

UNIVERSITY OF NAPLES FEDERICO II



DEPARTMENT OF PHARMACY

PhD programme in Pharmaceutical Sciences

***Discovery of novel Direct-Acting Antiviral Agents to
counteract emerging viruses***

PhD student

Dr.

IRINA STEFANELLI

**Coordinator
Prof.
ROSARIA MELI**

**Tutor
Prof.
VINCENZO SUMMA**

XXXV CYCLE (2019-2022)

“Variability is not actually caused by man; he only unintentionally exposes organic beings to new conditions of life, and then nature acts on the organization and causes it to vary”.

-Charles Darwin

INDEX

ABBREVIATIONS AND ACRONYMS.....	7
ABSTRACT.....	10
1. INTRODUCTION.....	15
1.1. RNA viruses: putative agents for epidemic outbreaks.....	16
1.2. ZIKA Virus (ZIKV).....	19
1.2.1. Life cycle of ZIKV.....	22
1.2.2. Therapeutic approaches against ZIKV.....	23
1.2.2.1. Essential viral enzymes as potential molecular targets..	24
1.2.2.2. ZIKV Non-Structural 5 RNA-Dependent-RNA Polymerase (NS5 RdRp).....	30
1.3. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2).....	34
1.3.1. Life cycle of SARS-CoV-2.....	37
1.3.2. Therapeutic approaches against SARS-CoV-2	41
1.3.2.1. Vaccines	43
1.3.2.2. Direct-Acting Antiviral Drugs	46
1.3.2.2.1. Drug repurposing.....	46
1.3.2.2.2. Drug design.....	49
1.3.2.3. 3-Chymotrypsin-Like protease (3CL ^{pro}).....	51
1.4. References Introduction.....	57
2. BACKGROUND AND AIMS.....	81
2.1. ZIKV: background and aims of the project.....	82
2.2. SARS-CoV-2: background and aims of the project.....	83
2.3. References Background and Aims.....	84

CHAPTER 3.

SYNTHESIS OF ANTI-ZIKV AGENTS IDENTIFIED BY RATIONAL DESIGN AND STRUCTURE-BASED VIRTUAL SCREENING.....85

3.1. Rational design and synthesis of ZIKA NS5 RdRp Non-Nucleoside Inhibitors.....	86
3.1.1. Hopping of privileged dihydroxypyrimidine nucleus.....	86
3.1.2. Synthesis of new dihydroxypyrimidine derivatives.....	89
3.2. Discovery of antiviral agents against ZIKV through computational-driven polypharmacology approach.....	91
3.3. Biological evaluation against ZIKV infection in cell culture.....	98
3.4. Summary and conclusions.....	101
3.5. Experimental section.....	102
3.5.1. Synthesis and characterization of compounds.....	102
3.5.2. Molecular docking.....	114
3.5.3. Biological study.....	114
3.6. References Chapter 3.....	115

CHAPTER 4.

DESIGN AND SYNTHESIS OF BROAD SPECTRUM COV_s 3CL^{PRO} REVERSIBLE COVALENT PEPTIDOMIMETIC INHIBITORS.....117

4.1. Development of peptidomimetic aldehydes as SARS-CoV-2 3CL ^{pro} reversible covalent inhibitors.....	118
4.1.1. Rational design and molecular docking.....	118
4.1.2. Synthesis of tripeptides with aldehyde warhead.....	123
4.2. Development of peptidomimetic nitriles as SARS-CoV-2 3CL ^{pro} covalent reversible inhibitors	126
4.2.1. Rational design	126
4.2.2. Synthesis of peptidomimetics with nitrile warhead.....	129

4.3. Biological evaluation	133
4.3.1. Enzymatic assays against SARS-CoV-2 and MERS-CoV 3CL ^{pro} s.....	133
4.3.2. Phenotypic assays against SARS-CoV-2 infection.....	138
4.3.3. Binding and kinetic assays	141
4.4. Crystallographic studies.....	142
4.5. Molecular modelling on MERS-CoV 3CL ^{pro}	145
4.6. Summary and conclusions.....	148
4.7. Experimental section.....	151
4.7.1. Synthesis and characterization of compounds.....	151
4.7.2. Molecular docking.....	210
4.7.3. Enzymatic, cell-based and binding assays.....	211
4.7.4. Crystallization studies.....	214
4.8. References Chapter 4.....	215

CHAPTER 5.

SYNTHESIS OF NIRMATRELVIR FOR <i>IN VITRO</i> AND <i>IN VIVO</i> STUDIES.....	219
5.1. Synthesis and scale-up of benchmark inhibitor Nirmatrelvir 12	221
5.2. <i>In vitro</i> biological evaluation of Nirmatrelvir 12	223
5.3. <i>In vivo</i> efficacy studies of Nirmatrelvir 12 in an innovative COVID-19 mouse model.....	227
5.3.1. Administration of aerosolized SARS-CoV-2 to K18-hACE2 transgenic mice as innovative COVID-19 mouse model.....	227
5.3.2. <i>In vivo</i> efficacy studies.....	230
5.3.3. Impact of mice treatment with Nirmatrelvir 12 on the development of antiviral adaptive immunity.....	236

5.4. Stability tests of Nirmatrelvir 12 in different formulations.....	239
5.5. Summary and conclusions.....	241
5.6. Experimental section.....	243
5.6.1. Synthesis and characterization of Nirmatrelvir 12	243
5.6.2. Formulation study of Nirmatrelvir 12	249
5.6.3. <i>In vitro</i> enzymatic and antiviral cell-based assays.....	251
5.6.4. <i>In vivo</i> efficacy studies of Nirmatrelvir 12	253
5.7. References Chapter 5.....	257
GENERAL CONCLUSIONS.....	261
APPENDIX.....	266
A.1. External training.....	267

ABBREVIATIONS AND ACRONYMS

Throughout this thesis, abbreviations and acronyms recommend by the American Chemical Society in the Medicinal Chemistry area have been employed (revised in the Journal of Medicinal Chemistry in April 2021; http://pubs.acs.org/paragonplus/submission/jmcmr/jmcmr_abbreviations.pdf) Furthermore, those indicated below have also been used:

AR	Aerosolized
C	Capsid
CL ^{pro}	Chymotrypsin-like protease
CNS	Central nervous system
COMU	1-[(1-(Cyano-2-ethoxy-2-oxoethylideneaminoxy)-dimethylaminomorpholino)] uranium
CoV	Coronavirus
COVID	Coronavirus disease
CPE	Cytopathic effect
DAA	Direct-acting antiviral
DENV	Dengue virus
DMAD	Dimethylacetylene dicarboxylate
DMAP	4-dimethylaminopyridine
ds	Double-stranded
E	Envelope
EDCI	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EEA	European economic area
ER	Endoplasmic reticulum
EU	European Union
FRET	Fluorescence resonance energy transfer
GTP	Guanosine triphosphate

HCV	Hepatitis C virus
HEK-293	Human embryonic kidney 293
HIV	Human Immunodeficiency Virus
HOBt	Hydroxybenzotriazole
HPC	High performance computing
IN	Intranasal
INGM	Instituto Nazionale di Genetica Molecolare
JEV	Japanese-encephalitis virus
LDH	Lactate dehydrogenase
M	Membrane
MERS	Middle east respiratory syndrome
MTase	Methyltransferase
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide
N	Nucleocapsid
N-CoV-2	SARS-CoV-2 nucleoprotein
NCR	Non-coding flanking region
NHC	N4-hydroxycytidine
NI	Nucleotide inhibitor
NNI	Non-nucleoside inhibitor
NS	Non-structural
NTP	Nucleoside triphosphate
NTPase	Nucleotide TriPhosphatase
ORF	Open reading frame
pi	Post infection
PL	Papain-like
pp	Polyprotein
prM	Membrane precursor

RBD	Receptor-binding domain
RdRp	RNA-dependent-RNA-polymerase
RTC	Replicase–transcriptase complex
S	Spike
SAM	S-adenosyl methionine
SARS	Severe acute respiratory syndrome
SIM	Safe in man
ss	Single-stranded
TCEP	Tris(2-carboxyethyl)phosphine
TCID ₅₀	Tissue culture infectious dose 50
TEA	Triethylamine
TMPRSS2	Transmembrane protease serine 2
UNICA	University of Cagliari
UTR	Untranslated region
VLP	Virus-like particles
VOC	Variant of concern
VOHC	Variant of high consequence
VOI	Variant of interest
VS	Virtual screening
VSSR	Vita-Salute San Raffaele
WHO	World health organization
WNV	West Nile virus
YFV	Yellow Fever virus
ZIKV	Zika virus

ABSTRACT

Emerging and re-emerging infections have threatened the humanity from several decades. The globalization, the international trade and the expansion of human population increase the risk to contracting infections and contribute to speed-up their spreading. In particular, RNA viruses such as Flaviviruses and Coronaviruses, represent critical zoonotic agents, which are involved in inter-species transmission and periodically cause epidemic outbreaks around the World. The most recent COVID-19 pandemic, caused by novel coronavirus Severe Acute Respiratory Syndrome (SARS-CoV)-2, highlights the main role of RNA viruses as etiological agents of human diseases and the major gaps in the control of new pathogens. In this context, this thesis work is focused on the discovery of novel Direct-Acting Antivirals (DAAs) to counteract emerging RNA viruses. DAAs are defined as small molecules that specifically target viral proteins with essential roles in the replication and no counterparts in the human host, thus resulting in potent and safe antiviral activity. Usually, a combination of at least two DAAs with different mode of actions ensures a higher efficacy and limits the occurrence of viral resistant mutants. Indeed, the introduction of DAAs in the antiviral therapy allowed to successfully treat or cure viral infections, such as Human Immunodeficiency Virus (HIV) and Hepatitis C virus (HCV), respectively. Thus, DAAs represent efficacious therapeutic agents against viral infections.

Among RNA viruses, ZIKA virus (ZIKV) is an arising flavivirus that causes neurodevelopmental congenital diseases and the neuroinflammatory Guillain–Barré syndrome. In 2016 the World Health Organization (WHO) declared ZIKV a “Public Health Emergency of International Concern”. There are neither vaccines nor drugs for ZIKV, thus European Union invested millions of Euros to finance research against ZIKV infection. According to the aim of my

doctoral project, I focused on the discovery of new small molecules as DAAs active against ZIKA virus.

In chapter 3, two strategies are described for the ZIKV DAAs identification:

- rational design and synthesis of non-nucleoside inhibitors of ZIKV polymerase, that is an ideal target for the development of effective and safe DAAs;
- synthesis of potential polypharmacologic ZIKV candidates, selected through an ultra-fast multitargeting virtual screening by an innovative drug discovery platform, developed by Dompè Pharmaceuticals to address pandemic crisis.

After biological evaluation in antiviral phenotypic assays performed by our collaborators, two promising candidates (**34**, **37**) were identified, active in the nanomolar range and not toxic for cells.

As consequence of the COVID-19 pandemic, in the early 2020, my efforts were turned immediately to identify novel DAAs active against SARS-CoV-2, when no drugs or vaccines were available. Moreover, despite the recent approval of vaccines, monoclonal antibodies and few DAAs, the demand for new efficacious and safe antivirals is compelling to complete the arsenal against the COVID-19. The viral 3-Chymotrypsin-Like Protease (3CL^{pro}) is an essential enzyme for replication with high homology in the active site across CoVs and variants, showing an almost unique specificity for Leu-Gln as P2-P1 residues and allowing the development of broad-spectrum inhibitors.

The design, synthesis, biological evaluation and cocrystal structural information of newly conceived peptidomimetic covalent reversible inhibitors are described in chapter 4. A series of designed inhibitors displayed an aldehyde warhead, while another series was endowed with a nitrile. Both series featured a Gln mimetic at P1 and modified P2-P3 residues. Particularly, functionalized proline

residues were inserted at P2. The most potent compounds displayed low/sub-nM potency against the 3CL^{pro} of SARS-CoV-2 and MERS-CoV and inhibited SARS-CoV-2 replication in cells, with derivative **46** exhibiting nM EC₅₀ and the highest selectivity index. Some compounds were co-crystallized with SARS-CoV-2 3CL^{pro}, confirming our design. Altogether, these results foster future work toward broad-spectrum 3CL^{pro} inhibitors to challenge CoVs related pandemics.

To date, Nirmatrelvir is the first-in-class 3CL^{pro} inhibitor to receive approval by FDA and EMA in combination with the metabolic booster Ritonavir (Paxlovid[®]). Nirmatrelvir has been shown to reduce the risk of progression to severe COVID-19, but a rebound of the infection with the current treatment regimen has been reported in several recent studies. The in-house multigram synthesis of Nirmatrelvir, thereafter called compound **12**, and its evaluation in *in vitro* and *in vivo* assays developed by our collaborators were reported in chapter 5. In particular, it was described:

- an efficient synthetic route to afford Nirmatrelvir in gram scale for *in vitro/in vivo* evaluation, performed prior the publication of synthesis and antiviral activity by Pfizer team;
- validation of biochemical and phenotypic assays developed by our collaborators at the University of Cagliari and the “Istituto Nazionale di Genetica Molecolare” of Milan for the screening of new 3CL^{pro} inhibitors, through the use of Nirmatrelvir as benchmarker;
- validation of *in vivo* efficacy in an innovative and tractable mouse model of SARS-CoV-2 infection developed by our collaborators at the “Vita-Salute San Raffaele” University of Milan, and identification of a more suitable dosage and vehicle of po (oral administration) formulations of Nirmatrelvir respect to data published;

- *in vivo* investigation of impact of Nirmatrelvir treatment on the development of SARS-CoV-2-specific adaptive immune responses, which was still unknown.

CHAPTER 1

INTRODUCTION

1. Introduction

1.1. RNA viruses: putative agents for epidemic outbreaks

Emerging and re-emerging infectious diseases represent a huge threat to public health and a great challenge to global socio-economic equilibria. A large percentage of emerging infectious disease are traced to zoonotic viruses, that spread out from animal reservoir to humans and cause the onset of illness syndromes of varying severity.^{1,2} The progressive growth of the human population, closely associated to globalization, travelling and commerce, facilitates the exposure to zoonoses and spread of infection among human population, that can evolve into outbreaks restricted to specific geographical areas or, in the worst cases, epidemic or pandemic.³ In this scenery, RNA viruses have shown to play a relevant role in pathogenesis of human diseases.

RNA viruses are submicroscopic infectious agents, characterized by a single-stranded (ss) or a double-stranded (ds) RNA, often with associated basic proteins, surrounded by a protective and symmetric protein capsid. RNA viruses include a large group of important human pathogens, that periodically cause epidemic outbreaks. The last decades are marked by several outbreaks by RNA viruses, due to re-emerging of infectious diseases belonging to the Influenza,⁴ Ebola⁵ and Zika⁶ viruses, or emerging of new zoonotic viruses, such as the SARS,⁷ Middle East Respiratory Syndrome (MERS)⁸ and other CoronaViruses (CoVs).^{9,10} Lastly, the new coronavirus SARS-CoV-2, etiological agent of the pandemic CoronaVirus Disease 19 (COVID-19) highlighted the importance, as well major gaps, in the control of new pathogens (Fig. 1).

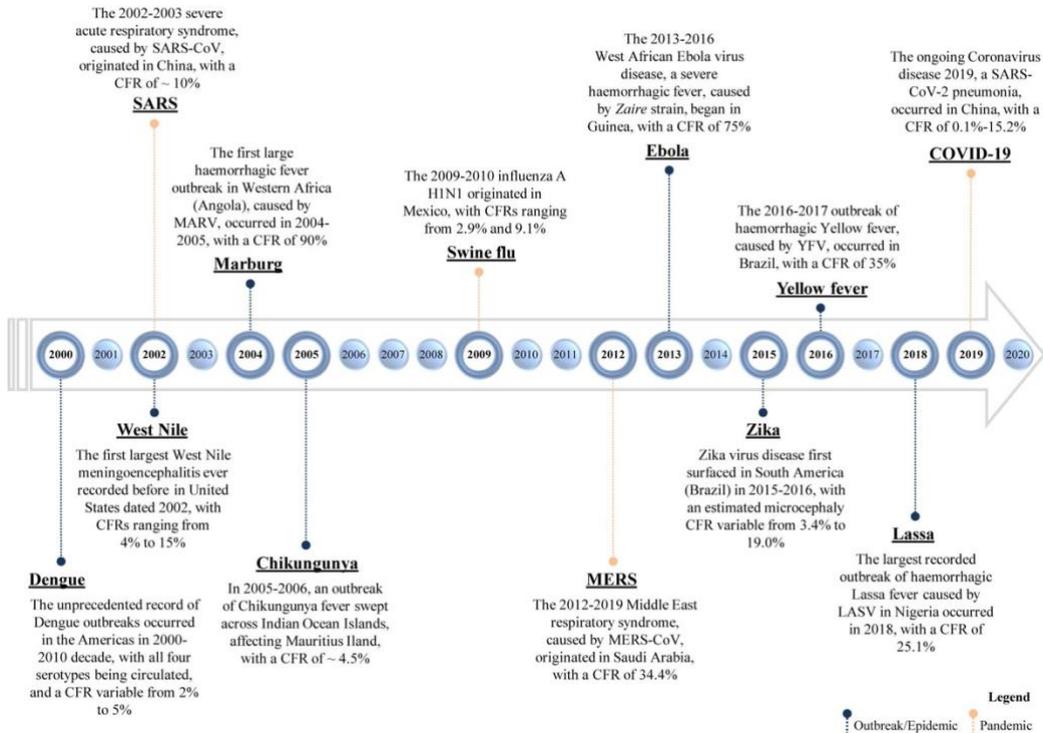


Figure 1. Timeline of recent viral epidemic outbreaks. The overall given values of CFR (case fatality rate) refer to “the proportion of cases of a specified condition that are fatal within a specified time,” according to Dictionary of Epidemiology.

In order to focus the scientific efforts on greatest public health risk, since 2016 WHO and the National Institutes for Allergy and Infectious Diseases periodically published a list of pathogens to be prioritized for research and development on basis of epidemic potential.¹¹ Not surprisingly, the priority diseases concern several infectious agents including different RNA viruses. Among these, ZIKV is an arising human pathogen that causes neurodevelopmental congenital diseases and the neuroinflammatory Guillain–Barré syndrome. ZIKV belongs to the *Flavivirus* genus, already known for comprising other human infectious agents, such as Dengue Virus with four distinct serotypes (DENV1-4), West Nile Virus (WNV), Yellow Fever Virus (YFV), Japanese-Encephalitis Virus (JEV), tick-borne-encephalitis and Usutu viruses, just to cite the most known members.¹²

These viruses are mostly transmitted by mosquitoes and ticks in the (sub)tropical areas, but the global urbanization, climate change and intercontinental travel allow the spread of flaviviruses outside the original habitat, reaching European and North American countries. Flavivirus diseases show features of epidemics with DENV associated to millions cases each year, while the recent ZIKV outbreak during 2016 Olympic Games in Brazil has highlighted the pandemic risk associated to these viruses.¹³ To date, vaccines are available for few flaviviruses (DENV, YFV, JEV), while no antiviral therapeutics are approved to treat infected individuals, thus European Union (EU) invested millions of Euros to finance research against ZIKV infection.¹³

Starting from late 2019, the global health authorities were faced with another emergency, represented by COVID-19 outbreak. SARS-CoV-2, as well the closest relatives CoVs, causes respiratory infection such common cold to critical respiratory issues. Despite five vaccines are currently available, the emergence of new SARS-CoV-2 variants and global access challenges make COVID-19 a still looming threat. The availability of effective antiviral drugs would have an impact on the control of these health emergency. Indeed, the introduction of DAA agents in the antiviral therapy allowed to successfully treat viral diseases, both seasonal, as in the case of Influenza Virus, and chronic, as in the case of HIV, or reach the complete eradication of the virus from the patients, as in the case of HCV.¹⁴ The DAA agents act by targeting essential viral proteins for the life cycle and inhibiting virus replication, counteracting the evolution of the infection causing the development of critical conditions for the patients and the risk of hospitalisation and death. Moreover, the possibility to target well conserved simultaneously different viral targets allows to overcome the serious issue of the high mutation rate own of RNA viruses.¹⁵ In consideration of the efficacy, safety and tolerability of DAA therapy, the scientific community is putting a lot of energy in the discovery of new DAAs to respond at the emerging of high-risk

pathogens, including Flaviviruses and CoVs, as it will be described in the next paragraphs.

1.2. Zika Virus (ZIKV)

ZIKV is a mosquito-borne flavivirus belonging to *Flaviviridae* family, already known to encompass important human pathogens, that are responsible of morbidity and mortality around the World. ZIKV was first isolated from rhesus monkey in 1947 in Uganda and from human in the United Republic of Tanzania and Uganda in 1952.¹⁶ The primary life cycle was mosquito-monkey-mosquito, then the urban transmission is mostly maintained in an infection cycle between humans and mosquitoes across tropical/sub-tropical regions. Following first occasional human infections, the major outbreaks in the South Pacific in 2007,¹⁷ 2013–2014,^{18,19} and the most recent one in Brazil in early 2015 drew the global concern on ZIKV pathogenicity.²⁰ Since the outbreak in Brazil, ZIKV rapidly spread across South and Central American countries, with 578,148 suspected cases in 45 countries and territories in the Americas and confirmed cases as of August 2016.²¹ The rapid and unpredicted spreading of infection to new areas, especially due to trade and travel, connected to lack of vaccines for the prevention and effective treatments, make Zika outbreaks more challenging. Thus, the WHO declared “Public Health Emergency of International Concern” on 1 February 2016.²²

Zika is transmitted by the bite of infected female mosquitoes of the *Aedes* species, most commonly *Aedes aegypti* and *Aedes albopictus*, and is spread from human to human through different ways: sexual contact, blood transfusion and perinatal transmission from mother to fetus.^{23,24} The disease often proceeds asymptotically or show mild symptoms after a period of incubation of 3-10 days, characterized by low-grade fever, conjunctivitis, widespread pruritic

maculopapular rash, arthralgia and myalgia.²⁵ These symptoms are non-specific and similar to infection caused by other arboviruses such as DENV, consequently the early diagnosis of ZIKV infection is complicated. This issue causes even more concern, considering that ZIKV has been linked to severe neurological complications, including microcephaly in infants²⁶ and Guillain-Barré syndrome in adults.^{27,28} In details, microcephaly is a birth defect and represents a permanent condition. The infants with microcephaly are characterized by a much smaller head than a normal, indicating that the brain has not also developed properly.²⁹ In adults, the Guillain-Barré syndrome usually affects the feet or hands, subsequently spreading to other compartments of the body within ~4 weeks. They start to show general symptoms such as numbness, pins and needles, muscle weakness, pain and problems with balance and co-ordination, until to develop severe conditions, including paralysis of face, legs and/or arms, persistent pain and serious difficulty to breath and speak.³⁰

In comparison with other flaviviruses, ZIKV shows a similar structural organization (Fig. 2). ZIKV is an enveloped and icosahedral virus, with a positive-sense, ssRNA genome of approximately 10,794 nucleobases.³¹ The genome comprises two Non-Coding flanking Regions (NCRs), the 5' NCR and the 3'NCR, respectively with a methylated nucleotide cap and a non-polyadenylated region. The genome is characterized by a single long Open Reading Frame (ORF). The ORF encodes for a polyprotein that subsequently is cleaved by host and viral proteases into seven Non-Structural (NS) proteins and three Structural (S) proteins, designated as the capsid (C), membrane precursor/membrane (prM/M) and envelope (E) proteins (Fig. 2).³²

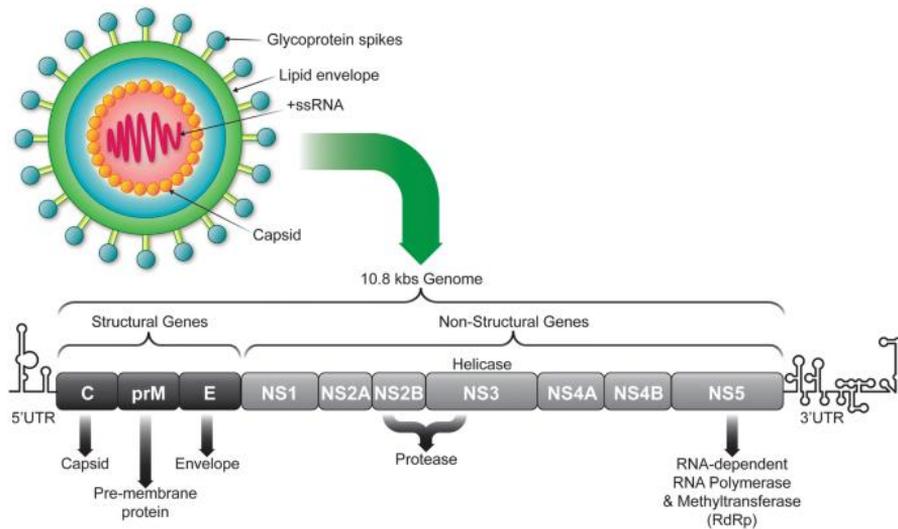


Figure 2. Structure of ZIKV and genome organization. The (+) ssRNA comprised seven NS genes and three S genes (C, M, E). M is present as immature form prM, which is cleaved by furin enzyme to get pr and M.³³

The seven NS proteins play an important role in the virus life cycle; in addition, they are involved in cell-mediated immune response.³⁴ The protein C is a basic protein that binds the viral RNA to form the nucleocapsid, which is surrounded by a lipid bilayer, containing the prM and E proteins. The protein E is an envelope glycoprotein and is the largest structural protein. Cryo-electron microscopic studies showed the Zika E protein is similar to the DENV, WNV and JEV.³⁵ The E protein function is involved in host cell binding, membrane fusion and entry, so it is fundamental for viral pathogenesis.³⁶ The viral membrane is formed by protein M, which is first translated as immature prM complex and then cleaved into mature protein M by cellular enzyme (furin) once virion is released from cell. The ‘pr’ segment has the task to protect protein E from degradation during the process of viral assembly (Fig. 2).

1.2.1. Life cycle of ZIKV

The life cycle of ZIKV is similar to other flaviviruses (Fig. 3). The E protein of mature Zika virion binds host cell membrane receptors, subsequently the virus gets internalized by endocytosis.³⁷ The acid pH of the endosome environment induces conformational changes of the E protein, which lets to the fusion of viral envelope with the host endosome membrane and then the release of the RNA genome into the cytoplasm.³⁸ The (+) ssRNA is translated into a long polyprotein by host ribosomes located on the Endoplasmic Reticulum (ER), which is cleaved into S and NS proteins by host proteases in the ER lumen and viral NS2B/NS3 protease in the cytoplasm.³⁹ The NS proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) form a replication complex at the ER membrane. In particular, the NS5 RNA-Dependent-RNA-Polymerase (RdRp) transcribes the viral (+) ssRNA into a complementary (-) ssRNA strand, forming a transient dsRNA.^{40,41} The dsRNA is then remodeled and separate by NS3 helicase into (-) and (+) ssRNA strands. The resulting (-) ssRNA strand act as template for transcription of new multiple (+) ssRNA strands, while the (+) ssRNA is capped by methyltransferase of NS5 on the N-terminus. The methylated (+) ssRNA strand with C, E and prM proteins at the ER membrane and the resulting immature virion buds out of the ER into to transport to the Golgi apparatus. Once through the trans-Golgi network, the prM protein is cleaved by host furin to release the pr segment and get mature the virion, which exits the cell via endocytosis.

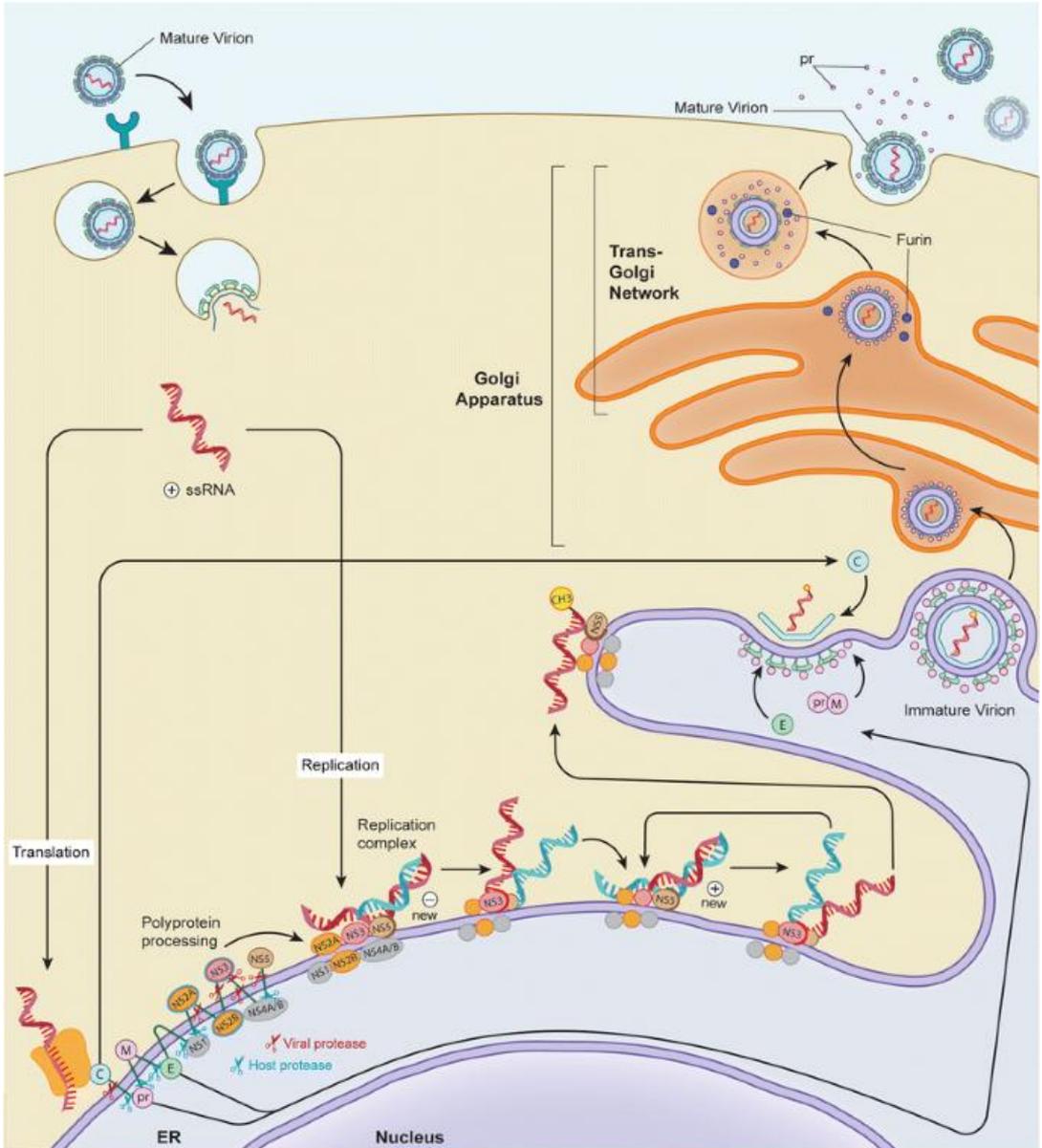


Figure 3. ZIKV life cycle. Virus enters the cell through receptor-mediated endocytosis. The synthesis of RNA and viral protein leads to genome replication and assembly of the new virus particles. Finally, the mature virions are released from the host cell via exocytosis.⁴²

1.2.2. Therapeutic approaches against ZIKV

Despite the several outbreaks and the severity of neuropathological diseases associated with ZIKV infection, no effective vaccines or specific treatments are

currently approved for clinical use. In view of lack of therapeutic strategy, the Center for Disease Control (CDC) and Prevention suggest a guideline for the treatment of mild symptoms for ZIKV infection, including rest, good hydration and taking acetaminophen for fighting the fever or pain.

However, efforts have been made for avoiding the spread of virus through the control of mosquito populations and to develop vaccines. Since WHO declared ZIKV emergency, the scientific community has been committed to give a prompt reply and hopefully a solution. Several ZIKV vaccine candidates have been developed in a short time and tested in clinical trials, but only a DNA vaccine sponsored by NIH-NIAID, named VRC-ZKAD- NA085-00-VP, has completed the clinical in phase II (NCT03110770).⁴³ Noteworthy, ZIKV cocirculates in the same endemic regions of other arboviruses, such as DENV, YFV, especially in Southeast Asia and Latin America.^{44,45} Since the symptoms of the disease are similar and the viral antigens are highly conserved, the challenge is developing a specific serological diagnosis, avoiding the antigen cross-reactivity.

In this context, the development of a pan-flavivirus chemotherapy could be a great option to bypass vaccine-cross-reactivity, besides a successful strategy to target the viral infection and prevent or treat the neurological manifestations. A promising option could be targeting viral proteins with a high degree of conservation across flavivirus genus to achieve a broad-spectrum inhibition. The development of DAAs targeting viral proteins, essential for the virus replication and absent in mammalian cells or deeply different from human proteins, is a mainly focus of antiviral chemotherapy, including ZIKV and other flaviviruses.

1.2.2.1. Essential viral enzymes as potential molecular targets

Among the essential viral enzymes, proteases and polymerases represent the most favorite targets for the development of novel therapeutic agents. Inhibitors of

these enzymes are already approved for use in the treatment of HIV and cure of HCV, and more recently for SARS-CoV-2, thus the antiviral drug research focused on this direction. NS5 and NS3 proteins are structurally well characterized, and several inhibitors of protease and polymerase have been reported for ZIKV and other flaviviruses. The NS3 and NS5 proteins exert a dual enzymatic function, respectively as protease/helicase and RdRp/MethylTransferase (MTase), resulting a pivotal role for the viral life cycle, and not homologues are present in humans. In addition, they are characterized from the highest degree of conservation along flaviviruses and represent potential molecular targets to discover novel broad-spectrum inhibitor.

NS2B/NS3 protease. The N-terminal NS3 domain is a trypsin-like serine protease with the catalytic triad Ser¹³⁵-His⁵¹-Asp⁷⁵, which cooperates with the factor NS2B to catalyze the cleavage of viral polyprotein into three S proteins and seven NS proteins. Interestingly, ZIKV protease shows a high similarity with other flaviviral proteases, with a sequence identity of 40-79% compared to DENV1-4 and 70% compared to WNV protease.⁴⁶ Similar to DENV and WNV proteases, ZIKV protease recognizes and cleaves substrates with a highly polar consensus sequence of (Arg/Lys)Arg-(Ser/Gly) protease.⁴⁶ On basis of the substrate binding site, several peptidomimetic inhibitors were designed against ZIKV, even supported by the availability of many crystal structures of free or inhibitor-bound ZIKV NS2B-NS3.⁴⁷⁻⁴⁸ The compound **1** (cn-716) is a dipeptide boronic acid, that showed a potent inhibition against ZIKV NS2B-NS3, with IC₅₀ = 0.25 μM and Ki = 0.040 μM (Fig. 4). The boronic acid function is able to engage a covalent bond with the catalytic Ser¹³⁵ in the active site, as confirmed by the X-ray studies.⁴⁹ Different peptidomimetic inhibitors were then investigated, functionalized with boronic acid as serine warhead.⁵⁰ Notably, they

showed to be active also against DENV and WNV, highlighting the potential to develop efficient broad-spectrum anti-flavivirus agents.

The serine protease inhibitor Aprotinin, also known as bovine pancreatic trypsin inhibitor and used for prophylactic use to reduce perioperative blood loss and the need for blood transfusion in patients undergoing cardiopulmonary disease, was also evaluated against ZIKV protease, showing a potent inhibition in a nanomolar range ($IC_{50} = 70 \text{ nM}$)^{51,52} as well as DENV and WNV proteases.^{53,54} Although the promising activity against different flaviviruses, the selective action in host cell must be improved to avoid off-target effects and to develop safe inhibitors.

Interestingly, the compounds **2** and **3** represent the first macrocyclic peptide-based inhibitors of ZIKV NS2B-NS3, discovered *de novo* through in vitro screening of a genetically reprogrammed library (Fig. 4).⁵⁵ The compounds **2** and **3** displayed a low/submicromolar inhibition of ZIKV protease in biochemical assays, with IC_{50} values of 1.5 and 0.24 μM respectively. Differently from the above compounds, **2** and **3** act as reversible peptide inhibitors and are characterized by a noncompetitive inhibition mechanism, suggesting allosteric inhibition.⁵⁵

Allosteric inhibition strategy was further investigated to target ZIKV protease. **4** (NSC135618) was first identified as allosteric inhibitor of DENV2 protease (Fig. 4) and subsequently tested against other flaviviruses, such as YFV, WNV and ZIKV, showing *in vitro* potency across several flaviviruses.⁵⁶ In addition, the compound **4** was tested in neuronal progenitor cells to evaluate the antiviral activity against ZIKV and showed to inhibit the viral replication, suggesting that it could be used as drug prophylactic and prevent the vertical transmission from mother to the fetus. This compound was further tested in mouse models. Treatment of ZIKV-infected mice with compound **4** (30 mg/kg/day) showed a reduction of viral titer in plasma, but high metabolism issue (half-time of only 20 min) and the high polar nature remain aspects to improve.

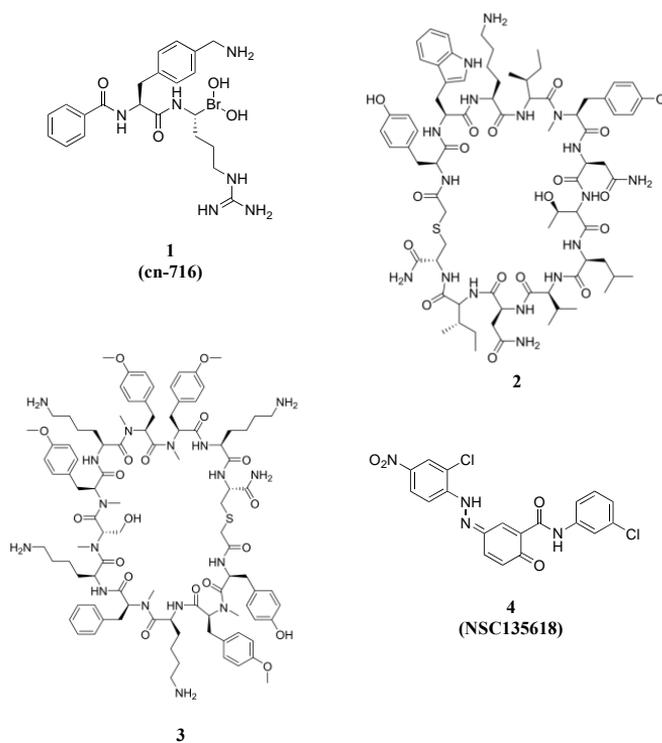


Figure 4. Chemical structures of anti-ZIKV NS2B/NS3 protease agents, including covalent (**1**) and reversible (**2, 3**) peptidomimetic inhibitors, and allosteric small molecule inhibitor (**4**).

NS3 Helicase. The C-terminal NS3 domain exerts ATPase/helicase activity, performing the unwinding in 3'–5' direction of newly synthesized dsRNA and achieving the translation and replication of viral genome. The helicase enzyme presents a Nucleotide TriPhosphatase (NTPase) activity to obtain chemical energy to unwind and remodel the dsRNA. In addition, ZIKV NS3 has the function to remove the 5'-phosphate group for the 5'-capping of the newly synthesized (+) ssRNA.⁵⁷ Considering its role in the viral life cycle, the NS3 helicase could be targeted at the NTP or RNA binding site for generating new therapeutic agents against ZIKV. Different NS3 ZIKV crystal structures are available, revealing the presence of three distinct subdomains (DI-DIII) and an

ATP-binding pocket between DI and DII well conserved across flaviviruses. Despite the key role of NS3 and the potential for the identification of broad-spectrum inhibitors, limited efforts have been carrying out within drug discovery against NS3 helicase yet.^{42,58} The common bioavailability concerns of helicase inhibitors and the selectivity issues for RNA and NTP binding sites make more complicated the discovery of new helicase inhibitors.

NS5 RdRp/MTase. ZIKV NS5 is the largest NS protein and the most conserved within the replication complex. NS5 is essential for the replication of the viral genome and performs a dual essential activity of RdRp and MTase. In addition, NS5 is also involved into suppression of interferon signaling by engagement of signal transducer and activator of transcription 2 protein, explicating its role in ZIKV pathogenesis. As described in the section 1.2.1, RdRp domain is located at the C-terminus of NS5 and is involved in the *de novo* RNA synthesis. Essential for the replication and production of new virions, the inhibition of the NS5 RdRp represents an efficient strategy to develop DAAs. In addition, the availability of full-length NS5 protein X-ray structures supported the design of new nucleoside and non-nucleoside RdRp inhibitors, as it will be discussed deeply in the following section.^{59,60}

Differently, the N-terminus of NS5 protein exhibits the MTase activity, responsible of the capping at the 5' end of ssRNA, essential for the genomic stability, splicing, transport and translation. The MTase activity of NS5 protein comprises the N-7 methylation, which is relevant for viral replication, and 2'-O methylation, involved into RNA escape from the host immune response. Many crystal structures of Zika N-terminal MTase NS5 have been solved, supporting the development of selective viral MTase inhibitors through the analysis of binding interactions and comparison with human homologue. The structural analysis of NS5 MTase allowed to identify different binding sites, including the

S-adenosyl Methionine (SAM) and the Guanosine TriPhosphate (GTP) pockets, and a positively charged RNA binding site, which is highly conserved among flaviviruses.⁶¹ While mutagenesis of residues into GTP site leads to a repeal of methylation at 2'-OH and a reduction of methylation N7 but without blocking viral replication, mutagenesis of residues into SAM pocket almost inhibits the methylation activity at 2'-OH and N7, leading to the block of viral replication if associated to Arg replacement in RNA binding site. This makes the NS5 MTase an attractive target to develop broad-spectrum flavivirus inhibitors, besides potential anti-ZIKV agents.

E protein. Targeting of proteins involved in the early stage of viral replication is another promising strategy, but often more challenging due to the antigenic feature of these proteins. Therefore, viral fusion inhibitors are known in the antiviral therapy, i.e. Enfuvirtide approved for HIV infection.⁶² Interestingly, the E protein shows a 40% of amino acid identity across flaviviruses and is absent in host cells, proposing a potential antiviral target.⁶³ The E protein is a class II fusion protein, which enables the fusion of the viral and endosomal membrane through different conformational changes. The E protein presents a hydrophobic β -N-octyl-glucoside pocket, responsible for the major conformational changes during fusion. Targeting β -N-octyl-glucoside pocket would mean to interfere with the conformational rearrangements of protein E and, in addition, could achieve broad-spectrum anti-flavivirus activity, since the amino acidic residues proximal to the β -N-octyl-glucoside pocket are highly conserved among ZIKV, DENV1-4, WNV and JEV.^{62,64}

1.2.2.2. ZIKV Non-Structural 5 RNA-Dependent-RNA Polymerase (NS5 RdRp)

NS5 RdRp initiates the *de-novo* RNA synthesis of (+) ssRNA genome by introduction of a single nucleotide triphosphate, which serves as primer for nucleotide polymerization.^{41,65} Similarly to other members of *Flavivirus* genus, ZIKV RdRp showed a canonical right-hand conformation, organized in fingers, palm and thumb subdomains (Fig. 5). The Fingers subdomain controls the *de novo* RdRp activity, participating in the formation of the active site and NTP entry channel. The Fingers subdomain is characterized by four flexible loops, important to modulate the RdRp conformation during the enzymatic activities and the entry of the RNA template in the tunnel. This tunnel shows positive charges, which interact with viral genome and favor the binding of RNA.

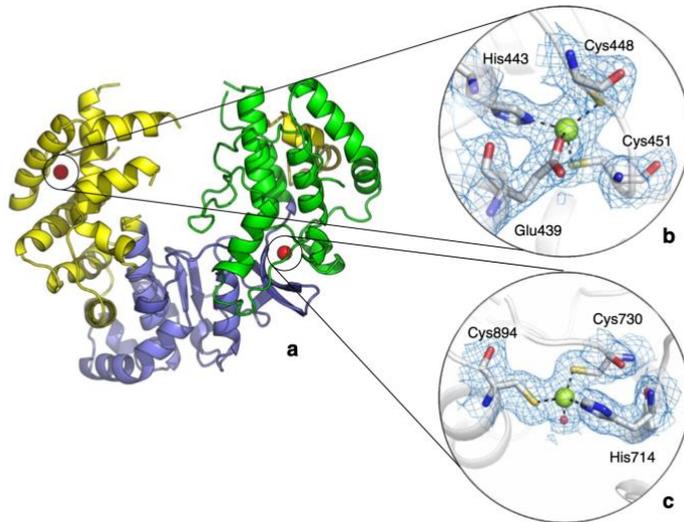


Figure 5. The detailed structure of ZIKV NS5 RdRp. (a) The ribbon representation of ZIKV RdRp (PDB: 5WZ3), organized in Fingers (yellow), Palm (purple) and Thumb (green) subdomains. (b) The zinc ion (represented as green sphere) binding pocket in the Fingers subdomain. (c) The zinc ion (represented as grey sphere) binding pocket in the Thumb subdomain.⁶⁶

The Palm subdomain accommodates the catalytic site, that comprises the catalytic residues Asp⁵³³ and Asp⁶⁶⁴ and mediates the RNA template binding,

translocation, and/or NTP specificity.^{67,68} As for other viruses like HCV, DENV and JEV, the ZIKV RdRp presents two zinc ions, which act as structure stabilizing agents. The structural analysis of the NS5 RdRp domain showed two different zinc-binding pockets: a zinc atom is located at the Fingers subdomain and is coordinated by Glu⁴³⁹, His⁴⁴³, Cys⁴⁴⁸, and Cys⁴⁵¹, while another zinc atom is coordinated by three amino acids (His⁷¹⁴, Cys⁷³⁰, Cys⁸⁴⁹) and a water molecule in the Thumb domain (Fig. 5). The Thumb domain presents a conserved extension, called “priming loop”, which is responsible of *de novo* initiation of RNA synthesis and is essential for the polymerase activity.⁶⁹

NS5 RdRp is the most conserved protein across flaviviruses. JEV and WNV share the higher sequence homology with ZIKV polymerase (70% sequence identity).⁶⁶ In addition, the thumb, palm and fingers subdomains of DENV2 and DENV3 polymerase are well superimposed on the equivalent domains of ZIKV polymerase, with a high sequence identity even of priming loops (76% and 81%, respectively).⁷⁰ Overall, targeting the NS5 RdRp can be an efficient strategy for the discovery of broad-spectrum anti-flavivirus agents.

Several nucleoside/nucleotide inhibitors have been proposed so far. Generally, the Nucleotide Inhibitors (NIs) act as chain terminating inhibitors, competing with the nucleotides and stopping the RNA chain elongation. Structurally, they are nucleotide/nucleoside analogues with a substituent at the C1' or C2' position of the ribose ring, able to hind the 3'-OH upon incorporated by the RdRp and avoid the formation of a phosphodiester binding with the incoming NTP. Mindful of therapeutic success of this class of compounds and considering that HCV is member of the same *Flaviviridae* family, Sofosbuvir has been one of the first compounds tested against ZIKV as polymerase inhibitor. Sofosbuvir (**5**), previously approved for use in patients with HCV with limited side effects,⁷¹ is a nucleoside inhibitor of HCV RdRp, orally administered as phosphoramidate nucleotide prodrug (Fig. 6). Therefore, Sofosbuvir is classified as pregnancy

category B drug by FDA, meaning it failed to demonstrate a risk to the fetus in animal reproduction studies and could be evaluated as clinical candidate for pregnant women. **5** showed to counteract the ZIKV infection in several neural cell lines with values of EC_{50} between 0.12 and 30.9 μM , without a significant cellular toxicity.⁷² Its efficacy and safety were also tested *in vivo* mouse models, proving to reduce the severity of viral disease and the mortality.^{73,74}

Among the class of C-2'-Methyl-nucleosides, the 7-deaza- 2'-C-methyl-adenosine **6** (Fig. 6), already investigated for HCV treatment but failed in clinical phase, exhibited a value of $EC_{50} = 20 \mu\text{M}$ in CytoPathic Effect (CPE) assay, using Vero cells infected with ZIKV. Promising results were shown by *in vivo* studies: the treatment with **6** of mice, previously infected with ZIKV, reduced the viremia, morbidity and mortality.⁷⁵ Unfortunately, the compound failed in clinical phase I.⁷⁶

Another adenosine analog, **7** (NITD008) displayed a promising activity against ZIKV polymerase. **7** is a 2'-C-ethynyl β -substituted nucleoside (Fig. 6) and was developed by Novartis Institute for Tropical Disease as inhibitor of DENV.⁷⁷ Therefore, **7** showed to inhibit ZIKV infectivity in phenotypic assay, with a value of $EC_{50} = 0.64 \mu\text{M}$.⁷⁸ Evaluation of **7** (oral dose 50 mg/kg daily for 1 week) in AG129 mice infected with ZIKV revealed promising results, such as improving of survival and decreasing of peak viremia.⁷⁷ However, long-term toxicity issues showed in other animal models limited **7** as clinical candidate.⁷⁶

Following the repurposing strategy of drugs approved for other viral diseases, the adenoside analogue **8** (BCX4430) was evaluated against ZIKV. **8** was already known as broad-spectrum antiviral agent, active against Ebola virus (EBOV), Marburg virus (MARV), and YFV.⁷⁹ Interestingly, **8** revealed *in vitro* efficacy and *in vivo* suppression of ZIKV infection. **8**, based on imino-sugar scaffold (Fig. 6), inhibited ZIKV infection in Vero cells ($EC_{50} = 3.8 \mu\text{M}$), without showing a significant toxicity ($CC_{50s} > 100 \mu\text{M}$, $SI > 26$). Additionally, treatment of AG129

mouse model infected with ZIKV with **8** (300 mg/kg/day) reduced viral RNA titer in serum and increased survival.^{79,80} Moreover, **8** could be a promising antiviral broad-spectrum candidate, considering that it displayed a good activity against YFV in hamster model.⁸⁰

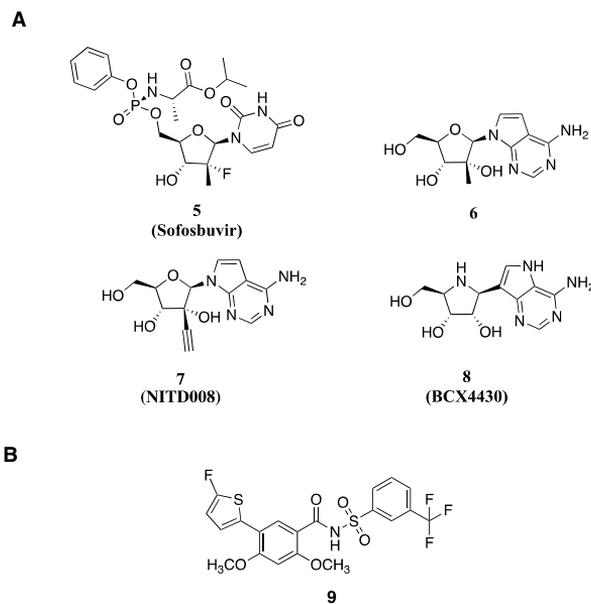


Figure 6. Structures of ZIKV NS5 RdRp inhibitors. (A) Nucleoside analogue inhibitors of ZIKV polymerase. (B) Non-nucleoside inhibitor of ZIKV polymerase.

Recently, a fragment-based screening against DENV NS5 RdRp disclosed the presence of an allosteric binding pocket, called “N-pocket”, located in the thumb subdomain, close to the catalytic site.^{70,81} The N-pocket ligands subsequently proposed were able to inhibit the DENV1-4 NS5 RdRp with a low micromolar potency. Interestingly, Lim *et al.* reported that N-pocket is highly conserved also in ZIKV polymerase, opening up the possibility to identify Non-Nucleoside Inhibitors (NNIs).^{66,82,83} In this context, a series of small NNIs of ZIKV RdRp have been designed on basis of known NNIs of DENV NS5 RdRp.⁸⁴ The most potent compound of series **9** (Fig. 6) showed a IC_{50} value of 7.3 μ M against Zika

polymerase and EC₅₀ of 24.3 μM in cell-based assays. Therefore, further efforts are still needed to identify potent NNIs compounds able to inhibit ZIKV NS5 polymerase and efficiently the ZIKV infection in cell line.

1.3. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2)

The novel SARS-CoV-2 is a highly infectious human RNA virus, responsible of the current pandemic COVID-19 emerged in late December 2019 in Wuhan, China, causing an outbreak of severe respiratory disease.^{85,86} Its rapid spread across the World led the WHO to declare a Public Health Emergency of International Concern on January 30 2020, and a pandemic status by March 11 2020.⁸⁷ Therefore, there are still millions of new cases reported daily from different countries and regions, with a considerable number of deaths.⁸⁸

CoVs are known to cause diseases in animals and humans. Belonging to the *Coronaviridae* family, *Nidovirales* order, the CoVs include enveloped (+) ssRNA viruses, classified into four genera: α, β, γ, and δ.⁸⁹ In humans, four CoVs (human CoVs 229E, NL63, OC43 and HKU1) are widely spread and used to infect the upper respiratory tract, causing mild symptomatic manifestations.⁹⁰ However, three zoonotic β-CoVs originating from bats have emerged in the last two decades and caused severe outbreaks in humans: SARS-CoV,^{91,92} MERS⁹³ and the most recent SARS-CoV-2.⁹⁴ These viruses are highly pathogenic respiratory, which attach the lower respiratory tract and can cause severe pneumonia, even fatal. Among human viruses, SARS-CoV-2 share the higher genetic sequence identity of 79.6% with SARS-CoV.^{95,96} Although the origin of SARS-CoV-2 is not known yet, the bat CoV RaTG13, detected in Yunnan Province, China, share 98% similarity with SARS-CoV-2 and represent the closest relative of SARS-CoV-2 to date.⁸⁶ Notably, the high genetic similarity between SARS-CoV-2 and related bat CoVs likely represents more than two

decades of evolution, suggesting that these bat CoVs are the most probable evolutionary progenitor of SARS-CoV-2, while other intermediate hosts might have played a crucial role in the process of transmission to humans.^{95,96} Like other CoVs, SARS-CoV-2 is transmitted person-to-person through direct contact or via respiratory droplets.^{97,98} Over the infection, the period of incubation takes around 4-5 days, but it can extend to 14 days.^{99,100} Upon the infection, the most common symptoms are fever, fatigue and dry cough, besides other less common symptoms, such as headache, sore throat, myalgia, diarrhea, vomiting, chills, loss of smell, and loss of taste.^{99,101,102} 81% of COVID-19 patients show mild to moderate symptoms, while in 14% of cases the infection progresses to severe pneumonia up to require assisted ventilation or develop other critical manifestations, that could affect the gastrointestinal system, heart, kidney, liver, and central nervous system.^{99,103,104} Although a lower mortality rate than SARS-CoV has been estimated until now, the higher transmissibility of SARS-CoV-2 is a determinant factor of its pathogenic danger.¹⁰⁵

Moreover, SARS-CoV-2, as well as RNA viruses in general, used to accumulate mutation frequently due to the low capacity of the RdRp to correct errors during genome replication. Over two years, SARS-CoV-2 changed and is still changing, and several variants were emerged so far, resulting in a different impact to the progress of pandemic.¹⁰⁶ The variants arise from one or more different genetic mutations on initial SARS-CoV-2 lineage, which affected the virus's properties, like the transmissibility, and the severity of associated disease.¹⁰⁷ In collaboration with multidisciplinary expertise and national authorities, WHO track the evolution of SARS-CoV-2 since January 2020,¹⁰⁸ in order to inform the global trend of the pandemic and priorities global research to provide validated countermeasures. The emerged variants are classified in three categories by CDC, on basis of public health action required: Variant Of Concern (VOCs) Variant Of Interest (VOIs) and Variant Of High Consequence (VOHC). For

VOCs, clear evidence is available indicating a significant impact on transmissibility, severity and/or immunity that could have an impact on the epidemiological situation in the EU/European Economic Area (EEA). For VOIs, evidence is available on genomic properties, epidemiological evidence or *in vitro* evidence that could imply a significant impact on transmissibility, severity and/or immunity, realistically having an impact on the epidemiological situation in the EU/EEA. For VOHCs, prevention measures or medical countermeasures have reduced effectiveness. The control measures and methods for reducing transmission through population remains a crucial point of the global strategy to reduce the emergence of new variants and the spread of existing variants across countries.

Several SARS-CoV-2 variants have been classified as VOCs (alpha, beta, gamma, delta, and omicron) at different times during the pandemic (Fig. 7),^{109,110} while any variant has been classified as VOHC by CDC so far.

VOC	Transmissibility	Disease severity	Neutralization by convalescent sera	Neutralization by vaccine-elicited sera	Neutralization by therapeutic antibodies	Global distribution ¹
B.1.1.7 <i>Alpha</i>	43% to 90% higher	Higher viral loads Higher risk of hospitalization and death in some age groups	Mostly maintained	Mostly maintained	Reduced for some antibodies	All continents
B.1.351 <i>Beta</i>	50% higher	Higher risk of hospitalization	Reduced	Reduced for mRNA-inactivated virus- and non-replicative viral vector-based vaccines	Reduced for some antibodies	All continents
P.1 <i>Gamma</i>	1.7 to 2.5 times higher	Higher viral loads Higher risk of hospitalization and death in some age groups	Reduced	Reduced for mRNA-based vaccines and AZD1222	Reduced for some antibodies	Asia Europe Oceania North America Central America South America
B.1.617.2 <i>Delta</i>	97% higher	Higher viral loads Higher risk of hospitalization	Reduced	Reduced. Effect more pronounced after partial vaccination than full vaccination	Reduced for some antibodies	All continents
B.1.1.529 <i>Omicron</i>	10 times higher Than Wuhan strain	Lower risk of hospitalization and severe illness	Reduced	Strongly reduced. This effect is minimized after the third dose of the vaccine	Reduced for most antibodies	All continents

Figure 7. Phenotypic characteristics of VOCs. ¹Global distributions are according to the PANGO lineages website.¹¹⁰

In particular, Delta (B.1.617.2), first identified in India in late 2020, has rapidly spread throughout the World and became the predominant variant of the SARS-CoV-2, until Omicron took its place in December 2021. B.1.617.2 infectivity was estimated to be twice as contagious than other earlier variants, causing severe disease in people who were not vaccinated. Therefore, the available vaccines were considered highly effective against severe illness, hospitalizations, and death from Delta.¹¹¹ The subsequent Omicron variant (BA.1) resulted 2.8 times more infectious than the Delta variant.¹¹² Omicron BA.1, first identified in Botswana and South Africa in late November 2021,¹¹³ was quickly recognized as a VOC due to more than 30 mutations and arising of case reported in South Africa.¹¹⁴ Nevertheless, Omicron is continuously evolving, leading to the emergence of various subvariants, including BA.2, BA.2.75, BA.4, BA.4.6 and BA.5. Notably, Omicron BA.1 and subvariants BA.2, BA.2.75 and BA.5 have demonstrated strong neutralization evasion capability, posing severe challenges to the efficacy of vaccine to control the infection.¹¹⁵

In this context, the rapid spread of SARS-CoV-2 and its variants around the World, besides the significant consequences on public health and the global economy, have mobilized the scientific community to shed light on virus morpho-structural characterization, and pathophysiology and diagnosis of related COVID-19, in order to develop efficient countermeasures to halt the spread of SARS-CoV-2.

1.3.1. Life cycle of SARS-CoV-2

CoVs share morphological, structural and functional features, making a relevant contribution in knowledge of aspects of SARS-CoV-2 infection, starting from pathogenesis to treatment options.⁹⁵ CoVs are enveloped viruses with a diameter relatively large (50-200 nm), and contain a non-segmented (+) ssRNA genome

of ~30 kb, which is the largest identified RNA genome to date.¹¹⁶ Similarly to other CoVs, SARS-CoV-2 appears in a spherical shape by cryo-electron microscopy, with Spike (S) glycoproteins that emanate from surface of the virion and give the characteristic appearance of a crown.¹¹⁶ The S protein constitutes the viral envelope together with other two main structural proteins, namely envelope and membrane, while another viral structural protein, nucleocapsid (N), binds and protects the RNA (Fig. 8A).¹¹⁷ The S glycoprotein drives tissue tropism and mediates attachment to the host cell through the interaction between the Receptor-Binding Domain (RBD) within the S1 region and the human angiotensin-converting enzyme 2 (ACE2) receptor.¹¹⁸ It has been estimated that binding affinity of the RBD of SARS-CoV-2 S1 subunit toward ACE2 is about 20-fold higher than SARS-CoV, thus explaining the major degree of infectivity that characterizes the novel CoV-19.⁹⁵ Following receptor binding, the access to the host cellular cytosol is realized by proteolytic cleavage of the S protein at the S1/S2 and S2' sites, due to a the transmembrane protease serine 2 (TMPRSS2) or a cathepsin B/L.^{118,119} The cleavage at site S2' exhibits a fusion peptide that inserts into the host cell membrane and leads to the formation of an antiparallel six-helix bundle, allowing the fusion process and therefore the uncoating and release of the viral RNA in the cytoplasm (Fig. 8B).¹²⁰ *In vitro* the simultaneous inhibition of TMPRSS2 and cathepsin dramatically reduces the entry of SARS-CoV-2 into host cell,¹¹⁸ while it is completely blocked for SARS-CoV.¹²¹ This evidence suggests the possible involvement of other endogenous proteases in the priming of SARS-CoV-2 S, but not of SARS-CoV S. Interestingly, the comparison of SARS-CoV-2 S and SARS-like CoVs S protein sequence showed a peculiar furin-like cleavage site in the S protein of the novel CoV-19, despite of other CoVs.¹²² Furin is highly expressed in human airway epithelial cells and has proved to be a relevant factor for proteolytically activation of SARS-CoV-2 S protein, hence a potential drug target.¹²³ Thus, the inhibition of viral entry can

represent a valid antiviral strategy with the identification of both viral and host factors targeting agents.¹¹⁹ The next step in the CoV lifecycle is the translation of the replicase gene from the viral RNA, which takes place in the host cytoplasm (Fig. 8B). The full genome contains a 5' cap structure with a 3' poly (A) tail, allowing it to act as an mRNA, and a leader sequence. The 5'- and the 3'-ends contain UnTranslated Region (UTR) with multiple stem loop structures, necessary for RNA replication and translation.⁸⁹ Besides the replicative region, the genome presents a structural region closer to 3'-end, with various accessory genes interposed. The replicase gene comprises two overlapping ORFs, ORF1a and ORF1b, that encode for polyproteins (pp)1a and pp1ab, respectively. The replicase polyproteins are cleaved by two viral proteases, 3CL^{pro}, also named main protease (M^{pro}), and papain-like protease (PL^{pro}), to release non-structural proteins (NS1-16) and form the replicase–transcriptase complex (RTC).¹²⁴ The residual 10kb affects the encoding of structural proteins and accessory proteins.⁹⁵ In line with other CoVs, the accessory proteins appear not essential for RNA replication in tissue culture, rather play a role in viral pathogenesis and host immunity suppression.¹⁰⁵ A lot of NS proteins assemble in the RTC and perform functions in RNA synthesis and maturation. In particular, NS12 encodes the RNA dependent RNA polymerase, that cooperates with NS13-RNA helicase domain/RNA 5'-triphosphatase, NS14-exoribonuclease and NS16- 2'-O-MTase for a correct replication of viral genome.¹²⁴ Genomic (+) ssRNA transcription produces genomic and sub-genomic RNAs, both capped and polyadenylated, through (-) strand intermediates that act as templates. The sub-genomic RNAs derives from a process of dissociation/reassociation of RTC complex with the respective template, and serve as mRNAs for the translation of structural and accessory proteins (Fig. 8B).¹²⁴

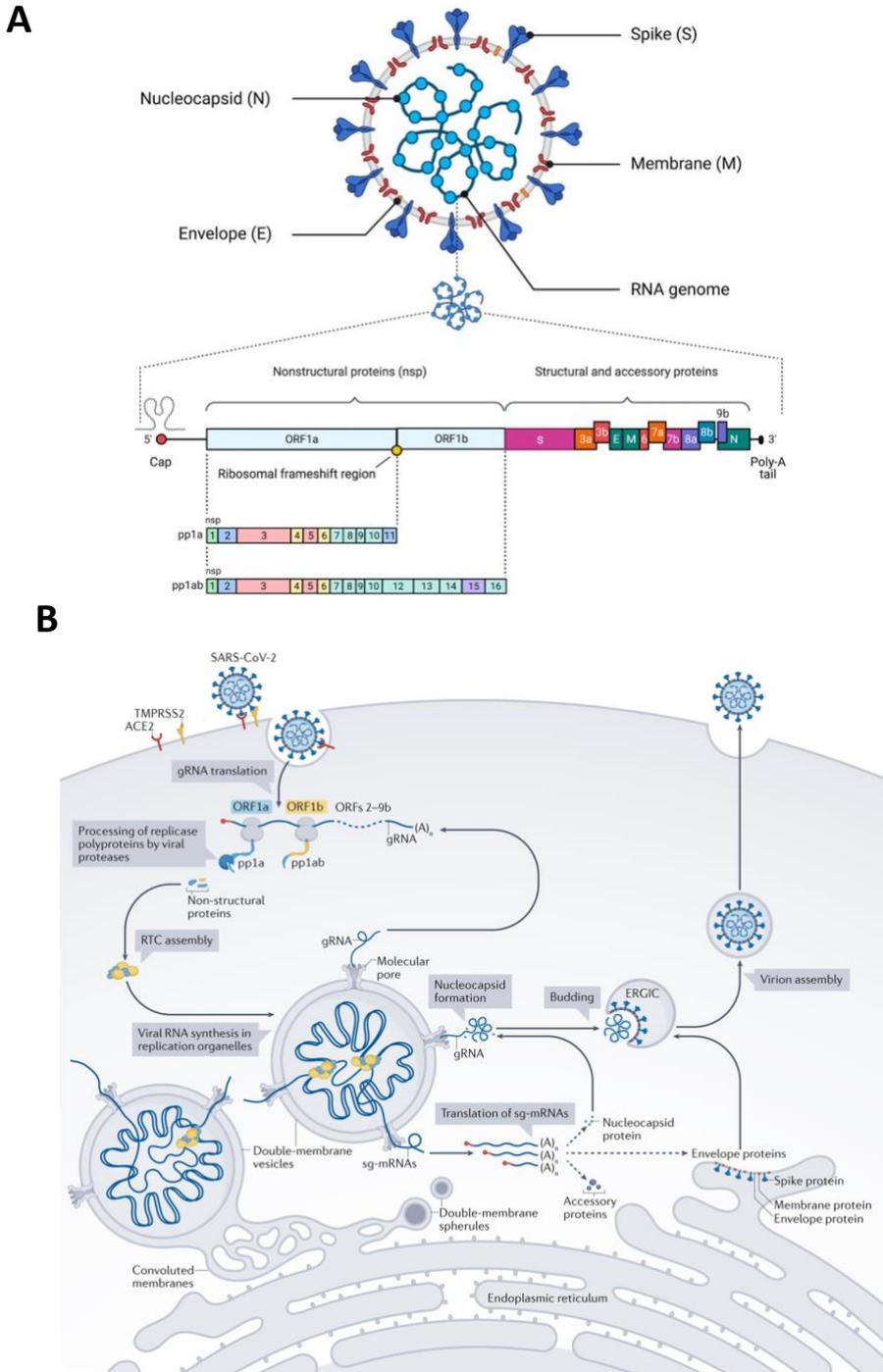


Figure 8. (A) Schematic representation of SARS-CoV-2 virion and RNA genome organization, including the 5' cap, the regions that encode the nonstructural proteins and structural proteins,

and the poly-A tail.¹²⁵ (B) Representation of SARS-CoV-2 life cycle to delineate the key steps of the viral infection into host cell.¹²⁶

Following sub-genomic RNAs translation, the structural proteins S, E and M enter the ER and move towards ER-Golgi intermediate compartment. The N protein binds to genomic RNA, thus encapsidated viral genome joins the structural proteins in the ER-Golgi intermediate compartment to form mature virions. Subsequent assembly, virions are transported to the cell surface in vesicles and released by exocytosis (Fig. 8B).¹²⁴

1.3.2. Therapeutic approaches against SARS-CoV-2

Most COVID-19 patients show mild or moderate symptoms of infection, but severe manifestations are present into up to 5-10% cases,⁹⁹ closely connected to risk of hospitalization and death. In this regard, the public health authorities committed fundings and global expertise to counteract the progress of the infection and find strategies for the control of SARS-CoV-2. As first line of action, social distancing and mask wearing were the key points of guidelines disseminated by public entities in each country for limiting the viral transmissibility. Indeed, the CDC recommends:

- wearing mask with the best fit and protection;
- maintaining the social distancing of at least 2 meter;
- avoiding crowd and poorly ventilated areas;
- washing hands with appropriate detergents, especially in a public place or before to touch eyes, nose and mouth;
- wearing mask when you cough or sneeze.

These countermeasures are effective at reducing the spread of virus across the population, but the vaccination plays an essential role in controlling and preventing COVID-19.¹²⁷ An intense vaccine program has been employed globally to prevent severe illness and, consequently, reduce the burden on the

healthcare system, as it will be deeply discussed in the section 1.3.2.1. However, the need of effective treatment is still present, especially for countries where the vaccination rate is slow and for the emergence of new variants that may escape the vaccine.

At the beginning of the pandemic, the repurposing drug was the more immediate strategy to discover effective agents against SARS-CoV-2. Antiviral agents already active against influenza virus, HIV, Ebola virus and SARS-CoV/MERS viruses, as well as antibiotics, antiprotozoals and anthelmintics have been evaluated against SARS-CoV-2. These antivirals were proposed to inhibit the virus life cycle by targeting specific human host receptors or viral proteins, including receptor ACE2 and human proteases such as TMPRSS4, TMPRSS2, furin and cathepsin L for the virus cellular entry, or viral enzymes such as RdRp, the helicase complex, 3CL^{pro} and PL^{pro}. Despite the huge efforts to discover efficient antiviral agent, only three DAAs have been approved so far: Remdesivir (intravenous), Molnupinavir (po) and Nirmatrelvir (po).¹²⁸

Close to antiviral drugs, several anti-SARS-CoV-2 antibody products are approved for COVID-19, with specific recommendations. These include monoclonal antibody for the treatment of COVID-19 (bebtelovimab), anti-SARS-CoV-2 monoclonal antibodies as pre-exposure prophylaxis (tixagevimab plus cilgavimab) and as post-exposure prophylaxis for SARS-CoV-2 infection (bamlanivimab plus etesevimab and casirivimab plus imdevimab, COVID-19 convalescent plasma and SARS-CoV-2-specific immunoglobulins (trials ongoing)).¹²⁹ The monoclonal /polyclonal antibodies act by modulating the host response to the viral infection and represent an additional part of the COVID-19 therapy. Therefore, health authorities have set criteria for the use of monoclonal antibody in the treatment and prevention of COVID-19. Importantly, these criteria also include the impact of Omicron variant on antibody efficacy. Monoclonal antibodies such as Casirivimab-imdevimab and Bamlanivimab-

etesevimab resulted inactive against Omicron and they are not used in countries where the variant is predominant, while Sotrovimab and Tixagevimab-cilgavimab showed a different activity respect to the Omicron sublineage. Bebtelovimab resulted to be active against the variant Omicron.¹³⁰

Anti-inflammatory and respiratory therapies were additionally introduced to treat patients hospitalized at an advanced stage of the disease. The National Institutes of Health (NIH) recommends the following immunomodulators drugs for the capability to counteract the pro-inflammatory cytokine storm: corticosteroids (dexamethasone), IL-6-inhibitors (tocilizumab or sarilumab) and JAK inhibitors (baricitinib or tofacitinib).¹³¹

Among different therapeutic approaches for COVID-19, a deeper focus has been reported on the vaccines approved in the prevention of COVID-19 and the DAAs available in the following sections.

1.3.2.1. Vaccines

In response to the COVID-19 pandemic, enormous efforts have been made worldwide to develop effective vaccines in an efficient and timely manner. As of October 28 2022, 49 vaccines have been approved across 201 countries around the World.¹³² Particularly, six vaccines have been approved by EMA,¹³³ while FDA has authorized the use of four vaccines to date.¹³⁴ Moreover, there are 207 vaccine candidates and nearly 800 vaccine trials ongoing.¹³²

Vaccines work stimulating the immune host response to recognize virus or pieces of virus, named antigens, and to produce antibodies against SARS-CoV-2. In the specific case of SARS-CoV-2, the antigen is the S glycoprotein on the viral surface. Vaccines approved so far respond to several platforms. In summary, the vaccine platforms are classified into two categories: component viral vaccines and whole virus vaccines.

The component viral vaccines include:

- Protein subunit, which contains isolated and purified viral proteins as antigenic pathogen fragments.¹³⁵ Nuvaxovid (Novavax) and Covovax (manufactured by Serum Institute of India, Novavax formulation) are characterized by a recombinant nanoparticle S protein, in association with the matrix M as adjuvant. However, the few viral fragments included are not able to reproduce the complexity of the virus.¹³⁶
- Virus-like particles, which contains viral proteins that mimic the structure of the virus, but no the genetic material.
- RNA-based, which contains viral mRNA that encodes full-length SARS-CoV-2 S protein (modified by two mutations). This platform was approved for the first time and showed a significant safety and efficacy. Comirnaty (Pfizer/BioNTech) and Spikevax (Moderna) contain a nucleoside modified mRNA packed in drops of fat (lipid nanoparticles) to protect it and avoid its immediate degradation before cellular entry.^{137,138}
- DNA-based, which contains viral DNA to provide the instructions for making viral proteins. To date, only one was approved for emergency use in India, named Zydus Cadila (ZyCoV-D). ZyCoV-D is based on DNA plasmid vector that encodes the S protein of SARS-CoV-2. Differently from mRNA vaccines, it is administered intradermally via a needle-free injection system.¹³⁹ Therefore, in Italy a phase I/II study to assess the safety and immunogenicity of DNA vaccine COVID-eVax was recently stopped (NCT04788459), since the good progress of the Italian vaccination campaign made impossible to complete the clinical study as planned.
- Non-replicated viral vector, which contains viral genetic material packaged inside another harmless. The approved vaccines Vaxzevria

(Oxford/AstraZeneca) and Covishield (manufactured with Oxford and AstraZeneca formulation by Serum Institute of India and Fiocruz-Brazil) are characterized by a chimpanzee adenovirus, that encoded the SARS-CoV-2 S protein.

- Replicating viral vector, which contains viral genetic material packaged inside another harmless virus that can copy itself. Differently, Ad26.COVS.S (Janssen/Johnson & Johnson) is based on a recombinant human adenovirus, able to express the S protein.¹⁴⁰

The whole virus vaccines include:

- Inactivated: Covaxin (Bharat Biotech), Covilo (Sinopharm), and CoronaVac (Sinovac) contain copies replicated in cells and inactivated chemically, mixed with compounds which act as immune stimulants and immune response amplifiers. Notably, the chemical inactivation is carried out by heat or irradiation, but this process could influence the immunogenicity of pathogens, reducing the efficacy of this platform of the virus that have been killed (inactivated).¹⁴¹
- Live-Attenuated, which contains copies of the virus that have been weakened (attenuated).

Although national vaccination campaigns are moving as fast as possible to reach the highest percentage of immunized people and billions of vaccines doses have been administrated until now, there is still a significant number of people not vaccinated, due to pre-existing medical conditions or unwilling to be vaccinated. In addition, global access challenges limit the vaccination rate in developing countries.

The emerging variants and strains that evade the immune system represent an additional risk, considering that most of these vaccines have been developed on

basis of the Wuhan-Hu-1 strain. For instance, several national health systems are proceeded with additional dosing of vaccine to boost the antibody response in the most vulnerable subjects.¹⁴² These boosters are called “monovalent”, because their manufactory is based on the Whuan-Hu-1 strain and is developed to protect against the original virus, even if provide some protection against Omicron. Recently, updated boosters or “bivalent”, active against original SARS-CoV-2 and the Omicron variant BA.4 and BA.5, were developed to ensure a better protection against COVID-19. To date, only two bivalent COVID-19 boosters were manufactured by Pfizer and Moderna companies, both authorized in September 2022.¹³³

1.3.2.2. Direct-Acting Antiviral Drugs

The availability of effective antiviral drugs offers a significant contribution to the control of the health emergency COVID-19.¹⁴³ Despite several vaccines approved, the need of small molecule therapeutics to treat/prevent CoVs infection is still present for counteracting emerging SARS-CoV-2 variants and future CoV outbreaks, beyond COVID-19. Indeed, drugs have the advantage to improve patient management and are the only way for immunocompromised patients. Moreover, their manufacturing, distribution, and administration are easier than for vaccines, thus likely reaching also developing countries where vaccination rate is still very low. To date, only two repurposed drugs and an oral first-in-class DAA agent are authorized for the use, as it will be explained in next sections.

1.3.2.2.1. Drug repurposing

The drug repurposing was the first approach investigated to identify faster a drug to prevent or control SARS-CoV-2, although clinical validations are required.¹⁴⁴

In this context, WHO launched “Solidarity”, a comprehensive international program of clinical trials,¹⁴⁵ with the aim to identify bona fide therapeutic options against COVID-19. The clinical trial was a milestone in the drug repurposing program against SARS-CoV-2, considering that the result eliminated several putative active repurposing drugs such as 4-Hydroxychloroquine, HIV protease inhibitors, interferon α . In contrast, Remdesivir was tested in this trial and it was later approved in another clinical trial, leading to the first drug authorized for emergency use for COVID-19.

Remdesivir (**10**) is a RdRp inhibitor, discovered by Gilead Science as prodrug of an adenine-C-nucleoside analogue (before named GS-5737) for the treatment of Ebola virus¹⁴⁶ and evaluated in phase II/III clinical studies (NCT02818582, NCT04385719, NCT03719586). Therefore, **10** showed an *in vitro* broad-spectrum inhibition of viruses from different families, including SARS-CoV, MERS-CoV, and other human and bat CoVs.¹⁴⁶⁻¹⁴⁷ Structurally, Remdesivir is a bis(S-acyl-2-thioethyl) monophosphate prodrug and its triphosphate derivative (Fig. 9) act as a chain-terminator through the incorporation into the nascent viral RNA, which leads to a delayed RNA chain termination.^{148,149} It showed to inhibit SARS-CoV-2 both *in vitro* ($EC_{50} = 0.28 \mu\text{M}$ in Calu3 cells, $EC_{50} = 1.65 \mu\text{M}$ in Vero cell line)¹⁵⁰ and *in vivo*.^{151,152} Many clinical trials were followed to evaluate the efficacy of Remdesivir for COVID-19. Although a significant mortality decrease was not revealed, **10** reduces serious adverse events of hospitalized COVID-19 patient and can speed up the recovery time in some patients.^{153,154} Remdesivir is administrated via intravenous infusion in adults and adolescents (from 12 years of age and weighing at least 40 kilograms) with pneumonia requiring supplemental oxygen (low- or high-flow oxygen or other non-invasive ventilation at the start of treatment). The medicine can also be used in adults who do not require supplemental oxygen and who are at increased risk of developing severe COVID-19.¹⁵⁵

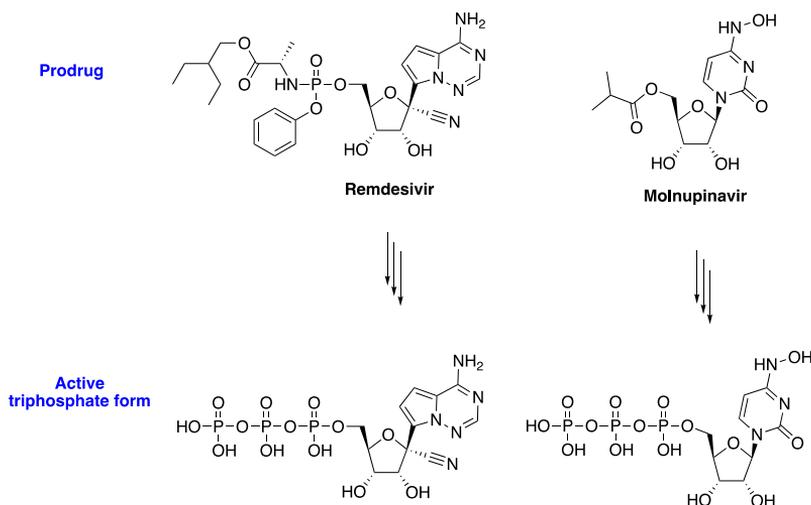


Figure 9. Structures of RdRp inhibitor drugs **10** and **11** for the treatment of COVID-19 and their metabolic activation.

Molnupiravir (**11**) is another repurposed drug approved for the treatment of COVID-10 (also known as MK-4482, proposed trade name: Lagevrio®). As well as Remdesivir, Molnupiravir is a RdRp inhibitor prodrug, initially developed against influenza A and B. Structurally, it is a 5'-isopropyl ester of the nucleoside analogue N4-hydroxycytidine, which is administered po and hydrolyzed *in vivo* after absorption to release the active molecule in the plasma. Once converted into its active triphosphate form (Fig. 9), it is incorporated into viral RNA and induce RNA mutagenesis.¹⁵⁶ NHC showed to inhibit SARS-CoV-2 and other CoVs *in vitro* with an EC₅₀ against SARS-CoV-2 of 0.3 μM in Vero cells and 0.08 μM in human lung epithelial cell line Calu-3.¹⁵⁷ *In vivo* study on Syrian hamsters infected with SARS-CoV-2 reported that **11** was able to reduce the lung pathology, decreasing viral loads in the lung.^{156,158} Treatment infected ferrets with **11** led to a reduction of viral loads in nasal lavages, in addition to inhibition of spread to other animals not treated with the drug.¹⁵⁹ Studies on hamster infected with the B.1.1.7 (α) and B.1.351 (β) variants of virus displayed

Molnupiravir was able to inhibit the viral infection as well.¹⁶⁰ Several clinical trials are performed to evaluate the safety and efficacy of **11**, and ten clinical trials are still ongoing to test **11** for COVID-19 disease at present. It is known that Molnupiravir reduces the risk of hospitalization or death by approximately 50% in non-hospitalized adult patients with mild-to-moderate COVID-19 disease.¹⁶¹ The U.K.'s Medicines and Healthcare Products Regulatory Agency was the first agency to temporary authorize Molnupiravir for the treatment of mild to moderate COVID-19 in adults with at least one risk factor for severe illness.¹⁶² Following, on 5 August 2022 FDA approved Lagevrio[®] with the same therapeutic indications, while EMA has not yet authorized Lagevrio[®] specifically for patients with COVID-19.

1.3.2.2.2. Drug design

As described above, repurposing of approved drugs failed to find effective therapeutic options for COVID-19. Differently from the antiviral drugs described above, Nirmatrelvir is the unique drug identified through a rational drug design and is first-in-class approved inhibitor of 3CL^{pro}.¹⁶³

Nirmatrelvir is the active compound of Paxlovid[®], oral medicine used for treating COVID-19 in adults who do not require supplemental oxygen and who are at increased risk of the disease becoming severe. Paxlovid[®] provides two tablets, each containing 150 mg Nirmatrelvir, plus one tablet containing 100 mg ritonavir, a cytochrome inhibitor used as pharmacokinetic booster blocking the high metabolic degradation of Nirmatrelvir (Fig. 10).

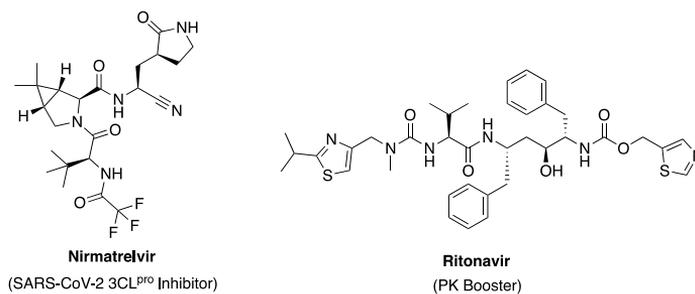


Figure 10. Nirmatrelvir, the first-in-class SARS-CoV-2 3CL^{pro} inhibitor, is packaged with ritonavir (as Paxlovid[®]), a strong cytochrome P450 3A4 inhibitor and pharmacokinetic boosting agent.

Since Nirmatrelvir works inhibiting an enzyme which is essential for the viral replication and is involved at early stage of the virus life cycle, it can help the body to surmount the SARS-CoV-2 infection and avoid the disease becomes severe. Indeed, Nirmatrelvir showed to reduce by 89% hospitalization and death of non-hospitalized adult patients with COVID-19 who are at high risk of progressing to severe illness.¹⁶⁴

The global scientific community is making further efforts to develop selective antivirals against SARS-CoV-2, even considering there may be the emergence of resistance against Paxlovid once widely deployed in the clinic.¹⁶⁵

For this purpose, structural/nonstructural viral proteins are evaluated as potential SARS-CoV-2 druggable targets, including the S protein, RdRp, NTPase/helicase, 3CL^{pro} and PL^{pro}.^{166,167} Among these, the 3CL^{pro} represents the most attractive druggable target. Although both proteases are essential for the virus replication, the paramount role of 3CL^{pro} in polyprotein processing and the lack of a human homolog focused research to identify new DAAs against SARS-CoV-2 3CL^{pro}.¹⁶⁷

1.3.2.3. 3-Chymotrypsin-Like protease (3CL^{PRO})

The 3CL^{PRO}, also known as M^{PRO}, is a cysteine protease, that cleaves the polyproteins pp1a and pp1ab at different 11 conserved sites to release 12 NS proteins, involved into replication/transcription process and essential for the viral replication (see section 1.3.1).^{168,169} Consequently, the inhibition of 3CL^{PRO} can stall the production of virions into host cell, providing a target to develop effective DAAs against SARS-CoV-2.

Viral proteases are already validated molecular targets in the discovery of safe and efficient antiviral drugs, for example against HIV and HCV.¹⁷⁰ Moreover, no human homologues proteases are present in mammalian cells, defining the 3CL^{PRO} as ideal drug target for the development of selective antiviral inhibitors. In details, 3CL^{PRO} recognizes the glutamine amino acid residues and cleaves the peptide bond between glutamine and the adjacent amino acid.^{166,167}

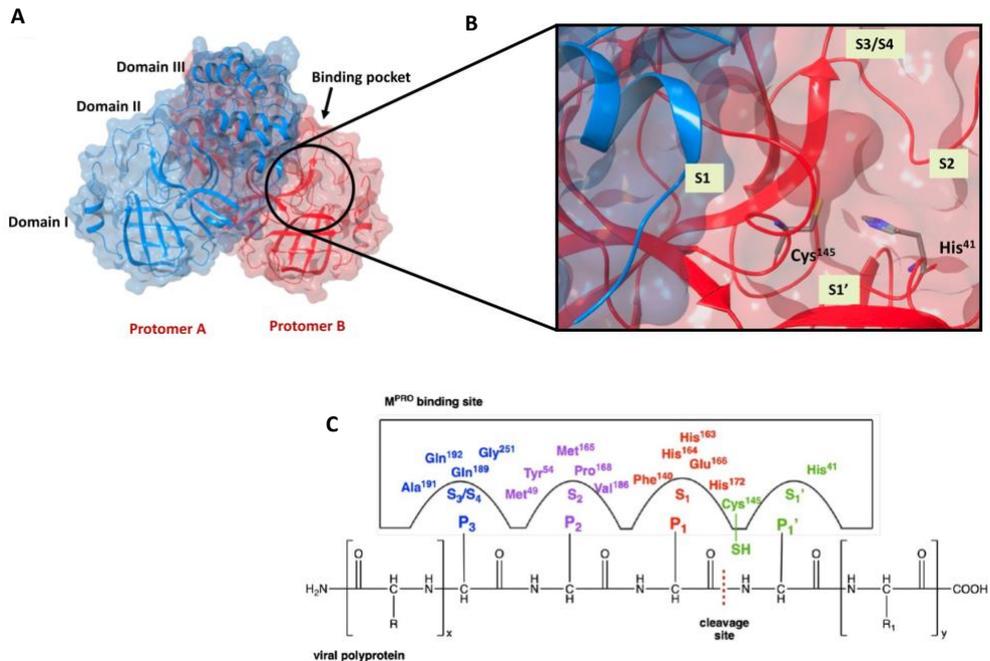


Figure 11. (A) X-ray structure of dimeric SARS-CoV-2 3CL^{PRO} (PDB code 6Y2F). (B) Focus on the catalytic site, with the four regions S1, S1', S2, and S3/S4 highlighted. (C) Warhead P1' is

shown in green, while fragments P1, P2, and P3 are shown in red, purple, and blue, respectively. Subregions of the binding pocket are labeled with S numbering complementary to the fragments of the inhibitor.¹⁷¹

Studies of homology sequence revealed that SARS-CoV-2 3CL^{pro} share a high percentage of identity with other CoVs, indicating ~99% identity with BatCoV RaTG13 3CL^{pro} and ~96% with the previous SARS-CoV 3CL^{pro}.¹⁷² In contrast, only ~50% of whole sequence identity is shared with MERS-CoV M^{pro}.

Structurally, 3CL^{pro} is a homodimer, and each protomer is composed by 306 amino acid, divided into three domains (I, II and III).^{167,173} The domains I and II comprise six antiparallel β -barrels, similar to trypsin-like serine proteases, while the domain III is characterized by a cluster of α -helices. A long loop region connects domain II to domain III. A cleft between domain I and II hosted the substrate binding site with the Cys¹⁴⁵-His⁴¹ catalytic dyad, that is responsible of proteolytic process.^{167,174,175} In details, His⁴¹ deprotonates the thiol of Cys¹⁴⁵ and the resulting thiolate carries out a nucleophile addition to the amide bond of the substrate, leading to the formation of an oxyanion, stabilized by the Gly¹⁴³ and Cys¹⁴⁵ backbones. The N-terminal peptide product is released by proton extraction from His⁴¹ and the thioester is hydrolyzed to release the C-terminal product and recover the catalytic dyad (Fig. 11).^{176,177,178}

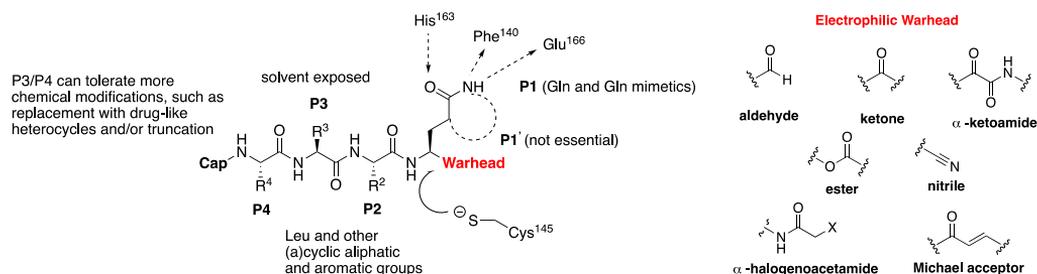
3CL^{pro} is active in its homodimer form, while the 3CL^{pro} monomer is basically inactive.^{177,179} An ionic bond involves Arg⁴ and Glu²⁹⁰ of each protomer to drive the formation of the dimer, while the N-terminal tail, named “N-finger”, between domain II and III allows the formation of substrate-binding site in a cleft located between domains I and II. Four subsites are identified into binding pocket, named S1', S1, S2, and S3/S4, respectively occupied by the portions P1', P1, P2, and P3 of the viral polyproteins (Fig. 11).^{180,181} Generally, 3CL^{pro} recognizes a glutamine in P1, which is highly conserved among polyproteins cleavage sites of SARS-CoV, MERS-CoV and SARS-CoV-2. P2 position accepts small and hydrophobic

amino acids, with preference for leucine. P1' tolerates small amino acidic residues, such as serine or alanine.^{166,180,182,183} This substrate specificity is conserved across other coronaviruses, including SARS-CoV and MERS-CoV.^{176,184,185} Notably, X-Ray studies highlighted a high degree of structural similarity and conservation of active site across SARS-CoV-2, SARS-CoV and MERS-CoV, providing a promising target to identify pan-coronavirus agents. On basis of this similarity, initial efforts to discover new inhibitors against SARS-CoV-2 3CL^{pro} have been used on basis of the compounds already known to target SARS-CoV and MERS-CoV main proteases.

Over the past two years, considerable efforts were pursued by academia and pharmaceutical companies to discover new compounds against SARS-CoV-2 3CL^{pro}. A variety of peptidomimetics were investigated and many crystal structures are now available in complex with such compounds (643 PBD structures of SARS-CoV-2 3CL^{pro}, as of October 26, 2022).¹⁷¹ Most of these act as covalent inhibitors of SARS-CoV-2 3CL^{pro},^{186,187} being functionalized at C-terminus with an electrophilic warhead able to react with the catalytic sulfur of Cys¹⁴⁵ and form a covalent bond. Different chemical warheads have been explored, including Michael acceptors, α -ketoamides, aldehydes, ketones, nitrile, ester, α -haloacetamide (Fig. 12A).¹⁷¹ To trap the Cys¹⁴⁵, these inhibitors need to bind the catalytic site of 3CL^{pro} and bring the warhead to the cleavage site of the viral polyproteins. With this aim, different groups from P1 to P4 positions were investigated to reproduce the substrate like sequence and establish a considerable number of specific interactions with the binding site (Fig. 12A). Generally, the inhibitors were characterized by γ -lactam in P1, able to mimetic the glutamine and engage H-bonds with amino acids Phe¹⁴⁰, His¹⁶³ and Glu¹⁶⁶; in line with the sequence of the natural substrate, amino acids with an aliphatic side chain (i.e. leucine) were preferred at P2 positions, which can occupy S2 subpocket and form favorable interactions; different chemical groups at P3/P4 could be tolerated,

even including replacement of amino acids with heterocycles of various molecular size.¹⁴³ As shown in Fig. 12B, the covalent enzymatic inhibition involves two steps: i) the inhibitor (I) binds to the active site of the enzyme (E) and forms a reversible noncovalent enzyme – inhibitor complex (EI); ii) a more stable complex (E-I) was formed by the nucleophilic attack of the catalytic Cys¹⁴⁵. Analyzing the kinetic scheme, the potency of the first phase is defined by the equilibrium-binding constant k_i ($k_1 = k_2/k_1$), while the second phase is correlated to the inactivation constant k_{inact} (k_3) due to covalent bond formation.¹⁸⁸ It is worth noting that the covalent inhibition strategy was already proposed against other viral protease, resulting a successful approach to the drug discovery.¹⁸⁹ Indeed, the electrophilic warhead inhibitors ensured a better efficacy and higher potency toward molecular target, compared to the non-covalent agents.^{186,187}

A



B

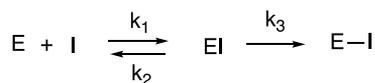


Figure 12. (A) Schematization of the main features of 3CL^{pro} inhibitors and protease subsite specificity, with electrophilic warheads characterizing covalent SARS-CoV-2 3CL^{pro} inhibitors. (B) Kinetic scheme of covalent inhibition. E, I, EI, and E-I stand for enzyme, inhibitor, noncovalent enzyme–inhibitor complex, and covalent enzyme–inhibitor complex, respectively.

The first X-ray structure of SARS-CoV-2 3CL^{pro} was in complex with the tetrapeptide N3 (**13**), already known to be active against SARS-CoV and MERS-

CoV 3CL^{pro}.¹⁹⁰ **13** presents an α,β -unsaturated Michael acceptor group as electrophilic warhead (Fig. 13), which is able to irreversibly binds the thiol group of catalytic Cys¹⁴⁵ by a conjugate addition mechanism (PDB 6LU7).^{190,191} It showed to inhibit SARS-CoV-2 replication with an EC₅₀ = 16.77 μ M.^{192,193} Therefore, **13** has modest selective index with SARS-CoV-2 respect to other CoVs (EC₅₀ = 4.0, 8.8, 2.7, and 3.4 μ M against HCoV-229E, FIPV, MHV-A59, and MHV, respectively).

Subsequently, considerable reversible covalent peptidomimetic have been proposed, showing a potent inhibition against SARS-CoV-2 3CL^{pro}.^{194,195} A significant relevance was achieved by GC-376 (**14**). Initially designed to target the 3CL^{pro} of feline CoV, **14** was able to potently inhibit SARS-CoV-2 3CL^{pro} (IC₅₀ = 30 nM) and viral replication (EC₅₀ = 3.37 μ M in CPE assay), without showing cytotoxicity up to 100 μ M. Structurally, GC-376 is a α -hydroxy-bisulfite dipeptide (Fig. 13), able to rapidly release the aldehyde warhead.¹⁹⁶ *In vivo* studies were performed in K18 hACE2/SARS-CoV-2 transgenic mouse model to evaluate the activity of GC-376 against SARS-CoV-2 infection. Although **GC-376** gave a positive outcome in terms of clinical symptoms and survival, a significant reduction of tissue lesions and inflammation was not observed.¹⁹⁷ However, the covalent reversible inhibition revealed a successful strategy for the drug discovery. To date, compounds belonging to this category, such as PF-00835231 (**15**),¹⁹⁸ PBI-0451 (**17**),¹⁹⁹ ALG-097161 (**16**)²⁰⁰, EDP-235 (undisclosed structure, patent WO 2022/020242 A1)²⁰¹ are current clinical candidates, while Nirmatrelvir is the only inhibitor authorized for the treatment of COVID-19 (Fig. 13).²⁰²

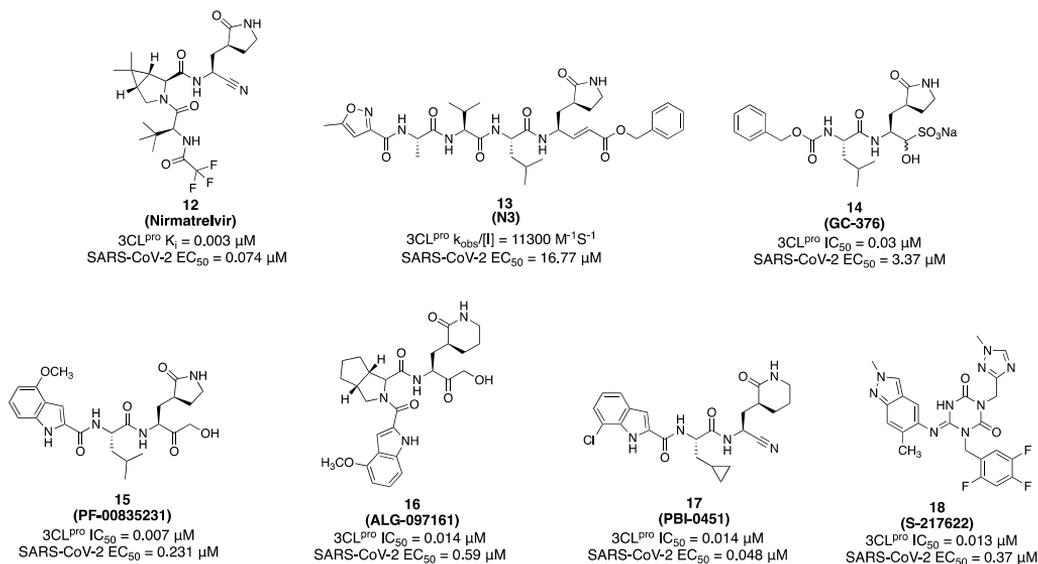


Figure 13. Chemical structures of represented inhibitors of SARS-CoV-2 3CL^{PRO} with biological activities. Antiviral activity evaluated by CPE reduction in Vero cell line.

Therefore, Uho *et al.* described a virtual screening study followed by a biological screening and hit optimization, that allowed to discover the first oral noncovalent, nonpeptidic SARS-CoV-2 3CL^{PRO} inhibitor clinical candidate: S-217622 (**18**).²⁰³ Interestingly, **18** also showed a good antiviral activity *in vitro* against a range of SARS-CoV-2 variants and CoVs (Fig. 13). The favorable preclinical profile that **18** exhibited *in vivo* allowed to start clinical trial studies, advanced to phase III so far (NCT05305547).

However, the ongoing pandemic poses the urgent demand for effective treatments for COVID-19 disease. Moreover, the emergence of possible resistance against Paxlovid raised the need to discover additional efficient inhibitors of SARS-CoV-2 3CL^{PRO} . A recent study, conducted by Dirk Jochmans *et al.* showed that SARS-CoV-2 accumulated mutations at amino acid sequence of 3CL^{PRO} after 12 rounds of Nirmatrelvir treatment – specifically the substitutions L50F, E166A, L167F – and their combination decreased of 20-fold the sensibility

to the drug treatment.²⁰⁰ Nearly at the same time, another *in vitro* study carried out by Yuyong Zhou *et al.* showed that the co-presence of mutations L50F and E166V in 3CL^{pro} revealed up to 80-fold resistance to Nirmatrelvir treatment.²⁰⁴ Both studies conclude that relevant loss of enzymatic 3CL^{pro} activity was correlated to a reduction of the number of inhibitor/enzyme interactions due to these substitutions, while the binding of the substrate was conserved. However, it is interesting to note that the mutations to amino acids 166 and 167 of 3CL^{pro} were already circulating in human population, even before the introduction of Nirmatrelvir in the therapeutic regime of COVID-19.²⁰⁵ This proved that SARS-CoV-2 presents some flexibility at the positions 166 and 167 of 3CL^{pro} amino acid sequence and poses a risk for a potential emergence of drug-resistance to Nirmatrelvir.

1.4. References Introduction

1. Nii-Trebi, N. I. Emerging and Neglected Infectious Diseases: Insights, Advances, and Challenges. *Biomed Res. Int.* **2017**, (2017).
2. Bloom, D. E. & Cadarette, D. Infectious disease threats in the twenty-first century: Strengthening the global response. *Front. Immunol.* **10**, 549 (2019).
3. Grubaugh, N. D. *et al.* Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* **2018 41 4**, 10–19 (2018).
4. Chowell, G. *et al.* Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico. *PLOS Med.* **8**, e1000436 (2011).
5. Cenciarelli, O. *et al.* Ebola virus disease 2013-2014 outbreak in West Africa: An analysis of the epidemic spread and response. *Int. J. Microbiol.*

- 2015**, (2015).
6. Zanoluca, C. *et al.* First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz* **110**, 569–572 (2015).
 7. Drosten, C. *et al.* Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. <https://doi.org/10.1056/NEJMoa030747> **348**, 1967–1976 (2003).
 8. Zumla, A., Hui, D. S. & Perlman, S. Middle East respiratory syndrome. *Lancet (London, England)* **386**, 995–1007 (2015).
 9. Bloom, D. E., Black, S. & Rappuoli, R. Emerging infectious diseases: A proactive approach. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4055–4059 (2017).
 10. Smith, K. M., Machalaba, C. C., Seifman, R., Feferholtz, Y. & Karesh, W. B. Infectious disease and economics: The case for considering multi-sectoral impacts. *One Heal.* **7**, 100080 (2019).
 11. Prioritizing diseases for research and development in emergency contexts. <https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts>.
 12. Holbrook, M. R. Historical Perspectives on Flavivirus Research. *Viruses* **2017**, Vol. 9, Page 97 **9**, 97 (2017).
 13. Bernatchez, J. A. *et al.* Drugs for the Treatment of Zika Virus Infection. *J. Med. Chem.* **63**, 470–489 (2020).
 14. Spengler, U. Direct antiviral agents (DAAs) - A new age in the treatment of hepatitis C virus infection. *Pharmacol. Ther.* **183**, 118–126 (2018).
 15. Mackenzie, J. Wrapping things up about virus RNA replication. *Traffic* **6**, 967–977 (2005).
 16. GW, D., SF, K. & AJ, H. Zika virus. I. Isolations and serological

- specificity. *Trans. R. Soc. Trop. Med. Hyg.* **46**, 509–520 (1952).
17. Duffy, M. R. *et al.* Zika Virus Outbreak on Yap Island, Federated States of Micronesia. <https://doi.org/10.1056/NEJMoa0805715> **360**, 2536–2543 (2009).
 18. Cao-Lormeau, V. M. *et al.* Zika Virus, French Polynesia, South Pacific, 2013 - Volume 20, Number 6—June 2014 - Emerging Infectious Diseases journal - CDC. *Emerg. Infect. Dis.* **20**, 1085–1086 (2014).
 19. Rapid risk assessment: Zika virus epidemic in the Americas: potential association with microcephaly and Guillain-Barré syndrome - 4th update, 10 December 2015. <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-zika-virus-epidemic-americas-potential-association>.
 20. Zanluca, C. *et al.* First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz* **110**, 569–572 (2015).
 21. Zika Cumulative Cases - 25 August 2016 - PAHO/WHO | Pan American Health Organization. <https://www.paho.org/en/documents/zika-cumulative-cases-25-august-2016>.
 22. Zika virus infection – Dominican Republic. <https://www.who.int/emergencies/disease-outbreak-news/item/27-january-2016-zika-dominican-republic-en>.
 23. Suzuki, Y. *et al.* Uncovering the Repertoire of Endogenous Flaviviral Elements in Aedes Mosquito Genomes. *J. Virol.* **91**, 571–588 (2017).
 24. Zhao, L., Alto, B. W., Smartt, C. T. & Shin, D. Transcription Profiling for Defensins of *Aedes aegypti* (Diptera: Culicidae) During Development and in Response to Infection With Chikungunya and Zika Viruses. *J. Med. Entomol.* **55**, 78–89 (2018).

25. Ahmad, S. S. Y., Amin, T. N. & Ustianowski, A. Zika virus: management of infection and risk. *BMJ* **352**, (2016).
26. Johansson, M. A., Mier-y-Teran-Romero, L., Reefhuis, J., Gilboa, S. M. & Hills, S. L. Zika and the Risk of Microcephaly. *N. Engl. J. Med.* **375**, 1–4 (2016).
27. Acosta-Ampudia, Y. *et al.* Autoimmune neurological conditions associated with zika virus infection. *Front. Mol. Neurosci.* **11**, (2018).
28. Pinto-Díaz, C. A. *et al.* Autoimmunity in Guillain-Barré syndrome associated with Zika virus infection and beyond. *Autoimmun. Rev.* **16**, 327–334 (2017).
29. Mai, C. T. *et al.* Selected Birth Defects Data from Population-Based Birth Defects Surveillance Programs in the United States, 2006 to 2010: Featuring Trisomy Conditions. *Birth Defects Res. A. Clin. Mol. Teratol.* **97**, 709 (2013).
30. Guillain-Barré syndrome - Symptoms - NHS. <https://www.nhs.uk/conditions/guillain-barre-syndrome/symptoms/>.
31. White, M. K., Wollebo, H. S., David Beckham, J., Tyler, K. L. & Khalili, K. Zika virus: An emergent neuropathological agent. *Ann. Neurol.* **80**, 479–489 (2016).
32. Kuno, G. & Chang, G. J. J. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch. Virol.* **152**, 687–696 (2007).
33. Singh, R. K. *et al.* Zika virus – emergence, evolution, pathology, diagnosis, and control: current global scenario and future perspectives – a comprehensive review. <http://dx.doi.org/10.1080/01652176.2016.1188333> **36**, 150–175 (2016).

34. Blitvich, B. J. & Firth, A. E. Insect-Specific Flaviviruses: A Systematic Review of Their Discovery, Host Range, Mode of Transmission, Superinfection Exclusion Potential and Genomic Organization. *Viruses* 2015, Vol. 7, Pages 1927-1959 **7**, 1927–1959 (2015).
35. Kostyuchenko, V. A. *et al.* Structure of the thermally stable Zika virus. *Nat.* 2016 5337603 **533**, 425–428 (2016).
36. Neal, J. W. Flaviviruses are neurotropic, but how do they invade the CNS? *J. Infect.* **69**, 203–215 (2014).
37. Smit, J. M., Moesker, B., Rodenhuis-Zybert, I. & Wilschut, J. Flavivirus cell entry and membrane fusion. *Viruses* **3**, 160–171 (2011).
38. Perera-Lecoin, M., Meertens, L., Carnec, X. & Amara, A. Flavivirus Entry Receptors: An Update. *Viruses* 2014, Vol. 6, Pages 69-88 **6**, 69–88 (2013).
39. Murray, C. L., Jones, C. T. & Rice, C. M. Architects of Assembly: roles of Flaviviridae nonstructural proteins in virion morphogenesis. *Nat. Rev. Microbiol.* **6**, 699 (2008).
40. Wu, J., Lu, G., Zhang, B. & Gong, P. Perturbation in the Conserved Methyltransferase-Polymerase Interface of Flavivirus NS5 Differentially Affects Polymerase Initiation and Elongation. *J. Virol.* **89**, 249–261 (2015).
41. Welsch, S. *et al.* Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell Host Microbe* **5**, 365–375 (2009).
42. Abrams, R. P. M., Solis, J. & Nath, A. Therapeutic Approaches for Zika Virus Infection of the Nervous System. *Neurother.* 2017 144 **14**, 1027–1048 (2017).
43. Pattnaik, A., Sahoo, B. R. & Pattnaik, A. K. Current Status of Zika Virus

- Vaccines: Successes and Challenges. *Vaccines* 2020, Vol. 8, Page 266 **8**, 266 (2020).
44. Duffy, M. R. *et al.* Zika Virus Outbreak on Yap Island, Federated States of Micronesia. <https://doi.org/10.1056/NEJMoa0805715> **360**, 2536–2543 (2009).
 45. Gould, E. & Solomon, T. Pathogenic flaviviruses. *Lancet (London, England)* **371**, 500–509 (2008).
 46. Yao, Y. *et al.* Discovery, X-ray Crystallography and Antiviral Activity of Allosteric Inhibitors of Flavivirus NS2B-NS3 Protease. *J. Am. Chem. Soc.* **141**, 6832–6836 (2019).
 47. Zhang, Z. *et al.* crystal structure of unlinked NS2B-NS3 protease from Zika virus. *Science (80-.)*. **354**, 1597–1600 (2016).
 48. Phoo, W. W. *et al.* Structure of the NS2B-NS3 protease from Zika virus after self-cleavage. *Nat. Commun.* 2016 71 **7**, 1–8 (2016).
 49. Li, Y. *et al.* Structural Insights into the Inhibition of Zika Virus NS2B-NS3 Protease by a Small-Molecule Inhibitor. *Structure* **26**, 555-564.e3 (2018).
 50. Nitsche, C. *et al.* Peptide-Boronic Acid Inhibitors of Flaviviral Proteases: Medicinal Chemistry and Structural Biology. *J. Med. Chem.* **60**, 511–516 (2017).
 51. Chen, X. *et al.* Mechanisms of activation and inhibition of Zika virus NS2B-NS3 protease. *Cell Res.* **26**, 1260–1263 (2016).
 52. Shiryayev, S. A. *et al.* Characterization of the Zika virus two-component NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists. *Antiviral Res.* **143**, 218–229 (2017).

53. Leung, D. *et al.* Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. *J. Biol. Chem.* **276**, 45762–45771 (2001).
54. Lee, H. *et al.* Identification of novel small molecule inhibitors against NS2B/NS3 serine protease from Zika virus. *Antiviral Res.* **139**, 49–58 (2017).
55. Nitsche, C. *et al.* De Novo Discovery of Nonstandard Macrocyclic Peptides as Noncompetitive Inhibitors of the Zika Virus NS2B-NS3 Protease. *ACS Med. Chem. Lett.* **10**, 168–174 (2019).
56. Brecher, M. *et al.* A conformational switch high-throughput screening assay and allosteric inhibition of the flavivirus NS2B-NS3 protease. *PLoS Pathog.* **13**, e1006411 (2017).
57. Bollati, M. *et al.* Structure and functionality in flavivirus NS-proteins: perspectives for drug design. *Antiviral Res.* **87**, 125–148 (2010).
58. Swarbrick, C. M. D. *et al.* NS3 helicase from dengue virus specifically recognizes viral RNA sequence to ensure optimal replication. *Nucleic Acids Res.* **45**, 12904–12920 (2017).
59. Ahidjo, B. A., Loe, M. W. C., Ng, Y. L., Mok, C. K. & Chu, J. J. H. Current Perspective of Antiviral Strategies against COVID-19. *ACS Infect. Dis.* **6**, 1624–1634 (2020).
60. Lim, S. P., Noble, C. G. & Shi, P. Y. The dengue virus NS5 protein as a target for drug discovery. *Antiviral Res.* **119**, 57–67 (2015).
61. Dong, H., Zhang, B. & Shi, P. Y. Flavivirus methyltransferase: a novel antiviral target. *Antiviral Res.* **80**, 1–10 (2008).
62. Kim, S. Y., Li, B. & Linhardt, R. J. Pathogenesis and Inhibition of Flaviviruses from a Carbohydrate Perspective. *Pharmaceuticals* **10**,

- (2017).
63. Zhang, X. *et al.* Structures and Functions of the Envelope Glycoprotein in Flavivirus Infections. *Viruses* **9**, (2017).
 64. de Wispelaere, M. *et al.* Inhibition of Flaviviruses by Targeting a Conserved Pocket on the Viral Envelope Protein. *Cell Chem. Biol.* **25**, 1006-1016.e8 (2018).
 65. Wu, J., Lu, G., Zhang, B. & Gong, P. Perturbation in the Conserved Methyltransferase-Polymerase Interface of Flavivirus NS5 Differentially Affects Polymerase Initiation and Elongation. *J. Virol.* **89**, 249–261 (2015).
 66. Godoy, A. S. *et al.* Crystal structure of Zika virus NS5 RNA-dependent RNA polymerase. *Nat. Commun.* 2017 81 **8**, 1–6 (2017).
 67. Castro, C. *et al.* Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* 2009 162 **16**, 212–218 (2009).
 68. Yap, T. L. *et al.* Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution. *J. Virol.* **81**, 4753–4765 (2007).
 69. Wu, J., Liu, W., Gong, P. & Gong, P. A Structural Overview of RNA-Dependent RNA Polymerases from the Flaviviridae Family. *Int. J. Mol. Sci.* 2015, Vol. 16, Pages 12943-12957 **16**, 12943–12957 (2015).
 70. Lim, S. P. *et al.* Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling. *PLOS Pathog.* **12**, e1005737 (2016).
 71. Sofia, M. J., Chang, W., Furman, P. A., Mosley, R. T. & Ross, B. S. Nucleoside, nucleotide, and non-nucleoside inhibitors of hepatitis C virus NS5B RNA-dependent RNA-polymerase. *J. Med. Chem.* **55**, 2481–2531

- (2012).
72. Mumtaz, N. *et al.* Cell-line dependent antiviral activity of sofosbuvir against Zika virus. *Antiviral Res.* **146**, 161–163 (2017).
 73. Ferreira, A. C. *et al.* Sofosbuvir protects Zika virus-infected mice from mortality, preventing short- and long-term sequelae. *Sci. Reports 2017 71* **7**, 1–9 (2017).
 74. Mesci, P. *et al.* Blocking Zika virus vertical transmission. *Sci. Reports 2018 81* **8**, 1–13 (2018).
 75. Zmurko, J. *et al.* The Viral Polymerase Inhibitor 7-Deaza-2'-C-Methyladenosine Is a Potent Inhibitor of In Vitro Zika Virus Replication and Delays Disease Progression in a Robust Mouse Infection Model. *PLoS Negl. Trop. Dis.* **10**, (2016).
 76. Munjal, A. *et al.* Advances in developing therapies to combat zika virus: Current knowledge and future perspectives. *Front. Microbiol.* **8**, 1469 (2017).
 77. Yin, Z. *et al.* An adenosine nucleoside inhibitor of dengue virus. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20435–20439 (2009).
 78. Deng, Y. Q. *et al.* Adenosine Analog NITD008 Is a Potent Inhibitor of Zika Virus. *Open Forum Infect. Dis.* **3**, (2016).
 79. Warren, T. K. *et al.* Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430. *Nature* **508**, 402–405 (2014).
 80. Julander, J. G. *et al.* Efficacy of the broad-spectrum antiviral compound BCX4430 against Zika virus in cell culture and in a mouse model. *Antiviral Res.* **137**, 14–22 (2017).
 81. Noble, C. G. *et al.* A Conserved Pocket in the Dengue Virus Polymerase

- Identified through Fragment-based Screening. *J. Biol. Chem.* **291**, 8541–8548 (2016).
82. Wang, B. *et al.* The structure of Zika virus NS5 reveals a conserved domain conformation. *Nat. Commun.* 2017 **8**, 1–6 (2017).
 83. Duan, W. *et al.* The crystal structure of Zika virus NS5 reveals conserved drug targets. *EMBO J.* **36**, 919–933 (2017).
 84. Sáez-Álvarez, Y. *et al.* Novel nonnucleoside inhibitors of Zika virus polymerase identified through the screening of an open library of antikinoplastid compounds. *Antimicrobial Agents and Chemotherapy* vol. 65 at <https://doi.org/10.1128/AAC.00894-21> (2021).
 85. Weiss, S. R. & Leibowitz, J. L. Coronavirus pathogenesis. *Adv. Virus Res.* **81**, 85–164 (2011).
 86. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
 87. Coronavirus disease 2019 (COVID-19): situation report, 51. <https://apps.who.int/iris/handle/10665/331475>.
 88. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. <https://covid19.who.int/>.
 89. Chan, J. F. W. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **9**, 221–236 (2020).
 90. Su, S. *et al.* Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).
 91. Peiris, J. S. M. Severe Acute Respiratory Syndrome (SARS). *J. Clin. Virol.* **28**, 245 (2003).

92. Chan-Yeung, M., Xu, R., Chan-Yeung, M. & Chan-yeung, M. SARS: epidemiology. *Respirology* **8**, S9–S14 (2003).
93. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
94. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
95. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
96. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nat.* 2020 5797798 **579**, 270–273 (2020).
97. Zhang, R., Li, Y., Zhang, A. L., Wang, Y. & Molina, M. J. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14857–14863 (2020).
98. Falahi, S. & Kenarkoohi, A. Transmission routes for SARS-CoV-2 infection: review of evidence. *New Microbes New Infect.* **38**, 100778 (2020).
99. Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
100. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
101. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet (London, England)* **395**, 507–513 (2020).

102. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)* **395**, 497–506 (2020).
103. Xu, Z. *et al.* Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet. Respir. Med.* **8**, 420–422 (2020).
104. Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020).
105. Sanche, S. *et al.* High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1470 (2020).
106. Silva, S. J. R. da, Lima, S. C. de, Silva, R. C. da, Kohl, A. & Pena, L. Viral Load in COVID-19 Patients: Implications for Prognosis and Vaccine Efficacy in the Context of Emerging SARS-CoV-2 Variants. *Front. Med.* **8**, (2022).
107. Luring, A. S. & Hodcroft, E. B. Genetic Variants of SARS-CoV-2—What Do They Mean? *JAMA* **325**, 529–531 (2021).
108. Tracking SARS-CoV-2 variants. <https://www.who.int/activities/tracking-SARS-CoV-2-variants>.
109. Silva, S. J. R. da & Pena, L. Collapse of the public health system and the emergence of new variants during the second wave of the COVID-19 pandemic in Brazil. *One Heal. (Amsterdam, Netherlands)* **13**, (2021).
110. Cov-Lineages. https://cov-lineages.org/lineage_list.html.
111. Dhawan, M. *et al.* Delta variant (B.1.617.2) of SARS-CoV-2: Mutations, impact, challenges and possible solutions. *Hum. Vaccin. Immunother.* **18**, (2022).

112. Chen, J., Wang, R., Gilby, N. B. & Wei, G. W. Omicron Variant (B.1.1.529): Infectivity, Vaccine Breakthrough, and Antibody Resistance. *J. Chem. Inf. Model.* **62**, 412–422 (2022).
113. Vaughan, A. Omicron emerges. *New Sci.* **252**, 7 (2021).
114. Callaway, E. Heavily mutated Omicron variant puts scientists on alert. *Nature* **600**, 21 (2021).
115. Cao, Y. *et al.* Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *bioRxiv* 2022.09.15.507787 (2022) doi:10.1101/2022.09.15.507787.
116. Shereen, M. A., Khan, S., Kazmi, A., Bashir, N. & Siddique, R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* **24**, 91–98 (2020).
117. Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **1282**, 1–23 (2015).
118. 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).
119. Cannalire, R. *et al.* SARS-CoV-2 Entry Inhibitors: Small Molecules and Peptides Targeting Virus or Host Cells. *Int. J. Mol. Sci.* 2020, Vol. 21, Page 5707 **21**, 5707 (2020).
120. Bosch, B. J., van der Zee, R., de Haan, C. A. M. & Rottier, P. J. M. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**, 8801–8811 (2003).
121. Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F. & Matsuyama, S. Simultaneous Treatment of Human Bronchial Epithelial Cells with Serine

- and Cysteine Protease Inhibitors Prevents Severe Acute Respiratory Syndrome Coronavirus Entry. *J. Virol.* **86**, 6537–6545 (2012).
122. Coutard, B. *et al.* The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* **176**, (2020).
 123. Bestle, D. *et al.* TMPRSS2 and furin are both essential for proteolytic activation and spread of SARS-CoV-2 in human airway epithelial cells and provide promising drug targets. *bioRxiv* 2020.04.15.042085 (2020) doi:10.1101/2020.04.15.042085.
 124. Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **1282**, 1–23 (2015).
 125. Da Silva, S. J. R. *et al.* Two Years into the COVID-19 Pandemic: Lessons Learned. *ACS Infectious Diseases* at <https://doi.org/10.1021/acsinfecdis.2c00204> (2022).
 126. 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.* **2021** *231* **23**, 21–39 (2021).
 127. Cheng, Y. *et al.* Face masks effectively limit the probability of SARS-CoV-2 transmission. *Science* **372**, 1339–1343 (2021).
 128. Antiviral Therapy Summary Recommendations | COVID-19 Treatment Guidelines.
<https://www.covid19treatmentguidelines.nih.gov/therapies/antiviral-therapy/summary-recommendations/>.
 129. Anti-SARS-CoV-2 Antibody Products | COVID-19 Treatment Guidelines.

- <https://www.covid19treatmentguidelines.nih.gov/therapies/anti-sars-cov-2-antibody-products/>.
130. Criteria for monoclonal antibodies COVID-19 - UpToDate. <https://www.uptodate.com/contents/image/print?imageKey=PC%2F131725>.
 131. Immunomodulators | COVID-19 Treatment Guidelines. <https://www.covid19treatmentguidelines.nih.gov/therapies/immunomodulators/>.
 132. Vaccines – COVID19 Vaccine Tracker. <https://covid19.trackvaccines.org/vaccines/approved/#vaccine-list>.
 133. COVID-19 vaccines: authorised | European Medicines Agency. <https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-authorised>.
 134. COVID-19 Vaccines | FDA. <https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines>.
 135. Yang, S. *et al.* Safety and immunogenicity of a recombinant tandem-repeat dimeric RBD-based protein subunit vaccine (ZF2001) against COVID-19 in adults: two randomised, double-blind, placebo-controlled, phase 1 and 2 trials. *Lancet Infect. Dis.* **21**, 1107–1119 (2021).
 136. Enjuanes, L. *et al.* Molecular Basis of Coronavirus Virulence and Vaccine Development. *Adv. Virus Res.* **96**, 245–286 (2016).
 137. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
 138. Dagan, N. *et al.* BNT162b2 mRNA Covid-19 Vaccine in a Nationwide

- Mass Vaccination Setting. *N. Engl. J. Med.* **384**, 1412–1423 (2021).
139. Mallapaty, S. India's DNA COVID vaccine is a world first - more are coming. *Nature* **597**, 161–162 (2021).
 140. Granados-Riveron, J. T. & Aquino-Jarquín, G. Engineering of the current nucleoside-modified mRNA-LNP vaccines against SARS-CoV-2. *Biomed. Pharmacother.* **142**, (2021).
 141. Kyriakidis, N. C., López-Cortés, A., González, E. V., Grimaldos, A. B. & Prado, E. O. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ vaccines* **6**, (2021).
 142. Fiolet, T., Kherabi, Y., MacDonald, C. J., Ghosn, J. & Peiffer-Smadja, N. Comparing COVID-19 vaccines for their characteristics, efficacy and effectiveness against SARS-CoV-2 and variants of concern: a narrative review. *Clin. Microbiol. Infect.* **28**, 202–221 (2022).
 143. 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.* **65**, 2716–2746 (2022).
 144. Pushpakom, S. *et al.* Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2019).
 145. Alpern, J. D. & Gertner, E. Off-Label Therapies for COVID-19-Are We All In This Together? *Clin. Pharmacol. Ther.* **108**, 182–184 (2020).
 146. Siegel, D. *et al.* Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo[2,1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. *J. Med. Chem.* **60**, 1648–1661 (2017).
 147. Agostini, M. L. *et al.* Coronavirus susceptibility to the antiviral remdesivir

- (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *MBio* **9**, (2018).
148. Warren, T. K. *et al.* Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nat. 2016 5317594* **531**, 381–385 (2016).
 149. Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Götte, M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses 2019, Vol. 11, Page 326* **11**, 326 (2019).
 150. Pruijssers, A. J. *et al.* Remdesivir potently inhibits SARS-CoV-2 in human lung cells and chimeric SARS-CoV expressing the SARS-CoV-2 RNA polymerase in mice. *bioRxiv* 2020.04.27.064279 (2020) doi:10.1101/2020.04.27.064279.
 151. Pruijssers, A. J. *et al.* Remdesivir Inhibits SARS-CoV-2 in Human Lung Cells and Chimeric SARS-CoV Expressing the SARS-CoV-2 RNA Polymerase in Mice. *Cell Rep.* **32**, (2020).
 152. Williamson, B. N. *et al.* Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *Nat. 2020 5857824* **585**, 273–276 (2020).
 153. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *N. Engl. J. Med.* **383**, 1813–1826 (2020).
 154. Wang, Y. *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* **395**, 1569–1578 (2020).
 155. Gilead’s Investigational Antiviral Remdesivir Receives U.S. Food and Drug Administration Emergency Use Authorization for the Treatment of COVID-19. <https://www.gilead.com/news-and-press/press-room/press-releases/2020/5/gileads-investigational-antiviral-remdesivir-receives-us->

food-and-drug-administration-emergency-use-authorization-for-the-treatment-of-covid19.

156. Rosenke, K. *et al.* Orally delivered MK-4482 inhibits SARS-CoV-2 replication in the Syrian hamster model. *Nat. Commun.* 2021 121 **12**, 1–8 (2021).
157. Tian, L. *et al.* Molnupiravir and Its Antiviral Activity Against COVID-19. *Front. Immunol.* **13**, 1387 (2022).
158. Wahl, A. *et al.* SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nat.* 2021 5917850 **591**, 451–457 (2021).
159. Toots, M. *et al.* Characterization of orally efficacious influenza drug with high resistance barrier in ferrets and human airway epithelia. *Sci. Transl. Med.* **11**, (2019).
160. Abdelnabi, R. *et al.* Molnupiravir Inhibits Replication of the Emerging SARS-CoV-2 Variants of Concern in a Hamster Infection Model. *J. Infect. Dis.* **224**, 749–753 (2021).
161. Merck and Ridgeback’s Investigational Oral Antiviral Molnupiravir Reduced the Risk of Hospitalization or Death by Approximately 50 Percent Compared to Placebo for Patients with Mild or Moderate COVID-19 in Positive Interim Analysis of Phase 3 Study - Merck.com. <https://www.merck.com/news/merck-and-ridgebacks-investigational-oral-antiviral-molnupiravir-reduced-the-risk-of-hospitalization-or-death-by-approximately-50-percent-compared-to-placebo-for-patients-with-mild-or-moderat/>.
162. Regulatory approval of Lagevrio (molnupiravir) - GOV.UK. <https://www.gov.uk/government/publications/regulatory-approval-of-lagevrio-molnupiravir>.

163. Zhao, Y. *et al.* Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein Cell* **13**, 689–693 (2022).
164. Pfizer's Novel COVID-19 Oral Antiviral Treatment Candidate Reduced Risk of Hospitalization or Death by 89% in Interim Analysis of Phase 2/3 EPIC-HR Study | Pfizer. <https://www.pfizer.com/news/press-release/press-release-detail/pfizers-novel-covid-19-oral-antiviral-treatment-candidate>.
165. Service, R. F. Bad news for Paxlovid? Resistance may be coming. *Science* (80-.). **377**, 138–139 (2022).
166. Hilgenfeld, R. & Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res.* **100**, 286–295 (2013).
167. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* (80-.). **368**, 409–412 (2020).
168. Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853–879 (2000).
169. Du, Q. S. *et al.* Polyprotein cleavage mechanism of SARS CoV Mpro and chemical modification of the octapeptide. *Peptides* **25**, 1857–1864 (2004).
170. Agbowuro, A. A., Huston, W. M., Gamble, A. B. & Tyndall, J. D. A. Proteases and protease inhibitors in infectious diseases. *Med. Res. Rev.* **38**, 1295–1331 (2018).
171. La Monica, G., Bono, A., Lauria, A. & Martorana, A. Targeting SARS-CoV-2 Main Protease for Treatment of COVID-19: Covalent Inhibitors Structure–Activity Relationship Insights and Evolution Perspectives. *J.*

- Med. Chem.* **2022**, 12500–12534 (2022).
172. Stoermer, M. J. Homology Models of Coronavirus 2019-nCoV 3CLpro Protease. (2020) doi:10.26434/CHEMRXIV.11637294.V3.
 173. Tan, J. *et al.* pH-dependent Conformational Flexibility of the SARS-CoV Main Proteinase (Mpro) Dimer: Molecular Dynamics Simulations and Multiple X-ray Structure Analyses. *J. Mol. Biol.* **354**, 25–40 (2005).
 174. Yang, H. *et al.* The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190–13195 (2003).
 175. Chou, C. Y. *et al.* Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* **43**, 14958–14970 (2004).
 176. Jin, Z. *et al.* Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, (2020).
 177. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S. H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **59**, 6595–6628 (2016).
 178. Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism†. *Biochemistry* **43**, 4568–4574 (2004).
 179. Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S. C. & Mesecar, A. D. Evaluating the 3C-like protease activity of SARS-Coronavirus: Recommendations for standardized assays for drug discovery. *Virus Res.* **133**, 63–73 (2008).
 180. Fan, K. *et al.* Biosynthesis, Purification, and Substrate Specificity of

- Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J. Biol. Chem.* **279**, 1637–1642 (2004).
181. Chen, C. N. *et al.* Inhibition of SARS-CoV 3C-like Protease Activity by Theaflavin-3,3'-digallate (TF3). *Evidence-Based Complement. Altern. Med.* **2**, 209–215 (2005).
182. Thiel, V. *et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315 (2003).
183. Rut, W. *et al.* Substrate specificity profiling of SARS-CoV-2 main protease enables design of activity-based probes for patient-sample imaging. *bioRxiv* 2020.03.07.981928 (2020) doi:10.1101/2020.03.07.981928.
184. Bzówka, M. *et al.* Structural and Evolutionary Analysis Indicate That the SARS-CoV-2 Mpro Is a Challenging Target for Small-Molecule Inhibitor Design. *Int. J. Mol. Sci.* 2020, Vol. 21, Page 3099 **21**, 3099 (2020).
185. Zhang, C. H. *et al.* Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **7**, 467–475 (2022).
186. Shindo, N. & Ojida, A. Recent progress in covalent warheads for in vivo targeting of endogenous proteins. *Bioorganic Med. Chem.* **47**, (2021).
187. Gehringer, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **62**, 5673–5724 (2019).
188. Yang, H. *et al.* Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. *PLOS Biol.* **3**, e324 (2005).
189. Sharma, A., Kaliya, K. & Maurya, S. K. Recent Advances in the Discovery

- of Potent Proteases Inhibitors Targeting the SARS Coronaviruses. *Curr. Top. Med. Chem.* **21**, 307–328 (2020).
190. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nat.* 2020 5827811 **582**, 289–293 (2020).
191. Wang, F., Chen, C., Tan, W., Yang, K. & Yang, H. Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Sci. Reports* 2016 61 **6**, 1–12 (2016).
192. Ren, Z. *et al.* The newly emerged SARS-Like coronavirus HCoV-EMC also has an “Achilles’ heel”: current effective inhibitor targeting a 3C-like protease. *Protein Cell* 2013 44 **4**, 248–250 (2013).
193. Xue, X. *et al.* Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design. *J. Virol.* **82**, 2515–2527 (2008).
194. Konno, S. *et al.* 3CL Protease Inhibitors with an Electrophilic Arylketone Moiety as Anti-SARS-CoV-2 Agents. *J. Med. Chem.* **65**, 2926–2939 (2022).
195. Dampalla, C. S. *et al.* Structure-Guided Design of Potent Spirocyclic Inhibitors of Severe Acute Respiratory Syndrome Coronavirus-2 3C-like Protease. *J. Med. Chem.* **65**, 7818–7832 (2022).
196. Pedersen, N. C. *et al.* Efficacy of a 3C-like protease inhibitor in treating various forms of acquired feline infectious peritonitis. *J. Feline Med. Surg.* **20**, 378–392 (2018).
197. Cáceres, C. J. *et al.* Efficacy of GC-376 against SARS-CoV-2 virus infection in the K18 hACE2 transgenic mouse model. *Sci. Reports* 2021 111 **11**, 1–13 (2021).
198. Owen, D. R. *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate

- for the treatment of COVID-19. *Science* (80-.). **374**, 1586–1593 (2021).
199. Pipeline - Pardes BioSciences. <https://www.pardesbio.com/pipeline/#our-lead-program>.
 200. Jochmans, D. *et al.* The substitutions L50F, E166A and L167F in SARS-CoV-2 3CLpro are selected by a protease inhibitor in vitro and confer resistance to nirmatrelvir. *bioRxiv* 2022.06.07.495116 (2022) doi:10.1101/2022.06.07.495116.
 201. Chia, C. S. B. & See, Y. Y. Novel Coronavirus Main Protease Di- and Tripeptide Inhibitors for Treating COVID-19. *ACS Med. Chem. Lett.* **13**, 1388–1389 (2022).
 202. Owen, D. R. *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* (80-.). **374**, 1586–1593 (2021).
 203. Unoh, Y. *et al.* Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022**, 6499–6512 (2022).
 204. Zhou, Y. *et al.* Nirmatrelvir Resistant SARS-CoV-2 Variants with High Fitness in Vitro. *bioRxiv* 2022.06.06.494921 (2022) doi:10.1101/2022.06.06.494921.
 205. Sedova, M., Jaroszewski, L., Iyer, M. & Godzik, A. Monitoring for SARS-CoV-2 drug resistance mutations in broad viral populations. *bioRxiv* 2022.05.27.493798 (2022) doi:10.1101/2022.05.27.493798.

CHAPTER 2

BACKGROUND AND AIMS

2. BACKGROUND AND AIMS

2.1. ZIKV: background and aims of the project

As mentioned in the introduction, neither vaccines nor drugs and clinical candidates are available to treat ZIKV and most relevant flaviviruses, highlighting the urgency to identify new anti-ZIKV therapeutics (1.2.2). Only very recently, a phase II clinical candidate (JNJ-64281802)¹ has been reported for DENV1-4.

Mindful of safe and effective DAA-treatment, the development of small molecules specifically targeting viral proteins and interfering with the virus life cycle is an essential part of a successful antiviral strategy (1.2.2.1). In this context, ZIKV NS proteins represent promising targets for the discovery of novel therapeutic agents.² In particular, ZIKV NS5 RdRp is an ideal target, since it is fundamental for viral replication, is conserved across flaviviruses and is absent in mammalian cells (1.2.2.2).

In parallel, the discovery of drugs targeting multiple viral targets may be a promising and innovative strategy for the development of effective therapeutics. Employing drug repurposing with a polypharmacological approach represents a pragmatic strategy to tackle therapeutic emergency, even considering that computational approaches help to accelerate the search for multitargeting inhibitors.^{3,4}

In this regard, the aim of the project was focused on the discovery of new patentable small molecules as anti- flavivirus agents, including ZIKV as pilot example, by exploiting the DAA approach to select a preclinical candidate.

Two parallel strategies have been applied for the hit identification:

- 1) rational design and synthesis of new NNIs of ZIKV NS5 RdRp, by exploiting scaffold hopping of known NNIs of the related DENV and other flavivirus RdRp;
- 2) polypharmacological approach, utilizing the combination of *in silico* generation of a 5 billion compounds collection within a tangible chemical space, with an ultra-fast structure-based virtual screening and synthesis.

2.2. SARS-CoV-2: background and aims of the project

Since late 2019, COVID-19 poses a huge threat on the global public health.

Despite five vaccines available, the need of antiviral therapeutics is still urgent to counteract emerging SARS-CoV-2 variants and future CoV outbreaks, beyond COVID-19. To date, only one repurposed drug (iv) and two oral DAAs are authorized for the use, but with moderate efficacy or suboptimal pharmacokinetic (1.3.2.2).

As part of my doctoral project, aimed at identifying new DAAs against emerging viruses, I also focalized on the discovery of anti-SARS-CoV-2 agents within Exscalate4CoV,⁵ the first project funded by EU to counteract the pandemic (N° 101003551). To be noted, when I started this project in the early 2020, nether vaccines nor drugs were approved for the COVID-19 therapy.

The cardinal role of the 3CL^{pro} in the viral life cycle and its absence in human cells make it an attractive target to develop novel anti-SARS-CoV-2 DAA agents. Moreover, the sequence identity of catalytic site is highly conserved across CoVs, revealing 3CL^{pro} as a potential molecular target for the development of broad-spectrum DAA agents (1.3.2.3).

In this context, I focused on the identification of effective antiviral agents against SARS-CoV-2, dealing with two different projects:

- 1) rational design and synthesis of reversible covalent peptidomimetic inhibitors for broad spectrum inhibition of CoVs 3CL^{pro};
- 2) synthesis and scale-up of benchmark inhibitor Nirmatrelvir to validate an innovative mouse model of COVID-19 infection.

2.3. References Background and Aims

1. Kaptein SJF, Goethals O, Kiemel D, et al. A pan-serotype dengue virus inhibitor targeting the NS3–NS4B interaction. *Nat* 2021 5987881, **571**, 504-509 (2021).
2. Bernatchez JA, Tran LT, Li J, Luan Y, Siqueira-Neto JL, Li R. Drugs for the Treatment of Zika Virus Infection. *J Med Chem.* **63**, 470-489 (2020).
3. Reddy AS, Zhang S. Polypharmacology: drug discovery for the future. *Expert Rev Clin Pharmacol.* **6**, 41-47 (2013).
4. Proschak E, Stark H, Merk D. Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. *J Med Chem.* **62**, 420-444 (2019).
5. Home. Accessed November 2, 2022. <https://www.exscalate4cov.eu/>

CHAPTER 3

SYNTHESIS OF ANTI-ZIKV AGENTS
IDENTIFIED BY RATIONAL DESIGN AND
STRUCTURE-BASED VIRTUAL
SCREENING

Chapter 3.

Synthesis of anti-ZIKV agents identified by rational design and structure-based virtual screening

3.1. Rational design and synthesis of ZIKA NS5 RdRp Non Nucleoside Inhibitors

3.1.1. Hopping of privileged dihydroxypyrimidine nucleus

As mentioned above, I focused on the rational design of novel NNIs of ZIKV NS5 polymerase, in order to identify novel DAAs against ZIKV. When I started this project in late 2019, no ZIKV NNIs were reported yet; therefore, I decided to consider NNIs reported in literature to be active against other flaviviruses, such as DENV, with a potent antiviral activity and a robust characterization of binding mode. In particular, the Novartis Institute for Tropical Diseases reported a promising series of acyl sulfonamide inhibitors of DENV RdRp, that specifically bound to an allosteric site of the NS5 RdRp, called the N-pocket, and were able to inhibit the RNA initiation process.¹ Representative compound **19** showed potent enzyme inhibition ($IC_{50} = 0.17 \mu M$), even if a large shift in the antiviral activity was observed against all four DENV serotypes in cell-based assays ($EC_{50} = 1.8 - 2.3 \mu M$), without a significant cellular toxicity ($CC_{50} > 50 \mu M$) (Fig. 3.1a).¹ For this compound and relative analogs, available SAR information and X-ray cocrystal structure in complex with NS5 showed that the *N*-aryl sulfonamide moiety and the propargyl alcohol are pharmacophoric requirements. In its co-crystal pose in complex with DENV3 RdRp (PDB: 531P), compound **19** interacts with the residues of the N-pocket, projecting the propargyl alcohol into the allosteric binding-site to form two important H-bond interactions with His⁸⁰⁰ and Glu⁸⁰², while the benzenesulfonamide forms three H-bond interactions

with Thr⁷⁹⁴ and Arg⁷²⁹ and the backbone of Trp⁷⁹⁵, and in addition offers an increase of lipophilicity, substantial for cellular activity. A detailed structural analysis of DENV3 NS5 and ZIKV NS5 has shown that the critical residues of N-pocket for the binding of acyl sulfonamide inhibitors are conserved between the two homologue enzymes and arranged in a similar structural environment.² Indeed, the overlay of the RdRps of DENV3 (PDB 5I3P) and ZIKV (PDB 6LD3) bound to compound **19** shows the high structural similarity between the polymerases. Moreover, the compound **19** well inserts into the conserved N-pocket of ZIKV NS5 RdRp, delimited by the priming loop (Fig. 3.1b).

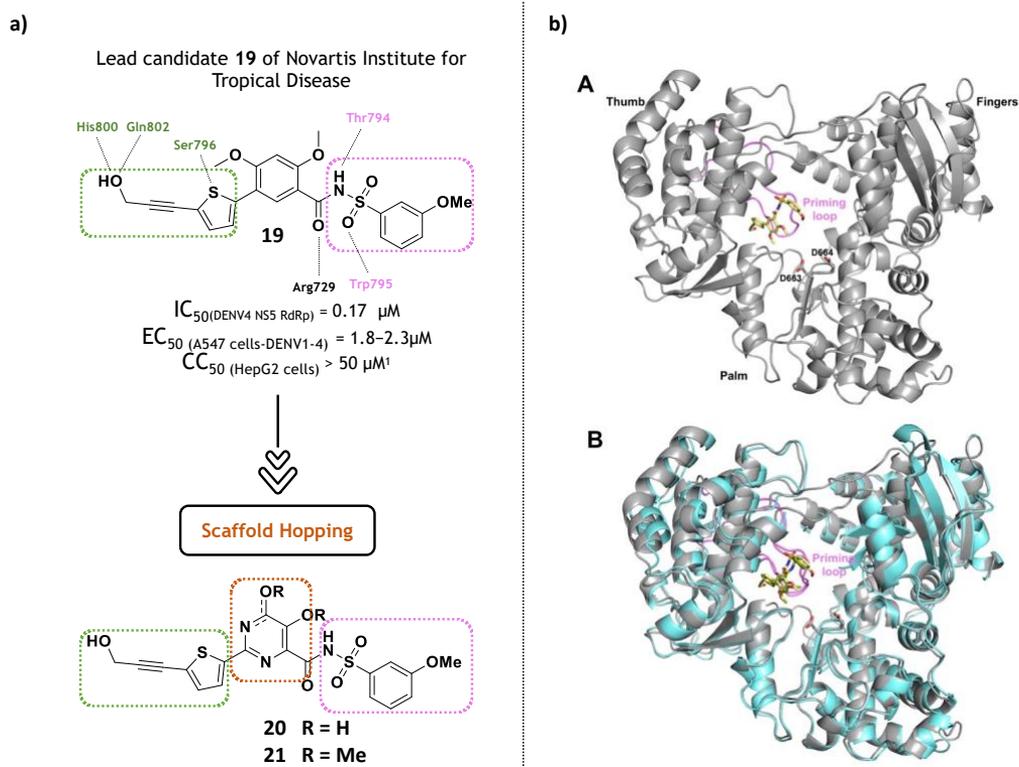


Figure 3.1. a) Design of new NNIs of ZIKV NS5 RdRp by exploiting scaffold hopping. 2D representation and biological activity of compound **19**/DENV3 RdRp (5I3P) crystal structure, highlighting key intermolecular interactions.¹ b) Comparison of the RdRp structure of ZIKV and DENV. (A) Crystal structure of compound **19** (yellow) bound to the RdRp of DENV3, shown as grey cartoon (PDB 5I3P). The palm subdomain contains the active-site residues (Asp⁶⁶³ and

Asp⁶⁶⁴ represented as grey sticks); the priming loop is colored violet. (B) Overlay of the RdRps of DENV3 (PDB 5I3P, grey) and ZIKV (PDB 6LD3, cyan) bound to compound **19** (yellow).

On basis of *in silico*-aided results, the design of ZIKV NNIs was focused on the ability of the new molecules to bind to allosteric pocket and, in addition, to have favorable properties to achieve activity in the cell-based assay. In this regard, the propargyl alcohol and benzenesulfonamide moiety were maintained, while the 2,4-dimethoxyphenyl ring was replaced by a privileged dihydroxypyrimidine scaffold, recurring in different known antiviral agents, such as HCV NS5B RdRp NNIs,^{3,4} and HIV integrase inhibitor drug Raltegravir (Isentress®).⁵ Indeed, since the 2,4-dimethoxyphenyl ring only appeared to act as a rigid scaffold to properly orient the two harms of the interaction moieties, and furthermore, the methoxy groups did not have any role in the interaction in the binding site, we thought to apply a scaffold hopping approach to increase the drug-like properties and bridge the shift of antiviral activity between the enzymatic and cell-based assays. The new designed small molecules **20** and **21** were respectively characterized by a dihydroxypyrimidine and a less polar dimethoxypyrimidine scaffold (Fig. 3.1a). In collaboration with Dr. Carmen Cerchia of University of Naples Federico II, *in silico* docking studies on ZIKV NS5 polymerase (PDB 6LD3) were carried out to predict their ability to bind the N-pocket and the possible intermolecular interactions (Fig. 3.2). Similar to original compound **19**, compound **20** nicely fits into the ZIKV N-pocket: the sulfonamide moiety forms strong interactions with two arginine residues, Arg⁷³⁹ and Arg⁷³¹, whereas the propargyl alcohol engages two H-bonds with Lys⁸⁰² and Glu⁸⁰⁴. Instead, compound **21** adopts a slightly different binding mode, with the sulfonamide moiety shifting downward and forming two H-bonds with Ser⁷⁹⁸ and Ser⁷¹², and the propargyl alcohol engaging in a single H-bond with Glu⁸⁰⁴.

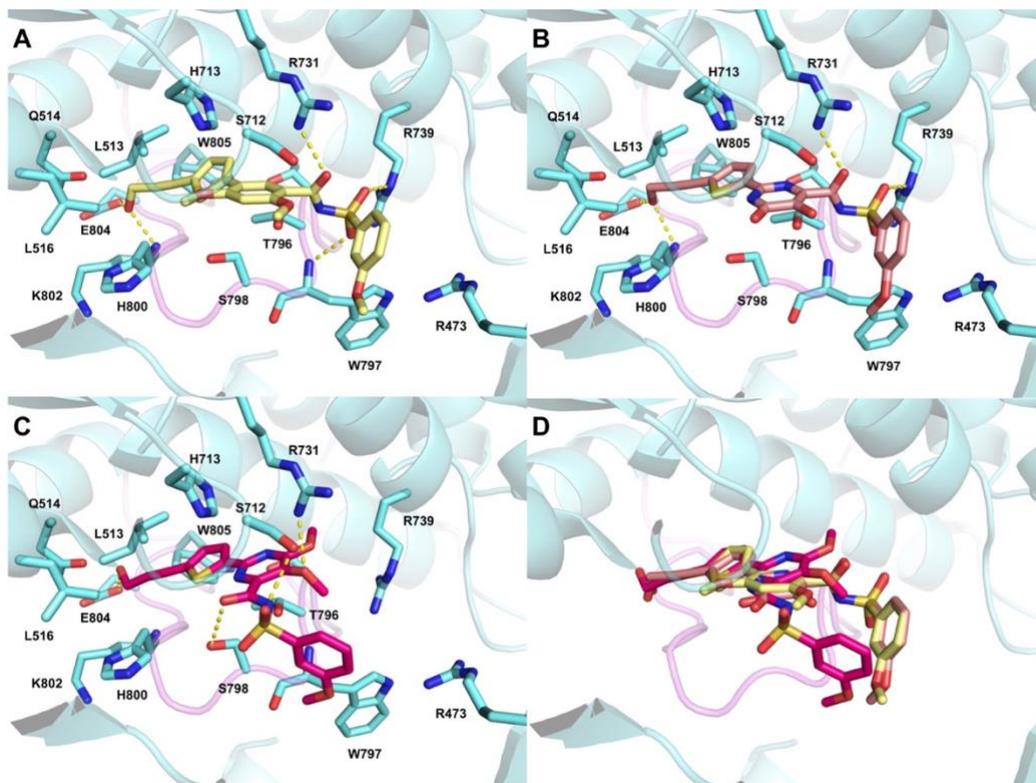


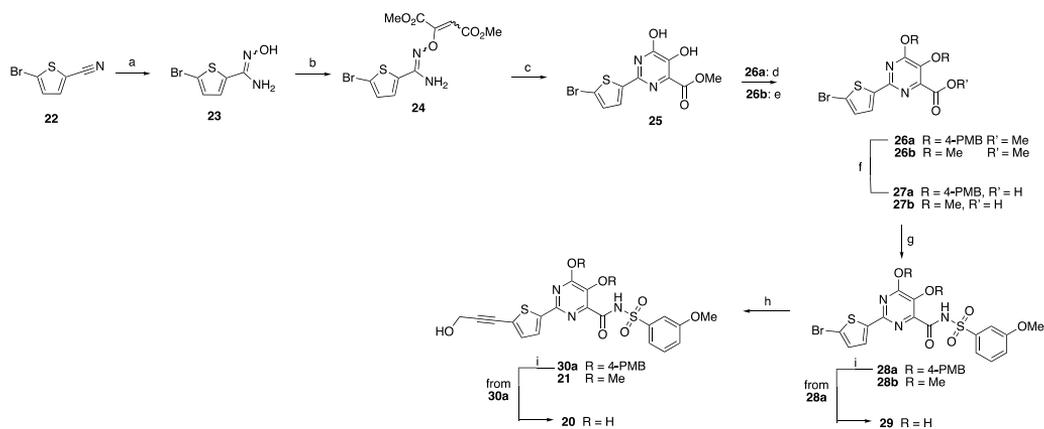
Figure 3.2. Docked poses of compound **19** (A, yellow), **20** (B, salmon) and **21** (C, hotpink) into the N-pocket of ZIKV RdRp (PDB 6LD3, cyan). The key residues are displayed as cyan sticks and labeled. H-bonds are depicted as dashed yellow lines. (D) Overlay of compound **19**, **20** and **21** docked poses.

3.1.2. Synthesis of new dihydroxypyrimidine derivatives

In order to validate the design, a linear multistep procedure was applied to obtain the compounds **20** and **21**, as reported in Scheme 1. The commercial 5-bromothiophene-2-carbonitrile **22** was converted into the corresponding amidoxime derivative **23**, which was reacted with dimethylacetylene dicarboxylate (DMAD) to yield adduct **24** as a mixture of E/Z isomers.³ This adduct was not isolated and directly cyclized in xylene at reflux to give the dihydroxypyrimidine methyl ester **25** in a discreet yield of 44%.⁴ The hydroxylic groups of key intermediate **25** were both protected by 4-methoxybenzyloxy

groups to give **26a**, by using 4-methoxybenzyl chloride in presence of potassium carbonate as base and catalytic amount of tetrabutylammonium bromide in DMF, or methylated into **26b**, by reaction with methyl iodide and caesium carbonate in DMF.

Scheme 1. Synthetic procedure for compounds **20** and **21^a**



^aReaction and conditions: a) $\text{NH}_2\text{OH}\cdot\text{Cl}$, Na_2CO_3 , EtOH, H_2O , 24 h, 91%; b) DMAD, Et_3N , CH_2Cl_2 , at reflux for 3 h, quant.; c) xylene, 140°C , 3 h, 44%; d) 4-methoxybenzyl chloride, K_2CO_3 , $\text{Bu}_4\text{N}^+\text{Br}^-$, DMF, 60°C , 16 h, 67%; e) CH_3I , Cs_2CO_3 , DMF, 5 h, 79%; f) NaOH , H_2O , THF, 2 h, quant.; g) 3-methoxybenzenesulfonamide, EDCI, HOBT, DMAP, dry DMF, 16 h, 43-63%; h) Propargyl alcohol, $\text{PdCl}_2(\text{dppf})\text{CH}_2\text{Cl}_2$, CuI , Et_3N , dioxane, 80°C , 2 h, 58-83%; i) TFA, CH_2Cl_2 , 2 h, 84-99%.

The methyl esters **26a,b**, obtained respectively in 67% and 79% yield, were quantitatively hydrolysed, with an excess of sodium hydroxide, to carboxylic acid derivatives **27a,b**, which were coupled to the 3-methoxybenzenesulfonamide, using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) hydrochloride as a coupling agent in presence of hydroxybenzotriazole (HOBT) and 4-dimethylaminopyridine (DMAP) as base, to afford the corresponding acylsulfonamides **28a,b**. Derivatives **28a,b** were reacted with propargyl alcohol under Sonogashira cross-

coupling conditions, using 1,1'-bis(diphenylphosphino)ferrocene]dichloropalladium(II) ($\text{PdCl}_2(\text{dppf})\cdot\text{CH}_2\text{Cl}_2$) and copper(I) iodide as catalytic system and triethylamine as a base in dry DMF at 80°C , in order to provide the target compound **21** and the derivative **30a**, respectively in 83% and 58% yield.¹ Finally, intermediates **28a** and **30a** were deprotected by using trifluoroacetic acid (TFA) in dichloromethane to afford the final dihydroxypyrimidine derivatives **20** and **29**.

3.2. Discovery of antiviral agents against ZIKV through computational-driven polypharmacology approach

In parallel, a polypharmacological strategy was pursued to discover new anti-ZIKV agents, by integrating artificial intelligence with chemical synthesis. The project was supported by the EXaSCale smArt pLatform Against paThogEns (Exscalate),⁶ which involves a large consortium, coordinated by Dompè pharmaceuticals. Exscalate was born from the ANTAREX project (AutoTuning and Adaptivity appRoach for Energy efficient eXascale HPC systems) under funding of EU to fight the ZIKV epidemic in 2015, and represents the first computational platform devoted to the need for promptly response to pandemic crisis. In this context, the Exscalate platform was exploited to perform High Performance Computing (HPC) simulations, with the final aim to select molecules active against the ZIKV. In particular, docking simulations were conducted to virtual screen the Safe In Man (SIM) library, containing commercialized and under development drugs already proved safe in man, against all the ZIKA proteins. The SIM library is a dataset of ~10,000 pharmaceutical compounds, corresponding to the list of drugs launched or under active development in several clinical phases, derived from an Integrity database search.⁷ All the compounds have been classified on basis of the mechanism of action, in particular by distinguishing therapeutic classes (antibacterial, antiviral,

antiparasitic, antifungal, antinfective). The simulation was performed using LiGen™ (Ligand Generator), the *de novo* structure-based virtual screening software, designed and developed to run on HPC architectures, which represents the most relevant tool of the Exscalate platform. LiGen™ is formed by a set of tools that can be combined in a user-defined manner to generate project centric workflows. In details, LiGenDock is a docking module, using LiGenScore to compute the scoring function and the LiGenPass and LiGenPocket modules to obtain the 3D structure of the binding site. Both the docking algorithm implemented in LiGen™, the pharmacophoric docking (LiGenPh4) and the geometrical docking (LiGenGeodock), as well as the different scoring functions calculated, were used in this study to explore different protocols of Virtual Screening (VS) and select the best one in terms of performance. Usually, the performance of a VS protocol is assessed by evaluating its capacity to recognize the active molecules within a large number of inactive decoys, where the active molecules represent ~1% of total number of compounds. In the work described herein, the performance of the VS strategy tested was evaluated by assessing its ability to correctly classify molecules with antiviral activity, in particular drugs against the HCV, considering that both HCV and ZIKV belong to *Flaviviridae* family. The crystal structures of the ZIKA proteome were obtained from the Protein Data Bank (PDB). In Table 3.1, the list of ZIKV proteins and identified binding sites was reported, with the corresponding PDB codes. More than one binding site was explored for each protein, including orthosteric and allosteric binding sites, that were identified for each protein and used to perform VS. Overall, 26 sites were used, with more than one crystal structure for each protein, selected among those available in the PDB. An ensemble docking procedure was used to combine the results obtained with the different PDB structures for each of the 13 unique binding sites identified in the ZIKA proteins.

Table 3.1. Selection of 13 binding sites from crystal structures of 5 ZIKV proteins.

Proteins	Sites	PDB code
Protease (NS2B-NS3)	Orthosteric, allosteric	5H6V, 5TFN
Helicase (NS3)	Orthosteric, allosteric, RNA	5VI7, 5Y4Z
Methyltransferase (NS5-MT)	Orthosteric, allosteric, RNA	5VIM, 5WXB, 5U0B
Polymerase (NS5-RdRp)	Orthosteric, allosteric, nucleosides	5U0B, 5U0C
Envelope	Orthosteric	5JHM
NS1	Orthosteric	5K6K

Following this approach, about 110 molecules were labelled as active compounds. The VS protocol performance was thus evaluated comparing the screening results of the SIM library on the ZIKA crystal structures, by using different docking algorithms and different scoring function. The conditions showing the best capacity to recognize the active molecules were used to prioritize the total list of screened molecules and select the top scored compounds, according to the score value that predicts the binding affinity of the molecules in the protein binding site. In particular, the results of the virtual screening on the 13 binding sites of the six viral proteins reported in Table 3.1. were combined, in order to select molecules with a potential polypharmacological effect against ZIKV. To this aim, a total score, corresponding to the sum of the docking scores obtained for each protein, was calculated for each molecule and used to prioritize the most interesting molecules. Approximately 80 top scored molecules, with a total score > 75 , were selected as the most promising multitargeting virtual hits. In collaboration with Prof. Mauro Martins Teixeira of Fluminense Federal University at Niterói, a phenotypic assay against ZIKV infection was carried out as a first *in vitro* screening of the virtual hits selected (for the details of assay, see Paragraph 3.3).

Among molecules tested, eight compounds showed to inhibit ZIKV infection in Vero cells with $EC_{50} < 10 \mu\text{M}$, without cellular toxicity at $30 \mu\text{M}$. The best compound was Raloxifene, a drug already approved for the prevention and treatment of osteoporosis (Fig. 3.3). Thus, Raloxifene has been used as template to generate a huge library of million analogues, exploring a tangible virtual chemical space through Exscalate platform. On this regard, Dompè collected and integrated millions of building blocks from different commercial databases (Mcule, Aldrich Market Select, Molport) in an entire building block database, which was then combined with a set of 80 chemical reactions (Fig. 3.4). In particular, Dompè selected an initial pool of 58 robust organic reactions used in drug discovery and collected in literature.⁸ Interestingly, the 50% of these are ring-forming reactions, a feature that increases the structural diversity of the chemical space, as ring closure allows to build new scaffolds. Further, Dompè collected our contribution, together with other synthetic chemists, for evaluating new reactions to enrich the chemical reaction database and for the annotation of each reaction on the basis of their synthetic knowledge and expertise. The reactions are written in the smirks format, which use the smarts language to encode the transforms needed to perform the reaction. The smarts language allows to specify substructures using specific rules that account for not only atoms and bonds but also the chemical environment of each atom, in a very detailed way.⁹ This detailed description allows to include in the reagent substructures definition all the chemical information collected about each reaction, taking it into account in the synthetic feasibility prediction (Fig. 3.4). Thus, by combining existing starting material and real chemical synthesis, a tangible virtual library was generated from Raloxifene, in order to identify novel chemical moieties through scaffold analysis at different abstraction levels (Fig. 3.3).

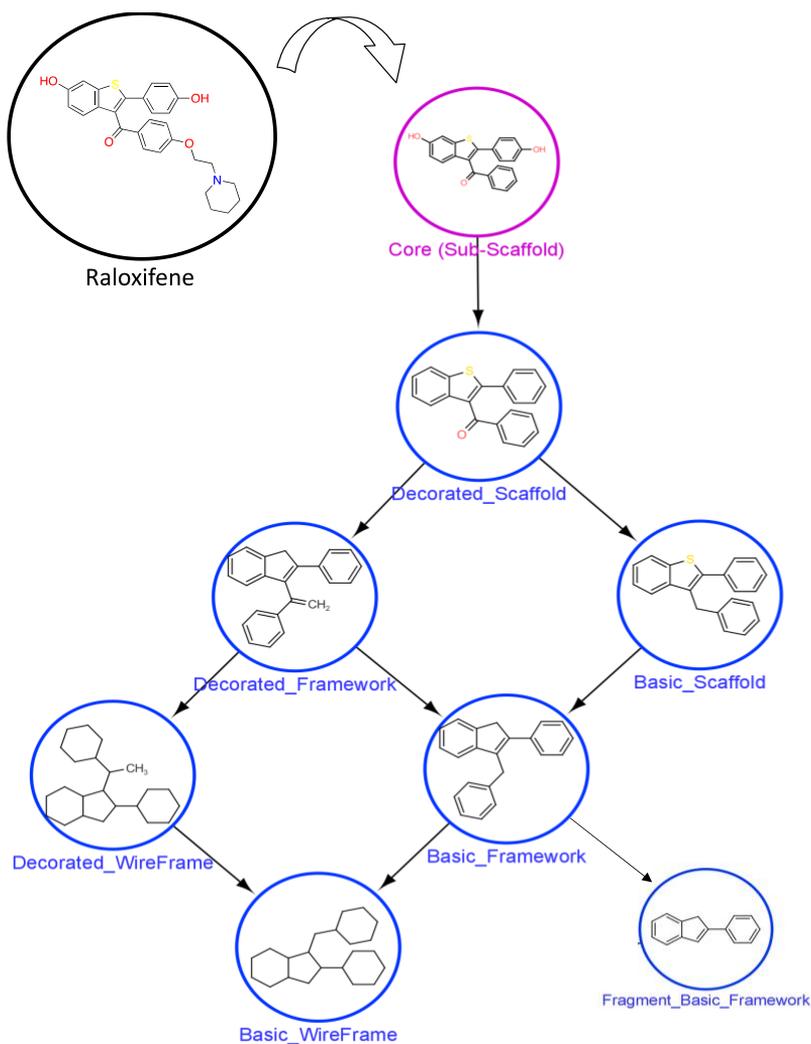


Figure 3.3. Generation of a tangible virtual library by Raloxifene scaffold hopping, based on a molecular representation at different abstraction levels. (Adapted from a figure courtesy of Dompè pharmaceuticals).

Dompè defined a novel and flexible approach, namely “Molecular Anatomy”, able to cluster large set of compounds starting from a molecular scaffold. This approach is based on hypothesis that it is possible to identify different levels of abstraction and rules for representing a scaffold.¹⁰ In particular, *wireframe* is the most abstracted representation, obtained removing both bond and atom type from

a scaffold, while *framework* retains only bond. As depicted in Fig. 3.3, starting from the Raloxifene decorated scaffold, which maintains terminal atoms with bond order greater than one, Dompè defined different molecular representations following two set of rules. The first one was based on an increased level of structural information respect to the basic scaffold, generated by removing all side chains and terminal atoms. The second one increased chemical abstraction by removing the atom type label and then the bond order, generating, respectively, a *framework* and a *wireframe* for each level of the scaffold (basic and decorated). Finally, seven molecular representations were produced within a hierarchical correlation (Fig. 3.3)

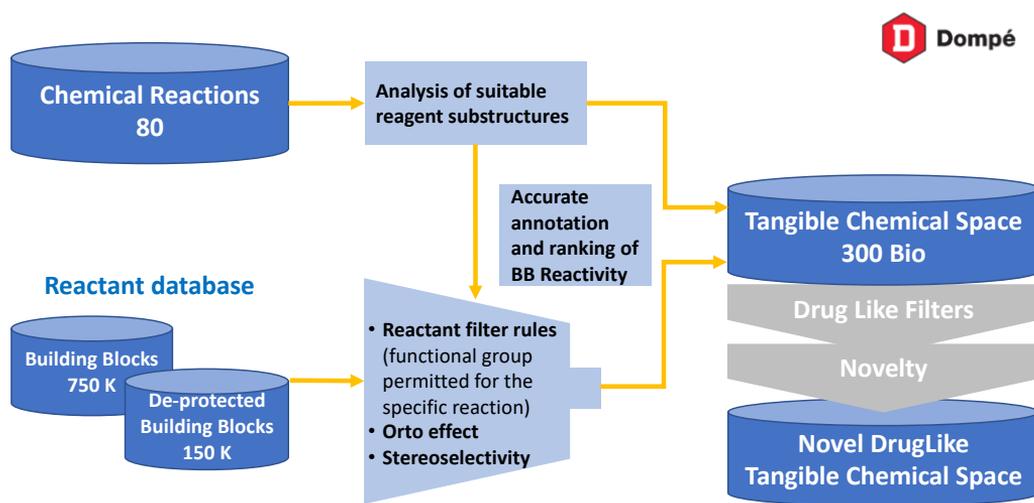


Figure 3.4. Novel tangible drug-like chemical space by Exscalate platform. Generation of a virtual library of hundred billion compounds by using real reagents and robust organic synthesis reactions to explore novel drug-like space, achievable with one reaction step. (Figure courtesy of Dompè pharmaceuticals).

In second instance, drug-like properties (i.e. molecular weight, pKa, logP) and novelty (excluding clinical candidates and approved drugs) filters were applied to build a tangible chemical space database, in order to generate a novel drug-

like tangible chemical space (Fig. 3.4), suitable for following docking calculations. Indeed, an ultra-fast molecular docking was performed to evaluate the interaction with ZIKV proteins, as described above. A total score was calculated for each member of library and used to identify the most interesting molecules on basis of polypharmacological profile. Among the best performing multitargeting virtual hits, we selected an initial set of eight compounds for the chemical synthesis (**31-38**), considering two criteria: i) the docking scores against ZIKV proteins; ii) the synthetic feasibility, in terms of cost of reagents and time for synthesis. It is noteworthy that patentability restriction limited the disclosing of the compound's structure and, consequently, the synthetic routes. Therefore, the general structure of virtual hits selected and synthesized was represented in Fig. 3.5.

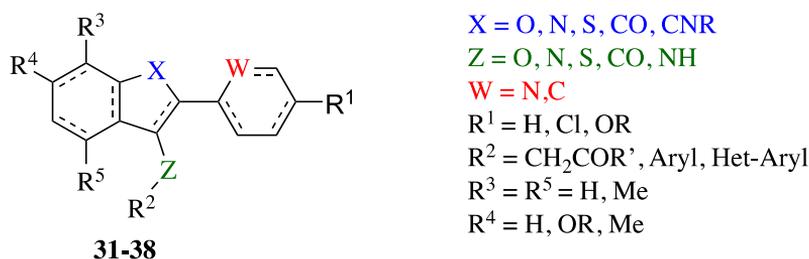


Figure 3.5. General structure of virtual hits **31-38** selected through ZIKV multitargeting virtual screening by Exscalate software.

Generally, I carried out multistep synthesis and multicomponent reactions to obtain the new molecules. Reliable synthetic procedures, based on one or two steps, were applied to form the various scaffolds and modify the different accessory groups, such as Groebke-Blackburn-Bienaymè reaction, Friedel-crafts acylation, Pd-catalyzed cross-coupling reaction, nucleophilic substitutions and ester hydrolysis, just to cite a few. Notably, these reactions are in line with the principles of tangible chemical space explored by Exscalate platform.

3.3. Biological evaluation against ZIKV infection in cell culture

The compounds **20**, **21**, **28b**, **29** and **31-38** were submitted to the biological evaluation, carried out by the Prof. Mauro Martins Teixeira of Fluminense Federal University at Niterói, Brazil. A phenotypic assay was performed to determine the potential antiviral activity of compounds against ZIKV.

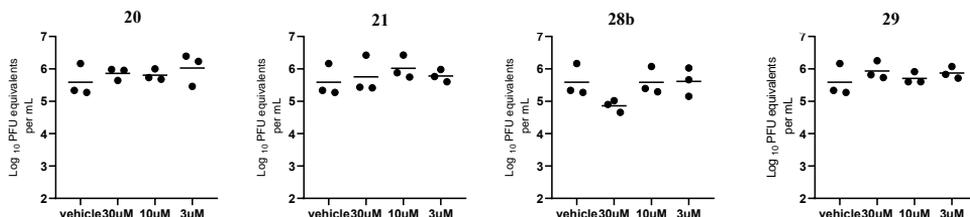
ZIKV-infected Vero cells were exposed to three concentrations compound (30 μ M, 10 μ M and 3 μ M) following the ZIKV infection at MOI = 0.01. The viral quantification was analyzed by plaque assay, while the cellular viability was evaluated by (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide) (MTT) and lactate dehydrogenase (LDH) assays.

Awaiting enzyme assays, the dihydroxypyrimidine **20**, **29** and dimethoxypyrimidine **21**, **28b** analogues were first evaluated in cell-based assay (Fig. 3.6). Unfortunately, the compounds have not shown a relevant antiviral activity, with EC₅₀ values > 30 μ M or < 10 μ M only for **21**, although they were not toxic for cell (CC₅₀ MTT > 30 μ M, CC₅₀ LDH > 30 μ M).

However, while the antiviral activity results were still pending, Gharbhi Ayachi *et. al.* reported a novel series of NNIs of ZIKA polymerase targeting N-pocket allosteric site.¹¹ High-resolution crystallographic studies were performed, allowing a comparison of the N pocket in the RdRp between ZIKV and DENV3. Interestingly, the high-resolution free ZIKV RdRp structure revealed the volume of N-pocket of ZIKV RdRp is 2 to 3 times smaller than DENV RdRp N-pocket. This is due to the substitution of Met⁷⁶⁵ with Leu⁷⁶⁷: the presence of a branched methyl group in the ZIKV RdRp pocket leads to a reduction in the overall volume of the ZIKV RdRp N pocket, having an impact on ligand binding. The Met-to-Leu substitution results in a shrinkage of the channel, providing steric hindrance with several substitutions introduced in the thiophene moiety. Indeed, a DENV-3 NNI, similar to compound **19**, inhibited the ZIKV NS5 polymerase activity

with a potency ~2,000-fold lower than that for DENV NS5 polymerase activity,¹ highlighting how the differences in the size and shape of N-pocket are crucial for antiviral activity of inhibitors.

A



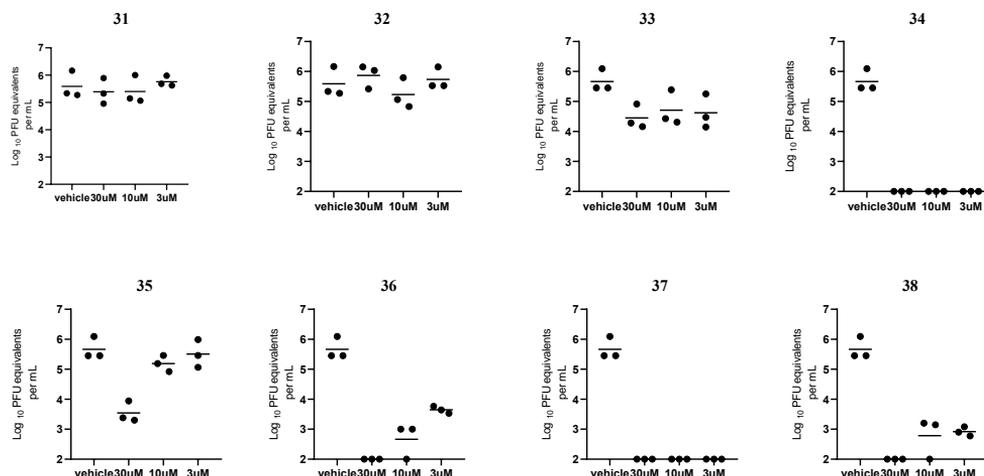
B

Cpd	EC ₅₀ (μM) ^a ZIKV	CC ₅₀ (μM) ^b Vero E6	CC ₅₀ (μM) ^c Vero E6
20	> 30	> 30	> 30
21	30-10	> 30	> 30
28b	> 30	> 30	> 30
29	> 30	> 30	> 30

Figure 3.6. Biological evaluation of **20**, **21**, **28b** and **29** against ZIKV infection in cell culture. (A) Viral loads (plaque assay) of compounds **20**, **21**, **28b** and **29**, testing each compound in three concentrations: 30 μM, 10 μM and 3 μM. Viral load was expressed by unit forming plaques (PFU) per mL. (B) Table 3.2. Biological activities of target compounds **20**, **21**, **28b** and **29**. ^aCompound concentration required to reduce the ZIKV infection in Vero E6 cell by 50%. ^bCompound concentration required to reduce Vero E6 cell viability by 50% by MTT assay. ^cCompound concentration required to reduce Vero E6 cell viability by 50% by LDH assay.

On basis of this evidence, and considering the not promising antiviral activity results, the designed compounds were not subjected to further biological assays. Similar to the designed NNIs, the ZIKV multitargeting virtual hits **31-38** were evaluated in cell-based assay (Fig. 3.7).

A



B

Cpd	EC ₅₀ (μM) ^a ZIKV	CC ₅₀ (μM) ^b Vero E6	CC ₅₀ (μM) ^c Vero E6
Raloxifene	10 - 3	30 - 10	30 - 10
31	> 30	> 30	> 30
32	10	> 30	> 30
33	< 3	> 30	> 30
34	< 3	> 30	> 30
35	10 - 3	> 30	> 30
36	< 3	> 30	> 30
37	< 3	> 30	> 30
38	< 3	> 30	> 30

Figure 3.7. Biological evaluation compounds **31-38** and Raloxifene as reference compound against ZIKV infection in cell culture. (A) Viral loads (plaque assay) of compounds **31-38**, testing each compound in three concentrations: 30 μM, 10 μM and 3 μM. Viral load was expressed by unit forming plaques (PFU) per mL. (B) Table 3.3. Biological activities of target compounds **31-38** and reference compound Raloxifene. ^aCompound concentration required to reduce the ZIKV infection in Vero E6 cell by 50%. ^bCompound concentration required to reduce Vero E6 cell viability by 50% by MTT assay. ^cCompound concentration required to reduce Vero E6 cell viability by 50% by LDH assay.

Along the series, some compounds (**33**, **34**, **36-38**) showed a relevant activity against ZIKV replicating in cell lines, with value of EC₅₀ equal or below 3 μM and no significant host cell toxicity in MTT and LDH assays (Fig. 3.7). In particular, compounds **34** and **37** were able to totally block viral replication, thus suggesting a very potent antiviral activity, with EC₅₀ value most likely in the nanomolar range. Therefore, cell-based assays at lower concentrations are ongoing to evaluate the specific EC₅₀ value of **34** and **37**.

3.4. Summary and conclusions

The absence of vaccines and therapeutic treatments for ZIKV infection poses a significant medical need of effective antiviral agents. On this regard, this part of project was focused on the discovery of new small molecules active against ZIKV, by integrating rational design and parallelly artificial intelligence with chemical synthesis.

Rational design aimed to discover NNIs ZIKV RdRp was based on dihydroxypyrimidine scaffold hopping of potent NNIs DENV3 polymerase **19**. This approach has not identified promising agents. Indeed, all compounds are inactive in cell, although without a detectable toxicity for the cellular viability. In meantime, the study conducted by Gharbhi Ayachi *et. al* revealed structural differences of N-pocket between ZIKV NS5 polymerase and DENV3 homologue enzyme, which influences on binding mode and potency of inhibitors. Indeed, the presence of propargyl alcohol on the thiophene moiety is not tolerated into cavity of N-pocket allosteric site of ZIKV polymerase. In view of this and the antiviral activity resulted in phenotypic assay, compounds **20** and **21** are uninteresting for further biological investigations.

Differently, the ZIKV multitargeting approach through the innovative platform of Exscalate allowed to identify new potent antiviral agents, within tangible chemical space generated around repurposed Raloxifene. Among the polypharmacological virtual hits selected for the synthesis (**31-38**) and evaluated in phenotypic assay, the most potent derivatives **34, 37** showed a potent antiviral activity in cell-based plaque assay, with EC₅₀ value in a nanomolar range, without a significant cellular toxicity.

Based on the potency in the phenotypic screening, **34, 37** are currently under evaluation in the mouse *in vivo* model of ZIKV infection, in order to evaluate parameters such as lethality and clinical signs, blood test, viral load and inflammation markers in the brain and neurodegeneration. In addition, biochemical and biophysical assays are planned to evaluate ability of compound to inhibit and bind the putative targets, thus providing insight in the mode of actions and apply a structure-based approach exploiting docking pose(s) as possible binding mode(s).

3.5. Experimental section

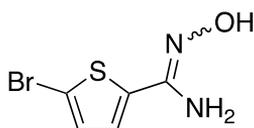
3.5.1. Synthesis and characterization of compounds

General methods. Reagents and solvents were purchased from commercial sources and used without further purification. Reactions were carried out at ambient temperature, unless otherwise specified. Moisture-sensitive reactions were performed under a positive pressure of dry nitrogen in oven-dried glassware. Analytical thin-layer chromatography (TLC) on silica gel 60 F254 plates (250 μm thickness) was performed to monitor the reaction progress, using UV as revelation method. Flash chromatography on silica gel (70—230 mesh)

was performed for purification. All products were characterized by their NMR and MS spectra. (ESI)-MS spectra were performed on a LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) by infusion into the ESI source using MeOH as solvent. ^1H and ^{13}C NMR spectra were recorded in DMSO- d_6 at 25°C on Bruker Avance NEO 400 MHz instruments equipped with a RT-DR-BF/ ^1H -5mm-OZ SmartProbe. Chemical shifts (δ) are reported in part per million (ppm) downfield from tetramethylsilane, using residual solvent signal as the internal reference.

Synthesis of target compounds and intermediates

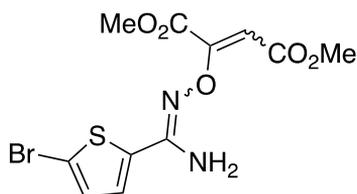
5-bromo-N'-hydroxythiophene-2-carboximidamide (23).



Chemical Formula: $\text{C}_5\text{H}_5\text{BrN}_2\text{OS}$
Molecular Weight: 221.07

2-Bromocyanothiophene (1.77 mL, 15.953 mmol) was suspended in mixture water/ethanol (ratio 7:1, 0.2 M). Sodium carbonate and hydroxylamine were added, and the mixture was left stirring at room temperature for 24 h. The white suspension was filtered under vacuum and the solid filtered was dried under vacuum to obtain the pure product **23** (3.722 g, 91%) as a white solid. The product was obtained as a mixture E/Z and used for the next step without purification. ^1H NMR (400 MHz, DMSO): δ 7.54 (d, $J = 4.1$ Hz, 1H), 7.30 (d, $J = 4.1$ Hz, 1H), 5.88 (s, 2H). MS (ESI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_5\text{H}_6\text{BrN}_2\text{OS}^+$ 222.08 (Br^{79}), 224.08 (Br^{81}), found 222.0 (Br^{79}), 224.1 (Br^{81}).

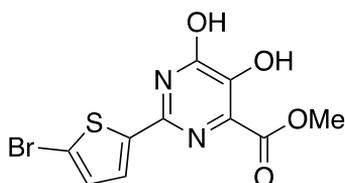
Dimethyl 2-(((amino(5-bromothiophen-2-yl)methylene)amino)oxy)but-2-enedioate (**24**).



Chemical Formula: C₁₁H₁₁BrN₂O₅S
Molecular Weight: 363.18

To a suspension of 5-bromo-N'-hydroxythiophene-2-carboximidamide **23** (2.802 g, 12.738 mmol) in dichloromethane (0.3 M), triethylamine (0.100 mL, 0.713 mmol) and dimethyl acetylenedicarboxylate (1.7 mL, 14.011 mmol) were added, and the resulted mixture was left stirring at reflux for 3 h. The reaction mixture was diluted with distilled water (20 mL) and extracted with dichloromethane (20 mL x 3). The collected organic layers were washed with brine (15 mL x 1), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum to give the clean product **24** (5.285 g, 100%) as a reddish oil. The compound was used without further purification. ¹H NMR (400 MHz, DMSO): δ 7.54 (d, *J* = 4.1 Hz, 1H), 7.18 (d, *J* = 4.1 Hz, 1H), 6.55 (s, 1H), 5.68 (s, 2H), 3.82 (s, 3H), 3.70 (s, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₁₁H₂₁BrN₂O₅S⁺ 364.19 (Br⁷⁹), 366.19 (Br⁸¹), found 364.2 (Br⁷⁹), 366.2 (Br⁸¹).

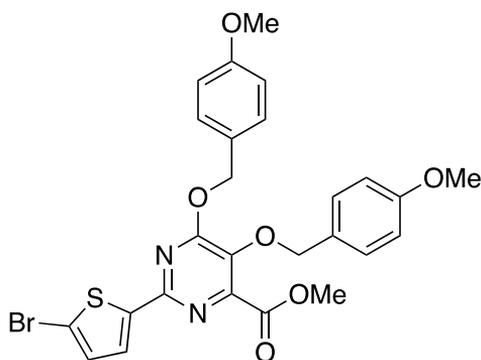
Methyl 2-(5-bromothiophen-2-yl)-5,6-dihydropyrimidine-4-carboxylate (**25**).



Chemical Formula: C₁₀H₇BrN₂O₄S
Molecular Weight: 331.14

To a flask containing dimethyl 2-(((5-bromo-2-thiophenyl)methylene)amino)oxy)but-2-enedioate **24** (1.124 g, 3.095 mmol) was added xylene (0.4 M), and the reaction was left at 140 °C for 3 h until the disappearance of the starting material. The reaction mixture was stored in a refrigerator overnight and the precipitate isolated by filtration. The brown solid was washed with ethyl acetate (20 mL x 1) and petroleum ether (20 mL x 1), and the collected organic layers were washed with brine (10 mL x 1), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum to give the clean product **25** (450 mg, 44%) as a brown solid. ¹H NMR (400 MHz, DMSO): δ 7.78 (d, *J* = 4.1 Hz, 1H), 7.30 (d, *J* = 4.1 Hz, 1H), 3.83 (s, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₁₀H₈BrN₂O₄S⁺ 332.15 (Br⁷⁹), 334.15 (Br⁸¹), found 332.2 (Br⁷⁹), 334.2 (Br⁸¹).

Methyl 2-(5-bromothiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)pyrimidine-4-carboxylate (26a).

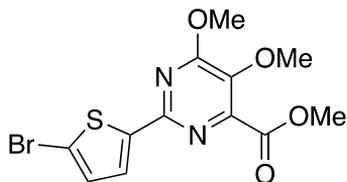


Chemical Formula: C₂₆H₂₃BrN₂O₆S
Molecular Weight: 571.44

To a solution of methyl 2-(5-bromothiophen-2-yl)-5,6-dihydroxypyrimidine-4-carboxylate **25** (686 mg, 2.071 mmol) in anhydrous DMF (0.04 M) in a three-necked oven-dried flask, potassium carbonate (859 mg, 6.213 mmol) and tetrabutylammonium bromide (334 mg, 1.036 mmol) were added, following 1-chloro-4-methoxybenzene (0.842 mL, 6.213 mmol) was added. The reaction mixture was stirred at room temperature for 16 h, then it was diluted with

ice–water (7 mL) and a solution of 2N HCl (15 mL) was added dropwise at 0°C until pH = 5. The mixture was extracted with EtOAc (2 x 20 mL), then the organic phase was washed with a saturated NaHCO₃ solution (2 x 10 mL) and brine (8 x 5 mL). The collected organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum to give the product **26a** (337 mg, 67%) as a brownish solid. ¹H NMR (400 MHz, DMSO): δ 7.75 (d, *J* = 4.0 Hz, 1H), 7.57 (d, *J* = 8.6 Hz, 2H), 7.37 (d, *J* = 4.0 Hz, 1H), 7.23 (d, *J* = 8.6 Hz, 2H), 7.03 (d, *J* = 8.7 Hz, 2H), 6.90 (d, *J* = 8.6 Hz, 2H), 5.59 (s, 2H), 5.02 (s, 2H), 3.86 (s, 3H), 3.80 (s, 3H), 3.77 (s, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₆H₂₄BrN₂O₆S⁺ 572.45 (Br⁷⁹), 574.45 (Br⁸¹), found 572.4 (Br⁷⁹), 574.4 (Br⁸¹).

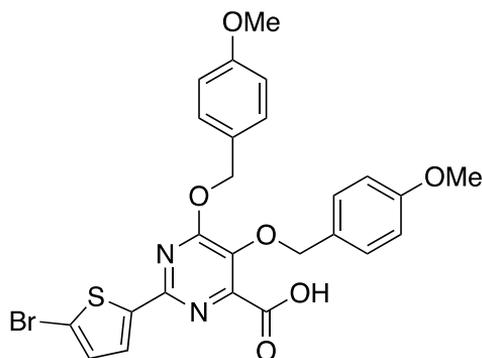
Methyl 2-(5-bromothiophen-2-yl)-5,6-dimethoxypyrimidine-4-carboxylate (**26b**).



Chemical Formula: C₁₂H₁₁BrN₂O₄S
Molecular Weight: 359.19

To a solution of Methyl 2-(5-bromothiophen-2-yl)-5,6-dihydroxypyrimidine-4-carboxylate **25** (500 mg, 1.510 mmol) in anhydrous DMF (0.3 M), caesium carbonate (1.5 g, 4.530 mmol) and iodomethane (0.282 mL, 4.530 mmol) were added and the reaction mixture was left stirring at room temperature for 5 h. The mixture was diluted with distilled water (6 mL) and the precipitated was filtered under vacuum to give the product **26b** (430 mg, 79%) as a white solid. ¹H NMR (400 MHz, DMSO): δ 7.68 (d, *J* = 4.0 Hz, 1H), 7.33 (d, *J* = 4.0 Hz, 1H), 4.10 (s, 3H), 3.92 (s, 3H), 3.84 (s, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₁₂H₁₂BrN₂O₄S⁺ 360.20 (Br⁷⁹), 362.20 (Br⁸¹), found 360.2 (Br⁷⁹), 362.2 (Br⁸¹).

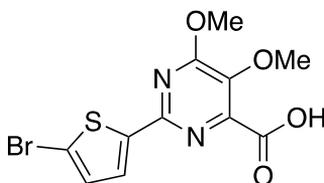
2-(5-bromothiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)pyrimidine-4-carboxylic acid (**27a**).



Chemical Formula: C₂₅H₂₁BrN₂O₆S
Molecular Weight: 557.42

To a solution of methyl 2-(5-bromothiophen-2-yl)-5,6-dimethoxypyrimidine-4-carboxylate **26a** (360 mg, 0.630 mmol) in THF (0.1 M), a solution of 0.4 N NaOH (9 mL, 3.150 mmol) was added dropwise at 0°C and the reaction left stirring at room temperature for 2h. The reaction mixture was diluted with ice-water (5 mL) and a solution of 2N HCl was added dropwise until pH = 5. The mixture was extracted with EtOAc (3 x 15 mL), and the collected organic layers were washed with brine (1 x 10 mL), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum to give the product **27a** (337 mg, 96%) as a brownish solid. ¹H NMR (400 MHz, DMSO): δ 7.75 (d, *J* = 4.0 Hz, 1H), 7.57 (d, *J* = 8.6 Hz, 2H), 7.37 (d, *J* = 4.0 Hz, 1H), 7.23 (d, *J* = 8.6 Hz, 2H), 7.03 (d, *J* = 8.7 Hz, 2H), 6.90 (d, *J* = 8.6 Hz, 2H), 5.59 (s, 2H), 5.02 (s, 2H), 3.86 (s, 3H), 3.80 (s, 3H). MS (ESI) *m/z*: [M-H]⁻ calcd for C₂₅H₂₀BrN₂O₆S⁻ 556.41 (Br⁷⁹), 558.41 (Br⁸¹), found 556.4 (Br⁷⁹), 558.4 (Br⁸¹).

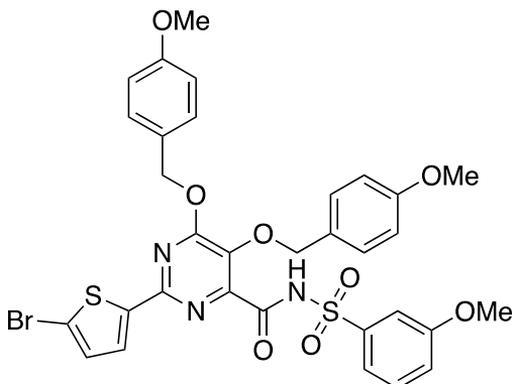
2-(5-bromothiophen-2-yl)-5,6-dimethoxypyrimidine-4-carboxylic acid (**27b**).



Chemical Formula: C₁₁H₉BrN₂O₄S
Molecular Weight: 345.17

To a solution of methyl 2-(5-bromothiophen-2-yl)-5,6-dimethoxypyrimidine-4-carboxylate **26b** (420 mg, 1.169 mmol) was dissolved in THF (0.1 M), a solution of 0.4 N NaOH (17 mL, 5.846 mmol) was added dropwise at 0°C and the reaction left stirring at room temperature for 2h. The reaction mixture was diluted with ice-water (5 mL) and a solution of 2N HCl was added dropwise until a pH = 5. The solid that precipitated was filtered under vacuum to give the product **27b** (383 mg, 95%) as a light brown solid. ¹H NMR (400 MHz, DMSO): δ 7.68 (d, *J* = 4.0 Hz, 1H), 7.33 (d, *J* = 4.0 Hz, 1H), 4.10 (s, 3H), 3.92 (s, 3H). MS (ESI) m/z: [M-H]⁻ calcd for C₁₁H₈BrN₂O₄S⁻ 344.16 (Br⁷⁹), 346.16 (Br⁸¹), found 344.2 (Br⁷⁹), 346.2 (Br⁸¹).

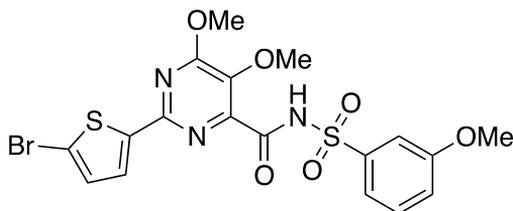
2-(5-bromothiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)-N-((3-methoxyphenyl)sulfonyl)pyrimidine -4-carboxamide (**28a**).



Chemical Formula: C₃₂H₂₈BrN₃O₈S₂
Molecular Weight: 726.61

To a solution of 2-(5-bromothiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)pyrimidine-4-carboxylic acid **27a** (120 mg, 0.215 mmol) in anhydrous CH₂Cl₂ (0.1 M) in a three-necked oven-dried flask, EDCI (62 mg, 0.323 mmol) and DMAP (79 mg, 0.645 mmol) were added under a positive anhydrous nitrogen pressure. After 10 minutes, 3-methoxybenzenesulfonamide (59 mg, 0.323 mmol) was added and the reaction mixture was left stirring at room temperature for 16 h. The reaction mixture was diluted with CH₂Cl₂ (10 mL) and washed with a solution of 2N HCl (2 x 15 mL), a saturated Na₂CO₃ solution (2 x 15 mL) and brine (1 x 10 mL). The collected organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by column chromatography (CH₂Cl₂/MeOH, 99:1 v/v) to afford the pure product **28a** (98 mg, 63%) as a white solid. ¹H NMR (400 MHz, DMSO): δ 7.79 (d, *J* = 4.1 Hz, 1H), 7.59 – 7.54 (m, 4H), 7.49 (d, *J* = 1.5 Hz, 1H), 7.31 (d, *J* = 4.1 Hz, 1H), 7.30 – 7.26 (m, 1H), 7.23 (d, *J* = 8.6 Hz, 2H), 7.03 (d, *J* = 8.7 Hz, 2H), 6.90 (d, *J* = 8.6 Hz, 2H), 5.59 (s, 2H), 5.02 (s, 2H), 3.86 (s, 3H), 3.84 (s, 3H), 3.80 (s, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₃₂H₂₉BrN₃O₈S₂⁺ 727.62 (Br⁷⁹), 729.62 (Br⁸¹), found 727.6 (Br⁷⁹), 729.6 (Br⁸¹).

2-(5-bromothiophen-2-yl)-5,6-dimethoxy-*N*-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide (**28b**).

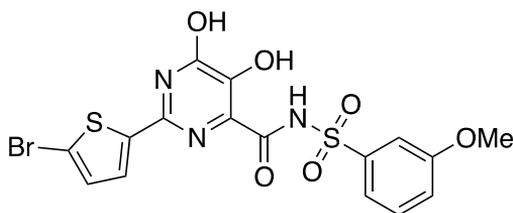


Chemical Formula: C₁₈H₁₆BrN₃O₆S₂
Molecular Weight: 514.37

To a solution of 2-(5-bromothiophen-2-yl)-5,6-dimethoxypyrimidine-4-carboxylic acid **27b** (250 mg, 0.724 mmol) in anhydrous DMF (0.2 M) in a three-

necked oven-dried flask, EDCI (208 mg, 1.086 mmol) and DMAP (265 mg, 2.172 mmol) were added under a positive anhydrous nitrogen pressure. After 10 minutes, 3-methoxybenzenesulfonamide (203 mg, 0.724 mmol) was added and the reaction mixture was left stirring at room temperature for 16 h. The reaction mixture was diluted with EtOAc (10 mL) and washed with a solution of 2N HCl (2 x 15 mL), a saturated Na₂CO₃ solution (2 x 15 mL) and brine (1 x 10 mL). The collected organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by column chromatography (CH₂Cl₂/MeOH, 98:2 v/v) to afford the pure product **28b** (162 mg, 43%) as a white solid. ¹H NMR (400 MHz, DMSO): δ 7.71 (d, J = 3.9 Hz, 1H), 7.58 (dd, J = 14.3, 7.8 Hz, 2H), 7.46 (s, 1H), 7.39 (d, J = 7.7 Hz, 1H), 7.37 – 7.31 (m, 4H), 7.16 (d, J = 7.9 Hz, 1H), 4.06 (s, 3H), 3.85 (s, 3H), 3.82 (s, 3H), 3.68 (s, 3H). ¹³C (400 MHz, DMSO) δ 169.3, 160.9, 159.3, 151.0, 146.6, 141.2, 140.7, 131.1, 130.0, 129.1, 119.6, 117.5, 111.6, 104.1, 55.8, 55.7, 54.5. MS (ESI) m/z: [M+H]⁺ calcd for C₁₈H₁₇BrN₃O₆S₂⁺ 515.37 (Br⁷⁹), 517.37 (Br⁸¹), found 515.4 (Br⁷⁹), 517.4 (Br⁸¹).

2-(5-bromothiophen-2-yl)-5,6-dihydroxy-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide (29).

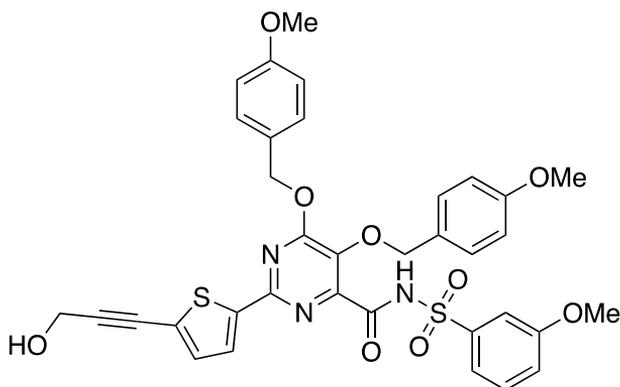


Chemical Formula: C₁₆H₁₂BrN₃O₆S₂
Molecular Weight: 486.31

To a solution of 2-(5-bromothiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide **28a** (165 mg, 0.227 mmol) in CH₂Cl₂ (0.1 M), trifluoroacetic acid (0.75 mL, 9.8 mmol) was added

dropwise at 0°C; then, the reaction was warmed to room temperature and left stirring for 2 h. The reaction solvent was evaporated under vacuum and the crude solid was triturated with diethyl ether to give the product **29** (92 mg, 84%) as a light brown solid. ¹H NMR (400 MHz, DMSO): δ 7.79 (d, *J* = 4.1 Hz, 1H), 7.59 – 7.54 (m, 2H), 7.49 (d, *J* = 1.5 Hz, 1H), 7.31 (d, *J* = 4.1 Hz, 1H), 7.30 – 7.26 (m, 1H), 3.84 (s, 3H). ¹³C (400 MHz, DMSO) δ 169.3, 162.9, 158.9, 146.4, 145.0, 144.8, 144.6, 136.0, 135.7, 131.1, 123.8, 129.1, 111.6, 86.8, 54.2, 51.6. MS (ESI) *m/z*: [M-H]⁻ calcd for C₁₆H₁₁BrN₃O₆S₂⁻ 485.30 (Br⁷⁹), 487.30 (Br⁸¹), found 485.2 (Br⁷⁹), 487.2 (Br⁸¹).

2-(5-(3-hydroxyprop-1-yn-1-yl)thiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide (**30a**).

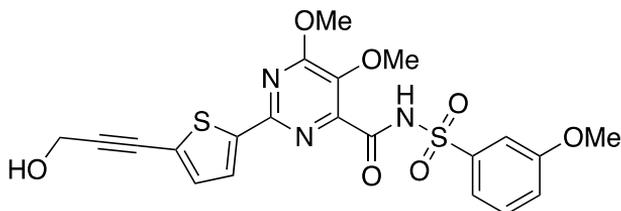


Chemical Formula: C₃₅H₃₁N₃O₉S₂
Molecular Weight: 701.77

In a three-necked oven-dried flask, 2-(5-bromothiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide **28a** (100 mg, 0.138 mmol) was dissolved in anhydrous 1,4-dioxane (0.05 M) under a positive pressure of anhydrous nitrogen. The catalyst PdCl₂(dppf) (10 mg, 0.014 mmol), copper(I) iodide (5 mg, 0.028 mmol) and triethylamine (0.192 mL, 1.380 mmol) were added, and the mixture was purged with anhydrous nitrogen for 5 minutes. The propargyl alcohol (0.010 mL, 0.179 mmol) was

added, and the reaction mixture was warmed to 80°C and stirred for 2 h. After cooling to room temperature, the reaction mixture was filtered over celite and washed with EtOAc (10 mL). The filtrate was washed with distilled water (2 x 10 mL), brine (1 x 5 mL), and the collected organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The crude solid was triturated with diethyl ether to afford the product **30a** (56 mg, 58 %) as a light brown solid. ¹H NMR (400 MHz, DMSO): δ 7.74 (d, J = 2.9 Hz, 1H), 7.48 – 7.40 (m, 4H), 7.30 (dd, J = 8.9, 6.1 Hz, 2H), 7.17 (d, J = 8.4 Hz, 2H), 7.01 (d, J = 8.1 Hz, 1H), 6.95 (d, J = 8.5 Hz, 2H), 6.79 (d, J = 8.5 Hz, 2H), 5.48 – 5.40 (m, 3H), 4.79 (s, 2H), 4.35 (d, J = 5.9 Hz, 2H), 3.75 (s, 3H), 3.74 (s, 3H), 3.71 (s, 3H). MS (ESI) m/z: [M+H]⁺ calcd for C₃₅H₃₂N₃O₉S₂⁺ 702.77, found 702.6.

2-(5-(3-hydroxyprop-1-yn-1-yl)thiophen-2-yl)-5,6-dimethoxy-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide (21).

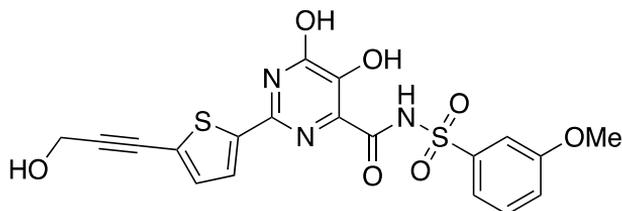


Chemical Formula: C₂₁H₁₉N₃O₇S₂
Molecular Weight: 489.52

In a three-necked oven-dried flask, 2-(5-bromothiophen-2-yl)-5,6-dimethoxy-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide **28b** (100 mg, 0.194 mmol) was dissolved in anhydrous 1,4-dioxane (0.05 M) under a positive pressure of anhydrous nitrogen. The catalyst PdCl₂(dppf) (14 mg, 0.019 mmol), copper(I) iodide (7 mg, 0.039 mmol) and triethylamine (0.270 mL, 1.940 mmol) were added, and the mixture was purged with anhydrous nitrogen for 5 minutes. The propargyl alcohol (0.015 mL, 0.253 mmol) was added, and the reaction mixture was warmed to 80°C and stirred for 2 h. After cooling to room

temperature, the reaction mixture was filtered over celite and washed with EtOAc (10 mL). The filtrate was washed with distilled water (2 x 10 mL), brine (1 x 5 mL), and the collected organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The crude solid was triturated with diethyl ether to afford the product **21** (79 mg, 83%) as a white solid. ¹H NMR (400 MHz, DMSO): δ 7.71 (d, *J* = 3.9 Hz, 1H), 7.58 (dd, *J* = 14.3, 7.8 Hz, 2H), 7.46 (s, 1H), 7.39 (d, *J* = 7.7 Hz, 1H), 7.37 – 7.31 (m, 4H), 7.16 (d, *J* = 7.9 Hz, 1H), 5.19 (t, *J* = 5.9 Hz, 1H), 4.27 (d, *J* = 5.9 Hz, 2H), 4.06 (s, 3H), 3.85 (s, 3H), 3.82 (s, 3H). ¹³C (400 MHz, DMSO): δ 169.3, 160.9, 159.3, 151.0, 144.1, 143.8, 141.2, 140.7, 132.8, 130.0, 127.6, 121.1, 119.6, 117.5, 111.2, 88.9, 80.0, 55.8, 55.7, 54.4, 50.7. MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₁H₂₀N₃O₇S₂⁺ 490.53, found 490.5.

5,6-dihydroxy-2-(5-(3-hydroxyprop-1-yn-1-yl)thiophen-2-yl)-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide (20).



Chemical Formula: C₁₉H₁₅N₃O₇S₂
Molecular Weight: 461.46

To a solution of 2-(5-(3-hydroxyprop-1-yn-1-yl)thiophen-2-yl)-5,6-bis((4-methoxybenzyl)oxy)-N-((3-methoxyphenyl)sulfonyl)pyrimidine-4-carboxamide **30a** (50 mg, 0.071 mmol) in CH₂Cl₂ (0.05 M), trifluoroacetic acid (0.5 mL, 6.5 mmol) was added dropwise at 0°C; then, the reaction was warmed to room temperature and left stirring for 2 h. The reaction solvent was evaporated under vacuum and the crude solid was triturated with diethyl ether to give the product **20** (32 mg, 99%) as a brown solid. ¹H NMR (400 MHz, DMSO): δ 7.73 (d, *J* = 3.9 Hz, 1H), 7.59 (dd, *J* = 14.3, 7.8 Hz, 2H), 7.42 (s, 1H), 7.40 (d, *J* = 7.7 Hz,

1H), 7.36 – 7.30 (m, 4H), 7.17 (d, $J = 7.9$ Hz, 1H), 5.19 (t, $J = 5.9$ Hz, 1H), 4.27 (d, $J = 5.9$ Hz, 2H), 3.78 (s, 3H). ^{13}C (400 MHz, DMSO): δ 169.3, 162.9, 159.3, 145.0, 144.8, 144.6, 143.8, 140.7, 130.0, 129.0, 128.6, 119.6, 117.5, 111.2, 88.9, 80.0, 55.8, 50.7. MS (ESI) m/z : $[\text{M}-\text{H}]^-$ calcd for $\text{C}_{19}\text{H}_{14}\text{N}_3\text{O}_7\text{S}_2^-$ 460.46, found 460.4.

3.5.2. Molecular docking

All compounds were converted to 3D and prepared with Schrödinger's LigPrep tool. This process generated multiple states for stereoisomers, tautomers, ring conformations (1 stable ring conformer by default) and protonation states. In particular, another Schrödinger package, Epik, was used to assign tautomers and protonation states that would be dominant at a selected pH range ($\text{pH}=7\pm 1$). Ambiguous chiral centers were enumerated, allowing a maximum of 32 isomers to be produced from each input structure. Then, an energy minimization was performed with the OPLS3 force. The protein structure was prepared using Maestro Protein Preparation Wizard. Hydrogen atoms were added, and water molecules were removed from the protein structure.

3.5.3. Biological study

Cells and Virus. African green monkey (Vero CCL81 – BCRJ 0245) cells were cultured in RPMI 1640 medium (CultiLab, São Paulo), and SH-SY5Y (BCRJ 0223) cells were grown in DMEM/F-12 medium supplemented with non-essential amino acids (1%), sodium pyruvate (1 mM), L-glutamine (2 mM), both supplemented with 10% FBS and maintained at 37 °C with 5% of CO_2 . C6/36 cells (BCRJ 0343) was grown in Leibovitz L-15 medium containing 10% FBS at

28 °C. In this study, we used a low-passage-number clinical isolate of ZIKV (HS-2015-BA-01; GenBank accession no. KX520666) that was isolated from a viremic patient with symptomatic infection in Bahia State, Brazil, in 2015. Viral stocks were propagated in C6/36 cell and titrated in Vero CCL81 cells.

Cytotoxic assay. Vero CCL81 or SH-SY5Y cells were placed on 96-well plates (1×10^5 cells/well) and treated with different concentrations of selected compounds or vehicle (RPMI supplemented with FBS at 2%) and plates maintained at 37 °C with 5% of CO₂ for 48h. Then, supernatant was harvested for lactate-dehydrogenase (LDH) assay and cells were incubated with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT; Himedia, United states) solution (0.5mg/mL in PBS buffer) for 2h. After incubation of 2h at 37°C, plates were read in a spectrophotometer (BioTek, United states) at 490 nm. LDH assay was performed according to manufacturer's procedure (Bioclin LDH Kit).

***In vitro* anti-ZIKV activity.** Vero CCL-81 or SYSH5Y cells were placed in 96-well plates (1×10^5 cells/well) and infected with ZIKV at a multiplicity of infection (MOI) of 0.01 and 0.001, respectively. Plates were maintained at 37°C and 5% of CO₂ for 1 hour for viral absorption. Then, ZIKV was removed and cells treated with different concentrations of selected compounds or vehicle (RPMI supplemented with FBS at 2%). Plates were maintained for 48 hours at 37°C and 5% of CO₂. Then, supernatant was harvested for lactate-dehydrogenase (LDH) assay and stored at -80°C for further viral titration by plaque assay. Cells were incubated with MTT as described above.

3.6. References Chapter 3

1. Yokokawa, F. *et al.* Discovery of Potent Non-Nucleoside Inhibitors of

- Dengue Viral RNA-Dependent RNA Polymerase from a Fragment Hit Using Structure-Based Drug Design. *J. Med. Chem.* **59**, 3935–3952 (2016).
2. Lim, S. P. *et al.* Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling. *PLOS Pathog.* **12**, e1005737 (2016).
 3. Koch, U. *et al.* 2-(2-Thienyl)-5,6-dihydroxy-4-carboxypyrimidines as inhibitors of the hepatitis C virus NS5B polymerase: discovery, SAR, modeling, and mutagenesis. *J. Med. Chem.* **49**, 1693–1705 (2006).
 4. Summa, V. *et al.* HCV NS5b RNA-dependent RNA polymerase inhibitors: From α,γ -diketoacids to 4,5-dihydroxypyrimidine- or 3-methyl-5-hydroxypyrimidinonecarboxylic acids. Design and synthesis. *J. Med. Chem.* **47**, 5336–5339 (2004).
 5. Summa, V. *et al.* Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* **51**, 5843–5855 (2008).
 6. Home. <https://www.exscalate4cov.eu/>.
 7. Cortellis Drug Discovery Intelligence - Clarivate. <https://clarivate.com/products/biopharma/discovery-clinical-regulatory/pre-clinical-intelligence-analytics/>.
 8. Hartenfeller, M. *et al.* A collection of robust organic synthesis reactions

- for in silico molecule design. *J. Chem. Inf. Model.* **51**, 3093–3098 (2011).
9. Daylight Theory: SMARTS - A Language for Describing Molecular Patterns.
<https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>.
 10. Manelfi, C. *et al.* “Molecular Anatomy”: a new multi-dimensional hierarchical scaffold analysis tool. *J. Cheminform.* **13**, 1–19 (2021).
 11. Gharbi-Ayachi, A. *et al.* Non-nucleoside Inhibitors of Zika Virus RNA-Dependent RNA Polymerase. *J. Virol.* **94**, (2020).

CHAPTER 4

DESIGN AND SYNTHESIS OF BROAD
SPECTRUM COV_s 3CL^{PRO} REVERSIBLE
COVALENT PEPTIDOMIMETIC
INHIBITORS

Chapter 4.

Design and synthesis of broad spectrum CoVs 3CL^{pro} reversible covalent peptidomimetic inhibitors

4.1. Development of peptidomimetic aldehydes as SARS-CoV-2 3CL^{pro} reversible covalent inhibitors

4.1.1. Rational design and molecular docking

Following the unexpected pandemic COVID-19, the second part of my doctoral project was focused on the identification of novel effective DAA agents against SARS-CoV-2 by targeting 3CL^{pro}. Noteworthy, this work was started in the early 2020, when pandemic was just at the beginning, the related biology was under investigation and, more importantly, no drugs were either in development or approved for the COVID-19 treatment.

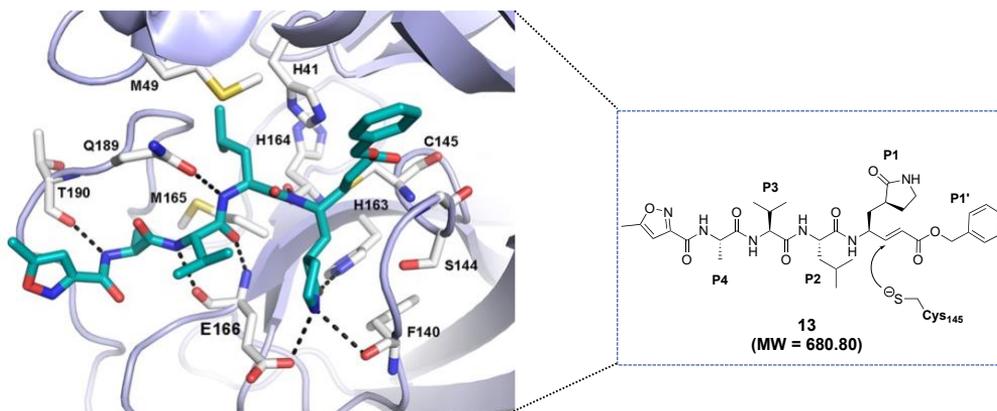
In February 2020, Zhenming Jin *et al.* released the first X-ray structure of SARS-CoV-2 3CL^{pro} in complex with the repurposed tetrapeptide **13** (Fig. 4.1), an irreversible inhibitor endowed with a vinyl carboxyl ester as Michael acceptor warhead, while sharing with other peptidomimetic inhibitors a glutamine mimetic γ -lactam and a leucine residue at P1 and P2 positions, respectively (PDB 6LU7, superseded by 7BQY). Previously, **13** was known to inhibit SARS-CoV and MERS-CoV 3CL^{pro}s in enzymatic assays, respectively with $K_i = 9 \mu\text{M}^1$ and $\text{IC}_{50} = 0.3 \mu\text{M}$,² and was then repurposed also against SARS-CoV-2, showing $\text{EC}_{50} = 16.77 \mu\text{M}$ in phenotypic assay. Unfortunately, the authors do not report data of **13** in SARS-CoV-2 3CL^{pro} biochemical assays to make a homogenous comparison of data with the other two viruses.³

Starting from this compound and its X-Ray structure, we designed a truncated analogue by removing the alanine residue in P4 and its peculiar capping group,

in order to reduce the molecular size and retaining the natural substrate sequence; moreover, replacing the vinyl ester with an aldehyde allowed to obtain a more desirable covalent reversible inhibitor (Fig. 4.1).



➤ First Co-crystalized SARS-CoV-2 Inhibitor



Aims of rational design

- Identification of reversible covalent inhibitor
- Natural substrate sequence
- MW and size reduction

Figure 4.1. Aims of rational design of tripeptide aldehydes as SARS-CoV-2 3CL^{pro} inhibitors starting from **13** as template. Crystallographic pose of compound **13** (cyan sticks) covalently bound to the active site of SARS-CoV-2 3CL^{pro} (PDB: 7BQY) is shown at the top left. The enzyme is represented as light-blue cartoons. Residues important for ligand binding are displayed as white sticks. H-bonds are depicted as dashed black lines. Hydrogens are omitted for clarity.

The resulting tripeptide **39** proved to inhibit the SARS-CoV-2 3CL^{pro} with an $IC_{50} = 0.7$ nM (Table 4.1), comparable to the most potent inhibitors reported in literature at the time, i.e. the compound **14** (for the chemical structure, see Fig. 12, 1.3.2.3 section; in-house data $IC_{50} = 0.22$ nM; reported $IC_{50} = 30$ nM),⁴ a commercial broad-spectrum inhibitor commonly included as reference compound in the biological assays (Table 4.1). Notably, the difference between

the IC₅₀ values was due to different assay conditions (in-house data: enzymatic assay carried out with buffer at pH = 7.3, 5 mM TCEP, 0.1 % BSA, and the proteins were preincubated for 30 minutes at 37°C; lit. data⁴: enzymatic assay carried out with buffer at pH = 6.5, without TCEP and BSA, and pre-incubation of 30 'at 30 °C). Due to its very interesting activity, the compound **39** was taken as a starting hit for further modifications in the peptide sequence.

By visual inspection of the X-ray structure of compound **13**, we observed that the ligand, upon binding, undergoes a conformational folding within the catalytic site, characterized by a turn-like motif around P1-P3 residues. This turn-like motif resulted from the bidentate H-bond between the backbones of Glu¹⁶⁶ and the P3 valine residue (Fig. 4.1). In this regard, we sought to insert a proline at the P2 position, in order to induce a conformational constraint for stabilizing its bioactive conformation. Thus, we designed the new tripeptide **40** (for the chemical structure, see Fig. 4.3) by replacing the P2 leucine of the starting hit **13** with a proline. Covalent docking calculations suggested that the newly designed compound **40** could effectively bind at the 3CL^{pro} catalytic site, establishing all the primary interactions observed for the co-crystallized derivative **13** (Fig. 4.2). In detail, docking predicted that essential hydrogen bonds can be established between: i) the thiohemiacetal moiety resulting by the nucleophilic attack of the reactive cysteine over the aldehyde warhead and the backbone of Gly¹⁴³ and Cys¹⁴⁵; ii) the Gln-mimetic γ -lactam and the side chains of His¹⁶³ and Glu¹⁶⁶ in the S1 pocket; iii) the ligand's P3 backbone atoms and the main chain of Glu¹⁶⁶. Additionally, according to docking, proline could allow the P3 lipophilic valine to establish van der Waals contacts with P1 γ -lactam, thus further stabilizing the ligand's binding conformation.

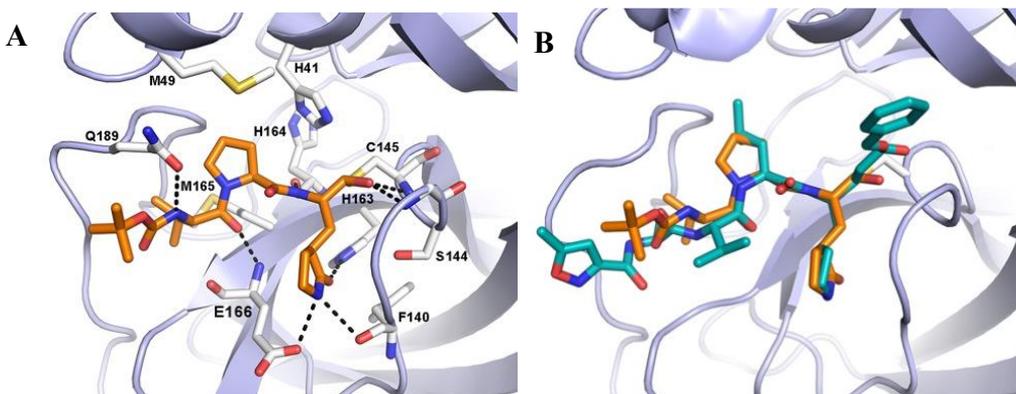
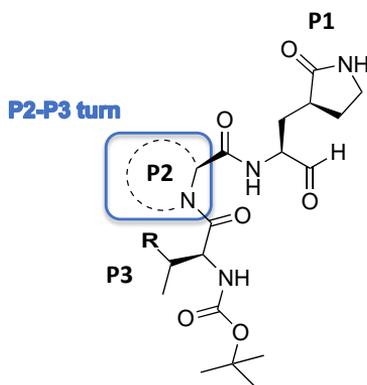


Figure 4.2. A) Docking pose of **40** (orange sticks) bound to the active site of SARS-CoV-2 3CL^{pro} (PDB: 7BQY). B) Docking pose of the designed tripeptide aldehyde **40** (orange sticks) overlaid with the X-ray conformation of **13**. The enzyme is represented as light-blue cartoons. Residues important for ligand binding are displayed as white sticks. H-bonds are depicted as dashed black lines. Hydrogens are omitted for clarity.

In addition, the introduction of a rigid and hydrophobic proline in the peptidomimetic sequence could be an efficient strategy to also improve the pharmacokinetic properties of this class of compounds, decreasing peptidomimetic metabolism.^{5,6} Moreover, the tertiary amide bond between P2-P3 can only act as an acceptor of H-bonds, avoiding the formation of ordered secondary structures and avoiding aggregations. Furthermore, the presence of a proline residue in the peptide sequence may increase its proteolytic stability.

According to the docking of **40**, compared to the natural leucine, a smaller proline may not be able to fulfil the enzyme S2 subpocket of SARS-CoV-2 3CL^{pro}. In this regard, we first explored a series of modifications of the proline-based P2 residue, while keeping constant the *N*-Boc-alanine at the P3, the glutamine mimetic γ -lactam and the aldehyde warhead. (Fig. 4.3). Considering synthetic accessibility, in this set we included compounds with a proline functionalized at the 4 position with groups of variable size and lipophilicity, such as Me (**41**), MeO (**42**) and *tert*-BuO (**43**) in *S* configuration, MeO (**44**) and *tert*-BuO (**45**) in *R* configuration, or a bicycloproline (i.e. (1*S*,3*aR*,6*aS*)-

octahydrocyclopenta[c]pyrrole-1-formamide) (**46**), which could suitably occupy the S2 subpocket of the 3CL^{pro} (Fig. 4.3). Moreover, we designed two additional tripeptides, by either replacing the proline with a 6-term piperidine analogue (**47**), in order to get further insights into the optimal ring size of the cyclic P2 amino acid, or interrogating the outcome of a P3 *tert*-BuO-threonine moiety, as in derivative **48**, used instead of the valine residue present in all the other compounds with the aim to better explore the S3/S4 sites and to modulate the compound lipophilicity (Fig. 4.3).



	40	41	42	43	44	45	46	47	48
P2									
R	Me	<i>t</i> -BuO							

Figure 4.3. Chemical structures of designed and synthesized aldehyde tripeptides as SARS-CoV-2 3CL^{pro} inhibitors, highlighting the introduction of proline and different proline derivatives at the P2 position in order to induce a P2-P3 turn.

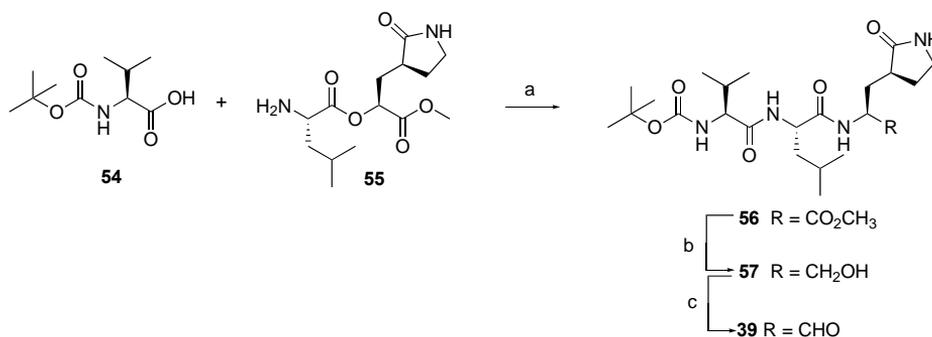
Remarkably, during our studies, the crystal structures of the proline-based repurposed HCV protease inhibitor Boceprevir in complex with the SARS-CoV-2 3CL^{pro} was released,⁷ which further corroborated our design strategy. Moreover, a work related to the focus of the present research has been published while our studies were ongoing, reporting on a series of dipeptide aldehydes

having only two different proline residues in P2.⁸ The most potent of the series, MI-30 (**49**), showed to inhibit the SARS-CoV-3CL^{pro} with a nanomolar potency ($IC_{50} = 17.2 \pm 0.6$ nM), in addition to an excellence antiviral activity ($EC_{50} = 0.54 \pm 0.13$ μ M in Vero cell line), without a significant cellular toxicity (CC_{50} (Vero) > 500 μ M). Therefore, compound **49** was synthesized (Scheme 4.3) to provide a potent reference compound for *in vitro* antiviral screenings of our designed inhibitors (in-house data $IC_{50} = 48 \pm 0.004$ nM, $EC_{50} = 3.2 \pm 0.4$ μ M).

4.1.2. Synthesis of tripeptides with aldehyde warhead

The synthesis of tripeptides **39-49** described herein, was obtained by a peptide solution synthesis via Boc-based strategy. As shown in Scheme 4.1, the target compound **39** was prepared by a linear synthetic approach, starting from a coupling reaction between the commercially available *N*-Boc-valine **54** and the intermediate **55**, synthesized as reported in literature,⁹ to afford the compound **56**.

Scheme 4.1. Synthesis of compound **39**^a



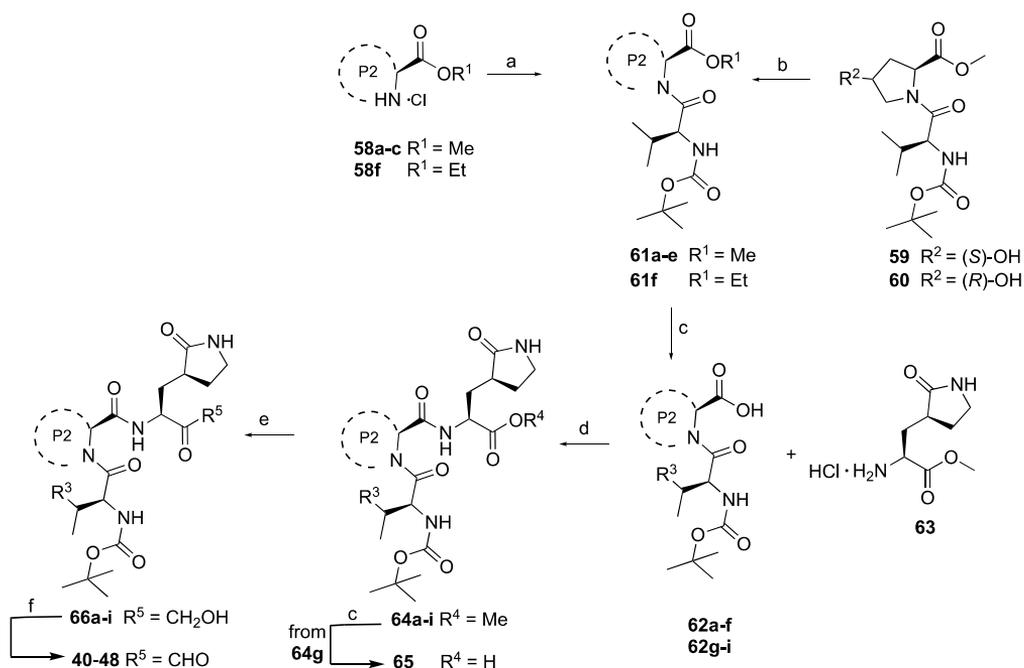
^aReagents and conditions: (a) EDCl, DIPEA, HOBT, DMF, N₂, overnight 40%; (b) LiBH₄, THF, N₂, 5 h 63%; (c) Py/SO₃, pyridine, DIPEA, DMSO, CH₂Cl₂, -5°C to RT, N₂, 24 h, 21%.

The ester derivative **56** was first reduced with lithium borohydride to primary alcohol **57**, which was oxidized into the desired aldehyde **39** under Parikh-

Doering conditions, using DMSO as oxidant activated by the sulfur trioxide pyridine complex in the presence of diisopropylethylamine as base.

The synthesis of compounds **40-48** was performed with a convergent synthetic procedure, as shown in Scheme 4.2.

Scheme 4.2. Synthesis of compound **40-48**^a



	a	b	c	d	e	f	g	h	i
P2									
R³	Me	t-BuO							

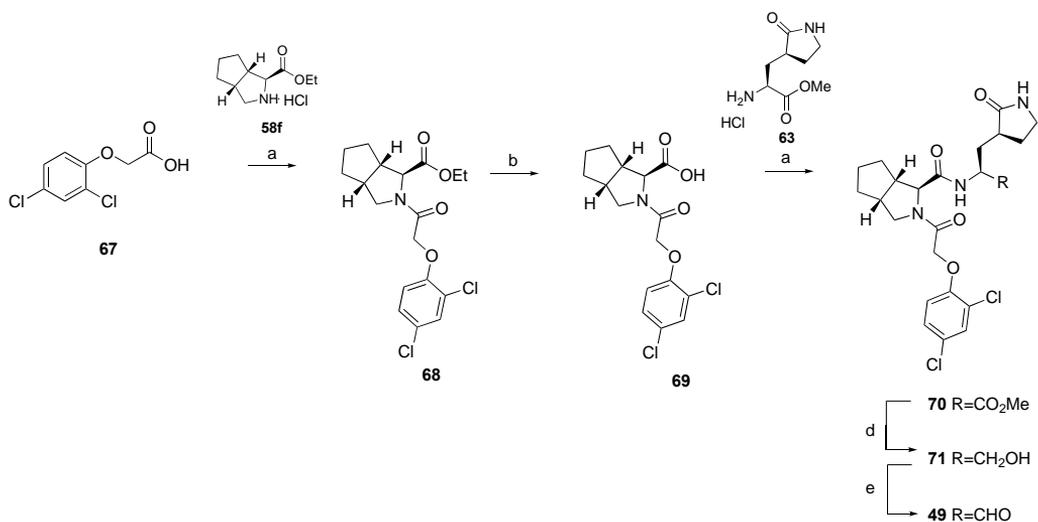
^aReagents and conditions: a) for the synthesis of **61a-c**, **61f**: **54**, HBTU, DIPEA, dry CH₂Cl₂, rt, 4 h, 45-74%; b) for the synthesis of **61d,e**: Boc₂O, Yb(OTf)₃, CH₂Cl₂, reflux, 28 h, 40-46%; c) 1N LiOH, THF, rt, 3 h, quant.; d) EDCI, DIPEA, HOBT, dry DMF, N₂, rt, overnight, 37-83%; e) LiBH₄, dry THF, N₂, 0 °C to rt, 5 h, 40-75%; f) Py/SO₃, pyridine, DIPEA, DMSO, CH₂Cl₂, -5 °C to rt, N₂, 24 h, 17-60%.

A condensation reaction, using HBTU in presence of DIPEA, between the commercial proline analogues **58a-c**, **58f** and *N*-Boc-valine **54** provided the dipeptide intermediates **61a-c** and **61f**. The dipeptide intermediates **59** and **60** were functionalized at the C-4 hydroxyl group of the P2 proline residue by reaction with di-*tert*-butyl dicarbonate under Lewis acid promoted conditions in presence of ytterbium (III) trifluoromethanesulfonate to obtain the corresponding *tert*-butyl ether intermediates **61d,e** with ~ 45% yield.

Alkaline hydrolysis of the ester derivatives **61a-f** gave the correspondent carboxylic acids **62a-f** in quantitative yields, which were reacted under coupling conditions, together with the dipeptide carboxylic acids **62g-i**, by using EDCI in presence of HOBT and DIPEA, with the commercial P1 amine hydrochloride **63** to afford the tripeptides **64a-i** in yields ranging from 37 to 83%. Subsequently, the ester derivatives **64a-i** were reduced into the alcohol intermediates **66a-i**, which were finally oxidized by Parikh-Doering reaction to obtain the final aldehydes **40-48**, while alkaline hydrolysis of the ester intermediate **64g** afforded the corresponding acid **65**.

For the synthesis of compound **49**, I followed the synthetic procedure reported in literature.⁸ Similar to the synthesis of our designed compounds **40-48**, a convergent approach was applied, as depicted in Scheme 4.3. The ethyl ester derivative **68** was quantitatively hydrolyzed, under basic conditions, to give the correspondent carboxylic acid **69**, which was coupled with the P1 amine hydrochloride **63** to obtain the intermediate **70**. The dipeptide methyl ester **70** was then reduced to primary alcohol, by using an excess of sodium borohydride, to afford the intermediate **71**, which was finally oxidized to the final target aldehyde **49** in presence of Dess-martin reagent as oxidizing agent in anhydrous dichloromethane.

Scheme 4.3. Synthesis of compound 49^a



^aReagents and conditions: (a) HATU, DIPEA, dry DMF, 16 h, 85-100%; (b) NaOH, MeOH, H₂O, 2h, 81%; (d) NaBH₄, MeOH, 0°C to 25°C, 2 h, 47%; (e) Dess-martin reagent, dry CH₂Cl₂, 25°C, 3.5 h, 40%.

The building block P2-P3 was obtained through a first amidation step between the commercially available 2-(2,4-dichlorophenoxy)acetic acid **67** and the amine hydrochloride **58f**, using HATU as coupling agent and DIPEA as base in anhydrous DMF, in order to afford the intermediate **68** in high yield.

4.2. Development of peptidomimetic nitriles as SARS-CoV-2 3CL^{pro} reversible covalent inhibitors

4.2.1. Rational design

The results obtained from this first series of compounds **39-48** showed that the introduction of a proline in P2 is a well-tolerated modification and the insertion of fairly large groups, such as bicycloproline ((1*S*,3*aR*,6*aS*)-octahydrocyclopentan[*c*]pyrrolo-1-formamide), were able to occupy the S2 sub-pocket and establish additional hydrophobic interactions (Table 4.1).

As mentioned above, these results predate the Pfizer revelation of the first-in-class oral SARS-CoV-2 inhibitor, Nirmatrelvir **12**, to be entered in clinical trials, targeting the virus's main protease. When in April 2021 the chemical structure of **12** was disclosed, we noticed structural characteristics similar to our compounds. Indeed, Nirmatrelvir is a reversible covalent tripeptide inhibitor, which presents the classic γ -lactam in P1 and the P2-P3 residues selected from the HCV protease inhibitor Boceprevir, respectively corresponding to a bicycleproline derivative and a *tert*-butyl glycine, an aliphatic non-natural amino acid. In contrast to our compounds, the P1 residue is functionalized with a nitrile group as electrophilic warhead, which is able to trap the catalytic Cys¹⁴⁵ and engage a reversible covalent bond through the formation of thioimidate adduct.¹⁰ The milder electrophilic nature of nitrile respect to aldehyde entails the advantage to achieve a higher selectivity towards cysteine proteases than other proteases characterized by less nucleophilic catalytic residues. Furthermore, nitrile improves the chemical-physical properties of the molecule, besides to be more accessible by a synthetic route and to avoid the epimerization issue at the stereocenter in P1. On the other hand, Nirmatrelvir presents a trifluoroacetyl as capping group. The replacement of the Boc group with a trifluoroacetyl capping improves the molecular hydrophilic/lipophilic balance and the pharmacokinetic profile. However, before synthesis and biological data were released by Pfizer, in April 2021 we immediately planned the synthesis of compound Nirmatrelvir **12**, in order to provide a reference compound for *in vitro* and *in vivo* screening, as it will be described in Chapter 5. For homogeneity of biological data reported here, values of IC₅₀ and EC₅₀ in Vero cell line of **12** have also been reported in Table 4.1.

Based on the advantages of the nitrile group over the aldehyde as warhead, we designed and synthesized a new series of nitrile peptides as SARS-CoV-2 3CL^{pro}

(**50-53**), by exploring novel substituents at the P2 and P3 positions and keeping constant the classical γ -lactam analogue of glutamine in P1 (Fig. 4.4).

Considering the results of the previous series (**39-48**) and the advantages of the introduction of the proline in P2 position, we decided to investigate a novel proline analogue in P2, functionalized with two methyl groups at position 4 (**50, 51**).

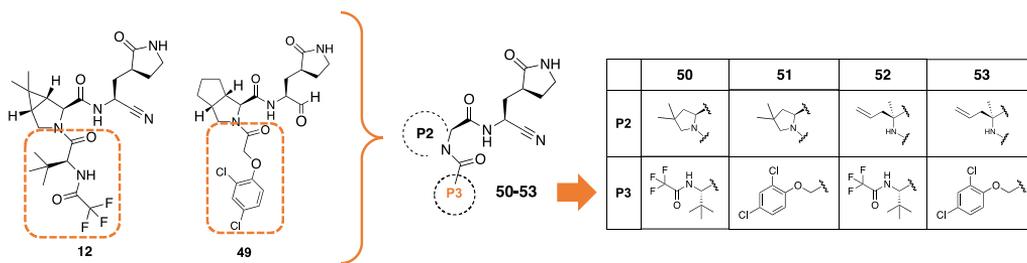


Figure 4.4. Chemical structures of compounds **50-53** with nitrile warhead. Selection of P3 residues (orange dashed rectangular squares) from literature hits of SARS-CoV-2 3CL^{pro} inhibitors (**12** and **49**).

Unlike the proline derivatives already investigated, the (*S*)-4,4-dimethylpyrrolidin-2-carboxylic acid has neither a stereogenic centre nor an ether function at position 4 of the pyrrolidinic ring to act as a hydrogen-bond acceptor, but offers the possibility of establishing hydrophobic interactions, fitting into the S2 sub-pocket due to its small size. In addition, preserving the (*S*)-4,4-dimethylpyrrolidin-2-carboxylic acid in P2, the compounds **50** and **51** aimed to investigate the effect of different residues in P3 on antiviral activity, by merging structural motif of potent SARS-CoV-2 3CL^{pro} inhibitors known in literature, with our designed compounds (Fig. 4.4). In line with Nirmatrelvir **12**, compound **50** presents a *tert*-butyl glycine residue, capable to establish valine-like interactions at the catalytic site, which is capped with a trifluoroacetyl group, in order to improve the hydrophilic/lipophilic balance of the molecule. On the other hand, compound **51** presents a P3 2,4-dichlorophenoxyacetic acid, which

represents the best non-amino acid capping group reported to date, on basis of the Qiao *et al* studies and the excellent results obtained for compound **49**.⁸ Indeed, the hydrophobic nature and the medium size of 2,4-dichlorophenoxyacetic acid in P3 resulted to be well tolerated into S4 subpocket of 3CL^{pro} catalytic site; in addition, 4-dichlorophenoxyacetic acid presents the advantages to enhance the potency and pharmacokinetic properties of the compound, even demonstrated by *in vivo* studies in a transgenic mouse model of SARS-CoV-2.⁸

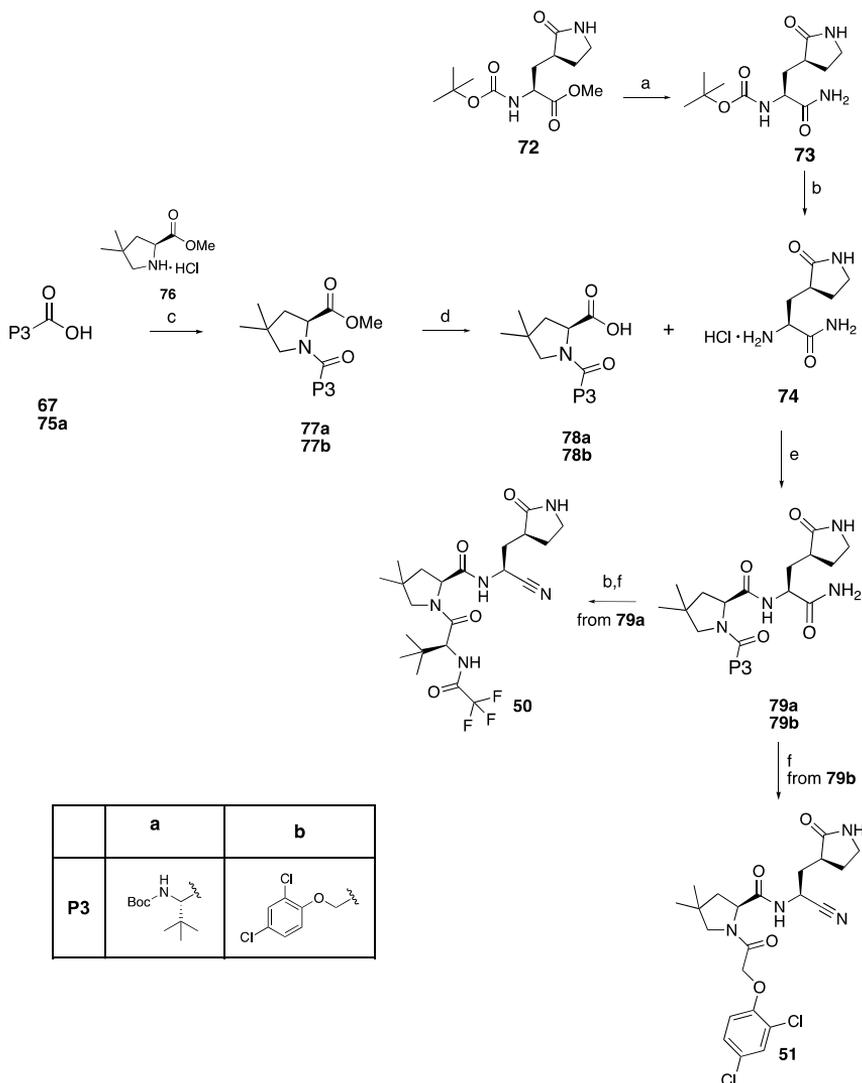
Differently, the compounds **52** and **53** aimed to explore a new residue in P2, namely methyl-allylglycine, which is not characterized by a pyrrolidine ring (Fig. 4.4). Similar to the isopropyl group of leucine in the natural protease substrate, the methyl-allylglycine could establish hydrophobic interactions in the S2 subpocket through its allyl residue. In addition, the methyl in the *S* configuration represents a structural constrain center, that may reproduce the turn between P1-P3 residues and replay the bioactive conformation of the molecule at the catalytic site of the protease. As well as for compounds **50** and **51**, the *tert*-butyl glycine and 4-dichlorophenoxyacetic acid were investigated as P3 residue in **52** and **53** compounds, respectively.

4.2.2. Synthesis of peptidomimetics with nitrile warhead

A *tert*-butyloxycarbonyl peptide solution synthesis was applied for the synthesis of tripeptides **50-53**. As shown in Scheme 4.4, the synthesis of compounds **50**, **51** was performed by a convergent synthetic procedure. The synthesis of the first building block to obtain the P1 residue was started from the commercially available *N*-Boc-amino ester **72**, which was subjected to an ammonolysis reaction in 33% aqueous NH₃/MeOH 1:1 mixture at room temperature, to afford the

corresponding amide intermediate **73** in quantitative yield. No racemization of any of the chiral centers in the product was observed by NMR analysis of intermediate **73**.

Scheme 4.4. Synthesis of compounds **50**, **51**^a



^aReagents and conditions: a) 33% aq. NH₃, MeOH, RT, 12 h, quantitative; b) 4N HCl in 1,4-dioxane, CH₂Cl₂, overnight, quantitative; c) HBTU, DIPEA, dry DMF, N₂, RT, 16 h, 94, 96%; d) 1N aq. LiOH/THF, 1:1, 0°C to rt, 5 h, quantitative; e) COMU, oxyma, DIPEA, dry DMF, N₂, rt., 16 h, 53%, 64%; f) dry Pyr, TFAA, dry CH₂Cl₂, 0°C to rt., 6 h, 46, 79%.

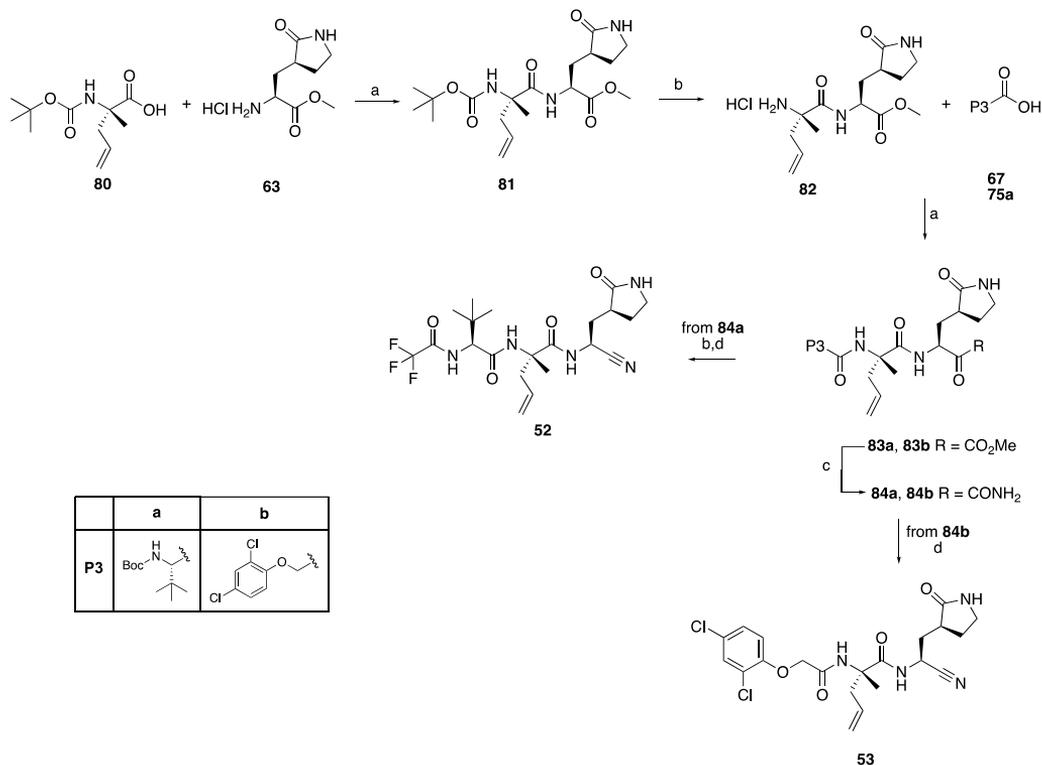
Compound **73** was deprotected, by the removal of the Boc group under acidic conditions, to quantitatively give the desired building block hydrochloride amide salt **74**.

The P3-P2 dipeptide acid was built starting from a condensation reaction between the commercially available proline **76** and compounds **67** and **75a**, by using HBTU in presence of DIPEA to obtain in high yields the dipeptide intermediates **77a** and **77b**, which were hydrolyzed under basic conditions to the corresponding carboxylic acids **78a** and **78b**. The P3-P2 acid intermediates **78a** and **78b** were reacted with P1 amine hydrochloride salt **74**, using 1-[(1-(Cyano-2-ethoxy-2-Oxoethylideneaminoxy)- dimethylamino-Morpholino)] Uranium (COMU) and oxyma [ethyl-2-cyano-2-(hydroxyamino)acetate] as coupling reagents and DIPEA as base, and afforded the tripeptides **79a** and **79b** in a yield of 53% and 64%, respectively. After the quantitative removal of the *tert*-butoxycarbonyl group by using 4N HCl in dioxane, the intermediate **79a** was submitted to a “two-steps-one-pot” procedure: a first treatment with anhydrous trifluoroacetic anhydride (TFAA), in presence of anhydrous pyridine as a base, to promote the functionalization of the primary *N*-terminal amine with the trifluoroacetyl group, and then using again TFAA/pyridine to promote the dehydration of the amide into final nitrile **50**. A step of dehydration by TFAA/pyridine allowed to afford the target nitrile **51** with a 79% yield, starting from amide intermediate **79b**.

Differently from the synthetic procedures described so far, the synthesis of **52** and **53** has been achieved by a linear synthetic approach (Scheme 4.5). As shown in Scheme 4.5, the dipeptide methyl ester **81** was synthesized starting from the commercial carboxylic acid **80** and the hydrochloride amine **63**, using COMU and oxyma and DIPEA as base in anhydrous DMF. The acidic Boc-group removal from intermediate **81** provided the primary amine derivative **82** as hydrochloride salt for the next amidation step with the commercially available

compounds **67** and **75a**, in order to obtain the correspondent tripeptide **83a** and dipeptide **83b** in good yields, respectively.

Scheme 4.5. Synthesis of compounds 52, 53^a



^aReagents and conditions: a) COMU, Oxyma, DIPEA, dry DMF, N₂, rt, 50, 55%; b) 4N HCl in dioxane, DCM, 0°C to rt, 8 h, quantitative; c) 33% aq. NH₃, MeOH, rt, 12 h, quantitative; d) dry Pyr, TFAA, dry THF, 0°C to rt, 28, 30%.

The derivatives **83a** and **83b** were then subjected to ammonolysis to give the correspondent amides **84a** and **84b**. Intermediate **84a** was deprotected by 4N HCl in dioxane and treated with dry TFAA and dry pyridine to afford target compound **52**, while intermediate **84b** was dehydrated to target **53**.

4.3. Biological evaluation

4.3.1. Enzymatic assays against SARS-CoV-2 and MERS-CoV 3CL^{pro}s

All the synthesized peptides **39-53** were tested in a Fluorescence Resonance Energy Transfer (FRET) enzymatic assay, in order to evaluate their inhibitory activity against the isolated SARS-CoV-2 3CL^{pro} processing a fluorescent substrate. Also tripeptides **64g**, **65** and **66g**, namely the corresponding C-terminal ester, acid, and alcohol analogues (structures in Scheme 4.2), respectively, of aldehyde **40**, were tested to assess the key role of the electrophilic warhead as a requirement for the inhibitory activity. The experiment carried out with buffer at pH = 7.3, including Tris(2-CarboxyEthyl)Phosphine (TCEP) (5mM) as reducing agent of disulfide bonds, and Bovin Serum Albumin (BSA) (0.1%) for stabilization of enzyme. The commercial compound **14** and in-house synthesized Nirmatrelvir **12** and MI-30 **49** were included as positive controls in the biochemical assays.

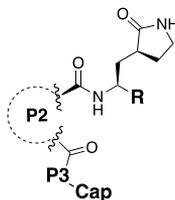
All the compounds of the tripeptide aldehyde series **39-48** were tested at fixed 30 μ M concentration and only active compounds ($\geq 70\%$ enzyme inhibition) were tested for dose-effect curves, to obtain the IC₅₀ values.

All aldehydes effectively inhibited the SARS-CoV-2 3CL^{pro} ($>70\%$) and almost all showed IC₅₀ values in the sub- to low-nM range (0.24-6.23 nM) (Table 4.1), while derivatives **43**, **44**, and **47** resulted much less potent with IC₅₀ in the μ M range (4.2, 20.43 and 3.3 μ M, respectively). The compound **40**, having a non-functionalized proline in P2, showed a potent enzymatic inhibition (IC₅₀ = 6.23 nM), even if not yet comparable to the close analogue **39** (IC₅₀ = 0.70 nM) and the reference **14** (IC₅₀ = 0.22 nM), both characterized by a P2 leucine. On the contrary, compounds **64g**, **65** and **66g** (whose structures are reported in Scheme

4.2), that differ from **40** only for the replacement of the aldehyde with the not reactive C-terminal groups, such as carboxylic acid or the simple primary alcohol, resulted completely inactive at 30 μM (data not reported in Table 4.1). These data confirm that the electrophilic warhead is essential to observe enzyme inhibition in this series. Furthermore, the introduction of (*S*)-pipecolinic acid as P2 turned out to be very detrimental for activity, with the corresponding analogue **47** ($\text{IC}_{50} = 3.35 \mu\text{M}$) resulting around 500-fold less potent than the proline-based analogue **40**. This suggested that a larger 6-membered ring could either be less tolerated by the S2 pocket or not allow the P1 and P3 residues to properly accommodate in the enzyme active site. More interesting are the effects of the substituents on the proline in P2. Here, the introduction at the C4 position of small substituents, such as a methyl (**41**) and a methoxy (**42**) with *S* configuration, provided the most potent inhibitors with picomolar activity ($\text{IC}_{50} < 1 \text{ nM}$), while a bulkier *tert*-butoxy (**43**) substituent was very detrimental ($\text{IC}_{50} = 4.2 \mu\text{M}$) for the activity. These results are opposite for the corresponding *R* epimers, with the *R*-MeO derivative **44** resulting in low inhibition ($\text{IC}_{50} = 20.29 \mu\text{M}$) of the protease, while the (*R*)-*tert*-BuO analogue **45** displayed nM potency ($\text{IC}_{50} = 2.5 \text{ nM}$). The introduction of a bulkier bicycloproline analogue as P2 residue (**46**) also provided a potent compound with $\text{IC}_{50} = 5.0 \text{ nM}$, in agreement with the potency reported for the Nirmatrelvir **12** that is endowed with a similar bicycloproline.

These results confirmed that the P2 proline residue is a key moiety to modulate the potency. Finally, compound **48**, having a *tert*-Bu-threonine instead of a valine as P3 residue, demonstrated a potency ($\text{IC}_{50} = 5.3 \text{ nM}$) similar to that of **40**, suggesting a marginal or neutral contribution to the binding of this larger substituent to the protein.

Table 4.1. Biological activities of target compounds **39-48**, **50-53** and reference compounds **12**, **14**, **49**.



General structure of compounds **39-48**, **50-53**

Cpd	R	P2	P3	Cap	3CL ^{pro} IC ₅₀ (nM) ^a SARS-CoV-2	3CL ^{pro} IC ₅₀ (nM) ^a MERS	EC ₅₀ (μM) ^b SARS-CoV-2 (SI) ^c	EC ₅₀ (μM) ^d SARS-CoV-2 (SI) ^c	CC ₅₀ (μM) ^e Vero E6
12 (Nirmatrelvir)					50 ± 0.003	0.22 ± 0.001	1.02 ± 1.4 (> 98)	0.04 ± 0.12 (> 2500)	> 100
14 (GC-376)					0.22 ± 0.05	0.24 ± 0.06	3.55 ± 1.4 (> 28)	0.40 ± 0.12 (> 250)	> 100
49 (MI-30)					48 ± 0.004	NT ^f	3.2 ± 0.4 (> 31)	0.21 ± 0.09 (> 476)	> 100
39	CHO				0.70 ± 0.02	0.02 ± 0.003	40.1 ± 4.4 (> 2.5)	3.6 ± 0.5 (> 27.7)	> 100
40	CHO				6.23 ± 0.70	0.02 ± 0.0007	22.5 ± 8.0 (> 4.4)	16.0 ± 5.2 (> 6.2)	> 100
41	CHO				0.47 ± 0.10	2.33 ± 0.50	47.2 ± 11.2 (> 2)	16.9 ± 3.9 (> 5.9)	> 100
42	CHO				0.24 ± 0.02	0.007 ± 0.0005	23.6 ± 10.8 (> 4.2)	8.7 ± 0.1 (> 11.5)	> 100
43	CHO				4196 ± 944	NT ^f	19.6 ± 3.2 (> 5.1)	2.8 ± 0.7 (> 36)	> 100
44	CHO				20290 ± 7000	NT ^f	>100	>100	> 100
45	CHO				2.5 ± 0.46	26.0 ± 2.0	33.4 ± 5.2 (> 3)	34.0 ± 1.2 (> 29.4)	> 100
46	CHO				5.0 ± 0.25	143.4 ± 12.79	5.3 ± 1.7 (> 19)	0.21 ± 0.13 (> 476)	> 100
47	CHO				3350 ± 280	NT ^f	>100	5.83 ± 0.81 (> 17)	> 100
48	CHO				5.3 ± 0.73	0.022 ± 0.002	22.9 ± 8.18 (> 4.3)	1.86 ± 0.96 (> 54)	> 100

50	CN				1.4 ± 0.0001	400 ± 0.17	ND ^g	2.2 ± 0.5 (> 45)	> 100
51	CN				17 ± 0.003	125 ± 0.04	ND ^g	> 100	> 100
52	CN				8.0 ± 0.001	90 ± 0.005	ND ^g	> 100	> 100
53	CN				1590 ± 0.94	NT ^f	ND ^g	2.2 ± 0.5 (> 45)	> 100

^aCompound concentration required to reduce the 3CL^{pro} activity by 50%. ^bCompound concentration required to reduce the SARS-CoV-2 cytopathic effect in Vero E6 cell by 50% (EC₅₀) in absence of P-gp inhibitor (CP-100356). ^cSI is the selectivity index calculated as the CC₅₀/EC₅₀ ratio. ^dEC₅₀ in presence of P-gp inhibitor (CP-100356) at the fixed concentration of 2 μM. ^eCompound concentration required to reduce Vero E6 cell viability by 50%. ^fNT = not tested. ^gND = not determined.

Replacing the aldehyde warhead with a nitrile function (**50-52**) provides potent compounds as well, despite its less pronounced electrophilic nature. These compounds showed IC₅₀ values in the low nM range (1.4-17 nM) (Table 4.1), while **53** (IC₅₀ = 1.6 μM) resulted the less potent of the series. The compound **50** showed the most potent enzymatic inhibition among the tested nitrile peptides, with an IC₅₀ value of 1.4 nM. This result suggested that two methyl groups at the 4-position of the P2 proline analogue could add additional hydrophobic interactions in the S2 pocket; furthermore, the introduction of the *tert*-butyl glycine residue in P3 and the trifluoroacetyl capping, present on the Nirmatrelvir **12** prototype, proved to be a positive contribution to the ligand binding affinity. Indeed, changing the *N*-(trifluoroacetyl)-*tert*-butyl-glycine by the 2,4-dichlorophenoxyacetyl in P3 afforded the compound **51**, that was approximately 10-fold less potent than analogue **50**. The same activity profile was shown for **52** and **53**. Compound **52** showed significant inhibitory activity, with an IC₅₀ = 8 nM, supporting that the substitution of a leucine-like residue in P2 may work, especially if supported by a spatial constriction center such as methyl in the *S*

configuration. Therefore, the analogue **53** is almost 200-fold less potent than **52**, further proving that the P3 *tert*-butyl glycine is better tolerated respect to 2,4-dichlorophenoxyacetyl group.

Overall, almost all peptide synthesized resulted potent inhibitors of SARS-CoV-2 3CL^{pro}, albeit with different degree of potencies, prompting their evaluation as broad-spectrum CoV 3CL^{pro} inhibitors. Thus, for compounds showing $IC_{50} < 1 \mu M$ on SARS-CoV-2 3CL^{pro}, the screening was also extended to MERS-CoV 3CL^{pro}, taken as representative homologue protein in order to evaluate the potential broad inhibition within the same virus family (Table 4.1). The selected compounds (**39-42**, **45**, **46** and **48-52**) resulted in a broad-spectrum inhibition of CoVs 3CL^{pro}, being active also against the MERS homologue. The tripeptide aldehydes were generally found potent against MERS 3CL^{pro}, with IC_{50} ranging from 0.007 to 143 nM (Table 4.1). In particular, derivative **42** showed an $IC_{50} = 0.007$ nM, resulting to our knowledge the most potent MERS-CoV 3CL^{pro} inhibitor so far reported. Only derivative **46** exhibited a lower but still significant inhibitory activity on MERS 3CL^{pro} ($IC_{50} = 143$ nM), if compared to the data obtained on SARS-CoV-2 ($IC_{50} = 5$ nM).

Nitrile peptides **50-52** showed a potent enzymatic inhibition against MERS 3CL^{pro} as well, with IC_{50} values ranging from 90 to 400 nM. Notably, derivative **50** showed the most potent inhibition against SARS-CoV-2 3CL^{pro} among the designed nitrile peptides ($IC_{50} = 1.4$ nM), but its inhibitory potency decreases significantly (~ 300-fold) against the MERS homologue protease ($IC_{50} = 400$ nM) (Table 4.1).

In order to rationalize these results, *in silico* studies were performed to compare co-crystallographic structures obtained for some of our compounds in complex with the SARS-CoV-2 3CL^{pro} and data reported in literature for the MERS homologue protein, as described in the next section 4.6.

4.3.2. Phenotypic assays against SARS-CoV-2 infection

All the active compounds (**39-53**) were tested in a phenotypic assay, using Vero E6-GFP stably expressing a GFP protein under the control of cytomegalovirus promoter. Results were expressed as EC₅₀ and determined through CPE method. The host cell toxicity was also evaluated by quantifying the viability as proportional to the GFP signal of untreated cells and expressed as CC₅₀. Indeed, active compounds causing cytotoxicity, for instance non-specific cellular protease inhibitors, may show antiviral activity coupled to reduced viability, resulting in a low selectivity index ($SI = CC_{50}/EC_{50}$). Active compounds with low or no cytotoxicity exert a genuine antiviral effect resulting in suitable SI (≥ 10), which is a very important feature for developable antiviral agents. In the assay, compound **14**, and then **12** and **49**, were again included as positive controls. All the compounds were tested at 100 μ M and those able to reduce > 95% viral replication were tested in concentration-effect curves for EC₅₀ values determination (Table 4.1).

With the only exceptions of derivatives **44** and **47**, which already showed null or modest inhibition in the biochemical assay against SARS-CoV-2 3CL^{pro}, the other aldehyde compounds displayed EC₅₀ values ranging from 5 to 47 μ M, comparable to the positive controls **14** (EC₅₀ = 3.55 μ M), Nirmatrelvir **12** (EC₅₀ = 1.02 μ M), and MI-30 **49** (EC₅₀ = 3.2 μ M), and no significant cell toxicity (CC₅₀ > 100 μ M) (Table 4.1). However, a significant drop in the compounds' potency from the biochemical to the cell-based assays was observed. Considering that Vero E6 cells express high level of active membrane transporters such as the P-glycoprotein (P-gp), we speculated that drug efflux could contribute to the high antiviral EC₅₀/enzyme IC₅₀ ratio, as already observed for other peptidomimetic inhibitors of CoVs 3CL^{pro} (*i.e.* Nirmatrelvir **12**).¹¹ The underestimation of the antiviral potency may lead to discard compounds likely able to potently inhibit

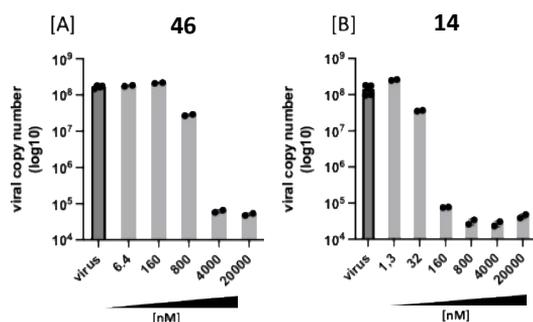
SARS-CoV-2 replication in human lung cells, the relevant tissue for human CoVs and COVID-19. Thus, according to recent literature, all the compounds were co-dosed with a commercial non-toxic P-gp inhibitor (CP-100356) at the fixed concentration of 2 μM ,¹¹ resulting in a shift for the antiviral activity without any detectable impact on cell viability/toxicity ($\text{CC}_{50\text{s}} > 100 \mu\text{M}$) (Table 4.1). This effect was considerably especial for derivative **46**, which showed a potent antiviral activity with an $\text{EC}_{50} = 0.21 \mu\text{M}$ (w/o CP-100356: $\text{EC}_{50} = 5.3 \mu\text{M}$), a value closer to its IC_{50} and 2-fold lower than reference **14** (with CP-100356: $\text{EC}_{50} = 0.4 \mu\text{M}$). Consequently, the SI values of all the compounds were also significantly improved, with **46** having the highest SI (> 476). Excluding derivatives **40-42**, the remaining compounds (**39, 43, 45, 47, and 48**) gained more than 3-fold improvement in their antiviral potency, displaying EC_{50} values in the low μM range and SI values > 11.5 (Table 4.1). In summary, derivative **46** resulted the most active inhibitor in cellular assay, showing EC_{50} and a SI values comparable to the most potent SARS-CoV-2 3CL^{pro} inhibitors reported in literature when evaluated in the same cellular system, including the oral inhibitor Nirmatrelvir **12**.¹¹

Interestingly, although Nirmatrelvir **12** is about 10 times more potent than compound **46** in the enzyme inhibition assay on 3CL^{pro} of SARS-CoV-2, the in-cell activities in the CPE assays are comparable, with EC_{50} values = 0.21 and 0.04 μM for compound **46** and Nirmatrelvir **12** respectively, in the presence of the P-gp inhibitor CP100356 (in the absence of this inhibitor the EC_{50} values = 5.3 and 1.02 μM respectively).

Unlikely, nitrile peptides **50-53** did not show to be active in cell-based assays, showing EC_{50} values $> 100 \mu\text{M}$, although they were not toxic to the cell with CC_{50} values $> 100 \mu\text{M}$. Therefore, the biological assays were also performed in the presence of a fixed concentration at 2 μM of the P-gp inhibitor CP100356. No increase in cell toxicity was reported for any molecule, with $\text{CC}_{50} > 100 \mu\text{M}$.

However, compounds **51**, **52** remained inactive at $EC_{50} > 100 \mu\text{M}$. Compound **50** increased its activity almost 25-fold ($EC_{50} = 2.2 \pm 0.5 \mu\text{M}$; $SI > 45$), proving an antiviral activity comparable to the tripeptide aldehydes previously synthesized. Therefore, a considerable reduction of antiviral activity was observed for **50** in comparison to its enzymatic activity ($IC_{50} = 1.4 \text{ nM}$), observed in biochemical assay. Similarly, the compound **53** increased its activity almost 25-fold ($EC_{50} = 2.2 \pm 0.5 \mu\text{M}$, $SI > 45$), with an antiviral activity comparable to **50**, although **53** showed an enzymatic potency significantly lower than **50** in biochemical assay ($IC_{50} = 1.6 \mu\text{M}$).

Considering the overall results from biochemical and antiviral cell-based assays, compounds **46** resulted the best inhibitor of this series, thus its antiviral activity was investigated in a more physio-pathological relevant cell line for COVID-19 research. Viral load reduction assays were performed in human epithelial lung adenocarcinoma Calu-3 cells, that is more relevant for a respiratory virus such as SARS-CoV-2 and presents a low expression of efflux pumps, able to eliminate the artifacts of drug extrusion on compound activity. The extracellular viral load was determined by qPCR, titrating S gene copy number in the RNA extracted from the supernatant 48 hours post infection. Results showed that **46** effectively reduced viral titers, causing a 3 Log decrease of viral titer in the cell supernatant at $20 \mu\text{M}$ and $4 \mu\text{M}$, and showing an $EC_{50} = 0.749 \pm 0.085$ without sign of cytotoxicity (Fig. 4.5).



	Calu-3 EC_{50} (μM)
14	0.026 ± 0.011
46	0.749 ± 0.085

Figure 4.5. Evaluation of antiviral activity of compounds **46** and GC376 (**14**) against SARS-CoV-2 in human lung adenocarcinoma cell line Calu-3. Calu-3 cells were infected with SARS-CoV-2 strain BetaCov/Belgium/GHB-03021/2020. At 48 hours post-infection, viral RNA was extracted and SARS-CoV-2 S-gene copy number was determined by RT-qPCR. Results report the mean and standard deviation of two independent biological replicates in duplicate.

Therefore, these results confirmed the potent antiviral activity observed in the Vero cell assays in presence of the P-gp inhibitor.

4.3.3. Binding and kinetic assays

Compounds **46** was further characterized by additional experiments. In particular, a differential scanning fluorimetry assay was carried out to validate the specific binding of this compounds to the SARS-CoV-2 3CL^{pro}, and **14** was included as reference compound. Compound **46** and the reference **14** increased the thermal stability of the target protein with considerable thermal shifts (Fig. 4.6). In detail, derivative **46** caused the highest increase in the protein melting point (+ 11°C), thus suggesting that bicycloproline could well fill the S2 pocket of this enzyme, while the reference **14** showed a less pronounced effect (+ 4°C).

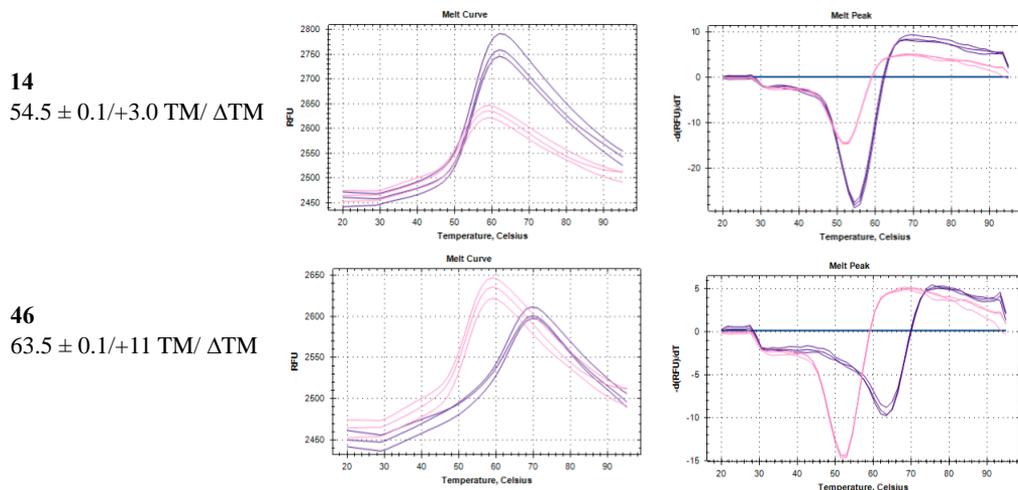


Figure 4.6. The differential scanning fluorimetry assay of **46** was performed on SARS-CoV-2 3CL^{pro}, including GC-376 (**14**) as reference compound. The melt temperature (T_m) and the thermal shift (ΔT_m) was calculated as described in literature.¹² All experiments were performed in triplicate, and the values are presented as mean \pm SD.

In order to elucidate the mechanism of action of **46**, a kinetics study was performed by reading the signal generated by the proteolytic cleavage of the FRET substrate for 60 minutes (Fig. 4.7). The results for **46** showed a biphasic enzymatic progression curve in the presence but not in absence of inhibitors, which is a characteristic behavior of a slow covalent binding inhibitor, more evident in the first 30 minutes of reaction. The same profile is reported for reference compound **14** (Fig. 4.7). Moreover, binding affinity of derivative **46** for SARS-CoV-2 3CL^{pro} was estimated, resulting in an apparent $K_i = 3.2 \pm 0.96$ nM.

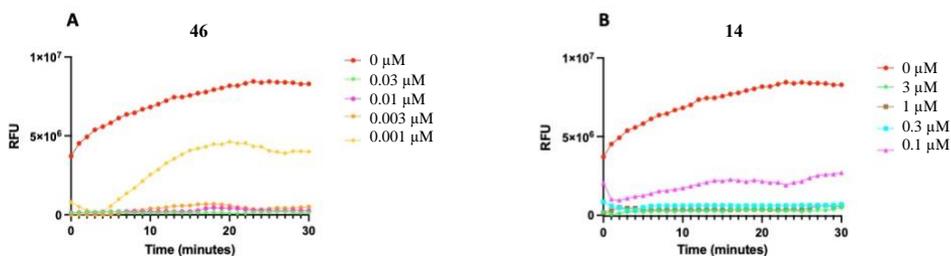


Figure 4.7. 3CL^{pro} reaction curves in presence and absence of inhibitors. The kinetics studies were performed with the addition of various concentrations of compounds **46** (panel A) and **14** (panel B).

4.4. Crystallographic studies

Single crystal X-ray crystallography was exploited to solve the structures of the most potent compound **42** and some other selected inhibitors (**41**, **46** and **48**) bound in the catalytic site of SARS-CoV-2 3CL^{pro}. Crystallization experiments were carried out through seeding in the Morpheus screening (Molecular

Dimensions) as previously described.¹³ For the four compounds, well diffracting crystals grew in several conditions in space group $P2_12_12_1$, reaching resolutions in the range of 1.35-1.66 Å (7: 1.63 Å, 8: 1.5 Å, 12: 1.66 Å, 14: 1.35 Å). In all the resulting crystal structures, the well-known heart-shape dimer of SARS-CoV-2 3CL^{pro} is present in the asymmetric unit.^{13,14} Refinement was carried out alternating manual model building cycles in coot¹⁵ and automated refinement cycles in Phenix.¹⁶ Accordingly, each inhibitor was unambiguously modelled covalently bound to the catalytic Cys¹⁴⁵ in both protomers (Fig. 4.9).

The four solved crystal structures show that a new chiral center is formed following the attack of the catalytic Cys¹⁴⁵ to the aldehyde group of the inhibitors, and that the resulting thiohemiacetal is in the *S* configuration, confirming that this reaction is typically enantioselective, as previously reported.⁷

In the covalent adducts, the thiohemiacetal occupies the oxyanion hole formed by the backbone amide groups of Gly¹⁴³, Ser¹⁴⁴, and Cys¹⁴⁵, where it forms hydrogen bonds with the amide groups of Gly¹⁴³ and Cys¹⁴⁵. According to our initial design hypothesis, the four inhibitors adopt a turn-like motif around P1-P2 groups and extend along the S1-S4 binding subsites, interacting with active site residues through several hydrogen bonds and hydrophobic contacts. The S1 cavity is occupied by the P1 γ -lactam ring, that forms hydrogen bonds with the backbone of Phe¹⁴⁰ and with the side chains of His¹⁶³ and Glu¹⁶⁶. The proline ring (**48**) occupies the S2 pocket, where substituted residues can also be hosted, as expected. Indeed, both the prolines functionalized at the 4 position with either Me (**41**) or MeO (**42**) and the bicycloproline (*i.e.*, (1*S*,3*aR*,6*aS*)-octahydrocyclopenta[*c*]pyrrole-1-formamide) (**46**) fit in this cleft establishing, albeit to a different extent, hydrophobic interactions with the side chains of residues, such as His⁴¹, Met⁴⁹, Met¹⁶⁵ and Gln¹⁸⁹. On the other hand, these residues could partly hinder an optimal accommodation of the bulkiest groups,

explaining the lower affinity of compounds **46** and especially **43**, compared to **41** and **42**.

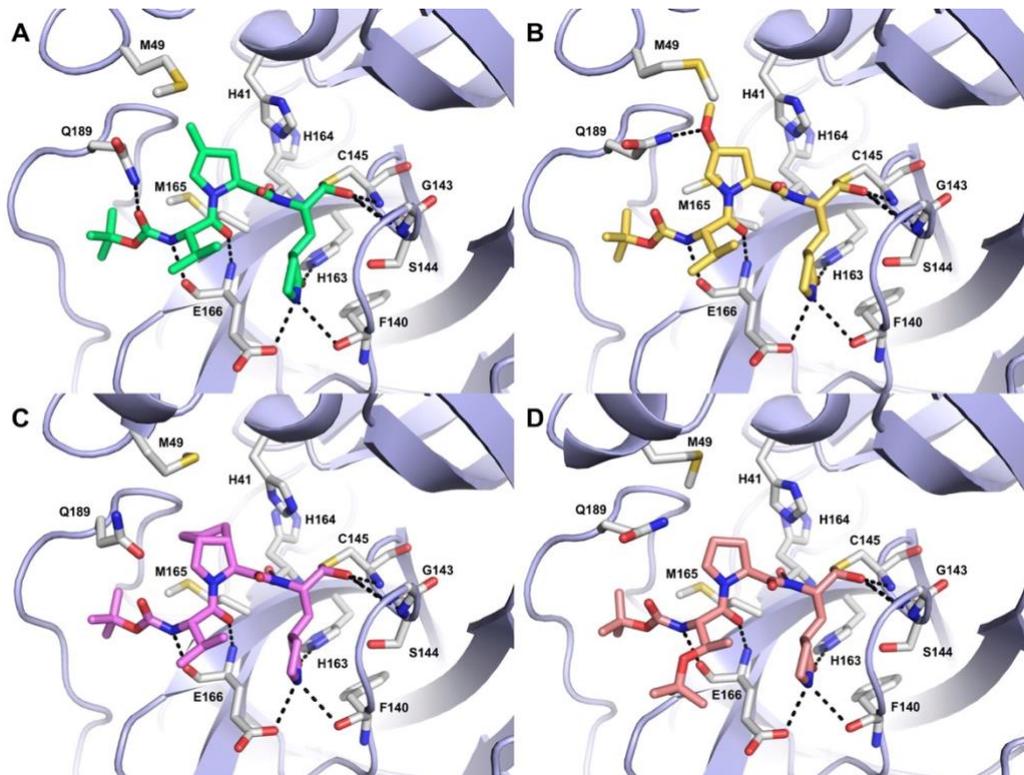


Figure 4.9. Crystal structures of compounds **41** (A, green), **42** (B, yellow), **46** (C, violet), and **48** (D, salmon) into SARS-CoV-2 3CL^{pro} active site represented as light-blue cartoons. Residues important for ligand binding are displayed as white sticks and labeled. H-bonds discussed in the text are depicted as dashed black lines.

It is also interesting to remark that: i) the MeO oxygen of **42** forms an additional H-bond with the Gln¹⁸⁹ side chain, which might explain why this compound is the most potent within the series, and ii) the *S* configuration of the proline γ -carbon is required to ensure the optimal positioning of the substituent within the S2 pocket, as demonstrated by the very low affinity displayed by **44**. In the S3-S4 pockets, each ligand forms two hydrogen bonds with the backbone of Glu¹⁶⁶ through the backbone of its P3 group, while **41** forms an additional H-bond with

the Gln¹⁸⁹ side chain through its terminal *N*-Boc cap. Finally, the side chain of the P3 residue (valine in **41**, **42** and **46**, *tert*-But-threonine in **48**) can form intramolecular hydrophobic contacts with the P1 γ -lactam, that contribute to stabilizing the ligand binding conformation.

Altogether, the four solved X-ray structures, confirm the predictions from the docking studies and the validity of our rational design approach.

4.5. Molecular modelling on MERS-CoV 3CL^{pro}

CoVs 3CL^{pro} possess a well-conserved active site, with highly similar substrate recognition profiles. Particularly, SARS-CoV-2 and MERS-CoV 3CL^{pro}s share overall 50% identity. Indeed, the overlay of their X-ray structures indicates a high degree of structural similarity and conservation, with very few mutations occurring at the substrate binding site. Herein, the S1', S1 and S2 subsites show mostly conservative substitutions, whereas the main differences involve the residues lining the S3-S4 region and the lid covering the S2 (Fig. 4.10). Therefore, these mutations result in slight differences in the overall active site architecture. In order to gain insights into the possible binding mode of the most potent compound **42** into MERS 3CL^{pro} active site and to help interpretation of SAR data, docking studies were performed through the same covalent docking procedure used for the SARS-CoV-2 homologue (see Experimental Section 4.7.2 for details). According to docking, the binding mode of **42** observed in the crystal structure of SARS-CoV-2 3CL^{pro} is essentially maintained also in MERS-CoV 3CL^{pro} active site.

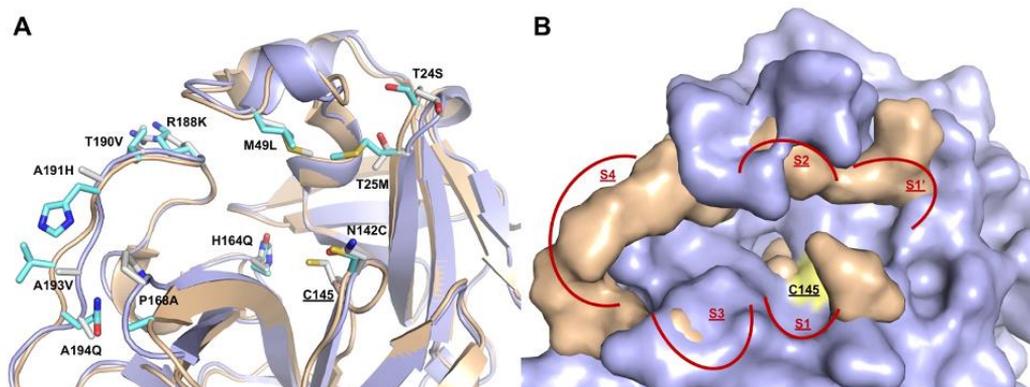


Figure 4.10. Superimposed active sites of 3CL^{pro} from SARS-CoV-2 (light blue, PDB 7BQY), and MERS (wheat, PDB 4RSP). (A) The amino acid mutations are shown as sticks (white and aquamarine for SARS-CoV-2 and MERS, respectively), and labeled according to SARS-CoV-2 numbering. (B) Surface representation of SARS-CoV-2 3CL^{pro} active site (light blue), with amino acid mutations from MERS-CoV (displayed as wheat surface). The S1–S4 and S1' subsites are indicated with red lines and labeled accordingly. The catalytic residue Cys¹⁴⁵ is highlighted in yellow. The S1', S1 and S2 subsites show mostly conservative substitutions, whereas the main differences involve the residues lining the S3–S4 region and the lid covering the S2, namely: the replacement of Thr25 in S1' with Met²⁵; Asn¹⁴² with Cys¹⁴⁵ in S1; His¹⁶⁴ with Gln¹⁶⁷ at the S1/S2 borders; Met⁴⁹ with Leu⁴⁹ in S2.

The docked pose of **42** displays the thiohemiacetal adduct in the *S* configuration, with the hydroxy group accepting two H-bonds from the backbone amide groups of the oxyanion hole residues Gly¹⁴⁶ and Cys¹⁴⁸ (Fig. 4.11A). The P1 γ -lactam moiety engages the canonical H-bonds with His¹⁶⁶ and Glu¹⁶⁹ side chains and with Phe¹⁴³ main chain within the S1 pocket, whereas the ligand backbone interacted with Gln¹⁶⁷ and Glu¹⁶⁹ main chain. The P2 fragment is deeply inserted into the S2 pocket, with the Leu⁴⁹ side chain clamping the MeO substituent through tight hydrophobic interactions. This latter, in turn, could potentially accept a H-bond from the side chain of Gln¹⁹², likely accounting for the higher potency of **42** also towards MERS 3CL^{pro} compared to compounds such as **40**, bearing an unsubstituted proline, and **39**, characterized by the P2 leucine. Finally, the Boc cap can establish further hydrophobic interactions with Leu¹⁷⁰ and Gln¹⁹⁵.

At variance with **42**, compound **46**, bearing a bicycloproline as P2 fragment, displayed a significant drop in potency towards MERS 3CL^{pro}. To rationalize these data, we also performed covalent docking calculations of this compound in the active site of the MERS enzyme. Notably, the presence of the bicycloproline caused unfavorable close contacts with Leu⁴⁹ and His⁴¹ (Fig. 4.11B). Thus, the replacement of Met⁴⁹ in SARS-CoV-2 with Leu in MERS exerts a certain impact on the shape and size of the S2 subsite, thereby influencing the inhibitory activity of functionalized P2 proline fragments. Indeed, also the lower potency of **41**, bearing a 4-methyl group on the proline ring could be due to the potential clashes that this substituent can form with the side chain of Leu⁴⁹. On the other hand, the presence of the oxygen atom at position 4 of the P2 proline rescues activity towards MERS, allowing the ligand i) to better orient the methyl **41** to form profitable hydrophobic interactions with Leu⁴⁹; ii) to potentially establish a H-bond with Gln¹⁹². Importantly, the *S* configuration ensures the optimal positioning of the substituents within the S2 pocket.

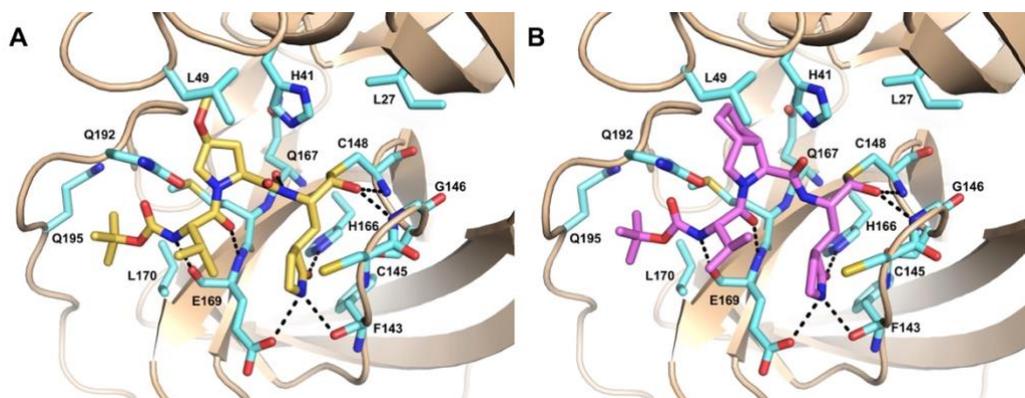


Figure 4.11. Predicted binding mode of **42** (A, yellow sticks) and **46** (B, violet sticks) into MERS 3CL^{pro} (PDB ID: 4RSP)¹⁵ active site represented as wheat ribbon model. Residues important for ligand binding are displayed as aquamarine sticks and labeled. H-bonds discussed in the text are depicted as dashed black lines.

4.6. Summary and conclusions

The heavy impact of COVID-19 pandemic rapidly boosted at an unprecedented speed the research on the disease, leading to the approval of vaccines and DAAs within a very short time frame, if compared to any other disease including viral infections. However, only three DAAs are available to date, all of which show limitations in efficacy and/or PK properties. An elective target to identify DAAs against SARS-CoV-2 is the 3CL^{pro}; indeed, many research groups reported peptidomimetic and non-peptide inhibitors of this enzyme. These gigantic efforts from big pharma, medium size company biotech and academia culminated in the development of Nirmatrelvir **12**, the first-in-class approved DAA targeting 3CL^{pro}. Compound **12** was designed as a proline-based peptidomimetic reversible covalent inhibitor, similar to the previously approved first-in-class α -ketoamide inhibitors of HCV NS3/4A serine protease, Telaprevir and Boceprevir.

In 2020, within the multicenter and multidisciplinary project Exscalate4CoV, funded by the European Commission, my research efforts were focused on the discovery of novel SARS-CoV-2 3CL^{pro} inhibitors, with the aim to identify effective and safe antiviral agents active against SARS-CoV-2.

In this regard, a series of proline-based tripeptides was designed and synthesized as potential reversible covalent inhibitors of SARS-CoV-2 3CL^{pro} (**39-48**), bearing an aldehyde warhead as cysteine trap and a cyclic analogue of glutamine in P1, similarly to other potent inhibitors reported in that period. In addition, a deep analysis of the X-ray structures of inhibitors, released at the beginning of the pandemic, led to the idea of replacing the P2 leucine of the natural substrate with a proline. Furthermore, the effect of different substitution patterns at the position 4 of pyrrolidine ring was investigated. In second instance, a series of di- and tripeptides (**50-53**) was designed and synthesized, by replacing the aldehyde

warhead with a nitrile functional group to achieve a reversible covalent inhibitory activity against 3CL^{pro}, similarly to Nirmatrelvir. Moreover, novel P2 derivatives were exploited on basis of the results of previous series of peptidomimetic inhibitors, and two P3 residues were introduced by applying a “mix and match” approach with P3 residues of known potent SARS-CoV-2 3CL^{pro} inhibitors, such as Nirmatrelvir and MI-30.

Notably, most of our newly synthesized compounds showed excellent inhibition of the SARS-CoV-2 3CL^{pro}, with some displaying IC₅₀ values in the sub-nM range. Nonetheless, the inhibitory potency was influenced by the size, the nature and the chirality of the substituents at the P2 position, indicating that:

- i) a 5-membered cyclic amino acid (i.e. proline) is preferred over a 6-membered residue (i.e. pipercolinic acid);
- ii) small substituents at position 4 of the proline ring are preferred in *S* configuration, while bigger moieties are better tolerated in *R* configuration;
- iii) linear derivative with a small aliphatic side chain and a stereogenic constriction center, such as methyl-allylglycine, is tolerated as well as proline.

Moreover, the inhibitory potency was also influenced by P3 residue. The introduction of *tert*-butyl glycine and 2,4-dichlorophenoxyacetic acid in P3 afford potent peptide inhibitors, even if the antiviral activity results showed that the *t*-butylglycine appears to be better tolerated in presence of proline derivative in P2, while P3 2,4-dichlorophenoxyacetyl is preferred when linear derivatives occupy the P2 position of peptide sequence.

Most of the active compound in the biochemical assays effectively inhibited SARS-CoV-2 replication in Vero cells, however showing a large shift in the antiviral potency without a clear correlation with the enzymatic activity. To be noted, this behavior is common to other peptidomimetic inhibitors reported in

this field, including Nirmatrelvir **12**, being such molecules substrates of cellular efflux pumps like P-gp. Indeed, compounds' potencies in cell-based assays improved when co-dosed with a P-gp inhibitor, with derivative **46** displaying the higher antiviral activity, in the nanomolar range. These results were further confirmed in human epithelial lung adenocarcinoma Calu-3 cells, a cell type not only more relevant to evaluate the infection of a respiratory virus such as SARS-CoV-2, but also characterized by low expression of efflux pumps. In addition to the 3CL^{pro} of SARS-CoV-2, the majority of the compounds presented high affinity towards the MERS homologue (IC₅₀ values ranged from mid- to sub-nM). In light of the high sequence and structural similarity of the main proteases within the *Coronaviridae* family, they could thus represent new leads for the development of broad-spectrum anti-coronavirus agents challenging useful in the current and in the future pandemics. This work showed that the exploiting of molecules bearing an aldehyde warhead can be advantageous in terms of target engagement kinetics for the formation of the reversible covalent adduct. Moreover, our derivatives, particularly compound **46**, demonstrated a good in-cell safety index, also superior to nitrile peptide, thus overcoming the toxicity warning related to the reactivity of aldehydes. Particularly, the most potent compound **46** (SI > 476) showed an in-cell safety index superior to the nitrile peptides **50** (SI > 45) and **53** (SI > 45).

These results constitute a promising starting point for physicochemical optimization, in order to obtain new derivatives with improved PK properties and provide suitable candidates for *in vivo* efficacy studies against COVID-19 and other CoV-related diseases.

4.7. Experimental section

4.7.1. Synthesis and characterization of compounds

General methods. Reagents and solvents were purchased from commercial sources and used without further purification. Reactions were carried out at ambient temperature, unless otherwise specified. Moisture-sensitive reactions were performed under a positive pressure of dry nitrogen in oven-dried glassware. Analytical thin-layer chromatography (TLC) on silica gel 60 F254 plates (250 μm thickness) was performed to monitor the reaction progress, using UV and KMnO_4 as revelation method. Flash chromatography on silica gel (70—230 mesh) was performed for purification. All products were characterized by their NMR and MS spectra. (ESI)-MS spectra were performed on a LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) by infusion into the ESI source using MeOH as solvent. ^1H and ^{13}C NMR spectra were recorded in CDCl_3 or $\text{DMSO-}d_6$ at 25°C on Bruker Avance NEO 400 MHz and 700 MHz instruments equipped with a RT-DR-BF/ ^1H -5mm-OZ SmartProbe. Chemical shifts (δ) are reported in part per million (ppm) downfield from tetramethylsilane, using residual solvent signal as the internal reference.

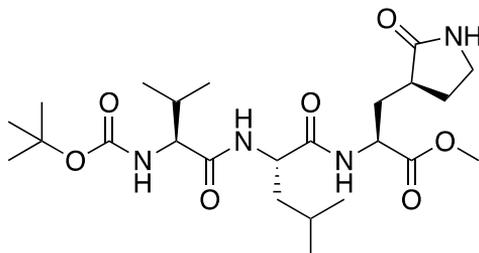
The final compounds were characterized by HPLC-MS/MS, using a Dionex ULTIMATE 3000 (Thermo Fisher Scientific) HPLC module and a LTQ XL mass spectrometer with electrospray ionization in positive mode and an Ion-Trap detector. Separation was performed with a Kinetex column C18 Polar column (250 mm \times 4.6 mm; particle size 5 μm , Phenomenex, Torrance, CA, USA) at 30°C , using a 17 min gradient, 5% [0.1% TFA/ CH_3CN]/95% [0.1% TFA/ H_2O] to 95% [0.1% TFA/ CH_3CN].

Analytical HPLC was performed on Shimadzu-1100 HPLC using a Kinetex C18 column (4.6 mm x 150 mm, 5 μ m, 100 Å) with an acetonitrile (0.1%TFA) – water (0.1% TFA) custom gradient. The purities of the final compounds were all > 95%, as determined by HPLC (UV λ = 220 and 254 nm).

A Jasco P-2000 digital polarimeter with a sodium lamp at 589 nm and a photomultiplier tube detector was used for the measurement of specific optical rotation ($[\alpha]_D$), using a 100 mm cell, sample concentration 1 mg/1 mL in MeOH, at 25°C.

Synthesis of target compounds and intermediates

Methyl (6S,9S,12S)-9-isobutyl-6-isopropyl-2,2-dimethyl-4,7,10-trioxo-12-(((S)-2-oxopyrrolidin-3-yl)methyl)-3-oxa-5,8,11-triazatridecan-13-oate (56).

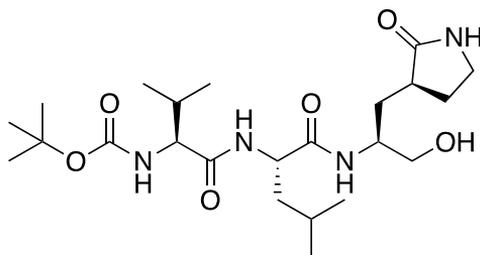


Chemical Formula: C₂₄H₄₂N₄O₇
Molecular Weight: 498.62

To a solution of Boc-L-Val-OH (189 mg, 0.870 mmol) in anhydrous DMF (2 mL) in a three-necked oven-dried flask, EDCI (208.5 mg, 1.09 mmol, 1.2 eq.) and HOBt (167 mg, 1.09 mmol, 1.2 eq.) were added under a positive anhydrous nitrogen pressure, and the mixture was left stirring for 30 min. In a separate flask, a solution of amine hydrochloride intermediate **55**¹⁷ (292 mg, 0.870 mmol, 1 eq.) in anhydrous DMF (3 mL) was cooled to 0-5°C and DIPEA (0.61 mL, 3.48 mmol, 4 eq.) was added dropwise. After 30 min, this solution was added to the reaction mixture containing activated intermediate of **54** and was left stirring for 16 h at room temperature. The reaction mixture was diluted with brine and

extracted with EtOAc (25 mL x 4). The collected organic layers were washed with brine (15 mL x 10), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by column chromatography (CH₂Cl₂/MeOH, 97:3 v/v) to afford the pure product **56** (138 mg, 40%) as a white solid. ¹H NMR (400 MHz, CDCl₃): δ 7.79 (d, *J* = 6.3 Hz, 1H), 6.94 (d, *J* = 8.1 Hz, 1H), 6.46 (s, 1H), 5.03 (d, *J* = 8.7 Hz, 1H), 4.64 (td, *J* = 9.1, 4.7 Hz, 1H), 4.52 (s, 1H), 3.90 – 3.81 (m, 1H), 3.72 (s, 3H), 3.33 (dd, *J* = 13.3, 7.7 Hz, 2H), 2.39 (dd, *J* = 12.1, 5.5 Hz, 2H), 2.25 – 2.05 (m, 2H), 1.94 – 1.78 (m, 2H), 1.75 – 1.61 (m, 2H), 1.55 – 1.46 (m, 1H), 1.43 (s, 9H), 1.00-0.90 (m, 12H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₄H₄₃N₄O₇⁺ 499.3, found 499.2.

tert-Butyl ((*S*)-1-(((*S*)-1-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)carbamate (**57**).

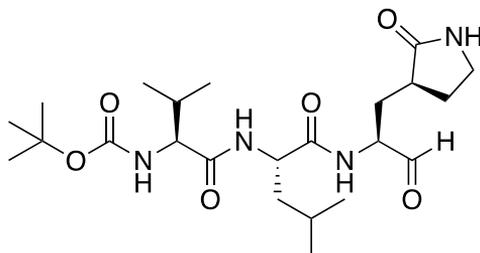


Chemical Formula: C₂₃H₄₂N₄O₆
Molecular Weight: 470.61

In oven-dried three-necked round bottom flask, the tripeptide methyl ester **56** (138 mg, 0.277 mmol) was dissolved in anhydrous THF (1.6 mL) under a positive anhydrous nitrogen pressure, then a 2 M solution of LiBH₄ (18 mg, 0.830 mmol, 3 eq.) in anhydrous THF (0.415 mL) was added dropwise at 0°C. The reaction mixture was stirred at room temperature for 5 h, then quenched by acidifying with a saturated citric acid solution to pH 2 and extracted with EtOAc (25 mL x 3). The collected organic layers were washed with brine (15 mL x 10), dried over

anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by column chromatography (CH₂Cl₂/MeOH, 98:2 to 95:5 v/v) to afford the desired alcohol **57** (82 mg, 63%) as a white solid. ¹H NMR (400 MHz, CDCl₃): δ 7.65 (d, *J* = 7.0 Hz, 1H), 7.13 (d, *J* = 7.7 Hz, 1H), 6.43 (s, 1H), 5.12 (d, *J* = 8.4 Hz, 1H), 4.56 (td, *J* = 9.0, 4.7 Hz, 1H), 4.01 (s, 1H), 3.92 (t, *J* = 7.4 Hz, 1H), 3.62 (d, *J* = 15.7 Hz, 2H), 3.45 – 3.24 (m, 3H), 2.38 (d, *J* = 5.7 Hz, 2H), 2.07 (dt, *J* = 23.8, 8.6 Hz, 2H), 1.90 – 1.73 (m, 3H), 1.71 – 1.57 (m, 3H), 1.55 – 1.49 (m, 1H), 1.43 (s, 9H), 0.99 – 0.86 (m, 12H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₃H₄₃N₄O₆⁺ 471.3, found 471.3.

tert-Butyl((*S*)-3-methyl-1-(((*S*)-4-methyl-1-oxo-1-(((*S*)-1-oxo-3-(((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)pentan-2-yl)amino)-1-oxobutan-2-yl)carbamate (**39**).



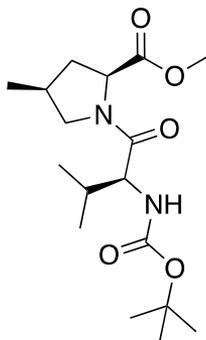
Chemical Formula: C₂₃H₄₀N₄O₆
Molecular Weight: 468.60

Compound **57** (82 mg, 0.174 mmol) was dissolved in anhydrous DMSO (0.06 mL, 0.870 mmol, 5 eq.), anhydrous CH₂Cl₂ (0.43 mL) under a positive anhydrous nitrogen pressure, and DIPEA (0.106 mL, 0.609 mmol, 3.5 eq.) was added at -5°C. In another flask, pyridine sulfur trioxide complex (55.4 mg, 0.348 mmol, 3 eq.) and pyridine (0.028 mL, 0.384 mmol, 3 eq.) were suspended in anhydrous DMSO (0.06 mL, 0.870 mmol, 7.5 eq.) at room temperature under stirring for 10 min, and then the resulting mixture was added to the previously formed **57**/DMSO mixture at -5°C. After 24 h, the reaction mixture was poured into

ice/water (1:1) and extracted with CH₂Cl₂ (25 mL x 3). The collected organic layers were washed with a saturated citric acid solution (15 mL x 1), water (15 mL x 1), saturated solution of NaHCO₃ (15 mL x 1) and brine (15 mL x 1), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by column chromatography (CH₂Cl₂/MeOH, 95:5 v/v) to afford the target aldehyde **39** (17 mg, 21%) as a white solid. $[\alpha]_{\text{D}}^{20}$ -50.0 (c = 1.0, MeOH). ¹H NMR (700 MHz, CDCl₃): δ 9.50 (s, 1H), 8.22 (s, 1H), 6.77 (d, *J* = 7.3 Hz, 1H), 6.32 (s, 1H), 5.05 (d, *J* = 8.6 Hz, 1H), 4.63 (td, *J* = 8.9, 4.9 Hz, 1H), 4.35 (s, 1H), 3.93 – 3.86 (m, 1H), 3.41 – 3.29 (m, 2H), 2.49 – 2.30 (m, 2H), 2.13 (dd, *J* = 13.4, 6.7 Hz, 1H), 2.00 (ddd, *J* = 14.5, 10.5, 6.6 Hz, 1H), 1.92 (ddd, *J* = 14.4, 6.6, 4.4 Hz, 1H), 1.87 – 1.78 (m, 2H), 1.75 – 1.61 (m, 1H), 1.60 – 1.55 (m, 1H), 1.43 (s, 9H), 1.01 – 0.88 (m, 12H). ¹³C NMR (176 MHz, CDCl₃): δ 199.7, 180.1, 173.3, 171.9, 156.2, 129.9, 80.3, 60.4, 57.7, 51.8, 42.0, 40.7, 38.2, 30.7, 29.9, 28.6, 24.9, 23.1, 22.0, 19.4, 18.0. LC-MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₃H₄₁N₄O₆⁺ 469.3, found 469.2 (*r*_t: 7.07).

General procedure for the synthesis of dipeptide ester intermediates 61a-c, 61f, 77a, 77b (Method A). In a three-necked oven-dried round-bottom flask, *N*-Boc-Valine **54** (1 eq) was dissolved in anhydrous CH₂Cl₂ (0.3 M) under a positive anhydrous nitrogen pressure, then HBTU (1.1 eq.), the appropriate amine hydrochloride (1.3 eq.) and DIPEA (2 eq.) were added at 0°C. The reaction mixture was stirred for 4 h at room temperature. A saturated aqueous solution of NaHCO₃ was added to the reaction mixture and the resulting aqueous mixture was extracted with CH₂Cl₂ (25 mL x 3), the collected organic layers were washed with a saturated citric acid solution (15 mL x 1), brine (15 mL x 1), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a colorless oil.

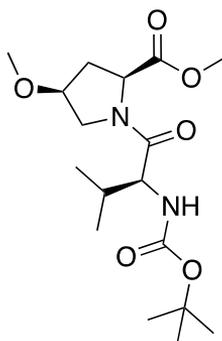
Methyl(2S,4S)-1-((tert-butoxycarbonyl)-l-valyl)-4-methylpyrrolidine-2-carboxylate (61a).



Chemical Formula: C₁₇H₃₀N₂O₅
Molecular Weight: 342.44

Compound **61a** was prepared following the general procedure Method A and using methyl (2*S*,4*S*)-4-methylpyrrolidine-2-carboxylate hydrochloride **58a** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 149 mg, 45 %). ¹H NMR (400 MHz, CDCl₃): δ 5.22 (d, *J* = 9.1 Hz, 1H), 4.41 (dd, *J* = 9.7, 7.6 Hz, 1H), 4.29 (dd, *J* = 9.3, 6.0 Hz, 1H), 4.03 – 3.89 (m, 1H), 3.71 (d, *J* = 3.7 Hz, 3H), 3.12 (dd, *J* = 16.6, 6.7 Hz, 1H), 2.41 (dd, *J* = 12.2, 6.6 Hz, 1H), 2.33 (d, *J* = 6.6 Hz, 1H), 2.02 (dd, *J* = 13.4, 6.4 Hz, 1H), 1.56 – 1.48 (m, 1H), 1.43 (s, 9H), 1.09 (d, *J* = 6.4 Hz, 3H), 1.02 (d, *J* = 6.8 Hz, 3H), 0.96 – 0.89 (m, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₁₇H₃₁N₂O₅⁺ 343.2, found 343.1.

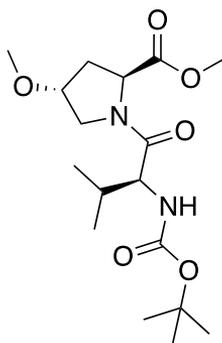
Methyl(2S,4S)-1-((tert-butoxycarbonyl)-l-valyl)-4-methoxypyrrolidine-2-carboxylate (61b).



Chemical Formula: $C_{17}H_{30}N_2O_6$
Molecular Weight: 358.44

Compound **61b** was prepared following the general procedure Method A and using methyl (2*S*,4*S*)-4-methoxypyrrolidine-2-carboxylate hydrochloride **58b** (purification method: flash chromatography eluting $CH_2Cl_2/MeOH$ 98:2 v/v; yield: 1.45 g, 70 %). 1H NMR (400 MHz, $CDCl_3$): δ 5.20 (d, $J = 9.4$ Hz, 1H), 4.70 (dd, $J = 8.8, 3.8$ Hz, 1H), 4.22 (dd, $J = 9.3, 6.3$ Hz, 1H), 3.99 (ddd, $J = 16.1, 9.3, 4.3$ Hz, 1H), 3.69 (s, 3H), 3.59 (dd, $J = 10.6, 2.8$ Hz, 1H), 3.28 (s, 3H), 2.37 – 2.28 (m, 1H), 2.25 – 2.14 (m, 2H), 2.08 – 1.98 (m, 1H), 1.42 (s, 9H), 1.05 (d, $J = 6.8$ Hz, 3H), 0.96 (d, $J = 6.7, 2.4$ Hz, 3H). MS (ESI) m/z : $[M+H]^+$ calcd for $C_{17}H_{31}N_2O_6^+$ 359.2, found 359.1.

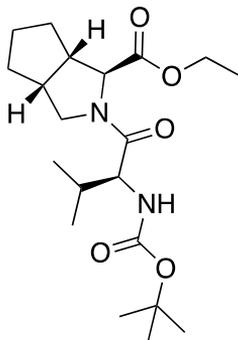
Methyl(2S,4R)-1-((tert-butoxycarbonyl)-l-valyl)-4-methoxypyrrolidine-2-carboxylate (61c).



Chemical Formula: C₁₇H₃₀N₂O₆
Molecular Weight: 358.44

Compound **61c** was prepared following the general procedure Method A and using methyl (2*S*,4*R*)-4-methoxypyrrolidine-2-carboxylate hydrochloride **58c** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 98:2 v/v; yield: 986 mg, 74%). ¹H NMR (400 MHz, CDCl₃): δ 5.18 (d, *J* = 9.4 Hz, 1H), 4.69 (dd, *J* = 8.8, 3.8 Hz, 1H), 4.21 (dd, *J* = 9.3, 6.3 Hz, 1H), 4.01 (ddd, *J* = 16.1, 9.3, 4.3 Hz, 1H), 3.68 (s, 3H), 3.60 (dd, *J* = 10.6, 2.8 Hz, 1H), 3.28 (s, 3H), 2.36 – 2.27 (m, 1H), 2.25 – 2.14 (m, 2H), 2.09 – 1.98 (m, 1H), 1.42 (d, *J* = 5.3 Hz, 9H), 1.06 (d, *J* = 6.8 Hz, 3H), 0.96 (dd, *J* = 6.7, 2.4 Hz, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₁₇H₃₁N₂O₆⁺ 359.2, found 359.1.

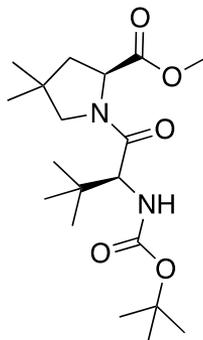
Ethyl(1S,3aR,6aS)-2-((tert-butoxycarbonyl)-l-valyl)octahydrocyclopenta[c]pyrrole-1-carboxylate (61f).



Chemical Formula: $C_{20}H_{34}N_2O_5$
Molecular Weight: 382.50

The compound **61f** was prepared following the general procedure Method A and using ethyl (1S,3aR,6aS)-octahydrocyclopenta[c]pyrrole-1-carboxylate hydrochloride **58f** (purification method: flash chromatography eluting $CH_2Cl_2/MeOH$ 97:3 v/v; yield: 580 mg, 87 %). 1H NMR (400 MHz, $CDCl_3$): δ 5.21 (d, $J = 9.3$ Hz, 1H), 4.35 (d, $J = 3.9$ Hz, 1H), 4.27 (dd, $J = 9.3, 6.2$ Hz, 1H), 4.17 (q, $J = 7.1$ Hz, 2H), 3.81 (dd, $J = 10.3, 7.7$ Hz, 1H), 3.66 (dd, $J = 10.4, 3.5$ Hz, 1H), 2.74 (dd, $J = 7.7, 3.8$ Hz, 1H), 2.65 (dd, $J = 8.0, 3.9$ Hz, 1H), 2.09 – 1.81 (m, 3H), 1.75 (dd, $J = 12.2, 6.1$ Hz, 1H), 1.68 – 1.56 (m, 3H), 1.39 (s, 9H), 1.26 (t, $J = 7.1$ Hz, 3H), 1.01 (d, $J = 6.8$ Hz, 3H), 0.91 (d, $J = 6.7$ Hz, 3H). MS (ESI) m/z: $[M + H]^+$ calcd for $C_{20}H_{35}N_2O_5^+$ 383.2, found 383.1.

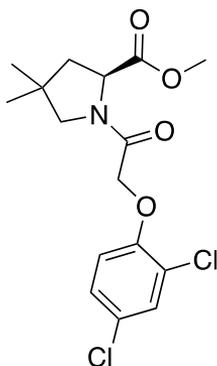
Methyl(S)-1-((S)-2-((tert-butoxycarbonyl)amino)-3,3-dimethylbutanoyl)-4,4-dimethylpyrrolidine-2-carboxylate (77a).



Chemical Formula: C₁₉H₃₄N₂O₅
Molecular Weight: 370.49

The compound **77a** was prepared following the general procedure Method A, using ethyl (*S*)-4,4-dimethylpyrrolidine-2-carboxylate hydrochloride **76** and (*S*)-2-((*tert*-butoxycarbonyl)amino)-3,3-dimethylbutanoic acid **75a** (purification method: flash chromatography eluting Hex/EtOAc 40:60 v/v; yield: 544 mg, 96%). ¹H NMR (700 MHz, CDCl₃): δ 5.23 (d, *J* = 9.9 Hz, 1H), 4.49 (t, *J* = 8.7 Hz, 1H), 4.26 (d, *J* = 10.0 Hz, 1H), 4.18 (q, *J* = 7.1 Hz, 2H), 3.68 (d, *J* = 9.7 Hz, 1H), 3.34 (d, *J* = 9.8 Hz, 1H), 2.05 – 2.03 (m, 1H), 1.72 (dd, *J* = 12.3, 9.6 Hz, 1H), 1.41 (s, 9H), 1.26 (t, *J* = 3.5 Hz, 3H), 1.22 (s, 3H), 1.14 (s, 3H), 1.02 (s, 9H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₁₉H₃₅N₂O₅⁺ 371.50, found 371.5.

Methyl(S)-1-(2-(2,4-dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxylate (77b).



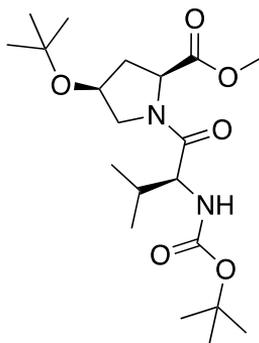
Chemical Formula: $C_{16}H_{19}Cl_2NO_4$
Molecular Weight: 360.23

The compound **77b** was prepared following the general procedure Method A, using ethyl (*S*)-4,4-dimethylpyrrolidine-2-carboxylate hydrochloride **76** and 2-(2,4-dichlorophenoxy)acetic acid **67** (yield: 233 mg, 94%). 1H NMR (700 MHz, $CDCl_3$): δ 7.36 (d, $J = 2.3$ Hz, 1H), 7.14 (dd, $J = 8.8, 2.3$ Hz, 1H), 6.89 (d, $J = 8.8$ Hz, 1H), 4.72 (d, $J = 6.8$ Hz, 2H), 4.53 (d, $J = 8.4$ Hz, 1H), 4.20 (q, $J = 7.1$ Hz, 2H), 3.54 (d, $J = 10.3$ Hz, 1H), 3.36 (d, $J = 10.3$ Hz, 1H), 2.05 (dd, $J = 12.0, 9.0$ Hz, 1H), 1.90 – 1.84 (m, 1H), 1.25 (t, $J = 3.5$ Hz, 3H), 1.16 (s, 3H), 1.04 (s, 3H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{16}H_{20}Cl_2NO_4^+$ 361.24, found 361.2.

General procedure for the synthesis of dipeptide ester intermediates 61d and 61e (Method B). In an oven-dried three-necked round-bottom flask, the dipeptidyl alcohol (1 eq.) and $Yb(OTf)_3$ (0.2 eq.) were left stirring in dry CH_2Cl_2 (0.1 M) under anhydrous nitrogen atmosphere until complete dissolution. Subsequently, di-*tert*-butyl carbonate (3.3 eq.) was added and the reaction was stirred at reflux for 24 h. The mixture reaction was quenched with distilled water, filtered over celite, and the filtrate was extracted with CH_2Cl_2 (25 mL x 3). The collected organic layers were washed with brine (15 mL x 1), dried over anhydrous Na_2SO_4 , filtered and concentrated under vacuum. The reaction crude

was purified by flash chromatography to yield the desired product as a white solid.

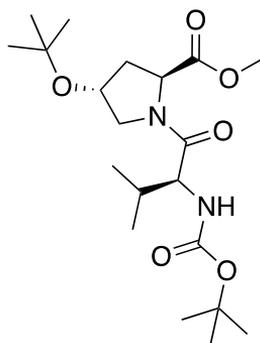
Methyl(2S,4S)-4-(tert-butoxy)-1-((tert-butoxycarbonyl)-l-valyl)pyrrolidin-2-carboxylate (61d).



Chemical Formula: $C_{20}H_{36}N_2O_6$
Molecular Weight: 400.52

Compound **61d** was prepared following the general procedure Method B and using methyl-(2*S*,4*S*)-1-((*tert*-butoxycarbonyl)-*l*-valyl)-4-hydroxypyrrolidine-2-carboxylate **59** (purification method: flash chromatography eluting $CHCl_3/MeOH$ 99:1 v/v; yield: 384 mg, 40%). 1H NMR (400 MHz, $CDCl_3$): δ 5.26 – 5.06 (m, 1H), 4.74 (dd, $J = 8.7, 3.9$ Hz, 1H), 4.49 (t, $J = 7.8$ Hz, 1H), 4.12 (dd, $J = 11.5, 5.7$ Hz, 1H), 3.95 (dd, $J = 10.0, 6.5$ Hz, 1H), 3.71 (s, 3H), 3.36 (dd, $J = 10.1, 6.4$ Hz, 1H), 2.47 – 2.31 (m, 2H), 2.09 – 1.91 (m, 1H), 1.45 (s, 9H), 1.17 (s, 9H), 0.95 (d, $J = 6.8$ Hz, 3H), 0.91 (d, $J = 6.7$ Hz, 3H). MS(ESI) m/z : $[M + H]^+$ calcd for $C_{20}H_{37}N_2O_6^+$ 401.3, found 401.2.

Methyl(2S,4R)-4-(tert-butoxy)-1-((tert-butoxycarbonyl)-l-valyl)pyrrolidin-2-carboxylate (61e).

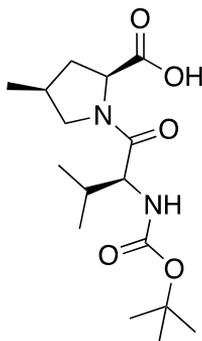


Chemical Formula: $C_{20}H_{36}N_2O_6$
Molecular Weight: 400.52

Compound **61e** was prepared following the general procedure Method B and using methyl-(2*S*,4*R*)-1-((*tert*-butoxycarbonyl)-*l*-valyl)-4-hydroxypyrrolidine-2-carboxylate **60** (purification method: flash chromatography eluting $CHCl_3/MeOH$ 99:1 v/v; yield: 516 mg, 46 %). 1H NMR (400 MHz, $CDCl_3$): δ 5.28 – 5.04 (m, 1H), 4.73 (dd, $J = 8.7, 3.9$ Hz, 1H), 4.50 (t, $J = 7.8$ Hz, 1H), 4.11 (dd, $J = 11.5, 5.7$ Hz, 1H), 3.93 (dd, $J = 10.0, 6.5$ Hz, 1H), 3.69 (s, 3H), 3.35 (dd, $J = 10.1, 6.4$ Hz, 1H), 2.48 – 2.31 (m, 2H), 2.10 – 1.92 (m, 1H), 1.44 (s, 9H), 1.15 (s, 9H), 0.95 (d, $J = 6.8$ Hz, 3H), 0.92 (d, $J = 6.7$ Hz, 3H). MS(ESI) m/z : $[M + H]^+$ calcd for $C_{20}H_{37}N_2O_6^+$ 401.3, found 401.2.

General procedure for the synthesis of acid derivatives 62a-f, 65, 78a, 78b (Method C). A solution of 1 N aq. LiOH (2 eq.) was added dropwise to a solution of ester intermediate (1 eq.) in THF (0.2 M) at 0°C, and the resulting mixture (1:3) was stirred for 3 h at room temperature. The reaction mixture was acidified to pH 3 with a saturated solution of citric acid and extracted with EtOAc (25 mL x 4). The collected organic layers were dried over anhydrous Na_2SO_4 , filtered and concentrated under vacuum to quantitatively yield the product as a white solid, used in the next reaction step without further purification.

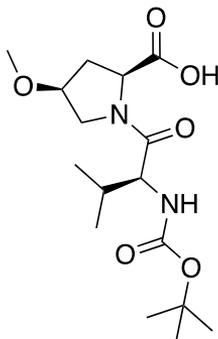
(2*S*,4*S*)-1-((*tert*-Butoxycarbonyl)-*l*-valyl)-4-methylpyrrolidine-2-carboxylic acid (**62a**).



Chemical Formula: C₁₆H₂₈N₂O₅
Molecular Weight: 328.41

Compound **62a** was prepared following the general procedure Method C and using ester intermediate **61a** (yield: 130 mg) ¹H NMR (400 MHz, CDCl₃): δ 5.23 (d, *J* = 8.9 Hz, 1H), 4.53 (t, *J* = 8.5 Hz, 1H), 4.31 (dd, *J* = 9.0, 6.4 Hz, 1H), 4.07 – 3.97 (m, 1H), 3.08 (dd, *J* = 13.4, 6.2 Hz, 1H), 2.43 – 2.32 (m, 2H), 1.99 (dd, *J* = 13.4, 6.7 Hz, 1H), 1.90 (dd, *J* = 21.3, 9.8 Hz, 1H), 1.41 (s, 9H), 1.11 (d, *J* = 6.3 Hz, 3H), 0.98 (d, *J* = 6.6 Hz, 3H), 0.94 (s, 3H). MS (ESI) *m/z*: [M]⁻ calcd for C₁₆H₂₇N₂O₅⁻ 327.2, found 327.1.

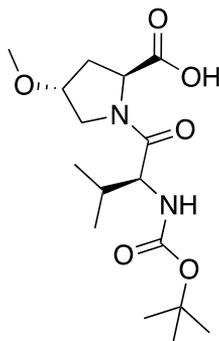
(2*S*,4*S*)-1-((*tert*-Butoxycarbonyl)-*l*-valyl)-4-methoxypyrrolidine-2-carboxylic acid (**62b**).



Chemical Formula: C₁₆H₂₈N₂O₆
Molecular Weight: 344.41

Compound **62b** was prepared following the general procedure Method C and using ester intermediate **61b** (yield: 1.36 g) ^1H NMR (400 MHz, CDCl_3): δ 5.22 (d, $J = 9.2$ Hz, 1H), 4.71 (dd, $J = 9.2, 2.8$ Hz, 1H), 4.22 (dd, $J = 9.0, 6.9$ Hz, 1H), 3.96 (dd, $J = 10.8, 4.9$ Hz, 1H), 3.64 (d, $J = 10.4$ Hz, 1H), 3.33 (d, $J = 7.7$ Hz, 3H), 2.60 (d, $J = 13.7$ Hz, 1H), 2.15 (ddd, $J = 13.8, 9.3, 4.6$ Hz, 1H), 2.09 – 1.97 (m, 2H), 1.41 (s, 9H), 1.02 (d, $J = 6.8$ Hz, 3H), 0.96 (d, $J = 6.8$ Hz, 3H). MS (ESI) m/z : $[\text{M}]^-$ calcd for $\text{C}_{16}\text{H}_{27}\text{N}_2\text{O}_6^-$ 343.2, found 343.3.

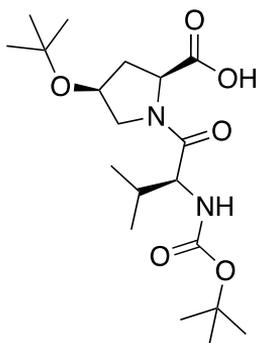
(2*S*,4*R*)-1-((*tert*-Butoxycarbonyl)-*l*-valyl)-4-methoxypyrrolidine-2-carboxylic acid (**62c**).



Chemical Formula: $\text{C}_{16}\text{H}_{28}\text{N}_2\text{O}_6$
Molecular Weight: 344.41

Compound **62c** was prepared following the general procedure Method C and using ester intermediate **61c** (yield: 1.36 g) ^1H NMR (400 MHz, CDCl_3): δ 5.25 (d, $J = 9.2$ Hz, 1H), 4.69 (dd, $J = 9.2, 2.8$ Hz, 1H), 4.25 (dd, $J = 9.0, 6.9$ Hz, 1H), 3.94 (dd, $J = 10.8, 4.9$ Hz, 1H), 3.65 (d, $J = 10.4$ Hz, 1H), 3.34 (d, $J = 7.7$ Hz, 3H), 2.58 (d, $J = 13.7$ Hz, 1H), 2.16 (ddd, $J = 13.8, 9.3, 4.6$ Hz, 1H), 2.11 – 1.95 (m, 2H), 1.40 (s, 9H), 1.02 (d, $J = 6.8$ Hz, 3H), 0.95 (d, $J = 6.8$ Hz, 3H). MS (ESI) m/z : $[\text{M}]^-$ calcd for $\text{C}_{16}\text{H}_{27}\text{N}_2\text{O}_6^-$ 343.2, found 343.3.

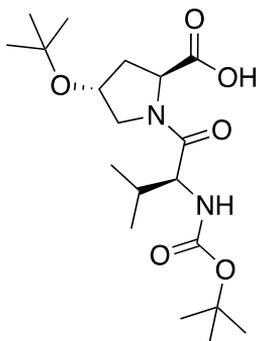
(2*S*,4*S*)-4-(*tert*-Butoxy)-1-((*tert*-butoxycarbonyl)-*l*-valyl)pyrrolidin-2-carboxylic acid (**62d**).



Chemical Formula: C₁₉H₃₄N₂O₆
Molecular Weight: 386.49

Compound **62d** was prepared following the general procedure Method C and using ester intermediate **61d** (yield: 300 mg) ¹H NMR (700 MHz, CDCl₃): δ 5.30 – 5.23 (m, 1H), 4.53 (dd, *J* = 8.4, 7.1 Hz, 1H), 4.29 – 4.21 (m, 1H), 3.97 (dd, *J* = 10.2, 6.1 Hz, 1H), 3.41 (dd, *J* = 10.2, 6.0 Hz, 1H), 2.39 (ddd, *J* = 14.1, 10.7, 6.3 Hz, 2H), 2.14 (dt, *J* = 13.1, 6.6 Hz, 1H), 2.04 – 1.99 (m, 1H), 1.42 (s, 9H), 1.20 (s, 9H), 1.00 (d, *J* = 6.6 Hz, 3H), 0.92 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M]⁻ calcd for C₁₉H₃₃N₂O₆⁻ 385.2, found 385.3.

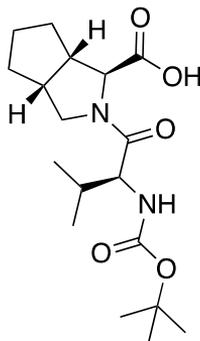
(2*S*,4*R*)-4-(*tert*-Butoxy)-1-((*tert*-butoxycarbonyl)-*l*-valyl)pyrrolidin-2-carboxylic acid (**62e**).



Chemical Formula: C₁₉H₃₄N₂O₆
Molecular Weight: 386.49

Compound **62e** was prepared following the general procedure Method C and using ester intermediate **61e** (yield: 497 mg). ¹H NMR (700 MHz, CDCl₃): δ 5.31 – 5.22 (m, 1H), 4.54 (dd, *J* = 8.4, 7.1 Hz, 1H), 4.30 – 4.19 (m, 1H), 3.98 (dd, *J* = 10.2, 6.1 Hz, 1H), 3.39 (dd, *J* = 10.2, 6.0 Hz, 1H), 2.35 (ddd, *J* = 14.1, 10.7, 6.3 Hz, 2H), 2.16 (dt, *J* = 13.1, 6.6 Hz, 1H), 2.06 – 1.97 (m, 1H), 1.41 (s, 9H), 1.22 (s, 9H), 1.01 (d, *J* = 6.6 Hz, 3H), 0.94 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M]⁻ calcd for C₁₉H₃₃N₂O₆⁻ 385.2, found 385.3.

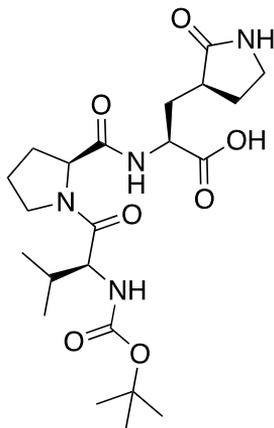
(1*S*,3*aR*,6*aS*)-2-((*tert*-Butoxycarbonyl)-*l*-valyl)octahydrocyclopenta[*c*]pyrrole-1-carboxylic acid (**62f**).



Chemical Formula: C₁₈H₃₀N₂O₅
Molecular Weight: 354.45

Compound **62f** was prepared following the general procedure Method C and using ester intermediate **61f** (yield: 531 mg). ¹H NMR (400 MHz, CDCl₃): δ 5.25 (d, *J* = 9.5 Hz, 1H), 4.41 (d, *J* = 3.5 Hz, 1H), 4.26 (dd, *J* = 9.1, 7.0 Hz, 1H), 3.82 – 3.69 (m, 1H), 2.92 – 2.84 (m, 1H), 2.79 (dd, *J* = 7.5, 4.0 Hz, 1H), 2.05 – 1.95 (m, 2H), 1.92 – 1.82 (m, 1H), 1.81 – 1.71 (m, 1H), 1.68 – 1.47 (m, 4H), 1.41 (s, 9H), 0.97 (d, *J* = 6.7 Hz, 3H), 0.92 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M calcd]⁻ for C₁₈H₂₉N₂O₅⁻ 353.2, found 353.3.

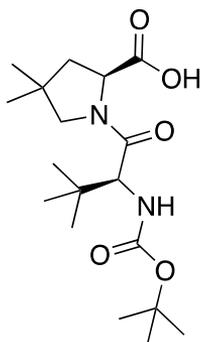
(*S*)-2-((*S*)-1-((*tert*-Butoxycarbonyl)-*l*-valyl)pyrrolidine-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoic acid (**65**).



Chemical Formula: $C_{22}H_{36}N_4O_7$
Molecular Weight: 468.55

Compound **65** was prepared following the general procedure Method C and using ester intermediate **64g** (yield: 180 mg) 1H NMR (400 MHz, DMSO- d_6): δ 8.19 (s, 1H), 7.54 (s, 1H), 6.79 (d, $J = 8.1$ Hz, 1H), 4.30 (d, $J = 28.8$ Hz, 1H), 3.99 (d, $J = 8.1$ Hz, 1H), 3.70 (s, 1H), 3.56 – 3.52 (m, 1H), 3.12 – 3.04 (m, 2H), 2.10 – 1.73 (m, 8H), 1.55 (d, $J = 11.2$ Hz, 2H), 1.36 (s, 9H), 0.89 (d, $J = 6.2$ Hz, 3H), 0.84 (d, $J = 6.1$ Hz, 3H). MS (ESI) m/z : $[M]^-$ calcd for $C_{22}H_{35}N_4O_7^-$ 467.2, found 467.2.

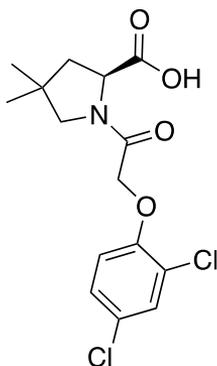
(*S*)-1-((*S*)-2-((*tert*-Butoxycarbonyl)amino)-3,3-dimethylbutanoyl)-4,4-dimethylpyrrolidine-2-carboxylic acid (**78a**).



Chemical Formula: C₁₈H₃₂N₂O₅
Molecular Weight: 356.46

The compound **78a** was prepared following the general procedure Method C and using methyl (*S*)-1-((*S*)-2-((*tert*-butoxycarbonyl)amino)-3,3-dimethylbutanoyl)-4,4-dimethylpyrrolidine-2-carboxylate **77a** (yield: 547 mg). ¹H NMR (700 MHz, CDCl₃): δ 5.23 (d, *J* = 9.9 Hz, 1H), 4.49 (t, *J* = 8.7 Hz, 1H), 4.26 (d, *J* = 10.0 Hz, 1H), 4.18 (q, *J* = 7.1 Hz, 2H), 3.68 (d, *J* = 9.7 Hz, 1H), 3.34 (d, *J* = 9.8 Hz, 1H), 2.05 – 2.03 (m, 1H), 1.72 (dd, *J* = 12.3, 9.6 Hz, 1H), 1.41 (s, 9H), 1.26 (t, *J* = 3.5 Hz, 3H), 1.22 (s, 3H), 1.14 (s, 3H), 1.02 (s, 9H). MS (ESI) *m/z*: [M - H]⁻ calcd for C₁₈H₃₁N₂O₅⁻ 355.46, found 355.5.

(*S*)-1-(2-(2,4-Dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxylic acid (**78b**).



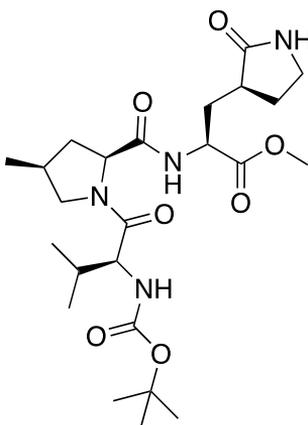
Chemical Formula: C₁₅H₁₇Cl₂NO₄
Molecular Weight: 346.20

The compound **78b** was prepared following the general procedure Method C and using methyl (*S*)-1-(2-(2,4-dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxylate **77b** (yield: 212 mg). ¹H NMR (700 MHz, CDCl₃): δ 7.36 (d, *J* = 2.3 Hz, 1H), 7.14 (dd, *J* = 8.8, 2.3 Hz, 1H), 6.89 (d, *J* = 8.8 Hz, 1H), 4.72 (d, *J* = 6.8 Hz, 2H), 4.53 (d, *J* = 8.4 Hz, 1H), 3.54 (d, *J* = 10.3 Hz, 1H), 3.36 (d, *J* = 10.3 Hz, 1H), 2.05 (dd, *J* = 12.0, 9.0 Hz, 1H), 1.90 – 1.84 (m, 1H), 1.16 (s, 3H), 1.04 (s, 3H). MS (ESI) *m/z*: [M - H]⁻ calcd for C₁₅H₁₆Cl₂NO₄⁻ 345.20, found 345.2.

General procedure for the convergent synthesis of tripeptide intermediates 64a-i (Method D). To a solution of the appropriate dipeptide acid intermediate (1 eq.) in anhydrous DMF (0.4 M) in an oven-dried three-necked round-bottom flask, EDCI (1.2 eq.) and HOBT (1.2 eq.) were added under a positive anhydrous nitrogen pressure, and the reaction mixture was left under stirring at room temperature for 30 min. In a separate flask, a solution of amine hydrochloride **63**²²⁸ (1 eq.) in anhydrous DMF (0.3 M) was cooled to 0-5 °C and DIPEA (4 eq.) was added dropwise. After 30 min, this solution was added to the reaction mixture containing the activated acid, and the resulting mixture was left under stirring at room temperature for 16 h. The resulting mixture was washed with

brine and extracted with EtOAc (25 mL x 4). The collected organic layers were washed with brine (15 mL x 10), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a white solid.

Methyl-(S)-2-((2S,4S)-1-((tert-butoxycarbonyl)-l-valyl)-4-methylpyrrolidine-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (64a).

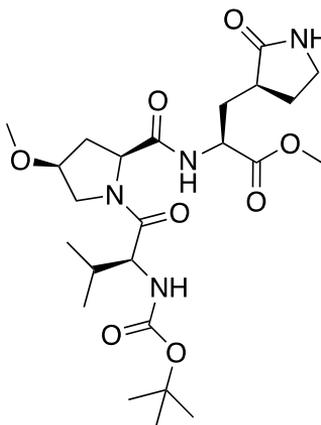


Chemical Formula: C₂₄H₄₀N₄O₇
Molecular Weight: 496.61

Compound **64a** was prepared following the general procedure Method D and using (2*S*,4*S*)-1-((*tert*-butoxycarbonyl)-*l*-valyl)-4-methylpyrrolidine-2-carboxylic acid **62a** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 97:3 v/v; yield: 72 mg, 39 %). ¹H NMR (400 MHz, CDCl₃): δ 7.51 (d, *J* = 7.4 Hz, 1H), 5.64 (s, 1H), 5.27 (d, *J* = 9.2 Hz, 1H), 4.55 (ddd, *J* = 11.3, 7.3, 4.0 Hz, 1H), 4.37 – 4.25 (m, 2H), 3.98 – 3.90 (m, 1H), 3.72 (s, 3H), 3.31 (dd, *J* = 8.8, 6.2 Hz, 2H), 3.12 (t, *J* = 10.1 Hz, 1H), 2.60 (d, *J* = 8.1 Hz, 1H), 2.49 – 2.32 (m, 2H), 2.27 (d, *J* = 6.6 Hz, 1H), 2.14 (ddd, *J* = 14.2, 11.4, 5.7 Hz, 1H), 2.00 (dd, *J* = 13.0, 6.5 Hz, 1H), 1.92 – 1.78 (m, 2H), 1.71 (dd, *J* = 16.8, 6.6 Hz, 1H), 1.41 (s, 9H), 1.09 (d, *J* = 6.4 Hz, 3H), 0.99 (d, *J* = 6.8 Hz, 3H), 0.88 (d,

$J = 6.7$ Hz, 3H). MS (ESI) m/z : $[M+H]^+$ calcd for $C_{24}H_{41}N_4O_7^+$ 497.3, found 497.3.

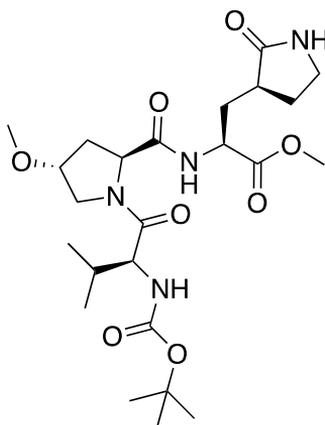
Methyl(S)-2-((2S,4S)-1-((tert-butoxycarbonyl)-l-valyl)-4-methoxypyrrolidine-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (64b).



Chemical Formula: $C_{24}H_{40}N_4O_8$
Molecular Weight: 512.60

Compound **64b** was prepared following the general procedure Method D and using (2S,4S)-1-((tert-butoxycarbonyl)-l-valyl)-4-methoxypyrrolidine-2-carboxylic acid **62b** (purification method: flash chromatography eluting $CH_2Cl_2/MeOH$ 95:5 v/v; yield: 890 mg, 45 %). 1H NMR (400 MHz, $CDCl_3$): δ 7.29 (brs, 1H), 5.61 (s, 1H), 5.23 (d, $J = 9.2$ Hz, 1H), 4.62 (ddd, $J = 9.3, 7.2, 3.5$ Hz, 2H), 4.27 (dd, $J = 9.3, 5.8$ Hz, 1H), 4.01 (d, $J = 2.6$ Hz, 1H), 3.92 (dd, $J = 10.7, 5.0$ Hz, 1H), 3.71 (s, 3H), 3.63 (d, $J = 10.6$ Hz, 1H), 3.35 – 3.27 (m, 5H), 2.59 – 2.39 (m, 3H), 2.20 – 2.00 (m, 2H), 1.84 (ddd, $J = 14.0, 8.1, 3.9$ Hz, 2H), 1.41 (s, 9H), 1.05 (d, $J = 6.8$ Hz, 3H), 0.96 (d, $J = 6.7$ Hz, 3H). MS (ESI) m/z : $[M+H]^+$ calcd for $C_{24}H_{41}N_4O_8^+$ 513.3, found 513.3.

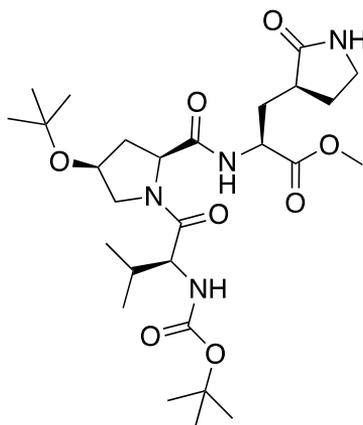
Methyl(S)-2-((2S,4R)-1-((tert-butoxycarbonyl)-l-valyl)-4-methoxypyrrolidine-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (64c).



Chemical Formula: C₂₄H₄₀N₄O₈
Molecular Weight: 512.60

Compound **64c** was prepared following the general procedure Method D and using (2*S*,4*R*)-1-((*tert*-butoxycarbonyl)-*l*-valyl)-4-methoxypyrrolidine-2-carboxylic acid **62c** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 366 mg, 52 %). ¹H NMR (400 MHz, CDCl₃): δ 7.28 (brs, 1H), 5.59 (s, 1H), 5.25 (d, *J* = 9.2 Hz, 1H), 4.61 (ddd, *J* = 9.3, 7.2, 3.5 Hz, 2H), 4.28 (dd, *J* = 9.3, 5.8 Hz, 1H), 4.0 (d, *J* = 2.6 Hz, 1H), 3.92 (dd, *J* = 10.7, 5.0 Hz, 1H), 3.73 (s, 3H), 3.61 (d, *J* = 10.6 Hz, 1H), 3.33 – 3.29 (m, 5H), 2.57 – 2.36 (m, 3H), 2.19 – 1.99 (m, 2H), 1.82 (ddd, *J* = 14.0, 8.1, 3.9 Hz, 2H), 1.41 (s, 9H), 1.03 (d, *J* = 6.8 Hz, 3H), 0.94 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₄H₄₁N₄O₈⁺ 513.3, found 513.3.

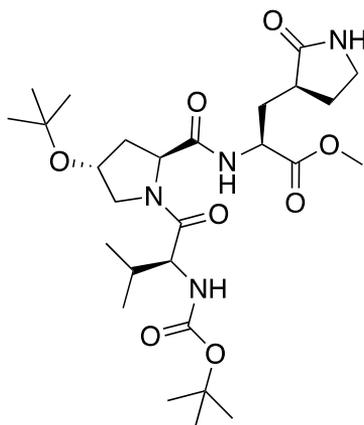
Methyl(S)-2-((2S,4S)-4-(tert-butoxy)-1-((tert-butoxycarbonyl)-l-valyl)pyrrolidin-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (64d).



Chemical Formula: $C_{27}H_{46}N_4O_8$
Molecular Weight: 554.69

Compound **64d** was prepared following the general procedure Method D and using (2*S*,4*S*)-4-(*tert*-butoxy)-1-((*tert*-butoxycarbonyl)-*l*-valyl)pyrrolidin-2-carboxylic acid **62d** (purification method: flash chromatography eluting $CH_2Cl_2/MeOH$ 95:5 v/v; yield: 147 mg, 37 %). 1H NMR (700 MHz, $CDCl_3$): δ 7.39 (d, $J = 7.2$ Hz, 1H), 5.58 (d, $J = 20.1$ Hz, 1H), 4.58 (ddd, $J = 11.4, 7.6, 3.9$ Hz, 1H), 4.36 (t, $J = 8.0$ Hz, 1H), 4.28 (dd, $J = 9.1, 5.5$ Hz, 1H), 4.22 – 4.18 (m, 1H), 3.92 (dd, $J = 10.0, 6.3$ Hz, 1H), 3.71 (s, 3H), 3.40 (dd, $J = 10.0, 6.8$ Hz, 1H), 3.31 (s, 3H), 2.53 (dd, $J = 9.8, 5.1$ Hz, 1H), 2.49 – 2.42 (m, 2H), 2.41 – 2.36 (m, 1H), 2.17 – 2.02 (m, 2H), 1.89 – 1.82 (m, 2H), 1.42 (s, 9H), 1.20 (s, 9H), 1.01 (d, $J = 6.8$ Hz, 3H), 0.89 (d, $J = 6.7$ Hz, 3H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{27}H_{47}N_4O_8^+$ 555.3, found 555.3.

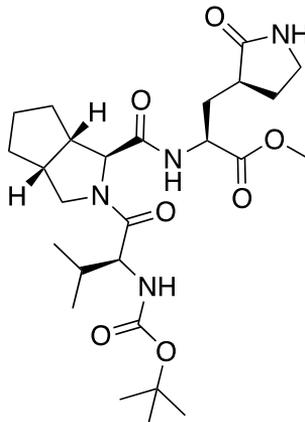
Methyl(*S*)-2-((2*S*,4*R*)-4-(*tert*-butoxy)-1-((*tert*-butoxycarbonyl)-*l*-valyl)pyrrolidin-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate (**64e**).



Chemical Formula: C₂₇H₄₆N₄O₈
Molecular Weight: 554.69

Compound **64e** was prepared following the general procedure Method D and using (2*S*,4*R*)-4-(*tert*-butoxy)-1-((*tert*-butoxycarbonyl)-*l*-valyl)pyrrolidin-2-carboxylic acid **62e** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 600 mg, 83 %). ¹H NMR (700 MHz, CDCl₃): δ 7.41 (d, *J* = 7.2 Hz, 1H), 5.59 (d, *J* = 20.1 Hz, 1H), 4.56 (ddd, *J* = 11.4, 7.6, 3.9 Hz, 1H), 4.38 (t, *J* = 8.0 Hz, 1H), 4.27 (dd, *J* = 9.1, 5.5 Hz, 1H), 4.21 – 4.17 (m, 1H), 3.91 (dd, *J* = 10.0, 6.3 Hz, 1H), 3.73 (s, 3H), 3.41 (dd, *J* = 10.0, 6.8 Hz, 1H), 3.33 (s, 3H), 2.54 (dd, *J* = 9.8, 5.1 Hz, 1H), 2.51 – 2.42 (m, 2H), 2.39 – 2.35 (m, 1H), 2.15 – 2.04 (m, 2H), 1.89 – 1.83 (m, 2H), 1.43 (s, 9H), 1.21 (s, 9H), 1.03 (d, *J* = 6.8 Hz, 3H), 0.89 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₇H₄₇N₄O₈⁺ 555.3, found 555.3.

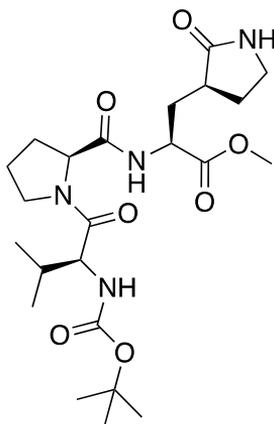
Methyl-(*S*)-2-((1*S*,3*aR*,6*aS*)-2-((*tert*-butoxycarbonyl)-*l*-valyl)octahydrocyclopenta[*c*]pyrrole-1-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate (**64f**).



Chemical Formula: C₂₅H₄₀N₄O₇
Molecular Weight: 508.62

Compound **64f** was prepared following the general procedure Method D and using (1*S*,3*aR*,6*aS*)-2-((*tert*-butoxycarbonyl)-*l*-valyl)octahydrocyclopenta[*c*]pyrrole-1-carboxylic acid **62f** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 427 mg, 62 %). ¹H NMR (400 MHz, CDCl₃): δ 7.54 (d, *J* = 7.1 Hz, 1H), 5.61 (s, 1H), 5.22 (d, *J* = 9.3 Hz, 1H), 4.52 (ddd, *J* = 11.2, 7.2, 4.2 Hz, 1H), 4.31 – 4.23 (m, 2H), 3.82 (dd, *J* = 10.2, 7.4 Hz, 1H), 3.72 (s, 3H), 3.67 – 3.60 (m, 1H), 3.37 – 3.25 (m, 2H), 2.80 (s, 3H), 2.55 – 2.35 (m, 2H), 2.13 (ddd, *J* = 14.3, 10.4, 5.0 Hz, 1H), 2.06 – 1.94 (m, 2H), 1.93 – 1.70 (m, 5H), 1.50-1.41 (m, 10H), 0.98 (d, *J* = 6.8 Hz, 3H), 0.89 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₆H₄₃N₄O₇⁺ 523.3, found 523.4.

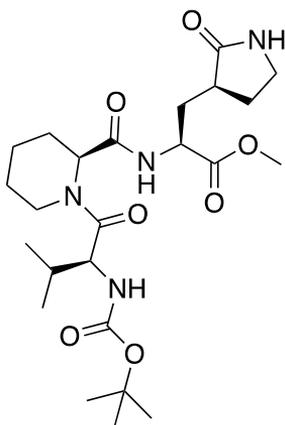
Methyl (S)-2-((S)-1-((tert-butoxycarbonyl)-l-valyl)pyrrolidine-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (**64g**).



Chemical Formula: C₂₃H₃₈N₄O₇
Molecular Weight: 482.58

Compound **64g** was prepared following the general procedure Method D and using (*tert*-butoxycarbonyl)-l-valyl-l-proline **62g**¹⁹ (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 180 mg, 65 %). ¹H NMR (400 MHz, DMSO-d₆): δ 8.43 (d, *J* = 8.4 Hz, 1H), 7.59 (s, 1H), 6.80 (d, *J* = 8.4 Hz, 1H), 4.43 – 4.29 (m, 2H), 3.97 (t, *J* = 8.0 Hz, 1H), 3.75 – 3.66 (m, 1H), 3.62 (s, 3H), 3.55 (d, *J* = 8.6 Hz, 1H), 3.14 – 3.02 (m, 2H), 2.16 – 1.70 (m, 8H), 1.63 – 1.50 (m, 2H), 1.36 (s, 9H), 0.88 (d, *J* = 6.6 Hz, 3H), 0.84 (d, *J* = 6.4 Hz, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₃H₃₉N₄O₇⁺ 483.3, found 483.3.

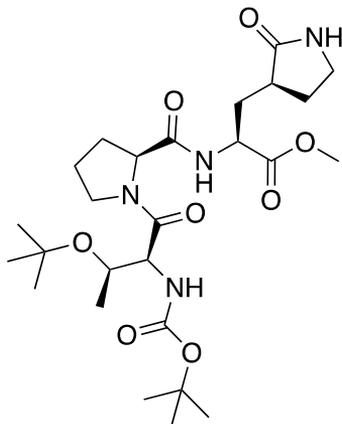
Methyl (S)-2-((S)-1-((tert-butoxycarbonyl)-l-valyl)piperidine-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (**64h**).



Chemical Formula: C₂₄H₄₀N₄O₇
Molecular Weight: 496.61

Compound **64h** was prepared following the general procedure Method D and using (S)-1-((tert-butoxycarbonyl)-l-valyl)piperidine-2-carboxylic acid **62h**²⁰ (purification method: flash chromatography eluting CH₂Cl₂/MeOH 98:2 to 95:5 v/v; yield: 397 mg, 49%). ¹H NMR (400 MHz, CDCl₃): δ 7.62 (d, *J* = 6.3 Hz, 1H), 5.81 (t, *J* = 23.4 Hz, 1H), 5.57 (d, *J* = 7.9 Hz, 1H), 5.32 (dd, *J* = 35.4, 12.9 Hz, 1H), 4.67 – 4.55 (m, 1H), 4.53 – 4.45 (m, 1H), 4.41 – 4.31 (m, 1H), 3.86 (t, *J* = 12.5 Hz, 1H), 3.71 (s, 3H), 3.38 – 3.25 (m, 2H), 2.42 (t, *J* = 13.2 Hz, 4H), 2.00 (dddd, *J* = 26.7, 22.7, 16.5, 8.9 Hz, 8H), 1.43 (s, 9H), 0.99 (d, *J* = 5.4 Hz, 3H), 0.86 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₄H₄₁N₄O₇⁺ 497.3, found 497.1.

Methyl-(S)-2-((S)-1-(N-(tert-butoxycarbonyl)-O-(tert-butyl)-l-threonyl)pyrrolidin-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (**64i**).



Chemical Formula: C₂₆H₄₄N₄O₈
Molecular Weight: 540.66

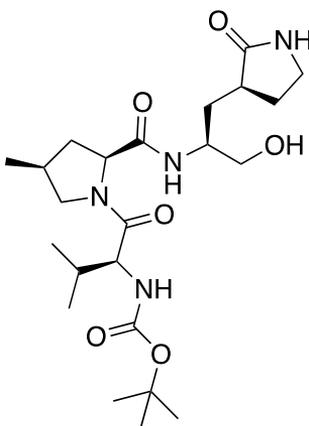
Compound **64i** was prepared following the general procedure Method D and using *N*-(*tert*-butoxycarbonyl)-*O*-(*tert*-butyl)-*l*-threonyl-*l*-proline **62i** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 97:3 v/v; yield: 296 mg, 41 %). ¹H NMR (400 MHz, CDCl₃): δ 7.43 (d, *J* = 7.8 Hz, 1H), 5.76 (s, 1H), 5.40 (d, *J* = 8.1 Hz, 1H), 4.66 – 4.51 (m, 2H), 4.46 (dd, *J* = 8.1, 5.3 Hz, 1H), 4.00 – 3.84 (m, 2H), 3.71 (s, 3H), 3.32 (dd, *J* = 18.7, 9.1 Hz, 2H), 2.42 (ddd, *J* = 12.4, 9.0, 3.4 Hz, 2H), 2.26 – 1.76 (m, 8H), 1.41 (s, 9H), 1.23 (s, 9H), 1.15 (d, *J* = 5.2 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₆H₄₅N₄O₈⁺ 541.3, found 541.3.

General procedure for the synthesis of tripeptide alcohols 66a-i (Method E).

In oven-dried three-necked round-bottom flask, the tripeptide methyl ester (1 eq.) was dissolved in anhydrous THF (0.2 M) under a positive anhydrous nitrogen pressure, then a 2M solution of LiBH₄ (3 eq.) in anhydrous THF was added dropwise at 0°C, and the reaction mixture was stirred at room temperature for 5h.

The reaction was quenched by acidifying with a saturated citric acid solution to pH 2 and extracted with EtOAc (25 mL x 3). The organic phase was washed with brine (15 mL x 10), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a white solid.

tert-Butyl ((*S*)-1-((2*S*,4*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-4-methylpyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66a**).

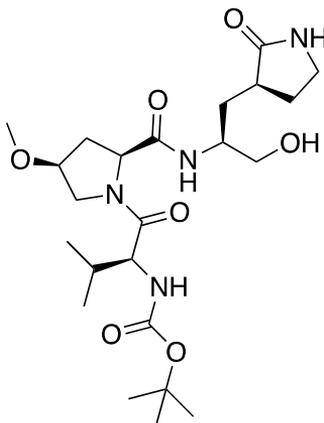


Chemical Formula: C₂₃H₄₀N₄O₆
Molecular Weight: 468.60

Compound **66a** was prepared following the general procedure Method E and using methyl (*S*)-2-((2*S*,4*S*)-1-((*tert*-butoxycarbonyl)-1-valyl)-4-methylpyrrolidine-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64a** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 30 mg, 54%). ¹H NMR (400 MHz, CDCl₃): δ 7.50 (d, *J* = 7.2 Hz, 1H), 5.58 (s, 1H), 5.27 (d, *J* = 9.2 Hz, 1H), 4.56 (ddd, *J* = 11.3, 7.3, 4.0 Hz, 1H), 4.35 – 4.25 (m, 2H), 3.96 – 3.90 (m, 1H), 3.31-3.20 (m, 4H), 3.10 (t, *J* = 10.1 Hz, 1H), 2.61 (d, *J* = 8.1 Hz, 1H), 2.47 – 2.30 (m, 2H), 2.25 (d, *J* = 6.6 Hz, 1H), 2.12 (ddd, *J* = 14.2, 11.4, 5.7 Hz, 1H), 2.00 (dd, *J* = 13.0, 6.5 Hz, 1H), 1.92 – 1.79 (m, 2H),

1.71 (dd, $J = 16.8, 6.6$ Hz, 2H), 1.39 (s, 9H), 1.08 (d, $J = 6.4$ Hz, 3H), 0.98 (d, $J = 6.8$ Hz, 3H), 0.86 (d, $J = 6.7$ Hz, 3H). MS (ESI) m/z : $[M+H]^+$ calcd for $C_{23}H_{41}N_4O_6^+$ 469.3, found 469.3.

tert-Butyl ((*S*)-1-((2*S*,4*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate-2-yl)carbamoyl)-4-methoxypyrrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66b**).

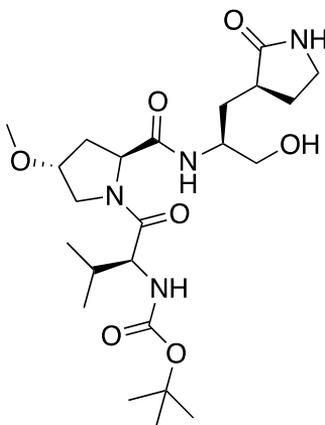


Chemical Formula: $C_{23}H_{40}N_4O_7$
Molecular Weight: 484.59

Compound **66b** was prepared following the general procedure Method E and using methyl (*S*)-2-((2*S*,4*S*)-1-((*tert*-butoxycarbonyl)-1-valyl)-4-methoxypyrrrolidine-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64b** (purification method: flash chromatography eluting $CH_2Cl_2/MeOH$ 93:7 v/v; yield: 624 mg, 74%). 1H NMR (700 MHz, $CDCl_3$): δ 7.29 (d, $J = 8.0$ Hz, 1H), 7.04 (d, $J = 6.5$ Hz, 1H), 5.94 (s, 1H), 5.24 (d, $J = 9.0$ Hz, 1H), 4.53 (dd, $J = 9.1, 3.7$ Hz, 1H), 4.15 (dd, $J = 8.8, 5.5$ Hz, 1H), 3.91 (dd, $J = 9.8, 4.9$ Hz, 2H), 3.80 (dd, $J = 11.9, 2.4$ Hz, 1H), 3.72 (dd, $J = 13.7, 4.9$ Hz, 1H), 3.51 – 3.45 (m, 2H), 3.33 – 3.29 (m, 5H), 2.40 (dd, $J = 17.0, 7.1$ Hz, 2H), 2.07 (dd, $J = 11.8, 6.7$ Hz, 2H), 2.00 – 1.94 (m, 2H), 1.81 (dd, $J = 21.0, 9.4$ Hz, 2H), 1.42 (s, 9H), 1.03 (d, J

= 6.6 Hz, 3H), 0.97 (d, $J = 6.8$ Hz, 3H). MS (ESI) m/z : $[M+H]^+$ calcd for $C_{23}H_{41}N_4O_7^+$ 485.3, found 485.3.

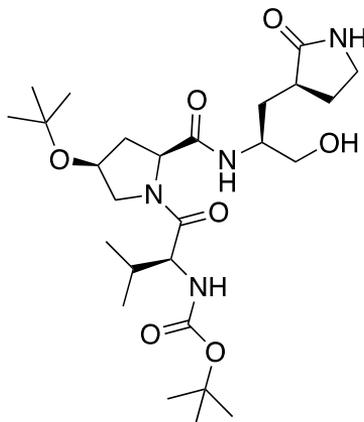
tert-Butyl ((*S*)-1-((2*S*,4*R*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-4-methoxypyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66c**).



Chemical Formula: $C_{23}H_{40}N_4O_7$
Molecular Weight: 484.59

Compound **66c** was prepared following the general procedure Method E and using methyl ((*S*)-2-((2*S*,4*R*)-1-((*tert*-butoxycarbonyl)-1-valyl)-4-methoxypyrrolidine-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64c** (purification method: flash chromatography eluting $CH_2Cl_2/MeOH$ 95:5 v/v; yield: 160 mg, 47%). 1H NMR (700 MHz, $CDCl_3$): δ 7.28 (d, $J = 8.0$ Hz, 1H), 7.02 (d, $J = 6.5$ Hz, 1H), 5.95 (s, 1H), 5.26 (d, $J = 9.0$ Hz, 1H), 4.54 (dd, $J = 9.1$, 3.7 Hz, 1H), 4.17 (dd, $J = 8.8$, 5.5 Hz, 1H), 3.92 (dd, $J = 9.8$, 4.9 Hz, 2H), 3.79 (dd, $J = 11.9$, 2.4 Hz, 1H), 3.71 (dd, $J = 13.7$, 4.9 Hz, 1H), 3.52 – 3.44 (m, 2H), 3.35 – 3.31 (m, 5H), 2.42 (dd, $J = 17.0$, 7.1 Hz, 2H), 2.09 (dd, $J = 11.8$, 6.7 Hz, 2H), 1.99 – 1.92 (m, 2H), 1.82 (dd, $J = 21.0$, 9.4 Hz, 2H), 1.41 (s, 9H), 1.05 (d, $J = 6.6$ Hz, 3H), 0.99 (d, $J = 6.8$ Hz, 3H). MS (ESI) m/z : $[M+H]^+$ calcd for $C_{23}H_{41}N_4O_7^+$ 485.3, found 485.3.

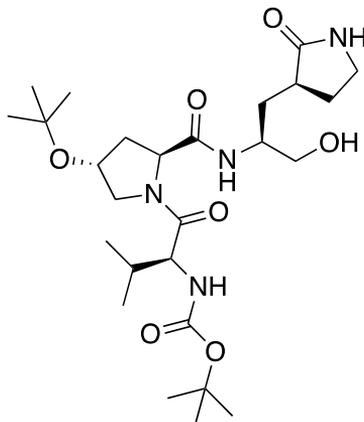
tert-Butyl ((*S*)-1-((2*S*,4*S*)-4-(*tert*-butoxy)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66d**).



Chemical Formula: C₂₆H₄₆N₄O₇
Molecular Weight: 526.68

Compound **66d** was prepared following the general procedure Method E and using methyl (*S*)-2-((2*S*,4*S*)-4-(*tert*-butoxy)-1-((*tert*-butoxycarbonyl)-1-valyl)pyrrolidin-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64d** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 53 mg, 40%). ¹H NMR (400 MHz, CDCl₃): δ 7.31 (d, *J* = 7.6 Hz, 1H), 5.93 (s, 1H), 5.33 – 5.25 (m, 1H), 4.29 (dt, *J* = 9.1, 7.0 Hz, 2H), 4.22 – 4.14 (m, 1H), 3.99 (d, *J* = 3.4 Hz, 1H), 3.95 – 3.85 (m, 1H), 3.65 (s, 1H), 3.52 – 3.39 (m, 3H), 3.31 (dd, *J* = 9.0, 4.4 Hz, 2H), 2.57 – 2.47 (m, 1H), 2.46 – 2.35 (m, 2H), 2.13 – 1.93 (m, 3H), 1.82 (dd, *J* = 27.7, 17.3 Hz, 2H), 1.42 (s, 9H), 1.19 (s, 9H), 1.01 (d, *J* = 6.7 Hz, 3H), 0.90 (d, *J* = 6.6 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₆H₄₇N₄O₇⁺ 527.3, found 527.3.

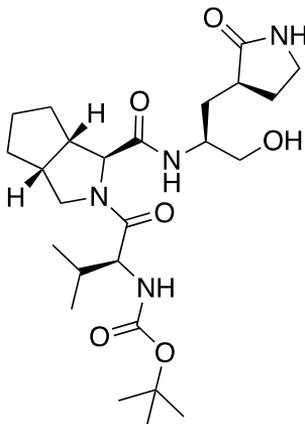
tert-Butyl ((*S*)-1-((2*S*,4*R*)-4-(*tert*-butoxy)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66e**).



Chemical Formula: C₂₆H₄₆N₄O₇
Molecular Weight: 526.68

Compound **66e** was prepared following the general procedure Method E and using methyl ((*S*)-2-((2*S*,4*R*)-4-(*tert*-butoxy)-1-((*tert*-butoxycarbonyl)-l-valyl)pyrrolidin-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64e** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 225 mg, 75%). ¹H NMR (400 MHz, CDCl₃): δ 7.31 (d, *J* = 7.6 Hz, 1H), 5.95 (s, 1H), 5.34 – 5.23 (m, 1H), 4.27 (dt, *J* = 9.1, 7.0 Hz, 2H), 4.24 – 4.15 (m, 1H), 4.01 (d, *J* = 3.4 Hz, 1H), 3.94 – 3.87 (m, 1H), 3.67 (s, 1H), 3.53 – 3.41 (m, 3H), 3.32 (dd, *J* = 9.0, 4.4 Hz, 2H), 2.59 – 2.45 (m, 1H), 2.48 – 2.36 (m, 2H), 2.14 – 1.92 (m, 3H), 1.81 (dd, *J* = 27.7, 17.3 Hz, 2H), 1.43 (s, 9H), 1.17 (s, 9H), 1.02 (d, *J* = 6.7 Hz, 3H), 0.89 (d, *J* = 6.6 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₆H₄₇N₄O₇⁺ 527.3, found 527.3.

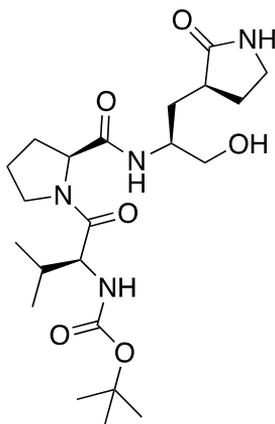
tert-Butyl ((*S*)-1-((1*S*,3*aR*,6*aS*)-1-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)hexahydrocyclopenta[*c*]pyrrol-2(1*H*)-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66f**).



Chemical Formula: C₂₅H₄₂N₄O₆
Molecular Weight: 494.63

Compound **66f** was prepared following the general procedure Method E and using methyl-((*S*)-2-((1*S*,3*aR*,6*aS*)-2-((*tert*-butoxycarbonyl)-1-valyl)octahydrocyclopenta[*c*]pyrrole-1-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64f** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 to 93:7 v/v; yield: 250 mg, 62%). ¹H NMR (700 MHz, CDCl₃): δ 7.64 – 7.60 (m, 1H), 5.81 (d, *J* = 14.7 Hz, 1H), 5.30 (dd, *J* = 11.7, 6.1 Hz, 1H), 4.30 – 4.24 (m, 1H), 4.21 (d, *J* = 3.7 Hz, 1H), 4.06 (dd, *J* = 9.0, 4.6 Hz, 1H), 3.96 (dt, *J* = 10.6, 5.6 Hz, 1H), 3.86 (dd, *J* = 10.0, 8.1 Hz, 1H), 3.68 (d, *J* = 10.0 Hz, 1H), 3.64 – 3.55 (m, 2H), 3.51 (ddd, *J* = 32.2, 15.7, 9.0 Hz, 1H), 3.32 (dd, *J* = 9.0, 4.5 Hz, 2H), 2.80 (dd, *J* = 7.6, 3.9 Hz, 1H), 2.76 – 2.72 (m, 1H), 2.55 – 2.48 (m, 1H), 2.40 (ddd, *J* = 12.4, 8.6, 4.3 Hz, 1H), 2.02 – 1.96 (m, 1H), 1.95 – 1.90 (m, 1H), 1.84 (ddd, *J* = 26.6, 17.7, 9.8 Hz, 2H), 1.65 – 1.58 (m, 3H), 1.47 (dt, *J* = 11.8, 5.5 Hz, 2H), 1.42 (d, *J* = 5.8 Hz, 9H), 0.98 (t, *J* = 6.0 Hz, 3H), 0.89 (d, *J* = 6.7 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₅H₄₃N₄O₆⁺ 495.3, found 495.2.

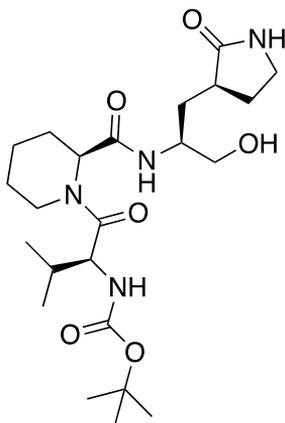
tert-Butyl ((*S*)-1-((*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**66g**).



Chemical Formula: C₂₂H₃₈N₄O₆
Molecular Weight: 454.57

Compound **66g** was prepared following the general procedure Method E and using methyl ((*S*)-2-((*S*)-1-((*tert*-butoxycarbonyl)-1-valyl)pyrrolidine-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64g** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 90 mg, 50%). ¹H NMR (400 MHz, DMSO-*d*₆): δ 7.66 (d, *J* = 8.9 Hz, 1H), 7.45 (s, 1H), 6.80 (d, *J* = 8.4 Hz, 1H), 4.64 (d, *J* = 5.7 Hz, 1H), 4.23 (d, *J* = 7.8 Hz, 1H), 3.98 (s, 1H), 3.77 (s, 1H), 3.69 (s, 2H), 3.13 – 3.04 (m, 2H), 3.03 – 2.95 (m, 2H), 2.16 (d, *J* = 14.8, 2H), 2.05 – 1.90 (m, 6H), 1.50 (d, *J* = 11.4 Hz, 2H), 1.36 (s, 9H), 0.88 (d, *J* = 6.7 Hz, 3H), 0.84 (d, *J* = 6.6 Hz, 3H). MS (ESI) *m/z*: [M+H]⁺ calcd for C₂₂H₃₉N₄O₆⁺ 455.3, found 455.2.

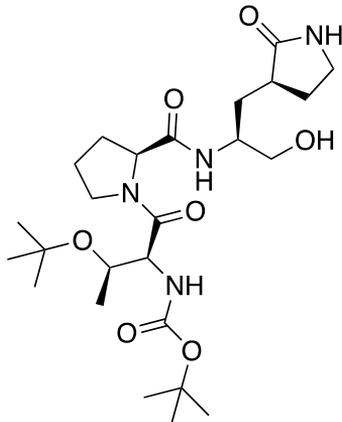
tert-Butyl ((S)-1-((S)-2-(((S)-1-hydroxy-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)piperidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (66h).



Chemical Formula: C₂₃H₄₀N₄O₆
Molecular Weight: 468.60

Compound **66h** was prepared following the general procedure Method E and using methyl (S)-2-(((S)-1-((tert-butoxycarbonyl)-1-valyl)piperidine-2-carboxamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate **64h** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 97:3 v/v; yield: 274 mg, 73%). ¹H NMR (400 MHz, CDCl₃): δ 7.87 (d, *J* = 7.9 Hz, 1H), 5.80 (m, 1H), 5.23 (m, 1H), 4.57 (m, 1H), 4.49 (d, *J* = 5.2 Hz, 1H), 4.25 - 4.15 (m, 1H), 4.03 - 3.71 (m, 1H), 3.59 (dd, *J* = 22.9, 12.1 Hz, 2H), 3.61-3.53 (m, 2H), 3.40 - 3.32 (m, 3H), 2.57 - 2.30 (m, 3H), 1.93 - 1.77 (m, 8H), 1.63 (s, 9H), 1.01 (d, *J* = 3.8 Hz, 3H), 0.89 (d, *J* = 7.6 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₃H₄₁N₄O₆⁺ 469.3, found 469.2.

tert-Butyl ((2*S*,3*R*)-3-(*tert*-butoxy)-1-((*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-1-oxobutan-2-yl)carbamate (**66i**).



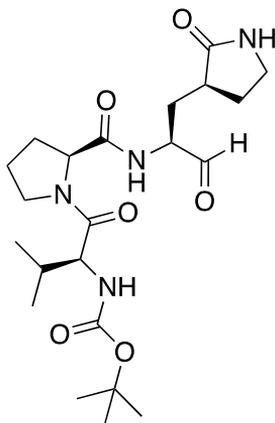
Chemical Formula: C₂₅H₄₄N₄O₇
Molecular Weight: 512.65

Compound **66i** was prepared following the general procedure Method E and using methyl-(*S*)-2-((*S*)-1-(*N*-(*tert*-butoxycarbonyl)-*O*-(*tert*-butyl)-*l*-threonyl)pyrrolidin-2-carboxamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **64i** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 98:2 to 95:5 v/v; yield: 190 mg, 62%). ¹H NMR (400 MHz, CDCl₃): δ 6.84 (d, *J* = 8.3 Hz, 1H), 5.71 (s, 1H), 5.38 (d, *J* = 7.6 Hz, 1H), 4.59 – 4.44 (m, 2H), 4.04 (ddd, *J* = 23.9, 18.6, 5.5 Hz, 2H), 3.90 (dt, *J* = 12.6, 6.4 Hz, 1H), 3.85 – 3.71 (m, 1H), 3.70 – 3.61 (m, 1H), 3.59 – 3.41 (m, 1H), 3.37 – 3.29 (m, 2H), 3.18 (s, 1H), 2.54 – 2.29 (m, 2H), 2.27 – 1.92 (m, 5H), 1.88 – 1.76 (m, 2H), 1.43 (s, 9H), 1.26 (s, 9H), 1.18 (d, *J* = 9.0 Hz, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₅H₄₅N₄O₇⁺ 513.3, found 513.3.

General procedure for the synthesis of target aldehydes 40-48 (Method F). The tripeptide alcohol intermediate (1 eq.) was dissolved in anhydrous DMSO (5 eq.) and anhydrous CH₂Cl₂ (0.4 M) under anhydrous nitrogen atmosphere, and

DIPEA (3.5 eq.) was added at -5°C. In another flask, pyridine sulfur trioxide complex (3 eq.) and anhydrous pyridine (3 eq.) were suspended in anhydrous DMSO (7.5 eq.) at room temperature. The suspension was left under stirring for 10 min and then added to the previously formed tripeptide alcohol/DMSO mixture at -5°C. After 24 h, the reaction mixture was poured into ice/water (1:1) and extracted with CH₂Cl₂ (25 mL x 3). The collected organic layers were washed with a saturated citric acid solution (15 mL x 1), water (15 mL x 1), saturated solution of NaHCO₃ (15 mL x 1), brine (15 mL x 1), dried over anhydrous Na₂SO₄, filtered, and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a white solid.

tert-Butyl ((S)-3-methyl-1-oxo-1-((S)-2-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)butan-2-yl)carbamate (40).

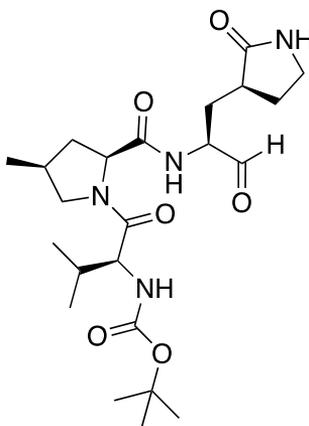


Chemical Formula: C₂₂H₃₆N₄O₆
Molecular Weight: 452.55

Compound **40** was prepared following the general procedure Method F and using *tert*-butyl ((S)-1-((S)-2-(((S)-1-hydroxy-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate **66g** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 to 93:7

v/v; yield: 22 mg, 25%). $[\alpha]^{20}_{\text{D}}$ -84.5 ($c = 1.0$, MeOH). ^1H NMR (400 MHz, DMSO- d_6): δ 9.42 (s, 1H), 8.44 (d, $J = 7.9$ Hz, 1H), 7.60 (s, 1H), 6.82 (d, $J = 8.2$ Hz, 1H), 4.32 (s, 1H), 4.26 (d, $J = 7.6$ Hz, 1H), 3.99 (d, $J = 7.7$ Hz, 1H), 3.71 (s, 1H), 3.56 (s, 1H), 3.18 – 3.03 (m, 2H), 2.19 – 2.05 (m, 2H), 1.90 (dd, $J = 33.5$, 14.7 Hz, 6H), 1.67 – 1.55 (m, 2H), 1.36 (s, 9H), 0.88 (d, $J = 6.5$ Hz, 3H), 0.84 (d, $J = 6.5$ Hz, 3H). ^{13}C NMR (176 MHz, CDCl_3): δ 199.9, 180.1, 172.8, 172.0, 156.0, 131.2, 79.7, 60.4, 57.9, 57.0, 47.8, 40.6, 37.9, 31.4, 30.0, 29.1, 28.9, 25.4, 19.6, 17.4. LC-MS (ESI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{22}\text{H}_{37}\text{N}_4\text{O}_6^+$ 453.3, found 453.2 (rt: 6.65).

tert-Butyl ((*S*)-3-methyl-1-((2*S*,4*S*)-4-methyl-2-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-1-oxobutan-2-yl)carbamate (**41**).

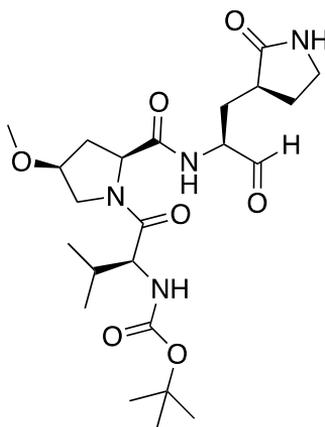


Chemical Formula: $\text{C}_{23}\text{H}_{38}\text{N}_4\text{O}_6$
Molecular Weight: 466.58

Compound **41** was prepared following the general procedure Method F and using *tert*-butyl ((*S*)-1-((2*S*,4*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-4-methylpyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate **66a** (purification method: flash chromatography eluting $\text{CH}_2\text{Cl}_2/\text{MeOH}$ 95:5; yield: 18 mg, 60%). $[\alpha]^{20}_{\text{D}}$ -38.9 ($c = 1.0$, MeOH). ^1H NMR (400 MHz, CDCl_3):

δ 9.50 (s, 1H), 8.16 – 8.07 (m, 1H), 5.37 – 5.28 (m, 1H), 4.43 – 4.35 (m, 1H), 4.30 (dd, $J = 9.0, 6.0$ Hz, 1H), 4.08 (dd, $J = 14.7, 7.5$ Hz, 1H), 3.99 – 3.89 (m, 1H), 3.37 – 3.29 (m, 2H), 3.15 (t, $J = 10.2$ Hz, 1H), 2.71 – 2.54 (m, 1H), 2.48 – 2.35 (m, 2H), 2.28 (td, $J = 12.9, 6.5$ Hz, 1H), 2.10 – 1.75 (m, 6H), 1.42 (s, 9H), 1.09 (t, $J = 6.5$ Hz, 3H), 0.99 (t, $J = 5.9$ Hz, 3H), 0.90 – 0.83 (m, 3H). ^{13}C NMR (176 MHz, CDCl_3): δ 199.8, 180.1, 172.9, 171.0, 155.9, 129.9, 79.6, 61.3, 58.0, 57.0, 54.9, 40.7, 37.6, 34.3, 29.8, 29.5, 28.6, 28.5, 19.5, 17.3, 16.6. LC-MS (ESI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{23}\text{H}_{39}\text{N}_4\text{O}_6^+$ 467.3, found 467.2 (r: 6.96).

tert-Butyl ((*S*)-1-((2*S*,4*S*)-4-methoxy-2-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**42**).

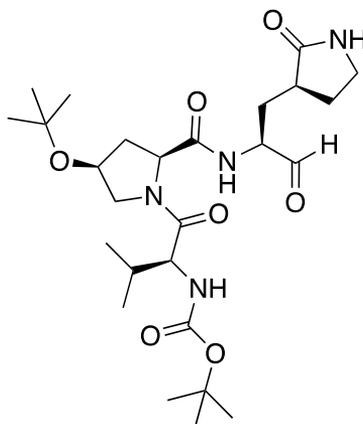


Chemical Formula: $\text{C}_{23}\text{H}_{38}\text{N}_4\text{O}_7$
Molecular Weight: 482.58

Compound **42** was prepared following the general procedure Method F and using *tert*-butyl ((*S*)-1-((2*S*,4*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-4-methoxypyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate **66b** (purification method: flash chromatography eluting $\text{CH}_2\text{Cl}_2/\text{MeOH}$ 95:5 v/v, yield: 76 mg, 35%). $[\alpha]_D^{20}$ -49.6 ($c = 1.0$, MeOH). ^1H NMR (700 MHz, CDCl_3): δ 9.45 (s, 1H), 7.31 (d, $J = 8.0$ Hz, 1H), 7.05 (d, $J =$

6.5 Hz, 1H), 5.90 (s, 1H), 5.24 (d, $J = 9.0$ Hz, 1H), 4.53 (dd, $J = 9.1, 3.7$ Hz, 1H), 4.15 (dd, $J = 8.8, 5.5$ Hz, 1H), 3.91 (dd, $J = 9.8, 4.9$ Hz, 2H), 3.80 (dd, $J = 11.9, 2.4$ Hz, 1H), 3.72 (dd, $J = 13.7, 4.9$ Hz, 1H), 3.51 – 3.45 (m, 2H), 3.33 – 3.29 (m, 2H), 2.40 (dd, $J = 17.0, 7.1$ Hz, 2H), 2.07 (dd, $J = 11.8, 6.7$ Hz, 2H), 2.00 – 1.94 (m, 2H), 1.81 (dd, $J = 21.0, 9.4$ Hz, 2H), 1.42 (s, 9H), 1.03 (d, $J = 6.6$ Hz, 3H), 0.97 (d, $J = 6.8$ Hz, 3H). ^{13}C NMR (176 MHz, CDCl_3): δ 199.5, 180.0, 172.9, 172.1, 156.1, 129.9, 79.8, 79.2, 59.5, 57.4, 57.0, 52.9, 40.5, 37.5, 33.2, 31.2, 30.4, 28.6, 28.5, 19.8, 17.4. LC-MS (ESI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{23}\text{H}_{39}\text{N}_4\text{O}_7^+$ 483.3, found 483.2 (rt: 6.60).

tert-Butyl ((*S*)-1-((2*S*,4*S*)-4-(*tert*-butoxy)-2-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**43**).

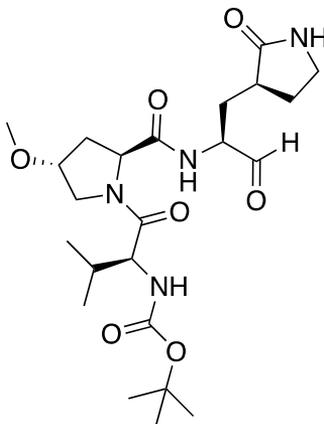


Chemical Formula: $\text{C}_{26}\text{H}_{44}\text{N}_4\text{O}_7$
Molecular Weight: 524.66

Compound **43** was prepared following the general procedure Method F and using *tert*-butyl ((*S*)-1-((2*S*,4*S*)-4-(*tert*-butoxy)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl) carbamate **66d** (purification method: flash chromatography eluting $\text{CH}_3\text{Cl}/\text{MeOH}$ 93:7 v/v, yield: 13 mg, 27%). $[\alpha]_D^{20}$ -12.1 ($c = 1.0$, MeOH).

^1H NMR (400 MHz, CDCl_3): δ 9.50 (s, 1H), 7.98 (d, $J = 5.3$ Hz, 1H), 5.67 (s, 1H), 5.24 (d, $J = 8.3$ Hz, 1H), 4.47 – 4.35 (m, 2H), 4.34 – 4.16 (m, 1H), 3.95 (dd, $J = 10.0, 6.3$ Hz, 1H), 3.43 (dd, $J = 10.0, 6.9$ Hz, 1H), 3.34 (dd, $J = 8.6, 6.3$ Hz, 2H), 2.61 – 2.50 (m, 1H), 2.47 – 2.36 (m, 2H), 2.15 – 2.04 (m, 2H), 1.94 (t, $J = 6.7$ Hz, 2H), 1.90 – 1.80 (m, 2H), 1.43 (s, 9H), 1.19 (s, 9H), 1.01 (d, $J = 6.7$ Hz, 3H), 0.89 (d, $J = 6.7$ Hz, 3H). ^{13}C NMR (176 MHz, CDCl_3): δ 199.9, 179.9, 172.4, 171.9, 156.1, 129.9, 79.7, 74.5, 69.7, 59.2, 57.8, 57.0, 53.8, 40.6, 38.0, 37.0, 31.2, 30.2, 28.9, 28.4, 19.7, 17.2. LC- MS (ESI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{26}\text{H}_{45}\text{N}_4\text{O}_7^+$ 525.3, found 525.3 (r: 7.45).

tert-Butyl ((*S*)-1-((2*S*,4*R*)-4-methoxy-2-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate (**44**).

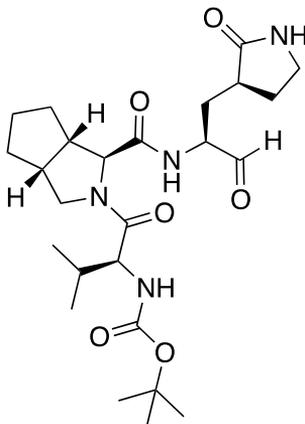


Chemical Formula: $\text{C}_{23}\text{H}_{38}\text{N}_4\text{O}_7$
Molecular Weight: 482.58

Compound **44** was prepared following the general procedure Method F and using *tert*-butyl ((*S*)-1-((2*S*,4*R*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-4-methoxypyrrolidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate **66c** (purification method: flash chromatography eluting $\text{CH}_2\text{Cl}_2/\text{MeOH}$ 95:5 v/v, yield: 10 mg, 17%). $[\alpha]_D^{20}$ -54.2 ($c = 1.0$, MeOH). ^1H

^1H NMR (700 MHz, CDCl_3): δ 9.50 (s, 1H), 8.07 (dd, $J = 22.8, 8.9$ Hz, 1H), 6.05 (t, $J = 43.5$ Hz, 1H), 5.41 – 5.21 (m, 1H), 4.68 – 4.53 (m, 1H), 4.42 (dd, $J = 28.9, 24.1$ Hz, 1H), 4.28 – 4.21 (m, 1H), 3.83 – 3.74 (m, 1H), 3.64 – 3.51 (m, 1H), 3.36 – 3.27 (m, 2H), 2.64 – 2.54 (m, 1H), 2.43 – 2.30 (m, 2H), 2.01 – 1.89 (m, 3H), 1.87 – 1.77 (m, 3H), 1.42 (s, 9H), 1.18 (d, $J = 5.3$ Hz, 9H), 0.96 (d, $J = 5.6$ Hz, 3H), 0.88 (d, $J = 6.4$ Hz, 3H). ^{13}C NMR (176 MHz, DMSO): δ 201.1, 178.6, 172.3, 170.5, 155.6, 149.6, 77.9, 73.5, 69.6, 65.0, 58.5, 57.4, 56.0, 54.5, 46.0, 40.0, 37.0, 28.2, 28.0, 27.4, 19.0, 18.4. LC- MS (ESI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{26}\text{H}_{45}\text{N}_4\text{O}_7^+$ 525.3, found 525.3 (r: 7.13).

tert-Butyl ((*S*)-3-methyl-1-oxo-1-((1*S*,3*aR*,6*aS*)-1-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)hexahydrocyclopenta[*c*]pyrrol-2(1*H*)-yl)butan-2-yl)carbamate (**46**).

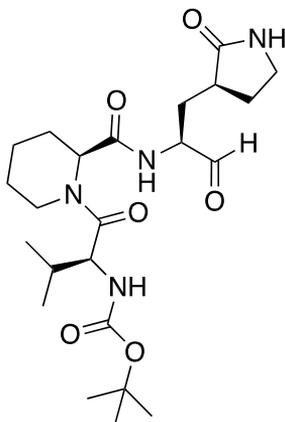


Chemical Formula: $\text{C}_{25}\text{H}_{40}\text{N}_4\text{O}_6$
Molecular Weight: 492.62

Compound **46** was prepared following the general procedure Method F and using *tert*-butyl ((*S*)-1-((1*S*,3*aR*,6*aS*)-1-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)hexahydrocyclopenta[*c*]pyrrol-2(1*H*)-yl)-3-methyl-1-oxobutan-2-yl)carbamate **66f** (purification method: flash chromatography eluting $\text{CH}_2\text{Cl}_2/\text{MeOH}$ 93:7 v/v, yield: 44 mg, 36%). $[\alpha]_D^{20}$ -40.6 ($c = 1.0$, MeOH). ^1H

NMR (700 MHz, CDCl₃): δ 9.50 (s, 1H), 8.12 – 8.09 (m, 1H), 6.08 (s, 1H), 5.27 (d, $J = 9.3$ Hz, 1H), 4.38 – 4.34 (m, 1H), 4.32 (d, $J = 3.6$ Hz, 1H), 4.27 (dd, $J = 9.2, 6.3$ Hz, 1H), 3.86 (dd, $J = 10.2, 7.8$ Hz, 1H), 3.66 (dd, $J = 10.3, 3.6$ Hz, 1H), 3.39 – 3.32 (m, 3H), 2.84 – 2.76 (m, 2H), 2.57 – 2.53 (m, 1H), 2.39 (ddd, $J = 15.6, 8.7, 4.2$ Hz, 1H), 2.02 (dd, $J = 13.5, 6.9$ Hz, 1H), 2.00 – 1.81 (m, 1H), 1.79 – 1.73 (m, 2H), 1.62 (ddd, $J = 14.5, 11.5, 6.1$ Hz, 3H), 1.51 – 1.47 (m, 2H), 1.42 (s, 9H), 0.97 (d, $J = 6.8$ Hz, 3H), 0.88 (d, $J = 6.7$ Hz, 3H). ¹³C NMR (176 MHz, CDCl₃): δ 199.9, 180.3, 173.0, 171.8, 156.0, 129.9, 79.6, 66.8, 57.8, 57.0, 53.6, 47.5, 43.5, 40.7, 38.0, 32.5, 31.7, 31.1, 30.0, 28.9, 25.4, 19.3, 17.4. LC-MS (ESI) m/z : [M+H]⁺ calcd for C₂₅H₄₁N₄O₆⁺ 493.3, found 493.3 (r_t : 7.26).

tert-Butyl ((*S*)-3-methyl-1-oxo-1-((*S*)-2-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)piperidin-1-yl)butan-2-yl)carbamate (**47**).



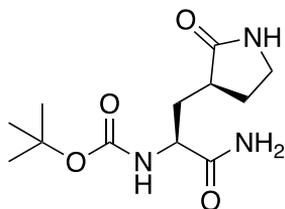
Chemical Formula: C₂₃H₃₈N₄O₆
Molecular Weight: 466.58

Compound **47** was prepared following the general procedure Method F and using *tert*-butyl ((*S*)-1-((*S*)-2-(((*S*)-1-hydroxy-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)piperidin-1-yl)-3-methyl-1-oxobutan-2-yl)carbamate **66h** (purification method: flash chromatography eluting CH₃Cl/MeOH 96:4 v/v, yield: 74 mg, 60%). [α]_D²⁰ -9.3 (c = 1.0, MeOH). ¹H NMR (400 MHz, CDCl₃): δ

5.80 – 5.72 (m, 1H), 5.38 (d, $J = 7.8$ Hz, 1H), 4.65 – 4.60 (m, 1H), 4.44 (ddd, $J = 12.8, 10.0, 3.8$ Hz, 1H), 3.99 – 3.90 (m, 1H), 3.71 (dd, $J = 16.6, 7.2$ Hz, 2H), 3.39 (d, $J = 9.1$ Hz, 2H), 2.49 – 2.30 (m, 2H), 2.15 (dt, $J = 13.1, 6.4$ Hz, 4H), 2.01 – 1.93 (m, 4H), 1.24 (s, 9H), 1.22 (s, 9H), 1.17 (d, $J = 1.8$ Hz, 3H). ^{13}C NMR (176 MHz, CDCl_3): δ 199.6, 179.7, 172.8, 170.2, 155.4, 79.9, 75.2, 68.9, 60.4, 57.2, 56.4, 48.5, 40.4, 37.7, 30.6, 29.3, 28.5, 28.3, 25.1, 22.6, 18.6. MS (ESI) m/z : $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{25}\text{H}_{43}\text{N}_4\text{O}_7^+$ 511.3, found 511.2 (r_t : 7.09).

General procedure for the synthesis of amide derivatives 73, 79a, 79b, 84a and 84b (Method G). The intermediate methyl ester was left stirring in a mixture of MeOH and ammonia solution 33% (0.2 M, ratio 1:1) at room temperature for 12 h. The reaction mixture was concentrated under reduced pressure, coevaporated with hexane (3×10 mL) and dried under reduced pressure to quantitatively yield the amide derivative as a white solid. This material was used without further purification in the next step.

tert-Butyl ((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamate (**73**).

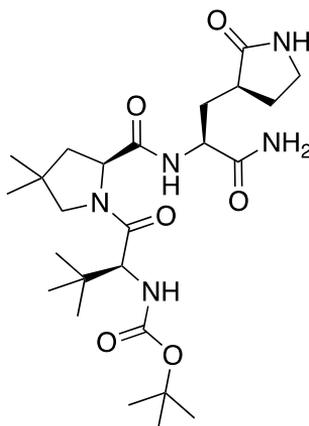


Chemical Formula: $\text{C}_{12}\text{H}_{21}\text{N}_3\text{O}_4$
Molecular Weight: 271.32

Compound **73** was prepared following the general procedure Method G and using methyl ((*S*)-2-((*tert*-butoxycarbonyl)amino)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **72** (yield: 850 mg, 96%). ^1H NMR (400 MHz, Methanol- d_4): δ 4.10 (dd, $J = 11.0, 3.8$ Hz, 1H), 3.40 - 3.31 (m, 2H), 2.48 (d, $J = 9.8, 4.5$ Hz,

1H), 2.41 - 2.28 (m, 1H), 2.04 (ddd, $J = 14.1, 11.1, 4.4$ Hz, 1H), 1.87 (dq, $J = 12.5, 8.8$ Hz, 1H), 1.74 (ddd, $J = 14.0, 10.3, 4.1$ Hz, 1H), 1.45 (s, 9H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{12}H_{22}N_3O_4^+$ 272.32, found 272.3.

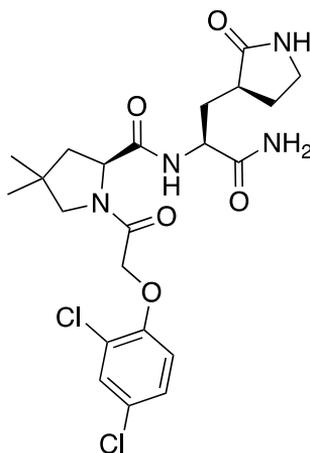
tert-Butyl ((*S*)-1-((*S*)-2-(((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-4,4-dimethylpyrrolidin-1-yl)-3,3-dimethyl-1-oxobutan-2-yl)carbamate (**79a**).



Chemical Formula: $C_{25}H_{43}N_5O_6$
Molecular Weight: 509.65

Compound **79a** was prepared following the general procedure Method G, using (*S*)-2-amino-3-((*S*)-2-oxopyrrolidin-3-yl)propanamide hydrochloride **74** and (*S*)-1-((*S*)-2-((*tert*-butoxycarbonyl)amino)-3,3-dimethylbutanoyl)-4,4-dimethylpyrrolidine-2-carboxylic acid **78a** (yield: 92 mg, 64%). 1H NMR (700 MHz, $CDCl_3$): δ 8.30 (s, 1H), 7.23 (s, 1H), 5.92 (s, 1H), 5.55 (s, 1H), 4.38 (dd, $J = 11.5, 4.1$ Hz, 1H), 4.30 (d, $J = 10.1$ Hz, 1H), 3.81 (d, $J = 9.9$ Hz, 1H), 3.69 – 3.66 (m, 1H), 3.52 (d, $J = 9.8$ Hz, 1H), 3.35 (d, $J = 8.3$ Hz, 2H), 3.26 – 3.19 (m, 1H), 2.49 (d, $J = 5.8$ Hz, 1H), 2.39 (dd, $J = 8.3, 3.6$ Hz, 1H), 2.14 – 2.09 (m, 1H), 2.01 (dd, $J = 12.1, 7.3$ Hz, 1H), 1.94 (d, $J = 9.2$ Hz, 1H), 1.90 – 1.84 (m, 2H), 1.41 (s, 9H), 1.13 (s, 3H), 1.01 (s, 3H), 1.00 (s, 9H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{25}H_{44}N_5O_6^+$ 510.66, found 510.7.

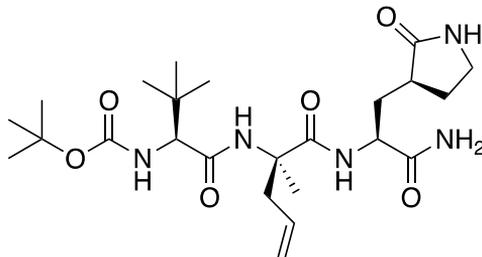
(S)-*N*-((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)-1-(2-(2,4-dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxamide (**79b**).



Chemical Formula: $C_{22}H_{28}Cl_2N_4O_5$
Molecular Weight: 499.39

Compound **79b** was prepared following the general procedure Method G, using (*S*)-2-amino-3-((*S*)-2-oxopyrrolidin-3-yl)propanamide hydrochloride **74** and (*S*)-1-(2-(2,4-dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxylic acid **78b** (yield: 76 mg, 53%). 1H NMR (700 MHz, $CDCl_3$): δ 8.85 (d, $J = 4.9$ Hz, 1H), 7.33 (d, $J = 2.5$ Hz, 1H), 7.13 (s, 1H), 7.08 (dd, $J = 8.8, 2.5$ Hz, 1H), 6.89 (t, $J = 6.7$ Hz, 1H), 4.80 (s, 2H), 4.46 – 4.41 (m, 1H), 4.24 – 4.17 (m, 1H), 3.55 – 3.44 (m, 2H), 3.38 – 3.33 (m, 2H), 2.43 – 2.36 (m, 2H), 2.13 (dt, $J = 14.7, 7.8$ Hz, 1H), 2.09 (dd, $J = 12.5, 8.0$ Hz, 1H), 1.92 – 1.86 (m, 2H), 1.85 – 1.78 (m, 2H), 1.71 – 1.67 (m, 1H), 1.13 (s, 3H), 1.04 (s, 3H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{22}H_{29}Cl_2N_4O_5^+$ 500.40, found 500.4.

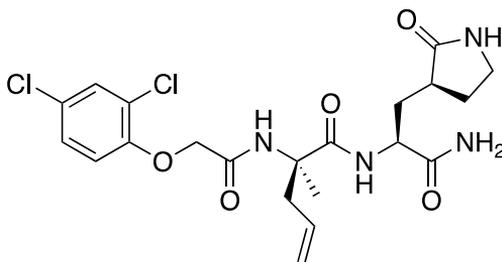
tert-Butyl ((*S*)-1-(((*S*)-1-(((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-2-methyl-1-oxopent-4-en-2-yl)amino)-3,3-dimethyl-1-oxobutan-2-yl)carbamate (**84a**).



Chemical Formula: C₂₄H₄₁N₅O₆
Molecular Weight: 495.62

Compound **84a** was prepared following the general procedure Method G and using methyl (6*S*,9*S*,12*S*)-9-allyl-6-(*tert*-butyl)-2,2,9-trimethyl-4,7,10-trioxo-12-(((*S*)-2-oxopyrrolidin-3-yl)methyl)-3-oxa-5,8,11-triazatridecan-13-oate **83a** (yield: 178 mg, 100%). ¹H NMR (700 MHz, CDCl₃): δ 8.09 (d, *J* = 6.4 Hz, 1H), 6.68 (s, 1H), 5.87 (s, 1H), 5.77 – 5.69 (m, 1H), 5.45 (s, 1H), 5.20 (t, *J* = 12.8 Hz, 2H), 5.06 (d, *J* = 5.2 Hz, 1H), 4.39 (dd, *J* = 9.9, 5.4 Hz, 1H), 3.69 (t, *J* = 5.7 Hz, 1H), 3.36 (t, *J* = 9.0 Hz, 1H), 3.33 (dd, *J* = 17.1, 8.5 Hz, 1H), 3.13 (s, 1H), 2.53 (t, *J* = 7.6 Hz, 2H), 2.37 (d, *J* = 2.4 Hz, 2H), 2.20 – 2.14 (m, 1H), 2.08 (d, *J* = 11.2 Hz, 1H), 1.90 (dd, *J* = 8.7, 6.3 Hz, 1H), 1.57 (s, 3H), 1.45 (s, 9H), 1.03 (s, 9H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₄H₄₂N₅O₆⁺ 496.63, found 496.6.

(*S*)-*N*-((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)-2-(2-(2,4-dichlorophenoxy)acetamido)-2-methylpent-4-enamide (**84b**).



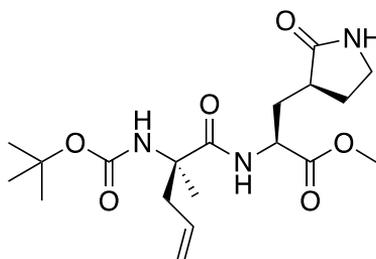
Chemical Formula: C₂₁H₂₆Cl₂N₄O₅
Molecular Weight: 485.36

Compound **84b** was prepared following the general procedure Method G and using methyl (*S*)-2-((*S*)-2-(2-(2,4-dichlorophenoxy)acetamido)-2-methylpent-4-enamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate **83b** (yield: 57 mg, 100%). ¹H NMR (700 MHz, CDCl₃): δ 8.75 (d, *J* = 5.4 Hz, 1H), 7.42 – 7.38 (m, 2H), 7.23 – 7.20 (m, 1H), 6.87 (d, *J* = 8.8 Hz, 1H), 5.78 (d, *J* = 3.8 Hz, 1H), 5.76 – 5.70 (m, 1H), 5.48 (s, 1H), 5.22 (dd, *J* = 26.7, 13.5 Hz, 2H), 4.60 (d, *J* = 14.3 Hz, 1H), 4.47 (t, *J* = 14.0 Hz, 2H), 4.32 (dt, *J* = 10.9, 5.3 Hz, 1H), 3.35 (d, *J* = 9.0 Hz, 2H), 2.60 (d, *J* = 7.3 Hz, 2H), 2.40 (dd, *J* = 17.0, 8.7 Hz, 2H), 2.17 – 2.08 (m, 1H), 2.05 – 1.96 (m, 1H), 1.91 – 1.83 (m, 1H), 1.61 (s, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₁H₂₇Cl₂N₄O₅⁺ 486.37, found 486.4.

General procedure for *tert*-butoxycarbonyl protecting group deprotection to afford intermediates **74 and **82**** (Method H). To a solution of the appropriate *N*-(*tert*-butoxycarbonyl) analogue **73** or **81** (1 eq.) in CH₂Cl₂ (0.2 M), 4 M HCl in 1,4-dioxane (10 eq.) was added dropwise at 0 °C. Then, the mixture was warmed to room temperature and left stirring overnight. The reaction mixture was concentrated under reduced pressure, coevaporated with hexane (3 × 10 mL) and dried under reduced pressure to quantitatively yield the amine hydrochloride as a white solid. This material was used without further purification in the next step.

General procedure for the synthesis of peptide esters **81, **83a** and **83b**** (Method I). In a three-necked oven-dried round-bottom flask, the acid intermediate (1 eq) was dissolved in anhydrous DMF (0.1 M) under a positive anhydrous nitrogen pressure, then COMU (1 eq.) and oxyma (1 eq.) were added. After 10 minutes, the appropriate amine hydrochloride (1.3 eq.) and DIPEA (2 eq.) were added to the mixture, which was left stirring at room temperature overnight. The reaction mixture was diluted with EtOAc (15 mL) and the organic phase was washed with a saturated citric acid solution (10 mL x 3), a saturated aqueous solution of NaHCO₃ (10 mL x 3) and brine (5 mL x 1). The collected organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a white solid.

*Methyl (S)-2-((S)-2-((tert-butoxycarbonyl)amino)-2-methylpent-4-enamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (**81**)*.

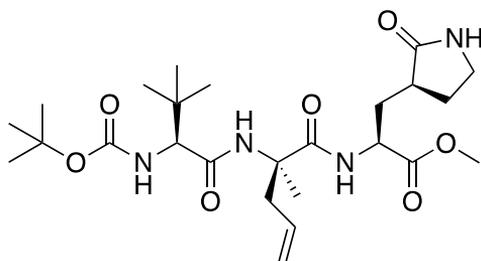


Chemical Formula: C₁₉H₃₁N₃O₆
Molecular Weight: 397.47

Compound **81** was prepared following the general procedure Method I, using methyl (S)-2-amino-3-((S)-2-oxopyrrolidin-3-yl)propanoate hydrochloride **63** and (S)-2-((tert-butoxycarbonyl)amino)-2-methylpent-4-enoic acid **80** (yield: 200 mg, 50%). ¹H NMR (700 MHz, CDCl₃): δ 5.97 (s, 1H), 5.74 (ddt, *J* = 17.5, 10.2, 7.4 Hz, 1H), 5.20 – 5.12 (m, 2H), 5.09 (s, 1H), 4.54 (s, 1H), 4.45 – 4.37 (m,

1H), 3.72 (s, 3H), 3.41 – 3.32 (m, 2H), 2.74 (s, 1H), 2.66 (s, 1H), 2.54 – 2.45 (m, 2H), 2.15 (ddd, $J = 14.3, 11.3, 5.5$ Hz, 1H), 1.95 – 1.83 (m, 2H), 1.47 (s, 3H), 1.41 (s, 8H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{19}H_{32}N_3O_6^+$ 398.48, found 398.5.

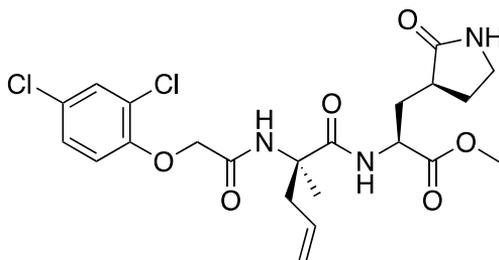
Methyl (6S,9S,12S)-9-allyl-6-(tert-butyl)-2,2,9-trimethyl-4,7,10-trioxo-12-(((S)-2-oxopyrrolidin-3-yl)methyl)-3-oxa-5,8,11-triazatridecan-13-oate (83a).



Chemical Formula: $C_{25}H_{42}N_4O_7$
Molecular Weight: 510.63

Compound **83a** was prepared following the general procedure Method I, using methyl (*S*)-2-((*S*)-2-((*tert*-butoxycarbonyl)amino)-2-methylpent-4-enamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate hydrochloride **82** and (*S*)-2-((*tert*-butoxycarbonyl)amino)-3,3-dimethylbutanoic acid **75a** (yield: 170 mg, 55%). 1H NMR (700 MHz, $CDCl_3$): δ 8.15 (s, 1H), 6.67 (s, 1H), 5.72 (s, 1H), 5.37 (s, 1H), 5.10 (d, $J = 13.8$ Hz, 2H), 4.40 (s, 1H), 3.72 (s, 3H), 3.35 (d, $J = 3.2$ Hz, 2H), 3.13 (s, 1H), 2.95 (s, 1H), 2.67 (d, $J = 12.4$ Hz, 1H), 2.46 (d, $J = 37.9$ Hz, 2H), 2.16 – 2.07 (m, 1H), 1.96 – 1.90 (m, 1H), 1.89 – 1.80 (m, 1H), 1.58 (s, 3H), 1.42 (s, 9H), 0.97 (s, 9H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{25}H_{43}N_4O_7^+$ 511.64, found 511.6.

Methyl (S)-2-((S)-2-(2-(2,4-dichlorophenoxy)acetamido)-2-methylpent-4-enamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (**83b**).



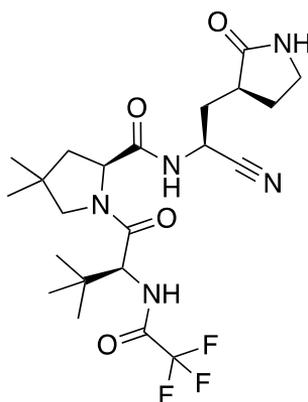
Chemical Formula: $C_{22}H_{27}Cl_2N_3O_6$
Molecular Weight: 500.37

Compound **83b** was prepared following the general procedure Method I, using methyl (S)-2-((S)-2-((*tert*-butoxycarbonyl)amino)-2-methylpent-4-enamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate hydrochloride **82** and 2-(2,4-dichlorophenoxy)acetic acid **67** (yield: 83 mg, 51%). 1H NMR (700 MHz, $CDCl_3$): δ 8.15 (s, 1H), 7.41 (d, $J = 14.1$ Hz, 2H), 7.20 (d, $J = 8.7$ Hz, 1H), 6.84 (d, $J = 8.6$ Hz, 1H), 5.79 (s, 1H), 5.77 – 5.68 (m, 1H), 5.16 (d, $J = 21.3$ Hz, 1H), 5.13 (d, $J = 12.3$ Hz, 1H), 4.46 (s, 1H), 4.45 (s, 2H), 3.73 (s, 3H), 3.35 (d, $J = 8.2$ Hz, 2H), 2.91 – 2.83 (m, 1H), 2.77 (s, 1H), 2.46 (s, 2H), 2.10 (s, 1H), 1.98 – 1.83 (m, 2H), 1.64 (s, 3H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{22}H_{28}Cl_2N_3O_6^+$ 501.41, found 501.4.

General procedure for the synthesis of tripeptide nitriles 50 and 52 (Method J). To a solution of amide (1 eq.) in CH_2Cl_2 (0.2 M), a solution of 4 M HCl in 1,4 dioxane (10 eq.) was added dropwise at $0^\circ C$. Upon complete addition, the reaction mixture was warmed to room temperature and stirred for 1 h. Then, the solvent mixture was removed under vacuum and the residue was washed with hexane (10 mL x 2) to quantitatively yield the amine hydrochloride as a white solid, that was used for next step without further purification. In an oven-dried three-necked round bottom flask the amine hydrochloride (1 eq.) was suspended

in anhydrous THF (0.2 M) under a positive pressure of anhydrous nitrogen and cooled to 0°C, then anhydrous pyridine (2 eq.) was added. After 30 min, trifluoroacetic anhydride (1 eq.) was added dropwise and the reaction mixture was stirred for 2 h at room temperature. Once observed by TLC monitoring, the complete conversion of the amino hydrochloride intermediate, anhydrous pyridine (4 eq.) and anhydrous THF (0.5 M) were added, and the mixture was cooled at 0°C. After 5 min, trifluoroacetic anhydride (2 eq.) was added and the reaction was left stirring for 18 h. The solvent mixture was removed under vacuum, the resulting crude was diluted with EtOAc and the organic phase was washed with 0.5 N HCl (20 mL x 3), saturated solution NaHCO₃ (20 mL x 1), dried over anhydrous Na₂SO₄, filtered, and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a white solid.

(S)-*N*-((*S*)-1-Cyano-2-((*S*)-2-oxopyrrolidin-3-yl)ethyl)-1-((*S*)-3,3-dimethyl-2-(2,2,2-trifluoroacetamido)butanoyl)-4,4-dimethylpyrrolidine-2-carboxamide (**50**).

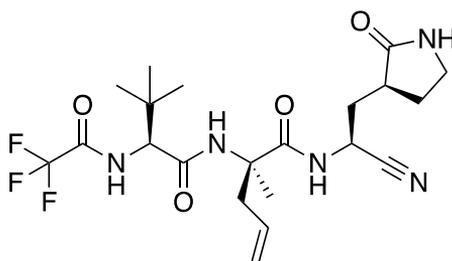


Chemical Formula: C₂₂H₃₂F₃N₅O₄
Molecular Weight: 487.52

Compound **50** was prepared following the general procedure Method J and using *tert*-butyl ((*S*)-1-((*S*)-2-(((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-

yl)propan-2-yl)carbamoyl)-4,4-dimethylpyrrolidin-1-yl)-3,3-dimethyl-1-oxobutan-2-yl)carbamate **79a** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 96:4 v/v; yield: 21 mg, 46%). ¹H NMR (700 MHz, CDCl₃): δ 8.44 (d, *J* = 5.9 Hz, 1H), 7.00 (d, *J* = 9.0 Hz, 1H), 4.77 (dt, *J* = 11.2, 5.5 Hz, 1H), 4.61 (d, *J* = 9.2 Hz, 1H), 4.35 (dd, *J* = 9.5, 7.9 Hz, 1H), 3.56 (d, *J* = 9.8 Hz, 1H), 3.44 – 3.37 (m, 4H), 2.65 – 2.57 (m, 1H), 2.47 – 2.41 (m, 1H), 2.32 (ddd, *J* = 14.4, 11.2, 7.9 Hz, 1H), 2.06 – 2.01 (m, 2H), 1.94 – 1.89 (m, 2H), 1.18 (s, 3H), 1.04 (s, 3H), 1.03 (s, 9H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₂H₃₃F₃N₅O₄⁺ 488.53, found 488.5.

(*S*)-*N*-((*S*)-1-Cyano-2-((*S*)-2-oxopyrrolidin-3-yl)ethyl)-2-((*S*)-3,3-dimethyl-2-(2,2,2-trifluoroacetamido)butanamido)-2-methylpent-4-enamide (**52**).



Chemical Formula: C₂₁H₃₀F₃N₅O₄
Molecular Weight: 473.50

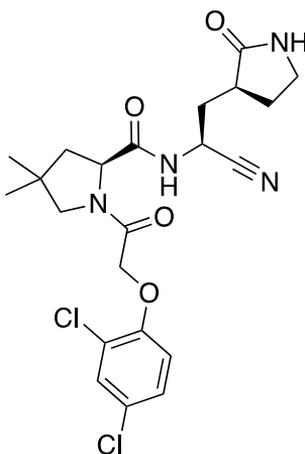
Compound **52** was prepared following the general procedure Method J and using tert-butyl ((*S*)-1-(((*S*)-1-(((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-2-methyl-1-oxopent-4-en-2-yl)amino)-3,3-dimethyl-1-oxobutan-2-yl)carbamate **84a** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 16 mg, 30%). ¹H NMR (700 MHz, CDCl₃): δ 9.00 (s, 1H), 7.22 (d, *J* = 8.3 Hz, 1H), 6.54 (d, 1H), 5.91 (s, 1H), 5.70 (dd, *J* = 16.6, 9.1 Hz, 1H), 5.15 (dd, *J* = 20.2, 13.6 Hz, 2H), 4.63 – 4.57 (m, 1H), 4.27 (d, *J* = 8.6 Hz, 1H), 3.39 (dd, *J* = 20.6, 12.0 Hz, 2H), 2.94 – 2.85 (m, 2H), 2.45 (t, *J* = 11.4 Hz, 2H), 2.27 (dd, *J* = 19.4, 13.8 Hz, 1H), 1.97 – 1.89 (m, 2H),

1.55 (s, 3H), 1.02 (s, 9H). MS (ESI) m/z: $[M + H]^+$ calcd for $C_{21}H_{31}F_3N_5O_4^+$ 474.50, found 474.5.

General procedure for the synthesis of peptide nitriles 51 and 53 (Method K).

In an oven-dried three-necked round bottom flask the amine hydrochloride (1 eq.) was suspended in anhydrous THF (0.2 M) under a positive pressure of anhydrous nitrogen and cooled to 0°C, then anhydrous pyridine (4 eq.) was added. After 5 min, trifluoroacetic anhydride (2 eq.) was added dropwise and the reaction mixture was stirred for 18 h at 0°C. The solvent mixture was removed under vacuum, the resulting crude was diluted with EtOAc and the organic phase was washed with 0.5 N HCl (20 mL x 3), saturated solution $NaHCO_3$ (20 mL x 1), dried over anhydrous Na_2SO_4 , filtered, and concentrated under vacuum. The reaction crude was purified by flash chromatography to yield the desired product as a white solid.

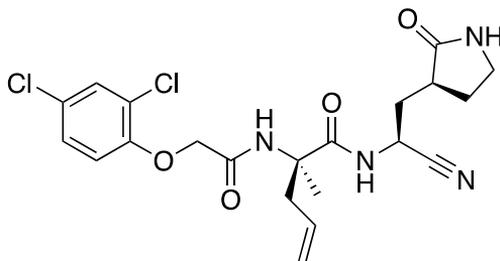
(S)-N-((S)-1-Cyano-2-((S)-2-oxopyrrolidin-3-yl)ethyl)-1-(2-(2,4-dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxamide (51).



Chemical Formula: $C_{22}H_{26}Cl_2N_4O_4$
Molecular Weight: 481.37

Compound **51** was prepared following the general procedure Method K and using (*S*)-*N*-((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)-1-(2-(2,4-dichlorophenoxy)acetyl)-4,4-dimethylpyrrolidine-2-carboxamide **79b** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 97:3 v/v; yield: 23 mg, 79%). ¹H NMR (700 MHz, CDCl₃): δ 8.38 (d, *J* = 6.1 Hz, 1H), 7.36 (d, *J* = 2.5 Hz, 1H), 7.18 (dd, *J* = 8.8, 2.5 Hz, 1H), 6.84 (d, *J* = 8.8 Hz, 1H), 5.65 (s, 1H), 4.80 (dt, *J* = 11.6, 6.0 Hz, 1H), 4.68 (dd, *J* = 42.6, 14.3 Hz, 2H), 4.39 (d, *J* = 8.4 Hz, 1H), 3.41 (dd, *J* = 45.5, 10.1 Hz, 2H), 3.31 (dt, *J* = 9.4, 8.0 Hz, 2H), 2.61 – 2.54 (m, 1H), 2.38 – 2.33 (m, 1H), 2.27 (ddd, *J* = 14.3, 10.7, 7.9 Hz, 1H), 2.01 – 1.98 (m, 1H), 1.98 – 1.94 (m, 2H), 1.88 – 1.81 (m, 1H), 1.17 (s, 3H), 1.03 (s, 3H). MS (ESI) *m/z*: [M + H]⁺ calcd for C₂₂H₂₇Cl₂N₄O₄⁺ 482.38, found 482.4.

(*S*)-*N*-((*S*)-1-Cyano-2-((*S*)-2-oxopyrrolidin-3-yl)ethyl)-2-(2-(2,4-dichlorophenoxy)acetamido)-2-methylpent-4-enamide (**53**).



Chemical Formula: C₂₁H₂₄Cl₂N₄O₄
Molecular Weight: 467.35

Compound **53** was prepared following the general procedure Method K and using (*S*)-*N*-((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)-2-(2-(2,4-dichlorophenoxy)acetamido)-2-methylpent-4-enamide **84b** (purification method: flash chromatography eluting CH₂Cl₂/MeOH 95:5 v/v; yield: 16 mg, 30%). ¹H NMR (700 MHz, CDCl₃): δ 9.00 (s, 1H), 7.22 (d, *J* = 8.3 Hz, 1H), 6.54 (d, 1H), 5.91 (s, 1H), 5.70 (dd, *J* = 16.6, 9.1 Hz, 1H), 5.15 (dd, *J* = 20.2, 13.6 Hz, 2H), 4.63 – 4.57 (m, 1H), 4.27 (d, *J* = 8.6 Hz, 1H), 3.39 (dd, *J* = 20.6, 12.0 Hz,

2H), 2.94 – 2.85 (m, 2H), 2.45 (t, $J = 11.4$ Hz, 2H), 2.27 (dd, $J = 19.4, 13.8$ Hz, 1H), 1.97 – 1.89 (m, 2H), 1.55 (s, 3H), 1.02 (s, 9H). MS (ESI) m/z : $[M + H]^+$ calcd for $C_{21}H_{25}Cl_2N_4O_4^+$ 468.35, found 468.3.

4.7.2. Molecular Docking

For docking calculations, we selected the crystal structures of SARS-CoV-2 3CL^{pro} (PDB ID: 7BQY)¹³ and MERS-CoV 3CL^{pro} (PDB ID: 4RSP)²¹ in complex with compound **13** and a peptide inhibitor, respectively. Prior to docking, the protein structure was prepared with the Protein Preparation Wizard tool within Maestro. In particular, bond orders were assigned and missing hydrogens added. A prediction of the receptor side chains ionization and tautomeric states was performed using Epik. Then, an optimization of the hydrogen-bonding network was carried out and the positions of the hydrogen atoms were minimized using the OPLS3e force field. Finally, both the co-crystallized water molecules and inhibitor were removed. Compounds' 3D structures were generated using the graphical interface of Maestro and prepared for docking using LigPrep. Calculations were performed using the Covalent Docking procedure implemented in Maestro in the Pose Prediction (thorough) mode. A grid box of $30 \text{ \AA} \times 30 \text{ \AA} \times 30 \text{ \AA}$ was centered on the reactive Cys¹⁴⁵ in SARS-CoV-2 3CL^{pro} and Cys¹⁴⁵ in MERS-CoV 3CL^{pro}. The 'nucleophilic addition to a double bond' reaction type was selected to account for the aldehyde warhead. For the refinement phase, the maximum number of poses to retain was set to 250, while the cutoff for minimization was set to 5.0 \AA . Macrocycle sampling and MM-GBSA scoring were disabled. Otherwise, default parameters were applied. For each ligand, 5 poses were finally retained and visually inspected.

4.7.3. Enzymatic, cell based and binding assays

Protein purification. The SARS-CoV-2 and MERS-CoV 3CL^{pro} proteins for the biochemical assay were expressed using a pET vector in *E. coli* cells BL21 (DE3) and the expression pellets were clarified by ultracentrifugation. The proteins were purified as previously described.²² Eluted fractions containing the 3CL^{pro} SARS-2 were pooled and subjected to buffer exchange (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.8) by using Amicon Ultra 15 centrifugal filters at 4000 x g, at 4 °C. 3CL^{pro} MERS-CoV protein was purified using a Ni-Sepharose column and by HiTrap Q HP column. Eluted fractions containing the target protein were pooled and subjected to buffer exchange (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.8) by using Amicon Ultra 15 centrifugal filters at 4000 x g, at 4 °C. Proteins purity was checked by SDS-PAGE analysis. The SARS-CoV02 3CL^{pro} for crystallographic studies was expressed in BL21 (DE) *E. coli* grown in YT medium using the pGEX vector provided by the Hilgenfeld's group and described in literature.¹⁴ Protein was purified using the protocol reported in literature.¹³

3CL^{pro} SARS-CoV-2 and MERS-CoV enzymatic assays. The 3CL^{pro} SARS-CoV-2 and MERS-CoV enzymes inhibition assays was performed with the substrate DABCYL-KTSAVLQ↓SGFRKM-EDANS (Bachem) in an assay buffer containing 20 mM Tris (pH 7.3), 100 mM NaCl, 5 mM TCEP, 0.1 % BSA and 1 mM EDTA.^{13,10} The proteins were preincubated for 30 minutes at 37 °C with different concentrations of compounds before the addition of the substrate. The signal was monitored after 15 minutes and 30 minutes of incubation for 3CL^{pro} SARS-CoV-2 and 3CL^{pro} MERS, respectively. Following enzymatic cleavage, generation of the fluorescent product was monitored (Ex 340 nm, Em 490 nm). Compound GC376 was used as positive control.

In the Kinetics assay for the determination of the MOA and K_i of compound **46** the fluorescence signal was followed for 60 minutes and acquired every 60 seconds. The apparent K_i was determined with the single-point equation as reported,²³ in triplicate and averaged. Initial enzymatic velocities in presence and absence of the inhibitor were determined by linear regression in the first 15 minutes of the generated curves.

SARS-CoV-2 viral replication assay. The African green monkey kidney cell line was previously engineered to constitutively express GFP (Vero E6-GFP) and were kindly provided by Janssen Pharmaceutical. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% v/v fetal beef serum (FBS; Gibco), 0.075% Sodium Bicarbonate (7.5% solution, Gibco) and 1x Pen-strep (Euroclone) and kept under 5% CO₂ on 37°C. SARS-CoV-2 strain BetaCov/Belgium/GHB-03021/2020 was provided by KU Leuven. All virus-related work was carried out in certified, high-containment biosafety level-3 facilities at the University of Cagliari. Cells were seeded at 10,000 cells/well in 96-well black cell-treated plates. The following day, cells were incubated with the control compounds at different concentrations and the virus at MOI 0.01. GC376 compound was used as positive control,⁴ in presence of 2 μ M Pgp inhibitor CP-100356.⁵ 72 hours post infection the media was removed and total well GFP fluorescence was measured with a Victor 3 with 485/535 nm excitation wavelength. The inhibition of viral replication was calculated as percentage of viral induced cytopathic effect on infected untreated controls, minus blanks (empty wells). EC₅₀ value was calculated with Prism 9. Version 9.1.2 via non-linear regression. The experiments represent average and standard deviation of at least two independent experiments in triplicate.

For quantification of viral copy number, by RT-PCR human epithelial lung adenocarcinoma cells Calu-3 cells were plated in 96 well plates (20,000

cells/well). The next day, drugs were added to cells and cells were infected with SARS-CoV-2 (MOI = 0.3) for one hour and subsequently, the virus inoculum was removed. The cells were placed into fresh media with the indicated drugs. At 48 hours post-infection, viral RNA was extracted from the supernatant with the QIAamp Viral RNA Mini Kit (Qiagen) following manufacturer instructions. One-step RT-qPCR was performed in 20 μ L to detect SARS-CoV-2 S-gene copy number using the primers: forward `_GTGTTTATTTTGCTTCCACTG`; reverse `_GGCTGAGAGACATATTCAAAA` with Luna Universal One-Step RT-qPCR Kit (New England Biolabs) according to manufacturer's instructions in a CFX-96 RT-PCR (Biorad). Results report the mean and standard deviation of two independent replicates.

Cytotoxicity assay. Vero E6-GFP cells were seeded at 10,000 cells/well in 96-well black cell-treated plates, the following day, cells were incubated with the compounds and virus MOI of 0.002, 72 hours post-infection the media was removed and total well GFP fluorescence was measured with a Victor 3 with 485/535 nm excitation wavelength. Cytotoxicity was calculated as a percentage of fluorescence of untreated controls, minus blanks (empty wells). CC_{50} value was calculated with Prism 9. Version 9.1.2 (225) via non-linear regression. The experiments represent average and standard deviation of at least two independent experiments in triplicate.

Differential scanning fluorimetry assay. The differential scanning fluorimetry (DSF) assay was performed in a buffer containing 20 mM HEPES pH 7.8 and 120 mM NaCl and 2 μ M of 3CL^{pro} protein was pre-incubated with 40 μ M of compounds for of 30 min. 5 \times SYPRO Orange dye was added to probe the thermal denaturation from 20 $^{\circ}$ C to 95 $^{\circ}$ C at a scan rate of 1.5 $^{\circ}$ C/min. The melt temperature (T_m) and the thermal shift (ΔT_m) was calculated as described.⁸ All

experiments were performed in triplicate, and the values are presented as mean \pm SD.

4.7.4. Crystallization studies

Crystallization of SARS-CoV-2 3CL^{pro} in complex with compounds, was carried out as previously described.¹³ Briefly, 3CL^{pro} at a concentration of 5 mg/mL in 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.8, 1 mM DTT was incubated for 1 h at RT with compounds at a final concentration of 5 mM. Crystallization experiments were then set up by seeding in sitting drops using the Morpheus® kit (Molecular Dimensions) with a Mosquito robot (STPlabtech Ltd., Melbourn Hertfordshire, UK). Crystals were flash frozen in liquid nitrogen after a few days of growth.

For compounds **41** and **42** the best diffracting crystals grew in condition G8: 0.1 M sodium formate, 0.1 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate, 0.1 M potassium sodium tartrate tetrahydrate, 0.1 M sodium oxamate, 0.1 M HEPES/MOPS, pH 7.5, 12.5% v/v MPD; 12.5% PEG 1000, 12.5% w/v PEG 3350.

For compound **46** the best diffracting crystals grew in condition B8: 0.09 M sodium fluoride, 0.09 M sodium bromide, 0.09 M sodium iodide, 0.1 M HEPES/MOPS, pH 7.5, 0.1 M HEPES/MOPS pH 7.5, 12.5% v/v MPD; 12.5% PEG 1000, 12.5% w/v PEG 3350.

For compound **48** the best diffracting crystals grew in condition D12: 0.12 M 1,6-Hexanediol, 0.12 M 1-butanol, 0.12 M 1,2-propanediol, 0.12 M 2-propanol, 0.12 M 1,4-butanediol, 0.12 M 1,3-propanediol, 0.1 M Tris/BICINE, pH 8.5, 12.5% v/v MPD; 12.5% PEG 1000, 12.5% w/v PEG 3350.

4.8. References Chapter 4

1. Xue, X. *et al.* Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design. *J. Virol.* **82**, 2515–2527 (2008).
2. Yang, H. *et al.* Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. *PLOS Biol.* **3**, e324 (2005).
3. Jin, Z. *et al.* Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nat.* / **582**, (2020).
4. Ma, C. *et al.* Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* 2020 308 **30**, 678–692 (2020).
5. Unoh, Y. *et al.* Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022**, 6499–6512 (2022).
6. Shereen, M. A., Khan, S., Kazmi, A., Bashir, N. & Siddique, R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* **24**, 91–98 (2020).
7. Fu, L. *et al.* Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* 2020 111 **11**, 1–8 (2020).
8. Qiao, J. *et al.* SARS-CoV-2 M pro inhibitors with antiviral activity in a

- transgenic mouse model . *Science* (80-.). **371**, 1374–1378 (2021).
9. 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.* **65**, 2716–2746 (2022).
 10. Owen, D. R. *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* (80-.). **374**, 1586–1593 (2021).
 11. Owen, D. R. *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* (80-.). **374**, 1586–1593 (2021).
 12. Leuw, P. de & Stephan, C. Protease inhibitors for the treatment of hepatitis C virus infection. *GMS Infect. Dis.* **5**, Doc08 (2017).
 13. Costanzi, E. *et al.* Structural and Biochemical Analysis of the Dual Inhibition of MG-132 against SARS-CoV-2 Main Protease (Mpro/3CLpro) and Human Cathepsin-L. *Int. J. Mol. Sci.* 2021, Vol. 22, Page 11779 **22**, 11779 (2021).
 14. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* (80-.). **368**, 409–412 (2020).
 15. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).

16. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D, Struct. Biol.* **75**, 861–877 (2019).
17. Prior, A. M. *et al.* Design, synthesis, and bioevaluation of viral 3C and 3C-like protease inhibitors. *Bioorg. Med. Chem. Lett.* **23**, 6317–6320 (2013).
18. Mandadapu, S. R. *et al.* Potent inhibition of norovirus 3CL protease by peptidyl α -ketoamides and α -ketoheterocycles. *Bioorg. Med. Chem. Lett.* **22**, 4820–4826 (2012).
19. Agbowuro, A. A. *et al.* Stereochemical basis for the anti-chlamydial activity of the phosphonate protease inhibitor JO146. *Tetrahedron* **74**, 1184–1190 (2018).
20. Jiang, X., Meyer, D., Baran, D., Cortés González, M. A. & Szabó, K. J. Trifluoromethylthiolation, Trifluoromethylation, and Arylation Reactions of Difluoro Enol Silyl Ethers. *J. Org. Chem.* **85**, 8311–8319 (2020).
21. Tomar, S.; Johnston, M. L.; st. John, S. E.; Osswald, H. L.; Nyalapatla, P. R.; Paul, L. N.; Ghosh, A. K.; Denison, M. R.; Mesecar, A. D. Ligand-Induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus Nsp5 Protease (3CL^{pro}). *J Biol Chem* **32**, 19403–19422 (2015).
22. Kuzikov, M.; Costanzi, E.; Reinshagen, J.; Esposito, F.; Vangeel, L.; Wolf, M.; Ellinger, B.; Claussen, C.; Geisslinger, G.; Corona, A.; Iaconis, D.; Talarico, C.; Manelfi, C.; Cannalire, R.; Rossetti, G.; Gossen, J.;

- Albani, S.; Musiani, F.; Herzog, K.; Ye, Y.; Giabbai, B.; Demitri, N.; Jochmans, D.; Jonghe, S. de; Rymenants, J.; Summa, V.; Tramontano, E.; Beccari, A. R.; Leyssen, P.; Storici, P.; Neyts, J.; Gribbon, P.; Zaliani, A. Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen. *ACS Pharmacol Transl Sci* **3**, 1096–1110 (2021).
23. Kuzmič, P.; Sideris, S.; Cregar, L. M.; Elrod, K. C.; Rice, K. D.; Janc, J. W. High-Throughput Screening of Enzyme Inhibitors: Automatic Determination of Tight-Binding Inhibition Constants. *Anal Biochem* **1**, 62–67 (2000).

CHAPTER 5

SYNTHESIS OF NIRMATREL VIR FOR
IN VITRO AND *IN VIVO* STUDIES

Chapter 5.

Synthesis and scale up of Nirmatrelvir for *in vitro* and *in vivo* efficacy studies

As mentioned above, Nirmatrelvir (**12**) is the first oral 3CL^{pro} inhibitor of SARS-COV-2 to enter clinical trials for COVID-19 and granted with emergency use authorization in December 2021/January 2022 (in a combination with metabolic booster Ritonavir, called Paxlovid[®]). Since its discovery, Nirmatrelvir represents the reference molecule for all research projects in the field. Structurally, Nirmatrelvir presents a C-terminal nitrile warhead for the reversible covalent bond formation with the protease catalytic Cys¹⁴⁵, and a trifluoroacetyl group as N-terminal capping (for the chemical structure, see Fig. 4.4 and Scheme 5.1). The P1 residue is a cyclic mimetic of glutamine having a γ -lactam ring in the side chain, that is a structural motif common to all potent inhibitors reported so far on Coronaviruses 3CL^{pro}. As P2-P3 sequence, there is a dipeptide based on bicycloproline and *tert*-butyl glycine, respectively, already present in the first generation HCV drug Boceprevir.¹

At the moment of the disclosure by Pfizer of the structure of the compound **12**, known at that time as PF-07321332, during the “American Chemical Society - ACS Spring Meeting in April 2021”, the synthesis as well as the *in vitro/in vivo* data were not yet published. In this context, we decided to synthesize compound **12** (Nirmatrelvir) in order to have in our laboratory the benchmark inhibitor of 3CL^{pro} to be used as reference in the development of new compounds. In particular, in this chapter we describe:

- the planning and implementation of a suitable synthetic route, in order to obtain the compound **12** in gram scale;

- *in vitro* antiviral activity in enzymatic and cell-based assays, evaluated by our collaborators at University of Cagliari (UNICA) and “Istituto Nazionale di Genetica Molecolare” (INGM) of Milan in their assay systems;
- *in vivo* antiviral studies on an innovative animal model of COVID-19, carried out by our collaborators at “Vita-Salute San Raffaele” (VSSR) University of Milan.
- formulation studies, with the aim of obtaining a vehicle suitable for the formulation of new compounds to be used for the oral or intraperitoneal treatment of SARS-CoV-2 infected mice.

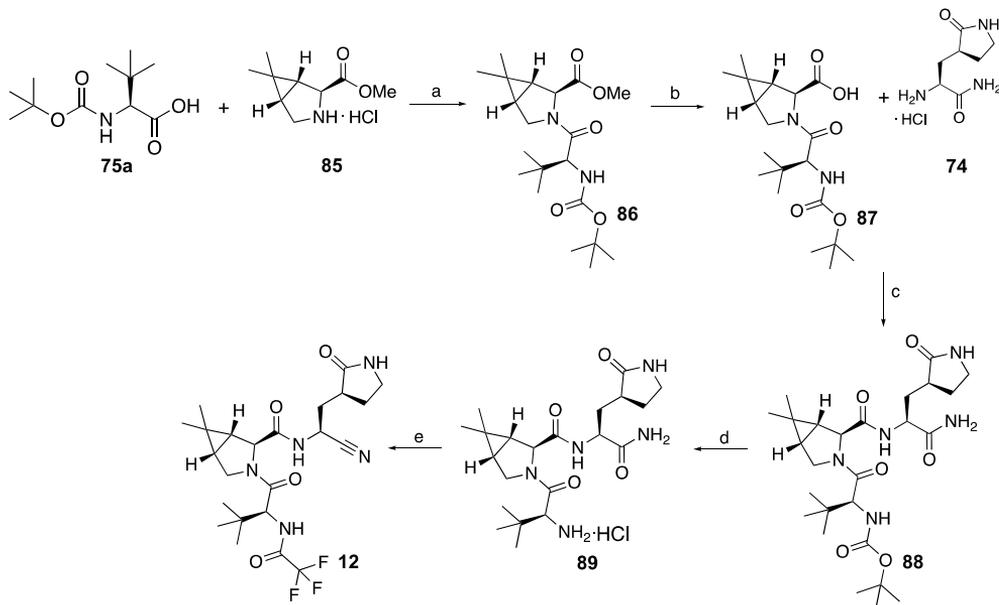
5.1. Synthesis and scale-up of benchmark inhibitor Nirmatrelvir **12**

Based on previous experience acquired by the synthesis of similar peptidomimetic inhibitors investigated in our laboratory (Chapter 4), I applied a "multistep" convergent approach, by a Boc-based procedure in solution, to afford the target compound **12**. In details, two "building blocks", respectively the dipeptide acid P3-P2 and the residue P1, were firstly generated separately and then assembled into a single step to generate an advanced intermediate, that has been further modified to introduce the nitrile electrophilic warhead and the trifluoroacetyl group as N-terminal capping.

The description of the synthesis of **12** is depicted in scheme 5.1 The bicycloproline hydrochloride amine **85** and the *tert*-butyl glycine acid **75a**, which respectively represent the commercially available precursors of the amino acids P2 and P3, were coupled in an amide bond formation reaction, by using HBTU as coupling reagent and DIPEA as a base in dry CH₂Cl₂ at room temperature, to provide the intermediate dipeptide **86**, with a high yield of 80% after flash chromatography purification. Saponification of ester **86**, by using 1N LiOH in

THF at room temperature, afforded the corresponding carboxylic acid **87** in quantitative yield.

Scheme 5.1. Synthesis of target tripeptide **12** (approved drug Nirmatrelvir)^a



^aReagents & conditions: a) HBTU, DIPEA, dry CH₂Cl₂, rt, 16 h, 78%; b) 1N aq. LiOH/THF (1:1), rt, 2 h, quantitative; c) HBTU, DIPEA, dry CH₂Cl₂/DMF, rt, 3 h, 75%; d) 4N HCl in 1,4-dioxane/CH₂Cl₂ (1:1), 0 °C to RT, 2 h, 98%; e) a) TFAA, dry Py, dry CH₂Cl₂, 0 °C to rt, 2 h; b) TFAA, dry Py, dry CH₂Cl₂, 0 °C to RT, 15 h, 40%.

The amine **74**, previously prepared as shown in Scheme 4.4 (Chapter 4, Paragraph 4.2.2), and the dipeptide acid **87** were then assembled by a peptide coupling reaction, conducted using the same conditions described above, to provide the tripeptide **88** with a yield of 75% after purification. Intermediate **88** was then subjected to an acid deprotection reaction of the N-terminal portion to remove Boc protecting group and quantitatively afford the corresponding tripeptide amine hydrochloride salt **89**. Finally, I set up a "two-steps-one-pot" procedure to simultaneously allow the capping of the N-terminal amine with the trifluoroacetyl moiety and the dehydration of the C-terminal carboxamide to the

desired nitrile. In particular, the intermediate **89** was acylated in the presence of TFAA, using dry pyridine as a base in CH₂Cl₂ in a nitrogen atmosphere at 0 ° C, and subsequently, additional TFAA and pyridine were added to promote the dehydration reaction and obtain the final compound **12**, with a 40% yield over two-steps and after purification by preparative HPLC. Moreover, the overall yield calculated for the whole convergent process was 30%.

Subsequently, in November 2021 to be precise, Owen and co-workers at Pfizer published the discovery of Nirmatrelvir, reporting on the design, synthesis, antiviral efficacy on enzymes and cell based assays, selectivity against a panel of cathepsins enzymes, pharmacokinetic in different species, toxicity and *in vivo* efficacy in animal model.¹ Regarding the synthetic process, the authors described a convergent approach very similar to our procedure described above. In particular, the synthesis of the P1 building block is almost identical between the two approaches, while some differences were observed in the preparation of the P2-P3 building block and in the nitrile warhead formation reaction.¹ In a head-to-head comparison, the in-house procedure has a lower number of steps, showing an overall process yield of 30%, compared to an overall yield of 45% of the Pfizer procedure. However, our planned synthesis was suitable to a milligram-to-gram scale-up of Nirmatrelvir **12**, thus allowing to successfully achieve the initial target amount for *in vitro* antiviral activities and PK characterization (~100 mg) and, subsequently, the large amount of Nirmatrelvir **12** required for *in vivo* efficacy study in the mouse model (2.5 g).

5.2. *In vitro* biological evaluation of Nirmatrelvir **12**

Before the Pfizer disclosure of biological results of Nirmatrelvir, *in vitro* assays were performed to evaluate the biological activity of **12** and to have a reference

compound for *in vitro* screening of our designed 3CL^{pro} inhibitors, as shown in the previous Chapter 4 (Table 4.1).

In particular, enzymatic assays against SARS-CoV-2 and MERS-CoV 3CL^{pro}s were performed at UNICA, by team of Prof. Enzo Tramontano, to evaluate enzymatic inhibitory potency and the broad-spectrum activity of **12**. At this regard, a FRET-based biochemical assay was carried out by preincubating the protease with the compound for 30 minutes at 37°C (as reported in Chapter 4, Paragraph 4.6), and by directly adding the substrate to the reaction, following the published procedure.¹ The commercially available SARS-CoV-2 3CL^{pro} inhibitor GC376 **14** was used as positive control in both procedures, showing an IC₅₀ of 0.14 nM and 4.8 nM, respectively. In both assays, the compound **12** was able to inhibit the activity of the SARS-CoV-2 3CL^{pro} in a dose-dependent manner, respectively showing an IC₅₀ of 50 nM and 14 nM, in line to published data (lit. IC₅₀ = 15 nM). Moreover, **12** resulted significantly more potent on the MERS 3CL^{pro} with an IC₅₀ = 0.22 nM respect to the literature data (lit. IC₅₀ = 100 nM) (Table 5.1).

Table 5.1. Enzymatic and antiviral activity in Vero E6 cell-based assays for **12**^a

3CL ^{pro} IC ₅₀ (nM) ^a		EC ₅₀ (μM) Vero E6 (SI) ^b		CC ₅₀ (μM) ^c Vero E6
SARS-CoV-2	MERS	SARS-CoV-2 ^c	SARS-CoV-2 ^d	
50 ± 3 ^f	0.22 ± 0.001 ^g	1.02 ± 1.4 ^h (> 98)	0.04 ± 0.12 ⁱ (> 2500)	> 100 ^j

^aCompound concentration required to reduce the 3CL^{pro} activity by 50%. Data obtained in presence of TCEP (5 mM) and BSA (0.1%). ^bSI is the selectivity index calculated as the CC₅₀/EC₅₀ ratio. ^cCompound concentration required to reduce the SARS-CoV-2 cytopathic effect in Vero E6 cell by 50%. ^dEC₅₀ in presence of P-gp inhibitor (CP-100356) at the fixed concentration of 2 μM. ^eCompound concentration required to reduce Vero E6 cell viability by

50%. ^fLit. IC₅₀: 15 nM. ^gLit. IC₅₀: 100 nM. ^hLit EC₅₀: 4.3 μM. ⁱLit. EC₅₀: 0.075 μM. ^jLit. CC₅₀ > 100 μM.¹

The teamwork of Prof. Enzo Tramontano then carried out a phenotypic assay, in order to: i) evaluate antiviral activity of **12** at a time when literature data were not yet known; ii) provide a potent reference compound for *in vitro* screening of our designed compounds, as introduced in previous Chapter 4, Paragraph 4.6. Thus, the antiviral activity of **12** against SARS-CoV-2 was evaluated in Vero E6 cells (CPE assay), infected with the ancestral SARS-CoV-2 strain, also namely as wild-type or D614G. The experiments were carried out in absence and in presence of the commercial P-gp inhibitors CP-100356 at 2 μM, due to high levels of P-gp efflux pump in the Vero E6 cells. The compound **12** showed to inhibit the viral replication with a EC₅₀ value of 1.02 μM (w/o CP-100356) and 0.04 μM (w/ CP-100356), without toxicity up to 100 μM (Table 5.1). These results are in line to literature data, with an EC₅₀ 4.48 μM and EC₅₀ of 0.0745 μM, respectively in absence and in presence of CP-100356.

Further phenotypic assays were performed into Prof. De Francesco lab at INGM, considering a different cellular line, namely Human Embryonic Kidney 293 (HEK-293)T-hACE2 cells. Specifically designed for COVID-19 studies, the HEK-293T-hACE2 cells are more sensitive to SARS-CoV-2, since they have been engineered to stably overexpress the host SARS-CoV-2 receptor, human ACE2, and host TMPRSS2 protease. In addition to a major sensitive towards SARS-CoV-2, the use of HEK-293T-hACE2 cells avoid the issue related to Vero cells, which are known to express high levels of an efflux pump P-gp and can pump compounds out from cells, thus resulting in a lower amount of drug available intracellularly. The antiviral activity was evaluated through two different parameters: i) quantification of SARS-CoV-2 RNA in the supernatant

after 72 h of infection (shown to be the best time point for qPCR detection); ii) protection from CPE after 5 days of infection (shown to be the best time point for CPE protection assay). Moreover, the antiviral activity was evaluated against three different SARS-CoV-2 strains: D614G (wild-type), B.1.617.2 (Delta), selected on basis of its high pathogenesis, and B.1.1.529 (Omicron BA.1), which was the most widespread SARS-CoV-2 variant at the time.

Table 5.2. Antiviral activity in HEK-293T-hACE2 cell-based assay for **12**^a

EC ₅₀ (μM) HEK-293T-hACE2 (SI) ^a						CC ₅₀ (μM) ^c HEK-293T-hACE2
SARS-CoV-2 ^b						
D614G		B.1.617.2		B.1.1.529		
qPCR	CPE assay	qPCR	CPE assay	qPCR	CPE assay	
0.058 (> 172)	0.044 (> 227)	0.026 (> 385)	0.046 (> 217)	0.071 (> 141)	0.026 (> 385)	> 10

^aSI is the selectivity index calculated as the CC₅₀/EC₅₀ ratio. ^bEC₅₀ to inhibit viral replication and cytopathic effect of SARS-CoV-2 in HEK-293T-hACE2 cell. ^cCompound concentration required to reduce HEK-293T-hACE2 cell viability by 50%.

The experiments showed that **12** prevents death of HEK293T-hACE2 cells infected with the SARS-CoV-2 variants D614G, B.1.617.2 (Delta) and B.1.1.529 (Omicron BA.1) with a mean IC₅₀ value of 33 nM, and prevented SARS-CoV-2 RNA release with a mean IC₅₀ value of 54 nM. All the compounds no showed cellular toxicity at 10 μM (Table 5.2). The CPE protection assay provided similar EC₅₀ values and equivalent result for all the variants. Therefore, the experiments described above on HEK293T-hACE2 cells allowed our collaborators to define the final protocol to test newly designed 3CL^{pro} inhibitors, using a more reliable cell line for the SARS-CoV-2 infection.

5.3. *In vivo* efficacy studies of Nirmatrelvir **12** in an innovative COVID-19 mouse model

Our collaborators Prof. Luca Guidotti and Prof. Matteo Iannacone at VSSR University developed a novel mouse model of COVID-19 in order to provide a tractable small animal model, capable to faithfully reproduce the human disease and carry out a reliable *in vivo* screening of new compounds active against SARS-CoV-2 *in vitro* assays. Therefore, they tested the compound **12** in their animal model. In particular, in this paragraph it will describe:

- an innovative novel K18-hACE2 transgenic mouse model of SARS-CoV-2 infection by aerosol exposure;
- *in vivo* efficacy at different concentration of **12** in the novel K18-hACE2 transgenic mouse model of COVID-19;
- assessment of the impact of **12** treatment on the development of antiviral adaptive immunity in the same animal model.

5.3.1. Administration of aerosolized SARS-CoV-2 to K18-hACE2 transgenic mice as innovative COVID-19 mouse model

As mentioned above, Prof. Luca Guidotti and Prof. Matteo Iannacone developed in their laboratories at VSSR an innovative mouse model of COVID-19, able to faithfully reproduce the human clinical signs of COVID-19 and provide a reliable model for testing new potential clinical candidates against SARS-CoV-2.

Specifically, the novelty of this mouse model was related to a controlled administration of aerosolized (AR) SARS-CoV-2 respect to the conventional intranasal (IN) inoculation (Fig. 5.1).² The issue linked to the typical IN-infect was related to severe cases of fatal encephalitis among infected mice,³ due to the high level of virus in the central nervous system (CNS). Therefore, this

significant viral neuroinvasion rarely occurs in patients with COVID-19, representing a critical limit to the reliability of the IN-infected mouse model, as well as manifesting problems for long-term studies due to the early death of mice. In this context, the study at VSSR described here was undertaken. Thus, K18-hACE2 transgenic mice, which express hACE2 under the control of the cytokeratin 18 (KRT18) promoter,⁴ were exposed to aerosolized virus through a nose-only inhalation tower system, under controlled pressure, temperature and humidity conditions. Moreover, the mice breathe spontaneously, since they are held by a neck clip positioned between the base of the skull and the shoulders. A target dose of 1×10^5 tissue culture infectious dose 50 (TCID₅₀) was used for the infection of K18-hACE2 mouse, through 20-30 minutes of aerosol administration or a liquid suspension to be inoculated via intranasal under deep anesthesia.

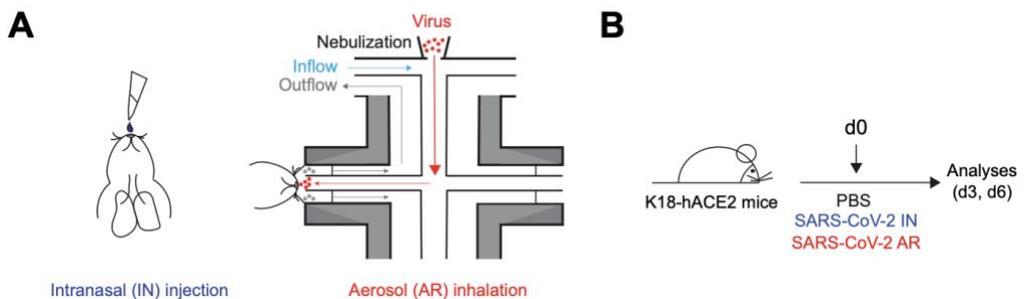


Figure 5.1. (A) Schematic representation of the two modalities of infection on K18-hACE2 mice: IN injection (on the left) and AR inhalation (on the right). On the right, representation of administration of aerosolized virus to mouse placed in the nose-only allay restrainer on the inhalation chamber is shown. Viral particles are shown in red; primary flow set to (0.5 liter/min per port) and, mouse breathing outflow are shown in light-blue and in grey, respectively.²

Following IN-infection, a significant body weight loss was observed by day 6 post infection. Indeed, ~80% of them had died, while the remaining showed severe clinical signs. The severe disease was associated to a deep viral neuroinvasion due to high levels of virus RNA and viral antigens in the brain, in

addition to a significant recruitment of immune cells, such as B-cells, T-cells, monocytes and eosinophils. Immunohistochemical and immunofluorescence assays highlight the diffuse presence of the SARS-CoV-2 nucleoprotein (N-CoV-2) in the brain of IN-infected, excluding only the cerebellum. These assays confirmed the high CNS infection, responsible of fatal encephalitis in IN-infected mice. On the other hand, SARS-CoV-2 RNA and viral antigens were not detected in the brain of AR-infected mice, that did not show weight loss or critical clinical signs. Therefore, high amounts of SARS-CoV-2 RNA and infectious virus were detected in the lower respiratory tract of AR-infected mice, comparable to IN-infection model, showing that AR administration is able to cause an efficient respiratory infection in the lung of SARS-CoV-2 infected K18-hACE2 transgenic mice. The robust viral replication in the respiratory tract was also associated to anosmia, signs of airway obstruction and platelet aggregation with fibrin deposition in the lung, which is related to inflammation processes typical of the most sever forms of COVID-19. In addition, AR-infected mice showed histopathological changes, immune response and transcriptional signs similar to SARS-CoV-2 patients. The immune infiltration was increased in the lung, showing a high number of CD4⁺ and CD8⁺ T cells and in monocytes, while genes up-regulated in patients with COVID-19 were found in the lungs of AR-infected mice, especially related to leukocyte trafficking, antiviral response induced by type I IFN and tumor necrosis factor.⁵⁻⁷

In summary, the administration of aerosolized SARS-CoV-2 to K28-hAC2 transgenic mice is able to reproduce a reliable *in vivo* model of COVID-19, characterized by a more robust infection, pulmonary cell infiltration and inflammation, and morbidity, without showing a fatal viral neuroinvasion, typical of IN administration.

5.3.2. *In vivo* efficacy studies

A novel *in vitro* or *in vivo* model requires a validation step to provide a useful tool for the drug discovery. The validation step is needed to verify i) reliability, i.e. that it yields 'reproducible' results; ii) relevance and pertinence, i.e. that it is meaningful and useful for the intended purpose; iii) efficacy, i.e. that it is truly predictive. In this regard, we decided to test **12** on the K18-hACE2 transgenic mouse model developed by Prof. Luca Guidotti and co-workers, in order to compare experimental data obtained in-house with data published in literature.² *In vivo* study of Nirmatrelvir was reported by Pfizer on BALB/c mice, previously infected with SARS-CoV-2 by IN inoculation and then treated twice daily with Nirmatrelvir at two different doses: 300 mg/kg and 1000 mg/kg. Notably, both doses reported by Pfizer require large amount of **12** for *in vivo* treatment. Considering that an antiviral activity was shown at 300 mg/kg, our collaborators selected the minor dosage at 300mg/kg of **12** to carry out the efficacy study on novel K18-hACE2 transgenic mouse model. Although the dosage and the vehicle of formulation (0.5% methylcellulose in water, 2% Tween80) were equal to the data reported in literature,¹ solubility issues were observed about the formulation, considerably due to the lipophilic nature of the compound and the high concentration of **12** in the vehicle. In this regard, stability tests of **12** were investigated in different formulations (Paragraph 5.4), in parallel to *in vivo* studies.

As depicted in Fig. 5.2, the experiment was carried out on 8-week-old male mice, divided into three groups, as described below:

- mock-treated group, including two uninfected male mice, used as control;
- placebo group, including four male mice infected through aerosol administration of a target dose of 2×10^5 TCID₅₀ of SARS-CoV-2 D614G and treated with the vehicle, in absence of **12**;

- treated group: four male mice infected with SARS-CoV-2 and treated with 300 mg/kg (mpk) of **12** via oral gavage, as reported in literature.

The effects of the infection and treatment with the drug were assessed through different methods: body weighting by plethysmography, histopathological evaluation of the lung tissue and quantification of the viral titer at different compartments, including lung, brain, nasal turbinate, olfactory bulb, blood. While the first two methods were useful to assess the inhibitor ability to protect against the damage of infection, the last one allowed to quantify the antiviral activity exerted by compound *in vivo*. The measure of mice weight was carried out at zero time (t_0) before the virus exposure, after 4 h pi, and then daily for four days until sacrifice, as shown in Fig. 5.2.

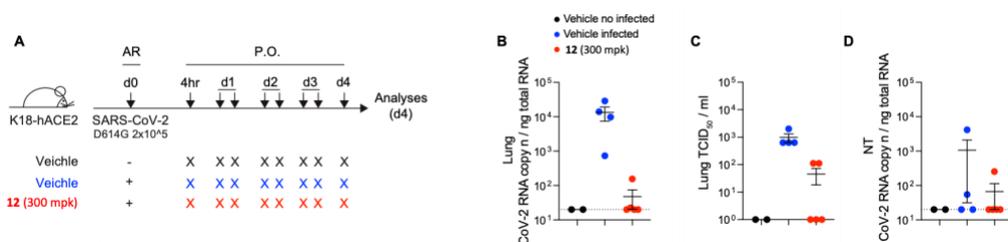


Figure 5.2. A) Schematic representation of the experimental set up. Non-anesthetized K18-hACE2 transgenic mice were infected with a target dose of 2×10^5 TCID₅₀ of SARS-CoV-2 D614G through aerosol exposure.² Infected mice were treated with 300 mg/kg (mpk) of Nirmatrelvir (red symbols, n = 4) or vehicle (blue symbols, n = 4) for six times by oral gavage starting 4h post infection (pi), and every 12 hours thereafter. Mock-treated mice were used as control (black symbols, n = 2). (B) Quantification of SARS-CoV-2 RNA and (C) viral titers in the lung 3 days after infection. RNA values are expressed as copy number per ng of total RNA and the limit of detection is indicated as a dotted line. Viral titers were determined by median tissue culture infectious dose (TCID₅₀). (D) Quantification of SARS-CoV-2 RNA in the nasal turbinates 3 days after infection. RNA values are expressed as copy number per ng of total RNA and the limit of detection is indicated as a dotted line.

As expected, the mock-mice, neither infected nor treated with the drug nor placebo, showed a constant body weight from t_0 to day 4. At the same time range, the infected and untreated mice (placebo group) showed a marked loss in weight

over time, compared to the drug-treated mice. Notably, the virus infection caused a weight loss in the antiviral-treated mice as well, but this was still modest or with a lower trend respect to the placebo group of mice, which showed a constant and marked loss over the four days of measurement. Histopathological analyses were performed to evaluate the efficacy of **12** to protect the lung tissue from damage due to viral infection. The results of analysis were expressed as histopathological score, considering a range from 0 to 4: the value of 0 indicated a healthy lung condition, while a score of 4 was characterized by inflammatory state and areas of tissue necrosis. Consistently, the mock-mice showed a clinical score of 0 for whole time of experiment, whereas in the infected and untreated (placebo) mice showed a progressive necrosis of the lung tissues, with 50% mice at an inflammatory state. However, the treated mice showed different cases: i) healthy lungs in 3 mice (score = 0), indicating that the drug has completely protected against infection; ii) a mouse showed a tissue a score of 2, thus there was partial efficacy; iii) a mouse showed a score of 4, indicating that the drug did not work.

Finally, the viral titer was measured after sacrificing the mice, and at autopsy the virus concentration in the lungs (Fig. 5.2B), brain, nasal turbinate (Fig. 5.2D), olfactory bulbs and plasma was quantified by using qPCR method. In addition, the viral titer in the lungs was expressed as TCID₅₀, correspondent to the amount of virus required to cause a cytopathic effect in 50 % of the cells (Fig. 5.2C).

Consistent with the above description of the AR infection method in mouse model, the TCID₅₀ measured in the brain showed that the neuroinvasion was not developed, indeed viral RNA was not detected in the neuronal cells of three groups of animals. On the other hand, high levels of viral RNA were found in the lung cells of the placebo group, sign of a robust respiratory infection, while a drastic reduction of viral RNA was observed in the drug-treated mice, consistent with the data reported in literature. In details, the RNA copy number was for three

mice above 100,000 and for a mouse around 1,000 in the lungs of the untreated infected mice, while an appreciable reduction in copy number was detected in the treated infected mice. Indeed, the amount of RNA is undetectable in three mice, while it is below 1,000 in a mouse (Fig. 5.2B,C).

As well in the brain, the analysis of viral titers in blood and other tissue, such as nasal turbinate (Fig. 5.2D), showed a number of RNA copies below the limits of detectability for each mice group. These results showed that the viral infection was predominantly developed in the respiratory tract and lungs, and confirmed further the data supporting the development of AR-mice model.

Consistent with the data described above, the treatment with **12** (po 300 mpk) allowed to successfully verify the efficacy of the COVID-19 animal model developed at VSSR University. However, in accordance with solubility issues observed when Nirmatrelvir is dosed at 300 mpk in the vehicle of formulation, a following *in vivo* study was performed by Prof. Luca Guidotti and co-workers by dosing Nirmatrelvir at a lower dosage of 150 mpk. Specifically, anesthetized mice were treated po twice daily with 150 mpk, using the same vehicle of the previous experiment and reported in literature¹ (0.5% methylcellulose in water, 2% Tween80). Although the decrease of the dose was not sufficient to achieve a complete solubility of the compound in the vehicle, the formulation appeared as a sharp micro-suspension, without a visible precipitate. Similar to the experiment described above, 8-week-old K18-hACE2 transgenic mice were infected by aerosol exposure of a SARS-CoV-2 B.1.1.529 at TCID₅₀ of 2×10^5 , under controlled pressure, temperature and humidity conditions (Fig. 5.3).

A group of 4 mice was treated twice daily with 150 mpk of **12** via oral gavage, while another group of 4 mice was treated with vehicle, starting at 4 hours pi and twice daily until day 3 pi. In addition, 3 mock-treated mice were used as control. As expected, K18-hACE2 mice did not show body weight loss. In line with the previous study, no detectable viral RNA was found in blood and in the brain.

High levels of SARS-CoV-2 RNA were detected in the lung tissue and in the nasal turbinates (Fig. 5.3B,D) consistent with a robust SARS-CoV-2 replication. In opposite, undetectable viral RNA was detected in the same tissues of Nirmatrelvir-treated mice. The results described herein supported that the treatment of Nirmatrelvir at 150 mpk is able to show an efficient antiviral activity, compared to the higher dose of 300 mpk, in addition to advantages of solubility for the formulation of **12**. According to this, the assay conditions described above were set out to provide a well-characterized mouse model of COVID-19, based on the controlled administration of aerosolized SARS-CoV-2 to K18-hACE2 transgenic mice and treatment with **12** at 150 mpk po. This model was used for further studies about the effects of Nirmatrelvir treatment, as it will be discussed in the next section.

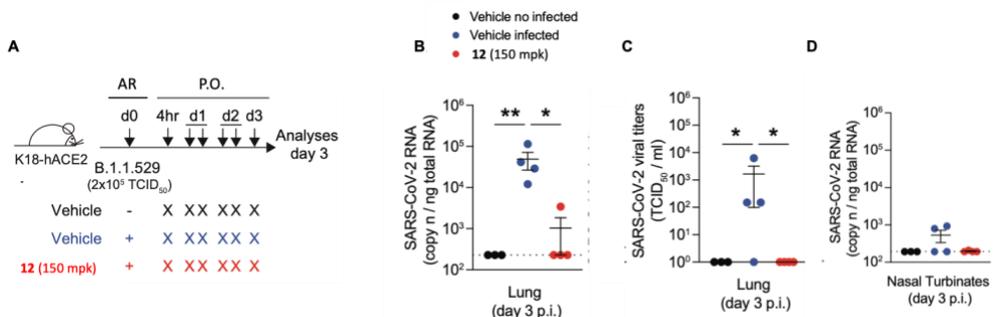


Figure 5.3. A) Schematic representation of the experimental set up. Non-anesthetized K18-hACE2 transgenic mice were infected with a target dose of 2×10^5 TCID₅₀ of SARS-CoV-2 D614G through aerosol exposure.² Infected mice were treated with 150 mpk of Nirmatrelvir (red symbols, n = 4) or vehicle (blue symbols, n = 4) for six times by oral gavage starting 4 h post infection, and every 12 h thereafter. Mock-treated mice were used as control (black symbols, n = 2). (B) Quantification of SARS-CoV-2 RNA and (C) viral titers in the lung 3 days after infection. RNA values are expressed as copy number per ng of total RNA and the limit of detection is indicated as a dotted line. Viral titers were determined by median tissue culture infectious dose (TCID₅₀). (D) Quantification of SARS-CoV-2 RNA in the nasal turbinates 3 days after infection. RNA values are expressed as copy number per ng of total RNA and the limit of detection is indicated as a dotted line. * *p*-value < 0.05, ** *p*-value < 0.01.

In parallel, the PK experiment was performed by Dr. Camilla Montesano of the “La Sapienza” University, in order to determine the plasmatic concentration of **12** in samples derived from treated animals during the two experiments at 300 mpk and 150 mpk carried out at VSSR. Firstly, an Ultrapformance Liquid Chromatography-Mass Spectrometry (UPLC-MS) method has been set up to obtain the calibration curve and allowed quantification of the molecule found in the plasma samples (using as analytical standard **12**, synthetized as described in Scheme 5.1).

Table 5.1. Plasmatic concentration of **12**^a

K18hACE2 MOUSE	TREATMENT	LUNG	PLASMA EXPOSURE
	+ 4 h	Virus RNA copy n/ng tot RNA	Concentration (μM)
1	-	ND	-
2	-	ND	-
3	Vehicle	9.72 x 10 ³	-
4	Vehicle	2.88 x 10 ⁴	-
5	Vehicle	1.40 x 10 ⁴	-
6	Vehicle	7.35 x 10 ²	-
7	12 300 mpk po	ND	11.96
8	12 300 mpk po	ND	1.613
9	12 300 mpk po	1.56 x 10 ²	5.369
10	12 300 mpk po	1.56 x 10 ²	4.046

K18hACE2 MOUSE	TREATMENT	LUNG	PLASMA EXPOSURE
	+ 4 h	Virus RNA copy n/ng tot RNA	Concentration (μM)
1	-	ND	-
2	-	ND	-
3	Vehicle	4.35 x 10 ³	-
4	Vehicle	ND	-
5	Vehicle	ND	-
6	Vehicle	2.79 x 10 ³	-
7	12 150 mpk po	ND	3
8	12 150 mpk po	ND	4.11
9	12 150 mpk po	ND	3.47
10	12 150 mpk po	ND	3.54

^aPlasmatic concentration of **12** in plasma samples derived from 2 different *in vivo* experiments, compared to viral RNA titers. ND = not detected.

Results from exposure measurement highlighted high concentrations of **12** after 4 h pi in the samples derived from experiment with the compound dosed at 300 mpk po twice a day (Table 5.1). No viral RNA was detected in two mice of the treated group, showing a potent antiviral activity of compound **12**. Therefore, two treated mice showed a detectable quantity of viral RNA, also compared to the vehicle. On the other hand, all four K18hACE2 mice treated with 150 mpk of **12** showed a similar plasmatic concentration, which appeared to significantly inhibit the viral replication in the lung (Table 5.1). In this experiment, the *in vivo* antiviral effect of the inhibitor was very strong, although the administrated dose of **12** was lower than previous experiment at 300 mpk.

5.3.3. Impact of mice treatment with Nirmatrelvir **12** on the development of antiviral adaptive immunity

With this established system, we next set out to study the consequences of Nirmatrelvir treatment on antiviral immune responses. Indeed, Nirmatrelvir showed to reduce viral titers and the risk of progressing to severe COVID-19,^{8,9} but there are no studies about the impact of Nirmatrelvir treatment on the development of adaptive immunity to SARS-CoV-2. Since Paxlovid was authorized for emergency use from FDA and EMA, it is not clear why some patients showed a virologic rebound and an associated release of COVID-19 symptoms following an early and effective Paxlovid treatment for COVID-19.^{10,11} Therefore, analysis of viral sequencing revealed that the release is not correlated to treatment-resistant mutations, or to an infection with different SARS-CoV-2 variants.^{10,12} In this context, the study performed at VSSR aimed to determine if treatment rebound is due to an impairment in the development of adaptive immunity necessary to complete SARS-CoV-2 clearance.

A group of K18-hACE2 transgenic mice were infected and treated po with **12** at 150 mpk, exactly as before described.^{1,2} In this experiment, the mice were monitored until 24 days pi, in order to measure the SARS-CoV-2 specific antibody response in the sera. Therefore, a homologous re-challenge with a higher dose (1×10^6 TCID₅₀) of aerosolized SARS-CoV-2 B.1.1.529 was administrated to the mice in order to monitor the eventual recruitment of memory T (and B) cells to the infected lung and lung-draining mediastinal lymph nodes, since SARS-CoV-2 T cells are not readily detectable in the blood of K18-hACE2 transgenic.

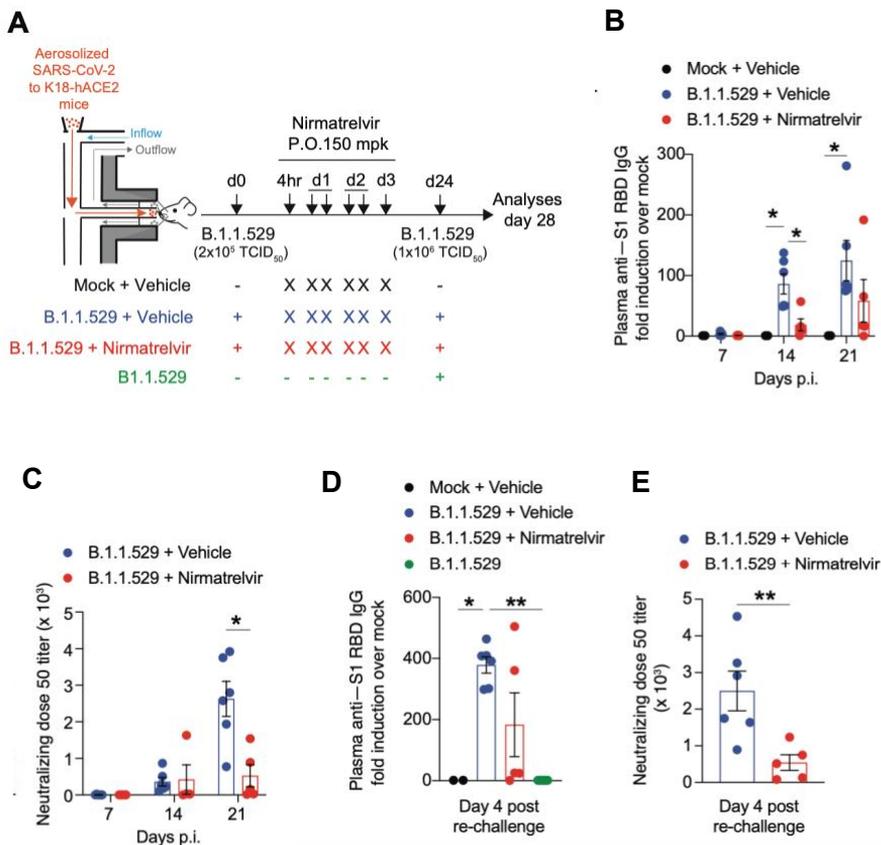


Figure 5.4. (A) Schematic representation of the experimental set up. Non-anesthetized K18-hACE2 mice were exposed to a target dose of 2×10^5 TCID₅₀ of aerosolized SARS-CoV-2

B.1.1.529. Twenty-four days after infection, mice were re-challenged with a target dose of 1×10^6 TCID₅₀ of SARS-CoV-2 B.1.1.529 through aerosol exposure. A group of naïve mice challenged with 1×10^6 TCID₅₀ of SARS-CoV-2 B.1.1.529 served as additional controls (green symbols, n = 5). (B, D) Quantification of anti-S1 RBD IgG levels by ELISA in the plasma of the indicated mice (B) 7, 14 and 21 days pi or (D) 4 days post re-challenge. (C, E) Neutralization dose 50 (ND₅₀) against SARS-CoV-2 B.1.1.529 pseudovirus in the plasma of the indicated mice (C) 7, 14 and 21 days pi or (E) 4 days post re-challenge.^{242*} *p-value* < 0.05, ** *p-value* < 0.01.

No significant weight loss was observed for whole duration of the experiment. Quantification of anti-S1 RBD IgG levels by ELISA in the plasma of the indicated mice (Fig. 5.4B) 7, 14 and 21 days pi or (Fig. 5.4D) 4 days post re-challenge were measured. Notably, the levels of total IgG specific for the spike S1 subunit and the levels of anti-B.1.1.529 neutralizing antibodies were significantly reduced by **12** treatment 14 and 21 days pi respectively, and 4 days after re-challenge (Fig. 5.4D,E). Indeed, B cells recovered from the mediastinal lymph-nodes of **12**-treated mice 4 days after re-challenge exhibited a lower expression of the activation marker CD95, while in one vehicle- treated mouse RBD-specific B cells were detected. Intracellular IFN- γ and TNF- α expression upon was stimulated *in vitro* to evaluate SARS- CoV-2-specific CD8⁺ and CD4⁺ T cells recovered from lung homogenates, using a pool of SARS-CoV-2 peptides covering the complete nucleocapsid, matrix, and spike proteins.¹³ In line with the previous results, the frequency and absolute number of IFN- γ ⁺ and IFN- γ ⁺, TNF- α ⁺, SARS-CoV-2-specific CD8⁺ and CD4⁺ T cells were found significantly lower in the lungs of Nirmatrelvir-treated mice compared to vehicle-treated mice, 4 days after homologous re-challenge (Fig. 5.4E)

In summary, the data obtained indicated that treatment with **12** at early stage of infection has a negative impact on the development of adaptive immune response to SARS-CoV-2 K18-hACE2 transgenic mice. This study presents an important implication on the use of Nirmatrelvir in therapy. Indeed, while the clinical data continue to support Nirmatrelvir treatment for the prevention of severe COVID-19 in high-risk individuals,²³⁸ the data exposed here draw attention to a potential

negative impact of this therapy and might explain the virological and/or symptomatic relapse after treatment completion reported in some patients.

Although these results are obtained in a mouse model that resemble the physiopathology in human this observation needs to be evaluated in patients taking Nirmatrelvir in the early phase of the infection.

5.4. Stability tests of Nirmatrelvir **12** in different formulations

As reported in literature, for oral administration the inhibitor is dosed 300 mpk in 0.5% methylcellulose in water, containing 2% Tween80, and administered twice daily.¹ This vehicle is not optimal to dissolve lipophilic molecules such as Nirmatrelvir **12**, but offers a clear advantage when the compound has to proceed in the development phase. As expected, the compound showed poor solubility in this vehicle at the tested concentration in experiments previously described (5.3.2). In order to identify a suitable formulation for *in vivo* administration of **12** both intraperitoneal (ip) and po, we carried out stability test of the compound in four different formulations for oral/IP administration by analyzing samples at different concentrations and different time points (Fig. 5.5). Below, the composition of the formulations is described:

- formulation 1: NMP 6% v/v, PEG400 64% v/v, normal saline 30% v/v;
- formulation 2: DMSO 10% v/v, NMP 10% v/v, Solutol 10% w/v, PEG400 40% v/v, normal saline 30% v/v;
- formulation 3: MCT oil 40% v/v, Solutol 20% w/v, PEG400 40% v/v;
- formulation 4: MCT oil 40% v/v, Solutol 20% w/v, PEG400 40% v/v (different from Formulation 3 for the preparation method).

The tested compound was formulated in the four vehicles at different concentrations, corresponding to different administration doses for the *in vivo*

assays, *i.e.* 15 mg/ml (60 mpk), 37 mg/ml (150 mpk), and 75 mg/ml (300 mpk), and analyzed by HPLC analysis immediately after sample preparation (t_0) and thereafter twice a day (every 12 hours) for 7 days. The time schedule was set up to cover a hypothetical 7 days twice a day administration.

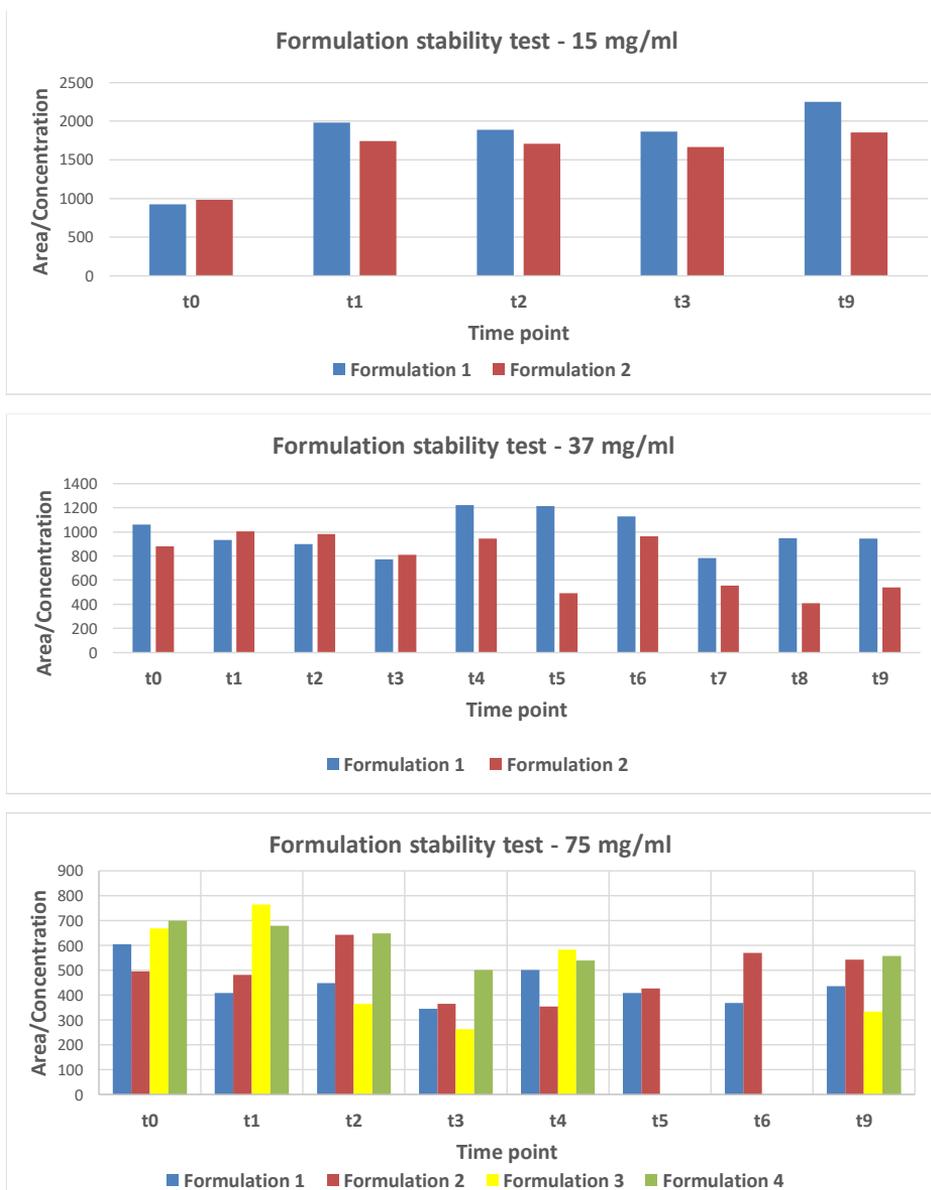


Figure 5.5. Graphics derived from HPLC analysis for formulation stability tests of **12**; where t_0 : 0h; t_1 : 12 h; t_2 : 24 h; t_3 : 36 h; t_4 : 48 h; t_5 : 60 h; t_6 : 72h; t_7 : 84 h; t_8 : 96 h; t_9 : 108 h.

Generally, the compound appeared soluble in all the tested formulations without precipitation by visual inspection. Below the description of the stability and solubility profiles by HPLC is reported:

- formulation 1 was analyzed using three concentrations (15 mg/mL, 37 mg/mL, and 75 mg/mL). It was stable during all the experiment. It is worth noting that formulation 1 was prepared without DMSO.
- formulation 2 was analyzed using three concentrations (15 mg/mL, 37 mg/mL, and 75 mg/mL). It was not stable during the experiment at the concentration of 37 mg/mL beyond the t_4 . However, it appeared stable at the concentrations of 15 mg/mL and 75 mg/mL.
- formulation 3 was only analyzed at the concentration of 75 mg/mL. It was not stable during the experiment beyond the t_1 .
- formulation 4 was only analyzed at the concentration of 75 mg/mL. It was stable during all the experiment.

In summary, formulation 1 and 4 resulted the best vehicles for **12**, which was stable in each analyzed concentrations at all the time points. It is worth nothing that these formulations do not contain DMSO.

The stability tests of **12** allowed us to identify a suitable vehicle to be used for the formulation of our future compounds subjected to *in vivo* study at VSSR University and avoid solubility issues, critical for the drug administration and drug absorption processes.

5.5. Summary and conclusions

Over two years, the development of vaccines and antiviral drugs allowed to provide important tools for our response to the COVID-19 pandemic. In

particular, efforts to identify effective SARS-CoV-2 DAAs have resulted in the discovery of Nirmatrelvir (PF-07321332), a reversible covalent peptidomimetic inhibitor developed by Pfizer and disclosed in April 2021. Nirmatrelvir was the first-in-class 3CL^{pro} inhibitor to receive approval by FDA in combination with the metabolic booster Ritonavir (Paxlovid[®]). Notably, the announcement from Pfizer about the discovery of Nirmaltrelvir was due to the extraordinary efforts of 200 chemist working on the project.

In this context, the experimental work performed in collaboration with INGM and VSSR aimed to use Nirmatrelvir as benchmarker for *in vitro/in vivo* assays, which were developed for an accurate screening of new compounds against SARS-CoV-2.

Before the disclosure of Nirmatrelvir synthesis by Pfizer, a synthetic route was planned in-house and carried out to afford the compound **12** in gram scale.

The compound **12** was used for *in vitro* antiviral screenings, including biochemical assays against SARS-CoV-2 and MERS 3CL^{pro}s, and phenotypic assays on different cell lines. In particular, the compound **12** was used as benchmark to validate a novel *in vitro* infectivity assay, developed at INMG of Milan in order to perform a screening campaign of newly designed 3CL^{pro} inhibitors. The evaluation of Nirmatrelvir antiviral activity against three different SARS-CoV-2 variants, using HEK-293T-hACE2 cells, allowed to define the final protocol of a phenotypic assay based on a more physiologically relevant cellular system.

In vivo efficacy and PK studies were performed to validate an innovative and tractable mouse model of COVID-19, developed at VSSR to provide a tractable animal model to mechanistically dissect virological, immunological, and pathogenetic aspects of the infection with SARS-CoV-2. Moreover, tests at

different dosage of Nirmatrelvir were carried out in the same animal model, allowing to identify a lower dosage of drug treatment respect to the literature data, which also solves the solubility problem of the high dosage formulation of **12**. At this regard, stability tests of **12** in different po formulations allowed to identify a more suitable vehicle to dissolve the drug, defining the best conditions to prepare the formulations of our clinical candidates for *in vivo* studies.

On basis of set experimental conditions, a long-term study on the same animal model was carried out with the aim to investigate the impact of Nirmatrelvir treatment on adaptative immunity, which was still completely unknown. Surprisingly, the negative impact emerged from obtained data could have important implications for clinical management and give an explanation to the virological and/or symptomatic rebound after treatment completion reported in some patients.

5.6. Experimental section

5.6.1. Synthesis and characterization of Nirmatrelvir **12**

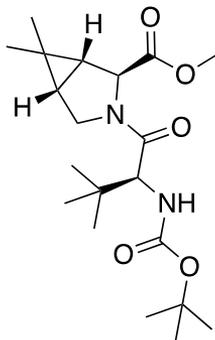
General methods. Reagents and solvents were purchased from commercial sources and used without further purification. Reactions were carried out at room temperature (RT), unless otherwise specified. Moisture-sensitive reactions were performed under a positive pressure of dry nitrogen in oven-dried glassware. Analytical thin-layer chromatography (TLC) on silica gel 60 F254 plates (250 μm thickness) was performed to monitor the reaction progress, using UV and KMnO_4 as revelation method. Analytical HPLC was performed to monitor the reaction progress and the purity of target compound. Flash chromatography on

silica gel (70—230 mesh) and preparative HPLC were performed for purification. All products were characterized by their NMR and MS spectra. (ESI)-MS spectra were performed on a LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) by infusion into the ESI source using MeOH as solvent. ¹H NMR spectra were recorded in deuterated DMSO at 25°C on Bruker Avance NEO 400 MHz and 700 MHz instruments. equipped with a RT-DR-BF/1H-5mm-OZ SmartProbe. Chemical shifts (δ) are reported in part per million (ppm) downfield from tetramethylsilane, using residual solvent signal as the internal reference.

The final compound was characterized by HPLC-MS/MS, using a Dionex ULTIMATE 3000 (Thermo Fisher Scientific) HPLC module and a LTQ XL mass spectrometer with electrospray ionization in positive mode and an Ion-Trap detector. Separation was performed with a Kinetex column C18 Polar column (250 mm \times 4.6 mm; particle size 5 μ m, Phenomenex, Torrance, CA, USA) at 30°C, using a 17 min gradient, 5% [0.1% TFA/CH₃CN]/95% [0.1% TFA/H₂O] to 95% [0.1% TFA/CH₃CN]. Analytical HPLC was performed on Shimatzu-1100 HPLC using a Kinetex C18 column (4.6 mm x 150 mm, 5 μ m, 100 Å) with an acetonitrile (0.1% HCOOH) – water (0.1% HCOOH) custom gradient. The purity of the final compound was >95%, as determined by HPLC (UV λ = 200 nm). Preparative HPLC was performed on Shimatzu LC-20AP using a Sunfire C18 column (19 mm x 100 mm, 5 μ m, 100 Å) with an acetonitrile (0.1% HCOOH) – water (0.1% HCOOH) custom gradient.

Synthesis of Nirmatrelvir 12

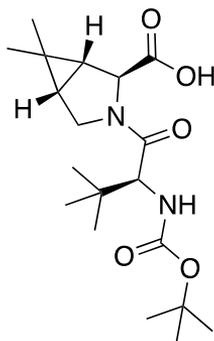
Methyl (1*R*,2*S*,5*S*)-3-((*S*)-2-((*tert*-butoxycarbonyl)amino)-3,3-dimethylbutanoyl)-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carboxylate (**86**).



Chemical Formula: C₂₀H₃₄N₂O₅
Molecular Weight: 382.50

tert-Butylglycine-OH **75a** (1 g, 4.32 mmol) was dissolved in dry CH₂Cl₂ (5 mL) and amine hydrochloride **85** (1.15 g, 5.62 mmol), HBTU (1.8 g, 4.75 mmol) and DIPEA (1.5 mL, 8.64 mmol) were added at 0 ° C under a nitrogen atmosphere. The resulting solution was kept under magnetic stirring at RT for 16 h. Then, the reaction mixture was washed with saturated aq. NaHCO₃ (x1), 1N HCl (x1), brine (x1), dried over anhydrous Na₂SO₄, filtered and concentrated under vacuum. The reaction crude was purified by flash column chromatography (Hexane / EtOAc 8: 2) to obtain **86** as a colorless oil (reaction time: 16 h; yield: 1.38 g, 78%). ¹H NMR (400 MHz, DMSO-d₆): δ 6.73 (d, *J* = 9.3 Hz, 1H), 4.21 (s, 1H), 4.05 (d, *J* = 9.4 Hz, 1H), 3.93 (d, *J* = 10.4 Hz, 1H), 3.79 (dd, *J* = 10.3, 5.3 Hz, 1H), 3.65 (s, 3H), 1.55 - 1.49 (m, 1H), 1.41 (d, *J* = 7.5 Hz, 1H), 1.35 (s, 9H), 1.01 (s, 3H), 0.93 (s, 9H), 0.85 (s, 3H). MS (ESI) *m/z* calcd: [M + H]⁺ for C₂₀H₃₅N₂O₅⁺ 383.51, found [M + H]⁺ 383.50.

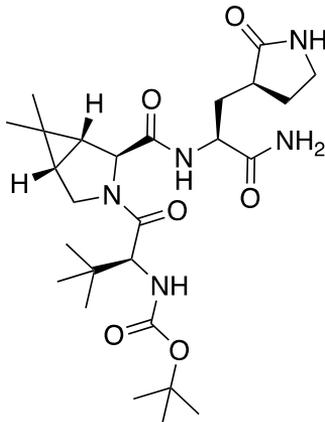
(1*R*,2*S*,5*S*)-3-((*S*)-2-((*tert*-Butoxycarbonyl)amino)-3,3-dimethylbutanoyl)-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carboxylic acid (**87**).



Chemical Formula: C₁₉H₃₂N₂O₅
Molecular Weight: 368.47

The methyl ester intermediate **86** (1.35 g 3.6 mmol) was dissolved in THF (18 mL), then 1N aq. LiOH was added (18 mmol, 18 mL), and the reaction mixture was kept under stirring at RT for 3h. The reaction mixture was cooled to 0 °C, placed in water / ice, acidified with 1 N HCl to pH = 4, then extracted with EtOAc (x3). Then, the collected organic layers were washed with brine (x1), dried over Na₂SO₄, filtered and concentrated under vacuum to afford dipeptide acid **87** as white solid, which was used in the following step without further purification (reaction time: 2 h, yield: 1.30 mg, 100%). ¹H NMR (400 MHz, DMSO-d₆): δ 12.54 (s, 1H), 6.67 (d, *J* = 9.4 Hz, 1H), 4.13 (s, 1H), 4.04 (s, 1H), 3.91 (d, *J* = 10.4 Hz, 1H), 3.77 (dd, *J* = 10.2, 5.3 Hz, 1H), 1.54 - 1.46 (m, 1H), 1.40 (s, 1H), 1.35 (s, 9H), 1.01 (s, 3H), 0.93 (s, 9H), 0.84 (s, 3H). MS (ESI) *m/z* calcd: [M + H]⁺ for C₁₉H₃₃N₂O₅⁺ 369.48, found [M + H]⁺ 369.50.

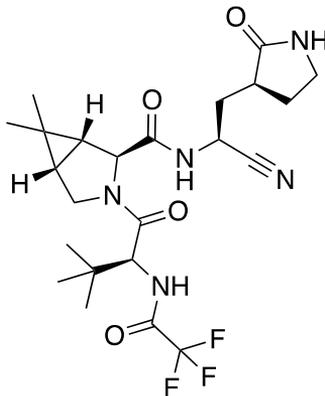
tert-Butyl ((*S*)-1-((1*R*,2*S*,5*S*)-2-(((*S*)-1-amino-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)carbamoyl)-6,6-dimethyl-3-azabicyclo[3.1.0]hexan-3-yl)-3,3-dimethyl-1-oxobutan-2-yl)carbamate (**88**).



Chemical Formula: C₂₆H₄₃N₅O₆
Molecular Weight: 521.66

Dipeptide acid **87** (1.22 g, 3.3 mmol) was dissolved in dry CH₂Cl₂ (6 mL) and the amine hydrochloride **74** (898 mg, 4.3 mmol), HBTU (1.25 g, 3.6 mmol), and DIPEA (1.4 ml, 8.25 mmol) were added at 0 °C, then DMF (3 mL) was added, and the reaction mixture was kept under stirring at RT for 3 h. The reaction mixture was washed with 1N HCl (x3), saturated aq. NaHCO₃ (x3), brine (x3), dried over Na₂SO₄, filtered, and concentrated under vacuum. The crude was purified by flash column chromatography (CHCl₃/MeOH 5 to 10%) to obtain the tripeptide **88** as a white solid (reaction time: 3 h, yield: 1.3 g, 75%). ¹H NMR (400 MHz, DMSO-d₆): δ 9.41 (d, *J* = 8.6 Hz, 1H), 8.29 (d, *J* = 8.9 Hz, 1H), 7.55 (s, 1H), 7.31 (s, 1H), 7.03 (s, 1H), 4.43 (d, *J* = 8.6 Hz, 1H), 4.35 – 4.25 (m, 2H), 3.91 – 3.84 (m, 1H), 3.67 (d, *J* = 10.4 Hz, 1H), 3.13 (t, *J* = 9.0 Hz, 1H), 3.06 – 2.97 (m, 1H), 2.45 – 2.34 (m, 1H), 2.14 (dt, *J* = 10.5, 7.4 Hz, 1H), 1.97 – 1.86 (m, 1H), 1.70 – 1.57 (m, 1H), 1.54 – 1.45 (m, 2H), 1.38 (d, *J* = 7.7 Hz, 1H), 1.10 (s, 3H), 1.05 (s, 9H), 0.98 (s, 9H), 0.84 (s, 3H). MS (ESI) *m/z* calcd: [M + H]⁺ for C₂₆H₄₄N₅O₆⁺ 522.67, found [M + H]⁺ 522.70.

(1*R*,2*S*,5*S*)-*N*-((*S*)-1-cyano-2-((*S*)-2-oxopyrrolidin-3-yl)ethyl)-3-((*S*)-3,3-dimethyl-2-(2,2,2-trifluoroacetamido)butanoyl)-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carboxamide (**12**).



Chemical Formula: C₂₃H₃₂F₃N₅O₄
Molecular Weight: 499.54

Tripeptide intermediate **88** (300 mg, 0.57 mmol) was dissolved in CH₂Cl₂ (3 mL), the solution was cooled to 0 °C, 4N HCl d in 1,4-dioxane (1.5 mL, 5.7 mmol) was added and the reaction mixture was stirred at RT for 2 h. Then, the solvent mixture was evaporated in vacuo and the crude was treated with hexane to obtain the desired compound **89** as HCl salt white solid, which was used without further purification in the following step (reaction time: 2 h, yield: 261 mg, 98%). Intermediate **89** (230 mg, 0.5 mmol) was suspended in dry CH₂Cl₂ (2 mL), under a nitrogen atmosphere, and dry pyridine (0.10 mL, 1.43 mmol) was added. After 30 min, the resulting mixture was cooled to 0 °C, TFFA (0.08 mL, 0.57 mmol) was added, and the reaction mixture was kept under stirring at RT for 2 h. Once observed, by TLC monitoring, the complete conversion of the intermediate **89**, anhydrous pyridine (0.18 mL, 2.28 mmol) was added, and the mixture was cooled at -5°C. After 5 min, TFFA (0.16 mL, 1.14 mmol) was added and the reaction was kept under stirring at RT for 15 h. The solvent was removed under vacuum, the resulting crude was diluted with EtOAc, and the organic phase was washed with 0.5 N HCl (x3), saturated aq. NaHCO₃ (x1), dried over anhydrous Na₂SO₄,

filtered, and concentrated under vacuum. The crude was purified by preparative HPLC (Shimadzu LC-20AP; column: Kinetex, 5 μ m, C18, 100 Å, 150 x 4,6 mm, C18 with TMS endcapping; mobile phase gradient: 10-90_15 min [H₂O 0,1% HCOOH, MeCN 0,1% HCOOH]; time course: 30 min; t_R = 16.9 min) to afford the target **12** (yield: 115 mg, 40%) as a white solid. ¹H NMR (700 MHz, DMSO-d₆): δ 9.43 (d, J = 8.4 Hz, 1H), 9.03 (d, J = 8.6 Hz, 1H), 7.68 (s, 1H), 4.97 (ddd, J = 10.9, 8.6, 5.1 Hz, 1H), 4.41 (d, J = 8.4 Hz, 1H), 4.15 (s, 1H), 3.91 (dd, J = 10.4, 5.5 Hz, 1H), 3.69 (d, J = 10.4 Hz, 1H), 3.17 - 3.11 (m, 1H), 3.04 (td, J = 9.4, 7.1 Hz, 1H), 2.40 (tdd, J = 10.4, 8.4, 4.4 Hz, 1H), 2.14 (ddd, J = 13.4, 10.9, 4.4 Hz, 1H), 2.11 - 2.03 (m, 1H), 1.76 - 1.65 (m, 2H), 1.57 (dd, J = 7.6, 5.5 Hz, 1H), 1.32 (d, J = 7.6 Hz, 1H), 1.03 (s, 3H), 0.98 (s, 9H), 0.85 (s, 3H). MS (ESI) *m/z* calcd: [M + H]⁺ for C₂₃H₃₃F₃N₅O₄⁺ 500.53, found [M+H]⁺ 500.40.

5.6.2. Formulation study of Nirmatrelvir **12**

LC-MS/MS Analysis. The Nirmatrelvir **12** stock solutions were prepared in DMSO at 1 mg/mL and further diluted to obtain a working solution (WS) at 20 μ g/mL. The drug JWH250 was used as internal standard. The internal standard working solutions (IS-WS) was prepared at 20 ng/mL in methanol:acetonitrile (50:50, v/v) acidified with 0.1 % formic acid. Plasma of mice were collected, centrifuged at 10,000 rpm for 10 minutes and incubated 10 at 60°C for 30 minutes. The mixture of 30 μ L of plasma, 105 μ L of IS-WS and 15 μ L of WS was vortex for 1 minute and centrifuged at 12,500 g for 10 minutes at 4 °C. The supernatant was collected and 100 μ L were injected into the liquid chromatography tandem mass spectrometry (LC–MS/MS) system. The HPLC equipment consisted of an LC AC System from AB Sciex (Toronto, ON, Canada). A Triple Quadrupole Mass Spectrometer (API 2000) from AB-Sciex

(Toronto, ON, Canada) was used for detection. The analytes were separated using an Acquity UPLC BEH C18 1.7 μm Column (2.1 x 50 mm ID) from Waters. The mobile phases were (B) MeOH containing 0.2% formic acid and (A) water containing 0.1 % formic acid, at a flow rate of 0.4 mL/min and were entirely transferred into the mass spectrometer source. The gradient elution was as follows: increase of the organic phase from 10 to 100% in 2 minutes and after 1.5 minute of 100 % B the column was led to the original conditions; equilibration of the column was achieved in 2 minutes. Both analytes were detected in positive ionization with a capillary voltage of 4500 V, nebulizer gas (air) at 45 psi, turbo gas (nitrogen) at 70 psi and 450 °C. The other ion source parameters were set as follows: curtain gas (CUR) 25 psi; collision gas (CAD) 6 psi; declustering potential 80 V; entrance potential 8 V. Instrument conditions optimization was performed by direct infusion and manual tuning. Data collection and elaboration were performed by means of Analyst 1.4 software (AB-Sciex). The quantitative data were acquired using Multi Reaction Monitoring (MRM) mode. Two MRM transitions (precursor ion > fragment ion) were selected for the analytes. The parameters used for each analyte are listed in the following table.

	Q1	Q3	CE (V)	CXP (V)
Nirmatrelvir	500.2	110.0	40	16
		68.0	80	8
JWH-250 (IS)	336.1	121.0	31	18

The analytical method was validated according to FDA guidelines for bioanalytical method validation. Linearity, precision, accuracy, limits of detection (LODs) and limits of quantification (LOQ) were evaluated. Calibration standard solutions were prepared in blank plasma by spiking 15 μL of a standard

mixture at appropriate concentration to 30 μ L of plasma and by adding 105 μ L of methanol:acetonitrile (50:50, v/v). Calibrators were then treated similarly to the animal samples. The calibration range was 2 to 750 ng/mL and the calibrators were prepared at nine level of concentration. Precision, recovery and accuracy were evaluated at three level of concentrations (25, 100, 750 ng/ml) and resulted within the acceptable limits. LOD was defined as the lowest concentration with a signal-to-noise (S/N) ratio greater than 3. LOQ was defined as the concentration at which both precision (RSD %) and accuracy were less than 20 %. LOQ resulted to be 2 ng/mL while LOD was 1 ng/mL for both analytes.

5.6.3. *In vitro* enzymatic and antiviral cell-based assays

Biochemical assay. M^{pro} SARS-CoV-2 was expressed in E. coli cells BL21 (DE3) and purified as described.^{14,15} Briefly, the protein was purified in two steps using a Ni-Sepharose column and by HiTrap Q HP column and the fractions containing the M^{pro} SARS-2 were pooled and concentrated using Amicon Ultra 15 centrifugal filters, at 4000 x g, at 4 °C, in a buffer exchange (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.8). Protein purity was verified by SDS-PAGE analysis and the proteins were stored at -80°C.

The M^{pro} SARS-CoV-2 biochemical assays was performed in 386 wells plate in 20 μ L of assay buffer containing diluted protein, 20 mM Tris (pH 7.3), 100 mM NaCl and 1mM EDTA, with the addition of 5 mM TCEP, 0.1 % BSA. In Table 5.1, the protein was preincubated for 30 minutes at 37 °C with different concentrations of Nirmatrelvir or GC376, as positive control of inhibition.^{16,17} The substrate DABCYL-KTSAVLQ↓SGFRKM-EDANS (Bachem) was added, and the generation of the fluorescent product was monitored after 15 minutes of incubation (Ex 340 nm, Em 490nm). The protein and the compound were not

preincubated and the enzymatic reaction was immediately initiated with the addition of the substrate in the assay buffer, as described.¹ The reaction was allowed to progress for 60 minutes at 23° C and then monitored (Ex 340 nm, Em 490nm). Dose response curve were generated by nonlinear regression curve fitting with GraphPad Prism to calculate IC₅₀.

***In vitro* antiviral assays.** For *in vitro* antiviral assay HEK293T-hACE were plated in 96 well plates at 5,000 cells/well in complete DMEM plus 2% FBS. After 24 h, cells were treated with 7 concentrations of 5-fold serially diluted Nirmaltrevir and infected at 0.1 MOI of SARS-CoV-2 virus. DMSO was used as vehicle for compound serial dilution and not treated control (final concentration 0.25%). Not infected condition was inserted as negative control of infection. Each condition was assayed in three replicates. Antiviral activity was evaluated by qPCR quantification of secreted SARS-CoV-2 RNA and/or by cytopathic effect protection assay (CPE) after 72 h of incubation at 37°C under 5% CO₂. For the quantification of SARS-CoV-2 RNA by qPCR, 10 µL of cell supernatants were subjected to direct lysis with the addition of 10 µL ViRNAex solution (Cabru) and heated at 70°C for 15 min. After addition of distilled water (1:1), samples were used as template for PCR amplification using TaqPath™ 1-Step RT-qPCR Master Mix (Thermofisher Scientific) and specific SARS-CoV-2 primers/probe (2019-nCoV RUO Integrated DNA Technologies). Obtained Ct were normalized to untreated infected wells and a dose response curve were generated by nonlinear regression curve fitting with GraphPad Prism to calculate the concentration that inhibits 50% of viral replication (EC₅₀).

For CPE assays, CellTiter-Glo® Luminescent Cell Viability Assay (promega), was used. Relative luciferase units (RLUs) were normalized to infected or uninfected controls in order to obtain the percentage of inhibition of cytopathic effect using the following formula: % CPE inhibition = 100*(Test Cmpd - Avg.

$Virus)/(Avg. Cells - Avg. Virus)$, where *Avg. virus* is the RLU average obtained from infected and not treated wells, while *Avg. Cells* is the RLU average obtained from not infected and not treated wells. Dose response curve were generated by nonlinear regression curve fitting with GraphPad Prism to calculate EC₅₀.

5.6.4. *In vivo* efficacy studies of Nirmaltrelvir 12

Mice. B6 Cg-Tg(K18-ACE2)^{2Pr1mn}/J mice (referred to in the text as K18-hACE2) were purchased from The Jackson Laboratory. Mice were housed under specific pathogen-free conditions and heterozygous mice were used at 8-10 weeks of age. All experimental animal procedures were approved by the Institutional Animal Committee of the San Raffaele Scientific Institute and all infectious work was performed in designed BSL-3 workspaces.

Mouse infection through aerosol exposure. Infection of K18-hACE2 transgenic mice with aerosolized SARS-CoV-2 was performed as described.² Briefly, non-anesthetized K18-hACE2 mice were placed in a nose-only Allay restrainer on the inhalation chamber (DSI Buxco respiratory solutions, DSI). To reach a target accumulated inhaled aerosol (also known as delivered dose) of 2×10^5 TCID₅₀ mice were exposed to aerosolized SARS-CoV-2 B.1.1.529 for 30-60 minutes (depending on the total volume of diluted virus and on the number of mice simultaneously exposed). In selected experiments, mice were exposed to a target accumulated inhaled aerosol of 1×10^6 TCID₅₀. Primary inflows and pressure were controlled and set to 0,5 L/minute/port and -0,5 cm H₂O, respectively. As control, K18-hACE2 mice received the same volume of aerosolized PBS (125 µL per mouse). Infected mice were monitored daily to record body weight, clinical and respiratory parameters.

***In vivo* treatment.** K18-hACE2 mice were treated by oral gavage with Nirmaltrevir at 150 mg/kg or vehicle [0,5% Methylcellulose (Methocel A4M, Sigma #94378), 2% Tween80 (Sigma #8170611000) in purified water] for six times starting 4 h post infection, and every 12 hours thereafter.

Tissue homogenate and viral titers. Tissues homogenates were prepared by homogenizing perfused lungs using gentleMACS Octo Dissociator (Miltenyi BioTec, #130-096-427) in M tubes (#130-093-335) containing 1 ml of DMEM 0% FBS. Samples were homogenized for three times with program m_Lung_01_02 (34 seconds, 164 rpm). The homogenates were centrifuged at 3'500 rpm for 5 minutes at 4°C. The supernatant was collected and stored at -80°C for viral isolation and viral load detection. Viral titer was calculated by 50% tissue culture infectious dose (TCID₅₀). Briefly, Vero E6 cells were seeded at a density of 1.5×10^4 cells per well in flat-bottom 96-well tissue culture plates. The following day, 10-fold dilutions of the homogenized tissue were applied to confluent cells and incubated 1 h at 37°C. Then, cells were washed with phosphate-buffered saline (PBS) and incubated for 72 h at 37°C in DMEM 2% FBS. Cells were fixed with 4% paraformaldehyde for 30 min and stained with 0.05% (wt/vol) crystal violet in 20% ethanol.

RNA extraction and qPCR. Tissues homogenates were prepared by homogenizing perfused lung and nasal turbinates using gentleMACS dissociator (Miltenyi BioTec, #130-096-427) with program RNA_02 in M tubes (#130-096-335) in 1 ml of Trizol (Invitrogen, #15596018). The homogenates were centrifuged at 2000 g for 1 min at 4°C and the supernatant was collected. RNA extraction was performed by combining phenol/guanidine-based lysis with silica membrane-based purification. Briefly, 100 µL of Chloroform was added to 500 µL of homogenized sample and total RNA was extracted using ReliaPrep™ RNA

Tissue Miniprep column (Promega, Cat #Z6111). Total RNA was isolated according to the manufacturer's instructions. qPCR was performed using TaqMan Fast virus 1 Step PCR Master Mix (Lifetechnologies #4444434), standard curve was drawn with 2019_nCoV_N Positive control (IDT#10006625), primer used are: 2019-nCoV_N1- Forward Primer (5'-GAC CCC AAA ATC AGC GAA AT-3'), 2019-nCoV_N1- Reverse Primer (5'-TCT GGT TAC TGC CAG TTG AAT CTG-3') 2019-nCoV_N1-Probe (5'-FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1-3') (Centers for Disease Control and Prevention (CDC) Atlanta, GA 30333). All experiments were performed in duplicate.

ELISA. Individual sera were titrated in parallel for the presence of SARS-CoV-2 S1 RBD-specific antibody by end-point ELISA. The ELISA plates were functionalized by coating with recombinant Sars-CoV-2 S1 subunit protein (RayBiotech, #230-30162) at a concentration of 2 g/mL and incubated overnight (O/N) at 4°C. Subsequently, the plates were blocked with 3% fat-free milk, 0,05% Tween20 in PBS for 1 h at RT. The sera were then added at a dilution of 1/20 (sera from day 7) or 1/500 (sera from day 14, 21, 28) and diluted 1:10 up to 1/1280 or 1/32000, respectively, in duplicate, and the plates were incubated for 2h at RT. After five washes with 0,05% Tween20 in PBS, the secondary anti-murine IgG conjugated with horseradish peroxidase (HRP, PerkinElmer, #NEF822001EA) (1:2000) was added and the plates were incubated for 1 h at RT. After washing, the binding of the secondary was detected by adding the substrate 3,3',5,5'-tetramethylbenzidine (TMB, BD Biosciences). The reaction was blocked with 0,5 M H₂SO₄ and the absorbance at 450 nm and reference 630 nm was measured.

SARS-CoV-2 pseudovirus neutralization assay. SARS-CoV-2 pseudovirus

neutralization assay was performed as previously described.¹⁸ Briefly, lentiviral vector containing luciferase reporter were pseudotyped with B.1.1.529 SARS-CoV-2 spike protein and used for neutralization assay. HEK293T-hACE2 receptor were plated in 96 well plates and transduced with 0.05 MOI of SARS-CoV-2 pseudovirus that were subjected to 1 h at 37°C of preincubation with 3-fold serially diluted mice plasma. After 24h of incubation, pseudoparticle cell transduction was measured by luciferase assay using Bright-Glo™ Luciferase System (Promega) and dose response curves were generated by nonlinear regression curve fitting to calculate Neutralization dose 50 (ND₅₀).

Cell Isolation and Flow Cytometry. Mice were euthanized by cervical dislocation. At the time of autopsy, mice were perfused through the right ventricle with PBS. Nasal turbinates were removed from the nose cavity. Lung tissue was digested in RPMI 1640 containing 3.2 mg/ml Collagenase IV (Sigma, #C5138) and 25 U/ml DNase I (Sigma, #D4263) for 30 minutes at 37°C. Homogenized lungs were passed through 70 µm nylon meshes to obtain a single cell suspension. Cells were resuspended in 36% percoll solution (Sigma #P4937) and centrifuged for 20 minutes at 2000 rpm (light acceleration and low brake). The remaining red blood cells were removed with ACK lysis. Peripheral blood was collected in PBS 0,5 mM EDTA and lysed two times with ACK.

For analysis of *ex-vivo* intracellular cytokine production, 1 mg/mL of brefeldin A (Sigma #B7651) was included in the digestion buffer. All flow cytometry stainings of surface-expressed and intracellular molecules were performed as described.¹⁹⁻²¹ Briefly, cells were stimulated for 4 h at 37°C in the presence of brefeldin A, monensin (life technologies, #00-4505-51) and a pool of overlapping peptides (1 µg/ml per peptide), including MHC class I- and MHC class II-restricted peptides (9-22 aminoacids) covering the S, S1, S+, M and N protein of SARS-CoV-2 (Milteny, #130-126-700; #130-127-041; #130-127-311; #130-

126-702, #130-126-698).¹³ Cell viability was assessed by staining with Viability™ 405/520 fixable dye (Miltenyi, Cat #130-109-814). Antibodies (Abs) are indicated in the table below. Flow cytometry analysis was performed on BD FACS Symphony A5 SORP and analyzed with FlowJo software (Treestar).

Histology and histochemistry. Mice were euthanized and perfused transcardially with PBS. One left lobe of the lung was fixed in zinc formalin and transferred into 70% ethanol 24h later. For hematoxylin and eosin (H&E) staining, tissues were stained as previously described.^{11,22,23} Bright-field images were acquired with an Aperio Scanscope System CS2 microscope and the ImageScope program (Leica Biosystem) following the manufacturer's instructions.

5.7. References Chapter 5

1. Owen, D. R. *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **374**, 1586–1593 (2021).
2. Fumagalli, V. *et al.* Administration of aerosolized SARS-CoV-2 to K18-hACE2 mice uncouples respiratory infection from fatal neuroinvasion. *Sci. Immunol.* **7**, (2022).
3. Jiang, R. Di *et al.* Pathogenesis of SARS-CoV-2 in Transgenic Mice Expressing Human Angiotensin-Converting Enzyme 2. *Cell* **182**, 50-58.e8 (2020).
4. Dong, W. *et al.* The K18-Human ACE2 Transgenic Mouse Model Recapitulates Non-severe and Severe COVID-19 in Response to an

- Infectious Dose of the SARS-CoV-2 Virus. *J. Virol.* **96**, (2022).
5. Delorey, T. M. *et al.* COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113 (2021).
 6. Karki, R. *et al.* Synergism of TNF- α and IFN- γ Triggers Inflammatory Cell Death, Tissue Damage, and Mortality in SARS-CoV-2 Infection and Cytokine Shock Syndromes. *Cell* **184**, 149-168.e17 (2021).
 7. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).
 8. Najjar-Debbiny, R. *et al.* Effectiveness of Paxlovid in Reducing Severe COVID-19 and Mortality in High Risk Patients. *Clin. Infect. Dis.* (2022) doi:10.1093/CID/CIAC443.
 9. Hammond, J. *et al.* Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N. Engl. J. Med.* **386**, 1397–1408 (2022).
 10. Gupta, K. *et al.* Rapid Relapse of Symptomatic SARS-CoV-2 Infection Following Early Suppression with Nirmatrelvir/Ritonavir. (2022) doi:10.21203/rs.3.rs-1588371/v1.
 11. Ranganath, N. *et al.* Rebound Phenomenon after Nirmatrelvir/Ritonavir Treatment of Coronavirus Disease-2019 in High-Risk Persons. *Clin. Infect. Dis.* (2022) doi:10.1093/CID/CIAC481.
 12. Boucau, J. *et al.* Characterization of virologic rebound following nirmatrelvir-ritonavir treatment for COVID-19. *Clin. Infect. Dis.* (2022)

doi:10.1093/CID/CIAC512.

13. Silva-Cayetano, A. *et al.* A booster dose enhances immunogenicity of the COVID-19 vaccine candidate ChAdOx1 nCoV-19 in aged mice. *Med* **2**, 243-262.e8 (2021).
14. Costanzi, E. *et al.* Structural and biochemical analysis of the dual inhibition of mg-132 against sars-cov-2 main protease (Mpro/3clpro) and human cathepsin-l. *Int. J. Mol. Sci.* **22**, 22 (2021).
15. Kuzikov, M. *et al.* Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen. *ACS Pharmacol. Transl. Sci.* **4**, 1096–1110 (2021).
16. 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.* **65**, 2716–2746 (2022).
17. Luan, X. dong *et al.* Structure basis for inhibition of SARS-CoV-2 by the feline drug GC376. *Acta Pharmacol. Sin.* 2022 1–3 (2022) doi:10.1038/s41401-022-00929-z.
18. Conforti, A. *et al.* COVID-eVax, an electroporated DNA vaccine candidate encoding the SARS-CoV-2 RBD, elicits protective responses in animal models. *Mol. Ther.* **29**, 1–16 (2021).
19. Soares, H., Lynn Baniecki Pzer Rhonda Cardin Pzer Heidi Leister-Tebbe Pzer Yuao Zhu Pzer Shunjie Guan Pzer Craig Hyde Pzer Wen He Pzer

Zhenyu Wang Pzer Li Hao Pzer Scott Perrin, M. B. & Bao Pzer Phylinda Chan Pzer Bharat Damle Pzer Sandeep Menon Pzer, W. Viral Load Rebound in Placebo and Nirmatrelvir-Ritonavir Treated COVID-19 Patients is not Associated with Recurrence of Severe Disease or Mutations. doi:10.21203/rs.3.rs-1720472/v2.

20. Bénéchet, A. P. *et al.* Dynamics and genomic landscape of CD8+ T cells undergoing hepatic priming. *Nat.* 2019 5747777 **574**, 200–205 (2019).
21. De Simone, G. *et al.* Identification of a Kupffer cell subset capable of reverting the T cell dysfunction induced by hepatocellular priming. *Immunity* **54**, 2089-2100.e8 (2021).
22. Halfmann, P. J. *et al.* SARS-CoV-2 Omicron virus causes attenuated disease in mice and hamsters. *Nature* **603**, 687–692 (2022).
23. Mercado, R. *et al.* Early programming of T cell populations responding to bacterial infection. *J. Immunol.* **165**, 6833–6839 (2000).

GENERAL CONCLUSIONS

RNA emerging viruses, such as ZIKV and SARS-CoV-2, are serious threats to public health and have become a global concern.

Despite the “Public Health Emergency” declared by WHO in 2016, scientific efforts have not resulted in approval of vaccines and therapeutics for ZIKV infection. In response to the pressing medical need of effective antiviral agents against ZIKV, my work was based on the discovery of new ZIKV DAAs by integrating rational design and parallelly artificial intelligence with chemical synthesis, as described in the chapter 3.

Since no NNIs ZIKV RdRp were known at the time of starting this project, rational design aimed to discover such compounds by applying scaffold hopping approach of NNIs of DENV RdRp, another member of the Flavivirus genus. In particular, compound **19** was selected as template, as it was the most potent DENV polymerase inhibitor reported and its binding site, known as allosteric N-pocket, was conserved in ZIKV homologue protein, according to the sequence alignment and comparison of the apo protein crystal structures. At this regard, compounds **20** and **21**, maintaining the interactions moieties of **19** for the binding to the allosteric pocket, and modifying the central aryl with a dihydroxy- or dimethoxypyrimidine, a privileged scaffold found in potent HCV RdRp inhibitors and in Raltegravir, the first-in-class approved HIV integrase inhibitor, were designed. Unfortunately, no significant antiviral activity was shown by **20** and **21** in cell-based assays, although they were not toxic for the cellular viability. However, a study recently published by Gharbhi Ayachi *et. al* revealed structural differences of N-pocket between ZIKV NS5 polymerase and DENV3 homologue enzyme, critical for the binding mode of our inhibitors. In particular, the introduction of a propargyl alcohol on the thiophene moiety is not tolerated into the allosteric site of ZIKV polymerase and may cause a significant drop of the potency of inhibitors. In view of this and the antiviral activity resulted in

phenotypic assay, compounds **20** and **21** are uninteresting for further biological investigations.

In parallel, a computer-aided drug discovery program carried out by Dompè Pharmaceuticals was set to identify potential poly-viral pharmacological virtual hits against ZIKV. Through the innovative platform of Exscalate, a large library of clinical candidates and approved drugs was screened against 13 binding sites identified from 5 ZIKV structural and not structural proteins. Raloxifene, a drug approved for the osteoporosis, was identified as the most promising virtual hit though a first phenotypic screening. A second library of compounds was generated around repurposed Raloxifene by exploiting a tangible chemical space and screened against ZIKV proteins. Among the polypharmacological virtual hits selected for the synthesis (**31-38**) and then evaluated in phenotypic assay, the most potent derivatives, **34** and **37**, showed a potent antiviral activity in cell-based plaque assay, with EC₅₀ values in nanomolar range, without significant cellular toxicity. Thus, the compounds **34** and **37** have been selected for the *in vivo* screening on a mouse model of ZIKV infection, in order to evaluate the antiviral efficacy and PK profiles. In addition, further *in vitro* assays are planned to experimentally investigate the ability of **34** and **37** to bind or inhibit their putative viral targets.

SARS-CoV-2 is the etiological agent of COVID-19 and its recent outbreak represent the worst pandemic over the last 100 years. As opposed to ZIKV, the heavy impact of COVID-19 pandemic rapidly boosted the research on this disease, leading to the approval of vaccines and DAAs within a short time frame, if compared to any other infections. However, only three DAAs are available to date, which show limitations in efficacy and/or PK properties. An elective target

to identify DAAs against SARS-CoV-2 is the 3CL^{pro}, and Nirmaltrevir **12** is the first-in-class approved DAA targeting 3CL^{pro} for COVID-19 treatment.

In the chapter 4, it is described the rational design, synthesis and *in vitro* antiviral activity of a novel series of proline-based tripeptides as reversible covalent inhibitors of SARS-CoV-2 3CL^{pro} (**39-48**), bearing an aldehyde warhead as cysteine trap and a cyclic analogue of glutamine in P1. In particular, by replacing the classic leucine in P2 with different functionalized proline residues, we investigated the impact on the activity of inducing a turn between the P2-P3 positions. Subsequently, another series of di- and tripeptides (**50-53**) were designed and synthesized by replacing the aldehyde warhead with a nitrile functional group, similarly to Nirmatrelvir, and inserting different residues in P2 and P3 by applying a “mix and match” approach of chemical moieties present in potent SARS-CoV-2 3CL^{pro} inhibitors reported in literature.

Most of our newly synthesized compounds showed excellent inhibition of the SARS-CoV-2 3CL^{pro} with some displaying IC₅₀s in the sub-nanomolar range, in addition to a potent antiviral activity in the low micromolar to nanomolar range, with **46** showing the most potent antiviral cell activity. Nonetheless, the inhibitory effect was influenced by the size, the nature and the chirality of the substituents at the P2 position, indicating that a 5-membered cyclic amino acid (i.e. proline) is tolerated in P2, and small substituents at position 4 of the proline ring are preferred in *S* configuration, while bigger moieties are better tolerated in *R* configuration. Moreover, the majority of the compounds also showed a high potency against the MERS 3CL^{pro} (IC₅₀ values ranged from mid- to sub-nM), highlighting a potential anti-CoV broad-spectrum activity. X-ray crystallographic studies on representative compound of the aldehyde series (**41**, **42**, **46**, **48**) allowed to corroborate the initial rational design, that was supported by *in silico*-aided studies. These results will prompt us to design new SARS-CoV-2 3CL^{pro} peptidomimetic inhibitors, in order to optimize physicochemical

properties of the most potent compound **46** and provide a promising preclinical candidate, suitable for *in vivo* efficacy study.

As described in the chapter 5, the in-house synthesis of Nirmatrelvir **12**, planned and carried out prior the publication of Owen *et al.* (Owen *et al.*, *Science* 10.1126/science.abl4784 (2021)), allowed to afford the target compound in multigram scale and to test it as benchmark for the *in vitro/in vivo* screening, developed by our collaborators. In this regard, the protocol of a phenotypic assay, based on a more physiologically relevant cellular system (HEK-293T-hACE2 cells), was set out by the evaluation of antiviral activity of **12** against three different SARS-CoV-2 variants. In addition, an innovative and tractable K18-hACE2 transgenic mice of SARS-CoV-2 infection by aerosol exposure was validated by *in vivo* efficacy studies of **12** at different dosage. A lower dosage of **12** (150 mpk) respect to the literature data (300 mpk) showed antiviral efficacy in the same humanized animal model. On the basis of these experimental conditions, a long-term study on the same animal model was carried out, with the aim to investigate the impact of Nirmatrelvir treatment on adaptative immunity. The results indicated that treatment with Nirmatrelvir, early after infection, negatively impacts the development of adaptive immune response to SARS-CoV-2 in K18-hACE2 transgenic mice. Although the mechanistic bases behind this observation were not addressed in this study, it is conceivable that this is due to insufficient antigen exposure (quantity and/or duration) of naive B and T cells. While the clinical data continue to support Nirmatrelvir treatment for the prevention of severe COVID-19 in high-risk individuals, the data reported here draw attention to a potential negative impact of this therapy. Whether this effect is an exclusive feature of Nirmatrelvir or whether forthcoming antivirals acting on SARS-CoV-2 would have similar effects, should be addressed in future studies.

APPENDIX

A.1. External training

As part of my research activities forecast by the Ph.D. course in “Pharmaceutical Science” at the University of Naples Federico II, I spend six months secondment at the Institute of Cancer Research, Division of Cancer Therapeutics in London, 15 Cotswold Road Sutton SM2 5NG.

I carried out my secondment activities under the supervision of Prof. Ian Collins for the whole period of approximately six months, starting on 1st February and ending on 25th July, 2022. During my secondment, I have contributed to research in cancer medicinal chemistry, by designing and synthesizing small molecules against Ependymoma paediatric brain tumour, and chemical probes to identify protein target(s). The project was supported by Cancer Research UK - Children’s Brain Tumour Centre of Excellence, hosted at the University of Cambridge and the Institute of Cancer Research.

However, I had full access to all the required lab equipment (synthetic and analytic) and infrastructure to perform my research, besides the Institute library and online databases necessary to carry out the intended projects. Importantly, I benefited from collaborations and interactions with scientists from the Institute of Cancer Research and other institutions to maximize the successful outcome of my PhD research activities.

As the project is restricted confidential, I cannot report the research activity and results obtained during my period at the Institute of Cancer Research in London. Therefore, I have attached below a letter from the novel head of the project, prof. Gary Newton, in support of my research activity.



13th October 2022

Dear University of Naples Federico II,

I am writing to confirm that Irina Stefanelli completed 6 months research in our laboratories at the Institute of Cancer Research in Sutton, as part of the expected activities of the PhD course "Pharmaceutical Science" that Irina has undertaken at your institution.

During this time, she contributed to an ongoing research programme, which I am responsible for, that is part of the Children's Brain Tumour Centre of Excellence. Her work involved the design and synthesis of molecules targeting rare children's brain tumours. The molecules that she made were tested in variety of assays to assess their metabolic stability and activity against tumour cell lines. This led to several compounds with improved properties. This work was part of a multi-disciplinary collaboration and Irina gave regular updates on her work at team and project meetings. She had full access to modern synthetic and analytical equipment to perform her role. We would like to thank her for the valuable contribution she has made to our research programme.

Please do not hesitate to get in touch if further information is required.

Yours Sincerely,

A handwritten signature in black ink, appearing to read 'Gary Newton', is written over a light blue horizontal line.

Dr Gary Newton
Team Leader Medicinal Chemistry 3
Centre for Cancer Drug Discovery, Division of Cancer Therapeutics
Institute of Cancer Research

The Institute of
Cancer Research

Dr Gary Newton, PhD
Team Leader Medicinal Chemistry 3

The Institute of Cancer Research
Centre for Cancer Drug Discovery
Division of Cancer Therapeutics
15 Cotswold Road
Sutton SM2 5NG
T +44 203 437 6404
E gary.newton@icr.ac.uk

Registered office
The Institute of Cancer Research:
Royal Cancer Hospital
123 Old Brompton Road
London SW7 3RP

A Charity. Not for Profit.
Company Limited by Guarantee.
Registered in England No. 534147.
VAT Registration No. 849 0581 02