



**UNIVERSITÀ DEGLI STUDI DI PAVIA**

**DOTTORATO IN SCIENZE CHIMICHE E FARMACEUTICHE E  
INNOVAZIONE INDUSTRIALE  
(XXXVI Ciclo)**

**Coordinatore: Chiar.mo Prof. Giorgio Colombo**

**Antiviral Compounds Against SARS-CoV-2 (ACASA)**

Tesi di Dottorato di  
**Marco Leusciatti**

AA 2022/2023

**Tutor**

Chiar.mo Prof. Paolo Quadrelli

**Co-tutor**

Chiar.ma Dott.ssa Giulia Morra

# Index

<b>INTRODUCTION</b> .....	<b>1</b>
BUT WHAT IS A VIRUS? .....	1
VIRUS ORIGIN AND EVOLUTION .....	3
RNA VIRUS MUTATION RATES ARE VERY HIGH .....	6
SARS-CoV-2 .....	9
SPIKE PROTEIN: S1 MUTATIONS .....	12
CONSERVED REGIONS: APPEALING TARGETS .....	15
THE RNA-DEPENDENT RNA POLYMERASE (RdRp) NSP12 .....	17
3C-LIKE OR MAIN PROTEASE (M <sup>PRO</sup> ) NSP5 .....	21
REFERENCES .....	24
<b>CHAPTER 1</b> .....	<b>27</b>
1.1 TARGETING SARS-CoV-2's RdRp (NSP12) .....	27
1.1.1 Nitrosocarbonyl .....	28
1.1.2 Nitrosocarbonyl Generation and Reactivity .....	28
1.1.3 Nitrosocarbonyl as Key Intermediates to Carbocyclic Nor-Nucleosides .....	30
1.2 RESULTS: COMPUTATIONAL ANALYSIS: BINDING PREDICTION .....	32
1.2.1 Methods .....	32
1.2.2 Modeling .....	32
1.2.3 Docking .....	32
1.2.4 MD Simulations .....	33
1.2.5 Trajectory Analysis .....	33
1.2.6 Receptor Modeling and Docking .....	34
1.2.7 Molecular Dynamics .....	38
1.2.8 Chemistry .....	40
1.2.9 Biological Assays .....	45
1.3 DISCUSSION .....	47
1.4 CONCLUSIONS .....	48
1.5 EXPERIMENTAL SECTION .....	49
1.5.1 General .....	49
1.5.2 Starting and Reference Materials .....	49
1.5.3 Synthesis of compounds <b>15a,b</b> through Bromination .....	49
1.5.4 Synthesis of Hydroxylamines <b>16a,b</b> .....	50
1.5.5 Synthesis of Aminols <b>17a,b</b> .....	51
1.5.6 Synthesis of Pyrimidine derivatives <b>18a,b</b> .....	51
1.5.7 Synthesis of Nucleosides <b>10a,b</b> .....	52
1.5.8 Synthesis of Nucleotide <b>11a,b</b> .....	53
1.6 REFERENCES .....	54
1.7 APPENDIX .....	57
1.7.1. Ligand efficiencies and trajectory analyses .....	57
1.7.2. Docking poses for triphosphate derivatives in 7CTT .....	58
1.7.3. Docking poses for monophosphate derivatives in 7BV2 .....	61
<b>CHAPTER 2</b> .....	<b>64</b>
2.1 TARGETING SARS-CoV-2's M <sup>PRO</sup> (NSP5) .....	64



2.2 RESULTS: COMPUTATIONAL ANALYSIS .....	70
2.2.1 Methods .....	70
2.2.2 Clustering .....	70
2.2.3 Modeling .....	70
2.2.4 Binding sites identification .....	70
2.2.5 Database .....	70
2.2.6 Docking .....	71
2.2.7 MD Simulations .....	71
2.2.8 Trajectory Analysis .....	71
2.2.9 Clustering and Allosteric Sites selection .....	72
2.2.10 Pocket 1 – Docking and Molecular Dynamics.....	75
2.2.11 Pocket 2 – Docking and Molecular Dynamics.....	83
2.2.12 Chemical approach to the scaffolds .....	88
2.2.16 Biological Assays .....	96
2.3 DISCUSSION.....	97
2.4 CONCLUSION .....	99
2.5 EXPERIMENTAL SECTION .....	100
2.5.1 General.....	100
2.5.2 Starting and Reference Materials .....	100
2.5.3 Synthesis of compounds <b>7a</b> .....	100
2.5.4 Synthesis of compound <b>8a</b> .....	101
2.5.5 Synthesis of compound <b>62</b> .....	101
2.5.6 Synthesis of cycloadducts <b>12a,b</b> .....	102
2.5.7 Synthesis of hydroxylamines <b>13a,b</b> .....	102
2.5.8 Synthesis of the aminols <b>177</b> and <b>193</b> .....	103
2.5.9 Protection of aminol <b>177</b> .....	104
2.5.10 Coupling of amine <b>14</b> with Fmoc-Gly-OH .....	104
2.5.11 Deprotection of <b>15</b> .....	105
2.5.12 Coupling of compound <b>181</b> with Fmoc-Gly-OH .....	105
2.5.13 Deprotection of <b>16</b> .....	106
2.5.14 Synthesis of cycloadducts <b>19a,b</b> .....	106
2.5.15 Synthesis of hydroxylamine <b>20</b> .....	107
2.5.16 Synthesis of the aminol <b>179</b> .....	108
2.5.17 Protection of aminol <b>179</b> .....	108
2.5.18 Coupling of amine <b>21</b> with Fmoc-Gly-OH .....	109
2.5.19 Deprotection of <b>22</b> .....	109
2.5.20 Synthesis of the 5-hydroxyisoxazolidine <b>11</b> .....	110
2.6 REFERENCES .....	111
2.7 APPENDIX .....	113
2.7.1 Docking Score, Ligand Efficiency, Structures.....	113
2.7.2 Trajectory analysis P1 .....	114
<b>CHAPTER 3 .....</b>	<b>118</b>
3.1 GENERAL DISCUSSION .....	118
HDA Cycloaddition of Nitrosocarbonyls .....	119
1,3-Dipolar Cycloadditions .....	120
In the present work .....	121
3.2 REFERENCES .....	123

<b>CHAPTER 4 .....</b>	<b>124</b>
4.1 FINAL CONCLUSIONS .....	124
4.2 REFERENCES .....	128
<b>CHAPTER 5 .....</b>	<b>129</b>
5.1 PUBBLICATIONS .....	129
<b>CHAPTER 6 .....</b>	<b>130</b>
6.1 ATTENDED COURSES .....	130
<b>CHAPTER 7 .....</b>	<b>131</b>
7.1 FUNDING .....	131

# Abbreviations

**(-)RNA** negative strand RiboNucleic Acid  
**3CL<sup>Pro</sup>** ChimoTripsin Like protease  
**3D** Three-Dimensional  
**ACE2** Angiotensin Converting Enzyme 2  
**AcOEt** Ethyl Acetate  
**Asp** Aspartate  
**AZT** AZidoThymidine  
**BtCoV** Bat CoronaVirus  
**BtRf-BetaCoV** Bat Beta CoronaVirus  
**CC<sub>50</sub>** Cytotoxic Concentration (50%)  
**COVID-19** COronaVirus Disease 19  
**CYP3A** CYtochrome P450 family3 subfamilyA  
**DCM** DiChloro Methane  
**DEAE-C** DiEthylAminoEthyl Cellulose  
**DIPEA** N,N-DiIsoPropylEthylAmine  
**DMSO** DiMethyl SulfOxide  
**DNA** DeoxyriboNucleic Acid  
**dsDNA** double-stranded DeoxyriboNucleic Acid  
**E** envelope  
**ENE** Enic  
**FTP** Favipiravir TriPhosphate  
**HBTU** Hexafluorophosphate Benzotriazole Tetramethyl Uronium  
**HCoV** Human CoronaVirus  
**HCV** Hepatitis C Virus  
**HDA** Hetero Diels-Alder  
**HEp-2** Human Epithelioma-2  
**HepG2** Human hepatoma G2  
**HIV** Human Immunodeficiency Virus  
**HRMS** High Resolution Mass Spectroscopy  
**IC<sub>50</sub>** Inhibition Concentration (50%)  
**LID** Ligand Interactions Diagram  
**M** membrane  
**MD** Molecular Dynamic  
**MERS-CoV** Middle East Respiratory Syndrome CoronaVirus  
**Mo** Molnupiravir  
**MO** MOrtar  
**M<sup>Pro</sup>** Main protease

**mRNA** messenger RiboNucleic Acid  
**MSA** Multiple Sequence Alignment  
**MTS** [3,4-(5- dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium salt]  
**N** nucleocapsid  
**NA** Nucleoside Analogue  
**nAb** neutralizing Antibodies  
**NC** NitrosoCarbonyl  
**NI** Nucleoside Inhibitor  
**NMO** N-methylMorpholine N-Oxide  
**NMR** Nuclear Magnetic Resonance  
**NNi** Non Nucleoside Inhibitor  
**NNRTI** Non Nucleoside Reverse Transcriptase Inhibitor  
**NSP** Non Structural Protein  
**NTP** Nucleoside TriPhosphate  
**ORF** Open Reading Frame  
**PCR** Polymerase Chain Reaction  
**PDB** Protein Data Bank  
**PL<sup>Pro</sup>** Papain Like protease  
**POC** Proof Of Concept  
**Pp** Polyprotein  
**QTOF** Quadrupole Time-of-Flight  
**R** Remdesivir  
**RBD** Receptor Binding Domain  
**RdRp** RNA-dependent RNA-polymerase  
**RMP** Remdesivir MonoPhosphate  
**RMSD** Root Mean Square Deviation  
**RNA** RiboNucleic Acid  
**RTP** Remdesivir TriPhosphate  
**S** spikSARS-CoV-2Severe Acute Respiratory Syndrome CoronaVirus 2  
**SPC** Simple Point Charge  
**ssDNA** single-stranded DeoxyriboNucleic Acid  
**TBDMS** tert-Butyldimethylsilyl  
**THF** Tetrahydrofuran  
**TRIR** Time-Resolved IR spectroscopy  
**VOC** Variant Of Concern  
**WGS** Whole-genome sequencing  
**WHO** World Health Organization  
**WIV1-CoV** Bat SARS-like coronavirus  
**WP** Working Package

## Riassunto

Nel 2020 è stato un fatto un salto a ritroso nel tempo, dall'era della medicina moderna alla grande influenza del 1918. La recente pandemia ci ha colti di sprovvisa, con i medesimi strumenti di difesa del secolo scorso. Come un unico gruppo diversi ricercatori da tutto il mondo hanno rivolto la loro attenzione alla scoperta di nuovi antivirali. A Pavia dal 1997 è attiva la linea di ricerca basata sulla chimica dei nitrosocarbonili generati a partire dai nitrilossidi. Tramite l'utilizzo di metodi computazionali si è voluto indagare a quali molecole potenzialmente attive come antivirali si potesse accedere tramite queste due vie sintetiche. Si sono studiati due diversi target enzimatici: RNA polimerasi e proteasi. Sul primo target sono stati condotti esperimenti di docking di derivati nucleotidici, caratterizzati da diverse modifiche dello scaffold e dell'eterobase, ottenuti mediante la cicloadizione 1,3-dipolare di diversi nitrilossidi all'*N*-benzoyl-2,3-ossazanorborn-5-ene, nelle strutture cristalline di SARS-CoV-2 RdRp nello stato pre- e post-catalitico. Le pose di docking sono state classificate e confrontate, le migliori sono state sottoposte a simulazioni di dinamica molecolare e le traiettorie analizzate focalizzandosi sui requisiti necessari alla catalisi. Con questi strumenti sono state proposte modifiche strutturali, per migliorare l'affinità dei composti alla tasca catalitica. Per quanto riguarda il secondo target, M<sup>pro</sup>, in seguito alle analisi eseguite sulle strutture più rappresentative identificate attraverso dinamica molecolare del dimero, si sono individuati due siti allosterici con buone caratteristiche di *druggability*. Caratterizzando le tasche allosteriche così identificate con un approccio di clustering, è stato eseguito sulle strutture rappresentative un *docking screening* di un database contenente composti ottenibili tramite le vie chimiche studiate in questa unità di ricerca. Ciò ha permesso la selezione di *candidate molecules* da sintetizzare. Test di attività antivirale sono in corso per valutare l'accordo tra previsioni computazionali ed i sistemi biologici studiati.

## Thesis Summary

In 2020, we took a leap back in time, from the era of modern medicine to the great flu of 1918. The recent pandemic caught us off guard, with the same defense tools as the last century. As one group, several researchers from all over the world have turned their attention to the discovery of new antivirals. The research line based on the chemistry of nitrosocarbonyls generated from nitriloxides has been active in Pavia since 1997. Through the use of computational methods we wanted to investigate which molecules potentially active as antivirals could be accessed via these two synthetic routes. Two different enzymatic targets have been studied: RNA polymerase and protease. Docking experiments of nucleotide derivatives were conducted on the first target, characterized by different modifications of the scaffold and the heterobase, obtained through the 1,3-dipolar cycloaddition of different nitriloxides to *N*-benzoyl-2,3-oxazanorborn-5-ene, in the crystal structures of SARS-CoV-2 RdRp in the pre- and post-catalytic state. The docking poses were classified and compared, the best ones were subjected to molecular dynamics simulations

and the trajectories analyzed focusing on the requirements necessary for catalysis. With these tools, structural modifications have been proposed to improve the affinity of compounds to the catalytic pocket. As regards the second target, M<sup>PrO</sup>, following the analyzes performed on the most representative structures identified through molecular dynamics of the dimer, two allosteric sites with good druggability characteristics were identified. By characterizing the allosteric pockets thus identified with a clustering approach, a docking screening of a database containing compounds obtainable via the chemical pathways studied in this research unit was performed on the representative structures. This allowed the selection of candidate molecules to synthesize. Antiviral activity tests are ongoing to evaluate the agreement between computational predictions and the biological systems studied.

## Riferimenti:

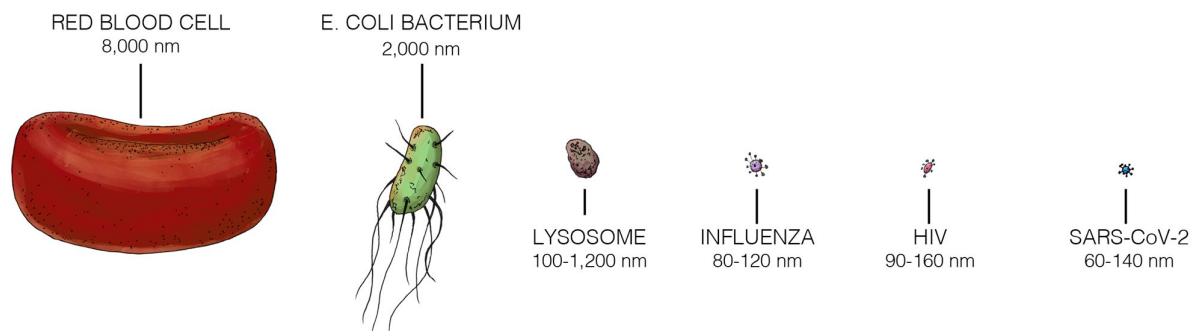
1. Memeo, M.G.; Lapolla, F.; Bovio, B.; Quadrelli, P. *Tetrahedron* **2004**, *60*, 3643-3651.
2. Memeo, M.G.; Quadrelli, P. *Chemical Reviews* **2017**, *117*, 2109-2200.
3. Yin, W.; Mao, C.; Luan, X.; Shen, D.D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y.C.; Tian, G.; Jiang, H.W.; Tao, S.C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Xu, H.E. *Science* **2020**, *368*, 1499-1504.
4. Morra, G.; Neves, M.A.C.; Plescia, C.J.; Tsustsumi, S.; Neckers, L.; Verkhivke, G.; Altieri, D.C.; Colombo, G. *Journal of Chemical Theory and Computation* **2010**, *6*, 2978-2989.
5. Alzyoud, L.; Ghattas, M.A; Atatreh, N. *Drug Design Development and Therapy* **2022**, *16*, 2463-2478.

# Introduction

The 21st century can be remembered divided into two parts: before and after the SARS-CoV-2 virus. Despite decades of warnings about the possibility of a deadly global pandemic, health systems around the world have been overwhelmed. In 2020, millions of people around the world radically changed their lifestyle to avoid contact with other people and, consequently, with the new coronavirus. Despite social distancing, many still got sick. As scientists are discovering, many viruses inhabit the human body as silent hosts, hidden in the cells of the lungs, blood, and nerves and within the multitudes of microbes that colonize the gut. Some viruses can cause disease while others can coexist with humans without problems. The continuous discoveries of scientists clearly indicate that our organism is not so much a complex of "human" cells sometimes invaded by microbes, but rather a superorganism in which cells, bacteria, fungi and viruses coexist, the latter being the most numerous around 380,000 billion, a figure ten times higher than that of bacteria.<sup>1</sup>

























## But What Is A Virus?

A virus (from the Latin *vīrus*, "poison") is a biological entity with characteristics of a symbiont or obligate parasite, as it replicates exclusively within the cells of organisms. Viruses can infect all life forms, from animals, to plants, to microorganisms (including other infectious agents such as bacteria) and even other viruses.<sup>2</sup> They coexist within host living systems in a symbiotic way and individuals of every animal species, including humans, normally host a high number of symbiotic viral species, which form a population called virome.<sup>3</sup> Viruses in their isolated form, called virion, are invisible under the optical microscope, unlike microorganisms; in fact, their dimensions range from tens to hundreds of nanometers, while an optical microscope reaches a maximum of 0.2  $\mu\text{m}$  (**Figure 1**).



**Figure 1.** Dimensions scale comparing red blood cell, Escherichia coli bacterium, lysosome, influenza, HIV, SARS-CoV-2.<sup>4</sup>

They are composed of DNA or RNA strands enclosed within a protein shell, called a "capsid"; they also contain enzymes, are surrounded by a further external envelope, formed by a lipid layer, and can be equipped with other types of proteins. The shapes of viruses can vary greatly, but some are more common: these include spherical, polyhedral, as well as helical, thread-like, amorphous or composite types (**Figure 2**).<sup>5</sup>

 <b>CORONAVIRUS   2019</b> Severe acute respiratory syndrome coronavirus 2 <b>SARS-CoV-2</b>	 <b>MERS   2012</b> Middle-East Respiratory Syndrome <b>MERS-CoV</b>	 <b>SARS   2002</b> Severe Acute Respiratory Syndrome <b>SARS-CoV</b>	 <b>AIDS   1981</b> Acquired Immune Deficiency Syndrome <b>HIV</b>	 <b>EBOLA   1976</b> (Nome di un Fiume del Congo) <b>EVD</b>	 <b>PAPILLOMA   1976</b> PaPoVa virus Pa (pilioma), Po (iloma), V (aculante) <b>HPV</b>	 <b>HANTA   1970</b> (Nome di un fiume della Corea del Sud) <b>HantaVirus</b>	 <b>INF. HONG KONG   1968</b> Seconda influenza <b>H3N2</b>
 <b>INF. ASIATICA   1957</b> Primo esempio di Influenza Aviaria <b>H2N2</b>	 <b>ZIKA   1947</b> Foresta Zika riserva naturale in Uganda <b>ZIKV</b>	 <b>INF.SPAGNOLA   1918</b> Delta la "Grande Influenza" <b>H1N1</b>	 <b>INF. RUSSA   1889</b> Soprannominata "L'asiatica" <b>H3N8</b>	 <b>FEBBRE GIALLA   1635</b> Tifo itericoide o febbre delle Antille <b>Flaviviridae</b>	 <b>POLIOMIELITE ARCAICO</b> Polio o Paralisi infantile <b>PolioVirus</b>	 <b>ROSOLIA   ARCAICO</b> Esantematica maulopapulosa <b>Rubivirus</b>	 <b>HERPES SYMPLEX ARCAICO</b> Herpes <b>HSV</b>
 <b>VARICELLA HERPES ARCAICO</b> Delta "Fuoco di Sant'Antonio" <b>VZV - HHV-3</b>	 <b>PAROTITE ARCAICO</b> Parotite epidemica ("orecchioni" o "gattoni") <b>Rubulavirus</b>	 <b>CYTOMEGALOVIRUS ARCAICO</b> Human Herpes virus 5 <b>CMV - HHV-5</b>	 <b>EPATITE ARCAICO</b> Ittero (dal greco ipar) <b>HxV A-B-C-D-E</b>	 <b>MORBILLO   500</b> Nel medioevo "Piccolo Morbo" <b>Morbillivirus</b>	 <b>DENGUE   276</b> Dalla frase Swahili "Ka-dinga" (spirito maligno) <b>DENV 1-2-3-4-5</b>	 <b>VAIOLO   2000 A.C.</b> Variola vera (dal latino "chiazzato") <b>Variola virus</b>	 <b>RABBIA   2900 A.C.</b> Dal sanscrito "Rabbash" (fare violenza) <b>Rhabdovirus</b>

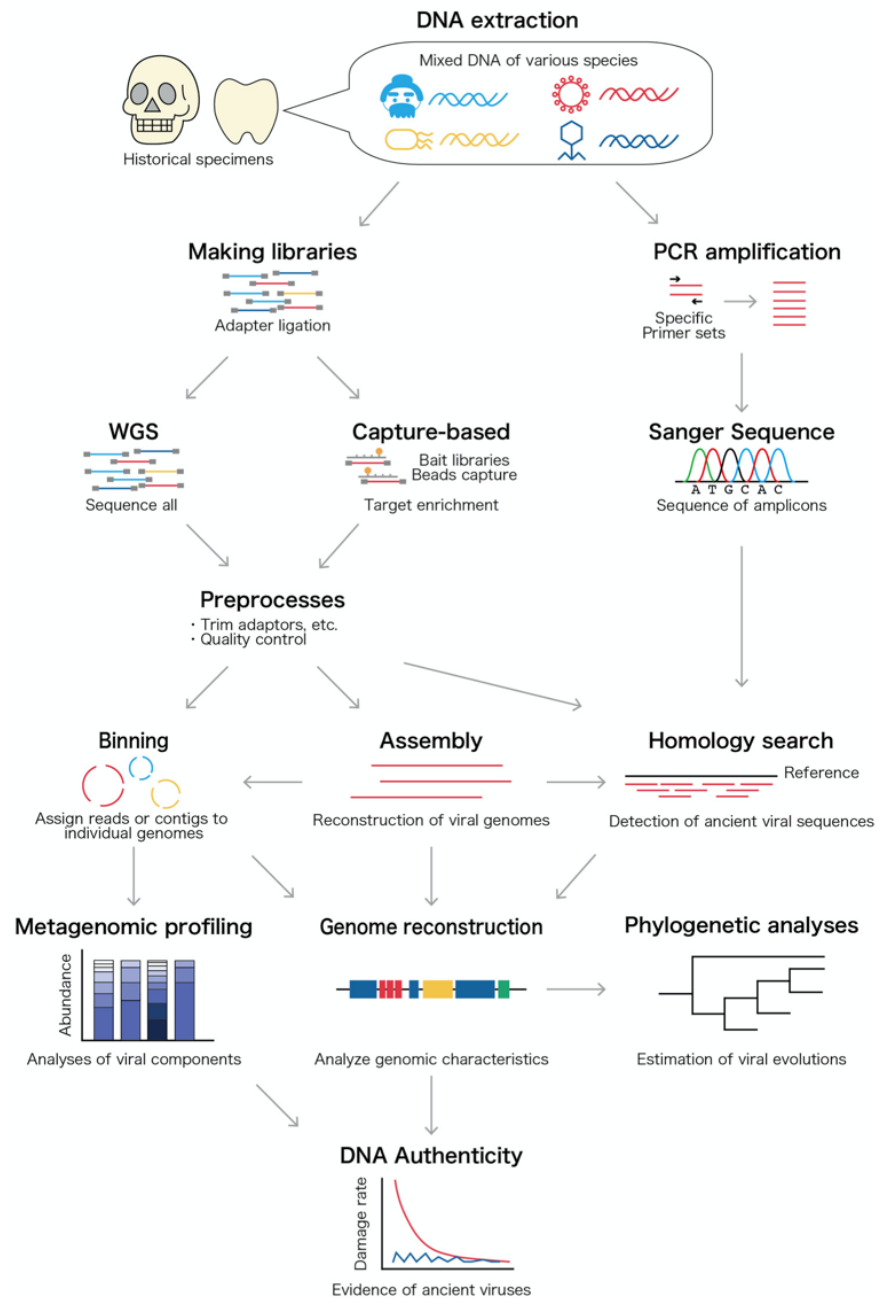
**Figure 2.** Different types of viruses, including common and formal names (WHO).<sup>6</sup>

Viruses have different characteristics of living beings as they possess a genome, reproduce and evolve through natural selection. However, according to most authors, although they contain DNA or RNA, they are not considered living beings as they are not organized into cells, not endowed with autonomous life and not capable of replicating autonomously. Since they possess only some characteristics, they have also been described as "organisms on the edge of life".<sup>7</sup>

But as we all well know they could have far too widespread effects on life.

# Virus Origin and Evolution

It is impossible to pinpoint the exact origin of a virus. Some more recent viral genome sequences will be discovered by PCR and other sensitive molecular techniques, but very ancient viral genomes will only be the subject of conjecture (**Figure 3**).<sup>8</sup>



**Figure 3.** Outline of the research and bioinformatic evaluations of the genomes of ancient viruses. [Reproduced with permission from ref 9. Creative Common CC BY 4.0 DEED license]



Comparative sequence research, however, clearly indicates that the evolutionary origins of both DNA and RNA viruses are deep and archaic, influencing both the structural organization of the genome and specific genomic and protein domains. Furthermore, it is evident that a range of processes, including as mutation, recombination, and reassortment, can lead to the emergence and evolution of both DNA and RNA viruses. Point mutations, insertions and deletions, the acquisition or loss of genes (and gene domains, or groupings of genes), genome rearrangements, and the use of alternative or inverted reading frames can all be involved in this. Although the nature of the first viruses remains unknown, it is most likely that they appeared very early in the history of life on Earth. The likelihood of RNA replicons being involved in elemental living forms is very high.<sup>8</sup>

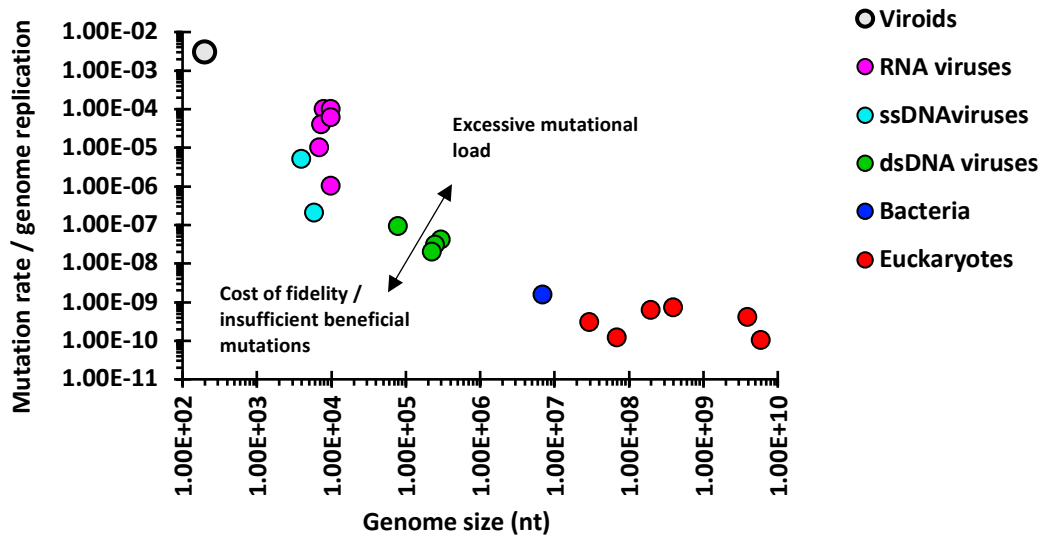
Although the nature of the first viruses remains unknown, it is most likely that they appeared very early in the history of life on Earth. The likelihood of RNA replicons being involved in elemental living forms is very high.<sup>8</sup>

According to Robertson's theory, RNA replicons that are primitive, autonomous, self-cleaving, and resemble modern viroids may have acquired extra genes to create conjoined replicons. These replicons then evolved into mosaic DNA-based entities, and they may even be responsible for the current, widespread prevalence of split genes, introns, and RNA-catalyzed cleavage and ligation splicing systems.<sup>10,11</sup>

The modular evolution theory of DNA viruses proposed by Botstein is a plausible theory. It suggests the recombinational arrangement of genetic elements or modules that can be swapped out to evolve viruses. Such modular evolution has clear advantages. It permits the independent evolution of virus genes, protein domains, regulatory systems, etc. under a broad range of selection circumstances. Clearly, this kind of modular mobility can loosen the evolutionary restrictions that would apply if everyone had to co-evolve inside a single genetic unit. Naturally, it will be an exceptionally uncommon occurrence that could result in a fortunate compatible recombination of independently evolving modules to produce a novel virus with strong biological adaptability. However, major virus emergences are also incredibly unusual events, and the likelihood of a radically different virus emerging from a single genome through mutational alterations is typically orders of magnitude lower.<sup>12</sup>

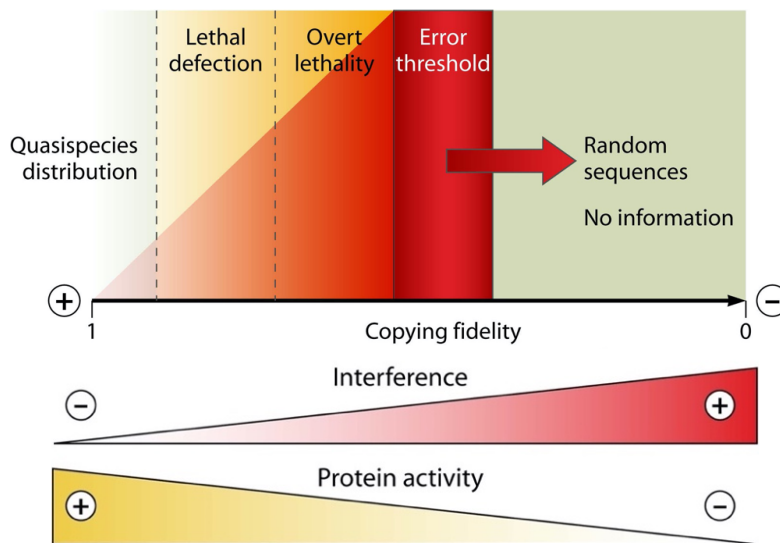
Therefore, it is probable that the origins of viruses date from great antiquity and continue to the present. This aligns with the inferred deep evolutionary branches for both DNA and RNA viruses and the widely acknowledged ability of viruses to obtain genetic components from host cell nucleic acids and generate more mobile replicons.<sup>8,12</sup>

The significant inverse relationship between genome size and mutation rate per genome replication is one of the most important discoveries in evolutionary genetics in recent years (**Graph 1**). Because of this, the tiniest viroids ( $\approx 400$  nucleotides long) have been found to have the highest error rates per nucleotide of any system, whereas bacteria and eukaryotes have mutation rates that are orders of magnitude lower.<sup>13</sup>



**Graph 1.** Connection between genome size and mutation rate in various genetic systems, such as viruses. Also displayed are the opposing evolutionary processes that could account for the small range of recorded error rates and genome sizes. Adapted from ref. 13.

Strong evidence that mutation rate is a trait optimized by competing selection forces may be seen in the conspicuous lack of data points with too high or abnormally low mutation rates for a given genome size. Mutation rates too low either slow down the rate of adaptive evolution or impose a physiological cost on increased fidelity that prevents the evolution of a zero mutation rate. On the other hand, mutation rates too high are likely to be selected against because they result in an excessively high number of deleterious mutations per replication and lose fitness (**Figure 4**).<sup>14</sup>

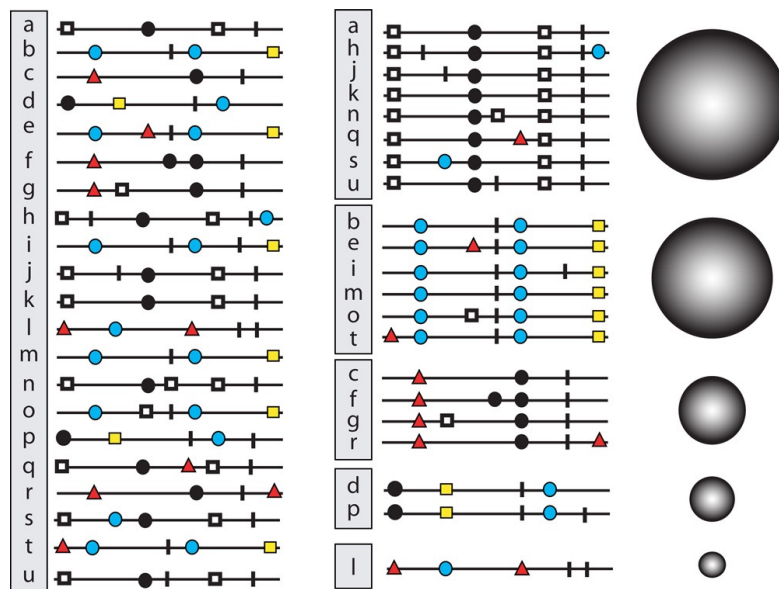


**Figure 4.** Concept of error threshold of quasispecies theory, modified by experimental findings with RNA viruses. [Reproduced with permission from ref 15. Copyright © 2012, American Society for Microbiology]

## RNA Virus Mutation Rates are Very High

The incredibly high mutation rates of all or almost all RNA viruses are now evident. While nucleotide base misincorporation rates range from  $10^{-4}$  to  $10^{-5}$  on average, mutation rates at specific base locations can vary significantly. Even in cases when the viral population has recently emerged from a clone, this leads to the creation of quasispecies mutant swarms. Clonal quasispecies virus populations, which can be thought of as a cloud in sequence space, are a heterogeneous combination of virus mutants that differ from one another at one or more genomic locations.<sup>8</sup>

The master sequence(s) represent the most fit member(s) of the swarm in any given determined selection environment, whereas the consensus sequence will reflect the average sequence at each genomic location. The master sequence(s) and the quasispecies swarm's overall composition will change when the selective environment does (**Figure 5**).

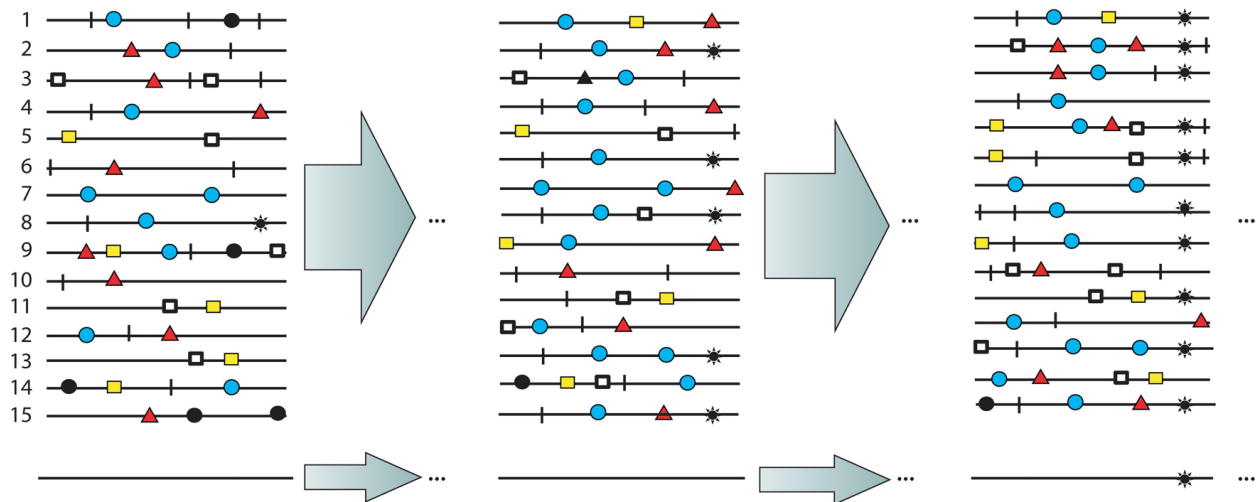


**Figure 5.** Mutations are shown as distinct colored symbols on horizontal lines that represent viral genomes. Differentiable viral subpopulations may be concealed by a quasispecies distribution. Mutant classes are depicted as spheres of sizes proportional to the number of genomes in each class. [Reproduced with permission from ref 15. Copyright © 2012, American Society for Microbiology]

Of course, the creation of quasispecies mutant swarms can give RNA viruses a great deal of flexibility in environments that are subject to change. In fact, viruses constantly face conditions that are changing in complex mammalian hosts, such as distinct cell types, inflammatory and immune responses, fever temperatures, and interferons. Note that under relatively constant

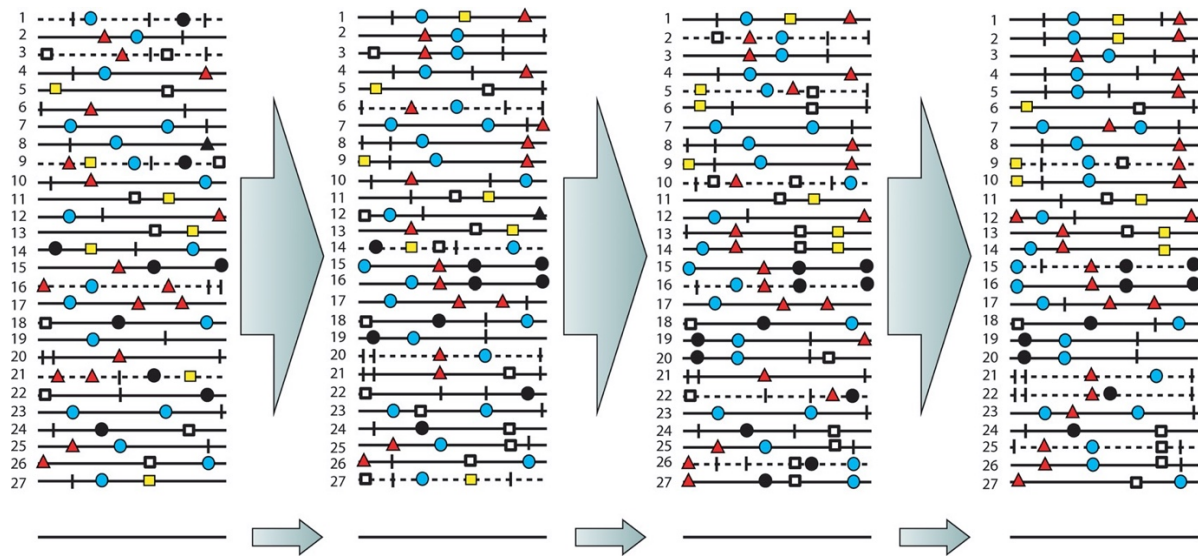
environmental conditions, the most adapted master sequence(s) and closely related variants will be the most abundant and important variants; under rapidly changing conditions (e.g., adaptation to a new mammalian host or a new arthropod vector), however, the opposite will be true.<sup>8</sup>

The highest chances for quick adaptation to the new environment are typically found in variants that are on the edge of the quasispecies mutant distribution, or those that are most atypical with respect to the prior master sequence. In order to maximize adaptation in the new environment and produce new master sequences, the quasispecies will often migrate through sequence space, selecting peripheral variations from the previous mutant distribution to become dominant variants in the new distribution (**Figure 6**).<sup>8</sup>



**Figure 6.** The rate of dominance in a population is changed by the selective advantage (relative fitness) of a mutant genome and its cloud. Large arrows indicate next generation of genomes. After a certain number of replication rounds, the mutation in genome 8 of the first distribution indicated by a black asterisk gains a selection advantage that leads to its dominance and modification of the consensus sequence (below the upper right distribution). [Reproduced with permission from ref 15. Copyright © 2012, American Society for Microbiology]

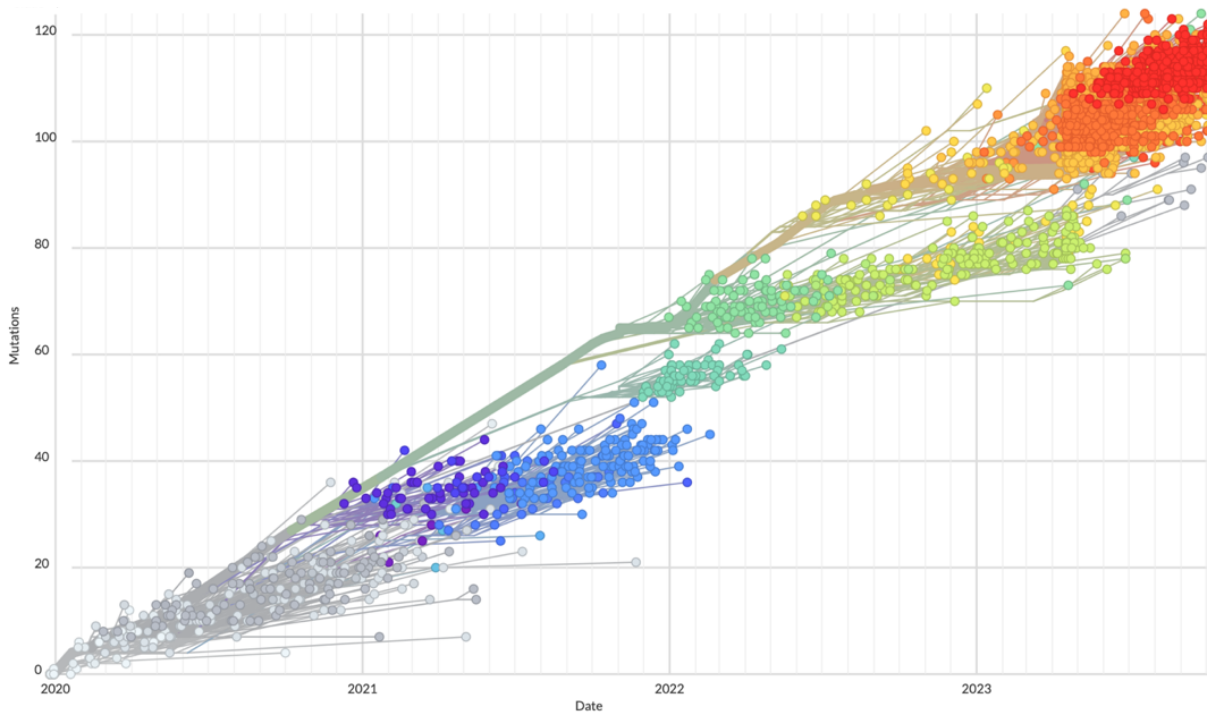
Error rates in RNA virus replicase that are on the verge of catastrophic errors offer the wide reach through sequence space that is fundamental to the quasispecies theory. As with all evolution, it should be highlighted that the quick appearance of new RNA variants is balanced by the quick disappearance of existing ones (**Figure 7**).<sup>8,16</sup>



**Figure 7.** Evolution (compositional change) of a virus quasispecies without affecting the consensus sequence (single line below). Genomes with five or more mutations that are unable to survive and serve as a template for the next generation of genomes are shown by discontinuous lines. Mutations can be incorporated into other genomes during replication; for example, genome 2 in the distribution on the left can produce genomes 2 and 3 in the second distribution, until an overabundance of mutations in genome 2 of the third distribution hampers. [Reproduced with permission from ref 15. Copyright © 2012, American Society for Microbiology]

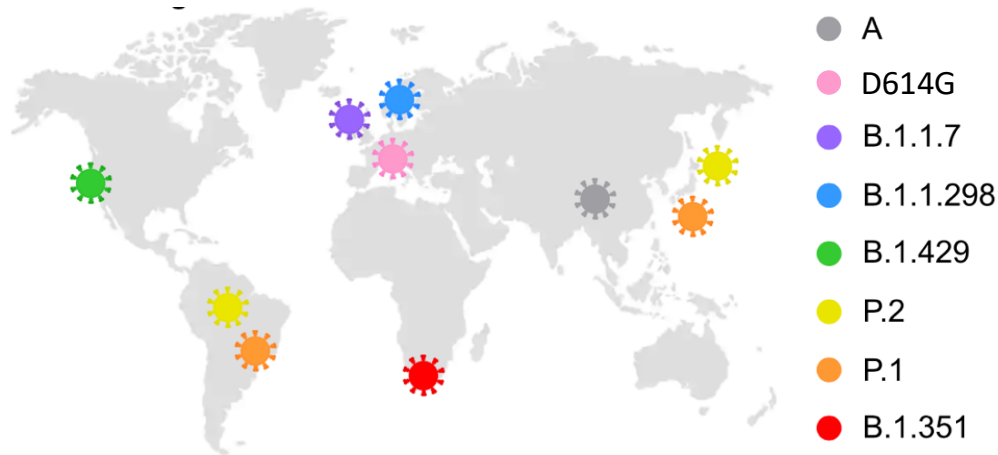
## SARS-CoV-2

What was previously discussed in the case of SARS-CoV-2 is highlighted in the timeline that follows (**Graph 2**). Clouds of virus mutants make up quasispecies mutant swarms that cover the sequence space.



**Graph 2.** Timeline of RNA mutations. Each color represents a different clade (variant) stabilized around a fixed number of mutations. [Reproduced with permission from ref 17,18. Creative Common CC BY 4.0 license]

Since the outbreak of SARS-CoV-2 in Wuhan, China in December 2019, the virus has rapidly started to evolve. The virus will eventually diverge into new strains due to these mutations; these strains will initially resemble the ancestral strain greatly, but they will gradually evolve further to create more diverse mutants. Numerous variations of SARS-CoV-2 have been documented subsequent to the first virus epidemic. The World Health Organization (WHO) has designated certain variants as variants of concern (VOC) due to their influence on world health, through evaluation of an increase in transmissibility or an increase or modification in clinical presentation. The primary VOCs are listed in (**Figure 8**), however throughout time, more sub-variants have developed from them.<sup>19</sup>

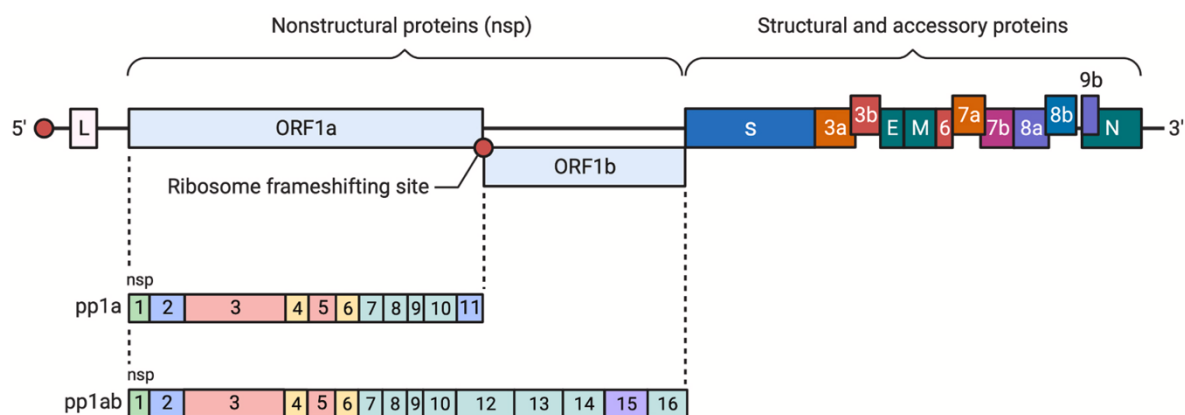


**Figure 8.** Emergence of SCoV-2 variant worldwide. Each color represents a different variant. [Reproduced with permission from ref 20. Creative Common CC BY-NC-ND 4.0 DEED license]

NSPs, accessories, and structural genes are encoded in the SARS-CoV-2 genome. The genome of SARS-CoV-2 has about 30,000 nucleotides and 14 ORFs that encode for at least 27 different proteins. Polyproteins produced by the translation of these ORFs by the machinery of the host cell are broken down by both host and viral proteases.

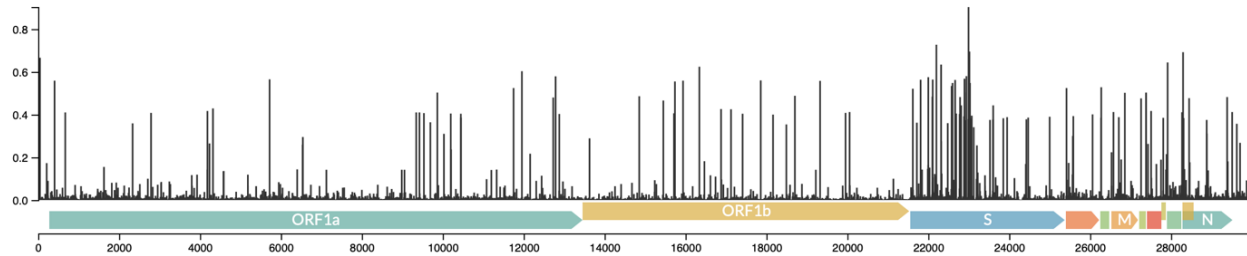
Two open reading frames, ORF1a and ORF1ab, on the viral RNA are translated into two long polyproteins, PP1a and PP1ab (**Figure 9**). The papain-like protease (PL<sup>Pro</sup>) and chymotrypsin-like protease (3CL<sup>Pro</sup>, also call M<sup>Pro</sup>) from the virus that are located inside the polyprotein complex are responsible for cleaving the proteins. These two enzymes are responsible for producing the mature NSPs.<sup>19</sup>

The structural proteins encoded at the 3' terminus include spike (S), nucleocapsid (N), membrane (M), and envelope (E) proteins (**Figure 9**).<sup>21</sup>



**Figure 9.** Polyprotein-1ab, which is cleaved into 16 NSPs in addition to auxiliary and structural proteins. [Reproduced with permission from ref 19. Creative Commons CC-BY 4.0 DEED license]

Even though several mutations have surfaced during the pandemic (**Graph 3**), careful investigation has revealed a general pattern: a large number of the arising and surviving mutants improve the infectivity and immune escape of SARS-CoV-2.



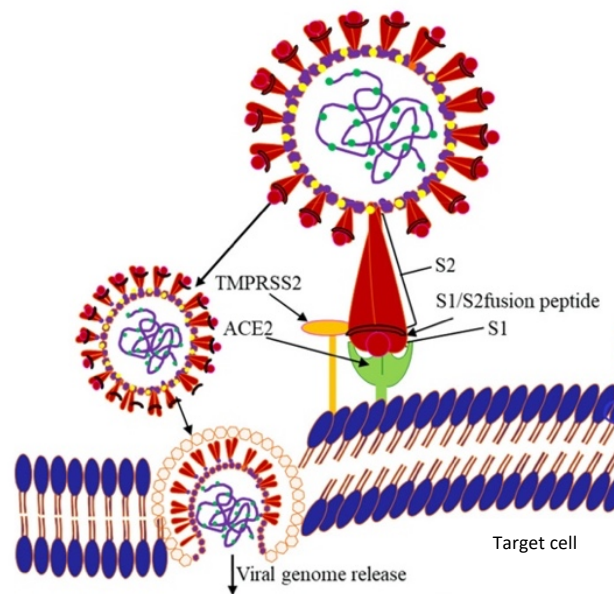
**Graph 3.** Plot of Nucleotide diversity of genome (Nucleotide substitution normalized to 1 per base position in SARS-CoV-2 genome). [Reproduced with permission from ref 17,18. Creative Common CC BY 4.0 license]

During host cell entry, the spike protein of SARS-CoV-2 helps in binding to the receptor on the surface of the host cell. While the S2 subunit of the S protein induces membrane fusion, the S1 subunit of the S protein aids in ACE2-mediated virus attachment.<sup>21,22</sup>



## Spike Protein: S1 Mutations

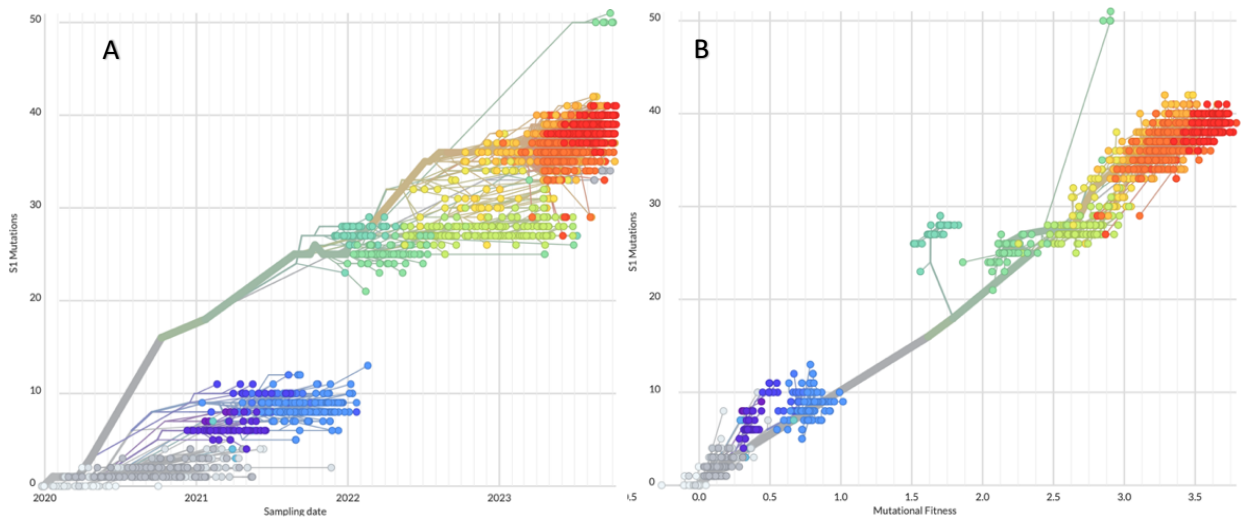
The envelope of corona-virion contains protruding projections from its surface called the large surface glycoproteins or spike proteins responsible for recognizing the host's receptor followed by its binding to it (S1) and fusing with its membrane (S2). S proteins also aid in the attachment of infected cells to nearby non-infected cells, which facilitates the virus's transmission. A Receptor Binding Domain (RBD) is present in the S1 subunit that recognizes and binds to the human ACE2 with 10 – 20 times more affinity than the SARS-CoV-1 domain (**Figure 10**).<sup>19</sup>



**Figure 10.** First, SARS-CoV-2 binds to the host cell receptor ACE2 via its S protein (RBD/S1), which causes the S2 subunit to undergo a conformational shift that where the subunit's fusion with the plasma membrane. [Reproduced with permission from ref 23. Creative Commons CC-BY license].

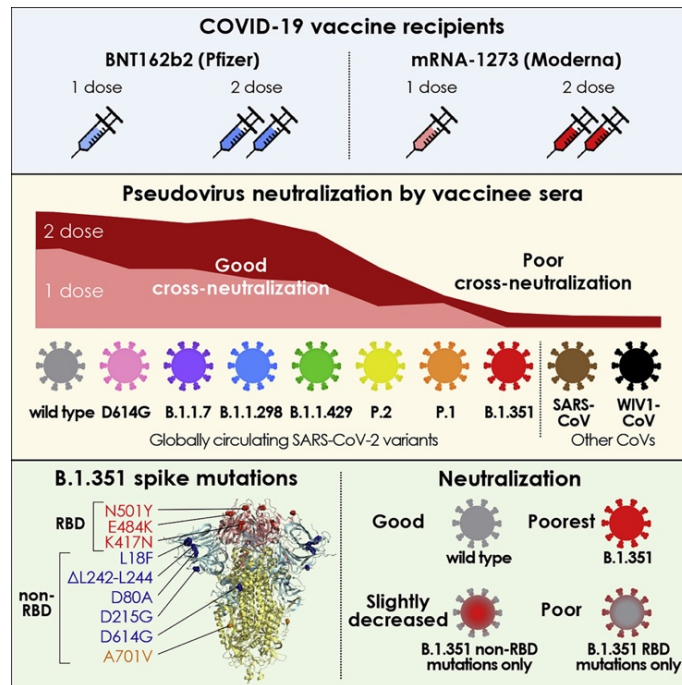
Merely focusing on mutation rates provides a restricted understanding of the virus's capacity to produce effective escape mutations. Certain vaccine-preventable viruses, like smallpox, have very low mutation rates ( $\approx 10^{-6}$  sub/nuc/yr), whereas other viruses, like poliovirus, have large mutation rates ( $\approx 10^{-2}$  sub/nuc/yr). The low antigenic evolvability of viruses like poliovirus, which have proven highly tractable to antibody-mediated prophylaxis via vaccines despite a high evolutionary rate, contrasts sharply with the high antigenic evolvability of viruses like influenza, which are noteworthy for their evolutionary capacity for immune evasion. Natural selection where antigenic evolvability, according to research on several viral illnesses.<sup>24</sup>

A positive selection of mutations within the SARS-CoV-2 spike protein has resulted in the genesis of new SARS-CoV-2 variants with significantly increased overall fitness (**Graph 4**). The spike protein mediates target recognition, cellular entry, and ultimately the viral infection that leads to different levels of COVID-19 severities.<sup>21</sup>



**Graph 4.** A) Timeline of S1 Mutations. B) Plot of S1 Mutations against Mutational Fitness. [Reproduced with permission from ref 17,18. Creative Common CC BY 4.0 license].

The circulating strains of SARS-CoV-2 are evolving at a rapid pace, and mutations in their S protein are expected to occur more often as selection pressures from host immunity from prior infections and/or vaccinations continue to drive this rapid evolution (**Figure 10**).<sup>21</sup>



**Figure 10.** Limited neutralization by vaccine-induced humoral immunity is observed in P.1 and B.1.351 in particular, according to analyses of sera from people who received one or two doses of mRNA vaccines against ten circulating forms of SARS-CoV-2. It was discovered that the receptor-binding domain alterations of SARS-CoV-2 spike were mostly responsible for this escape. [Reproduced with permission from ref 20. Creative Common CC BY-NC-ND 4.0 DEED license]

When considered collectively, the ability of RBD nAb escape mutations to produce infectious and viable viral particles capable of ACE2 binding, the immunodominance of the spike protein (immunodominant epitopes will curtail immune responses against non-dominant epitopes), and the immune response's restricted focus on a particular RBD region all point to a low evolutionary cost for the virus in producing nAb escape mutants.<sup>24</sup>

With new variants constantly emerging, future therapies, as well as vaccinations, will be more successful if they demonstrate effectiveness over a broad range of S variants. Given these rapid changes, it may also be beneficial to give particular consideration to therapies that function independently of S structure and functionality.

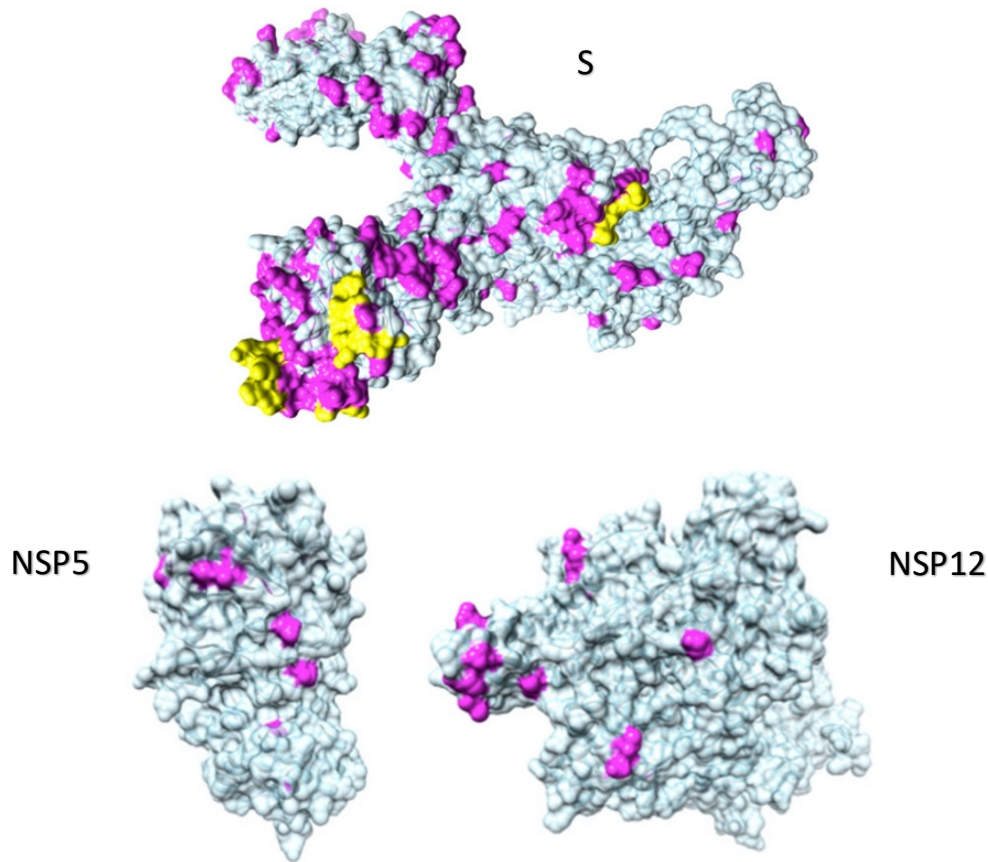
## Conserved Regions: Appealing Targets

Nucleotide or amino acid sequences that hold steady relative to one another across different species or far back in the phylogenetic tree are known as conserved sequences. Highly conserved sequences imply that natural selection has preserved the sequence across time. Multiple sequence alignment (MSA) studies can be used to identify the conserved areas found within the sequences, which may indicate structural or functional significance to the protein.<sup>19</sup>

In a study published in 2021, Almubaid and Al-Mubaid examined 1200 genomes sampled over the first seven months of 2020, taking into account the various SARS-CoV-2 variations. They experimentally demonstrated that, contrary to common belief, some mutations in the ScoV-2 genome do not arise at random. To adapt to changing environmental conditions, climatic conditions, and population changes, the ratios (frequency %) of certain mutations change over time. One of the paper's intriguing discoveries is that, among the genetic variants found, the ORF1a region predominantly contained synonymous mutations that, because of the degeneracy of the genetic code, kept the amino acid that the codon encoded. The third nucleotide of each codon triplet in this area was altered by the mutations discovered, although the reference genome's amino acids were preserved. NSP12 and NSP13 were also discovered to have a few synonymous variants.<sup>19,25</sup> There is more than 70% genetic similarity between SARS-CoV-2 and SARS-CoV-1, which was responsible for the 2003 SARS outbreak.

The ORF1ab polyprotein of SARS-CoV-2 and SARS-CoV-1 exhibited 94.6% sequence identity in seven conserved replicase domains, according to a pairwise amino acid sequence alignment.<sup>19</sup> The conserved enzymes 3Cl<sup>Pro</sup>/M<sup>Pro</sup> (NSP5), and RdRp (NSP12) share sequence similarities with viruses such as SARS-CoV-1 (98.7% and 98.1%, respectively), making them attractive targets for medical therapy (**Figure 11**). Given the sequence similarities between SARS-CoV-1, MERS-CoV, and SARS-CoV-2, it is conceivable that therapeutic compounds intended to target SARS-CoV-1 and MERS-CoV may also be effective against SARS-CoV-2.<sup>19</sup>

Given the similarities of constituent proteins within quasispecies swarm and moreover within RNA virus, in this thesis we propose the targeting of the conserved NSP regions of SARS-CoV-2 with the aim to develop, through robust and reliable chemical pathways, small molecule inhibitors obtainable in the gram scale, which may potentially work as pan inhibitors against future disease outbreaks caused by coronaviruses.



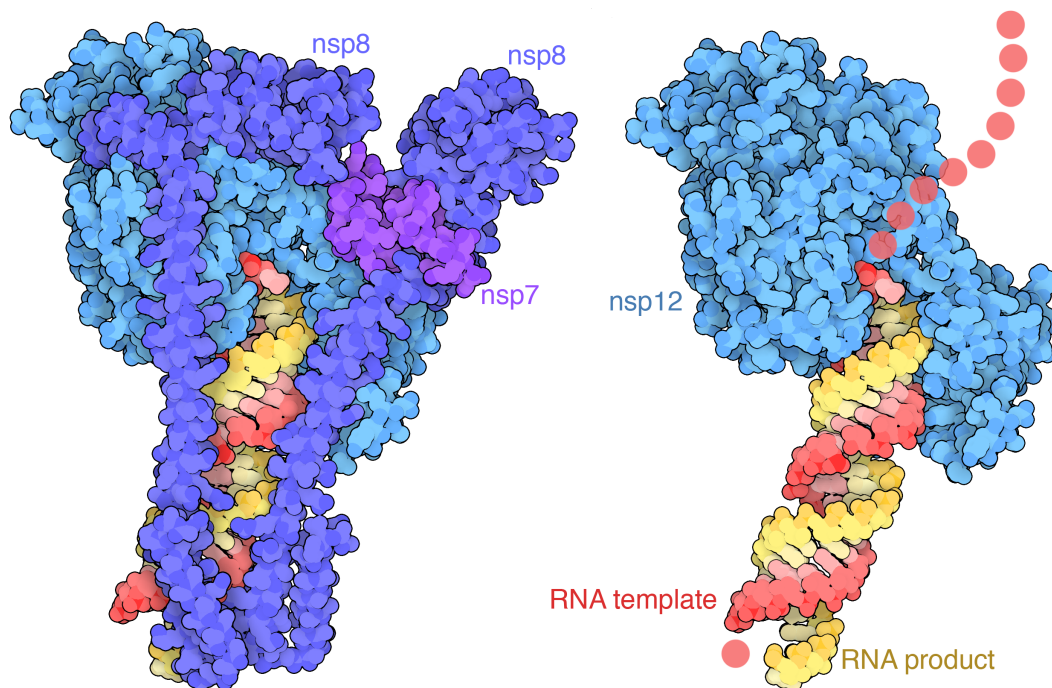
**Figure 11.** Highlighted in pink are mutations found when aligning the proteins against their homologs from the closest related coronaviruses: human SARS-CoV-1, bat coronavirus BtCoV, and another bat betacoronavirus BtRf-BetaCoV. Novel inserts are yellow. Sequence similarities: S protein 75%, NSP5 98.7%, NSP12 98.1%.<sup>19,27</sup> [Reproduced with permission from ref 27. Creative Common CC BY license]

One of the major achievements in the field of antiviral drug discovery has been the inhibition of viral protease and polymerase functions, which has led to the development of antiviral drugs that target various virus families and produced majority of the antiviral medications that are currently used to treat HIV and hepatitis C virus (HCV) infections. Remarkably, the WHO list of critical medications includes both polymerase inhibitors and viral proteases.<sup>26</sup>

There is a pressing need for medications to treat COVID-19 and other coronavirus infections worldwide, hence research is currently being done on several approaches that target the virus or the host cell. Protease and polymerase functions are the targets of direct-acting pharmaceuticals, which mark a significant advancement in antiviral therapy. The most well-characterized SARS-CoV-2 targets are the 3C-like (or Main) protease (3CL<sup>Pro</sup>) and the NSP12 RNA-dependent RNA-polymerase (RdRp). These targets exhibit the highest level of conservation across coronaviruses, which facilitates the discovery of broad-spectrum inhibitors.<sup>26</sup>

## The RNA-dependent RNA polymerase (RdRp) NSP12

The RNA-dependent RNA polymerase (NSP12) (**Figure 12**), which is essential for RNA replication, is known to be conserved within RNA viruses so targeting this protein may be an effective therapeutic approach.<sup>19</sup> 231 residues make up the common structural core of viral RdRps, which is found in 23% of SARS-CoV-2 polymerase structures, over 50% of polioviruses, and over 30% of influenza A viruses, according to a study by Möttönen.<sup>28</sup> Therefore, even though there is very little sequence similarity, there is a significant amount of structural conservation among all viral RdRp structures. This suggests that it would be feasible to pinpoint the target locations for broad-spectrum antivirals. The theory that all viral RdRps, including those from (-)RNA viruses, descended from a common ancestor is further supported by the great structural conservation.<sup>28</sup> Moreover, the RdRp of SARS-CoV-1 and the RdRp of flaviviruses like Zika, West Nile Virus, and Dengue Virus show 66–69% amino acid sequence similarities.<sup>19</sup>



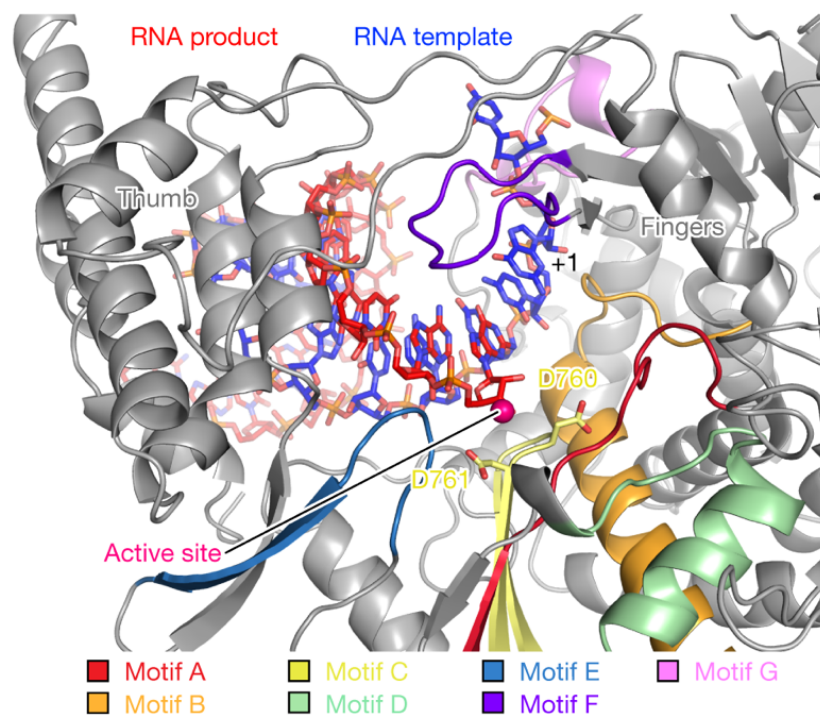
**Figure 12.** SARS-CoV-2's RdRp, PDB:6YTT. The virus proteins non-structural protein 12 (NSP12), NSP8, and NSP7, as well as more than two rotations of the RNA template-product duplex, compose the structure. With conserved residues, the active-site cleft of NSP12 mediates RdRp activity by binding to the first turn of RNA. The second RNA turn is positioned by two copies of NSP8 binding to opposing sides of the cleft. NSP8's long helical extensions emerge along the RNA that is leaving the enzyme to provide positively charged "sliding poles".<sup>30</sup> [Reproduced with permission from ref 29. Creative Common CC BY 4.0 license]



For the transcription of its genes and the replication of its genome, the novel coronavirus severe acute respiratory syndrome coronavirus requires an RNA dependent RNA polymerase (RdRp). With its fingers, palm, and thumb subdomains found in all single-subunit polymerases the RdRp domain resembles a right hand. The thumb is bound by subunits NSP7 and NSP8, while the fingers domain is bound by a second copy of NSP8. Specifically, RdRp activity is found in NSP12 and catalyses the primer-dependent production of a phosphodiester link between nucleoside triphosphates (NTPs). Two metal ions ( $Mg^{2+}$ ) bound by conserved Asp residues coordinate the NTP substrates engaged in this process. Strong amino acid conservation is seen in areas that are directly involved in nucleotide selection or catalysis, which is consistent with this conserved process.<sup>26,30</sup>

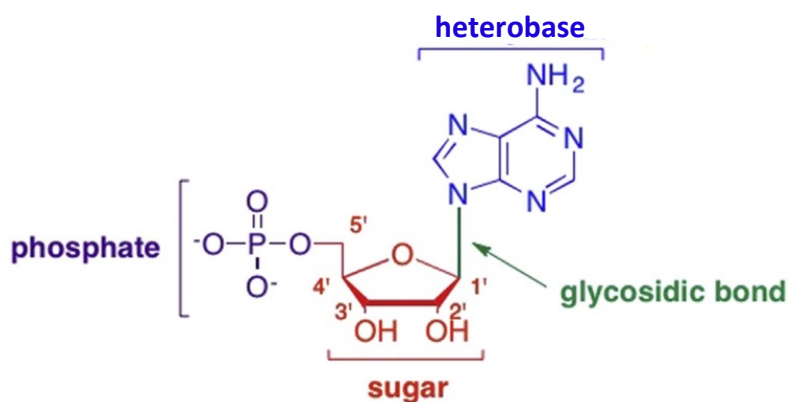
In fact, five conserved NSP12 components that make up the active site, referred to as motifs A–E, are found on the palm subdomain. The residues D760 and D761 found in motif C, which binds to the 3' end of RNA, are necessary for RNA synthesis. The RNA template is positioned by the extra NSP12 motifs F and G, which are found in the fingers subdomain (**Figure 13**).<sup>30</sup>

The crucial function of the SARS-CoV-2 RdRp in viral replication and its lack of an enzyme homolog in the host cell make it an excellent target for medication. As a matter of fact, the majority of medications that are authorised for the treatment of viral infections, such as HIV and HCV, are viral polymerases inhibitors, which constitute the cornerstone of antiviral therapies.<sup>26</sup>



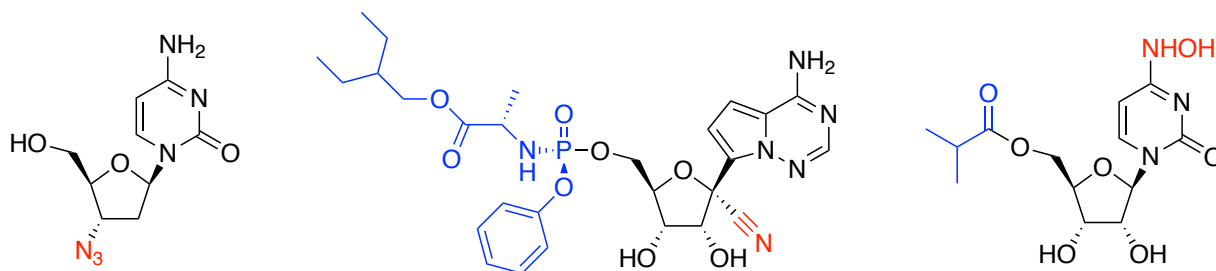
**Figure 13.** Interactions between the RNA's initial turn and the RdRp active site. Grey represents subunit NSP12, whereas colored represents conserved motifs A–G. The sticks represent the active-site residues D760 and D761. The magenta sphere shows a  $Mg^{2+}$  ion. [Reproduced with permission from ref 30. Copyright © 2020, Springer Nature Limited]

Based on their structure and method of action, inhibitors of viral polymerases can be divided into two main groups: non-nucleoside inhibitors (NNIs) which interact with allosteric binding sites, and nucleoside inhibitors (NIs) (**Figure 14**), which act at the substrate site. Despite the possibility that medicinal chemistry efforts could enhance the activity of the currently reviewed candidates, given the abundance of structural information that becomes available, only NIs have reached the most advanced stages of development thus far. On the other hand, NNIs for SARS-CoV-2 and other human pathogenic CoVs have not yet been described.<sup>26</sup> Since the active sites of RdRp have a high degree of conservation and a relatively low rate of mutation, NIs offer an appealing approach to inhibiting viral RNA replication. This allows for a broad antiviral activity and a high barrier to resistance.<sup>19,26</sup> Following the drug's binding by the viral RdRp in the form of NTP, the drug's hydrolysis and subsequent pyrophosphate group release supply the energy needed for the incorporation of nucleotide monophosphate into the developing RNA chain, hence inhibiting viral replication and strand elongation.<sup>26</sup>



**Figure 14.** Potential alteration sites for nucleoside/tide drugs include variations to the phosphate group, sugar moiety, glycosidic bond, and heterocyclic base. [Reproduced with permission from ref 31. © 2018 Elsevier B.V. All rights reserved]

Antiviral NIs fall into three categories: obligate chain terminators, nonobligate chain terminators, and mutagenic.<sup>26</sup> Below is reported an example of each category (**Figure 15**).



**Figure 15.** From left to right Azidothymidine (AZT), Remdesivir, Molnupiravir.<sup>19,31</sup>

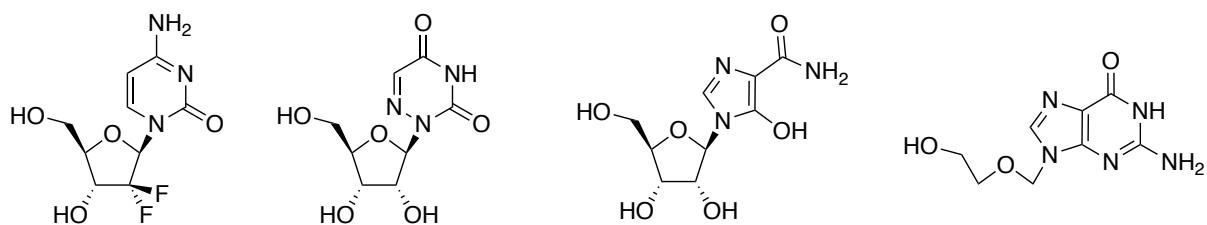
The 3'-hydroxyl group is absent from the riboside moiety of obligate chain terminators like AZT. Rather than having a natural base and a 3'-hydroxyl on the sugar, nonobligate chain terminators



like Remdesivir exhibit an extra substituent at the C-1' or C-2' locations of the ribose ring, preventing the formation of the phosphodiester bond with the incoming NTP. Lethal mutagenesis is caused by an inability to identify nucleoside analogues, such as Molnupiravir, as ordinary nucleobases (**Figure 15**). This results in base pairing errors and an increase in mutations, which eventually lead to nonviable genomes.<sup>19,31</sup>

Five NIs, Remdesivir,  $\beta$ -D-N4 hydroxycytidine (**Figure 15**), Gemcitabine, 6-Azauridine, Mizoribine, and Acyclovir-Fleximer (**Figure 16**), exhibit moderate-to-potent efficacy against three human pathogenic CoVs (HCoV-NL63, MERS-CoV, and SARS-CoV), making up the spectrum of NAs active against human pathogenic. One striking feature of this activity spectrum is the structural peculiarity of analogues relative to their potency towards CoVs as well as other RNA viruses. Of course, the activity spectrum can be associated with a modified base, a modified ribose, or both.

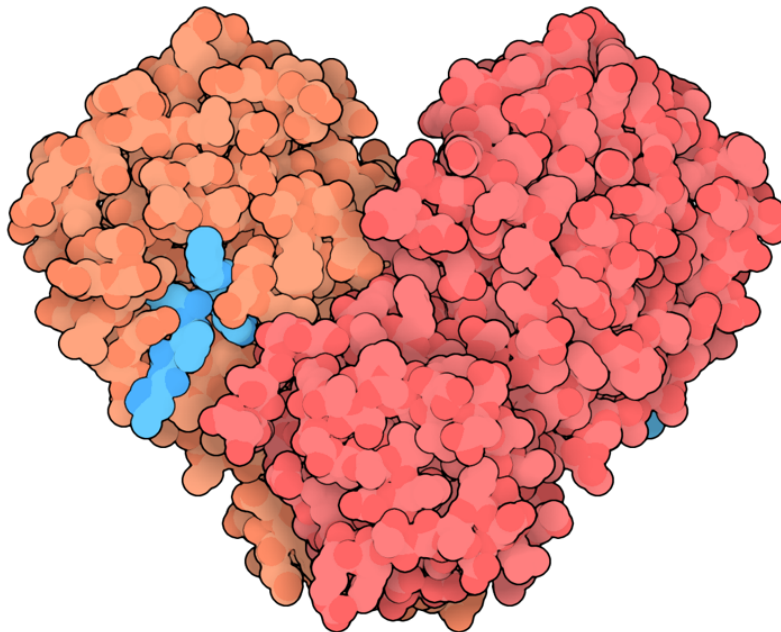
32



**Figure 16.** From left to right Gemcitabine, 6-Azauridine, Mizoribine, and Acyclovir-Fleximer.<sup>32</sup>

## 3C-like or Main protease (M<sup>Pro</sup>) NSP5

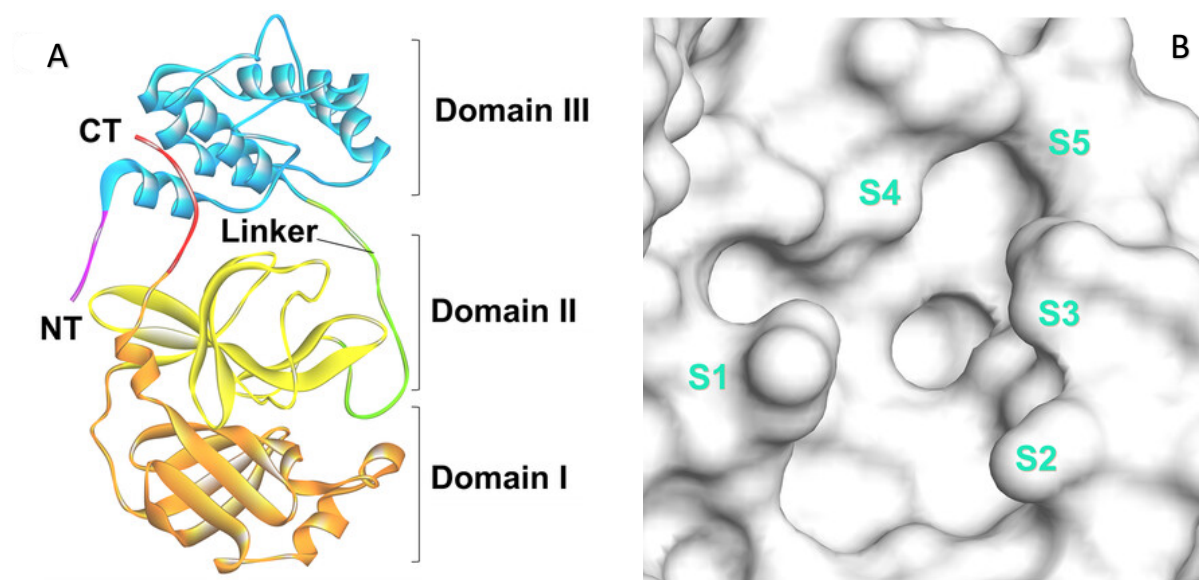
The viral protease NSP5, is highly conserved, has a chymotrypsin-like fold (3CLpro) and is also known as the main protease (M<sup>Pro</sup>) (**Figure 17**). Cross-coronavirus structural conservation is strong at the tertiary and quaternary levels, particularly in domains 1 and 2 of M<sup>Pro</sup>. Coronaviruses are known to exhibit a high degree of conservation, with the N-terminal domains (1 and 2) forming a chymotrypsin-like fold made up of  $\beta$ -barrels. The catalytic site of SARS-CoV-2 and SARS-CoV-1 proteases contains two highly conserved residues, namely His41-Cys145. As a broad-spectrum treatment approach, Goyal and Goyal reviewed the significance of focusing on the primary protease's dimerization. Among beta coronaviruses, the residues required for dimerization are substantially conserved.<sup>19,33</sup>



**Figure 17.** SARS-CoV-2's M<sup>Pro</sup>, PDB:6LU7. Protomer A and B, two identical polyproteins, combine to produce a dimer that comprises two catalytic sites. The substrate-binding site of M<sup>Pro</sup> is situated in a cleft between domain I and domain II, and the protein has a Cys-His catalytic dyad. These characteristics resemble those of M<sup>Pro</sup> from other coronaviruses that have been previously identified.<sup>34</sup> [Reproduced with permission from ref 29. Creative Common CC BY 4.0 license]

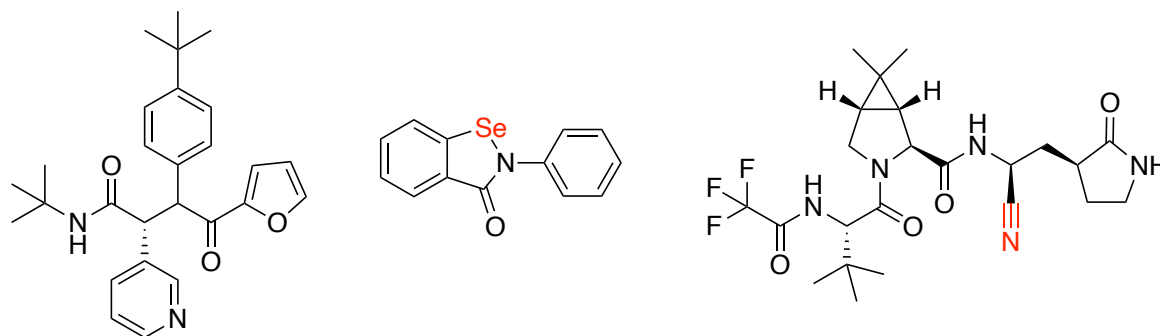
The mature form of M<sup>Pro</sup> is dimeric, with individual monomers having reduced enzymatic activity. The monomers are made up of three domains: domain I, domain II, and domain III (**Figure 18A**). Domain III is an extra helical domain among them, and it is through its aggregation that M<sup>Pro</sup> dimerizes. The dimeric form of M<sup>Pro</sup> functions as a functional unit with the maximum hydrolytic activity, whereas the monomer is typically a transient state that has been shown to be less enzymatically active if none. One of the essential prerequisites for the formation of dimeric M<sup>Pro</sup>

is a strong binding between the N-finger and C-terminus, particularly the salt bridges of Arg4 and Arg298. The M<sup>Pro</sup> catalytic site is situated where domains I and II converge. This region can be further subdivided into five primary sub-pockets: S1, S2, S3, S4, and S5 (**Figure 18B**). The dimensional chemical environment of key facial residues of the five sub-pockets matches five specific substrate-binding positions. The locations P4, P3, and P3' improve substrate identification and stable binding, whereas P1, P2, and P1' positions primarily control M<sup>Pro</sup>'s substrate specificity. Glutamate's O<sup>β</sup> atom may attach to the oxyanion hole (residues 143–145) of S1, allowing the thiol of Cys145 to act as a nucleophile against the C atom of glutamine. Consequently, P1 needs glutamine or lactam warhead virtually invariably. Interestingly, in the homodimer, only one catalytic site has a hydrolytic action.<sup>34</sup>



**Figure 18.** A) Each protomer consists of three domains: Domain I, which comprises residues 8–101, and domain II, which comprises residues 102–184, exhibit an antiparallel  $\beta$ -barrel structure. Meanwhile, Domain III, which comprises residues 201–303, is related to domain II by a lengthy loop region that has five  $\alpha$ -helices grouped into a globular cluster that is predominantly antiparallel. B) Surface illustration of SARS-CoV-2 M<sup>Pro</sup>'s catalytic pocket (sub-pockets: S1–S5). [Reproduced with permission from ref 35. Copyright © 2020, Springer Nature Limited]

With the ability to process the polyprotein at no fewer than 11 conserved sites, beginning with the autolytic cleavage from pp1a and pp1ab, 3CL<sup>Pro</sup>, also referred to as M<sup>Pro</sup>, is in charge of releasing the majority of the NSPs CoVs functional proteins. The identification of anti-SARS-CoV-2 medicines finds 3CL<sup>Pro</sup> to be a highly interesting target due to its key function in the viral life cycle and the absence of comparable proteins in human cells. The therapeutic effectiveness of HIV and HCV proteases inhibitors indicates that targeting proteases is a viable avenue for the identification of antiviral drugs. Several strategies can be used to inhibit these enzymes; in fact, the development of reversible, covalently reversible, and irreversible binders has been utilized in a variety of therapeutic contexts, including the development of antivirals (**Figure 19**).<sup>26</sup>



**Figure 19.** From left to right ML188 a reversible noncovalent inhibitor, Ebselen an irreversible covalent inhibitor, Nirmatrelvir a covalent reversible inhibitor.<sup>36,37,38</sup>

Finding strong noncovalent inhibitors is usually a laborious and time-consuming procedure, but it's the preferred method for obtaining medications that are both safe and effective. The identification of irreversible inhibitors, on the other hand, seems like a more straightforward and appealing strategy, but it has a major flaw in that it lacks target selectivity, which can result in unpredictable, severe toxicity. A less hazardous but comparable strategy is the creation of covalent reversible inhibitors, which depends on natural substrate modification and necessitates understanding of the catalytic mechanism and substrate specificity. Such a strategy generally entails the transformation of good substrates into good covalent reversible inhibitors. This is accomplished by substituting in the natural substrate a moiety, referred to as a "warhead," which is an electrophilic reactive group capable of forming a covalent reversible bond with the catalytic site (ie. Michael acceptor).<sup>26</sup>

With this new drug design approach emerged in last decade we completed the depiction from a medicinal-chemistry point of view on viruses in general and specifically on SARS-CoV-2.<sup>39-41</sup>

## References

- 1) Pride, D.; I Virus che vivono in noi *Le Scienze* **2021**, *4*, 32-39.
- 2) La Scola, B.; Desnues, C.; Pagnier, I.; Robert, C.; Barrassi, L.; Fournous, G.; Merchat, M.; Suzan-Monti, M.; Forterre, P.; Koonin, E.; Raoult, D. The virophage as a unique parasite of the giant mimivirus *Nature* **2008**, *455*, 100–104.
- 3) Abeles, S.R; Pride, D.T. Molecular Bases and Role of Viruses in the Human Microbiome . *Mol. Biol.* **2014**, *426*, 3892-3906.
- 4) Dajose, L. Researchers Focus on the Novel Coronavirus that Causes COVID-19 *Caltech Mag.* **2020**.
- 5) Harrison, S.C. Principles of Virus Structure *Virology* **1990**, 2<sup>nd</sup> Edition, *Chapter 3*, 37-61.
- 6) Leonardi, L.; Sequi, E. Collezione delle Cere Anatomiche “Luigi Cattaneo”.
- 7) Koonin, E.V.; Starokadomskyy, P. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question *Stud. Hist. Philos. Biol. Biomed. Sci.* **2016**, *59*, 125-134.
- 8) Holland, J.; Domingo, E. Origin and evolution of Viruses *Virus Genes* **1988**, *16*, 13-21.
- 9) Nishimura, L.; Fujito, N.; Sugimoto, R.; Inoue I. Detection of Ancient Viruses and Long-Term Viral Evolution *Viruses* **2022**, *14*, 1336-1360
- 10) Robertson, H.D. in Holland J.J. (ed.) Genetic Diversity of RNA Viruses *Curr. Top. Microbiol. Immunol.* **1992**, *176*.
- 11) Robertson, H.D. How Did Replicating and Coding RNAs First Get Together? *Science* **1996**, *274*, 66-67.
- 12) Botstein, D.A. Theory of Modular Evolution for Bacteriophages *Ann. N. Y. Acad. Sci.* **1980**, *354*, 484-491.
- 13) Forterre, P.; Prangishvili, D. The origin of Viruses *Res. Microbiol.* **2009**, *160*, 466-472.
- 14) Holmes, E.C. What Does Virus Evolution Tell Us about Virus Origins? *J. Virol.* **2011**, *85*, 5257-5251.
- 15) Domingo, E.; Sheldon, J.; Perales, C. Viral Quasispecies Evolution *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 159-216.
- 16) Domingo, E. Molecular Basis of Genetic Variation of Viruses: Error-Prone Replication *Vir. Popul.* **2016**, 35–71.
- 17) Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: real-time tracking of pathogen evolution *Bioinformatics* **2018**, *34*, 4121–4123. <https://nextstrain.org/ncov>
- 18) Sagulenko, P.; Puller, V.; Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis *Virus Evol.* **2018**, *4*, vex042 9 pages.
- 19) Kandwal, S.; Fayne, D. Genetic conservation across SARS-CoV-2 non-structural proteins – Insights into possible targets for treatment of future viral outbreaks *Virology* **2023**, *581*, 97-115.
- 20) Garcia-Beltran, W.F.; Lam, E.C.; Denis, St.K.; Nitido, A.D.; Garcia, Z.H.; Hauser, B.M.; Feldman, J.; Pavlovic, M.N.; Gregory, D.J.; Poznansky, M.C.; Sigal, A.; Schmidt, A.G.;

- lafrate, A.J.; Naranbhai, V.; Balazs A.B. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity *Cell* **2021**, *184*, 2372-2383.
- 21) Magazine, N.; Zhang, T.; Wu, Y.; McGee, M.C.; Veggiani, G.; Huang, W. Mutations and Evolution of the SARS-CoV-2 Spike Protein *Viruses* **2022**, *14*, 640-651.
  - 22) Satarker, S.; Nampoothiri, M. Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2 *Arch. Med. Res.* **2020**, *51*, 482-491.
  - 23) Yadav, R.; Chaudhary, J.K.; Jain, N.; Chaudhary, P.K.; Khanra, S.; Dhamija, P.; Sharma, A.; Kumar, A.; Handu, S. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19 *Cells* **2021**, *10*, 821-837.
  - 24) Egeren, D.V.; Novokhodko, A.; Stoddard, M.; Tran, U.; Zetter, B.; Rogers, M.; Pentelute, B.L.; Carlson, J.M.; Hixon, M.; Joseph-McCarthy, D.; Chakravarty, A. Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein *PLOS ONE* **2021**, *16*, e0250780 17 pages.
  - 25) Almubaid, Z.; Al-Mubaid, H. Analysis and comparison of genetic variants and mutations of the novel coronavirus SARS-CoV-2 *Gene Rep.* **2021**, *23*, 101064 10 pages.
  - 26) 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.* **2022**, *65*, 2716-2746.
  - 27) Srinivasan, S.; Cui, H.; Gao, Z.; Liu, M.; Lu, S.; Mkandawire, W.; Narykov, O.; Sun, M.; Korkin, D. Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins *Viruses* **2020**, *12*, 360-377.
  - 28) Mönttinen, H.A.M.; Ravantti, J.J.; Poranen, M.M. Structure Unveils Relationships between RNA Virus Polymerases *Viruses* **2021**, *13*, 313-330.
  - 29) David S. Goodsell and the RCSB PDB doi:10.2210/rcsb\_pdb/mom\_2020\_9.
  - 30) Hillen, H.S.; Kobic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase *Nature* **2020**, *584*, 154-167.
  - 31) Seley-Radtke, K.L.; Yates, M.K. The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold *Antivir. Res.* **2018**, *154*, 66-86.
  - 32) Shannon, A.; Le, N.T.T.; Selisko, B.; Eydoux, C.; Alvarez, K.; Guillemot, J.C.; Decroly, E.; Peersen, O.; Ferron, F.; Canard, B. Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites *Antivir. Res.* **2020**, *178*, 104793 9 pages.
  - 33) Goyal, B.; Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy *ACS Comb. Sci.* **2020**, *22*, 297-305.
  - 34) Hu, Q.; Xiong, Y.; Zhu, G.H.; Zhang, Y.N.; Zhang, Y.W.; Huang, P.; Ge, G.B. The SARS-CoV-2 main protease (M<sup>pro</sup>): Structure, function, and emerging therapies for COVID-19 *MedComm* **2022**, *3*, e151 27 pages .
  - 35) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L.W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors *Nature* **2020**, *582*, 289-293.

- 36) Lockbaum, G.J.; Reyes, A.C.; Lee, J.M.; Tilvawala, R.; Nalivaika, E.A.; Ali, A.; Yilmaz, K.N.; Thompson, P.R.; Schiffer, C.A. Crystal Structure of SARS-CoV-2 Main Protease in Complex with the Non-Covalent Inhibitor ML188 *Viruses* **2021**, *13*, 174 9 pages.
- 37) Amporndanai, K.; Meng, X.; Shang, W.; Jin, Z.; Rogers, M.; Zhao, Y.; Rao, Z.; Liu, Z.J.; Yang, H.; Zhang, L.; O'Neill, P.M.; Hasnain, S.S Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives *Nat. Commun.* **2021**, *12*, 3061-3068.
- 38) Greasley, S.E.; Noell, S.; Plotnikova, O.; Ferre, R.A.; Liu, W.; Bolanos, B.; Fennell, K.; Nicki, J.; Craig, T.; Zhu, Y.; Stewart, A.E.; Stepan, C.M. Structural basis for the in vitro efficacy of nirmatrelvir against SARS-CoV-2 variants *J. Biol. Chem.* **2022**, *298*, 101972 8 pages.

# Chapter 1

Over the last three years, we have been experiencing the world health crisis triggered by a devastating pandemic, which challenged us both in public health policies and in pharmacological tools. Starting from the end of 2019, SARS-CoV-2 rapidly spread worldwide since the first localized epicenter in China's Hubei Province, causing COVID-19, a disease associated with a severe form of interstitial pneumonia.<sup>1</sup> The pandemic advanced with exponential growth resulting in millions of cases and deaths.<sup>2,3</sup> With an unprecedented effort, anti-COVID-19 vaccination campaigns started at the end of 2020 in several countries, promising to end the pandemic.<sup>4</sup> Despite vaccination success, a worldwide uneven distribution of doses, vaccine hesitancy in the population, emergence of virus variants, and noticing of limited duration of protective immunity made eradication an unlikely outcome and highlighted the need for efficacious pharmacological treatments.<sup>5-7</sup> Since the beginning of the pandemic, a critical aspect has been the lack of effective antiviral therapeutics to treat COVID-19 patients. Repurpose of antivirals already approved exhibiting inhibitory activity against either proteases (Lopinavir, Ritonavir) or RNA-dependent RNA-polymerases (RdRp) of other viruses (Remdesivir, Favipiravir), was initially proposed as a possible strategy and inspired a number of studies.<sup>8</sup> Various outcomes pushed toward the development of new therapeutics specifically targeted toward SARS-CoV-2. The use of Molnupiravir, prodrug of  $\beta$ -D-N4-hydroxycytidine (from Merck), was approved as a competitive inhibitor of SARS-CoV-2 RdRp.<sup>9</sup> These recent determinations have oriented the development of antiviral drugs targeting RdRp that is an essential enzyme for the viral replication process, catalyzing the viral RNA synthesis using a metal ion-dependent mechanism.

An updated overview has reported the main sequence and structural features of the RdRp of emerging RNA viruses such as Coronaviruses, Flaviviruses, and HCV, as well as inhibition strategies implemented so far.<sup>11</sup>

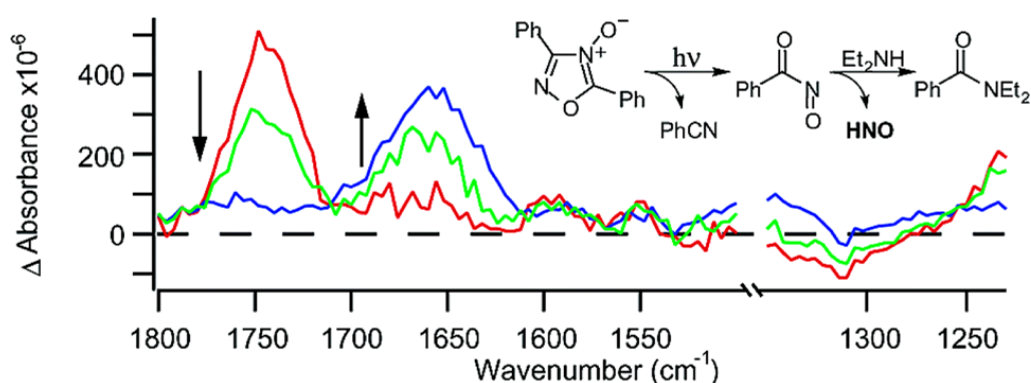
## 1.1 Targeting SARS-CoV-2's RdRp (NSP12)

For more than 20 years, the research group I worked in have been investigating the synthetic methodologies in organic chemistry to achieve reliable advancement in the knowledge of chemical structures apt to behave as antiviral compounds to contrast infectious diseases.<sup>12</sup> We strongly promoted the pericyclic reaction approach to several compounds and in particular the exploitation of the chemistry of nitrosocarbonyls, a family of fleeting intermediates suitable for the stereochemical ordinate introduction on carbocyclic scaffolds of functional groups suitable for the linear construction of purine and pyrimidine heterobases as well as introduction of phosphate units.<sup>13</sup> Nitrosocarbonyl chemistry offers indeed rapid and reliable access to a variety of organic compounds, often capable of remarkable biological activities in different fields.



### 1.1.1 Nitrosocarbonyl

The nitrosocarbonyl (NC) intermediates R-CONO are highly reactive species that were discovered at the beginning of the 1970s by inferring their existence from their demonstrated reactivity, by generating them in situ in the presence of a trapping agent. We have to wait until the year 2003 to find a definitive direct observation of an NC in solution by using time-resolved IR spectroscopy (TRIR) **Figure 1**. Toscano and co-workers had the intuition to use the 3,5-diphenyl-1,2,4-oxadiazole-4-oxide (Wieland heterocycle) to generate photochemically the nitrosocarbonyl benzene, in accordance with the protocol proposed by the Pavia group.<sup>13-15</sup>



**Figure 1.** TRIR spectrum of nitrosocarbonyl. [Reproduced from ref 15. Copyright 2003 American Chemical Society.]

When the nitrosocarbonyl is formed by photochemical cleavage of the heterocycle, the red line of the TRIR spectrum is shown; upon addition of diethylamine the signal of NC is progressively quenched to leave the amide represented by the blue line. The stability of nitrosocarbonyl benzene was determined to be at least 180  $\mu\text{s}$ . These results are the cornerstone and the ultimate evidence of the NC intermediates.<sup>13,15</sup>

### 1.1.2 Nitrosocarbonyl Generation and Reactivity

The range of applications of NCs in organic synthesis has been widened in recent years by the discovery (and tuning) of new generation methods allowing for the use of these transient species to achieve many multifunctionalized molecules. The nitrosocarbonyl intermediates (NC) R-CONO have a marked dienophilic, enophilic character, this is due to the notable electrophilicity of the nitroso group, enhanced by the presence of the electron-withdrawing carbonyl group. If a

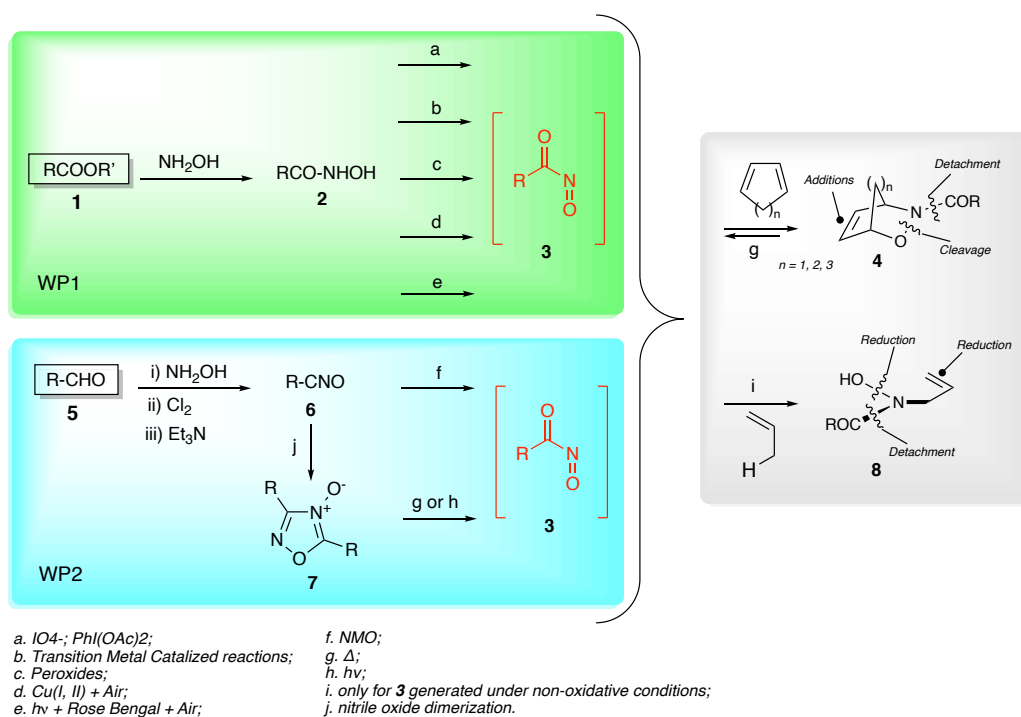
trapping agent such as a diene or olefin is not present they undergo dimerization with the formation of anhydrides and N<sub>2</sub>O.

At first, NC generated by oxidation in the presence of an amine were used to simply obtain amides. Years after their powerful dienophilic and enophilic characters were recognized, this result represented the very starting point of the chemistry of these fleeting intermediates.<sup>13</sup>

These fleeting compounds **3** can be easily obtained according to two main synthetic approaches, as shown by the WP1 and WP2 in **Scheme 1**.<sup>16</sup>

WP1 shows the use of hydroxamic acids **2**, easily prepared from esters **1**, whose oxidation is performed either with periodate salts or PhI(OAc)<sub>2</sub> (path a).<sup>17,18</sup> Other oxidative conditions were developed such as the use of transition metal catalyzed reactions and peroxides (paths b and c).<sup>19,20</sup> Aerobic oxidation with copper salts (path d) or photochemical oxidation (path e) can be alternatively employed.<sup>16,21,22</sup>

In WP2, nitrile oxides **6** are used instead. This alternative towards nitrosocarbonyls **3** stands on the mild oxidation of nitrile oxides **6** with N-methylmorpholine N-oxide (path f, NMO).<sup>23</sup> In alternative, thermolysis (path g) and photolysis (path h) of 1,2,4-oxadiazole 4-oxides **7**, the most reactive dimers of nitrile oxides **6**, can be applied.<sup>14,16,24</sup>

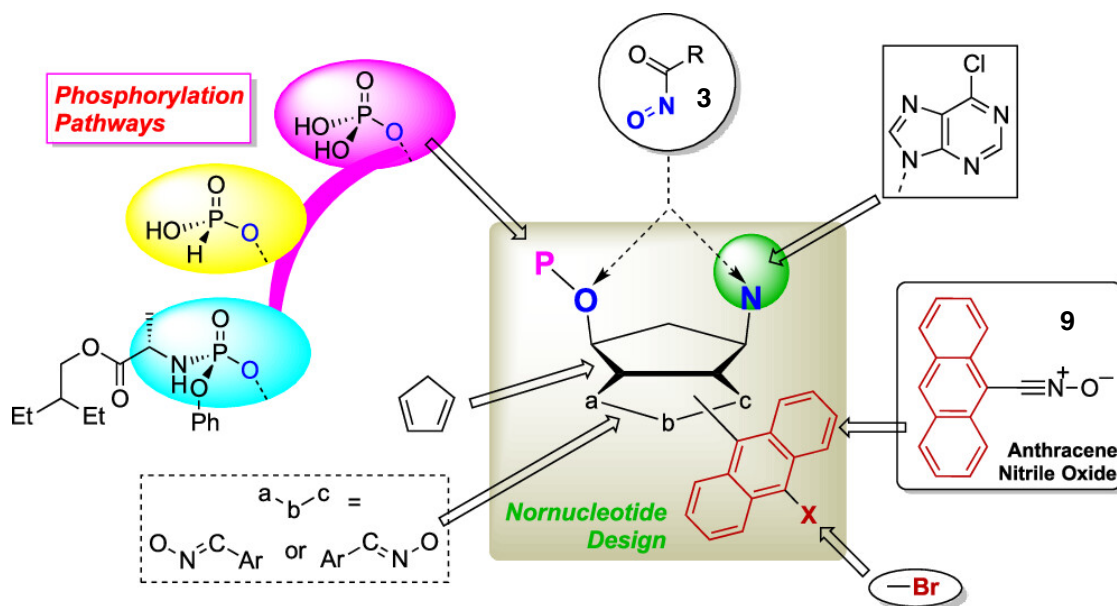


**Scheme 1.** Nitrosocarbonyl generation working packages (WPs) and reactivity in hetero Diels-Alder (HDA) and ENE reactions. [Adapted with permissions from ref 16]

Nitrosocarbonyls **3** do not survive for long and must be instantly trapped on the basis of their high dienophilic as well as enophilic power.<sup>25</sup> Several dienes and alkenes can be used to prepare highly functionalized HDA cycloadducts **4** or ene adducts **8** in very high yields (60-99%).<sup>16,26</sup>

### 1.1.3 Nitrosocarbonyl as Key Intermediates to Carbocyclic Nor-Nucleosides

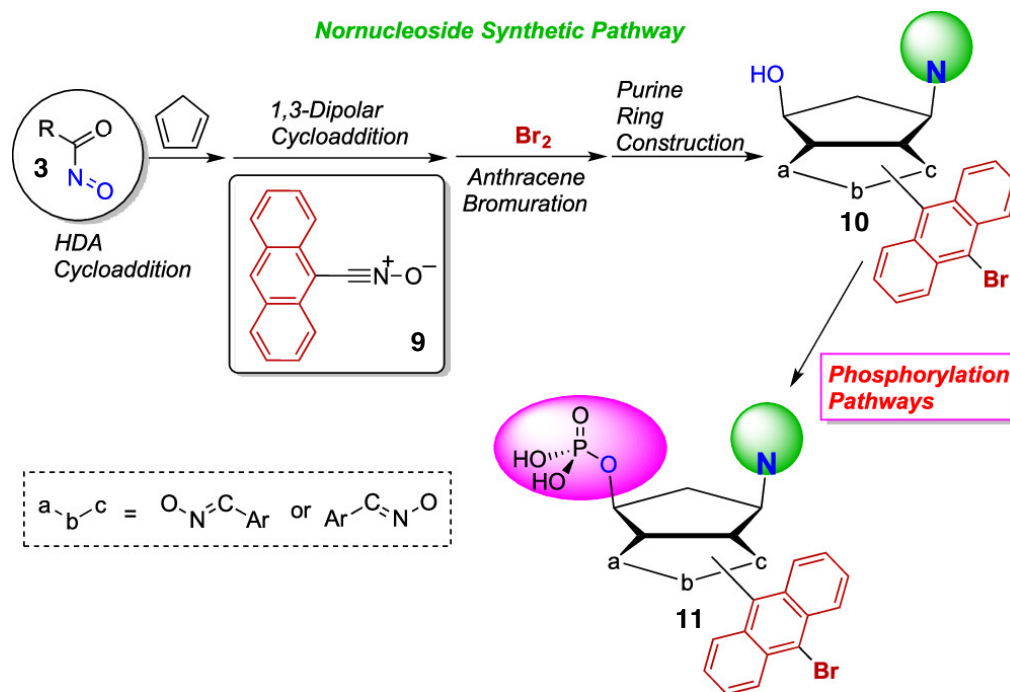
By using this powerful synthetic instrument, we designed a special type of carbocyclic nor-nucleotide in the monophosphate form to conduct a computational analysis aiming to investigate the behavior of selected compounds as possible binders for SARS- CoV-2 RdRp and to verify the relationship between molecular structures and potential activity. Modeling studies will direct/orientate the synthetic strategies and suggest eventual/consequent structural design changes. Second, biological evaluations through antiviral and cytotoxicity tests will give immediate and preliminary responses on the activity of the prepared compounds, providing helpful feedback for concentrating the efforts of modeling and synthesis toward the most promising molecules. It clearly appears the pivotal combination of these three pillars [modeling⇌chemistry⇌in vitro tests] that are highly connected and integrated and the results of any advancement in one area generate effects in the others in a continuous way, thus fulfilling the application of pericyclic reactions in this Proof of Concept (POC) devoted to contrast SARS-CoV-2. In fact, **Scheme 2** reports the setup chemical design obtained by means of those synthetic instruments which typically operate in the field of pericyclic reactions: hetero Diels–Alder (HDA) and 1,3-dipolar cycloadditions; in addition, the derivatization of aromatic substituents, the construction of specific heterobases (6-chloropurine) and the choice of the best phosphorylation method will complete the building protocol of the newly proposed antiviral candidates.



**Scheme 2.** Chemical Design of Carbocyclic Nor-Nucleotides.

The syntheses of the nor-nucleosides start from the hetero Diels–Alder (HDA) cycloaddition of suitably prepared nitrosocarbonyl intermediates and cyclopentadiene, followed by the 1,3-dipolar cycloaddition reaction of the obtained dipolarophile with the stable anthracenenitrile oxide **9** (**Scheme 3**). The derivatization of the anthracene ring with bromine conducted on the

regioisomeric cycloadducts and the linear construction of the 6-chloropurine rings will afford the nucleosides of type **10**. These are the precursors of the nor-nucleotides of type **11** that can be obtained through phosphorylation according to properly chosen protocols. Phosphorylated compounds are essential to conduct first a computational study for determining the best candidate for SARS-CoV-2 inhibition and second to perform the biological assays on viral polymerases.



**Scheme 3.** General Synthetic Strategy toward Brominated/Phosphorylated Nor-Nucleotides of Type **11**.

The study also aims to set the experimental conditions of the new synthetic steps introduced in our nucleoside protocol. In particular, the phosphorylation steps need to become a standard in our pathway. Before these latter steps, the derivatization of the anthracene moiety with bromine, a large atom, modifies the steric demand of the polycyclic system in a specific position; in addition, bromine atom offers the way for further functionalizations on the aromatic ring through simple substitution affording suitably designed more complex derivatives, if needed/suggested by the modeling.

Additionally, the choice of the 6-chloropurine ring as a heterobase allows for the valuable possibility to further functionalize the heterocyclic ring at the position 6 with replacement of the chlorine atom with several nucleophiles (e.g., RR'N, OH, RO, SH, etc.). These reactions can be conducted in mild and efficient ways, thus widening the product family available for different purposes, as the modeling will be able to indicate a better inhibition of SARS or other respiratory viruses.

## 1.2 Results: Computational Analysis: Binding Prediction

### 1.2.1 Methods

Computational experiments were conducted inside Schrodinger's Maestro v13.1.141 release 2022-1.

### 1.2.2 Modeling

Downloaded proteins from the Protein Data Bank (PDB: 7CTT; 7BV2) were prepared by using the Protein Preparation Wizard. Water was removed, hydrogens and missing side chains were added, and the states were generated using Epik at pH  $7 \pm 2$ . PROPKA at pH7 was used for H-bond assignment, and then restrained minimization, with OPLS4 force field to converge heavy atoms to Root-Mean-Square-Deviation (RMSD) of  $0.30 \text{ \AA}$ , was conducted. In the crystal structure of PDB: 7BV2, the diphosphate group and its coordinated magnesium ion were also removed. Remdesivir was modeled detached from chain P. The volume of the binding pocket was calculated using the script volume\_calc.py provided by Schrodinger.

### 1.2.3 Docking

In order to optimize the size and shape of the binding site, each ligand was initially built starting from the pre-existing crystal ligand in an iteration scheme where protein minimization steps were performed after adding each new group to accommodate it in the pocket. Finally, the structure was minimized, as described above. In the crystal structure of PDB 7CTT, cytosine residue C10 was converted to uracil prior to the minimization. Grids for docking were generated using Glide, with default settings. Ligands were built with the 3D Builder tool of Maestro and then prepared with LigPrep using force field OPLS4, and Epik was used to generate possible states at target pH  $7 \pm 2$ . Ligands were docked with standard precision in the predefined grids by using Glide. Ten poses per ligand were generated, and postdocking minimization was performed. Optimal poses were chosen based on docking score and base-pairing with the complementary base.

### 1.2.4 MD Simulations

The construct used for the DESMOND simulations of the precatalytic state consists of the 7CTT protein structure complex after removing chain D, residues 22–26 of chain T and residues 5–9 of chain Q to make it comparable to the 7BV2 construct. The system was prepared using Desmond's System Builder with the OPLS4 force field. Protein–ligand complexes or apo proteins (i.e., protein without the prostetic group) were put into an orthorhombic box, surrounded by around 27,000 water molecules and SPC solvent model, with a buffer distance of 10 Å. The volume of the system was then minimized to around  $1 \times 10^6$  Å. Negative charges in the system were neutralized by the addition of Na<sup>+</sup> ions. NaCl at a concentration of 0.15 M was also added.

Molecular dynamics trajectories were generated by using Schrodinger's Desmond. NPT ensemble was used at a temperature of 300.0 K and pressure of 1.01325 bar. Prior to the simulation, the system was relaxed with the standard relaxation protocol of Desmond with default settings resulting in around 200 ps equilibration. Nose-Hoover thermostat with relaxation time 1 ps and Martyna-Tobias-Klein barostat with relaxation time 2 ps were used. The coupling was set to isotropic. Simulations were run for 50 ns and the trajectory was recorded each 20 ps.

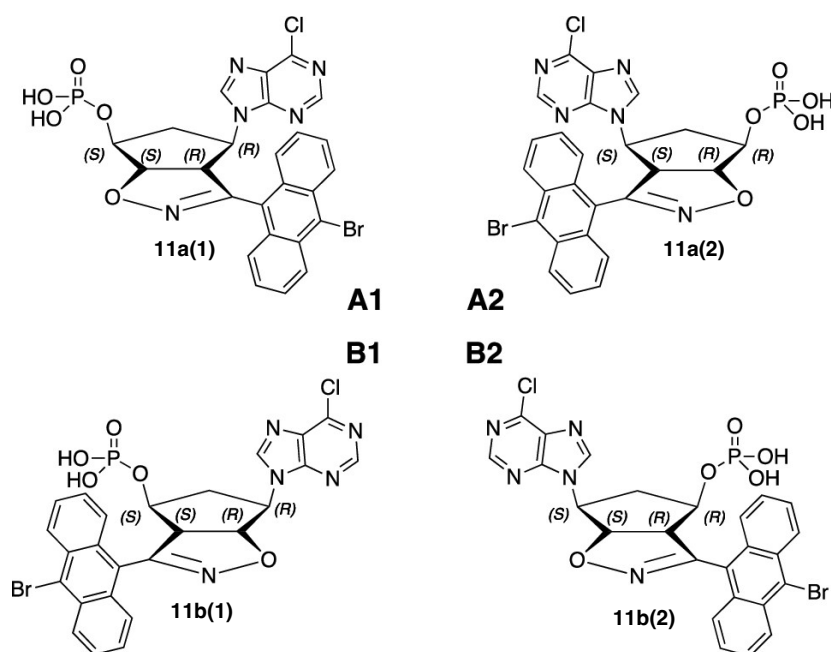
### 1.2.5 Trajectory Analysis

Interactions with the protein and RMSD were obtained through the simulation interactions diagram plugin of Maestro. Hydrogen bonds and distances were calculated using VMD, the protein was first centered in the box, and then frames were aligned. Donor–acceptor distance was set to 3.0 Å and the angle cutoff to 20° in the VMD H-Bond plugin.

Distances in Å were measured throughout the course of the trajectory from the C $\alpha$  of the residue pairs Asp761-Ala550, Asp761-Lys798 and OH3' of the elongating strand of RNA and P $\alpha$  of the ligand. Mean values are reported.

## 1.2.6 Receptor Modeling and Docking

**Figure 1** shows the regioisomeric monophosphate nucleotide structures **11a** and **11b** in their enantiomeric forms that were used for modeling and computational analyses. Since computational studies are conducted on the SARS-CoV-2 RNA polymerase, the monophosphate structures of **Figure 1** have been modified as triphosphate derivatives to fit the mechanism of action of nucleotides when interacting with polymerases. Subsequently, the same compounds were re-examined by deleting the pyrophosphate group to restore them in the monophosphate forms, aiming to verify the behavior of simpler derivatives.



**Figure 1.** Chemical Structure of Designed Compounds **11a** and **11b**, Couples of Enantiomers

The set of the four assumed compounds **11a(1)**, **11a(2)**, **11b(1)**, and **11b(2)** was examined as possible binders for SARS-CoV-2 RdRp by means of a computational strategy combining molecular docking and molecular dynamics (MD) simulations. To this end, two different inhibitor-bound high-resolution experimental structures of SARS-CoV-2 RdRp were considered as receptors. The first one is the precatalytic state (7CTT.pdb) in complex with the triphosphate form of Favipiravir before incorporation into the chain.<sup>27</sup> The second structure (7BV2.pdb) represents the Remdesivir-bound postcatalytic, pretranslocation state, where the inhibitor has been just incorporated into the primer RNA strand and therefore has lost its pyrophosphate moiety.<sup>28</sup> The overall volume of the nucleotide binding pocket is only slightly reduced by 20% in the postcatalytic one (see Modelling). The nucleotide binding pockets in the pre- and postcatalysis forms mainly differ by the positioning of several positively charged Lys and Arg residues, namely, K551 R553 R555 K621, and K798. In the precatalytic structure 7CTT those

residues are oriented to coordinate the negatively charged triphosphate group: despite being limited by the low density map, it is hypothesized that residues R553 and R555 might interact with the  $\beta$ - and  $\gamma$ -phosphate groups of Favipiravir triphosphate, while the terminal  $\gamma$ -phosphate is involved in a salt bridge contributed by K798. These interactions together are hypothesized to stabilize the incoming nucleotide and 3' terminus of primer RNA in close vicinity to facilitate the nucleophilic attack.

Hence, this structure allows us to investigate the recognition of the triphosphate form of the designed compounds and generate hypotheses about the possibility of incorporating them into the chain. In contrast, in the postcatalytic structure 7BV2, which still contains the pyrophosphate after the incorporation reaction, the positively charged residues are far away from the pyrophosphate as its negative charge is partially neutralized by the proximity of the second  $Mg^{2+}$  ion.

This structure, after modeling the detachment of the nucleotide analogue ligand from the RNA chain and deleting the pyrophosphate group, is used to test by molecular docking the binding of the monophosphate form of each compound in parallel to targeting 7CTT. Also, the docking of the monophosphate form of each compound in the postcatalytic structure can cross validate the binding pose found for its triphosphate form, under the hypothesis that chain incorporation might take place.

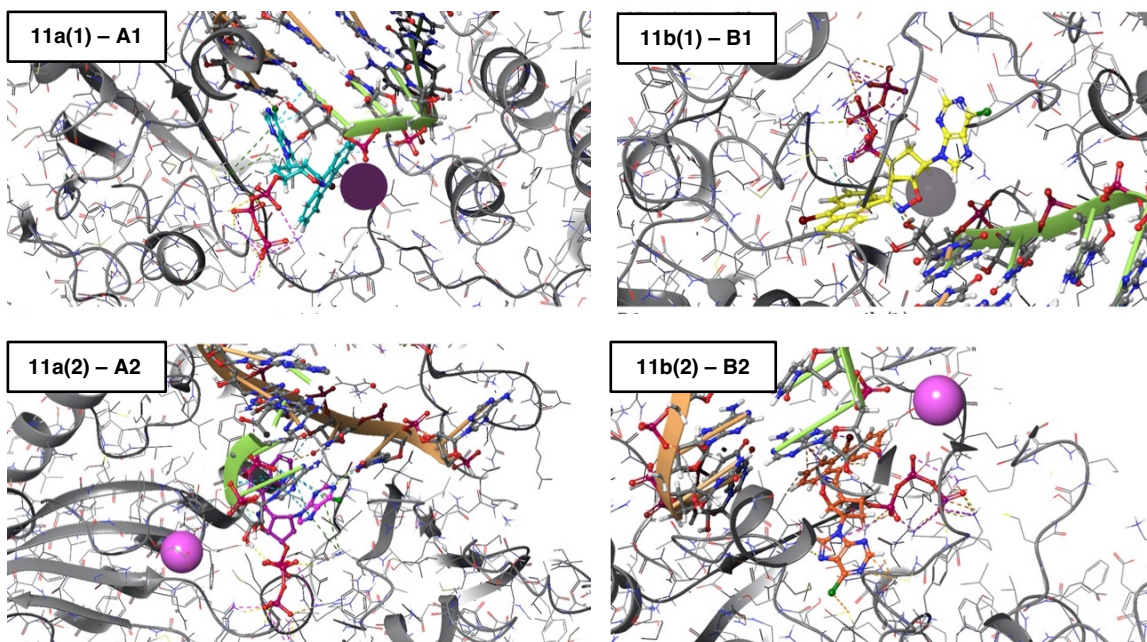
Each one of the four ligands was docked to each one of the structures, either in triphosphate (to 7CTT) or in monophosphate form (to 7CTT and 7BV2). To be able to pair Remdesivir starting from the Favipiravir bound structure, a mutant C  $\rightarrow$  U was modeled in the complementary base of template RNA of 7CTT. The matrix of receptor–ligand pair combinations was used to generate a set of complex models, which are filtered in terms of docking score and pose.

Due to the bulky character of the anthracene moiety, for each ligand, the binding site was first relaxed by gradually growing the compound inside the pocket and correspondingly minimizing the local environment, and then, once equilibrated, the compound was removed and rigid docking was performed (see Docking).

Docking poses were first evaluated by comparing the normalized binding energy (or ligand efficiency) to the native ligands Favipiravir or Remdesivir and then by inspecting the presence of base pairing between the 6-chloropurine moiety and the complementary base. The correct positioning of the triphosphate is also considered an evaluation criterion.

**Figure 3** shows the docking poses of the enantiomeric couples of compounds **11a** and **11b** in their triphosphate forms. The results of the docking experiment reveal that **11a(2)** (**Figure 3, A2**) docked to 7CTT shows the best arrangement in terms of both geometries, when compared to the reference ligand Favipiravir, and ligand efficiency. **Table S1** in the Appendix reports the ligand efficiencies of selected binding poses of the triphosphate derivatives in the precatalytic structure.





**Figure 3.** Docking poses (**A1,2, B1,2**) of enantiomeric compounds **11a** and **11b** in the triphosphate forms in the precatalytic structure.

In **11a(2)**, the anthracene moiety interacts through stacking with the n-1 base, while the 6-chloropurine moiety is engaged in base pairing to the template RNA (**Figure S1** in the Appendix).

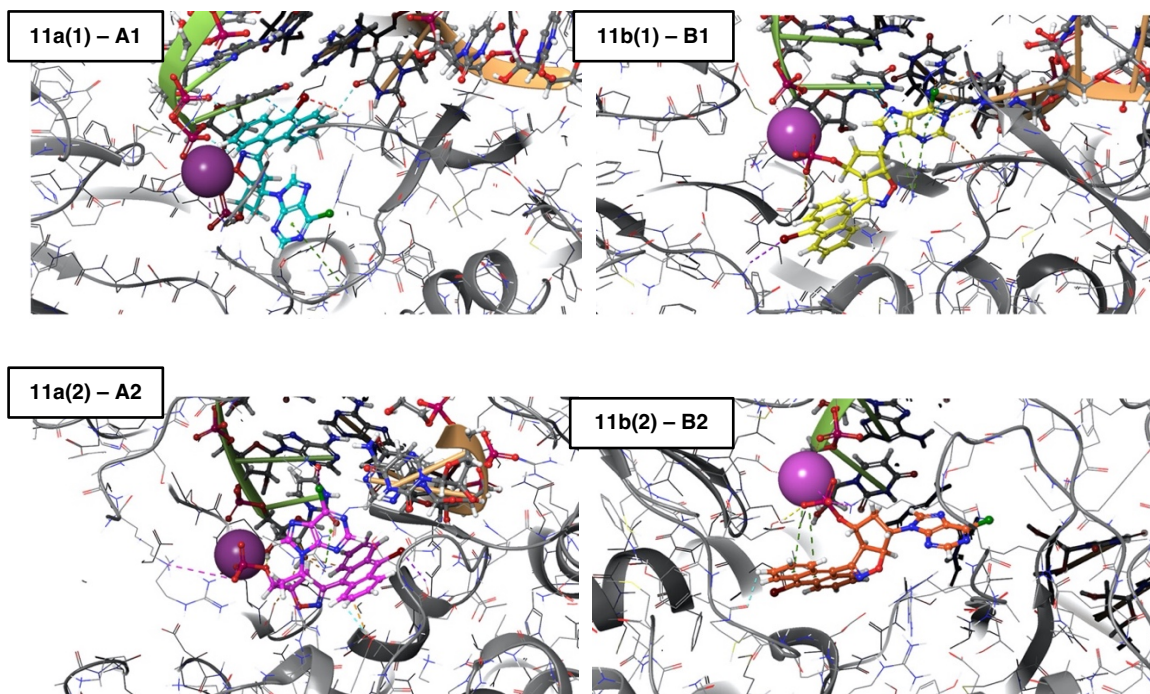
In contrast, for compound **11a(1)** (**Figure 3, A1**) no pose was found, showing a correct base pairing in either receptor structure. Also, no  $\pi$ -stacking interactions are formed by the anthracene group, which cannot be accommodated in the pocket (see **Figure S1** in the Appendix). This compound was therefore discarded in the following analysis.

Compounds **11b(1)** and **11b(2)** (**Figure 3, B1** and **B2**) show different arrangements in their triphosphate form docked to target 7CTT. On the one hand, **11b(2)** keeps the anthracene group in a stacking interaction with base n-1 but abolishes the base pairing by the 6-chloropurine moiety. Conversely, triphosphate **11b(1)** does not establish any stabilizing interactions apart from the phosphate group (see also **Figure S1** in the Appendix).

**Figure 4** shows the selected binding poses of enantiomeric couples **11a** and **11b** in the monophosphate forms docked to 7BV2 (see also **Figure S2** in the Appendix).

In pose **A1** of **Figure 4**, the anthracene ring in compound **11a(1)** is located erratically in the position normally occupied by the heterobase, presenting a  $\pi$ -stacking with base n-1; the phosphate group is coordinated to the  $Mg^{2+}$  ion, also coordinated by the oxygen atom belonging to the isoxazoline ring. The 6-chloropurine ring of **11a(1)** is orientated toward the loop where the Ala550 is located.

Analogously, **11a(2)** in the monophosphate form behaves as its corresponding triphosphate, with the anthracene ring positioned under the elongating RNA chain while the 6-chloropurine ring is correctly located, although in the absence of base-pairing. Again, the monophosphate group is coordinated to the  $Mg^{2+}$  ion.



**Figure 4.** Docking poses (**A1,2, B1,2**) of enantiomeric compounds **11a** and **11b** in the monophosphate forms in the postcatalytic structure.

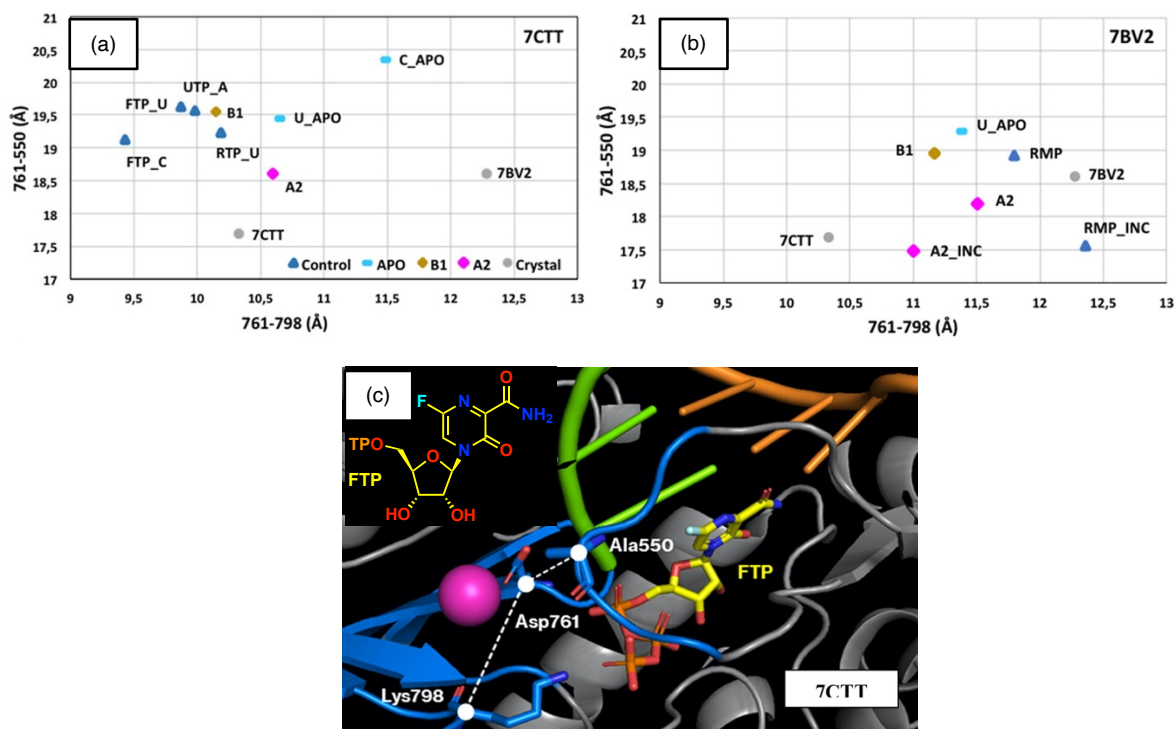
Interestingly though, the monophosphate form of **11b(1)** docked to 7BV2 shows coordination of the monophosphate to the catalytic  $Mg^{2+}$  ion, correct base pairing, and  $\pi$ -stacking with base n-1 while the anthracene moiety fills the now available triphosphate subpocket (**Figure 4, B1**). This suggests a possible binding mode of this compound that would be compatible with a non-nucleosidic reverse transcriptase inhibitors (NNRTI) mechanism.

Finally, in compound **11b(2)** (**Figure 2, B2**) the 6-chloropurine ring is completely out of the pocket; there is a void space where the base should be positioned, but the base misses the target. Furthermore, the anthracene is somewhat hung out in an empty zone thinly connected to the  $Mg^{2+}$  ion. On the basis of the computational results reported above, the structures of compounds **11a(2)** and **11b(1)** will be considered in terms of molecular dynamic studies.

## 1.2.7 Molecular Dynamics

Triphosphate and monophosphate forms of compounds **11a(2)** and **11b(1)** in complex with RdRp in the precatalytic and the postcatalytic states were further analyzed by means of Molecular Dynamics. To assess possible local structural rearrangements in the protein, we focused on the closure of the binding pocket by measuring the average distances between the C $\alpha$  atoms of selected residue pairs along the dynamics.

Complexes with **11b(1)** and **11a(2)** were compared to Favipiravir and Remdesivir complexes and the apo form of RdRp and subjected to control MD simulations. Specifically, the distances between Asp761 and Lys798, residues of a  $\beta$ -sheet responsible for the coordination of the Mg<sup>2+</sup> ions and the triphosphate group of the incoming nucleotide, and the distance between Asp761 and Ala550, residues contained in the so-called finger of RdRp, were measured (**Figure 5**, Scatter plot).



**Figure 5.** Closure of the binding pocket observed in the MD simulations. Scatter plot of the average distances between Asp 671 and Lys 798 and the distance between Asp 761 and Ala 550 in (a) complexes obtained with the 7CTT structure and (b) 7BV2 (refer to **Figure 1** for structures). The label 'Control' indicate average for MD simulations of the X-ray complex structures of Remdesivir and/or Favipiravir bound RdRp. These are (a) FTP bound complex with C or U as complementary base; RTP bound complex with U as complementary base. (b) RMP bound complex; RMP bound complex after incorporation into the chain. The pdb code label indicates crystal structure values as further reference. (c) RdRp in complex with FTP in 7CTT.

MD trajectories (50 ns) were run, and a few parameters were evaluated. Specifically, the RMSD of the ligand was calculated, and the persistence of base pairing was analyzed focusing on Watson–Crick base pairs hydrogen bonding patterns to complementary RNA. For the triphosphate forms complexed to 7CTT the angles are **11a(2)** 55.2° and **11b(1)** 107.8°, respectively. In Favipiravir triphosphate (FTP) the angle is 135.4° between OH3', the phosphate P $\alpha$  and the pyrophosphate group along dynamics and can suggest how well the nucleotide analogue is positioned for the nucleophilic attack of the OH3' moiety. **11b(1)** is closer to the reference values for FTP, which suggests a better positioning.

In the 7CTT scatter plot (**Figure 5a**) the pose **B1** relative to compound **11b(1)** is located in the control zone, while pose **A2** relative to compound **11a(2)** shows a Asp761–Lys798 distance >1 Å. In 7BV2 both compounds are located in the control group (**Figure 5b**).

The results are summarized in **Table S2** in the Appendix. While **11a(2)** shows a low RMSD (average 1.33 Å) and a reasonably low mean distance between OH3' and the phosphate (4.7 Å), **11b(1)** shows increased mobility, higher RMSD (3.78 Å) and the phosphate group is displaced away from the position required for catalysis (8.8 Å distance OH3'–P $\alpha$ ) although, similarly to **11a(2)**, it retains contact to the positively charged residues that coordinate the phosphate group in the Favipiravir crystal structure K551 R553 R555 K621 and K798 (see **Figure 5c**).

The triphosphate forms of **11a(2)** and **11b(1)** ligands cause the binding pocket parameters to assume the same range of values that are found in the presence of the reference ligands Remdesivir and Favipiravir, which determine a closure of the pocket when compared to the crystal structure values. In contrast, the apo state of both 7CTT variants with uracile and cytosine leads to a wider pocket.

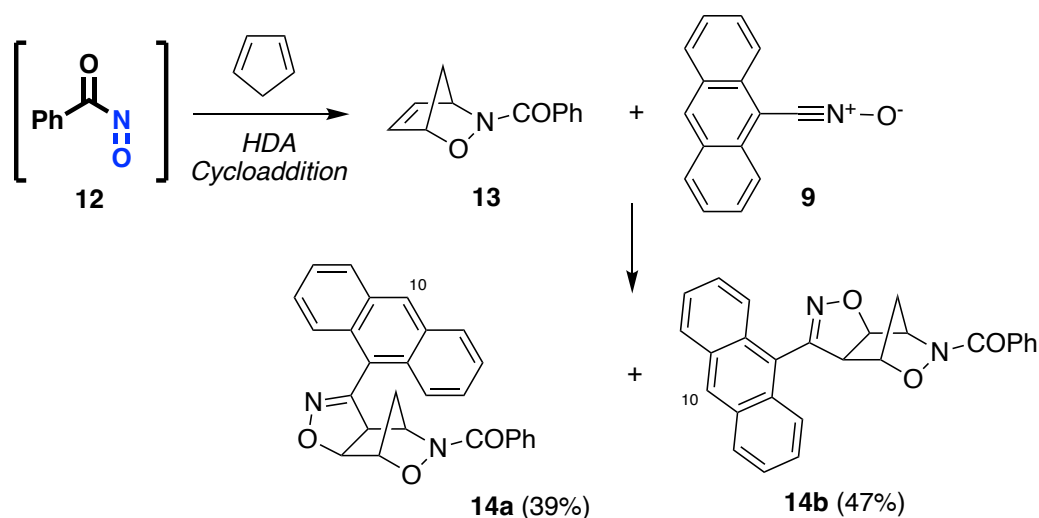
The monophosphate forms were simulated in complex with the postcatalytic structure 7BV2. The same parameters were evaluated here to predict the stability of the poses (see **Table S2** in the Appendix). Results show that monophosphates **11a(2)** and **11b(1)** confirm what was found in the docking experiments. Namely, while **11b(1)**, which had a better ligand efficiency when docked to 7BV2, stably maintains base pairing and an overall low RMSD, **11a(2)** loses its pairing, reflecting the less favorable binding pose. It is worth mentioning that monophosphate **11a(2)** docked to the precatalytic structure despite the absence of the triphosphate group is able to maintain favorable interactions with the RNA showing a ligand efficiency of –0.317, comparable to the best binders of the target (see Appendix). On the other hand, the pocket distances (**Figure 5**) reveal a conformational response of the protein in line with the control ligands monophosphate Remdesivir and Favipiravir, similarly to what is observed in the triphosphate form, resulting in a closer binding pocket than the apo state, as expected.

Overall, this analysis predicts that among the pool of designed compounds, **11a(2)** and **11b(1)** are the ones that could, with different binding modes, productively interact with RdRp. Specifically, for the first one, we expect that both the triphosphate and the monophosphate forms might be active, while for the second one, the monophosphate form shows the best

results, due to the occupation of the triphosphate subpocket by the anthracene group. With this ranking in mind, we proceeded with the synthesis of the monophosphate compounds **11a** and **11b** and in vitro testing as a preliminary evaluation of the biological activities in view of further implementation of the synthetic process.

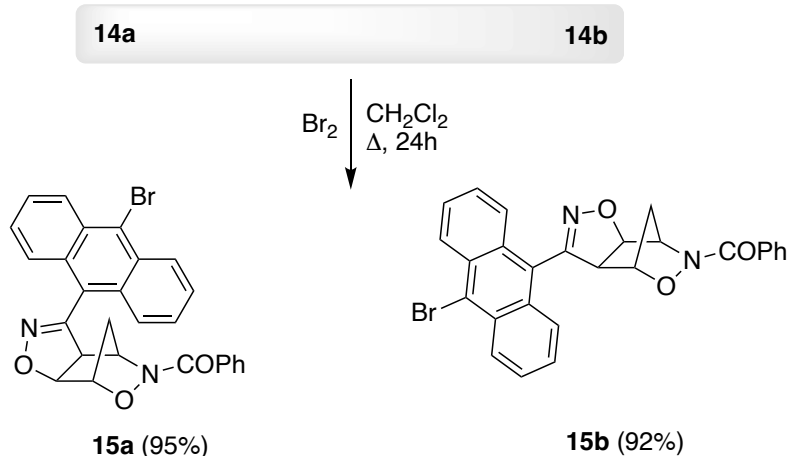
### 1.2.8 Chemistry

Path *a* (see **Scheme 1**) was the route of choice to afford the key intermediate benzo-nitrosocarbonyl **12** (**Scheme 4**) then reacted with cyclopentadiene; the starting regioisomeric compounds **14a,b** were prepared, according to the published procedure, from the *N*-benzoyl-2,3-oxazanoborn-5-ene **13** as dipolarophile in the presence of a slight excess of anthracenenitrile oxide **9** and isolated in nearly 1:1 ratio by column chromatography.<sup>29</sup>



**Scheme 4.** Synthesis of cycloadducts **14a,b**.

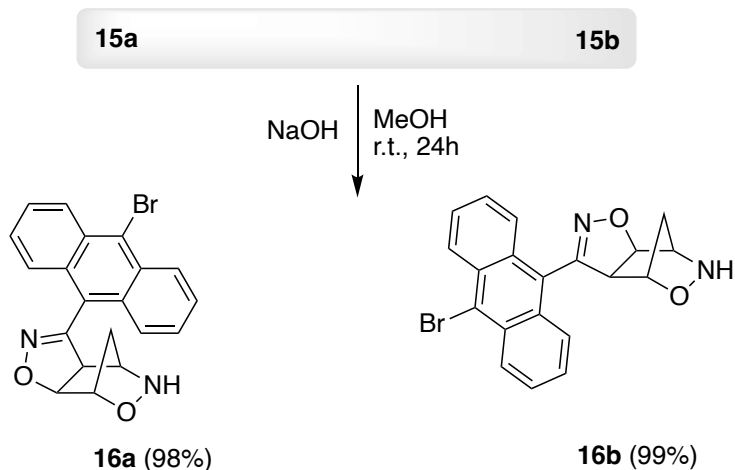
The bromination protocol was adapted from the literature procedures and the regioisomeric cycloadducts **15a,b** (**Scheme 5**) were obtained as bromo-derivatives upon treatment with bromine under mild conditions (DCM, at reflux for 24 h).<sup>30</sup>



**Scheme 5.** Bromination of cycloadducts **14a,b**.

The following synthetic steps run again through the classical pathway required for the linear construction of purine heterobases as we have shown in previous works with few adaptations and experimental improvements.<sup>13</sup>

Alkaline hydrolysis of the cycloadducts **15a,b**, conducted with grinded NaOH in methanol solutions, afforded the corresponding hydroxylamine derivatives in nearly quantitative yield (**Scheme 6**).<sup>13,31</sup>

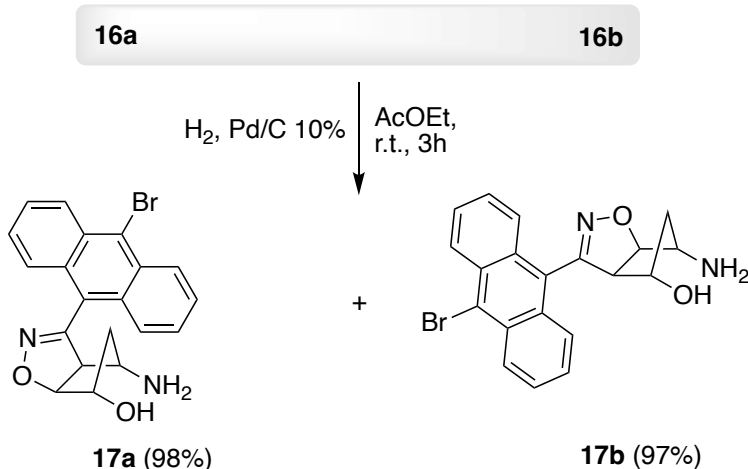


**Scheme 6.** Hydrolysis of cycloadducts **15a,b**.

The <sup>1</sup>H NMR spectra (DMSO-d<sub>6</sub>) show the absence of the 5 proton signals corresponding to the benzoyl group, and the presence of signals at  $\delta$  6.61 ppm and  $\delta$  3.71 ppm, respectively for **16a** and **16b**, corresponding to the NH protons, further attesting to the benzoyl group detachment.

Hydrogenolysis of the hydroxylamine derivatives under standard conditions (H<sub>2</sub>, Pd/C 10%, AcOEt) afforded the desired aminols **17a,b** (yields: 98 and 97%, respectively, **Scheme 7**).<sup>13,31</sup>

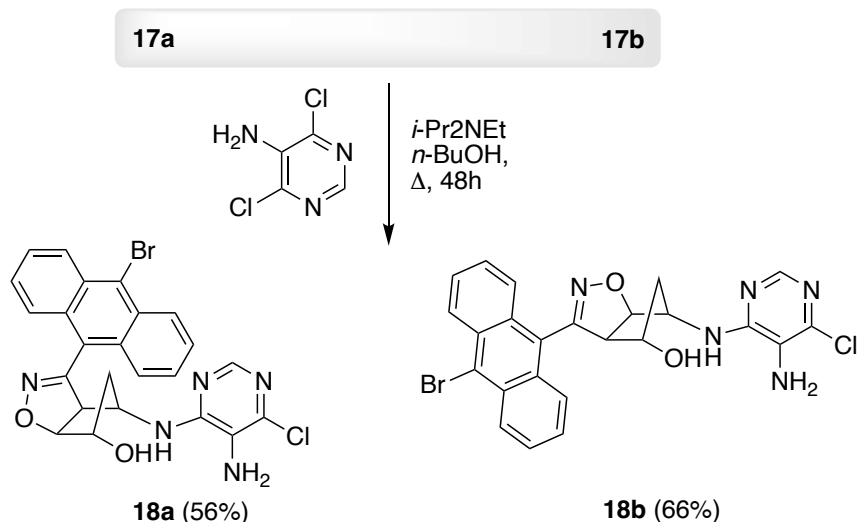




**Scheme 7.** Hydrogenolysis of the hydroxylamine derivatives **16a,b**.

The reaction time must be carefully contained within three hours because a prolonged hydrogenation process was detrimental to the stability of the anthracene moiety and a plethora of degradation products was observed in the reaction mixtures. The structures of **17a,b** were confirmed from their analytical and spectroscopic data. In particular, in the  $^1\text{H}$  NMR spectra the NH signals had disappeared and the spectrum displayed a more complex fine structure with couplings due to the increased conformational mobility of the cyclopentane moieties following strain relief after the hydrogenolytic ring opening. Significantly, the  $\text{NH}_2$  and OH groups (linked to each other by H-bonding) furnished signals, respectively, at  $\delta$  5.94 ppm and  $\delta$  8.05 ppm for compound **17a** while for compound **17b** the corresponding groups gave signals at  $\delta$  3.32 ppm and  $\delta$  8.03 ppm.

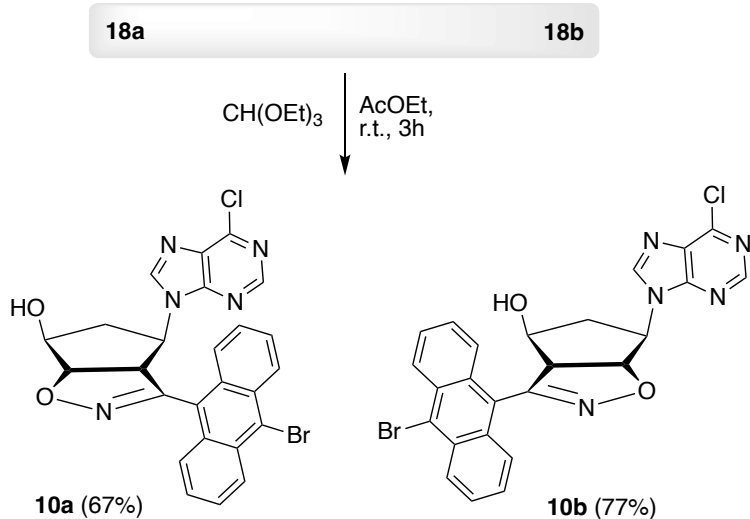
By using known procedures for the construction of the purine nucleus, the aminols **17a,b** were converted into the 6-chloropurine derivatives **10a,b** by substitution of 5-amino-4,6-dichloropyrimidine followed by acid-catalyzed condensation with ethyl orthoformate (**Scheme 8,9**).<sup>32,33</sup> The pyrimidine derivatives **18a,b** were obtained in good yields (**18a**, 56%; **18b**, 66%) in the most delicate synthetic step of the entire pathway, by heating solutions of the aminols **17a,b** and 5-amino-4,6-dichloropyrimidine (2 equiv.) in *n*-BuOH (bp 117 °C) in sealed flasks in the presence of an excess of *i*-Pr<sub>2</sub>NEt (5 equiv.) for 48 h. The use of sealed flasks (Schlenk tubes) for these synthetic steps was essential for the success of the reactions since the sealed reactors create a moderate overpressure; control experiments conducted at room pressure resulted in failure of the processes.



**Scheme 8.** Synthesis of the pyrimidine derivatives **18a,b**.

The structures of **18a,b** rely upon their analytical and spectroscopic data. The IR spectra of pyrimidines **18a,b** exhibit a complex series of bands between 3180-3370  $\text{cm}^{-1}$  due to the presence of OH, NH and  $\text{NH}_2$  groups and the  $^1\text{H-NMR}$  spectra ( $\text{DMSO-d}_6$ ) were unambiguously consistent for the assigned structures where the typical signals corresponding to the pyrimidine moieties ( $\text{CH=N}$  group) can be found linked to those belonging to the aminol structures.

The conversion of the stereoisomeric pyrimidines **18a,b** into the chloropurines **10a,b** was finally achieved in excellent yield (87% for both the products) by treatment with triethyl orthoformate in the presence of catalytic *p*-TsOH, maintaining the reaction mixtures at r.t. for six days and in accordance with a previously reported protocol.<sup>32</sup> Isolation and purification of **10a,b** were secured by hydrolysis of triethyl orthoformate, evaporation of the organic solvent and extraction of the water solution with dichloromethane, followed by crystallization of the final products.



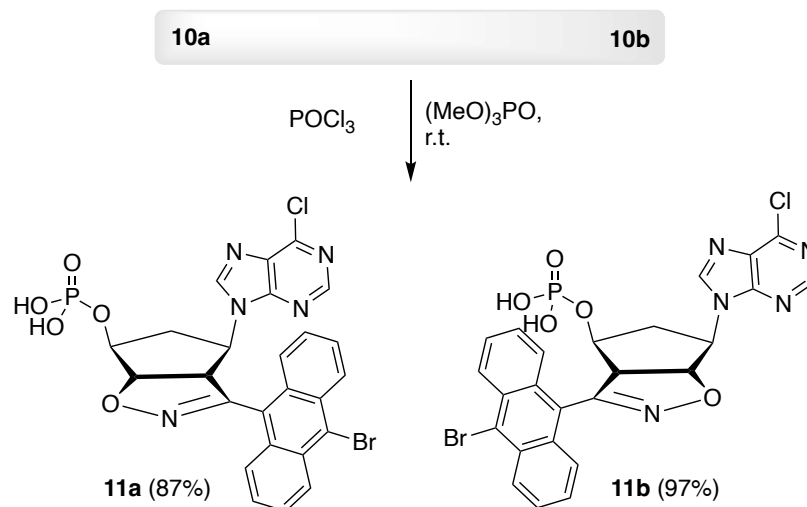
**Scheme 9.** Construction of the 6-chloropurine derivatives **10a,b**.



The chloropurines **10a,b** were fully characterized. The infrared spectra of **10a,b** showed a single broad band at  $3328\text{ cm}^{-1}$  (**10a**) and  $3413\text{ cm}^{-1}$  (**10b**) corresponding to the OH absorptions. In the  $^1\text{H-NMR}$  spectra (DMSO- $d_6$ ) the two N=CH protons of the purine rings occur as singlets at  $\delta$  8.57 and 8.59 ppm for **10a** and at  $\delta$  8.88 and 8.90 ppm for **10b**, these signals being diagnostic of the presence of a purine ring.

Finally, the conversion of the nor-nucleosides **10a,b** into the corresponding monophosphate derivatives **11a,b** has been selected among the proposed methods shown in **Scheme 2** since these structures represent valuable starting compounds for the synthesis of the relative triphosphate compounds; furthermore, tests are also available for conducting biological evaluation also in the monophosphate derivatives as preliminary evaluation. Hence, we setup the methodology for converting the nucleosides **10a,b** into the regioisomeric monophosphates **11a,b** by adapting the known methods reported in literature.<sup>34,35</sup>

Treatment of the nucleosides **10a,b** with excess  $\text{POCl}_3$  (3 equiv) in an ice-cooled solution of trimethyl phosphate afforded the desired products **11a,b** in nearly quantitative yield after an overnight reaction at room temperature (**Scheme 10**).



**Scheme 10.** Syntheses of Regioisomeric Nor-Nucleotides of Type **11a,b**.

The obtained products were fully characterized spectroscopically. In the  $^1\text{H NMR}$  (DMSO- $d_6$ ), the structures of the nucleosides are confirmed upon comparison with the spectra of the starting material. Significantly, the absence of the signal relative to the OH group indicates functionalization with the phosphorus. The  $^{31}\text{P NMR}$  spectrum (DMSO- $d_6$ ) revealed the presence of the phosphate group as a singlet at  $\delta$   $-0.10$  ppm for compound **11a** and at  $\delta$   $-2.10$  ppm for compound **11b**, in full accordance with the literature on phosphorus derivatives.<sup>36</sup>

### 1.2.9 Biological Assays

Samples of the synthesized monophosphate compounds **11a** and **11b**, representative of the regioisomeric families (A) and (B) (see **Figure 1**), were biologically assayed to preliminarily assess their potential as antiviral compounds capable of inhibiting SARS-CoV-2 RdRp activity.

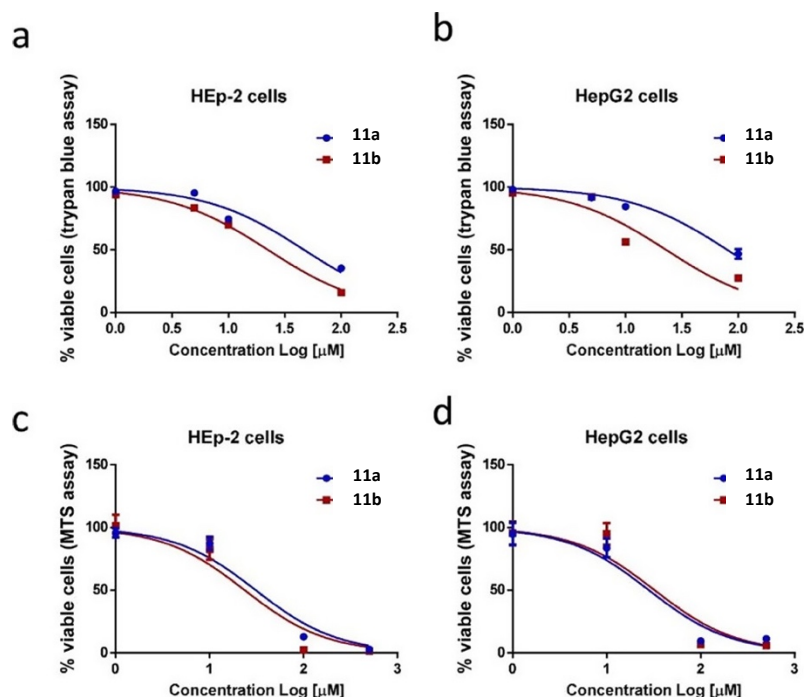
For this purpose, first their cytotoxicity was assayed in terms of inhibition of cell viability and of cell metabolic activity in epithelial HEP-2 and HepG2 cells. Cell viability and death were evaluated by the classical trypan blue assay. The cell metabolic activity was detected through formazan product formation, using a commercial colorimetric kit (MTS [3,4-(5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium salt]). Concerning the cell viability, a dose effect assay using three different concentrations of the compounds, within the range from 100 to 0.1  $\mu\text{M}$ , assessed levels of cytotoxicity, expressed as  $\text{CC}_{50}$ .

Results showed that the **11b** was endowed with a higher ability to induce cell death with respect to **11a** toward both cell lines tested, with  $\text{CC}_{50}$  values of about 22  $\mu\text{M}$  for HEP-2 and HepG2 cells (**Figure 6a,b**). Conversely, **11a** and **11b** compounds, assayed within the range from 500 to 1  $\mu\text{M}$ , showed a quite similar ability in inhibiting the metabolic activity, with **11b** being more effective toward HEP-2 cells, but not toward HepG2 cells, with respect to **11a** (**Figure 6c,d**). In any case, even in this assay, the lower  $\text{CC}_{50}$  value obtained was about 23  $\mu\text{M}$  for **11b** toward HEP-2 cells (**Figure 6c,d**).

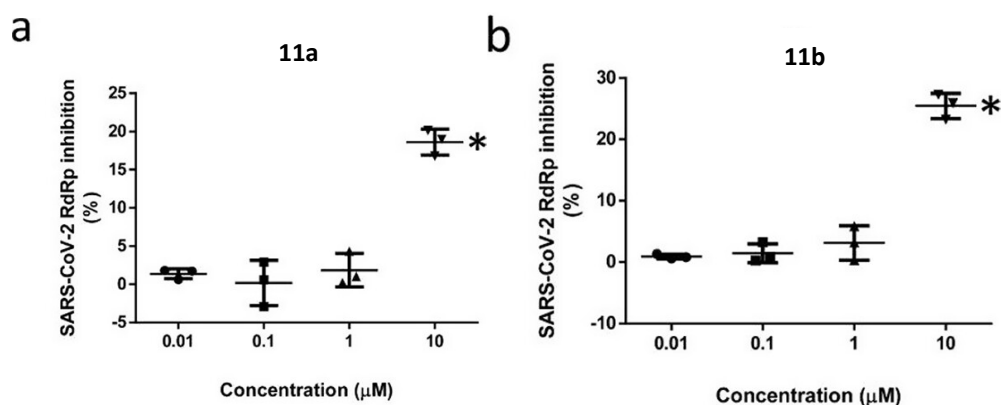
Thus, the comparative analysis of the cell viability and of the cell metabolic activity inhibition indicated that the two compounds tested, in their current structure, were endowed with a certain cytotoxicity at concentrations higher than 20  $\mu\text{M}$  (**Figure 6**).

We then focused our attention on the capacity of the compounds to inhibit the SARS-CoV-2 RdRp activity in the range from 10 to 0.01  $\mu\text{M}$ . For this purpose, a commercial, virus-cell-free assay based on detection of the RNA molecules synthesized by the SARS-CoV-2 recombinant RNA polymerase complex in the presence of an RNA template and NTPs, was utilized. The results showed that **11a** and **11b** were actually capable of significantly reducing RNA synthesis by part SARS-CoV-2 RdRp by about 18 and 25%, respectively, but only at the higher concentration assayed of 10  $\mu\text{M}$  (**Figure 7a,b**).

Overall, the results of the biological assays experimentally confirmed what was predicted by the computational analysis about the capability of the newly synthesized nor-nucleotides to efficiently interact with the functional activity of SARS-CoV-2 RdRp, making **11b** more interesting for future design of structural modifications. On the other hand, the results indicated that the currently designed structure of the compounds is also endowed with a quite high cytotoxic potential. These drawbacks must be taken into account in relation to the use of isoxazoline-carbocyclic synthetic protocol as the strategic way to select the candidates for the SARS-CoV-2 polymerase inhibition.



**Figure 6.** Inhibition of cell viability assessed by the trypan blue assay (a, b) and of metabolic activity assessed by MTS assay (c, d) in HEp-2 cells (a, c) and HepG2 cells (b, d) treated with nucleotides **11a** and **11b** for 24 h (experiments in triplicate). Cytotoxicity, calculated as concentration of compounds needed to reduce the viability or metabolic activity by 50% with respect to control cells.  $\text{CC}_{50}$  values (expressed as mean  $\pm$  SD: (a) **11a** =  $46.79 \pm 1.18$ , **11b** =  $22.36 \pm 1.04$ ; (b) **11a** =  $79.67 \pm 1.10$ , **11b** =  $22.85 \pm 1.19$ ; (c) **11a** =  $30.96 \pm 1.2$ , **11b** =  $23.64 \pm 1.3$ ; (d) **11a** =  $28.14 \pm 1.3$ , **11b** =  $32.37 \pm 1.4$ ).



**Figure 7.** Inhibition of SARS-CoV-2 RdRp activity by nucleotides (a) **11a** and (b) **11b**, assessed by a cell/virus-free fluorometric RdRp activity assay. Data presented as the mean  $\pm$  SD. Statistical significance analyzed using the 1 way ANOVA Bonferroni's multiple comparisons test: \* $p < 0.001$  vs all (experiments in triplicate). Inhibitory activity was calculated as the concentration of compounds needed to reduce RdRp activity by 50%.  $\text{IC}_{50}$  values, expressed as mean  $\pm$  SD, were: **11a** =  $43.83 \pm 1.07$ ; **11b** =  $29.31 \pm 1.05$ .

## 1.3 Discussion

In this POC study, we have synthesized the regioisomeric nucleosides **10a,b** obtained through the chemistry of nitrosocarbonyl intermediates of type **3**, taking advantage of the use of the stable anthracenenitrile oxide **9**. The synthetic pathway relies upon a well-established protocol, setup in our laboratories and well-documented in several papers and reviews.<sup>12,13</sup> These methodologies are potentially suitable for introducing the phosphorus moiety needed for the biological tests on SARS-CoV-2 polymerases.

We also considered various phosphorylation methods to be applied to functionalize compounds **10a,b**. Among the phosphorylation methods, we set up the experimental conditions to a fast track to monophosphate derivatives through POCl<sub>3</sub> treatment, which is the preliminary step for eventual triphosphorylation. The phosphorylation processes gave valuable and reliable results and are pivotal in our strategy oriented to the selection of the best candidate as antiviral compound and specifically dedicated to the contrast of the SARS-CoV-2 virus or other viral emerging infectious diseases. Regarding the SARS topic, a first evaluation has been performed through docking computational studies. Although the nature of nucleoside/nucleotides analogs, they resemble NNRTI that have not yet been found/synthesized to contrast SARS-CoV-2. In principle, they could recognize allosteric NSP12 sites and interact with sites other than those occupied by Remdesivir. NNRTI are considered potential targets by those molecules inducing little cell death since they should be antiviral and not antitumor candidates.<sup>37</sup> The computational studies revealed the possibility for compounds of types **11a** and **11b** to give positive responses in terms of contrast to SARS-CoV-2, with compounds of type **11b** being the most promising.

The biological results showed that **11a** and **11b** were capable of significantly reducing RNA synthesis by part SARS-CoV-2 RdRp by about 18 and 25%, respectively. One of the main problems concerns the concentration required to obtain these performances (10 μM). Overall, the results of the biological assays confirmed what predicted by the computational analysis. Nevertheless, the results indicated that the currently designed structures of monophosphate compounds are endowed with a quite high cytotoxicity.

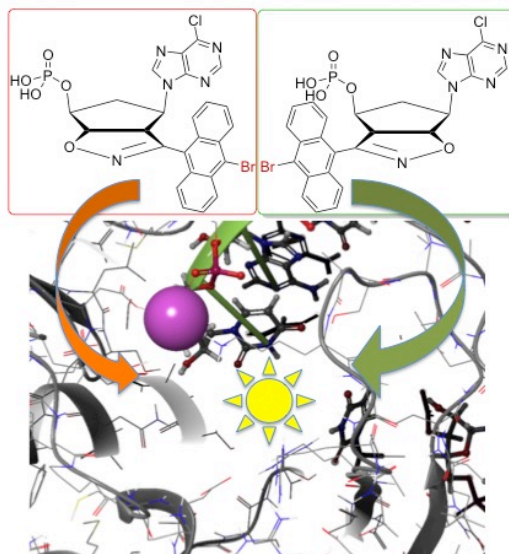
## 1.4 Conclusions

Prominent in the current stage of drug development, antiviral compounds can be efficiently prepared through cycloaddition reactions. In recent reviews we reported the use of 1,3-dipolar cycloaddition reactions of selected 1,3-dipoles, from azides to nitrones, passing through nitrile oxides and other ylides, in the light of their application for the preparation of key intermediates in the design and synthesis of compounds that were tested for their antiviral activities against a variety of viruses.<sup>38,39</sup> The products obtained from these pericyclic reaction approaches were tested for their activities in terms of blocking the virus replication, and the relevant biological data are highlighted.

The ability of the 1,3-dipolar cycloaddition reactions and the hetero Diels–Alder cycloadditions of nitrosocarbonyl intermediates (easily obtained from nitrile oxides through mild oxidation) to produce a variety of heterocyclic scaffolds extends its importance to many areas of organic synthesis, and the synthesis of antiviral compounds represents a topic in constant growing. The use of 1,3-dipoles offers a unique method for the preparation of a variety of compounds with potential biological activity. It must be taken into account that heteroatom-containing cycloadducts may be transformed in other functionalized molecules, cyclic or acyclic. Furthermore, the ability of the 1,3-dipole to generate rings even containing stereocenters in a single synthetic operation is a pivotal step when a valuable compound has to be prepared.

In conclusion, these are surely precious synthetic methods that can be applied for the preparation of antiviral candidates targeting respiratory viruses, SARS-CoV-2 included. The computational screening is also a pivotal step in selecting the most relevant structures, but the biological assays will definitively direct the synthesis toward the desired goal.

The chemistry of pericyclic reactions is still at work!



## 1.5 Experimental Section

### 1.5.1 General

All melting points (mp) are uncorrected.  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{31}\text{P}$  NMR spectra were recorded on a 300 MHz spectrometer (solvents specified). Chemical shifts are expressed in ppm from internal tetramethylsilane ( $\delta$ ), and coupling constants (J) are in Hertz (Hz): b, broad; s, singlet; bs, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet. IR spectra (nujol mulls) were recorded on a spectrophotometer available at the Department and absorptions ( $\nu$  are in  $\text{cm}^{-1}$ ). HRMS were done on a X500B QTOF system (Sciex, Framingham, MA 01701 USA) available at the CGS of the University of Pavia. Column chromatography and tlc: silica gel H60 and GF254, respectively; eluants: cyclohexane/ethyl acetate 9:1 to pure ethyl acetate.

### 1.5.2 Starting and Reference Materials

Compounds **15a,b** were synthesized according to the procedure reported in the literature.<sup>29,31,32</sup> Other reagents and solvents were purchased from chemical suppliers and used without any further purification following or adapting the methods reported in the cited literature.

### 1.5.3 Synthesis of compounds 15a,b through Bromination.

Compounds **14a,b** (1.5 g, 3.57 mmol) were dissolved in anhydrous  $\text{CH}_2\text{Cl}_2$  (100 mL) and a bromine solution (0.21 mL, 1.1 equiv.) in  $\text{CH}_2\text{Cl}_2$  (30 mL) was added dropwise. The reaction was kept in the dark under stirring at reflux for 24 h. Next, the organic phase was washed with an aqueous solution of thiosulfate and dried over anhydrous  $\text{Na}_2\text{SO}_4$ . Upon evaporation of the solvent the crude solid products **15a,b** were collected and recrystallized from EtOH/*i*-Pr<sub>2</sub>O and fully characterized.

**Compound 15a.** 1.69 g (95%), mp 226-228 °C from EtOH/*i*-Pr<sub>2</sub>O. IR:  $\nu$  1649 (C=O), 1578 (C=N)  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  2.16 and 2.60 (d, 1H+1H, CH<sub>2</sub>), 4.37 (d, 1H, *J* 8 Hz, H4<sub>isox.</sub>), 4.83 (s, 1H, CH-N), 5.13 (s, 1H, CH-O), 5.37 (d, 1H, *J* 8 Hz, H5<sub>isox.</sub>), 7.36 (m, 2H, arom.), 7.48 (m, 1H,

arom.), 7.70 (m, 6H, arom.), 8.00 (d, 2H, arom.), 8.68 (d, 2H, arom.).  $^{13}\text{C}$  NMR (75 MHz, DMSO- $d_6$ , 25 °C):  $\delta$  33.1, 57.6, 60.6, 80.6, 82.6, 121.6, 124.3, 125.9, 126.9, 127.0, 127.5, 128.4, 128.7, 129.9, 130.1, 131.6, 131.7, 154.2, 171.6.  $\text{C}_{27}\text{H}_{19}\text{BrN}_2\text{O}_3$  (499.36): HRMS: calcd. (MW+H) 499.0652; found 499.0647.

**Compound 15b.** 1.64 g (92%), mp 228-230 °C from EtOH *i*-Pr $_2$ O. IR:  $\nu$  1650 (C=O), 1577 (C=N)  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ , 25 °C):  $\delta$  2.31 and 2.65 (d, 1H+1H, CH $_2$ ), 4.28 (d, 1H,  $J$  8 Hz, H $_{4\text{isox}}$ ), 4.54 (s, 1H, CH-N), 5.38 (s, 1H, CH-O), 5.49 (d, 1H,  $J$  8 Hz, H $_{5\text{isox}}$ ), 7.45 (m, 2H, arom.), 7.61 (m, 1H, arom.), 7.66 (m, 4H, arom.), 7.77 (d, 2H, arom.), 7.95 (d, 2H, arom.), 8.66 (d, 2H, arom.).  $^{13}\text{C}$  NMR (75 MHz, DMSO- $d_6$ , 25 °C):  $\delta$  33.3, 61.6, 62.6, 79.8, 82.5, 122.1, 124.7, 126.2, 127.3, 127.4, 128.2, 128.5, 128.7, 128.8, 130.2, 130.3, 132.0, 154.1, 170.0.  $\text{C}_{27}\text{H}_{19}\text{BrN}_2\text{O}_3$  (499.36): HRMS: calcd. (MW+H) 499.0652; found 499.0648.

#### 1.5.4 Synthesis of Hydroxylamines 16a,b

Compounds **15a,b** (1.9 g, 3.80 mmol) were dissolved in MeOH (100 mL) and powdered NaOH (0.18 g, 4.50 mmol) was added portionwise at rt. The reaction was left under stirring for 48 h. After this period of time, the solvent was evaporated at reduced pressure and the residue taken up with CH $_2$ Cl $_2$  and the organic phase washed with brine and finally dried over anhydrous Na $_2$ SO $_4$ . Upon evaporation of the solvent the crude solid products **8a,b** were collected and properly recrystallized from EtOAc and fully characterized.

**Compound 16a.** 1.47 g (98%), mp 180-182 °C from EtOAc. IR:  $\nu$  3235 (NH), 1654 (C=N)  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ , 25 °C):  $\delta$  1.93 and 2.43 (d, 1H+1H, CH $_2$ ), 3.45 (bs, 1H, CH-N), 3.86 (b, 1H, H $_{4\text{isox}}$ ), 4.80 (s, 1H, CH-O), 5.09 (d, 1H,  $J$  8 Hz, H $_{5\text{isox}}$ ), 6.61 (s, 1H, NH), 7.77 (m, 4H, arom.), 8.10 (d, 2H, arom.), 8.56 (d, 2H, arom.).  $^{13}\text{C}$  NMR (75 MHz, DMSO- $d_6$ , 25 °C):  $\delta$  35.0, 58.3, 61.3, 76.1, 83.7, 123.8, 124.5, 127.2, 127.4, 128.0, 128.7, 154.8.  $\text{C}_{20}\text{H}_{15}\text{BrN}_2\text{O}_2$  (395.26): HRMS: calcd. (MW+H) 395.0390; found 395.0396.

**Compound 16b.** 1.48 g (99%), mp 140-142 °C from EtOAc. IR:  $\nu$  3239 (NH), 1624 (C=N)  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz, CDCl $_3$ , 25 °C):  $\delta$  2.19 and 2.71 (d, 1H+1H, CH $_2$ ), 3.71 (s, 1H, NH), 4.13 (d, 1H,  $J$  8 Hz, H $_{4\text{isox}}$ ), 3.31 (s, 1H, CH-N), 4.46 (s, 1H, CH-O), 5.42 (d, 1H,  $J$  8 Hz, H $_{5\text{isox}}$ ), 7.67 (m, 4H, arom.), 7.93 (d, 2H, arom.), 8.63 (d, 2H, arom.).  $^{13}\text{C}$  NMR (75 MHz, CDCl $_3$ , 25 °C):  $\delta$  34.8, 61.4, 63.0, 77.3, 82.8, 122.3, 124.7, 124.8, 126.0, 127.3, 128.2, 128.6, 130.2, 154.3.  $\text{C}_{20}\text{H}_{15}\text{BrN}_2\text{O}_2$  (395.26): HRMS: calcd. (MW+H) 395.0390; found 395.0392.

### 1.5.5 Synthesis of Aminols 17a,b

Compounds **16a,b** (1.4 g, 3.54 mmol) were dissolved in EtOAc (70 mL) along with 0.4 g C/Pd 10% and the mixture was submitted to hydrogenation at rt. The reaction was left under stirring with hydrogen stream for 3 h. After this period of time, the solution was filtered and the solvent was evaporated to dryness. The crude solid products **17a,b** were recrystallized from *i*-Pr<sub>2</sub>O and fully characterized.

**Compound 17a.** 1.38 g (98%), mp 75-78 °C from *i*-Pr<sub>2</sub>O. IR:  $\nu$  3348, 3284 (NH<sub>2</sub>), 3176 (OH), 1622 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  1.93 and 2.25 (m, 1H+1H, CH<sub>2</sub>), 3.22 (m, 1H, CH-N), 4.41 (m, 1H, H<sub>4</sub><sub>isox.</sub>), 4.48 (m, 1H, CH-O), 5.42 (m, 1H, H<sub>5</sub><sub>isox.</sub>), 5.94 (bs, 2H, NH<sub>2</sub>), 7.64 (m, 4H, arom.), 8.05 (bs, 1H, OH), 8.21 (d, 2H, arom.), 8.57 (d, 2H, arom.). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  37.2, 59.7, 63.5, 76.5, 91.3, 106.8, 122.3, 125.5, 125.6, 127.3, 127.7, 128.2, 128.9, 129.1, 129.6, 130.8, 154.7. C<sub>20</sub>H<sub>17</sub>BrN<sub>2</sub>O<sub>2</sub> (397.27): HRMS: calcd. (MW+H) 397.0546; found 397.0533.

**Compound 17b.** 1.37 g (97%), mp 78-80 °C from *i*-Pr<sub>2</sub>O. IR:  $\nu$  3366, 3303 (NH<sub>2</sub>), 3180 (OH), 1654 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  1.71 and 2.11 (m, 1H+1H, CH<sub>2</sub>), 3.32 (bs, 2H, NH<sub>2</sub>), 3.63 (m, 1H, CH-N), 3.83 (m, 1H, CH-O), 4.28 (d, 1H, *J* 8 Hz, H<sub>4</sub><sub>isox.</sub>), 5.29 (d, 1H, *J* 8 Hz, H<sub>5</sub><sub>isox.</sub>), 7.61 (m, 4H, arom.), 8.03 (d, 1H, *J* 4 Hz, OH), 7.96 (d, 2H, arom.), 8.19 (d, 2H, arom.). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  39.8, 59.5, 67.1, 74.9, 92.9, 123.1, 124.7, 125.6, 127.1, 127.6, 128.2, 128.7, 129.1, 129.6, 130.8, 154.9. C<sub>20</sub>H<sub>17</sub>BrN<sub>2</sub>O<sub>2</sub> (397.27): HRMS: calcd. (MW+H) 397.0546; found 397.0531.

### 1.5.6 Synthesis of Pyrimidine derivatives 18a,b

Compounds **17a,b** (1.9 g, 4.78 mmol) were dissolved in *n*-BuOH (40 mL) and 5-amino-4,6-dichloropyrimidine (1.57 g, 9.57 mmol) were added to the solution along with excess *i*-Pr<sub>2</sub>NEt (3.09 g, 23.91 mmol). The mixtures were heated at 117 °C in a sealed flask for 48 h. After this period of time, the reactions were quenched with brine and extracted with CH<sub>2</sub>Cl<sub>2</sub> (2 x 15 mL) and the organic phases dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. Upon evaporation of the solvent, the residues were submitted to chromatographic purification to isolate the crude solid products **18a,b** that were recrystallized from EtOH/*i*-Pr<sub>2</sub>O and fully characterized.

**Compound 18a.** 1.40 g (56%), mp 246-248 °C from EtOH/*i*-Pr<sub>2</sub>O. IR:  $\nu$  3362, 3320 (NH<sub>2</sub>), 3180 (OH), 1571 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  1.99 and 2.28 (m, 1H+1H, CH<sub>2</sub>), 4.08 (m, 1H, CH-N), 4.41 (m, 1H, H<sub>4</sub><sub>isox.</sub>), 4.51 (m, 1H, CH-O), 4.85 (s, 2H, NH<sub>2</sub>), 5.33 (m, 1H, H<sub>5</sub><sub>isox.</sub>), 5.50 (d, 1H, *J* 4 Hz, OH), 6.46 (d, 1H, *J* 7 Hz, NH), 7.60 (m, 4H, arom.), 8.09 (d, 2H, arom.), 8.20 (d, 2H, arom.), 8.77 (s, 1H, CH=N). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  30.7, 55.1, 64.0, 76.8, 91.6, 123.4, 124.8, 125.8, 126.9, 127.3, 127.5, 128.0, 128.7, 129.5, 130.8, 136.8, 144.1, 150.3, 155.8. C<sub>24</sub>H<sub>19</sub>BrClN<sub>5</sub>O<sub>2</sub> (524.80): HRMS: calcd. (MW-H) 522.0338; found 522.0322.



**Compound 18b.** 1.66 g (66%), mp 116-118 °C from EtOH/*i*-Pr<sub>2</sub>O. IR:  $\nu$  3373, 3327 (NH<sub>2</sub>), 3181 (OH), 1570 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  1.91 and 2.39 (m, 1H+1H, CH<sub>2</sub>), 4.02 (m, 1H, CH-N), 4.33 (d, 1H, *J* 8 Hz, H<sub>4</sub><sub>isox.</sub>), 4.78 (m, 1H, CH-O), 5.04 (s, 2H, NH<sub>2</sub>), 5.13 (d, 1H, *J* 4 Hz, OH), 5.50 (d, 1H, *J* 8 Hz, H<sub>5</sub><sub>isox.</sub>), 6.71 (d, 1H, *J* 7 Hz, NH), 7.66 (m, 4H, arom.), 8.02 (d, 2H, arom.), 8.21 (d, 2H, arom.), 8.81 (s, 1H, CH=N). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  30.7, 58.7, 66.7, 73.9, 89.7, 122.8, 123.7, 124.3, 124.8, 125.7, 127.2, 127.6, 128.8, 130.8, 137.9, 146.1, 151.7, 155.3. C<sub>24</sub>H<sub>19</sub>BrClN<sub>5</sub>O<sub>2</sub> (524.80): HRMS: calcd. (MW-H) 522.0338; found 522.0325.

### 1.5.7 Synthesis of Nucleosides 10a,b

Pyrimidine derivatives **18a,b** (1.0 g, 1.91 mmol) were dissolved in triethyl orthoformate (50 mL) and the solution stirred for a couple of hours. After this period of time, a catalytic amount of *p*-TsOH was added and the solutions left at rt for 6 days. The reactions were quenched with H<sub>2</sub>O and extracted with CH<sub>2</sub>Cl<sub>2</sub> (2 x 20 mL) and the organic phases dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. Upon evaporation of the solvent, the solid crude products **10a,b** were recrystallized from EtOH/*i*-Pr<sub>2</sub>O and fully characterized.

**Compound 10a.** 0.68 g (67%), mp 183-185 °C from EtOH/*i*-Pr<sub>2</sub>O. IR:  $\nu$  3328 (OH), 1589 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  2.40 and 2.59 (m, 1H+1H, CH<sub>2</sub>), 4.61 (bs, 1H, CH-N), 4.78 (m, 1H, CH-O), 5.10 (d, 1H, *J* 8 Hz, H<sub>4</sub><sub>isox.</sub>), 5.54 (d, 1H, *J* 8 Hz, H<sub>5</sub><sub>isox.</sub>), 5.71 (b, 1H, OH), 7.71 (m, 4H, arom.), 8.15 (m, 4H, arom.), 8.57 (s, 1H, CH=N), 8.59 (s, 1H, CH=N). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  37.8, 57.2, 63.3, 76.6, 92.2, 123.1, 124.5, 125.4, 125.6, 127.6, 127.7, 128.0, 128.8, 129.5, 130.7, 146.1, 150.3, 150.6, 154.2. C<sub>25</sub>H<sub>17</sub>BrClN<sub>5</sub>O<sub>2</sub> (534.80): HRMS: calcd. (MW+FA-H) 578.0236; found 578.0238.

**Compound 10b.** 0.79 g (77%), mp 202-204 °C from EtOH/*i*-Pr<sub>2</sub>O. IR:  $\nu$  3413 (OH), 1558 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  2.50 and 2.68 (m, 1H+1H, CH<sub>2</sub>), 4.16 (bs, 1H, CH-N), 4.56 (d, 1H, *J* 8 Hz, H<sub>4</sub><sub>isox.</sub>), 5.29 (bs, 1H, OH), 5.48 (bs, 1H, CH-O), 6.04 (dd, 1H, *J* 8, 2 Hz, H<sub>5</sub><sub>isox.</sub>), 7.79 (m, 4H, arom.), 8.16 (m, 2H, arom.), 8.59 (d, 2H, arom.), 8.87 (s, 1H, CH=N), 8.90 (s, 1H, CH=N). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  38.2, 61.7, 67.2, 72.9, 89.3, 123.6, 124.5, 125.7, 127.7, 128.3, 129.6, 130.4, 131.2, 146.7, 149.1, 151.5, 151.9, 155.4. C<sub>25</sub>H<sub>17</sub>BrClN<sub>5</sub>O<sub>2</sub> (534.80): HRMS: calcd. (MW+FA-H) 578.0236; found 578.0229.

### 1.5.8 Synthesis of Nucleotide 11a,b

To an ice-cooled solution of POCl<sub>3</sub> (0.45 mmol) in trimethyl phosphate (0.50 mL) were added nucleosides **10a,b** (80 mg, 0.15 mmol). The resulting solutions were left under stirring at room temperature overnight. After this period of time, the reactions were then diluted with diethyl ether (50 mL) and water was added (50 mL); upon shaking solid compounds precipitate and allowing for their collection by filtration under vacuum. The solid crude products **11a,b** were recrystallized from ethanol/water and fully characterized.

**Compound 11a.** 80.2 mg (87%), mp 220 °C (dec.) from EtOH/H<sub>2</sub>O. IR:  $\nu$  3245 (OH), 1582 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  2.40 and 2.57 (m, 1H+1H, CH<sub>2</sub>), 4.61 (bs, 1H, CH-N), 4.79 (m, 1H, CH-O), 5.10 (d, 1H, *J* 8 Hz, H4<sub>isox.</sub>), 5.54 (d, 1H, *J* 8 Hz, H5<sub>isox.</sub>), 7.75 (m, 4H, arom.), 8.15 (m, 4H, arom.), 8.57 (s, 1H, CH=N), 8.77 (s, 1H, CH=N). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  37.8, 57.2, 63.3, 76.6, 92.2, 123.1, 124.5, 124.7, 125.4, 125.6, 127.3, 127.6, 128.1, 128.8, 129.3, 129.5, 130.5, 130.7, 146.1, 148.4, 150.3, 151.0, 154.2. <sup>31</sup>P NMR (121.5 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  -0.10 [s, -PO(OH)<sub>2</sub>]. C<sub>25</sub>H<sub>18</sub>BrClN<sub>5</sub>O<sub>5</sub>P (614.78): HRMS: calcd. (MW-H) 611.9845; found 611.9836.

**Compound 11b.** 89.5 mg (97%), mp >250 °C (dec.) from EtOH/H<sub>2</sub>O. IR:  $\nu$  3265 (OH), 1582 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  2.46 and 2.68 (m, 1H+1H, CH<sub>2</sub>), 4.15 (bs, 1H, CH-N), 4.56 (d, 1H, *J* 8 Hz, H4<sub>isox.</sub>), 5.48 (bs, 1H, CH-O), 6.04 (dd, 1H, *J* 8, 2 Hz, H5<sub>isox.</sub>), 7.81 (m, 4H, arom.), 8.15 (m, 2H, arom.), 8.59 (d, 2H, arom.), 8.88 (s, 1H, CH=N), 8.90 (s, 1H, CH=N). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  38.2, 61.7, 67.2, 72.9, 89.3, 123.6, 124.6, 125.7, 127.3, 127.7, 128.3, 128.9, 129.1, 130.8, 131.1, 146.7, 151.6, 151.9, 155.4. <sup>31</sup>P NMR (121.5 MHz, DMSO-*d*<sub>6</sub>, 25 °C):  $\delta$  -2.10 [s, -PO(OH)<sub>2</sub>]. C<sub>25</sub>H<sub>18</sub>BrClN<sub>5</sub>O<sub>5</sub>P (614.78): HRMS: calcd. (MW-H) 611.9845; found 611.9836.

## 1.6 References

- 1) Lai, C. C.; Shih, T. P.; Ko, W. C.; Tang, H. J.; Hsueh, P. R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents* **2020**, *55*, 105924 9 pages.
- 2) WHO announces COVID-19 outbreak a pandemic. <http://www.euro.who.int/en/health-topics/healthemergencies/coronavirus-covid-19/news/news/2020/3/whoannounces-covid-19-outbreak-a-pandemic> (accessed Sep 7, 2020)
- 3) Coronavirus (COVID-19) Events As They Happen; WHO, 2019; <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> (accessed Sep 28, 2020).
- 4) (a) COVID-19 Vaccine Development Pipeline; Vaccine Centre, London School of Hygiene and Tropical Medicine. [https://vac-lshtm.shinyapps.io/ncov\\_vaccine\\_landscape/](https://vac-lshtm.shinyapps.io/ncov_vaccine_landscape/) (accessed Sep 7, 2020). (b) COVID-19 Vaccine Tracker. Lay summary. <https://covid-19tracker.milkeninstitute.org/> (accessed Sep 7, 2020).
- 5) Maxmen, A. More than 80 clinical trials launch to test coronavirus treatments *Nature* **2020**, *578*, 347–348.
- 6) (a) Mercorelli, B.; Palu', G.; Loregian, A. Drug Repurposing for Viral Infectious Diseases: How Far Are We? *Trends Microbiol.* **2018**, *26*, 865-876. (b) Pushpakom, S.; Iorio, F.; Eyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; Norris, A.; Sanseau, P.; Cavalla, D.; Pirmohamed, M. Drug repurposing: progress, challenges and recommendations *Nat. Rev. Drug Discov.* **2019**, *18*, 41–58.
- 7) Rosa, S. G. V.; Santos, W. C. Clinical trials on drug repositioning for COVID-19 treatment. *Rev. Panam. Salud Publica* **2020**, *44*, No. e40 7 pages.
- 8) Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; Ge, J.; Zheng, L.; Zhang, Y.; Wang, H.; Zhu, Y.; Zhu, C.; Hu, T.; Hua, T.; Zhang, B.; Yang, X.; Li, J.; Yang, H.; Liu, Z.; Xu, W.; Guddat, L. W.; Wang, Q.; Lou, Z.; Rao, Z. Structure of the RNA-dependent RNA polymerase from COVID-19 virus *Science* **2020**, *368*, 779 4 pages.
- 9) WHO Solidarity Trial Consortium Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results *N. Engl. J. Med.* **2020**, *384*, 497-511.
- 10) Hill, A.; Wang, J.; Levi, J.; Heath, K.; Fortunak, J. Minimum costs to manufacture new treatments for COVID-19 *J. Virus Erad.* **2020**, *6*, 61-69.
- 11) Picarazzi, F.; Vicenti, I.; Saladini, F.; Zazzi, M.; Mori, M. Targeting the RdRp of Emerging RNA Viruses: The Structure-Based Drug Design Challenge *Molecules* **2020**, *25*, 5695 25 pages.
- 12) Quadrelli, P. Modern Applications of Cycloaddition Chemistry; Elsevier: Amsterdam, **2019**; pp. 1–152.

- 13) Memeo, M. G.; Quadrelli, P. Generation and trapping of nitrosocarbonyl intermediates *Chem. Rev.* **2017**, *117*, 2108–2200.
- 14) Quadrelli, P.; Mella, M.; Caramella, P. A Photochemical Generation of Nitrosocarbonyl Intermediates *Tetrahedron Lett.* **1999**, *40*, 797–800.
- 15) Cohen, A. D.; Zeng, B. B.; King, S. B.; Toscano, J. P. Direct Observation of an Acyl Nitroso Species in Solution by Time-Resolved IR Spectroscopy *J. Am. Chem. Soc.* **2003**, *125*, 1444–1445.
- 16) Al-Saad, D.; Memeo, M. G.; Quadrelli, P. Pericyclic Reactions for Antivirals: Synthesis of 4-Bromo-*N*-[(1*R*\*,4*S*\*)-4-hydroxy-2-cyclohexen-1-yl]-2-thiazolecarboxamide *Lett. Org. Chem.* **2016**, *13*, 757-763.
- 17) Kirby, G.W. Tilden Lecture. Electrophilic C-nitroso-compounds *Chem. Soc. Rev.* **1977**, *6*, 1-24.
- 18) Adam, W.; Bottke, N.; Krebs, O.; Saha-Moller, C. R. Allylic amidation of olefins by ene reaction of acylnitroso compounds generated in situ by oxidation of hydroxamic acids *Eur. J. Org. Chem.* **1999**, *1999*, 1963-1965.
- 19) Pulacchini, S.; Sibbons, K. F.; Shastri, K.; Motevalli, M.; Watkinson, M.; Wan, H.; Whiting, A.; Lightfoot, A.P. Synthesis of C2-symmetric aza- and azaoxa-macrocyclic ligands derived from (1*R*,2*R*)-1,2-diaminocyclohexane and their applications in catalysis *Dalton Trans. J.* **2003**, *10*, 2043-2052.
- 20) Atkinson, D.; Kabeshov, M. A.; Edgar, M.; Malkov, A. V. Intramolecular carbonyl nitroso ene reaction catalyzed by iron (III) chloride/hydrogen peroxide as an efficient tool for direct allylic amination *Adv. Synth. Catal.* **2011**, *353*, 3347-3351.
- 21) Chaiyaveij, D.; Cleary, L.; Batsanov, A. S.; Marder, T. B.; Shea, K. J.; Whiting, A. Copper(II)-Catalyzed Room Temperature Aerobic Oxidation of Hydroxamic Acids and Hydrazides to Acyl- Nitroso and Azo Intermediates, and Their Diels–Alder Trapping *Org. Lett.* **2011**, *13*, 3442-3445.
- 22) Memeo, M. G.; Quadrelli, P. A new life for nitrosocarbonyls in pericyclic reactions *ARKIVOC* **2013**, *1*, 418-423.
- 23) Quadrelli, P.; Mella, M.; Gamba, I. A.; Caramella, P. The mild oxidation of nitrile oxides affords a convenient entry to nitrosocarbonyl intermediates, versatile tools in organic syntheses *Tetrahedron* **1999**, *55*, 10497-10510.
- 24) Quadrelli, P.; Campari, G.; Mella, M.; Caramella, P. A thermal fragmentation of 1,2,4-oxadiazole-4-oxides to nitriles and nitrosocarbonyls *Tetrahedron Lett.* **2000**, *41*, 2019-2022.
- 25) Quadrelli, P.; Romano, S.; Piccanello, A.; Caramella, P. The remarkable cis effect in the ene reactions of nitrosocarbonyl intermediates *J. Org. Chem.* **2009**, *74*, 2301-2310.
- 26) Bodnar, B. S.; Miller, M. J. The Nitrosocarbonyl Hetero-Diels-Alder Reaction as a Useful Tool for Organic Syntheses *Angew. Chem. Int. Ed.* **2011**, *50*, 5630-5647.
- 27) Peng, Q.; Peng, R.; Yuan, B.; Wang, M.; Zhao, J.; Fu, L.; Qi, J.; Shi, Y. Structural Basis of SARS-CoV-2 Polymerase Inhibition by Favipiravir *Innovation* **2021**, *2*, No. 100080 8 pages.

- 28) Yin, W.; Mao, C.; Luan, X.; Shen, D. D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y.C.; Tian, G.; Jiang, H.W.; Tao, S.C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H. E. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir *Science* **2020**, *368*, 1499–1504.
- 29) Moggio, Y.; Legnani, L.; Bovio, B.; Memeo, M. G.; Quadrelli, P. Synthesis of novel anthracene derivatives of isoxazolino-carbocyclic nucleoside analogues *Tetrahedron* **2012**, *68*, 1384–1392.
- 30) Zhang, Y.; Jiao, Z.; Xu, W.; Fu, Y.; Zhu, D.; Xu, J.; He, Q.; Cao, H.; Cheng, J. Design, synthesis and properties of a reactive chromophoric/fluorometric probe for hydrogen peroxide detection *New J. Chem.* **2017**, *41*, 3790–3797.
- 31) Memeo, M. G.; Lapolla, F.; Maga, G.; Quadrelli, P. Synthesis and antiviral activity of anthracene derivatives of isoxazolino- carbocyclic nucleoside analogues *Tetrahedron Lett.* **2015**, *56*, 1986– 1990.
- 32) Quadrelli, P.; Scrocchi, R.; Caramella, P.; Rescifina, A.; Piperno, A. From cyclopentadiene to isoxazoline-carbocyclic nucleosides: a rapid access to biological molecules through nitrosocarbonyl chemistry *Tetrahedron* **2004**, *60*, 3643–3651.
- 33) Ishikura, M.; Murakami, A.; Katagiri, N. First synthesis of 2',3'-epimino-carbocyclic nucleosides. *Org. Biomol. Chem.* **2003**, *1*, 452–453.
- 34) Yoshikawa, M.; Kato, T.; Takenishi, T. Studies of Phosphorylation. III. Selective Phosphorylation of Unprotected Nucleosides. *Bull. Chem. Soc. Jpn* **1969**, *42*, 3505–3508.
- 35) Caton-Williams, J.; Lin, L.; Smith, M.; Huang, Z. Convenient synthesis of nucleoside 5'-triphosphates for RNA transcription *Chem. Comm.* **2011**, *47*, 8142–8144.
- 36) Berger, S.; Braun, S.; Kalinowski, H.-O. *NMR Spectroscopy of the Non-Metallic Elements*; John Wiley And Sons, **1997**.
- 37) 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.* **2022**, *65*, 2716–2746.
- 38) Faita, G.; Leusciatti, M.; Quadrelli, P. Synthesis and biological activity of potential antiviral compounds through 1,3-dipolar cycloadditions. Part 1: general aspects and reactions of azides *ARKIVOC* **2022**, *part viii*, 19–78.
- 39) Faita, G.; Mella, M.; Quadrelli, P. Synthesis and biological activity of potential antiviral compounds through 1,3-dipolar cycloadditions. Part 2: nitrones, nitrile oxides and imines, and other 1,3-dipoles *ARKIVOC* **2022**, *part viii*, 257–294.

## 1.7 Appendix

### 1.7.1. Ligand efficiencies and trajectory analyses

**Table S1.** a) Ligand efficiencies of selected binding poses of triphosphates (TPs) in the precatalytic structure. Submitted to molecular dynamics. b) Ligand efficiencies of selected binding poses of monophosphates (MPs) in the post-catalytic structure.

(a)

	7CTT_TPs	Ligand Efficiency
	RTP	-0,288
	FTP	-0,229
11a(1)	A1	-0,187
11a(2)	A2	-0,311
11b(1)	B1	-0,228
11b(2)	B2	-0,229

(b)

	7BV2_MPs	Ligand Efficiency
	RMP	-0,389
11a(1)	A1	-0,197
11a(2)	A2	-0,317
11b(1)	B1	-0,300
11b(2)	B2	-0,294

RTP, Remdesivir triphosphate; FTP, Favipiravir triphosphate; Refer to **Figure 1** in the text for structures.

**Table S2.** a) and b) analysis of the trajectories, mean values of RMSD of the ligand respect to the protein and distance O3'-P $\alpha$  are reported. Color code for hydrogen bonds to RNA: green H-Bond maintained for 70<100% of the time, yellow 70<50%, orange 10<50%, red <10%.

(a)

	7CTT	Ligand Efficiency	RMSD (Å)	d(Å) OH3' P $\alpha$	H-bond RNA
	RTP	-0,288	1,94	5,9	Green
	FTP	-0,229	3,01	9,4	Yellow
11a(2)	A2	-0,311	1,33	4,70	Green
11b(1)	B1	-0,228	3,78	8,80	Red

(b)

	7BV2	Ligand Efficiency	RMSD (Å)	d(Å) OH3' P $\alpha$	H-bond RNA
	RMP	-0,389	1,37	5,4	Green
	RMP_INC	-	2,34	1,7	Green
11a(2)	A2	-0,317	1,15	7,3	Red
11a(2)	A2_INC	-	0,85	1,7	Green
11b(1)	B1	-0,300	2,87	3,6	Green

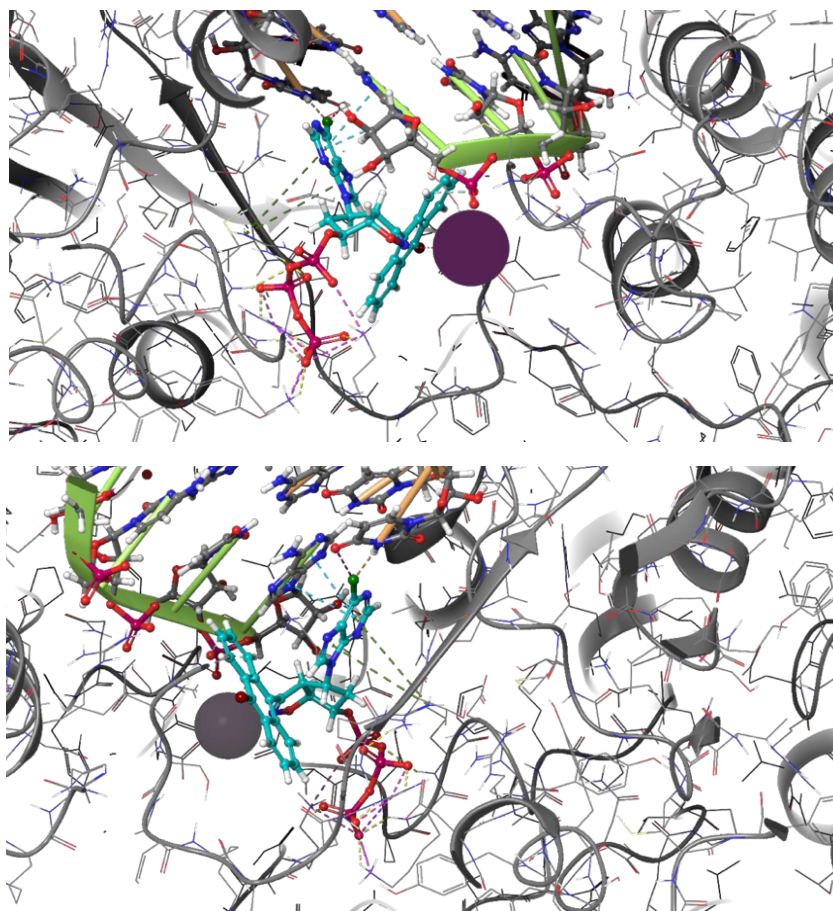
RTP, Remdesivir triphosphate; FTP, Favipiravir triphosphate; Refer to **Figure 1** in the text for structures.

### 1.7.2. Docking poses for triphosphate derivatives in 7CTT

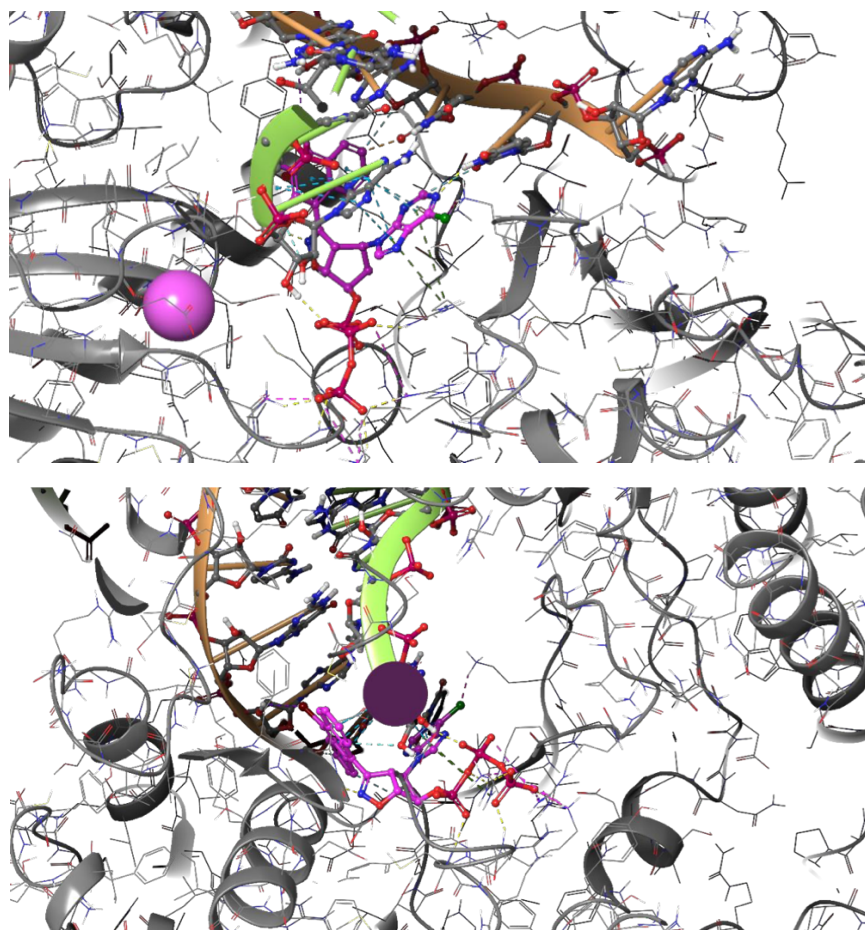
Refer to **Figure 1** in the text for structures.

**Figure S1.**

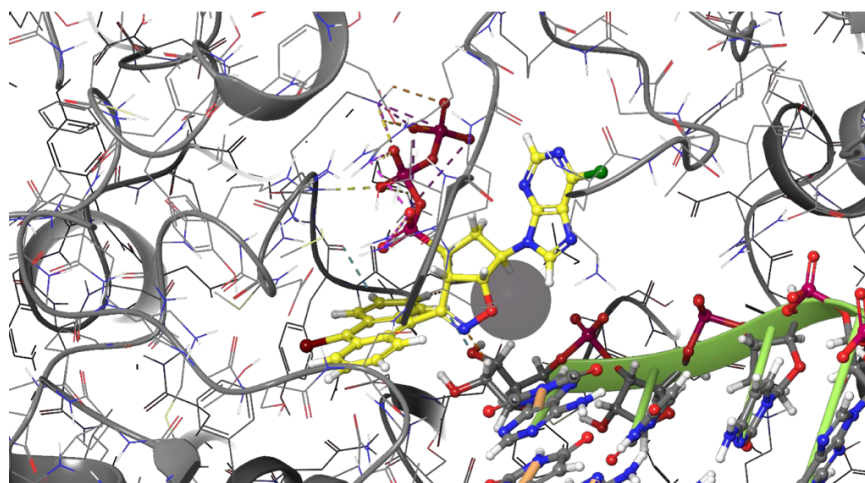
**A1**



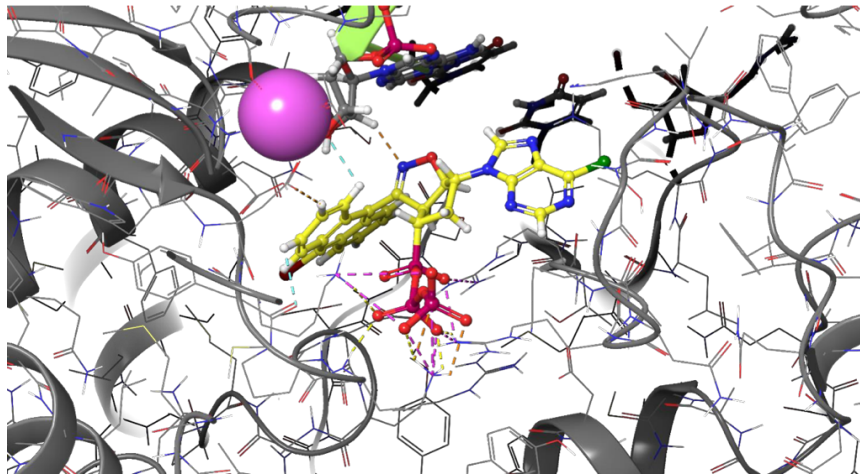
**A2**



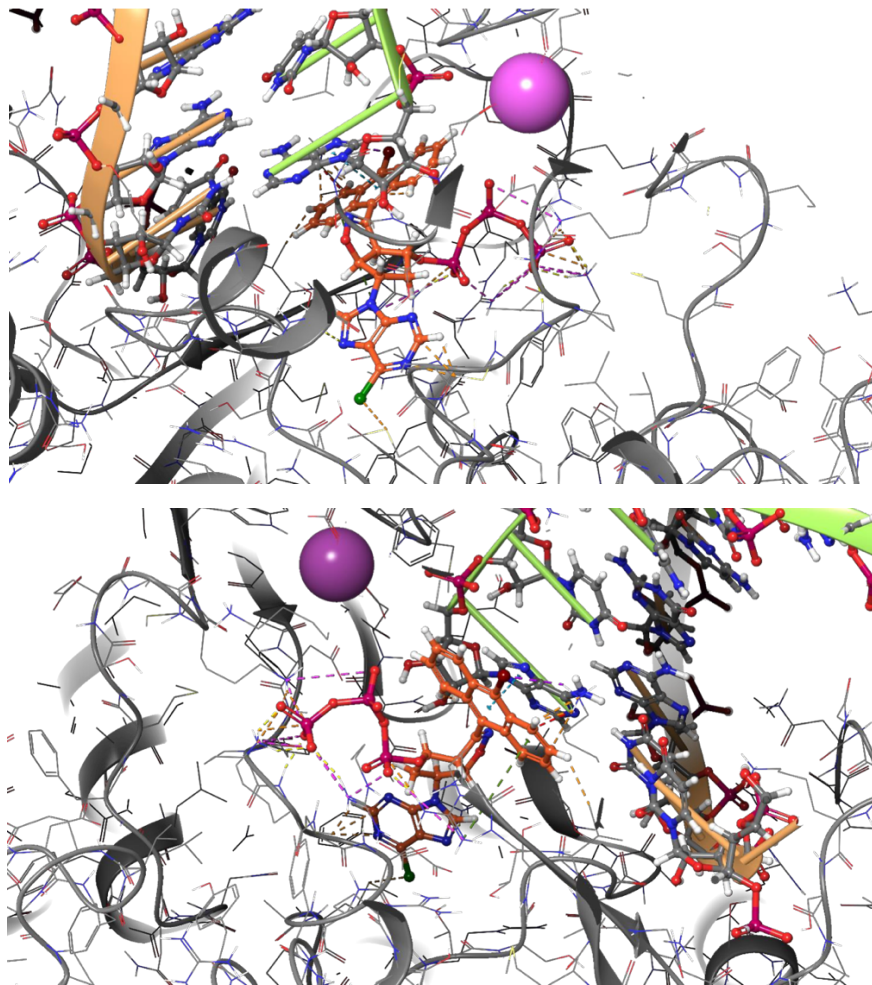
**B1**



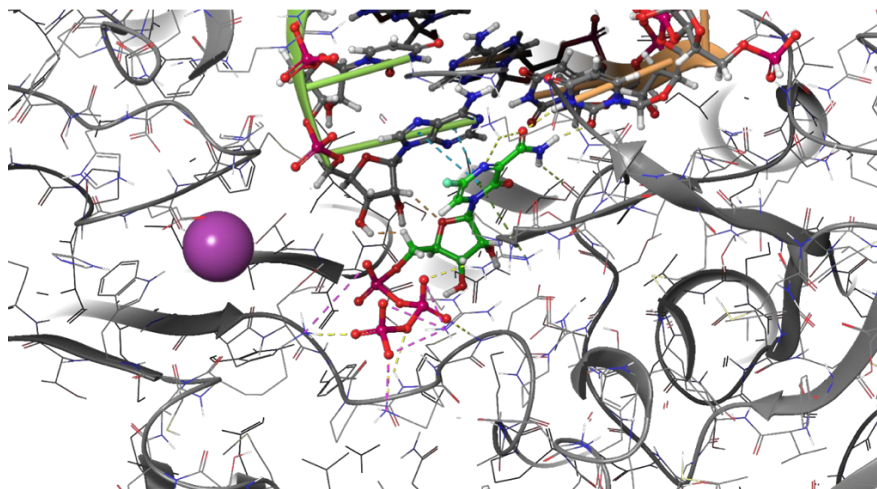




**B2**



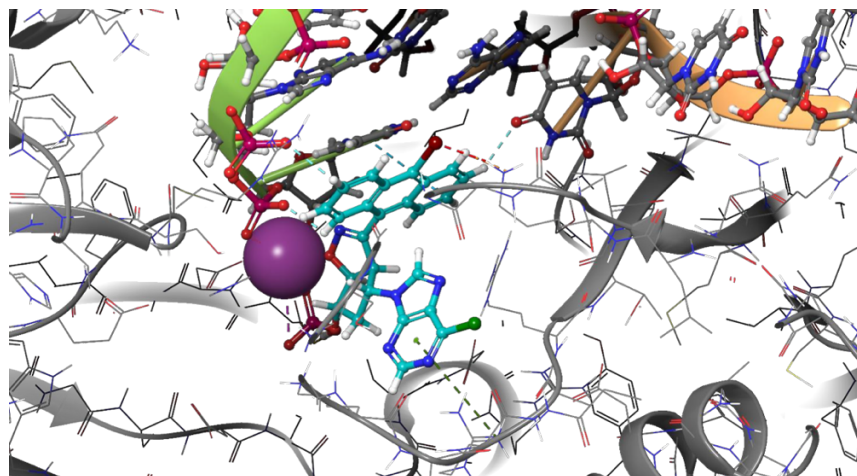
FTP

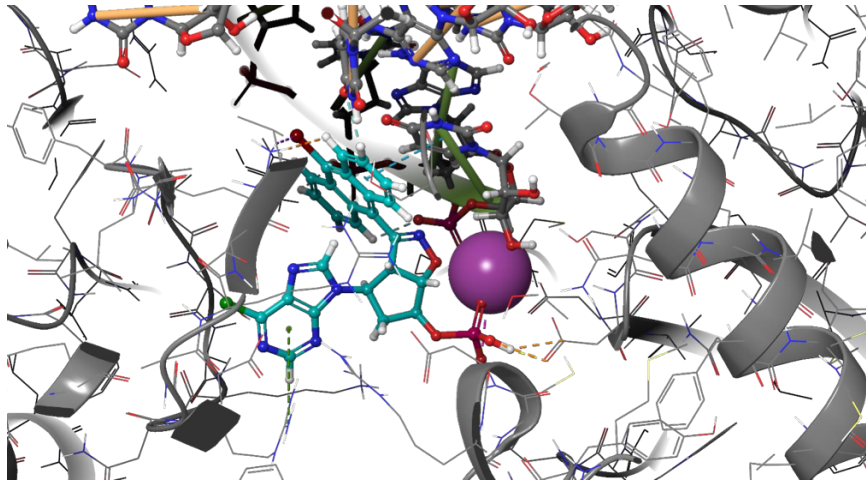


### 1.7.3. Docking poses for monosphate derivatives in 7BV2

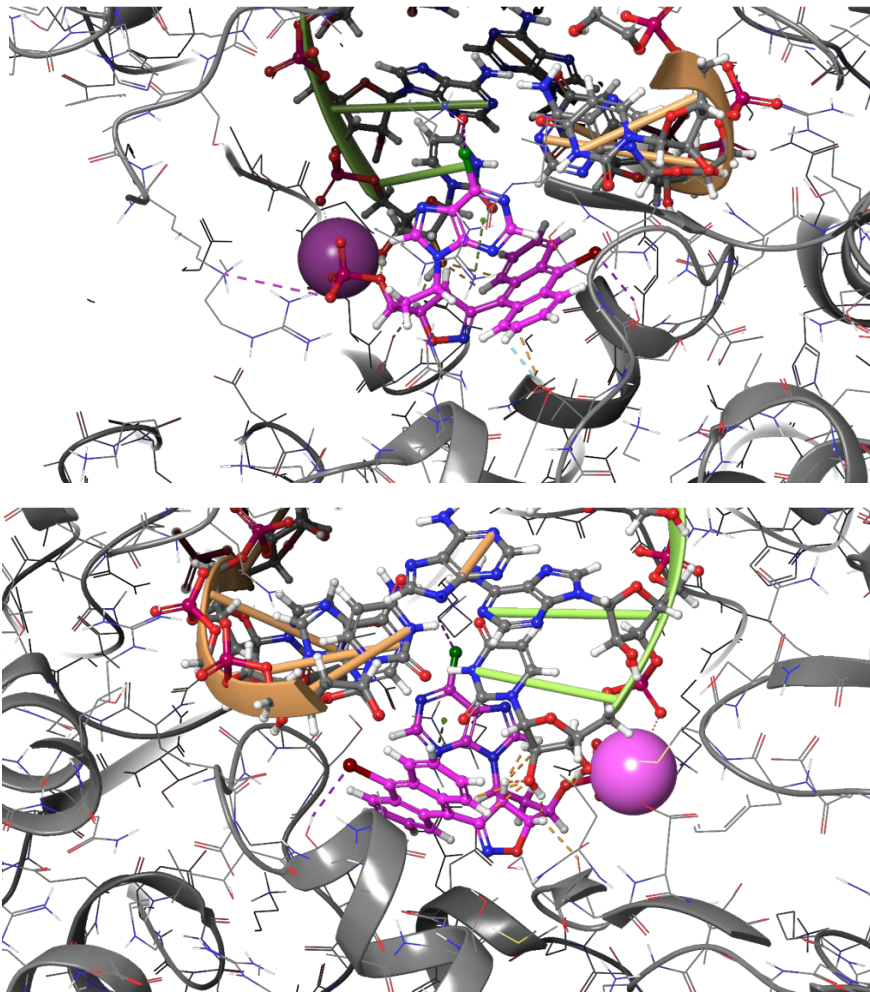
Figure S2.

A1



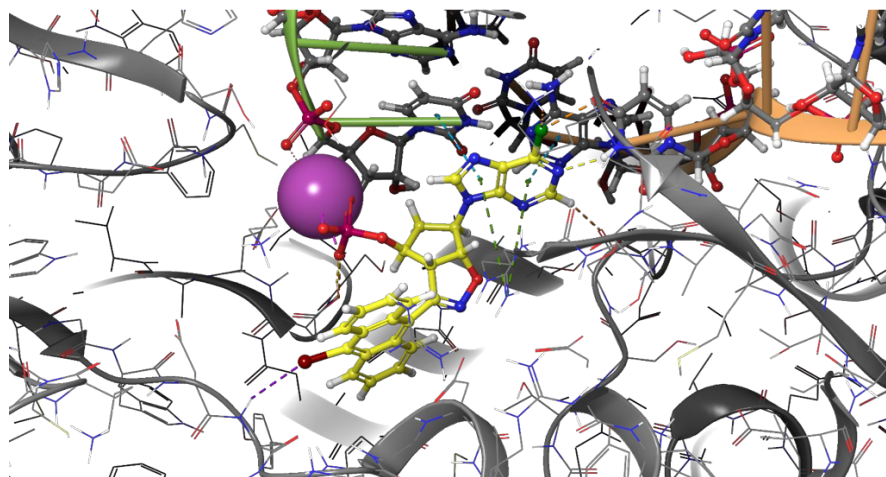


A2

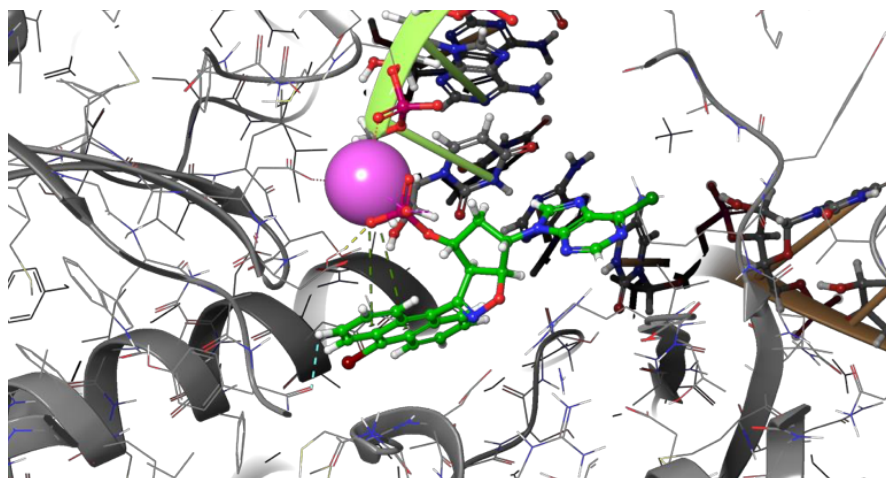




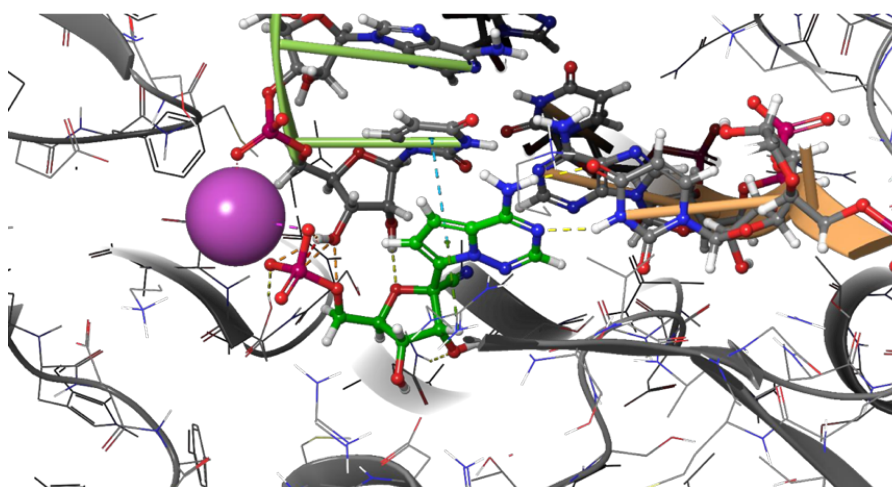
**B1**



**B2**



**RMP**



# Chapter 2

## 2.1 Targeting SARS-CoV-2's M<sup>Pro</sup> (NSP5)

Due to its great conservation throughout coronaviruses, the non-structural SARS-CoV-2 major protease, NSP5 or M<sup>Pro</sup>, is a relevant biological target (in contrast to the spike structural protein, which has a high rate of mutation). Moreover, it is essential to viral multiplication because it cleaves polypeptides to create shorter, non-structural proteins that are needed for replication.<sup>1</sup>

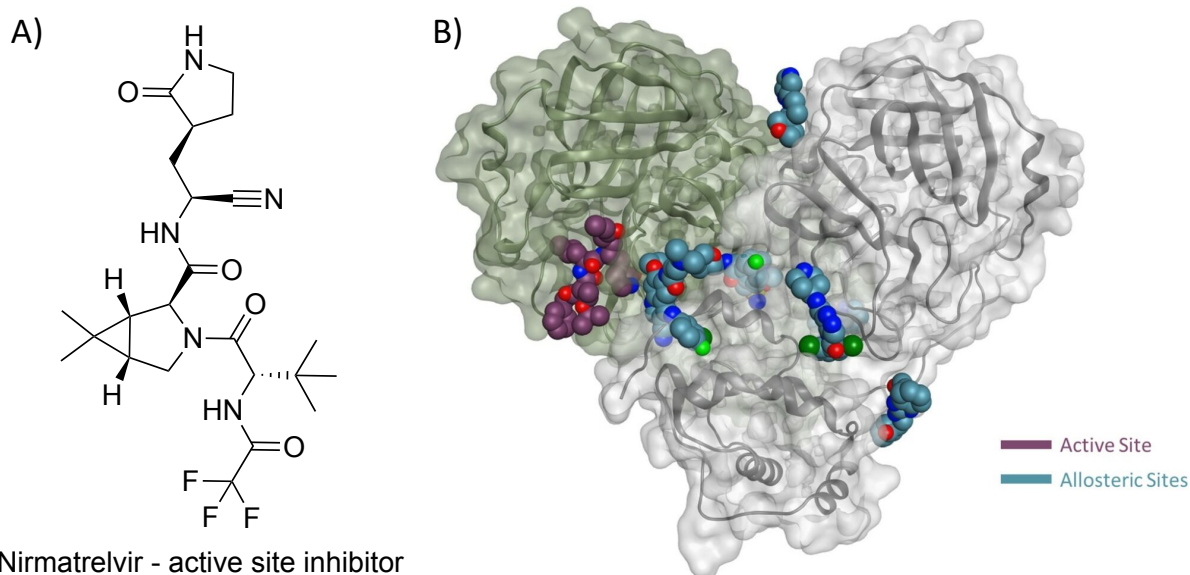
In January 2022, nirmatrelvir plus ritonavir received authorization in the EU for the treatment of COVID-19 in adults who do not require supplemental oxygen and who are at increased risk for progression to severe COVID-19. It represented one of the earliest oral antiviral therapies for COVID-19, which Pfizer discovered to be a strong SARS-CoV-2 M<sup>Pro</sup> inhibitor. Co-packaged with a low dose of ritonavir, it is meant to be administered in conjunction with it. Though ritonavir lacks direct action against SARS-CoV-2 M<sup>Pro</sup>, it slows down nirmatrelvir's metabolism to optimize its therapeutic efficacy. Ritonavir is a CYP3A inhibitor. Nirmatrelvir specifically and reversibly inhibits SARS-CoV-2 M<sup>Pro</sup> activity by binding directly to the catalytic site. [inhibition constant (K<sub>i</sub>) 3.1 nM; half-maximal inhibitory concentration (IC<sub>50</sub>) 19.2 nM]. This inhibition hinders the protein from processing polyprotein precursors, which in turn stops the spread of the virus.<sup>1</sup>

So far, peptidomimetics that directly target the active site, such as nirmatrelvir (**Figure 1A**), have reached the most advanced stages of development; moreover, the activity of the currently evaluated candidates could be improved by medicinal chemistry efforts, considering the wealth of structural information that has become available.<sup>1,2</sup>

Conversely, allosteric inhibitors have not yet been deeply studied for SARS-CoV-2 and for other human pathogenic CoVs. Allosteric inhibitors generally show lower resistance barrier with respect to molecules directly targeting the active site and are characterized by drug-like heterocyclic scaffold; thus, the identification of such agents may offer the advantage to handle molecules with higher optimization potential.<sup>2</sup>

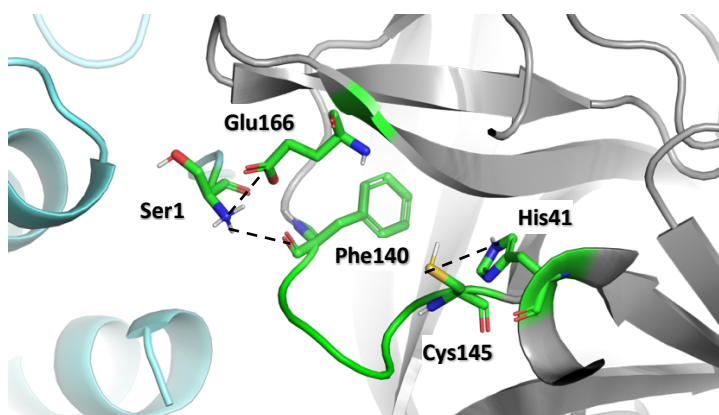
The integrated computational and experimental analysis of the NSP5 surface may point toward the discovery of potential allosteric sites. In solution, the functional replicase complex exists as a dimer owing to the stable association of highly conserved motifs, some of which may be promising targets. The research field of allosteric functional modulation of protein dynamics is becoming more and more popular. Discovery of new allosteric sites, for instance, may open up new avenues for drug development and help to overcome the limitations of traditional active-site-oriented drug design.<sup>1,2</sup>

In fact, some allosteric sites have been recently mapped on the M<sup>Pro</sup> protein for CoVs and a few complexes with co-crystallized ligand, specifically binding one of the sites, have begun to emerge (**Figure 1B**).<sup>3</sup>



**Figure 1.** A) Nirmatrelvir. B) Representation of the M<sup>Pro</sup> enzyme in the dimeric form, along with co-crystallized ligands in either active site and in experimentally proven allosteric sites.<sup>3</sup> [Reproduced with permission from Ref 3. Copyright © 2022 Alzyoud et al., CC BY-NC 3.0 DEED]

M<sup>Pro</sup> monomers arrange themselves in the proper conformation to carry out the catalytic function upon dimerization and activation, and their N fingers (see **Figure 2**) interact with each other's Glu166.



**Figure 2.** One protomer's N terminal engages interaction with the other protomer's active site region. The side chain of Glu166 and the backbone oxygen of Phe140 establish hydrogen bonds with the N-terminal Ser1. This affects the oxyanion hole (green loop) and the area next to the catalytic residues Cys145 and His41 (PDB: 5RFA).

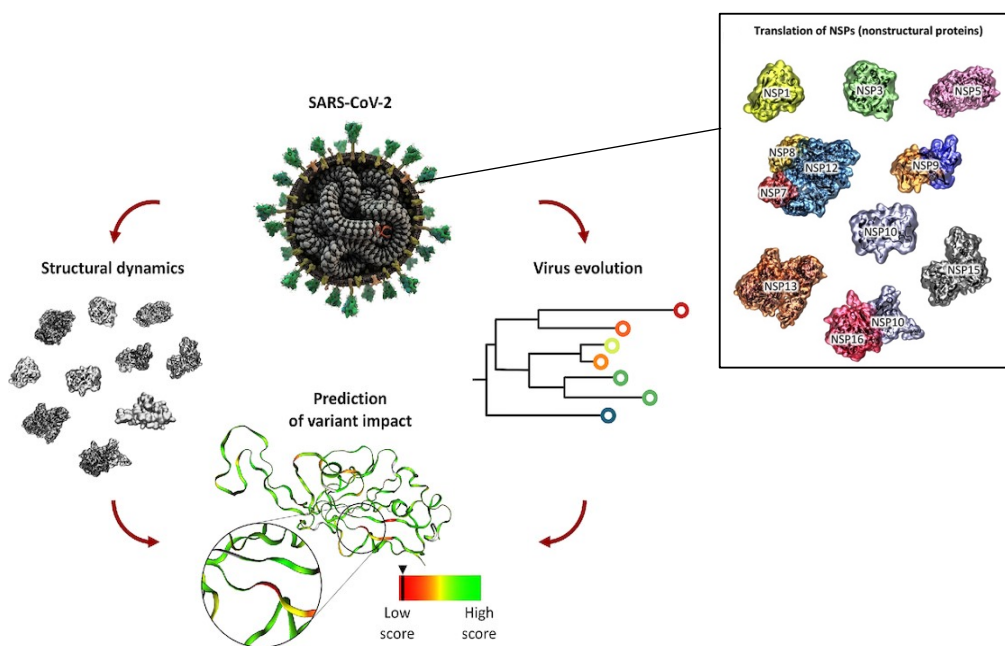
As dimerization molds the S1 pocket of the substrate-binding site, which houses the catalytic dyad, His41, and Cys145, this step is essential to the enzyme's catalytic function. Previous research using MD simulations and mutational analysis identified important residues, such as Arg4, Ser10, Gly11, Glu14, Asn28, Ser139, Phe140, Ser147, Glu290, and Arg298 that are crucial in stabilizing the catalytically active dimeric structure of the SARS-CoV-2 M<sup>Pro</sup> enzyme.<sup>3,5</sup> Together, these residues regulate dimerization and preserve the integrity of the dimer interface, both of which are critical for the stability of the dimer.<sup>3,6</sup> It has been determined that several inhibitors bind to the dimer location and cause mild to moderate antiviral activity. Therefore, disrupting the residues on the dimer site is probably going to prevent the enzymes from catalyzing reactions and from dimerizing.<sup>3,4</sup>

From molecular recognition and binding to enzymatic activity, the dynamic characteristics of proteins are essential to every facet of protein functions. The intricate interplay of covalent and noncovalent interactions defines the relative population of three-dimensional (3D) structures (defined by their free energies) and the potential kinetic pathways for interconversion among them (defined by the heights of the free energy barriers between them). These interactions determine the functional dynamics of a system.<sup>7</sup>

Catalytic function can be turned on, adjusted, regulated, or prevented as a result of this altering the structural and/or dynamic features of the protein. Numerous methods, including ligand binding, covalent changes, or mutations, might cause the protein's conformational ensemble to be perturbed. Certain accessible conformations with particular functional characteristics may be adopted by binding of a ligand or substrate at an active site or of a protein partner at a particular zone of the structure. The precise mix of dynamic modes that underpins molecule recognition and function can be altered by allostery-induced molecular disturbances by affecting the covalent and noncovalent forces. The binding affinity in a remote functional region, such the active site or a protein contact surface, may therefore be impacted by changes in the protein state at a particular location.<sup>7,8</sup>

In the last years there has been a growing interest in the study of allosteric functional regulation in dynamic proteins. Specifically, the identification of new allosteric sites may open up new avenues for drug development and help to overcome the constraints of traditional active-site-oriented drug design.<sup>7</sup> Reasonable targeting of alternative sites might therefore provide innovative ways to disrupt protein-protein interactions, which are widely acknowledged as difficult targets, as well as novel chemotypes for possible inhibitors.<sup>7</sup> Massive research endeavors have yielded high-resolution structural data on the majority of the SARS-CoV-2 3D-proteome, which is extensively available through the PDB databank. The structural data obtained from the experimental methods used (X-ray crystallography and cryogenic electron microscopy) has proven to be a valuable foundation for subsequent efforts to use MD simulations to acquire time-resolved data regarding the functional dynamics of viral proteins.<sup>9</sup>

Published at the end of 2021, SCoV2-MD is a new online resource that systematically arranges atomistic simulations of the SARS-CoV-2 proteome.<sup>9</sup> The library includes simulations developed by top research teams employing molecular dynamics (MD) techniques to investigate the relationships between viral protein dynamics, structure, and function. Understanding the functional dynamics of the viral proteome, which is typically not deduced from the static structure that has been solved experimentally, depends critically on the derived molecular dynamics (MD) data. The platform is centered around the dynamics of noncovalent protein contacts and facilitates the mapping of external data, the investigation of individual subunits, interfaces, and protein-ligand contacts, as well as the interpretation of allostery.<sup>9</sup>



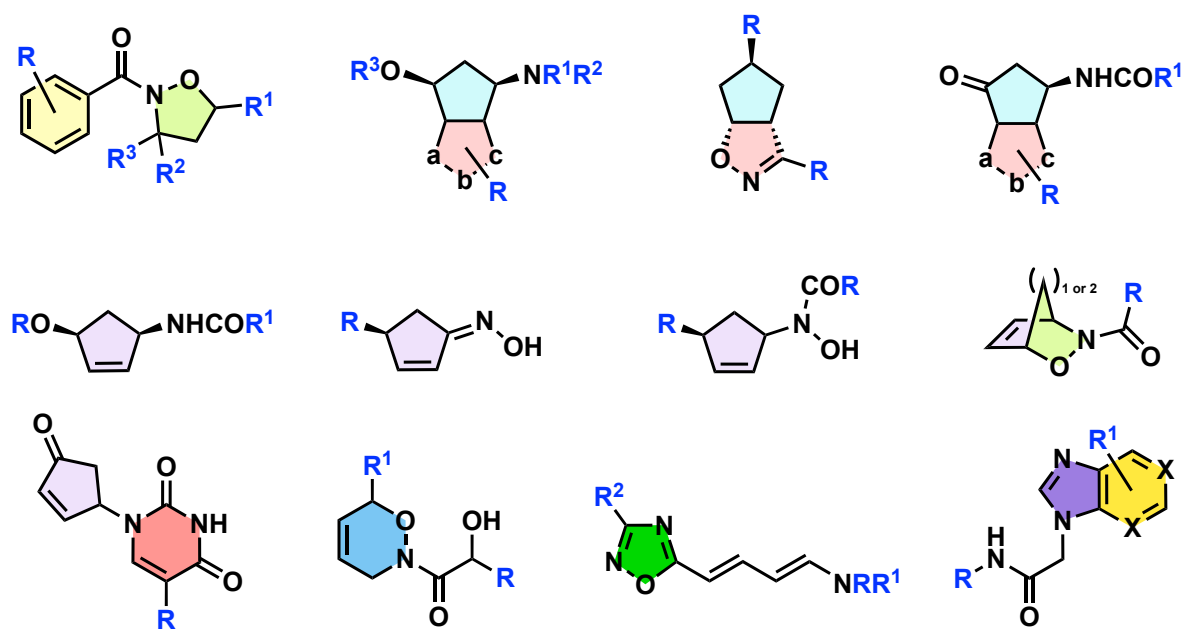
**Figure 3.** The diagrams provide a unified overview of SCoV2-MD database, it integrates structural dynamics of SARS-CoV-2 proteins with mutation events from virus evolution. Static and dynamic properties of each mutation can be combined to obtain predictions of its impact on viral proteins. Three-dimensional models of the available NSPs are also reported. [Adapted with permission from ref 9. Copyright © 2021 Torrens-Fontanals et al., CC BY-NC 4.0 DEED]

Analyses of these data allowed us at first to select the most promising binding sites then retrieving receptor's conformations with particular functional characteristics. Lastly, after computational analysis (Docking experiments, MD simulations), to purpose candidate molecules that would work as novel antagonists of NSP5 protease function. These compounds would target a different site than the active one, destabilizing several important protein-protein interactions, interfering with the active site state. In solution, the protein exists as a dimer owing to the stable association of highly conserved motifs some of which may be promising targets.



The search for new compounds for the treatment of viral diseases and for selective antiviral chemotherapy is being actively pursued in our Pavia's group, and various scaffolds have been used in working towards this goal. Here below are reported a few examples (**Figure 4**).

As can be seen in **Figure 4** the chemical space explored in the past years by our research group comprehends mostly five- and six-membered heterocyclic rings: the most frequently used ring systems in small molecule drugs. The great deal of interest in these scaffolds derives from the biological activities of their derivatives and the frequent use in drug discovery for example as bioisosteric replacements or spacers. So, the reported scaffolds seem to fit nicely with the application of computational experiments for selecting the structural modification to prompt biological activity.



**Figure 4.** Overview of the affordable scaffolds by the pathways developed by the Pavia's group (a,b,c = C=N-O or O-C=N).

Among the five-membered heterocyclic rings appear isoxazolidines, isoxazolines fused to a cyclopentane ring, cyclopentene, 1,2,4-oxadiazoles, norbornenes and hydroxycyclopentenones.<sup>11-17</sup> Whilst in the six-membered family we can find nucleobases of the purine or pyrimidine type, cyclohexenes, oxazines.<sup>18-21</sup> All of which can be extensively and in most of the cases stereospecifically decorated with a plethora of functional groups, starting from the simplest and smallest amino, hydroxyl, carbonyl, oxime groups, reaching expanded polyaromatic groups that may be active through their ability to establish  $\pi$ - $\pi$  stacking interactions. Another idea would be to link the polycyclic aromatic residue to a heterocyclic ring in a rigid or flexible manner. The replacement of traditional or substituted heterobases, such as purines or pyrimidines, with selected and easily available heterocycles bearing a single

heteroatom (the pyridine ring) in which the nitrogen atom orientation can be easily changed can also be sought. Alternatively, the quinoline ring, a naphthalene-like heterocycle, possesses the structural features to display both the effects of an extended aromatic system and the capabilities of donating a lone pair of electrons.<sup>22</sup> The introduction of carbohydrate-based chiral substituents or aminoacidic chains represent another promising plan.<sup>21</sup>

All these possibilities allow for a viral target-oriented production of small molecules able to display some inhibition activity and suitable for further development under structural modification suggested by computational analysis and biological tests. We expect that these decorated scaffolds, obtained by the chemical pathways traditionally pursued by our research group, could find interesting applications in the present investigation on drug design/development due to the intrinsic characteristics of their structural and electronic properties from a medicinal chemistry point of view (i.e. minor degrees of freedom of movement are allowed for a ring system compared to linear system, elevated number of intrinsic pharmacophoric features, regio- and stereo-ordered introduction of substituents).

## 2.2 Results: Computational Analysis

### 2.2.1 Methods

Computational experiments were conducted inside Schrodinger's Maestro v13.1.141 release 2022-1. Clustering analyses were performed using GROMACS 2021.

### 2.2.2 Clustering

Downloaded trajectory from the MD-SARS-CoV2 database (PDB: 6LU7; model file ID: 520; trajectory file ID: 521; 500  $\mu$ s; software and version: GROMACS 2020.2; FF: AMBER 99) were visualized with VMD and analyzed with GROMACS 2021.<sup>9</sup> An index file was generated from the model file and clustering analysis was performed (on C $\alpha$  or selected residues indicated in *paragraph 2.2.9*) using the method *gromos* with a cutoff of 0.2 nm. The most representative structure of each cluster was used in the subsequent analysis.

### 2.2.3 Modeling

The most representative conformations of the protein, identified through the clustering analysis, were prepared using Maestro: hydrogens were added, and the protonation states were generated using PROPKA at pH 7.

### 2.2.4 Binding sites identification

On the most representative structure of the protein, obtained through clustering analysis, pockets were searched using SiteMap, a module of Schrodinger's Maestro.

### 2.2.5 Database

The library, of around 600 compounds, was designed based on the chemical pathways developed by our research unit.<sup>11-22</sup>

### 2.2.6 Docking

Grids for docking were generated using Glide, with default settings. Ligands were built with the 3D Builder tool of Maestro and then prepared with LigPrep using force field OPLS4, and Epik was used to generate possible states at target pH  $7 \pm 2$ . Ligands were docked with extra precision in the predefined grids by using Glide. Ten poses per ligand were generated, and post-docking minimization was performed. Optimal poses were chosen based on docking score, ligand efficiency and interactions with selected residues, Arg298 for pocket 1, Ser1, Arg4, Lys5 and Glu290 for pocket 2. Ligand interactions diagram (LID) of docked compounds were generated in 2D workspace.

### 2.2.7 MD Simulations

Systems were prepared using Desmond's System Builder with the OPLS4 force field. Protein-ligand complexes were put into an orthorhombic box, surrounded by around 22000 water molecules and SPC solvent model, with a buffer distance of 15 Å. The volume of the system was then minimized to around  $1 \times 10^6$  Å. Negative charges in the system were neutralized by the addition of Na<sup>+</sup> ions. NaCl at a concentration of 0.15 M was also added.

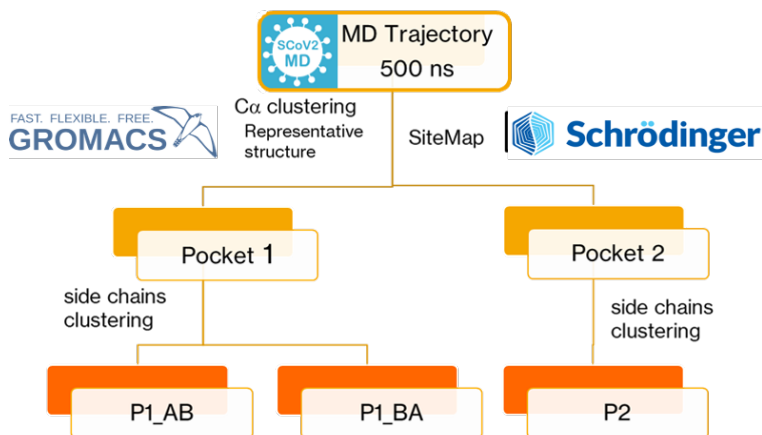
Molecular dynamics trajectories were generated by using Schrodinger's Desmond. NPT ensemble was used at a temperature of 300.0 K and pressure of 1.01325 bar. Prior to the simulation, the system was relaxed with the standard relaxation protocol of Desmond with default settings resulting in around 200 ps equilibration. Nose-Hoover thermostat with relaxation time 1 ps and Martyna-Tobias-Klein barostat with relaxation time 2 ps were used. The temperature and pressure coupling were set to isotropic. Simulations were run for 50 ns and the trajectory was recorded each 20 ps.

### 2.2.8 Trajectory Analysis

Interactions with the protein and RMSD were obtained through the simulation interactions diagram plugin of Maestro. Distances were calculated using VMD, the protein was first centered in the box, and then frames were aligned. For the APO protein measurements were done one each 10 frames. Distances in Å were measured throughout the course of the trajectory from the CZ and CD of the residue pair Arg4-Glu290, CZ and O (backbone) for Arg298-Ser123 pair and S atom of Cys145 to N H-donor nitrogen of His41 of the dimer complex. Superposition of docking poses and frames of the trajectory were achieved by aligning the binding sites, considering residues within 8 Å from the pocket, through the align binding sites plugin of Maestro. Pairwise mean-square distance variations were calculated as described in Ref. 10, matrices were plotted with gnuplot.

## 2.2.9 Clustering and Allosteric Sites selection

**Figure 5** shows the workflow that was used to obtain the protein structures object of the screening and computational analysis. Since crystals only provide a static depiction of the polymerase, the 500  $\mu$ s dynamics available on SCoV2-MD.org represented a great mole of additional information concerning our target. The unraveling of the knot begun with the  $C\alpha$  clustering analysis performed with GROMACS, to identify the dynamics-derived structures of SARS-CoV-2 M<sup>Pro</sup> considered as receptors on which Schrödinger's SiteMap was run.



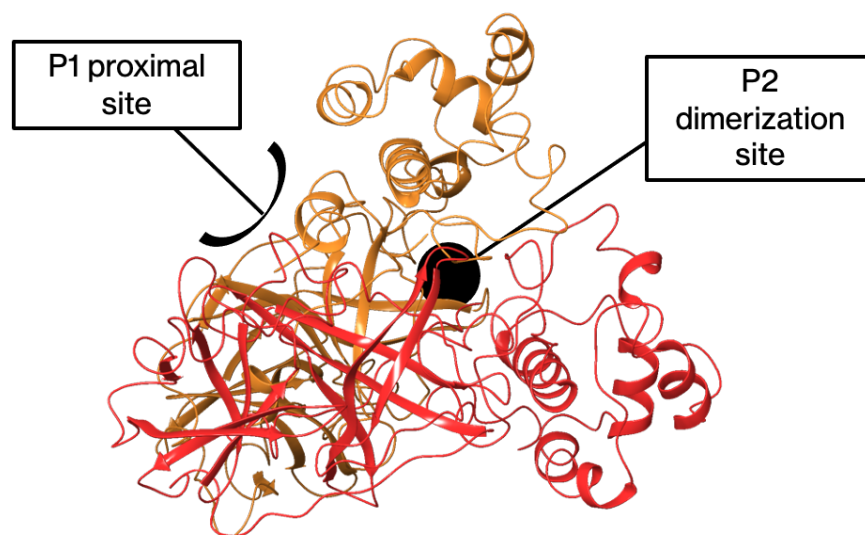
**Figure 5.** Workflow used to obtain the protein structures object of the screening and computational analysis.

According to our hypothesis of inhibiting protease activity through the destabilization of the dimer interface, among the pockets found by SiteMap we selected the ones closer to the protein-protein interface, having the best druggability characteristics in terms of Dscores (Druggability scores generated by SiteMap according to pocket size, hydrophobicity and enclosure) and comprising essential conserved residues. So, we focused our efforts on two interesting region of the protein, what we call “Proximal Site” representing *Pocket 1* (*P1*, **Figure 6**) and the “Dimerization Site” namely *Pocket 2* (*P2*, **Figure 6**).

*P2* presented a slightly lower Dscore,  $D_s = 1.035$ , but bigger volume ( $788 \text{ \AA}^3$ ), more enclosure  $e_x = 0.44$  and higher philic character of  $p = 1.25$  respect to  $p = 0.88$  of *P1*.

Accordingly to the workflow depicted in **Figure 5** we then proceeded with the subsequent side chain clustering analysis, considering for *P1*-A:B residues 4-11, 104-113, 151-159, 290-306 of chain A and 103-128, 141-147 of chain B. The same residues but of opposite chains were selected for *P1*-B:A.

For *P2* residues 1-7, 125-129, 135-141, 238-290 of both chains were selected. By these means we obtained the protein conformations that were used as receptor structures (**Figure 7**) in the docking screening.



**Figure 6.** Representation *P1* and *P2* in the NSP5's dynamics-derived structure. Chain A in red, chain B in orange.

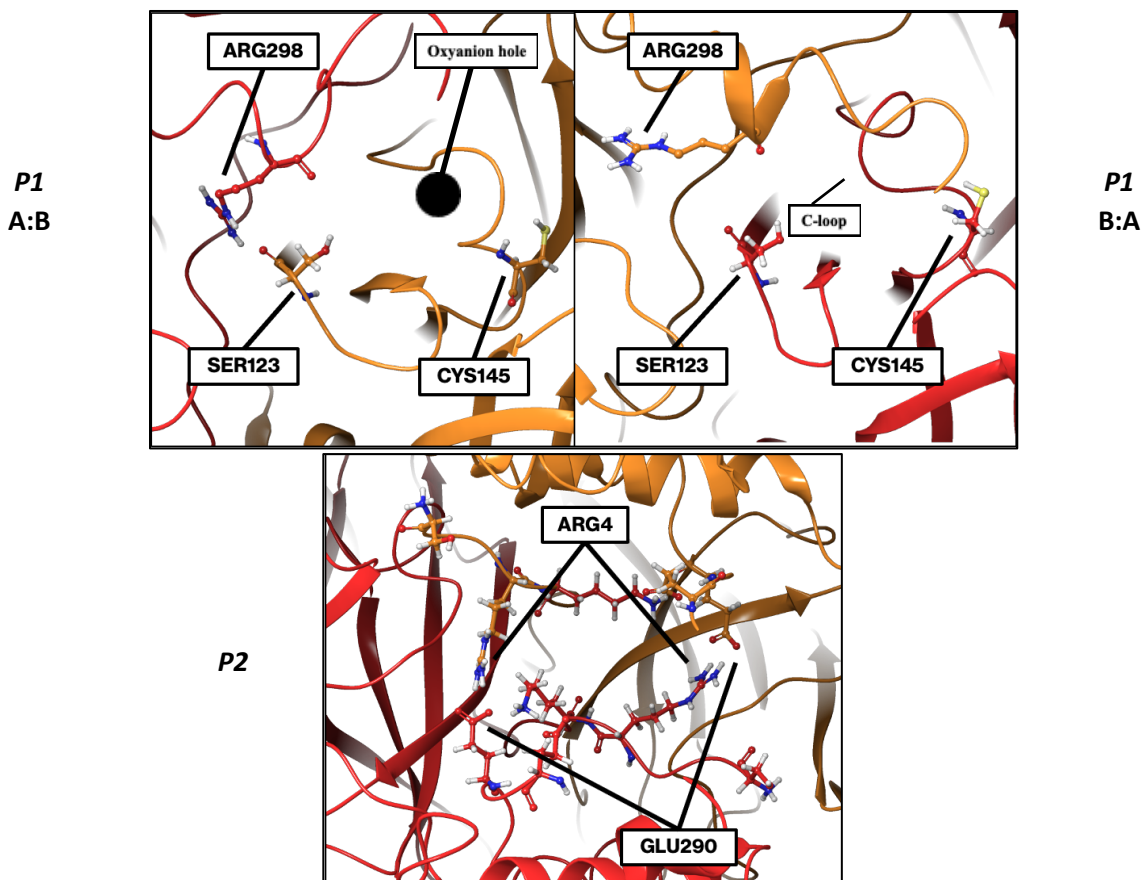
Interestingly, *P1* was found in a different conformation at each chain interface (Chain A:B; chain B:A) (**Figure 7**). The Dscores were similar between the two, *P1*-A:B Ds= 1.059, *P2*-B:A Ds= 1.057 (Pockets that score higher than 1.0 are generally regarded to highly druggable, while sites that score lower than 0.8 are categorized as challenging non-druggable binding sites). The volumes instead pointed to a more disclosed cavity at the B:A interface than at A:B, 624 Å<sup>3</sup> and 431 Å<sup>3</sup> respectively, due to the lifting of Arg298 of chain B loosing the H-bond with Ser123 of chain A. In fact, exposure to the solvent resulted a little higher in *P1*-A:B  $e_x$ = 0.65 against B:A  $e_x$ = 0.6.

*Pocket 1, Proximal Site (Figure 7)*, is a superficial pocket located near the PP interface close to the catalytic site (Cys145). Thanks to the molecular dynamics analysis, two conformations of the pocket were identified, one linked to an active and one to an inactive catalytic site, due to the switch of the C-loop (Residues 138-143) by forming and collapsing the oxyanion hole stabilizing the catalytic intermediate.<sup>6</sup> Interestingly, in SARS-CoV M<sup>Pro</sup> the residue Arg298 uncovered to be exceptionally important for both dimerization and enzymatic activity by point mutation experiments. Replacement of Arg298 by Ala suddenly rendered the enzyme into an inactive and monomeric form.<sup>6</sup>

The second, *Dimerization Site (Figure 7)*, is deeper, symmetric and located at the dimer interface. There are a lot of high-energy hot spots within the dimerization interface. Certain residues (Ser1, Arg4, Lys5, and Glu290) create salt bridges between the two monomers, which are necessary for M<sup>Pro</sup> dimerization.<sup>5</sup>

Our pocket selection was independently validated by structural evidence, consisting of complexes with co-crystallized ligands available from the literature and biochemical data yielding the list of conserved residues mentioned in the introduction.<sup>3,5</sup>

Hence, these structures constitute a starting point for our process of drug discovery.



**Figure 7.** Cartoon representation of *P1* A:B on the left, *P1* B:A on the right and *P2* at the bottom; chain A is colored in red, chain B in orange. In *P1* residues Arg298 and Ser123 are indicated, C-loop and oxyanion hole are also highlighted. In *P2* the salt-bridge forming residues Arg4 and Glu290 are indicated, Ser1 and Lys5 are drawn.

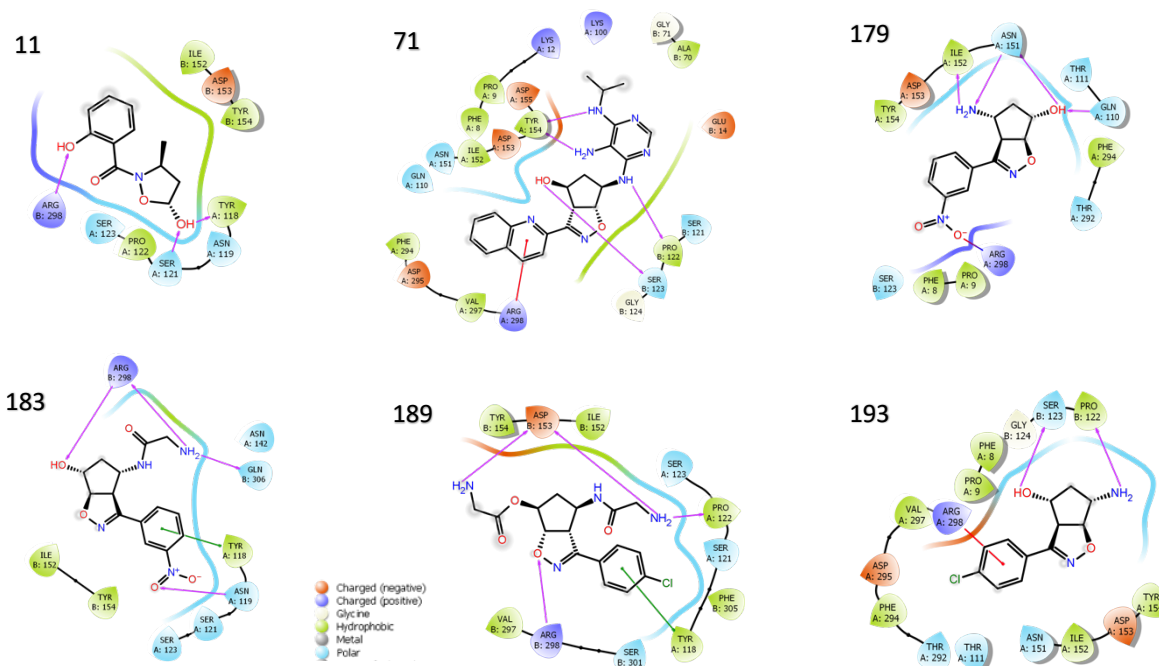
## 2.2.10 Pocket 1 – Docking and Molecular Dynamics

The database generated was examined for possible binders for the previously identified pockets of SARS-CoV-2 M<sup>Pro</sup> by means of a computational strategy combining molecular docking and molecular dynamics simulation to evaluate elicitation of allosteric effects. To this end, both dynamic-derived conformations of *P1* were considered as receptors.

From here on compounds will be numbered accordingly to the respective entry in the database we compiled (i.e. 11, 71, 179, 183, etc..). Each ligand was rigidly docked into *P1*.

Docking poses were selected by docking score, ligand efficiency and interaction with Arg298 given its reported role in dimerization. **Figure 8** shows the LIDs of selected designed compound obtained from the docking screening (Docking score and ligand efficiency are reported in **Table S1** in the Appendix for both receptor structures).

The results show that several designed molecules can establish interactions of various nature with Arg298, such as stacking in the case of quinoline derivatives **71** and the para-chlorine substituted aminol **193**, a salt bridge with the meta-nitro group of aminol **179**, hydrogen bonds also with heteroatoms of the scaffold in the case of AA (amino acid) mono- or bi-substituted aminols **183** and **189**. **Table S1** in the Appendix reports the docking scores and ligand efficiencies of selected binding poses in both *P1* receptor's structure.



**Figure 8.** LIDs in 2D workspace of selected compounds; **71**, **193** and **179** docked in *P1*-A:B; **189** and **183** docked in *P1*-B:A. Magenta arrows: H-bond; red lines:  $\pi$ -cation; green lines:  $\pi$ -stacking; red to blue gradient line: salt bridge.



To further investigate their potential as allosteric inhibitors, the best docking pose of the selected compounds was subjected to 50 ns of MD and a few structural and dynamic parameters were evaluated. To assess the presence of allosteric effects in the protein, we focused on the effects on dimer stability by measuring the distances of Arg298-Ser123, Arg4-Glu290 and Cys145-His41 (picked atoms are specified in *paragraph 2.2.8*) along the dynamics.

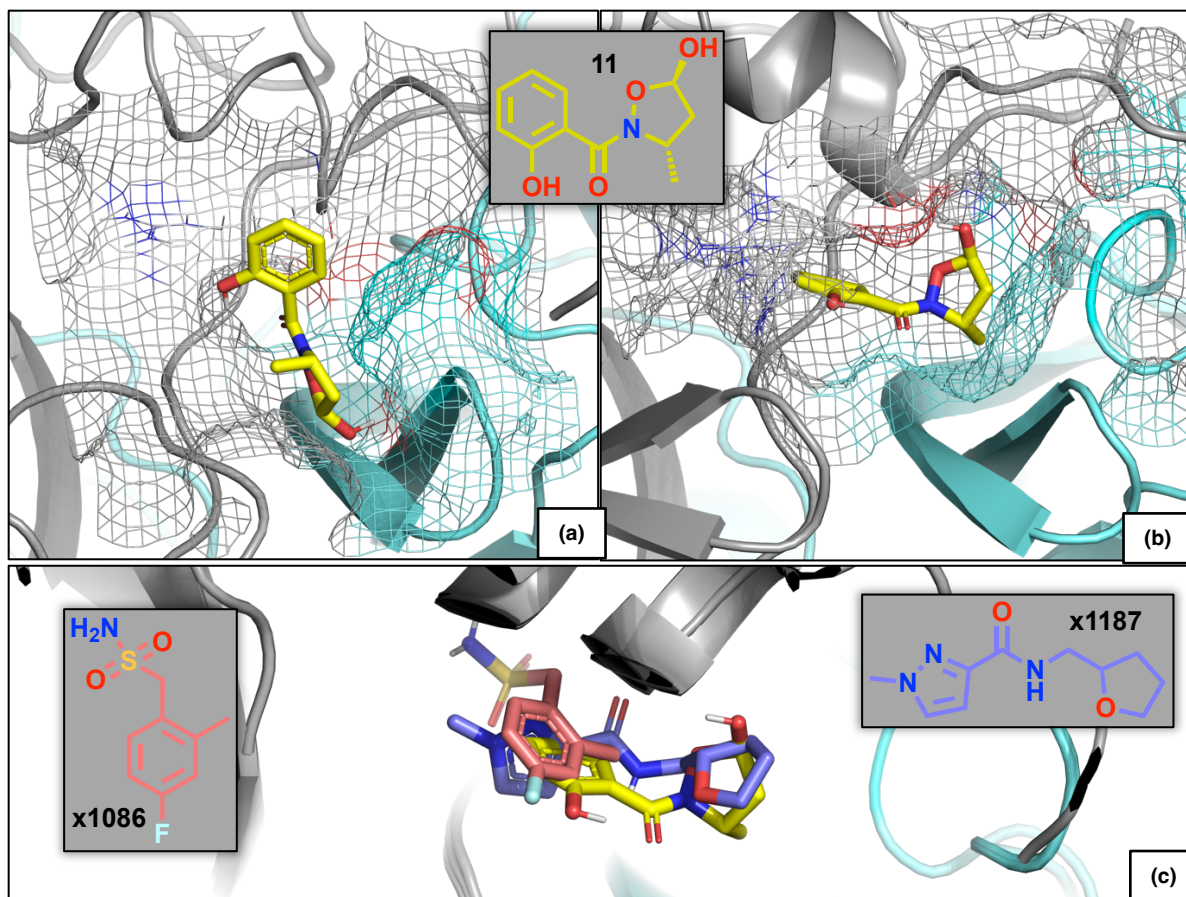
We began our MD analysis starting from compound **11**, which gave encouraging results.

As can be seen in **Figure 9**, upon binding, the isoxazolidine **11**, after an initial rotation, buries itself under Arg298 further revealing the true landscape of *Pocket 1*. A snapshot at the end of the simulation revealed that **11** was able to reproduce the binding pose of two known co-crystallized fragments: a sulfamide **x1086** (PDB: 5RGQ) and 1-Me-pyrazole **x1187** (PDB: 5RFA) derivatives. One of these compounds, the pyrazole one, is reported to disrupt the dimer, leading to a decrease of turnover rate of M<sup>Pro</sup>.<sup>23</sup>

Binding sites of the docking pose and of final MD frame were aligned and volumes calculated for the ligand and for the pocket before and after the simulation (using a square box of 6 Å<sup>3</sup>, centered on compound **11** in the last frame, resolution 0.2 Å).

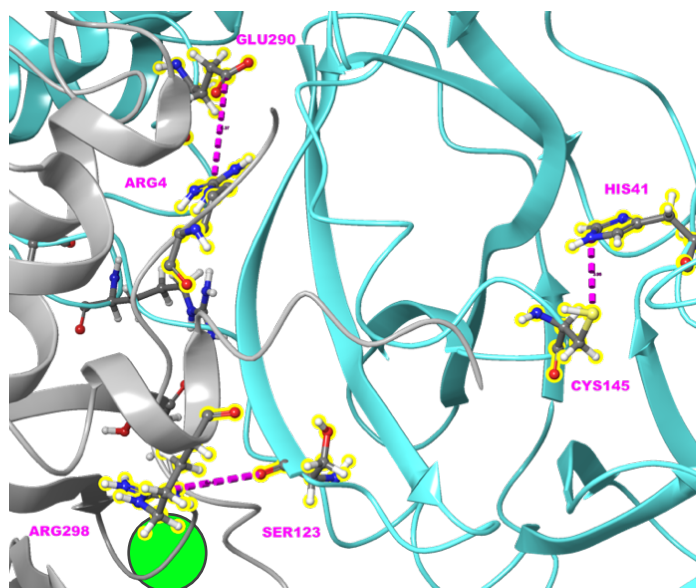
Initially, the volume of the pocket was too small to accommodate the ligand being 138.6 Å<sup>3</sup> compared to 171.9 Å<sup>3</sup> of the ligand. At the end of the trajectory the pocket has widened to 191.3 Å<sup>3</sup>: more than the volume of **11** and more than what was found for the crystal of **x1187** (PDB: 5RFA, 169.5 Å<sup>3</sup>) having a ligand volume of 164.5 Å<sup>3</sup>.

In line with our objective of pinpointing active hits by utilizing allosteric pocket information straight from the analysis of the complex dynamics, the final frame of the trajectory served as a novel receptor structure for docking tests, which were subsequently carried out throughout the entire ligand database.<sup>7</sup> The results are summarized in **Table S2** in the Appendix. After visual inspection, poses of compounds able to better accommodate in the newly formed cavity and having the best ligand efficiency (**Table S2** in Appendix, refer to **Figure S1** for structures) were submitted to a subsequent run of 50 ns MD simulations with Desmond. To assess the presence of allosteric effects on the dimerization state of the protein or on the stability of the catalytic site upon binding of the ligands, three key distances were evaluated, comparing apo state and complexes (**Figure 10**).



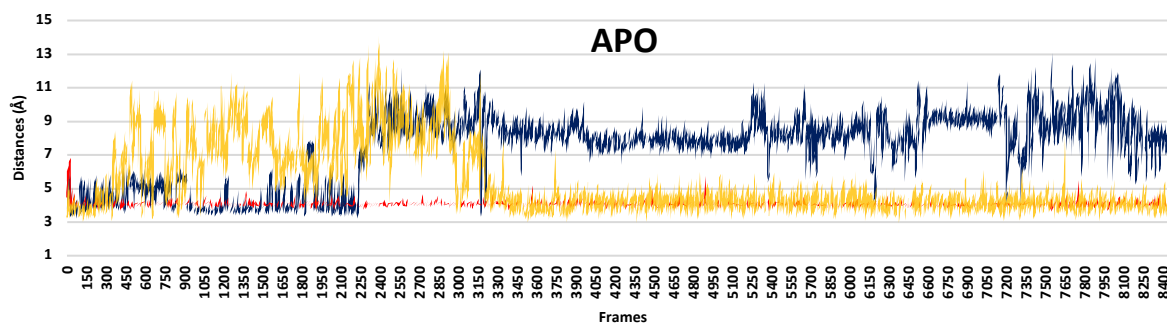
**Figure 9.** a) Docking pose of compound **11**. b) binding mode of **11** at the end of the 50 ns trajectory. c) Superimposition of the complex of **11** with **x1086** (PDB: 5RGQ) and **x1187** (PDB: 5RFA).

In **Figure 9a** the phenolic ring of **11** is placed almost perpendicular to the protein surface with the hydroxyl group pointing toward the lifted Arg298 establishing an H-bond. The amidic carbonyl faces the empty space under Arg298 while the isoxazolidine falls in the valley present at monomers interface, with the hydroxyl group bridging with H-bonds between residues Tyr118 and Ser121 of a turn in chain A. At the end of the trajectory (**Figure 9b**) the H-bond of the phenolic -OH with the side chain of Arg298 is replaced by the donation of an H-bond from the 5-hydroxyl group to the backbone carbonylic group of the Arginine. The polar oxygen atoms of the phenol and the amide are pointing to the outward of the pocket. The aromatic part makes good contacts with Met6, Phe8, Gln299 and the back of Arg298. On the other hand, aliphatic residues are not repelled by the side chains of Met6 of monomer B and interact with monomer A's Ala116, Tyr118, Ser123, Asn142 and the protruding methyl reaching the backbone of Cys177. In **Figure 9c** the positioning of compound **11** reflects the same alternance aromatic-linker-aliphatic that is present in **x1187**. Ligand **x1086** occupies the same region with its aromatic group, but compound **11** lacks the sulfamidic anchor.



**Figure 10.** Outline of the distances evaluated in the analysis. Highlighted residues Arg4 and Arg298 are part of chain B while His41, Ser123, Cys145 and Glu290 belong to chain A. Location of the allosteric site is indicated by the green circle.

We were looking for correlation between fluctuations of these distances with the aim to make predictions about how destabilizing the local environment of *P1* (distance in blue) could lead to effects on the other two (**Graph 1**).



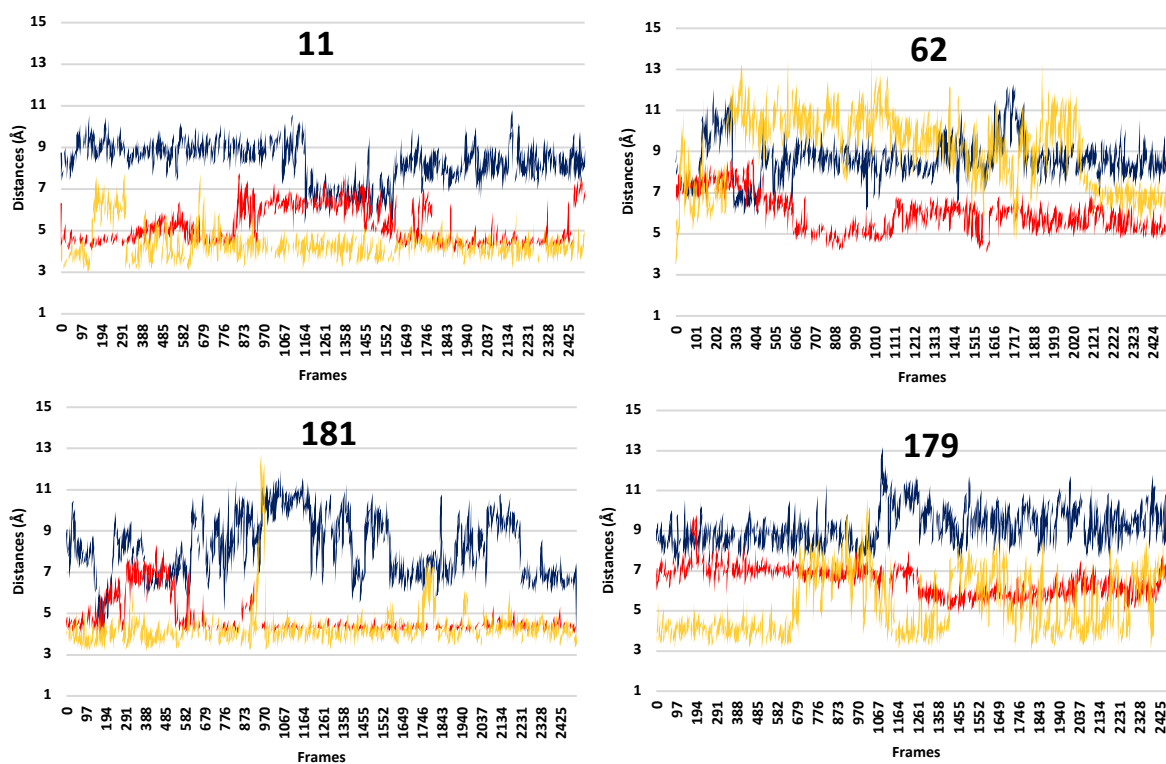
**Graph 1.** Molecular dynamics simulations timeline of key distances between residues Arg298-Ser123 (blue), Arg4-Glu290 (red), Cys145-His41 (yellow) of the apo  $M^{Pro}$ .

The blue distance represents local effects at the dimer interface of *P1* binding site (green circle in Figure 8); red is the distance between two important residues linked to dimer stability and yellow reports on the productive state of the catalytic dyad.

In the APO timeline (500  $\mu$ s, **Graph 1**) after initial rearrangements whilst Arg298 kept fluctuating, residues involved in dimerization and catalytic activity remained stable.

From the timelines of complexes with compound **11** and **62** (**Graph 2**), in the latter more than in the former, effects on all the three distances are visible. Both compounds were able to occupy

the pocket keeping Arg298 in the lifted position (blue distance). For **11** the decrease (Frame 1164, **Graph 2**) corresponds to the rotation of the ligand before reaching the final binding pose (b, **Figure 9**). Ligand **11** was slightly less performant in destabilizing the dimer distance (red) from the starting 4 Å to a mean of 5.2 Å, while for **62** was 0.7 Å higher. The other difference is represented by the significantly higher effects on yellow distance exerted by **62** impacting on the catalytic dyad, keeping Cys145 an average of 9.1 Å away from the helping His41 compared to an average distance of 4.2 Å for isoxazolidine **11**. By these considerations we predict **62** as binder more likely capable of eliciting allosteric effects impairing the protease productivity, than **11**. Differently, compound **71** (refer to **Graph S1** in Appendix), despite having similar initial quinoline positioning to **62**, could not keep a stable face to face p-stacking interactions, but rather edge to face, remaining struck at the entrance of the pocket, with the isoxazoline bicyclic scaffold and pyrimidine substituent that kept fluctuating towards the external. Nevertheless, compound **71** was able to interfere with dimerization stability throughout the course of the simulation (mean Arg4-Glu290 distance: 6.5 Å) but apparently without affecting the catalytic dyad (mean distance Cys145-His41 of 3.9 Å).



**Graph 2.** Molecular dynamics simulations timelines of key distances between residues Arg298-Ser123 (blue), Arg4-Glu290 (red), Cys145-His41 (yellow) of complexes with compounds **11**, **62**, **181** and **179**.

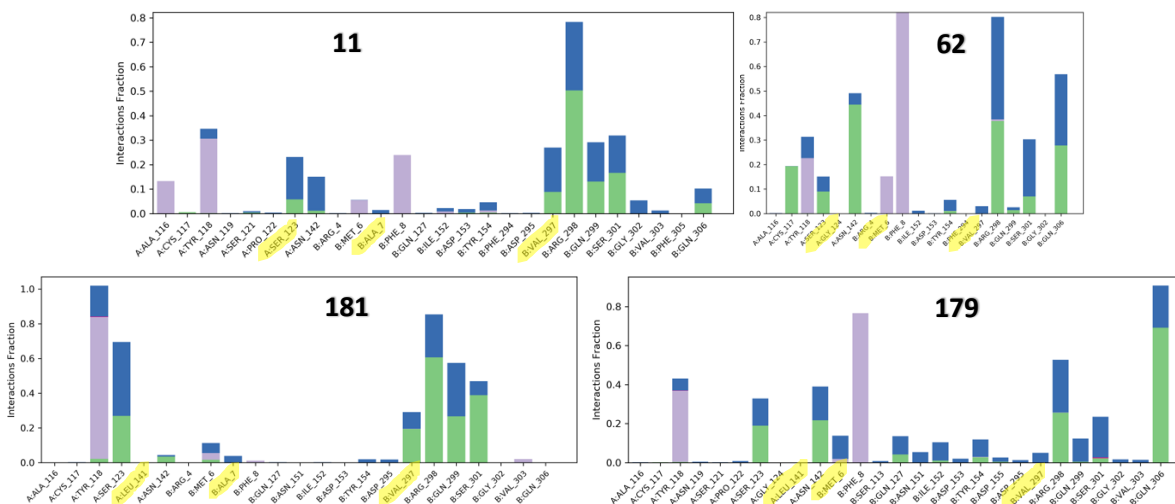
The set is completed by the timeline of compound **181** (**Graph 2**) and **177**, **189** (**Graph S1** in Appendix) that worked as control, allowing validation of the method by confirming that the observed effects were due to the presence of the ligands. Namely, in **181** and **177** timelines as

the ligand shifts off the binding site, at 4.0 ns (Frame 201, **Graph 2**) and at 1.3 ns (Frame 65, **Graph S1** in Appendix) respectively, the original red and yellow distances, following a decreasing trend, were soon after recovered, while Arg298 kept fluctuating like in the APO timeline (**Graph 1**) (**181**: average Arg4-Glu290 4.7 Å, average Cys145-His41 4.3 Å; **177**: average Arg4-Glu290 5.7 Å, average Cys145-His41 3.9 Å).

Compound **189** behaved similarly to **181**, but without ever destabilizing neither dimer interface nor catalytic dyad (Average Arg4-Glu290 4.9 Å, average Cys145-His41 4.5 Å) Arg298 and Ser123 drove near in correspondence of the expulsion of the ligand from the binding site, later than **181** and **177**, at 26.5 ns of simulation (frame 1234, **Graph S1** in Appendix). After that Arg298 kept fluctuating as observed before.

Compound **179** (**Graph 2**) induces moderate allosteric effects on the catalytic site (yellow distance, average 5.3 Å), compared to the higher fluctuations observed for the quinoline aminol **62**, but still perturbing Arg4-Glu290 interactions leading to an average distance of 6.5 Å. Another aminol **193**, like **62** and **179**, induces a mean distance Arg4-Glu290 of 7 Å and dividing Cys145 from His41 an average of 6.5 Å, placing it into the hypothesized active group.

This analysis shows that interacting only with Arg298, maintaining it in the lifted position, is apparently not sufficient to induce effects. To rationalize the observed induced destabilization and relate it to specific ligand interactions with residues having allosteric modulation capabilities, we first compared the protein-ligand interactions fingerprints of the considered active and inactive compounds (**Graph 3** and **Graph S2** in the Appendix).



**Graph 3.** Protein-ligand interaction diagrams of compound **11**, **62**, **181** and **179**. Highlighted in yellow interactions we hypothesized related to activity.

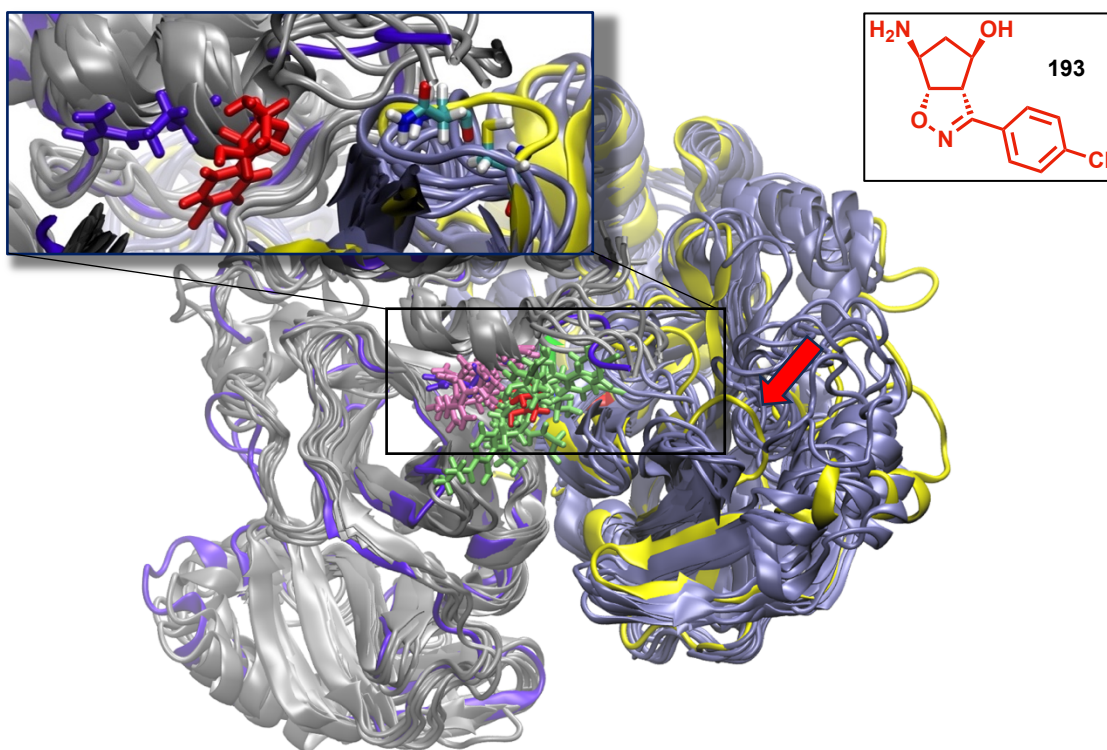
The interactions that stand out from **Graph 3** are the ones involving Asn142 of chain A, Phe8 and Arg298 of the opposite chain. Interestingly, these interactions were present at similar interaction fractions in the compounds predicted to be active (**62**, **179** and **193**; refer to **Graph S2** in the



Appendix). Conversely, they were absent or present in lower percentage in the expected less active or inactive compounds (**11**, **71**, **177**, **181**, **189**; refer to **Graph S2** in the Appendix).

Finally, visual inspection of the dynamics of the evaluated compounds corroborated these findings. Lifting of Arg298 gives access to the pocket, where Phe8, closely connected to the conserved Arg4, is pushed outwardly destabilizing the dimer interface. Furthermore, interactions with Asn142 inhibit the switch of the C-loop (**Figure 7**), preventing formation of the oxyanion hole required for catalysis and by doing so, pulling the entire C-loop away from the helix where His41 is located.<sup>6</sup>

Special consideration must be given to compound **193**. Despite lower interaction fractions (see **Graph S2** in the Appendix) as can be seen from the superimposition of all the trajectories (**Figure 9**), remarkably higher conformational changes appeared in the important C-loop region, bringing the catalytic Cys145 to face Asn142, differently from all the other dynamics of complexes in both catalytic sites and apo. Also, the turn Tyr118-Ser123 of chain A is shifted away from the protein protein interface (red arrow in **Figure 9**).



**Figure 11.** Superimposition of the final frame of the trajectories generated.

In **Figure 11**, for all complexes (except **193**) chain B is gray, chain A is colored iceblue; ligands are in green, Arg298 is mauve. For complex of **193** Arg298 is violet together with chain B, chain A is colored in yellow instead. Compound **193**, in red, occupies a central position respect to all other

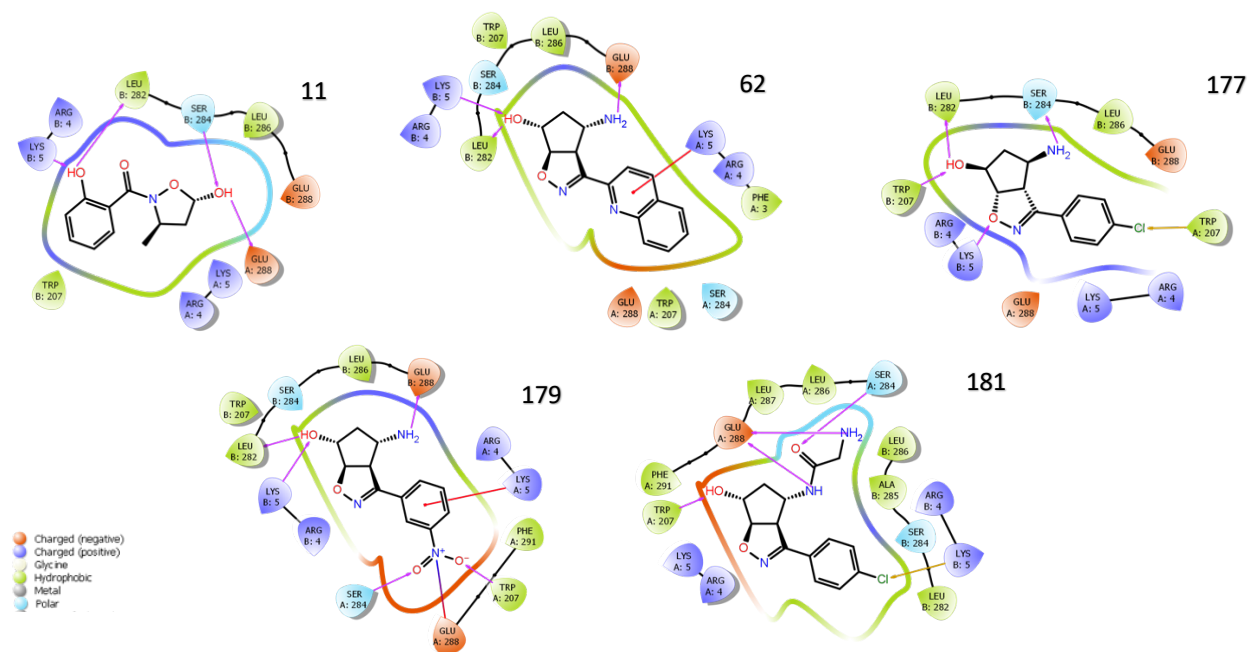
ligands, in green. At top left is reported a magnification of *P1*, residues Asn142 and Cys145 are also shown. It is evident the distortion of the C-loop, in yellow, compared to the other complexes.

Overall, this analysis predicts that among the dataset of compounds we modeled, the designed compound **62**, **193** and to less extent **179**, are the best candidates to inhibit SARS-CoV-2 NSP5 by inducing allosteric effects at the proximal site. We predict that this might lead to disruption of proteolytic activity acting both on dimerization stability (impairing the fundamental salt bridge between Arg4 and Glu290) and catalytic state (inducing variations in the relative position of the aiding His41 and the catalytic Cys145 and moreover distorting C-loop conformation).

## 2.2.11 Pocket 2 – Docking and Molecular Dynamics

The dimerization interface is densely packed with high-energy hot spots. We hypothesized that perturbation of these hot spots could lead the transition of the protein from dimeric to monomeric together with loss of activity.<sup>3-7</sup> For *Pocket 2* the computational approach was similar to the one carried out in *Pocket 1*, a docking screening of the compiled database was conducted at first, followed by molecular dynamics of chosen complexes. The resulting trajectories were analyzed by different means to get insights of conformational effects on the protein driven by interactions with the ligand.

Docking poses were selected by docking score, ligand efficiency and interaction with Arg4, Lys5, Ser10, Gly11, Glu14, Glu288, Glu290 given their reported role in dimerization.<sup>3,5</sup> **Figure 12** shows the LIDs of designed compound selected from the docking screening (Docking score and ligand efficiency are reported in **Table S3** in the Appendix). *Compounds are numbered accordingly to the respective entry in the database we compiled (i.e. 11, 62, 177, etc..).*



**Figure 12.** LIDs in 2D workspace of selected compounds, namely **11**, **62**, **177**, **179**, **181** docked in P2. Magenta arrows: H-bond; red lines:  $\pi$ -cation; green lines:  $\pi$ -stacking; red to blue gradient line: salt bridge; yellow: halogen bond.

Here smaller structures prevail, with an abundance of aminols and all with excellent ligand efficiency. The type of interactions with AA responsible for stabilizing the dimer is of various types. The quinoline aminol **62** has two different interactions, hydrogen bond and  $\pi$ -cation, with each Lys5 of the monomers and an H-bond with Glu288. A network of H-bonds with the heteroatoms of the aminolic scaffold and a halogen bond are present in the case of para-chlorine



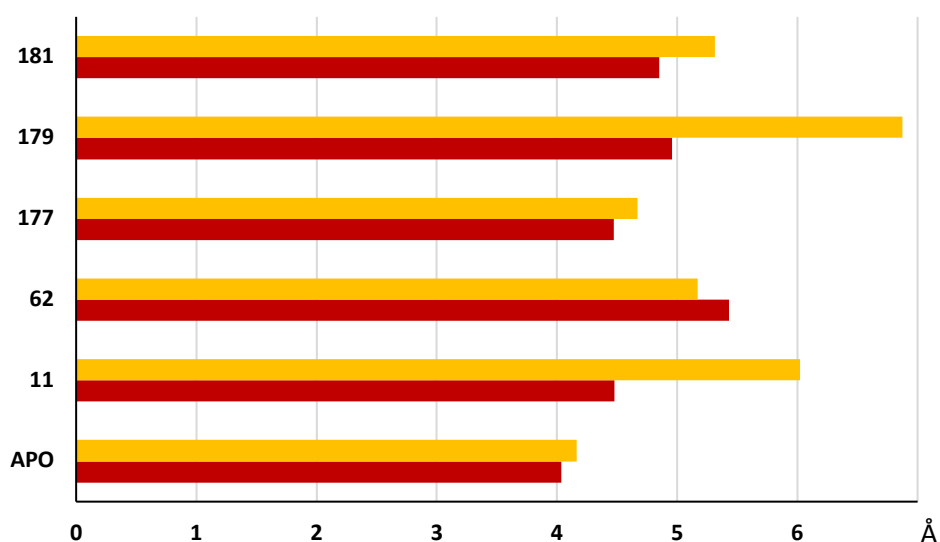
substituted aminol **177**,  $\pi$ -cation and a salt bridge are established by the meta-nitro substituted one **179**. Hydrogen bonds with the aminoacidic residue in case of isoxazoline **181**. Ending with the hydrogen bonds of the 5-hydroxyazolidines **11** with the hydroxyl group derived from the allylic alcohol bridging between Ser284 and Glu288.

The best docking pose of the chosen compounds was submitted to 50 ns of MD, and a few parameters were investigated to further explore their potential as allosteric inhibitors. As before, a couple of distances were evaluated comparing apo and complexes to figure out whether allosteric effects are present on the protein's dimerization state or the stability of the catalytic site after ligand interaction.

In order to estimate the potential consequences of destabilizing *Pocket 2*'s local environment on the distances Arg4-Glu290 and Cys145-His41, we sought for a trend between these distance shifts. The average distances of the residues, considering both couples for each dimer, mentioned before are reported in **Graph 4**.

As previously described, red is the distance between conserved residues linked to dimer stability and yellow stands for the state of the catalytic dyad.

The APO bars (note: frames before Arg298 lifting were removed, **Graph 1**) have the smaller values, as expected, as result the reference mean distances for sustainable catalytic activity are 4.04 Å for Arg4-Glu290 and 4.17 Å between Cys145 and His41.



**Graph 4.** Bar graph of average distances in Å (Cys145-His41, yellow; Arg4-Glu290, red) for each entry.

From **Graph 4** compounds **11** and **179** stand out for their high values for Cys145-His41 distance. Effects are noticeable among the simulation on both inspected distances whilst the ligand was occupying P2. While both compounds differ by a mean of 0.48 Å, in disfavor of **11** in increasing Arg4-Glu290 distance, thus destabilizing the dimer interface, the relative difference of effects considering the catalytic dyad was almost double being of 0.85 Å in advantage of **179**. It is also

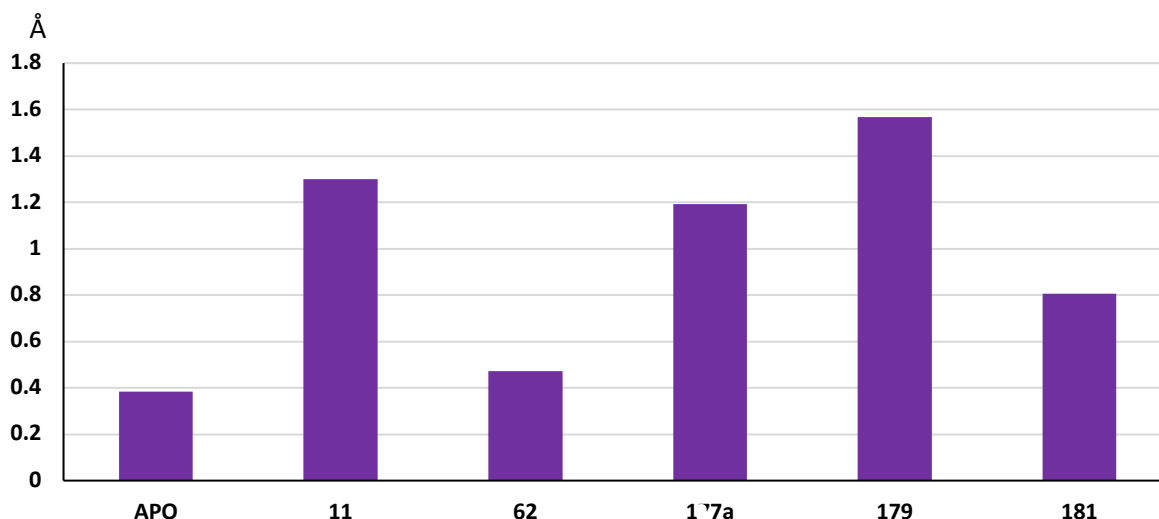
worth to mention that **179** presented the highest separation between the residues involved in catalytic activity, reaching 15.21 Å, while **11** maxed out at 9.88 Å.

In contrast, ligand **177** despite destabilizing the dimer to the same extent as **11** is significantly less able to affect the catalytic dyad (yellow bars, **Graph 4**).

Finally, compounds **62** and **181** behave similarly to each other and neither one reaches **11**'s and **179**'s destabilizing effects. Nevertheless, **62** is the most impacting on Arg4-Glu290 with the highest mean value of 5.17 Å.

This data point out two sets of compounds **11** and **179** in one group, **62**, **177** and **181** in a second one, which differ from the first by a predicted lower impact on catalytic dyad respect to dimer interface destabilization.

To get a global point of view of ligand-dependent perturbations on the dimerization state, we calculated the fluctuations of pairwise amino acid distances between the two monomers in the MD trajectories of complexes and apo. This holds as a starting point in the investigation of the short- and long-ranged ligand-dependent modulation (see Methods); mean values are reported for each entry in **Graph 5**.



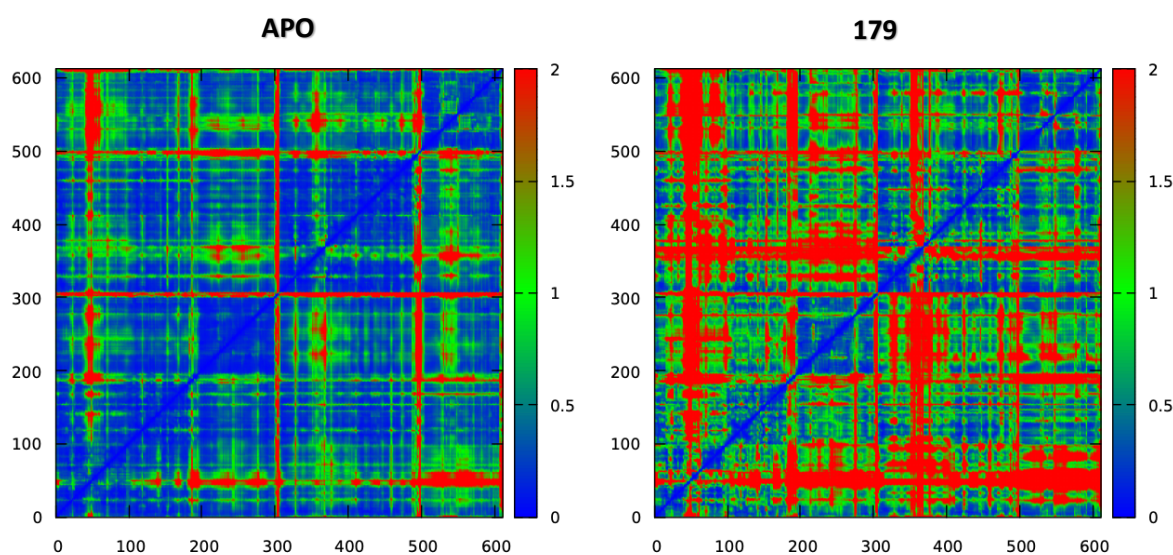
**Graph 5.** Global mean fluctuations of monomer-monomer pairwise amino acid (Å) for each complex.

There are some similarities to the trend observed before (**Graph 4**), especially for compounds **11** and **179**, which caused the highest degree of fluctuations in the dimer (**Graph 5**). **177** and **181** followed but with an inverted rating compared to the previous analysis. Unexpectedly **62** falls behind all of them, with the lowest average of induced fluctuations, even though it was the higher-ranking compound in the docking score (**Table S3** in the Appendix) and also performed well considering the effects on Arg4-Glu290 and Cys145-His41 distances.

**Figure 13** displays the pairwise mean-square distance variations as color-coded residue-residue matrices. All matrices (see **Figure S2** in the Appendix) share a common block character overall, which is indicative of the alternating zones with modest (rigid subdomains) and significant (loops and flexible regions) changes in inter-residue distance fluctuations.

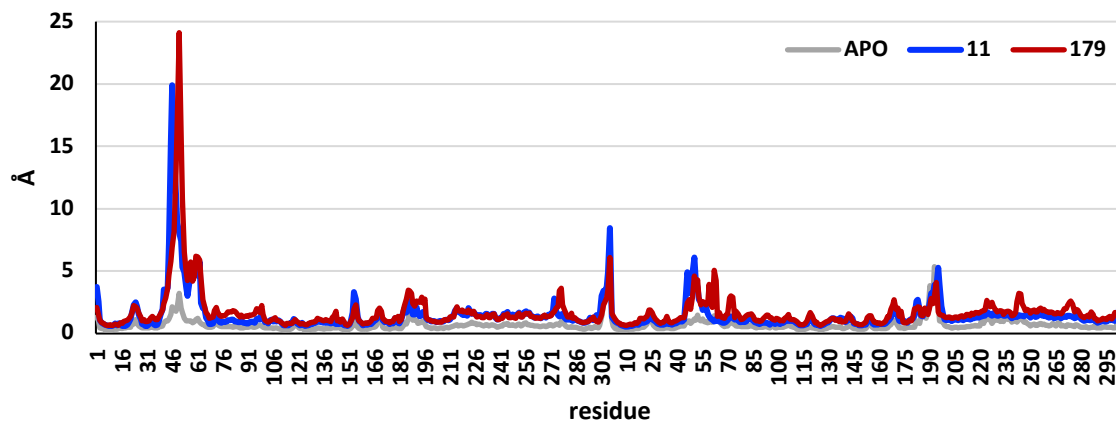
The matrices allow us to map at a residue level where the pair distance perturbations are specifically located.

In **Figure 13**, it is readily seen that the fluctuations of the distances of segments 40-60, 185-200 of each chain (segments are highlighted in **Figure S3** in the Appendix), respectively in the helices containing His41 and C-terminal domain and in the interdomain-linker region, are larger in the **179**-bound-state compared to the apo one, indicating an increased relative motion of the two domains.



**Figure 13.** Distance fluctuation matrices for dimeric NSP5 in apo form on the left, dimeric NSP in complex with **179** on the right.

Protein internal mobility is recapitulated in the absolute magnitude of fluctuations per residue that is modulated between complexes and apo (**Graph 6, Figure S2** in the Appendix). More in detail, the presence of **179** increases the mobility of the catalytically aiding crucial residue His41 compared to the apo, furthermore the entire segment where His41 is located segment 40–60 fluctuates more (average 7.6 Å in **179** and 1.4 Å in the apo), as indicated by higher values of distance fluctuation with respect to the rest of the dimer (**Graph 6**, for the other compounds consult **Graph S3, S4** in the Appendix). Compound **11** closely follows, demonstrating lower segment 40-60 fluctuations (average 7.0 Å), but with the effects more localized on His41 (3.5 Å).



**Graph 6.** Plot of absolute magnitude of pairwise fluctuation per residue (Å) in complexes **11** (blue), **179** (red) and apo (grey). Each value is obtained by averaging the distance fluctuations of the residue with any other residue in the rest of the protein.

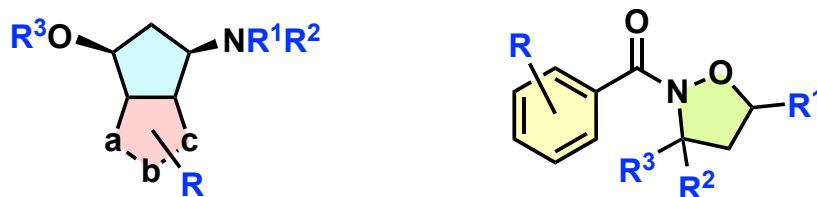
Results for the other compounds can be found in the Appendix, in **Graph S3** and **Graph S4**. Fluctuation average fingerprints of **177** were almost the same of compound **181**, but with higher magnitude. While no compound shows higher peak than that of the apo (5.3 Å) at position Phe-185 of the second chain, located on the linker between domains II and III of the second chain, compound **177** does, presenting an additional signal respect to all the other compound (7.3 Å). This may be due to the drift of the ligand from the starting docking position to the outward, as can be seen from RMSD (**Figure S4** in the Appendix, reported all compounds).

Lastly, compound **62** performed almost as the apo but displaying some small extra sporadic peaks such as the one at Asn274 position, present in both chains, in the zone where the regulatory domain III of each monomer faces each other.

Based on our analysis, it can be concluded that out of the compounds in the database that we modelled, compounds **11**, **179** and **62** (**Figure 12**, this last one being the highest scoring in distancing Arg4 from Glu290) are the most likely to perturb SARS-CoV-2 NSP5 by causing allosteric effects at the dimerization site. This, we predict, might disrupt proteolytic activity by generating perturbations at the protein-protein interface that will be transmitted through each chain, possibly leading to conformational rearrangements that will inactivate the catalytic dyad.

## 2.2.12 Chemical approach to the scaffolds

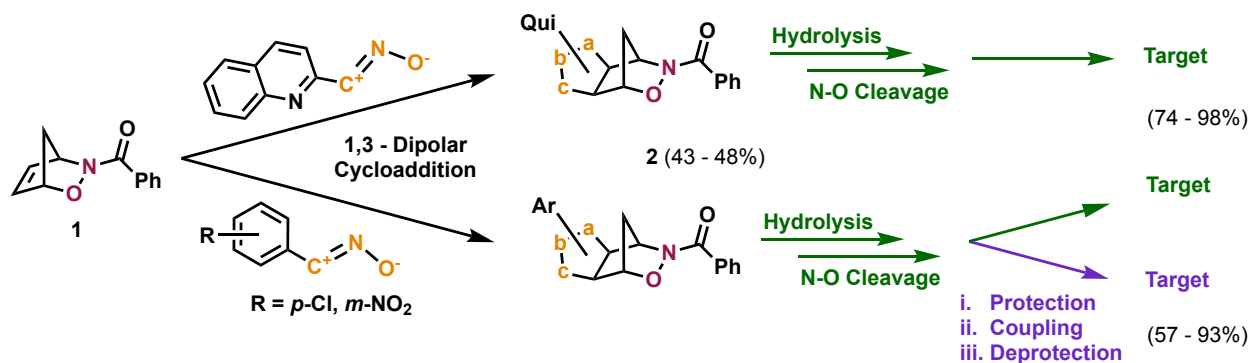
Two recurring scaffolds, **Figure 14**, have emerged from the computational analysis.



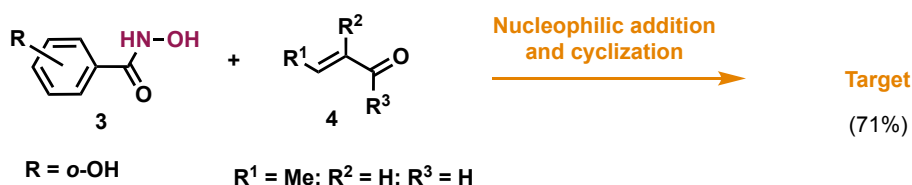
**Figure 14.** Isoxazolidine and cyclopentane-fused isoxazoline scaffold (a,b,c = C=N-O or O=N=C).

Most of the selected entries fall under the group of the cyclopentane-fused isoxazoline scaffold, being this group represented by the different decorations of compounds **62**, **177**, **179**, **181**, **189** and **193**. Compound **11** is the representative of the group of the isoxazolidine scaffold instead.

For the cyclopentane-fused isoxazoline scaffold, similarly to the previous chapter, we found ourselves exploiting the chemistry of pericyclic reactions one more time, but with a different development, this time we focused on the highly variable functionalities that can be introduced on nitrosocarbonyl-derived scaffolds by the regioselective 1,3-dipolar cycloadditions and further synthetic elaborations of the obtainable aminols (**Scheme 1**).



**Scheme 1.** Synthetic pathway toward compounds **62**, **177**, **179**, **181**, **189** and **193**.

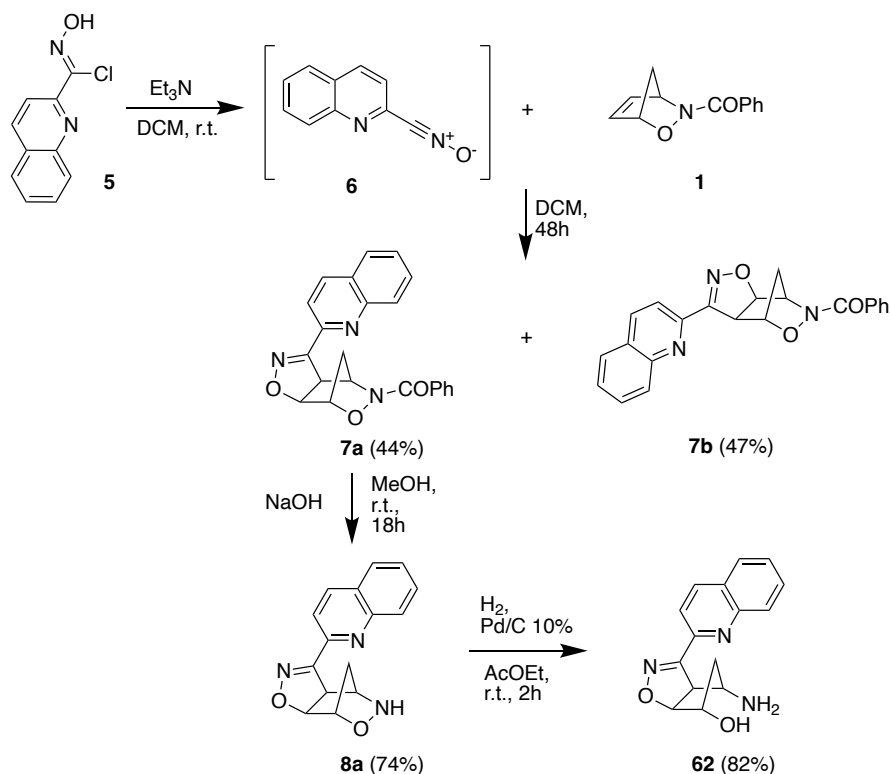


**Scheme 2.** Synthetic pathway toward compound **11**.<sup>11</sup>

The same WP1 route described in **Chapter 1** was used to afford the benzo-nitrosocarbonyl then reacted with cyclopentadiene; The starting regioisomeric compounds of type **2** were prepared, according to the published procedure, from the *N*-benzoyl-2,3-oxazanorborn-5-ene **1** as dipolarophile in the presence of a slight excess of the desired nitrile oxide and isolated in nearly 1:1 ratio each by column chromatography.<sup>13,14</sup>

### 2.2.15.1 Quinoline derivatives

A solution of 2,3-Oxazanorborn-5-ene **1** and Et<sub>3</sub>N was allowed to react with a slight excess (1.2 equiv) of 2-quinolinenitrile oxide **6**, generated in situ from the hydroxymoyl chloride **5** according to the published procedure, in dichloromethane (DCM) solution for 2 days.<sup>13</sup> After this period of time, the residue, obtained after evaporation of the solvent, was submitted to chromatographic separation to isolate the two regioisomeric exo-cycloadducts **7a** and **7b** in 44% and 47% yields, respectively (**Scheme 3**).



**Scheme 3.** Synthesis of compound **62**.

The structures of cycloadducts **7a** and **7b** were determined on the basis of their analytical and spectroscopic data as previously reported.<sup>14</sup> The synthesis of the planned aminol containing the quinoline residue **62** started with benzoyl group detachment from the cycloadduct **7a** through alkaline hydrolysis, which took place in the presence of 1.1 equiv. NaOH in methanol, stirring at

room temperature for 18 h.<sup>24</sup> The hydroxylamine derivatives **8a** was obtained in 74% yield. Hydrogenolysis of derivatives **8a** under standard conditions (H<sub>2</sub>, Pd/C 10%, AcOEt), afforded the desired aminol **62**, in high yields 82%.<sup>24</sup> The structures were established on the basis of their analytical and spectroscopic data as previously stated.<sup>14</sup>

### 2.2.15.2 *para*-Chloro derivatives

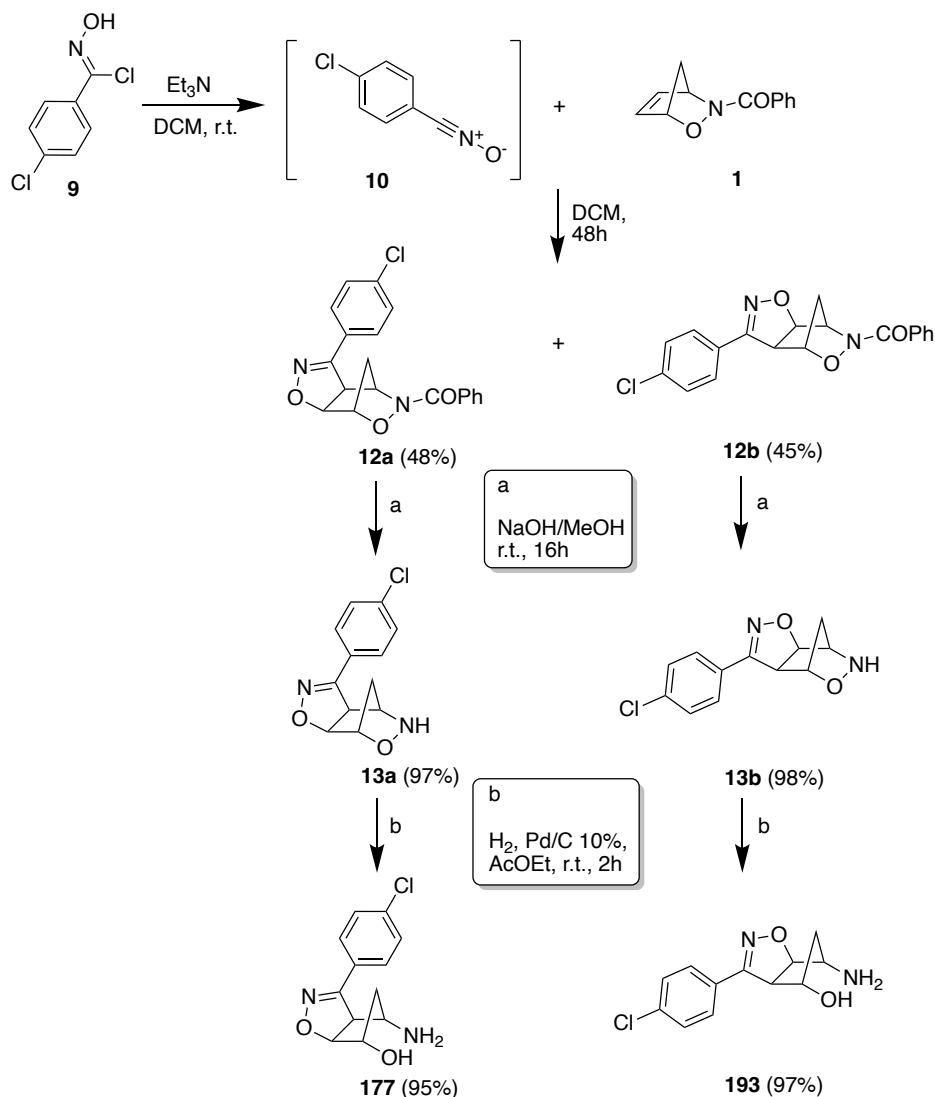
A solution of 2,3-Oxazanorborn-5-ene **1** and Et<sub>3</sub>N was allowed to react with a slight excess (1.2 equiv) of 4-chloro-benzonitrile oxide **10**, generated in situ from the hydroxymoyl chloride **9** according to the published procedure, in DCM solution for 2 days.<sup>13</sup> After this period of time, the residue, obtained after evaporation of the solvent, was submitted to chromatographic separation to isolate the two regioisomeric exo-cycloadducts **12a** and **12b** in 48% and 45% yields, respectively (Scheme 4).

The structures of cycloadducts **12a** and **12b** were determined on the basis of their analytical and spectroscopic data. The <sup>1</sup>H NMR spectrum (CDCl<sub>3</sub>) of **12a** showed the isoxazoline proton H5 at  $\delta = 5.04$  ppm (d,  $J = 8$ Hz), coupled with the H4 proton at  $\delta = 4.32$  ppm (d,  $J = 8$ Hz). The <sup>1</sup>H NMR (CDCl<sub>3</sub>) spectrum of regioisomer **12b** showed the isoxazoline proton signal H5 at  $\delta = 5.22$  ppm (d,  $J=8$ Hz), coupled with the H4 proton at  $\delta = 4.21$  ppm (d,  $J = 8$  Hz). The aromatic protons are clearly shown in the <sup>1</sup>H NMR spectra (CDCl<sub>3</sub>) of both the regioisomeric cycloadducts, with a series of deshielded signals in the expected range.

The following synthetic steps run again through the classical pathway required for the synthesis of 1,3-aminols as we have shown in previous works.<sup>13</sup> Alkaline hydrolysis of the cycloadducts **12a,b**, conducted with grinded NaOH in methanol solutions, afforded the corresponding hydroxylamine derivatives in nearly quantitative yield (Scheme 4).<sup>13,31</sup>

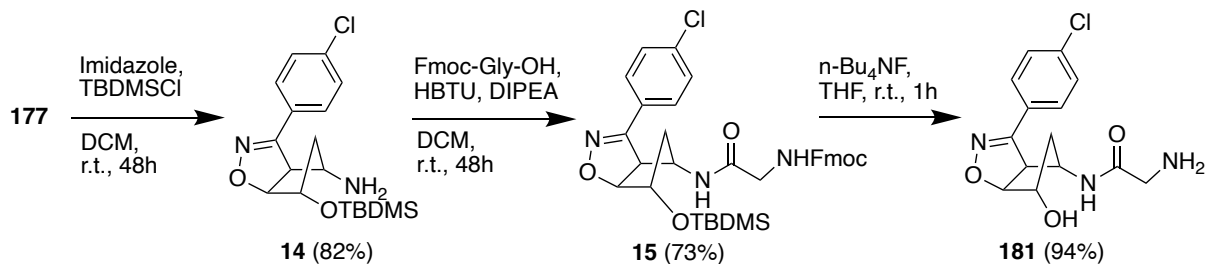
The <sup>1</sup>H NMR spectra (CDCl<sub>3</sub>) show the absence of the 5 proton signals corresponding to the benzoyl group for both **13a** and **13b**. IR spectra show bands at  $\nu_{\text{NH}} = 3209$  cm<sup>-1</sup> and  $\nu_{\text{NH}} = 3243$  cm<sup>-1</sup>, respectively, corresponding to the NH protons, further attesting to the benzoyl group detachment. Hydrogenolysis of the hydroxylamine derivatives under standard conditions (H<sub>2</sub>, Pd/C 10%, AcOEt) afforded the desired aminols **177** and **193** (yields: 95 and 97%, respectively, Scheme 4).<sup>13,31</sup> The reaction time must be carefully contained within two hours because a prolonged hydrogenation process was detrimental to the stability of the aromatic moiety and a plethora of degradation products was observed in the reaction mixtures. The structures of **177** and **193** were confirmed from their analytical and spectroscopic data. In particular, the <sup>1</sup>H NMR spectra (CDCl<sub>3</sub>) displayed a more complex fine structure with couplings due to the increased conformational mobility of the cyclopentane moieties following strain relief after the hydrogenolytic ring opening. Significantly, the NH<sub>2</sub> and OH groups IR spectra furnished signals, respectively, at  $\nu_{\text{NH}_2} = 3336$  cm<sup>-1</sup>, 3280 cm<sup>-1</sup> and  $\nu_{\text{OH}} = 3041$  cm<sup>-1</sup> for compound **177** while for

compound **193** the corresponding groups gave bands at  $\nu_{\text{NH}_2} = 3347 \text{ cm}^{-1}$ ,  $3284 \text{ cm}^{-1}$  and  $\nu_{\text{OH}} = 3094 \text{ cm}^{-1}$ .



**Scheme 4.** Synthesis of the aminols **177** and **193**.

Compound **177** was further elaborated through established procedures that quickly allowed obtainment of compounds **181** and **189**.<sup>25</sup> Those steps involved -OH protection, coupling with Fmoc-Gly-OH followed by removal of both protecting groups to obtain **181** (**Scheme 5**).



**Scheme 5.** Synthesis of the glycine derivative **181**.



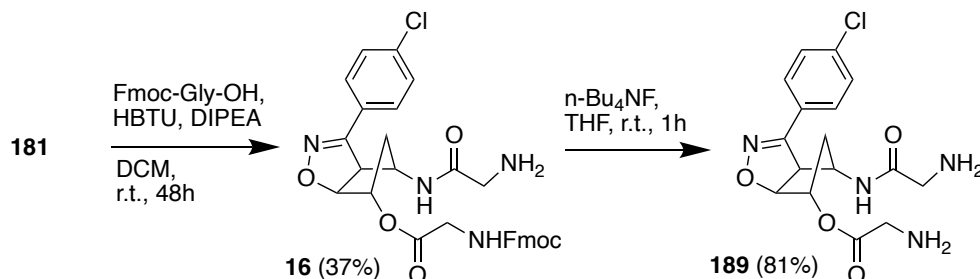
The hydroxy group protection was performed under standard procedure affording the amine **14** (82%) that was fully characterized (**Scheme 5**).<sup>24</sup>

The appearance of shielded signals in the <sup>1</sup>H-NMR spectra (CDCl<sub>3</sub>) at  $\delta = 0.15 - 0.93$  of the hydrogens of tert-butyl-dimethyl silyl group and the absence of the -OH band in the IR spectra confirmed the good outcome of the protection.

Compound **14** was coupled with the commercially available Fmoc-Gly-OH and the adduct **15** was isolated in good yields (73 %) and fully characterized. Fmoc signals crowded the aromatic region of the <sup>1</sup>H-NMR spectra (DMSO-*d*<sub>6</sub>), the IR spectra displayed two carbonyl stretching peaks at  $\nu_{\text{CO amide}} = 1681 \text{ cm}^{-1}$  and  $\nu_{\text{CO ester}} = 1703 \text{ cm}^{-1}$ .

The hydroxy functionalities deprotection was secured by standard *n*-Bu<sub>4</sub>NF treatment; during this step the Fmoc protection is also removed without detrimental effects in the subsequent synthetic step leading to the obtainment of compound **181** (**Scheme 5**) in good yields (88 %) which was fully characterized. The Fmoc and TBDMS signals disappeared from the <sup>1</sup>H-NMR spectra (DMSO-*d*<sub>6</sub>) while the IR signals mutated from the sharp signals of -NH groups to a broad band peaking at  $\nu = 3310 \text{ cm}^{-1}$  comprehending all signals of the groups -OH -NH and NH<sub>2</sub>.

Consequently, the coupling and deprotection steps were repeated to synthesize **189** (**Scheme 6**).



**Scheme 6.** Synthesis of the double glycine-substituted derivative **189**.

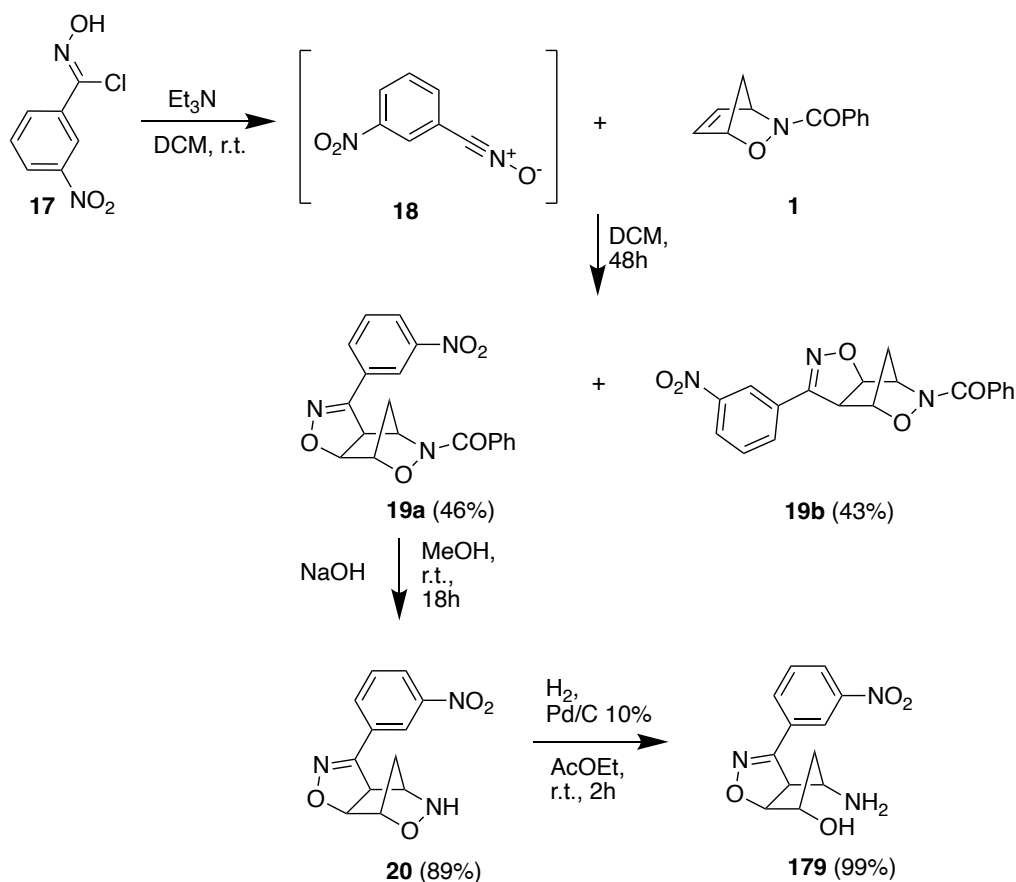
All the compounds reported were separated from unreacted starting materials or decomposition products and fully characterized. The adduct was obtained in fair yields (37%), the appearance in the <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) of Fmoc signals in the aromatic range and of the glycine CH<sub>2</sub>-NHCO, together with the disappearance of the broad signal of the -OH group from the IR, indicated the good outcome of the reaction. In the next step Fmoc removal with *n*-Bu<sub>4</sub>NF treatment led to the obtainment of the target compound **189** in good yields (81 %). Fmoc signals were no longer present in the <sup>1</sup>H NMR spectra (DMSO-*d*<sub>6</sub>) and coherently was instead appreciable the broad signal at  $\delta = 3.35 \text{ ppm}$ , integrating for four protons, indicating NH<sub>2</sub> groups. The IR spectra accordingly displayed a broad band peaking at  $\nu = 3235 \text{ cm}^{-1}$  comprehending signals of the groups -NH and NH<sub>2</sub>.

### 2.2.15.3 *meta*-Nitro derivatives

A solution of 2,3-Oxazanorborn-5-ene **1** and Et<sub>3</sub>N was allowed to react with a slight excess (1.2 equiv) of 3-nitro-benzonitrile oxide **18**, generated in situ from the hydroxymoyl chloride **17** according to the published procedure, in DCM solution for 2 days.<sup>13</sup> After this period of time, the residue, obtained after evaporation of the solvent, was submitted to chromatographic separation to isolate the two regioisomeric exo-cycloadducts **19a** and **19b** in 43% and 46% yields, respectively (**Scheme 7**).

The structures of cycloadducts **19a** and **19b** were determined on the basis of their analytical and spectroscopic data. The <sup>1</sup>H NMR spectrum (CDCl<sub>3</sub>) of **19a** showed the isoxazoline proton H5 at  $\delta$  = 5.13 ppm (d, *J* = 8 Hz), coupled with the H4 proton at  $\delta$  = 4.40 ppm (d, *J* = 8 Hz). The <sup>1</sup>H NMR spectrum (CDCl<sub>3</sub>) of regioisomer **19b** showed the isoxazoline proton signal H5 at  $\delta$  = 5.30 ppm (d, *J* = 8 Hz), coupled with the H4 proton at  $\delta$  = 4.28 ppm (d, *J* = 8 Hz).

The *m*-nitro-benzo aromatic protons are clearly shown in the <sup>1</sup>H NMR spectra (CDCl<sub>3</sub>) with a series of deshielded signals in the expected range and also the IR band at  $\nu_{\text{NO}_2}$  = 1521 cm<sup>-1</sup>.

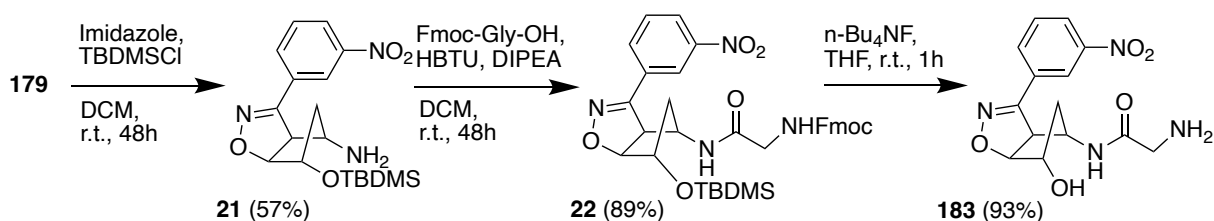


**Scheme 7.** Synthesis of the aminol **179**.

The synthesis of the selected aminol containing the meta-nitro group **179** started with benzoyl group detachment from the cycloadduct **19a** through alkaline hydrolysis, which took place in the

presence of 1.7 equiv. NaOH in methanol, stirring at room temperature for 18 h. The hydroxylamine derivatives **20** was obtained in 78% yield, and fully characterized. The  $^1\text{H}$  NMR spectra ( $\text{CDCl}_3$ ) show the absence of the five protons corresponding to one aromatic ring, and the presence of a broad signal at  $\delta = 4.79$  originating from the NH proton. The new functional group is also clearly shown in the IR spectra  $\nu_{\text{NH}} = 3235 \text{ cm}^{-1}$ . Hydrogenolysis of derivatives **20** under standard conditions ( $\text{H}_2$ , Pd/C 10%, EtOAc), afforded the desired aminol **179**, in high yields 91%.<sup>24</sup> The structure was established on the basis of their analytical and spectroscopic data. In particular, in the  $^1\text{H}$  NMR ( $\text{CDCl}_3$ ) spectra the isoxazoline signals were not affected by hydrogenation:  $\delta = 4.01$  (d,  $J = 8\text{Hz}$ , H4), 5.39 ppm (d,  $J = 8\text{Hz}$ , H5). Cleavage of the bicyclo[2.2.1] ring affords the OH and  $\text{NH}_2$  groups on the cyclopentane spacer, however, it was not easy to find both these two groups in the proton spectra recorded in  $\text{CDCl}_3$ , being seen just the alcoholic group as broad signal at  $\delta = 5.15$  ppm whereas they were pretty evident in the IR spectra at  $\nu_{\text{NH}_2} = 3358 \text{ cm}^{-1}$ ,  $3291 \text{ cm}^{-1}$  and  $\nu_{\text{OH}} = 3086 \text{ cm}^{-1}$ .

Compound **179** was further elaborated through the procedure used before, those steps involved -OH protection, coupling with Fmoc-Gly-OH followed by removal of both protecting groups to obtain **183** (Scheme 8).



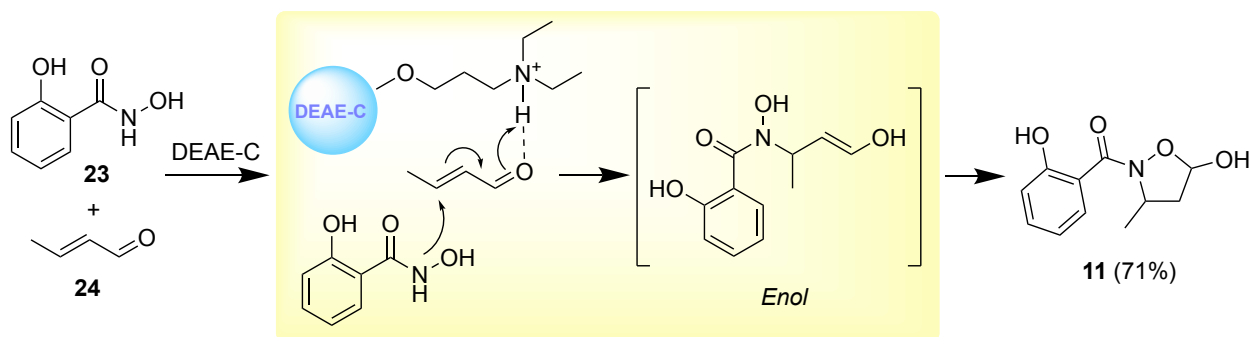
### Scheme 8. Synthesis of the glycine derivative **183**.

The hydroxy group protection was performed under standard procedure affording the amine **21** (57%) that was fully characterized (Scheme 8).<sup>25</sup> The appearance of shielded signals in the  $^1\text{H}$ -NMR spectra ( $\text{CDCl}_3$ ) at  $\delta = 0.11 - 1.03$  of the hydrogens of tert-butyl-dimethyl silyl group and the absence of the -OH band in the IR spectra confirmed the good outcome of the reaction. Compound **21** was coupled with the commercially available Fmoc-Gly-OH and the adduct **22** was isolated in fair yields (57 %) and fully characterized. Fmoc signals crowded the aromatic region of the  $^1\text{H}$ -NMR spectra ( $\text{CDCl}_3$ ), the IR spectra displayed two carbonyl stretching peaks at  $\nu_{\text{CO}}$  amide =  $1722 \text{ cm}^{-1}$  and  $\nu_{\text{CO}}$  ester =  $1674 \text{ cm}^{-1}$ .

The hydroxy functionalities deprotection was secured by standard  $n\text{-Bu}_4\text{NF}$  treatment; during this step the Fmoc protection is also removed leading to the obtainment of compound **183** (Scheme 8) in good yields (93 %) which was fully characterized. The Fmoc and TBDMS signals disappeared from the  $^1\text{H}$ -NMR spectra ( $\text{DMSO-}d_6$ ) and the -OH group was evident at  $\delta = 5.77$  ppm. the IR signals broadened and shown additional peaks at  $\nu_{\text{NH}_2} = 3343 \text{ cm}^{-1}$ ,  $3295 \text{ cm}^{-1}$ ;  $\nu_{\text{OH}} = 3086 \text{ cm}^{-1}$ .

### 2.2.15.1 5-hydroxy isoxazolidine derivative

The synthesis was conducted accordingly to a recent published method which involves conducting the reaction in the solid state with DEAE-C (Diethylaminoethyl cellulose) by grinding/mixing in a mortar the hydroxamic acid **23** and the  $\alpha,\beta$ -unsaturated aldehyde **24** together with DEAE-C with pestle in a mortar, affording the 5-hydroxy-substituted isoxazolidine **11** in good yields.<sup>11</sup>



**Scheme 9.** Synthesis and proposed mechanism to the 5-hydroxy-isoxazolidine **11**.<sup>11</sup>

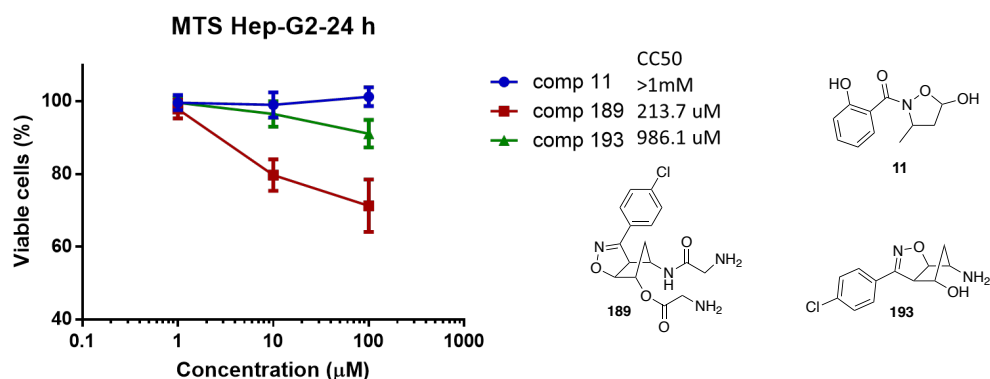
Due to its positive charge, DEAE-C has the ability to activate unsaturated aldehydes or ketones via hydrogen bonding, rendering them susceptible to nucleophilic addition by the weakly donating nitrogen atom of the hydroxamic acid (**Scheme 9**, yellow box). This results in the formation of the Enol intermediate, which is recognized as the immediate precursor to isoxazolidine **11**. The exchange of protons among different intermediates restores the characteristic nature of DEAE-C, allowing it to initiate the cyclization process once again.<sup>11</sup>

The isoxazolidine derivatives **11** was fully characterized and resulted identical to previously synthesized samples available in our laboratory.<sup>11</sup>

## 2.2.16 Biological Assays

Samples of the synthesized compounds representative of the scaffold families cyclopentane-fused isoxazoline and isoxazolidine (see **Figure 14**), are being biologically assayed by the Messina group (Prof. Francesca Marino-Merlo) to preliminarily assess their potential as antiviral compounds capable of inhibiting SARS-CoV-2's M<sup>Pro</sup> activity.

For this purpose, first their cytotoxicity was assayed in terms of inhibition of cell viability and of cell metabolic activity in epithelial HepG2 cells. The cell metabolic activity was detected through formazan product formation, using a commercial colorimetric kit (MTS [3,4-(5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium salt]). Concerning the cell viability, a dose effect assay using three different concentrations of the compounds, within the range from 100 to 1  $\mu$ M, assessed levels of cytotoxicity, expressed as CC<sub>50</sub> (**Figure 15**).



**Figure 15.** Evaluation of the metabolic activity assessed by the MTS assay in HepG2 cells treated with the M<sup>Pro</sup> inhibitors **11**, **189**, **193**. Data refer to three experiments performed in duplicate.

Results showed that **189** was endowed with a higher ability to induce cell death with respect to **11** and **193** toward cell line tested, with CC<sub>50</sub> values of about 214  $\mu$ M for HepG2 cells (**Figure 15**) but none of them displayed high levels of cytotoxicity as the compounds designed to target the polymerase (*Chapter 1*) demonstrated.

Overall, concerning the cell viability, preliminary tests, assessed low levels of cytotoxicity this represents a promising initial result that points toward the absence of off-target activity of the proposed compounds.

Inhibitory tests are ongoing, new result will be implemented as soon as possible.

## 2.3 Discussion

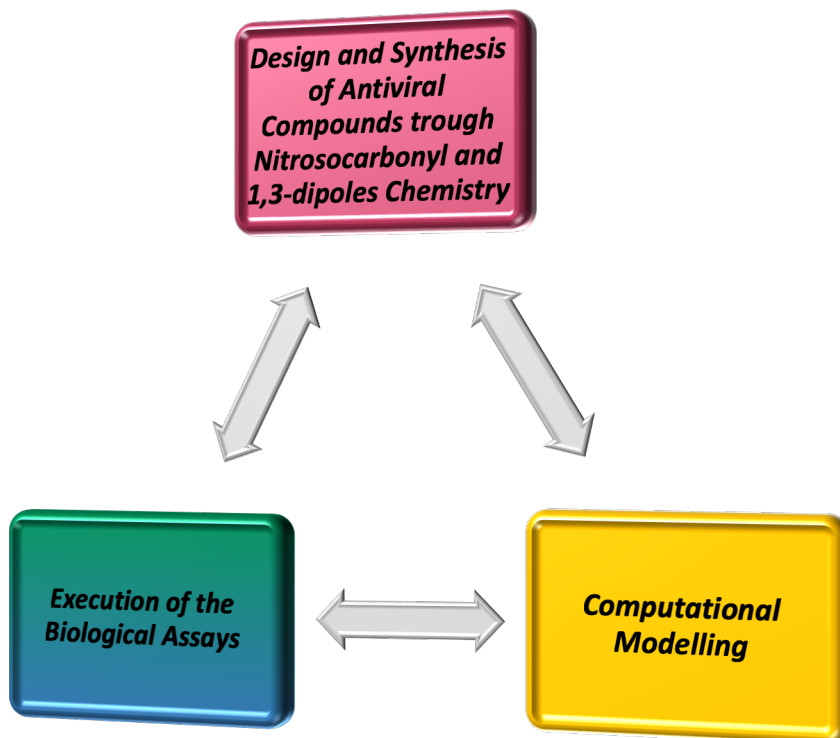
In the present work we have designed and synthesized possible M<sup>Pro</sup> allosteric inhibitors exploiting the publicly accessible Molecular Dynamics data on SCoV2-MD and employing the chemical pathways developed by our Pavia's research group.<sup>9</sup> The protein is a dimer in solution because of the stable pairing of extremely conserved motifs, some of which could make good targets. Clustering analysis on C $\alpha$  assisted by SiteMap runs, to rank the putative binding sites on the protein surface, led us to identify the most druggable allosteric pockets of SARS-CoV-2 M<sup>Pro</sup>. The comparison between our findings and the literature led us to validate our selection of allosteric sites whose regulation (perturbation) may be able to impact on protein-protein interactions and destabilize the active site.<sup>3,5,6</sup> Furthermore, a second clustering analysis focusing on the side chains that define the pockets was fundamental to obtain receptor conformations suited for the drug design phase. We created a database of differently decorated scaffolds to be computationally evaluated for their potential inhibitory activity. The docking screening was the first step in the selection of candidate molecules to be synthesized. The accordance between MD simulations of our ligand-protein modelled complexes, conducted with Desmond, and the crystal data available for complexes (PDB: 5RFA, 5RGQ) have confirmed the consistency of our strategy. The computational studies revealed the possibility for two recurring scaffolds, namely the cyclopentane-fused isoxazoline and the isoxazolidine one both opportunely decorated, to give positive responses in terms of inhibiting SARS-CoV-2 NSP5. At the head of former scaffold group we found mostly aminols in forms of compounds **62**, **177**, **179** and **193**. Compounds **181**, **189** expanded this group including mono- or bi-substituted glycine isoxazoline derivatives. The last scaffold was represented by the 5-hydroxy isoxazolidine derivative **11**. In principle, these compounds could recognize allosteric NSP5 sites and interact with sites other than those occupied by Nirmatrelvir and other peptidomimetics.

The reported synthetic pathway towards targeted potential compounds that could operate as novel NSP5 protease function antagonists was found to be robust and reliable as well-documented in several papers and reviews. All the nitrile oxides **6**, **10**, **18** displayed a remarkable reactivity, and their trapping by the dipolarophile *N*-benzoyl-2,3-oxazanorborn-5-ene **1** to the corresponding regioisomeric cycloadducts intermediates provided a unique opportunity to use the formers as keys 1,3-dipole building-blocks allowing introduction of different aromatic or heterocyclic substituents, syn and anti, bearing edge or ewg functional groups (quinoline, *para*-chlorine and *meta*-nitro substituted benzene).<sup>15</sup> Single or double amino acid residues were then introduced on the selected aminols through an already used chemistry, which allowed the synthesis to be developed quickly.<sup>25</sup> In term of yields, all the synthetic steps gave from fair to very good results, and the quality of the products was found to be encouraging. Finally, compound **11** was synthesized taking advantage of the use of an unconventional absorbent based on DEAE-cellulose (DEAE-C) that is able to promote the cyclization of hydroxamic acids on  $\alpha,\beta$ -unsaturated

aldehydes. The reaction proceeded smoothly affording the 5-hydroxy-isoxazolidine **11** in good yields and, more intriguingly, in extremely mild way at room temperature and in the solid phase.<sup>11</sup>

Overall, we exploited synthetic methods for producing a variety of heterocyclic scaffolds, on which different functionalities can be introduced in pivotal steps both regioselectively and stereospecifically. Hence, discovering and targeting allosteric sites, relying on the here presented versatile synthetic methodology, can lead us to the expansion of the chemical space of leads and to new classes of drugs.

Therefore, we hope that this computationally aided research of allosteric SARS-CoV-2 inhibitors may shed new light on the applications of the chemistry of pericyclic reactions!



## 2.4 Conclusion

In conclusion, this scientific investigation promises a significant advancement in the battle against SARS-CoV-2 by unveiling a strategic approach that seamlessly integrates computational analyses with synthetic chemistry. The discovery of allosteric pockets within the SARS-CoV-2 main protease, achieved through computational methods and corroborated by literature data, has illuminated a promising avenue for targeted antiviral intervention.<sup>2-6</sup> The subsequent design and synthesis of novel antiviral molecules, harnessing the distinctive chemistry developed by our research group, showcase a pioneering effort in structure-based drug design. As a matter of fact, identifying and focusing on allosteric sites can lead to the discovery of new therapeutic classes and the expansion of lead compounds chemical space. Most crucially, the identification of novel compounds that target allosteric regions offered a workable approach to the quest for novel inhibitors of protein-protein interactions. Our goal, though, was to find active hits by employing data on an allosteric pocket that we directly gathered from the analysis of a dynamic molecular machine. In fact, the key to our strategy of discovery of inhibitors of protein-protein interactions or of possible drug-binding sites was to explicitly take into account the flexibility of the target. Protein's molecular dynamic data can be used to reveal potential binding states of the receptor that may exist on the protein's energy landscape but may not be readily visible in a single-structure depiction. Thus, the identification of alternative states can lead to the identification of potential allosteric binding sites, permit the occupancy of a site by structurally distinct ligands, or direct efforts towards the construction of novel leads that functionally and structurally modify established leads to target-specific receptor geometries. As we navigate the dynamic landscape of viral infections, the collaborative synergy between computational efforts and synthetic chemistry showcased in this study emphasizes the importance of a holistic approach. The deliberate exploitation of pericyclic reactions, ultimately in the form of 1,3-dipolar cycloadditions and catalyzed nucleophilic addition to  $\alpha,\beta$ -unsaturated aldehydes, in the synthetic process adds a layer of renewal, enabling the creation of compounds with tailored structures designed to interact specifically with the identified allosteric sites. With this approach we not only demonstrated the versatility of the presented chemistry, but we also highlight its utility in crafting molecules with enhanced antiviral properties. The successful translation of computational insights into tangible antiviral candidates underscores the efficacy of this multidisciplinary strategy.

This research not only contributes to our understanding of SARS-CoV-2 but also sets a precedent for future endeavors in antiviral drug discovery. The fusion of computational analyses with the innovative application of chemical pillars holds promise for the development of effective antiviral therapeutics, reinforcing the significance of scientific collaboration in addressing global health challenges.



## 2.5 Experimental Section

### 2.5.1 General

All melting points (mp) are uncorrected.  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{31}\text{P}$  NMR spectra were recorded on a 300 and 400 MHz spectrometer (solvents specified). Chemical shifts are expressed in ppm from internal tetramethylsilane ( $\delta$ ), and coupling constants ( $J$ ) are in Hertz (Hz): b, broad; s, singlet; bs, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet. IR spectra were recorded on a Agilent Technologies Cary 630 ZnSe Engine spectrophotometer available at the Department and absorptions ( $\nu$  are in  $\text{cm}^{-1}$ ). HRMS were done on a X500B QTOF system (Sciex, Framingham, MA 01701 USA) available at the CGS of the University of Pavia. Column chromatography and tlc: silica gel H60 and GF254, respectively; eluants: cyclohexane/ethyl acetate 9:1 to pure ethyl acetate.

### 2.5.2 Starting and Reference Materials

Compounds **1**, **5**, **9**, **16** were synthesized according to the procedure reported in the literature.<sup>26</sup> Other reagents and solvents were purchased from chemical suppliers and used without any further purification following or adapting the methods reported in the cited literature.

### 2.5.3 Synthesis of compounds 7a

2-Quinolinenitrile oxide **6** was generated in situ by dehydrohalogenation of the corresponding hydroximoyl chloride **5** with triethylamine.<sup>26</sup> To a stirred solution of dipolarophile **1** (10.0 g, 49.7 mmol) in anhydrous  $\text{CH}_2\text{Cl}_2$  (150 mL) and triethylamine (9.95 mL), a solution of hydroximoyl chloride **5** (12.3 g, 59.6 mmol) in the same solvent (50 mL) was added while stirring at  $0\text{ }^\circ\text{C}$  over 0.5 h. After keeping the reaction mixture for 2 days at room temperature, the organic phase was washed twice with water and dried with  $\text{Na}_2\text{SO}_4$ . The filtrate was evaporated under reduced pressure leaving a residue, which was separated by column chromatography.<sup>14</sup>

**Compound 7a.** Yield 8.1 g (44 %); mp 179–180  $^\circ\text{C}$  (EtOH). IR:  $\nu = 1652$  (C=O), 1548 (C=N)  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ ,  $25^\circ\text{C}$ ):  $\delta=1.99$  and  $2.08$  (d,  $J=12\text{Hz}$ , 2H,  $\text{CH}_2$ ), 4.60 (d,  $J = 8$  Hz, 1H,  $\text{H}_{4\text{isox.}}$ ), 5.06 (s, 1H, CH-O), 5.21 (d,  $J = 8$  Hz, 1H,  $\text{H}_{5\text{isox.}}$ ), 5.58 (s, 1H, CH-N), 7.61 (s, 4H, ArH), 7.43 (t,  $J = 7$  Hz, 1H, ArH), 7.84 (d,  $J = 7$  Hz, 1H, ArH), 7.92 (m, 3H, ArH), 8.04 (d,  $J = 8.6$  Hz, 1H, ArH), 8.17 (d,  $J$

= 8.6 Hz, 1H, ArH) ppm.  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ , 25 °C):  $\delta$  = 32.8, 48.1, 56.6, 61.5, 80.6, 83.5, 118.9, 127.5, 127.6, 128.1, 128.4, 128.8, 129.7, 129.8, 131.4, 131.8, 133.4, 136.5, 147.5, 147.9, 157.2 ppm.

$\text{C}_{22}\text{H}_{17}\text{N}_3\text{O}_3$  (371.39): calcd. C 71.15, H 4.61, N 11.31; found C 71.18, H 4.60, N 11.29.

#### 2.5.4 Synthesis of compound 8a

Cycloadduct **7a** was deacylated by dissolving the adduct (5 g, 13.5 mmol) in a stirred solution of NaOH (0.9 g, 20.3 mmol) in methanol (100 mL). After keeping overnight at room temp., the solutions were concentrated under reduced pressure, taken up with  $\text{CH}_2\text{Cl}_2$ , and the organic phase was washed three times with water and finally dried with  $\text{Na}_2\text{SO}_4$ . Filtration and evaporation of the solvent afforded oily residues of compounds **18a/b**, which were crystallized from diisopropyl ether/ethanol.<sup>14</sup>

**Compound 8a.** Yield 2.7 g (74 %); mp 118–120 °C (*i*-Pr<sub>2</sub>O/ ethanol). IR:  $\nu$  = 3218 (NH), 1600 (C=N)  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300 MHz,  $\text{DMSO}-d_6$ , 25 °C):  $\delta$  = 1.77 (m, 2H,  $\text{CH}_2$ ), 3.99 (s, 1H, CH-O), 4.08 (d,  $J$  = 8Hz, 1H,  $\text{H}_{4\text{isox.}}$ ), 4.92 (d,  $J$  = 8Hz, 1H,  $\text{H}_{5\text{isox.}}$ ), 5.00 (s, 1H, CH-N), 6.83 (bs, 1H, NH), 7.68 (m, 1H, ArH), 7.83 (m, 1H, ArH), 8.02 (m, 2H, ArH), 8.11 (d,  $J$  = 8.6Hz, 1H, ArH), 8.42 (d,  $J$  = 8.6Hz, 1H, ArH) ppm.  $^{13}\text{C}$  NMR (75 MHz,  $\text{DMSO}-d_6$ , 25 °C):  $\delta$  = 39.9, 76.2, 84.9, 119.2, 128.0, 128.1, 128.3, 129.6, 130.6, 137.3, 147.5, 148.9, 157.9 ppm.

$\text{C}_{15}\text{H}_{13}\text{N}_3\text{O}_2$  (267.28): calcd. C 67.40, H 4.90, N 15.72; found C 67.38, H 4.91, N 15.79.

#### 2.5.5 Synthesis of compound 62

A solution of the hydroxylamine **8a** (2 g, 7.5 mmol) and 10 % Pd/C (0.4 g) in ethyl acetate (75 mL) absorbed 1 equiv. of hydrogen in 2 h. The catalyst was filtered off and the filtrate was evaporated under reduced pressure. Crystallization from chloroform afforded aminol **62**.<sup>14</sup>

**Compound 62.** Yield 1.7 g (82 %); mp 171–174 °C (chloroform). IR:  $\nu$  = 3348, 3283 ( $\text{NH}_2$ ), 3130 (OH), 1601 (C=N)  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ , 25 °C):  $\delta$  = 1.82 (m, 2H,  $\text{CH}_2$ ), 2.0 – 3.5 (bs, 3H,  $\text{NH}_2$  and OH), 4.11 (d,  $J$  = 2Hz, 1H, CH-O), 4.31 (d,  $J$  = 8 Hz, 1 H,  $\text{H}_{4\text{isox.}}$ ), 4.44 (d,  $J$  = 2 Hz, 1H, CH-N), 5.35 (d,  $J$  = 8 Hz, 1 H,  $\text{H}_{5\text{isox.}}$ ), 7.60 (m, 1H, ArH), 7.76 (m, 1H, ArH), 7.84 (d,  $J$  = 8 Hz, 1H, ArH), 8.08 (d,  $J$  = 8 Hz, 2H, ArH), 8.16 (d,  $J$  = 8 Hz, 1H, ArH) ppm.

$^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ , 25 °C):  $\delta$  = 38.1, 57.5, 60.0, 78.7, 93.2, 119.3, 126.4, 127.3, 127.6, 128.0, 129.6, 129.7, 136.2, 147.6, 149.0, 158.6 ppm.

C<sub>15</sub>H<sub>15</sub>N<sub>3</sub>O<sub>2</sub> (269.30): calcd. C 66.90, H 5.61, N 15.60; found C 66.90, H 5.60, N 15.58.

### 2.5.6 Synthesis of cycloadducts 12a,b

4-chloro-benzonitrile oxide **10** was generated in situ by dehydrohalogenation of the corresponding hydroximoyl chloride **9** with triethylamine.<sup>23</sup> To a stirred solution of dipolarophile **1** (1.5 g, 7.5 mmol) in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (150 mL) and triethylamine (1.25 mL), a solution of hydroximoyl chloride **9** (1.56 g, 8.2 mmol) in the same solvent (50 mL) was added while stirring at 0 °C over 2 h. After keeping the reaction mixture for 2 days at room temperature, the organic phase was washed twice with water and dried with Na<sub>2</sub>SO<sub>4</sub>. The filtrate was evaporated under reduced pressure leaving a residue, which was separated by column chromatography.

**Compound 12a.** Yield 1.27 g (48%); mp 154-155°C (acetone). IR:  $\nu = 1632$  (C=O), 1595 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 2.05$  and  $2.15$  (m, 2H, CH<sub>2</sub>), 4.32 (d, J=8Hz, 1H, H<sub>4</sub><sub>isox.</sub>), 4.96 (s, 1H, CH-N), 5.04 (dt, J= 8, 1 Hz, 1H, H<sub>5</sub><sub>isox.</sub>), 5.11 (s, 1H, CH-O), 7.43 (AA'BB' syst., 2H, ArH), 7.48 (AA'BB' syst., 2H, ArH), 7.58 (m, 1H, ArH), 7.70 (m, 2H, ArH), 7.85 (m, 2H, ArH) ppm. <sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C):  $\delta = 32.7, 55.2, 81.2, 83.0, 126.1, 128.1, 128.2, 129.1, 129.5, 132.3, 136.8, 154.5, 171.2$  ppm.

C<sub>19</sub>H<sub>15</sub>ClN<sub>2</sub>O<sub>3</sub> (354.79): HRMS: calcd. (MW+Na) 377.0663; found 377.0657.

**Compound 12b.** Yield 1.19 g (45%); mp 129-130°C (acetone). IR:  $\nu = 1651$  (C=O), 1599 (C=N) cm<sup>-1</sup>. <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>, 25°C):  $\delta = 2.08$  (m, 2H, CH<sub>2</sub>), 4.21 (d, J=8Hz, 1H, H<sub>4</sub><sub>isox.</sub>), 4.99 (s, 1H, CH-N), 5.15 (bs, 1H, CH-O), 5.22 (d, J=8Hz, 1H, H<sub>5</sub><sub>isox.</sub>), 7.43 (d, J=8Hz, 2H, ArH), 7.47 (t, J=8Hz, 2H, ArH), 7.55 (t, J=8Hz, 1H, ArH), 7.67 (d, J=8 Hz, 2H, ArH), 7.80 (d, J=8Hz, 2H, ArH) ppm. <sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 32.7, 57.2, 79.9, 83.0, 126.3, 127.9, 128.4, 128.8, 129.5, 132.1, 132.8, 136.9, 154.1, 170.4$  ppm.

C<sub>19</sub>H<sub>15</sub>ClN<sub>2</sub>O<sub>3</sub> (354.79): HRMS: calcd. (MW+Na) 377.0663; found 377.0658.

### 2.5.7 Synthesis of hydroxylamines 13a,b

Compounds **12a,b** (2.1 g, 5.9 mmol) were dissolved in MeOH (100 mL) and powdered NaOH (0.40 g, 10.1 mmol) was added portionwise at rt. The reaction was left under stirring for 48 h. After this period of time, the solvent was evaporated at reduced pressure and the residue taken up with CH<sub>2</sub>Cl<sub>2</sub> and the organic phase washed with brine and finally dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. Upon evaporation of the solvent the crude solid products **13a,b** were collected and properly recrystallized from EtOAc and fully characterized.

**Compound 13a.** Yield 1.44 g (97%); mp 148 - 149 °C (acetone-*i*Pr<sub>2</sub>O).

IR:  $\nu = 3209$  (NH),  $1588$  (C=N)  $\text{cm}^{-1}$ .

<sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 1.82$  and  $2.09$  (m, 2H, CH<sub>2</sub>),  $4.00$  (d, J=8Hz, H<sub>4</sub><sub>isox.</sub>),  $4.03$  (s, 1H, CH-N),  $4.71$  (s, 1H, CH-O),  $5.90$  (d, J=8Hz, 1H, H<sub>5</sub><sub>isox.</sub>),  $7.42$  (AA'BB' sys., 2H, ArH),  $7.68$  (AA'BB' sys., 2H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 35.0, 55.9, 59.2, 77.5, 84.0, 126.7, 128.0, 129.3, 136.5, 155.0$  ppm.

C<sub>12</sub>H<sub>11</sub>ClN<sub>2</sub>O<sub>2</sub> (250.68): HRMS: calcd. (MW+H) 251.0582; found 251.0579.

**Compound 13b.** Yield 1.45 g (98%); mp 182 - 183 °C (acetone-*i*Pr<sub>2</sub>O).

IR:  $\nu = 3243$  (NH),  $1592$  (C=N)  $\text{cm}^{-1}$ .

<sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 1.84$  and  $2.11$  (d, J=16Hz, 2H, CH<sub>2</sub>),  $3.98$  (d, J=8Hz, 1H, H<sub>4</sub><sub>isox.</sub>),  $4.10$  (s, 1H, CH-N),  $4.73$  (s, 1H, CH-O),  $5.00$  (d, J=8Hz, 1H, H<sub>5</sub><sub>isox.</sub>),  $7.41$  (AA'BB' sys., 2H, ArH),  $7.67$  (AA'BB' sys., 2H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 34.7, 57.9, 61.6, 76.4, 84.4, 126.8, 128.0, 129.3, 136.5, 154.6$  ppm.

C<sub>12</sub>H<sub>11</sub>ClN<sub>2</sub>O<sub>2</sub> (250.68): HRMS: calcd. (MW+H) 251.0582; found 251.0578.

### 2.5.8 Synthesis of the aminols 177 and 193

Compounds **16a,b** (1.9 g, 7.6 mmol) were dissolved in EtOAc (65 mL) along with 0.4 g C/Pd 10% and the mixture was submitted to hydrogenation at rt. The reaction was left under stirring with hydrogen stream until it absorbed 170 mL. After this, the solution was filtered and the solvent was evaporated to dryness. The crude solid products **177** and **193** were recrystallized from *i*-Pr<sub>2</sub>O and fully characterized.

**Compound 177.** Yield 1.82 g (95 %); mp 159 - 160 °C (MeOH-*i*Pr<sub>2</sub>O).

IR:  $\nu = 3336, 3280$  (NH<sub>2</sub>),  $3041$  (OH),  $1607$  (C=N)  $\text{cm}^{-1}$ .

<sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 1.84$  (m, 2H, CH<sub>2</sub>),  $3.81$  (d, J=6Hz, 1H, CH-N),  $4.02$  (d, J=8Hz 1H, H<sub>4</sub><sub>isox.</sub>),  $4.43$  (d, J=6Hz, 1H, CH-O),  $5.29$  (d, J=8Hz, 1H, H<sub>5</sub><sub>isox.</sub>),  $7.43$  (AA'BB' sys., 2H, ArH),  $7.65$  (AA'BB' sys., 2H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 32.5, 57.2, 77.2, 92.4, 126.7, 127.9, 128.8, 128.9, 129.2, 130.1, 156.2$  ppm.

C<sub>12</sub>H<sub>13</sub>ClN<sub>2</sub>O<sub>2</sub> (152.70): HRMS: calcd. (MW+H) 253.0738; found 253.0736

**Compound 193.** Yield 1.86 g (97 %); mp 194 - 195 °C (acetone-*i*Pr<sub>2</sub>O).

IR:  $\nu = 3347, 3284$  (NH<sub>2</sub>),  $3094$  (OH),  $1580$  (C=N)  $\text{cm}^{-1}$ .

<sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>, 25 °C):  $\delta = 1.85$  (m, 2H, CH<sub>2</sub>),  $3.97$  (m, 1H, CH-N),  $4.29$  (s, 1H, CH-O),  $4.32$  (d, J=8Hz, 1H, H<sub>4</sub> isox.),  $5.04$  (d, J=8Hz, 1H, H<sub>5</sub> isox.),  $7.41$  (AA'BB' sys., 2H, ArH),  $7.74$  (AA'BB' sys., 2H, ArH) ppm.

$^{13}\text{C}$  NMR (75MHz,  $\text{CDCl}_3$ , 25°C):  $\delta$  = 38.6, 59.1, 62.6, 78.0, 91.8, 127.0, 127.2, 128.2, 128.9, 129.2, 136.1, 155.6 ppm.

$\text{C}_{12}\text{H}_{13}\text{ClN}_2\text{O}_2$  (152.70): HRMS: calcd. (MW+H) 253.0738; found 253.0735

### 2.5.9 Protection of aminol 177

In a 100 mL flask, 1.40 g (5.5 mmol) of aminol **177** were dissolved in 50 mL of anhydrous DCM along with 0.83 mg (12.2 mmol) of imidazole. TBDMSCl (0.92 g, 6.1 mmol) was then added under stirring at room temperature and the solution was left to react for 48 h. After this period of time, the solution was diluted with DCM and the washed with brine (3 x 20 mL). The organic phase was dried over anhydrous  $\text{Na}_2\text{SO}_4$ . Upon evaporation of the solvent, the pale yellow oils relative to the protected compound **14** was obtained and fully characterized.

**Compound 14.** Yield 1.67 g (82 %); mp 83°C (n-hexane). IR:  $\nu$  = 3388 ( $\text{NH}_2$ ), 1592 ( $\text{C}=\text{N}$ )  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300MHz,  $\text{CDCl}_3$ , 25°C):  $\delta$  = 0.15 (s, 3H,  $\text{CH}_3$ ), 0.17 (s, 3H,  $\text{CH}_3$ ), 0.93 (s, 9H, tBu), 1.74 and 1.91 (m, 2H,  $\text{CH}_2$ ), 2.21 (bs, 2H,  $\text{NH}_2$ ) 3.46 (d, J=6, 1H, CH-N), 4.02 (d, J=9, 1H,  $\text{H}_{4\text{isox.}}$ ), 4.46 (d, J=4, 1H, CH-O), 5.05 (d, J=9Hz, 1H,  $\text{H}_{5\text{isox.}}$ ), 7.39 (AA'BB' sys., 2H, ArH), 7.73 (AA'BB' sys., 2H, ArH) ppm.

$^{13}\text{C}$  NMR (75MHz,  $\text{CDCl}_3$ , 25°C):  $\delta$  = -5.0, 17.8, 25.7, 41.0, 58.6, 62.1, 78.8, 92.7, 127.3, 128.2, 129.0, 135.9, 156.9 ppm.

$\text{C}_{18}\text{H}_{27}\text{ClN}_2\text{O}_2\text{Si}$  (366.96): HRMS: calcd. (MW+H) 367.1603; found 367.1598

### 2.5.10 Coupling of amine 14 with Fmoc-Gly-OH

In a 250 mL flask, 1.20 g (3.3 mmol) of amine **14** was dissolved in 120 mL of anhydrous DCM along with 1.43 g (3.6 mmol) HBTU and 2.3 mL of DIPEA.

Fmoc-Gly-OH (0.97 g, 3.3 mmol) were then added under stirring at room temperature and the solution was left to react for 48 h. After this period of time, the organic phase is washed with brine and dried over anhydrous  $\text{Na}_2\text{SO}_4$ . Upon evaporation of the solvent, oily residues were collected and the product **15** was isolated from column chromatography and fully characterized.

**Compound 15.** Yield 1.54 g (73 %); mp 75 - 76 °C (acetone). IR:  $\nu$  = 3396 (NH), 1703 ( $\text{C}=\text{O}$ ), 1681 ( $\text{C}=\text{O}$ ), 1591 ( $\text{C}=\text{N}$ )  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300MHz,  $\text{DMSO}-d_6$ , 25°C):  $\delta$  = 0.10 (s, 3H,  $\text{CH}_3$ ), 0.12 (s, 3H,  $\text{CH}_3$ ), 0.89 (s, 9H, tBu), 1.69 and 1.90 (m, 2H,  $\text{CH}_2$ ), 3.62 (d, J=6,  $\text{CH}_2\text{NH}$ ), 4.26 (m, 6H, CHO, CHN,  $\text{H}_{4\text{isox.}}$ ,  $\text{CH}_2\text{O}$ , CHFmoc), 4.96 (d, J=9Hz, 1H,  $\text{H}_{5\text{isox.}}$ ), 7.34 (t, J=7, 2H, ArH), 7.43 (t, J=7, 2H, ArH), 7.50 (d, J=8, 2H, ArH), 7.63 (t, 1H, NH), 7.72 (d, J=7, 3H, 2ArH and 1NH), 7.80 (d, J=8, 2H, ArH), 7.90 (d, J=7, 2H, ArH) ppm.

$^{13}\text{C}$  NMR (75MHz, DMSO- $d_6$ , 25°C):  $\delta$  = -5.0, 17.8, 25.7, 43.9, 46.6, 52.6, 57.3, 65.8, 78.2, 92.3, 120.1, 125.2, 127.1, 127.2, 127.6, 128.8, 128.9, 134.6, 140.7, 143.7, 143.8, 156.4, 156.6, 168.3 ppm.

$\text{C}_{35}\text{H}_{40}\text{ClN}_3\text{O}_5\text{Si}$  (646.26): HRMS: calcd. (MW+Na) 668.2318; found 668.2356.

### 2.5.11 Deprotection of 15

To solution of the compound **15** (0.65 g, 1.0 mmol) in anhydrous THF (50 mL), a 1 M solution of  $n\text{-Bu}_4\text{N}^+\text{F}^-$  (1.2 mL, 1.2 mmol) was added under stirring at room temperature. The mixtures are left under stirring for 1 h. Then the solutions were concentrated and taken up with DCM (150 mL) and washed with brine (3x30 mL) and finally dried over  $\text{Na}_2\text{SO}_4$ . The crude solid product **181** was recrystallized from acetone/ $i\text{-Pr}_2\text{O}$  and fully characterized.

**Compound 181.** Yield 0.29 g (94 %); mp 139-140°C (acetone). IR:  $\nu$  = 3310 - 3050 bs ( $\text{NH}_2$ , OH), 1636 (C=O), 1592 (C=N)  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300MHz, DMSO- $d_6$ , 25°C):  $\delta$  = 1.71 (s, 2H,  $\text{NH}_2$ ), 1.83 (s, 2H,  $\text{CH}_2$ ), 3.75 (m, 2H,  $\text{CH}_2$ ), 4.17 (d,  $J=9$ , 1H,  $\text{H}_{4\text{isox.}}$ ), 4.29 (s, 1H, CH-N), 4.45 (m, 1H, CH-O), 5.02 (d,  $J=9$ , 1H,  $\text{H}_{5\text{isox.}}$ ), 5.77 (d,  $J=2\text{Hz}$ , 1H, OH), 7.51 (AA'BB' sys., 2H, ArH), 8.01 (AA'BB' sys., 2H, ArH), 8.55 (d,  $J=9$ , 1H, NHCO) ppm.

$^{13}\text{C}$  NMR (75MHz, DMSO- $d_6$ , 25°C):  $\delta$  = 29.0, 36.4, 53.4, 53.6, 59.7, 76.8, 91.8, 127.5, 128.8, 134.7, 155.7, 169.1, 169.4.

$\text{C}_{14}\text{H}_{16}\text{ClN}_3\text{O}_3$  (309.75): HRMS: calcd. (MW+H) 310.0953; found 310.0961

### 2.5.12 Coupling of compound 181 with Fmoc-Gly-OH

In a 250 mL flask, 0.30 g (1.0 mmol) of compound **181** was dissolved in 100 mL of anhydrous DCM. DIC (0.25 g, 2.0 mmol) was added along with a catalytic amount of DMAP and Fmoc-Gly-OH (0.32 g, 1.1 mmol) under stirring at room temperature and the solution was left to react for 48 h. After this period of time, the organic phase is washed with a saturated solution  $\text{NaHCO}_3$  and brine and dried over anhydrous  $\text{Na}_2\text{SO}_4$ . Upon evaporation of the solvent, oily residue was collected and the products **16** was isolated from column chromatography and fully characterized.

**Compound 16.** Yield 0.21 g (37%); mp 123 - 124°C (acetone- $i\text{Pr}_2\text{O}$ ). IR:  $\nu$  = 3295 bs ( $\text{NH}_2$ ), 1655 (C=O), 1595 (C=N)  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300MHz, DMSO- $d_6$ , 25°C):  $\delta$  = 1.80 – 2.10 (m, 2H+2H,  $\text{CH}_2$  and  $\text{CH}_2\text{NH}_2$ ), 2.90 (bs, 2H,  $\text{NH}_2$ ), 3.17 (2H,  $\text{CH}_2\text{NHCO}$ ), 3.41 (bs, 2H,  $\text{CH}_2\text{O}$ ), 4.30 (d,  $J=8$  Hz, 1H,  $\text{H}_{4\text{isox.}}$ ), 4.37 (bs, 1H CHN), 5.01 (t,  $J=3\text{Hz}$ , 1H, CHFmoc), 5.18 (d,  $J=8\text{Hz}$ ,  $\text{H}_{5\text{isox.}}$ ), 5.23 (bs, 1H, CHO), 6.58 (d,  $J=6$  Hz, 1H, NHCO), 7.45 (m, 6H, ArH), 7.85 (m, 4H, ArH), 8.00 (d,  $J=8$  Hz, 2H, ArH), 8.11 (d,  $J=6\text{Hz}$ , NHCO)

$^{13}\text{C}$  NMR (75MHz, DMSO- $d_6$ , 25°C):  $\delta$  = 26.3, 38.7, 53.5, 89.3, 106.7, 109.7, 120.0, 121.4, 127.1, 127.3, 128.9, 134.9, 137.4, 139.4, 142.6, 149.2, 154.0, 169.8, 171.9 ppm.

$\text{C}_{31}\text{H}_{29}\text{ClN}_4\text{O}_6$  (589.05): HRMS: calcd. ( $\text{C}_{16}\text{H}_{19}\text{ClN}_4\text{O}_4+\text{H}$ ) 367.1168; found 367.1169

### 2.5.13 Deprotection of 16

To solution of the compound **16** (0.28 g, 0.5 mmol) in anhydrous THF (50 mL), a 1 M solution of  $n\text{-Bu}_4\text{N}^+\text{F}^-$  (0.48 mL, 0.5 mmol) was added under stirring at room temperature. The mixtures are left under stirring for 1 h. Then the solutions were concentrated and taken up with DCM (100 mL) and washed with brine (3x20 mL) and finally dried over  $\text{Na}_2\text{SO}_4$ . The crude solid product **189** was recrystallized from acetone/*i*-Pr $_2$ O and fully characterized.

**Compound 189.** Yield 0.14 g (81 %); mp 153 - 155°C (acetone-*i*Pr $_2$ O). IR:  $\nu$  = 3235 bs (NH $_2$ , NH), 1655 (C=O), 1603 (C=N)  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300MHz, DMSO- $d_6$ , 25°C):  $\delta$  = 1.60 and 1.73 (m, 2H, CH $_2$ ), 3.17 (m, 2H+2H, CH $_2$ ), 3.35 (bs, 4H, NH $_2$ ), 4.13 (d, J=8, 1H, H $_{4\text{ISOX}}$ ), 4.97 (d, J=8, 1H, H $_{5\text{ISOX}}$ ), 5.17 (bs, 1H, CHN), 5.19 (bs, 1H, CHO), 7.49 (AA'BB' sys., 2H, ArH), 7.88 (AA'BB' sys., 2H, ArH), 8.90 (b, 1H, NH) ppm.

$^{13}\text{C}$  NMR (75MHz, DMSO- $d_6$ , 25°C):  $\delta$  = 30.4, 57.49, 57.5, 58.5, 76.8, 77.6, 92.6, 127.5, 128.7, 128.8, 134.5, 156.1, 167.4, 173.7

$\text{C}_{31}\text{H}_{29}\text{ClN}_4\text{O}_6$  (366.80): HRMS: calcd. (MW+H) 367.1168; found 367.1169.

### 2.5.14 Synthesis of cycloadducts 19a,b

3-nitro-benzonitrile oxide **18** was generated in situ by dehydrohalogenation of the corresponding hydroximoyl chloride **17** with triethylamine.<sup>23</sup> To a stirred solution of dipolarophile **1** (2.0 g, 9.9 mmol) in anhydrous  $\text{CH}_2\text{Cl}_2$  (150 mL) and triethylamine (1.66 mL), a solution of hydroximoyl chloride **17** (1.56 g, 8.2 mmol) in the same solvent (50 mL) was added while stirring at 0 °C over 2 h. After keeping the reaction mixture for 2 days at room temperature, the organic phase was washed twice with water and dried with  $\text{Na}_2\text{SO}_4$ . The filtrate was evaporated under reduced pressure leaving a residue, which was separated by column chromatography.

**Compound 19a.** Yield 1.67 g (46%); mp 173 – 174 °C (acetone). IR:  $\nu$  = 1614 (C=O), 1595 (C=N), 1521 (NO $_2$ )  $\text{cm}^{-1}$ .

$^1\text{H}$  NMR (300MHz,  $\text{CDCl}_3$ , 25°C):  $\delta$  = 2.04 (d, J=12Hz, 1H, CH $_2$ ), 2.15 (dd, J=12Hz, J=5Hz, 1H, CH $_2$ ) 4.40 (d, J=8Hz, H $_{4\text{ISOX}}$ ), 5.00 (s, 1H, CH-N), 5.12 (s, 1H, CH-O), 5.13 (d, J=8Hz, 1H, H $_{5\text{ISOX}}$ ), 7.50 (m,

3H, ArH), 7.67 (t, J=8Hz, 1H, ArH), 7.86 (d, J=8Hz, 2H, ArH), 8.10 (d, J=8Hz, 1H, ArH), 8.32 (d, J=8Hz, 1H, ArH), 8.61 (s, 1H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C): δ = 32.6, 54.9, 58.9, 80.8, 83.6, 121.3, 125.0, 128.3, 128.9, 129.4, 130.3, 132.2, 132.4, 148.6, 153.8, 171.4 ppm.

C<sub>19</sub>H<sub>15</sub>N<sub>3</sub>O<sub>5</sub> (365.35): HRMS: calcd. (MW+H) 366.1084; found 366.1071

**Compound 19b.** Yield 1.56 g (43%); mp °C (acetone). IR: ν = 1638 (C=O), 1599 (C=N), 1517 (NO<sub>2</sub>) cm<sup>-1</sup>.

<sup>1</sup>H NMR (300MHz, CDCl<sub>3</sub>, 25°C): δ = 2.13 (m, 2H, CH<sub>2</sub>), 4.28 (d, J = 8Hz, 1H, H<sub>4</sub><sub>isox.</sub>), 5.04 (s, 1H, CH-N), 5.22 (s, 1H, CH-O), 5.30 (d, J = 8Hz, 1H, H<sub>5</sub><sub>isox.</sub>), 7.49 (m, 2H, ArH), 7.57 (m, 1H, ArH), 7.67 (t, J= 8Hz, 1H, ArH), 7.81 (m, 2H, ArH), 8.11 (d, J=8Hz, 1H, ArH), 8.32 (m, 1H, ArH), 8.54 (s, 1H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C): δ = 32.7, 56.7, 77.1, 79.6, 83.6, 121.3, 125.0, 128.3, 128.8, 129.6, 130.2, 132.0, 132.1, 132.5, 148.6, 153.4, 170.6 ppm.

C<sub>19</sub>H<sub>15</sub>N<sub>3</sub>O<sub>5</sub> (365.35): HRMS: calcd. (MW+H) 366.3533; found 366.3537

### 2.5.15 Synthesis of hydroxylamine 20

Compounds **19a** (1.35 g, 3.7 mmol) was dissolved in MeOH (100 mL) and powdered NaOH (0.25 g, 6.3 mmol) was added portionwise at rt. The reaction was left under stirring for 48 h. After this period of time, the solvent was evaporated at reduced pressure and the residue taken up with CH<sub>2</sub>Cl<sub>2</sub> and the organic phase washed with brine and finally dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. Upon evaporation of the solvent the crude solid products **20** was collected and properly recrystallized from EtOAc and fully characterized.

**Compound 20.** Yield 0.86 g (89%); m.p. 193 - 194 °C (acetone-iPr<sub>2</sub>O). IR: ν = 3235 (NH), 1591 (C=N), 1521. (NO<sub>2</sub>) cm<sup>-1</sup>.

<sup>1</sup>H NMR (300MHz, CDCl<sub>3</sub>, 25°C): δ = 1.91 and 2.11 (d, J=12Hz, 2H, CH<sub>2</sub>), 4.13 (d, J=8Hz, H<sub>4</sub><sub>isox.</sub>), 4.14 (s, 1H, CH-N), 4.77 (s, 1H, CH-O), 4.79 (b, 1H, NH), 4.99 (d, J=8Hz, 1H, H<sub>5</sub><sub>isox.</sub>), 7.65 (t, J=8Hz, 1H, ArH), 8.13 (d, J=8Hz, 1H, ArH), 8.30 (dd, J=8Hz, J=1Hz, 1H, ArH), 8.56 (d, J=1Hz, 1H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C): δ = 34.7, 55.3, 59.0, 77.6, 84.4, 121.4, 124.8, 130.0, 130.1, 132.2, 148.5, 154.2 ppm.

C<sub>12</sub>H<sub>11</sub>N<sub>3</sub>O<sub>4</sub> (261.24): HRMS: calcd. (MW+H) 262.0822; found 262.0829



### 2.5.16 Synthesis of the aminol 179

Compound **20** (0.76 g, 2.9 mmol) were dissolved in EtOAc (65 mL) along with 0.3 g C/Pd 10% and the mixture was submitted to hydrogenation at rt. The reaction was left under stirring with hydrogen stream until it absorbed 70 mL. After this, the solution was filtered and the solvent was evaporated to dryness. The crude solid products **179** was recrystallized from *i*-Pr<sub>2</sub>O and fully characterized.

**Compound 179.** Yield 0.76 g (99 %); mp 180 – 182 °C (acetone-*i*Pr<sub>2</sub>O). IR:  $\nu$  = 3358 and 3291 (NH<sub>2</sub>), 3086 (OH), 1588 (C=N), 1528 (NO<sub>2</sub>) cm<sup>-1</sup>.

<sup>1</sup>H NMR (300MHz, CDCl<sub>3</sub>, 25°C):  $\delta$  = 1.79 and 1.88 (m, 2H, CH<sub>2</sub>), 3.99 (d, J=8Hz, H<sub>4</sub><sub>isox.</sub>), 4.24 (d, J=4Hz, 1H, CH-N), 4.46 (s, 1H, CH-O), 5.15 (b, 1H, OH), 5.38 (d, J=8Hz, 1H, H<sub>5</sub><sub>isox.</sub>), 7.63 (t, J=8Hz, 1H, ArH), 8.16 (d, J=8Hz, 1H, ArH), 8.29 (d, J=8Hz, 1H, ArH), 8.34 (s, 1H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C):  $\delta$  = 38.6, 59.1, 65.0, 78.4, 93.7, 121.2, 124.5, 130.0, 130.9, 132.3, 154.7, 166.6 ppm.

C<sub>12</sub>H<sub>13</sub>N<sub>3</sub>O<sub>4</sub> (263.25): HRMS: calcd. (MW+H) 264.0979; found 264.0987.

### 2.5.17 Protection of aminol 179

In a 250 mL flask, 1.0 g (3.8 mmol) of aminol **179** were dissolved in 100 mL of anhydrous DCM along with 0.57 mg (8.4 mmol) of imidazole. TBDMSCl (0.63 g, 4.2 mmol) was then added under stirring at room temperature and the solution was left to react for 48 h. After this period of time, the solution was diluted with DCM and the washed with brine (3 x 20 mL). The organic phase was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. Upon evaporation of the solvent, the pale yellow oil relative to the protected compound **21** was obtained and fully characterized.

**Compound 21.** Yield 0.82 g (57 %); mp 95 - 96°C (*i*Pr<sub>2</sub>O). IR:  $\nu$  = 3384 (NH<sub>2</sub>), 1592 (C=N), 1528 (NO<sub>2</sub>) cm<sup>-1</sup>.

<sup>1</sup>H NMR (300MHz, CDCl<sub>3</sub>, 25°C):  $\delta$  = 0.11 (s, 3H, CH<sub>3</sub>), 0.13 (s, 3H, CH<sub>3</sub>), 0.90 (s, 9H, CH<sub>3</sub>), 1.60 and 1.88 (m, 2H, CH<sub>2</sub>), 3.21 (m, 1H, CH-N), 3.33 (s, 2H, NH<sub>2</sub>), 4.07 (dd, J=10Hz, J=4Hz H<sub>4</sub><sub>isox.</sub>), 4.26 (m, 1H, CH-O), 5.02 (dd, J=10Hz, J=2Hz, 1H, H<sub>5</sub><sub>isox.</sub>), 7.77 (t, J=8Hz, 1H, ArH), 8.29 (dd, J=8Hz, J=2Hz, 2H, ArH), 8.69 (t, J=2Hz, 1H, ArH), ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C):  $\delta$  = -4.9, 17.7, 25.7, 42.2, 56.8, 60.5, 78.4, 93.2, 121.5, 124.4, 130.5, 133.2, 148.1, 157.1, 168.0 ppm.

C<sub>18</sub>H<sub>27</sub>N<sub>3</sub>O<sub>4</sub>Si (377.52): HRMS: calcd. (MW+H) 378.1844; found 378.1832.

### 2.5.18 Coupling of amine 21 with Fmoc-Gly-OH

In a 250 mL flask, 0.20 g (0.5 mmol) of amine **21** was dissolved in 120 mL of anhydrous DCM along with 0.23 g (0.6 mmol) HBTU and 0.4 mL of DIPEA.

Fmoc-Gly-OH (0.16 g, 0.5 mmol) were then added under stirring at room temperature and the solution was left to react for 48 h. After this period of time, the organic phase is washed with brine and dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. Upon evaporation of the solvent, oily residues were collected and the product **22** was isolated from column chromatography and fully characterized.

**Compound 22.** Yield 0.31 g (89 %); mp 82 - 83°C (acetone-*i*Pr<sub>2</sub>O)

IR:  $\nu$  = 3392, 3310 (NH<sub>2</sub>), 1722 (C=O), 1674 (C=O), 1592 (C=N), 1528 (NO<sub>2</sub>) cm<sup>-1</sup>.

<sup>1</sup>H NMR (300MHz, CDCl<sub>3</sub>, 25°C):  $\delta$  = 0.11 (s, 3H, CH<sub>3</sub>), 0.13 (s, 3H, CH<sub>3</sub>), 0.90 (s, 9H, tBu), 1.72 and 1.95 (m, 2H, CH<sub>2</sub>), 3.62 (m, 2H, CHN and CHFmoc.), 4.28 (m, 6H, CH<sub>2</sub>O, COCH<sub>2</sub>N, CHO, H<sub>4</sub><sub>isox.</sub>), 5.03 (d, J=10Hz, H<sub>5</sub><sub>isox.</sub>), 7.38 (m, 4H, ArH), 7.59 (m, 1H, NH), 7.74 (m, 3H, ArH), 7.82 (d, J=8Hz, 2H, ArH NH), 7.90 (d, J=7, 1H, ArH), 8.22 (d, J=8Hz, 2H, ArH), 8.28 (d, J=8Hz, 2H, ArH), 8.54 (s, 1H, ArH), ppm.

<sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>, 25°C):  $\delta$  = -5.0, 22.8, 25.7, 46.6, 52.3, 57.1, 65.8, 67.3, 78.1, 92.8, 120.1, 121.3, 124.4, 125.2, 127.1, 127.6, 130.0, 130.5, 133.3, 140.7, 143.8, 148.0, 156.2, 156.6, 168.4 ppm.

C<sub>35</sub>H<sub>40</sub>N<sub>4</sub>O<sub>7</sub>Si (656.81): HRMS: calcd. (MW+H) 657.2739; found 657.2729

### 2.5.19 Deprotection of 22

To solution of the compound **22** (0.26 g, 0.4 mmol) in anhydrous THF (50 mL), a 1 M solution of n-Bu<sub>4</sub>N<sup>+</sup>F<sup>-</sup> (0.3 mL, 0.3 mmol) was added under stirring at room temperature. The mixtures are left under stirring for 1 h. Then the solutions were concentrated and taken up with DCM (150 mL) and washed with brine (3x20 mL) and finally dried over Na<sub>2</sub>SO<sub>4</sub>. The crude solid product **183** was recrystallized from acetone/*i*-Pr<sub>2</sub>O and fully characterized.

**Compound 183.** Yield 0.12 g (93 %); mp 159-160°C (acetone). IR:  $\nu$  = 3343 and 3295 bs (NH<sub>2</sub>) 3086 (OH), 1636 (C=O), 1592 (C=N), 1528 (NO<sub>2</sub>) cm<sup>-1</sup>.

<sup>1</sup>H NMR (300MHz, DMSO-*d*<sub>6</sub>, 25°C):  $\delta$  = 1.84 (s, 2H, CH<sub>2</sub>), 3.77 (m, 2H, CH<sub>2</sub>), 4.29 (m, 2H, H<sub>4</sub><sub>isox.</sub>, CH-N), 4.44 (m, 1H, CH-O), 5.09 (d, J=9, 1H, H<sub>5</sub><sub>isox.</sub>), 5.77 (d, J=3Hz, 1H, OH), 7.77 (t, J=8, 1H, ArH), 8.29 (d, J=8, 1H, ArH), 8.47 (m, J=8, 1H, NH), 8.53 (d, J=9, 1H, ArH), 8.72 (t, J=2Hz, 1H, ArH) ppm.

<sup>13</sup>C NMR (75MHz, DMSO-*d*<sub>6</sub>, 25°C):  $\delta$  = 19.4, 22.8, 29.9, 36.4, 53.3, 53.7, 59.3, 66.3, 76.7, 92.4, 121.3, 124.4, 130.3, 130.4, 133.4, 148.2, 155.5, 169.1, 169.6 ppm.

C<sub>14</sub>H<sub>16</sub>N<sub>4</sub>O<sub>5</sub> (320.31): HRMS: calcd. (MW+H) 321.1193; found 321.1191

### 2.5.20 Synthesis of the 5-hydroxyisoxazolidine 11

DEAE-C (5 g) is placed in a mortar and 0.77 g (5 mmol) of the hydroxamic acid **23** and crotonaldehyde **24** (0.35 g) is added. The mixture is mixed with a pestle carefully for 2 hours. The mixed solid material is left at room temperature for additional 2 hours before dissolving in DCM (100 mL) and water (100 mL); extraction allowed for the collection of the organic product that was purified by column chromatography.<sup>11</sup>

**Compound 11.** Yield 0.79g (71 %); mp 130 - 132 °C (EtOAc). IR  $\nu$  = 3311 (OH), 1624 (C=O)  $\text{cm}^{-1}$ .

<sup>1</sup>H NMR (300MHz, DMSO-*d*<sub>6</sub>, 25°C):  $\delta$  = 1.18 and 1.25 (d, 3H, J=7 Hz, CH<sub>3</sub>), 1.74 and 2.75 (m, 2H, CH<sub>2</sub>), 4.11 (m, 1H, CH-CH<sub>3</sub>), 5.81 (m, 1H, O-CH-OH), 6.87 (m, 2H, ArH), 7.29 (m, 2H, ArH), 10.45 (s, 1H, OH) ppm.

<sup>13</sup>C NMR (75MHz, DMSO-*d*<sub>6</sub>, 25°C):  $\delta$  = 17.8, 44.1, 78.2, 97.9, 116.2, 116.3, 118.5, 120.4, 167.5 ppm.

Anal. Calcd for C<sub>11</sub>H<sub>13</sub>NO<sub>4</sub> (223.23): C, 59.19; H, 5.87; N, 6.27. Found: C, 59.17; H, 5.86; N, 6.28.

## 2.6 References

- 1) Lamb, Y. N. Nirmatrelvir Plus Ritonavir: First Approval *Drugs* **2022**, *82*, 585-591.
- 2) 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.* **2022**, *65*, 2716-2746.
- 3) Alzyoud, L.; Ghattas M.A.; Atatreh, N. Allosteric Binding Sites of the SARS-CoV-2 Main Protease: Potential Targets for Broad-Spectrum Anti-Coronavirus Agents *Drug Des. Devel. Ther.* **2022**, *16*, 2463-2478.
- 4) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors *Science* **2020**, *368*, 409-412.
- 5) Goyal B.; Goyal, D. Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy *ACS Comb. Sci.* **2020**, *22*, 297-305.
- 6) Hu T.; Zhang Y.; Li L.; Wang K.; Chen S.; Chen J.; Ding J.; Jiang H. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure *Virology* **2009**, *388*, 324-334.
- 7) Morra, G.; Neves, M. A. C.; Plescia, C. J.; Tsustsumi, S.; Neckers, L.; Verkhivker, G.; Altieri, D. C.; Colombo, G. Dynamics-Based Discovery of Allosteric Inhibitors: Selection of New Ligands for the C-terminal Domain of Hsp90 *J. Chem. Theory Comput.* **2010**, *6*, 2978-2989.
- 8) Sol, A.; Tsai, C. J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways *Structure* **2009**, *17*, 1042-1050.
- 9) Torrens-Fontanals, M.; Peralta-García, A.; Talarico, C.; Guixà-González, R.; Giorgino, T.; Selent, J. SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions *Nucleic Acids Res.* **2021**, *50*, D858-D866.
- 10) Morra, G.; Potestio, R.; Micheletti, C.; Colombo, G. Corresponding Functional Dynamics across the Hsp90 Chaperone Family: Insights from a Multiscale Analysis of MD Simulations *PLOS Comput. Biol.* **2012**, *8*, e1002433 16 pages.
- 11) Aljaf, K.; Amin A.; Hussain F.H.S.; Quadrelli, P. DEAE-cellulose-catalyzed synthesis of h-hydroxy-isoxazolidines and their synthetic uses towards nucleoside analogues *Arkivoc* **2020**, *part vi*, 73-83.
- 12) Presenti, P.; Moiola, M.; Quadrelli, P. 5-Hydroxy-isoxazolidine: A New Synthetic Approach to a Privileged Heterocycle for Organic Synthesis *Chemistry Select* **2020**, *5*, 5367-5373.
- 13) Scagnelli, L.; Memeo, M.; G; Carosso, S.; Bovio, B.; Quadrelli, P. Synthesis of New Carbanucleosides by Pericyclic Reactions *Eur. J. Org. Chem.* **2013**, 3835-3846.
- 14) Quadrelli, P.; Mella, M.; Legnani, L.; Al-Saad, D. From Cyclopentadiene to Isoxazoline-Carbocyclic Nucleosides; Synthesis of Highly Active Inhibitors of Influenza A Virus H1N1 *Eur. J. Org. Chem.* **2013**, 4655-4665.
- 15) Al-Saad, D.; Memeo, M. G.; Quadrelli, P. Pericyclic Reactions for Anti-HPV Antivirals: Unconventional Nucleoside Analogue Synthesis via Nitrosocarbonyl Chemistry *Chemistry Select* **2017**, *2*, 10340-10346.

- 16) Carella, S.; Memeo M. G.; Quadrelli, P. Electrocyclic Ring-Opening of 1,2,4-Oxadiazole[4,5-*a*]pyridinium Chloride: a New Route to 1,2,4-Oxadiazole Dienamino Compounds *Chemistry Open* **2019**, *8*, 1209-1221.
- 17) Mantione, D.; Aizpuru, O. O.; Memeo M. G.; Bovio, B.; Quadrelli, P. 4-Heterosubstituted Cyclopentenone Antiviral Compounds: Synthesis, Mechanism, and Antiviral Evaluation *Eur. J. Org. Chem.* **2016**, 983-911.
- 18) Moggio, Y.; Legnani, L.; Bovio, B.; Memeo, M., G.; Quadrelli, P. Synthesis of novel anthracene derivatives of isoxazolino-carbocyclic nucleoside analogues *Tetrahedron* **2012**, *68*, 1384-1392.
- 19) Savion, M.; Memeo, M. G.; Bovio, B.; Grazioso, G.; Legnani, L.; Quadrelli, P. Synthesis and molecular modelling of novel dihydroxycyclopentane-carbonitrile nor-nucleosides by vromonitrile oxide 1,3-dipolar cycloaddition *Tetrahedron* **2012**, *68*, 1854-1852.
- 20) Al-Saad, D.; Memeo, M., G.; Quadrelli, P. Pericyclic Reactions for Antivirals: Synthesis of 4-Bromo-*N*-[(1*R*\*,4*S*\*)-4-hydroxy-2-cyclohexen-1-yl]-2-thiazolecarboxamide *Lett. Org. Chem.* **2016**, *13*, 757-763.
- 21) Corti, M.; Leusciatti, M.; Moiola, M.; Mella, M.; Quadrelli, P. Nitrosocarbonyl Carbohydrate Derivatives: Hetero Diels-Alder and Ene Reaction Products for Useful Organic Synthesis *Synthesis* **2020**, *52*, A-M.
- 22) Al-Saad, D.; Memeo, M. G.; Quadrelli, P. #Nitrosocarbonyls 1: Antiviral Activity of *N*-(4-Hydroxycyclohex-2-en-1-yl)quinoline-2-carboxamide against the Influenza A Virus H1N1 *Sci. World J.* **2014**, Article ID 472373, 10 pages.
- 23) El-Baba, T. J.; Lutomski, C. A.; Kantsadi, A. L.; Malla, T. R.; John, T.; Mikhailov, V.; Bolla, J. R.; Schofield, C. J.; Zitzmann, N.; Vakonakis, I.; Robinson, C. V. Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays *Angew. Chem. Int. Ed.* **2020**, *59*, 23544–23548.
- 24) a) Quadrelli P.; Scrocchi, R., Caramella P.; Rescifina, A.; Piperno, A. *Tetrahedron* **2004**, *60*, 3643–3651; b) Quadrelli, P.; Mella, M.; Carosso, S.; Bovio, B.; Caramella, P. *Eur. J. Org. Chem.* **2007**, 6003–6015; c) Quadrelli, P.; Mella, M.; Assanelli, G. *Tetrahedron* **2007**, *64*, 7312–7317; d) Quadrelli, P.; Piccanello, A.; Vazquez-Martinez, N.; Bovio, B.; Mella, M.; Caramella, P. *Tetrahedron* **2006**, *62*, 7370–7379; e) Savion, M.; Memeo, M.G.; Bovio, B.; Grazioso, G.; Legnani, L.; Quadrelli, P. *Tetrahedron* **2012**, *68*, 1845–1852.
- 25) Memeo, M. G.; Mella, M.; Montagna, V.; Quaderelli, P. Design, Synthesis, and Conformational Analysis of Proposed  $\beta$ -Turn Mimics from Isoxazoline-Cyclopentane Aminols *Chem. Eur. J.*, **2015**, *21*, 16374-16378.
- 26) a) Grundmann, C.; Grünanger, P. *The Nitrile Oxides* Springer-Verlag, Heidelberg, **1971**; b) Caramella P., Grünanger, P. *1,3-Dipolar Cycloaddition Chemistry* (Ed.: A. Padwa), John Wiley & Sons, Inc. New York, **1984**, vol. 1, p. 291–392; c) Grünanger, P.; Finzi P., V. *Isoxazoles*, John Wiley & Sons, Inc., New York, **1999**, vol. 1 and 2.

## 2.7 Appendix

### 2.7.1 Docking Score, Ligand Efficiency, Structures

**Table S1.** Docking-based Virtual Screening – Pocket 1

Compound	Docking Score	Ligand Efficiency	Pocket
71	-6.913	-0.223	1 AB
193	-5.317	-0.313	1 AB
179	-5.135	-0.27	1 AB
189	-6.863	-0.275	1 BA
11	-5.252	-0.328	1 BA
183	-3.495	-0.152	1 BA

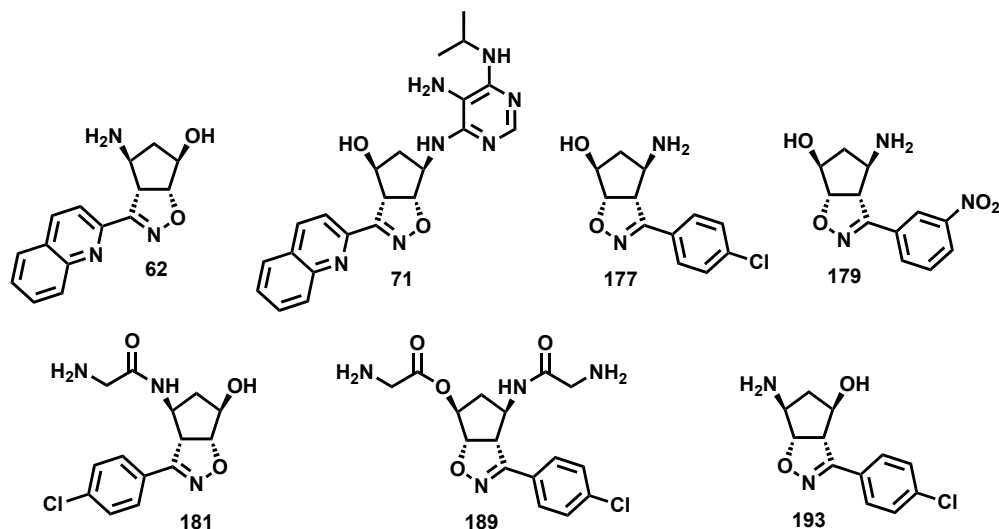
**Table S2.** Docking-based Virtual Screening – Pocket 1 – Receptor: last frame of compound 11's trajectory.

Compound	Docking Score	Ligand Efficiency	Pocket
62	-6.697	-0.335	1 BA
71	-4.812	-0.155	1 BA
177	-6.715	-0.395	1 BA
179	-5.914	-0.311	1 BA
189	-5.873	-0.235	1 BA
193	-4.972	-0.292	1 BA
x1187	-4.679	-0.311	1 BA

**Table S3.** Docking-based Virtual Screening – Pocket 2

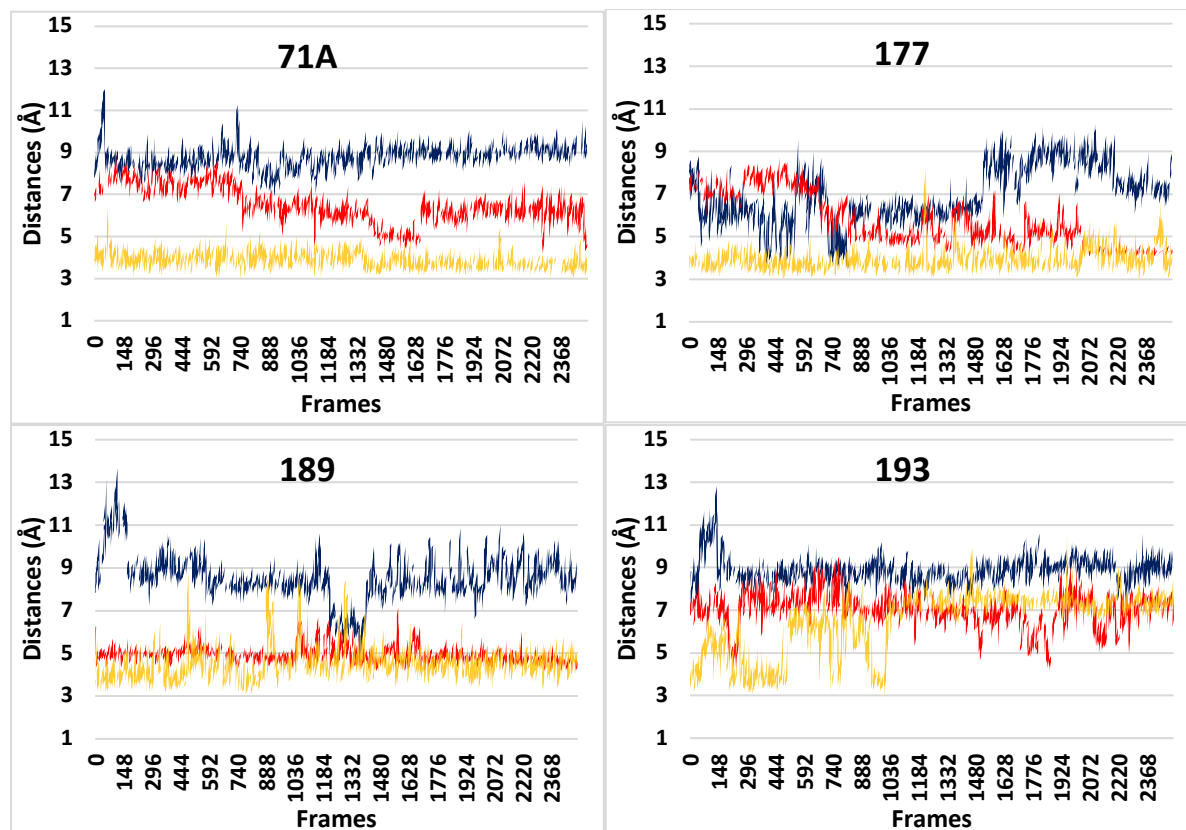
Compound	Docking Score	Ligand Efficiency	Pocket
62	-9.116	-0.456	2
181	-9.34	-0.445	2
177	-7.778	-0.458	2
11	-7.38	-0.461	2
179	-7.214	0.38	2

**Figure S1.** Structures of selected compounds from the second docking screening.

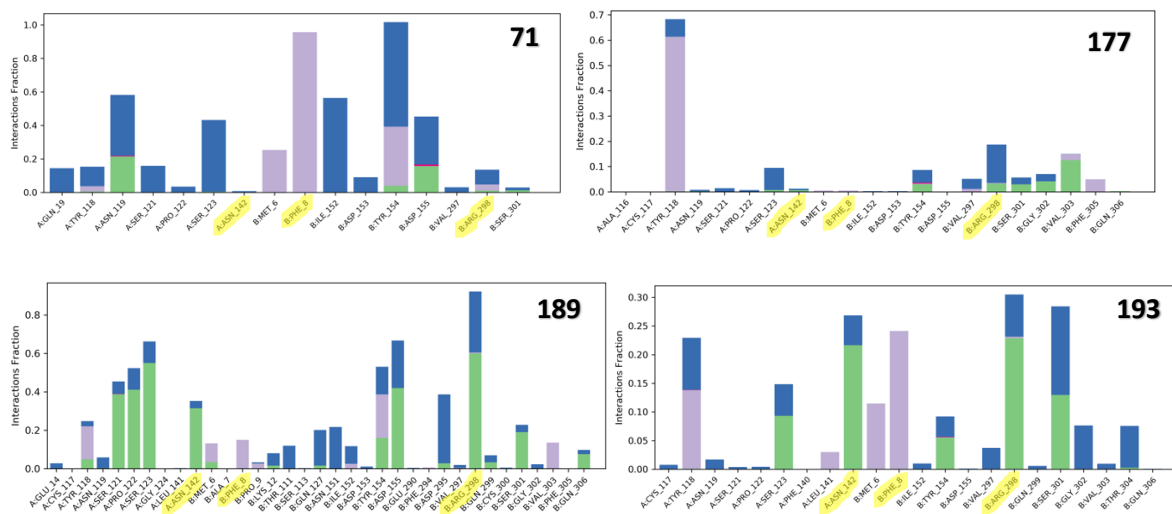


## 2.7.2 Trajectory analysis P1

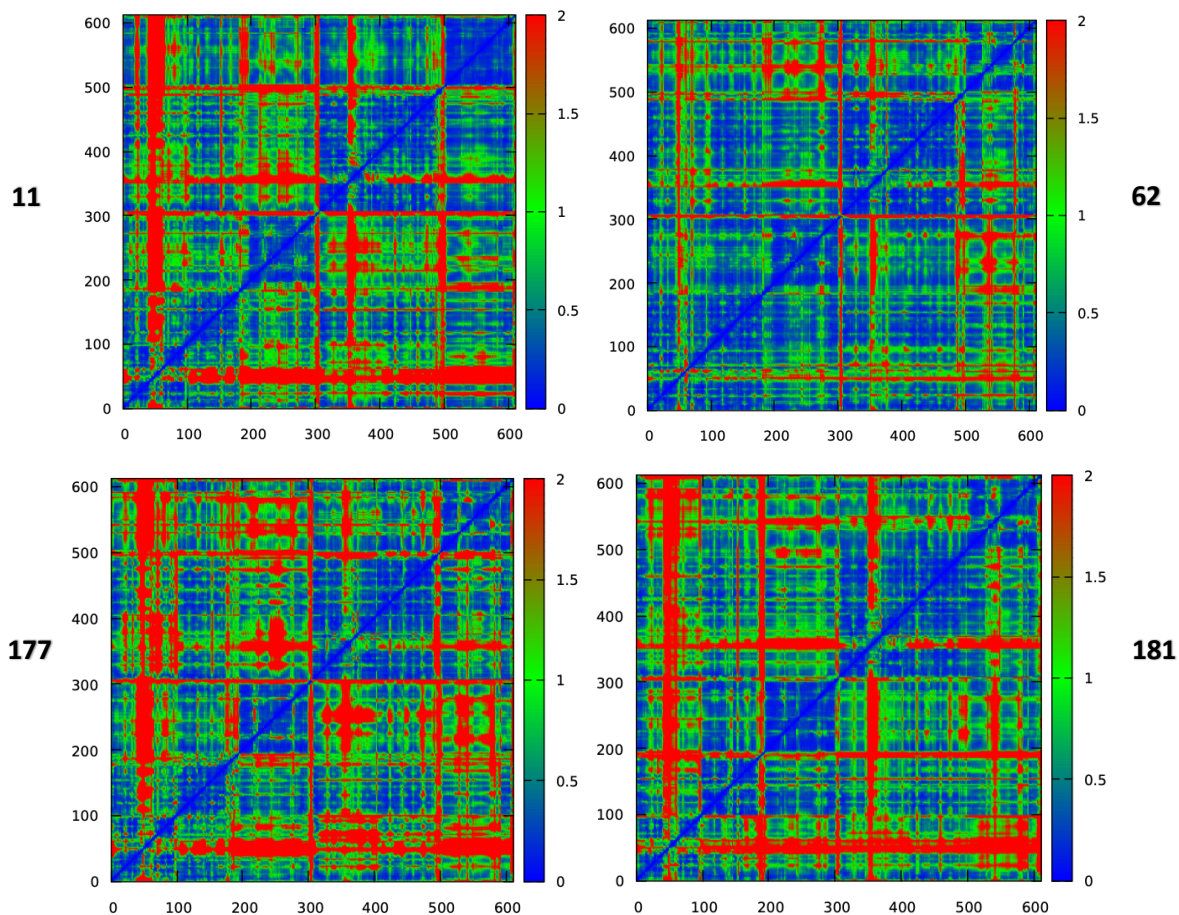
**Graph S1.** Timelines of compounds **71**, **177**, **189** and **193**.



**Graph S2.** Protein-ligand interaction diagrams of compound **71**, **177**, **189** and **193**. Highlighted in yellow interactions we hypothesized related to activity.



**Figure S2.** Distance fluctuation matrices for dimeric NSP5-ligand complexes, values are in Å.

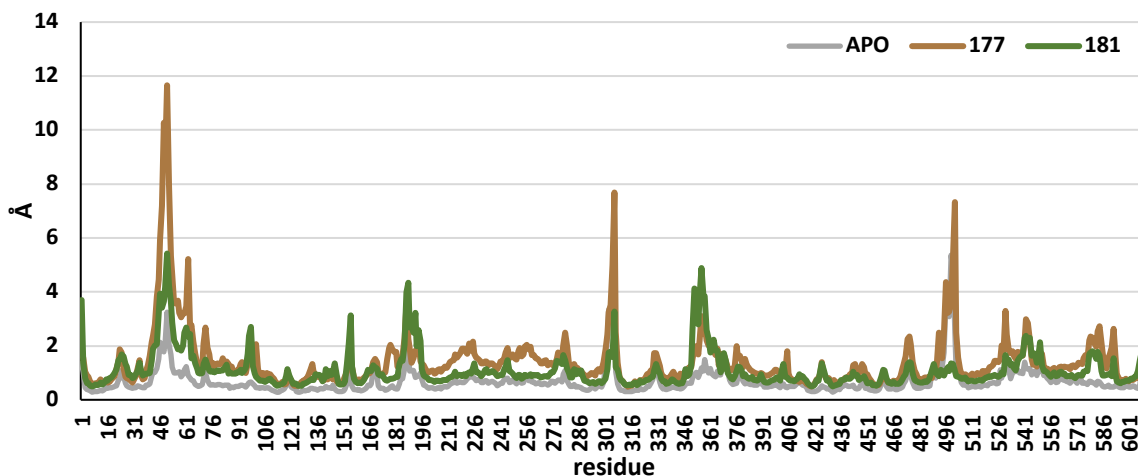




**Figure S3.** Dimeric NSP5, chain A is green, chain B is colored magenta. Segment 40-60 is highlighted in yellow, segment 185-200 is depicted in cyan.



**Graph S3.** Plot of absolute magnitude of pairwise changes ( $\text{\AA}$ ) between complexes **177** (bronze), **181** (green) and apo (grey).



**Graph S4.** Plot of absolute magnitude of pairwise changes ( $\text{\AA}$ ) between complex **62** (water blue) and apo (grey).

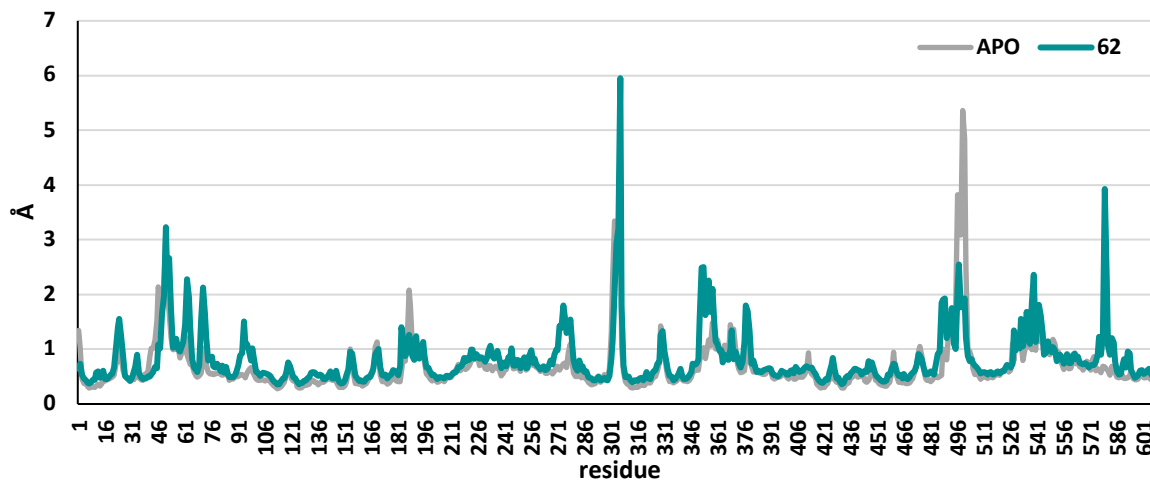
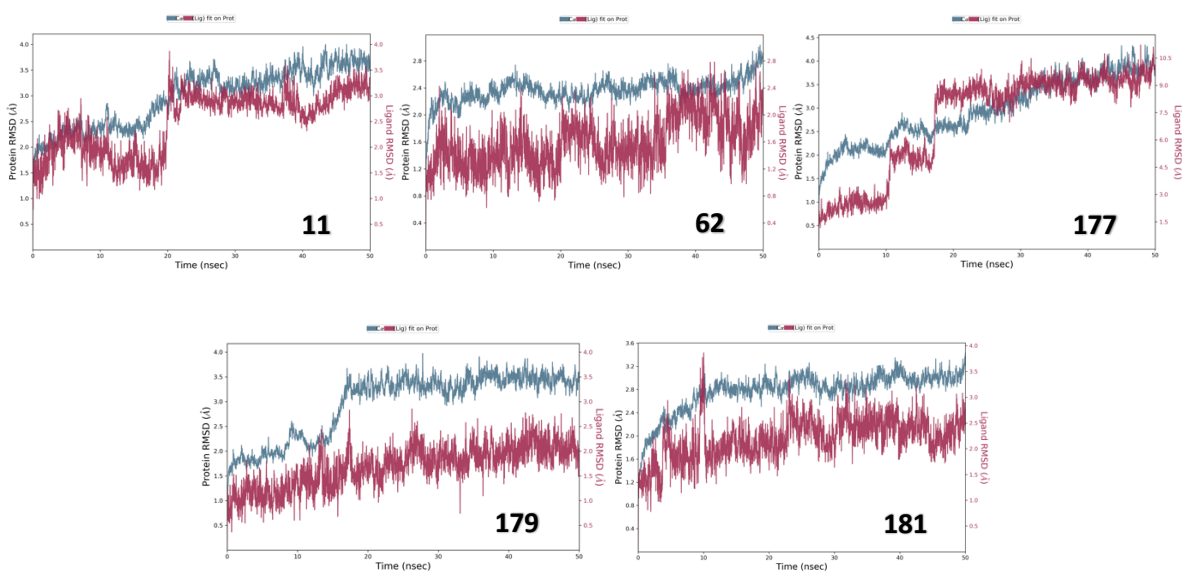


Figure S4. RMSD plots of protein C $\alpha$  and compounds **11**, **62**, **177**, **179** and **181** fit on protein.



# Chapter 3

## 3.1 General Discussion

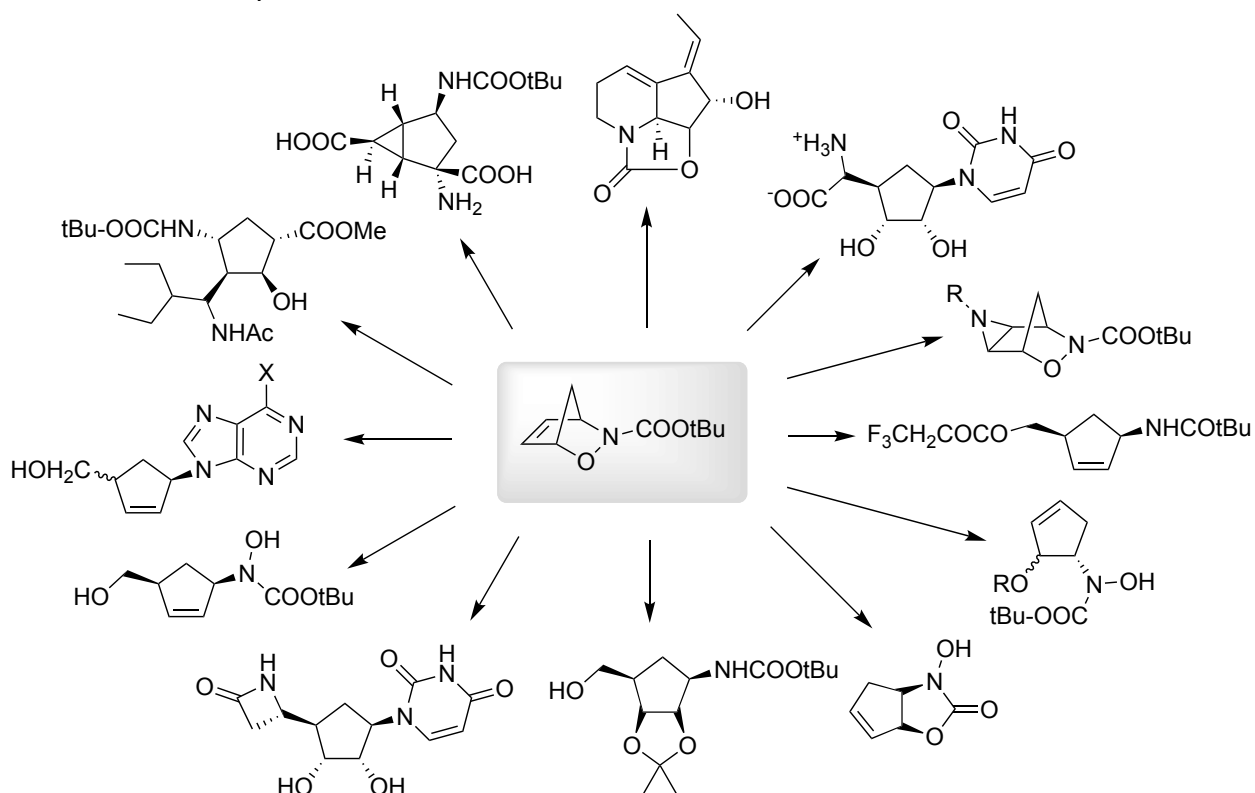
In the scenario of pressing need for new drugs to face future outbreaks, the difficult challenge of creating novel medications that combine high activity and selectivity, drug-likeness, and good pharmacokinetic qualities falls on medicinal chemists. Chemists are attempting to find ways to simplify synthetic processes through the use of techniques such as flow chemistry, polymer-assisted solution-phase synthesis, and microwave-assisted organic synthesis. Organic synthesis benefits from quick and reliable reactions; novel approaches often make unique use of click chemistry, ring-closing metathesis, and multicomponent reactions. Over the last ten years, the preparation of novel and intricate chemical structures has been made possible by the advancement of synthetic methodologies.<sup>1</sup>

Cycloaddition reactions are a mature topic of organic chemistry, but abundantly used in different areas where chemistry represents the pivotal step for reaching the desired target. Books and reviews dedicated to cycloaddition reactions deal typically with either the synthesis of heterocycles, or general synthetic applications, or specific uses in natural product synthesis, or describe the use of a class of organic compounds as partners in the cycloaddition reactions.<sup>2</sup>

In this Thesis to reach our objective we exploited two breakthrough approaches that have entered in the portfolio of common chemical reactions used by medicinal chemists: Hetero Diels-Alder transformations and 1,3-dipolar cycloadditions. The improvements that these reactions have introduced include high chemical efficiency and selectivity, good versatility and operational simplicity.<sup>1</sup>

## HDA Cycloaddition of Nitrosocarbonyls

Nitrosocarbonyl chemistry has emerged as a versatile and powerful tool in synthetic organic chemistry, offering a plethora of opportunities for the construction of diverse molecular scaffolds. A recent review provides a comprehensive overview of the unique reactivity and multifaceted applications of nitrosocarbonyl compounds in the synthesis of complex molecular architectures discussing key methodologies, recent advances, and notable examples that showcase the utility of nitrosocarbonyl chemistry in the creation of novel scaffolds.<sup>3</sup> After the HDA cycloaddition, diverse transformations such as rearrangements, and additions to multiple functional groups, nitrosocarbonyl compounds offer a wide range of possibilities for scaffold elaboration (**Figure 1**). This versatility allows chemists to access an extensive library of molecular architectures, facilitating the synthesis of compounds with diverse properties and applications. The unique reactivity of nitrosocarbonyl compounds opens up avenues for creative and efficient synthetic routes, making them invaluable in the design and construction of novel molecules in medicinal chemistry.<sup>3</sup>

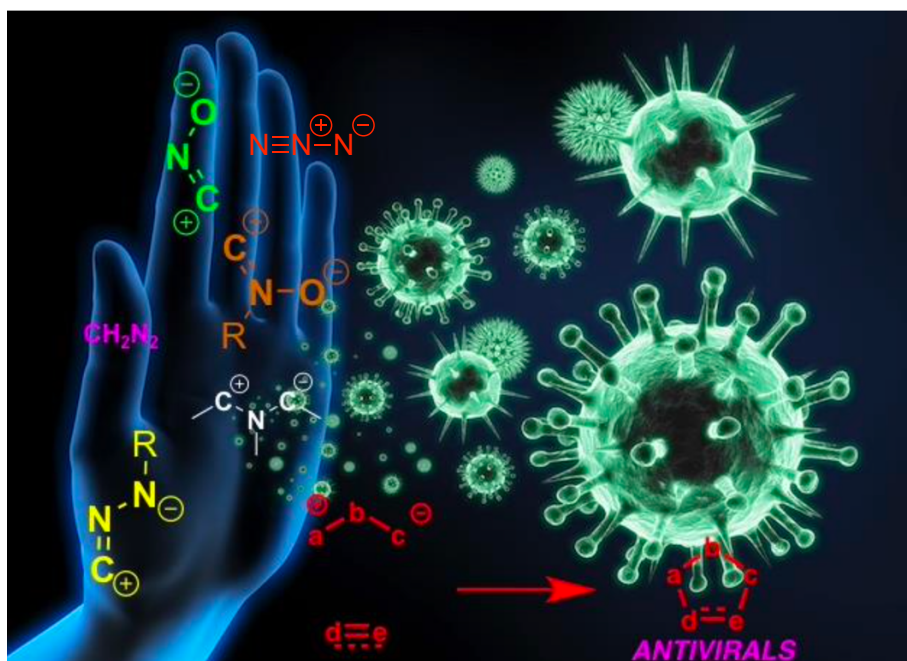


**Figure 1.** Synthetic uses of *N*-carboxyt.butyl-2,3-oxazanorborn-5-ene obtained from nitrosocarbonyl chemistry. Adapted from Ref. 3

## 1,3-Dipolar Cycloadditions

The 1,3-dipolar cycloaddition somewhat represents a valuable example of “ancient methodology” modernly used in chemistry and in modern organic synthesis.

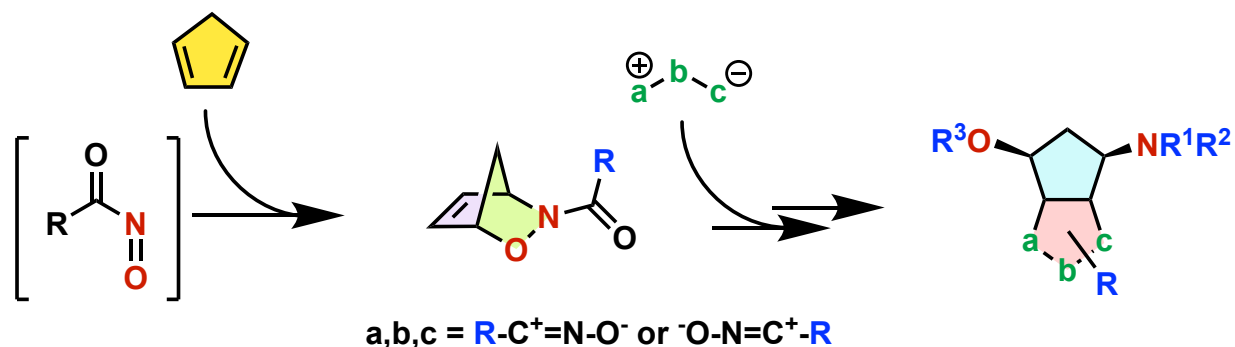
1,3-Dipolar cycloadditions offer a highly versatile and precise methodology for the decoration of molecular scaffolds, presenting an invaluable tool in the realm of synthetic organic chemistry. The controlled addition of 1,3-dipoles, of propargyl-allenyl type or allyl type, to dipolarophiles allows for the selective incorporation of functional groups onto predefined positions within a molecular framework.<sup>2</sup> This strategic approach enables chemists to intricately modify and decorate scaffolds with a diverse array of substituents, tailoring the overall structure to meet specific design requirements. The regio- and stereochemistry inherent in 1,3-dipolar cycloadditions provide an exquisite level of control, ensuring the accurate placement of functional moieties. As a result, this synthetic strategy has proven instrumental in the synthesis of complex molecules, pharmaceuticals, and materials, showcasing its broad applicability and significance in the creative design and functionalization of molecular scaffolds. Recent review reported the use of 1,3-dipolar cycloaddition reactions of azides, nitrones, nitrile oxides, nitrile imines, nitrosocarbonyls, azomethine ylides, diazoalkanes, in the light of their application for the preparation of key intermediates in the design and synthesis of compounds that were tested for their antiviral activities against a variety of viruses.<sup>2,4</sup> The products obtained from these pericyclic reaction approaches were tested for their activities in terms of blocking the virus replication and the relevant biological data are highlighted.<sup>2,4</sup>



**Figure 2.** Azides, nitrones, nitrile oxides, nitrile imines, azomethine ylides, diazoalkanes, in the light of their synthetic potential in developing antivirals. Adapted from Ref. 2,4.

## In the present work

In this thesis, we strategically harnessed the reactivity of nitrosocarbonyls through Hetero Diels-Alder reactions to obtain a foundational scaffold. This key synthetic step served as the precursor for further molecular elaboration; subsequently, we employed the versatility of 1,3-dipolar cycloadditions to decorate the obtained scaffold with precision and efficiency (**Figure 3**).



**Figure 3.** General pathway to the bicyclic scaffold.

The robustness of these sequential processes was demonstrated through the reliable generation of diverse molecular architectures. Leveraging the regio- and stereochemical control inherent in both nitrosocarbonyl Hetero Diels-Alder reactions and 1,3-dipolar cycloadditions, we successfully tailored the structures of the resulting compounds. This methodology not only showcases the synthetic impact of nitrosocarbonyls but also underscores the reliability and strategic utility of 1,3-dipolar cycloadditions as powerful tools for scaffold decoration in the context of organic synthesis. The synergistic integration of these reactions in our work to hit two different SARS-CoV-2 targets, the polymerase and the protease, exemplifies a systematic and effective approach to molecular design and functionalization.

In the context of our investigation targeting the polymerase, our synthetic strategy, starting with stable anthracene nitrile oxide as the precursor, yielded a bicyclic scaffold that proved to be instrumental in mimicking nucleoside structures. The incorporation of stereoordered 1,3-amino alcohol groups, facilitated through hetero-Diels-Alder (HDA) reactions, played a pivotal role in shaping this scaffold. Subsequently, the strategic construction of the 6-chloropurine heterobase and the introduction of the phosphate group further enriched the molecular architecture. The resulting bicyclic scaffold not only mirrored key features of nucleosides but also demonstrated a robust and modular platform, through possible substitutions of the chlorine and bromine atoms, for the rational design of molecules targeting the polymerase. This synthetic pathway, rooted in stable starting materials and guided by precise reactions, showcases the potential for tailored molecular design in the pursuit of novel antiviral agents or other therapeutic interventions involving the polymerase enzyme.<sup>4</sup> Nevertheless, synthetic steps concerning phosphorylation process were tedious and many reaction's conditions had to be tested before reaching our goal.

The attempts to heighten structural complexity by incorporating the 6-chloropurine base and phosphate group did not achieve a favorable balance with increased selectivity. Instead, the resulting compounds exhibited cytotoxicity levels nearly equivalent to that of inhibitory activity. From a computational perspective, comparable considerations apply as well: the RdRp system's complexity is evident in its substantial dimensions, to expedite calculations NSP7 and NSP8s auxiliary subunits, completing the NSP12 polymerase complex, were not considered.<sup>6</sup>

In the second case, a distinct paradigm emerges as a result of the notable variability of functional groups that can be introduced to the substituent of the isoxazoline ring. This variability is achieved through the strategic utilization of diverse nitrile oxides, ultimately leading to the formation of a regio-defined scaffold, SYN or ANTI (referring to relative orientations of the amino and R-substituent of the corresponding nitrile oxide), whose way to the aminol proved once again to be reliable, with high yields and easy workups. The significance of this approach becomes particularly evident in its heightened relevance to the process of targeting the protease. The gentle introduction of diverse functional groups imparts a variability to the molecular architecture, allowing for a tailored and adaptable response to the target as exemplified also by the easy introduction of aminoacidic branches in some compounds. Overall, the synthetic flexibility conferred by the diverse nitrile oxides employed underscores a nuanced return to prominence of 1,3-dipolar reactions and offers a promising avenue for the development of targeted inhibitors with enhanced efficacy in protease modulation.

Concerning the other scaffold, the choice of the solid-state method, coupled with the catalytic influence of DEAE-C resin, not only facilitated a linear progression of the synthesis of the 5-hydroxiisoxazolidine but also contributed to its rapid completion and rapid workup.

From a computational standpoint, the exploration of allosteric pockets and the methods employed to acquire receptor structures proved to be more intriguing. Fortunately, in the present scenario, allowed saving resources the freely available 500  $\mu$ s-long simulation from the SCoV2-MD database exploited to take into account the flexibility of the target.<sup>7</sup> Even though the majority of the compounds selected to target the protease share a common cyclopentane-fused isoxazoline scaffold with those directed at the polymerase, the introduction of distinct decorations has a pronounced effect on cytotoxicity. Notably, these variations in structural embellishments result in a significant reduction in cytotoxic levels indicated by the available preliminary tests. This observation underscores the intricate relationship between molecular structure and biological activity, emphasizing the importance of subtle modifications in compound design to achieve desired therapeutic effects while mitigating undesirable cytotoxicity.

Tests to assess inhibition are currently in progress, and the outcomes will be available for inclusion in the final dissertation.

Because of the low cytotoxicity and the ease of synthesis of new aminolic compounds and derivatives, it may be worth to test them against other viruses such as flu or even West Nile.

## 3.2 References

- 1) Colombo, M.; Peretto, I. Chemistry strategies in early drug discovery: an overview of recent trends *Drug Discov. Today* **2008**, *13*, 677-684.
- 2) Faita, G.; Leusciatti, M.; Quadrelli, P. Synthesis and biological activity of potential antiviral compounds through 1,3-dipolar cycloadditions; Part1: general aspects and reactions of azides *Arkivok* **2022**, *part viii*, 19-78.
- 3) Memeo, M. G.; Quadrelli, P. Generation and Trapping of Nitrosocarbonyl Intermediates *Chem. Rev.* **2017**, *117*, 2108-2200.
- 4) Faita, G.; Mella, M.; Quadrelli, P. Synthesis and biological activity of potential antiviral compounds through 1, 3-Dipolar cycloadditions. Part 2: nitrones, nitrile oxides and imines, and other 1,3-dipoles *Arkivok* **2022**, *part viii*, 257-294
- 5) Leusciatti, M.; Macchi, B.; Marino-Merlo, F.; Mastino, A.; Morra, G.; Quadrelli, P. Inhibition of the RNA-Dependent RNA-Polymerase from SARS-CoV-2 by 6-Chloropurine Isoxazoline-Carbocyclic Monophosphate Nucleotides *ACS Omega* **2023**, *8*, 36311-36320.
- 6) Hillen, H. S.; Kocic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase *Nature* **2020**, *584*, 154-156.
- 7) Torrens-Fontanals, M.; Peralta-García, A.; Talarico, C.; Guixà-González, R.; Giorgino, T.; Selent, J. SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions *Nucleic Acids Res.* **2021**, *50*, D858-D866.

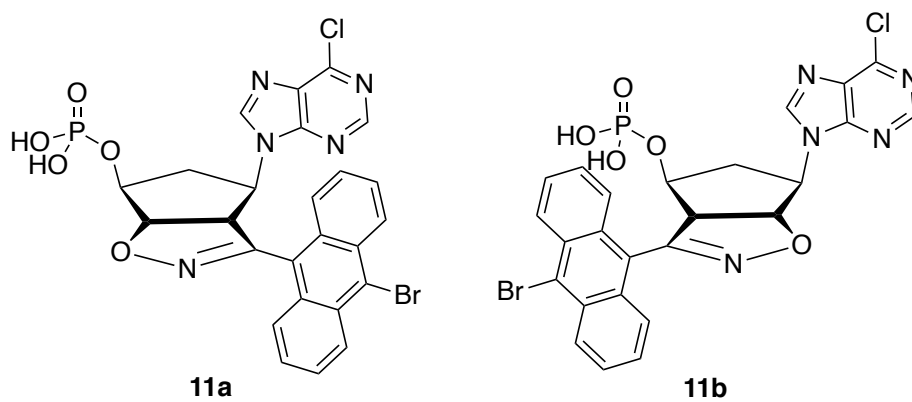


# Chapter 4

## 4.1 Final Conclusions

In conclusion, our interdisciplinary approach encompassing computational studies, synthesis, and biological tests has yielded significant insights into the development of potential antiviral candidates, with a focus on targeting SARS-CoV-2. The computational phase played a pivotal role in identifying promising candidate molecules, demonstrating its effectiveness in streamlining the synthesis process.

The RNA-dependent RNA-polymerase has been targeted leveraging the chemistry of nitrosocarbonyl, through a well-established protocol we successfully utilized stable nitrile oxide and implemented a phosphorylation protocol, showcasing innovative synthetic methodologies. The ability of the 1,3-dipolar cycloaddition reactions and the hetero Diels–Alder cycloadditions of nitrosocarbonyl intermediates (easily obtained from nitrile oxides through mild oxidation) to produce a variety of heterocyclic scaffolds extends its importance to many areas of organic synthesis, and the synthesis of antiviral compounds represents a topic in constant growing. A first evaluation has been performed through docking computational studies. Although the nature of nucleoside/nucleotides analogs, they resemble NNRTI that have not yet been found/synthesized to contrast SARS-CoV-2. The computational studies revealed the possibility for compounds of types **11a** and **11b** to give positive responses in terms of contrast to SARS-CoV-2, with compounds of type **11b** being the most promising (**Figure 1**). While the synthesized compounds exhibited inhibitory activity in accordance with our computational predictions with one compound (**11b**) being more active than the other (**11a**), an unexpected observation arose during the biological tests, revealing a noteworthy cytotoxicity (**Table 1**).<sup>1</sup>



**Figure 1.** Structures of synthesized compounds of types **11a** and **11b**.

	IC <sub>50</sub> (cell/virus-free fluorometric assay) (μM)	CC <sub>50</sub> Hep-2 cells (Tryphan blue assay) (μM)	CC <sub>50</sub> HepG2 cells (Tryphan blue assay) (μM)	CC <sub>50</sub> Hep-2 cells (MTS assay) (μM)	CC <sub>50</sub> HepG2 cells (MTS assay) (μM)
<b>11a</b>	43.83 ± 1.07	46.79 ± 1.18	79.67 ± 1.10	30.96 ± 1.2	28.14 ± 1.3
<b>11b</b>	29.31 ± 1.05	22.36 ± 1.04	22.85 ± 1.19	23.64 ± 1.3	32.37 ± 1.4

**Table 1.** Summary of biological results (for details see **Figure 6** in **Chapter 1**).

This finding emphasizes the complexity of the relationship between molecular structure and biological response, highlighting the importance of integrating computational and experimental approaches. The results underscore that computational screening is an indispensable initial step in selecting structurally relevant candidates, but biological assays are essential in steering the synthesis towards compounds with desired properties.

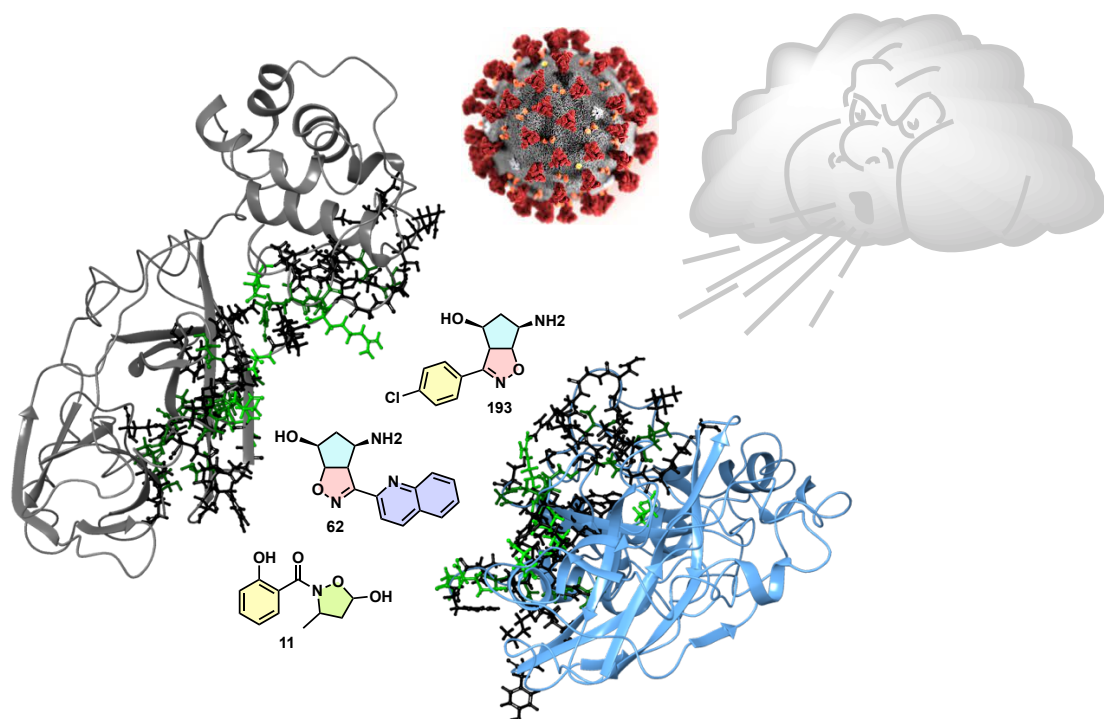
These findings have broader implications for antiviral drug discovery, suggesting that our synthetic methods can be applied to the preparation of antiviral candidates targeting respiratory viruses, in the case at hand the current global health concern SARS-CoV-2. As we continue to refine our understanding of structure-activity relationships through computational predictions, making **11b** more interesting for future design of structural modifications, and validate these insights through biological assays, our work contributes to the ongoing efforts to develop effective antiviral therapies.

The scientific investigation concerning the second target, the protease (M<sup>Pro</sup>), marks a significant advancement in the battle against SARS-CoV-2 by unveiling a strategic approach that seamlessly integrates computational analyses with synthetic chemistry. The discovery of allosteric pockets within the SARS-CoV-2 main protease, achieved through computational methods<sup>2</sup> and corroborated by experimental findings,<sup>3-5</sup> has illuminated a promising avenue for targeted antiviral intervention. The subsequent design and synthesis of novel antiviral molecules, harnessing the distinctive chemistry developed by our research group, showcase a pioneering effort in allosteric inhibitors drug design.<sup>6-11</sup>

As a matter of fact, identifying and focusing on allosteric sites can lead to the discovery of new therapeutic classes and the expansion of lead compounds' chemical space. Most crucially, the identification of novel compounds that target allosteric regions offered a workable approach to the quest for novel inhibitors of protein-protein interactions. Our goal, though, was to find active hits by employing data on an allosteric pocket that we directly gathered from the analysis of a dynamic molecular machine. In fact, the key to our discovery strategy of possible drug-binding sites and protein-protein interactions inhibitors was explicitly taking into account the flexibility

of the target. Protein's molecular dynamic data can be used to reveal potential binding states of the receptor that may exist on the protein's energy landscape but may not be readily visible in a single-structure depiction. Thus, the identification of alternative states can lead to the identification of potential allosteric binding sites, permit the occupancy of a site by structurally distinct ligands, or direct efforts towards the construction of novel leads that functionally and structurally modify established leads to target-specific receptor geometries. As we navigate the dynamic landscape of viral infections, the collaborative synergy between computational efforts and synthetic chemistry showcased in this study emphasizes the importance of a holistic approach. The deliberate exploitation of pericyclic reactions, ultimately in the form of 1,3-dipolar cycloadditions and catalyzed nucleophilic addition to  $\alpha,\beta$ -unsaturated aldehydes, in the synthetic process adds a layer of renewal, enabling the creation of compounds with tailored structures designed to interact specifically with the identified allosteric sites.

The use of 1,3-dipoles offers a unique method for the preparation of a variety of compounds with potential biological activity. It must be taken into account that heteroatom-containing cycloadducts may be transformed in other functionalized molecules, cyclic or acyclic. Furthermore, the ability of the 1,3-dipole to generate rings even containing stereocenters in a single synthetic operation is a pivotal step when a valuable compound has to be prepared.



**Figure 2.** Designed compounds inducing SARS-CoV-2 M<sup>Pro</sup> dimer dissociation. Residues participating dimerization interface are shown as spheres and sticks colored from black to green by the increasing number of specific interactions of each residue (black was set to 0 and green at 3).

Preliminary biologic assays indicate low cytotoxicity of the compounds tested (**Table 2**). In vitro inhibitory activity tests are ongoing in collaboration with Doctress Francesca Marino-Merlo from the University of Messina.

	CC <sub>50</sub> HepG2 cells (MTS assay) (μM)
<b>11</b>	>1000
<b>189</b>	213.7
<b>193</b>	986.1

**Table 2.** Summary of preliminary biological results (for details see **Figure 15** in **Chapter 2**).

With this approach we not only demonstrated the versatility of discussed chemistry, but we also highlight its utility in crafting molecules with enhanced antiviral properties. The successful translation of computational insights into tangible antiviral candidates underscores the efficacy of this multidisciplinary strategy. This research not only contributes to our understanding of SARS-CoV-2 but also sets a precedent for future endeavors in antiviral drug discovery. The fusion of computational analyses with the innovative application of chemical pillars holds promise for the development of effective antiviral therapeutics, reinforcing the significance of scientific collaboration in addressing global health challenges.

Ultimately, this research exemplifies the synergy between computational and experimental approaches, paving the way for the rational design and synthesis of novel therapeutic agents.

## 4.2 References

- 1) Leusciatti, M.; Macchi, B.; Marino-Merlo, F.; Mastino, A.; Morra, G.; Quadrelli, P. Inhibition of the RNA-Dependent RNA-Polymerase from SARS-CoV-2 by 6-Chloropurine Isoxazoline-Carbocyclic Monophosphate Nucleotides *ACS Omega* **2023**, *8*, 36311-36320.
- 2) Morra, G.; Neves, M. A. C.; Plescia, C. J.; Tsustsumi, S.; Neckers, L.; Verkhivker, G.; Altieri, D, C.; Colombo, G. Dynamics-Based Discovery of Allosteric Inhibitors: Selection of New Ligands for the C-terminal Domain of Hsp90 *J. Chem. Theory Comp.* **2010**, *6*, 2978-2989.
- 3) Alzyoud, L., Ghattas M. A., Atatreh N. Allosteric Binding Sites of the SARS-CoV-2 Main Protease: Potential Targets for Broad-Spectrum Anti-Coronavirus Agents *Drug Des. Devel. Ther* **2022**, *16*, 2463-2478.
- 4) Goyal B.; Goyal D. Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy *ACS Comb. Sci.* **2020**, *22*, 297–305.
- 5) Hu T.; Zhang Y.; Li L.; Wang K.; Chen S.; Chen J.; Ding J.; Jiang H. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure *Virology* **2009**, *388*, 324–334.
- 6) Quadrelli, P.; Mella, M.; Legnani, L.; Al-Saad, D. From Cyclopentadiene to Isoxazoline-Carbocyclic Nucleosides; Synthesis of Highly Active Inhibitors of Influenza A Virus H1N1 *Eur. J. Org. Chem.* **2013**, 4655-4665.
- 7) Al-Saad, D.; Memeo, M. G.; Quadrelli, P. Pericyclic Reactions for Anti-HPV Antivirals: Unconventional Nucleoside Analogue Synthesis via Nitrosocarbonyl Chemistry *Chemistry Select* **2017**, *2*, 10340-10346.
- 8) Corti, M.; Leusciatti, M.; Moiola, M.; Mella, M.; Quadrelli, P. Nitrosocarbonyl Carbohydrate Derivatives: Hetero Diels-Alder and Ene Reaction Products for Useful Organic Synthesis *Synthesis* **2020**, *52*, A-M.
- 9) a) P. Quadrelli, R. Scrocchi, P. Caramella, A. Rescifina, A. Piperno, *Tetrahedron* **2004**, *60*, 3643–3651; b) P. Quadrelli, M. Mella, S. Carosso, B. Bovio, P. Caramella, *Eur. J. Org. Chem.* **2007**, 6003–6015; c) P. Quadrelli, M. Mella, G. Assanelli, *Tetrahedron* **2007**, *64*, 7312– 7317; d) P. Quadrelli, A. Piccanello, N. Vazquez Martinez, B. Bovio, M. Mella, P. Caramella, *Tetrahedron* **2006**, *62*, 7370–7379; e) M. Savion, M. G. Memeo, B. Bovio, G. Grazioso, L. Legnani, P. Quadrelli, *Tetrahedron* **2012**, *68*, 1845–1852.
- 10) Memeo, M. G.; Mella, M.; Montagna, V.; Quaderelli, P. Design, Synthesis, and Conformational Analysis of Proposed b-Turn Mimics from Isoxazoline-Cyclopentane Aminols *Chem. Eur. J.* **2015**, *21*, 16374-16378.
- 11) a) Grundmann, C.; Grünanger, P. *The Nitrile Oxides* Springer- Verlag, Heidelberg, **1971**; b) Caramella P., Grünanger, P. *1,3- Dipolar Cycloaddition Chemistry* (Ed.: A. Padwa), John Wiley & Sons, Inc. New York, **1984**, vol. 1, p. 291–392; c) Grünanger, P.; Finzi P., V. *Isoxazoles*, John Wiley & Sons, Inc., New York, **1999**, vol. 1 and 2.

# Chapter 5

## 5.1 Publications

the following articles were published:

- Moiola, M.; Leusciatti, M.; Quadrelli, P. Reactions of 1,2,4-Oxadiazole[4,5-*a*]pyridinium Salts with Alcohols: the Synthesis of Alkoxybutadienyl 1,2,4-Oxadiazoles *Chemistry Open* **2020**, *9*, 195-199.
- Corti, M.; Leusciatti, M.; Moiola, M.; Mella, M.; Quadrelli, P. Nitrosocarbonyl Carbohydrate Derivatives: Hetero Diels-Alder and Ene Reaction Products for Useful Organic Synthesis *Synthesis* **2021**, *53*, 574-586.
- Leusciatti, M.; Mannucci, B.; Recca, T.; Quadrelli, P. Turn-folding in fluorescent anthracene-substituted cyclopenta[*d*]isoxazoline short peptides *RCS Adv.* **2021**, *11*, 19551-19559.
- Ahmed, S. M.; Hussain, F. H. S.; Leusciatti, M.; Mannucci, B.; Mella, M.; Quadrelli, P. Phosphorylation of 10-bromoanthracen-9-yl-cyclopenta[*d*]isoxazole-6-ols: chemistry suitable for antivirals *Arkivoc* **2022**, *part v*, 20 pages.
- Faita, G.; Leusciatti, M.; Quadrelli, P. Synthesis and biological activity of potential antiviral compounds through 1,3-dipolar cycloadditions; Part 1: general aspects and reactions of azides *Arkivoc* **2022**, *part viii*, 19-78.
- Leusciatti, M.; Macchi, B.; Marino-Merlo, F.; Mastino, A.; Morra, G.; Quadrelli, P. Inhibition of the RNA-Dependent RNA-Polymerase from SARS-CoV-2 by 6-Chloropurine Isoxazoline-Carbocyclic Monophosphate Nucleotides *ACS Omega* **2023**, *8*, 36311-36320.

the following communications were given at the conferences:

- Oral communication “New Fluorescent  $\beta$ -Turn Mimics” at 1<sup>st</sup> Virtual Symposium on Pericyclic Reactions and Synthesis of Carbo- and Heterocyclic Systems (CIRP). November 25, 2020
- Poster “Design and Synthesis of (Nor)Nucleotide Analogs Inhibitors of SARS-CoV-2 RdRp through Nitrosocarbonyls Chemistry” at 23rd EuroQSAR: 23rd European Symposium on Quantitative Structure-Activity Relationship; Integrative Data-Intensive Approaches to Drug Design. Heidelberg, Germany - September 26-30, 2022.
- Oral communication “Nitrosocarbonili: Design e Sintesi di derivati Antivirali” at congress CIRP 2023: XIX Convegno Nazionale sulle Reazioni Pericicliche e Sintesi di Etero e Carbocicli. Reggio Calabria with the patronage of the University of Messina – June 29-30, 2023.

# Chapter 6

## 6.1 Attended courses

Throughout the Ph.D. program, the following courses were successfully completed, with exams being passed:

- **Analisi e Chimica degli Alimenti (6CFU)**, Prof. Daniele Merli, Grade: 30L/30
- **Progettazione e Sviluppo di Farmaci (4CFU)**, Prof.ssa Daniela Rossi, Grade: 30/30
- **Proprietà Fisiche delle Polveri (2CFU)**, Prof.ssa Franca Ferrari, Grade: 30/30
- **Tecniche NMR in Chimica Organica (6 CFU)**, Prof.ssa Mariella Mella, Grade: 30/30
- **Farmaci Peptidici e Peptidomimetici (6 CFU)**, Prof.ssa Simona Collina, Grade: 30/30

# Chapter 7

## 7.1 Funding

Financial support by the **University of Pavia**, **University of Rome**,  
**University of Messina** and **CNR** is gratefully acknowledged.



Consiglio Nazionale  
delle Ricerche

We also thank **Steroid S.p.A.** and **Curia S.r.l.** for their financial support to  
research activities.

