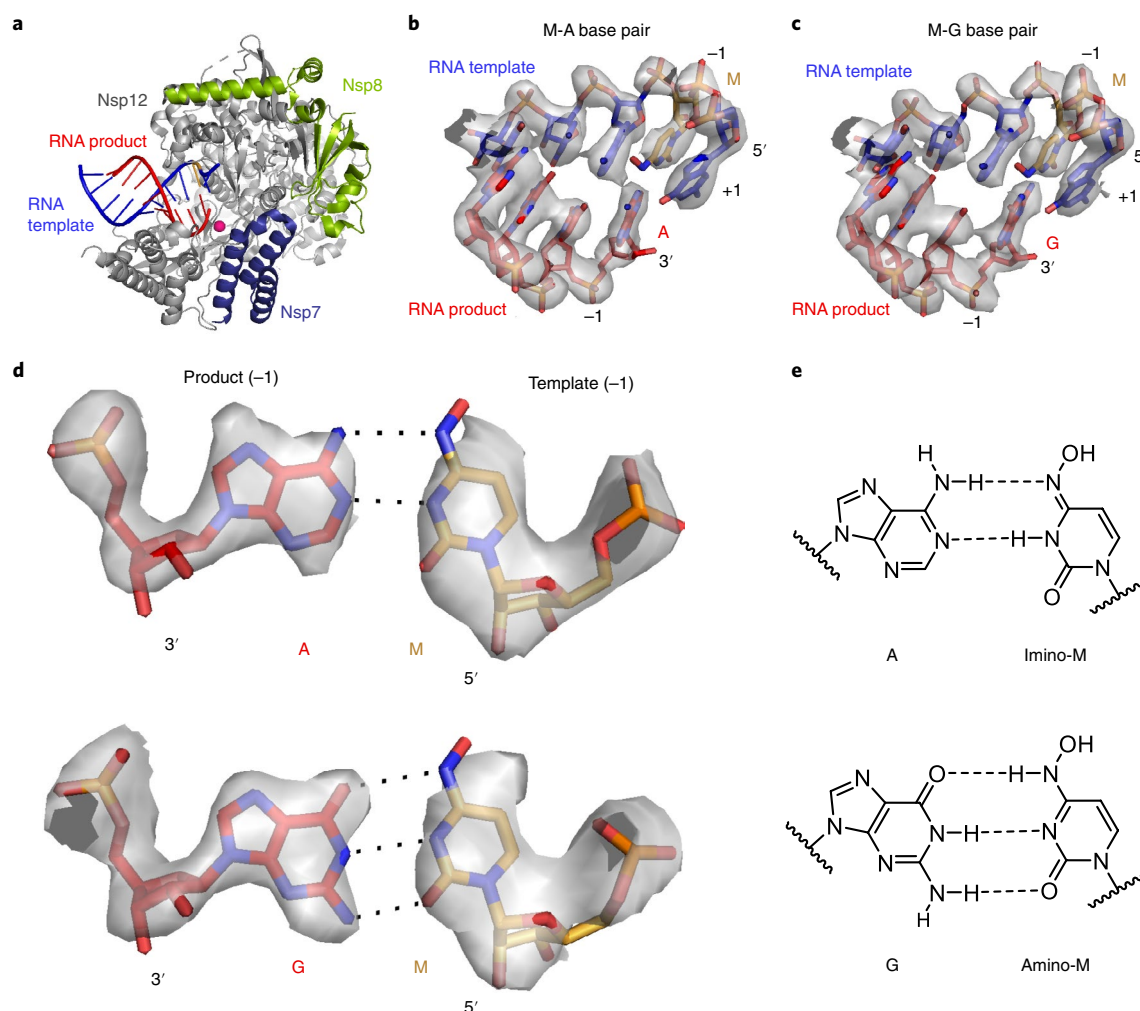
Fig. 4 | Structures of RdRp–RNA product complexes after NHC-induced mutagenesis. a, Overview of RdRp–RNA structure with an M residue (orange) at position −1 in the RNA template strand. RdRp subunits nsp7, nsp8 and nsp12 are in dark blue, green and gray, respectively. The RNA template and product are in blue and red, respectively. The active site is indicated by a magenta sphere. Depicted is the structure containing the M-A base pair.
b, RNA duplex containing the M-A base pair in the RdRp active center. The +1 position (templating nucleotide, NTP substrate site) and the −1 position (post-translocation position of the nascent base pair) are indicated. **c**, RNA duplex containing the M-G base pair in the RdRp active center. **d**, Cryo-EM density for the nascent M-A (top) and M-G (bottom) base pairs in position −1, viewed along the RNA duplex axis in the direction of RNA translocation.
e, M-A (top) and M-G (bottom) base pairing relies on different tautomeric forms of NHC[45], as predicted[46].

using the convertible nucleoside approach from a ribose-protected O^4-chlorophenyluridine (Fig. 3a, Methods, Extended Data Fig. 1 and Supplementary Dataset 1). The presence of M in the obtained RNA as well as RNA purity were confirmed by denaturing HPLC and HR-ESI-MS (Fig. 3b). The M-containing RNA oligo was annealed with a fluorescently labeled product RNA such that the M nucleotide occupied templating position +1 (Fig. 3c and Supplementary Table 1).

Elongation assays showed that the M residue at the +1 position of the template strand directed incorporation of either G or A into nascent RNA, but not C or U (Fig. 3d,e). This can be explained by the formation of M-GTP or M-ATP base pairs in the RdRp active center. Consistent with this, thermal melting experiments with RNA duplexes containing M-G or M-A base pairs located at terminal or internal positions showed similar RNA duplex stabilities that were slightly lower than for duplexes containing a C-G base pair (Extended Data Fig. 2 and Supplementary Table 2). Thus, when the RdRp uses RNA containing NHC monophosphate as a template, either the correct or the incorrect nucleotide is incorporated into the RNA product, and thus mutagenesis will occur.

Structural basis of NHC-induced RNA mutagenesis. The above data indicate that the key aspect of the mutagenesis mechanism is the formation of stable M-G and M-A base pairs in the RdRp active center. To investigate this, we solved two structures of RdRp–RNA complexes that correspond to mutagenesis products after M-templated incorporation of either G or A (Methods). We formed RdRp–RNA complexes containing M in the template strand and either G or A at the 3′ end of the product strand. This was predicted to result in the formation of nascent M-G or M-A base pairs in position −1, which is occupied after successful M-templated nucleotide incorporation and RdRp translocation. We prepared RNA duplex scaffolds with M-containing oligonucleotides (Extended Data Fig. 3), formed RdRp–RNA scaffold complexes, and subjected these to cryo-EM analysis as described in ref. [15].
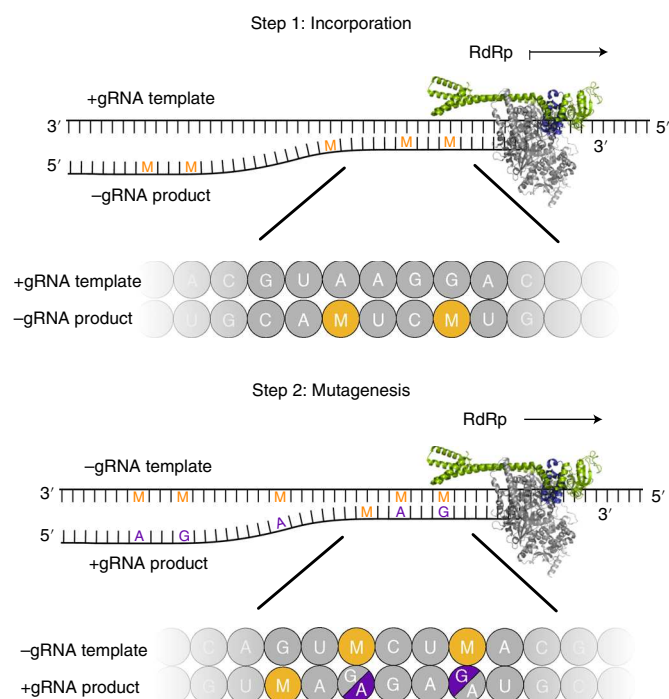
We indeed obtained RdRp–RNA structures that contained either an M-A or an M-G base pair at position −1 (Fig. 4 and Table 1). The structures showed an overall resolution of 3.3 Å and 3.2 Å, respectively, with the active center region resolved at ~2.9 Å in both cases (Extended Data Fig. 4). As expected from the scaffold design, the structures showed the post-translocation state with a

Table 1 | Cryo-EM data collection, refinement and validation statistics

	RdRp–RNA with M-A base pair (EMD-13135) (PDB 7OZU)	RdRp–RNA with M-G base pair (EMD-13138) (PDB 7OZV)
Data collection and processing		
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e−/Å²)	59.6	59.6
Defocus range (μm)	0.4–2.2	0.4–2.7
Pixel size (Å)	0.834	0.834
Symmetry imposed	C1	C1
Initial particle images (no.)	2,528,775	2,183,996
Final particle images (no.)	373,938	851,168
Map resolution (Å)	3.3	3.2
FSC threshold	0.143	0.143
Map resolution range (Å)	2.7–4.5	2.8–3.9
Refinement		
Initial model used (PDB code)	7B3D	7B3D
Model resolution (Å)	3.3	3.3
FSC threshold	0.5	0.5
Map sharpening B factor (Å²)	−142	−159
Model composition		
Nonhydrogen atoms	8,405	8,406
Protein residues	991	991
Ligands	3	3
B factors (Å²)		
Protein	66.68	61.32
Ligand	68.17	60.27
R.m.s. deviations		
Bond lengths (Å)	0.004	0.005
Bond angles (°)	0.587	0.581
Validation		
MolProbity score	1.53	1.57
Clashscore	6.28	8.42
Poor rotamers (%)	0.11	0.11
Ramachandran plot		
Favored (%)	97.34	97.45
Allowed (%)	2.66	2.55
Disallowed (%)	0.00	0.00

Fig. 5 | Two-step model of molnupiravir-induced RNA mutagenesis. In the presence of NTPs and MTP, M nucleotides can be incorporated by SARS-CoV-2 RdRp instead of C or U into the negative-strand genomic (−gRNA) or subgenomic RNA (−sgRNA) during copying of the positive-strand genomic RNA template (+gRNA). The obtained M-containing negative-strand RNAs can then be used as a template for the production of mutagenized +gRNA and positive-strand subgenomic mRNA (+sgmRNA). These RNA products are predicted to be mutated and not to support formation of functional viruses. RNA of random sequence is shown, with M and mutated residues indicated as orange and violet letters, respectively.

tautomeric forms of NHC[45] (Fig. 4e). However, the observed hydrogen-bonding geometries were not optimal, possibly explaining our biochemical observations that suggest that M can mimic C well, and mimic U to some extent, but neither mimicry is perfect (Figs. 1 and 2). These results represent the first direct visualization of NHC in a polymerase enzyme and show that stable M-G and M-A base pairs can be formed and accommodated in the RdRp active center, readily explaining our biochemical results.

Discussion

Our systematic biochemical analysis suggests a two-step model for the mechanism of molnupiravir-induced coronavirus RNA mutagenesis (Fig. 5). When the molnupiravir prodrug enters the cell, it is converted to NHC triphosphate (MTP), which can be used by the RdRp of SARS-CoV-2 as a substrate instead of CTP or UTP. Therefore, in a first step, the RdRp is predicted to frequently incorporate M instead of C or U when it uses the positive-strand genomic RNA (+gRNA) as a template to synthesize negative-strand genomic (−gRNA) and subgenomic RNA (−sgRNA). In a second step, the resulting M-containing RNA can be used as a template for the synthesis of +gRNA or positive-strand subgenomic mRNA (+sgmRNA). The presence of M in the −gRNA then leads to mutations in the positive-strand RNA products, which do not support formation of intact new viruses, as predicted by the 'error catastrophe' model[30,37,43].

Our structural studies confirm the key aspect of this model, namely the formation of M-G and M-A base pairs in the RdRp active center. When we interpreted our cryo-EM densities, we

free NTP-binding site at position +1 (Fig. 4b,c). Comparison of the two structures with each other and with our original RdRp–RNA structure[15] and with remdesivir-containing RdRp–RNA structures[23] did not reveal major differences, neither in the protein subunits nor in the nucleic acids, except that the protruding, second turn of RNA and the sliding poles of the nsp8 subunits were poorly ordered and not retained in the final model.

The cryo-EM densities at the −1 position of the structures could readily be interpreted by modeling M-A or M-G base pairs (Fig. 4b,c). The densities were so detailed that we could clearly distinguish G and A bases (Fig. 4d). The densities were also consistent with the proposed base pairing[46] that is enabled by different

assumed that one stable tautomer predominates in each case; that is, the amino-M tautomer forms a base pair with G and the imino-M tautomer forms a base pair with A. These base pairs do not impair RdRp progression, explaining the observed mutagenic pattern seen in in vivo studies of NHC-exposed coronaviruses[37]. This antiviral mechanism is conceptually similar to the recently suggested mutagenesis mode of action of favipiravir[47,48], but is entirely distinct from that of remdesivir, which impairs RdRp progression[23]. However, like remdesivir, molnupiravir escapes viral RNA proofreading because M incorporation and M-directed misincorporation are apparently not recognized by the viral exonuclease[18,19]. Such proofreading escape may also be due to the stability of the M-G and M-A base pairs, which are predicted not to induce or favor backtracking of RdRp, which is probably required for exposing the RNA 3′ end to the proofreading exonuclease[49,50].

Finally, the two-step model can explain how molnupiravir or NHC monophosphate leads to RNA mutagenesis by polymerases of other viruses. For influenza, a possible two-step mutagenesis mechanism had been inferred from sequencing a molnupiravir-experienced virus population[30]. Also consistent with the model, MTP does not inhibit RNA synthesis by hepatitis C polymerase[36] or the RdRp of respiratory syncytial virus[32]. Furthermore, the reverse transcriptase of human immunodeficiency virus can incorporate G or A opposite the NHC located in the template[51]. The two-step mutagenesis model rests on the base-pairing properties of NHC that we have structurally defined here and can explain why molnupiravir and NHC exhibit broad-spectrum antiviral activity against a wide variety of RNA viruses.

Beside the high antiviral potency of NHC, potential risks have to be considered. Host RNA polymerases may use MTP as substrate, and indeed the mitochondrial DNA-dependent RNA polymerase can use EIDD-1931 and incorporate NHC monophosphate into RNA in vitro[52]. In addition, possible mutagenic effects of NHC were recently described in mammalian cells[53]. Therefore, it will be important to characterize the effects of molnupiravir and NHC on cellular polymerase function in future studies.

After the submission of our manuscript, independently derived biochemical data about the molnupiravir mechanism of action became available[54] that are consistent with our findings. On the basis of steady-state kinetic measurements, the authors suggest the formation of M-A and M-G base pairs, a longstanding hypothesis that could now be confirmed by the structural data presented in our study.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-021-00651-0.

References

1. Dolgin, E. The race for antiviral drugs to beat COVID—and the next pandemic. *Nature* **592**, 340–343 (2021).
2. Hilgenfeld, R. & Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antivir. Res.* **100**, 286–295 (2013).
3. Snijder, E. J., Decroly, E. & Ziebuhr, J. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv. Virus Res.* **96**, 59–126 (2016).
4. Posthuma, C. C., Te Velthuis, A. J. W. & Snijder, E. J. Nidovirus RNA polymerases: complex enzymes handling exceptional RNA genomes. *Virus Res.* **234**, 58–73 (2017).
5. Romano, M., Ruggiero, A., Squeglia, F., Maga, G. & Berisio, R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* **9**, 1267 (2020).
6. Jiang, Y., Yin, W. & Xu, H. E. RNA-dependent RNA polymerase: structure, mechanism and drug discovery for COVID-19. *Biochem. Biophys. Res. Commun.* **538**, 47–53 (2021).
7. Cannalire, R., Cerchia, C., Beccari, A. R., Di Leva, F. S. & Summa, V. Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: state of the art and future opportunities. *J. Med. Chem.* https://doi.org/10.1021/acs.jmedchem.0c01140 (2020).
8. Vicenti, I., Zazzi, M. & Saladini, F. SARS-CoV-2 RNA-dependent RNA polymerase as a therapeutic target for COVID-19. *Experrt Opin. Ther. Pat.* **31**, 325–337 (2021).
9. Tian, L. et al. RNA-dependent RNA polymerase (RdRp) inhibitors: the current landscape and repurposing for the COVID-19 pandemic. *Eur. J. Med. Chem.* **213**, 113201 (2021).
10. Subissi, L. et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl Acad. Sci. USA* **111**, E3900–E3909 (2014).
11. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 2342 (2019).
12. Gao, Y. et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779–782 (2020).
13. Yin, W. et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499–1504 (2020).
14. Wang, Q. et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* **182**, 417–428 (2020).
15. Hillen, H. S. et al. Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154–156 (2020).
16. Jochheim, F. A. et al. Dimeric form of SARS-CoV-2 polymerase. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.23.436644 (2021).
17. Hillen, H. S. Structure and function of SARS-CoV-2 polymerase. *Curr. Opin. Virol.* **48**, 82–90 (2021).
18. Bouvet, M. et al. RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl Acad. Sci. USA* **109**, 9372–9377 (2012).
19. Robson, F. et al. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell* **79**, 710–727 (2020).
20. Ferron, F. et al. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl Acad. Sci. USA* **115**, E162–E171 (2018).
21. Gordon, C. J., Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Gotte, M. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J. Biol. Chem.* **295**, 4773–4779 (2020).
22. Gordon, C. J. et al. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* **295**, 6785–6797 (2020).
23. Kokic, G. et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **12**, 279 (2021).
24. Bravo, J. P. K., Dangerfield, T. L., Taylor, D. W. & Johnson, K. A. Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. *Mol. Cell* **81**, 1548–1552 (2021).
25. Beigel, J. H. et al. Remdesivir for the treatment of Covid-19—preliminary report. *N. Engl. J. Med.* **383**, 1813–1826 (2020).
26. Grein, J. et al. Compassionate use of remdesivir for patients with severe Covid-19. *N. Engl. J. Med.* **382**, 2327–2336 (2020).
27. Teoh, S. L., Lim, Y. H., Lai, N. M. & Lee, S. W. H. Directly acting antivirals for COVID-19: where do we stand? *Front. Microbiol.* **11**, 1857 (2020).
28. WHO Solidarity Trial Consortium. Repurposed antiviral drugs for COVID-19; interim WHO SOLIDARITY trial results. *N. Engl. J. Med.* **384**, 497–511 (2021).
29. WHO Solidarity Trial Consortium Repurposed antiviral drugs for Covid-19—interim WHO solidarity trial results. *N. Engl. J. Med.* **384**, 497–511 (2021).
30. Toots, M. et al. Quantitative efficacy paradigms of the influenza clinical drug candidate EIDD-2801 in the ferret model. *Transl. Res.* **218**, 16–28 (2020).
31. Painter, G. R. et al. The prophylactic and therapeutic activity of a broadly active ribonucleoside analog in a murine model of intranasal Venezuelan equine encephalitis virus infection. *Antivir. Res.* **171**, 104597 (2019).
32. Yoon, J. J. et al. Orally efficacious broad-spectrum ribonucleoside analog inhibitor of influenza and respiratory syncytial viruses. *Antimicrob. Agents Chemother.* **62**, e00766-18 (2018).
33. Ehteshami, M. et al. Characterization of β-D-N^4-hydroxycytidine as a novel inhibitor of Chikungunya virus. *Antimicrob. Agents Chemother.* **61**, e02395-16 (2017).
34. Reynard, O. et al. Identification of a new ribonucleoside inhibitor of Ebola virus replication. *Viruses* **7**, 6233–6240 (2015).
35. Costantini, V. P. et al. Antiviral activity of nucleoside analogues against norovirus. *Antivir. Ther.* **17**, 981–991 (2012).

36. Stuyver, L. J. et al. Ribonucleoside analogue that blocks replication of bovine viral diarrhea and hepatitis C viruses in culture. *Antimicrob. Agents Chemother.* **47**, 244–254 (2003).

37. Sheahan, T. P. et al. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci. Transl. Med.* **12**, eabb5883 (2020).

38. Agostini, M. L. et al. Small-molecule antiviral β-D-N^4-hydroxycytidine inhibits a proofreading-intact coronavirus with a high genetic barrier to resistance. *J. Virol.* **93**, 24 (2019).

39. Wahl, A. et al. SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature* **591**, 451–457 (2021).

40. Cox, R. M., Wolf, J. D. & Plemper, R. K. Therapeutically administered ribonucleoside analogue MK-4482/EIDD-2801 blocks SARS-CoV-2 transmission in ferrets. *Nat. Microbiol.* **6**, 11–18 (2021).

41. MERCK. Merck and Ridgeback Biotherapeutics provide update on progress of clinical development program for molnupiravir, an investigational oral therapeutic for the treatment of mild-to-moderate COVID-19. Press release, https://www.merck.com/news/merck-and-ridgeback-biotherapeutics-provide-update-on-progress-of-clinical-development-program-for-molnupiravir-an-investigational-oral-therapeutic-for-the-treatment-of-mild-to-moderate-covid-19/ (Merck & Ridgeback Biotherapeutics, 15 April 2021).

42. Painter, W. P. et al. Human safety, tolerability and pharmacokinetics of molnupiravir, a novel broad-spectrum oral antiviral agent with activity against SARS-CoV-2. *Antimicrob. Agents Chemother.* **65**, 5 (2021).

43. Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl Acad. Sci. USA* **98**, 6895–6900 (2001).

44. Urakova, N. et al. β-D-N^4-hydroxycytidine is a potent anti-alphavirus compound that induces a high level of mutations in the viral genome. *J. Virol.* **92**, e0165-17 (2018).

45. Les, A., Adamowicz, L. & Rode, W. Structure and conformation of N^4-hydroxycytosine and N^4-hydroxy-5-fluorocytosine. A theoretical ab initio study. *Biochim. Biophys. Acta* **1173**, 39–48 (1993).

46. Jena, N. R. Role of different tautomers in the base-pairing abilities of some of the vital antiviral drugs used against COVID-19. *Phys. Chem. Chem. Phys.* **22**, 28115–28122 (2020).

47. Shannon, A. et al. Rapid incorporation of favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat. Commun.* **11**, 4682 (2020).

48. Naydenova, K. et al. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc. Natl Acad. Sci. USA* **118**, e2021946118 (2021).

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

50. Malone, B. et al. Structural basis for backtracking by the SARS-CoV-2 replication–transcription complex. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.13.435256 (2021).

51. Suzuki, T., Moriyama, K., Otsuka, C., Loakes, D. & Negishi, K. Template properties of mutagenic cytosine analogues in reverse transcription. *Nucleic Acids Res.* **34**, 6438–6449 (2006).

52. Sticher, Z. M. et al. Analysis of the potential for N^4-hydroxycytidine to inhibit mitochondrial replication and function. *Antimicrob. Agents Chemother.* **64**, e01719-19 (2020).

53. Zhou, S. et al. β-D-N^4-hydroxycytidine (NHC) inhibits SARS-CoV-2 through lethal mutagenesis but is also mutagenic to mammalian cells. *J. Infect. Dis.* **224**, 3 (2021).

54. Gordon, C. J., Tchesnokov, E. P., Schinazi, R. F. & Gotte, M. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J. Biol. Chem.* **297**, 100770 (2021).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein preparation. Preparation of SARS-CoV-2 RdRp, composed of nsp12, nsp7 and two copies of the nsp8 subunits, was carried out as described in ref. [15], with some modifications. Nsp12 protein was expressed from pFastBac vector 438C (Addgene 154759) in Hi5 insect cells. Production of the bacmid, V0 and V1 viruses was carried out as described in ref. [15]. Sixty hours after transfection with the corresponding V1 virus, the cells producing nsp12 with a 6xHis-MBP N-terminal tag were collected by centrifugation (3,000 r.p.m., 10 min at 4 °C) and lysed by sonication in lysis buffer A1 (400 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 30 mM imidazole pH 8.0, 5 mM β-mercaptoethanol, 0.284 µg ml^{-1} leupeptin, 1.37 µg ml^{-1} pepstatin, 0.17 mg ml^{-1} PMSF and 0.33 mg ml^{-1} benzamidine). The lysate was clarified by centrifugation at 74,766g for 60 min and ultracentrifugation at 100,000g at 4 °C for 60 min, followed by filtration through a 0.45-µm filter (Amicon Ultra centrifugal filter, Merck). The protein was bound to HisTrap HP prepacked columns (GE Healthcare) pre-equilibrated in lysis buffer, washed with high salt buffer B2 (1 M NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 5 mM β-mercaptoethanol) followed by washing with buffer A1, and eluted with a gradient of 0–80% A1-B1 buffers over 30 column volumes (CV) (B1 buffer: 200 mM NaCl, 25 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 400 mM imidazole pH 8.0, 3 mM MgCl$_2$ and 5 mM β-mercaptoethanol). Fractions containing nsp12 were pooled and the tag was cleaved with His-tagged tobacco etch virus (TEV) protease overnight during dialysis against buffer D (200 mM NaCl, 25 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 5 mM β-mercaptoethanol). The protein solution was applied to a HisTrap HP prepacked column to remove uncleaved protein, tag and TEV protease. The flow-through containing nsp12 was further purified by ion exchange chromatography using Hi TRAP Q HP and SP HP prepacked columns (GE Healthcare) equilibrated with buffer D. The unbound protein was concentrated and further purified via size exclusion chromatography on a HiLoad S200 16/60 column (GE Healthcare) in buffer A2 (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 1 mM MgCl$_2$, 1 mM Tris(2-carboxyethyl)phosphine (TCEP)). The peak contained monomeric fractions of nsp12 that were pooled, concentrated to 55 µM, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

Nsp7 and nsp8 were prepared as described in ref. [15]. Briefly, both proteins were expressed in *Escherichia coli* BL21(DE3) RIL from the pET-derived vector 14-B (a gift from S. Gradia; Addgene 48308) in LB medium individually. Cells were grown to an optical density at 600 nm of 0.4 at 30 °C and protein expression was induced with 0.5 mM IPTG at 18 °C for 16 h. After collection, cells were resuspended in lysis buffer A1 (400 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 30 mM imidazole pH 8.0, 5 mM β-mercaptoethanol, 0.284 µg ml^{-1} leupeptin, 1.37 µg ml^{-1} pepstatin, 0.17 mg ml^{-1} PMSF and 0.33 mg ml^{-1} benzamidine). Nsp8 and nsp7 were purified separately using the same purification procedure. The cells were lysed using a French press and three cycles. Lysates were subsequently cleared by centrifugation (74,766g, 4 °C, 30 min). The supernatant was applied to a HisTrap HP column (GE Healthcare), pre-equilibrated in lysis buffer. The column was washed with high-salt buffer B2 (1 M NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 5 mM β-mercaptoethanol) and with buffer A1 (400 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 30 mM imidazole pH 8.0 and 5 mM β-mercaptoethanol). The protein sample was then eluted using a gradient of 0–80% B1 over 30 CV (B1 buffer: 300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 400 mM imidazole pH 8.0, 3 mM MgCl$_2$, 5 mM β-mercaptoethanol). The protein sample was dialysed in buffer D (200 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (vol/vol) glycerol, 5 mM β-mercaptoethanol) in the presence of His-tagged TEV protease at 4 °C. The dialysed sample was subsequently applied to a HisTrap HP column (GE Healthcare), pre-equilibrated in A1 buffer. The flow-through that contained the untagged protein of interest was applied to an ion exchange chromatography HiTrap Q column (GE Healthcare). The unbound protein sample, containing nsp8 or nsp7, was concentrated using a 10,000 molecular weight cutoff Amicon Ultra Centrifugal Filter (Merck) and applied to a HiLoad S200 16/600 system (GE Healthcare) equilibrated in buffer A2 (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 5% (vol/vol) glycerol, 1 mM TCEP). Peak fractions containing proteins were pooled, concentrated to 430 µM (nsp7) and 400 µM (nsp8), aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

RNA extension assays. RNA scaffolds were designed according to published SARS-CoV-2 RNA extension assays[15,23]. All assays except those shown in Fig. 3c were performed with RNA duplexes composed of short RNA template and product strands. The 5′ end of the RNA product strands contained a 6-carboxyfluorescein (FAM) fluorescent label that allowed us to monitor RNA extension. The 3′ end of the RNA template strands harbored a mismatched dideoxy-C to prevent extension. The RNA extension assay shown in Fig. 3c was performed with a single-stranded looped RNA scaffold that contained a FAM fluorescent label at the 5′ end. For all assays that required synthesis of M-containing oligonucleotides, RNA strands had to be kept short for technical reasons.

All unmodified and 5′ FAM-labeled RNA oligonucleotides (Supplementary Table 1) were purchased from Integrated DNA Technologies (IDT). MTP was purchased from MedChemExpress and NTPs from Thermo Scientific. The assay was performed as described in ref. [15], except for the following changes. The final concentrations of nsp12, nsp8, nsp7 and RNA were 3 µM, 9 µM, 9 µM and 3 µM, respectively, except that 1 µM of RNA was used in the assays in Figs. 1f,g and 2d. The concentration of NTPs was 37.5 µM except for the assays shown in Figs. 1f,g and 2d, where 4 µM was used. RNA in annealing buffer (50 mM NaCl, 10 mM Na-HEPES pH 7.5) was annealed by heating it to 75 °C for 1 min and gradually cooling to 4 °C. Annealed RNA and pre-mixed RdRp were incubated in reaction buffer (100 mM NaCl, 20 mM Na-HEPES pH 7.5, 5% (vol/vol) glycerol, 10 mM MgCl$_2$, 5 mM β-mercaptoethanol) for 10 min at 30 °C and the reactions were started by addition of NTPs. After 20 min of incubation at 30 °C, the reactions were stopped with 2× stop buffer (7 M urea, 50 mM EDTA, 1× TBE buffer). RNA products were resolved on 20% denaturing polyacrylamide-urea gels in 0.5× TBE running buffer and visualized with a Typhoon 95000 FLA Imager (GE Healthcare Life Sciences). Bands were quantified with Image Lab (Bio-Rad). The obtained data were plotted and analyzed with GraphPad Prism version 9.

Preparation and analysis of NHC-containing RNA oligonucleotides. 5′-*O*-DMT-2′-*O*-TOM-*O*^{4}-chlorophenyluridine was prepared as described in ref. [55] and converted to the corresponding *N*^{4}-hydroxy-*N*^{4}-benzoylcytidine 3′-(2-cyanoethyl) diisopropyl phosphoramidite (M-PA) in three steps. Details of the synthetic procedures are provided in the following, and NMR spectra of the isolated compounds are provided in the Supplementary Information. RNA oligonucleotides were prepared by solid-phase synthesis on a controlled pore glass support (0.6-µmol scale) using 2′-*O*-TOM-protected ribonucleoside phosphoramidites (70 mM in CH$_3$CN) and ethylthiotetrazol (ETT, 250 mM in CH$_3$CN) as activator, with 4-min coupling time, as previously described[23,55]. The oligonucleotides were deprotected with 25% NH$_4$OH/EtOH 3/1 at 55 °C for 6 h, followed by 1 M TBAF in THF for 12 h, and carefully purified by denaturing polyacrylamide gel electrophoresis to remove a minor fraction, in which NHC was converted to U during deprotection.

The purity and identity of the RNA oligonucleotides were analyzed by anion-exchange HPLC (Dionex DNAPac PA200, 2 × 250 mm, at 60 °C; solvent A: 25 mM Tris-HCl (pH 8.0), 6 M urea; solvent B: 25 mM Tris-HCl (pH 8.0), 6 M urea, 0.5 M NaClO$_4$; gradient: linear, 0–48% solvent B, 4% solvent B per 1 CV) and HR-ESI-MS (Bruker micrOTOF-Q III, negative ion mode, direct injection).

Thermal melting experiments were performed in 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl, at an RNA duplex concentration of 20 µM. Absorbance versus temperature profiles were recorded at 260 nm on a Varian Cary 100 spectrometer equipped with a Peltier temperature controller, at a heating rate of 0.5 °C min^{-1}, for two heating and two cooling ramps between 10 and 90 °C. Melting curves were normalized to the absorbance at 95 °C, fitted to a two-state transition model with linearly sloping lower and upper baselines, and the melting temperatures were determined at the inflection point of the curves.

Synthesis and characterization of M-PA. All reactions were performed under an inert nitrogen atmosphere with dry solvents (CH$_2$Cl$_2$, CH$_3$CN). For workup and purification, distilled solvents (technical quality) were used. Column chromatography was performed on silica gel (Kieselgel 60, Merck) with a particle size of 0.040–0.063. Thin-layer chromatography was performed on Alugram aluminium sheets (Machery-Nagel, UV visualization, 254 nm). NMR spectra were recorded using Bruker Avance III (400 MHz) spectrometers. Chemical shifts (δ) are given in ppm, relative to the residual solvent signals as internal standards (CDCl$_3$: ^{1}H = 7.26, ^{13}C = 77.16). Data are reported as s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. Coupling constants (J) are given in Hz. HR-ESI-MS results were recorded on a Bruker micrOTOF-Q III spectrometer. The detected mass-to-charge ratio (m/z) is given, as well as the calculated monoisotopic mass (Supplementary Information, Extended Dataset 1 and Extended Data Fig. 1).

To synthesize compound **1** (5′-*O*-(4,4′-dimethoxytrityl)-*N*^{4}-hydroxy-2′-*O*-(triisopropyl-silyloxymethylcytidine), 5′-*O*-DMT-2′-*O*-TOM-*O*^{4}-chlorophenyluridine (400 mg, 474 µmol, 1.0 equiv.) was dissolved in anhydrous CH$_3$CN (4 ml) under a nitrogen atmosphere. 4-Dimethylaminopyridine (DMAP; 174 mg,1.42 mmol, 3.0 equiv.) and NEt$_3$ (331 µl, 2.37 mmol, 5.0 equiv.) were added, followed by hydroxyl amine hydrochloride (165 mg, 2.37 mmol, 5.0 equiv.). After stirring for 21 h at ambient temperature, the reaction mixture was diluted with CH$_2$Cl$_2$ and washed with saturated aqueous NaHCO$_3$ (2×). The organic phase was dried over Na$_2$SO$_4$ and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (*n*-hexane:EtOAc + 1% NEt$_3$ 1:2 to 2:1) to yield the product (compound **1**, 230 mg, 308 µmol, 64%) as a colorless foam.

To synthesize compound **2** (*N*^{4}-benzoyl-5′-*O*-(4,4′-dimethoxytrityl)-*N*^{4}-hydroxy-2′-*O*-(tri-isopropylsilyloxy)methylcytidine), a solution of compound **1** (200 mg, 267 µmol, 1.0 equiv.) in anhydrous CH$_2$Cl$_2$ (4 ml) was treated with DMAP (65.3 mg, 535 µmol, 2.0 equiv.) and NEt$_3$ (149 µl, 1.07 mmol, 4.0 equiv.) under a nitrogen atmosphere. Benzoic anhydride (59.9 mg, 265 µmol, 0.99 equiv.) was added in three portions within 3 h and the resulting reaction mixture was stirred for one more hour at ambient temperature. Volatiles were removed under

reduced pressure. The crude residue was purified by column chromatography (*n*-hexane:EtOAc + 1% NEt$_3$ 1:2) to yield compound **2** (183 mg, 215 μmol, 80%) as a colorless foam.

To synthesize compound **3**, (*N*^{4}-benzoyl-5′-*O*-(4,4′-dimethoxytrityl)-*N*^{4}-hydroxy-2′-*O*-(tri-isopropylsilyloxy)methylcytidine 3′-cyanoethyl-*N,N*-diisopropyl phosphoramidite), compound **2** (150 mg, 176 μmol, 1.0 equiv.) was dissolved in anhydrous CH$_2$Cl$_2$ (2 ml) and cooled to 4 °C. 2-Cyanoethyl *N,N,N′,N′*-tetraisopropyl phosphoramidite (67.3 μl, 212 μmol, 1.2 equiv.) and 4,5-dicyanoimidazol (23 mg, 194 μmol, 1.1 equiv.) were added in two portions within 1 h. After one additional hour at 4 °C, the reaction mixture was allowed to warm to ambient temperature, then stirred for one more hour. The solvent was evaporated and the crude residue was purified by column chromatography (*n*-hexane:EtOAc + 1% NEt$_3$ 2:1 to 1:1) to yield compound **3** (75.0 mg, 71.3 μmol, 40%) as a colorless foam.

Cryo-EM sample preparation and data collection. RNA scaffolds for structural studies were prepared by annealing two RNA oligonucleotides, because the length of the NHC-containing RNA was limited for technical reasons. The first RNA oligo was designed to form a template–product hybrid with a hairpin leaving a 10 nt product overhang at the 3′ end. A second oligo containing 8 nt complementary to the overhang was annealed to obtain an RNA template–product scaffold with a single nick at position −10 in the template RNA (Extended Data Fig. 3 and Table 1). The second RNA oligo had NHC incorporated at position −1 and a short 5′ overhang of three G nucleotides. RNA scaffolds for RdRp–RNA complex formation were prepared by mixing equimolar amounts of two RNA strands in annealing buffer (10 mM Na-HEPES pH 7.4, 50 mM NaCl) and heating to 75 °C, followed by stepwise cooling to 4 °C. RdRp–RNA complexes were formed by mixing purified nsp12 (1.25 nmol) with an equimolar amount of annealed RNA scaffold and threefold molar excess each of nsp8 and nsp7. After 10 min of incubation at 30 °C, the mixture was applied to a Superdex 200 Increase 3.2/300 size exclusion chromatography column (GE Healthcare), equilibrated in complex buffer (20 mM Na-HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl$_2$, 1 mM TCEP) at 4 °C. Peak fractions corresponding to RdRp–RNA complex were pooled and diluted to 1.5 mg ml^{-1}. An additional 0.2 nmol of the annealed RNA scaffold were spiked into each sample before grid preparation. A 3 μl volume of the concentrated RdRp–RNA complex was mixed with 0.5 μl of octyl β-D-glucopyranoside (0.003% final concentration) and applied to freshly glow-discharged R 2/1 holey carbon grids (Quantifoil). The grids were blotted for 7 s with blot force 5 using a Vitrobot Mark IV instrument (Thermo Fischer Scientific) at 4 °C and 95% humidity and plunge-frozen in liquid ethane.

Cryo-EM data were collected with SerialEM[56] on a Titan Krios transmission electron microscope (Thermo Fischer Scientific) operated at 300 keV. Inelastically scattered electrons were filtered out with a GIF Quantum energy filter (Gatan) using a slit width of 20 eV. Images were acquired using a K3 direct electron detector in counting mode (non-super resolution) at a nominal magnification of ×105,000, resulting in a calibrated pixel size of 0.834 Å px^{-1}. Images were exposed for a total of 2.0 s with a dose rate of 20.5 e$^-$ px^{-1} s^{-1}, resulting in a total dose of 59.6 e$^-$/Å^2 that was fractionated into 80 frames. Our previous cryo-EM analysis of SARS-CoV-2 RdRp–RNA complexes showed strong preferred particle orientation in ice[15]. To obtain more particle orientations, all data were collected with 30° stage tilt. Motion correction, CTF estimation and particle picking and extraction were performed using Warp[57]. A total of 11,060 and 10,230 videos were collected for M-A- and M-G-containing structures, respectively.

Cryo-EM data processing and structural modeling. For the RdRp–RNA complex containing the M-A base pair, 2.5 million particles were extracted using Warp[57] 1.0.9. Particles were imported to cryoSPARC[58] 2.15 and subjected to two-dimensional (2D) classification. 2D classes representing contamination or broken particles were selected and used for calculating two ab initio structures. All particles were then used for supervised 3D classification against five references, where four were originating from the ab initio reconstruction of contamination and broken particles and one was a previous RdRp–RNA complex structure (EMD-11995[23]). The class containing RNA-bound RdRp yielded ~800,000 particles, which were subjected to homogeneous 3D refinement. The refined particles were then exported to RELION 3.1[59] and focus-refined in 3D with an initial local angular sampling of 3.7° and a mask around RdRp that omitted the nsp8 sliding poles and the upstream, second RNA turn. To improve the quality of the density in the active site, particles were 3D-classified without image alignment (*T* = 4, four classes) using the same mask. The best class was focus-refined to an overall resolution of 3.3 Å. Local resolution was estimated with Relion 3.1 using a kernel size of 10 Å. For the RdRp–RNA complex containing the M-G base pair, 2.2 million particles were extracted using Warp[57] 1.0.9. Further processing was as for the first complex, except that the class containing RNA-bound RdRp yielded 850,000 particles and the refinement resulted in a reconstruction at an overall resolution of 3.2 Å.

Atomic models were built using our previously published SARS-CoV-2 RdRp–RNA complex structure (PDB 7B3D[23]) as starting model. The model was first rigid-body-fitted into the density and then manually adjusted in Coot[60]. The protruding second RNA turn and the nsp8 extensions were removed from the model due to discontinuous cryo-EM density for these regions, indicating mobility. Restraints for molnupiravir monophosphate (M) were generated in phenix.elbow[61] and the structures were refined using phenix.real_space_refine[62] with restraints

imposed on secondary structure and the backbone phosphate geometry of M. Model quality was assessed using MolProbity within Phenix[63], which revealed excellent stereochemistry for both structural models (Table 1). The figures and the Supplementary video were prepared with PyMol (Schrödinger) and ChimeraX[64].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability
The cryo-EM reconstructions and structure coordinates for the RdRp–RNA structures containing M-A or M-G base pairs have been deposited with the Electron Microscopy Database (EMDB) under accession codes EMD-13135 and EMD-13138 and with the Protein Data Bank (PDB) under accession codes 7OZU and 7OZV, respectively. Source data are provided with this paper. Other data are available from corresponding authors upon reasonable request.

References

55. Buttner, L., Seikowski, J., Wawrzyniak, K., Ochmann, A. & Hobartner, C. Synthesis of spin-labeled riboswitch RNAs using convertible nucleosides and DNA-catalyzed RNA ligation. *Bioorg. Med. Chem.* **21**, 6171–6180 (2013).
56. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
57. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
58. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
59. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
60. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
61. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).
62. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
63. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
64. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

Acknowledgements

Author contributions
F.K. and J.S. designed and carried out biochemical experiments and analyzed biochemical data. C.S. synthesized and analyzed M-containing RNA oligonucleotides. F.K. and C.D. carried out cryo-EM data collection and processing. H.S.H. assisted with model building. G.K. helped with interpretation of biochemical and structural data. C.H. designed and supervised RNA synthesis and analysis. P.C. designed and supervised research and wrote the manuscript, with input from all authors.

Funding

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41594-021-00651-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-021-00651-0.
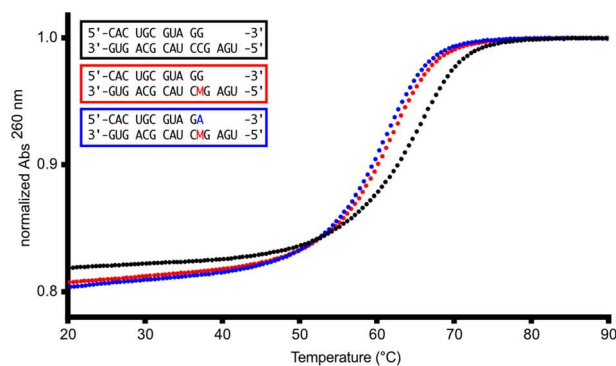
Correspondence and requests for materials should be addressed to C.H. or P.C.

Peer review information *Nature Structural & Molecular Biology* thanks Elizabeth Campbell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Editor recognition statement: Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.
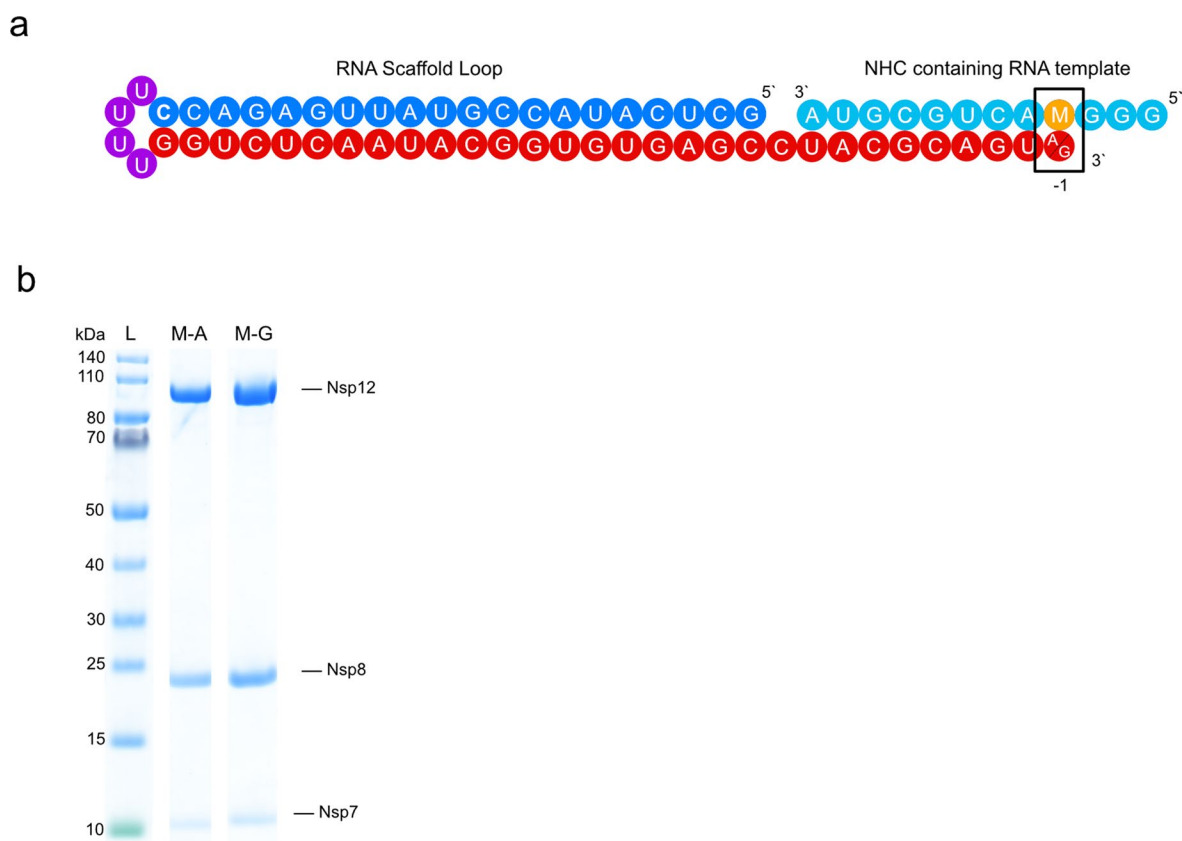
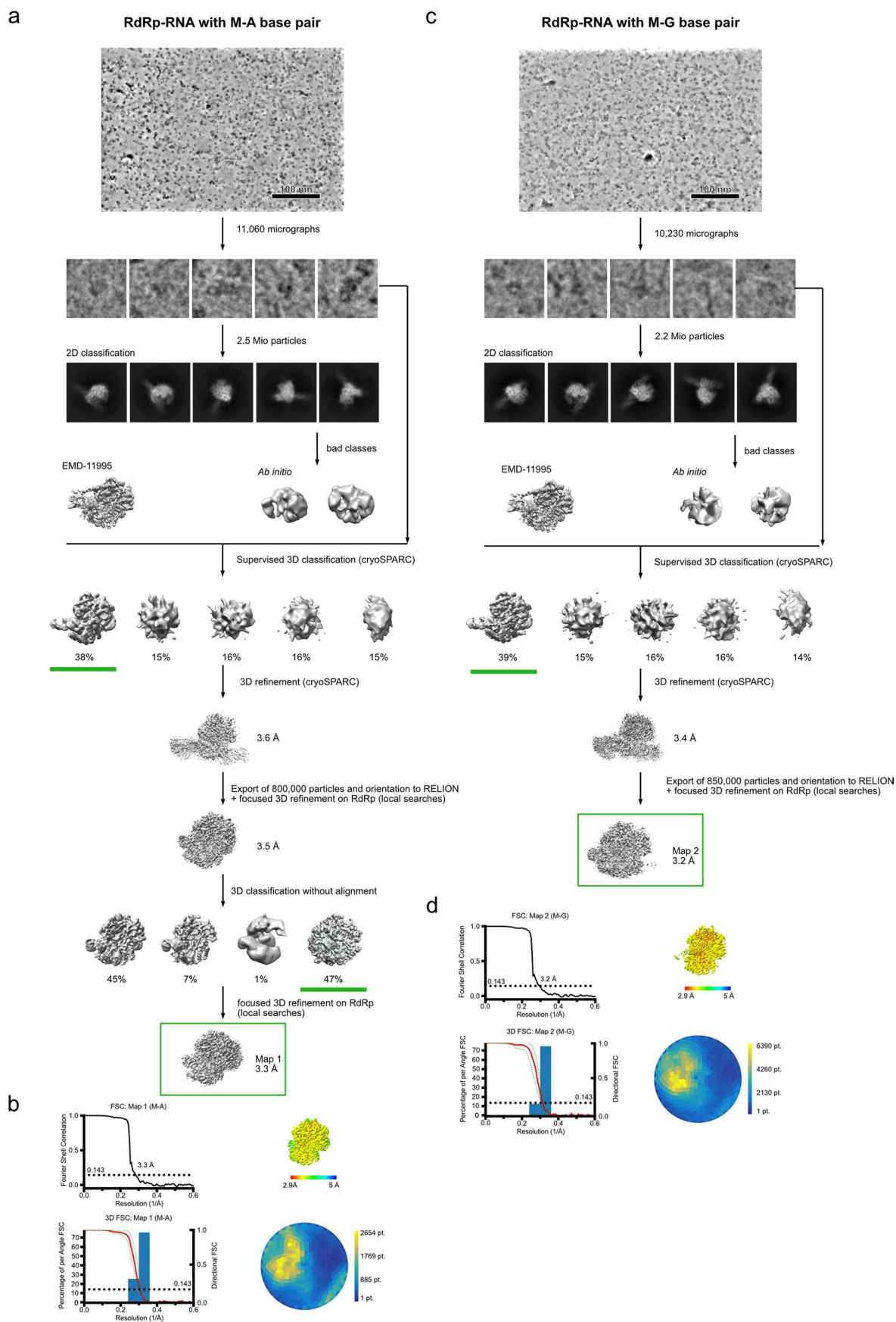**Reprints and permissions information** is available at www.nature.com/reprints.

Extended Data Fig. 1 | Synthesis of NHC-phosphoramidite 3 (M-PA). 5'-O-DMT-2'-O-TOM-O^4-chlorophenyluridine (**A**) was synthesized from uridine as previously reported (L. Buttner, J. Seikowski, K. Wawrzyniak, A. Ochmann, C. Hobartner, Synthesis of spin-labeled riboswitch RNAs using convertible nucleosides and DNA-catalyzed RNA ligation. Bioorg Med Chem 21, 6171-6180 (2013)). The chlorophenol group was displaced by hydroxylamine to give new compound **1**. After selective benzoyl protection at N4 with benzoic anhydride, compound **2** was converted to the phosphoramidite M-PA (**3**) using 2-cyanoethyl-N,N,N',N'-tetraisopropylphosphorodiamidite and 4,5-dicyano-imidazol (DCI) in analogy to a previous report (J. Lu, L. Nan-Sheng, J. A. Piccirilli, Efficient Synthesis of N 4-Methyl- and N 4-Hydroxycytidine Phosphoramidites. Synthesis 16, 2708-2712 (2010)). DMT-Cl = 4,4'-dimethoxy-trityl chloride, TOM-Cl = triisopropylsilyloxymethyl chloride, DBU = 1,8-diazabicyclo [5.4.0]undec-7-ene, DMAP = 4-(N,N-dimethylamino)- pyridine.

Extended Data Fig. 2 | Melting curves for RNA duplexes containing M–G or M–A base pairs. UV thermal melting monitored at 260 nm for 20 µM duplexes (11 bp, 4 nt single‾stranded overhang) in 100 mM NaCl, 10 mM Na-phosphate buffer pH 7.0. black: unmodified duplex C–G (Tm = 64.7 °C), red terminal M–G (Tm = 61.2 °C): blue: terminal M–A base pair (Tm = 60.6 °C). See also Supplementary Table 2.

Extended Data Fig. 3 | Protein preparation and RNA scaffold for structural studies. a, RNA scaffold was obtained by annealing a short M-containing oligonucleotide to a hairpin RNA duplex. b, SDS-PAGE of purified RdRp-RNA complexes used for cryo EM. Purified proteins were run on 4-12 % Bis-Tris SDS-PAGE gels in 1x MOPS buffer and stained with Coomassie Blue. M-A corresponds to the RdRp complex with RNA scaffold, where M in template base pairs to A. M-G corresponds to the RdRp complex with RNA scaffold, where M in template base pairs to G. L: PageRuler Prestained Protein Ladder (Thermo Scientific). The experiment was performed once, source data are provided as a Source Data file.

Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Cryo-EM data processing trees and quality of reconstructions. a, Cryo-EM data processing tree for M–A bp-containing RdRp-RNA structure. Scale bar, 100 nm. b, local resolution, FSC plot, angular distribution and directional FSC calculated according to Tan et al., Nature Methods 14, 793–796 (2017). Sphericity of M-A containing structure is 0.976; c, Cryo-EM data processing tree for M–G bp-containing RdRp-RNA structure. Scale bar, 100 nm. d, Local resolution, FSC plot, angular distribution and directional FSC calculated according to Tan et al., Nature Methods 14, 793–796 (2017). Sphericity of M-G containing structure is 0.966.

Corresponding author(s):   Patrick Cramer; Claudia Höbartner

Last updated by author(s):   Jul 26, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Serial EM 3.8 beta 8
Data analysis	RELION 3.1, USCF ChimeraX v0.8, Pymol 2.2.2, Coot 0.9, Warp v1.0.9, PHENIX 1.18, crypSPARC 2.15, Prism 9, Biorad Image Lab v6.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cryo-EM reconstructions and structure coordinates for the RdRp-RNA structures containing M–A or M–G base pairs were deposited with the Electron Microscopy Database (EMDB) under accession codes EMD-13135 and EMD-13138 and with the Protein Data Bank (PDB) under accession codes 7OZU and 7OZV, respectively. Source data are provided with this paper. Other data are available from corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. For cryo-EM samples, nine grids of each RdRp-RNA complex (M-G and M-A) were pre-screened to identify the optimal grid for data collection. The number of grids screened was random and was not limited by any experimental parameter.
Data exclusions	No data were excluded from the analyses.
Replication	All attempts of replication were succesful. Cryo-EM single particle analysis inherently relies on averaging a large number of independent observations. All biochemical experiments that were quantified were performed in independent triplicates. Results shown in figure 2b and 2d were performed once under exact same conditions.
Randomization	Samples were not allocated to groups. All cryo-EM particles used for structure determination adopt random orientations in the ice on the grid. Division of particles into random halves was automatically performed during 3D reconstruction by Relion 3.1. Other experiments did not involve randomization
Blinding	Blinding is not applicable for this study, as group allocation is not used.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☐ ☒	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Hi 5 cells: Expression System, Tni Insect cells in ESF921 media
Authentication	None of the cell lines were authenticated.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.