Uncovering the mode of action of small molecule inhibitors against the RNA polymerase of SARS-CoV-2

Dissertation

for the award of the degree

“Doctor rerum naturalium”

by the Georg-August-Universität Göttingen

within the graduate program

“International Max Planck Research School for Genome Science”

of the Georg-August University School of Science (GAUSS)

submitted by

Florian Kabinger

from Vienna, Austria

Göttingen 2024

Thesis Advisory Committee:

Prof. Dr. Patrick Cramer
Department of Molecular Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Prof. Dr. Matthias Dobbelstein
Institute of Molecular Oncology
University Medical Center Göttingen

Prof. Dr. Hauke Hillen
Department for Cellular Biochemistry
University Medical Center Göttingen

Members of the Examination Boards:

Referee: Prof. Dr. Patrick Cramer
Department of Molecular Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

2nd Referee: Prof. Dr. Matthias Dobbelstein
Institute of Molecular Oncology
University Medical Center Göttingen

Further members of the Examination Boards:

Prof. Dr. Hauke Hillen
Department for Cellular Biochemistry
University Medical Center Göttingen

Prof. Dr. Stefan Pöhlmann
Infection Biology Unit
German Primate Center, Göttingen

Prof. Dr. Herbert Jäckle
Department of Molecular Developmental Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Prof. Dr. Johannes Söding
Computational Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Date of oral examination: 26th of August, 2024

0. Acknowledgements

First of all, I would like to thank Prof. Patrick Cramer for allowing me to work in his research group. He is an inspiring mentor and supported me in numerous ways. I am thankful for his ideas and his vision, which allowed us to tackle intriguing scientific questions. I am honored to have worked with him, and I am convinced that the scientific community will benefit from his new role as president of the Max-Planck-Society.

Next, I would like to express my gratitude to Michael Lidschreiber. Especially towards the end of my PhD, he stepped in as a great supervisor and helped me in multiple ways. I admire his personality, not only as a curious scientist but also as a person who always cares about the well-being of his colleagues.

I am also very thankful for a supportive TAC committee. Prof. Matthias Dobbelstein was always interested in discussing my latest results, and I really appreciated his scientific advice. Prof. Hauke Hillen was not only a TAC member but also a close collaborator. It was great working with him, and he always found time to help. I would also like to thank the members of my examination board, Prof. Stefan Pöhlmann, Prof. Herbert Jäckle and Prof. Johannes Söding, for their time.

The research described within this thesis would not have been possible without huge support from colleagues. A special thank you hereby goes to Christian Dienemann. He spent countless hours with me in front of the microscope and was a great teacher. He was always happy to answer my questions and solve issues, even providing remote troubleshooting instructions. I would also like to thank Jana Schmitzová for welcoming me into the lab and introducing me to the RdRp project. It was great to learn from her experience. A big thank you to Valerie Doze. She performed countless gel-based experiments and was an important part of the RdRp inhibitor project. It was a great experience mentoring her, and I really appreciate the scientific discussions we had. Thank you also for proofreading my thesis. Further, I would like to thank Louisa Menges, who was my first student in the Cramer lab, for her important contributions to the RdRp project. Thanks to Maik Engelholm for sharing the office with me and for his scientific advice. Additionally, I would like to thank Alexander Rotsch, and all members of the “weird polymerase” subgroup, for having breakfast together and for the great scientific discussions. Thank you to all Äkta team members, especially Issac Fianu and Julio Abril Garrido, for the fun we had during many hours spent with

troubleshooting. Thanks to all “genomic warriors”, the volleyball team of the Cramer lab, for many great beach volleyball games. In general, thank you to all current and past members of the Cramer lab for creating a very supportive research environment.

In addition, I am grateful for the help of all collaboration partners. Prof. Claudia Höbartner and Carina Stiller from the Universität Würzburg were especially instrumental in deciphering the mode of action of Molnupiravir.

I would also like to express my gratitude towards agencies that supported my PhD project. Receiving generous support from the Boehringer Ingelheim Fonds (BIF) allowed me to perform the here described research and gave me access to a very inspiring group of scientists, the BIF Family. Thanks to the whole BIF team, especially Stephan Formella, for providing this support. In addition, I would like to thank the coordinator of the IMPRS for genome science, Henriette Irmer, and Frauke Bergmann, for their support, and a big thanks to all students of the IMPRS-GS for the motivating retreats.

My scientific journey did not start at the beginning of my PhD but rather began much earlier. Thus, I would like to thank all teachers and mentors who inspired me to do research and allowed me to grow as a person and scientist. Prof. Angela Koehler’s support during my internship in Boston was especially integral to my career, and Prof. Christoph Loderer’s mentoring during my internship in Stockholm provided me with a strong skillset and knowledge base. Lecturers from my bachelors and masters studies at the University for applied sciences Vienna, and teachers from the “AHS Friesgasse”, were instrumental in sparking my interest in science.

Additionally, I would like to thank many people outside of the lab for reminding me that there is more than the scientific world. Thank you to all members of the Frisbee team, “Göttinger 7”, and all members of the Volleyball team, “Grelli 3”, for welcoming me onto their team and for the great times we spent together playing sports. In addition, I am very grateful to all my close friends in Göttingen and in Vienna for listening to my complaints and motivating me.

In the end, I would like to express my deepest gratitude towards my girlfriend, my parents, my grandparents and all other family members. They allowed me to follow my curiosity and always supported me, even in cases when I was far away from home. Thank you for believing in me and for providing a place in Vienna and in Göttingen that feels like home.

1. Manuscripts and Contribution

Part of this thesis have been published in a peer-reviewed journal or are currently in the process of being published:

Manuscript 1:

Florian Kabinger*, Carina Stiller*, Jana Schmitzová*, Christian Dienemann, Goran Kokic, Hauke S. Hillen, Claudia Höbartner & Patrick Cramer. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* ***28***, 740–746 (2021). https://doi.org/10.1038/s41594-021-00651-0

* Authors contributed equally

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. **F.K.** designed and created all figures. C.H. designed and supervised RNA synthesis and analysis. P.C. designed and supervised research and wrote the manuscript, with input from all authors.

Manuscript 2:

Florian Kabinger, Valerie Doze, Jana Schmitzová, Christian Dienemann, Michael Lidschreiber, Patrick Cramer. Structural basis of SARS-CoV-2 polymerase inhibition by the non-nucleoside inhibitor HeE1-2Tyr. *in preparation* (2024)

Author contributions: F.K. and V.D. designed and carried out biochemical experiments and analyzed biochemical data. J.S. performed protein production. **F.K.** and J.S. designed and performed biophysical experiments. F.K. carried out cryo-EM data collection and processing. C.D. assisted with interpretation of biochemical and structural data. **F.K.** designed and created all figures. **F.K.**, C.D., M.L. wrote the manuscript, with input from all authors. P. C. designed and supervised research

Chapter 6. and chapter 7. of this thesis were extracted from Manuscript 1 and 2, respectively. Text and figures were modified to adapt them to the format of this work.

2. Summary

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19).[1,2] The pandemic has claimed more than 7 million lives since its outbreak at the beginning of 2020, and the number of deaths is still raising.[3] SARS-CoV-2 is a positive-sense single-stranded RNA virus that encodes an RNA-dependent RNA polymerase (RdRp) for replication of its genome.[4,5] Inhibition of this enzyme with small molecules blocks the viral life cycle and therefore presents a highly promising therapeutic strategy against COVID-19.[6,7] Despite extensive efforts by the scientific community, potent and safe RdRp inhibitors are still lacking. Further improvement of potentially promising small molecules is often hampered by a lack of knowledge about their mode of action.[8] In this doctoral thesis I present the biochemical and structural basis of RdRp inhibition by the nucleoside inhibitor Molnupiravir and by the non-nucleoside inhibitor HeE1-2Tyr.

In order to perform quantitative analysis of compound-induced RdRp inhibition, I advanced the development of an existing minimal biochemical system recapitulating viral RNA replication *in vitro.* In addition, I established a fluorescence polarization assay that allows investigation of the interaction between RdRp and its RNA substrate. Combining these and other biochemical assays enabled us to quantify the inhibitory effects of small molecules on RdRp. To complement the biochemical data, single particle cryogenic electron microscopy (cryo-EM) was utilized to visualize the inhibitor bound to RdRp. I was able to solve cryo-EM structures of RdRp in the presence of Molnupiravir and HeE1-2Tyr with an overall resolution of 3 Å. This allowed us to visualize key drug-target interactions and to deduce the molecular mechanism of inhibition.

Molnupiravir is an orally available antiviral drug that was in phase III clinical trials for treatment of COVID-19 patients at the time this research project was initiated.[9] It was known that Molnupiravir increases the frequency of viral RNA mutations[10] and impairs SARS-CoV-2 replication in animal models[11] and in humans[12,13], but the underlying mode of action remained elusive. We were able to establish the molecular mechanisms underlying Molnupiravir-induced RNA mutagenesis by RdRp. Biochemical assays show that the RdRp uses the active form of Molnupiravir, β-D-N4-hydroxycytidine (NHC) triphosphate, as a substrate instead of CTP or UTP. When the RdRp uses the resulting RNA as a template, NHC directs incorporation of either G or A, leading to mutated RNA products. In order to

understand the mutagenesis on a molecular level, I solved two high-resolution cryo-EM structures of RdRp-RNA complexes containing the mutagenesis products. These show 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 likely applies to various viral polymerases and can explain the broad-spectrum antiviral activity of Molnupiravir.

HeE1-2Tyr is a non-nucleoside inhibitor of the dengue virus RdRp[14], which was also shown to be effective against SARS-CoV-2 in cell culture.[15] However, it remains unclear how HeE1-2Tyr facilitates inhibition of the SARS-CoV-2 RdRp. Combining biochemical and structural data, we were able to elucidate the mode of action of HeE1-2Tyr-mediated SARS-CoV-2 RdRp inhibition. Biochemical assays confirm that HeE1-2Tyr inhibits RdRp with an IC_{50} of 5 µM and show that the compound interferes with RNA binding to RdRp *in vitro*. Structural analysis using cryo-EM reveals that a stack of three HeE1-2Tyr molecules binds to the RNA binding site of RdRp. The stack is stabilized by inter-compound π-π interactions and by a clamp of three arginine residues that are highly conserved across coronaviruses. This suggests that HeE1-2Tyr inhibits the RNA synthesis of SARS-CoV-2 by sterically blocking the interaction between RdRp and RNA, which readily explains the biochemical data.

In summary, my doctoral thesis contributes to the global drug development efforts to combat SARS-CoV-2 in multiple ways. The here established biochemical assays provide a valuable tool for quantitative characterization of RdRp inhibition, and the applied cryo-EM strategy can be used for efficient structural elucidation of RdRp-inhibitor complexes. Understanding the two-step mutagenesis mechanism of Molnupiravir was highly relevant for its clinical usage and studying HeE1-2Tyr-mediated RdRp inhibition revealed a novel mechanism to inhibit RdRp. Altogether, this work has the potential to drive further development of potent pan-corona inhibitors adding to pandemic preparedness.

3. Table of contents

1. Acknowledgements ... 3

2. Manuscripts and Contribution ... 5

3. Summary ... 6

4. Table of contents ... 8

5. General introduction: The biology of SARS-CoV-2 ... 10

5.1. Taxonomy ... 12

5.2. Genome architecture and gene products ... 13

5.3. The life cycle of SARS-CoV-2 ... 15

5.4. Transcription and replication ... 18

5.4.1. RNA synthesis by RdRp ... 19

5.4.2. Backtracking by RdRp – nsp13 complexes ... 21

5.4.3. Proofreading by nsp10 and nsp14 ... 22

5.4.1. Capping complexes of SARS-CoV-2 ... 23

5.4.1. Nsp15 and the formation of a super-complex ... 24

5.5. Antivirals ... 25

5.5.1. Nucleoside analogues ... 26

5.5.2. Non-nucleoside analogues ... 30

6. Manuscript 1: Mechanism of Molnupiravir-induced SARS-CoV-2 mutagenesis ... 32

6.1. Abstract ... 32

6.2. Introduction ... 32

6.3. Results ... 34

6.3.1. SARS-CoV-2 RdRp readily incorporates NHC into RNA ... 34

6.3.2. RdRp does not stall after NHC incorporation ... 36

6.3.3. RdRp uses NHC-containing templates to direct RNA mutagenesis ... 37

6.3.4. Structural basis of NHC-induced RNA mutagenesis ... 38

6.4. Discussion ... 41

6.5. Methods ... 43

6.5.1. Protein preparation ... 43

6.5.2. RNA extension assays ... 45

6.5.3. Preparation and analysis of NHC-containing RNA oligonucleotides ... 45

6.5.4. Synthesis and characterization of M-PA ... 46

6.5.5. Cryo-EM sample preparation and data collection ... 47

6.5.6. Cryo-EM data processing and structural modeling ... 49

7. Manuscript 2: Structural basis of SARS-CoV-2 polymerase inhibition by non-nucleoside inhibitor HeE1-2Tyr 50

7.1. Abstract 50

7.2. Introduction 51

7.3. Results 52

7.3.1. HeE1-2Tyr is a potent non-nucleoside inhibitor of SARS-CoV-2 RdRp 52

7.3.2. HeE1-2Tyr interferes with RNA binding to SARS-CoV-2 RdRp 54

7.3.3. Cryo-EM structural analysis of HeE1-2Tyr binding to RdRp 57

7.3.4. HeE1-2Tyr binds to RdRp as a stack of three molecules 60

7.3.5. The HeE1-2Tyr stack is stabilized by an arginine bracket 62

7.3.6. RNA and HeE1-2Tyr both stabilize the RdRp complex 64

7.3.7. The arginine bracket is required for RNA binding and conserved across corona viruses 66

7.4. Discussion 69

7.5. Methods 71

7.5.1. Protein preparation 71

7.5.2. RNA extension assays 72

7.5.3. RNA fluorescence polarization assay 73

7.5.4. Mass photometry 75

7.5.5. Cryo-EM sample preparation and data collection 76

7.5.6. Cryo-EM data processing and structural modeling 76

7.5.7. Ligand interactions 77

7.5.8. Sequence conservation 77

8. General discussion: Targeting the RdRp of SARS-CoV-2 with small molecule inhibitors 78

8.1. Advantages and drawbacks of Molnupiravir-induced mutagenesis 78

8.1.1. Tautomerism of Molnupiravir and its implications for base pairing 78

8.1.2. Similarity of Molnupiravir to endogenous molecules and its consequences 79

8.1.3. Concerns around Molnupiravir´s lack of selectivity 80

8.2. HeE1-2Tyr: A novel mode of action to inhibit RdRp 81

8.2.1. HeE1-2Tyr is an atypical example for positive cooperativity 81

8.2.2. Stacking of multiple inhibitor molecules; a promising strategy? 82

8.2.3. Further advancement of HeE1-2Tyr into a potential pan-corona RdRp inhibitor 83

8.3. The future of drug development against corona virus RdRp 84

8.3.1. Do we still need antivirals against corona viruses? 84

8.3.2. Designing an efficient pipeline for drug discovery against SARS-CoV-2 RdRp 86

9. References 88

10. List of figures 101

11. Appendix 102

4. General introduction: The biology of SARS-CoV-2

Viruses are ubiquitous. We are surrounded by them, and they are an essential part of a healthy ecosystem.[16] Viruses are nonliving infectious entities that bridge the gap between inert chemicals and living organisms.[17] They fulfill some criteria of living systems, such as the presence of a genome, but they lack the ability to self-replicate and have no metabolism. Thus viruses are defined as nonliving obligate intracellular parasites, which depend on the metabolism of the host cell to replicate.[18] Therefore, it is also said that viruses „lead a kind of borrowed life".[17]

Viruses are able to infect every form of cellular life.[19] They are the most abundant biological entity on planet earth with approximately 10^{31} viral particles in the biosphere, outnumbering all forms of cellular life combined.[20,21] Most of these viruses only infect microorganisms and do not harm humans, such as bacteriophages, which only infect bacterial cells.[22] However, some viruses do infect humans, presenting a major challenge for global human health.[23]

There are approximately 270 known viral species that can infect humans.[24] Some of these cause only mild symptoms such as coughing and sneezing (e.g. Rhinovirus) while others have a high virulence and cause serious life-threatening conditions in infected patients (e.g. Ebola virus). Unfortunately, there is a constant increase in the number of viruses that can infect humans due to zoonotic spillovers. Such events happen when a virus that is circulating in a wild-life population or livestock animals mutates and acquires the ability to infect human cells.[25,26] Depending on severity of the caused disease and the infectivity of the virus, zoonotic spillovers can lead to devastating effects on a global scale. A recent infamous example of this risk is the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in January 2019.[1] This virus is the causative agent of coronavirus disease 2019 (COVID-19), and it has claimed more than 7 Million lives.[2,3] By the time this thesis is written (June 2024), the acute phase of the SARS-CoV-2 pandemic has been overcome by the deployment of various pharmaceutical and non-pharmaceutical interventions. These measures helped to save millions of lives and allowed the population to build a baseline immunity against SARS-CoV-2.[27,28] Despite the progress made, SARS-CoV-2 is still spreading by reinfecting patients with mutated variants and causing premature death due to COVID-19. Thus, the pandemic is not over and it is likely that SARS-CoV-2 will become an endemic virus, such as Influenza.[29]

Predicting and preventing zoonotic viruses and their potential detrimental effects on human health is a highly complex topic that has to be tackled by various scientific disciplines.[30] In the past, it has been neglected to study these viruses due to lack of economic incentives. Now, after the shock of SARS-CoV-2, it has been realized that zoonotic viruses are a real threat for humans and that we need to invest more resources in order to be prepared for the next pandemic.

The aim of this doctoral thesis is to contribute to the global efforts to fight SARS-CoV-2 and to generate knowledge which potentially helps us during future pandemics. Therefore, the thesis is focused on the molecular replication machinery of SARS-CoV-2 and its interplay with small molecule inhibitors. Understanding potential mechanisms to inhibit the molecular weak point of corona viruses, namely RNA replication, allows us to exploit it for the development of efficient antiviral treatments.[31] To accomplish this, key viral components are reconstituted *in vitro* and subjected to biochemical and structural studies in the presence of potential inhibitors. This defined biochemical system allows us to decipher the mode of action of molecules that block the replication of SARS-CoV-2.

All experiments are performed with *in vitro* reconstituted systems, strictly avoiding the usage of infectious SARS-CoV-2 viral particles. In addition, research questions that involve the interplay between the virus and the immune system or clinical development of the compounds would have gone beyond the scope of this thesis and were therefore omitted.

The thesis will start with a general introduction where I will address key aspects of the biology of SARS-CoV-2 with the main focus on the viral replication machinery. It will then continue with two main chapters that are comprised of self-contained manuscripts describing the mode of action of the inhibitors Molnupiravir and HeE1-2Tyr, respectively. The individual manuscripts contain an introduction focused on the according small molecule, a results part, a focused discussion and a methods section. The thesis will end with a general discussion where I describe the main conceptual advances and contextualize here described results with existing literature. In addition, I will provide an outlook that summarizes my perspective on the future of the drug development against corona virus RdRp.

4.1.Taxonomy

SARS-CoV-2 is a positive sense single stranded RNA virus that belongs to the order of Nidovirales.[2,32] The name of this order is derived from the Latin word "nidus" which stands for "nest", and describes a key molecular feature. All viruses of the order Nidovirales have in common that they produce a set of nested short RNAs during transcription, which are called subgenomic RNAs (sgRNAs)[33,34], a process that is discussed in more detail in chapter *4.4.1 RNA synthesis by RdRp*. Within the order of Nidovirales, SARS-CoV-2 belongs to the family of Coronaviridae.[2] Viral particles of this family have a very distinctive appearance when analyzed with electron microscopy (EM). The membrane of Coronaviridae viruses is decorated with large surface proteins, the so-called spike proteins, which are important for the interaction with the host.[35] The projections of these cause the typical annular pattern of protrusions in EM images, which is well reflected by the name "corona" standing for "crown".[36,37] The family of Coronaviridae is then further subdivided into the subfamily of Orthocoronavirinae, which consists of four genera, termed Alpha-, Beta-, Gamma- and Deltacoronavirus. Alpha- and Betacoronaviruses are of particular interest because they harbor viruses that can infect humans.[38] In contrast, so far no Gamma- or Deltacoronaviruses are known that are pathogenic to humans. Examples of Alphacoronaviruses that can infect humans are Human coronavirus 229E (HCoV-229E) and Human coronavirus NL63 (HCoV-NL63); both are causing the common cold but are generally not deadly. The genus of Betacoronaviruses also harbors member causing the common cold, such as Human coronavirus OC43 (HCoV-OC43) and Human coronavirus HKU1 (HCoV-HKU1). Besides that, some betacoronaviruses are far more virulent and infectious, such as Middle East Respiratory Syndrome Coronavirus (MERS-CoV), severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1) and SARS-CoV-2. SARS-CoV-1 and MERS-CoV are two betacoronaviruses that are closely related to SARS-CoV-2 and are known for causing severe disease outbreaks in 2003 and 2012, respectively.[39] SARS-CoV-2 has a sequence identity of 79% with SARS-CoV-1 and 50% MERS-CoV.[40] Subtle differences in the genome affecting the virulence and infectivity are probably the reason why SARS-CoV-1 and MERS-CoV only caused local epidemics compared to the global pandemic caused by SARS-CoV-2.

So far Alpha-, Gamma- and Deltacoronaviruses were not involved in severe human disease outbreaks due to their low virulence or their lack of ability to infect humans. However, chances are high that zoonotic spillover or gain of function mutations will happen

in the future, causing potential disease outbreaks. Thus, it can be concluded that coronaviruses and especially Betacoronaviruses present a substantial threat for human health tragically exemplified by past epidemics and the ongoing worldwide COVID-19 pandemic.

4.2.Genome architecture and gene products

Coronaviruses have in common that they are encoded by one of the largest known genomes among RNA viruses, consisting of a single strand of RNA with an approximate length of 30 000 nucleotides.[41,42] A schematic representation of the RNA genome of SARS-CoV-2 and its gene products is depicted in Fig. 1. The 3´ and 5´ ends are flanked with evolutionary conserved untranslated regions (UTR). These regions harbor a variety of secondary RNA structural elements and are known for their important functions during the transcription and replication of the RNA genome.[43-46]

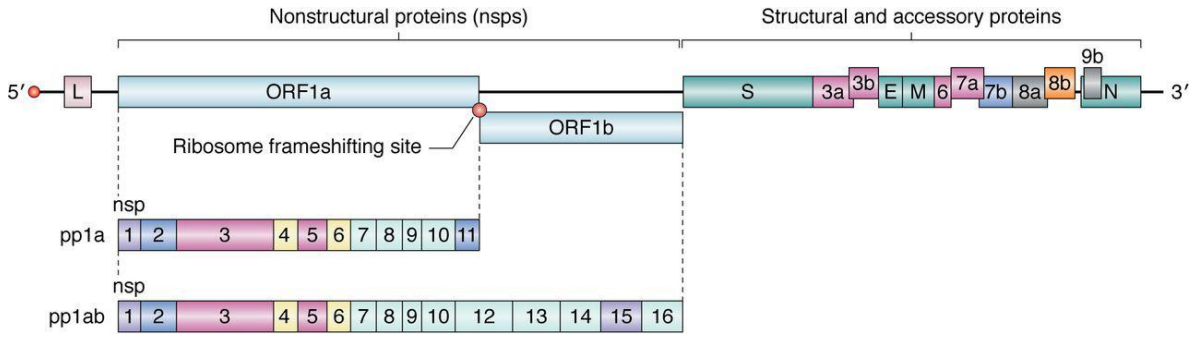


Figure 1: Genome organization of SARS-CoV-2

adopted from: Hartenian et. al.[47], The RNA genome encodes two categories of proteins: nsps and structural and accessory proteins. The nonstructural proteins are encoded in ORF1a and ORF1b. Cap-dependent translation begins at ORF1a and produces pp1a, encompassing nsp1–11, or pp1ab, a longer polypeptide that includes nsp12–16. The production of either polypeptide depends on whether the stop codon at ORF1a is recognized by the ribosome or is bypassed through a change in the reading frame by the ribosome frameshifting site. The structural and accessory proteins are synthesized by translation of their respective subgenomic mRNAs. The proteins have been color-coded by functional categories for SARS-CoV-2

The genome of SARS-CoV-2 encodes 14 partially overlapping open reading frames (ORFs).[48] Two thirds of the genome are comprised by the replicase gene (ORF1ab), while the rest encodes the information for structural- and accessory-proteins. Transcription and translation of ORF1ab results in two polyproteins, termed pp1a and pp1ab. The latter is a c-terminally extended form of pp1a, containing a programmed -1 ribosomal frameshift site to overcome the stop codon at the end of pp1a.[49,50] Proteolytic cleavage of the polyproteins by two virally encoded proteases results in 16 non-structural proteins (Nsps), named

Nsp1 - Nsp16.[42] There are numerous remaining questions about the biological functions of structural and especially non-structural proteins, which we cannot answer yet. Table 1 and table 2 list all non-structural and structural proteins from SARS-CoV-2, summarizing the current knowledge about their proposed functions.

Non-strucutral proteins	Function
Nsp1	Cellular mRNA degradation, host translation inhibitor[51,52]
Nsp2	Unknown
Nsp3	Polyprotein cleavage: Papain-like protease[53], component of the replication organelle pore[54,55]
Nsp4	Replication organelle formation[55]
Nsp5	Polyprotein cleavage: 3C-like protease[56,57]
Nsp6	Replication organelle formation[55]
Nsp7	Subunit of the RdRp[58,59]
Nsp8	Subunit of the RdRp[58,59], putative involvement in priming[60]
Nsp9	Nucleic acid binding protein, NiRAN UMP transferase substrate[61] involved in capping[62]
Nsp10	Stimulatory subunit of the nsp14 exonuclease and nsp16 methyltransferase activities[63-65]
Nsp11	Unknown
Nsp12	RdRp[58,59], NMPylase activity: NiRAN domain[61], putative involvement in capping[62]
Nsp13	RNA helicase[66,67], RNA 5'-phosphatase[68]
Nsp14	Proofreading: 3'-5' exonuclease[69,70], capping: N^7-methyltransferase[71]
Nsp15	Uridylate-specific endoribonuclease[72,73]
Nsp16	Capping: 2'-O-Methyltransferase[74]

Table 1: Non-strucutral proteins of SARS-CoV-2 and their proposed functions

Strucutral proteins	Function
Spike protein (S)	Host cell receptor binding for viral entry[75,76]
Envelope protein (E)	Viral assembly and release[77]
Membrane protein (M)	Virion shape, membrane curvature[78]
Nucleocapsid protein (N)	Packaging of genomic RNA[79,80], putative interaction with the RTC[81,82]

Table 2: Strucutral proteins of SARS-CoV-2 and their proposed functions

Accessory genes and their potential protein products are not included in the list because their sequence is highly variable among coronaviruses and their function in the viral life cycle is generally poorly understood.[83]

Nsps can be separated into two groups based on their main functions. Members of the first group are involved in the modulation of the host cell environment, such as immune suppression or compartmentalization of the host cell. Nsps of the second group are enzymes or cofactors that catalyze critical steps in the viral life circle such as RNA replication.[42] A detailed introduction of important Nsps is provided in the following chapters alongside the description of key molecular processes in the viral life cycle of SARS-CoV-2.

4.3.The life cycle of SARS-CoV-2

Viral replication occurs inside of host cells, as schematically depicted in Fig 2. Thus, viral infection starts with the attachment of viral particles to the host cell membrane and subsequent cell entry.[42] SARS-CoV-2 utilizes the spike protein (S), which is a glycoprotein that forms a homotrimeric complex, well-known for its iconic shape from many electron microscopy images of the virus. The spike protein can be divided into two functional parts, termed S1 and S2, separated by a proteolytic cleavage sequence.[84] S1 is surface-exposed and harbors the receptor binding domain (RBD) that physically interacts with cell surface proteins and defines the tropism of the virus. Structural and biochemical data show that the RBD of SARS-CoV-1 and SARS-CoV-2 interacts with the human angiotensin-converting enzyme 2 (ACE2) using a conserved binding interface.[85] Upon attachment of the RBD to ACE2, host cell proteases activate the S protein by proteolytic cleavage at two specific sites.

First, a furin protease cleaves the S1/S2 site, which is located on the boundary between the domains and only present in SARS-CoV-2. Then the type II transmembrane serine protease 2 (TMPRSS2) cleaves the S2´ site which resides in the S2 domain and is conserved among corona viruses.[75,86] This leads to extensive conformational rearrangements of the S protein that activate the membrane fusion subunit of S2 and expose the fusion peptide, resulting in membrane fusion.[87]

Fusion of the viral membrane and the host cell membrane allows the release of the viral genome into the cytoplasm. There the N protein dissociates from the RNA, and cellular ribosomes can directly use the positive sense RNA genome for the translation of the viral polyproteins.[42] In 45% - 70% of translation events, a -1 frameshift occurs between ORF1a and ORF1b, leading to a 1.4 – 2.2-fold expression of pp1a compared to pp1ab.[50,88] In other words, nsps1-11 (located on pp1a) are approximately twice as abundant as nsp12-16 (located on pp1ab). Frameshifting of the ribosome is induced by the formation of a pseudoknot in the viral RNA that acts as a road block and generates tension. In addition, the energy barrier of backtracking of the ribosome is lowered by a slippery site and specific interactions between the viral polyprotein and the ribosomal tunnel.[88] This presumably allowed SARS-CoV-2 to finetune the stoichiometric ratios between the according nsps. Upon translation of pp1a or pp1ab, the polyprotein is cleaved into its individual proteins by viral protease activity, residing in the cysteine proteases nsp3 and nsp5. Nsp3 is a papain-like protease which gave rise to its name, Plpro. Nsp5, also frequently referred to as the main protease (Mpro) because it processes most cleavage sites of the polyprotein, is a chymotrypsin-like protease with similarities to the picornaviral 3C protease, which explains the abbreviation 3CLpro.[42]

Newly produced nsps reorganize the cellular environment and form compartments acting like molecular factories for the production of viral components.[89] Inside these compartments, full-length positive sense RNA genomes and structural viral proteins are produced for the assembly of new infectious viral particles. The production of both components requires replication and transcription of the genome by the RNA-dependent RNA polymerase (RdRp) and other RNA processing enzymes.[5] This process is highly complex and involves several distinct steps, described in detail in chapter *4.4. Transcription and replication.*

Briefly, RdRp replicates the viral RNA and produces full length copies acting as genomes for new viral particles. In addition, RdRp performs discontinued transcription which produces short subgenomic RNAs (sgRNAs).[90] These molecules are comparable to messenger RNA (mRNA) and harbor usually only one ORF of a structural protein flanked by the 3´ ladder sequence of SARS-CoV-2. This allows cellular ribosomes to use sgRNAs as template for translation.

Once the genome is replicated and the structural proteins are expressed, assembly of new viral particles can begin. The genomic RNA is encapsulated by N protein molecules to form a large ribonucleoprotein complex, creating a helical symmetrical structure.[80] In addition, the structural proteins S, M, and E are assembled in the ER–Golgi intermediate compartment (ERGIC), reshaping the intracellular lipid membranes.[91] The encapsulated RNA is then loaded and new virions are formed by budding into the lumen at ERGIC. The viral particles are then released by exocytosis or lysosomal trafficking pathways and the viral life cycle can start again.[92]

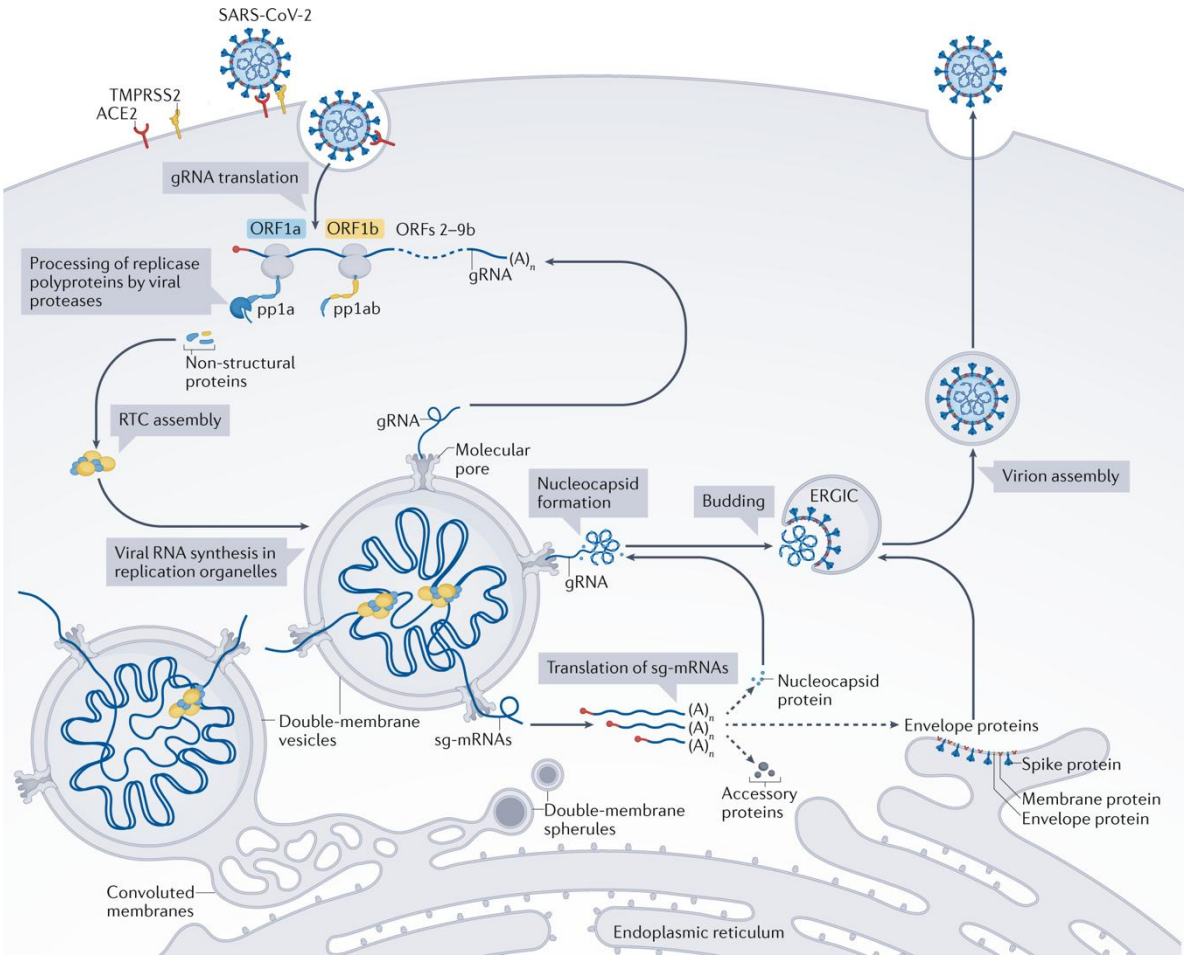


Figure 2: The life cycle of SARS-CoV-2

adopted from: Malone et. al.[93], critical steps of the viral life cycle are described in the text

4.4.Transcription and replication

Propagation of a 30 kb long RNA genome with high fidelity presents a major challenge for a virus. Viral RNA-dependent RNA polymerases are more error-prone than DNA polymerases and have an error frequency of 10^{-4} on average.[94] In other words, typical viral RdRps incorporate one erroneous nucleotide every 1000 transcribed bases. This would create a natural limit of the RNA genome length between 15 kb – 20 kb. [95] Beyond this size, too many mutations would be introduced by each round of replication, and the virus could not propagate efficiently. Therefore, SARS-CoV-2 and corona viruses in general had to develop a sophisticated molecular machinery that deviates from the average RdRp and allows them to overcome the size limiting factor of genome length. This happened during viral evolution, and corona viruses are now equipped with a versatile replication and transcription complex (RTC).[95,96] The RTC is formed by several nsps and harbors multiple functions allowing RNA genome replication with high efficiency and fidelity. This includes but is not limited to the synthesis of new RNA, a proofreading mechanism, sgRNA synthesis and RNA capping.

Replication and transcription of the RNA genome happens in replication organelles, which provide a shielded environment for the synthesis of RNA. These organelles are termed double-membrane vesicles (DMVs) and are formed by the interaction of SARS-CoV-2 nsps with the host ER membrane and potentially other host proteins.[91] DMVs were studied using cryogenic electron tomography (cryoET) and other microscopy techniques, which revealed the existence of DMV pores. The pores are formed by nsp3-nsp4 and connect the inside of DMVs with the host cell cytoplasm.[54,97] Besides that, it was shown that double stranded RNA (dsRNA), the product of RNA replication, is present inside the DMVs and that proteins of the RTC are associated with DMVs.[89] However, it remains elusive where and how RNA is processed in the vicinity of the DMVs. For example, one key question focuses on whether the RNA synthesis takes place on the inner (luminal) or outer (cytoplasmic) side of the DMV pore. In addition, we lack the complete molecular picture of the RTC.

Despite its name, the RTC is not a well-defined protein complex. Instead, it is used as an umbrella term for nsps that come together and catalyze critical RNA processing steps such as RNA synthesis or capping. So far, nobody was able to isolate or characterize a complete RTC, which led to two hypotheses. First, nsps might interact biophysically and

form a stable super-complex bound to or inside DMVs, that has not been found yet. Second, it is plausible that the local concentration of nsps subcomplexes inside DMVs is sufficiently high that processing steps are performed without the need of a stable protein complex.[81,93]

Impressive progress has been made by the scientific community to investigate these RTC subcomplexes structurally and biochemically. The following chapter will provide a summary of our current knowledge about the molecular details of SARS-CoV-2s RNA processing steps, including RNA synthesis, RNA capping and the proofreading function.

4.4.1. RNA synthesis by RdRp

The RdRp is the core enzyme complex of the RTC. Several high resolution cryo-EM structures of RdRp were solved in the absence and presence of RNA, providing detailed molecular insights.[98-100] Fig. 3 depicts the structure of a replicating RdRp bound to RNA.[99] The RdRp complex is comprised of the large catalytic subunit nsp12 bound to one molecule of the accessory subunit nsp7 and two molecules of accessory subunit nsp8. Nsp12 harbors a C-terminal polymerase domain that is homologous to other viral RdRp and adopts the typical right-hand fold of single subunit RNA polymerases. The appearance of such polymerases has similarities with a right hand grabbing on to RNA. Accordingly, the polymerase domains can be divided into the finger, palm and thumb subdomains.[58] The RNA binding site, also termed RNA binding cleft, is positioned between the finger and thumb domain, and the active site of the polymerase is located at the interface of the finger and palm domain. The active site is comprised of evolutionary highly conserved residues, such as threonine 619, aspartic acid 760 and aspartic acid 761, which coordinate two metal ion essential for catalysis.[99] In addition, nsp12 accommodates an N-terminal signature domain of nidoviruses. It was shown that this domain resembles the nucleotide binding domain of kinases, and it was therefore termed nidovirus RdRp-associated nucleotidyl transferase (NiRAN) domain.[5,101]

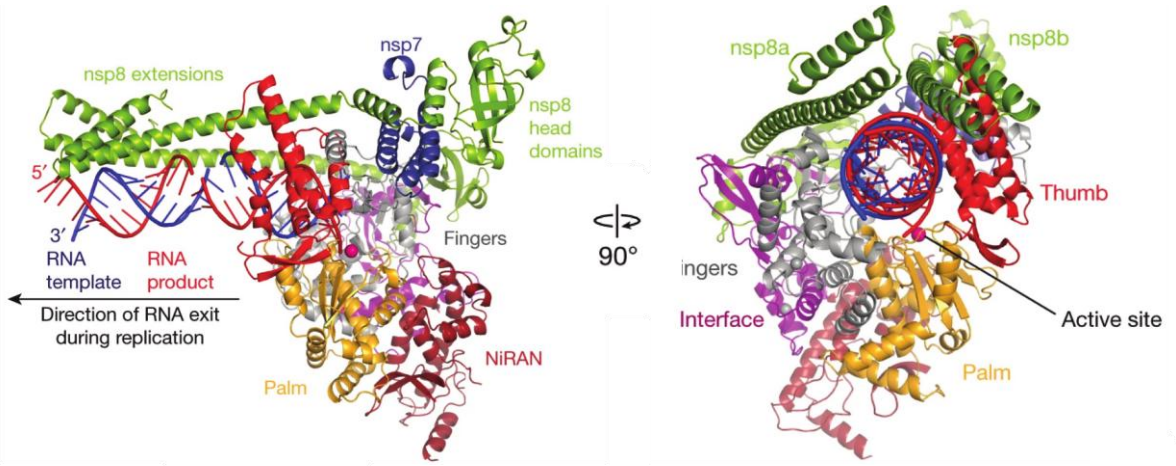


Figure 3: Structure of replicating SARS-CoV-2 polymerase

adopted from Hillen, H.S., Kokic, G., Farnung, L. et al[99], Colour code for nsp12 (NiRAN, interface, fingers, palm and thumb), nsp8, nsp7, RNA template (blue) and RNA product (red). The magenta sphere depicts a modelled[102] metal ion in the active site.

Besides nsp12, the accessory subunits nsp7 and nsp8 are needed for efficient RNA replication. Nsp7 is a single domain protein, while nsp8 is comprised of a long N-terminal protrusion domain and a C-terminal head domain. The structure of RdRp bound to dsRNA revealed that nsp12 binds to one molecule of nsp7 and two nsp8 molecules, which stabilize the interaction between RdRp and RNA.[98] $Nsp8_F$ (F stands for finger) binds to the finger domain of nsp12, and $nsp8_T$ (T stand for thumb) is locates on the opposite site of the RNA binding to the thumb domain and nsp7.[103] The N-terminal helical extension of nsp8 contact the emerging RNA duplex up to 28 bp upstream of the active site. Thus it was proposed that nsp8 act as sliding poles, helping replicating RdRp to stay engaged with RNA.[99]

The solved cryo-EM structures also helped to understand the structural basis for substrate selection and incorporation of the correct nucleotide.[104] RdRp holds approximately one turn of duplex RNA, composed of template and product strand, with the first single-stranded base of the template strand residing in the +1 site. This allows incorporation of the according nucleotide into the product strand and is termed post-translocated state. Once the correct nucleoside triphosphate (NTP) enters the NTP binding site, it is stabilized by Watson–Crick base pairing with the templating nucleotide and interactions with the two catalytic metal ions. This causes a conformational change that promotes a nucleophilic attack of the 3´OH of the product strand on the NTP α-phosphate. Pyrophosphate is released in a S_N2 condensation reaction, and the incoming nucleotide is covalently linked to the product strand, which reflects a typical two-metal ion mechanism

seen in many nucleic acid polymerases. Selection of the correct nucleotide is achieved by geometric constrains. Only when the geometry of the Watson–Crick base pairing is correct, the conformational change and the condensation reaction will happen. After completion of the condensation reaction, the +1 site is occupied with the incorporated nucleotide, a state termed pre-translocated. The RdRp needs to translocate to free up the +1 site and allow incorporation of the next nucleotide.[99,103,104]

The described condensation reaction requires the presence of a hydroxyl group at the 3´end of the product strand and is therefore only possible during the elongation phase.[93] Priming, which describes the process of starting transcription *de novo*, has to happen before the nucleotide addition cycle can start. Priming is still not fully understood for SARS-CoV-2, and several partially contradictory biochemical models have been published.[60,61,105] It is generally well accepted that the conserved secondary RNA structures found in the 5´and 3´UTR of the genome are important for priming. Apart from that two models emerged, which would both explain observed behavior of RdRp, and they are currently debated in the scientific community.[60] The first model suggests NiRAN dependent protein-priming. It is known that the NiRAN domain acts as nucleotidyl transferase, which would allow NMPylation of proteins. Indeed, biochemical data suggests that the NiRAN domain can covalently link nsp8 and UTP, creating a product that could be used to initiate RNA synthesis at the poly-A tail. The second model describes NiRAN independent priming and builds on *de novo* synthesis of dinucleotides by RdRp.[60] Both models are plausible, and it is possible that initiation of minus and plus strand synthesis requires different priming reactions.

4.4.2. Backtracking by RdRp – nsp13 complexes

Nsp13 is a dual-function protein that harbors a helicase and exhibits triphosphatase activity. The helicase is NTP-dependent and can unwind DNA and RNA in 5´to 3´direction.[93] The RNA 5'-triphosphatase activity can catalyze the removal of a 5' γ-phosphate group of pppA-RNA in order to generate ppA-RNA. This is the first step of RNA capping and suggests that nsp13 is involved in the capping process of SARS-CoV-2. The helicase activity was shown to be involved in several aspects of the viral life cycle and is essential for corona viruses.[106] Intriguing details about the molecular function of nsp13 were revealed by a high resolution cryo-EM structure of RdRp stability bound to nsp13.[67] RdRp binds to two helicase molecules, $nsp13_F$ and $nsp13_T$, which sit on top of the RNA binding site and interact with the N-terminal extensions of $nsp8_F$ and $nsp8_T$, respectively. The RNA binding channel of

$nsp13_T$ is directly in the path of downstream template RNA. Surprisingly, both nsp13 are oriented in way that the unwinding activity in 5´to 3´direction would pull the template RNA out of RdRp. This suggests that nsp13 can induce backtracking, a process well known from DNA polymerases. In case of the RNA replication and transcription by SARS-CoV-2, backtracking is proposed to be involved in the proofreading mechanism (discussed in the next chapter) and in template-switching during sgRNA synthesis.[107]

Discontinued RNA transcription describes the synthesis of a set of nested 3' and 5' co-terminal sgRNAs. It is a hallmark process of Nidoviruses and is dependent on the presence of transcription regulatory sequences (TRSs).[90] These sequences are located in several positions in the genome body (TRS-B) usually upstream of an ORF and in the leader sequence (TRS-L) at the 5´UTR. During negative sense strand RNA synthesis, the RTC transcribes one or several ORFs until it encounter a TRS-B which induces a pause. The RTC then scans the RNA strand for the TRS-L sequence, which is complementary to the TRS-B sequence. Once found, the TRS sequences can base pair and the RTC continues with transcription at UTR.[93] The degree of complementarity varies between different TRS-Bs which was shown to be a mechanism to control the amount of transcription of the according ORF.[108] The described template-switching step can span several kb of the genome, which raises the question how the RTC can accomplish this jump efficiently. The helicase activity of nsp13 would be well suited to help RdRp to move along the genome.[93] However, the exact mechanism remains elusive, and more studies are needed to explain the role of nsp13 in sgRNA synthesis.

4.4.3. Proofreading by nsp10 and nsp14

Proofreading describes the detection and removal of mismatched nucleotides during the replication of genetic material. It is therefore a quality control step that prevents the accumulation of mutations, which is especially important if large amounts of genetic material have to be amplified.[109] Proofreading is well-known for its crucial role in cellular RNA and DNA polymerases[110], but most RNA viruses do not harbor a proofreading function.[69,111] SARS-CoV-2, and the family of corona viruses in general, present an exception because they harbor a proofreading function crucial for their viral life cycle.[109]

Errors during RNA synthesis by RdRp lead to mismatched nucleotides that are unable to form a proper Watson-Crick base pair. This changes the geometry inside the active center and presumably induces backtracking of RdRp. The mismatched nucleotide can be then

removed by the 3'-to-5' exoribonuclease (ExoN) function, residing in the nsp10–nsp14 complex of SARS-CoV-2.[107,112] Nsp14 is a dual-function protein that contains a C-terminal N7-MTase domain involved in RNA capping and the N-terminal ExoN domain. The activity of the ExoN domain is controlled by the presence of the cofactor nsp10. If nsp14 is stably bound to nsp10, its exoribonuclease activity is 35-fold increase compared to the unbound state.[63,70] A high resolution cryo-EM structure of the nsp10-nsp14 complex bound to dsRNA showed that nsp10 stabilizes the active site of the ExoN domain, explaining the strong stimulatory effect.[65] In addition, it was shown that the ExoN can use dsRNA or ssRNA as substrate for the hydrolysis, since it only requires a free 3´OH to initiate a nucleophilic attack, resulting in phosphodiester bond cleavage.[65]

Characterization of mutant corona viruses that harbor an active site mutation in nsp14 causing a loss of the ExoN function have revealed severe negative effects on genome stability and less viable viral particles.[112] This underlines the importance of the proofreading mechanism for the life cycle of SARS-CoV-2. Besides that, ExoN has direct implications for the treatment of COVID-19 patients because it renders most nucleoside inhibitors of RdRp causing chain termination[112] inactive (discussed in more detail in chapter *4.5.1. Nucleoside* analogues*)*

Summarized, ExoN mediated proofreading by SARS-CoV-2 is an important process in the viral life cycle with implication for antiviral drug development. However, mechanistic insights into the spatial and temporal regulation of proofreading are still lacking.[93] It remains elusive how the nsp10-nsp14 is regulated to prevent undesired RNA degradation and how it is recruited to RdRp in order to remove mismatched nucleotides. Attempts to study the interaction biochemically and structurally suggest that the nsp10-nsp14 can interact with RdRp, but no stable complex or conclusive structure could be obtained.[113]

Capping complexes of SARS-CoV-2

RNA capping is defined as process that modifies the 5´end of primary RNA transcripts. It involves several steps of chemical reactions and is well known for its crucial function in eukaryotic cells.[114] The cap protects the 5´ end of RNA from exonuclease digestion and allows proper translation of the RNA during protein synthesis.[115] Similar to human mRNAs, the RNA genome of SARS-CoV-2 is capped with a $^{7Me}GpppA_{2'\text{-}O\text{-}Me}$ structure. This means that the 5´end is comprised of a methylated guanosine linked to the first nucleotide of the RNA through a reverse 5' to 5' triphosphate linkage and that the ribose of

the first nucleotide is methylated.[93] It has been shown that the 2'O methylation of first nucleotide serves as an important discriminatory signal of non-self for the innate immune system. Thus, the RNA cap of SARS-CoV-2 is also important for host immune evasion.[116]

RNA capping is an orchestrated process in cells that involves several enzymes to catalyze the chemical alteration of the 5´ RNA end.[114] In contrast to other RNA viruses such as Influenza, which hijacks the cellular capping pathway, SARS-CoV-2 brings its own capping machinery to the cell.[117] Based on a recent high resolution cryo-EM structure of nsp9 bound to the NiRNA domain of RdRp, a convincing and almost complete model of RNA capping by SARS-CoV-2 was proposed.[62] During transcription, nsp9 and the 5'-triphosphorylated RNA end of the nascent transcript bind to the NiRNA domain of nsp12. There, covalent linkage of nsp9 to 5´RNA is catalyzed, and a pyrophosphate is released. A GDP molecule, potentially generated by nsp13, is bound and attacks the RNAylated nsp9, which releases the GpppA capped RNA and unmodified nsp9. Methylation of the guanosine and the 2'O of the first nucleotide are then catalyzed by the N7-Mtase domain of nsp14 and the 2'-O-Mtase of nsp16, respectively.[62]

Hence, SARS-CoV-2 can catalyze the formation of a cap structure independent of the host cell, making it difficult for the cell to fight the infection.[115]

Nsp15 and the formation of a super-complex

In the previous chapters, the function of all nsps involved in the transcription and replication of the viral RNA were discussed with one exception, nsp15. The function of nsp15 and its potential involvement in the RTC remain mostly enigmatic.[93] It is described as a nidoviral RNA uridylate-specific endoribonuclease (NendoU) with an N-terminal oligomerization domain, a central domain and a C-terminal catalytic domain.[72] A high resolution cryo-EM structure of nsp15 bound to dsRNA revealed the presence of nsp15 hexamers and molecular details of dsRNA recognition and processing.[73] The uridine-specific endoribonuclease activity allows nsp15 to cleave RNA intermediates, which lowers the risk of detection by the host cells.[118] Nsp15 is highly conserved across coronaviruses, which suggests an important role in the viral life cycle. However, viruses that expresses a catalytically inactive nsp15 can still replicate and propagate. Thus, the question emerges whether nsp15 has an additional important role beyond cleavage of RNA intermediates.

Related to this question, an intriguing model of a large RTC assembly around the nsp15 hexamer was proposed.[81] The study used molecular modeling tools to investigate

potential interactions between individual nsps and the N protein. The authors propose a large hexameric complex that is composed of six subunits each of nsp12, nsp13, nsp14, nsp15, nsp16; 12 subunits of nsp8, nsp10 and two subunits of the N protein. The model fits well to solved cryo-EM structures of RTC subcomplexes and would explain key processes during viral RNA processing such as proofreading, capping and template switching.[81] However, biochemical and structural studies are needed to prove the existence of such a large complex and to investigate the role of nsp15 in RTC formation.

4.5.Antivirals

After the first incidences of COVID-19 were reported, it quickly became clear that SARS-CoV-2 poses a major risk to human health.[1] Determining the pathogenicity of SARS-CoV-2 was difficult at the time of emergence, but now in retrospect, several studies were able to investigate the hospitalization rate and lethality during the first two waves of COVID-19.[119,120] Approximately 10% of people who were infected with SARS-CoV-2 had to be admitted to hospital, and 30% of these required intensive care. Despite standard medical care, which includes symptomatic treatment, 1 – 2% of all infected patients passed away. This number increases significantly for patients with risk factors such as high-age, obesity or immune deficiency.[120] Besides the severity of COVID-19, it became clear that SARS-CoV-2 is highly infectious with a basic reproduction number (R_0) of 2.8 – 3.8.[121-123] This is comparably high and means that without intervention the virus spreads very quickly in an immunologically naïve population. Non-pharmaceutical interventions, such as lockdowns and social distancing, helped to reduce the spread of SARS-CoV-2.[27] But without a doubt, there was and still is a strong global need for potent therapeutic interventions against SARS-CoV-2.

There are several potential therapeutic strategies to prevent SARS-CoV-2 from replicating in the human body, including antivirals, vaccines and neutralizing antibodies. Antivirals and their molecular mechanisms of inhibition are the focus of this doctoral thesis. Other treatment modalities are beyond the scope of this thesis and are therefore only briefly mentioned in the discussion, chapter *7.3.1. Do we still need antivirals against corona viruses?*

Antivirals aim to stop viral propagation by interference with the viral life cycle.[124] Inhibition of essential viral enzymes will prevent the formation of key viral building blocks,

which impedes the assembly of new viral particles. Based on our understanding of the biology of SARS-CoV-2, two steps of the viral life cycle were identified as most promising for the purpose of drug development:[125] First, polyprotein cleavage by the viral protein, and second, RNA synthesis by the RdRp.[126] The enzymatic functions of other nsps, such as the helicase domain nsp13 or the exonuclease domain nsp14, could also be potential drug targets. However, their biological roles are less well understood, and it remains unclear whether inhibition would be detrimental for viral replication.

Proteolytic cleavage of the polyprotein is an essential step in the viral life cycle, and inhibition of one of the proteases, residing in nsp3 and nsp5, with small molecules, prevents propagation of the virus.[125] Paxlovid with the active pharmaceutical ingredient Nirmatrelvir (also termed PF-0732133), is a competitive peptidomimetic inhibitor of the 3Clpro enzyme. It is highly potent and FDA approved for usage in patients.[127] Paxlovid is supplied with Ritonavir as a second pharmaceutical active ingredient, which targets and irreversibly blocks a cytochrome P450 enzyme. Ritonavir is needed to prevent the oxidation of Nirmatrelvir but has severe side effects preventing long treatment duration.[128,129]

Another highly promising drug target is the RTC with RdRp at its core enzyme. Inhibition of RdRp prevents the replication of the genetic information and impedes the propagation of the virus.[126] In addition, RdRp is highly conserved across corona viruses, and it has no close homologues in human cells, which would allow the development of a highly selective pan-corona inhibitor. Thus, targeting the RdRp with small molecular inhibitors emerged as one the most promising strategies for the development of antivirals against SARS-CoV-2.

Based on the chemical structure, RdRp inhibitors can be broadly classified into nucleoside and non-nucleoside inhibitors.[130] A selection of the most relevant promising molecules of each class are summarized in the following two chapters.

4.5.1. Nucleoside analogues

Nucleoside inhibitors, also referred to as nucleoside analogues (NAs), mimic the naturally occurring substrate of polymerases, nucleoside triphosphates (NTPs), or their precursors.[131] Promiscuous polymerases, such as viral RNA polymerases, will incorporate nucleoside analogues into the growing RNA chain where they can either cause chain termination, stopping the elongation of the nascent RNA, or induce mutations that lead to a loss of genetic information.[132] NAs are usually administered as prodrugs because their

incorporation ready triphosphorylated state lacks chemical stability and membrane permeability. Cellular enzymes involved in the natural metabolism of NTPs are utilized to convert the prodrug into its triphosphate metabolite, which can be used as substrate by polymerases.[133]

NAs are a promising class of drugs with more than 30 FDA approved molecules successfully used against cancer and viral infections.[134] Soon after the emergence of COVID-19, researchers tried to repurpose already developed and clinically tested NAs for usage against SARS-CoV-2. If successful, repurposing of existing drugs can lead to significant cost and time savings.[135,136] However, SARS-CoV-2 is resistant to many NAs because of its proofreading mechanism. NAs that cause chain termination will create stalled RdRp complexes, which are rescued by exonucleolytic removal of the NA by the nsp10-nsp14 complex, rendering most NAs ineffective against SARS-CoV-2.[112]

Nevertheless, there are potent NAs which are resistant to the proofreading mechanism due to a sophisticated mode of action. The chemical structures of promising NAs with potent inhibitory effects against SARS-CoV-2 are depicted in Fig 4. Due to a vast number of publications that claim the discovery of small molecules with antiviral activities, the list is incomplete. It only contains molecules that are FDA approved or where mechanistic insides are published.

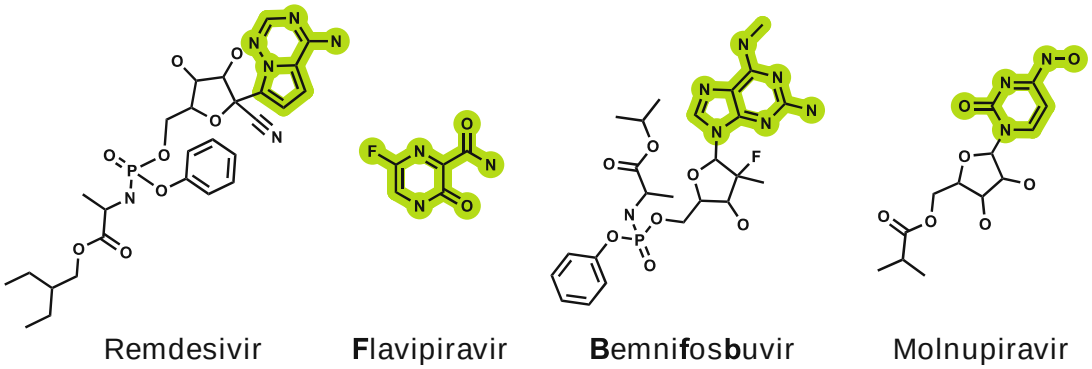


Figure 4: Promising nucleoside analogues against SARS-CoV-2 RdRp

summary of published nucleoside analogues against SARS-CoV-2 RdRp; nucleoside-like moieties highlighted in green

Remdesivir

Remdesivir is a 1'-cyano-modified adenosine nucleotide analogue that was initially developed against Ebola virus. Later, it was shown that it is also effective against SARS-CoV-2 RdRp.[137] A high resolution cryo-EM depicting the active triphosphate metabolite of Remdesivir incorporated into the product RNA of RdRp revealed how this NA circumvents the proofreading mechanism despite stalling RdRp.[138] Remdesivir is a delayed chain termination that, once incorporated, will allow the addition of three more nucleotides before the cyano-group clashes with the sidechain of serine 861 of RdRp, causing stalling. Remdesivir was the first FDA-approved NA drug against SARS-CoV-2.[139] However, it has to be administered shortly after the infection onset via intravenous injection, which requires medical attention and prevented extensive usage. In addition, its clinical efficacy is low and heavily debated in the scientific community.[140]

Flavipiravir

Flavipiravir is a derivative of pyrazine carboxamide that was originally developed to combat influenza viruses.[141] It is approved in Japan for the treatment of influenza patients and was shown to be also effective against numerous other viruses including Ebola and SARS-CoV-2. Despite the existence of a high-resolution cryo-EM structure showing the active metabolite of Flavipiravir bound to the active site of RdRp, its exact mode of action is still debated.[142] Some studies concluded that Flavipiravir is a mutagenic NA while others claim it terminates RNA extensions.[143] Its clinical usage was tested, but a recent review suggests that treatment of COVID-19 patients with Flavipiravir has no significant benefits.[144]

Sofosbuvir and Bemnifosbuvir

The active metabolites of Sofosbuvir and Bemnifosbuvir are uracil and guanosine nucleotide analogues carrying a 2'-fluoro-2'C-methyl modified ribose, respectively.[131] A high resolution cryo-EM structure of the active metabolite of Bemnifosbuvir (AT-527) revealed a dual inhibition mechanisms of RNA replication.[105] First, AT-527 acts as an immediate chain terminator that is to some extent resistant against the proofreading mechanism of the RTC. Second, AT-527 is stably bound to the NiRNA domain of nsp12, inhibiting its nucleotidyl transferase activity. Based on promising data from cell culture experiments, its inhibitory potential against SARS-CoV-2 was assessed in a clinical trial. It reached phase III but was

then discontinued because of better available treatments and a lack of clear benefits for patients.[145]

Molnupiravir

Molnupiravir is an orally available prodrug that is converted into β-D-N4-hydroxycytidine (NHC) triphosphate.[10] It interferes with the replication of various viruses including SARS-CoV-2.[12,146] Studies conducted with the Venezuelan equine encephalitis virus and Influenza virus have shown that exposure to NHC can introduce mutation into the viral RNA.[10,147] By the time this thesis was started, Molnupiravir was in a phase III trial for treatment of COVID-19 patients.[9] However, a molecular understanding of its mode of action was lacking. We were able to establish the molecular mechanisms underlying Molnupiravir-induced RNA mutagenesis, described in detail in chapter *5. Manuscript 1: Mechanism of Molnupiravir-induced SARS-CoV-2 mutagenesis*. Since then, numerous studies were published raising questions about the efficacy and safety of Molnupiravir, suggesting that clinical benefits are small if existent.[148] A discussion about recent developments regarding Molnupiravir is provided in chapter *7.1. Advantages and drawbacks of Molnupiravir-induced mutagenesis.*

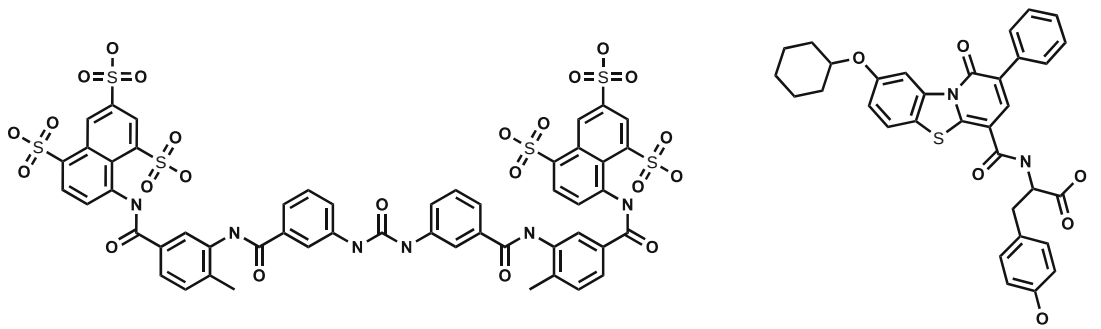
Other potential NA against SARS-CoV-2 RdRp

There are other NA molecules with potential inhibitory effects on the RdRp of SARS-CoV-2, such as Ribavirin, Tenofovir or Galidesivir.[131] However, all these molecules are not well understood and lack potency or selectivity. NAs are in general a promising class of small molecules for the development of potent antivirals, but per definition, their chemical nature is limited by the similarity to NTPs.[131] Related to this, the development of NAs against SARS-CoV-2 harbors several challenges, including NA bioavailability, selectivity and resistance against the proofreading mechanism.[8]

4.5.2. Non-nucleoside analogues

Non-nucleoside inhibitors are structurally distinct from NTPs and usually have a binding site outside the NTP binding pocket of the polymerase. Thus, such molecules are not constrained by their similarity to nucleosides and are not affected by the proofreading mechanism of SARS-CoV-2.[8] Because of that, non-nucleoside inhibitors could help to address the strong need for orally available, potent and safe RdRp inhibitors. Several examples of recently developed non-nucleoside inhibitors against the polymerases of other viruses can be found in literature, such as Baloxavir marboxil against Hepatitis C Virus or Doravirin against Human immunodeficiency virus.[149] Unfortunately, despite global efforts to find suitable small molecules against the RdRp of SARS-CoV-2, promising non-nucleoside inhibitors are still lacking.[8]

Many *in silico* studies were published that screened billions of small molecules in virtual compound libraries for their potential to bind to RdRp. inhibitors.[150-157] This led to the nomination of numerous potential RdRp inhibitors, but very few of these molecules showed an inhibitory effect in reproducible *in vitro* assays. Even for those compounds that were tested in biochemical assays, results should be interpreted with caution, because compounds were tested at very high concentrations, exceeding the limit of solubility, or poorly active RdRp was used. Suramin and HeE1-2Tyr are the only published non-nucleoside inhibitors that significantly inhibited RdRp in our assays at concentrations below 40 µM.



Suramin HeE1-2Tyr

Figure 5: Promising non-nucleoside analogues against SARS-CoV-2 RdRp

Suramin

Suramin is large molecule with a molecular weight of 1300 Dalton that was already synthesized in the 1920s.[158] Since then it is used as antiprotozoal agent in order to treat African sleeping sickness and river blindness.[159,160] Later it was shown that Suramin has antiviral effects against RNA viruses including SARS-CoV-2.[161] A high resolution cryo-EM structure visualized two binding sites of a partial Suramin molecule near the RNA binding site of RdRp and showed biochemical data supporting RNA displacement.[161] However, Suramin is known for its promiscuity due to the size of the molecule and the presence of six sulfonic acid groups.[162-164] Therefore, its properties as a small molecule probe are questionable, and clinical usage against SARS-CoV-2 is impossible.

HeE1-2Tyr

HeE1-2Tyr is a pyridobenzothaizol inhibitor that was initially discovered against dengue virus RdRp.[14] Later, it was also shown to be effective against SARS-CoV-2 RdRp *in vitro* and against SARS-CoV-2 replication in cell culture experiments.[15] However, it remained unclear how HeE1-2Tyr facilitates inhibition of the SARS-CoV-2. Combining biochemical and structural data, we were able to elucidate the molecular mechanism of HeE1-2Tyr-mediated SARS-CoV-2 RdRp inhibition. HeE1-2Tyr is a competitive RdRp inhibitor that prevents polymerase activity by displacing the RNA from RdRp, described in detail in chapter *6 Manuscript 2:* and further discussed in chapter *7.2 HeE1-2Tyr: A novel mode of action to inhibit RdRp*.

5. Manuscript 1: Mechanism of Molnupiravir-induced SARS-CoV-2 mutagenesis

5.1. Abstract

Molnupiravir is an orally available antiviral drug candidate currently in phase III trials for treatment of COVID-19 patients. 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-N4-hydroxycytidine (NHC) triphosphate, as a substrate instead of CTP or UTP. 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 containing 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 likely applies to various viral polymerases and can explain the broad-spectrum antiviral activity of Molnupiravir.

5.2. Introduction

Coronaviruses use an RNA-dependent RNA polymerase (RdRp) for the replication and transcription of their RNA genome[5,32,165-167]. The RdRp is an important target for the development of antiviral drugs against coronaviruses[59,168-170]. Structures of the RdRp were reported for SARS-CoV-1[58] and for SARS-CoV-2[99,171-174] and provided insights into the mechanisms of RNA-dependent RNA synthesis[103]. The structures also enable mechanistic studies that can rationalize the molecular processes underlying the antiviral activity of compounds targeting the 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[70,112,175]. The nucleoside analogue remdesivir can circumvent proofreading because

its incorporation does not terminate elongation but only stalls RdRp after the addition of three more nucleotides[137,138,173,176,177]. Remdesivir was the first FDA-approved drug for the treatment of COVID-19 patients[139,178-180], but its effectiveness is disputed[181], emphasizing the need to develop new antiviral drugs.

Another promising drug candidate for the treatment of COVID-19 patients is Molnupiravir (or EIDD-2801), which also targets the RdRp of SARS-CoV-2. Molnupiravir is an isopropylester prodrug of the nucleoside analogue β-D-N4-hydroxycytidine (NHC or EIDD-1931)[10,182]. Molnupiravir interferes with the replication of various viruses[10,182-187] including SARS-CoV-2[12,146]. Molnupiravir inhibits SARS-CoV-2 replication in human lung tissue[13], blocks SARS-CoV-2 transmission in ferrets[11], and reduces SARS-CoV-2 RNA in patients[9]. 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[188] for safety, tolerability and pharmacokinetics, and phase II/III studies are currently ongoing[9] (NCT04405739, NCT04405570, NCT04575597). Available data suggest that Molnupiravir acts as a mutagenizing agent that causes an 'error catastrophe' during viral replication[10,12,189]. Indeed, NHC can introduce mutations into viral RNA, as shown for Venezuelan equine encephalitis virus[190]. Also, sequencing of influenza virus populations indicated that NHC caused G-to-A and C-to-U transitions in viral RNA[10], and the same transitions were found for SARS-CoV-2[12].

Despite this progress, a systematic biochemical and structural analysis of Molnupiravir- or NHC-induced RNA mutagenesis by viral RNA polymerases is lacking. Here we quantify the effects of Molnupiravir/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 analogues.

5.3. Results

5.3.1. 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. 6a, 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 (Appendix), which directs binding of the incoming nucleoside triphosphate (NTP) substrate (Fig. 6c). The RNA product strand contained a fluorescent label at its 5'-end that allowed us to monitor and quantify RNA elongation.

When the 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. 6d, e). Time-dependent RNA elongation experiments showed that M was slightly less efficiently incorporated than the cognate nucleotide C (Fig. 6f). M incorporation opposite of A was also observed but substantially reduced compared to incorporation of the cognate nucleotide U (Fig. 6g). 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[191] that were predicted to allow for base pairing with either G or A[192].

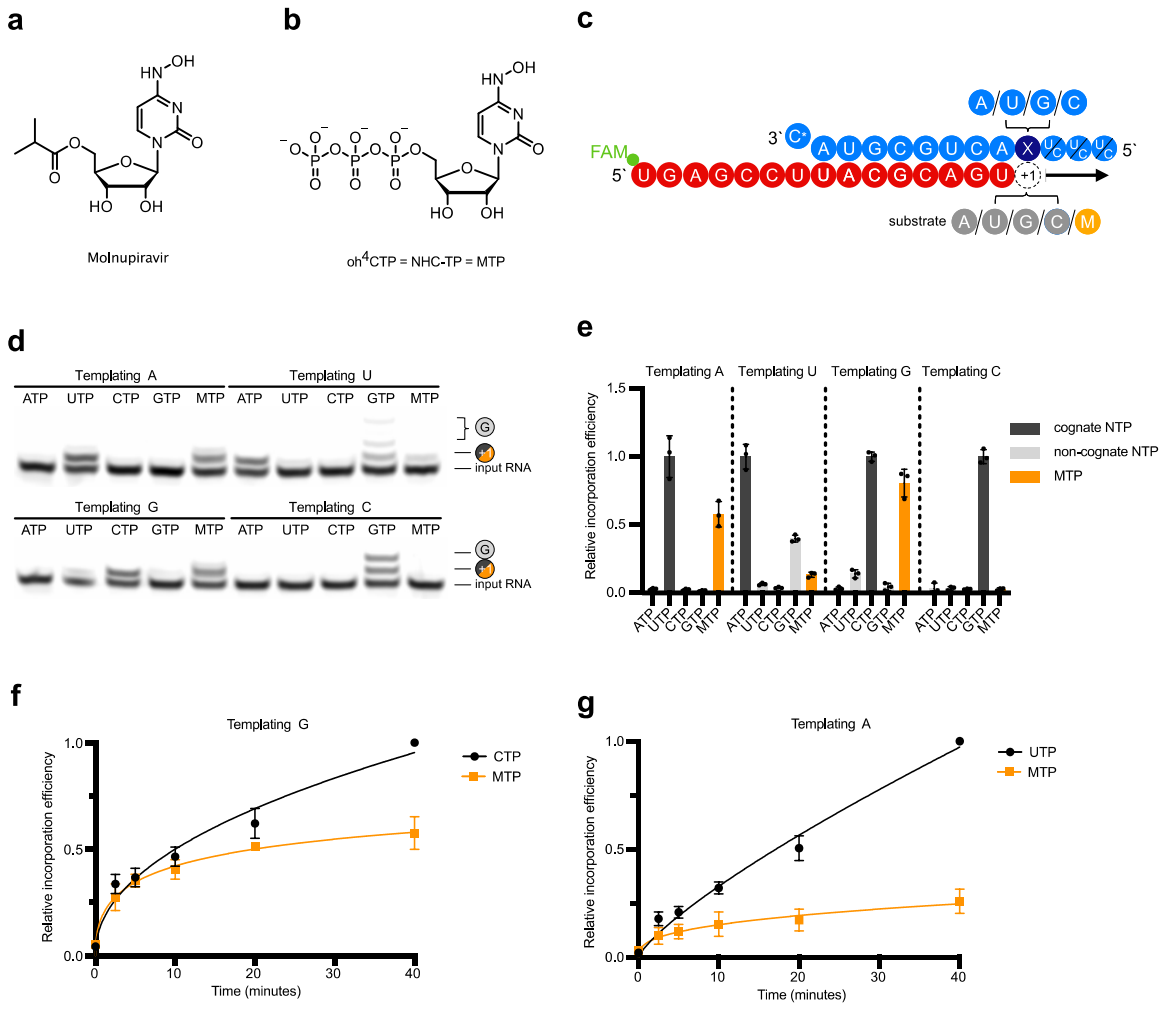


Figure 6: RdRp incorporates NHC opposite G and A in the template

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.

5.3.2. 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. 7a). We observed that incorporation of M instead of the cognate C or U did not prevent incorporation of three subsequent nucleotides (Fig. 7b). Further, we tested RNA extension with a scaffold that allowed for the incorporation of 11 nt (Fig. 7c). Also in this case the RdRp reached the end of the template when UTP or CTP were replaced by MTP, although again incorporation of M instead of U was less efficient than incorporation of M instead of C (Fig. 7d). 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 what happens when M-containing RNA is used as a template in a second step of RNA synthesis.

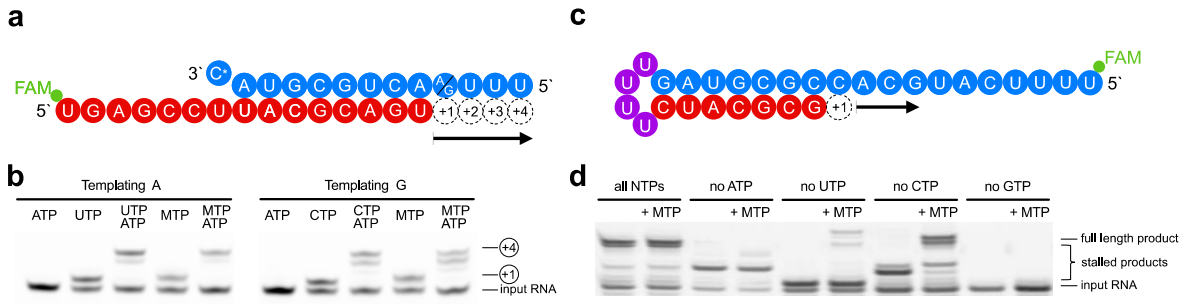


Figure 7: NHC incorporation does not stall SARS-CoV-2 RdRp

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.

5.3.3. 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 by the convertible nucleoside approach from a ribose-protected O^4-chlorophenyluridine (Fig. 8a, Methods, Appendix). The presence of M in the obtained RNA and RNA purity were confirmed by denaturing HPLC and HR-ESI-MS (Fig. 8b). The M-containing RNA oligo was annealed with a fluorescently labeled product RNA such that the M nucleotide occupied the templating position +1 (Fig. 8c, Appendix).

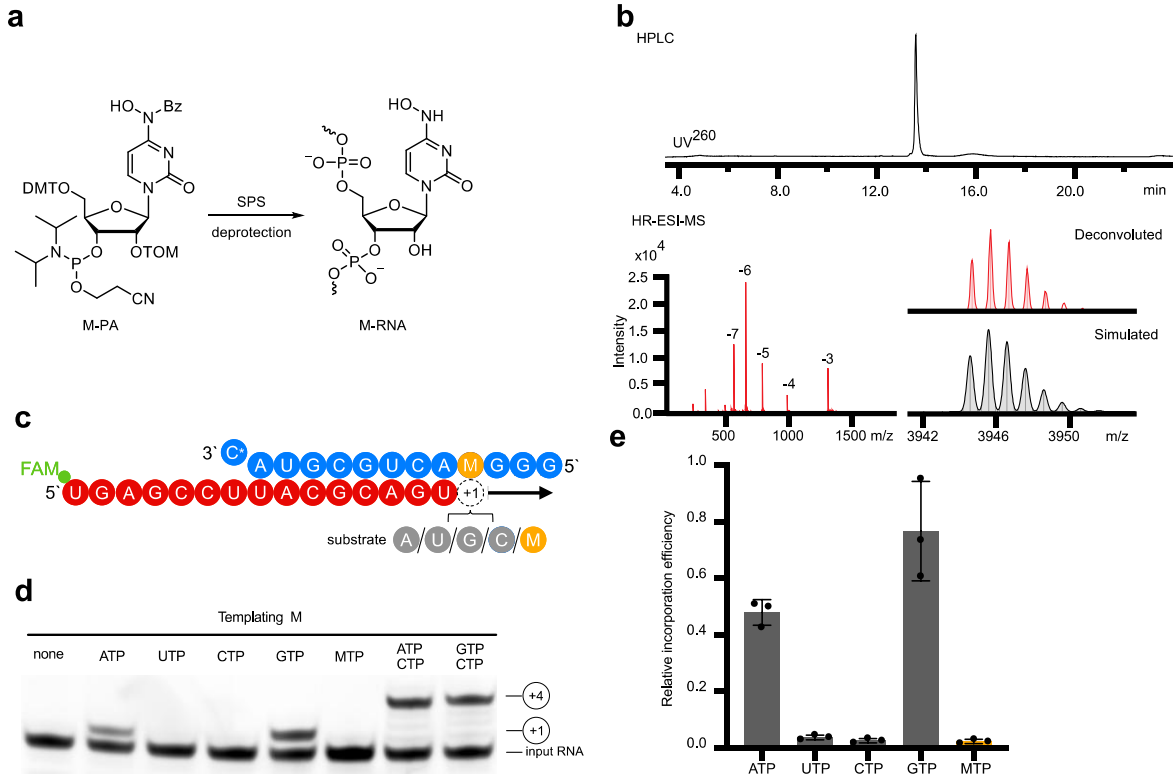


Figure 8: NHC can direct incorporation of G and A into RNA

a, Scheme of synthesis of RNA containing NHC monophosphate (M) at a defined position. 5′-O-DMT-2′-O-TOM-protected *N*4-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.

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. 8d, 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 (Appendix). 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.

5.3.4. 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 (Appendix), formed RdRp-RNA scaffold complexes and subjected these to cryo-EM analysis as described[99].

We indeed obtained RdRp-RNA structures that contained either a M–A or a M–G base pair at position –1 (Fig. 9, Appendix). 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. 1). As expected from the scaffold design, the structures showed the post-translocation state with a free NTP-binding site at position +1 (Fig. 9b, c). Comparison of the two structures with each other and with our original RdRp-RNA structure[99] and with remdesivir-containing RdRp-RNA structures[138] 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.

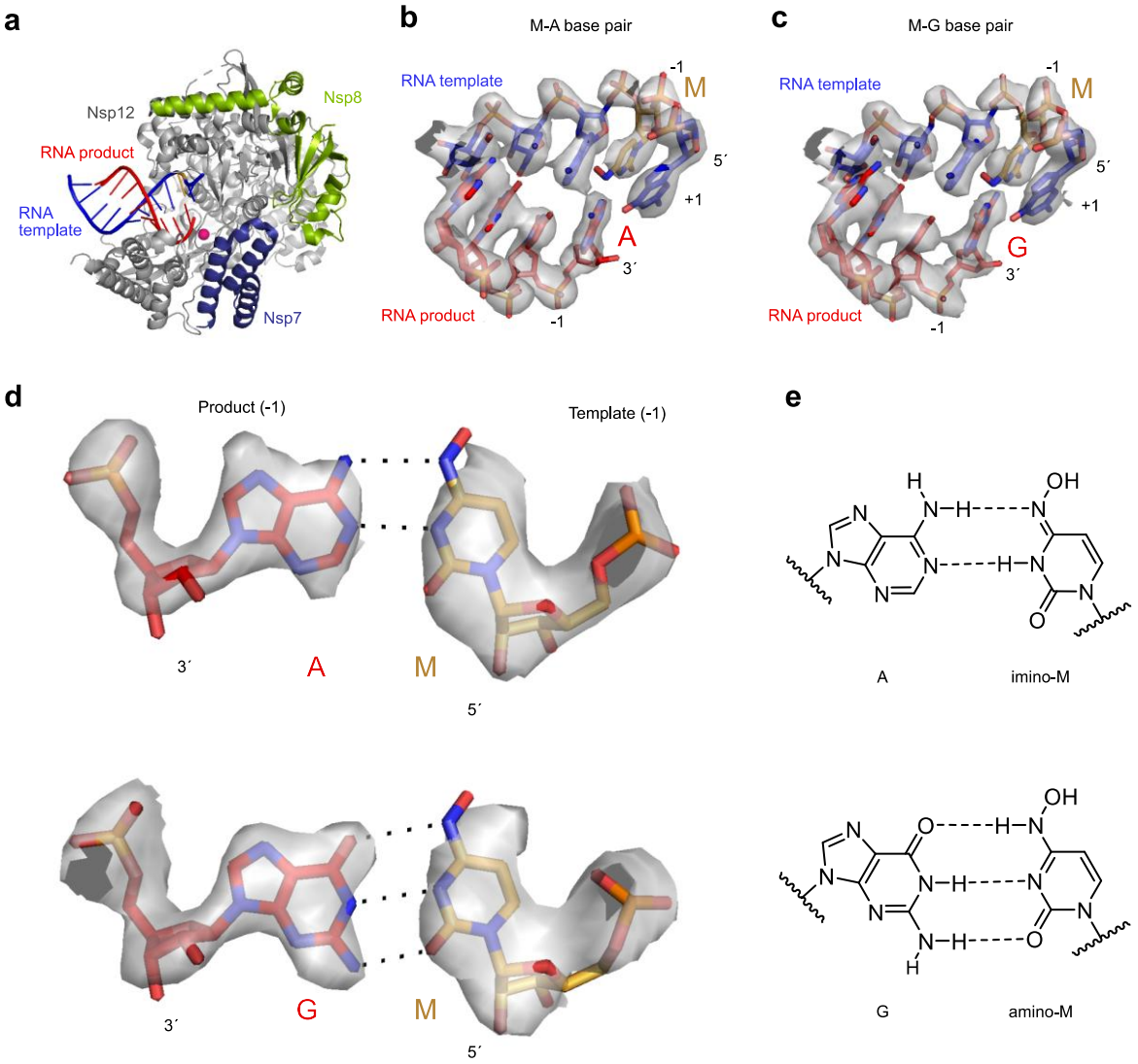
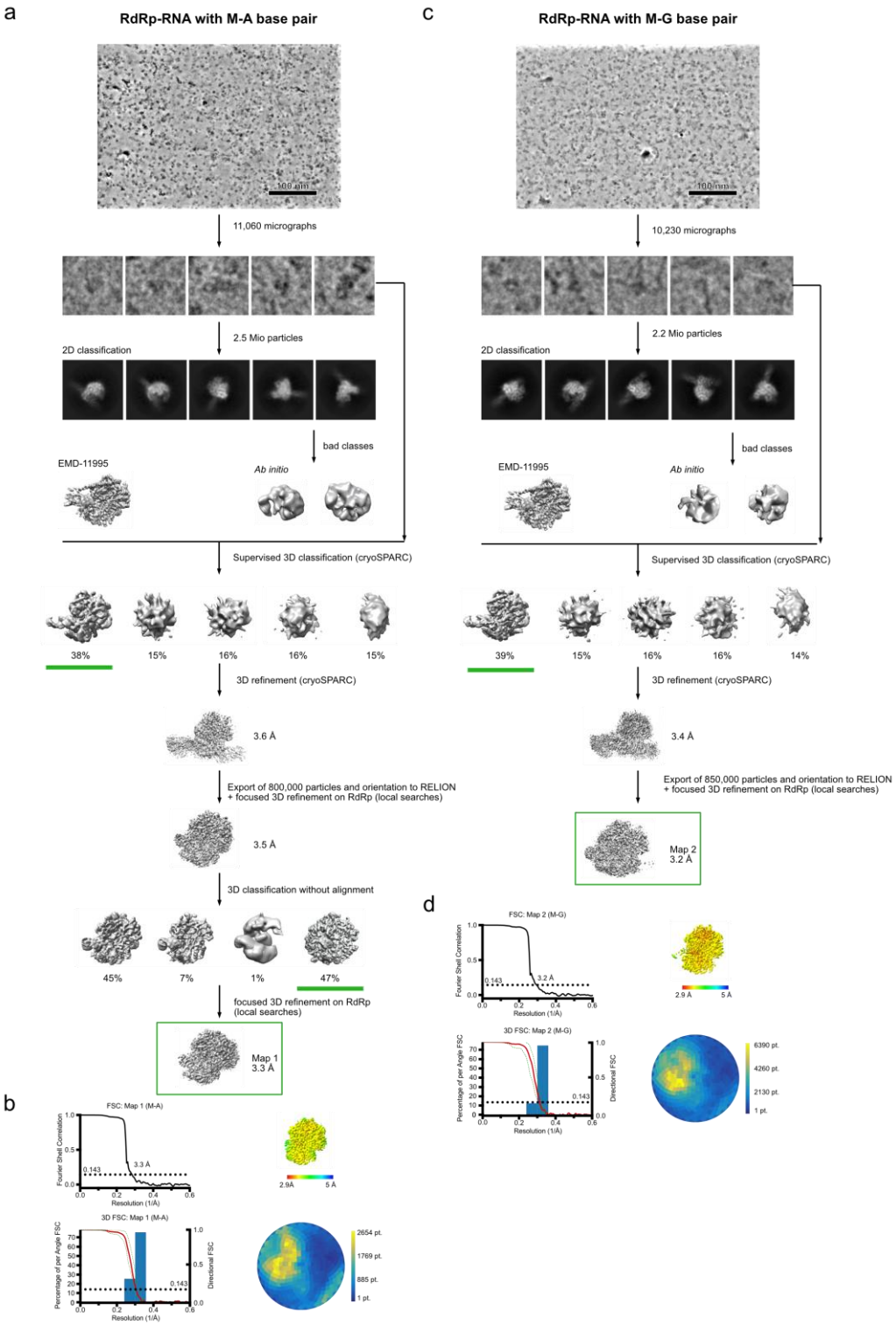


Figure 9: Structures of RdRp–RNA product complexes after NHC-induced mutagenesis

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 NHC45, as predicted[192]



Extended data figure 1: 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.

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. 9b, c). The densities were so detailed that we could clearly distinguish G and A bases (Fig. 9d). The densities were also consistent with the proposed base pairing[192] that is enabled by different tautomeric forms of NHC[191] (**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. 6, 7). 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.

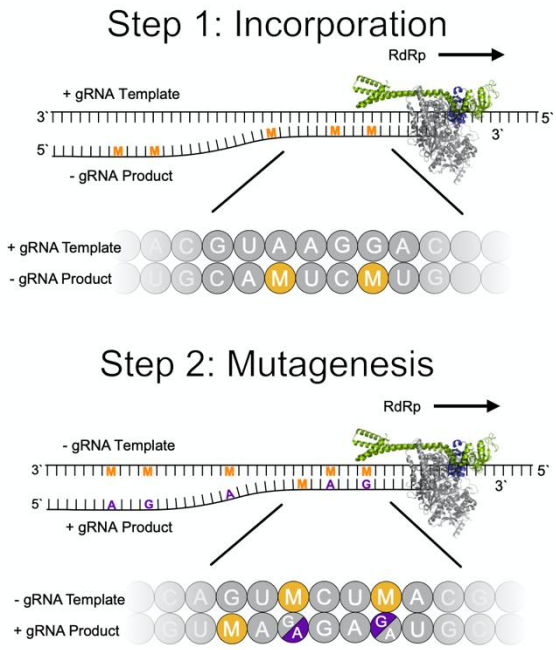


Figure 10: Two-step model of molnupiravir-induced RNA mutagenesis

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.

5.4. Discussion

Our systematic biochemical analysis suggests a two-step model for the mechanism of Molnupiravir-induced coronavirus RNA mutagenesis (Fig. 10). 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[10,12,189].

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 assumed that one stable tautomer predominates in each case, i.e. 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 and explain the observed mutagenic pattern seen in *in vivo* studies of NHC exposed coronaviruses[12]. This antiviral mechanism is conceptually similar to the recently suggested mutagenesis mode of action of favipiravir[193,194], but is entirely distinct from that of remdesivir, which impairs RdRp progression[138]. However, like remdesivir, Molnupiravir escapes viral RNA proofreading because M incorporation and M-directed misincorporation are apparently not recognized by the viral exonuclease[70,112]. Such proofreading escape may also be due to the stability of the M–G and M–A base pairs that are predicted not to induce or favor backtracking of RdRp, which is likely required for exposing the RNA 3'-end to the proofreading exonuclease[67,195].

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[10]. Also consistent with the model, MTP does not inhibit RNA synthesis by hepatitis C polymerase[187] or the RdRp of respiratory syncytial virus[183]. Further, the reverse transcriptase of human immunodeficiency virus can incorporate G or A opposite of NHC that is located in the template[196]. The two-step mutagenesis model resides on the base-pairing properties of NHC that we 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*[197]. In addition, possible mutagenic effects of NHC were recently described in mammalian cells[198]. 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[199] that are consistent with our findings. Based on 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 structural data presented in our study.

5.5. Methods

5.5.1. Protein preparation

Preparation of SARS-CoV-2 RdRp, composed of nsp12, nsp7 and two copies of the nsp8 subunits, was carried out as described[99], 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[99]. 60 hours after transfection with corresponding V1 virus, the cells producing nsp12 with a 6xHis-MBP N-terminal tag were harvested by centrifugation (3000 rpm, 10 min at 4°C) and lysed by sonication in lysis buffer A1 (400 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) 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). Lysate was clarified by centrifugation at 74,766 g for 60 min and ultracentrifugation at 100,000 g at 4°C for 60 min, followed by filtration through 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% (v/v) 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% (v/v) 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 TEV protease overnight during dialysis against buffer D (200 mM NaCl, 25 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 5 mM β-mercaptoethanol). The protein solution was applied to a HisTrap HP pre-packed 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 pre-packed columns (GE Healthcare) equilibrated with buffer D. The unbound protein was concentrated and further purified via

size exclusion chromatography on HiLoad S200 16/60 column (GE Healthcare) in buffer A2 (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 1 mM $MgCl_2$, 1 mM 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[99]. Briefly, both proteins were expressed in *E. 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 isopropyl β-D-1-thiogalactopyranoside at 18 °C for 16 h. After harvesting, cells were resuspended in lysis buffer A1 (400 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) 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 (74,766 g, 4 °C, 30 min). The supernatant was applied to a HisTrap HP column (GE Healthcare), preequilibrated in lysis buffer. The column was washed with high-salt buffer B2 (1 M NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 5 mM β-mercaptoethanol), and with buffer A1 (400 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) 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 column volumes (CV) (B1 buffer: 300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 400 mM imidazole pH 8.0, 3 mM MgCl2 and 5 mM β-mercaptoethanol). The protein sample was dialyzed in buffer D (200 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol and 5 mM β-mercaptoethanol) in the presence of His-tagged TEV protease at 4 °C. The dialyzed sample was subsequently applied to a HisTrap HP column (GE Healthcare), preequilibrated 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 MWCO 10,000 Amicon Ultra Centrifugal Filter (Merck) and applied to a HiLoad S200 16/600 (GE Healthcare) equilibrated in buffer A2 (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 5% (v/v) 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.

5.5.2. RNA extension assays

RNA scaffolds were designed according to published SARS-CoV-2 RNA extension assays[99,138]. All assays except assays shown in Fig. 7c were performed with RNA duplexes comprised of short RNA template and product strands. The 5´ end of the RNA product strands contained a FAM fluorescent label that allowed us to monitor RNA extension. The 3´ end of 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 due to technical reasons.

All unmodified and 5' FAM labelled RNA oligonucleotides (Appendix) were purchased from Integrated DNA Technologies (IDT). MTP was purchased from MedChemExpress and NTPs from ThermoScientific. The assay was performed as described[99], 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 figure panels 6f,g and 7d. The concentration of NTPs was 37.5 μM except for assays shown in figure panels 6f, 6g and 7d, 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% (v/v) 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 incubation at 30˚C the reactions were stopped with 2x stop buffer (7 M urea, 50 mM EDTA, 1x TBE buffer). RNA products were resolved on 20% denaturing polyacrylamide-urea gels in 0.5x TBE running buffer and visualized with a Typhoon 95000 FLA Imager (GE Healthcare Life Sciences). Bands were quantified with Image Lab (Bio-Rad). Obtained data was plotted and analyzed with GraphPad Prism version 9.

5.5.3. Preparation and analysis of NHC-containing RNA oligonucleotides

5'-*O*-DMT-2'-*O*-TOM-O^4-chlorophenyluridine was prepared as described[200], 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 below and NMR spectra of isolated compounds are given in the Supplementary

Information. RNA oligonucleotides were then prepared by solid-phase synthesis on CPG support (0.6 µmol scale) using 2'-*O*-TOM-protected ribonucleoside phosphoramidites (70 mM in CH_3CN) and ethylthiotetrazol (ETT, 250 mM in CH_3CN) as activator, with 4 min coupling time, as previously described[138,200]. The oligonucleotides were deprotected with 25% NH_4OH/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 was analyzed by anion-exchange HPLC (Dionex DNAPac PA200, 2x250 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 a 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, 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.

5.5.4. Synthesis and characterization of M-PA

All reactions were performed under inert nitrogen atmosphere with dry solvents (CH_2Cl_2, CH_3CN). 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. TLC 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 ($CDCl_3$: ^1H = 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. High-resolution (HR) electrospray ionization (ESI) mass spectra (MS) 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 (Appendix).

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 eq.) was dissolved in anhydrous CH_3CN (4 mL) under nitrogen atmosphere. DMAP (174 mg,1.42 mmol, 3.0 eq.) and NEt_3 (331 µL, 2.37 mmol, 5.0 eq.) were added, followed by hydroxyl amine hydrochloride (165 mg, 2.37 mmol, 5.0 eq). After stirring for 21 h at ambient temperature, the reaction mixture was diluted with CH_2Cl_2 and washed with saturated aq. $NaHCO_3$ (2x). The organic phase was dried over Na_2SO_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 eq.) in anhydrous CH_2Cl_2 (4 mL) was treated with DMAP (65.3 mg, 535 µmol, 2.0 eq.) and NEt_3 (149 µL, 1.07 mmol, 4.0 eq.) under nitrogen atmosphere. Benzoic anhydride (59.9 mg, 265 µmol, 0.99 eq.) 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*-(triisopropylsilyloxy)methyl cytidine 3'-cyanoethyl-*N*,*N*-diisopropyl phosphoramidite), compound **2** (150 mg, 176 µmol, 1.0 eq.) was dissolved in anhydrous CH_2Cl_2 (2 mL) and cooled to 4°C. 2-Cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite (67.3 µL, 212 µmol, 1.2 eq.) and 4,5-dicyanoimidazol (23 mg, 194 µmol, 1.1 eq.) were added in two portions within 1 hour. After one additional hour at 4°C, the reaction mixture was allowed to warm up to ambient temperature, and stirred for one more hour. Then, 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.

5.5.5. Cryo-EM sample preparation and data collection

RNA scaffolds for structural studies were prepared by annealing of 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. 1, Appendix). 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 step-wise 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 3-fold molar excess of each nsp8 and nsp7. After 10 minutes 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. An additional 0.2 nmol of the annealed RNA scaffold were spiked into each sample prior to grid preparation. 3 µL of the concentrated RdRp-RNA complex were mixed with 0.5 µl of octyl ß-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 seconds with blot force 5 using a Vitrobot MarkIV (Thermo Fischer Scientific) at 4 °C and 95 % humidity and plunge frozen in liquid ethane.

Cryo-EM data were collected with SerialEM[201] 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 105,000x, resulting in a calibrated pixel size of 0.834 Å/pixel. Images were exposed for a total of 2.0 seconds with a dose rate of 20.5 e^-/px/s, 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[99]. 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[202]. A total of 11,060 and 10,230 movies were collected for M–A and M–G containing structures, respectively.

5.5.6. 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[202] 1.0.9. Particles were imported to cryoSPARC[203] 2.15 and subjected to 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[138]). The class containing RNA-bound RdRp yielded ~800,000 particles that were subjected to homogenous 3D refinement. The refined particles were then exported to RELION 3.1[204] 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[202] 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[138]) as starting model. The model was first rigid-body fitted into the density and then manually adjusted in Coot[205]. 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[206] and the structures were refined using phenix.real_space_refine[207] with restraints imposed on secondary structure and the backbone phosphate geometry of M. Model quality was assessed using MolProbity within Phenix[208] which revealed excellent stereochemistry for both structural models (**Table 1**). Figures and the supplementary video were prepared with PyMol (Schrödinger) and ChimeraX[209].

6. Manuscript 2: Structural basis of SARS-CoV-2 polymerase inhibition by non-nucleoside inhibitor HeE1-2Tyr

6.1. Abstract

Targeting the RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 with small molecules is a promising therapeutic strategy against COVID-19, but potent and safe inhibitors are lacking. The non-nucleoside inhibitor HeE1-2Tyr has been shown to inhibit the RdRp of both SARS-CoV-2 and Dengue virus in cells, but the underlying mechanism remains unclear. Here, we elucidate the molecular mechanism of HeE1-2Tyr-mediated SARS-CoV-2 RdRp inhibition. Biochemical assays confirm that HeE1-2Tyr inhibits RdRp with an IC50 of 5 µM and show that it directly competes with RNA binding to RdRp in vitro. Structural analysis using cryo-EM reveals that a stack of three HeE1-2Tyr molecules binds to the RNA binding site of RdRp. The discovery of the HeE1-2Tyr binding site and its intriguing inhibition mechanism of three stacked molecules that outcompete RNA may facilitate further development of pan-corona non-nucleoside inhibitors.

6.2.Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19, a disease that has claimed more than 7 million lives worldwide since its outbreak in 2020.[1,3] SARS-CoV-2 is a positive-sense single-stranded RNA virus that encodes 16 non-structural proteins (nsps) that facilitate the viral life cycle.[5] The RNA-dependent RNA polymerase (RdRp) consists of the polymerase subunit nsp12 and the accessory subunits nsp7 and nsp8.[99,165] The RdRp is essential for replication and transcription of the viral genome and is therefore a highly promising drug target. Inhibition of RdRp with small molecules prevents the production of viral RNA and stops the propagation of the virus.[7,32,210]

RdRp inhibitors can be broadly classified into nucleoside and non-nucleoside inhibitors. Nucleoside inhibitors, also referred to as nucleoside analogues, mimic the naturally occurring substrate of polymerases, namely nucleoside triphosphates (NTPs) or their precursors.[130] In previous studies, we and others have characterized the mode of action of the nucleoside inhibitors Remdesivir[138,211] and Molnupiravir[212,213], which received an Emergency Use Authorization (EUA) for clinical usage against SARS-CoV-2.[214,215] However, due to the chemical nature of nucleoside inhibitors, they have numerous disadvantages related to their pharmacodynamics and pharmacokinetics. For example, Remdesivir and Molnupiravir are disputed due to issues with efficacy and specificity, respectively.[148,216,217] Non-nucleoside inhibitors are structurally distinct from nucleosides or NTPs and usually have a binding site outside the NTP binding pocket of the polymerase. Thus, such molecules are not constrained by their similarity to nucleosides and could help to address the strong need for potent and safe RdRp inhibitors. Unfortunately, despite global efforts to find suitable small molecules since the beginning of the pandemic, promising non-nucleoside inhibitors are still lacking.[8]

Many *in silico* studies have led to the nomination of numerous potential RdRp inhibitors.[150-157] However, very few of these molecules showed an inhibitory effect in reproducible *in vitro* assays. In addition, the further development of discovered small molecule inhibitors is often hampered by a lack of mechanistic insight and structural data of RdRp-inhibitor complexes.[8] There is only one published structure of RdRp bound to a non-nucleoside inhibitor, a natural molecule called suramin.[161] Suramin binds near the RNA

binding site of RdRp and displaces RNA. However, Suramin is a very large repurposed molecule that contains six sulfonic acid groups and is known for its promiscuity.[162-164] Therefore, its properties as a small molecule probe are questionable, and its clinical use to treat COVID-19 is impossible.

HeE1-2Tyr is a pyridobenzothiazole derivative that was first described as a flavivirus RdRp inhibitor (chemical structure shown in Fig. 11a).[14] HeE1-2Tyr inhibits Dengue virus polymerase (NS5 protein) in a dose-dependent manner *in vitro* and shows antiviral effects in cell assays.[14] The mode of action remains disputed, but it seems likely that HeE1-2Tyr binds directly to NS5 and prevents its activity.[14,218,219] Later, it was reported that HeE1-2Tyr is also effective against the RdRp of SARS-CoV-2 *in vitro* and in cell culture experiments.[15] However, the binding site of HeE1-2Tyr and mechanistic insights into HeE1-2Tyr-mediated RdRp inhibition remain unknown.

Here, we elucidate the mode of action underlying HeE1-2Tyr-mediated inhibition of the SARS-CoV-2 RdRp. Using RNA elongation and fluorescence polarization assays, we show that HeE1-2Tyr is a competitive inhibitor of RNA binding to SARS-CoV-2 RdRp and thereby prevents RNA extension. Building on this, we determine the high-resolution cryo-EM structure of HeE1-2Tyr bound to RdRp, which reveals that three molecules of HeE1-2Tyr are stacked on each other and are bound to the highly conserved RNA binding site. Our systematic biochemical and structural results not only provide detailed insights into the mechanism of action of HeE1-2Tyr, but also suggest HeE1-2Tyr as a starting molecule for the development of future pan-corona non-nucleoside RdRp inhibitors.

6.3. Results

6.3.1. HeE1-2Tyr is a potent non-nucleoside inhibitor of SARS-CoV-2 RdRp

To recapitulate the reported inhibitory effect of HeE1-2Tyr (Fig. 11a) against SARS-CoV-2 RdRp[15] in our biochemical system, we used RNA elongation assays as described[99] with recombinantly produced RdRp and a minimal RNA hairpin substrate (Fig 11b, Methods). RdRp was incubated with RNA and up to 20 µM HeE1-2Tyr, and the reaction was started by the addition of NTPs. RNA extension by RdRp was then analyzed by detecting the 5’ fluorescent label of the RNA hairpin after allowing RNA extension for 5 minutes.

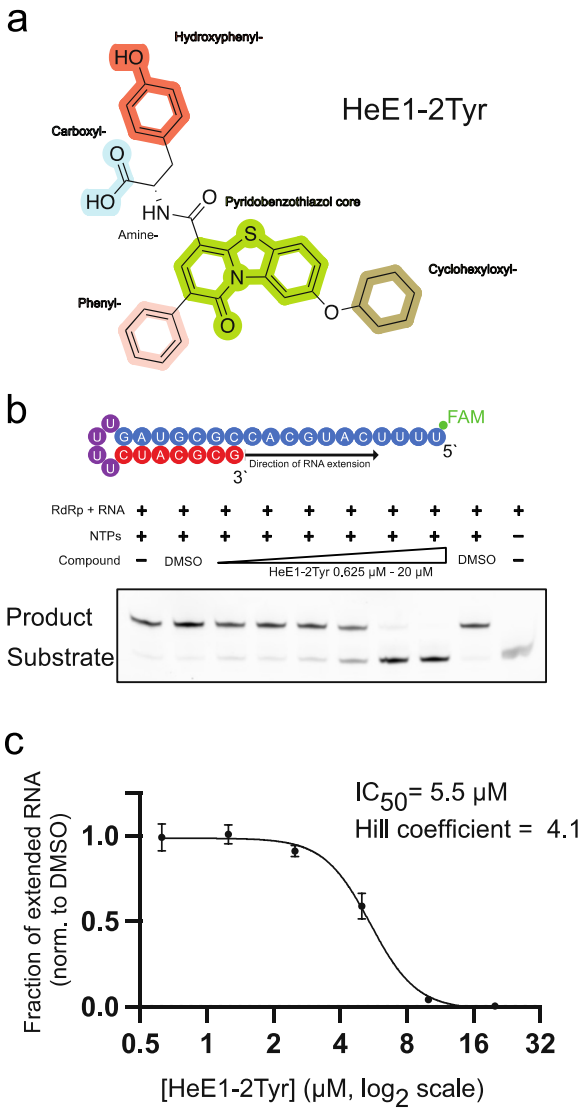
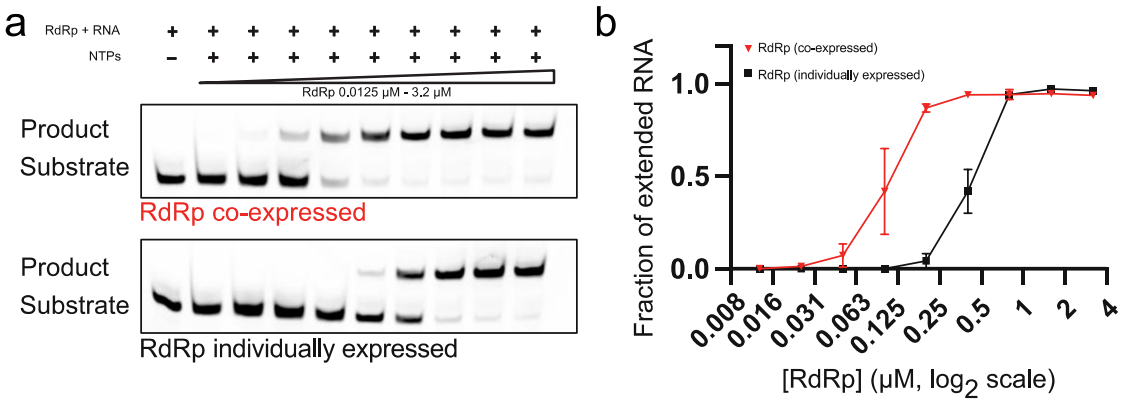


Figure 11: HeE1-2Tyr is a potent non-nucleoside inhibitor of the RdRp of SARS-CoV-2

a, Chemical structure of HeE1-2Tyr; relevant chemical functional groups are highlighted and labeled. **b**, top: minimal RNA substrate that folds into a hairpin with 'template' (blue), 'product' (red) and ´loop´ (purple) regions; The RNA is fluorescently labeled with 5' 6-Carboxyfluorescein (FAM). bottom: HeE1-2Tyr inhibits RNA extension by RdRp in a dose-dependent manner. **c**, Quantification of the fraction of extended RNA of HeE1-2Tyr exposed samples relative to the DMSO control. Mean ± s.d. of independent triplicate measurements are shown. HeE1-2Tyr concentration represented on a logarithmic scale with base 2. Half maximal inhibitory concentration (IC50) and Hill coefficient determined by fitting a non-linear dose-response curve.

Without HeE1-2Tyr, the RNA substrate was readily extended within 5 minutes (Fig. 11b). At HeE1-2Tyr concentrations of 0.625 to 2.5 µM, most of the RNA substrate was extended. However, when adding 5 µM or more HeE1-2Tyr, we observed a dose-dependent inhibitory effect that almost completely abolished RdRp activity at concentrations above 10 µM. Quantitative analysis of three independent experiments and the fit of a dose-response curve resulted in an IC_{50} of 5.5 µM (Fig. 11c), which is five times lower than reported before.[15] This is likely due to the higher activity of the co-expressed RdRp used here, which allowed us to use significantly lower protein concentration in the biochemical assay compared to RdRp reconstituted from individually expressed subunits (Extended data Fig. 2). Additionally, we determined a Hill coefficient of 4.1, suggesting positive cooperativity of RdRp inhibition by HeE1-2Tyr.

In summary, our results demonstrate that HeE1-2Tyr is a potent non-nucleoside inhibitor against RdRp of SARS-CoV-2 *in vitro*, consistent with prior reports.[15] Combining the minimal RNA substrate with highly active co-expressed RdRp provides a versatile system that allows us to further study the molecular mechanisms of SARS-CoV-2 RdRp inhibition by HeE1-2Tyr.

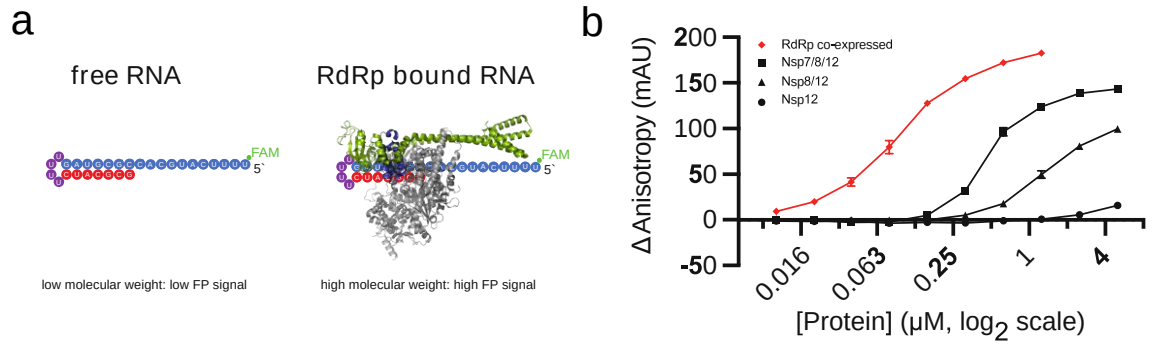


Extended data figure 2: Comparison of the enzymatic activity of RdRp co-expressed and reconstituted from individual subunits

a, Gel-based comparison of the enzymatic activity of RdRp co-expressed (top, red) and reconstituted from individual subunits (bottom, black). Minimal RNA substrate was incubated with NTPs and increasing concentration of the according RdRp (0.0125µM – 3.2µM, two-fold serial dilution); reaction products were separated on a denaturing acrylamide gel. **b**, Quantification of the experiment in **a** after triplicate measurements. Mean ± s.d. of the fraction of extended RNA are plotted; RdRp concentration represented on a logarithmic scale with base 2.

6.3.2. HeE1-2Tyr interferes with RNA binding to SARS-CoV-2 RdRp

Next, we investigated how HeE1-2Tyr inhibits SARS-CoV-2 RdRp. To test if HeE1-2Tyr has an effect on the RNA binding properties of RdRp we established a fluorescence polarization (FP) assay utilizing the same fluorescently labelled RNA hairpin scaffold as in the transcription assay (Extended data Fig. 3a). Since the polarization of the emitted fluorescence is proportional to the hydrodynamic radius of the fluorescently tagged molecule[220], the change of FP was used to calculate the fraction of RNA bound to RdRp (Methods). The FP assay showed that co-expressed RdRp binds the RNA scaffold with a Kd of 75 nM (Extended data Fig 3b.).



Extended data figure 3: Development of a fluorescence polarization (FP) assay to monitor the interaction between RdRp and RNA

a, Principle of the FP assay. RNA harbors a 5' 6-Carboxyfluorescein (FAM) label (green). Free RNA has a low molecular weight, low hydrodynamic radius and consequently a low FP signal whereas RNA bound to RdRp behaves the opposite. **b**, Quantification of the fluorescence anisotropy of RNA exposed to increasing concentrations of co-expressed RdRp (red diamonds), RdRp reconstituted from individual subunits (black squares), nsp8 and nsp12 (black triangles) and nsp12 alone (black circles). Mean ± s.d. of independent triplicate measurements are shown, protein concentration represented on a logarithmic scale with base 2

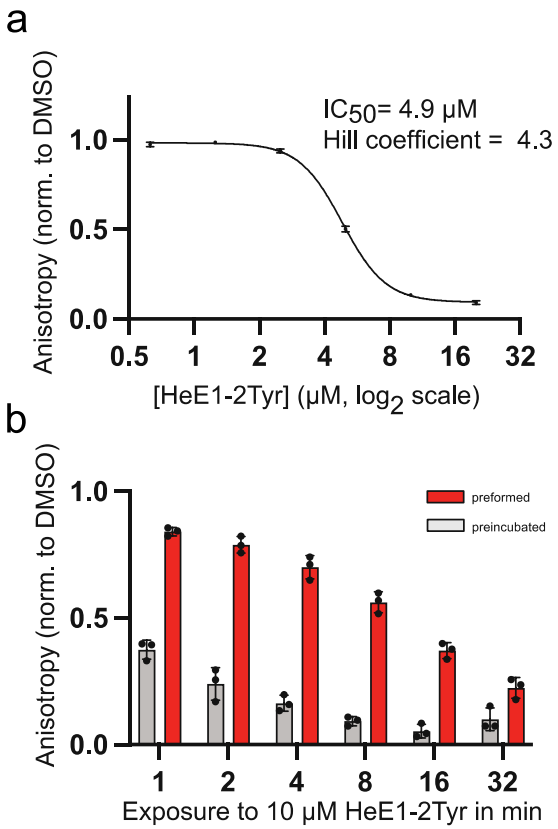


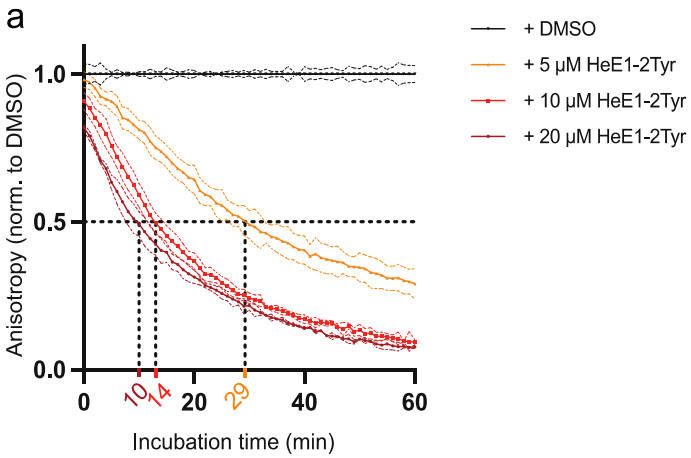
Figure 12: HeE1-2Tyr prevents the interaction between RdRp and RNA

a, HeE1-2Tyr inhibits the interaction between RdRp and RNA in a dose-dependent manner. Quantification of the decrease in fluorescence anisotropy of HeE1-2Tyr exposed RdRp samples relative to the DMSO control. Mean ± s.d. of independent triplicate measurements are shown; some error bars are too small to be depicted; raw values are provided in the source data. HeE1-2Tyr concentration represented on a logarithmic scale with base 2. Half maximal inhibitory concentration (IC_{50}) and Hill coefficient determined by fitting a non-linear dose-response curve. **b**, 10 μM HeE1-2Tyr has significantly stronger inhibitory effect when added before the RNA. RdRp was either preformed with RNA and then exposed to HeE1-2Tyr (red) or preincubated with HeE1-2Tyr and then allowed to interact with RNA (grey). Total time of exposure depicted on a logarithmic scale with base 2; individual data points and boxes represent mean ± s.d. of independent triplicate measurements.

Addition of HeE1-2Tyr caused a dose-dependent reduction in RNA binding to RdRp, where RNA binding was nearly abolished at HeE1-2Tyr concentrations above 10 µM (Fig. 12a). Accordingly, quantitative analysis of three independent experiments revealed that HeE1-2Tyr inhibits the interaction of RdRp and RNA with an IC_{50} of 4.9 µM and with a Hill coefficient of 4.3. These numbers are almost identical to the results obtained from the RNA extension assay (Fig. 11c), suggesting that HeE1-2Tyr inhibits RNA extension by interfering with RNA binding to RdRp.

To study the interference between RNA and HeE1-2Tyr in more detail, we repeated the FP assay while varying the order of addition of RNA and HeE1-2Tyr. For pre-incubated samples, RdRp and HeE1-2Tyr were incubated together before RNA was added. Whereas for pre-formed samples, RdRp was allowed to form a complex with RNA before HeE1-2Tyr was added. Comparison of the fraction of RdRp-bound RNA in the FP assay between pre-incubated and pre-formed reactions revealed that 10 µM HeE1-2Tyr is significantly more effective when added before the RNA (Fig. 12b). The difference is most pronounced for short incubation times and diminishes over time (Fig. 12b). Based on these results, we further tested if the rate of disruption of a preformed RdRp-RNA complex is dependent on the concentration of HeE1-2Tyr. Indeed, the time needed to disrupt 50% of the RdRp-RNA complex was significantly shorter for 20 µM than for 5 µM HeE1-2Tyr (Extended data Fig. 4).

Altogether, we show that HeE1-2Tyr at 5 µM or higher can displace RNA from RdRp irrespective of the order of addition. The data suggest that HeE1-2Tyr and RNA compete for the same binding site, and we conclude that HeE1-2Tyr is a competitive inhibitor of RNA binding to SARS-CoV-2 RdRp.

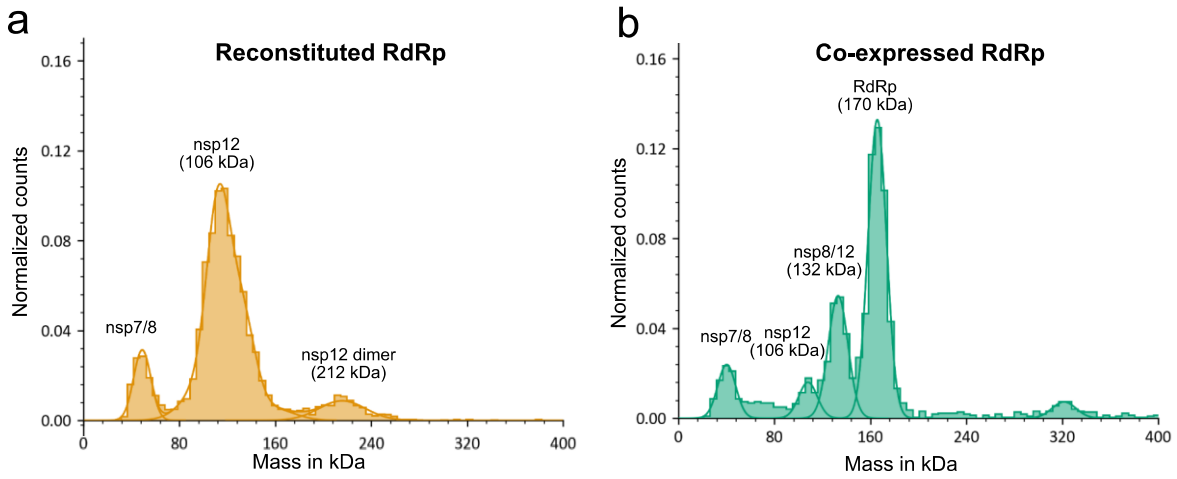


Extended data figure 4: Effect of HeE1-2Tyr on the RdRp-RNA interaction over time

a, Real-time quantification of the fluorescence anisotropy of preformed RdRp exposed to 5 µM (orange), 10 µM (red), and 20µM (dark red) HeE1-2Tyr relative to the DMSO control (black). Mean ± s.d. of independent triplicate measurements are shown; measurement interval: 1 minute. 50% reduction of the anisotropy relative to the DMSO control indicated by dashed lines and color-coded values on x-axis.

6.3.3. Cryo-EM structural analysis of HeE1-2Tyr binding to RdRp

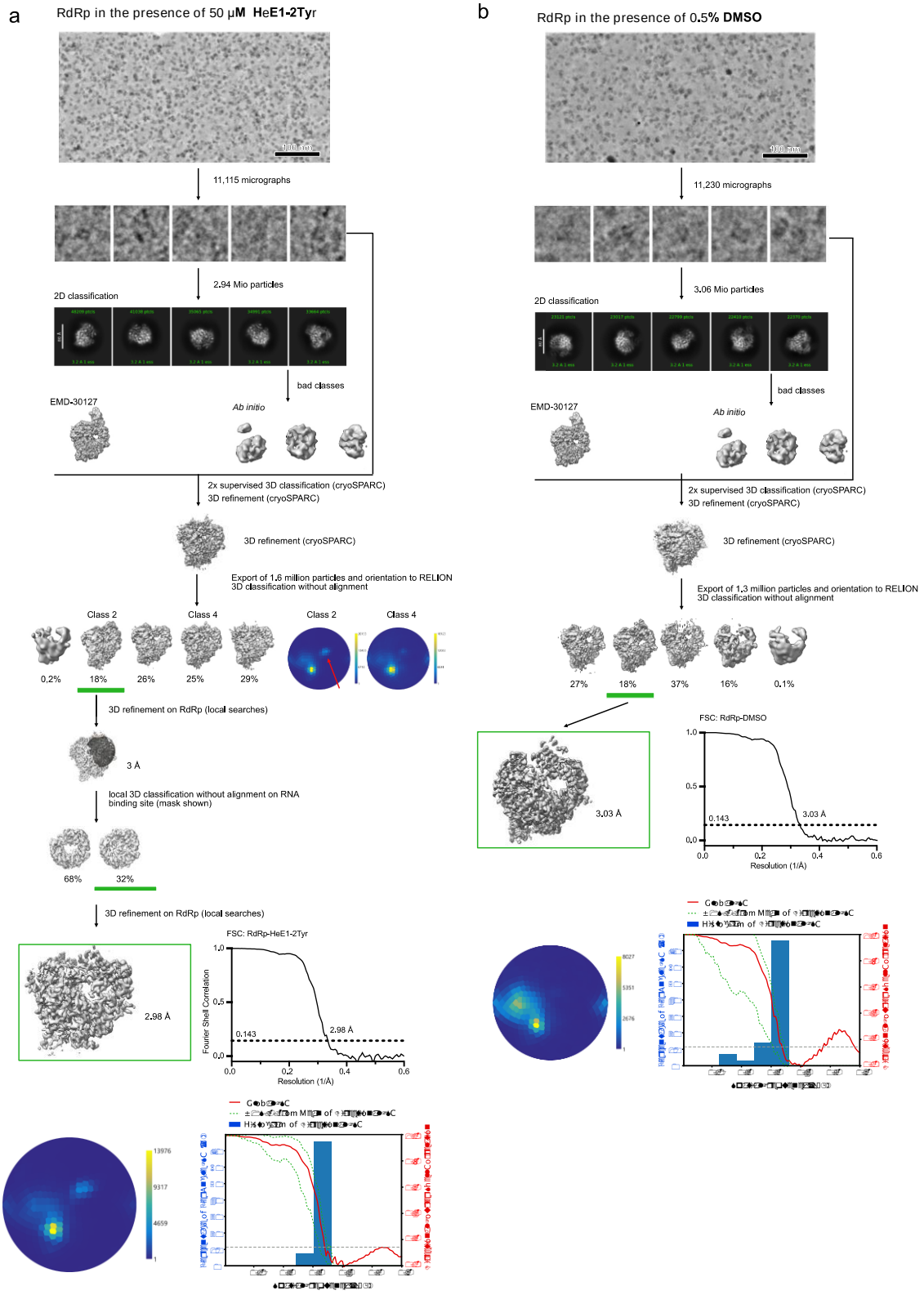
Next, we set out to determine the exact binding mode of HeE1-2Tyr to RdRp using single particle cryo-EM. Based on our finding that HeE1-2Tyr interferes with RNA binding to RdRp, we needed to prepare a stable RdRp complex in the absence of RNA. This has been difficult with RdRp reconstituted from individually expressed subunits because the yield of fully assembled RdRp complex is low (Extended data Fig. 5).[221] The co-expressed and co-purified RdRp that we used for our biochemical experiments, however, shows a significantly higher proportion of fully assembled RdRp (Extended data Fig. 5), which makes it highly suitable for structural investigation of RdRp-inhibitor complexes without RNA.



Extended data figure 5: Mass photometry based analysis of RdRp reconstituted from individual subunits and co-expressed

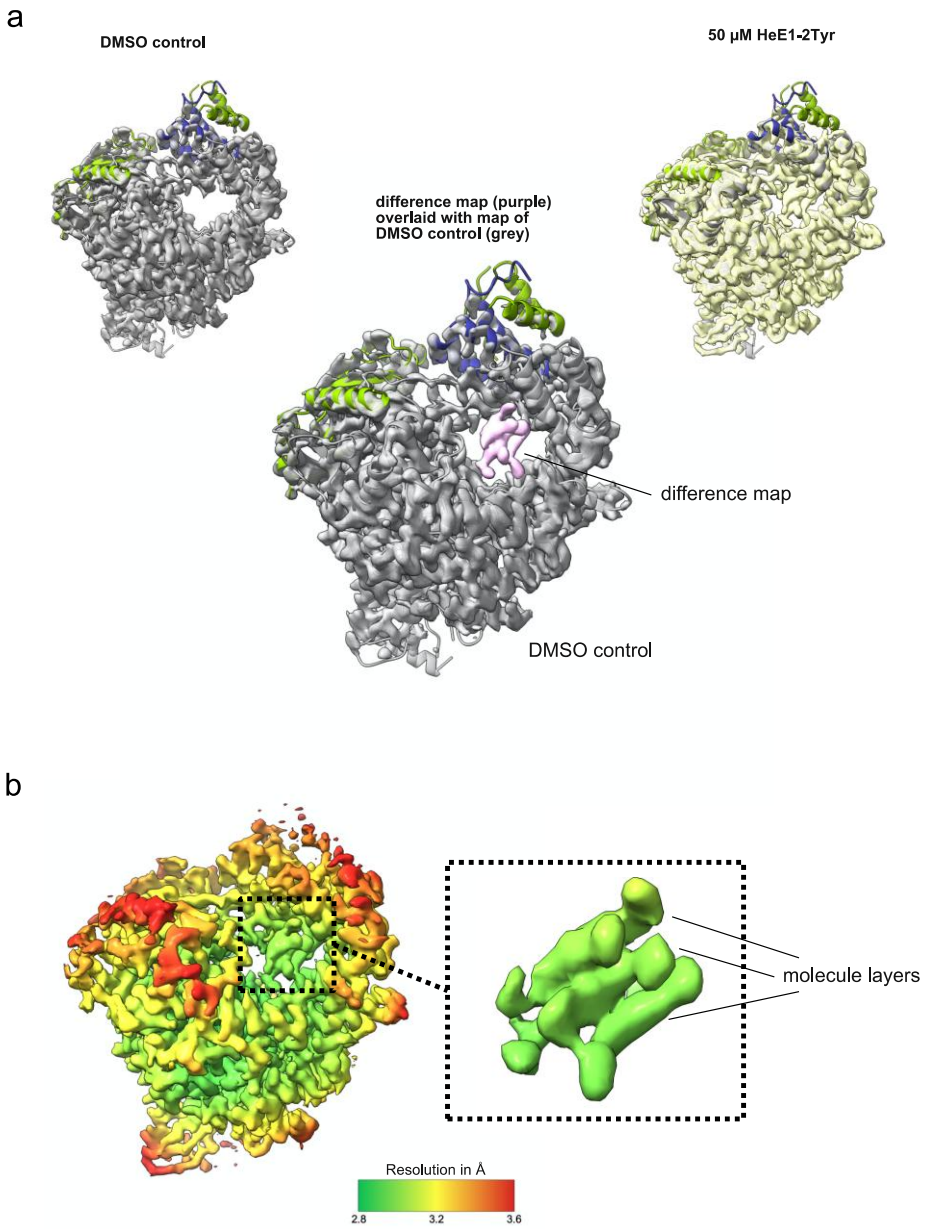
a, RdRp reconstituted from individual subunits (yellow). **b**, co-expressed RdRp (blue). Relevant peaks are annotated and labeled with the name of the respective proteins and the expected molecular weight of their complexes.

We acquired cryo-EM datasets of RdRp without RNA, but in presence of 50 µM HeE1-2Tyr, as well as RdRp with 0.5% DMSO as a solvent control. For the He1-2Tyr containing dataset, we first identified 1.65 million good particles by two rounds of heterogeneous refinements in cryoSPARC with a published RdRp volume (EMD-30572[161]) and three volumes created from junk particles (Extended data Fig. 6). The good particles were then subjected to 3D classification without image alignment in Relion to separate fully assembled RdRp from smaller sub-complexes, which resulted in 300,000 good particles of the RdRp complex.



Extended data figure 6: Cryo-EM data processing trees and quality of reconstructions

a, Cryo-EM data processing tree for RdRp in the presence of 50 µM HeE1-2Tyr. Scale bar, 100 nm. FSC plot, angular distribution and directional FSC calculated according to ref.[222] Global resolution of RdRp-HeE1-2Tyr structure is 2.98 Å and sphericity is 0.903. **b**, Cryo-EM data processing tree for RdRp in the presence of 0.5% DMSO. Scale bar, 100 nm. FSC plot, angular distribution and directional FSC calculated according to ref.[222] Global resolution of RdRp-DMSO structure is 3.03 Å and sphericity is 0.854.



Extended data figure 7: Difference map and map quality

a, reconstruction of the DMSO control (grey) and RdRp-HeE1-2Tyr (yellow) overlaid with atomic model of RdRp (PDB ID: 6M71[98]); reconstructions are shown as semi-transparent grey density; RdRp subunits nsp7, nsp8 and nsp12 are in dark blue, green and gray. middle: difference map of RdRp-DMSO and RdRp-HeE1-2Tyr reconstructions (pink) overlaid with the reconstruction of the DMSO control shown in a; surface dust removal size threshold: 14. **b**, local resolution of the RdRp-HeE1-2Tyr reconstructions with close-up on ligand density (dotted box).

During 3D classification, we identified a particle species containing extra density in the RNA binding site of RdRp. To clearly separate particles with and without the extra density, we performed focused 3D classification without image alignment. This resulted in 200,000 RdRp particles containing an extra density, which could be refined to 3.0 Å. Except for the last step, the DMSO control dataset was analyzed employing a similar strategy,

yielding a 3.1 Å reconstruction (Extended data Fig. 6, Methods). For both reconstructions, we observed an RdRp conformation that is very similar to previously determined structures.[98,161] Thus we could dock and refine a published molecular model of RdRp (PDB: 6M71[98]) comprising nsp12, nsp7 and two copies of nsp8, which yielded high model quality scores.

6.3.4. HeE1-2Tyr binds to RdRp as a stack of three molecules

To compare the cryo-EM maps derived from the dataset of the RdRp-DMSO control and RdRp in the presence of HeE1-2Tyr, we calculated a difference map of both (Methods). This revealed a defined extra density in the RNA binding site of nsp12 for the reconstruction from the compound containing sample (Extended data Fig. 7a). The density in this area is locally resolved to 2.8 Å and shows clear features for bulky chemical groups. The extra density is larger than expected for a single HeE1-2Tyr molecule and shows a three-layered arrangement (Extended data Fig. 7b). Based on the asymmetric chemical structure of HeE1-2Tyr (Fig. 11a) and the high local resolution, we could unambiguously model three stacked molecules of HeE1-2Tyr (Fig 13b), which we refer to as HeE1-2Tyr$_{1\text{-}3}$, with HeE1-2Tyr$_1$ being the molecule closest to the active site of nsp12.

The stacking between the three HeE1-2Tyr molecules is driven by the aromatic pyridobenzothiazole moieties of the compound molecules, which are arranged on top of each other with a planar antiparallel geometry (Fig. 13c). The spacing between the molecule planes is 3.6 Å for HeE1-2Tyr$_{1/2}$ and 3.4 Å for HeE1-2Tyr$_{2/3}$, which is in accordance with π-π interactions.[223,224] HeE1-2Tyr$_1$ and HeE1-2Tyr$_3$ are in the same orientation with the phenyl-group pointing towards the RdRp thumb domain (Fig. 13c), whereas HeE1-2Tyr$_2$ is flipped by 180°. The conformation of the sidechains of the three HeE1-2Tyr molecules differ in order to allow for interactions within the stack and with RdRp (Extended data Fig. 8).Besides π-π interactions, HeE1-2Tyr$_2$ and HeE1-2Tyr$_3$ are stabilized by the formation of two hydrogen bonds (H-bond) between the oxygen of the ether group and the nitrogen of the amine group of each molecule (Fig. 13c).

In summary, we solved the cryo-EM structure of SARS-CoV-2 RdRp in complex with HeE1-2Tyr. The compound is bound as a stack of three HeE1-2Tyr molecules that is stabilized by inter-compound π-π and H-bond interactions. Three-molecule stacking of HeE1-2Tyr within SARS-CoV-2 RdRp represents a novel binding mode of a small-molecule RdRp inhibitor.

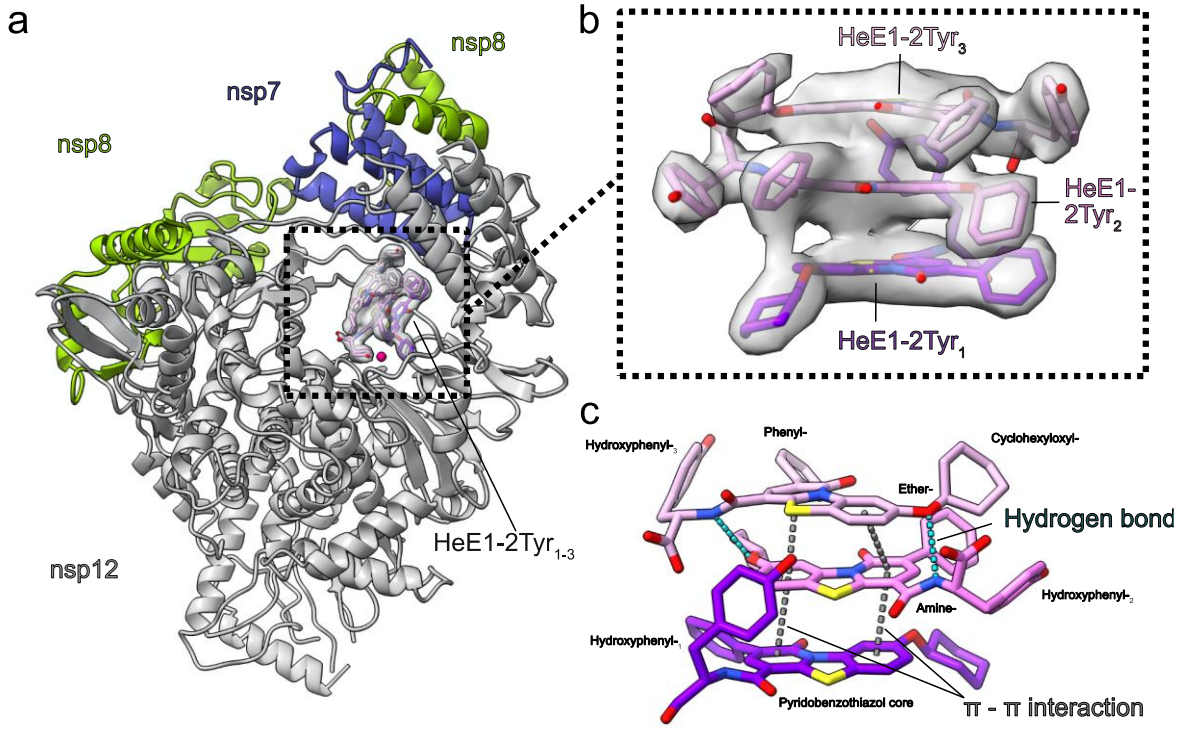
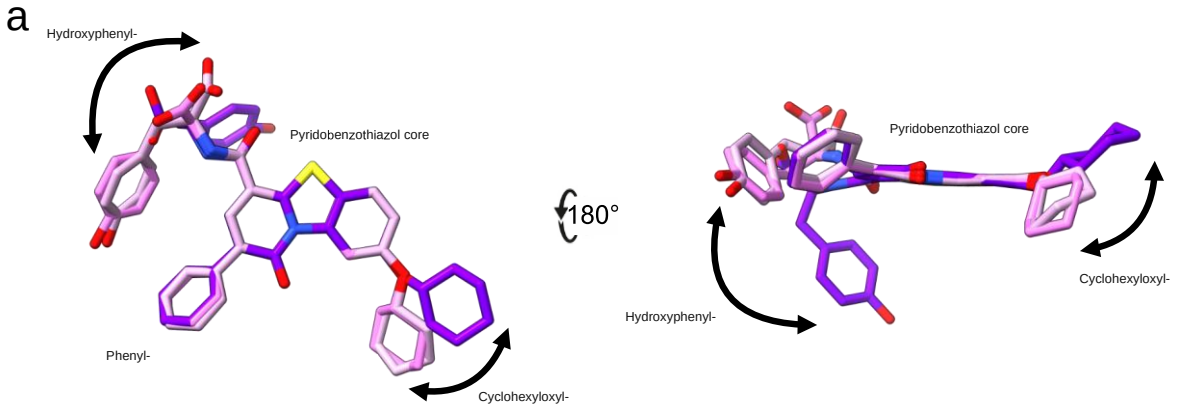


Figure 13: Structure of RdRp bound to a stack of three HeE1-2Tyr molecules

a, Overview of the cryo-EM structure of RdRp bound to a stack of three HeE1-2Tyr molecules (shades of purple). RdRp subunits nsp7, nsp8 and nsp12 are in dark blue, green and grey, respectively. Cryo-EM density shown in the proximity of HeE1-2Tyr molecules (dotted box). The active site ion is modelled according to ref.[102] and indicated by a magenta sphere. **b**, Close-up of the density fit of HeE1-2Tyr molecules inside the RNA binding site of RdRp. Extra density could be unambiguously interpreted by three stacked HeE1-2Tyr molecules, HeE1-2Tyr$_{1-3}$ are shown in violet, pink and light red, respectively. **c**, Stacking of the HeE1-2Tyr molecules is driven by inter-compound π-π interactions (grey) and the formation of two hydrogen bonds (cyan); relevant chemical functional group are labeled.



Extended data figure 8: Conformational changes of HeE1-2Tyr$_{1-3}$

a, HeE1-2Tyr molecules are superimposed and the conformational changes of the side chains are indicated by black arrows. HeE1-2Tyr1-3 are shown in violet, pink and light red, respectively; relevant chemical groups labeled.

6.3.5. The HeE1-2Tyr stack is stabilized by an arginine bracket

The stack of three HeE1-2Tyr molecules binds directly into the RNA binding site of nsp12 between the finger, palm and thumb domain (Fig. 14a). The three HeE1-2Tyr molecules are bound to RdRp by several types of non-covalent interactions including an ionic bond, hydrogen bonds, and cation-π interactions. HeE1-2Tyr$_1$ has the largest surface area exposed to RdRp and consequently has the most interactions with nsp12 (Fig. 14b). The carboxylate of HeE1-2Tyr$_1$ and the guanidinium group of Arg836 have a distance of 3.3 Å, which suggests the formation of an ionic bond (Fig. 14b). In addition, Ser814 is predicted to form a hydrogen bond with the tertiary amine of HeE1-2Tyr$_1$ (Fig. 14b). Besides that, it is known from other protein-ligand structures that guanidium groups of arginine residues can form relatively stable cation-π interactions with aromatic systems.[225,226] Such a cation-π interaction is formed between the hydroxyphenyl moiety of HeE1-2Tyr$_1$ and the guanidinium group of Arg555 (Fig. 14b). HeE1-2Tyr$_2$ has, apart from Van-der-Waals interactions and one potential weak polar interaction between its hydroxyphenyl moiety and a nitrogen in the backbone of Gly590, no significant interactions with RdRp. This suggests that HeE1-2Tyr$_2$ is mostly stabilized by the π-π interactions with the neighboring HeE1-2Tyr$_1$ and HeE1-2Tyr$_3$ molecules (Fig. 13c). HeE1-2Tyr$_3$, which is the molecule furthest away from nsp12, forms a cation-π interaction with Arg858 of RdRp (Fig. 14b).

From these interactions, the three arginines 555, 836 and 858 appear to be most important for the interaction between the HeE1-Tyr molecules and RdRp, and we therefore call them the arginine bracket. Interestingly, the arginine bracket residues also interact with RNA in the structure of replicating RdRp.[99] This suggests that RNA and HeE1-2Tyr binding to RdRp are mutually exclusive for steric reasons. Indeed, when the RdRp-RNA and RdRp-HeE1-2Tyr structures are superimposed, a steric clash between the stack of HeE1-2Tyr molecules and the RNA becomes apparent (Fig. 14c/d). This readily explains our biochemical data and further confirms that HeE1-2Tyr is a competitive inhibitor of RNA binding to SARS-CoV-2 RdRp.

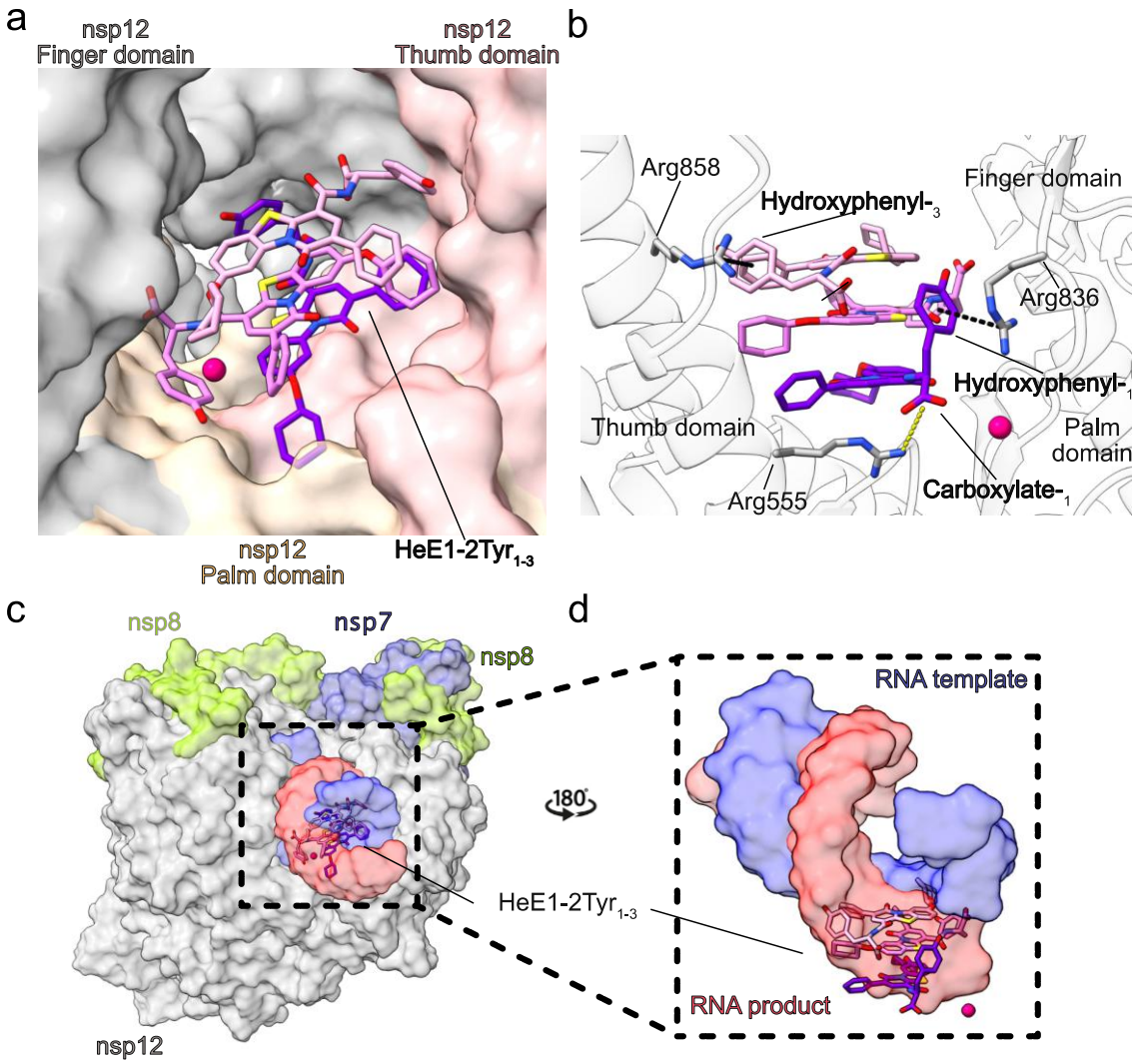


Figure 14: Molecular details of the interaction between HeE1-2Tyr and RdRp RNA binding site

a, HeE1-2Tyr molecules shown inside the RNA binding site of RdRp. Finger, palm and thumb domain of nsp12 are shown in grey, yellow and red surface representation, respectively. Molecules of HeE1-2Tyr$_{1-3}$ are depicted in violet, pink and light red, respectively. The active site ion is indicated by a magenta sphere. **b**, The arginine bracket stabilizes the stack of HeE1-2Tyr molecules inside the RNA binding site of RdRp, an ionic bond (yellow) is formed between carboxylate of HeE1-2Tyr$_1$ and the guanidinium group of Arg555. Two cation-π interactions (black) are observed between the hydroxyphenyl-$_1$ and the guanidinium group of Arg837, and hydroxyphenyl-$_3$ and the guanidinium group of Arg858. Nsp12 domains and relevant chemical functional groups are labeled. **c**, Overview of the structural overlay of RdRp bound HeE1-2Tyr and RdRp in the presence of RNA (PDB ID: 6yyt[99]). RdRp subunits nsp7, nsp8 and nsp12 are in dark blue, green and gray, respectively. The RNA template and product are shown as semi-transparent blue and red surface representation, respectively. HeE1-2Tyr$_{1-3}$ in violet, pink and light red, as before. **d**, The stack of three HeE1-2Tyr molecules would sterically clash with RNA, rendering the interaction mutually exclusive. Same color code as (**c**), view is rotated by 180° relative to (**c**).

In summary, the main interactions of the He1-2Tyr stack with RdRp are mediated through HeE1-2Tyr$_1$ and HeE1-2Tyr$_3$, whereas HeE1-2Tyr$_2$ is predominantly stabilized by π-π interactions and hydrogen bonds within the stack. Three RdRp arginine residues form an arginine bracket that stabilizes the stack of HeE1-2Tyr molecules within the RdRp RNA

binding site by the formation of two cation-π and one ionic interaction. Thus, our atomic structure of RdRp bound to a stack of three HeE1-2Tyr molecules explains how the inhibitor interferes with RNA binding and transcription of SARS-CoV-2 RdRp.

6.3.6. RNA and HeE1-2Tyr both stabilize the RdRp complex

To investigate if binding of the HeE1-2Tyr molecule stack induces conformational changes in RdRp, we compared our structure with structures of free RdRp (PDB ID: 6m71[98]) and RdRp bound to RNA (PDB ID: 6yyt[99]). All three residues of the arginine bracket show a substantial rearrangement compared to the free RdRp structure, adopting a conformation similar to RNA-bound RdRp (Fig. 15a). Interestingly, the arginine bracket residues are also better resolved in the cryo-EM density of the HeE1-2Tyr-bound RdRp when compared to the DMSO control, suggesting that these residues adopt a more defined conformation when they interact with HeE1-2Tyr.

Binding of HeE1-2Tyr structurally connects the finger, palm and thumb domains of RdRp using the same arginine residues as involved in RNA binding. Related to that, we and others observed that the RdRp complex is more stable when bound to RNA.[221] Thus, we hypothesized that HeE1-2Tyr might also have a stabilizing effect on the RdRp complex. To investigate this, we first incubated the RdRp complex with increasing concentrations of HeE1-2Tyr ranging from 1.25 µM to 20 µM and then measured the fraction of fully assembled RdRp using mass photometry (Fig. 15b). Indeed, HeE1-2Tyr had a dose-dependent stabilizing effect on RdRp as evident by the increase of the fraction of fully assembled RdRp in the sample. To compare the effect size with that of RNA, we next incubated RdRp with 1.25 µM RNA and then analyzed the fraction of RdRp and free nsp12 in the sample by mass photometry. RNA binding stabilized RdRp to a similar extent as binding of HeE1-2Tyr (Fig. 15b). In summary, when RdRp binds to HeE1-2Tyr or RNA, significantly more fully assembled RdRp comprising nsp7, nsp8 and nsp12 was observed. We propose that the arginine bracket is important for this effect as it interacts with both RNA and HeE1-2Tyr.

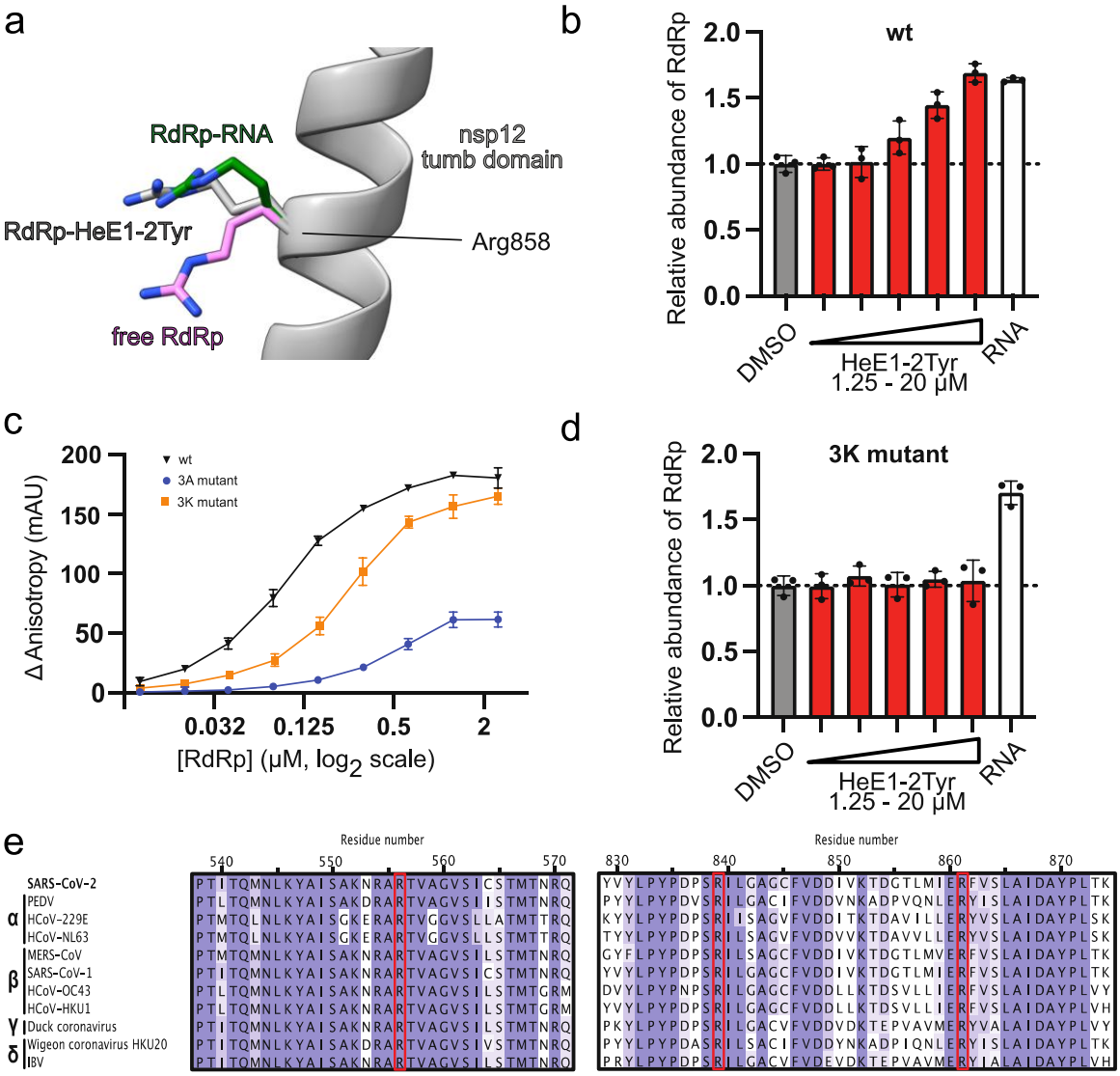
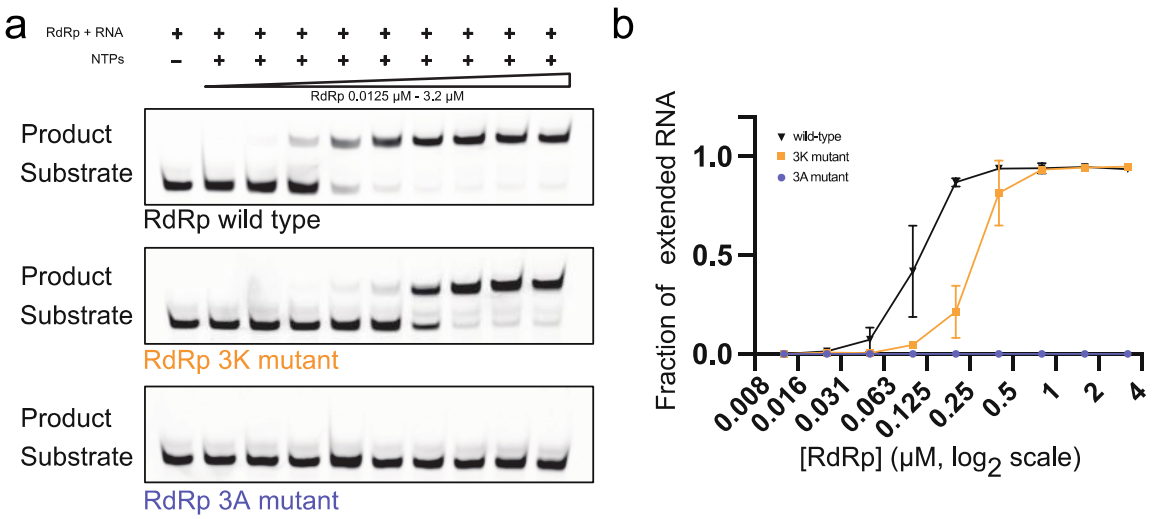


Figure 15: The importance of the arginine bracket for RdRp function and its interplay with HeE1-2Tyr

a, Comparison of the position of Arg858 in RdRp bound to RNA (green, PDB ID: 6yyt[99]), HeE1-2Tyr (grey) and without RNA (pink, PDB ID: 6M71[98]). **b**, HeE1-2Tyr stabilizes the RdRp complex similar to RNA. Mass photometry based quantification of the abundance of RdRp in HeE1-2Tyr (red) or RNA (white) exposed samples relative to the DMSO control (grey). Individual data points and boxes represent mean ± s.d. of independent triplicate measurements. **c**, Analysis of the RNA binding properties of RdRp wild-type (black), the 3K mutant (Nsp12 R555K, R836K, R858K; yellow) and 3A mutant (Nsp12 R555A, R836A, R858A; blue) using the fluorescence polarization assays as in **Fig. 12a**. Mean ± s.d. of independent triplicate measurements are depicted. RdRp concentration represented on a logarithmic scale with base 2. **d**, 3K mutant RdRp is stabilized by RNA but not by HeE1-2Tyr; same experimental setup and color code as in (**c**). **e**, Sequence conservation of the arginine bracket in different coronaviruses. Multiple sequence alignment for the ORF1ab of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Porcine epidemic diarrhea virus (PEDV), Human coronavirus 229E (HCoV-229E), Human coronavirus NL63 (HCoV-NL63), Middle East respiratory syndrome-related coronavirus (MERS-CoV), Severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), Human coronavirus OC43 (HCoV-OC43), Human coronavirus HKU1 (HCoV-HKU1), Duck corona virus and widgeon corona virus HKU20 and Chicken infectious bronchitis virus (IBV). Alignment colored by sequence conservation; residues of the arginine bracket are highlighted with a red box.

6.3.7. The arginine bracket is required for RNA binding and conserved across corona viruses

Next, we further characterized the binding site of HeE1-2Tyr and the importance of the arginine bracket for HeE1-2Tyr and RNA binding. For this purpose, we mutated the residues Arg555, Arg836 and Arg858 to lysines (3K mutant) or alanines (3A mutant). The lysine mutants should retain the positive charges of the arginine bracket required for RNA binding, but might interfere with the formation of cation-π interactions with HeE1-2Tyr due to a change in geometry.[225,227] In contrast, the 3A mutant should abolish all interactions with RNA and HeE1-2Tyr. Expression, purification and biophysical behavior of the mutants were virtually identical to the wild-type RdRp, confirming that the mutant RdRp complexes were correctly formed.



Extended data figure 9: Comparison of the activity of RdRp mutants

a, Gel-based comparison of the transcription activity of RdRp wild-type (top), 3K mutant (middle, yellow) and 3A mutant (bottom, blue). Minimal RNA substrate was incubated with NTPs and increasing concentration of the according RdRp (0.0125µM – 3.2µM, two-fold serial dilution); reaction products were separated on a denaturing acrylamide gel. **b**, Quantification of the experiment in **a** after triplicate measurements. Mean ± s.d. of the fraction of extended RNA are plotted; RdRp concentration is represented on a logarithmic scale with base 2.

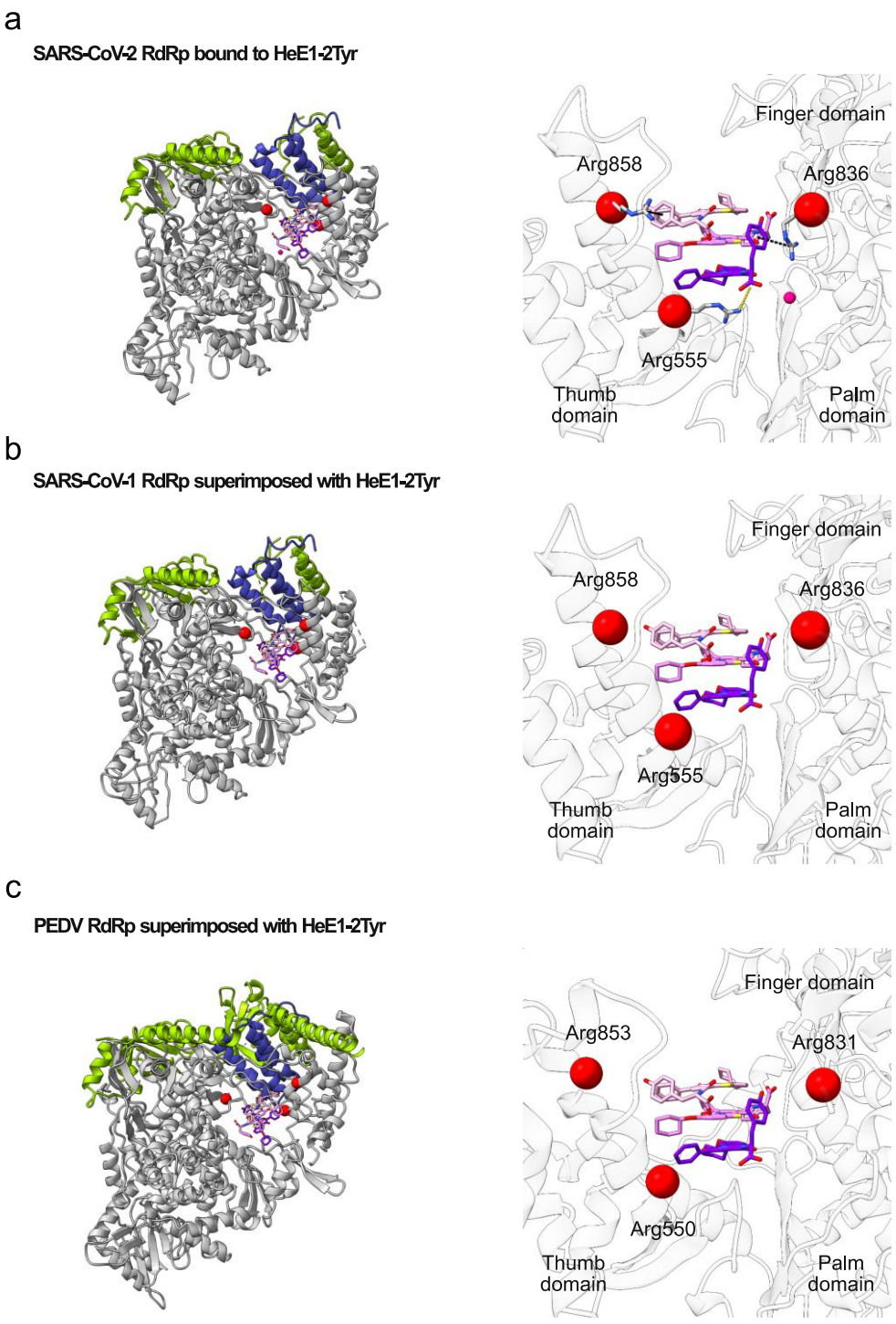
When analyzing the RNA binding properties of the mutant RdRp complexes using the FP assay, we observed that the 3K mutant is still able to bind RNA, but with two-fold decreased affinity compared to wild-type, whereas the 3A mutant shows very weak RNA binding (Fig 15c). The reduction in RNA binding is further reflected by a decreased

transcriptional activity of the 3K and 3A mutants compared to wild-type (Extended data Fig. 9a). Thus, the arginine bracket residues are important for RNA binding and transcription of SARS-CoV-2 RdRp.

Intrigued by our previous results which suggested a stabilizing effect of HeE1-2Tyr on wild-type RdRp, we tested the fraction of fully assembled 3K mutant RdRp in the presence of HeE1-2Tyr or RNA using mass photometry. In contrast to wild-type RdRp, HeE1-2Tyr had no effect on the stability of the 3K mutant, while RNA was still able to stabilize the complex (Fig. 15d). Taken together, this suggests that HeE1-2Tyr is unable to bind to RdRp when the residues of the arginine bracket are converted to lysines, and therefore provides further validation of the compound binding site.

Because of the importance of the arginine bracket for SARS-CoV-2 function, we further investigated the sequence conservation of the arginine residues in different coronaviruses. Our analysis contains representatives from all four coronavirus families, including pandemic causing viruses such as SARS-CoV-1 or MERS-CoV, seasonal human coronaviruses and animal coronaviruses (Fig. 15e). While the overall sequence identity of analyzed nsp12 subunits ranges from 51% - 96%, the residues of the arginine bracket are 100% conserved across all investigated coronaviruses. This conservation is also reflected at the structural level (Extended data Fig. 10), suggesting that HeE1-2Tyr may also be able to bind to RdRp of various coronaviruses and inhibit their activity.

In summary, the arginine bracket is crucial for activity of SARS-CoV2 RdRp and is highly conserved across various coronaviruses. Thus, HeE1-2Tyr or other small molecules with the same binding site could serve as pan-corona drugs capable of inhibiting a wide range of coronaviruses.



Extended data figure 10: Structural comparison of the RNA binding site in corona viruses

a, Cryo-EM structure of SARS-CoV-2 RdRp bound to HeE1-2Tyr. Active site ion indicated by magenta sphere; HeE1-2Tyr$_{1-3}$ are shown in violet, pink and light red, respectively; C-alpha atoms of Arg555, Arg837 and Arg858 depicted as red spheres; residues and RdRp domains are labeled. Close-up: HeE1-2Tyr binding site. b, Cryo-EM structure of SARS-CoV-1 RdRp (PDB ID: 6NUR[58]) superimposed with stack of three HeE1-2Tyr molecules from RdRp-HeE1-2Tyr structure. Same color code as **a;** c-alpha atoms of Arg555, Arg837 and Arg858 depicted as red spheres. **c**, Cryo-EM structure of PEDV RdRp (PDB ID:8URB[228]) superimposed with stack of three HeE1-2Tyr molecules from RdRp-HeE1-2Tyr structure. Same color code as **a;** c-alpha atoms of Arg550, Arg831 and Arg853 depicted as red spheres.

6.4. Discussion

HeE1-2Tyr was initially discovered as a flavivirus inhibitor[14] that also inhibits SARS-CoV-2 RdRp.[15] However, there is no understanding of the molecular mechanism of how HeE1-2Tyr binds and inhibits SARS-CoV-2 RdRp, which prevented further development of HeE1-2Tyr derivatives by structure-guided methods. In this study, we show that HeE1-2Tyr is a competitive RdRp inhibitor that prevents polymerase activity by displacing the RNA from RdRp. The atomic structure of HeE1-2Tyr bound to SARS-CoV-2 RdRp shows that HeE1-2Tyr binds as a stack of three molecules to the RNA binding site of RdRp. The compound stack is stabilized by a bracket of three arginine residues in nsp12 that are crucial for RNA binding and highly conserved across coronaviruses.

The functional importance of the arginine bracket for RNA binding suggests that mutations in SARS-CoV-2 RdRp that render the enzyme insensitive to HeE1-2Tyr will lead to a loss of viral fitness. This creates a high resistance barrier making it unlikely that HeE1-2Tyr-resistant RdRp variants evolve easily. Due to the high importance of the arginine bracket for RdRp function, the HeE1-2Tyr binding site is also structurally conserved. We therefore propose that HeE1-2Tyr or improved derivatives could bind to the same binding site in RdRp of other coronaviruses, such as the human pathogenic MERS-CoV, SARS-CoV-1 and the seasonal coronaviruses CoV-OC43 or CoV-NL63. Thus, HeE1-2Tyr is a promising starting point for the development of a potential pan-coronavirus drug.

Stacking of three small molecule inhibitors inside a binding site is rather unusual but has been reported before.[229,230] In one study, a triple stack of a polyphenolic ligand bound to an acetylcholine-binding protein was described.[229] The stack of ligand molecules is stabilized by inter-compound π-π stacking interactions, which is similar to what we observe for the stack of HeE1-2Tyr. Additionally, the structure of a helical tau filament bound to stacked PET ligands was reported, in which the stack is also stabilized by internal π-π interactions.[230] Nevertheless, binding of a stack of multiple compound molecules to a target protein is a rare binding mode and needs to be studied further.

Binding of several ligands to a target protein can cause cooperativity, usually because one compound influences the affinity for another, for example by modulating the protein conformation.[231] We also observed strong positive cooperativity in our biochemical characterization of HeE1-2Tyr-mediated RdRp inhibition. Considering that HeE1-2Tyr binds to RdRp as a stack of three molecules that interact with each other, the cooperativity can

be readily explained. We propose that once HeE1-2Tyr$_1$ binds to RdRp, a new binding interface for additional HeE1-2Tyr molecules is created. The inter-compound interactions will then further stabilize incoming HeE1-2Tyr molecules, explaining how binding of one molecule might increase the affinity for the others. This mode of positive cooperativity is partly built into the small molecule compound and therefore does not require a large conformational change of the protein. Thus, this mode of action could be exploited in the future to increase the surface area of small molecule inhibitors providing a potential starting point for other challenging targets such as protein-protein interactions.

In contrast to the stack of three HeE1-2Tyr molecules bound to SARS-CoV-2 RdRp, the compound binds as a monomer to the Dengue virus RdRp.[14] The RdRps of Dengue and SARS-CoV-2 are very distinct with a sequence identity of only 68%, thus their active sites and RNA binding are expected to differ. Indeed, the only common feature of the HeE1-2Tyr binding modes is that the molecule binds in the RNA binding site of both RdRps. However, the authors acknowledge that the crystal structure of the Dengue virus RdRp in complex with HeE1-2Tyr does not explain their biochemical data, raising the possibility that stacking or an alternative binding site of HeE1-2Tyr inside the Dengue virus RdRp may have been missed due to the use of protein crystallography as a structure determination method.[14]

Until now, the structure-activity relationship (SAR) of HeE1-2Tyr and RdRp inhibition was unknown, preventing the rational design of more potent derivatives. Initial attempts to modify HeE1-2Tyr affected the solubility of the compound, but did not provide clear insights into the SAR.[14] In a recent study Tóth et al. tried to simplify HeE1-2Tyr by replacing the pyridobenzothiazol core with pyridone or thiazolopyridone.[232] Interestingly, both truncations of the heterocyclic core resulted in significantly less potent derivatives, which can be readily explained by the loss of inter-compound π-π stacking interaction observed in our atomic structure. This underlines the importance of understanding the mode of action and suggests that future attempts to improve HeE1-2Tyr will benefit from our atomic structure. For example, it would be interesting to test if tethering of three HeE1-2Tyr molecules with a covalent linker leads to increased potency.

Altogether, we present biochemical data and a cryo-EM structure of SARS-CoV-2 RdRp bound to a stack of three HeE1-2Tyr molecules, revealing that HeE1-2Tyr inhibits SARS-CoV-2 RdRp by competing with RNA binding. The discovery of the unusual binding

mode of HeE1-2Tyr and the high conservation of the binding site across coronaviruses make this molecule a potent candidate for future structure-based pan-coronavirus drug design.

6.5. Methods

6.5.1. Protein preparation

Preparation of SARS-CoV-2 RdRp, composed of nsp12, nsp7 and two copies of the nsp8 subunits, was carried out as described by Madru et. al.[233], with small adaptations. Concise, the RdRp expression plasmid (Addgene 165451) was used to transform *Escherichia coli* BL21(DE3) RIL. Cells were grown at 30 °C and 150 rpm shaking in LB media supplemented with Kanamycin until an optical density at 600 nm of 0.4 was reached. Then the temperature was reduced to 18°C, and cells were cooled down for one hour. Protein expression was induced with 0.05 mM IPTG, and cells were further incubated for 16 h at 18°C. After collection, cells were resuspended in lysis buffer (50 mM Na-HEPES pH 8.0, 500 mM NaCl, 5% (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). Cells were lysed using a sonicator (80% intensity, 10 sec on, 50 sec off, total time 2 min). The lysate was clarified by centrifugation at 74,766g for 45 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 10 column volumes (CV) lysis buffer and eluted with a gradient of 0–100% lysis to elution buffers (50 mM Na-HEPES pH 8.0, 500 mM NaCl, 5% (vol/vol) glycerol, 500 mM imidazole pH 8.0 and 5 mM β-mercaptoethanol) over 30 CV. Fractions containing RdRp were pooled and 5-fold diluted with dilution buffer (50 mM Na-HEPES pH 8.0, 5% (vol/vol) glycerol and 5 mM β-mercaptoethanol) to reach a final sodium chloride concentration of 100 mM. The protein solution was applied to a HisTrap Q prepacked column preequilibrated in application buffer (50 mM Na-HEPES pH 8.0, 150 mM NaCl, 5% (vol/vol) glycerol and 5 mM β-mercaptoethanol) to remove impurities and free nsp8. Bound RdRp was washed with 10 CV application buffer and then eluted with a gradient of 0-100% application to elution buffers Q (50 mM Na-HEPES pH 8.0, 1 M NaCl, 5% (vol/vol) glycerol and 5 mM β-mercaptoethanol) over 20 CV. Elution peaks were analyzed using a 4-12 % Bis-Tris SDS-PAGE. Peaks containing RdRp without any significant contamination were pooled and

concentrated to 60 µM using a 15,000 molecular weight cutoff Amicon Ultra Centrifugal Filter (Merck). During the concentration procedure, a buffer exchange to storage buffer (20 mM Na-HEPES pH 8.0, 300 mM NaCl, 5% (vol/vol) glycerol, 1 mM $MgCl_2$ and 1 mM Tris(2-carboxyethyl)phosphine (TCEP)) was performed by diluting RdRp in storage buffer 5 times. The concentrated and buffer-exchanged protein was then aliquoted, flash-frozen in liquid nitrogen and stored at -70 °C until further usage.

6.5.2. RNA extension assays

Preparation of SARS-CoV-2 co-expressed RdRp was carried out as described above, and preparation of SARS-CoV-2 individually expressed RdRp was carried out as previously described.[99] The RNA scaffold was designed according to published SARS-CoV-2 RNA extension assays.[99,212] All assays were performed with a single-stranded looped RNA scaffold containing a 6-carboxyfluorescein (FAM) fluorescent label at the 5' end (Fig. 11b). The RNA sequence used for the assays was 5´-6-FAM/rUrUrU rUrCrA rUrGrC rArCrC rGrCrG rUrArG rUrUrU rUrCrU rArCrG rCrG-3´. The RNA template-product hairpin duplex allowed for RNA extension by 11 nucleotides. The 5' FAM-labeled RNA oligonucleotide was purchased from Integrated DNA Technologies (IDT).

For the HeE1-2Tyr dose curve, the final concentrations of RdRp and RNA in the reaction were 0.125 µM. The final HeE1-2Tyr concentrations ranged from 0.625 µM to 20 µM. As HeE1-2Tyr stocks were dissolved in DMSO, the DMSO concentration in the assays was held constant.

RNA was annealed in annealing buffer (50 mM NaCl, 10 mM Na-HEPES pH 7.5) by heating it to 75°C for 1 min and gradually cooling to 4°C. Annealed RNA was diluted in 50 mM NaCl reaction buffer (50 mM NaCl, 25 mM Na-HEPES pH 7.5, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol), and RdRp was diluted in 300 mM NaCl reaction buffer (300 mM NaCl, 25 mM Na-HEPES pH 7.5, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol). The annealed RNA and RdRp were combined, and the reaction buffer was diluted to 50 mM NaCl using reaction buffer without NaCl (25 mM Na-HEPES pH 7.5, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol). HeE1-2Tyr and DMSO controls were diluted in 300 mM NaCl reaction buffer with 0.2% DMSO and then added to the annealed RNA and RdRp. The reaction buffer was diluted to 50 mM NaCl using reaction buffer without NaCl. Annealed RNA, RdRp and compound were then incubated for 30 min at 30°C. For the RdRp expression system comparison (Extended data Fig. 2) and mutants tests (Extended data Fig. 9), RNA extension reactions contained RNA (0.4 µM), and

either co-expressed RdRp (0.0125 µM to 3.2 µM) or individually expressed nsp12 (0.0125 µM to 3.2 µM) with threefold molar excess each of nsp8 (0.0375 µM to 9.6 µM) and nsp7 (0.0375 µM to 9.6 µM) in 50 mM NaCl reaction buffer. Reactions were incubated for 30 min at 30°C to allow RdRp to bind RNA. For all extension assay tests, RNA extension was started by the addition of NTPs (150 µM UTP, GTP and CTP, and 300 µM ATP) in 50 mM NaCl reaction buffer. After 5 min of incubation at 30°C, the reactions were stopped with 2× stop buffer (6.4 M urea, 50 mM EDTA pH 8.0, 1× TBE buffer). RNA products were resolved on 15% denaturing polyacrylamide-urea gels in 1× TBE running buffer and visualized using a Typhoon 95000 FLA Imager (GE Healthcare Life Sciences). Bands were quantified with Image Lab (Bio-Rad). The fraction of RNA extended was calculated by obtaining the ratio of the adjusted volume (volume of band intensity with background subtracted) of the RNA product strand to the combined adjusted volume of the RNA substrate and product strands. For the HeE1-2Tyr dose curve, the fraction of RNA extended in the HeE1-2Tyr samples was normalized to that of the DMSO controls. The obtained data were plotted and analyzed with GraphPad Prism version 9.

6.5.3. RNA fluorescence polarization assay

The RNA fluorescence polarization (FP) assay was performed with the same single-stranded looped RNA scaffold as was used for the RNA extension assay. To study potential interactions of the RNA with its binding partners, the FP signal of the 6-carboxyfluorescein (FAM) label at the 5’ RNA end was utilized. Measurements of fluorescence and FP were performed with an Infinit M1000Pro plate reader (Tecan Group Ltd., Switzerland), with excitation and emission wavelengths of 470 nm (+/-5nm) and 517 nm (+/-5nm), respectively. All reactions were performed in assay buffer (20 mM Na-HEPES pH 8.0, 300 mM NaCl, 1 mM $MgCl_2$, 5 mM β-mercaptoethanol) with a final reaction volume of 20 µl per well. Black 384-well non-binding microplattes (Greiner Bio-One, Germany) were used to record data.

For the assay development, shown in Extended data Fig. 3, 25 nM of RNA was incubated with increasing concentrations of the according protein indicated in the figure legend. The reactions were incubated for 15 min at room temperature; subsequently, the FP was measured. To remove background FP signal caused by the RNA itself, the delta of the FP signal was calculated by subtracting the mean FP of RNA only samples from each

sample containing RNA and protein. Samples were measured in independent triplicates, and the mean +/- standard deviation is reported in the figure.

For the FP assay HeE1-2Tyr dose curve, shown in Fig. 12, 25 nM of RNA was incubated with 125 nM RdRp and increasing concentrations of HeE1-2Tyr. The final HeE1-2Tyr concentrations ranged from 0.625 µM to 20 µM, with 0.2% DMSO in all reactions. The reactions were incubated at room temperature (RT) for 30 min; subsequently, FP was measured. The delta FP was calculated as stated above and then normalized to the DMSO sample, resulting in relative FP signal. Samples were measured in independent triplicates, and the mean +/- standard deviation is reported in the figure.

For the FP assay shown in Fig. 12, the order of RNA and HeE1-2Tyr was varied by preforming RdRp or preincubating RdRp and HeE1-Tyr. For pre-formed samples, RdRp was allowed to form a complex with RNA for 5 min at RT before HeE1-2Tyr was added. For pre-incubated samples, RdRp and HeE1-2Tyr were incubated together for various timepoints before RNA was added. Time of compound exposure is indicted in the figure legend. RdRp and RNA interact happened very quickly (within seconds, exact quantification not possible due to technical reasons), and the time of exposure to RNA had no significant influence (data not shown). The delta FP was calculated as stated above and then normalized to the according DMSO control (DMSO controls performed for each condition and timepoint separately). Samples were measured in independent triplicates, and the mean +/- standard deviation is reported in the figure.

For the experiment shown in Extended data Fig. 4, RdRp was performed with RNA for 5 min at RT, and then different HeE1-2Tyr concentrations were added. The FP signal was measured every 60 sec for a total of 60 min. The delta FP was calculated as stated above and then normalized to the DMSO control of that specific timepoint. Samples were measured in independent triplicates, and the mean +/- standard deviation is reported in the figure.

The FP assay for mutant RdRps, shown in Fig. 15, utilizes the same protocol as the assay development experiment (Extended data Fig. 3), with a deviation in the RdRp proteins used (two RdRp mutant proteins) , as stated in the figure legend.

6.5.4. Mass photometry

All mass photometry measurements were performed on a Refeyn TwoMP mass photometer with the Refeyn AcquireMP and DiscoverMP software (both v. 2.2). Data acquisition records were done in normal mode with a regular image size for 60 s per sample. Measurements were carried out in silicone gaskets (3 mm × 1 mm, GBL103250, Grace Bio-Labs) on microscopy slides that had been cleaned by isopropanol, Milli-Q water and isopropanol. For each measurement, 15 µL of buffer was added to the gasket, and the autofocus function was switched on. After successful focusing, 5 µL of prediluted protein sample were added, and the measurement was started. The resulting movies of 60 s were then analyzed and graphed in the DiscoverMP software. For each measurement, at least 5,000 events were recorded and analyzed. Each condition was measured in independent triplicates.

For the comparison of RdRp preparation methods shown in Extended data Fig. 5, RdRp was either coexpressed as stated above or reconstituted from individual subunits as previously described.[99] 3 µM coexpressed RdRp or 1 µM nsp12, 3 µM Nsp7 and 3 µM nsp8 were incubated in SEC buffer (20 mM Na-HEPES pH 8.0, 300 mM NaCl, 5% (vol/vol) glycerol, 1 mM $MgCl_2$, 1 mM Tris(2-carboxyethyl)phosphine (TCEP)) for 30 min at 30°C. The samples were then cooled down to 4 °C and kept on ice. Shortly before the measurements, proteins were diluted with assay buffer to 40 nM, and each sample was applied to the mass photometer as stated above, resulting in a final measurement concentration of 10 nM.

In Fig. 15, the stability of RdRp in was investigated in the presence of HeE1-2Tyr or RNA. Therefore, 300 nM RdRp coexpressed wild-type or 3K mutant was incubated in assay buffer for 60 min at 4°C with various concentrations of HeE1-2Tyr or with 1.25 µM RNA template-product hairpin (same sequence as was used in the RNA extension assay and FP assay but without FAM). HeE1-2Tyr concentrations ranged from 0.625 µM to 20 µM, and DMSO was kept constant at 0.2%.

After the incubation, mixtures were diluted to 40 nM, and each sample was applied to the mass photometer as stated above. To analyze the abundance of RdRp, counts of events corresponding to the molecular weight of RdRp were divided by the total number of events. The abundance of RdRp in each sample was then normalized to the DMSO control, resulting in the relative abundance of RdRp.

6.5.5. Cryo-EM sample preparation and data collection

For cryo-EM sample preparation, 6.6 µM coexpressed RdRp was incubated with 50 µM HeE1-2Tyr or 0.5% DMSO for 30 min at 4°C. 3 µl of the reaction were mixed with 0.5 µl of octyl β-D-glucopyranoside (0.003% final concentration) and applied to freshly glow-discharged UltrAufoil® R 1.2/1.3, 300 mesh grids (Quantifoil). The grids were blotted for 6 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[201] 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 2.43 sec with a dose rate of 23.84 e^{-} px^{-1} s^{-1}, resulting in a total dose of 57.92 e−/Å2 that was fractionated into 80 frames. Motion correction, CTF estimation and particle picking and extraction were performed using Warp[202]. A total of 11,115 and 11,230 videos were collected for RdRp-HeE1-2Tyr and RdRp-DMSO, respectively.

6.5.6. Cryo-EM data processing and structural modeling

For the RdRp-HeE1-2Tyr dataset, 2.94 million particles were extracted using Warp[202] 1.0.9. Particles were imported to cryoSPARC[203] 2.15 and subjected to two-dimensional (2D) classification. 2D classes representing contamination or broken particles were selected and used for calculating three *ab initio* structures. All particles were then used for supervised 3D classification against four references, where three originated from the *ab initio* reconstruction of contamination and broken particles and one was a previous RdRp complex structure (EMD-30572[161]). The class containing RdRp yielded ~1.6 million particles, which were subjected to homogeneous 3D refinement. The refined particles were then exported to RELION 3.159. To separate fully assembled RdRp from smaller sub-complexes, particles were 3D-classified without image alignment (T = 4, six classes classes) using a mask around RdRp that omitted the nsp8 sliding poles. This resulted in two promising classes that appeared to contain fully-assembled RdRp particles. However, the viewing angles of class 2 were significantly better (red arrow Extended data Fig. 6), hence we continued with

~300,000 particles from class 2. To clearly separate particles with and without the extra density, we performed focused 3D classification without image alignment but with a mask around the active site of RdRp (mask: low pass filtered extra density with 20 px extension). This resulted in ~200,000 particles with extra density that we focus-refined to an overall resolution of 3.0 Å. Local resolution was estimated with Relion 3.1 using a kernel size of 10 Å. For the RdRp-DMSO dataset, 3.06 million particles were extracted. Further processing was similar to the first dataset with the exception that the focused classification of the active site was not performed. This resulted in ~200,000 particles of RdRp-DMSO that we focus-refined to an overall resolution of 3.1 Å.

Atomic models were built using the previously published SARS-CoV-2 RdRp structure (PDB 6M71[98]) as a starting model. The model was first rigid-body-fitted into the density and then adjusted with ISOLDE[234] flexible fitting. Based on the high local resolution of the extra density and the three observed layers, three HeE1-2Tyr molecules were manually placed into the extra density. HeE1-2Tyr molecules were then adjusted with the interactive molecular dynamics package from ISOLDE, which allowed us to unambiguously model three stacked molecules of HeE1-2Tyr. The protein chains was then subjected to phenix.real_space_refine. Restraints for HeE1-2Tyr molecules were generated in phenix.elbow[206], and the model quality was assessed using MolProbity within Phenix[208], which revealed excellent stereochemistry (Table 1). All figures were prepared with ChimeraX[235].

6.5.7. Ligand interactions

To analyze potential ligand-protein and ligand-ligand interactions, the Arpeggio tool[236] and LigPlot + v.2.2[237] were used. Proposed interactions were then interpreted based on distances measured with ChimeraX[235].

6.5.8. Sequence conservation

To analyze the sequence conservation of residues of the arginine bracket, a multiple sequence alignment (MSA) of ORF1ab across corona viruses was performed. Therefore, we assembled a small database of corona virus representatives from all four coronavirus families using the Virus Variation Resource[238]. An MSA was then calculated using clustal Omega.[239,240] Results were visualized with Jalview[241]

7. General discussion: Targeting the RdRp of SARS-CoV-2 with small molecule inhibitors

In the previous two main chapters I have described the molecular mechanisms of Molnupiravir-induced mutagenesis and HeE1-2Tyr-mediated RdRp inhibition, which we elucidated using a defined biochemical *in vitro* system. Understanding the molecular details of how a small molecule inhibits a biochemical process is critical for advancing the development of a drug. However, once a small molecule is applied to cells or even patients, the system becomes significantly more complex. Thus, additional influencing factors, such as specificity, solubility or metabolic stability of the small molecule have to be considered. Such factors cannot be addressed in our *in vitro* system, due to the lack of complexity. To put our results into context, I will discuss them based on available scientific literature.

Manuscript 1 was published in August 2021, and Molnupiravir was approved as a drug after the phase III clinical trial was completed in December 2021.[215] Since then, many ideas and concerns about Molnupiravir were raised. Thus the discussion will be focused on advantages and drawbacks of Molnupiravir-induced mutagenesis. HeE1-2Tyr is at a much earlier stage of drug development and therefore less literature is available. Because of that, its intriguing mode of action and the potential to develop it into a pan-corona inhibitor will be highlighted in the discussion. In the end, I will provide an outlook on the future of drug discovery against the RdRp of corona viruses, suggesting potential strategies for the development of pan-corona inhibitors.

7.1. Advantages and drawbacks of Molnupiravir-induced mutagenesis

7.1.1. Tautomerism of Molnupiravir and its implications for base pairing

Based on our biochemical and structural data, we proposed a two-step model of Molnupiravir-induced mutagenesis. In the first step, the active metabolite of Molnupiravir, NHC triphosphate, is used by the RdRp as a substrate instead of CTP or UTP. In the second step, the resulting RNA can be used as a template for the synthesis of new RNA. The presence of Molnupiravir in the template RNA will direct the incorporation of adenosine or guanosine, thus, causing G-A and C-U transition mutations. A key aspect of this model was revealed by our cryo-EM structures, which depicted the formation of stable M-G and M-A base pairs in the RdRp active center. We hypothesized that tautomerism of NHC enables

the formation of these base pairs, but experimental data about the tautomeric states of NHC and its implication for the base paring properties were still missing. Recently, Bessi et. al.[242] provided experimental evidence for the presence of the amino- and imino-tautomeric state of NHC using nuclear magnetic resonance spectroscopy (NMR). The authors showed that NHC forms a Watson–Crick base pair with G and a weak Watson–Crick base pair with A, which fits very well to our biochemical data and the two-step model. Besides that, a second conformation of the NHC:A base pair was revealed, where A is shifted toward the minor groove adopting a wobble-type base pair. This explains the slight geometry deviations from idealized Watson-Crick base pairs observed in our cryo-EM densities. Summarized, our model of Molnupiravir-induced mutagenesis is still valid and more supporting data is available, suggesting that the tautomeric changes NHC allow the formation of NHC:G and NHC:A base pairs in the active site of RdRp.

The promiscuity causes G-A or C-U transition mutations, which are mostly deleterious for the virus. However, some of these mutations can be advantageous for the virus, leading to concerns about the possibility of generating new SARS-CoV-2 variants. Zibat et al.[243] addressed this concern by propagating SARS-CoV-2 in cell culture in the presence of NHC and a selection pressure. The authors revealed that escape mutants of SARS-CoV-2 emerged significantly quicker when the virus was exposed to NHC. This observation is further supported by sequencing surveillance studies, which revealed higher genetic diversity of SARS-CoV-2 strains in patients treated with Molnupiravir[244] and signatures of Molnupiravir-induced mutagenesis found in emerging SARS-CoV-2 strains.[217] Altogether, this suggests that there is a real risk of generating escape mutants, especially when co-treating immune deficient patients with multiple antiviral drugs.

7.1.2. Similarity of Molnupiravir to endogenous molecules and its consequences

The active metabolite of Molnupiravir, NHC triphosphate, is, except for one additional hydroxy-group, chemically identical to endogenous CTP. This has important positive and negative implications for the properties of Molnupiravir as a drug.

Due to the high similarity it is virtually impossible for enzymes to distinguish between the drug and a naturally occurring molecule. As a consequence, Molnupiravir is not affected by the proofreading mechanism of SARS-CoV-2, and it is unlikely that escape mutants that are insensitive to Molnupiravir will emerge. In principal, every mutation that prevents binding of NHC triphosphate will also affect binding of endogenous NTPs, making

it very challenging for the virus to develop a resistance mechanism. This is supported by a recent *in vitro* study, which investigated the effects of prolonged exposure of NHC on cultivated SARS-CoV-2.[245] Despite 30 passages, no evidence of viral resistance emerged, suggesting that there is indeed a high barrier to developing resistance against NHC.

The negative consequence of the described chemical similarity is that NHC has to compete with endogenous NTPs for incorporation by RdRp. Considering the lower incorporation efficiency of NHC compared to the NTPs, as shown in our time-dependent NHC incorporation measurement opposite a templating G or A residue (Fig. 6), it appears unlikely that NHC incorporation events take place. Thus, the ratio of NHC triphosphate versus endogenous nucleoside triphosphates in infected cells is critical for efficient mutagenesis. This ratio can be modified in favor of NHC either by usage of higher doses of Molnupiravir, leading to higher local concentrations of NHC triphosphate, or by decreasing the concentration of NTPs. The latter strategy was investigated by Stegmann et. al.[246], who suppressed the cellular synthesis of pyrimidines with inhibitors of dihydroorotate dehydrogenase (DHODH). SARS-CoV-2 infected cells were cotreated with NHC and DHODH inhibitors, which in theory improves the probability that NHC gets incorporated into RNA. Indeed, the authors could show that NHC and DHODH inhibitors have strong synergistic effects in cells and in an animal model for SARS-CoV-2. Thus, it could be a potential therapeutic strategy to combine Molnupiravir and DHODH inhibitors to combat COVID-19.

7.1.3. Concerns around Molnupiravir´s lack of selectivity

Molnupiravir is per design mutagenic. It inhibits the replication of SARS-CoV-2 by introducing mutations into the viral RNA. In addition, it is unspecific due to its similarity to endogenous NTPs. For example, it was shown that human RNA polymerases, such as the mitochondrial DNA-dependent RNA polymerase (POLRMT), incorporate NHC into nascent RNA.[197] Thus, the question emerges how humans can tolerate the exposure to NHC. The main explanation lies in the fact that humans do not use RNA as template for the synthesis of nucleic acids. Humans and in fact all living organisms encode the genetic information as deoxyribonucleic acid (DNA) and only transcribe it into RNA when needed.[247] RNAs have multiple essential functions in human cells, but they are never replicated. In contrast, some viruses, such as SARS-CoV-2, store the genetic information in form of RNA and do require RNA replication. Thus, human cells can tolerate short-term exposure to Molnupiravir leading to mutations in the RNA, while it is fatal for RNA viruses.

Major concerns about the safety of Molnupiravir and its metabolites remain due to the potential risks of causing mutations in the DNA.[248] Per design Molnupiravir and NHC are incapable of interfering with DNA synthesis due to their chemical nature. However, nucleoside inhibitors are administered as prodrugs and are metabolized to their triphosphate derivative, as mentioned earlier in chapter *4.5.1 Nucleoside analogues*. The metabolic enzymes required for this are salvaged from cellular pathways for NTP production, a complicated process that overlaps with the synthesis of building blocks of DNA, deoxyribonucleoside triphosphate (dNTPs).[133] Critically, ribonucleoside 5´-diphosphate is an obligate intermediate product during the activation of nucleoside inhibitors and during the synthesis of dNTPs.[248] Thus, there is a potential metabolic pathway for Molnupiravir to become a precursor molecule for DNA synthesis. To investigate this risk, Molnupiravir was subjected to several assays measuring the mutagenic potential. Mutagenicity affecting the DNA was shown on cellular level but was not found in animal models.[198,249] Thus, it remains debated if exposure to Molnupiravir increases the risk of mutations in patient DNA.

Summarized, the mode of action of Molnupiravir-induced RNA mutagenesis is a promising strategy to combat SARS-CoV-2 and other emerging RNA viruses, due to the resistance barrier and its broad applicability. However, serious concerns regarding the possibility of generating new viral variants and the safety of Molnupiravir and its metabolic products remain. Thus, an in-depth risk benefit assessment has to be conducted considering the pathogenicity of the virus, other available treatment options for that virus and potential risks posed by usage of Molnupiravir.

7.2.HeE1-2Tyr: A novel mode of action to inhibit RdRp

7.2.1. HeE1-2Tyr is an atypical example for positive cooperativity

HeE1-2Tyr is a competitive RdRp inhibitor that prevents polymerase activity by displacing the RNA from RdRp. In our biochemical analysis of HeE1-2Tyr-mediated inhibition, we observed steep dose-response curves (Fig. 11 and Fig. 12), meaning that a small increase of the HeE1-2Tyr concentration can lead to a comparably large increase of the inhibitory effect. To quantify the slope of the dose-response curves, which is equivalent to the Hill coefficient, a non-linear four parameter logistic dose–response curve was fitted.

HeE1-2Tyr inhibited RNA extension and RNA binding with a Hill coefficient of 4.1 and 4.3, respectively, indicating strong positive cooperativity.

Cooperativity can occur if two or more ligands bind to a target protein. It describes the change in affinity for a ligand, depending if a ligand is already bound to the target protein or not. In most cases, cooperativity is caused by a ligand-induced conformational change of the target protein, which modulates the affinity for further ligand molecules, such as seen in the classical example of hemoglobin – oxygen binding.[231,250] Available data for RdRp does not suggest the presence of multiple distinct ligand binding sites or significant conformational changes. Thus, the observed positive cooperativity for HeE1-2Tyr-mediated RdRp inhibition was surprising for us. However, considering the cryo-EM structure of RdRp in the presence of HeE1-2Tyr, positive cooperativity can be readily explained.

HeE1-2Tyr binds to RdRp as a stack of three molecules that interact with each other. The binding sites of HeE1-2Tyr are in close proximity and partially overlapping, contrasting most textbook examples of cooperativity where distinct ligand binding sites are observed. Thus, we propose that binding of HeE1-2Tyr$_1$ to RdRp creates a new binding interface for additional HeE1-2Tyr molecules. Incoming ligand molecules are then stabilized by inter-compound π-π interactions and other non-covalent interactions, explaining the increased affinity.

Concise, positive cooperativity of HeE1-2Tyr inhibition is a powerful example highlighting the potential of combining biochemical and structural data in order to understand molecular processes.

7.2.2. Stacking of multiple inhibitor molecules; a promising strategy?

Targeting protein-protein or protein-nucleic acid interactions with small molecules remains highly challenging, due to the size of the interaction interface.[251] Thus, it is tempting to speculate that stacking of relatively small HeE1-2Tyr molecules is required to displace the large RNA molecule from RdRp. This hypothesis is supported by comparing the amount of buried surface area of RdRp and different binding partners. The RNA buries an area of 1600 Å^2 of nsp12 while HeE1-2Tyr$_1$ alone only buries an area of 450 Å^2. In contrast, the stack of three HeE1-2Tyr molecules occupies an area of 800 Å^2 on the surface of RdRp, which is comparable with the RdRp-Suramin structure, where 900 Å^2 of nsp12 are buried. This suggests that single small molecules might be simply too small to displace the large RNA molecule from RdRp, an issue that can be overcome by stacking of multiple ligand

molecules. It should be noted here that while the amount of buried surface area is proportional to the binding energy[252], it does not consider the difference of non-covalent interactions observed for RdRp-RNA and RdRp-HeE1-2Tyr. This difference might explain why the stack of three HeE1-2Tyr molecules is able to displace the RNA although it has a smaller interaction interface.

Inhibition mechanisms that require stacking of small molecules ligands are rather unusual, but a conceptually very similar example of three small molecule inhibitors stacked inside a large binding site has been reported before by Stornaioulo et al.[229], as described in *5.4 Discussion*. Altogether, this suggests that stacking of small molecules could be a promising strategy to target large interaction interfaces.

7.2.3. Further advancement of HeE1-2Tyr into a potential pan-corona RdRp inhibitor

Building on the revealed molecular mechanism of HeE1-2Tyr mediated RdRp inhibition, rational drug design can be employed for the development of HeE1-2Tyr derivatives with increased potency. In previous attempts to design HeE1-2Tyr derivatives, stacking interactions could not be considered because they were not discovered, explaining why they were largely unsuccessful.[15,232]

Exploiting the unusual binding mode of HeE1-2Tyr can be attempted with different chemical modifications. For example, the pyridobenzothiazol core of HeE1-2yr could be modified to allow for stronger inter-compound π-π stacking. Besides that, three HeE1-2Tyr molecules could be covalently linked to create a tethered inhibitor assembly with high affinity for RdRp. However, solubility of HeE1-2Tyr is already quite low, and tethering multiple molecules would presumably worsen the issue. Thus, a more radical option could be to develop a small molecule that interacts with the same residues as HeE1-2Tyr but has a different backbone, a strategy called scaffold hopping.[253] If successful, this could lead to the same inhibitory potential as HeE1-2Tyr but significantly improved bioavailability.

In addition to SARS-CoV-2 focused development of HeE1-2Tyr, its properties as a potential pan-corona virus inhibitor should be further explored. We have shown that the compound stack is stabilized by a clamp of three arginine residues that are crucial for RNA binding and highly conserved across coronaviruses. The conservation is evident on sequence- and structural-levels, which allowed us to model a stack of three HeE1-2Tyr bound to the RNA binding site of homologous RdRps from SARS-CoV-1 and PEDV (Fig. 15). The shape of the RNA binding sites and the position of the residues of the Arg clamp are

very similar, indicating that HeE1-2Tyr might be able to bind. However, potential inhibitory effects of HeE1-2Tyr on the RdRp of other corona viruses have to be confirmed in biochemical assays.

Summarized, the binding site of HeE1-2Tyr and its unusual binding mode are a promising starting point for future studies aiming to develop a potent and selective pan-corona inhibitor.

7.3. The future of drug development against corona virus RdRp

In the last chapter I will provide an outlook on the future of drug development against the RdRp of SARS-CoV-2 and other corona viruses. It will start with the questions of whether we still need antivirals and what kind of inhibitor we need. In the second part I will describe how I envision an efficient drug development pipeline against SARS-CoV-2 RdRp considering the methodical advancements described within this thesis.

7.3.1. Do we still need antivirals against corona viruses?

SARS-CoV-2 emerged at the beginning of 2020, and it has claimed more than 7 million lives since then.[3] Impressive progress has been made to combat COVID-19, employing a range of pharmaceutical- and non-pharmaceutical interventions, saving millions of lives.[27] Nevertheless, SARS-CoV-2 is still spreading, new variants are constantly emerging and it is yet more dangerous than influenza.[29] In addition, there are many lurking animal corona viruses with the potential for zoonotic spillovers. Thus, it is clear that SARS-CoV-2 and corona viruses in general are still a risk for human health.

Vaccines are a very effective pharmaceutical intervention to prevent viral infections or to protect from disease. Several vaccines were developed against SARS-CoV-2, saving more than 1.4 million lives. However, these vaccines should not be the only measure we have against SARS-CoV-2. Vaccine induced immunity is waning over time, and new immune evasion variants are emerging, causing breakthrough infections. In addition, vaccines are not applicable to every patient, such as immune compromised patients,^ and the development of vaccines against emerging viruses takes significant time. Because of these reasons it would be highly beneficial if potent and safe antivirals against SARS-CoV-2 and corona viruses in general were be available.

Despite extensive efforts by the scientific community, such antivirals are still lacking. Paxlovid, Remdesivir and Molnupiravir are the only antiviral drugs that are FDA-authorized

for usage in patients against SARS-CoV-2. All three have significant drawbacks, as described in chapter *4.5 Antivirals*. Thus, it can be concluded that there still is a need for antivirals against SARS-CoV-2 and corona viruses. An ideal antiviral drug would be a small molecule with pan-corona inhibitory potential that is orally available, highly potent, very selective and it should be unlikely that resistance mutations against it can emerge. Unfortunately, it remains highly questionable if such a "Swiss army knife"-like inhibitor exists. In my opinion, a combinatorial therapy of several antivirals will be required, as used in the highly active antiretroviral therapy (HAART) against HIV.[254]

Developing suitable antivirals is a time-consuming and costly endeavor, requiring a good choice of the target protein. Considering the life cycle of SARS-CoV-2 and the conservation of viral proteins across the corona virus family, RdRp is still one of the most promising targets. However, finding potent RdRp inhibitors turned out to be highly challenging. Most NAs that block transcription are rendered inactive by the viral proofreading mechanism, and mutagenic NAs such as Molnupiravir are associated with serious concerns about their safety. The discovery of potent non-nucleoside inhibitors is hampered by a lack of amenable binding pockets and incomplete mechanistic insights into the modes of action of potential inhibitors. In my opinion, a sophisticated molecular mechanisms, such as stacking of multiple small molecules in the RNA binding, will be required for a potent non-nucleoside RdRp inhibitor.

Another interesting aspect for drug development against SARS-CoV-2 can be deduced by looking at the rebound of viral infections in patients treated with Paxlovid. The exact frequency of such reoccurrences of SARS-CoV-2 infection after a 5-day treatment with Paxlovid is scientifically debated.[255,256] A recent study reported that up to 20% of Paxlovid-treated patients suffer from a rebound infection, especially in cases where the treatment was started immediately after patient were tested positive.[257] Perelson et. al.[258] and Ranard et. al.[259] provide an interesting theoretical model of SARS-CoV-2 immunity predicting Paxlovid rebound. Simplified, Paxlovid given in an early stage of infection will inhibit the replication of SARS-CoV-2 very efficiently, and the immune system will not be activated, due to the low number of viral particles. However, clearance of all viral particles is unlikely, and the treatment with Paxlovid has to be stopped after 5 days because of Ritonavir´s side effects. Thus, after a 5-day treatment is completed, remaining viral particles

will cause rebound infection. This suggests that not only the potency of small molecule antivirals is critical but also its tolerability for long-term treatment of a patient.

7.3.2. Designing an efficient pipeline for drug discovery against SARS-CoV-2 RdRp

There is a clear need for potent and selective RdRp inhibitors, but finding and characterizing potential small molecule inhibitors requires a pipeline consisting of reliable tools. Such biochemical and biophysical tools were largely missing when this research project was initiated. We and others addressed this need, and I have described in chapter 6. and 7. how developed tools can be utilized to study the mode of action of an inhibitor.

Condensed, there are a few critical aspects which significantly improved the characterization of RdRp inhibitors. Frist, an active protein complex in a near native state is a key requirement for every meaningful biochemical assay. Producing a stable RdRp complex without RNA was a serious challenge. Switching from reconstitution of individual subunits to a co-expression strategy reported by Madru et. al.[233], resulted in a significant improvement in the behavior and activity of the RdRp complex. Next, it was crucial to expand the repertoire of biochemical assays from low throughput gel-based assays to high-throughput plate based assays. The established and optimized FP assay allowed us to measure the interaction between RdRp and the RNA substrate, revealing insights into the mechanism of inhibition. Building on comprehensive biochemical characterization of the inhibitors, cryo-EM could be utilized to visualize the small molecule bound to RdRp. A key advancement here was the introduction of a mass photometry-based quality control step. Analyzing the behavior of samples before subjecting them to cryo-EM allowed elimination of conditions that cause protein aggregation and led to efficient usage of microscopy time.

Summarized, the here described pipeline presents an efficient way for biochemical and structural characterization of RdRp inhibitor in a highly defined *in vitro* system. Such simplified systems are important to deduce molecular mechanisms of inhibition, but the ultimate goal is that small molecule inhibitors are applicable to cells or even patients. Thus, drug development against corona virus RdRp should be a holistic endeavor combining simple but powerful *in vitro* systems with a more complex near native replication system. This could be achieved employing replicon systems of SARS-CoV-2, which omit infectious viral particles but include all key steps of the viral life cycle.

Taha et. al. [260] recently reported the development of a versatile SARS-CoV-2 replicon system based on a bacterial artificial chromosome (BAC) vector containing the full-length

genome of SARS-CoV-2. The authors replaced the spike protein of SARS-CoV-2 with a reporter gene, which prevents the formation of infectious particles and allows plate-based quantification of viral replication. Thus, such a replicon system is well suited to study the effects of potential small molecule inhibitors on viral replication in a native environment. In theory, this system could be also used to screen many small molecules for their inhibitory potential. However, false-positive hits have to be accepted due to cell-based setup. Thus, I propose using a defined biochemical assay during the screening stage and then testing promising molecules in the replicon system to study them further.

Besides the hunt for inhibitors, it will be important to tackle fundamental questions about the biology of corona viruses. Despite extensive efforts of laboratories around the world, we still lack the complete picture of replication and transcription of viral RNA within host cells. Reconstitution of RTC subcomplexes has yielded important biochemical and strucutral insights, but in my opinion, it is unlikely that a complete RTC can be reconstituted *in vitro*. Thus, it would be interesting to use the replicon system in order to study the RTC in its native environment. This could be attempted by infecting cells with the replicon and then subjecting them to cryoET. In addition, due the versatility of the system, it could be fruitful to add biochemical affinity tags to certain nsps and then perform pull-down experiments enriching for the RTC.

It remains to be seen, if we will ever be able to visualize a complete RTC, but I am certain, we will learn more about the molecular tricks of viral RNA replication.

8. References

1. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
2. Coronaviridae Study Group of the International Committee on Taxonomy of, V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536-544 (2020).
3. World Health Organization 2023 data.who.int, WHO Coronavirus (COVID-19) dashboard > Deaths. (https://data.who.int/dashboards/covid19/deaths).
4. Masters, P.S. The molecular biology of coronaviruses. *Adv Virus Res* **66**, 193-292 (2006).
5. Snijder, E.J., Decroly, E. & Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res* **96**, 59-126 (2016).
6. Zhu, W. et al. RNA-Dependent RNA Polymerase as a Target for COVID-19 Drug Discovery. *SLAS Discov* **25**, 1141-1151 (2020).
7. Dolgin, E. The race for antiviral drugs to beat COVID - and the next pandemic. *Nature* **592**, 340-343 (2021).
8. Bekheit, M.S., Panda, S.S. & Girgis, A.S. Potential RNA-dependent RNA polymerase (RdRp) inhibitors as prospective drug candidates for SARS-CoV-2. *Eur J Med Chem* **252**, 115292 (2023).
9. 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, Merck & RidgebackBio, April 15, 2021. (2021).
10. 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).
11. 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).
12. 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**(2020).
13. Wahl, A. et al. SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature* **591**, 451-457 (2021).
14. Tarantino, D. et al. Targeting flavivirus RNA dependent RNA polymerase through a pyridobenzothiazole inhibitor. *Antiviral Res* **134**, 226-235 (2016).
15. Dejmek, M. et al. Non-Nucleotide RNA-Dependent RNA Polymerase Inhibitor That Blocks SARS-CoV-2 Replication. *Viruses* **13**(2021).
16. Gao, Y. et al. The "Regulator" Function of Viruses on Ecosystem Carbon Cycling in the Anthropocene. *Front Public Health* **10**, 858615 (2022).
17. van Regenmortel, M.H. & Mahy, B.W. Emerging issues in virus taxonomy. *Emerg Infect Dis* **10**, 8-13 (2004).
18. Summers, W.C. Virus Infection. in *Encyclopedia of Microbiology (Third Edition)* (ed. Schaechter, M.) 546-552 (Academic Press, Oxford, 2009).

19. Harris, H.M.B. & Hill, C. A Place for Viruses on the Tree of Life. *Front Microbiol* **11**, 604048 (2020).
20. Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. & Hatfull, G.F. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* **96**, 2192-7 (1999).
21. Mushegian, A.R. Are There 10(31) Virus Particles on Earth, or More, or Fewer? *J Bacteriol* **202**(2020).
22. Edwards, R.A. & Rohwer, F. Viral metagenomics. *Nat Rev Microbiol* **3**, 504-10 (2005).
23. Woolhouse, M., Scott, F., Hudson, Z., Howey, R. & Chase-Topping, M. Human viruses: discovery and emergence. *Philos Trans R Soc Lond B Biol Sci* **367**, 2864-71 (2012).
24. Forni, D., Cagliani, R., Clerici, M. & Sironi, M. Disease-causing human viruses: novelty and legacy. *Trends Microbiol* **30**, 1232-1242 (2022).
25. Olival, K.J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646-650 (2017).
26. Wolfe, N.D., Dunavan, C.P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447**, 279-83 (2007).
27. Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257-261 (2020).
28. Network, T.W.E.R.S. Estimated number of lives directly saved by COVID-19 vaccination programs in the WHO European Region, December 2020 to March 2023. *medRxiv* (2024).
29. Xie, Y., Choi, T. & Al-Aly, Z. Risk of Death in Patients Hospitalized for COVID-19 vs Seasonal Influenza in Fall-Winter 2022-2023. *JAMA* **329**, 1697-1699 (2023).
30. Sanchez, C.A., Venkatachalam-Vaz, J. & Drake, J.M. Spillover of zoonotic pathogens: A review of reviews. *Zoonoses Public Health* **68**, 563-577 (2021).
31. Batool, M., Ahmad, B. & Choi, S. A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **20**(2019).
32. Hilgenfeld, R. & Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res* **100**, 286-95 (2013).
33. Pasternak, A.O., Spaan, W.J.M. & Snijder, E.J. Nidovirus transcription: how to make sense...? *J Gen Virol* **87**, 1403-1421 (2006).
34. de Vries, A.A.F., Horzinek, M.C., Rottier, P.J.M. & de Groot, R.J. The Genome Organization of the Nidovirales: Similarities and Differences between Arteri-, Toro-, and Coronaviruses. *Semin Virol* **8**, 33-47 (1997).
35. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **3**, 237-261 (2016).
36. Turonova, B. et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203-208 (2020).
37. Almeida, J.D. & Tyrrell, D.A. The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. *J Gen Virol* **1**, 175-8 (1967).
38. Chen, B. et al. Overview of lethal human coronaviruses. *Signal Transduct Target Ther* **5**, 89 (2020).
39. Steiner, S. et al. SARS-CoV-2 biology and host interactions. *Nat Rev Microbiol* **22**, 206-225 (2024).
40. Goyal, R. et al. Comparative highlights on MERS-CoV, SARS-CoV-1, SARS-CoV-2, and NEO-CoV. *EXCLI J* **21**, 1245-1272 (2022).

41. Su, S. et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol* **24**, 490-502 (2016).
42. V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155-170 (2021).
43. Yang, D. & Leibowitz, J.L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res* **206**, 120-33 (2015).
44. Zust, R., Miller, T.B., Goebel, S.J., Thiel, V. & Masters, P.S. Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *J Virol* **82**, 1214-28 (2008).
45. Miao, Z., Tidu, A., Eriani, G. & Martin, F. Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol* **18**, 447-456 (2021).
46. Zhao, J., Qiu, J., Aryal, S., Hackett, J.L. & Wang, J. The RNA Architecture of the SARS-CoV-2 3'-Untranslated Region. *Viruses* **12**(2020).
47. Hartenian, E. et al. The molecular virology of coronaviruses. *J Biol Chem* **295**, 12910-12934 (2020).
48. Snijder, E.J. et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* **331**, 991-1004 (2003).
49. Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* **7**, 439-50 (2009).
50. Finkel, Y. et al. The coding capacity of SARS-CoV-2. *Nature* **589**, 125-130 (2021).
51. Thoms, M. et al. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **369**, 1249-1255 (2020).
52. Schubert, K. et al. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol* **27**, 959-966 (2020).
53. Lee, H.J. et al. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567-82 (1991).
54. Wolff, G. et al. A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395-1398 (2020).
55. Angelini, M.M., Akhlaghpour, M., Neuman, B.W. & Buchmeier, M.J. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio* **4**(2013).
56. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. & Blinov, V.M. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res* **17**, 4847-61 (1989).
57. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-7 (2003).
58. 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).
59. Subissi, L. et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci U S A* **111**, E3900-9 (2014).
60. Shannon, A. et al. Protein-primed RNA synthesis in SARS-CoVs and structural basis for inhibition by AT-527. (2021).
61. Slanina, H. et al. Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. *Proc Natl Acad Sci U S A* **118**(2021).

62. Park, G.J. et al. The mechanism of RNA capping by SARS-CoV-2. *Nature* **609**, 793-800 (2022).
63. Bouvet, M. et al. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem* **289**, 25783-96 (2014).
64. Decroly, E. et al. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog* **7**, e1002059 (2011).
65. Liu, C. et al. Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science* **373**, 1142-1146 (2021).
66. Seybert, A., Hegyi, A., Siddell, S.G. & Ziebuhr, J. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA* **6**, 1056-68 (2000).
67. Chen, J. et al. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **182**, 1560-1573 e13 (2020).
68. Ivanov, K.A. et al. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol* **78**, 5619-32 (2004).
69. Minskaia, E. et al. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A* **103**, 5108-13 (2006).
70. 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 U S A* **109**, 9372-7 (2012).
71. Bouvet, M. et al. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* **6**, e1000863 (2010).
72. Ivanov, K.A. et al. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc Natl Acad Sci U S A* **101**, 12694-9 (2004).
73. Pillon, M.C. et al. Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat Commun* **12**, 636 (2021).
74. Decroly, E. et al. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J Virol* **82**, 8071-84 (2008).
75. Hoffmann, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e8 (2020).
76. Delmas, B. & Laude, H. Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J Virol* **64**, 5367-75 (1990).
77. Venkatagopalan, P., Daskalova, S.M., Lopez, L.A., Dolezal, K.A. & Hogue, B.G. Coronavirus envelope (E) protein remains at the site of assembly. *Virology* **478**, 75-85 (2015).
78. Neuman, B.W. et al. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* **174**, 11-22 (2011).
79. Chang, C.K. et al. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci* **13**, 59-72 (2006).
80. Wu, W., Cheng, Y., Zhou, H., Sun, C. & Zhang, S. The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Virol J* **20**, 6 (2023).
81. Perry, J.K. et al. An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15. *J Biol Chem* **297**, 101218 (2021).
82. Savastano, A., Ibanez de Opakua, A., Rankovic, M. & Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat Commun* **11**, 6041 (2020).

83. Redondo, N., Zaldivar-Lopez, S., Garrido, J.J. & Montoya, M. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Front Immunol* **12**, 708264 (2021).
84. Tortorici, M.A. & Veesler, D. Structural insights into coronavirus entry. *Adv Virus Res* **105**, 93-116 (2019).
85. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562-569 (2020).
86. Bestle, D. et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance* **3**(2020).
87. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221-224 (2020).
88. Bhatt, P.R. et al. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306-1313 (2021).
89. Knoops, K. et al. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol* **6**, e226 (2008).
90. Sawicki, S.G. & Sawicki, D.L. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv Exp Med Biol* **380**, 499-506 (1995).
91. Klein, S. et al. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885 (2020).
92. Ghosh, S. et al. beta-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* **183**, 1520-1535 e14 (2020).
93. Malone, B., Urakova, N., Snijder, E.J. & Campbell, E.A. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat Rev Mol Cell Biol* **23**, 21-39 (2022).
94. Drake, J.W. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci U S A* **90**, 4171-5 (1993).
95. Agranovsky, A.A. Structure and Expression of Large (+)RNA Genomes of Viruses of Higher Eukaryotes. *Biochemistry (Mosc)* **86**, 248-261 (2021).
96. Sevajol, M., Subissi, L., Decroly, E., Canard, B. & Imbert, I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res* **194**, 90-9 (2014).
97. Zimmermann, L. et al. SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a pore spanning replication organelle. *Nat Commun* **14**, 7894 (2023).
98. Gao, Y. et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779-782 (2020).
99. Hillen, H.S. et al. Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154-156 (2020).
100. Wang, Q. et al. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **182**, 417-428 e13 (2020).
101. 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**, 8416-34 (2015).
102. Zamyatkin, D.F., Parra, F., Machin, A., Grochulski, P. & Ng, K.K. Binding of 2'-amino-2'-deoxycytidine-5'-triphosphate to norovirus polymerase induces rearrangement of the active site. *J Mol Biol* **390**, 10-6 (2009).
103. Hillen, H.S. Structure and function of SARS-CoV-2 polymerase. *Curr Opin Virol* **48**, 82-90 (2021).

104. Malone, B.F. et al. Structural basis for substrate selection by the SARS-CoV-2 replicase. *Nature* **614**, 781-787 (2023).
105. Shannon, A. et al. A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. *Nat Commun* **13**, 621 (2022).
106. Lehmann, K.C., Snijder, E.J., Posthuma, C.C. & Gorbalenya, A.E. What we know but do not understand about nidovirus helicases. *Virus Res* **202**, 12-32 (2015).
107. Malone, B. et al. Structural basis for backtracking by the SARS-CoV-2 replication–transcription complex. *Proceedings of the National Academy of Sciences* **118**, e2102516118 (2021).
108. Wang, D. et al. The SARS-CoV-2 subgenome landscape and its novel regulatory features. *Mol Cell* **81**, 2135-2147 e5 (2021).
109. Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D. & Baric, R.S. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol* **8**, 270-9 (2011).
110. Sydow, J.F. & Cramer, P. RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol* **19**, 732-9 (2009).
111. Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. & Snijder, E.J. Nidovirales: evolving the largest RNA virus genome. *Virus Res* **117**, 17-37 (2006).
112. Robson, F. et al. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol Cell* **79**, 710-727 (2020).
113. Yan, L. et al. Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell* **184**, 3474-3485 e11 (2021).
114. Ramanathan, A., Robb, G.B. & Chan, S.H. mRNA capping: biological functions and applications. *Nucleic Acids Res* **44**, 7511-26 (2016).
115. Ferron, F., Decroly, E., Selisko, B. & Canard, B. The viral RNA capping machinery as a target for antiviral drugs. *Antiviral Res* **96**, 21-31 (2012).
116. Zust, R. et al. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol* **12**, 137-43 (2011).
117. De Vlugt, C., Sikora, D. & Pelchat, M. Insight into Influenza: A Virus Cap-Snatching. *Viruses* **10**(2018).
118. Otter, C.J. et al. SARS-CoV-2 nsp15 endoribonuclease antagonizes dsRNA-induced antiviral signaling. *Proc Natl Acad Sci U S A* **121**, e2320194121 (2024).
119. Griffin, I. et al. Estimates of SARS-CoV-2 Hospitalization and Fatality Rates in the Prevaccination Period, United States. *Emerg Infect Dis* **30**, 1144-1153 (2024).
120. Schilling, J. et al. [The different periods of COVID-19 in Germany: a descriptive analysis from January 2020 to February 2021]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **64**, 1093-1106 (2021).
121. Yang, X. et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* **8**, 475-481 (2020).
122. Rieg, S. et al. COVID-19 in-hospital mortality and mode of death in a dynamic and non-restricted tertiary care model in Germany. *PLoS One* **15**, e0242127 (2020).
123. Grasselli, G. et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* **323**, 1574-1581 (2020).

124. Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat Rev Microbiol* **19**, 685-700 (2021).
125. Lei, S., Chen, X., Wu, J., Duan, X. & Men, K. Small molecules in the treatment of COVID-19. *Signal Transduct Target Ther* **7**, 387 (2022).
126. Aiello, T.F., Garcia-Vidal, C. & Soriano, A. Antiviral drugs against SARS-CoV-2. *Rev Esp Quimioter* **35 Suppl 3**, 10-15 (2022).
127. Tanne, J.H. Covid-19: FDA authorises pharmacists to prescribe Paxlovid. *BMJ* **378**, o1695 (2022).
128. Mahase, E. Covid-19: Pfizer's paxlovid is 89% effective in patients at risk of serious illness, company reports. *BMJ* **375**, n2713 (2021).
129. Amani, B. & Amani, B. Efficacy and safety of nirmatrelvir/ritonavir (Paxlovid) for COVID-19: A rapid review and meta-analysis. *J Med Virol* **95**, e28441 (2023).
130. Seley-Radtke, K.L. & Yates, M.K. The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res* **154**, 66-86 (2018).
131. Xu, X. et al. An update on inhibitors targeting RNA-dependent RNA polymerase for COVID-19 treatment: Promises and challenges. *Biochem Pharmacol* **205**, 115279 (2022).
132. Shannon, A. & Canard, B. Kill or corrupt: Mechanisms of action and drug-resistance of nucleotide analogues against SARS-CoV-2. *Antiviral Res* **210**, 105501 (2023).
133. Sinokrot, H., Smerat, T., Najjar, A. & Karaman, R. Advanced Prodrug Strategies in Nucleoside and Non-Nucleoside Antiviral Agents: A Review of the Recent Five Years. *Molecules* **22**(2017).
134. Kamzeeva, P.N., Aralov, A.V., Alferova, V.A. & Korshun, V.A. Recent Advances in Molecular Mechanisms of Nucleoside Antivirals. *Curr Issues Mol Biol* **45**, 6851-6879 (2023).
135. Govender, K. & Chuturgoon, A. An Overview of Repurposed Drugs for Potential COVID-19 Treatment. *Antibiotics (Basel)* **11**(2022).
136. Warren, T.K. et al. Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* **531**, 381-5 (2016).
137. 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).
138. Kokic, G. et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun* **12**, 279 (2021).
139. Pan, H. et al. Repurposed antiviral drugs for COVID-19; interim WHO SOLIDARITY trial results. *medRxiv*, 2020.10.15.20209817 (2020).
140. Blair, H.A. Remdesivir: A Review in COVID-19. *Drugs* **83**, 1215-1237 (2023).
141. Furuta, Y., Komeno, T. & Nakamura, T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc Jpn Acad Ser B Phys Biol Sci* **93**, 449-463 (2017).
142. Peng, Q. et al. Structural Basis of SARS-CoV-2 Polymerase Inhibition by Favipiravir. *Innovation (Camb)* **2**, 100080 (2021).
143. Zhao, L. & Zhong, W. Mechanism of action of favipiravir against SARS-CoV-2: Mutagenesis or chain termination? *Innovation (Camb)* **2**, 100165 (2021).
144. Batool, S. et al. Efficacy and Safety of Favipiravir in Treating COVID-19 Patients: A Meta-Analysis of Randomized Control Trials. *Cureus* **15**, e33676 (2023).
145. Horga, A. et al. Oral bemnifosbuvir (AT-527) vs placebo in patients with mild-to-moderate COVID-19 in an outpatient setting (MORNINGSKY). *Future Virol* (2023).

146. Agostini, M.L. et al. Small-Molecule Antiviral beta-d-N (4)-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J Virol* **93**(2019).
147. Urakova, N. et al. beta-d-N(4)-Hydroxycytidine Is a Potent Anti-alphavirus Compound That Induces a High Level of Mutations in the Viral Genome. *J Virol* **92**(2018).
148. Focosi, D. Molnupiravir: From Hope to Epic Fail? *Viruses* **14**(2022).
149. Denel-Bobrowska, M. & Olejniczak, A.B. Non-nucleoside structured compounds with antiviral activity-past 10 years (2010-2020). *Eur J Med Chem* **231**, 114136 (2022).
150. Alici, H., Tahtaci, H. & Demir, K. Design and various in silico studies of the novel curcumin derivatives as potential candidates against COVID-19 -associated main enzymes. *Comput Biol Chem* **98**, 107657 (2022).
151. El Bakri, Y., Musrat Kurbanova, M., Ali Siddique, S., Ahmad, S. & Goumri-Said, S. One-pot synthesis, X-ray crystal structure, and identification of potential molecules against COVID-19 main protease through structure-guided modeling and simulation approach. *Arab J Chem* **15**, 104230 (2022).
152. Girgis, A.S. et al. Rational design, synthesis, and 2D-QSAR study of anti-oncological alkaloids against hepatoma and cervical carcinoma. *RSC Advances* **5**, 28554-28569 (2015).
153. Mun, C.S., Hui, L.Y., Sing, L.C., Karunakaran, R. & Ravichandran, V. Multi-targeted molecular docking, pharmacokinetics, and drug-likeness evaluation of coumarin based compounds targeting proteins involved in development of COVID-19. *Saudi J Biol Sci* **29**, 103458 (2022).
154. Nabati, F. et al. Virtual screening based on the structure of more than 10(5) compounds against four key proteins of SARS-CoV-2: M(Pro), S(RBD), RdRp, and PL(pro). *Inform Med Unlocked* **35**, 101134 (2022).
155. Nagar, P.R., Gajjar, N.D. & Dhameliya, T.M. In search of SARS CoV-2 replication inhibitors: Virtual screening, molecular dynamics simulations and ADMET analysis. *J Mol Struct* **1246**, 131190 (2021).
156. Rabie, A.M. Two antioxidant 2,5-disubstituted-1,3,4-oxadiazoles (CoViTris2020 and ChloViD2020): successful repurposing against COVID-19 as the first potent multitarget anti-SARS-CoV-2 drugs. *New Journal of Chemistry* **45**, 761-771 (2021).
157. Ramalingam, A. et al. Synthesis, spectroscopic, topological, hirshfeld surface analysis, and anti-covid-19 molecular docking investigation of isopropyl 1-benzoyl-4-(benzoyloxy)-2,6-diphenyl-1,2,5,6-tetrahydropyridine-3-carboxylate. *Heliyon* **8**, e10831 (2022).
158. Mehlhorn, H. *Encyclopedia of Parasitology*, (Springer, Berlin, 2008).
159. Brun, R., Blum, J., Chappuis, F. & Burri, C. Human African trypanosomiasis. *Lancet* **375**, 148-59 (2010).
160. Hawking, F. Chemotherapy of onchocerciasis. *Trans R Soc Trop Med Hyg* **52**, 109-11 (1958).
161. Yin, W. et al. Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. *Nature Structural & Molecular Biology* **28**, 319-325 (2021).
162. Dey, D., Ramakumar, S. & Conn, G.L. Targeted Redesign of Suramin Analogs for Novel Antimicrobial Lead Development. *J Chem Inf Model* **61**, 4442-4454 (2021).
163. Boniardi, I. et al. Suramin inhibits SARS-CoV-2 nucleocapsid phosphoprotein genome packaging function. *Virus Res* **336**, 199221 (2023).

164. Kwon, P.S. et al. Suramin binds and inhibits infection of SARS-CoV-2 through both spike protein-heparan sulfate and ACE2 receptor interactions. *Communications Biology* **6**(2023).
165. 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).
166. 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**(2020).
167. Jiang, Y., Yin, W. & Xu, H.E. RNA-dependent RNA polymerase: Structure, mechanism, and drug discovery for COVID-19. *Biochem Biophys Res Commun* (2020).
168. 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* (2020).
169. Vicenti, I., Zazzi, M. & Saladini, F. SARS-CoV-2 RNA-dependent RNA polymerase as a therapeutic target for COVID-19. *Expert Opin Ther Pat*, 1-13 (2021).
170. 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).
171. Gao, Y. et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, eabb7498 (2020).
172. Yin, W. et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* (2020).
173. Wang, Q. et al. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* (2020).
174. Jochheim, F.A. et al. Dimeric form of SARS-CoV-2 polymerase. *bioRxiv*, 2021.03.23.436644 (2021).
175. Ferron, F. et al. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci U S A* **115**, E162-E171 (2018).
176. 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* (2020).
177. 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 e4 (2021).
178. Beigel, J.H. et al. Remdesivir for the Treatment of Covid-19 - Preliminary Report. *N Engl J Med* (2020).
179. Grein, J. et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N Engl J Med* **382**, 2327-2336 (2020).
180. 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).
181. Consortium, W.H.O.S.T. et al. Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. *N Engl J Med* **384**, 497-511 (2021).
182. 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. *Antiviral Res* **171**, 104597 (2019).
183. Yoon, J.J. et al. Orally Efficacious Broad-Spectrum Ribonucleoside Analog Inhibitor of Influenza and Respiratory Syncytial Viruses. *Antimicrob Agents Chemother* **62**(2018).
184. Ehteshami, M. et al. Characterization of beta-d-N(4)-Hydroxycytidine as a Novel Inhibitor of Chikungunya Virus. *Antimicrob Agents Chemother* **61**(2017).

185. Reynard, O. et al. Identification of a New Ribonucleoside Inhibitor of Ebola Virus Replication. *Viruses* **7**, 6233-40 (2015).
186. Costantini, V.P. et al. Antiviral activity of nucleoside analogues against norovirus. *Antivir Ther* **17**, 981-91 (2012).
187. Stuyver, L.J. et al. Ribonucleoside analogue that blocks replication of bovine viral diarrhea and hepatitis C viruses in culture. *Antimicrob Agents Chemother* **47**, 244-54 (2003).
188. 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* (2021).
189. Crotty, S., Cameron, C.E. & Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci U S A* **98**, 6895-900 (2001).
190. Urakova, N. et al. beta-d-N (4)-Hydroxycytidine Is a Potent Anti-alphavirus Compound That Induces a High Level of Mutations in the Viral Genome. *J Virol* **92**(2018).
191. Les, A., Adamowicz, L. & Rode, W. Structure and conformation of N4-hydroxycytosine and N4-hydroxy-5-fluorocytosine. A theoretical ab initio study. *Biochim Biophys Acta* **1173**, 39-48 (1993).
192. 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).
193. 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).
194. Naydenova, K. et al. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc Natl Acad Sci U S A* **118**(2021).
195. Malone, B. et al. Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *bioRxiv*, 2021.03.13.435256 (2021).
196. Suzuki, T., Moriyama, K., Otsuka, C., Loakes, D. & Negishi, K. Template properties of mutagenic cytosine analogues in reverse transcription. *Nucleic Acids Res* **34**, 6438-49 (2006).
197. Sticher, Z.M. et al. Analysis of the Potential for N (4)-Hydroxycytidine To Inhibit Mitochondrial Replication and Function. *Antimicrob Agents Chemother* **64**(2020).
198. Zhou, S. et al. beta-D-N 4-hydroxycytidine (NHC) Inhibits SARS-CoV-2 Through Lethal Mutagenesis But Is Also Mutagenic To Mammalian Cells. *J Infect Dis* (2021).
199. Gordon, C.J., Tchesnokov, E.P., Schinazi, R.F. & Gotte, M. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J Biol Chem*, 100770 (2021).
200. 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**, 6171-80 (2013).
201. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
202. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152 (2019).
203. 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).

204. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
205. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
206. 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 Biol Crystallogr* **65**, 1074-80 (2009).
207. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).
208. Williams, C.J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315 (2018).
209. Pettersen, E.F. et al. UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. *Protein Sci* (2020).
210. 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).
211. Tchesnokov, E.P. et al. Template-dependent inhibition of coronavirus RNA-dependent RNA polymerase by remdesivir reveals a second mechanism of action. *J Biol Chem* **295**, 16156-16165 (2020).
212. Kabinger, F. et al. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* **28**, 740-746 (2021).
213. 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).
214. U.S. Food and Drug Administration, C.f.D.E.a.R. Emergency Use Authorization (EUA) for the emergency use of Veklury® (remdesivir). (retrieved from https://www.fda.gov/media/137564/download, 2022).
215. U.S. Food and Drug Administration, C.f.D.E.a.R. Emergency Use Authorization (EUA) for the emergency use of LAGEVRIO (molnupiravir). (retrieved from: https://www.fda.gov/media/155053/download, 2023).
216. Yasir, M. et al. An Updated Systematic Review on Remdesivir's Safety and Efficacy in Patients Afflicted With COVID-19. *Cureus* **15**, e43060 (2023).
217. Sanderson, T. et al. A molnupiravir-associated mutational signature in global SARS-CoV-2 genomes. *Nature* **623**, 594-600 (2023).
218. Milan Bonotto, R. et al. Virucidal Activity of the Pyridobenzothiazolone Derivative HeE1-17Y against Enveloped RNA Viruses. *Viruses* **14**(2022).
219. Cannalire, R. et al. Broad spectrum anti-flavivirus pyridobenzothiazolones leading to less infective virions. *Antiviral Res* **167**, 6-12 (2019).
220. Arkin, M.R., Glicksman, M.A., Fu, H., Havel, J.J. & Du, Y. Inhibition of protein-protein interactions: non-cellular assay formats. *Assay Guidance Manual [Internet]* (2012).
221. Wilamowski, M. et al. Transient and stabilized complexes of Nsp7, Nsp8, and Nsp12 in SARS-CoV-2 replication. *Biophys J* **120**, 3152-3165 (2021).
222. Tan, Y.Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796 (2017).
223. Huber, R.G. et al. Heteroaromatic pi-stacking energy landscapes. *J Chem Inf Model* **54**, 1371-9 (2014).
224. Janiak, C. A critical account on π–π stacking in metal complexes with aromatic nitrogen-containing ligands †. *Journal of the Chemical Society, Dalton Transactions*, 3885-3896 (2000).

225. Kumar, K. et al. Cation-pi interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem Sci* **9**, 2655-2665 (2018).
226. Cortopassi, W.A., Kumar, K. & Paton, R.S. Cation-pi interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity. *Org Biomol Chem* **14**, 10926-10938 (2016).
227. Gallivan, J.P. & Dougherty, D.A. Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* **96**, 9459-64 (1999).
228. Anderson, T.K., Hoferle, P.J., Lee, K.W., Coon, J.J. & Kirchdoerfer, R.N. An alphacoronavirus polymerase structure reveals conserved co-factor functions. *bioRxiv* (2023).
229. Stornaiuolo, M. et al. Assembly of a pi-pi stack of ligands in the binding site of an acetylcholine-binding protein. *Nat Commun* **4**, 1875 (2013).
230. Merz, G.E. et al. Stacked binding of a PET ligand to Alzheimer's tau paired helical filaments. *Nat Commun* **14**, 3048 (2023).
231. Stefan, M.I. & Le Novere, N. Cooperative binding. *PLoS Comput Biol* **9**, e1003106 (2013).
232. Tóth, L.J. et al. Novel analogues of a nonnucleoside SARS-CoV-2 RdRp inhibitor as potential antivirotics. *Beilstein Journal of Organic Chemistry* **20**, 1029-1036 (2024).
233. Madru, C. et al. Fast and efficient purification of SARS-CoV-2 RNA dependent RNA polymerase complex expressed in Escherichia coli. *PLoS One* **16**, e0250610 (2021).
234. Croll, T.I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530 (2018).
235. Meng, E.C. et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **32**, e4792 (2023).
236. Jubb, H.C. et al. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol* **429**, 365-371 (2017).
237. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-34 (1995).
238. Hatcher, E.L. et al. Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic Acids Res* **45**, D482-D490 (2017).
239. Madeira, F. et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* (2024).
240. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
241. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-91 (2009).
242. Bessi, I. et al. The Tautomeric State of N(4)-Hydroxycytidine within Base-Paired RNA. *ACS Cent Sci* **10**, 1084-1093 (2024).
243. Zibat, A. et al. N4-hydroxycytidine, the active compound of Molnupiravir, promotes SARS-CoV-2 mutagenesis and escape from a neutralizing nanobody. *iScience* **26**, 107786 (2023).
244. Alteri, C. et al. Genomic Evolution of Sars-Cov-2 in Molnupiravir-Treated Patients Compared to Paxlovid-Treated and Drug-Naïve Patients: A Proof-of-Concept Study. *Preprint* (2022).
245. Strizki, J.M. et al. Molnupiravir maintains antiviral activity against SARS-CoV-2 variants and exhibits a high barrier to the development of resistance. *Antimicrob Agents Chemother* **68**, e0095323 (2024).

246. Stegmann, K.M. et al. Inhibitors of dihydroorotate dehydrogenase cooperate with molnupiravir and N4-hydroxycytidine to suppress SARS-CoV-2 replication. *iScience* **25**, 104293 (2022).
247. Bruce Alberts, A.J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular biology of the cell (4th ed.)*, 212-214 (Garland Science, New York, 2002).
248. Swanstrom, R. & Schinazi, R.F. Lethal mutagenesis as an antiviral strategy. *Science* **375**, 497-498 (2022).
249. Painter, G.R., Natchus, M.G., Cohen, O., Holman, W. & Painter, W.P. Developing a direct acting, orally available antiviral agent in a pandemic: the evolution of molnupiravir as a potential treatment for COVID-19. *Curr Opin Virol* **50**, 17-22 (2021).
250. Bohr, C., Hasselbalch, K. & Krogh, A. Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt 1. *Skandinavisches Archiv Für Physiologie* **16**, 402-412 (1904).
251. Jin, L., Wang, W. & Fang, G. Targeting protein-protein interaction by small molecules. *Annu Rev Pharmacol Toxicol* **54**, 435-56 (2014).
252. Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* **22**, 510-5 (2013).
253. Sun, H., Tawa, G. & Wallqvist, A. Classification of scaffold-hopping approaches. *Drug Discov Today* **17**, 310-24 (2012).
254. Shafer, R.W. & Vuitton, D.A. Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. *Biomed Pharmacother* **53**, 73-86 (1999).
255. Boucau, J. et al. Characterization of Virologic Rebound Following Nirmatrelvir-Ritonavir Treatment for Coronavirus Disease 2019 (COVID-19). *Clin Infect Dis* **76**, e526-e529 (2023).
256. Smith, D.J., Lambrou, A. & Patel, P. SARS-CoV-2 Rebound With and Without Use of COVID-19 Oral Antivirals. *MMWR Morb Mortal Wkly Rep* **72**, 1357-1364 (2023).
257. Edelstein, G.E. et al. SARS-CoV-2 Virologic Rebound With Nirmatrelvir-Ritonavir Therapy : An Observational Study. *Ann Intern Med* **176**, 1577-1585 (2023).
258. Perelson, A.S., Ribeiro, R.M. & Phan, T. An explanation for SARS-CoV-2 rebound after Paxlovid treatment. *medRxiv* (2023).
259. Ranard, B.L. et al. A mathematical model of SARS-CoV-2 immunity predicts paxlovid rebound. *J Med Virol* **95**, e28854 (2023).
260. Taha, T.Y. et al. Rapid assembly of SARS-CoV-2 genomes reveals attenuation of the Omicron BA.1 variant through NSP6. *Nat Commun* **14**, 2308 (2023).

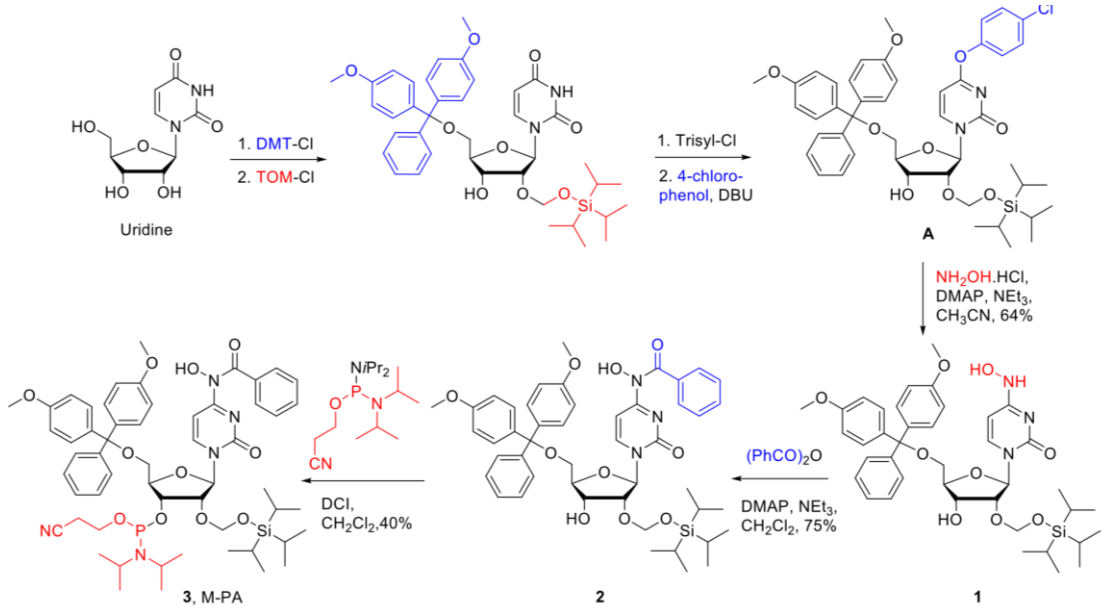
9. List of figures

Figure 1: Genome organization of SARS-CoV-2 13
Figure 2: The life cycle of SARS-CoV-2 17
Figure 3: Structure of replicating SARS-CoV-2 polymerase 20
Figure 4: Promissing nucleoside analogues against SARS-CoV-2 RdRp 27
Figure 5: Promissing non-nucleoside analogues against SARS-CoV-2 RdRp 30
Figure 6: RdRp incorporates NHC opposite G and A in the template 35
Figure 7: NHC incorporation does not stall SARS-CoV-2 RdRp 36
Figure 8: NHC can direct incorporation of G and A into RNA 37
Figure 9: Structures of RdRp–RNA product complexes after NHC-induced mutagenesis 39
Figure 10: Two-step model of molnupiravir-induced RNA mutagenesis 41
Figure 11: HeE1-2Tyr is a potent non-nucleoside inhibitor of the RdRp of SARS-CoV-2 53
Figure 12: HeE1-2Tyr prevents the interaction between RdRp and RNA 55
Figure 13: Structure of RdRp bound to a stack of three HeE1-2Tyr molecules 61
Figure 14: Molecular details of the interaction between HeE1-2Tyr and RdRp RNA binding site 63
Figure 15: The importance of the arginine bracket for RdRp function and its interplay with HeE1-2Tyr 65

Extended data figure 1: Cryo-EM data processing trees and quality of reconstructions 40
Extended data figure 2: Comparison of the enzymatic activity of RdRp co-expressed and reconstituted from individual subunits 54
Extended data figure 3: Development of a fluorescence polarization (FP) assay to monitor the interaction between RdRp and RNA 55
Extended data figure 4: Effect of HeE1-2Tyr on the RdRp-RNA interaction over time 56
Extended data figure 5: Mass photometry based analysis of RdRp reconstituted from individual subunits and co-expressed 57
Extended data figure 6: Cryo-EM data processing trees and quality of reconstructions 58
Extended data figure 7: Difference map and map quality 59
Extended data figure 8: Conformational changes of HeE1-2Tyr$_{1-3}$ 61
Extended data figure 9: Comparison of the activity of RdRp mutants 66
Extended data figure 10: Structural comparison of the RNA binding site in corona viruses 68

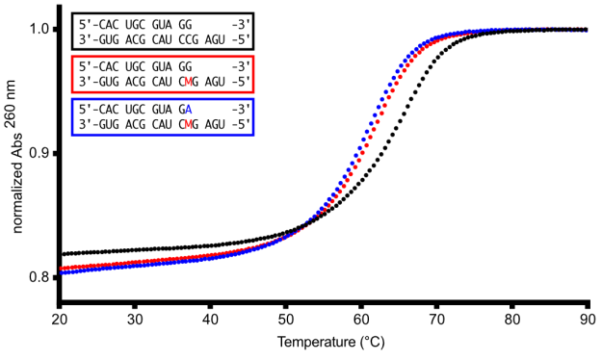
10. Appendix

Appendix 1: Synthesis of NHC-phosphoramidite 3 (M-PA). (Manuscript 1).....102
Appendix 2: Melting curves for RNA duplexes containing M–G or M–A base pairs. (Manuscript 1).....103
Appendix 3: Protein preparation and RNA scaffold for structural studies. (Manuscript 1).....103
Appendix 4: Cryo-EM data collection, refinement and validation statistics. (Manuscript 1) 104
Appendix 5: Supplementary Table RNA substrates (Manuscript 1).....105
Appendix 6: Supplementary Table Thermal melting analysis (Manuscript 1) 105
Appendix 7: Supplementary Table Sequences and high-resolution ESI-MS data of NHC/Mmodified RNA oligonucleotides (Manuscript 1) 105



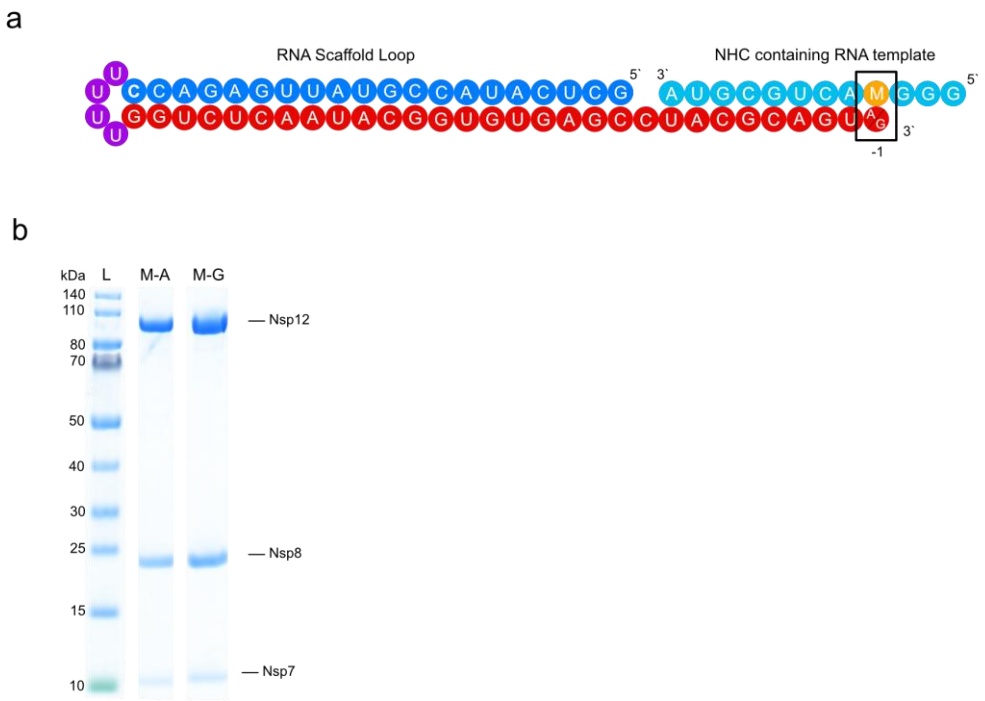
Appendix 1: Synthesis of NHC-phosphoramidite 3 (M-PA). (Manuscript 1)

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.



Appendix 2: Melting curves for RNA duplexes containing M–G or M–A base pairs. (Manuscript 1)

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



Appendix 3: Protein preparation and RNA scaffold for structural studies. (Manuscript 1)

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).

	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–/Å^2)	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 (Å^2)	-142	-159
Model composition		
Non-hydrogen atoms	8405	8406
Protein residues	991	991
Ligands	3	3
B factors (Å^2)		
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

Appendix 4: Cryo-EM data collection, refinement and validation statistics. (Manuscript 1)

RNA oligonucleotide	Sequence
Product	/56-FAM/rUrGrA rGrCrC rUrArC rGrCrA rGrU
Template 1	5´rUrUrU rArArC rUrGrC rGrUrA ddC 3´
Template 2	5 rCrCrC rUrArC rUrGrC rGrUrA ddC 3´
Template 3	5´rUrUrU rCrArC rUrGrC rGrUrA ddC 3´
Template 4	5´rUrUrU rGrArC rUrGrC rGrUrA ddC 3´
Template 5	5´rUrUrU MrArC rUrGrC rGrUrA ddC 3´
Template 6 (structure)	5´rGrGrG MrArC rUrGrC rGrUrA-3´
RNA-hairpin duplex (structure)	5´rGrCrU rCrArU rArCrC rGrUrA rUrUrG rArGrA rCrCrU rUrUrU rGrGrU rCrUrC rArArU rArCrG rGrUrA rUrGrA rGrCrC rUrArC rGrCrA rGrUrA 3´
RNA-hairpin duplex (structure)	5´rGrCrU rCrArU rArCrC rGrUrA rUrUrG rArGrA rCrCrU rUrUrU rGrGrU rCrUrC rArArU rArCrG rGrUrA rUrGrA rGrCrC rUrArC rGrCrA rGrUrG 3´
RNA-hairpin duplex	/56-FAM/rUrUrU rUrCrA rUrGrC rArCrC rGrCrG rUrArG rUrUrU rUrCrU rArCrG rCrG-3´

Appendix 5: Supplementary Table RNA substrates (Manuscript 1)

sequence 5'CACUGCGUAGR[a] 5'CACUGCGUAGRCUCA3' 5'CAYUGCGUAGGCUCA3' R – Y 3'GUGACGCAUCYGAGU5' 3'GUGACGCAUCYGAGU5' 3'GURACGCAUCCGAGU5'			
G – C	64.7 °C	77.6 °C	77.6 °C
G – M	61.2 °C	71.2 °C	73.2 °C
A – M	60.6 °C	68.2 °C	73.8 °C

Appendix 6: Supplementary Table Thermal melting analysis (Manuscript 1)

Name 5'-sequence-3'		nt	formula	Mass calc.	Mass found
M_T	GGGMACUGCGUAp	12	$C_{115}H_{144}N_{48}O_{86}P_{12}$	3944.5221	3944.5297
M_1	MACUGCGUAGGCUCA	15	$C_{142}H_{178}N_{56}O_{104}P_{14}$	4764.6688	4764.6760
M_3	CAMUGCGUAGGCUCA	15	$C_{142}H_{178}N_{56}O_{104}P_{14}$	4764.6688	4764.6797
M_5	UGAGMCUACGCAGUG	15	$C_{143}H_{178}N_{58}O_{104}P_{14}$	4804.6749	4804.6684

Appendix 7: Supplementary Table Sequences and high-resolution ESI-MS data of NHC/Mmodified RNA oligonucleotides (Manuscript 1)