

SARS-CoV-2 virus-host interaction: a molecular, genetic, and docking approach

Luís Emanuel Queirós dos Reis

D 2025



Luís Emanuel Queirós dos Reis. SARS-CoV-2 virus-host
interaction: a molecular, genetic, and docking approach

D.ICBAS 2024



SARS-CoV-2 virus-host interaction: a molecular, genetic, and docking approach

Luís Emanuel Queirós dos Reis

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR



DOCTORAL DEGREE
BIOMEDICAL SCIENCES

SARS-CoV-2 virus-host interaction: a molecular, genetic, and docking approach

Luís Emanuel Queirós dos Reis

D

2025

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR



Luís Emanuel Queirós dos Reis

SARS-CoV-2 virus-host interaction: a molecular, genetic, and docking approach

Candidature Thesis for the Degree of Doctor in
Biomedical Sciences

Doctoral Programme at the University of Porto
Abel Salazar Biomedical Sciences Institute

Supervisor – Professor João Rodrigo Mesquita
Associate Professor – Abel Salazar Biomedical
Sciences Institute - University of Porto

Co-Supervisor – Professor Marcella Bassetto
Senior Lecturer in Chemistry - Faculty of Science
and Engineering, Swansea University

Co-Supervisor – Professor Andrea Brancale
Chair of Medicinal Chemistry - University of
Chemistry and Technology, Prague

Declaration of honour

I declare that this thesis is my own work and has not been previously used in any other course or curricular unit, either at this or any other institution. References to other authors (statements, ideas, thoughts) strictly adhere to attribution rules and are properly cited in the text and in the bibliography, in accordance with referencing standards. I am aware that the practice of plagiarism and self-plagiarism constitutes an academic offense.

Declaro que a presente tese é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Porto, 20 de fevereiro de 2025

Luís Emanuel Queirós dos Reis

Institutional support for research

The research reported in this thesis was performed at:

- Abel Salazar Biomedical Sciences Institute - University of Porto
- Faculty of Science and Engineering – Swansea University
- School of Pharmacy and Pharmaceutical Sciences – Cardiff University
- Department of Microbiology - Oslo University Hospital



Financial support

This thesis was supported by a PhD fellowship, under the program DOCTORATES 4 COVID-19, with the reference SFRH/2020.10230.BD, granted by *Fundação para a Ciência e a Tecnologia*, through national funds and by the EEA/Norway GRANTS with the reference FBR_OC2_53 - COVinhib – “Inactivating the SARS-CoV-2 spike protein: drug discovery in the fatty acid binding pocket”.



Acknowledgments

First, I am deeply thankful to my supervisor, Professor João Mesquita, for enabling this PhD, without whom it would neither have been started nor finished. Thank you for showing me what science is about, for challenging me to this project, for always finding solutions to keep moving forward. Your mentorship and companionship throughout this journey have been invaluable, and I am truly grateful for everything I have learned under your supervision.

I am also massively grateful to my co-supervisors. To Professor Marcella Bassetto, for fully supporting our research, for going further than was required, in a true example of scientific excellence, openness, collaboration and dedication. Your guidance, support, and most importantly patience, during this entire project are truly valued. Thank you for teaching me so much. And for enabling my visit to Swansea University, where I explored an amazing culture, beautiful places, and met great people.

To Professor Andrea Brancale, your willingness to support a new project, to teach and share your expertise, and your commitment to seeing it through are a testament to the fact that science is, above all, made by people.

A special thanks must be directed at Dr. Mari Kaarbø, for accepting our invitation to join the later stages of this project, for being yet another example of scientific excellence, collaboration and dedication. For your guidance, patience and kindness during my time at the Oslo University Hospital, and for enabling two incredible experiences in the lovely country of Norway.

Additionally, to Professor Maria São José Alexandre and Professor Ana Colette Maurício for the opportunity to carry this research in the Virology Laboratory at FFUP and Department of Veterinary Science at ICBAS, respectively. A special thanks to all the staff at University of Porto and ICBAS.

I am especially thankful for the institutional support provided by the University of Porto, Swansea University, Cardiff University and Oslo University Hospital, for providing all the resources to create an environment that fosters scientific excellence.

Additionally, to the ERASMUS program and EEA grants, that enabled three scientific missions, and provided financial support to this project, as well as *Fundação para a Ciência e Tecnologia* for the PhD fellowship that supported the enrolment in the doctoral program and the completion of this thesis.

Throughout these four years, I had the opportunity to meet many amazing people along the way, with not only I learned a lot, whose help was invaluable, but also made the entire experience undoubtedly better by their company, friendship and support. To all the researchers and fellow PhD colleagues at ICBAS, Swansea University and Oslo University Hospital, thank you for the valuable knowledge exchange and for cultivating a positive work environment.

A special thanks to Martina, Rui, and Huda for their help, directly contributing to this project. None of this would, however, be possible without the unwavering support of my family and closest friends. To my parents and my brother, for the constant encouragement, support, and help—for being my safety net in all my projects—I could never thank you enough for all you have done for me. This PhD would never have happened without your help, for fostering an environment where my curiosity and dedication were rewarded. For always pushing me and believing in my potential, even when I doubted.

Finally, to my loving wife—words are not enough to describe how much you have done, as this achievement is as much yours as it is mine. Thank you for always pushing me forward, for being my guiding light in difficult moments, and for always bringing out the best in me. For being next to me in all the stages, and for making everything better. Your unwavering love, patience, and belief in me have been my greatest strength throughout this journey.

List of publications

In accordance with the *Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto*, the author of this thesis declares participation in the conception and execution of the experimental work, the analysis and interpretation of the data leading to the presented results, and the writing of the respective manuscripts. The manuscripts included in this thesis are original, have not been previously submitted in other thesis, and are provided in their complete form.

Within the framework of this thesis, the research conducted has resulted in the publication of the following manuscripts:

- 1) Queirós-Reis, L.; Gomes da Silva, P.; Gonçalves, J.; Brancale, A.; Bassetto, M.; Mesquita, J.R. SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *Int. J. Mol. Sci.* 2021, 22, 10836. <https://doi.org/10.3390/ijms221910836>
- 2) Queirós-Reis, L; Mesquita, J; Brancale, A; Bassetto, M. (2023). Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein - Confirmed and Potential Ligands. *Journal of chemical information and modeling.* 63. 10.1021/acs.jcim.3c00803
- 3) Queirós-Reis, L., Alvites, R.; Maurício, A.C.; Brancale, A.; Bassetto, M.; Mesquita, J.R. (2024). Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study. *International Journal of Molecular Sciences*, 26(1), 151. <https://doi.org/10.3390/ijms26010151>

Additionally, the following manuscript is currently under review:

- 1) Queirós-Reis, L; Alvites, R; Maurício, AC; Brancale, A; Bassetto, M; Mesquita, J.R. (2024). *In silico* discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding pocket. *ACS Omega*

Abstract

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic represents one of the most important global health crises in modern history, claiming millions of lives and straining healthcare systems worldwide. However, SARS-CoV-2 is not the first pathogenic coronavirus (CoV) to infect humans, nor even the first highly pathogenic one. Coronaviruses are widespread in nature, primarily infecting mammals and birds, and their ability to adapt to new hosts and cross interspecies barriers is remarkable. This adaptability is facilitated by their large RNA genome and advanced replication mechanisms, which promote rapid evolution and the emergence of novel strains capable of zoonotic transmission. Unlike earlier highly pathogenic human CoVs, SARS-CoV-2 demonstrated a unique combination of transmissibility and pathogenicity, leading to unprecedented challenges in prevention, treatment, and containment. The first reports of SARS-CoV-2 human infection appeared in December 2019, and by April 2020, half of the world's population was under lockdown, causing severe socioeconomic disruption. The health consequences, on the other hand, were staggering, as the virus overwhelmed healthcare systems worldwide. Vulnerable populations, particularly the elderly and those with pre-existing conditions, were disproportionately affected. The demand for effective antiviral therapies became critical, yet it took over 10 months for the first antiviral drug, remdesivir, to receive full approval for use in hospitalized patients.

The ultimate goal of this project was to provide a contribution on potential therapeutical alternatives, as this thesis aimed to investigate the virus-host interaction in SARS-CoV-2 to understand what distinguishes it from other CoVs and to explore potential targets for antiviral therapy, focusing on the Spike (S) protein as a key determinant of viral entry into host cells. This trimeric glycoprotein interacts with the angiotensin-converting enzyme 2 (ACE2) receptor, mediating attachment and promoting membrane fusion. The receptor-binding domain (RBD), present in each S protein monomer, directly interacts with the ACE2 receptor on host cells, facilitating viral attachment and subsequent entry. This interaction drives the conformational changes necessary for fusion between the viral and host cell membranes. While the RBD is a well-characterized site for therapeutic targeting, there is an intrinsic challenge, the receptor binding motifs are highly variable at the sequence level, as a viral evolution mechanism.

Among several targets in the S protein considered in this study, the Fatty Acid Binding Pocket (FABP) was selected as the most promising, formed at the interface between RBDs and independent from the binding motifs, while being conserved in highly pathogenic CoVs. The FABP plays a significant role in modulating the S protein's conformational dynamics, influencing its ability to recognize and infect host cells. To characterize the FABP and its potential as a therapeutic target, molecular dynamics simulations were conducted using a linoleic acid (LA)-FABP co-crystal structure as a model, providing insights into the structural stability of the FABP and key features of the binding pocket, including its accessibility and

flexibility. This structural information supported virtual screening studies, with commercially available compound libraries to identify small molecules capable of binding to the FABP.

Through docking studies, several potential candidates were identified, among which, one compound demonstrated micromolar antiviral activity against SARS-CoV-2 in experimental validation, in viral infection assays. In order to increase the potency of this hit compound, an optimization stage was performed, with structural analogues selected based on similarity and potential to explore new interactions and enhance pocket occupation, based on insights from preliminary docking results and molecular dynamics analyses.

Experimental validation showed an optimized compound achieving nanomolar activity against SARS-CoV-2 and allowed a structure-activity relationship study, that further improved the understanding of FABP binding dynamics, which could guide future activity improvement. This work highlights the therapeutic potential of the FABP as a novel target for antiviral intervention against SARS-CoV-2. The discovery of a new scaffold with antiviral activity, progressing from micromolar to nanomolar activity, emphasizes the use of computer-based approaches in drug development efforts. The knowledge and tools developed during this research not only retain their applicability to support the discovery of potential new SARS-CoV-2 therapies but also might be the basis to address new emerging CoVs.

Keywords: SARS-CoV-2, Spike Glycoprotein, Fatty Acid Binding Pocket, Antiviral drug discovery, Computer aided drug design

Resumo

A pandemia causada pelo coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2) representa uma das mais significativas crises globais de saúde na história moderna, causando milhões de mortes e sobrecarregando os sistemas de saúde em todo o mundo. No entanto, o SARS-CoV-2 não é o primeiro coronavírus (CoV) patogénico a infectar humanos, nem sequer o primeiro altamente patogénico. Os CoV estão amplamente disseminados na natureza, infectando principalmente mamíferos e aves, e a sua capacidade de se adaptarem a novos hospedeiros e ultrapassarem barreiras entre espécies é notável. Esta adaptabilidade é facilitada pelo seu extenso genoma de RNA e mecanismos de replicação avançados, que promovem a evolução rápida e a emergência de novas estirpes capazes de transmissão zoonótica. Ao contrário de outros CoVs humanos altamente patogénicos, o SARS-CoV-2 demonstrou uma combinação única de transmissibilidade e patogenicidade, conduzindo a desafios sem precedentes na prevenção, tratamento e contenção.

Os primeiros relatos de infeção entre humanos por SARS-CoV-2 surgiram em dezembro de 2019 sendo que, em abril de 2020, metade da população mundial encontrava-se em confinamento, causando uma grave perturbação socioeconómica. As consequências para a saúde foram, por outro lado, avassaladoras, uma vez que o vírus sobrecarregou os sistemas de saúde globalmente. As populações vulneráveis, particularmente os idosos e/ou com condições pré-existentes, foram desproporcionalmente afetadas. A procura por terapias antivirais eficazes tornou-se urgente, mas demorou mais de 10 meses até o primeiro fármaco antiviral, remdesivir, receber aprovação total para utilização em pacientes hospitalizados.

O objetivo deste projeto foi contribuir para a descoberta de alternativas terapêuticas. Assim, esta tese teve como foco estudar a interação vírus-hospedeiro no SARS-CoV-2, para compreender o que o distingue de outros CoVs e explorar potenciais alvos para terapias antivirais, com especial ênfase na proteína *Spike* (S), determinante na infeção viral de células hospedeiras. Esta glicoproteína trimérica interage com o recetor da enzima de conversão da angiotensina 2 (ACE2), mediando a ligação e promovendo a fusão das membranas. A região de ligação ao recetor (RBD), presente em cada monómero da proteína S, interage diretamente com o recetor ACE2 nas células hospedeiras, promovendo o reconhecimento e a infeção celular subsequente. Esta interação causa extensas alterações conformacionais, necessárias para a que a fusão entre as membranas viral e da célula hospedeira ocorra.

Embora o RBD seja um alvo bem caracterizado para intervenções terapêuticas, apresenta um desafio intrínseco, tendo em conta que é uma região são altamente variável, como mecanismo de evolução viral. Entre vários alvos na proteína S considerados neste projeto, foi selecionado o Local de Ligação de Ácidos Gordos (FABP), formado na interface entre RBDs e independente dos locais de ligação à ACE2, sendo este um alvo conservado em todos os CoVs altamente patogénicos. O FABP desempenha um papel significativo na

modulação da dinâmica conformacional da proteína S, influenciando a sua capacidade de reconhecer e infectar células hospedeiras.

Para caracterizar o FABP e o seu potencial como alvo terapêutico, foram realizadas simulações de dinâmica molecular usando uma estrutura de co cristalizada de ácido linoleico (LA)-FABP como modelo, com informações essenciais sobre a estabilidade estrutural do FABP e característica chave do local de ligação, incluindo a sua acessibilidade e flexibilidade. Estas informações estruturais sustentaram estudos de triagem virtual, utilizando bibliotecas de compostos comercialmente disponíveis, para identificar pequenas moléculas capazes de se ligarem ao FABP.

Através de estudos de ancoragem molecular, foram identificados vários candidatos potenciais, entre os quais um composto que demonstrou atividade antiviral a nível micromolar em ensaios experimentais de infecção viral. Para aumentar a potência deste composto inicial, foi realizada uma etapa de otimização, com a seleção de análogos estruturais baseados na semelhança estrutural, com potencial de explorar novas interações ou melhorar a ocupação do FABP.

A validação experimental revelou um composto com atividade a nível nanomolar contra o SARS-CoV-2, permitindo também um estudo de relação estrutura-atividade que aprofundou a compreensão das dinâmicas de ligação ao FABP e poderá orientar futuros estudos com vista a aumentar a atividade.

Este trabalho destaca o potencial terapêutico do FABP como um novo alvo para intervenções antivirais contra o SARS-CoV-2. A descoberta de uma nova estrutura com atividade antiviral, progressivamente melhorada de atividade micromolar para nanomolar, sublinha a relevância das abordagens computacionais nos esforços de desenvolvimento de fármacos. O conhecimento e as ferramentas desenvolvidos durante esta investigação não só mantêm a sua aplicabilidade no apoio à descoberta de potenciais novas terapias contra o SARS-CoV-2, como também podem servir de base para enfrentar novos CoVs emergentes.

Palavras-chave: SARS-CoV-2, Glicoproteína *Spike*, Local de Ligação de Ácidos Gordos, Descoberta de fármacos antivirais, Desenho de fármacos assistido por computador

Table of Contents

Chapter I – Introduction	2
1.1 General introduction	2
1.2 Coronavirus taxonomy	2
1.3 Coronavirus evolution and distribution.....	3
1.4 Human coronaviruses	4
1.5 COVID-19 pandemic and its origins	6
1.6 SARS-CoV-2 genome, structure, and lifecycle	8
1.6.1 RNA-dependent RNA polymerase	10
1.6.2 Viral proteases (Mpro and PLpro)	10
1.6.3 Structural proteins (M, N, E)	11
1.6.4 Spike glycoprotein.....	11
1.7 SARS-CoV-2 variants.....	13
1.8 COVID-19 treatments.....	14
1.8.1 Vaccines and monoclonal antibodies.....	14
1.8.2 Approved medicines against SARS-CoV-2.....	15
1.8.3 Investigational or experimental medicines	17
1.9 Computational approaches in antiviral research	18
1.9.1 Molecular modelling and molecular dynamics	19
1.9.2 Molecular docking	19
1.9.3 Scoring functions	20
1.9.4 Drug discovery strategies.....	20
References.....	22
Chapter II – Objectives and Outline of Thesis.....	28
2.1 Objectives	28
2.1.1 General objective	28
2.1.2 Specific Objectives.....	28
2.2 Outline of the thesis	29
2.2.1 Exploration of S protein and identification of pharmacological targets ..	29

2.2.2	Hit Molecule Identification and optimization.....	29
Chapter III – Results.....		32
3.1	Exploration of S protein structure and behaviour.....	32
3.1.1	SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response	32
3.2	Hit molecule identification and optimization	51
3.2.1	Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein - Confirmed and Potential Ligands	51
3.2.2	Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study.....	72
3.2.3	<i>In silico</i> discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding pocket.....	97
Chapter IV – Discussion		150
4.1	General discussion.....	150
4.1.1.	Exploration of S Protein and identification of pharmacological targets.....	150
4.1.2	Hit molecule identification and optimization.....	154
4.2	Future perspectives.....	162
References.....		162
5	Chapter V - Conclusions	168
5.1	Exploration of S protein structure and behaviour	168
5.2	Hit molecule identification and <i>in vitro</i> validation.....	169
5.3	Closing remarks	170
5.4	Personal remarks	171

List of Figures

Chapter I – Introduction

Figure 1 – Overview of the four genera within the subfamily Orthocoronavirinae and their respective example viruses. This figure illustrates the four genera of CoVs (alphacoronavirus, betacoronavirus, gammacoronavirus, and deltacoronavirus) along with notable viruses from each genus. The colours represent different genera, with example viruses listed for each group. The human-infecting CoVs are highlighted in bold and orange, and those associated with highly pathogenic outcomes are marked in bold and blue	3
Figure 2 – Map of Asia showing the points of origin for the three highly pathogenic human CoVs. SARS-CoV origin is marked in Guangdong, China (November 2002), MERS-CoV in Jeddah, Saudi Arabia (June 2012), and SARS-CoV-2 in Hubei, China (December 2019). These locations represent the initial reported cases in each viral outbreaks, which led to major health crises	5
Figure 3 – Global spread of SARS-CoV-2 from December 2019 to June 2020. World maps highlight the increasing number of countries with at least 100 confirmed COVID-19 cases at seven key timepoints (December 2019 - June 2020), indicated by blue colouring [33]	6
Figure 4 – Genome organization of SARS-CoV-2. Schematic representation of the SARS-CoV-2 genome, highlighting key regions including the 5' untranslated region (UTR), replicase gene (ORF1a/b), spike (S), envelope (E), membrane (M), nucleocapsid (N) proteins, and the 3' UTR. Each gene is represented by labelled segments along the genome, with the relative size of each gene shown in proportion to the entire viral genome	8
Figure 5 – SARS-CoV-2 lifecycle. Infection begins with the recognition and attachment of the virus to the human target receptor, ACE2 (1). The viral membrane then fuses with the host cell membrane, and uncoating occurs in the cytoplasm (2), releasing the viral +ssRNA genome. The virus hijacks the host's cellular machinery for primary translation (3), producing two polyproteins that are cleaved by viral proteases to generate 16 non-structural proteins (NSPs) (4). Among these NSPs, the replication-transcription complex (RTC) is assembled (5). The RTC then alternates between two functions: RNA replication (6) and the production of mRNA for structural and accessory proteins (6). In the main translation event, structural proteins -- such as the S protein, N protein, E protein, and M protein -- are synthesized (7). These proteins are transported to the endoplasmic reticulum (ER) (8), where they are assembled (9) before moving to the Golgi apparatus (10), along with newly synthesized +ssRNA bound to N protein. In the Golgi, the virions are fully assembled and are secreted from the infected cell by exocytosis (11)	9

Figure 6 – A) Pre-fusion conformation of the S protein, with three monomers coloured blue, yellow, and orange, and one ACE2 receptor bound coloured in red B) Monomeric structure of the S protein, showing the S1 subunit (blue), the S2 subunit (yellow), and the human ACE2 receptor (red) C) Post-fusion conformation of the S protein, with the S2 subunit isolated from the S1 subunit, which enables the major conformational change required for viral entry into the host cell	13
Figure 7 – Chemical structure of approved small molecule treatments for COVID-19	16
Figure 8 – A) Tautomeric forms of molnupiravir active metabolite. B) After integration in nascent genomic strain, molnupiravir can change between tautomeric forms, leading to binding and integration of either guanine or uridine, and to severe replication errors. Molnupiravir (A): Hydroxylamine form; Monupiravir (B): Oxime form	17

Chapter III – Results

3.1 Exploration of S protein structure and behaviour

3.1.1 SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response

Figure 1 – Representation of SARS-CoV-2 main structural proteins	34
Figure 2 – (A) Single chain in trimeric spike protein with subunit S1 (Tan) and S2 (Cyan) in 7A94. (B) Post-fusion conformation of all S2 units in the trimeric spike in 6LXT. (C) S1 subunit, with the RBD highlighted in 7A94	36
Figure 3 – Spike protein conformations, with each spike monomer is represented by a colour (blue, red, grey). (A) All RBD down (PDB 6VWX) (B) 1 RBD up (PDB 7A94) (C) 2 RBD up (PDB 7A97) (D) All RBD up (PDB 7A98)	37
Figure 4 – Organizational scheme of all structures found in this review	38
Figure 5 – Alignment and superposition of the models representing the interaction with no mutations (PDB 6VW1/6M0J/6LZG)	38
Figure 6 – Interface between SARS-CoV-2 S protein RBD (cyan) and hACE2 (tan), in the S protein/ACE2 interaction without mutations structure (PDB 6M0J). (A) Full view of RBD – hACE2 interface (B) N-terminal portion of the α 1 helix with a network of H-bonds (C) C-terminal interactions, with π -stacking interactions, h-bonds and Van der Waals forces	39
Figure 7 – Superimposition and alignment of wild-type RBD and hACE2 (PDB 6LZG) (light blue) and mutated RBD (PDB 7BH9) (hACE2 – dark blue ribbon/SARS-CoV-2 S protein – grey ribbon). The closer distance between the chains from the initial to the optimized model reflects the additional interactions established	41

Figure 8 – Structural superimposition between the wild-type RBD (Tan) (PDB 6M0J) and with the G485R RBD mutation (Cyan) (PDB 7LO4), with conformational change in the highlighted loop. The RMSD between structures is 0.42Å41

Figure 9 – Asparagine-501 from wild-type RBD (PDB 6M0J, carbon atoms in blue) and Tyr-501 from Gamma variant (PDB 7NXC, carbon atoms in yellow), capable of establishing an additional pi-stacking interaction with Tyr-41 from hACE2 (bottom) 42

Figure 10 – Mutation L452R, with the original variant from wild-type RBD (PDB 6M0J) (cyan) and the homology model of Delta variant (tan), obtained from genome MZ377116.1 (A) Superimposition of RBDs from original variant and Delta variant. Steric hindrance observed with Arg107 from Ab from structure 7CM4, due to a longer side chain in Delta variant (B) RBDs of original and delta variant structures with no steric hindrance regarding hACE2 46

3.2 Hit molecule identification and optimization

3.2.1 Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein - Confirmed and Potential Ligands

Figure 1 – Schematic representation of the SARS-CoV-2 virion structure, with the single-stranded, positive-sense RNA and the three structural proteins: spike glycoprotein, membrane protein, and envelope protein 53

Figure 2 – A) Ribbon representation of the S glycoprotein trimer, with the monomers represented as blue, orange, and yellow ribbons, respectively (PDB ID: 6VXX). B) Ribbon representation of an S glycoprotein monomer with the RBD highlighted in green. The RBD is part of the S1 subunit (orange), while the S2 subunit is represented in blue 53

Figure 3 – A) Ribbon representation of the S protein from a side view, with three FABPs defined near the RBD (PDB ID: 6ZB5). The S protein monomers are represented as yellow, orange, and blue ribbons, while three bound LA molecules are highlighted as the molecular surface (green: lipophilic; pink: hydrophilic; white: neutral). B) Zoom on one LA molecule (carbon atoms in cyan) cocrystallized in the FBP represented by the molecular surface, the hydrophilic (purple) portion near the carboxylate group, and the carbon chain buried in the hydrophobic portion (green) (PDB ID: 6ZB5) 54

Figure 4 – A) Protein surface representation of the S trimer in a three RBD down conformation (blue, orange, yellow) (PDB ID: 7DF3). B) Protein surface representation of the S trimer in a two RBD up (yellow and orange) bound to ACE2 (red and green) and one RBD down (blue) conformation (PDB ID: 7A93) 54

Figure 5 – A) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) in the FBP in a ribbon representation (blue), located between two adjacent RBDs. The pocket is formed by RBD-A (with hydrophobic residues) and hydrophilic residues in the adjacent RBD-B, interacting with

the carboxylate head. Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein, with distance in angstrom (Å)	55
Figure 6 – Ribbon representation of the S trimer (PDB ID: 6ZB5), in the side view with the base pair Asp398-Arg355 and residue Asp614 represented with carbon atoms in purple. The FABP is represented as a molecular surface	55
Figure 7 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with lifitegrast (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	59
Figure 8 – Crystallized SPC-14 (carbon atoms in orange) (PDB ID: 8H3E) in the FABP in a ribbon representation (blue). The green dashed line represents hydrophobic interactions, and the blue dashed line represents the distance (in angstrom (Å)) between SPC-14 and the side chain of Arg408	60
Figure 9 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with BAR 704 (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	60
Figure 10 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with dexamethasone (carbon atoms in orange) in the FABP in ribbon representation (blue)	60
Figure 11 – A) Structures of (A) LA (polyunsaturated), (B) oleic acid (ω -9), (C) palmitoleic acid (ω -7), and palmitic acid (unsaturated)	61
Figure 12 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with vitamin K (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	61
Figure 13 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with predicted pose for vitamin D (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	62
Figure 14 – A) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with the predicted binding pose for vitamin A (carbon atoms in orange) in the FABP in ribbon representation (blue). B) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with the predicted binding pose for tazarotene (carbon atoms in orange) in the FABP in ribbon representation (blue). C) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with the predicted binding pose for Tretinoin (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions	

(hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein 62

Figure 15 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with rotiorinol C (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein 64

Figure 16 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with hyoscyamine (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein 65

Figure S1 – PRISMA flow diagram for inclusion of studies 69

3.2.2 Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study

Figure 1 – (A) Protein surface representation of the S trimer in a three RBD down conformation (blue, orange, yellow) (PDB ID: 7KMS). (B) Protein surface representation of the S trimer in a one RBD up (yellow) bound to ACE2 (dark blue), and two RBDs down (blue and orange) conformation (PDB ID: 8IOU) 74

Figure 2 – (A) Side view of the S protein in ribbon representation with the monomers represented as blue, red, and grey ribbons, along with the three FABPs in yellow, green, and magenta molecular surfaces (PBD ID: 6ZB5). (B) Top view of the S protein in ribbon representation with the monomers represented as blue, orange, and grey ribbons, along with the three FABPs in yellow, green, and magenta molecular surfaces (PBD ID: 6ZB5) 75

Figure 3. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) in the FABP in a ribbon representation (blue), located between two adjacent RBDs (carbon atoms in white from RBD-B and carbon atoms in orange from RBD-C). Green solid lines represent polar interactions (hydrogen bonds) between the ligand and amino acid residues in the protein. All red atoms corresponds to oxygen atoms, while grey is due to shadows in the picture for depth perception 76

Figure 4. Chemical structure of compounds selected after the structure-based virtual screening and purchased from SPECS, identifying each compound with a number (1-26)77

Figure 5. Molecular docking results obtained for **19 (A)**, **20 (B)**, **24 (C)**, and **25 (D)**, represented with carbon atoms in sea green, in the FABP of SARS-CoV-2 S protein (PDB ID: 6ZB5, blue ribbon, carbon atoms in white from RBD-1 and orange from RBD-2). Green solid lines represent polar interactions (hydrogen bonds) between the ligand and amino acid residues in the protein 78

Figure 6. Molecular docking results obtained for 20 (carbon atoms in sea green) in the FABP of SARS-CoV-2 S protein Wuhan (PDB ID: 6ZB5, blue ribbon). (A) site 1 (RBD-B with carbon atoms in white, RBD-A with carbon atoms in orange); (B) site 2 (RBD-A with carbon atoms in white, RBD-C with carbon atoms in orange); (C) site 3 (RBD-C with carbon atoms in white, RBD-B with carbon atoms in orange). Green solid lines represent polar interactions (hydrogen bonds) between the ligand and amino acid residues in the protein	78
Figure 7. Inhibitory activity of the screened compounds tested at 100 μ M. Inhibition was determined as percentage calculated on the vehicle-treated cells (0.1% DMSO). The bars represent the mean $\pm$ SEM from three experimental repeats	79
Figure 8. Cytotoxicity of the screened compounds at 100 μ M, measured with Presto Blue™ Viability Assay. Cytotoxicity variation at timepoints 4 h and 48 h was estimated against an initial measurement (0 h) before presence of screened compounds. The bars represent the mean $\pm$ SEM from three experimental repeats	79
Figure 9. Molecular docking results obtained for 22 ((A), carbon atoms in sea green) and 23 ((B), carbon atoms in sea green) in the FABP of SARS-CoV-2 S protein (PDB ID: 6ZB5, blue ribbon and carbon atoms in white)	81

3.2.3 In silico discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding

Figure 1 - A) Top view of the S Protein in ribbon representation with the monomers represented as blue, orange, and grey ribbons, along with the FABP in yellow surface (PDB ID: 6ZB5). B) side view of the S Protein in ribbon representation with the monomers represented as blue, orange, and grey ribbons, along with the FABP in yellow surface (PDB ID: 6ZB5)	101
Figure 2 - All-atom RMSD of the S protein in the closed conformation, closed conformation with LA bound and open conformation (PDB accession codes 7DF3, 6ZB5 and 6VYB respectively)	102
Figure 3 - A) Ribbon representation of a single chain in the trimeric S protein structure (PDB ID: 6ZB5). Two FABPs are represented by orange molecular surfaces. Residues highlighted with a green circle correspond to reduced fluctuation, while residues highlighted with red circles are linked to increased residue movement. B) A) Co-crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) and FABP in ribbon representation (blue and white ribbons corresponding to separate RBDs), with the stabilized residues nearby LA (orange ribbon) ..	103
Figure 4 – Schematic representation of residues in pockets detected by Site Finder module in selected frames from MD of the closed conformation (A) and open (B). The colour intensity	

corresponds to detection of each FABP residue in zero, one, two or three pocket for each frame	104
Figure 5 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) in the FABP in a ribbon representation (blue), located between two adjacent RBDs. The pocket is formed by the hydrophobic area (carbon atoms in white) and hydrophilic area (carbon atoms in orange). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	105
Figure 6 – Chemical structure of compounds selected after the structure-based virtual screening and purchased from Enamine	106
Figure 7 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in light blue) superimposed with docking pose for compound 1 (carbon atoms in orange) in the FABP in ribbon representation (blue), obtained with GlideSP. Black lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	107
Figure 8 – Inhibitory activity of binding assays for the S-ACE2 interaction for the screened compounds in X-axis, tested at 100 μ M. Inhibition in the Y-axis determined as percentage of the vehicle-treated cells (2% DMSO)	108
Figure 9 – Inhibitory activity of viral induced CPE by the screened compounds in X-axis, at 67 μ M, after 72h of infection, in Vero E6 cells, measured using the CellTiter-Glo® method. CPE inhibition in Y-axis determined by comparing cells exposed to compounds with cells infected in the absence of compounds	109
Figure 10. A) Dose-response curve for compound 17 in Vero E6 cells, corresponding to the inhibitory activity of SARS-CoV-2 induced CPE in the Y-axis. The X-axis represents compound concentration expressed on a logarithmic scale B) Titration assay for compound 17 based on cell viability in the Y-Axis measured with CellTiter-Glo® method, for each dilution of viral supernatant in the X-axis	110
Figure 11 - Crystallized LA (PDB ID: 6ZB5) (carbon atoms in cyan) superimposed with LA in the FABP in ribbon representation (blue ribbon), along with predicted binding pose (obtained with GlideSP) for small molecule inhibitors with highest antiviral activity (compounds 1,4,15 and 17), and with chemical structure	111
Figure 12 – Chemical structure of analogues compounds to the strongest antiviral identified, divided in three categories and the original bioactive compound, with hydrophobic substitutions (light blue), hydrophilic substitutions (green), other changes (grey), original compound (blue)	114
Figure 13 – A) Inhibitory activity of viral induced CPE by analogue compounds in the X-axis at 33 μ M, after 72h of infection, in Vero E6 cells, measured using the CellTiter-Glo® method. The Y-axis represents CPE inhibition, determined by comparing cells exposed to compounds with cells infected in the absence of compounds. The bars represent the mean $\pm$ SEM from	

three experimental repeats. B) Dose-response curve for compound 36 in Vero E6 cells, corresponding to the inhibitory activity of SARS-CoV-2 induced CPE in the Y-axis. The X-axis represents compound concentration expressed on a logarithmic scale	115
Figure 14 – Quantitative/qualitative structure activity relationship concerning the first and second group of analogues	116
Figure 15 – Chemical structure of analogues compounds from the third category that resulted in compounds with inhibitory activity	117
Figure 16 – Predicted binding pose for compound 36 (carbon atoms in orange) obtained with GlideSP, in the FABP in ribbon representation (blue) (PDB ID: 6ZB5). Black lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein	117
Figure 17 – Dose-response curve for compound 36 in Vero E6 cells, corresponding to the inhibitory activity of SARS-CoV-2 – Delta Variant induced CPE in the Y-axis. The X-axis represents compound concentration expressed on a logarithmic scale to better visualize the dose-response relationship	118
Figure S1 – Preliminary cytotoxicity evaluation performed on Vero E6 cells, with the screened compounds, a growth control and a cytotoxicity control. The initial growth at 0h is a baseline against which growth at 4h and 96h is compared, as observed in the experimental protocol for viral infection in the proceeding cells assays	147
Figure S2 – Cytotoxicity of screened compounds (X-axis) in Vero E6 cells, measured by the CellTiter-Glo® method. Cytotoxicity is represented by CC ₅₀ in the Y-axis, the concentration of test compounds required to reduce cell viability by 50%. The bars represent the mean ± SEM from three experimental repeats	147
Figure S3 – Cytotoxicity of screened compounds (X-axis) in Vero E6 cells, measured by the CellTiter-Glo® method. Cytotoxicity is represented by CC ₅₀ in the Y-axis, the concentration of test compounds required to reduce cell viability by 50%. The bars represent the mean ± SEM from three experimental repeats	147
Figure S4 – Infection of Vero E6 cells with SARS-CoV-2 Omicron Variant at multiple concentrations. Cell viability was evaluated using the CellTiter-Glo® method	148

Chapter IV – Discussion

Figure 1 – Chemical structure of the strongest antiviral identified in the second virtual screening study (17)	159
Figure 2 – Chemical structure of the most active structural analogues to 17	160

List of Tables

Chapter I – Introduction

Table 1 – Characteristics of SARS-CoV-2 Variants: Mutations, Infectivity, Immune Evasion, and Disease Severity	14
---	----

Chapter III – Results

3.1 Exploration of S protein structure and behaviour

3.1.1 SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response

Table 1 – List of variants currently or in the past considered as variants of concern or variants of interest	35
--	----

Table 2 – List of mutations found in the spike protein of the Delta variant	45
--	----

3.2 Hit molecule identification and optimization

3.2.1 Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein - Confirmed and Potential Ligands

Table 1 – Confirmed and Potential Ligands of the FABP, with in Silico Studies and Complementing in Vitro Testing	56
---	----

Table 2 – Reported Compounds with Potential Activity without Validation in Experimental Assays	58
---	----

Table S1 – Chemical structures of all compounds reported (as SMILES)	70
---	----

3.2.2 Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study

Table S1 – Molecular formulas (SMILES), molecular weight, PAINS and BRENK analysis, and medicinal chemistry properties of tested compounds	87
---	----

Table S2 – Docking score from the scoring functions used during rescoring in FABP site 1 (Glide XP, Plants and Score Pose)	94
---	----

Table S3 – Docking score from the scoring functions used during rescoring in FABP site 2 (Glide XP, Plants and Score Pose)	95
---	----

Table S4 – Docking score from the scoring functions used during rescoring in FABP site 3 (Glide XP, Plants and Score Pose)	96
---	----

3.2.3 In silico discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding

Table 1 – ddPCR results, with detected viral genomic copies for SARS-CoV-2 infected cells treated with compound 36 compared to non-treated and infected cells	119
---	-----

Table 2 – Titration results for infection with SARS-CoV-2 Wuhan strain exposed to compound 36	120
--	-----

Table S1 – Chemical structures of compounds reported, including molecular Weight, calculated LogP, H-bond accepting groups, H-bond donating groups, as well as PAINS and BRENK analysis	136
--	-----

List of Abbreviations

ACE2 - Angiotensin converting enzyme 2
ADME - Absorption, Distribution, Metabolism, Excretion
CC₅₀ - Concentration 50% cytotoxic
CoV - Coronavirus
CoVs - Coronaviruses
COVID-19 - Coronavirus disease 19
CPE - Cytopathic Effect
DMEM - Dulbecco's Modified Eagle Medium
DMSO - Dimethylsulfoxide
E protein - Envelope Protein
EC₅₀ - half maximal effective concentration
EMA - European Medicines Agency
FABP - Fatty acid Binding Pocket
FCS - Fetal Calf Serum
FDA - Food and Drug Administration
LA - Linoleic acid
mAb - monoclonal Antibody
MD - Molecular Dynamics
M protein - Membrane Protein
MERS-CoV – Middle Eastern Respiratory Syndrome Coronavirus
Mpro - Main Protease
N - Nucleocapsid
NSP - Non-Structural Proteins
ORF - Open Reading Frames
P/S - Penicillin G/Streptomycin
PEA - Palmitoylethanolamide
Plpro - Papain-like Protease
RBD - Receptor Binding Domain
RdRp - RNA-dependent RNA Polymerase
RMSD - Root Mean Square Deviation
RMSF - Root mean Square Fluctuation
RT-ddPCR - Digital Droplet Polymerase Chain Reaction
RTC - Replication and Transcription Complex
S protein – Spike Glycoprotein
SARS-CoV-2 - Severe Acute Respiratory Syndrome Coronavirus 2

SARS-CoV - Severe Acute Respiratory Syndrome Coronavirus

TRS - Transcription-regulating Sequences

TCID₅₀ -50% tissue culture infectious dose

TMPRSS2 - Transmembrane Protease Serine 2

WHO - World Health Organization



Chapter I

Introduction

"If I have seen further, it is by standing on the shoulders of giants."

Isaac Newton

Chapter I – Introduction

1.1 General introduction

Coronaviruses (CoVs) are a large family of viruses, primarily infecting mammals and birds, capable of causing symptoms ranging from mild respiratory illness to severe disease [1]. While ubiquitous in nature, some can infect humans and typically cause mild disease with cold-like symptoms. However, a few can lead to severe illnesses, such as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) [1, 2]. In late 2019, a novel coronavirus, later named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), emerged in Wuhan, China, sparking a global pandemic [3]. SARS-CoV-2 causes Coronavirus Disease 2019 (COVID-19), an illness with a wide spectrum of clinical manifestations, from asymptomatic to severe respiratory distress and multisystem complications [4, 5]. This novel virus, with its rapid global spread and severe outcomes, led to widespread isolation measures and lockdowns, leaving a lasting impact on human history, sparking an unprecedented global effort to understand its biology, transmission, and potential treatments. [6, 7].

1.2 Coronavirus taxonomy

CoVs belong to the *Coronaviridae* family within the order *Nidovirales*. The primary subfamily, *Orthocoronavirinae*, is divided into four genera: *alphacoronavirus* and *betacoronavirus*, which primarily infect mammals (with bats serving as the main reservoir); *gammacoronavirus* and *deltacoronavirus*, which primarily infect birds (also their main reservoir) [8]. Among these, *betacoronavirus* includes the most clinically significant members, such as the highly pathogenic SARS-CoV, MERS-CoV, and SARS-CoV-2, as well as two additional CoVs associated with mild disease, OC43 and HKU1 [1]. The remaining human CoVs belong to the *Alphacoronavirus* genus, including 229E and NL63, both of which typically cause mild respiratory symptoms [1, 9].

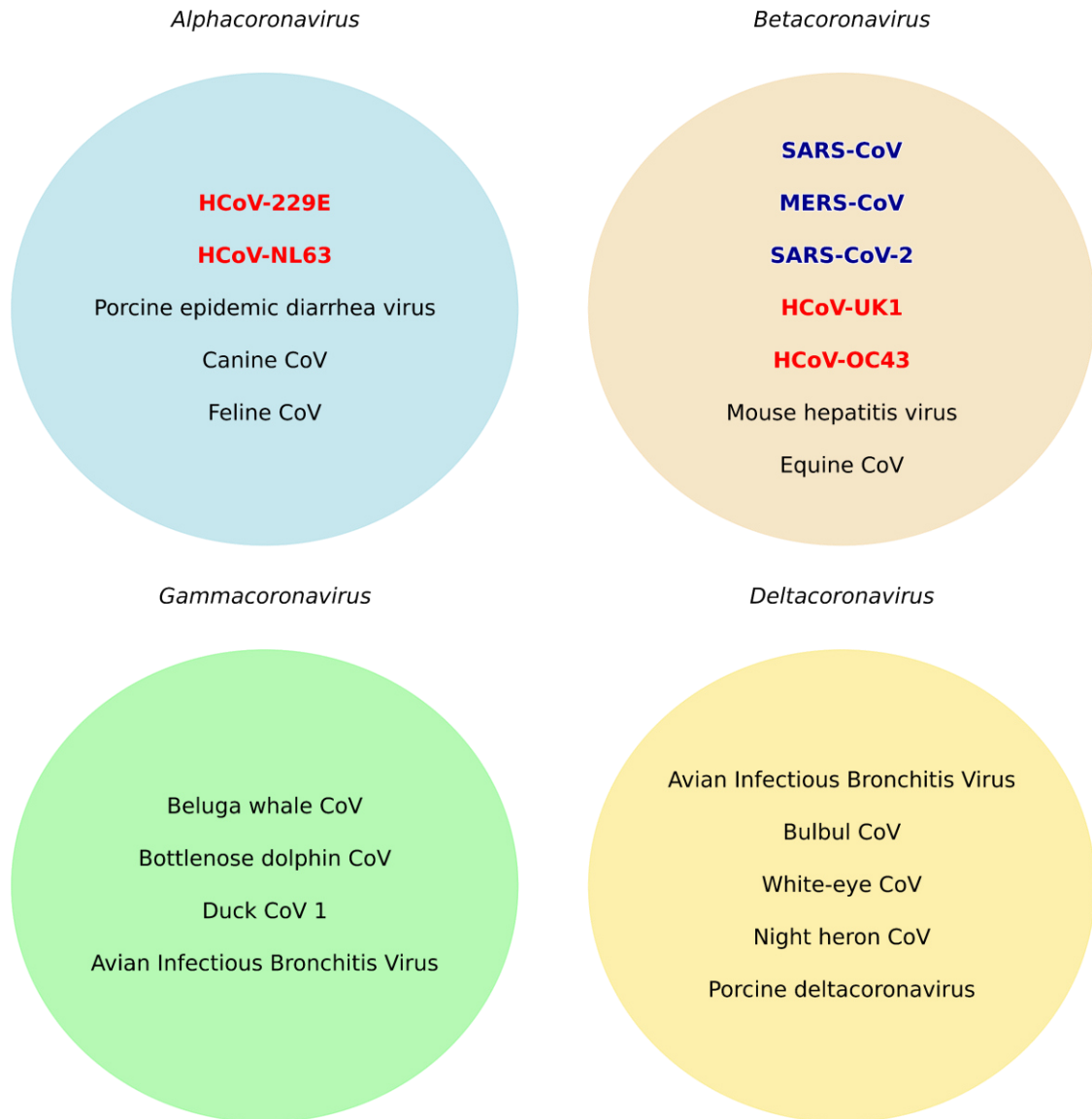


Figure 1 – Overview of the four genera within the subfamily *Orthocoronavirinae* and their respective example viruses. This figure illustrates the four genera of CoVs (*alphacoronavirus*, *betacoronavirus*, *gammacoronavirus*, and *deltacoronavirus*) along with notable viruses from each genus. The colours represent different genera, with example viruses listed for each group. The human-infecting CoVs are highlighted in bold and orange, and those associated with highly pathogenic outcomes are marked in bold and blue [8].

1.3 Coronavirus evolution and distribution

The taxonomy of coronaviruses (CoVs) suggests two distinct evolutionary pathways, distinguishing *alphacoronaviruses* and *betacoronaviruses*, which originate from bats, and *gammacoronaviruses* and *deltacoronaviruses*, which originate from birds [10, 11]. Multiple

factors have contributed to the wide distribution of CoVs, but critically they have the largest of known viral RNA genome [10]. This expanded RNA genome is a direct result of an advanced RNA-processing mechanism, which can increase replication fidelity, facilitating the acquisition of new genes, or reduce replication fidelity in ecologic stress, to boost adaptation [12, 13]. CoVs have two main reservoirs that further contribute to their adaption abilities, bats and birds [1]. Their wide diversity (bats amount to 20% of all mammalian species, while birds are the most numerous tetrapods) provides access to a range of cell types and molecular targets [1]. This variety has been further extended by the distribution of these hosts across biomes with different food sources, climate, shelter, and predatorial selective pressures. Combined, they appear in virtually every biological niche, due to their flying ability, removing obstacles faced by land-based animals. Finally, bats and birds exhibit habits of roosting and flocking, resulting in frequent opportunities for virus transmission among individuals, and for virus recombination events that drive CoV evolution [1, 11]. CoVs in both bats and birds show high genetic diversity and have a stable population size, while showing no symptoms of infection, characteristics usually associated to natural reservoirs for RNA viruses. In other species, there is typically an epidemic-like population growth (logistic or exponential growth) [14]. As a result, this adaptability leads to a diversity of strains and genotypes, with a wide range of hosts, infecting cattle, horses, camels, felines, canines, rodents, cetaceans, small mammals, birds, among others. Consequently, interspecies spillover events are frequent and can cause outbreaks with potentially disastrous consequences, both in humans and other animals [1, 15, 16].

1.4 Human coronaviruses

Human coronavirus (CoVs) infections have been known since the 1960's, generally associated with mild cold-like symptoms and a minor health threat, with HCoV-OC43 and HCoV-229E the first human CoVs discovered [17, 18]. This changed drastically in 2002 when a highly pathogenic CoV, causing severe respiratory disease, was identified in southern China, eventually named SARS-CoV, spreading to over 30 countries and infecting ~ 8000 people with a mortality rate of ~ 10% [10]. SARS-CoV was initially detected in Guangdong (China), epidemiologically associated with animal markets and food handlers [2]. While the host before crossover to human was determined to be civet cats, these proved only as intermediate hosts. Genetic analysis of SARS-CoV showed a split from bat CoV as late as 1986, while multiple closely related horseshoe bat viruses were also reported [14]. These bat, human and civet CoVs showed high similarity (over 95% in some major proteins), although only 70% similarity between human and bat Spike glycoprotein (S protein), responsible for target recognition, cell tropism and cell infection. On the other hand, human and civet had 90% similarity, suggesting that SARS-CoV gained the ability to infect human cells during the crossover event to civet cats

[1, 13, 14, 19]. Naturally, a surge in CoV related research followed and two more circulating human CoVs with cold-like symptoms were discovered during 2003, HCoV-NL63 and HCoV-HKU1 [1, 10]. In 2012, a new CoV emerged in Saudi Arabia, MERS-CoV, with 2,553 people infected and a mortality rate of 34% [20]. MERS-CoV was the second highly pathogenic CoV, directly identified in dromedary camels, showing mild disease and established as the direct origin, with again bats as the original reservoir [21-24]. Genomes found in human patients and camels share 99% similarity, and as with SARS-CoV, the virus has circulated in dromedary camels for 30 years, before the human outbreak, with nearly the entire dromedary population previously exposed to the virus [13, 21, 25]. To further this point, MERS-CoV like virus are not detected in camels exported to Australia earlier in the 20th century, pointing to a recent spillover event to camels that enabled the infection in human cells [23]. CoVs had shown the potential to become a significant danger to human health, a threat that escalated dramatically in December 2019 when a novel CoV was identified in Wuhan, China, and spread rapidly across the globe [26] (Figure 2).

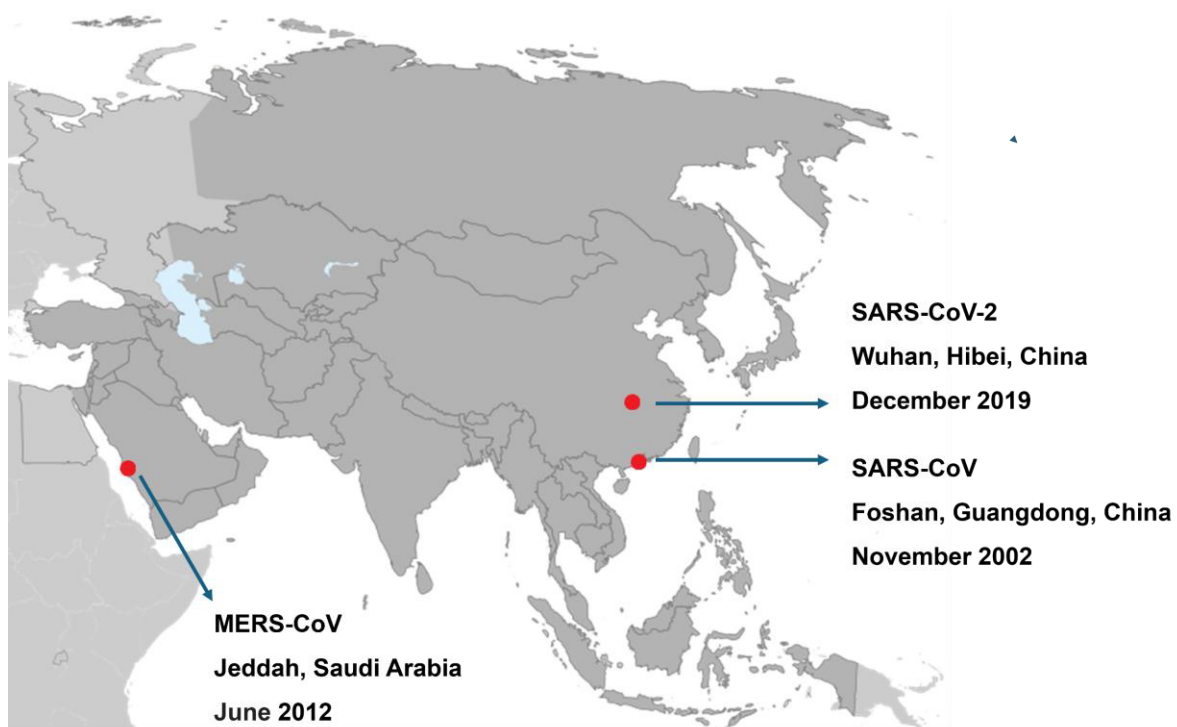


Figure 2 – Map of Asia showing the points of origin for the three highly pathogenic human CoVs. SARS-CoV origin is marked in Guangdong, China (November 2002), MERS-CoV in Jeddah, Saudi Arabia (June 2012), and SARS-CoV-2 in Hubei, China (December 2019). These locations represent the initial reported cases in each viral outbreaks, which led to major health crises.

By March 11th 2020 the World Health Organization (WHO) characterized this new virus as pandemic, which was the seventh human infecting CoV, and the third highly pathogenic, eventually named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [10]. By April 2021, The WHO estimated that one in every four people had anti-SARS-CoV-2 antibodies [27, 28]. Up to October 2024, SARS-CoV-2 is responsible for over 750 million people infected and 7 million deaths [29].

1.5 COVID-19 pandemic and its origins

Coronavirus disease 19 (COVID-19) is an infectious disease caused by SARS-CoV-2, initially detected in the city of Wuhan, Hubei province (China). In late December 2019, a cluster of pneumonia cases of unknown cause, and with similar symptoms previously identified in SARS-CoV (viral pneumonia, including fever and cough) was reported [5]. Among the first identified cases, most were epidemiologically linked to a seafood and wet animal wholesale market, in Wuhan [3]. Shortly after, new patients were identified without any link to the wet market, along with nosocomial infection in healthcare facilities, clear evidence of human-to-human transmission [30, 31]. During the month of January in 2020, it spread massively through China, with thousands of new daily cases, leading the WHO to declare this novel CoV a public health emergency of international concern [32]. Due to efficient transmission and international travel the virus rapidly spread across the world [32] (Figure 3).

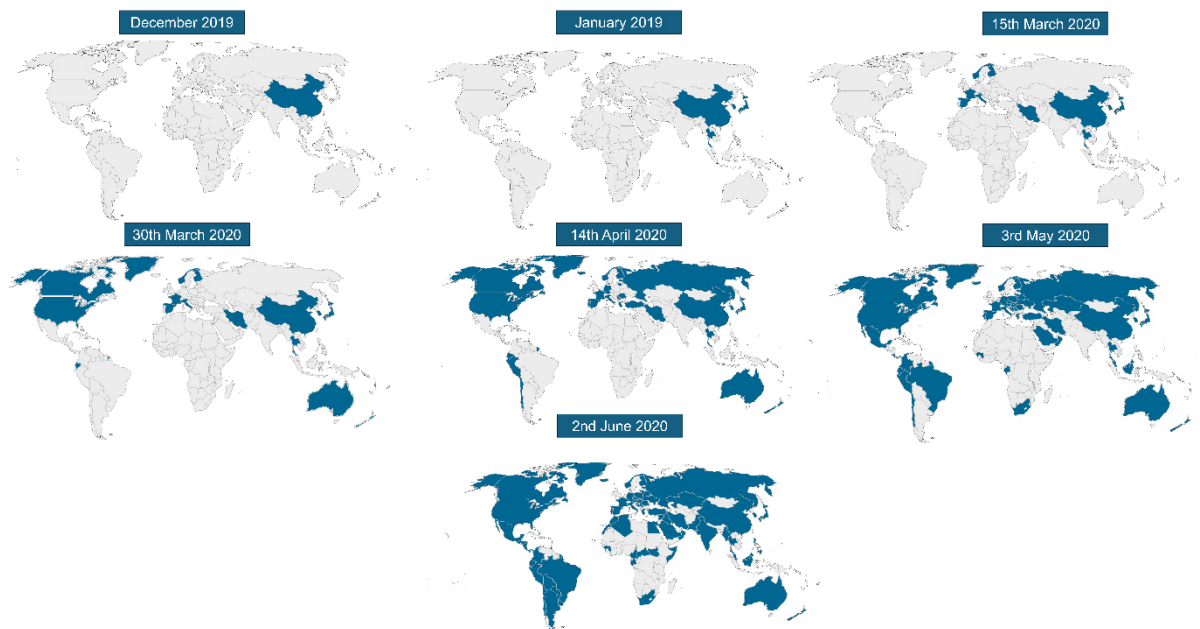


Figure 3 – Global spread of SARS-CoV-2 from December 2019 to June 2020. World maps highlight the increasing number of countries with at least 100 confirmed COVID-19 cases at seven key timepoints (December 2019 - June 2020), indicated by blue colouring [33]

Clinical symptoms are similar to SARS-CoV and MERS-CoV, with fever, cough, myalgia and asthenia [34]. Pneumonia occurs in virtually all patients, but for SARS-CoV-2 upper respiratory signs are rare, pointing to a preferred replication in the lower respiratory tract [34]. SARS-CoV-2 has a lower mortality rate (1.7% overall mortality) when compared with MERS-CoV or SARS-CoV (34% and 10%, respectively), although it increases with age, and reaches ~25% in people over 85 years of age [35, 36]. SARS-CoV-2 incubation lasts 7 to 14 days, with infection gradually progressing and worsening, and severity decided on the second week [37]. COVID-19 patients generally can be divided into asymptomatic, mild or severe cases, with complications including acute respiratory distress syndrome, acute heart injury, and secondary infections [34, 37]. SARS-CoV-2 main source of infection are aerosols with viral particles, created from cough and sneezing, although it can be transmitted by respiratory droplets, direct contact and fomites [34]. CoVs typically also replicate in the gastrointestinal tract, however these signs and symptoms are rare with SARS-CoV-2 (while frequent in MERS-CoV and SARS-CoV) [37]. Despite this, virions have been detected in faecal samples, opening the door for faecal-oral transmission [5, 34].

As previously observed for other coronaviruses (CoVs), a small mammal species was initially suspected as the animal host, in this case the pangolin, with an S protein strikingly similar to SARS-CoV-2 (a single aminoacidic variation in the receptor binding motifs) [38, 39]. Shortly after, a bat CoV genome was detected with 96% nucleotide similarity, although with large differences in the S protein [10, 40]. Phylogenetic analysis suggested that the SARS-CoV-2 lineage diverged from ancestral bat CoVs between 1948 and 1982 [41]. The evolutionary gap between SARS-CoV-2 and the closest bat CoV has a genetic distance of approximately 4% (around 1,150 nucleotide mutations), also indicating decades of evolutionary divergence [42]. As with SARS-CoV, the whole genome analysis points to a bat CoV origin, with the ability to bind human targets likely obtained in a crossover event [10, 43, 44]. However, the pangolin CoV is still lacking a critical feature in the S protein of SARS-CoV-2, a furin cleavage site that greatly increases infectivity [41, 42] and neither bat nor pangolin CoVs sampled thus far have polybasic cleavage sites [6]. Discovering the animal origins of human pathogens can be challenging, with many examples still unidentified (Ebola virus, hepatitis C virus, poliovirus, among others) [6, 42]. Even for SARS-CoV, it took over a decade to discover bat CoVs with high similarity and the same human target. As for SARS-CoV-2, while multiple animals are suspected natural reservoirs, the exact animal species or populations from where the main spillover occurred might not yet have been sampled [6, 42].

1.6 SARS-CoV-2 genome, structure, and lifecycle

SARS-CoV-2 is an RNA virus, with a single stranded positive sense genome, composed by 14 open reading frames (ORF), containing ~30kb, and encoding 29 viral proteins, including four structural proteins and a replicase (Figure 4) [34, 45, 46]. The genetic arrangement resembles SARS-CoV and MERS-CoV, sharing 79% and 50% nucleotide sequence identity, respectively [45]. In 3' terminus, four ORFs encode the structural proteins: Spike Glycoprotein (S-protein), Envelope (E), Membrane (M) and nucleocapsid (N), playing important roles in viral entry, fusion and survival, while almost two thirds of the 5' are occupied by the replicase gene, encoding a two polyproteins, as well as two proteases: main protease (Mpro) and papain-like protease (PLpro) [45, 46]. These viral proteases digest the polyproteins into 16 non-structural proteins (NSP) [45, 46]. Finally, several accessory genes can be found between these areas, encoding accessory proteins that regulate viral infection, helicase activity, immunomodulation, gene transactivation and host immune response [4]. Under electron microscopy, SARS-CoV-2 shows typical morphology for CoVs, with mostly spherical virions and some polygonal, and a diameter between 60 and 140 nm [34]. The virion has a coronal shape that gives the virus its name, due to the prominent spikes (9-12 nm) projected from the viral surface [34].

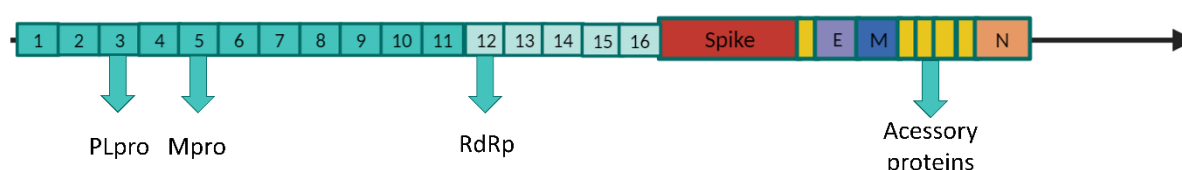


Figure 4 – Genome organization of SARS-CoV-2. Schematic representation of the SARS-CoV-2 genome, highlighting key regions including the 5' untranslated region (UTR), replicase gene (ORF1a/b), spike (S), envelope (E), membrane (M), nucleocapsid (N) proteins, and the 3' UTR. Each gene is represented by labelled segments along the genome, with the relative size of each gene shown in proportion to the entire viral genome.

The virus lifecycle starts when the S protein recognizes ACE2, eliciting cell fusion [44, 47] (Figure 5). The contents of the virion are then released into the cytoplasm, with RNA replication starting along, with synthesis of viral mRNA translated into viral structural and accessory proteins, that provide support functions, such as modulating intracellular membranes or host immune evasion [48]. The viral genome (+ssRNA) is attached to the host's ribosomes and translated into large polyproteins that are processed by proteases, originating non-structural proteins (NSP) [4, 5]. Progeny virions are assembled in the endoplasmic reticulum and Golgi body, being released by exocytosis to complete the virus lifecycle [46, 48]. Rapid replication

in the lungs triggers a strong immune response, and the ensuing cytokine storm may cause acute respiratory distress syndrome, the characteristic symptom of COVID-19 [34, 45].

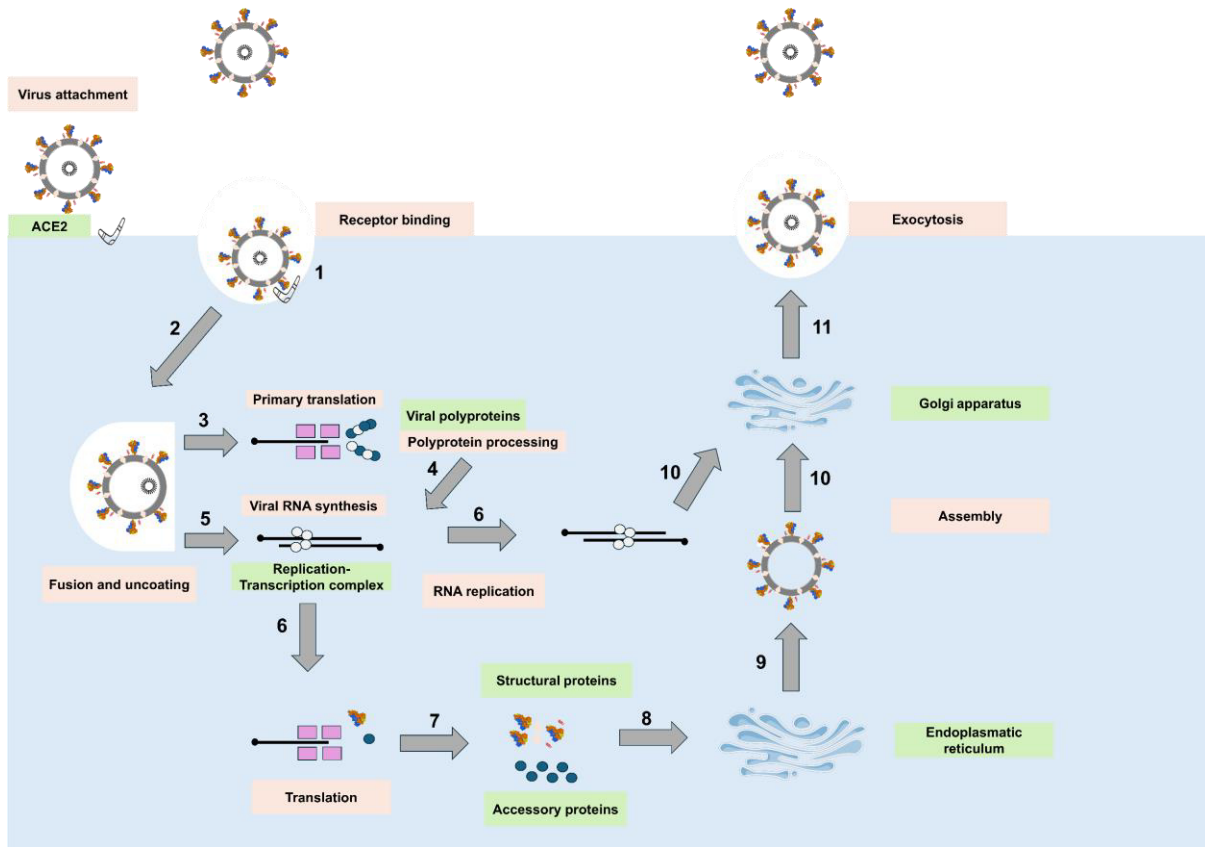


Figure 5 – SARS-CoV-2 lifecycle. Infection begins with the recognition and attachment of the virus to the human target receptor, ACE2 (1). The viral membrane then fuses with the host cell membrane, and uncoating occurs in the cytoplasm (2), releasing the viral +ssRNA genome. The virus hijacks the host's cellular machinery for primary translation (3), producing two polyproteins that are cleaved by viral proteases to generate 16 non-structural proteins (NSPs) (4). Among these NSPs, the replication-transcription complex (RTC) is assembled (5). The RTC then alternates between two functions: RNA replication (6) and the production of mRNA for structural and accessory proteins (6). In the main translation event, structural proteins—such as the S protein, N protein, E protein, and M protein—are synthesized (7). These proteins are transported to the endoplasmic reticulum (ER) (8), where they are assembled (9) before moving to the Golgi apparatus (10), along with newly synthesized +ssRNA bound to N protein. In the Golgi, the virions are fully assembled and are secreted from the infected cell by exocytosis (11). Adapted from [48]

1.6.1 RNA-dependent RNA polymerase

Replication in SARS-CoV-2 is performed by the replication and transcription complex (RTC), containing several subunits assembled from NSPs cleaved from the viral polyprotein [48, 49]. The core of the complex is NSP12 and along with NSP8 and NSP7, it makes up the complete RNA-dependent RNA polymerase (RdRp) [48]. The RdRp is distinctive of CoVs, as these viruses have the largest known positive sense RNA, requiring not only efficient replication of genomic material for new virions but also high fidelity to maintain the expanded genetic material. To achieve high replication efficiency, SARS-CoV-2 has a helicase unique to the order *Nidovirales*, capable of a remarkable template-switching step [46, 49, 50]. During negative-strand synthesis, the RdRp may pause at specific transcription-regulating sequences (TRS) and switch to a downstream TRS, leading to the production of subgenomic RNAs. These subgenomic RNAs serve as templates for the synthesis of various structural and accessory proteins, enabling efficient viral gene expression [46, 50]. Therefore, the virus can rapidly generate distinct mRNAs for various accessory proteins during genome replication, minimizing the energy and resources required, while ensuring coordinated expression of viral components necessary for new virions [50]. On the other hand, the RdRP guarantees high fidelity of replication due to NSP14, an exonuclease with proofreading activity, that backtracks mismatched template–product RNA chains, extruding the mismatched RNA [46, 50]. Mutations in this protein result in viruses with a lethal mutator phenotype [50].

1.6.2 Viral proteases (Mpro and PLpro)

SARS-CoV-2 has two proteases – Mpro (main protease, also known as 3CLpro) and PLpro (papain-like protease) – that share the same general function of cleaving viral polyproteins and are conserved in CoVs [51, 52]. Mpro cleaves the polyprotein at eleven sites to release non-structural proteins (NSPs), while PLpro cleaves three additional sites to release the remaining NSPs [53, 54]. Both proteases are essential to the formation of the RTC, genome replication and virion assembly. Mpro recognizes a consensus sequence P2P1–P1', where P1 is always glutamine, P1' is serine or alanine, and P2 is a hydrophobic residue such as leucine, phenylalanine, or valine [53, 54]. PLpro, on the other hand, cleaves after a glycine residue, a cleavage site that is not recognized by any known human cysteine proteases, providing a degree of selectivity for the viral protease over the human proteome [55]. Additionally, PLpro has a dual function: in addition to cleavage of the viral polyprotein, it also cleaves host proteins, including ubiquitin and ISG15, which modulate host cell metabolism. This function helps inhibit the antiviral interferon response, allowing the virus to better evade the host immune system [52].

1.6.3 Structural proteins (M, N, E)

The M protein, being the most abundant structural component of the virus, is crucial for virus assembly and membrane budding. As a transmembrane protein, the terminal domain is outside the virus [56, 57]. After being synthesized within host cells, the M protein is directed to the endoplasmic reticulum–Golgi intermediate compartment, where it acts as a scaffold for the recruitment of other viral structural proteins, binding all other structural proteins [56, 57]. Co-expression and combination of SARS-CoV-2 M protein and N protein is enough to produce virus-like particles, highlighting the role in viral assembly [56]. In regard to the other structural proteins, the M protein stabilizes the N protein-RNA complex in the virion and also cooperates with S protein [57]. As for the N protein, it is the only structural protein completely inside the virion, tightly bound to the viral RNA genome, packaging it into a ribonucleoprotein complex [58]. The main function is mediating RNA stabilization and packaging into new virions, although it also interacts with the RTC to facilitate RNA synthesis and translation, by recruiting host factors and promoting RNA template switching [58]. However, this protein performs an additional function by antagonizing host immune response, countering antiviral activities mediated by interfering RNA, and suppressing viral RNA silencing [46, 58]. The N protein is the most immunogenic protein in SARS-CoV-2, inducing a strong immune response by promoting phagocytosis, complement activation, cytotoxicity of infected cells by natural killer cells, among others [59, 60].

The E protein is a transmembrane protein, highly conserved in CoVs, with poorly understood functions, while being the smallest major structural protein [61]. It has been suggested a function viral lysis and the subsequent viral genome release, although it might also bind cellular adaptor proteins, affecting pathogenesis and inducing intracellular retention of S protein, regulating the location and conformation of the S protein [46]. By limiting cell surface expression of the S protein, it prevents the formation of syncytia, thereby ensuring adequate assembly of new virions [61]. Regarding viral assembly, recombinant CoVs lacking the E gene display aberrant morphology and a strong reduction in virion production, suggesting that the E protein is also involved in the budding of new virions [61].

1.6.4 Spike glycoprotein

The spike glycoprotein (S protein) is a class-1 fusion homotrimer protein, that extensively decorates the viral surface, creating a crown like shape visible under electron microscopy [46]. It is highly glycosylated, and its main function is to mediate infection of target cells, by facilitating cell fusion. However, it is also responsible for upstream target recognition and initial binding to the human functional receptor, ACE2 [46, 62]. Each monomer is divided into an S1 binding domain and an S2 fusion domain, covalently bound in the native pre-fusion

conformation [63, 64]. The S1 subunit, responsible for direct target recognition, contains the receptor binding domain (RBD), that interacts with ACE2 and the receptor binding motifs (RBM) corresponding to the residues that directly perform the interaction [64]. Each RBD can undergo hinge like conformational changes between two possible conformations, an up and active state, against a down and inactive state [65]. In the down conformation, the RBD adopts an angle that hides the RBM, blocking target recognition [46]. A critical feature of the S protein in SARS-CoV-2 is a polybasic furin cleavage site between subunits S1 and S2, cleaved by furin and other proteases, that separate the subunits [46, 66]. Before this cleaving step, the S protein is in a pre-fusion conformation, which blocks the S2 subunit from the extensive conformational changes required to achieve cell infection. The “activation” cleavage allows a separation of the S1 and S2 subunits, that initially remain bound by non-covalent interactions in the metastable conformation. Target recognition and binding to ACE2 disrupts this state, exposing the fusion peptide and allowing its insertion into the membrane, thereby eliciting cell fusion and infection [42, 64, 67]. Due to this function, the S protein is deeply intertwined with cell tropism and cell infection abilities [66]. Antibodies targeting the S protein have shown strong neutralizing activity, with most binding the RBD and competing with ACE2, although these can only bind to the up state when the RBM is exposed. Finally, it has the highest mutation rates in the viral genome and shows high variability within CoVs, since it must be adapted to each host [62, 66]. The S protein also allows cell-to-cell fusion between neighbouring infected and non-infected cells, facilitating direct spread of the virus, thereby avoiding virus-neutralizing antibodies, which leads to the formation of syncytia (i.e. giant, multinucleated cells) and increased pathogenesis [10].

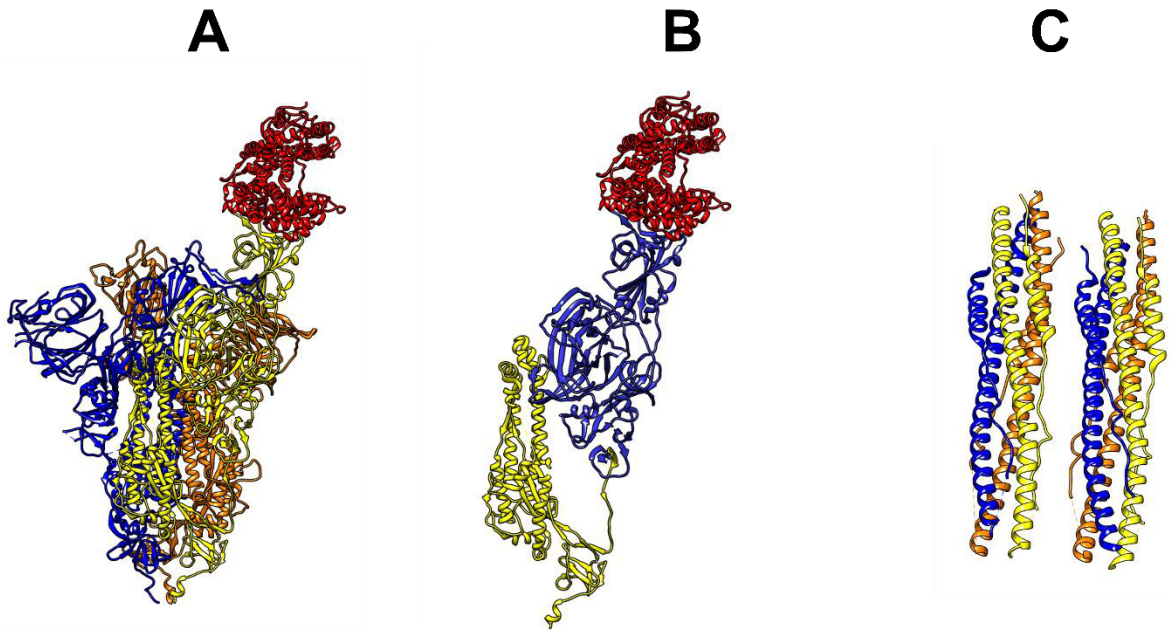


Figure 6 – A) Pre-fusion conformation of the S protein, with three monomers coloured blue, yellow, and orange, and one ACE2 receptor bound coloured in red B) Monomeric structure of the S protein, showing the S1 subunit (blue), the S2 subunit (yellow), and the human ACE2 receptor (red) C) Post-fusion conformation of the S protein, with the S2 subunit isolated from the S1 subunit, which enables the major conformational change required for viral entry into the host cell.

1.7 SARS-CoV-2 variants

Since the emergence of SARS-CoV-2, the virus has undergone numerous mutations, giving rise to distinct viral variants [68, 69]. Like other RNA viruses, this virus has a high mutation rate, with a heavy focus particularly in the gene encoding the S protein, given that it is the main target for neutralizing antibodies [68]. Some mutations can confer advantages, such as increased transmissibility or immune evasion, and over time viral variants with advantageous mutations can become dominant in the population, driving the evolutionary process [68, 69]. In the first year of the pandemic, SARS-CoV-2 had recently crossed over to human hosts and was still adapting, facing a largely immune-naïve population with minimal pre-existing immunity. This created a high potential for it to evolve and develop adaptations, especially to increase its transmissibility, although shortly after it evolved to evade immune response [68]. The WHO has classified certain SARS-CoV-2 variants as "Variants of Concern" when they meet specific criteria, such as increased transmission or reduced neutralization by antibodies from previous infection or vaccination [70, 71]. The first defined was variant alpha, associated with a 50-70% increase in transmission compared to earlier strains [72]. It spread rapidly to become dominant variant during 2021 [68]. At the same time the beta variant emerged

demonstrating increased resistance to neutralizing antibodies, particularly in individuals previously infected, showing virus adaptation to non-naïve populations. The gamma variant evolution revealed the combined ability to modulate immune response, and extensive mutations in the RBD leading to reinfection events [69, 73]. Among the variants that emerged during the pandemic, the delta variant was particularly significant, becoming the dominant global strain due to a combination of high transmissibility and severe disease [68, 72]. At this moment vaccination efforts were ramping and placed a strong pressure in the virus evolution, which lead to a new variant, omicron, with an unprecedented number of mutations in the S protein, and critically the ability to escape immune response, not only from prior infection but also vaccination [68]. However, disease became far less severe, with increased replication and symptoms in the upper respiratory tract. Multiple sub-lineages are still circulating, pressured mainly by newly introduced vaccines and antibodies [68].

Table 1 – Characteristics of SARS-CoV-2 Variants: Mutations, Infectivity, Immune Evasion, and Disease Severity

Variants	S protein mutations	RBD mutations	Infectivity (a)	Immune evasion (a)	Disease Severity (a)
Alpha	10	1	~50% more transmissible	Moderate	Similar
Beta	10	3	~50% more transmissible	High	Slightly increased
Gamma	12	3	~50% more transmissible	Moderate	Similar
Delta	10	2	~100% more transmissible	High Partial vaccine escape	Strongly increased
Omicron	26	15	~400% more transmissible	Very high Reinfection	Reduced

(a) relative to the Wuhan strain [68, 72, 73]

1.8 COVID-19 treatments

1.8.1 Vaccines and monoclonal antibodies

Among the various vaccines developed against SARS-CoV-2, the S protein became the primary target due to its key role in infection and immune recognition [33, 44]. This focus led

to the creation of vaccines designed to elicit strong immunogenic and antibody responses against the S protein. To date, four major vaccines have been approved by the Food and Drug Administration (FDA), and six by the European Medicines Agency (EMA) [74, 75], which have demonstrated the ability to successfully trigger protective immune responses against SARS-CoV-2 [76]. In the first year of widespread vaccination, the impact was profound, with an estimated 63% global reduction in potential deaths, preventing over 14 million fatalities during 2021 alone [77]. While many vaccine platforms and targets were explored in the early phases of the pandemic, the S protein targeted approaches proved to be the most effective, highlighting the importance of this viral structure in both vaccine design and pandemic control [74, 75]. With a similar focus on the S protein, monoclonal antibodies (mAbs) have emerged as a promising fast-track therapeutic approach for COVID-19, with a mechanism of action focused on the direct neutralization of viral entry, thereby preventing the virus from infecting host cells [78]. This precision targeting initially showed great potential in reducing viral load and mitigating disease severity. However, one major limitation of mAbs is their vulnerability to mutations, particularly in newer variants [79].

1.8.2 Approved medicines against SARS-CoV-2

Drug development requires very long periods of research and high investments, over 10-15 years [80]. Starting with the initial identification of a promising drug, it must proceed through preclinical and clinical trials [78, 80]. Therefore, drug development is not practical during a pandemic, and targeted drugs would not be available for use within a short time frame, although their development was still being actively pursued [78]. In this context, drug repurposing provides a rapid and effective approach, using drugs already approved for other indications to treat the current threat, in this case COVID-19. The main advantage is that these approved drugs are already well documented regarding drug safety, pharmacokinetics and have passed all steps required for regulatory approval [78]. During the pandemic several drugs were tested and even received emergency approvals, such as hydroxychloroquine (antimalaric) or remdesivir (antiviral drug originally developed for Hepatitis C). Since the beginning of the pandemic, multiple antivirals, antibiotics, antimalaric or immunomodulatory drugs have shown possible efficacy against SARS-CoV-2 [81].

Drug repurposing has resulted in two approved small molecule therapies for COVID-19 (molnupiravir and remdesivir), while an additional approved medicine (nirmatrelvir) was derived from previous research during SARS-CoV outbreak (Figure 7).

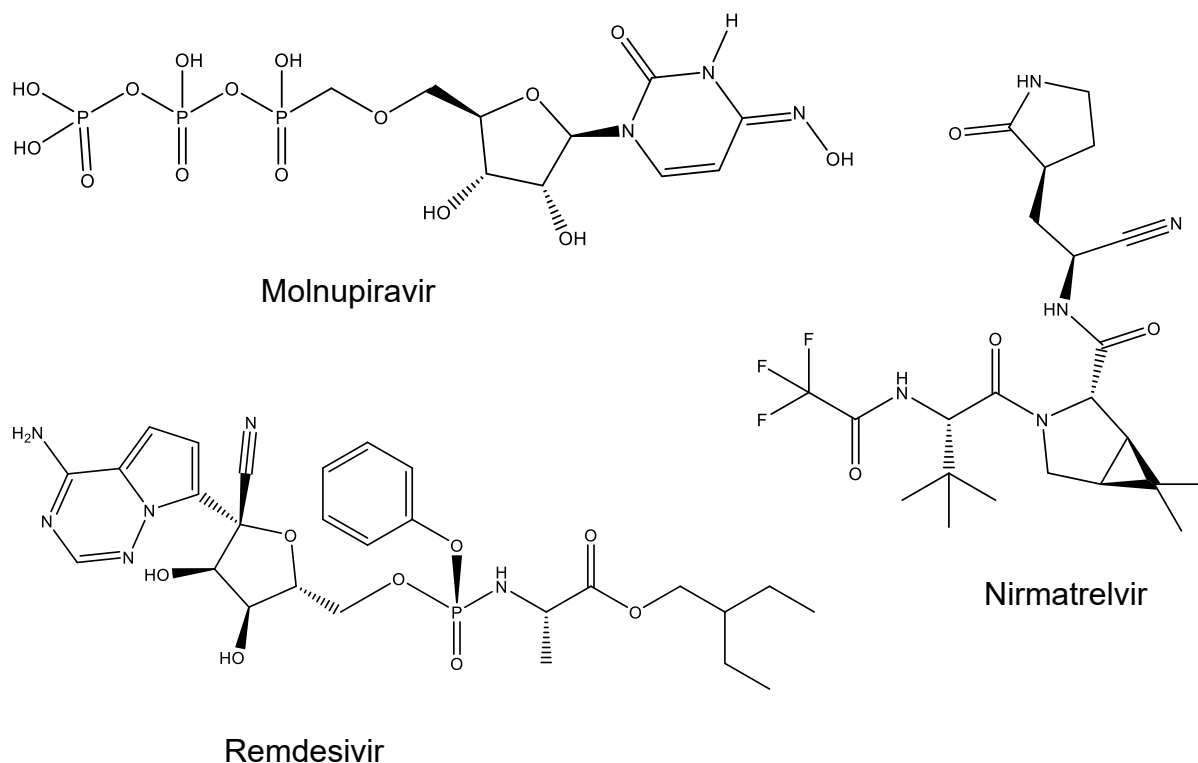


Figure 1 – Chemical structure of approved small molecule treatments for COVID-19.

Remdesivir acts by inhibiting the RdRp and is a prodrug [82], with the active metabolite being a nucleoside analogue of ATP, that competes with substrates to be efficiently incorporated into nascent RNA chains, causing delayed chain termination at the post-translocation position -3 . Therefore, it terminates RNA transcription and inhibits RNA replication [83]. However, current recommendations limit its use to early treatment in outpatients [84, 85]. Molnupiravir, which is also a prodrug, has an active metabolite that acts as a nucleotide derivative, initially developed to inhibit influenza A and B viruses [86]. As a nucleotide derivative, molnupiravir is incorporated in the nascent RNA strand and is resistant to the exonuclease activity of the RdRp [87]. Once successfully incorporated, the tautomeric nature of this molecule leads to severe replication errors (Figure 8) [87]. The two forms of this drug mimic cytosine or uridine triphosphate, which leads to ambiguous base-pairing, thus increasing the frequency of G-to-A and C-to-U transition errors in the viral RNA genome, in a lethal mutagenesis mechanism [87]. It also is limited by recommendations, with reduced benefits in hospitalized patients with severe disease [88], while the mutagenic effect is also observed in mammalian cells, hence why its approval is only for short-term use (less than 5 days) and is not recommended for pregnant women or young patients [89].

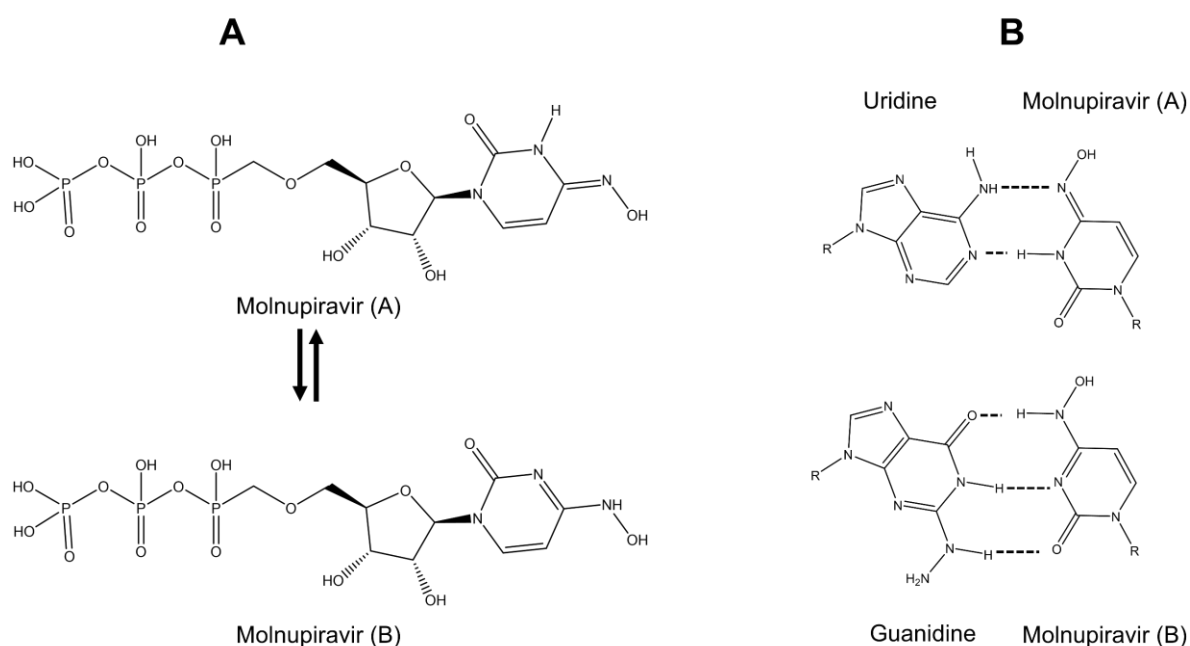


Figure 2 – A) Tautomeric forms of molnupiravir active metabolite. B) After integration in nascent genomic strain, molnupiravir can change between tautomeric forms, leading to binding and integration of either guanine or uridine, and to severe replication errors. Molnupiravir (A): Hydroxylamine form; Molnupiravir (B): Oxime form.

The mechanism of action of nirmatrelvir is based on its activity as a protease inhibitor targeting Mpro, harbouring a nitrile warhead that covalently binds to the catalytic pocket [55, 79]. Importantly, it is a broad-spectrum antiviral, inhibiting human CoVs, including multiple SARS-CoV-2 variants [55]. The discovery and development of this molecule was accelerated greatly by previous research focused against SARS-CoV, initially suspended when this outbreak subsided in 2003 [79]. Given the 100% sequence homology across substrate-binding sites in SARS-CoV and SARS-CoV-2 Mpro, nirmatrelvir was identified after pharmacokinetic optimization and submitted to clinical trials [79]. The state of public emergency declared worldwide justified the accelerated development of new drugs, which resulted in a short period of 22 months (March 2020 to December 2021) to market authorization [79]. The greatest limitation of nirmatrelvir is the frequency of drug–drug interactions, while resistance inducing mutations have already been identified [79].

1.8.3 Investigational or experimental medicines

Among other potential repurposed drugs that received attention earlier in the pandemic some can be highlighted, particularly, ivermectin, hydroxychloroquine and azithromycin [90]. A clear case of the challenges with repurposing drugs is highlighted by this triad of candidates that

were focused early on the pandemic [91]. In several studies, they showed a strong ability to inhibit viral infection in cell based *in vitro* studies, and they were used off-label due to this [92]. However, further research showed that cationic amphiphilic drugs (including hydroxychloroquine and azithromycin) induce phospholipidosis, which can be misinterpreted as antiviral efficacy [93]. As for ivermectin, the concentrations reported in successful experiments are far higher than what can be safely reached in human plasma, with standard dosages of ivermectin [94]. Further clinical trials have not supported ivermectin's efficacy as an antiviral treatment for COVID-19 [84, 95].

Implementation of repurposed drugs also shows some limitations, with candidate drugs requiring rigorous assessment in preclinical and clinical studies, particularly when the mechanism of action is not clear. Even when repurposing other antiviral drugs, the challenge remains difficult, with trial-and-error efforts to affect RNA synthesis, given the variability of mechanisms used by different virus [79]. For CoVs the unique exonuclease activity of the RdRp eliminated many candidates, while antivirals targeting DNA polymerases lack the rationale to be functional, with the same limitations present with protease inhibitors, where low protease similarity limits activity, such as HIV [79]. It is important to note that while repurposing candidates show promise, their use should be restricted to clinical trials, and off-label use should be avoided until sufficient and robust evidence supporting their efficacy and safety is available [93].

1.9 Computational approaches in antiviral research

The development of antiviral drugs, particularly for SARS-CoV-2, requires a detailed understanding of viral structures and the mechanisms by which they interact with host cells [80, 96]. While drug repurposing has provided a rapid solution to identify potential antiviral candidates, structure-based drug design, through computational methods, offers a more targeted and cost-effective pathway to develop specific therapies against SARS-CoV-2 [80, 96, 97]. These computational techniques enable the rational identification of binding sites on viral targets and the design of highly specific drugs to target these sites effectively [96]. Unlike repurposed drugs, which are typically characterized by suboptimal efficacy or off-target effects, structure-based drug design allows the identification of novel scaffolds tailored to viral targets, enhancing drug potency and reducing the potential for resistance development [97]. Furthermore, iterative cycles of computational simulations and structural refinement can enhance the pharmacodynamic and pharmacokinetic properties of drug candidates, thereby optimizing their efficacy and safety profiles. Past successes in antiviral drug design, such as Oseltamivir for *Influenza* virus and HIV protease inhibitors, highlighting the value of rational drug design aided by structural insights, from technologies like X-ray crystallography and cryo-

electron microscopy (Cryo-EM) [80, 98, 99]. These methodologies have been instrumental in identifying small molecules that disrupt viral functions, transforming an infectious disease like AIDS from a fatal condition to a manageable chronic illness [96, 99]. Lessons learned from these past efforts highlight the potential of structure-based drug design in the fight against SARS-CoV-2, providing hope for targeted and effective therapies.

1.9.1 Molecular modelling and molecular dynamics

Molecular modelling is the creation of models of molecules, involving all methods, theoretical or computational, to simulate and study their behaviour in molecular systems. In these systems, molecules are typically represented as mechanical entities in which particles are connected by bonds, that can rotate or vibrate into favourable conformations in response to external forces, such as van der Waals interactions or electrostatic forces [80, 96, 100]. These systems are dynamic, characterized by continuous movement and interactions, and rely on iterative predictions, with the assumption that any future state of the system is determined based on the current state. The process involves integrating equations of motion and performing calculations into very short time steps, typically on the order of one femtosecond. Although these calculations can be performed by hand, the complexity of even small molecular systems inevitably requires computational resources [101]. At each step, the forces acting on each atom are computed and combined with their current positions and velocities to generate updated positions and velocities. The result is a trajectory that captures the dynamic changes in the system over time [80, 101]. Molecular dynamics (MD) simulations are a computational approach to this problem, allowing estimations in complex systems, that cannot be solved analytically [80]. Although static representations from crystallography provide valuable structural snapshots, they may lack the dynamic information needed to understand biological activity fully. MD simulations gather data on the physical movements of atoms and molecules, including molecular geometries, mean atomic fluctuations, local fluctuations, binding interactions and free energy calculations, providing insights where static images are limited [80, 100].

1.9.2 Molecular docking

Docking is a molecular modelling technique used to predict the preferred orientation (or pose) of a molecule within a defined receptor target. Following principles similar to molecular modelling, molecular docking aims to identify the pose that minimizes total binding energy, while achieving favourable bond angles and torsions [80, 101, 102]. Given that organic molecules have many conformational degrees of freedom, it is challenging to find the optimal conformation that aligns with the receptor in a reasonable time frame [80]. Docking is

particularly valuable in small molecule structure-based drug design since it predicts the binding conformation of small molecules in target binding sites, as well as binding affinity, accelerating the search for effective drugs [102, 103]. Despite progress, no single algorithm efficiently balances docking quality and speed, so the most effective procedures are typically iterative and apply multiple docking techniques [103, 104].

1.9.3 Scoring functions

Scoring functions (SF) are essential in molecular docking, addressing two key challenges: identifying the ligand's binding mode on the site and predicting the binding affinity, enabling ranking of potential ligands for virtual screening analysis [96]. Each SF type approaches these tasks differently, and they can be broadly divided into four categories: physics-based, empirical, knowledge-based, and machine-learning-based [96, 105]. Physics-based SFs calculate binding energies using force fields, considering cumulative effects of van der Waals and electrostatic interactions between protein-ligand atom pairs, while empirical SFs estimate binding affinity by summing energetic factors (e.g., hydrogen bonding, hydrophobic interactions) from experimentally determined affinities, making them computationally more efficient but potentially less accurate for complex structural changes [96, 105]. Knowledge-based SFs use experimentally determined molecular structures and statistical potentials derived from known molecular interactions to refine models and predict interactions, leveraging patterns from atomic interactions and structural motifs [105], while machine learning-based SFs utilize large datasets of protein-ligand interactions, incorporating structural, physicochemical and binding affinity data. By learning complex data patterns, these models often achieve higher accuracy, though they require large, high-quality datasets and can be computationally demanding [96, 106]. In addition to types of scoring functions, consensus-scoring functions offer an improved approach by combining the strengths of multiple scoring types [96]. This technique involves rescoring predicted poses multiple times using different scoring functions and then integrating these results to rank solutions [96]. By pairing functions that excel at differentiating active and inactive compounds, with those that better identify correct binding poses, it enhances the reliability of predictions, by balancing strengths and compensating for individual limitations, making consensus scoring a valuable tool in structure-based drug design [96].

1.9.4 Drug discovery strategies

Virtual screening (VS) is a computational technique used in drug discovery to identify potential drug candidates from large chemical libraries by evaluating their likelihood to bind a target of interest, such as a protein or enzyme. It offers an efficient and cost-effective alternative to

high-throughput screening, which requires experimentally testing large numbers of compounds [103, 107]. The basis of structure-based VS is molecular docking, the most widely used method for prioritizing compounds based on their binding affinity to a biological target [80, 107]. The key for the VS is, then, the correct prediction of the binding mode and binding efficiency of the ligand [96, 107]. However, while VS serves as an effective tool to identify potential hits, the most costly and time-consuming phase in pre-clinical research occurs between the identification of a hit molecule and its optimization into a clinical candidate. This is an iterative process, with a strong empirical basis that requires the design, synthesis and testing of hundreds to thousands of compounds, over several years [80]. The key to identify new analogues with better properties are structure–activity relationships (SAR), relating structural modifications to biochemical properties [80]. The experimental synthesis of thousands of compounds and their subsequent evaluation for biological activity is both financially and logistically expensive. In contrast, computational methods, particularly the quantitative structure–activity relationship (QSAR) approach, can rationalize the synthesis process and significantly reduce costs in drug discovery [80, 108]. QSAR is a computational approach that quantifies the relationship between a drug's physicochemical properties and its biological activity, as well as its molecular properties, such as dipole moment and hydrophobicity, thereby evolving into an essential tool in computer-assisted drug discovery [108]. Furthermore, QSAR approaches can assess biological properties related to absorption, distribution, metabolism, excretion, and toxicity, as well as predict changes in potency based on structural modifications. This is based on the similarity-property principle, which posits that structurally similar compounds tend to exhibit similar biological effects [109]. This principle underlies the rationale for linear QSAR models, where gradual structural changes correlate with gradual alterations in potency. Similar compounds share a common core, which allows for predictable potency variations resulting from small substitutions [110].

References

1. Woo, P.C., et al., 2009, *Coronavirus diversity, phylogeny and interspecies jumping*. Exp Biol Med (Maywood). **234**(10): p. 1117-27.
2. Skowronski, D.M., et al., 2005, *Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus (SARS): a year in review*. Annu Rev Med. **56**: p. 357-81.
3. Wuhan Municipal Health Commission. *Report of clustering pneumonia of unknown etiology in Wuhan City*. 2019; Available from: <http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989>. Accessed : 2024/09/24
4. Naqvi, A.A.T., et al., 2020, *Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach*. Biochim Biophys Acta Mol Basis Dis. **1866**(10): p. 165878.
5. Wang, M.Y., et al., 2020, *SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development*. Front Cell Infect Microbiol. **10**: p. 587269.
6. Andersen, K.G., et al., 2020, *The proximal origin of SARS-CoV-2*. Nat Med. **26**(4): p. 450-452.
7. Naseer, S., et al., 2023, *COVID-19 outbreak: Impact on global economy*. Frontiers in Public Health. **10**.
8. Woo, P.C.Y., et al., 2023, *ICTV Virus Taxonomy Profile: Coronaviridae 2023*. J Gen Virol. **104**(4).
9. Lu, R., et al., 2020, *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding*. Lancet. **395**(10224): p. 565-574.
10. Schoeman, D., B. Gordon, and B.C. Fielding, 2021, *Pathogenic Human Coronaviruses*. Reference Module in Biomedical Sciences. p. B978-0-12-818731-9.00052-5.
11. Woo, P.C., et al., 2012, *Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus*. J Virol. **86**(7): p. 3995-4008.
12. Subissi, L., et al., 2014, *One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities*. Proc Natl Acad Sci USA. **111**(37): p. E3900-9.
13. Graham, R.L. and R.S. Baric, 2010, *Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission*. J Virol. **84**(7): p. 3134-46.
14. Vijaykrishna, D., et al., 2007, *Evolutionary insights into the ecology of coronaviruses*. J Virol. **81**(8): p. 4012-20.
15. Thakor, J.C., et al., 2022, *Swine coronaviruses (SCoVs) and their emerging threats to swine population, inter-species transmission, exploring the susceptibility of pigs for SARS-CoV-2 and zoonotic concerns*. Vet Q. **42**(1): p. 125-147.
16. Fan, Y., et al., 2019, *Bat Coronaviruses in China*. Viruses. **11**(3): p. 210.
17. Tyrrell, D.A. and M.L. Bynoe, 1965, *Cultivation of a Novel Type of Common-Cold Virus in Organ Cultures*. Br Med J. **1**(5448): p. 1467-70.
18. Hamre, D. and J.J. Procknow, 1966, *A new virus isolated from the human respiratory tract*. Proc Soc Exp Biol Med. **121**(1): p. 190-3.
19. Lau, S.K., et al., 2005, *Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats*. Proc Natl Acad Sci U S A. **102**(39): p. 14040-5.

20. Alyami, M.H., H.S. Alyami, and A. Warraich, 2020, *Middle East Respiratory Syndrome (MERS) and novel coronavirus disease-2019 (COVID-19): From causes to preventions in Saudi Arabia*. Saudi Pharm J. **28**(11): p. 1481-1491.
21. Meyer, B., et al., 2014, *Antibodies against MERS coronavirus in dromedary camels, United Arab Emirates, 2003 and 2013*. Emerg Infect Dis. **20**(4): p. 552-9.
22. Dawson, P., et al., 2019, *What Have We Learned About Middle East Respiratory Syndrome Coronavirus Emergence in Humans? A Systematic Literature Review*. Vector Borne Zoonotic Dis. **19**(3): p. 174-192.
23. Cotten, M., et al., 2013, *Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study*. Lancet. **382**(9909): p. 1993-2002.
24. Memish, Z.A., et al., 2013, *Middle East respiratory syndrome coronavirus in bats, Saudi Arabia*. Emerg Infect Dis. **19**(11): p. 1819-23.
25. Chu, D.K., et al., 2015, *Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels in Nigeria, 2015*. Euro Surveill. **20**(49).
26. WHO Timeline - COVID-19. 2020; Available from: <https://www.who.int/news/item/27-04-2020-who-timeline---covid-19>. Accessed: 2024/07/10
27. Organization, W.H., *True extent of SARS-CoV-2 Infection through seroprevalence studies*. February 3 ed. 2022: World Health Organization.
28. Bergeri, I., et al., 2022, *Global SARS-CoV-2 seroprevalence from January 2020 to April 2022: A systematic review and meta-analysis of standardized population-based studies*. PLOS Medicine. **19**(11): p. e1004107.
29. WHO. *Coronavirus Disease (COVID-19) Dashboard*. 2020; Available from: <https://covid19.who.int/>. Accessed: 2021/09/04
30. Deng, S.Q. and H.J. Peng, 2020, *Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China*. J Clin Med. **9**(2).
31. Chan, J.F., et al., 2020, *A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster*. Lancet. **395**(10223): p. 514-523.
32. WHO. *Coronavirus Disease (COVID-19) Situation in the WHO European Region*.]; Available from: <https://www.who.int/europe/emergencies/situations/covid-19>. Accessed: 2024/9/25
33. Data, O.W.i. *COVID-19 Data Explorer*. 2020; Available from: <https://ourworldindata.org/explorers/covid>. Accessed: 2024/9/25
34. Zhang, X.Y., et al., 2020, *Biological, clinical and epidemiological features of COVID-19, SARS and MERS and AutoDock simulation of ACE2*. Infect Dis Poverty. **9**(1): p. 99.
35. Griffin, I., et al., 2024, *Estimates of SARS-CoV-2 Hospitalization and Fatality Rates in the Prevaccination Period, United States*. Emerging Infectious Disease journal. **30**(6): p. 1144.
36. Pustake, M., et al., 2022, *SARS, MERS and CoVID-19: An overview and comparison of clinical, laboratory and radiological features*. J Family Med Prim Care. **11**(1): p. 10-17.
37. Huang, C., et al., 2020, *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. Lancet. **395**(10223): p. 497-506.
38. O'Toole, Á., et al. *Pangolin COVID-19 Lineage Assigner*. 2021. Available from: <https://pangolin.cog-uk.io/>. Accessed: 2021-09-01.
39. Kadam, S.B., et al., 2021, *SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights*. J Basic Microbiol.
40. Wacharapluesadee, S., et al., 2021, *Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia*. Nat Commun. **12**(1): p. 972.
41. Boni, M.F., et al., 2020, *Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic*. Nat Microbiol. **5**(11): p. 1408-1417.

42. Holmes, E.C., et al., 2021, *The origins of SARS-CoV-2: A critical review*. Cell. **184**(19): p. 4848-4856.
43. Liu, P., et al., 2020, *Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?* PLOS Pathogens. **16**(5): p. e1008421.
44. Schindell, B.G., et al., 2022, *SARS-CoV-2 and the Missing Link of Intermediate Hosts in Viral Emergence - What We Can Learn From Other Betacoronaviruses*. Frontiers in Virology. **2**.
45. Hu, B., et al., 2021, *Characteristics of SARS-CoV-2 and COVID-19*. Nat Rev Microbiol. **19**(3): p. 141-154.
46. Yang, H. and Z. Rao, 2021, *Structural biology of SARS-CoV-2 and implications for therapeutic development*. Nature Reviews Microbiology. **19**(11): p. 685-700.
47. Almehtdi, A.M., et al., 2021, *SARS-CoV-2 spike protein: pathogenesis, vaccines, and potential therapies*. Infection. **49**(5): p. 855-876.
48. V'kovski, P., et al., 2021, *Coronavirus biology and replication: implications for SARS-CoV-2*. Nature Reviews Microbiology. **19**(3): p. 155-170.
49. Wu, C.-r., et al., 2022, *Characteristics of SARS-CoV-2 and COVID: drug design templates for COVID-19*. Acta Pharmacologica Sinica. **43**(12): p. 3021-3033.
50. Chen, J., et al., 2020, *Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex*. Cell. **182**(6): p. 1560-1573.e13.
51. Hegyi, A. and J. Ziebuhr, 2002, *Conservation of substrate specificities among coronavirus main proteases*. Journal of General Virology. **83**(3): p. 595-599.
52. Mukherjee, R. and I. Dikic, 2023, *Proteases of SARS Coronaviruses*. Encyclopedia of Cell Biology. p. 930-41.
53. Li, X. and Y. Song, 2023, *Structure and function of SARS-CoV and SARS-CoV-2 main proteases and their inhibition: A comprehensive review*. European Journal of Medicinal Chemistry. **260**: p. 115772.
54. Lv, Z., et al., 2021, *Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development*. Front Chem. **9**: p. 819165.
55. Owen, D.R., et al., 2021, *An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19*. Science. **374**(6575): p. 1586-1593.
56. Zhang, Z., et al., 2022, *Structure of SARS-CoV-2 membrane protein essential for virus assembly*. Nature Communications. **13**(1): p. 4399.
57. Thomas, S., 2020, *The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET*. Pathog Immun. **5**(1): p. 342-363.
58. Lu, S., et al., 2021, *The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein*. Nat Commun. **12**(1): p. 502.
59. Izadi, A. and P. Nordenfelt, 2024, *Protective non-neutralizing SARS-CoV-2 monoclonal antibodies*. Trends in Immunology. **45**(8): p. 609-624.
60. Szymczak, A., et al., 2021, *Antibodies specific to SARS-CoV-2 proteins N, S and E in COVID-19 patients in the normal population and in historical samples*. J Gen Virol. **102**(11).
61. Zhou, S., et al., 2023, *SARS-CoV-2 E protein: Pathogenesis and potential therapeutic development*. Biomed Pharmacother. **159**: p. 114242.
62. Huang, Y., et al., 2020, *Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19*. Acta Pharmacologica Sinica. **41**(9): p. 1141-1149.
63. Li, F., 2016, *Structure, Function, and Evolution of Coronavirus Spike Proteins*. Annu Rev Virol. **3**(1): p. 237-261.
64. Walls, A.C., et al., 2020, *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein*. Cell. **181**(2): p. 281-292.e6.

65. Wrapp, D., et al., 2020, *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation*. Science. **367**(6483): p. 1260-1263.
66. Zhu, C., et al., 2021, *Molecular biology of the SARs-CoV-2 spike protein: A review of current knowledge*. J Med Virol. **93**(10): p. 5729-5741.
67. Johnson, B.A., et al., 2021, *Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis*. Nature. **591**(7849): p. 293-299.
68. Young, M., et al., 2022, *Covid-19: virology, variants, and vaccines*. BMJ Med. **1**(1): p. e000040.
69. Tao, K., et al., 2021, *The biological and clinical significance of emerging SARS-CoV-2 variants*. Nature Reviews Genetics. **22**(12): p. 757-773.
70. ECDC. *European Centre for Disease Prevention and Control*. 2021 2021/07/29]; Available from: <https://www.ecdc.europa.eu/en>.
71. WHO. *Tracking SARS-CoV-2 variants*. 2021; Available from: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Accessed: 2021/09/14
72. Girl, P., et al., 2024, *Comparative Analysis of Vaccine-Induced Neutralizing Antibodies against the Alpha, Beta, Delta, and Omicron Variants of SARS-CoV-2*. Vaccines. **12**(5): p. 515.
73. Esper, F.P., et al., 2022, *Alpha to Omicron: Disease Severity and Clinical Outcomes of Major SARS-CoV-2 Variants*. The Journal of Infectious Diseases. **227**(3): p. 344-352.
74. European Medicines Agency ,2024, *COVID-19 medicines*. Available from: <https://www.ema.europa.eu/en/human-regulatory-overview/public-health-threats/coronavirus-disease-covid-19/covid-19-medicines>. Accessed: 2024/09/15
75. Centers for Disease Control and Prevention, 2024, *Stay Up to Date with COVID-19 Vaccines*. Available from: <https://www.cdc.gov/covid/vaccines/stay-up-to-date.html>. Accessed: 2024/09/15
76. Padda, I.S. and M. Parmar, *COVID (SARS-CoV-2) Vaccine*, StatPearls, Editor. 2023, StatPearls Publishing: Treasure Island (FL).
77. Watson, O.J., et al., 2022, *Global impact of the first year of COVID-19 vaccination: a mathematical modelling study*. The Lancet Infectious Diseases. **22**(9): p. 1293-1302.
78. Kim, S., 2022, *COVID-19 Drug Development*. J Microbiol Biotechnol. **32**(1): p. 1-5.
79. Li, G., et al., 2023, *Therapeutic strategies for COVID-19: progress and lessons learned*. Nature Reviews Drug Discovery. **22**(6): p. 449-475.
80. *The Handbook of Medicinal Chemistry: Principles and Practice*, ed. S.E. Ward and A. Davis. 2023: The Royal Society of Chemistry.
81. Drożdżal, S., et al., 2021, *An update on drugs with therapeutic potential for SARS-CoV-2 (COVID-19) treatment*. Drug Resistance Updates. **59**: p. 100794.
82. Warren, T.K., et al., 2016, *Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys*. Nature. **531**(7594): p. 381-5.
83. Gordon, C.J., et al., 2020, *Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency*. J Biol Chem. **295**(20): p. 6785-6797.
84. 2022, *Remdesivir and three other drugs for hospitalised patients with COVID-19: final results of the WHO Solidarity randomised trial and updated meta-analyses*. Lancet. **399**(10339): p. 1941-1953.
85. Ader, F., et al., 2022, *Remdesivir plus standard of care versus standard of care alone for the treatment of patients admitted to hospital with COVID-19 (DisCoVeRy): a phase 3, randomised, controlled, open-label trial*. Lancet Infect Dis. **22**(2): p. 209-221.
86. Toots, M., et al., 2019, *Characterization of orally efficacious influenza drug with high resistance barrier in ferrets and human airway epithelia*. Sci Transl Med. **11**(515).

87. Malone, B. and E.A. Campbell, 2021, *Molnupiravir: coding for catastrophe*. Nature Structural & Molecular Biology. **28**(9): p. 706-708.
88. Brown, A.N., et al., 2022, *Why Molnupiravir Fails in Hospitalized Patients*. mBio. **13**(6): p. e02916-22.
89. Zhou, S., et al., 2021, *β -d-N4-hydroxycytidine Inhibits SARS-CoV-2 Through Lethal Mutagenesis But Is Also Mutagenic To Mammalian Cells*. J Infect Dis. **224**(3): p. 415-419.
90. Nittari, G., et al., 2020, *Current pharmacological treatments for SARS-COV-2: A narrative review*. European Journal of Pharmacology. **882**: p. 173328.
91. Caly, L., et al., 2020, *The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro*. Antiviral Research. **178**: p. 104787.
92. Pastick, K.A., et al., 2020, *Review: Hydroxychloroquine and Chloroquine for Treatment of SARS-CoV-2 (COVID-19)*. Open Forum Infectious Diseases. **7**(4).
93. Tummino, T.A., et al., 2021, *Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2*. Science. **373**(6554): p. 541-547.
94. Schmith, V.D., J.J. Zhou, and L.R.L. Lohmer, 2020, *The Approved Dose of Ivermectin Alone is not the Ideal Dose for the Treatment of COVID-19*. Clin Pharmacol Ther. **108**(4): p. 762-765.
95. López-Medina, E., et al., 2021, *Effect of Ivermectin on Time to Resolution of Symptoms Among Adults With Mild COVID-19: A Randomized Clinical Trial*. Jama. **325**(14): p. 1426-1435.
96. Sliwoski, G., et al., 2014, *Computational methods in drug discovery*. Pharmacol Rev. **66**(1): p. 334-95.
97. Pushpakom, S., et al., 2019, *Drug repurposing: progress, challenges and recommendations*. Nat Rev Drug Discov. **18**(1): p. 41-58.
98. Wlodawer, A., et al., 2008, *Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures*. Febs j. **275**(1): p. 1-21.
99. Wlodawer, A. and J.W. Erickson, 1993, *Structure-based inhibitors of HIV-1 protease*. Annu Rev Biochem. **62**: p. 543-85.
100. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*. 2010: Springer New York.
101. Leach, A.R., *Molecular Modelling: Principles and Applications*. 2001: Prentice Hall.
102. Lengauer, T. and M. Rarey, 1996, *Computational methods for biomolecular docking*. Current Opinion in Structural Biology. **6**(3): p. 402-406.
103. Kitchen, D.B., et al., 2004, *Docking and scoring in virtual screening for drug discovery: methods and applications*. Nature Reviews Drug Discovery. **3**(11): p. 935-949.
104. Chen, T., et al., 2023, *Algorithm selection for protein–ligand docking: strategies and analysis on ACE*. Scientific Reports. **13**(1): p. 8219.
105. Li, J., A. Fu, and L. Zhang, 2019, *An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking*. Interdiscip Sci. **11**(2): p. 320-328.
106. Ashraf, F.B., et al., 2023, *Bio-activity prediction of drug candidate compounds targeting SARS-Cov-2 using machine learning approaches*. PLoS One. **18**(9): p. e0288053.
107. Shoichet, B.K., 2004, *Virtual screening of chemical libraries*. Nature. **432**(7019): p. 862-865.
108. Saleh, N.A., H. Elhaes, and M. Ibrahim, *Chapter 2 - Design and Development of Some Viral Protease Inhibitors by QSAR and Molecular Modeling Studies*, in *Viral Proteases and Their Inhibitors*, S.P. Gupta, Editor. 2017, Academic Press. p. 25-58.
109. López-Pérez, K., et al., 2024, *Molecular similarity: Theory, applications, and perspectives*. Artificial Intelligence Chemistry. **2**(2): p. 100077.
110. Cruz-Monteaudo, M., et al., 2014, *Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde?* Drug Discovery Today. **19**(8): p. 1069-1080.



Chapter II

Objectives and Aims

*“The greatest discoveries often lie not in finding new things, but in seeing familiar things in
new ways”*

Alexander Fleming

Chapter II – Objectives and Outline of Thesis

2.1 Objectives

2.1.1 General objective

In early 2020, research on SARS-CoV-2 primarily focused on understanding its origins and genetic makeup, with a particular emphasis on viral phylogenetics and mutations, to better comprehend the viral crossover event. However, as the pandemic spread across the world and the socioeconomic impacts became more pronounced, the research focus quickly shifted to potential new therapies, with drug repurposing at the forefront of this effort. Although emergency approved drugs were used throughout the first months, many quickly had their efficacy disproven, while three received authorizations for COVID-19. Nevertheless, none of these targeted the most important protein involved in viral recognition and infection: the S protein. This target has been successfully explored in vaccine research and for antibody therapies, but no small molecules had been approved, highlighting a research gap that this thesis sought to address.

The main aim of this thesis was to explore the SARS-CoV-2 spike protein and its role in viral entry into host cells, with the ultimate goal of exploring new molecular targets for therapeutic development. By investigating these mechanisms, new bioactive scaffolds that could inhibit spike-mediated entry may provide potential antiviral strategies for SARS-CoV-2 and related coronaviruses.

2.1.2 Specific Objectives

At the outset of this study, the S protein had not been fully identified, with its behaviour remaining only partially understood. Therefore, the first objective of this thesis was to understand the virus host interaction, and based on this, discover novel scaffolds with antiviral activity. To address this overall objective, it was divided into two specific, sequential objectives, each designed to contribute to the identification of potential targets and guide the development of viral inhibitors.

Exploration of S protein and identification of pharmacological targets

The S protein of SARS-CoV-2 plays a critical role in viral entry by binding to the ACE2 receptor on host cells. Understanding the structure and dynamic behaviour of the S protein is crucial to identify regions that are essential for this interaction and potential promising sites for pharmacological exploration. This objective focused on an in-depth analysis of the virus-host interaction, namely the important contacts and how this interaction can be modulated.

Additionally, it aimed to elucidate how the binding of ACE2 affects S1/S2 interaction, the importance of this interaction in cell fusion and how mutations can influence infectivity.

Hit molecule identification and optimization

The most promising target identified previously was explored with virtual screening studies, using commercially available libraries of small molecules. The goal was to discover potential molecular scaffolds for further experimental evaluation, that could potentially bind the proposed targets and disrupt S protein-ACE2 and thereby affect the viral lifecycle, providing a foundation for the development of novel antiviral therapies targeting SARS-CoV-2.

2.2 Outline of the thesis

The research described in this thesis followed a step-by-step approach, with the results organized in four scientific manuscripts. Three of these are already published in peer-reviewed international scientific journals, while a fourth is currently in review stage.

The results of this thesis are reflected in the published and submitted articles, with a chapter corresponding to each objective. In the first chapter the structure of the S protein is discussed, along with multiple identified targets and their potential for further exploration, culminating in the selection of one target to be the focus on the project. In the second chapter, the selected target was explored through molecular dynamics simulations providing further insights into the stability and dynamics of the protein-ligand complexes. These insights enabled VS studies, which involved the identification of potential binding compounds based on their interaction profiles and experimental validation. This was followed by an optimization stage for the most active scaffold, with structural modifications were made to enhance binding affinity and study structure activity relationships.

2.2.1 Exploration of S protein and identification of pharmacological targets

- Queirós-Reis, L; Gomes da Silva, P; Gonçalves, J; Brancale, A; Bassetto, M; Mesquita, JR SARS-CoV-2 Virus-Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *Int. J. Mol. Sci.* 2021, 22, 10836. <https://doi.org/10.3390/ijms221910836>

2.2.2 Hit Molecule Identification and optimization

- Queirós-Reis, L; Mesquita, J; Brancale, A; Bassetto, M (2023). Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein - Confirmed and Potential Ligands. *Journal of chemical information and modeling.* 63. 10.1021/acs.jcim.3c00803

- Queirós-Reis, L; Alvites, R; Maurício, AC; Brancale, A; Bassetto, M; Mesquita, JR (2024). Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study. *International journal of molecular sciences*, 26(1), 151. <https://doi.org/10.3390/ijms26010151>
- Queirós-Reis, L; Alvites, R; Maurício, AC; Brancale, A; Bassetto, M; Mesquita, J.R. (2024). *In silico* discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding pocket. *ACS Omega*



Chapter III

Results

“The reward of the young scientist is the emotional thrill of being the first person in the history of the world to see something or understand something”

Cecilia Payne-Gaposchkin

Chapter III – Results

3.1 Exploration of S protein structure and behaviour

3.1.1 SARS-CoV-2 Virus-Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response



Review

SARS-CoV-2 Virus—Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response

Luís Queirós-Reis ¹, Priscilla Gomes da Silva ^{1,2,3}, José Gonçalves ⁴, Andrea Brancale ⁵, Marcella Bassetto ⁶ and João R. Mesquita ^{1,2,*}

¹ Abel Salazar Institute of Biomedical Sciences (ICBAS), University of Porto, 4050-313 Porto, Portugal; up201205115@up.pt (L.Q.-R.); priscilla@ua.pt (P.G.d.S.)

² Epidemiology Research Unit (EPIunit), Institute of Public Health, University of Porto, 4050-091 Porto, Portugal

³ LEPABE—Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal

⁴ Institute of Sustainable Processes, University of Valladolid, 47011 Valladolid, Spain; zemcg5@gmail.com

⁵ Cardiff School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff CF10 3NB, UK; brancalea@cardiff.ac.uk

⁶ Department of Chemistry, Faculty of Science and Engineering, Swansea University, Swansea SA2 8PP, UK; marcella.bassetto@swansea.ac.uk

* Correspondence: jrmesquita@icbas.up.pt



Citation: Queirós-Reis, L.; Gomes da Silva, P.; Gonçalves, J.; Brancale, A.; Bassetto, M.; Mesquita, J.R. SARS-CoV-2 Virus—Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *Int. J. Mol. Sci.* **2021**, *22*, 10836. <https://doi.org/10.3390/ijms221910836>

Academic Editor: Luisa Barzon

Received: 3 September 2021

Accepted: 1 October 2021

Published: 7 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Coronavirus disease 19, or COVID-19, is an infection associated with an unprecedented worldwide pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which has led to more than 215 million infected people and more than 4.5 million deaths worldwide. SARS-CoV-2 cell infection is initiated by a densely glycosylated spike (S) protein, a fusion protein, binding human angiotensin converting enzyme 2 (hACE2), that acts as the functional receptor through the receptor binding domain (RBD). In this article, the interaction of hACE2 with the RBD and how fusion is initiated after recognition are explored, as well as how mutations influence infectivity and immune response. Thus, we focused on all structures available in the Protein Data Bank for the interaction between SARS-CoV-2 S protein and hACE2. Specifically, the Delta variant carries particular mutations associated with increased viral fitness through decreased antibody binding, increased RBD affinity and altered protein dynamics. Combining both existing mutations and mutagenesis studies, new potential SARS-CoV-2 variants, harboring advantageous S protein mutations, may be predicted. These include mutations S13I and W152C, decreasing antibody binding, N460K, increasing RDB affinity, or Q498R, positively affecting both properties.

Keywords: SARS-CoV-2; Spike protein; hACE2; protein structures

1. Introduction

Coronavirus disease 19, or COVID-19, is an infection associated with an unprecedented global pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which resulted in more than 215 million infected people and more than 4.5 million deaths [1–3]. SARS-CoV-2 was first identified in December 2019 in Wuhan, Hubei province of China, in individuals exposed at a seafood/wet market and was successfully isolated and sequenced in January 2020 [1]. Genetic evidence suggests that the virus originates from animals, particularly bats, as SARS-CoV-2 has 96% sequence identity to a strain found in bats (BatCov RaTG13), possibly having pangolins, minks, turtles, or snakes as the intermediate host [1,4].

Coronaviruses (CoVs) are a group of RNA viruses belonging to the subfamily *Orthocoronaviridae* that cause disease in a variety of domestic and wild animals (e.g. porcine,

bovine, feline and avian, and bat and whale strains) [5]. A defining characteristic in all CoVs is the crown-like viral particle, from which their name is derived [6,7]. It consists of three major structural proteins (spike, membrane, and envelope), that protrude from the viral envelope and have a critical function in cell recognition and fusion [6]. Additionally, the nucleocapsid (protein N) binds and stabilizes the viral genome (Figure 1) [8].

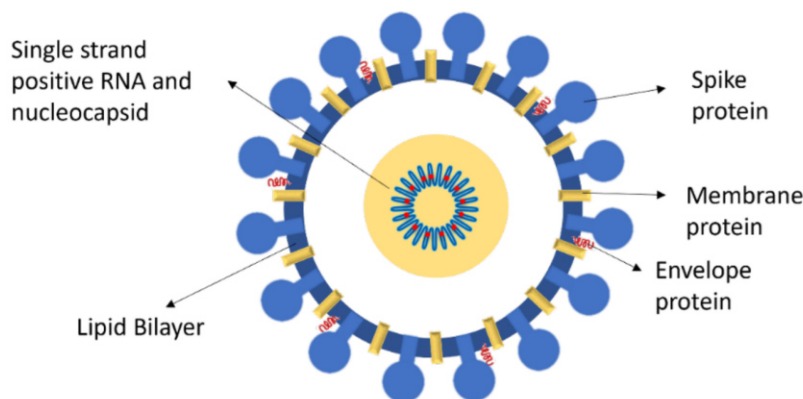


Figure 1. Representation of SARS-CoV-2 main structural proteins.

Furthermore, they have the second largest RNA viral genome, a positive sense genome of single-stranded RNA (+ssRNA), with an average size of 30 kb [9]. Importantly, CoVs have a large genome with a high mutation rate (one nucleotide per 1000 to 10,000 nucleotides replicated), associated with the RNA dependent RNA polymerase, and a random template switching during RNA duplications, that causes a high frequency of homologous RNA recombination [5]. These characteristics are responsible for the special plasticity of CoVs when it comes to accommodating and modifying genes, explaining the broad range of hosts infected by them [5].

Taken together, these factors have led to a diversity of strains and genotypes highly adaptable to new hosts and ecological niches, causing major zoonotic outbreaks [5].

SARS-CoV-2 is transmitted via aerosols or droplets, typically within one metre or in poorly ventilated and/or crowded indoor settings, and possibly by touching contaminated and eyes, nose, or mouth, with both symptomatic and asymptomatic patients being the main source of infection [1,10]. The immune response appears to be characteristic of each individual, with a unique collection of antibodies (Abs) and varying efficacy of viral neutralization potency observed [11].

Over time, numerous SARS-CoV-2 variants have been reported, differing by one or more mutations from the first virus sequenced, the so-called original virus variant, which corresponds to the Wuhan strain [12]. Some of these have been classified as variants of concern (VOC), defined by their features, such as increased transmissibility, virulence and disease severity, decreased Ab neutralization or vaccine efficiency and/or errors in current detection protocols [13–15]. The list of variants of concern is frequently updated according to new information, by the Centers for Disease Control and Prevention (CDC), World Health Organization or the European Centre for Disease Prevention and Control (ECDC) [14–18]. Additionally, some variants are classified as variants of interest (VOI), characterized by mutations with predicted negative features, although with insufficient evidence, by limited prevalence or epidemiological data suggesting an emerging risk to public health [14–16]. Variants can also be de-escalated, particularly if they cease to circulate, if, despite circulating, they have a negligible effect on the epidemiological situation or if scientific evidence shows no concerning features (Table 1) [14].

Table 1. List of variants currently or in the past considered as variants of concern or variants of interest.

WHO Label	Pango Lineage	Origin	Infectivity	Effects on Immune Response	Clinical Relevance	Current Classification (CDC/WHO/ECDC)	References
Alpha	B.1.1.7	United Kingdom Sep-2020	50% increased transmission	Minimal impact	Increased severity	VOC VOC DE	[14,16]
Beta	B.1.351	South Africa May-2020	50% increased transmission	Significant reduction in neutralization	Reduced susceptibility to treatment	VOC VOC VOC	[19–22]
Gamma	P.1	Japan/Brazil Nov-2020	Potential increased transmissibility ¹	Significant reduced neutralization	Significantly reduced susceptibility to treatment	VOC VOC VOC	[22–25]
Delta	B.1.617.2	India Oct-2020	Increased transmissibility	Significant reduction in neutralization	Reduced susceptibility to treatment	VOC VOC VOC	[14,16]
Epsilon	B.1.427/B.1.429	USA Sep-2020	Unclear	Reduced neutralization	No effect reported	NA NA DE	[14]
Zeta	P.2	Brazil Nov-2020	No effect reported	Potential reduced neutralization ¹	No effect reported	NA NA DE	[14]
Eta	B.1.525	Nigeria Dec-2020	No effect reported	Potential reduced neutralization ¹	Potential reduced susceptibility to treatment ¹	VOI VOI DE	[14]
Theta	P.3	Philippines Jan-2021	Potential increased transmissibility ¹	Potential reduced neutralization ¹	No effect reported	NA NA DE	[14]
Iota	B.1.526	USA Dec-2020	No effect reported	Reduced neutralization	Potential reduced susceptibility to treatment ¹	VOI VOI DE	[14,16]
Kappa	B.1.617.1	India Oct-2020	Potential increased transmissibility ¹	Potential reduced neutralization ¹	Potential reduced susceptibility to treatment ¹	VOI VOI DE	[14,16]
Lambda	C.37	Peru Dec-2020	Potential increased transmissibility ¹	Significantly reduction in neutralization	No effect reported	NA VOI VOI	[14,26]
Mu	B.1.621	Colombia Jan-2021	Potential increased transmissibility ¹	Reduced neutralization	No effect reported	NA VOI VOI	[14]

Abbreviations: VOC—variants of concern. VOI—variants of interest. DE—deescalated. NA- not available. Note: ¹ Potential effect is predicted from the mutations present and their effect in previous studies. Not directly observed in the variant.

Glycoprotein: Cell Recognition and Infection

As with all CoVs, SARS-CoV-2 cell infection is initiated by a densely glycosylated spike (S) protein, a trimeric class I fusion protein that binds to human angiotensin converting enzyme 2 (hACE2), acting as the functional receptor [1]. hACE2 expression is ubiquitous in the human body and although the degree of expression varies between organs, it is at its highest in type II alveolar epithelial cells in the lungs, which indicates the primary target for the virus [27]. In addition, hACE2 expression is also high in other tissues, such as myocardial cells, kidney and enterocytes, which explains less common symptoms as well as the multiorgan injuries observed in severe cases [27].

In general, the S protein exists in two main conformations, prefusion and post-fusion, requiring major structural rearrangements after recognition to promote cell fusion [28]. The S protein consists of two subunits (S1 and S2) responsible for cell recognition and membrane fusion, respectively [1]. Although S1 and S2 are originally covalently bound, there is a cleavage site between the subunits, where host serine proteases, such as furin, transmembrane protease serine protease 2 (TMPRSS2) or cathepsin L, cleave the protein [1,11,28]. Notably, TMPRSS2 is co-expressed with hACE2 in nasal epithelial cells, lungs and bronchial branches [11]. When the S1 and S2 subunits are cleaved, the protein is activated and can undergo the necessary conformational changes to achieve the post-fusion conformation and consequently enable cell infection (Figure 2) [1,28].

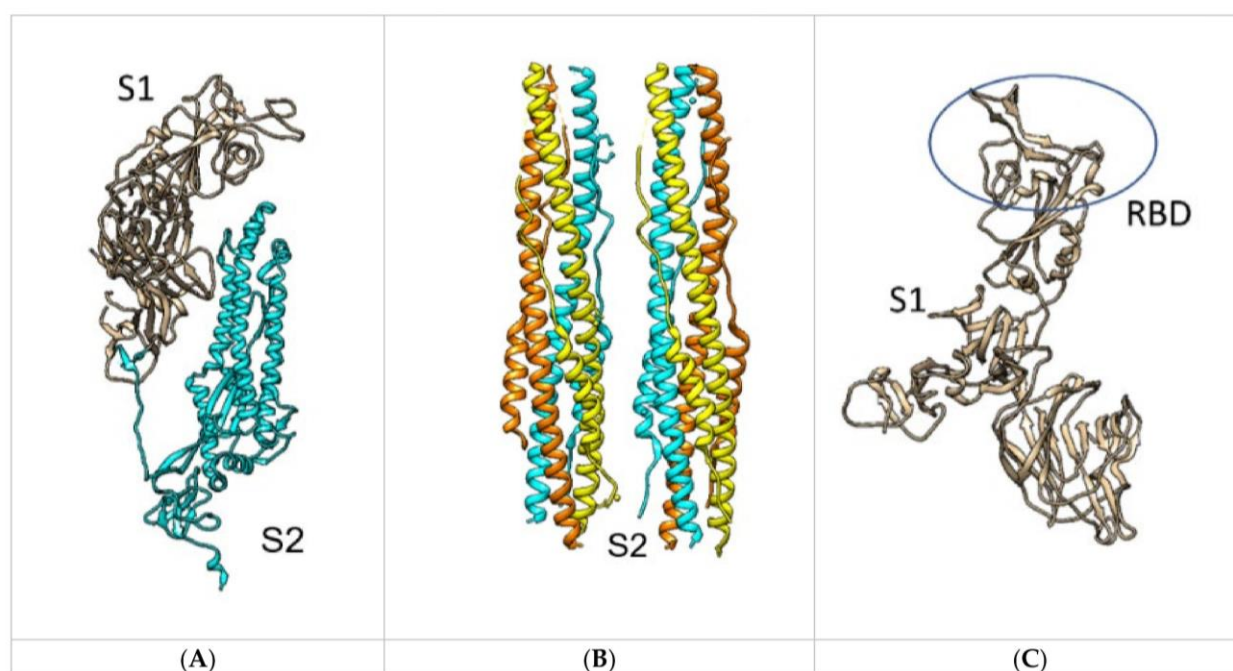


Figure 2. (A) Single chain in trimeric spike protein with subunit S1 (Tan) and S2 (Cyan) in 7A94. (B) Post-fusion conformation of all S2 units in the trimeric spike in 6LXT. (C) S1 subunit, with the RBD highlighted in 7A94 [29].

The S1 subunit formed of the N-terminal domain (NTD) and a receptor binding domain (RBD) that directly interacts with hACE2. Therefore, it is considered a key target for antivirals and neutralizing Abs, accounting for 90% of the neutralizing activity in convalescent sera [1,30–32]. Despite the low percentage of NTD-directed Abs, some are still potent inhibitors [32]. The NTD and two neighbouring regions, namely a β -hairpin (residues 140–158) and a loop (residues 245–264), are collectively termed an antigenic supersite, where NTD directed Abs bind [32]. The NTD-directed Abs neutralizing activity is due to both viral entry blocking and also limiting cell-to-cell spread of the virus [32].

Regarding the RBD, it can be further divided in two structural domains, a highly conserved core, and a variable external subdomain that binds hACE2 [1].

Subunit S2 contains two heptad repeats, a transmembrane domain, a cytoplasmic tail, a central helix, and the fusion peptide that is critical for the main function of S2 of merging the host and viral cell membranes (Figure 2) [31]. Additionally, N-linked glycans are extensively present along the molecule and are essential in proper folding [1,31].

Even if the S protein is activated through the cleavage process, the RBDs can still be inaccessible for interaction [28]. Each RBD can be either down (closed conformation), inaccessible for interaction, or up (open conformation). The open conformation is the necessary state for recognition, with the recognition motifs protruding from the S protein (Figure 3) [1,28].

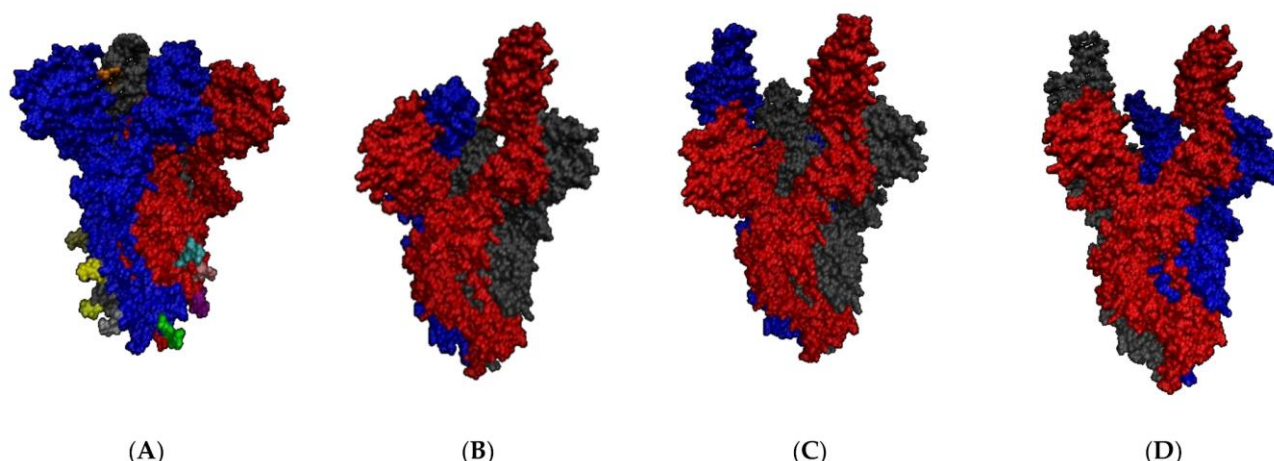


Figure 3. Spike protein conformations, with each spike monomer is represented by a colour (blue, red, gray). (A) All RBD down (PDB 6VXX) (B) 1 RBD up (PDB 7A94) (C) 2 RBD up (PDB 7A97) (D) All RBD up (PDB 7A98) [33].

To engage the host cell receptor, the RBD undergoes hinge-like conformational changes, transiently switching from a down conformation, to a less stable up conformation. Each RBD is capable of individually changing its conformation (Figure 3) [28]. When ACE2 interacts with the RBD, it binds tightly and stabilizes the up conformation [1,34].

In this review, the interaction of hACE2 with the RBD is explored, namely the important contacts between proteins and how this interaction can be modulated to increase affinity or reduce Ab binding. Additionally, we will also try to elucidate how the binding of hACE2 affects S1/S2 interaction, the importance of this interaction in cell fusion and how mutations can influence infectivity and immune response. This information may be important to predict or explain the changes in infectivity or immune response for new variants, which has been studied for the SARS-CoV-2 Delta variant, currently becoming the most prevalent variant worldwide [35].

A total of 37 structures were selected, three of which are complexes that, despite involving the SARS-CoV-2 S protein, also have an Ab or non-human ACE2, and therefore would not reliably represent the critical interactions between the virus and the human protein. The remaining structures can be broadly divided into two categories: studying the RBD-hACE2 interaction or full spike protein behaviour, focusing on the conformation changes from recognition to fusion (Figure 4) [21]. Furthermore, some structures also have hACE2 in a complex with broad neutral amino acid transporter 1 (B0AT1), which stabilizes hACE2 homodimers [36].

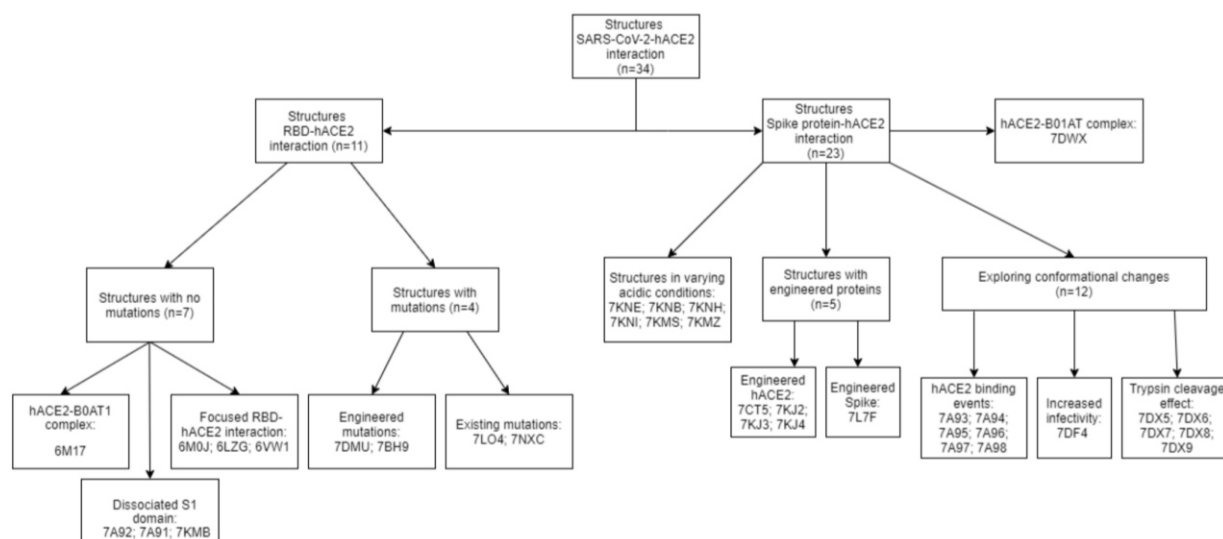


Figure 4. Organizational scheme of all structures found in this review. Molecular graphics analyses were performed with UCSF Chimera and Visual molecular dynamics [29,33]. A homology model for the S protein of SARS-CoV-2 Delta variant was created using SWISS-MODEL [37].

2. Discussion

2.1. SARS-CoV-2 Spike RBD – hACE2 Interaction

2.1.1. Structures with No Mutation

The RBD/hACE2 interaction structures represent the domains of both proteins that contribute to this interaction, providing the basis for computational protein-protein interaction studies, mutation analysis, drug design and computational models of the full spike protein, as it is harder to maintain a similar resolution for significantly larger structures [38–41].

The structures used for these analyses represent the wild-type SARS-CoV-2 S protein/ACE2 interaction without mutations (PDB 6VW1/6M0J/6LZG), with an Root Mean Square Distance (RMSD) of 0.32 Å between them (Figure 5) [39,42].

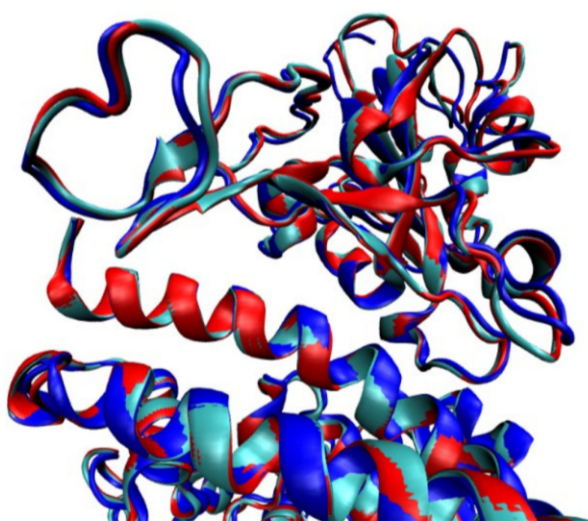


Figure 5. Alignment and superposition of the models representing the interaction with no mutations (PDB 6VW1/6M0J/6LZG) [33].

Consistently, in all structures focusing on this interaction in wild-type SARS-CoV-2 S protein, the contact interface is divided into two clusters at both ends of the $\alpha 1$ helix in hACE2, leaving a central cavity filled with water [40,43]. In these sections, there are over 10 H-bonds, multiple salt bridges, π -stacking and non-polar interactions [36,40,43].

In the N-terminal portion of the $\alpha 1$ helix, a network of H-bonds is present between Gln493, Gly496, Gln498, Thr500 and Asn501 from the RBD and Lys31, His34, Glu35, Tyr41, Gln42, Leu45, Lys353, and Arg357 from hACE2 (Figure 6B) [43]. Additionally, in the C-terminal portion there are H-bonds (Gln474-Gln24, Asn487-Tyr83), Van der Waals forces (Phe486- Met82) and π -stacking interactions (Tyr489-Phe28, Phe486-Tyr83), with more interactions through the middle cavity (Lys417/Tyr453 from RBD and Asp30/His34 from hACE2) (Figure 6C) [36,44].

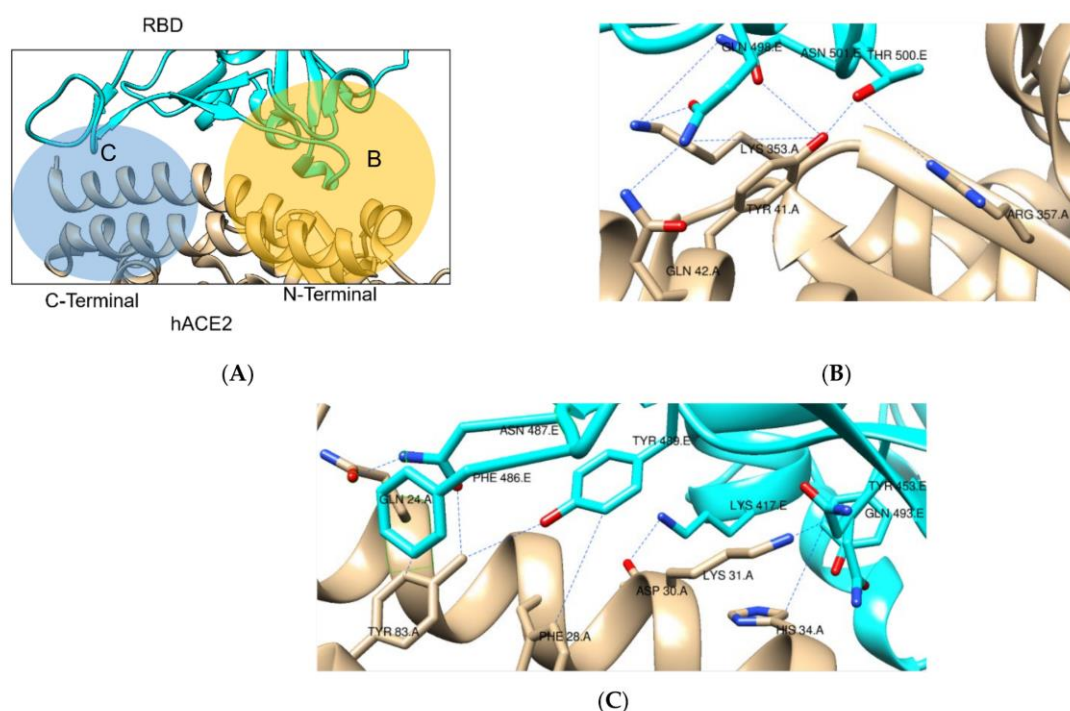


Figure 6. Interface between SARS-CoV-2 S protein RBD (cyan) and hACE2 (tan), in the S protein/ACE2 interaction without mutations structure (PDB 6M0J). (A) Full view of RBD—hACE2 interface (B) N-terminal portion of the $\alpha 1$ helix with a network of H-bonds (C) C-terminal interactions, with π -stacking interactions, h-bonds and Van der Waals forces [29].

Some structures also focus the RBD without mutations (PDB 7KMB/7A91/7A92), nevertheless they are reported in studies describing the behaviour of the entire spike protein, where the interaction region was built from non-mutated RBD focused structures (PDB 6VW1/6M0J/6LZG) [39,45].

Determining the structure of full-length human ACE2 bound to the RBD (PDB 6M17) requires the use of an amino acid transporter (B0AT1) to stabilize the protein. When bound to B0AT1, each hACE2 can bind an RBD, suggesting that native dimeric hACE2 can accommodate two S protein trimers [36]. Moreover, B0AT1 is not involved in dimerization, thus hACE2 may be a homodimer even in the absence of B0AT1 [36]. Critically, the distance between the ACE2 molecules, particularly for residue D615, is only 53 Å, and thus it appears unlikely that the dimeric ACE2 can engage more than one RBD from the same spike protein without substantial conformational changes, although it could bind two RBDs from separate viral particles [46].

2.1.2. Structures with Mutations

The main goal of determining the structure of a focused RBD without mutations is to understand the behaviour of the wild-type strain and to describe the main interactions of the S protein with hACE2 [43]. When introducing mutations or studying different strains, the goal may be to obtain new therapeutic agents based on either hACE2 or S protein, and to understand why new strains behave differently [23,24,41,47].

The four structures that focus the RBD and have mutations can be further categorized according to the aim of finding new anti-SARS-CoV-2 therapeutic agents (PDB 7DMU/7BH9), or to study mutations in new strains (PDB 7NXC/7LO4).

An hACE2 with mutations to increase the affinity to the viral protein was developed as an anti-SARS-CoV-2 therapeutic agent (PDB 7DMU), achieving more than 100-fold increase in blocking activity, when compared to the wild-type hACE2 [41].

The interaction pattern of engineered hACE2 with the RBD is similar to wild-type hACE2, extending a hydrophobic pocket and adding a new H-bond (K35 - Q493, from hACE2 and RBD, respectively), shortening the distance between hACE2 α 1-helix to the RBD by $\sim 1\text{\AA}$ [41]. The use of recombinant hACE2 instead of Abs has an important advantage, given that the resulting mutants with escape mutations have decreased binding affinity to both the engineered and wild-type hACE2, leading to a decrease in virulence [41].

On the other hand, developing a competitive inhibitor from the RBD also holds potential for significant inhibition, particularly as the affinity of the inhibitor with the hACE2 is increased. The RBD inhibitor with the highest binding affinity found (PDB 7BH9) has a RMSD of $0.66\text{\AA}$ when superimposed with the wild type [38]. It was developed from multiple libraries of SARS-CoV-2 clones generated by random mutagenesis and selection of best expressing clones, with several cycles of mutation and testing, attempting to mimic natural evolution in vitro [38].

The mutated RBD (PDB 7BH9) presents, among others, eight amino acid changes actively engaged in the interaction with hACE2, namely V445K, N460K, I468T, T470M, S477N, E484K, Q498R and N501Y. These changes result in additional H-bonds, salt bridges and increased networks of interactions, both stabilizing the structure and resulting in a more positive surface, complementary to the negatively charged hACE2 interaction interface (Figure 7) [38]. The engineered protein can block viral entry and replication, due to an optimized binding, while maintaining no negative effect in the hACE2 enzymatic physiologic activity [38].

Although the optimized structure shows a possible path to design inhibitors, it is noteworthy which mutations are appearing and the sequence in which they appear. Some of the mutations present in the optimized RBD are also present in new variants as critical mutations in variants of concern, appearing often and early in the mutagenesis process, particularly mutations N501Y (variants Alpha, Beta and Gamma) and E484K (variants Beta and Gamma) [48–50]. This may suggest that the in vitro process is possibly mimicking natural evolution, and that artificial selection may have important predictive capabilities over new emerging mutations [41,48,49]. Considering the main variants in circulation, only one mutation providing a clear survival advantage is located outside the RBD (mutation D614G, discussed in more detail ahead) [38,51]. The most common mutations appearing in the later phases, suggest that the next variants with increased infectivity might include mutation Q498R or N460K, as they both significantly increase the affinity and emerged late in the optimization process [38].

Besides predicting potential mutations, the optimized structure may also predict the effect on immune evasion. Analysis of 96 Abs revealed that 28 of them bind outside of the hACE2-binding domain, and the interaction was impaired in 9 Abs with major clashes. Additionally, mutation Q498R, besides increasing the affinity with hACE2, was frequently involved in clashes with Abs [38].

Regarding natural occurring mutations, two structures are available, a strain with mutation G485R and variant Gamma.

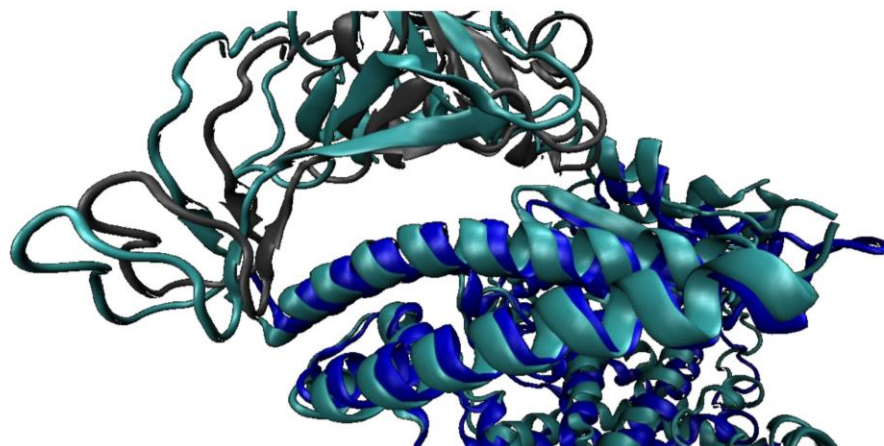


Figure 7. Superimposition and alignment of wild-type RBD and hACE2 (PDB 6LZG) (light blue) and mutated RBD (PDB 7BH9) (hACE2—dark blue ribbon/SARS-CoV-2 S protein—grey ribbon). The closer distance between the chains from the initial to the optimized model reflects the additional interactions established [33].

In the structure with mutation G485R (PDB 7LO4), residue 485, mutating from glycine to arginine, is not directly involved in the hACE2-SARS-CoV-2 S protein interaction, but close to Ab binding epitopes and in a loop region that makes multiple contacts with hACE2. This mutation leads to a rotation in the loop, affecting some interacting residues without significantly reducing the affinity [47]. Nevertheless, the conformational changes in the loop correlate with previous evidence that this mutation plays a role in antigenic escape from neutralising Abs, disrupting interactions involving E484 and F486 (Figure 8) [47].

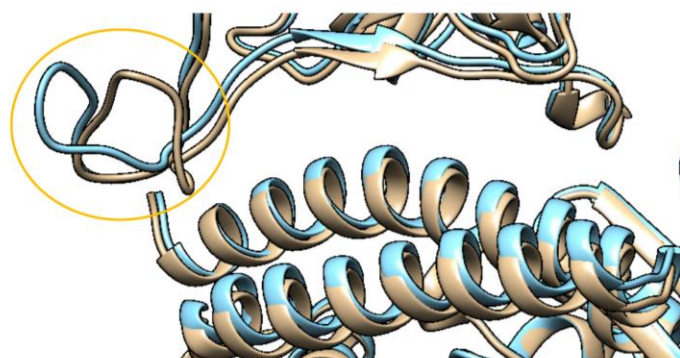


Figure 8. Structural superimposition between the wild-type RBD (Tan) (PDB 6M0J) and with the G485R RBD mutation (Cyan) (PDB 7LO4), with conformational change in the highlighted loop. The RMSD between structures is 0.42 Å [29].

Variant Gamma (PDB 7NXC), has been a cause for concern since it emerged in Brazil in the final months of 2020, showing increased transmissibility and low cross immunity [23,50]. One of the defining characteristics of this variant is the high number of mutations, particularly in the spike protein, with three mutations specific to the RBD (K417T, E484K and N501Y), that lead to a 19-fold increase in affinity compared to the Wuhan strain [50]. Mutation K417T deletes a salt bridge with D30 from hACE2, potentially leading to a decreased affinity. Critically, it hinders Ab binding, contributing to the immune escape observed with this strain [24,52]. The increased affinity is due to both mutation E484K, improving the electrostatic complementarity with the negatively charged hACE2, through the change from a carboxylic acid (E) to an amine group (K), and mutation N501Y creating

an additional ring stacking interaction, with the change from the short-chained asparagine to the aromatic chained tyrosine (Figure 9) [24].

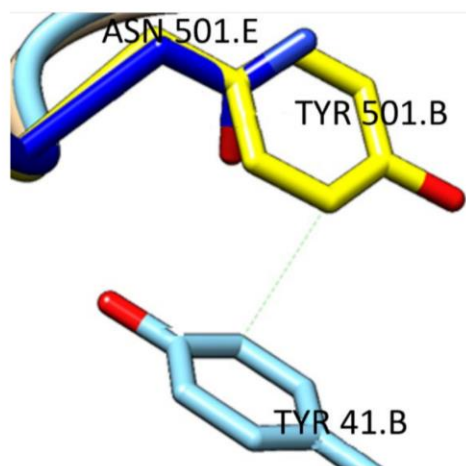


Figure 9. Asparagine-501 from wild-type RBD (PDB 6M0J, carbon atoms in blue) and tyrosine-501 from Gamma variant (PDB 7NXC, carbon atoms in yellow), capable of establishing an additional pi-stacking interaction with tyrosine-41 from hACE2 (bottom) [29].

2.1.3. RBD Only Structures—Final Remarks

The interaction interface and critical contacts between hACE2 and SARS-CoV-2 S protein are well understood. Besides having quality structures available, mutations to the RBD critical amino acids result in predictable and replicable changes in binding affinity [38,40]. When considering the differences between the wild-type RBD (PDB 6VW1/6M0J/6LZG) and the optimized RBD (PDB 7BH9), there is a path for increased binding, infectivity, and potential new variants. Some changes in the optimized RBD required two nucleotide mutations, which takes longer in natural selection [38]. Importantly, N501Y and E484K were present, emerged frequently in earlier stages and have already been found in emerging variants, suggesting some predicting potential for possible new variants [24,38]. The selection of new strains may also be related to the use of low-quality masks, as strains with a higher viral load or a tighter binding spike protein have an advantage over wild-type SARS-CoV-2 [36,38,39,44].

2.2. SARS-CoV-2 Full Spike Protein—hACE2 Interaction

2.2.1. Mutations in the Spike Protein

In various structures of the full spike protein, some mutations appear often, namely mutation of amino acids 986 and 987 to proline and mutations in the S1/S2 cleavage site. The introduction of two consecutive proline residues, at the beginning of the central helix (residues 986 and 987), is a general strategy to retain β -coronavirus S proteins in the prototypical prefusion conformation, which facilitates the study of this conformational state [53]. The same occurs with the cleavage site (RRAR) mutation, as mutating this sequence stabilizes the prefusion structure, since cleavage activates the protein [1,28]. This cleavage site, recognized by host proteases, is fundamental for the virus, allowing the conformational changes leading to the post-fusion structure [54]. In vivo, it increases infectivity, as proteases are ubiquitously present, and importantly, this effect is present not only in SARS-CoV-2 but also in Influenza viruses and human immunodeficiency viruses [55]. Furthermore, this is a distinct characteristic of SARS-CoV-2 and absent in other β -coronaviruses, particularly SARS-CoV [55].

2.2.2. SARS-CoV-2 Full Spike Protein—hACE2

The RBD-hACE2 focused structures highlight the main interactions leading to receptor recognition by SARS-CoV-2, and they are critical for spike protein function. Nevertheless, the spike protein has additional functions in budding and cell fusion, mainly through conformational changes caused by the interactions with hACE2 [1,28,56]. Structure determination for the full spike protein will necessarily have limitations due to its size, and usually other structures are used to build the RBD portion of the structure [39]. However, they give an important insight into the major conformational changes to the entire protein and into critical steps culminating in cell fusion and infection [39,56].

The process through which membrane fusion occurs, particularly the entry cell process, is still undefined. This seems to include the interaction between hACE2 and the spike protein, and after recognition, entry occurs via the endosomal pathway or direct cell surface interaction [45]. In both pathways, protein activation requires a protein, either cathepsin L or TMPRSS2 respectively, as full inhibition of both blocks viral entry [45]. When considering the endosomal pathway, it is important to notice the change in pH and the resulting effects in protein structure, particularly in amino acids containing multiple ionizable groups [45].

The main difference between the structures found at physiological pH (PDB 7KNE/7KNB/7KMS) and at pH 5.5 (PDB 7KNI/7KNH/7KNE) is the distribution of up and down RBDs. At acidic pH, no double or triple up conformations are observed, with small changes observed at pH 4.5 and pH 4.0. Starting at pH 5.5, a region in subunit S2 near the fusion peptide, rich in aspartic acid residues, undergoes a pH-dependent switch, twisting the orientation of disulphide bonds in the structure and causing RBD movement. The domain switch, caused by the pH in endosomal entry, represents an all-RBD-down locking strategy of immune evasion, that SARS-CoV-2, as other CoVs, potentially uses to avoid neutralization by RBD-up-recognizing Abs [45].

2.2.3. Before and after hACE2 Binding

The S protein gradually undergoes conformational changes as hACE2 binds to the each RBD. Particularly in the first hACE2 binding event, a major conformational change is observed, with a rigid body rotation and center of mass of the S protein movement, exposing the remaining unbound RBDs, further stabilized due to the high affinity with hACE2 [45]. The steps leading from a closed and unbound S protein to a full hACE2 binding decrease the contact area between each S1 and the corresponding S2, resulting in a weaker interaction, with the conformational changes driven by hACE2 binding events [39]. These rearrangements and restructuring events in subunits S1 and S2 may have important implications in fusion activation, as the fusion peptide is preceding the S1/S2 contact area [39]. Critically, the cleavage event increases the percentage of RBDs in the up conformation from the closed trimer (PDB 7DX5) to the open trimer (PDB 7DX9), increasing the ability to bind the first hACE2 [56].

Briefly, upon cleavage between S1 and S2 domains, an increase in spike trimers in open conformation is observed (unbound open trimer—PDB 7A93), followed by hACE2 binding (1 RBD bound, 2 RBD down—PDB 7A94) and a major conformational change, exposing further RBDs (1 RBD bound, 2 RBDs up - PDB 7A95/7A96) and hACE2 binding (3 RBD bound—PDB 7A97/7A98). The major conformational change and further binding weakens the S1–S2 interaction preceding the fusion peptide region, priming the S trimer for helical rearrangements in S2, required for viral–host cell membrane fusion [39].

The SARS-CoV-2 original viral S protein population consists mainly of RBD down structures, with S1/S2 uncleaved. The resulting tightly closed conformation of the spike protein is indicative of a conformation masking mechanism for immune evasion of SARS-CoV-2, similar to the changes found with low pH, potentially neutralizing RBD-binding Abs [39]. This feature is also described for other CoVs, such as SARS-CoV and MERS-CoV [42]. The ratio of spike proteins with up or down conformation before and after the first hACE2 binding event has a larger increase in SARS-CoV-2, indicating that the conformational masking may be more efficient [42].

Another important feature of CoVs, particularly regarding immune evasion, is glycosylation. The large number of N-linked glycans covering the surface of the spike protein of SARS-CoV-2, SARS-CoV and MERS-CoV could pose a challenge to antigen recognition, which thus may help the virus evade immune response [42]. However, in the residues close to the S1/S2 cleavage site, the SARS-CoV-2 trimer forms a glycan hole, a region with less glycosylation than the rest of the protein, resulting in an unhindered access by cleavage enzymes [42].

Taking in account all the events occurring from recognition to cell fusion, SARS-CoV-2 is overall very sensitive to the presence of hACE2, with the open state trimer existing only momentarily in the absence of hACE2 and frequently in its presence. This interaction shifts the protein population characteristics, favouring receptor binding, protein activation and the post-fusion state [42].

2.2.4. Spike Protein Mutation D614

From all the mutations in prevalent variants by June 2021 worldwide, mutation D614G is the only one located outside the RBD, and therefore with no direct effect in affinity (PDB 7DF4), but nevertheless, causing increased infectivity and transmission [42,57]. In the closed conformation, portions near the fusion peptide, such as the subdomain 2 and heptad repeat 1, form multiple hydrogen bonds and salt bridges. From the six bonds established in this zone, four have a direct contribution from the side chain of residue D614, suggesting that this amino acid may be essential for this interaction and subsequent stabilization of the prefusion conformation [42,57].

In general, mutation D614G greatly disturbs the interactions between subunits S1 and S2, due to the weakened interactions established by the side chain of glycine, in comparison with aspartic acid [42,57]. As a result, the energy barrier for the conformational change is reduced and SARS-CoV-2 D614G strains have increased sensitivity to the presence of hACE2, which could be the underlying mechanism for enhanced infectivity observed in these strains [42]. On the other hand, given the increase of RBDs in the up position, these strains are more susceptible to neutralization by RBD-directed Abs [58].

2.2.5. Full Spike Protein Neutralization Strategies

Soluble hACE2 has demonstrated a protective role against SARS-CoV2 infection, blocking viral entry in cell cultures and human organoids. Due to this activity, recombinant soluble hACE2 is already enrolled in several clinical trials, as a promising treatment for SARS-CoV-2 infection [59]. The use of ACE2 as a neutralizing agent can be achieved in several ways, either by increasing the affinity or engineering dimers and trimers capable of binding several RBDs. The spike protein can bind three hACE2 molecules and, as such, it is expected that a trimeric ACE2 with a linker can inhibit SARS-CoV-2. S proteins bound to one or more hACE2 have guided the design of ACE2-based viral inhibitors (PDB 7KJ2/7KJ3/7KJ4) [59]. The binding strength of constructs using a monomer, a dimer, and a trimer of hACE2, with engineered linkers, increases as expected. The trimer structure was then optimized, resulting in a trimeric ACE2 with a 200-fold increase in binding affinity, when compared with soluble wild-type hACE2, all without hindering the physiological functions of ACE2 in the angiotensin-renin system [59]. Notably, the neutralization potency of ACE2 constructs gradually increases with the size of the construct, and correlates strictly with the binding affinity, suggesting that the interaction between the RBD and hACE2 is the main component of neutralization [59]. Further improvements of these ACE2-based therapeutic candidates include modifications to enhance protein stability, to modulate peptidase activity and to increase its *in vivo* residence time in circulation [59].

An engineered ACE2-rigid-foldon, as a trimeric protein (PDB 7CT5) inhibits eight naturally occurring mutants, including D614G and seven other RBD domain mutants [46]. The inhibition of SARS-CoV-2 by the trimeric hACE2 occurs due to the simultaneous binding of all three RBD in each spike trimer, which stabilizes the viral protein in a three-

RBD-up conformation and, as this was the only conformation found, it shows the striking ability of these trimers to change the conformation of the spike protein population [46].

2.2.6. Spike Vaccine Candidate

Considering Ab potency for virus neutralizing effects, competition with hACE2 for RBD binding is a key predictor [30]. Hence, the spike protein is the prime candidate for vaccine development, in order to elicit the appearance of Abs directed at the spike protein. The S protein resulting from the coding sequence (mRNA) in the vaccine candidate BNT162b2 (PDB 7L7F), consists of a trimeric S protein capable of binding hACE2 with high affinity and structurally intact receptor binding sites. This is achieved due to a stabilization of the protein in the prefusion conformation, by substitution of residues 986 and 987 [60]. The altered spike protein induces a high functional Ab response both in mice and non-human primates (macaques), protecting them from infection with SARS-CoV-2, demonstrated by an absence of detectable RNA by RT-qPCR in the lower pulmonary tract [60].

2.3. Delta Variant

Spike protein mutations can be selected for either increased infectivity, generally from an increased interaction with hACE2, or decreased immune response, affecting Ab binding. As such, mutations in the spike protein are a fundamental aspect in variants of concern. Notably, one mutation is common to all prevalent strains worldwide, D614G, associated with increased infectivity, although with no direct effect to the RBD [42].

Variant Delta appeared in India in late 2020 and is becoming the dominant strain worldwide, responsible for most infections in the US, Europe, South Africa and Australia [61,62]. Importantly, this variant combines mutations affecting infectivity and potential immune escape (Table 2) [16]. Early data have shown that viral loads are 1,000 times higher with this strain, when compared with the Wuhan strain, which also contributes to increased transmission [63].

Table 2. List of mutations found in the spike protein of the Delta variant.

Mutation	Location	Effect on Affinity	Effect on Immune Response	Additional Information	References
T19R	S1/NTD	No effect reported	Affects NTD-directed Abs	Change in NTD antigenic site	[32,64]
E156-	S1/NTD	No effect reported	Affects NTD-directed Abs		
F157-	S1/NTD	No effect reported	Affects NTD-directed Abs		
R158G	S1/NTD	No effect reported	Affects NTD-directed Abs		
K417N	RBD	Potential decrease in affinity	Moderately decreases RBD-directed Abs	Affects only less important class of Abs	[24,52]
L452R	RBD	No effect reported	Severely affects RBD-directed Abs	None	[52,64]
T478K	RBD	Increases electrostatic complementarity	Moderately decreases RBD-directed Abs	Three additional H-bonds formed	[65–67]
D614G	S1-S2	No effect reported	Increases susceptibility to neutralization by RBD-directed Abs	Changes protein dynamics Increased viral loads and transmission	[52,58]
P681R	S2	No effect reported	No effect reported	Increased S1/S2 cleavage and viral fusion	[66]
D950N	S2	No effect reported	No effect reported	Change protein dynamics	[67]

Abbreviations: NTD—N-terminal domain, RBD—receptor binding domain.

Regarding infectivity, this strain possesses mutation D614G and one mutation in the RBD, T478K. Despite no effect with direct bonds or contacts, it increases the electrostatic complementary between the proteins [65]. As such, it is hypothesised that the Delta variant emerged mainly as a result of reduced immune recognition and enhanced transmissibility nonrelated with affinity (effects from mutations D614G, P681 and D950N) [68]. Importantly, mutation N501Y is not found in the Delta variant, in contrast with all other variants of concern. Mutation N501Y results in the highest increase in hACE2 affinity conferred by a single RBD mutation [52].

Potentially fundamental for immune escape are mutations K417N, and particularly L452R, both affecting Ab binding, and occurring in the RBD [64]. L452R has also appeared in variant Epsilon identified in California, reaching a maximum of 20% of new cases in the United States, before variant Delta became prevalent [61,69]. Although no structure on the Delta variant RBD is available, a homology model of the Delta variant RBD was built, using SWISS-MODEL, and the effect of specific and new mutations can be predicted [37]. Homology modelling is a method that predicts a 3D structure based on the aminoacidic sequence, with the only requirement of an existing experimental structure with at least 30% sequence identity [70]. The genome selected to represent the Delta strain was MZ377116, retrieved from the SARS-CoV-2 Data Hub (NCBI Virus) [71].

Mutation L452R causes a steric hindrance to Ab binding, retrieved from structure 7CM4, without affecting binding to hACE2 (Figure 10) but causing a greater than 10-fold decrease in neutralization potency for RBD-specific Abs [61,69]. Crucially, variant Epsilon has additional mutations, S13I and W152C, resulting in severe decrease of neutralization for NTD-directed Abs, due to a new disulphide bond inducing a conformational change [64]. In the Delta variant, mutations T19R, E156-, F157- and R158G severely affect NTD directed Abs, due to the adoption of an alpha-helical conformation, instead of the original β -strand [64,68].

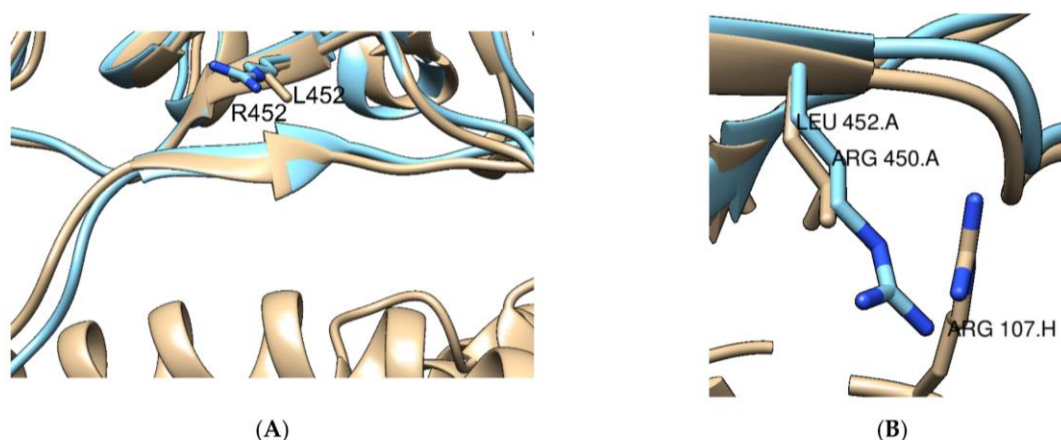


Figure 10. Mutation L452R, with the original variant from wild-type RBD (PDB 6M0J) (cyan) and the homology model of Delta variant (tan), obtained from genome MZ377116.1 (A) Superimposition of RBDs from original variant and Delta variant. Steric hindrance observed with Arg107 from Ab from structure 7CM4, due to a longer side chain in Delta variant (B) RBDs of original and delta variant structures with no steric hindrance regarding hACE2 [29].

Variant Epsilon has been deescalated from variant of concern due the decrease in community transmission, being replaced by the Delta variant [14,16].

3. Conclusions

The structures available in the Protein Data Bank provide important insights into SARS-CoV-2 spike protein behaviour, clarifying the conformational changes across the protein leading to post-fusion conformation and cell infection. In general, the spike protein is highly sensitive to the presence and binding of hACE2 to the RBDs, with an allosteric

effect, destabilizing the S1/S2 interaction and facilitating the final conformational change. The affinity of the complex formed in cell recognition is critical for infectivity, with consequences in the epidemiology of SARS-CoV-2. As such, the RBD is also an important target for neutralizing Abs.

The appearance of new variants has increased SARS-CoV-2 fitness, mainly by increasing infectivity, or decreasing immune response. The Delta variant carries particular mutations associated with increased viral fitness through decreased antibody binding, increased RBD affinity and altered protein dynamics, which justifies the increased prevalence worldwide [67,72]. Despite this, some of the most common mutations increasing affinity in other variants are absent, which still show potential for evolution [52,61].

The ability to predict infectivity and immune escape for new variants is crucial to decrease their impact. As such, mutagenesis studies have also shown an important predicting capacity for the major mutations in current variants of concern [38]. Combining both existing mutations and mutagenesis studies, new potential SARS-CoV-2 variants, harboring advantageous S protein mutations, may be predicted. These include mutations S13I and W152C, decreasing antibody binding, N460K, increasing RBD affinity, or Q498R, positively affecting both properties

Author Contributions: L.Q.-R.: conceptualization, data curation and investigation, formal analysis, methodology, and writing original draft preparation; P.G.d.S. and J.G.: conceptualization, review of the writing and substance of this article, supervision and validation M.B., A.B. and J.R.M.: conceptualization, data curation and investigation, supervision, and validation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: L.Q.-R. would like to acknowledge Fundação para a Ciência e para a Tecnologia for the grant “2020.10230.BD” under the program “DOCTORATES 4 COVID-19”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, M.Y.; Zhao, R.; Gao, L.J.; Gao, X.F.; Wang, D.P.; Cao, J.M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 587269. [CrossRef] [PubMed]
2. WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19—11 March 2020. Available online: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed on 3 September 2021).
3. WHO. Coronavirus Disease (COVID-19) Dashboard. Available online: <https://covid19.who.int/> (accessed on 4 September 2021).
4. Zhao, J.; Cui, W.; Tian, B.-p. The Potential Intermediate Hosts for SARS-CoV-2. *Front. Microbiol.* **2020**, *11*, 2400. [CrossRef] [PubMed]
5. Woo, P.C.; Lau, S.K.; Huang, Y.; Yuen, K.Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med. (Maywood)* **2009**, *234*, 1117–1127. [CrossRef] [PubMed]
6. Schoeman, D.; Gordon, B.; Fielding, B.C. Pathogenic Human Coronaviruses. *Ref. Modul. Biomed. Sci.* **2021**, B978-970-912-818731-818739.800052-818735. [CrossRef]
7. Huang, Y.; Yang, C.; Xu, X.-f.; Xu, W.; Liu, S.-w. Structural and functional properties of SARS-CoV-2 spike protein: Potential antiviral drug development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41*, 1141–1149. [CrossRef]
8. Cubuk, J.; Alston, J.J.; Incicco, J.J.; Singh, S.; Stuchell-Brereton, M.D.; Ward, M.D.; Zimmerman, M.I.; Vithani, N.; Griffith, D.; Wagoner, J.A.; et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **2021**, *12*, 1936. [CrossRef]
9. Fan, Y.; Zhao, K.; Shi, Z.-L.; Zhou, P. Bat Coronaviruses in China. *Viruses* **2019**, *11*, 210. [CrossRef] [PubMed]
10. WHO. Coronavirus disease (COVID-19): How is it transmitted? Available online: <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted> (accessed on 18 August 2021).
11. Hu, B.; Guo, H.; Zhou, P.; Shi, Z.L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021**, *19*, 141–154. [CrossRef]
12. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef]
13. Janik, E.; Niemcewicz, M.; Podogrocki, M.; Majsterek, I.; Bijak, M. The Emerging Concern and Interest SARS-CoV-2 Variants. *Pathogens* **2021**, *10*, 633. [CrossRef]

14. ECDC. SARS-CoV-2 Variants of Concern as of 9 September 2021. Available online: <https://www.ecdc.europa.eu/en/covid-19/variants-concern> (accessed on 15 September 2021).
15. WHO. Tracking SARS-CoV-2 Variants. Available online: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (accessed on 14 September 2021).
16. CDC. SARS-CoV-2 Variant Classifications and Definitions. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html#Concern> (accessed on 14 September 2021).
17. CDC. Centers for Disease Control and Prevention. Available online: <https://www.cdc.gov/> (accessed on 29 July 2021).
18. ECDC. European Centre for Disease Prevention and Control. Available online: <https://www.ecdc.europa.eu/en> (accessed on 29 July 2021).
19. Madhi, S.A.; Baillie, V.; Cutland, C.L.; Voysey, M.; Koen, A.L.; Fairlie, L.; Padayachee, S.D.; Dheda, K.; Barnabas, S.L.; Bhorat, Q.E.; et al. Efficacy of the ChAdOx1 nCoV-19 Vaccine against the B.1.351 Variant. *New Engl. J. Med.* **2021**, *384*, 1885–1898. [\[CrossRef\]](#)
20. Cele, S.; Gazy, I.; Jackson, L.; Hwa, S.-H.; Tegally, H.; Lustig, G.; Giandhari, J.; Pillay, S.; Wilkinson, E.; Naidoo, Y.; et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **2021**, *593*, 142–146. [\[CrossRef\]](#)
21. Pearson, C.A.; Russell, T.W.; Davies, N.G. Estimates of Severity and Transmissibility of Novel South Africa SARS-CoV-2 Variant 501Y.V2. Available online: https://cmmid.github.io/topics/covid19/reports/sa-novel-variant/2021_01_11_Transmissibility_and_severity_of_501Y_V2_in_SA.pdf (accessed on 14 September 2021).
22. Funk, T.; Pharris, A.; Spiteri, G.; Bundle, N.; Melidou, A.; Carr, M.; Gonzalez, G.; Garcia-Leon, A.; Crispie, F.; O'Connor, L.; et al. Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: Data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance* **2021**, *26*, 2100348. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.d.S.; Mishra, S.; Crispim, M.A.E.; Sales, F.C.S.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [\[CrossRef\]](#)
24. Dejnirattisai, W.; Zhou, D.; Supasa, P.; Liu, C.; Mentzer, A.J.; Ginn, H.M.; Zhao, Y.; Duyvesteyn, H.M.E.; Tuekprakhon, A.; Nutalai, R.; et al. Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **2021**, *184*, 2939–2954.e9. [\[CrossRef\]](#)
25. Eli-Lilly. FACT SHEET FOR HEALTH CARE PROVIDERS EMERGENCY USE AUTHORIZATION (EUA) OF BAMLANIVIMAB AND ETESEVIMAB. Available online: <https://www.fda.gov/media/145802/download> (accessed on 15 September 2021).
26. Uriu, K.; Kimura, I.; Shirakawa, K.; Takaori-Kondo, A.; Nakada, T.-a.; Kaneda, A.; Consortium, T.G.t.P.J.; Nakagawa, S.; Sato, K. Ineffective neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine sera. *bioRxiv* **2021**, 2021.2009.2006.459005. [\[CrossRef\]](#)
27. Ni, W.; Yang, X.; Yang, D.; Bao, J.; Li, R.; Xiao, Y.; Hou, C.; Wang, H.; Liu, J.; Yang, D.; et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Crit. Care* **2020**, *24*, 422. [\[CrossRef\]](#)
28. Wrapp, D.; Wang, N.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.-L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Petersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Ju, B.; Zhang, Q.; Ge, J.; Wang, R.; Sun, J.; Ge, X.; Yu, J.; Shan, S.; Zhou, B.; Song, S.; et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **2020**, *584*, 115–119. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **2020**, *30*, 343–355. [\[CrossRef\]](#) [\[PubMed\]](#)
32. McCallum, M.; Marco, A.; Lempp, F.; Tortorici, M.A.; Pinto, D.; Walls, A.C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F.; et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *bioRxiv* **2021**. [\[CrossRef\]](#)
33. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [\[CrossRef\]](#)
34. Huo, J.; Zhao, Y.; Ren, J.; Zhou, D.; Duyvesteyn, H.M.E.; Ginn, H.M.; Carrique, L.; Malinauskas, T.; Ruza, R.R.; Shah, P.N.M.; et al. Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host Microbe* **2020**, *28*, 445–454.e6. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Mahase, E. Delta variant: What is happening with transmission, hospital admissions, and restrictions? *BMJ* **2021**, *373*, n1513. [\[CrossRef\]](#)
36. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444–1448. [\[CrossRef\]](#)
37. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [\[CrossRef\]](#)
38. Zahradník, J.; Marciano, S.; Shemesh, M.; Zoler, E.; Chiaravalli, J.; Meyer, B.; Dym, O.; Elad, N.; Schreiber, G. SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv* **2021**, 2021.2001.2006.425392. [\[CrossRef\]](#)
39. Benton, D.J.; Wrobel, A.G.; Xu, P.; Roustian, C.; Martin, S.R.; Rosenthal, P.B.; Skehel, J.J.; Gamblin, S.J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588*, 327–330. [\[CrossRef\]](#)

40. Delgado Blanco, J.; Hernandez-Alias, X.; Cianferoni, D.; Serrano, L. In silico mutagenesis of human ACE2 with S protein and translational efficiency explain SARS-CoV-2 infectivity in different species. *PLoS Comput. Biol.* **2020**, *16*, e1008450. [CrossRef]
41. Higuchi, Y.; Suzuki, T.; Arimori, T.; Ikemura, N.; Mihara, E.; Kirita, Y.; Ohgitani, E.; Mazda, O.; Motooka, D.; Nakamura, S.; et al. High affinity modified ACE2 receptors protect from SARS-CoV-2 infection in hamsters. *bioRxiv* **2020**. [CrossRef]
42. Xu, C.; Wang, Y.; Liu, C.; Zhang, C.; Han, W.; Hong, X.; Wang, Y.; Hong, Q.; Wang, S.; Zhao, Q.; et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci. Adv.* **2021**, *7*. [CrossRef]
43. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581*, 215–220. [CrossRef]
44. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [CrossRef] [PubMed]
45. Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G.-Y.; Katsamba, P.S.; Sampson, J.M.; Schön, A.; Bimela, J.; et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **2020**, *28*, 867–879.e5. [CrossRef] [PubMed]
46. Guo, L.; Bi, W.; Wang, X.; Xu, W.; Yan, R.; Zhang, Y.; Zhao, K.; Li, Y.; Zhang, M.; Cai, X.; et al. Engineered trimeric ACE2 binds viral spike protein and locks it in “Three-up” conformation to potentially inhibit SARS-CoV-2 infection. *Cell Res.* **2021**, *31*, 98–100. [CrossRef] [PubMed]
47. Weekley, C.M.; Purcell, D.F.J.; Parker, M.W. SARS-CoV-2 Spike receptor-binding domain with a G485R mutation in complex with human ACE2. *bioRxiv* **2021**, 2021.2003.2016.434488. [CrossRef]
48. Hodcroft, E. Variant: 20I (Alpha, V1). Available online: <https://covariants.org/variants/20I.Alpha.V1> (accessed on 7 July 2021).
49. Hodcroft, E. Variant: 20H (Beta, V2). Available online: <https://covariants.org/variants/20H.Beta.V2> (accessed on 7 July 2021).
50. Hodcroft, E. Variant: 20J (Gamma, V3). Available online: <https://covariants.org/variants/20J.Gamma.V3> (accessed on 7 July 2021).
51. Juraszek, J.; Rutten, L.; Blokland, S.; Bouchier, P.; Voorzaat, R.; Ritschel, T.; Bakkers, M.J.G.; Renault, L.L.R.; Langedijk, J.P.M. Stabilizing the closed SARS-CoV-2 spike trimer. *Nat. Commun.* **2021**, *12*, 244. [CrossRef] [PubMed]
52. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S.J.; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [CrossRef] [PubMed]
53. Pallesen, J.; Wang, N.; Corbett, K.S.; Wrapp, D.; Kirchdoerfer, R.N.; Turner, H.L.; Cottrell, C.A.; Becker, M.M.; Wang, L.; Shi, W.; et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7348–E7357. [CrossRef]
54. Xia, X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. *Viruses* **2021**, *13*, 109. [CrossRef] [PubMed]
55. Johnson, B.A.; Xie, X.; Bailey, A.L.; Kalveram, B.; Lokugamage, K.G.; Muruato, A.; Zou, J.; Zhang, X.; Juelich, T.; Smith, J.K.; et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **2021**, *591*, 293–299. [CrossRef] [PubMed]
56. Yan, R.; Zhang, Y.; Li, Y.; Ye, F.; Guo, Y.; Xia, L.; Zhong, X.; Chi, X.; Zhou, Q. Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2. *Cell Res.* **2021**, *31*, 717–719. [CrossRef] [PubMed]
57. WHO. SARS-CoV-2 Variants. Available online: <https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/> (accessed on 3 September 2021).
58. Weissman, D.; Alameh, M.G.; de Silva, T.; Collini, P.; Hornsby, H.; Brown, R.; LaBranche, C.C.; Edwards, R.J.; Sutherland, L.; Santra, S.; et al. D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe* **2021**, *29*, 23–31.e24. [CrossRef] [PubMed]
59. Xiao, T.; Lu, J.; Zhang, J.; Johnson, R.I.; McKay, L.G.A.; Storm, N.; Lavine, C.L.; Peng, H.; Cai, Y.; Rits-Volloch, S.; et al. A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent in vitro. *bioRxiv* **2020**. [CrossRef]
60. Vogel, A.B.; Kanevsky, I.; Che, Y.; Swanson, K.A.; Muik, A.; Vormehr, M.; Kranz, L.M.; Walzer, K.C.; Hein, S.; Güler, A.; et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* **2021**, *592*, 283–289. [CrossRef]
61. Hodcroft, E. Variant: 21A (Delta). Available online: <https://covariants.org/variants/21A.Delta> (accessed on 7 July 2021).
62. ECDC. SARS-COV-2 Delta variant now dominant in much of the European Region and efforts must be reinforced to prevent transmission, warn WHO/Europe and ECDC. Available online: <https://www.ecdc.europa.eu/en/news-events/sars-cov-2-delta-variant-now-dominant-european-region> (accessed on 18 August 2021).
63. Li, B.; Deng, A.; Li, K.; Hu, Y.; Li, Z.; Xiong, Q.; Liu, Z.; Guo, Q.; Zou, L.; Zhang, H.; et al. Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant. *medRxiv* **2021**, 2021.2007.2007.21260122. [CrossRef]
64. McCallum, M.; Bassi, J.; De Marco, A.; Chen, A.; Walls, A.C.; Di Iulio, J.; Tortorici, M.A.; Navarro, M.-J.; Silacci-Fregni, C.; Saliba, C.; et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* **2021**, eabi7994. [CrossRef]
65. Di Giacomo, S.; Mercatelli, D.; Rakhimov, A.; Giorgi, F.M. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. *J. Med. Virol.* **2021**, *93*, 5638–5643. [CrossRef]
66. Cherian, S.; Potdar, V.; Jadhav, S.; Yadav, P.; Gupta, N.; Das, M.; Rakshit, P.; Singh, S.; Abraham, P.; Panda, S.; et al. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms* **2021**, *9*. [CrossRef]

67. Planas, D.; Veyer, D.; Baidaliuk, A.; Staropoli, I.; Guivel-Benhassine, F.; Rajah, M.M.; Planchais, C.; Porrot, F.; Robillard, N.; Puech, J.; et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **2021**, *596*, 276–280. [[CrossRef](#)] [[PubMed](#)]
68. McCallum, M.; Walls, A.C.; Sprouse, K.R.; Bowen, J.E.; Rosen, L.; Dang, H.V.; deMarco, A.; Franko, N.; Tilles, S.W.; Logue, J.; et al. Molecular basis of immune evasion by the delta and kappa SARS-CoV-2 variants. *bioRxiv* **2021**, 2021.2008.2011.455956. [[CrossRef](#)]
69. Hodcroft, E. Variant: 21C (Epsilon). Available online: <https://covariants.org/variants/21C.Epsilon> (accessed on 8 July 2021).
70. Sensoy, O.; Almeida, J.G.; Shabbir, J.; Moreira, I.S.; Morra, G. Chapter 16-Computational studies of G protein-coupled receptor complexes: Structure and dynamics. In *Methods in Cell Biology*; Shukla, A.K., Ed.; Academic Press: Cambridge, MA, USA, 2017; Volume 142, pp. 205–245.
71. NCBI-Virus. SARS-CoV-2 Data Hub. Available online: [https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202%20\(SARS-CoV-2\),%20taxid:2697049&SeqType_s=Nucleotide&Lineage_s=B.1.1.7](https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202%20(SARS-CoV-2),%20taxid:2697049&SeqType_s=Nucleotide&Lineage_s=B.1.1.7) (accessed on 31 July 2021).
72. Petra, M.; Steven, K.; Mahesh Shanker, D.; Guido, P.; Bo, M.; Swapnil, M.; Charlie, W.; Thomas, M.; Isabella, F.; Rawlings, D.; et al. SARS-CoV-2 B.1.617.2 Delta variant emergence and vaccine breakthrough. *Nat. Portf.* **2021**. [[CrossRef](#)]

3.2 Hit molecule identification and optimization

3.2.1 Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein - Confirmed and Potential Ligands

Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein – Confirmed and Potential Ligands

Luís Queirós-Reis,* João R. Mesquita, Andrea Brancale, and Marcella Bassetto


 Cite This: *J. Chem. Inf. Model.* 2023, 63, 7282–7298


Read Online

ACCESS |

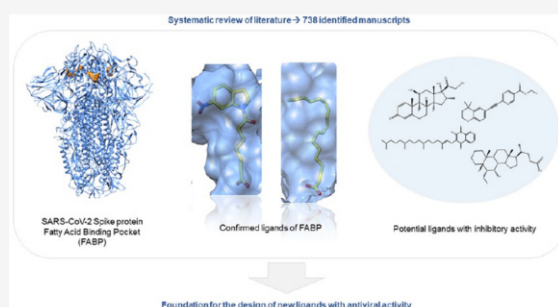
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Severe Acute Respiratory syndrome 2 (SARS-CoV-2) is a respiratory virus responsible for coronavirus disease 19 (COVID-19) and the still ongoing and unprecedented global pandemic. The key viral protein for cell infection is the spike glycoprotein, a surface-exposed fusion protein that both recognizes and mediates entry into host cells. Within the spike glycoprotein, a fatty acid binding pocket (FABP) was confirmed, with the crystallization of linoleic acid (LA) occupying a well-defined site. Importantly, when the pocket is occupied by a fatty acid, an inactive conformation is stabilized, and cell recognition is hindered. In this review, we discuss ligands reported so far for this site, correlating their activity predicted through *in silico* studies with antispike experimental activity, assessed by either binding assays or cell-infection assays. LA was the first confirmed ligand, cocrystallized in a cryo-EM structure of the spike protein, resulting in increased stability of the inactive conformation of the spike protein. The next identified ligand, lifitegrast, was also experimentally confirmed as a ligand with antiviral activity, suggesting the potential for diverse chemical scaffolds to bind this site. Finally, SPC-14 was also confirmed as a ligand, although no inhibition assays were performed. In this review, we identified 20 studies describing small-molecule compounds predicted to bind the pocket in *in silico* studies and with confirmed binding or *in vitro* activity, either inhibitory activity against the spike-ACE2 interaction or antiviral activity in cell-based assays. When considering all ligands confirmed with *in vitro* assays, a good overall occupation of the pocket should be complemented with the ability to make direct interactions, both hydrophilic and hydrophobic, with key amino acid residues defining the pocket surface. Among the active compounds, long flexible carbon chains are recurrent, with retinoids capable of binding the FABP, although bulkier systems are also capable of affecting viral fitness. Compounds able to bind this site with high affinity have the potential to stabilize the inactive conformation of the SARS-CoV-2 spike protein and therefore reduce the virus's ability to infect new cells. Since this pocket is conserved in highly pathogenic human coronaviruses, including MERS-CoV and SARS-CoV, this effect could be exploited for the development of new antiviral agents, with broad-spectrum anticoronavirus activity.

KEYWORDS: SARS-CoV-2, spike protein, fatty acid binding pocket (FABP), S glycoprotein, linoleic acid (LA), lifitegrast, SPC-14, ligand, virtual screening



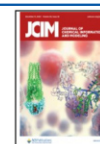
1. INTRODUCTION

1.1. SARS-CoV-2. Coronavirus disease 19, or COVID-19, is a respiratory infection caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), responsible for a global pandemic with more than 750 million people infected and close to 7 million deaths up to September 2023.^{1–4} Coronaviruses (CoVs) are a group of RNA viruses belonging to the subfamily *Orthocoronaviridae*, that cause disease in a variety of domestic and wild animals.⁵ The name “corona” is derived from a halo observed under microscopy, formed by three major structural proteins: spike, membrane, and envelope, projecting as spikes from the viral envelope (Figure 1).⁶

Cell infection for SARS-CoV-2 is initiated by the densely glycosylated spike (S) protein (Figure 2A), a trimeric fusion

protein that binds to human angiotensin converting enzyme 2 (ACE2), acting as the functional receptor.^{1,7} As the S protein mediates viral entry into host cells, it is the main target for antibodies and one of the primary focuses of therapeutics and vaccine design.^{8,9} All key areas for cell and antibody recognition are located in the S protein, with the receptor binding domain (RBD) defined as the portion that interacts with ACE2 (Figure 2B).^{10–12} Since the infection is initiated by

Received: May 30, 2023
Revised: October 27, 2023
Accepted: October 30, 2023
Published: November 22, 2023



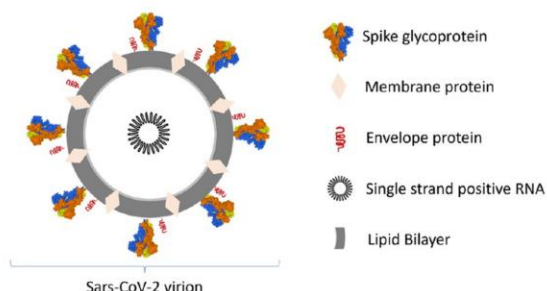


Figure 1. Schematic representation of the SARS-CoV-2 virion structure, with the single-stranded, positive-sense RNA and the three structural proteins: spike glycoprotein, membrane protein, and envelope protein.

receptor recognition between the RBD and ACE2, this interaction determines infectivity, host range, and pathogenesis of the virus.⁷

1.2. The Spike Glycoprotein as a Drug Target. Given the critical role of the S protein, it is the main target for vaccine development, antibody therapy, and small-molecule entry inhibitors.^{7,11} Indeed, vaccine development focused on the S protein and achieved a reduction in viral effectiveness.¹⁴ On the other hand, no small-molecule inhibitors of the S protein have been approved. When targeting the direct interaction between the RBD and ACE2, there is an intrinsic challenge, as the receptor binding motifs are highly variable at the sequence level.¹¹ Nevertheless, since it is effective in reducing viral fitness, research focusing on blocking the interaction has resulted in numerous virtual screening campaigns targeting the S protein since the pandemic started.^{15–19} Given the S protein

size (~1300 residues), variety of functions (from cell recognition to cell fusion), and conformational changes, the area selected for such analyses can vary significantly, with studies focusing only on the RBD, larger portions of the S protein, and searches for additional potential binding cavities.^{18,20–22}

1.3. Fatty Acid Binding Pocket. One of the hallmarks of viral infections is the imbalance they cause in normal cell functions. Specifically for coronavirus, lipid metabolism is deeply impacted, and the presence of fatty acids, such as linoleic acid (LA) and arachidonic acid, can significantly suppress viral replication.²³ With this in mind, the S protein was studied in the presence of LA, and a cryo-EM structure of the S protein with LA bound was determined (PDB 6ZBS), with the binding area identified as a fatty acid binding pocket (FABP). LA occupies a site formed between two adjacent RBDs, and considering that the S protein is a trimer, three similar pockets are defined, with up to three molecules of LA capable of binding simultaneously (Figure 3A).^{24,25}

Critically, despite the proximity to the RBD surface that interacts with ACE2, none of the FABP residues participate in the recognition event with the human receptor.^{24,25} Binding to this pocket appears to be enabled by a specific spatial occupation pattern, and LA, a molecule with a long carbon chain and two cis-unsaturated bonds, appears to have an ideal molecular shape to achieve an optimal fit of the curved cavity of the pocket (Figure 3B).²⁴

The presence of LA in the FABP affects conformational changes in the protein elicited by the cell recognition event. Each RBD has two possible conformations: down conformation (Figure 4A), inaccessible for interaction with ACE2, or up conformation, available for target recognition (Figure 4B).^{10,26} Additionally, changing between the up and down conformation

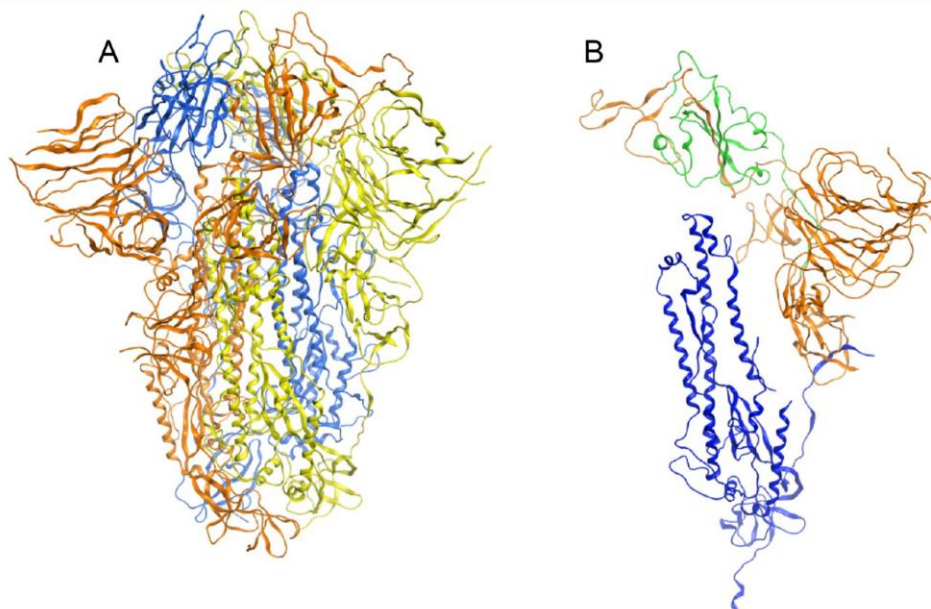


Figure 2. A) Ribbon representation of the S glycoprotein trimer, with the monomers represented as blue, orange, and yellow ribbons, respectively (PDB ID: 6VXX). B) Ribbon representation of an S glycoprotein monomer with the RBD highlighted in green. The RBD is part of the S1 subunit (orange), while the S2 subunit is represented in blue.¹³

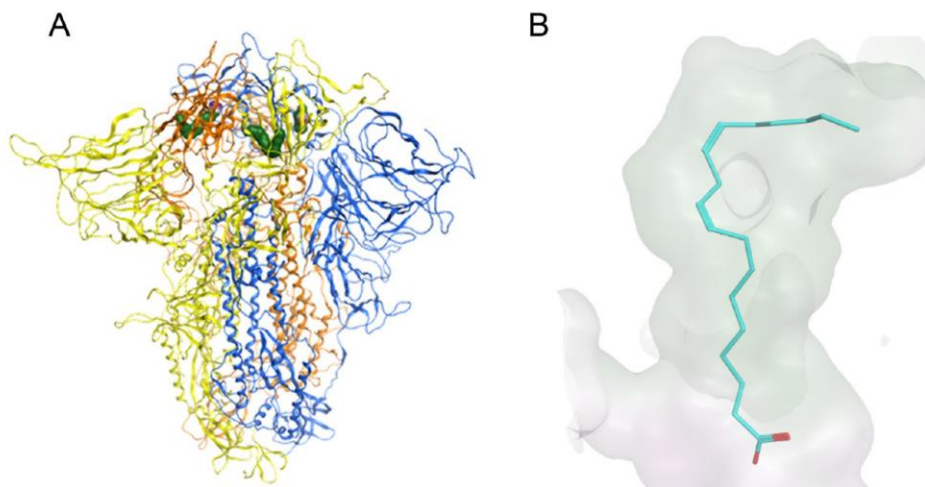


Figure 3. A) Ribbon representation of the S protein from a side view, with three FABPs defined near the RBD (PDB ID: 6ZBS). The S protein monomers are represented as yellow, orange, and blue ribbons, while three bound LA molecules are highlighted as the molecular surface (green: lipophilic; pink: hydrophilic; white: neutral). B) Zoom on one LA molecule (carbon atoms in cyan) cocrystallized in the FABP represented by the molecular surface, the hydrophilic (purple) portion near the carboxylate group, and the carbon chain buried in the hydrophobic portion (green) (PDB ID: 6ZBS).¹³

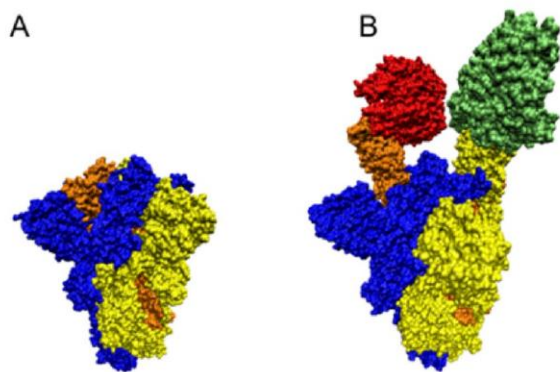


Figure 4. A) Protein surface representation of the S trimer in a three RBD down conformation (blue, orange, yellow) (PDB ID: 7DF3). B) Protein surface representation of the S trimer in a two RBD up (yellow and orange) bound to ACE2 (red and green) and one RBD down (blue) conformation (PDB ID: 7A93).¹³

is independent in each S monomer, with the down conformation being the most stable and frequent state.⁷ In the absence of LA, three-dimensional classification of S particle conformations in cryo-EM data reveals that ~70% of trimers have RBDs in the up conformation with the S protein primed for infection. On the other hand, when LA is present, ~70% of S trimers are in an all-down conformation, incapable of binding with the ACE2 receptor.^{24,27,28} The presence of LA in the FABP results in a rearrangement of the binding motifs, becoming ordered and buried between RBD interfaces, decreasing the ability of the RBD to interact with ACE2 (Figure 4A).²⁴ The increased rigidity and stabilization of the S protein in an inactive conformation changes the ratio between up- and down-conformation RBDs.^{24,27}

1.3.1. Fatty Acid Binding Pocket - Structure. The FABP is shaped in a long, curved tube fashion, with extensive

interaction surfaces formed by several phenylalanine residues (Phe338, Phe342, Phe374, Phe377, Phe392, Phe515), as well as other mainly hydrophobic amino acids (Cys336, Pro337, Ile358, Ala363, Tyr365, Leu368, Tyr369, Leu387, Val395, Ile434, Leu513, Val524).^{24,25} At the entrance of the pocket, the adjacent RBD provides several hydrophilic residues that can establish both electrostatic and hydrogen-bond interactions (Arg408, Gln409, Gly416, Lys417).^{24,25} Potential ligands can interact with residues in both portions, and the presence of the adjacent RBD is required to stabilize the ligand by forming direct interactions with the hydrophilic groups.^{24,29} Regarding LA, the main stabilizing interactions are a hydrogen bond between Lys417 and the carboxylic acid from LA and additional contacts across the pocket with Phe338, Tyr365, and Cys432 (Figure 5).²⁴

1.4. Occupation of the Fatty Acid Binding Pocket and Viral Infection. LA, as other similar ω -6 fatty acids, can affect virus binding to ACE2 and cell entry, and its increased intake has been linked to decreased fatality rates in epidemiological studies.^{30–32} A direct inhibitory effect in the RBD-ACE2 interaction has been detected, with a 18% reduction in binding assays at 0.29 mM, while linolenic acid displays a reduction of up to 60% infection in cell-based assays at 2.28 mM, which confirms the potential to modulate the S activity from this pocket.³⁰ The presence of fatty acids in the FABP, particularly LA, affects conformational rearrangements, infectivity, and immunogenicity.²⁹ In addition to the direct effects in recognition from the stabilization of the RBDs, there is also the potential for distal effects in the S protein, modulating the behavior of key residues involved in major conformational rearrangements preceding cell fusion.³³ Computational studies exploring the pH-dependence of S protein conformations, based upon pK_a calculations and molecular dynamics simulations at constant pH, have suggested that LA binding affects the pH-dependence of the RBD active/inactive conformation equilibrium.³⁴ The FABP is located in the vicinity of the base pair Asp398-Arg355, and the decreased flexibility upon fatty acid binding changes the solvent

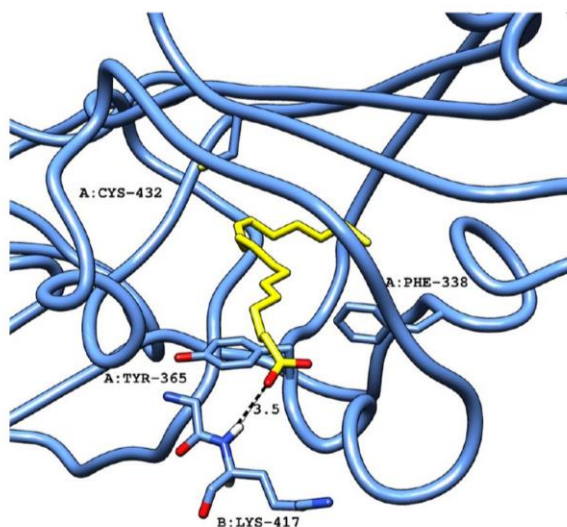


Figure 5. A) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) in the FABP in a ribbon representation (blue), located between two adjacent RBDs. The pocket is formed by RBD-A (with hydrophobic residues) and hydrophilic residues in the adjacent RBD-B, interacting with the carboxylate head. Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein, with distance in angstrom (Å).⁵²

accessibility of Arg355, since the regions adjacent to both amino acids are more closely packed, with a similar effect also observed for Asp614 (Figure 6).³⁴ The observed effect in Asp614 is particularly relevant, taking into account that a mutation in this residue is now prevalent in all circulating SARS-CoV-2 variants (Asp614Gly), increasing infectivity by reducing interactions within the S protein, and shifting the protein conformation toward active RBDs.^{28,34,35} LA has the opposite effect on the conformational states of the S protein, affecting the pH-dependence and shifting the protein conformation toward inactive RBDs.^{10,36}

In addition to a strong effect in the S trimer behavior and target recognition abilities, exploring the potential occupation of the FABP with small-molecule ligands presents an additional advantage, given that key residues in the pocket are highly conserved, not only among emerging SARS-CoV-2 variants of

concern but also in intermediate species, such as coronaviruses currently infecting bats and pangolins.^{25,29,37} While the pH-dependence of the S protein is highly relevant for the virus life cycle, particularly during the endosomal pathway for cell entry, the function of the FABP is not clear, although a modulation effect in the exposure of immunodominant sites has been suggested.^{29,36} During endosomal entry, as the environmental pH is gradually reduced, the open/closed RBD equilibrium is shifted to the inactive conformation, reducing not only the interaction with ACE2 but also with neutralizing antibodies, that commonly bind the RBD.³⁶ As a result, neutralizing antibodies are shed gradually, but partial interaction with ACE2 is maintained, along with the ability for cell fusion and infection.^{36,38} This has been described as a masking behavior, where all RBDs are in the inactive conformation.³⁶ Similarly, the presence of LA also stabilizes the inactive conformation, in a masking behavior, restricting viral replication.²⁹ As happens for a decreasing pH, the FABP appears to be essential for effective infection, allowing the virus to change from a high-infective status before a host immune response is established, with subsequent inflammation producing ω -3 and ω -6 molecules that activate the FABP, reducing viral recognition and clearance, while increasing viral titers.²⁹ Hence, the FABP modulates conformational changes that affect the immunogenicity of the S protein.²⁹ Taking into account the confirmed binding effect of ω -6 fatty acids, which decrease target recognition and viral replication,³⁰ the FABP is a promising target for the development of small-molecule binders, that could lock the S trimer and respective RBDs in an inactive conformation, inhibiting receptor interactions and cell infection.^{1,8,24} This is particularly important, as this binding pocket is not exclusive to SARS-CoV-2: it is found in other highly infectious human coronaviruses, being fully conserved in SARS-CoV and partially conserved in MERS-CoV. This pocket could therefore be exploited for the development of new antiviral agents with broad-spectrum anticoronavirus activity.²⁵

The main objective of this article is to compile currently confirmed and potential ligands, resulting from virtual screening studies that explore the FABP, either focusing on the pocket specifically or broader virtual screenings against the whole S glycoprotein. Given the potential to significantly affect the virus life cycle by binding to the FABP, a comprehensive understanding of which ligands are known to bind to this site, and their inhibitory activity of the S-ACE2 interaction or cell infection, may support the development of novel, potentially

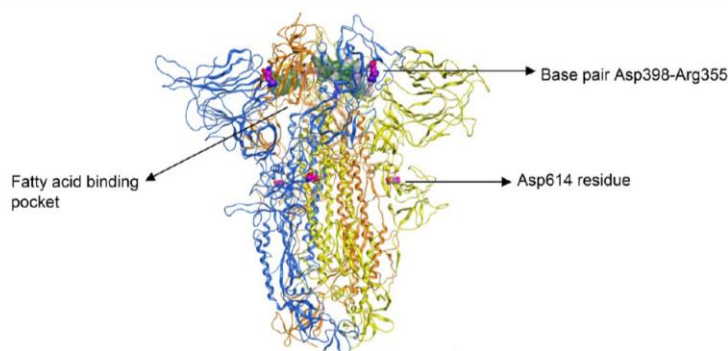


Figure 6. Ribbon representation of the S trimer (PDB ID: 6ZB5), in the side view with the base pair Asp398-Arg355 and residue Asp614 represented with carbon atoms in purple. The FABP is represented as a molecular surface.¹³

Table 1. Confirmed and Potential Ligands of the FABP, with *in Silico* Studies and Complementing *in Vitro* Testing

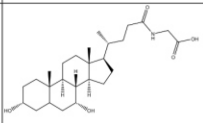
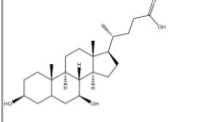
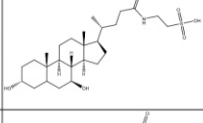
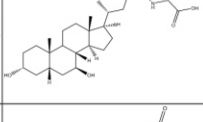
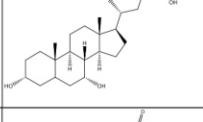
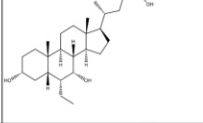
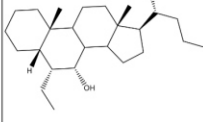
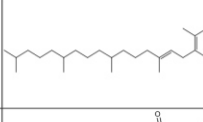
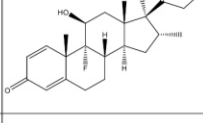
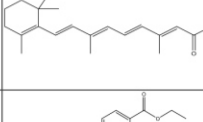
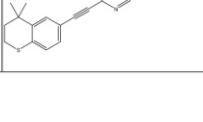
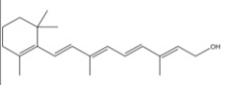
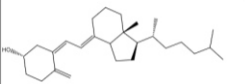
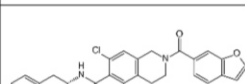
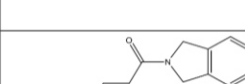
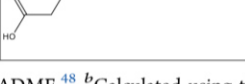
Compound	Structure	Molecular weight	cLogP ^a	LogD ^b (pH=7)	<i>In silico</i> studies	<i>In vitro</i> studies	Reference
Glyco-chenodeoxycholic acid		449.6	3.05	-0.5	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 15 % reduction in S-AI ^c in binding assays at 10 μ M	42
Ursodeoxycholic acid		392.57	3.60	0.4	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 15 % reduction in S-AI ^c in binding assays at 10 μ M	42
Tauro-ursodeoxycholic acid		499.7	2.76	-0.2	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 5 % reduction in S-AI ^c in binding assays at 10 μ M	42
Glyco-ursodeoxycholic acid		449.6	3.05	-0.5	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 20 % reduction in S-AI ^c in binding assays at 10 μ M	42
Chenodeoxycholic acid		392.6	3.60	1.3	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 10 % reduction in S-AI ^c in binding assays at 10 μ M	42
Obeticholic acid		420.6	4.16	2.1	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 15 % reduction in S-AI ^c in binding assays at 10 μ M	42
BAR 704		392.62	4.46	4.1	Binding pocket: Fpocket webserver with PDB ID: 6VSB Docking software: AutoDock	~ 35 % reduction in S-AI ^c in binding assays at 10 μ M	42
Vitamin K		450.7	8.34	9.7	Binding pocket: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	50% reduction in virion retention in cells at 4.7 μ M	19, 29
Dexamethasone		392.5	2.14	1.7	Binding pocket: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	50 % reduction in virion retention in cells at 3.2 μ M	19, 29
Tretinoin		300.4	5.01	3.0	Binding pocket: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	~ 60 % reduction in viral replication in cells at 1 μ M	19, 43
Tazarotene		351.5	4.61	5.2	Binding pocket: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	~ 50 % reduction in viral replication in cells at 1 μ M	19, 43

Table 1. continued

Vitamin A		286.5	5.04	4.7	Binding pocket: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	100% reduction in S-AI ^c in binding assays at 2.5 μM	19, 30
Vitamin D3		384.6	6.85	7.1	Binding pocket from: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	15% reduction in S-AI ^c in binding assays at 2.5 μM	19, 30
Linoleic acid		280.4	5.45	4.4	Binding pocket: LA in PDB ID: 6ZB5 Docking software: BUDE docking server	100% reduction in S-AI ^c in binding assays at 2.5 μM	24, 30
Myristic acid		228.4	4.45	3.3	Binding pocket: FABP in 6VXX Docking software: Glide XP	15% reduction in S-AI ^c in binding assays at 2.5 μM	30, 44
Stearic acid		284.5	5.93	5.1	Binding pocket: FABP in 6VXX Docking software: Glide XP	20% reduction in S-AI ^c in binding assays at 2.5 μM	30, 44
Oleic acid		282.5	5.71	4.8	Binding pocket: FABP in 6VXX Docking software: Glide XP	90% reduction in S-AI ^c in binding assays at 2.5 μM	30, 44
Lifitegrast		615.5	4.19	1.2	Binding pocket: S-ACE2 interaction surface in PDB ID: 6VW1 Docking software: Autodock Vina	50% reduction in cell infection at 1,295 μM	45, 46
Palmitic acid		256.5	5.20	4.2	Binding pocket: FABP in PDB ID 6VXX Docking software: Glide SP, Glide XP	15% reduction in S-AI ^c in binding assays at 2.5 μM	30, 44
Caprylic acid		144.2	2.23	0.9	Binding pocket: FABP in PDB ID 6VXX Docking software: Glide SP, Glide XP	20% reduction in S-AI ^c in binding assays at 2.5 μM	30, 44
SPC-14		306.3	1.29	-0.7	Binding Pocket: FABP in PDB ID 7E7B Docking software: DOCK3.7	Binding confirmed in SPR and cryo-EM No inhibition or cell-based assays	47

^aCalculated with SwissADME.⁴⁸ ^bCalculated using the Chemaxon LogD calculator.⁴⁹ ^cS-AI S-ACE2 interaction.

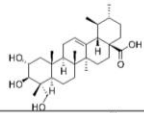
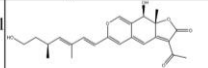
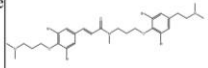
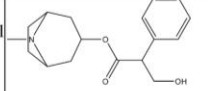
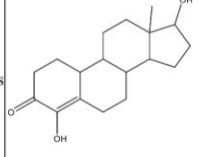
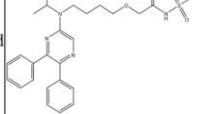
broad-spectrum therapeutic agents. Furthermore, a systematic elucidation of how each confirmed ligand binds, highlighting favored functional groups, scaffolds and essential interactions, can aid the design of new molecules with improved binding to this site, which may help in the inhibition of coronavirus infections. Compounds with potential or confirmed activity were evaluated with molecular docking analyses in the FABP when this data was unavailable to aid the discussion of potential interaction patterns.

2. METHODS

2.1. Literature Review. The systematic review followed the Preferred Reporting of Systematic Reviews and Metanalysis (PRISMA).³⁹ We concluded an extensive search for virtual screening studies focusing on the S protein through different databases, namely MEDLINE (PubMed), Scopus and Web of Science, to select literature published until May 15th, 2023. The main goal was to find compounds predicted to bind the

FABP, and as such, we collected all articles containing ("SARS-CoV-2" + "spike" + "virtual screening") in the title or abstract. Additionally, since LA, a fatty acid, is a known ligand, we also collected articles with the terms ("SARS-CoV-2" + "spike" + "linoleic acid") and ("SARS-CoV-2" + "spike" + "fatty acid"). As summarized in the PRISMA Diagram (Supporting Information Figure S1), a total of 738 articles were identified; filtering for duplicate articles resulted in a total of 387 different articles. From these, 159 articles were excluded since they were not virtual screening studies or had no experimental data regarding the S1 subunit or binding to the FABP. Finally, an additional 208 articles were also excluded since they were either virtual screening studies focusing exclusively on the S-ACE2 interaction or they reported no compounds predicted to interact with any FABP residues. Additional articles were added with complementary information, such as binding inhibition or cell-based assays. From the final selection of 20 articles, 21 compounds were identified with predicted binding

Table 2. Reported Compounds with Potential Activity without Validation in Experimental Assays

Compound	Total library	Structure	Interacting amino acids	cLogP ^a	LogD (pH=7) ^b	Molecular Weight	Reference
Asiatic acid	27 plant metabolites		Phe338, Tyr365, Leu368, Leu387, Phe392, Val395, Leu513, Val524	4.4	2.0	488.7	60
Rotinorinol C	92 natural compounds		Cys336, Phe338, Ala363, Phe374	2.6	1.2	398.4	59
Psammapplysene D	14 Marine Brominated Tyrosine Alkaloids		Gly339, Asn343, Ala344, Asp364, Ser366, Val367, Ser373, Phe374, Leu441, Arg509	6.6	0.9	783.24	61
Hyoscyamine	92 natural compounds		Thr376, Val407, Ala411, Val433, Tyr508	1.8	-0.6	289.4	59
Oxabolone	11,875 approved drugs		No data reported	2.75	2.5	290.4	62
Selexipag	2456 approved drugs		Leu368, Phe374, Pro384, Asn388, Asp405, Gln409, Arg408, Lys417	3.25	2.84	496.6	63

^aCalculated with SwissADME.⁴⁸ ^bCalculated using the Chemaxon LogD calculator.⁴⁹

to the S glycoprotein, along with complementary experimental results confirming either binding or functional activity. Additionally, 6 compounds were identified as possible binders, although no experimental assays were available. Identified structures with confirmed activity or potential activity were analyzed with docking simulations to further explore potential interactions.

2.2. Docking Simulations. All molecular modeling experiments were performed on an Asus WS X299 PRO Intel i9-10980XE CPU @ 3.00 GHz × 36 running Ubuntu 18.04 (graphic card: GeForce RTX 2080 Ti). Molecular Operating Environment (MOE) 2022.02¹³¹³ and Maestro (Schrödinger Release 2020-2) were used as molecular modeling software.⁴⁰ The structures of the compounds analyzed were built in MOE, saved in .sdf format, and prepared using the Maestro LigPrep tool by energy minimizing the structures (OPLS_2005 force field) and generating possible ionization states at pH 7 ± 2, tautomers, all possible stereoisomers per ligand, and low-energy ring conformers. The crystal structure of LA in the S glycoprotein was downloaded from the PDB (<http://www.rcsb.org/>) (accessed on 15 May 2023); PDB code 6ZB5). The protein was preprocessed with the MOE Protein Preparation tool, and the resulting protein–ligand complex was saved in mae format and prepared using the Schrödinger Protein Preparation Wizard by assigning bond

orders, adding hydrogens, and performing a restrained energy minimization of the added hydrogens using the OPLS_2005 force field. A 15 Å docking grid was prepared using the cocrystallized LA as the centroid. Molecular docking studies were performed using Glide SP, keeping the default parameters and setting 5 as the number of output poses per input ligand, performing for each pose a postdocking minimization. The output poses were saved as mol2 files. The docking results were visually inspected by using MOE 2022.02.

3. RESULTS

3.1. Compounds Interacting with the FABP. Predicted interactions with the FABP for selected compounds are reported in virtual screening studies; however, experimental confirmation of binding and its effect in receptor recognition/cell entry are not always available. Therefore, compounds with reported experimental data were prioritized (Table 1), with additional compounds identified when only predicted interactions are available (Table 2). The libraries from which these compounds are derived vary widely, from studies testing a single compound to libraries of a few thousand molecules. Selected compounds may originate from libraries with plant metabolites or other biological sources.⁴¹ Identification of compounds potentially active in the FABP was based on described interactions, either predicted in virtual screening

studies or directly confirmed by cryo-EM or NMR experiments. Regarding experimental validation, data was crossed with additional studies whenever complementary information was available. It should be noted that, when direct FABP activity is not accessed, validation was based on inhibition assays for the S-ACE2 interaction or cell-based infection assays, without, however, confirming specific interaction in the FABP. Finally, the selection of compounds without experimental validation (Table 2) was based on the patterns of interaction identified for LA, focusing on residues that compose the FABP.

As with LA, most compounds in Table 1 have a molecular weight over 250 and below 500 g/mol, likely required to adequately occupy the pocket. Despite the FABP being composed mainly of hydrophobic amino acids, the predicted logP of confirmed ligands varies between 1 and 8, while predicted logD (pH 7) varies between −0.5 and 9.7. These variations are due to the presence of functional groups (such as carboxylic acids and hydroxy groups) capable of interacting with positively charged amino acid residues. Another significant feature is the overall rigidity of the scaffold: while LA, the first confirmed ligand for this pocket, is flexible, this does not appear to be an essential requirement, as confirmed by lifitegrast, characterized by bulkier and rigid ring systems. As expected from the Cryo-EM structure with LA (PDB ID: 6ZB5), the predicted binding for these compounds always features two RBDs and interactions with both the hydrophilic and hydrophobic portions of the pocket.

3.2. Confirmed Ligands of the FABP. Despite the experimental effect in the virus-host interaction detected for multiple compounds, only two have been confirmed to directly bind the FABP and exert a biological activity, lifitegrast and LA.^{24,46} Lifitegrast is a drug (Xiidra) used for the treatment of dry eye disease as an ophthalmic solution,⁵⁰ and while LA has been cocrystallized in the S protein, lifitegrast was identified in a virtual screening study, for its predicted direct inhibition of the S-ACE2 interaction, unrelated to the FABP.⁴⁶ Lifitegrast predicted activity was confirmed later, as it was found to significantly reduce the S-ACE2 interaction in surface plasmon resonance (SPR) analyses.⁴⁶ Remarkably, additional studies have pointed to an effect through the FABP: measured with chemical shift perturbation (CSP) NMR, the chemical perturbation of lifitegrast matches the effect of LA, suggesting an action on the FABP instead of the protein surface.⁴⁵ Furthermore, saturation transfer difference (STD) NMR showed that all aromatic groups of lifitegrast are in contact with the RBD and buried in the FABP, as would be expected for binding to the hydrophobic portion of the pocket.⁴⁵ Binding assays determined a KD of 1.92 nM with the S protein for lifitegrast and a reduction of 99.8% of the RBD-ACE2 protein–protein interaction in SPR competition assays at 1.92 nM.⁴⁶ Consequently, further cell-infection studies determined an IC₅₀ of 1,295.3 μ M for *in vitro* infection of Vero E6 cells with no cytotoxicity up to 5 mM.⁴⁶ Given that lifitegrast was identified when screened against the RBD but its activity was later reconnected to binding in the FABP, we performed a docking analysis in the FABP with Glide SP,⁴⁰ to predict how it may bind the FABP (Figure 7).⁵¹

In line with the network of interactions with hydrophobic moieties detected in NMR assays, lifitegrast can be deeply buried in the hydrophobic region of the pocket.⁴⁵ As with LA, the predicted pose found for lifitegrast can be stabilized by interaction with hydrophilic residues (Arg408) and additional

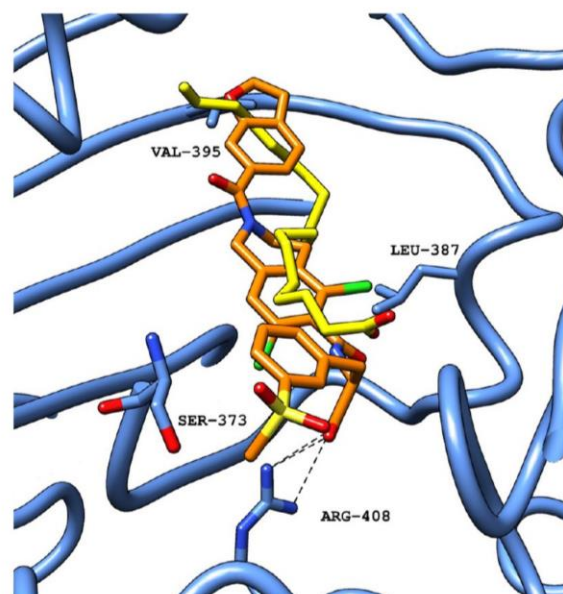


Figure 7. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with lifitegrast (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

contacts with the hydrophobic amino acids (Ala372, Ser373, Leu387, and Val395). Additionally, the predicted binding pose for lifitegrast indicates extensive pocket occupation. Taken together, LA and lifitegrast show that although molecules with long and flexible carbon chains can access the pocket and influence cell infection and the virus life cycle, more rigid and bulkier compounds can also adequately bind and exert an effect through occupation of the FABP. In addition to lifitegrast and LA, SPC-14 has also been confirmed to bind the FABP, although no inhibition assays on the interaction between the S protein and ACE2 have been reported.⁴⁷ Binding was determined in SPR assays with the S protein, resulting in a binding affinity (KD) of 9.5 μ M, with further cryo-EM structural determination confirming binding and clarifying how this compound binds to the FABP (Figure 8).⁴⁷

The cryo-EM structure for SPC-14 shows a different interaction pattern with LA, as SPC-14 is further buried in the pocket, and hydrophobic stabilizing interactions are prominent, particularly with the nitro group having π -stacking interaction with Phe338.⁴⁷ Although no interactions are directly established with the hydrophilic portion of the site, the carboxylic acid remains at a distance of 6 Å from hydrophilic residue Arg408 (Figure 8). Along with the cocrystallization of SPC-14, 48% of S protein population appeared in the closed conformation, with the remaining only showing one RBD exposed for interaction with ACE2.⁴⁷ The data collected for SPC-14 suggest that it is likely to affect the RBD-ACE2 interaction, and potentially the virus life cycle, although this has not yet been confirmed experimentally.

3.3. Compounds with Predicted Binding and Confirmed Activity. **3.3.1. Sterol Backbone.** Although LA and lifitegrast have been explicitly confirmed to bind the FABP and have experimental results for their direct binding to the FABP,

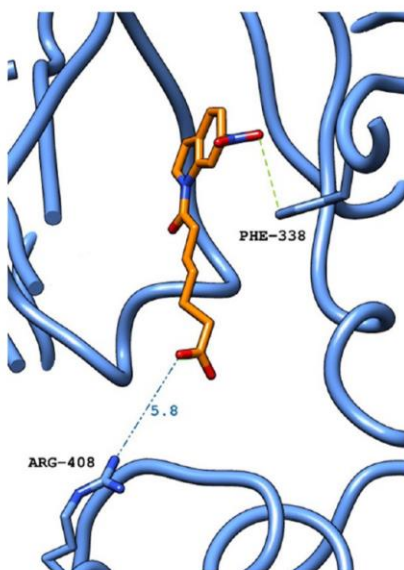


Figure 8. Crystallized SPC-14 (carbon atoms in orange) (PDB ID: 8H3E) in the FABP in a ribbon representation (blue). The green dashed line represents hydrophobic interactions, and the blue dashed line represents the distance (in angstrom (Å)) between SPC-14 and the side chain of Arg408.⁵²

leading to reduced interaction or cell infection, other compounds with exclusively binding or cell-based assays could potentially act through binding the FABP. In a virtual screening campaign, natural bile acids and semisynthetic derivatives were predicted to establish interactions in a proposed pocket, with common residues to the FABP, defined with the Fpocket online tool (using the S protein PDB ID: 6VSB).^{42,53} Proposed interactions are established with residues Lys378, Thr376, Phe377, Cys379, Tyr380, and Pro384, suggesting that the molecules selected could interact with FABP residues.⁴² Importantly, activity was confirmed for these compounds in S-ACE2 inhibitor screening assays, with the best ligand, BAR704, reaching up to 40% inhibition at a 10 μ M concentration.⁴² However, interaction with the FABP is not discussed, as some activity is detected, and in order to predict if BAR704 could interact with this pocket, we performed a docking analysis with Glide SP, obtaining the predicted binding pose illustrated in Figure 9.⁵¹

Bar704 is predicted to achieve a good overall occupation of the pocket and potentially establish interactions with residues Cys379, Cys432, Arg408, Thr415, and Phe377. Predicted interactions with Arg408, Phe377, and Cys432 are in common with LA.²⁴ Since, similar to lifitegrast, this bile acid derivative possesses a bulkier scaffold, it can be expected to have the same ability to enter the pocket and potentially influence the virus-host interaction, resulting in the reduced S-ACE2 interaction observed.

In addition to bile acids, further compounds with a steroid backbone have been identified in molecular dynamics studies, namely dexamethasone, with an irregular interacting pattern.¹⁹ Dexamethasone's activity has been confirmed in experimental assays against the S protein, in which it reduced S-mediated binding to the ACE2 receptor, with an IC₅₀ of 3.2 μ M for retention of virions in A549 human alveolar basal epithelial

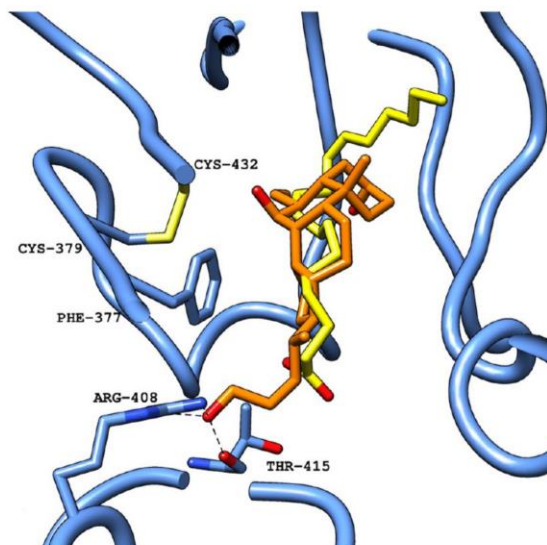


Figure 9. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with BAR 704 (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

cells.²⁹ A docking analysis using Glide SP has generated the predicted binding pose shown in Figure 10. The irregular predicted pattern (hydrophobic interaction with Phe377) of interaction suggests that dexamethasone is buried in the

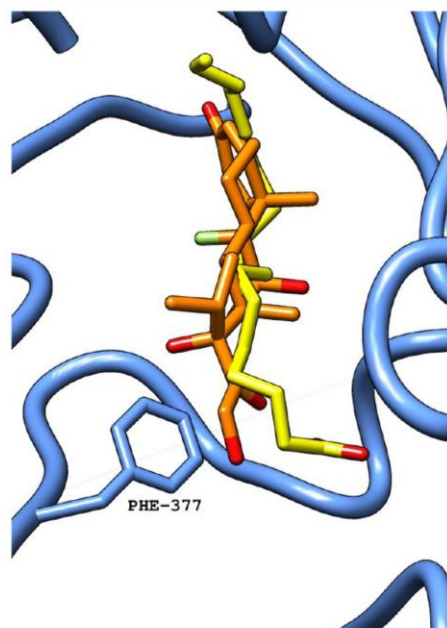


Figure 10. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with dexamethasone (carbon atoms in orange) in the FABP in ribbon representation (blue).⁵²

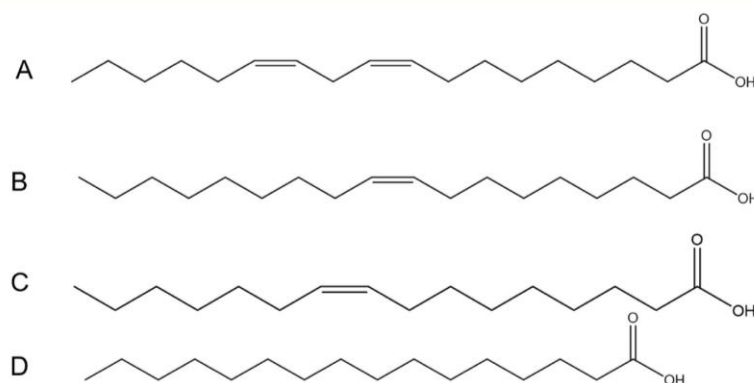


Figure 11. A) Structures of (A) LA (polyunsaturated), (B) oleic acid (ω -9), (C) palmitoleic acid (ω -7), and palmitic acid (unsaturated).

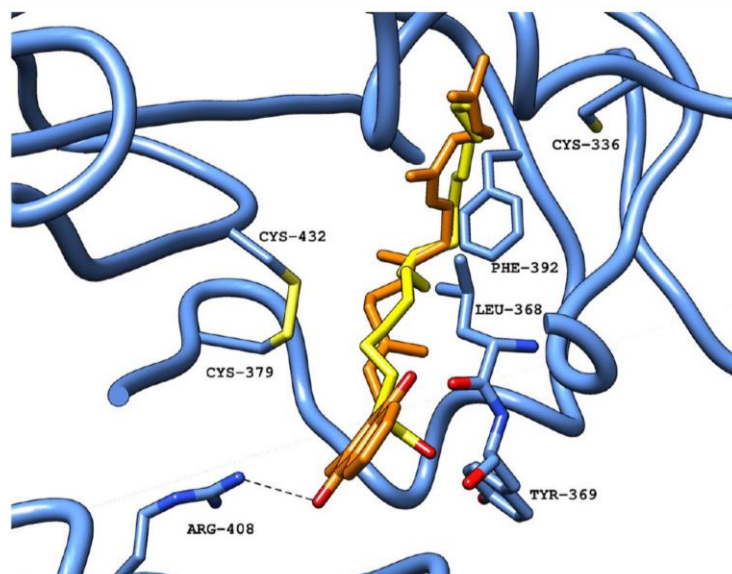


Figure 12. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with vitamin K (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

pocket, perhaps with a pattern similar to SPC-14 and no direct interaction with the hydrophilic region.^{19,47}

3.3.2. Fatty Acids. In addition to LA, other fatty acids are also expected to bind the FABP and affect the RBD-ACE2 interaction, with nonsaturated, mono-, and polyunsaturated fatty acids included in virtual screening campaigns against the pocket and predicted to interact similarly to LA.⁴⁴ Palmitic acid, a saturated fatty acid, has been described as a potential ligand.⁴⁴ However, experimental results disprove this, possibly due to the bent nature of the pocket, with the *cis*-bonds present in LA appearing to favor an optimal binding. Palmitic acid is associated with a modest 15% reduction in S-ACE2 binding assays at 9.8 mM and no direct binding to the S protein identified when monitored with liquid chromatography–mass spectrometry. On the other hand, oleic acid and LA were both detected bound to the S glycoprotein.^{29,30,38} This is further extended to other saturated fatty acids, with myristic and stearic acid resulting in 15% and 10% reduction in

binding assays at 10.9 mM and 8.8 mM, respectively.^{29,30} Multiple polyunsaturated fatty acids have shown a 100% reduction in binding assays at $\sim$ 8 mM, such as linolenic acid (9.0 mM), eicosapentaenoic acid (8.3 mM), and LA (8.9 mM), consistent with predicted binding poses.³⁰ As for monounsaturated fatty acids, these show varying behaviors in between polyunsaturated and saturated, with oleic acid achieving a strong effect at relatively high concentrations (90% reduction in binding assays at 8.9 mM) and palmitoleic acid producing instead a weak effect (10% reduction in binding assays at 9.8 mM) (Figure 11).^{29,30} For polyunsaturated fatty acids, as multiple double bonds are present, the molecule is able to accommodate and appropriately fit in the pocket, in contrast with the effect in activity for monounsaturated fatty acids, where oleic acid and palmitoleic acid show widely different activities.^{29,30} This may be attributed to the unsaturated bond, with oleic acid being an ω -9 fatty acid and palmitoleic an ω -7 fatty acid, with the extended carbon

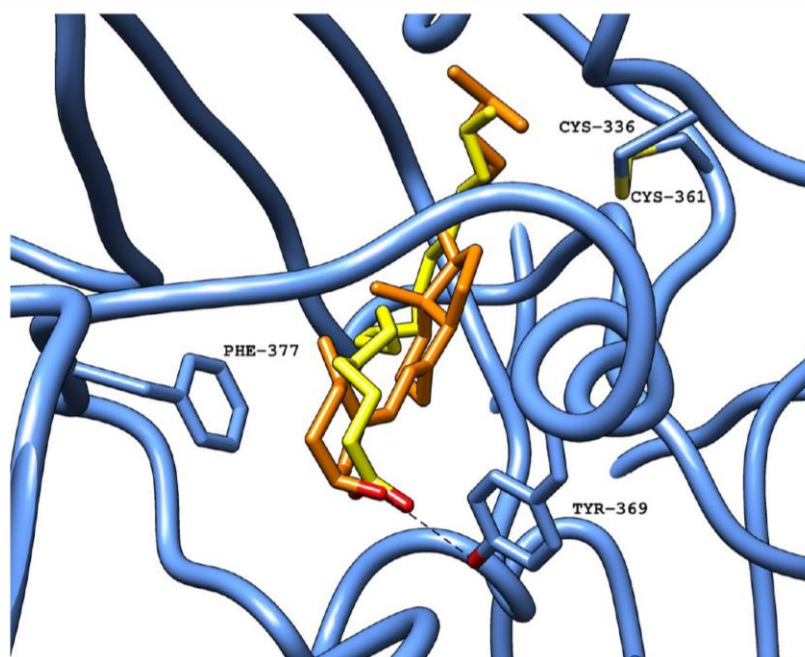


Figure 13. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with predicted pose for vitamin D (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

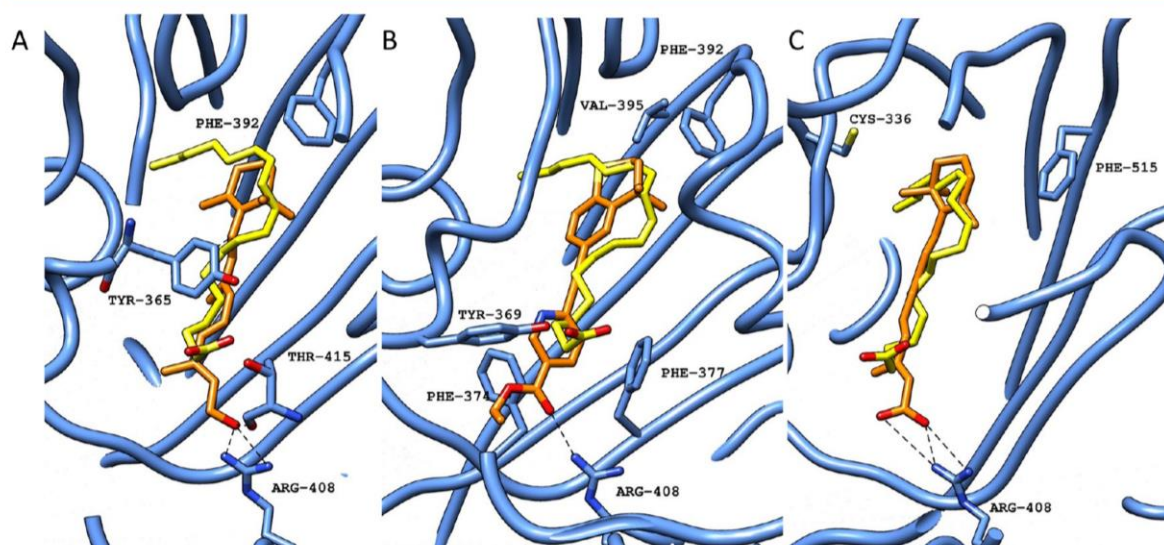


Figure 14. A) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with the predicted binding pose for vitamin A (carbon atoms in orange) in the FABP in ribbon representation (blue). B) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with the predicted binding pose for tazarotene (carbon atoms in orange) in the FABP in ribbon representation (blue). C) Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with the predicted binding pose for Tretinoin (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

chain after the double bond likely leading to different pocket occupation and established interactions.

3.3.3. Retinoids and Lipid-Soluble Vitamins. As with fatty acids, lipid soluble vitamins (A, D, and K), as well as

tazarotene and tretinoin, could be good ligands for the FABP, identified in virtual screening studies.¹⁹ Inhibitory activity has been confirmed for vitamins A and K, with vitamin A successfully blocking the interaction with ACE2 (100%

reduction in S-ACE2 binding assays at 8.7 mM), while vitamin D only reduced the interaction in S-ACE2 binding assays by 13% when tested at 6.0 mM.³⁰ Additional experimental assays for vitamin K showed S-mediated reduction of binding to ACE2, with an IC₅₀ of 4.7 μ M for retention of virions in A549 human alveolar basal epithelial cells.²⁰ To further elucidate the interaction with the FABP, we have generated a binding pose with a docking simulation using Glide SP (Figure 12).⁵¹

Vitamin K is predicted to occupy the pocket similarly to LA, establishing polar interactions with residue Arg408.¹⁹ In the generated pose, additional contacts with residues Leu368 and Tyr369 are predicted, and the long flexible chain extends deep into the hydrophobic portion of the pocket and establishes further interactions with residues Cys336, Cys379, Phe392, and Cys432. The predicted extensive network of interactions justifies the stabilization within the pocket and the ability to affect the virus-host interaction, with antiviral activity detected in virion cell infection assays.^{19,29} Vitamin D is also predicted to bind to the FABP, although S-ACE2 binding assays show limited binding (13% reduction at 6.5 mM).^{19,30} Predicted interactions are similar to LA, with direct contacts with Arg408 and Gln409 identified through molecular dynamics simulations.¹⁹ The best binding pose generated by Glide SP for vitamin D (Figure 13)⁵¹ is in line with this observation, both for the main stabilizing interactions and the overall pocket occupation.¹⁹ Other molecular properties, such as molecular weight (Vit D: 384.6 g/mol; LA: 280.4 g/mol), predicted log P values (Vit D: 6.85; LA: 5.45), or predicted LogD at pH 7 (Vit D: 7.1; LA: 4.4), could point to an increase of hydrophobic features, reflected in these values, as a potential trend in support of decreased binding. However, vitamin K surpasses vitamin D in all parameters (MW: 444.65 g/mol, LogP: 7.66, LogD: 9.7) but maintains both binding and FABP-mediated functional effects.²⁹

3.3.4. Retinoids. Given that vitamin A has been confirmed to bind the FABP and affect the S-ACE2 interaction, additional retinoids could be expected to have a similar effect. Particularly, tretinoin and tazarotene have been identified in both virtual screening studies and high-throughput screening studies, with predicted binding in the FABP and reduced viral infection by ~50% in Vero E6 cells at a test concentration of 1 μ M for both compounds.^{19,43} Taking into account the structural features of previously identified compounds binding the FABP, these retinoids are expected to bind strongly to the pocket.

Virtual screening studies predicted a polar interaction with Arg408 for these retinoids, anchoring the molecules in the pocket, an essential feature previously described for LA.¹⁹ Additional binding poses generated with Glide SP⁵¹ have, for vitamin A (Figure 14A), the hydroxyl group in direct contact with Thr415, whereas Tazarotene (Figure 14C) is predicted to interact with Tyr369. On the other hand, in the hydrophobic portion of the pocket, all three compounds are predicted to establish π -stacking or hydrophobic interactions with phenylalanine residues (Phe492 for vitamin A and tazarotene, Phe515 for tretinoin (Figure 14B), or Cys336 for tretinoin), with a predicted pocket occupation similar to LA.

To summarize, the compounds discussed so far have shown the ability to bind the FABP, confirmed for LA, lifitegrast, and SPC-14, and/or to affect the S-ACE2 interaction in binding assays, consequently resulting in reduced cell infection detected in cell-based assays. Binding trends include the ability to interact with hydrophobic residues buried in the pocket and

with polar functional groups of amino acid residues that define the pocket entrance. Additionally, there are two main structural features in the chemical scaffold of the active compounds. The first is the presence of bulky ring-based systems, as happens with lifitegrast, that achieves a 50% reduction in cell infection at 1.295 μ M, or SPC-14 with a binding affinity (K_D) of 6.9 μ M. Alternatively, active compounds also show the presence of only up to two aromatic rings, attached to a more flexible backbone, as observed with LA, retinoids, and vitamin K. Either group is predicted to have the ability to accommodate in the pocket, making different nonbonded interactions. Particularly, the presence of aromatic rings is likely essential so that π interactions can be established with the numerous phenylalanine residues, buried, and spread across the pocket. Scaffolds capable of interacting with other hydrophobic amino acids that coat the deeper portions of the site are also favored, as is the case for retinoids.

Regarding the hydrophilic entrance of the pocket, polar interactions with the amino acid residues present in this portion are common for stability, as initially identified with the cocrystallized LA. The carboxylic acid group appears to be an important feature to interact with the hydrophilic residues, but other functionalities, such as the sulfoxide in lifitegrast, also show the potential to make direct contacts with electrophilic residues in the vicinity. On the other hand, cocrystallized SPC-14 is buried in the pocket and unable to directly interact with these residues. The same behavior is predicted for dexamethasone.¹⁹ Dexamethasone and SPC-14 have the lowest predicted LogP values (2.14 and 1.29, respectively), along with low logD values (−0.7 and 1.7, respectively), which may lead to an interaction pattern more dependent on the hydrophobic residues lining the deeper portions of the pocket, further away from the hydrophilic residues in the entrance. Considering the active compounds highlighted in this review, their chemical variety appears to be quite limited, as many studies focused on compounds similar to LA. However, the activity observed for lifitegrast indicates that more complex chemical scaffolds could possibly be associated with an improved overall occupation of the FABP, as can also be observed for SPC-14, with a new interaction pattern, similar to that previously predicted for dexamethasone.¹⁹ Further studies, such as the resolution of a three-dimensional structure of the FABP in complex with lifitegrast or dexamethasone, are needed to further confirm interactions leading to increased activity and to elucidate if dexamethasone achieves the high activity observed with a pattern similar to SPC-14. Additionally, biological assays for SPC-14 are required to clarify if it is also capable of affecting the S-ACE2 interaction or the virus life cycle. The studies considered in this review, describing compounds identified as FABP ligands, can help to understand important features for FABP binding and how this pocket can be explored to design optimized ligands. The chemical space of possible candidates has been significantly expanded since the first ligand was identified, with the identification of bulkier scaffolds and new functional groups with biochemical and antiviral activity.

3.4. Compounds Predicted to Bind the FABP without Reported Experimental Activity. Although experimental data is essential to validate virtual screening studies, this is not always available in published works describing virtual screening campaigns,^{54–56} representing a significant limitation, as potential antiviral compounds cannot be selected for optimization before experimental confirmation.⁵⁷ For example,

as discussed previously, vitamin D and palmitic acid were both predicted to bind the FABP, but S-ACE2 binding inhibition assays proved otherwise.^{30,58}

Since experimental results are an essential requirement to validate potential ligands in virtual screening studies, a clear separation is needed from the compounds shown in Table 1. Although lacking experimental confirmation, the compounds listed in Table 2 have been predicted to bind to some residues in the FABP, based on their potential interactions in virtual screening studies. These compounds were selected considering the most common interaction features identified for confirmed ligands, namely, hydrophobic contacts with phenylalanine residues that coat the pocket, or polar contacts with the hydrophilic amino acids Arg408, Gly408, and Lys417. Virtual screening methodologies are applied for compound libraries using a wide variety of targets and structures, from portions of the S protein (closed and open states, as well as pre- and postfusion structures) to other SARS-CoV-2 proteins, such as the main protease or nucleocapsid protein.⁵⁹

Asiatic acid was included in a virtual screening study against the S glycoprotein, where multiple predicted residues described for interaction are also residues present in the FABP, such as Phe338, Tyr365, and Phe392.⁶⁰ Additionally, asiatic acid is a pentacyclic triterpenoid with a backbone similar to bile acid derivatives or dexamethasone, and it could therefore also bind to the pocket. Interactions with hydrophilic residues are not described, but multiple heteroatoms (within carbonyl and hydroxyl groups) could potentially establish these interactions. A docking simulation was run to predict how asiatic acid may bind the FABP and to assess if it has a similar behavior to other sterols. However, no favorable binding poses could be obtained for this compound using Glide SP, potentially indicating an inefficient interaction with the FABP.⁴⁰ Likewise, although Psammaphysene D has been described to interact with some FABP residues (Ser373, Phe374), no favorable binding poses could be obtained using the Glide SP docking tool.^{40,61,60} In the case of Psammaphysene D, the failure to obtain docking poses could be due to the larger size of the molecule, with a molecular weight of 783.2 g/mol, significantly surpassing confirmed ligands and likely not fitting the pocket properly. As for asiatic acid, taking into account the binding predicted for dexamethasone and bile acid derivatives, the absence of docking poses obtained is unexpected. This lack of favorable binding poses could be due to the additional ring, as it has a pentacyclic structure, or to the lack of a carbon chain capable of interacting with the electrophilic pocket entrance, extending from the main ring system, as observed in both dexamethasone and the bile acid derivatives.

Rotiorinol C was also identified in a virtual screening study not directed at the FABP, when screened against the RBD portion of the S protein (PDB ID: 6M17).⁵⁹ Several predicted interacting residues described in this study are in the FABP, particularly in the hydrophobic portion (Cys336, Phe338, Ser373 (H-bond), Phe374).⁵⁹ From a chemical point of view, rotiorinol C is similar to the retinoid backbone with a flexible carbon chain and a rigid, condensed ring system, suggesting potential activity due to FABP binding. The predicted binding pose obtained for rotiorinol C with molecular docking software Glide SP (Figure 15)⁴⁰ indicates extensive potential interactions in the FABP.

The main contacts are predicted with residues Tyr369, Gln409, Gly416 and Lys417, and additional predicted

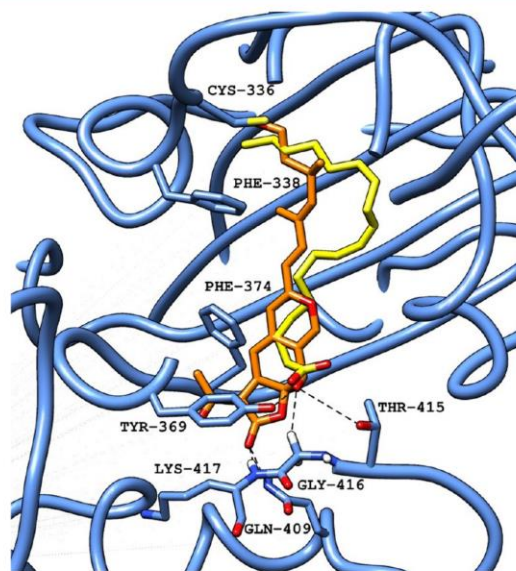


Figure 15. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with rotiorinol C (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

interactions with Phe374 and Phe338, as initially identified. Other interactions with Cys336 and Thr415 could further stabilize the molecule. When compared to known ligands, occupation of the pocket is similar to that of lifitegrast, with logP (lifitegrast: 4.19, rotiorinol C: 2.6) and LogD (lifitegrast: 1.2, rotiorinol C: 1.2), as well as molecular weight (lifitegrast: 615.5, rotiorinol C: 398.4) compatible with binding to the FABP and subsequent potential effects for the inhibition of the S-ACE2 interaction.

Similar to rotiorinol C, hyoscyamine was also identified in a virtual screening study not directed at the FABP, against multiple S protein conformations.⁵⁹ Despite only a few interacting residues highlighted in the FABP, a docking pose was generated for this compound using Glide SP (Figure 16).⁴⁰ This molecule has unique features when compared with the others discussed above as it is more compact and contains a tropane ring that could further extend the chemical space of potential binders to this site and explore new potential interactions that may strengthen binding.

Due to the compact structure of hyoscyamine, the predicted binding pose in the FABP is characterized by limited pocket occupation, although some stabilizing interactions, particularly two H-bonds with Arg408 and Thr415 and an additional contact with Phe374, are predicted. Further experimental assays are needed to elucidate whether hyoscyamine has any inhibitory activity of the S-ACE2 binding.

Oxabolone and Selexipag have been identified in two separate virtual screening studies directed at the FABP, based on the LA cocrystallized structure (PDB ID: 6ZB5 and 6ZB4), and libraries of approved drugs.^{62,63} However, none of these studies assessed biological activity or determined binding affinities with the S protein; therefore, they still require

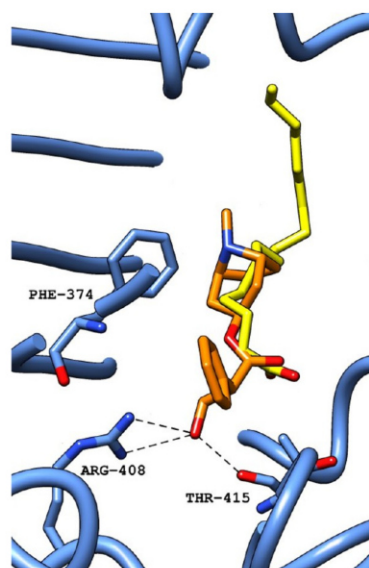


Figure 16. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) superimposed with hyoscyamine (carbon atoms in orange) in the FABP in ribbon representation (blue). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein.⁵²

validation for the virtual screening campaigns performed. As expected, the reported predicted binding and pocket occupation follows patterns previously identified, with oxabolone predicted to interact as SPC-14, whereas selexipag contains three aromatic rings and a flexible chain that extends toward the hydrophilic residues in the pocket.^{62,63}

In summary, six additional compounds were identified as potential ligands to the FABP, based on their predicted interactions with the S protein, described in the original virtual screening studies that explored these chemical entities. Upon analysis of these molecules for their binding to the FABP using a molecular docking procedure, only for two of these potential ligands, rotiorinol C and hyoscyamine, favorable binding poses to this site could be observed. The binding pattern predicted for these two additional compounds is similar to that previously observed for confirmed ligands. Moreover, the molecular scaffold of rotiorinol C shares features similar to those of retinoids, as it shows the presence of a ring system attached to a carbon chain that contains a heteroatom, available to interact with the hydrophilic portion of the pocket. However, only experimental validation of the predicted activity may confirm the accuracy of these observations. Similarly, selexipag and oxabolone are predicted to interact with the FABP, but only experimental studies can validate the potential ability to affect the S-ACE2 interaction. On the other hand, hyoscyamine presents a novel chemical scaffold, featuring a tropane ring. Although its predicted pocket occupation appears limited compared to that of the confirmed active compounds, the experimental validation of its activity may open new avenues for the design of new therapeutic agents with diverse structural features.

4. CONCLUSIONS

The FABP in the S surface glycoprotein is a promising target for drug discovery for SARS-CoV-2: compounds binding to this site have the potential to exert a strong effect in cell recognition and cell infection, stabilizing the inactive conformation of the S glycoprotein. The first known ligand for this pocket, LA, is capable of significantly affecting virus-host interaction and cell infection. Identifying new, more potent binders to this site can provide the basis to develop novel, broad-spectrum treatment options against current and future-emerging coronaviruses.

In this review, we selected 20 studies that describe 21 ligands with a confirmed binding and/or effect on the S-ACE2 interaction and six additional ligands with a predicted binding to the FABP, for which no experimental results have been reported yet.

In addition to LA, lifitegrast has also been confirmed not only to have a strong inhibitory effect on the S-ACE2 interaction and ability to reduce cell infection but also to directly bind to this pocket, showing that long, flexible carbon chains typical of fatty acids are not an essential requirement for binding. The confirmation of SPC-14 as a new ligand further indicates that different chemical scaffolds and interaction patterns can be accommodated in the pocket, although SPC-14 is still lacking experimental validation of the inhibitory effect of Spike-ACE2 interaction or antiviral cell-based assays.

Since the remaining compounds discussed are predicted to bind the pocket and have either confirmed inhibitory activity of the S-ACE2 interaction or antiviral activity, with a reduction of cell infection in cell-based assays, their observed activity is likely due to binding to the FABP. Among the compounds with inhibitory or antiviral activity, multiple ring systems, such as the sterol backbone, are a recurring feature, with retinoids, such as tretinoin, and unsaturated fatty acids with long carbon chains also affecting the virus-host interaction. The pattern of interactions is consistent, with common interactions established not only with the hydrophilic portion of the site but also critically throughout the hydrophobic region.

Additional compounds from virtual screening campaigns for which no experimental data is currently available have been analyzed with docking simulations in the FABP. Among them, rotiorinol C is associated with the best predicted binding profile, indicating a likely potential for this molecule to affect the virus-host interaction.

Overall, the observations summarized in this review may enable the design of more potent, broad-spectrum new compounds, able to strongly bind the FABP and therefore prevent cell recognition and entry of coronaviruses currently affecting humans, and of those possibly emerging in the future.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.3c00803>.

Chemical structures of all compounds reported (as SMILES) (XLSX)

PRISMA diagram for inclusion of studies (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Luis Queirós-Reis – Abel Salazar Institute of Biomedical Sciences (ICBAS), University of Porto, 4050-313 Porto,

Portugal; orcid.org/0000-0002-5204-5046;
Email: up201205115@up.pt

Authors

João R. Mesquita – Abel Salazar Institute of Biomedical Sciences (ICBAS), University of Porto, 4050-313 Porto, Portugal; Epidemiology Research Unit (EPlunit), Institute of Public Health, University of Porto, 4050-091 Porto, Portugal; orcid.org/0000-0001-8769-8103

Andrea Brancale – University of Chemistry and Technology, Prague, 166 28 Praha, Czechia; orcid.org/0000-0002-9728-3419

Marcella Bassetto – School of Pharmacy and Pharmaceutical Sciences, College of Biomedical and Life Sciences, Cardiff University, Cardiff CF10 3BN, U.K.; Department of Chemistry, Faculty of Science and Engineering, Swansea University, Swansea SA2 8PP, U.K.; orcid.org/0000-0002-2491-5868

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.3c00803>

Funding

L.Q.-R. would like to acknowledge Fundação para a Ciência e para a Tecnologia for the grant “2020.10230.BD” under the program “DOCTORATES 4 COVID-19” and EEA Grants/Norway Grants for the grant “FBR_OCS2_53”.

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Wang, M. Y.; Zhao, R.; Gao, L. J.; Gao, X. F.; Wang, D. P.; Cao, J. M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front Cell Infect Microbiol* **2020**, *10*, No. 587269.
- (2) WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed 09/03/2021).
- (3) WHO. Coronavirus Disease (COVID-19) Dashboard. WHO 2020. <https://covid19.who.int/> (accessed 2023/09/04).
- (4) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol* **2021**, *19* (3), 141–154.
- (5) Woo, P. C.; Lau, S. K.; Huang, Y.; Yuen, K. Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med* (Maywood) **2009**, *234* (10), 1117–1127.
- (6) Schoeman, D.; Gordon, B.; Fielding, B. C. Pathogenic Human Coronaviruses. *Reference Module in Biomedical Sciences*; 2021.
- (7) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260–1263.
- (8) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veeler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181* (2), 281–292.e6.
- (9) Huang, Y.; Yang, C.; Xu, X.-f.; Xu, W.; Liu, S.-w. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica* **2020**, *41* (9), 1141–1149.
- (10) Queirós-Reis, L.; Gomes da Silva, P.; Gonçalves, J.; Brancale, A.; Bassetto, M.; Mesquita, J. R. SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *International Journal of Molecular Sciences* **2021**, *22* (19), 10836.
- (11) Xia, X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. *Viruses* **2021**, *13* (1), 109.
- (12) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581* (7807), 221–224.
- (13) Chemical Computing Group, Inc. *Molecular Operating Environment* (MOE 2022.02); Chemical Computing Group, Inc. <http://www.chemcomp.com> (accessed 2023/05/07).
- (14) Vogel, A. B.; Kanevsky, I.; Che, Y.; Swanson, K. A.; Muik, A.; Vormehr, M.; Kranz, L. M.; Walzer, K. C.; Hein, S.; Güler, A.; et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* **2021**, *592*, 283.
- (15) Sinha, S. K.; Shakya, A.; Prasad, S. K.; Singh, S.; Gurav, N. S.; Prasad, R. S.; Gurav, S. S. An in-silico evaluation of different Saikosaponins for their potency against SARS-CoV-2 using NSP15 and fusion spike glycoprotein as targets. *J. Biomol Struct Dyn* **2020**, *39* (9), 3244–3255.
- (16) Romeo, A.; Iacovelli, F.; Falconi, M. Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors. *Virus Res.* **2020**, *286*, No. 198068.
- (17) Rodríguez, Y.; Cardoze, S. M.; Obineche, O. W.; Melo, C.; Persaud, A.; Fernández Romero, J. A. Small Molecules Targeting SARS-CoV-2 Spike Glycoprotein Receptor-Binding Domain. *ACS Omega* **2022**, *7* (33), 28779–28789.
- (18) Chan, W. K. B.; Olson, K. M.; Wotring, J. W.; Sexton, J. Z.; Carlson, H. A.; Traynor, J. R. In silico analysis of SARS-CoV-2 proteins as targets for clinically available drugs. *Sci. Rep.* **2022**, *12* (1), 5320.
- (19) Shoemark, D. K.; Colenso, C. K.; Toelzer, C.; Gupta, K.; Sessions, R. B.; Davidson, A. D.; Berger, I.; Schaffitzel, C.; Spencer, J.; Mulholland, A. J. Molecular Simulations suggest Vitamins, Retinoids and Steroids as Ligands of the Free Fatty Acid Pocket of the SARS-CoV-2 Spike Protein*. *Angew. Chem., Int. Ed. Engl.* **2021**, *60* (13), 7098–7110.
- (20) Mohebbi, A.; Askari, F. S.; Sammak, A. S.; Ebrahimi, M.; Najafimemar, Z. Druggability of cavity pockets within SARS-CoV-2 spike glycoprotein and pharmacophore-based drug discovery. *Future Virology* **2021**, *16* (6), 389–397.
- (21) Siragusa, L.; Menna, G.; Buratta, F.; Baroni, M.; Desantis, J.; Cruciani, G.; Goracci, L. CROMATIC: Cross-Relationship Map of Cavities from Coronaviruses. *J. Chem. Inf. Model.* **2022**, *62* (12), 2901–2908.
- (22) Ghosh, D.; Ghosh Dastidar, D.; Roy, K.; Ghosh, A.; Mukhopadhyay, D.; Sikdar, N.; Biswas, N. K.; Chakrabarti, G.; Das, A. Computational prediction of the molecular mechanism of statin group of drugs against SARS-CoV-2 pathogenesis. *Sci. Rep.* **2022**, *12* (1), 6241.
- (23) Yan, B.; Chu, H.; Yang, D.; Sze, K. H.; Lai, P. M.; Yuan, S.; Shuai, H.; Wang, Y.; Kao, R. Y.; Chan, J. F. Characterization of the Lipidomic Profile of Human Coronavirus-Infected Cells: Implications for Lipid Metabolism Remodeling upon Coronavirus Replication. *Viruses* **2019**, *11* (1), 73.
- (24) Toelzer, C.; Gupta, K.; Yadav, S. K. N.; Borucu, U.; Davidson, A. D.; Kavanagh Williamson, M.; Shoemark, D. K.; Garzoni, F.; Stauffer, O.; Milligan, R.; et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **2020**, *370* (6517), 725–730.
- (25) Toelzer, C.; Gupta, K.; Yadav, S. K. N.; Hodgson, L.; Williamson, M. K.; Buzas, D.; Borucu, U.; Powers, K.; Stenner, R.; Vasileiou, K.; et al. The free fatty acid binding pocket is a conserved hallmark in pathogenic coronavirus spike proteins from SARS-CoV to Omicron. *Science Advances* **2022**, *8* (47), No. ead9179.
- (26) Xu, C.; Wang, Y.; Liu, C.; Zhang, C.; Han, W.; Hong, X.; Wang, Y.; Hong, Q.; Wang, S.; Zhao, Q. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci. Adv.* **2021**, *7* (1), eabe5575.
- (27) Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKean, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; et al. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **2020**, *588* (7838), 498–502.

- (28) Benton, D. J.; Wrobel, A. G.; Roustan, C.; Borg, A.; Xu, P.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (9), No. e2022586118.
- (29) Staufer, O.; Gupta, K.; Hernandez Bücher, J. E.; Kohler, F.; Sigl, C.; Singh, G.; Vasileiou, K.; Yagüe Relimpio, A.; Macher, M.; Fabritz, S.; et al. Synthetic virions reveal fatty acid-coupled adaptive immunogenicity of SARS-CoV-2 spike glycoprotein. *Nat. Commun.* **2022**, *13* (1), 868.
- (30) Goc, A.; Niedzwiecki, A.; Rath, M. Polyunsaturated ω -3 fatty acids inhibit ACE2-controlled SARS-CoV-2 binding and cellular entry. *Sci. Rep.* **2021**, *11* (1), 5207.
- (31) Fonnesu, R.; Thunuguntla, V.; Veeramachaneni, G. K.; Bondili, J. S.; La Rocca, V.; Filippini, C.; Spezia, P. G.; Sidoti, M.; Plicanti, E.; Quaranta, P. Palmitoylethanolamide (PEA) Inhibits SARS-CoV-2 Entry by Interacting with S Protein and ACE-2 Receptor. *Viruses* **2022**, *14* (5), 1080.
- (32) Vivar-Sierra, A.; Araiza-Macias, M. J.; Hernández-Contreras, J. P.; Vergara-Castañeda, A.; Ramírez-Vélez, G.; Pinto-Almazán, R.; Salazar, J. R.; Loza-Mejía, M. A. In Silico Study of Polyunsaturated Fatty Acids as Potential SARS-CoV-2 Spike Protein Closed Conformation Stabilizers: Epidemiological and Computational Approaches. *Molecules* **2021**, *26* (3), 711.
- (33) Sofia F. Oliveira, A.; Shoemark, D. K.; Avila Ibarra, A.; Davidson, A. D.; Berger, I.; Schaffitzel, C.; Mulholland, A. J. The fatty acid site is coupled to functional motifs in the SARS-CoV-2 spike protein and modulates spike allosteric behaviour. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 139–147.
- (34) Warwicker, J. A model for pH coupling of the SARS-CoV-2 spike protein open/closed equilibrium. *Brief Bioinform* **2021**, *22* (2), 1499–1507.
- (35) Yurkovetskiy, L.; Wang, X.; Pascal, K. E.; Tomkins-Tinch, C.; Nyalile, T.; Wang, Y.; Baum, A.; Diehl, W. E.; Dauphin, A.; Carbone, C.; et al. SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain. *bioRxiv* **2020**, DOI: 10.1101/2020.07.04.187757.
- (36) Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G.-Y.; Katsamba, P. S.; Sampson, J. M.; Schön, A.; Bimela, J.; et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host & Microbe* **2020**, *28* (6), 867–879.e5.
- (37) Kadam, S. B.; Sukhrmani, G. S.; Bishnoi, P.; Pable, A. A.; Barvkar, V. T. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights. *J. Basic Microbiol* **2021**, *61*, 180.
- (38) Ma, J.; Su, D.; Sun, Y.; Huang, X.; Liang, Y.; Fang, L.; Ma, Y.; Li, W.; Liang, P.; Zheng, S. Cryo-electron Microscopy Structure of S-Trimer, a Subunit Vaccine Candidate for COVID-19. *Journal of Virology* **2021**, *95* (11), e00194–00121.
- (39) Shamseer, L.; et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ.* **2015**, *349*, g7647.
- (40) Schrödinger Release 2020-4: *Glide*; Schrödinger, LLC: New York, NY, USA, 2020.
- (41) Balunas, M. J.; Kinghorn, A. D. Drug discovery from medicinal plants. *Life Sciences* **2005**, *78* (5), 431–441.
- (42) Carino, A.; Moraca, F.; Fiorillo, B.; Marchianò, S.; Sepe, V.; Biagioli, M.; Finamore, C.; Bozza, S.; Francisci, D.; Distrutti, E. Hijacking SARS-CoV-2/ACE2 Receptor Interaction by Natural and Semi-synthetic Steroidal Agents Acting on Functional Pockets on the Receptor Binding Domain. *Frontiers in Chemistry* **2020**, *8*, 572885.
- (43) Riva, L.; Yuan, S.; Yin, X.; Martin-Sancho, L.; Matsunaga, N.; Pache, L.; Burgstaller-Muehlbacher, S.; De Jesus, P. D.; Teriete, P.; Hull, M. V.; et al. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **2020**, *586* (7827), 113–119.
- (44) Sharanya, C. S.; Sabu, A.; Haridas, M. Potent phytochemicals against COVID-19 infection from phyto-materials used as antivirals in complementary medicines: a review. *Future Journal of Pharmaceutical Sciences* **2021**, *7* (1), 113.
- (45) Creutzmacher, R.; Maass, T.; Veselkova, B.; Ssebyatika, G.; Krey, T.; Empting, M.; Tautz, N.; Frank, M.; Köbel, K.; Uetrecht, C.; et al. NMR Experiments Provide Insights into Ligand-Binding to the SARS-CoV-2 Spike Protein Receptor-Binding Domain. *J. Am. Chem. Soc.* **2022**, *144* (29), 13060–13065.
- (46) Day, C. J.; Bailly, B.; Guillon, P.; Dirr, L.; Jen, F. E.-C.; Spillings, B. L.; Mak, J.; Itzstein, M. v.; Haselhorst, T.; Jennings, M. P. Multidisciplinary Approaches Identify Compounds that Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2–ACE2 Receptor Interactions. *mBio* **2021**, *12* (2), e03681-20.
- (47) Wang, Q.; Meng, F.; Xie, Y.; Wang, W.; Meng, Y.; Li, L.; Liu, T.; Qi, J.; Ni, X.; Zheng, S.; et al. In Silico Discovery of Small Molecule Modulators Targeting the Achilles' Heel of SARS-CoV-2 Spike Protein. *ACS Cent Sci.* **2023**, *9* (2), 252–265.
- (48) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7* (1), 42717.
- (49) *LogD Predictor*; 2023.
- (50) Haber, S. L.; Benson, V.; Buckway, C. J.; Gonzales, J. M.; Romanet, D.; Scholes, B. Lifitegrast: a novel drug for patients with dry eye disease. *Ther. Adv. Ophthalmol* **2019**, *11*, 2515841419870366.
- (51) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
- (52) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
- (53) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **2009**, *10* (1), 168.
- (54) Gimeno, A.; Ojeda-Montes, M. J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; García-Vallvé, S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int. J. Mol. Sci.* **2019**, *20* (6), 1375.
- (55) Zhu, T.; Cao, S.; Su, P. C.; Patel, R.; Shah, D.; Chokshi, H. B.; Szukala, R.; Johnson, M. E.; Hevener, K. E. Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *J. Med. Chem.* **2013**, *56* (17), 6560–6572.
- (56) Durrant, J. D.; McCammon, J. A. NNScore 2.0: A Neural-Network Receptor–Ligand Scoring Function. *J. Chem. Inf. Model.* **2011**, *51* (11), 2897–2903.
- (57) Kolb, P.; Ferreira, R. S.; Irwin, J. J.; Shoichet, B. K. Docking and chemoinformatic screens for new ligands and targets. *Curr. Opin. Biotechnol.* **2009**, *20* (4), 429–436.
- (58) Bender, B. J.; Gahbauer, S.; Luttens, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balias, T. E.; Carlsson, J.; Irwin, J. J.; et al. A practical guide to large-scale docking. *Nat. Protoc.* **2021**, *16* (10), 4799–4832.
- (59) Skariyachan, S.; Gopal, D.; Muddebihalkar, A. G.; Uttarkar, A.; Niranjana, V. Structural insights on the interaction potential of natural leads against major protein targets of SARS-CoV-2: Molecular modelling, docking and dynamic simulation studies. *Comput. Biol. Med.* **2021**, *132*, No. 104325.
- (60) Azim, K. F.; Ahmed, S. R.; Banik, A.; Khan, M. M. R.; Deb, A.; Soman, S. R. Screening and druggability analysis of some plant metabolites against SARS-CoV-2: An integrative computational approach. *Informatics in Medicine Unlocked* **2020**, *20*, No. 100367.
- (61) El-Demerdash, A.; Hassan, A.; Abd El-Aziz, T. M.; Stockand, J. D.; Arafa, R. K. Marine Brominated Tyrosine Alkaloids as Promising Inhibitors of SARS-CoV-2. *Molecules* **2021**, *26* (20), 6171.
- (62) Piplani, S.; Singh, P.; Petrovsky, N.; Winkler, D. A. Identifying SARS-CoV-2 Drugs Binding to the Spike Fatty Acid Binding Pocket

Using In Silico Docking and Molecular Dynamics. *Int. J. Mol. Sci.* **2023**, *24* (4), 4192.

(63) Prajapat, M.; Sarma, P.; Shekhar, N.; Chauhan, A.; Kaur, G.; Bhattacharyya, A.; Avti, P.; Choudhary, G.; Bansal, S.; Sharma, S. Virtual screening and molecular dynamics simulation study of approved drugs as a binder to the linoleic acid binding site on spike protein of SARS-CoV-2 and double mutant (E484Q and L452R). *Indian J. Pharmacol* **2022**, *54* (6), 431–442.

Supplementary Material

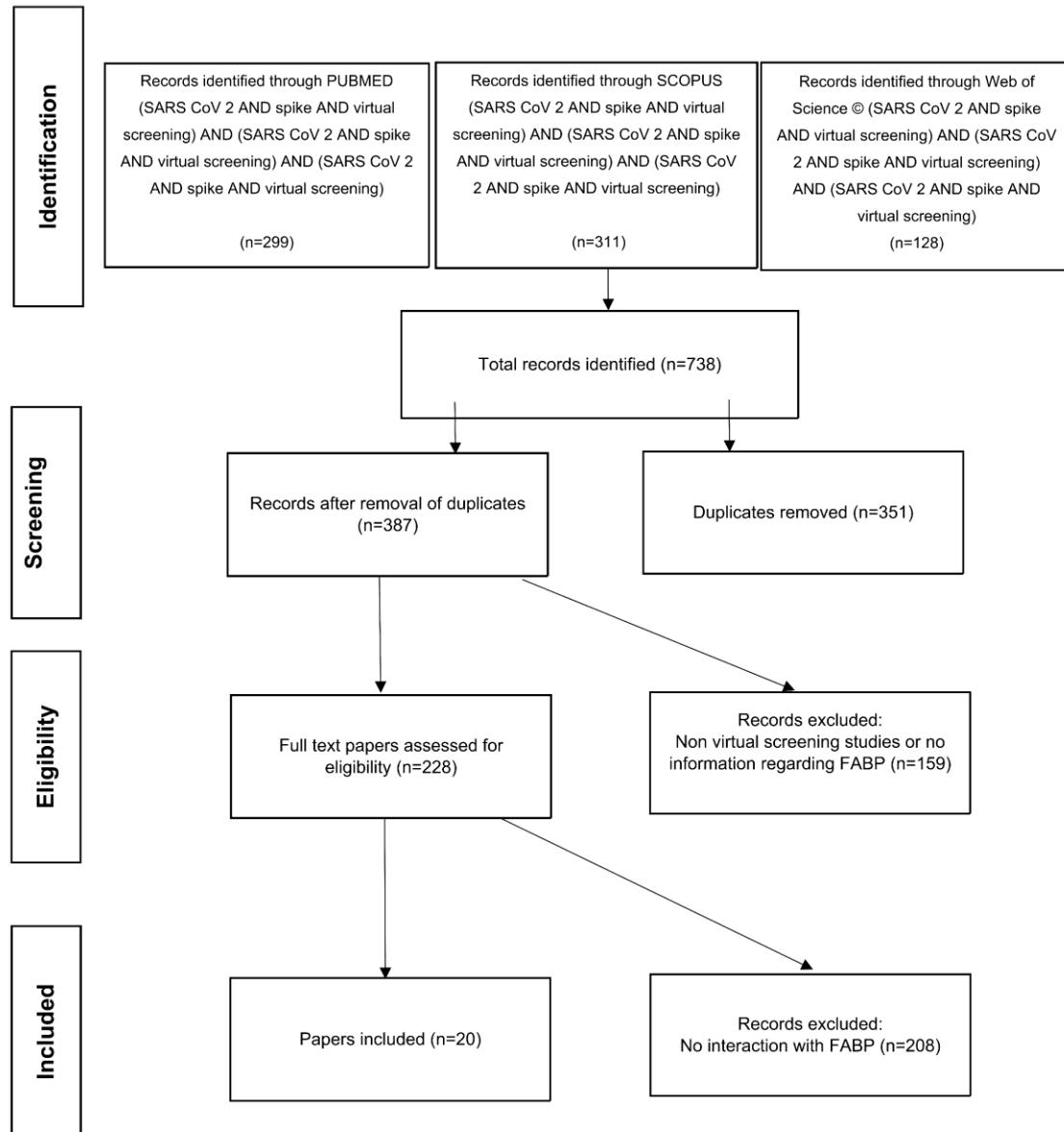


Figure S1 – PRISMA flow diagram for inclusion of studies

Table S1 – Chemical structures of all compounds reported (as SMILES)

Molecule	SMILES
Glyco-chenodeoxycholic acid	<chem>O[C@@H]1CC[C@]2(C)[C@@]3([H])CC[C@]4(C)[C@@H]([C@H](C)CCC(NCC(O)=O)=O)CC[C@@]4([H])[C@]3([H])[C@H](O)CC2C1</chem>
Ursodeoxycholic acid	<chem>O[C@H]1CC[C@]2(C)[C@@]3([H])CC[C@]4(C)[C@@H]([C@H](C)CCC(O)=O)CC[C@@]4([H])[C@]3([H])[C@@H](O)CC2C1</chem>
Tauro-ursodeoxycholic acid	<chem>O=S(O)(CCNC(CC[C@@H](C)[C@H]1CC[C@@]2([H])[C@]3([H])[C@@H](O)CC4C[C@H](O)CC[C@]4(C)[C@@]3([H])CC[C@]12C)=O)=O</chem>
Glyco-ursodeoxycholic acid	<chem>C[C@@H]([C@@]1([H])CC[C@]2([H])[C@]1(C)CCC3[C@@]2([H])[C@@H](O)C[C@@]4([H])[C@]3(C)C[C@@H](O)C4)CCC(NCC(O)=O)=O</chem>
Chenodeoxycholic acid	<chem>O[C@@H]1CC[C@]2(C)[C@@]3([H])CC[C@]4(C)[C@@H]([C@H](C)CCC(O)=O)CC[C@@]4([H])[C@]3([H])[C@H](O)CC2C1</chem>
Obeticholic acid	<chem>CCC1C2CC(CCC2(C3CCC4(C(C3C1O)CCC4C(C)CCC(=O)O)C)C)O</chem>
BAR 703	<chem>C[C@@]12C3CC[C@@]4([H])C(CC[C@@H]4[C@@H](CCCO)C)C3[C@@H](O)[C@@H](CC)[C@]1([H])C[C@H](O)CC2</chem>
Vitamin K	<chem>CC1=C(C(=O)C2=CC=CC=C2C1=O)CC=C(C)CCCC(C)CCCC(C)CCCC(C)C</chem>
Dexamethasone	<chem>O=C1C=C[C@]2(C)[C@@]3(F)[C@@H](O)C[C@]4(C)[C@@](O)(C(CO)=O)[C@H](C)C[C@@]4([H])[C@]3([H])CCC2=C1</chem>
Tretinoin	<chem>OC(/C=C(C)/C=C/C=C(C)/C=C/C1=C(C)CCCC1(C)C)=O</chem>
Tazarotene	<chem>CC1(C)C2=CC(C#CC3=CC=C(C(OCC)=O)C=N3)=CC=C2SCC1</chem>
Vitamin A	<chem>CC1=C(/C=C/C(C)=C/C=C/C(C)=C/CO)C(C)(C)CCC1</chem>
Vitamin D3	<chem>O[C@H](C/C1=C/C=C2C3CC[C@@H]([C@]3(CCC/2)C)[C@H](C)CCCC(C)C)CCC1=C</chem>
Linoleic acid	<chem>CCCCC/C=C\C/C=C\CCCCCCCC(O)=O</chem>
Myristic acid	<chem>CCCCCCCCCCCCCCC(O)=O</chem>
Stearic acid	<chem>CCCCCCCCCCCCCCCCCCC(O)=O</chem>
Oleic acid	<chem>CCCCCCCC=CCCCCCCC(=O)O</chem>
Lifitegrast	<chem>CS(=O)(=O)C1=CC=CC(=C1)CC(C(=O)O)NC(=O)C2=C(C=C3CN(CCC3=C2Cl)C(=O)C4=CC5=C(C=C4)C=CO5)Cl</chem>
Palmitic acid	<chem>CCCCCCCCCCCCCCCC(O)=O</chem>
Caprylic acid	<chem>CCCCCCCC(O)=O</chem>
SPC-14	<chem>O=C(CCCCC(O)=O)N(C1)CC2=C1C=CC([N+])([O-])=O=C2</chem>

Asiatic acid	<chem>CC1CCC2(CCC3(C(=CCC4C3(CCC5C4(CC(C(C5(C)CO)O)O)C)C)C2C1C)C)C(=O)O</chem>
Rotiorinol C	<chem>O=C1C(C(C)=O)=C2C=C3C=C(/C=C/C(C)=C/[C@@H](C)CCO)OC=C3[C@@H](O)[C@]2(C)O1</chem>
Psammaplysene D	<chem>O=C(N(CCCOC1=C(Br)C=C(CCN(C)C)C=C1Br)C)/C=C/C2=CC(Br)=C(OCCCN(C)C)C(Br)=C2</chem>
Hyosciamine	<chem>CN1C2CCC1CC(C2)OC(=O)C(CO)C3=CC=CC=C3</chem>
Oxabolone	<chem>CC12CCC3C(CCC4=C(O)C(CCC34)=O)C1CCC2O</chem>
Selexipag	<chem>CC(N(C1=CN=C(C2=CC=CC=C2)C(C3=CC=CC=C3)=N1)CCCCOCC(NS(=O)(C)=O)=O)C</chem>

3.2.2 Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study

Article

Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study

Luís Queirós-Reis ¹, Rui Alvites ^{1,2,3,4}, Ana Colette Maurício ^{1,2,3} , Andrea Brancale ⁵ , Marcella Bassetto ^{6,7} and João R. Mesquita ^{1,8,*} 

¹ Abel Salazar Institute of Biomedical Sciences (ICBAS), University of Porto, 4050-313 Porto, Portugal; up201205115@up.pt (L.Q.-R.); rmalvites@icbas.up.pt (R.A.); acmauricio@icbas.up.pt (A.C.M.)

² Animal Science Study Centre (CECA), University of Porto Agroenvironment, Technologies and Sciences Institute (ICETA), 4051-401 Porto, Portugal

³ Associate Laboratory for Animal and Veterinary Science (AL4AnimalS), 1300-477 Lisboa, Portugal

⁴ Advanced Polytechnic and University Cooperative, University Institute of Health Sciences (CESPU), Avenida Central de Gandra 1317, 4585-116 Gandra, Portugal

⁵ Department of Organic Chemistry, University of Chemistry and Technology Prague, 166 28 Prague, Czech Republic; andrea.brancale@vscht.cz

⁶ School of Pharmacy and Pharmaceutical Sciences, College of Biomedical and Life Sciences, Cardiff University, Cardiff CF10 3BN, UK; bassettom1@cardiff.ac.uk

⁷ Department of Chemistry, Faculty of Science and Engineering, Swansea University, Swansea SA2 8PP, UK

⁸ Epidemiology Research Unit (EPIunit), Institute of Public Health, University of Porto, 4050-091 Porto, Portugal

* Correspondence: jrmesquita@icbas.up.pt

Abstract: Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a respiratory virus that emerged in late 2019 and rapidly spread worldwide, causing the COVID-19 pandemic. The spike glycoprotein (S protein) plays a crucial role in viral target recognition and entry by interacting with angiotensin, converting enzyme 2 (ACE2), the functional receptor for the virus, via its receptor binding domain (RBD). The RBD availability for this interaction can be influenced by external factors, such as fatty acids. Linoleic acid (LA), a free fatty acid, has been shown to bind the S protein, modulating the viral infection by reducing initial target recognition. LA interacts with the fatty acid binding pocket (FABP), a potential drug target against SARS-CoV-2. In this study, we aimed to exploit the FABP as a drug target by performing a docking-based virtual screening with a library of commercially available, drug-like compounds. The virtual hits identified were then assessed in in vitro assays for the inhibition of the virus–host interaction and cytotoxicity. Binding assays targeting the spike–ACE2 interaction identified multiple compounds with inhibitory activity and low cytotoxicity.

Keywords: SARS-CoV-2; spike glycoprotein; virtual screening; in vitro assays



Academic Editor: Vincenzo Bramanti

Received: 24 October 2024

Revised: 19 December 2024

Accepted: 21 December 2024

Published: 27 December 2024

Citation: Queirós-Reis, L.; Alvites, R.; Maurício, A.C.; Brancale, A.; Bassetto, M.; Mesquita, J.R. Disrupting SARS-CoV-2 Spike Protein Activity: A Virtual Screening and Binding Assay Study. *Int. J. Mol. Sci.* **2025**, *26*, 151. <https://doi.org/10.3390/ijms26010151>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In late 2019, a new viral infection was detected in Wuhan, China, caused by a new coronavirus (CoV), Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). This virus rapidly spread worldwide in a few months, leading to the COVID-19 pandemic, infecting more than 750 million people and causing over seven million deaths [1]. SARS-CoV-2 is an RNA virus, belonging to the Coronaviridae family, originated in bats [2–4].

However, the course of the pandemic was changed by the introduction of vaccines against SARS-CoV-2 [5]. In fact, the recent development of vaccines was considered a powerful measure to save lives and minimize the impact on health, social systems, and global economics [6]. It is well known that SARS-CoV-2 genome mutations influence

the efficacy of the immune response induced by vaccination [5]. Since the beginning of the COVID-19 pandemic, numerous mutations of SARS-CoV-2 have been identified [5]. Periodic viral genomic sequencing helps to detect new genetic variants circulating in communities [5]. An updated version of the SARS-CoV-2 phylogenetic tree is shared on the GISAID platform (Global Initiative on Sharing Avian Influenza Data). A variant is recognized as a Variant of Concern (VOC) or Variant of Interest (VOI) by the World Health Organization (WHO) [5].

As overall observed in CoVs, cell entry is highly dependent on the spike glycoprotein (S protein), a structural surface protein, which recognizes the human angiotensin converting enzyme 2 (ACE2) [3,7]. The S protein is a trimeric fusion protein fully responsible for cell recognition and cell entry and is thus a main target for neutralizing antibodies, as well as therapeutics and vaccines [2,3]. The key area of interaction in each monomer of the S protein is the receptor binding domain (RBD) that interacts with ACE2 and initiates cell infection [7]. However, the RBDs have a dynamic nature, showing two conformations: a down conformation (inactive), inaccessible for interaction with ACE2, and an up conformation (active), required for target recognition, as well as neutralizing antibodies [3,8–10]. Each RBD can individually change between conformations, and the availability of RBDs can be influenced by external factors, such as the pH or the presence of ACE2 in the medium, as each binding event to ACE2 promotes further RBD change to the active conformation in the same S protein trimer (Figure 1) [8,11,12].

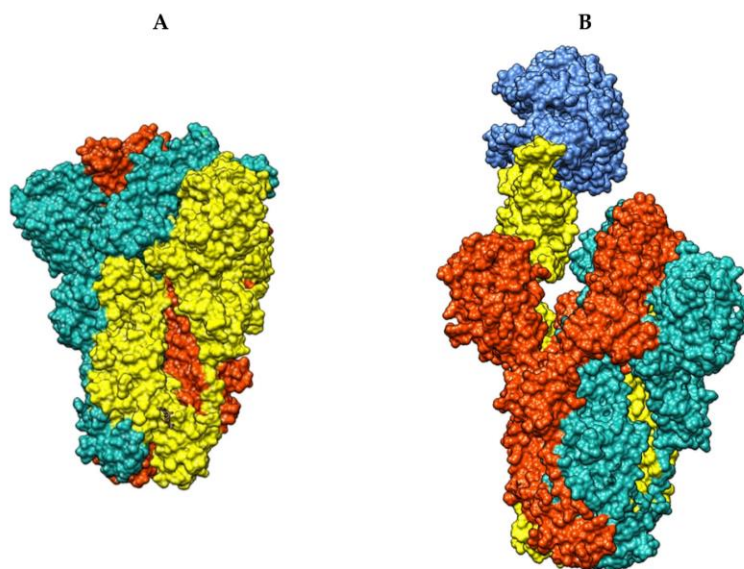


Figure 1. (A) Protein surface representation of the S trimer in a three RBD down conformation (blue, orange, yellow) (PDB ID: 7KMS). (B) Protein surface representation of the S trimer in a one RBD up (yellow) bound to ACE2 (dark blue), and two RBDs down (blue and orange) conformation (PDB ID: 8IOU).

Another external factor that can affect RBD accessibility to ACE2 recognition is the presence of free fatty acids (FFAs), since these affect the balance of open/closed RBDs. FFAs, such as linoleic acid (LA), are essential eicosanoid precursors and tissue inflammatory regulators [13], whose concentration is severely increased during infection and lung inflammation states [14]. LA affects the S protein and RBD behavior by binding a site within the protein, the fatty acid binding pocket (FABP) [14,15]. Physiologically, FFAs are proposed to function as molecular switches, enabling SARS-CoV-2 to adapt its immunogenicity to

inflammatory state by changing from a high infective status before a host immune response is established, to a reduced viral recognition and clearance state, increasing viral titers in inflammatory states [14]. Pharmacologically, this mechanism might be exploited by using compounds mimicking LA to permanently stabilize the S protein in an inactive conformation [15].

When LA is bound to the FABP, the stabilized inactive conformation hides the receptor binding motifs (RBM) between RBD interfaces, preventing interaction with ACE2, thereby reducing virus–host interactions, cellular recognition, and infection [15]. The FABP is formed between RBD pairs, being independent of the RBM. This structural arrangement results in three similar FABPs detected between S protein monomers (Figure 2) [15].

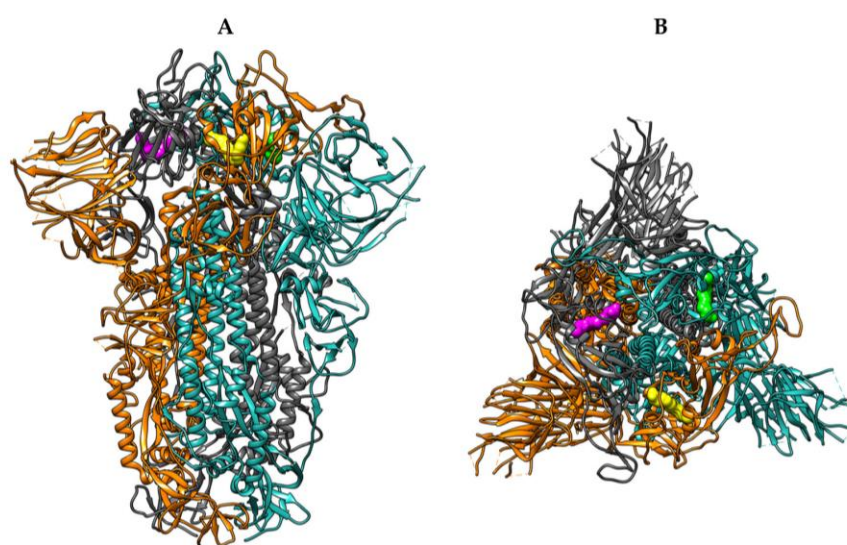


Figure 2. (A) Side view of the S protein in ribbon representation with the monomers represented as blue, red, and grey ribbons, along with the three FABPs in yellow, green, and magenta molecular surfaces (PDB ID: 6ZB5). (B) Top view of the S protein in ribbon representation with the monomers represented as blue, orange, and grey ribbons, along with the three FABPs in yellow, green, and magenta molecular surfaces (PDB ID: 6ZB5).

Structurally, the FABP features a bent hydrophobic pocket and a hydrophilic entrance. The hydrophobic pocket is lined with multiple phenylalanine residues from one RBD, while the entrance is formed by hydrophilic residues from the adjacent RBD. Each FABP can be divided into three main interaction areas: (1) a deep hydrophobic region, primarily formed by phenylalanine residues; (2) an intermediate region, characterized by phenylalanine and tyrosine residues that act as gating helices (Tyr365 and Tyr369); and (3) a hydrophilic entrance, with multiple residues from the adjacent RBD (Arg408, Gln409) [16]. LA can establish multiple interactions across the pocket, with the hydrocarbon chain interacting in deeper regions, while the carboxylic tail establishes electrostatic interactions with residues in the intermediate and entrance areas, locking the ligand in the final conformation (Figure 3) [15].

Overall, the FABP can affect the S protein and RBM exposure, although it does not interact directly, and therefore is not affected by the high frequency of mutations in this area of the protein [17]. On the other hand, the FABP can be detected in all SARS-CoV-2 variants, as well as in other highly pathogenic coronavirus [17].

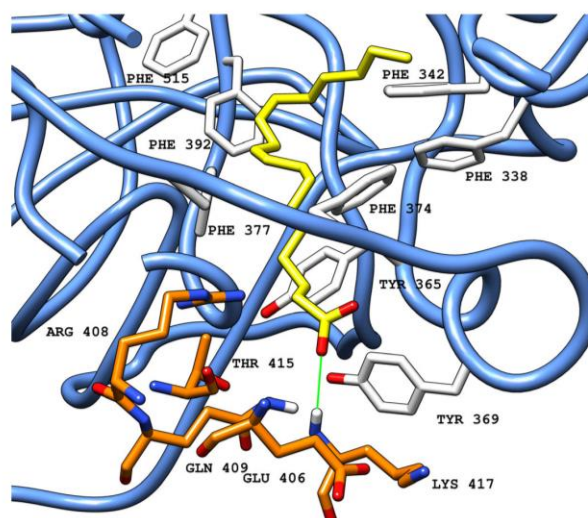


Figure 3. Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) in the FABP in a ribbon representation (blue), located between two adjacent RBDs (carbon atoms in white from RBD-B and carbon atoms in orange from RBD-C). Green solid lines represent polar interactions (hydrogen bonds) between the ligand and amino acid residues in the protein. All red atoms corresponds to oxygen atoms, while grey is due to shadows in the picture for depth perception.

In this study, we aimed to explore the biological effect of FFAs, shown by LA binding, using the FABP as a drug target to affect the virus lifecycle. A compound capable of binding the FABP and mimicking the effects of LA could significantly reduce the virus's ability to infect new cells, regardless of inflammatory state. The physiological role of the FABP may also present an additional barrier to resistance development, while the high conservation of this site in other highly pathogenic coronaviruses positions the FABP as a potential broad-spectrum anti-CoV target. To achieve this, a docking-based virtual screening was performed of a library of commercial, drug-like compounds, against the crystal structure of LA bound to the FABP. The virtual hits identified were then assessed in *in vitro* assays for the inhibition of the virus–host interaction and cytotoxicity.

2. Results

2.1. Identification of Compounds

To explore the FABP effects on the S protein, particularly the stabilization of an inactive conformation by a small molecule, the SPECS library of over 350,000 drug-like compounds was screened against the three FABPs in the crystal structure of LA bound to the S protein [15]. The virtual screening was performed with the Glide Standard Precision (SP) docking tool [18], with docking poses generated for each compound in each FABP, using LA molecules as centroids from SARS-CoV-2 Wuhan S protein (PDB ID: 6ZB5) [15]. Docking poses were then rescored with three scoring functions: Glide XP, CHEMPLP (PLANTS), and OpenEye (ScorePose) [19–21]. After applying an in-house optimized consensus scoring procedure [22], docking poses falling in the top 25% of the score range for all scoring functions were selected for visual inspection. A total of 5000 molecules for each binding site were chosen for visual inspection and combined into one set containing molecules common in the three selection sets. The selection of molecules was further reduced according to the predicted interactions, pharmacokinetics, and drug-like properties, resulting in 26 molecules (Figure 4). The final selection was purchased from SPECS and evaluated in S protein–ACE2 binding inhibition assays. Predicted poses shown are from the FABP formed by RBD-B and RBD-A.

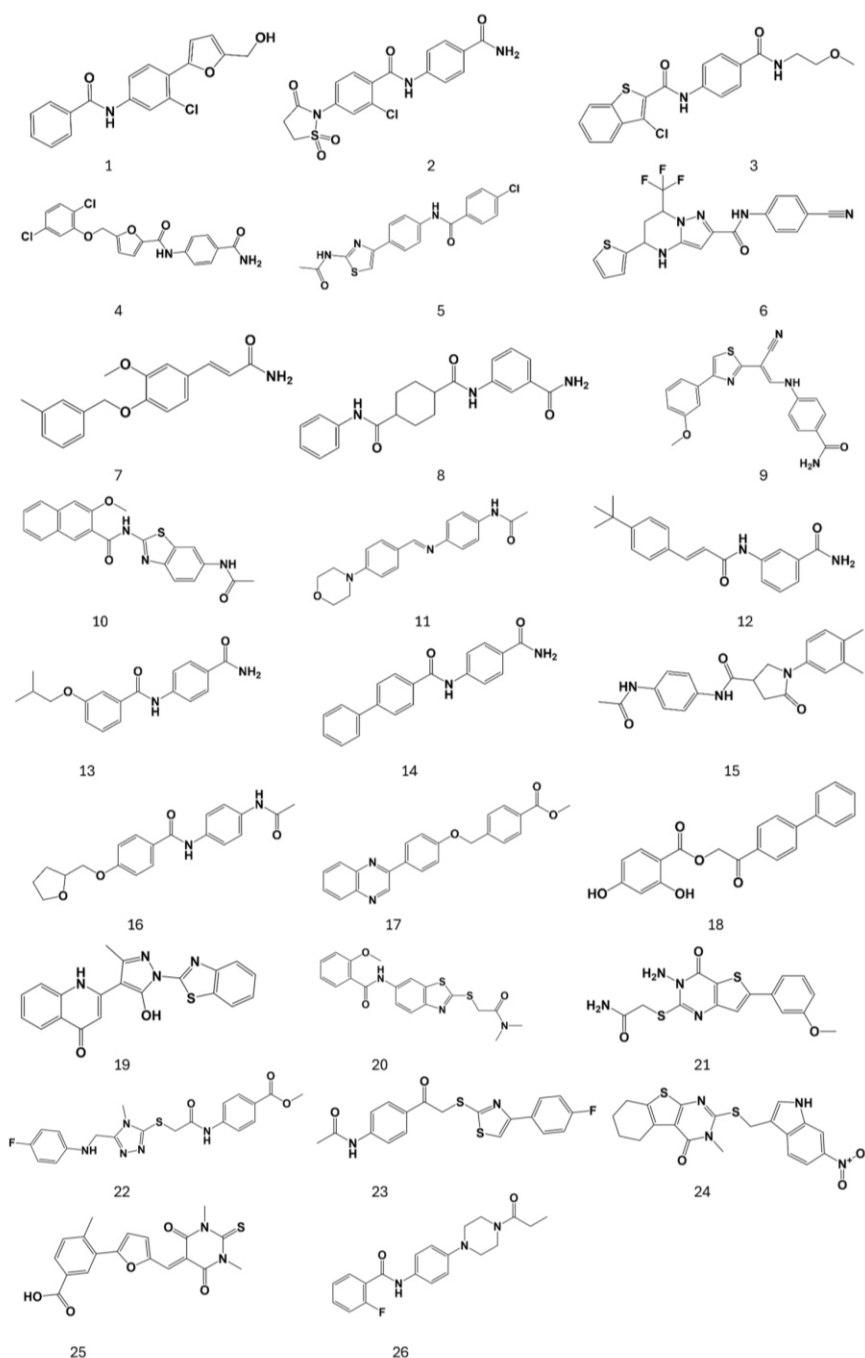


Figure 4. Chemical structure of compounds selected after the structure-based virtual screening and purchased from SPECS, identifying each compound with a number (1-26).

Overall, the presence of aromatic rings is the main feature observed in the selected compounds, along with highly hydrophilic groups in at least one terminal end of the molecule. Four representative compounds **19**, **20**, **24**, and **25** are shown (Figure 5), having the best activity profile among the screened compounds in our assays, detailed below.

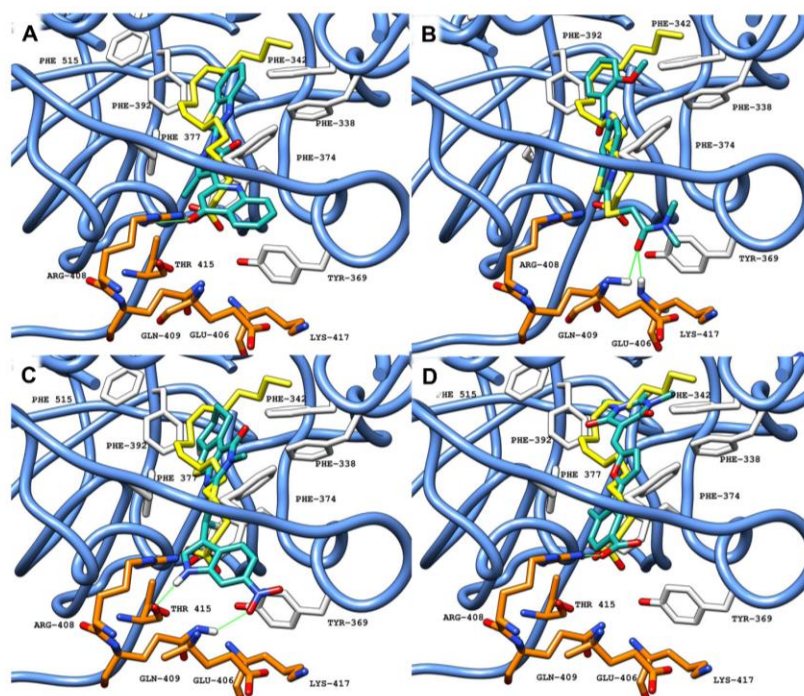


Figure 5. Molecular docking results obtained for **19** (A), **20** (B), **24** (C), and **25** (D), represented with carbon atoms in sea green, in the FABP of SARS-CoV-2 S protein (PDB ID: 6ZB5, blue ribbon, carbon atoms in white from RBD-1 and orange from RBD-2). Green solid lines represent polar interactions (hydrogen bonds) between the ligand and amino acid residues in the protein.

Using compound **20** as an example, binding is shown in the three FABP sites (Figure 6), consistent both regarding pocket occupation and main interaction residues.

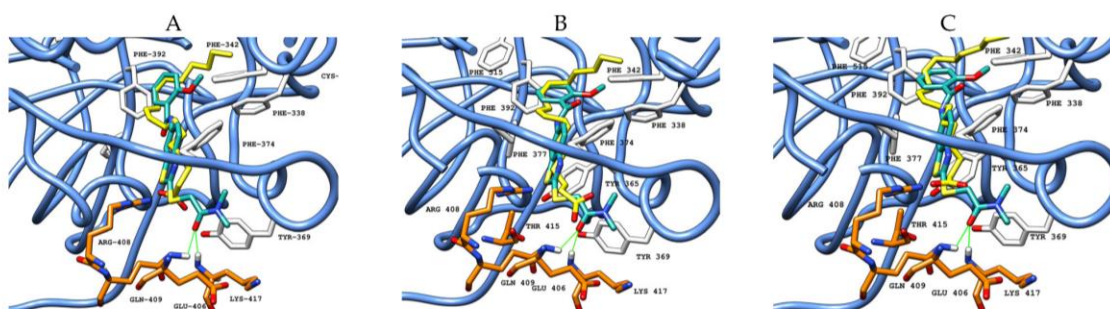


Figure 6. Molecular docking results obtained for **20** (carbon atoms in sea green) in the FABP of SARS-CoV-2 S protein Wuhan (PDB ID: 6ZB5, blue ribbon). (A) site 1 (RBD-B with carbon atoms in white, RBD-A with carbon atoms in orange); (B) site 2 (RBD-A with carbon atoms in white, RBD-C with carbon atoms in orange); (C) site 3 (RBD-C with carbon atoms in white, RBD-B with carbon atoms in orange). Green solid lines represent polar interactions (hydrogen bonds) between the ligand and amino acid residues in the protein.

2.2. In Vitro Validation Assays

The identified compounds' ability to block the virus–host interaction was tested using an ELISA-based inhibition assay, at 100 μ M, with activity compared with negative control (vehicle—1% dimethyl sulfoxide (DMSO)). LA and palmitoylethanolamide (PEA) were used as positive controls for inhibitory effect at 100 μ M, given that activity has been established for these compounds: LA has been shown to reduce RBD binding with ACE2

by 100% at 8.9 mM, and PEA reached ~50% inhibition of this interaction at 10 μ M [23,24]. At 100 μ M, test compounds inhibitory effect ranged from 0% to 34%, with seven compounds surpassing PEA activity (14%), although LA showed the strongest inhibitory activity (58%) (Figure 7).

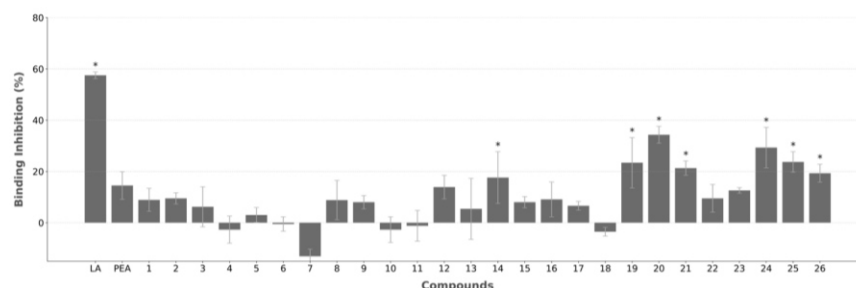


Figure 7. Inhibitory activity of the screened compounds tested at 100 μ M. Inhibition was determined as percentage calculated on the vehicle-treated cells (0.1% DMSO). The bars represent the mean $\pm$ SEM from three experimental repeats. *—Compounds with higher activity than PEA.

2.3. Cell Viability

A cytotoxicity evaluation of the 26 compounds was performed in L929 cells, fibroblastic-like cells extracted from mouse (*Mus musculus*) subcutaneous areolar and adipose tissue. This immortalized cell line provides fast, consistent, and uniform growth, facilitating acute and hyperacute toxicity testing, with high sensitivity to low concentrations. It is easy to culture and adapt to various conditions, widely used, and with predictive value validated in the literature [25]. Cytotoxicity was performed at 100 μ M, with two timepoints established and cell viability measured at 0 h and 48 h (Figure 8). In the virtual screening selection protocol, predicted toxicity was an important consideration, with ADME and PAINS (pan-assay interference compounds) analysis included to exclude potential toxic compounds (Supplementary Data file) [26].

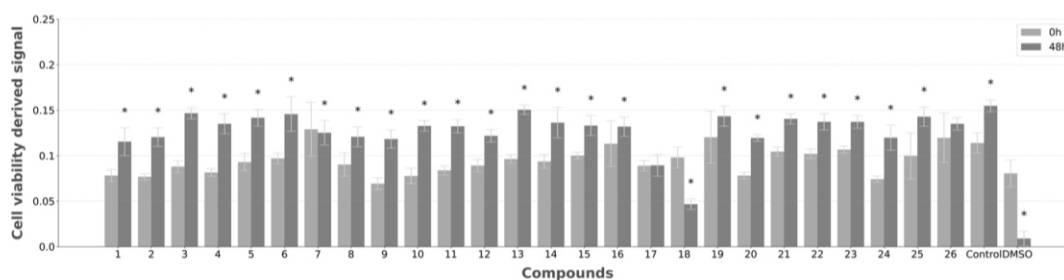


Figure 8. Cytotoxicity of the screened compounds at 100 μ M, measured with Presto BlueTM Viability Assay. Cytotoxicity variation at timepoints 4 h and 48 h was estimated against an initial measurement (0 h) before presence of screened compounds. The bars represent the mean $\pm$ SEM from three experimental repeats. *—Statistical significance difference between cell viability at both timepoints for each compound, determined by T-test for independent samples ($p < 0.05$).

Cell viability was measured with the Presto BlueTM viability assay method (Figure 8) [27]. In addition to the test compounds, two additional sets were tested, cells exposed only to growth medium (negative control), and cells treated with DMSO (positive control), since this compound is capable of altering cell membrane permeability and selectivity, justifying its use as evidence of cytotoxicity [28].

As observed in the negative control, in cells that were only exposed to growth medium (control), an increase in cell viability is reflected in a stronger signal detected. On the other

hand, DMSO-treated cells, acting as a cytotoxicity control, showed a stark decrease at 48 h, due to the expected cytotoxic effect of this substance. Apart from DMSO, compounds **7** and **18** showed a reduction in cell viability at this timepoint, while **17** and **26** showed a stabilization in cell growth (non-statistically significant differences (T-test for independent samples)). Each test group showed a statistically significant cell viability variation against DMSO-treated cells (one-way Anova with Dunnett's post hoc testing), as none of the compounds had a strong cytotoxic effect.

3. Discussion

From the initial set of 26 compounds, seven have shown the ability to affect the interaction more than PEA (14%), while none surpassed LA with 58% inhibitory activity. The maximum detected inhibition was 34% for compound **20**, with **14**, **19**, **21**, **24**, **25**, and **26** also surpassing PEA. Given that PEA has shown the ability to affect the virus life cycle through a reduction in the S protein–ACE2 interaction, the screened compounds might also harbor the potential to replicate this activity. In the presence of LA, the S protein is stabilized in an inactive conformation, with RBMs hidden, and the ability to interact with the human receptor severely reduced [15]. The previously reported inhibitory activity of LA was observed in these binding assays and, although the screened compounds showed weaker effects, they likely can affect the S protein behavior through the FABP. Additionally, the large variations in activity between compounds show a link to structural differences.

Overall, in the screened compounds, and unlike LA, large substituent groups are present in the deeper areas of the pocket, with aromatic rings deviating from the alkane chains in co-crystallized LA, potentially promoting aromatic interactions. This is particularly significant for **25**, where a six-membered ring with multiple heteroatoms shows pocket occupation similar to LA. Despite the fact that this large ring system is not aromatic, there are two carbonyl groups aligned with phenylalanine residues, overlapping the LA unsaturation, potentially establishing the similar stabilizing interactions. A carboxylic acid, as with LA, establishes a single H-bond with Arg408. While this compound shows some activity, the carboxylic acid is not the only terminal hydrophilic group in the active compounds (Figure 5D). The most active among screened compounds (LA achieves the highest activity), **20**, has a terminal amide group in this area, predicted to establish two H-bonds, as opposed to a single H-bond for LA, suggesting that it could be well suited to interact in this area.

However, the single H-bond established by LA is a charged H-bond, while this is not observed with the terminal amide group in **20**. Regarding the deeper areas of the pocket, while it is not as buried as LA, the terminal ring is predicted to overlap the unsaturation in LA. These two factors are likely the reason it was the best among the screened compounds, while still lacking when compared with LA (Figure 5B). Compound **24** has a nitro group, predicted as an H-bond acceptor, that extends to the hydrophilic area, and an H-bond predicted with the indole group (Figure 5C). Compound **19** has a bicyclic ring on both ends of the molecule, and groups capable of interactions in the intermediate portion of the FABP, adopting a conformation similar to LA (Figure 5A). However, **19** is predicted to establish a single H-bonds with Arg408, from the carbonyl group.

While the scoring functions and consensus protocol prioritized compounds based on predicted binding affinity, the ability to effectively engage with all three pockets served as an additional distinguishing factor. The pockets share the same amino acid composition, establishing similar interactions. Identifying compounds that exhibit strong and consistent binding across all three pockets was essential to validate the overall effectiveness of the virtual screening. Compound **20**, the strongest among screened compounds, adopts a

similar conformation in all pockets, as it establishes the same H-bonds, showing the ability to maintain key hydrogen bonds across all pockets (Figure 6).

Finally, the detected inhibitory activity observed for **22** and **23**, although limited (9% and 13%, respectively), could be associated with a feature unique to these compounds, among the screened selection. They both share a 4-fluorophenyl entity that is predicted to extend into the FABP and likely establishes aromatic interactions in this area, particularly with Phe338, Phe374, and Phe392, although **22** is predicted to be further buried in this area and to have significant overlap with LA (Figure 9).

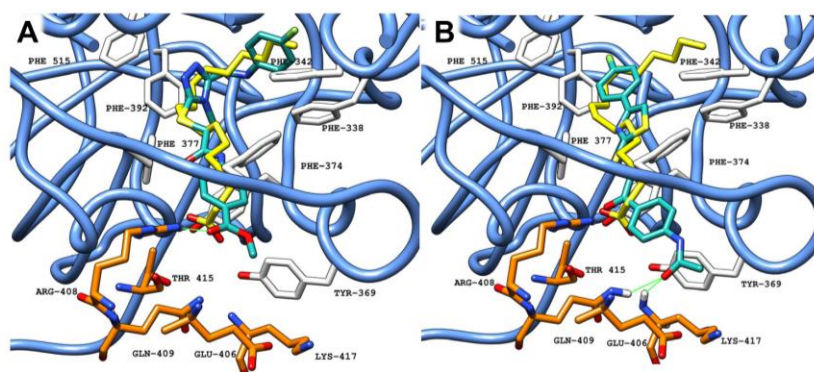


Figure 9. Molecular docking results obtained for **22** ((A), carbon atoms in sea green) and **23** ((B), carbon atoms in sea green) in the FABP of SARS-CoV-2 S protein (PDB ID: 6ZB5, blue ribbon and carbon atoms in white).

Furthermore, both compounds contain polar groups, such as an amide (**22**) and ketone (**23**), in the middle portion of the molecule, potentially capable of interacting with Tyr365 and Tyr369. An amide (**23**) and a methyl ester (**22**) are responsible for the stabilization in the pocket entrance, having predicted H-bonds with Gln409, Arg408, and Lys417.

Although strong cytotoxicity was not observed for any of the test compounds, **7** and **17** displayed weak cytotoxicity, maintaining the cell viability levels from the earlier timepoint (Figure 8). Cell viability for **18** had a strong reduction, although it did not reach the cytotoxicity levels showed by DMSO. During the virtual screening study, compound selection was performed considering toxicity predictions with the SWISS-ADME tool, and therefore strong cytotoxicity was less likely to arise [26]. This tool considers two complementary methods for pattern recognition, considering fragments with potential for cytotoxicity: pan assay interference compounds (PAINS), which show potent biological response irrespective of intended target, and Brenk et al.'s list of fragments or compounds that are putatively toxic, chemically reactive, metabolically unstable, or bear properties responsible for poor pharmacokinetics [26]. PAINS and Brenk warnings were, therefore, few by design, with only **21** showing a Brenk warning (Michael acceptor group) among the compounds showing cytotoxicity.

Overall, the initial virtual screening study targeting the S protein identified several compounds with activity against the virus–host interaction, reaching an inhibition of 34%, with no toxicity observed after 48 h. The reliability of computational methods, particularly virtual screening, is also observed in this study since, from an initial set of ~350,000 compounds, the selected molecules submitted to validity experimental assays resulted in four compounds with inhibitory activity of the target interaction over 25%. The observed effect against the S protein could be, at least partially, due to interactions in the FABP, stabilizing the inactive protein conformation, reducing binding between ACE2 and the S protein. Cytotoxicity studies provided important insights to exclude toxic compounds,

aiding the selection of the best candidate molecules for follow-up studies. The next stage for compound activity validation could involve conducting cell-based infection assays to assess antiviral effect and cytotoxicity, evaluating the ability to affect the virus–host interaction. This would be reflected in a reduced viral effect and increased cell viability due to reduced cell recognition and infection.

4. Materials and Methods

4.1. Virtual Screening

A library of commercially available drug candidates, the SPECS library, was screened against the fatty acid binding pocket using co-crystallized LA in the S protein (downloaded from the protein data bank PDB (<http://www.rcsb.org/> (accessed on 15 May 2023); PDB ID: 6ZB5). The structures of the compounds analyzed were built in MOE2019.10, saved in .sdf format, and prepared using the Maestro LigPrep tool by energy minimizing the structures (OPLS_2005 force field) and generating possible ionization states at pH 7 ± 2 , tautomers, all possible stereoisomers per ligand, and low-energy ring conformers. The protein was pre-processed with the MOE Protein Preparation tool, and the resulting protein–ligand complex was saved in .mae format and prepared using the Schrödinger Protein Preparation Wizard by assigning bond orders, adding hydrogens, and performing a restrained energy minimization of the added hydrogens using the OPLS_2005 force field. Additionally, the protein was also saved in .oedu format and .mol2 format to be used with scoring software CHEMPLP (PLANTS) 1.2 and ScorePose (OpenEye) 2.8.27.82.48.74, respectively [20,21]. The Glide Standard Precision virtual screening tool (SP) was used to virtually screen the commercial database against the binding site [15]. A 15 Å docking grid was prepared using the co-crystallized LA as the centroid, in parallel with the three FABPs in the S protein. The library was docked on the active sites using the Glide SP docking algorithm [18], keeping the default parameters, setting to three the number of output poses per input ligand to include in the solution, and performing a post-docking minimization of each of the poses kept. The output poses were saved as mol2 files. Docking poses obtained were then rescored (maintaining the identified pose) using Glide XP, CHEMPLP (PLANTS), and OpenEye (ScorePose) scoring functions [19–21]. Using a single docking program and scoring function might introduce potential bias, which justifies the use of three programs for rescoring. The values of each scoring function for each docking pose were then combined (consensus score) and only docking poses falling in the top 25% of the score value range for all the three scoring functions were selected for visual inspection in the three FABPs. The docking scores are reported in the Supplementary Data File (Tables S2–S4). The docking results were visually inspected in MOE 2022.02. The docking poses of the compounds obtained from the visual inspection were evaluated considering the following criteria: ability of a compound to adequately occupy the fatty acid binding site (similar to LA); and interactions predicted between compound and protein residues defining the site. Given that LA has been confirmed as a ligand and has shown antiviral activity in experimental assays, docked molecules were superimposed with a crystallographic structure of LA bound to the S protein (PDB accession code 6ZB5). In the next step, the set of molecules from each pocket was combined, with only molecules capable of good, predicted interactions in all pockets selected for further stages. Finally, the set of molecules identified for experimental validation was reduced to 26, by applying the Lipinski rule of five (selecting for good medicinal chemistry properties) and the SWISS-ADME webtool, to screen compound potential for toxicity (PAINS and BREAK analysis) [26].

4.2. Source of Small Molecules

All the compounds in this study, both the initial screened set and analogue compounds, were purchased from Specs Compound Handling B.V (Zoetermeer, the Netherlands). The library used for the virtual screening study was based on the Specs collection of screening compounds [29]. Molecular formulas (SMILES), molecular weight, and PAINS and BRENK analysis of tested compounds are reported in the Supplementary Data file.

4.3. Binding Assays

An Inhibitor Screening Assay Kit was used to screen inhibitors of the S–ACE2 interaction (BPS Bioscience Catalog # 78012) [30]. The kit includes the S protein in its native trimeric conformation from the Wuhan strain, providing the best physiologically relevant model for this interaction [31]. The assay kit also contains Biotinylated-ACE2, Streptavidin-HRP, and assay buffers. The assay procedure was performed as follows: SARS-CoV-2 S protein was first coated onto a 96-well plate. Following this, Biotinylated-ACE2 was incubated with the S protein on the plate, and Streptavidin-HRP was added to the plate. The interaction between Biotinylated-ACE2 and SARS-CoV-2 S protein was then detected using a colorimetric substrate. The resulting color change was quantified by measuring the absorbance using a UV–Vis microplate reader. Compounds were dissolved in DMSO and diluted until a testing concentration of 200 μ M, with each compound tested in triplicate. Finally, one negative control (vehicle–1% DMSO) and two positive controls were used, LA and PEA.

4.4. Cell Culture

L929 cells (a fibroblastic-like cell line derived from mouse connective tissue) used for determination of cytotoxicity were cultivated in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS, 1.5% penicillin G/streptomycin (P/S), and maintained in standard culture conditions, namely, in a humidified incubator at 37° and 5% CO₂. Cells were evaluated daily, and the culture medium was changed whenever necessary, with cell passaging performed when desirable confluences of 70–80% were observed.

4.5. Cytotoxicity—Presto Blue™ Viability Assay

The Presto Blue™ assay was performed to determine the cytocompatibility between the cellular system and test compounds [25,27]. This assay is based on a ready-to-use, commercially available water-soluble preparation and allows a live-cell evaluation. The resazurin solution was used to assess cell viability, based on the mitochondrial metabolism of this substance solution. Viable cells reduce the phenoxazine dye (resazurin), which results in color modification from blue to reddish over time that can be not only directly observed but also quantitatively measured using UV–VIS spectrophotometry, functioning as a cell viability indicator. L929 cells were seeded over a 96-well plate and maintained in incubation overnight (standard culture medium, 37 °C, 5% CO₂ environment, and 80% humidified atmosphere). To perform the Presto Blue™ assay, the culture medium was removed from each well at every timepoint (24, 28, and 96 h) and replaced by complete medium with 10% (v/v) of 10 Presto Blue™ cell viability reagent (Invitrogen, A13262, Thermo Fisher Scientific, Waltham, MA, USA). To perform the analysis, cells were incubated for 60 min at standard conditions to allow metabolism of the reagent. The supernatant medium was then collected and transferred to a 96-well plate and absorbance was read at 570 nm and 595 nm in a Multiskan™ FC Microplate Photometer (Thermo Scientific™, 51119000, Thermo Fisher Scientific, Waltham, MA, USA). Afterwards, wells were washed with Dulbecco's phosphate-buffered saline solution (DPBS, Gibco, 14190169) until the Presto Blue sediments were removed. Then, fresh culture medium was added to

each well, according to the timepoint specifications. Regular growth medium was used until the first timepoint (24 h), when it was replaced with DMEM (10% FBS, 1% P/S) supplemented with the test compound (determination of acute cytotoxicity). At the second timepoint (28 h), the medium was replaced with DMEM (2% FBS, 1% P/S) supplemented with test compounds, then left for 72 h until the last timepoint (96 h) (determination of acute cytotoxicity). For the Presto Blue assessment, both control group and test compounds were considered and, for each group, blank wells (without cell seeding), were included. The wavelength for excitation is 570 nm and, for emission, 595 nm. For that reason, the value obtained at 595 nm was subtracted from the value obtained for 570 nm (normalized value), for each well. In addition, the corrected absorbance for each experimental well, only considering seeded wells, was obtained by the subtraction of blank wells average from the normalized values of the respective sample group. The absorbance values were measured in triplicates. Data were further processed and normalized to the mean of the gold standard group and presented in a ratio between the 24 h timepoint, and both the 28 h and 92 h timepoints, representing variation against initial cell viability as a baseline. Statistical analysis was performed with a one-way Anova, with Dunnett's test post hoc.

5. Conclusions

SARS-CoV-2 is the third highly pathogenic coronavirus to infect humans, and although the COVID-19 pandemic has largely evolved to an endemic phase, the search for new antivirals is still relevant, both for current and potentially future coronavirus infections. Given the essential functions in the virus lifecycle, the S protein emerged as a target with strong potential. We explored a binding pocket in the S protein that stabilizes an inactive conformation, the fatty acid binding pocket, capable of affecting the virus lifecycle. Through an *in silico* virtual screening approach, 26 potentially active compounds were selected for experimental validation. Binding assays targeting the spike–ACE2 interaction revealed four compounds with activity over 25% and low cytotoxicity at 100 μ M. Further experimental testing, particularly cell-based infection assays, should be conducted to clarify the ability of the molecules to affect the virus lifecycle, derived from inhibitory activity in the virus–host interaction. These results highlight the reliability of computational techniques for discovering novel scaffolds and potentially bioactive compounds against a predetermined target in a rational methodology in drug discovery.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms26010151/s1>.

Author Contributions: Conceptualization, L.Q.-R., M.B. and J.R.M.; methodology, L.Q.-R., R.A. and M.B.; software, A.B., M.B. and J.R.M.; validation, A.C.M., M.B. and J.R.M.; formal analysis, M.B. and J.R.M.; investigation, L.Q.-R. and R.A.; writing—original draft preparation, L.Q.-R.; writing—M.B. and J.R.M.; supervision, A.C.M., A.B., M.B. and J.R.M. All authors have read and agreed to the published version of the manuscript.

Funding: L.Q.-R. would like to acknowledge Fundação para a Ciência e para a Tecnologia for the grant “2020.10230.BD” under the program “DOCTORATES 4 COVID-19”, and EEA Grants/Norway Grants for the grant “FBR_OC52_53”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

Acknowledgments: R.A. acknowledges the Centro de Estudos de Ciência Animal (CECA), Instituto de Ciências, Tecnologias e Agroambiente (ICETA), Porto University (UP), and Fundação para a Ciência e Tecnologia (FCT) for the funding and availability of all technical, structural, and human

resources necessary for the development of this work. His participation in this project was supported through project UIDB/00211/2020, merged by FCT/MCTES through national funds, and through project 2022.04501.PTDC (Olfabionerve—Olfactory Mucosa Mesenchymal Stem Cells and Biomaterials Promoting Peripheral Nerve Regeneration).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Coronavirus Disease (COVID-19) Dashboard. Available online: <https://covid19.who.int/> (accessed on 4 September 2021).
2. Naqvi, A.A.T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I.K.; Singh, A.; Atif, S.M.; Hariprasad, G.; Hasan, G.M.; Hassan, I. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. Acta Mol. Basis Dis.* **2020**, *1866*, 165878. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Wang, M.-Y.; Zhao, R.; Gao, L.-J.; Gao, X.-F.; Wang, D.-P.; Cao, J.-M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front. Cell Infect. Microbiol.* **2020**, *10*, 587269. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wrobel, A.G.; Benton, D.J.; Xu, P.; Roustan, C.; Martin, S.R.; Rosenthal, P.B.; Skehel, J.J.; Gamblin, S.J. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **2020**, *27*, 763–767. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Denaro, M.; Ferro, E.; Barrano, G.; Meli, S.; Busacca, M.; Corallo, D.; Capici, A.; Zisa, A.; Cucuzza, L.; Gradante, S.; et al. Monitoring of SARS-CoV-2 Infection in Ragusa Area: Next Generation Sequencing and Serological Analysis. *Int. J. Mol. Sci.* **2023**, *24*, 4742. [\[CrossRef\]](#)
6. Klimek, L.; Agache, I.; Cooke, E.; Jutel, M.; Akdis, C.A.; O’Hehir, R. COVID-19 vaccines—The way forward. *Allergy* **2022**, *77*, 15–16. [\[CrossRef\]](#)
7. Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292.e6. [\[CrossRef\]](#)
8. Benton, D.J.; Wrobel, A.G.; Xu, P.; Roustan, C.; Martin, S.R.; Rosenthal, P.B.; Skehel, J.J.; Gamblin, S.J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588*, 327–330. [\[CrossRef\]](#)
9. Almhedi, A.M.; Khoder, G.; Alchakee, A.S.; Alsayyid, A.T.; Sarg, N.H.; Soliman, S.S. SARS-CoV-2 spike protein: Pathogenesis, vaccines, and potential therapies. *Infection* **2021**, *49*, 855–876. [\[CrossRef\]](#)
10. Queirós-Reis, L.; Gomes da Silva, P.; Gonçalves, J.; Brancale, A.; Bassetto, M.; Mesquita, J.R. SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *Int. J. Mol. Sci.* **2021**, *22*, 10836. [\[CrossRef\]](#)
11. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S.J.; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [\[CrossRef\]](#)
12. Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G.Y.; Katsamba, P.S.; Sampson, J.M.; Schön, A.; Bimela, J.; et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **2020**, *28*, 867–879.e5. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Brash, A.R. Arachidonic acid as a bioactive molecule. *J. Clin. Investig.* **2001**, *107*, 1339–1345. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Staufer, O.; Gupta, K.; Hernandez Bücher, J.E.; Kohler, F.; Sigl, C.; Singh, G.; Vasileiou, K.; Yagüe Relimpio, A.; Macher, M.; Fabritz, S.; et al. Synthetic virions reveal fatty acid-coupled adaptive immunogenicity of SARS-CoV-2 spike glycoprotein. *Nat. Commun.* **2022**, *13*, 868. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Toelzer, C.; Gupta, K.; Yadav, S.K.; Borucu, U.; Davidson, A.D.; Kavanagh Williamson, M.; Shoemark, D.K.; Garzoni, F.; Staufer, O.; Milligan, R.; et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **2020**, *370*, 725–730. [\[CrossRef\]](#)
16. Queirós-Reis, L.; Mesquita, J.R.; Brancale, A.; Bassetto, M. Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein—Confirmed and Potential Ligands. *J. Chem. Inf. Model.* **2023**, *63*, 7282–7298. [\[CrossRef\]](#)
17. Toelzer, C.; Gupta, K.; Yadav, S.K.; Hodgson, L.; Williamson, M.K.; Buzas, D.; Borucu, U.; Powers, K.; Stenner, R.; Vasileiou, K.; et al. The free fatty acid-binding pocket is a conserved hallmark in pathogenic β -coronavirus spike proteins from SARS-CoV to Omicron. *Sci. Adv.* **2022**, *8*, ead9179. [\[CrossRef\]](#)
18. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. *Method and Assessment of Docking Accuracy*. *J. Med. Chem.* **2004**, *47*, 1739–1749.
19. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [\[CrossRef\]](#)

20. Korb, O.; Stützle, T.; Exner, T.E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96. [CrossRef]
21. OpenEye, OEDOCKING 4.3.1.0. Available online: <http://www.eyesopen.com>. (accessed on 12 September 2024).
22. Pasqualetto, G.; Schepelmann, M.; Varricchio, C.; Pileggi, E.; Khogali, C.; Morgan, S.R.; Boostrom, I.; Rozanowska, M.; Brancale, A.; Ferla, S.; et al. Computational Studies towards the Identification of Novel Rhodopsin-Binding Compounds as Chemical Chaperones for Misfolded Opsins. *Molecules* **2020**, *25*, 4904. [CrossRef]
23. Goc, A.; Niedzwiecki, A.; Rath, M. Polyunsaturated ω -3 fatty acids inhibit ACE2-controlled SARS-CoV-2 binding and cellular entry. *Sci. Rep.* **2021**, *11*, 5207. [CrossRef] [PubMed]
24. Fonnesu, R.; Thunuguntla, V.B.S.C.; Veeramachaneni, G.K.; Bondili, J.S.; La Rocca, V.; Filipponi, C.; Spezia, P.G.; Sidoti, M.; Plicanti, E.; Quaranta, P.; et al. Palmitoylethanolamide (PEA) Inhibits SARS-CoV-2 Entry by Interacting with S Protein and ACE-2 Receptor. *Viruses* **2022**, *14*, 1080. [CrossRef] [PubMed]
25. ISO 10993-5:2009; Biological Evaluation of Medical Devices—Part 5: Tests for In Vitro Cytotoxicity. International Organization for Standardization: Geneva, Switzerland, 2009.
26. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef] [PubMed]
27. Lall, N.; Henley-Smith, C.J.; De Canha, M.N.; Oosthuizen, C.B.; Berrington, D. Viability Reagent, PrestoBlue, in Comparison with Other Available Reagents, Utilized in Cytotoxicity and Antimicrobial Assays. *Int. J. Microbiol.* **2013**, *2013*, 420601. [CrossRef] [PubMed]
28. Gironi, B.; Kahveci, Z.; McGill, B.; Lechner, B.-D.; Pagliara, S.; Metz, J.; Morresi, A.; Palombo, F.; Sassi, P.; Petrov, P.G. Effect of DMSO on the Mechanical and Structural Properties of Model and Biological Membranes. *Biophys. J.* **2020**, *119*, 274–286. [CrossRef]
29. SPECS Library. Available online: <https://www.specs.net/> (accessed on 21 November 2024).
30. SARS-CoV-2 Spike Trimer (S1 + S2):ACE2 Inhibitor Screening Colorimetric Assay Kit. Available online: <https://bpsbioscience.com/sars-cov-2-spike-trimer-s1-s2-mutant-ace2-inhibitor-screening-colorimetric-assay-kit-79999> (accessed on 25 June 2024).
31. Wrapp, D.; Wang, N.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.-L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Supplementary Material

International Journal of
Molecular Sciences

Table 1 - Molecular formulas (SMILES), molecular weight, PAINS and BRENK analysis, and medicinal chemistry properties of tested compounds

Molecule	SMILES	MW	PAINS	BRENK	H-bond acceptors	H-bond donors	LogP
Linoleic acid		280.45	0	0	2	1	5.88
PEA		299.49	0	0	2	2	4.76
1		327.76	0	0	3	2	3.33
2		407.83	0	0	5	2	1.3
3		388.87	0	0	3	2	3.76

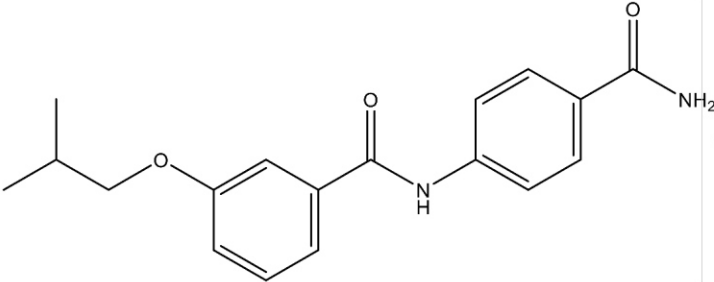
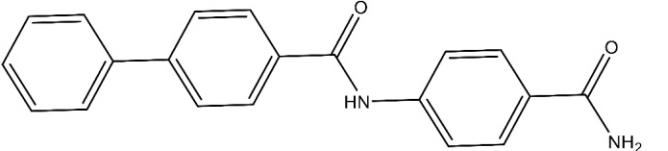
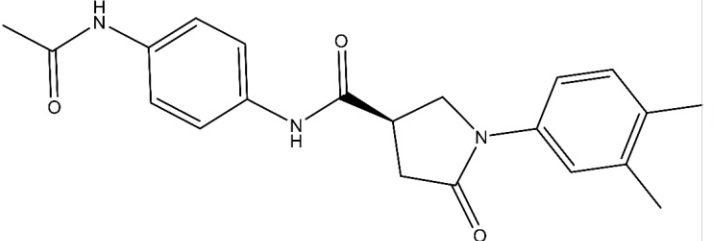


4		405.23	0	0	4	2	3.42
5		371.84	0	0	3	2	3.48
6		417.41	0	0	6	2	3.36
7		297.35	0	1	3	1	2.89
8		408.45	0	0	4	4	1.44

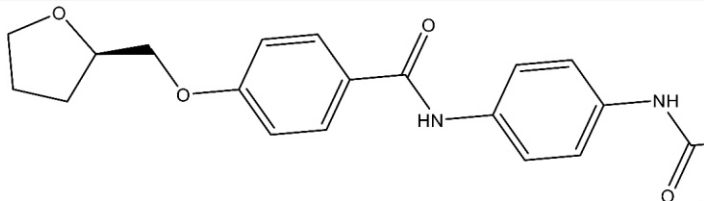
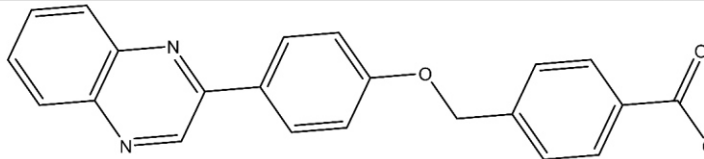
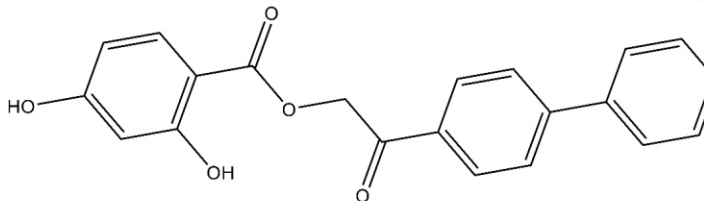
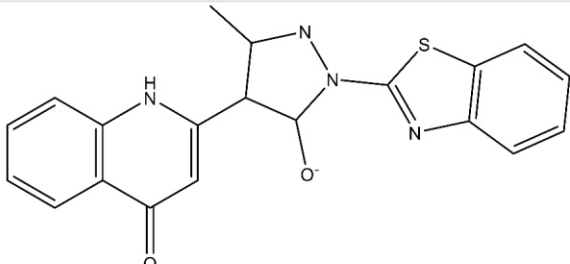


9		376.43	0	1	4	2	2.99
10		391.44	0	0	4	2	3.52
11		323.39	0	1	3	1	2.57
12		322.4	0	1	2	2	3.45

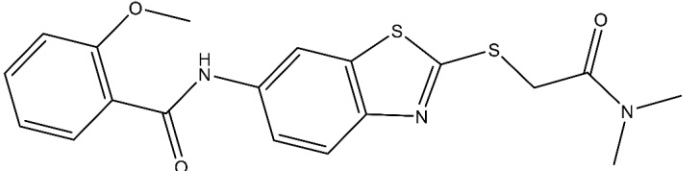
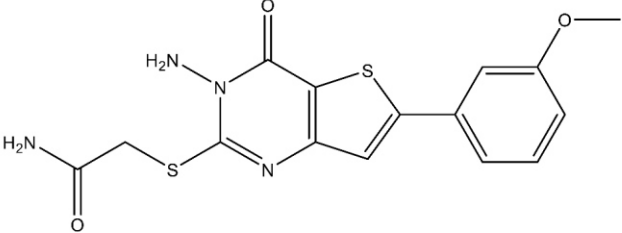
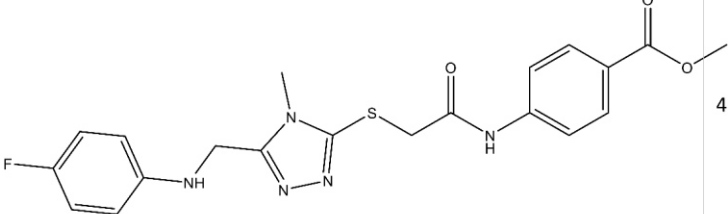
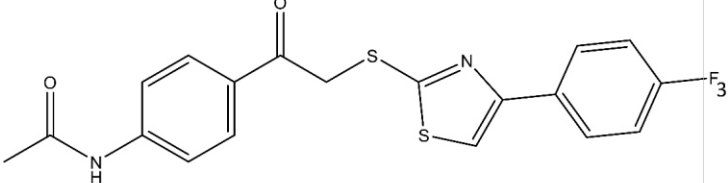


13		312.36	0	0	3	2	2.76
14		316.35	0	0	2	2	3.16
15		365.43	0	0	3	2	2.39



16		354.4	0	0	4	2	2.58
17		370.4	0	0	5	0	4.17
18		348.35	0	0	5	2	3.55
19		373.41	0	0	4	1	3.77



20		401.5	0	0	4	1	3.16
21		362.43	0	0	4	2	1.58
22		429.47	0	0	6	2	2.67
23		386.46	0	0	4	1	4.05



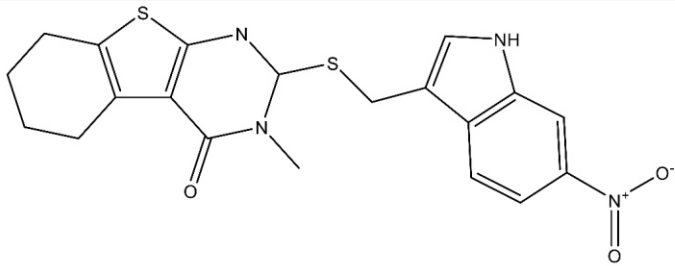
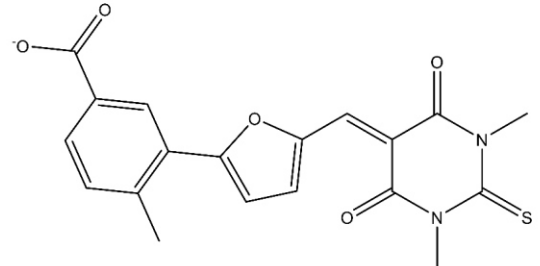
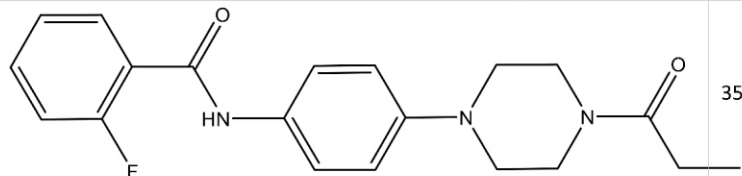
24		426.51	0	2	4	1	3.59
25		383.4	1	3	5	0	1.92
26		355.41	1	0	3	1	2.75

Table 2 - Docking score from the scoring functions used during rescoring in FABP site 1 (Glide XP, Plants and Score Pose)

Compound	Glide XP Score	Plants Score	Score Pose Score
1	-9.90934	-19.5896	-14.2772
2	-10.8551	-19.5701	-14.1237
3	-10.4287	-18.6332	-14.3903
4	-10.2246	-19.4036	-14.3926
5	-9.63088	-19.7809	-13.6593
6	-10.5832	-19.4332	-14.5225
7	-9.56014	-19.1504	-14.2456
8	-8.71374	-18.606	-15.5326
9	-10.4791	-19.4499	-14.8625
10	-10.7591	-19.6986	-15.0075
11	-10.5762	-18.5116	-14.2431
12	-10.8036	-19.0669	-15.048
13	-10.6837	-19.6938	-13.7471
14	-10.847	-19.9558	-14.9314
15	-9.84559	-19.3128	-14.5311
16	-11.1387	-19.7257	-15.2825
17	-11.0642	-19.1502	-15.1758
18	-10.8782	-18.5412	-16.8503
19	-9.40831	-19.8365	-13.6834
20	-9.88815	-18.5055	-15.2236
21	-9.69366	-18.5623	-13.7639
22	-9.86747	-19.1916	-15.3514
23	-8.91967	-18.5418	-16.1814
24	-11.064	-19.0385	-16.1967
25	-10.6393	-19.9047	-14.7905
26	-9.48367	-20.5197	-15.7748



Table 3 - Docking score from the scoring functions used during rescoring in FABP site 2 (Glide XP, Plants and Score Pose)

Compound	Glide XP Score	Plants Score	Score Pose Score
1	-9.60702	-84.0794	-14.316
2	-10.8539	-84.8198	-14.5393
3	-10.3137	-84.6865	-14.2819
4	-10.986	-96.7236	-16.302
5	-10.2643	-83.3837	-14.4076
6	-10.0921	-85.556	-15.1899
7	-9.69546	-84.4472	-14.2979
8	-10.9777	-87.697	-15.4979
9	-10.6759	-85.9965	-14.7041
10	-10.5987	-85.2048	-15.1432
11	-10.9371	-87.0268	-14.7195
12	-10.734	-84.8338	-15.2876
13	-10.5144	-87.0471	-13.7435
14	-10.6829	-88.7107	-15.1746
15	-9.49515	-84.4718	-14.8161
16	-10.9766	-93.3181	-15.1306
17	-12.254	-89.5624	-15.6337
18	-10.7716	-94.1199	-17.3446
19	-9.65984	-92.6943	-13.7081
20	-10.3088	-86.4465	-15.4393
21	-9.29881	-85.3654	-14.3168
22	-11.0059	-87.2998	-15.0335
23	-9.66039	-94.7483	-17.7919
24	-11.2866	-89.6218	-16.967
25	-10.6021	-84.3042	-15.0282
26	-9.93975	-91.2596	-16.8411

Table 4 - Docking score from the scoring functions used during rescoring in FABP site 3 (Glide XP, Plants and Score Pose)

Compound	Glide XP Score	Plants Score	Score Pose Score
1	-2.39452	-31.2631	-14.0827
2	-1.70741	-34.658	-13.7346
3	-0.793615	-35.306	-15.2858
4	-1.74717	-33.8839	-15.7387
5	-0.593194	-37.1896	-13.5115
6	-0.941544	-31.6506	-15.0051
7	-1.15977	-36.114	-13.413
8	-1.73428	-31.349	-14.6424
9	-0.766879	-34.3522	-14.6466
10	-2.10142	-35.3429	-14.6804
11	-2.85056	-32.8616	-14.3353
12	-1.71717	-34.0072	-15.2773
13	-1.87547	-33.1773	-13.7252
14	-1.97368	-34.4894	-15.0862
15	-1.15028	-33.4434	-14.0565
16	-0.93781	-33.8413	-15.1547
17	-1.79672	-36.7797	-16.7134
18	-1.24823	-36.4917	-17.4692
19	-0.986509	-44.2113	-13.449
20	-1.76133	-32.5783	-15.3749
21	-1.47461	-37.179	-13.8098
22	-0.716379	-32.5013	-14.5622
23	-1.87668	-42.4524	-15.2877
24	-0.96427	-41.0504	-16.3762
25	-0.814023	-39.0414	-14.0006
26	-0.495283	-31.3886	-15.8787

3.2.3 *In silico* discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding pocket

Submitted to ACS Omega

Under Revision

***In silico* discovery of a novel antiviral scaffold for SARS-CoV-2 targeting the spike glycoprotein through the fatty acid binding pocket**

Luís Queirós-Reis^{1*}, Mari Kaarbø², Huda Al-Baldawi³, Rui Alvites^{1,4,5,6}, Ana Colette Maurício^{1,4,5}, Andrea Brancale⁷, Marcella Bassetto^{8,9}, João R. Mesquita^{1,10}

1. Abel Salazar Institute of Biomedical Sciences (ICBAS), University of Porto, 4050-313 Porto, Portugal; up201205115@up.pt (L.Q.-R.)

2. Department of Microbiology, Oslo University Hospital, Oslo, Norway

3. Department of Microbiology, University of Oslo, Oslo, Norway

4. Animal Science Study Centre (CECA), University of Porto Agroenvironment, Technologies and Sciences Institute (ICETA), 4051-401 Porto, Portugal

5. Associate Laboratory for Animal and Veterinary Science (AL4AnimalS), 1300-477 Lisboa, Portugal

6. University Institute of Health Sciences (CESPU), Avenida Central de Gandra 1317, 4585-116 Gandra, Portugal

7. University of Chemistry and Technology, Prague 166 28 Praha, Czechia; andrea.brancale@vscht.cz

8. School of Pharmacy and Pharmaceutical Sciences, College of Biomedical and Life Sciences, Cardiff University, Cardiff, CF10 3BN, UK; bassettom1@cardiff.ac.uk

9. Department of Chemistry, Faculty of Science and Engineering, Swansea University, Swansea, SA2 8PP, UK

10. Epidemiology Research Unit (EPIunit), Institute of Public Health, University of Porto, 4050-091 Porto, Portugal

*Correspondence: L.Q.-R., up201205115@up.pt

Abstract

The key viral protein for infection by SARS-CoV-2 is the spike glycoprotein (S-protein), mediating entry into host cells, which therefore represents a strong focus for the development of targeted therapeutics. In this work, we have explored the fatty acid binding pocket within the S protein, which stabilizes an inactive conformation and disrupts cell recognition and infection. To explore the potential of this site as a drug target, molecular dynamics simulations were performed, followed by a docking-based virtual screening of commercial drug-like compounds. This *in-silico* procedure enabled the identification of potential inhibitors of SARS-CoV-2 cell infection, by stabilizing an inactive spike conformation. The antiviral effect of the virtual hits was analyzed in cell-based assays, and one molecule displayed low micromolar activity. Starting from the best antiviral compound found, structural analogues were purchased and evaluated in antiviral assays. An increase in activity was observed for multiple analogues, with the strongest antiviral compound showing sub-micromolar activity and low cytotoxicity. The successful identification of a new antiviral scaffold through *in silico* studies might pave the way for the further development of antivirals against SARS-CoV-2 and shows the reliability of the methodologies applied.

Introduction

Coronavirus disease 19, or COVID-19, is a respiratory infection caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), responsible for a global pandemic with more than 750 million people infected and seven million deaths ¹. Coronaviruses (CoVs) are RNA viruses, belonging to the *Coronaviridae* family, that infect a wide range of domestic and wild animals ²⁻⁵. The name “corona” is derived from a crown-like halo observed by microscopy, formed by three major structural proteins: the spike (S), membrane, and envelope proteins, projecting from the viral envelope ^{2,6,7}. CoV's cell recognition and infection are enabled by the densely glycosylated S protein, a trimeric fusion protein that, for SARS-CoV-2, recognizes the human target angiotensin converting enzyme 2 (ACE2) ⁸. Since the S protein facilitates viral entry into host cells, it is the main target for neutralizing antibodies, and it is a key focus in the development of therapeutics and vaccines. The S protein is composed of two subunits (S1 and S2), with S1 responsible for cell recognition and S2 containing the fusion machinery ^{8,9}. A particular portion of the S1 subunit, the receptor binding domain (RBD), interacts directly with the human receptor ACE2, and it is therefore a major epitope for neutralizing antibodies ^{4,9,10}. An RBD exists in each S protein monomer, and can have two possible conformations: a down conformation (inactive), inaccessible for interaction with ACE2, and an up conformation (active), available for target recognition ^{10,11}. Given the essential role of the S protein in infectivity, host range, and pathogenesis, any changes to its activity can have profound consequences for virus infection, with some mutations dramatically increasing the transmission and originating more dangerous viral strains ^{3,10}. On the other hand, vaccines targeted at the S protein have achieved significant success ^{10,12,13}, although no small-molecule inhibitors of the S protein have so far been approved for clinical use ¹⁴. Despite this, multiple compounds have shown the ability to affect the virus life cycle by targeting different sites in the S protein ¹⁵⁻²⁰. One such site is the Fatty acid Binding Pocket (FABP), which can bind fatty acids, particularly linoleic acid (LA) ^{18,21}. When LA binds the FABP, a stark change in S protein population dynamics is elicited, favoring the inactive conformation with all hidden RBDs ^{21,22}. The S protein is formed by three independent monomers and is able to interact with three ACE2 molecules. Therefore, three FABPs can be defined, formed between RBDs, and composed by two clearly defined hydrophilic and hydrophobic areas (Figure 1) ²¹. Critically,

the FABP is conserved in human CoVs and represents a potential target for pan-coronavirus activity ²³.

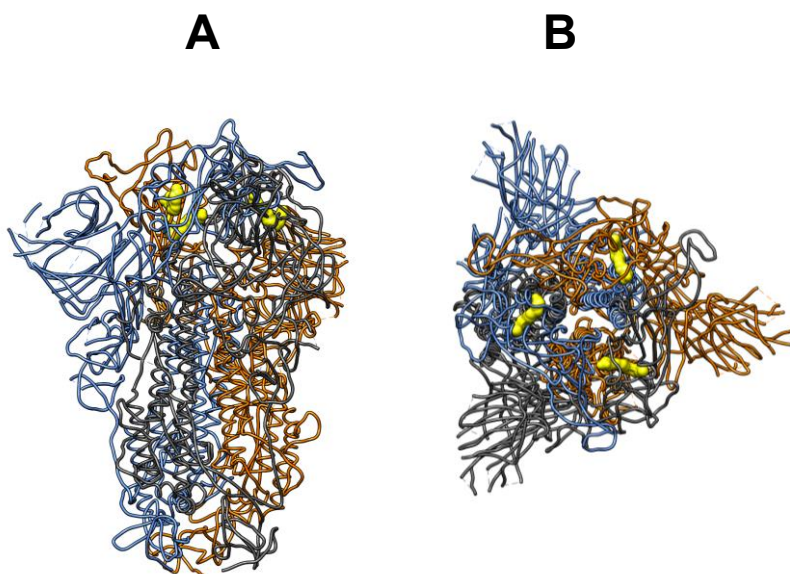


Figure 1 – A) Top view of the S Protein in ribbon representation with the monomers represented as blue, orange, and grey ribbons, along with the FABP in yellow surface (PDB ID: 6ZB5). B) side view of the S Protein in ribbon representation with the monomers represented as blue, orange, and grey ribbons, along with the FABP in yellow surface (PDB ID: 6ZB5). Abbreviations: S protein – Spike glycoprotein; FABP: Fatty Acid Binding Pocket; PDB - Protein Data Bank

These changes elicited by LA binding in the FABP significantly affect the virus lifecycle, by reducing the virus-host interaction and the virus' ability to infect new cells ¹⁸. Since this target site was discovered, research focused on the FABP has been significant, and multiple compounds have been identified capable of affecting virus-host interactions ²². However, in addition to LA, only lifitegrast and experimental compound SPC-14 have been confirmed to bind this pocket ^{16, 24}. Dexamethasone, multiple fatty acids, retinoids and liposoluble vitamins A and K have biological assays showing ability to affect the S protein – ACE2 interaction, and despite computational studies predicting FABP interaction, explicit binding was not confirmed ^{15, 25, 26}.

Regardless of the extensive research focused on the FABP, most confirmed bioactive compounds have limitations, such as fatty acids and retinoids, which display inadequate properties for translational or clinical applications. In this study, our main objective was to explore the FABP as a drug target, and to identify new bioactive scaffolds more suitable for drug development. To achieve this, the binding between the FABP and fatty acids was

analyzed by molecular dynamics, followed by a docking-based virtual screening of a library of commercial, drug-like compounds. The virtual hits identified were then assessed in *in vitro* inhibition assays (S-ACE2 interaction) and cell-based antiviral assays. An antiviral hit molecule was found with an EC₅₀ value in the low micromolar range. The subsequent evaluation of structural analogues enabled the identification of an antiviral compound with increased potency in the sub-micromolar range, and a preliminary evaluation of structure-activity relationships for this novel antiviral scaffold.

Results and Discussion

S protein dynamics and the FABP

To explore the FABP modulatory activity on the spike protein and key binding interactions in this site, the full-length 3D structure of the S protein in the presence and absence of LA (Protein Data Bank ²⁷ accession code PDB ID: 7DF3, 6VYB, 6ZB5 ^{3, 21, 28}) was analyzed via Molecular Dynamics (MD). Three different simulations were run in triplicate for the S protein in the open conformation, closed conformation, and closed conformation with bound LA. Analysis of Root Mean Square Deviation (RMSD) shows geometric convergence in MD simulations for all S protein systems at 100 ns (Figure 2). The open conformation showed the highest stability, while the presence of LA stabilized the closed conformation, achieving an intermediate stabilization level.

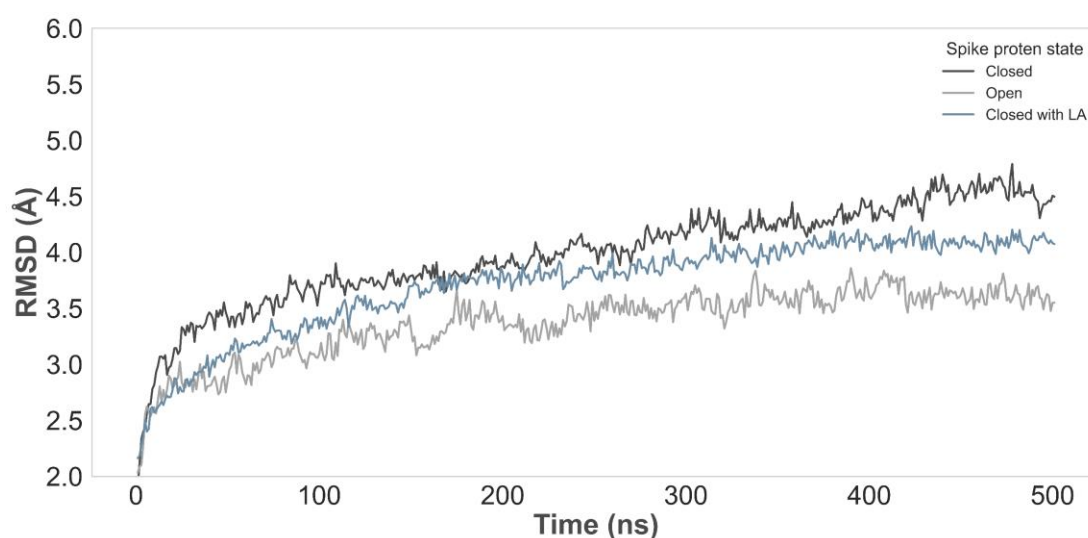


Figure 2 – All-atom RMSD of the S protein in the closed conformation, closed conformation with LA bound and open conformation (PDB accession codes 7DF3, 6ZB5 and 6VYB)

respectively). Abbreviations: RMSD – Root Mean Square Deviation; S protein – Spike glycoprotein; LA - linoleic acid; PDB - Protein Data Bank.

This stabilization is likely linked to the pattern of Root Mean Square Fluctuation (RMSF): the S1 subunit, containing the RBD, showed reduced fluctuation, while residues in the S2 subunit of the S protein were destabilized (Figure 3A). The most stabilized residues are concentrated near the FABP (Figure 3B), including Phe374, Ser375, Arg408, Thr415, Gly416 and Ile418, with some residues included in the RBD (Gly496, Phe497, Tyr505, Gln506). When LA binds the FABP, it directly stabilizes interacting residues, and indirectly, additional residues included in the RBD, therefore increasing the frequency of the closed conformation in the population of S proteins. On the other hand, the S2 subunit is destabilized, with multiple residues between Ser659 and Ala1065 with increased fluctuation, and particularly residue Asp614, which is critical for the S2 subunit stability ²⁹. This stabilization pattern is likely linked to the increase frequency of the S protein in the down conformation in the presence of LA, affecting binding and cell infection ²¹.

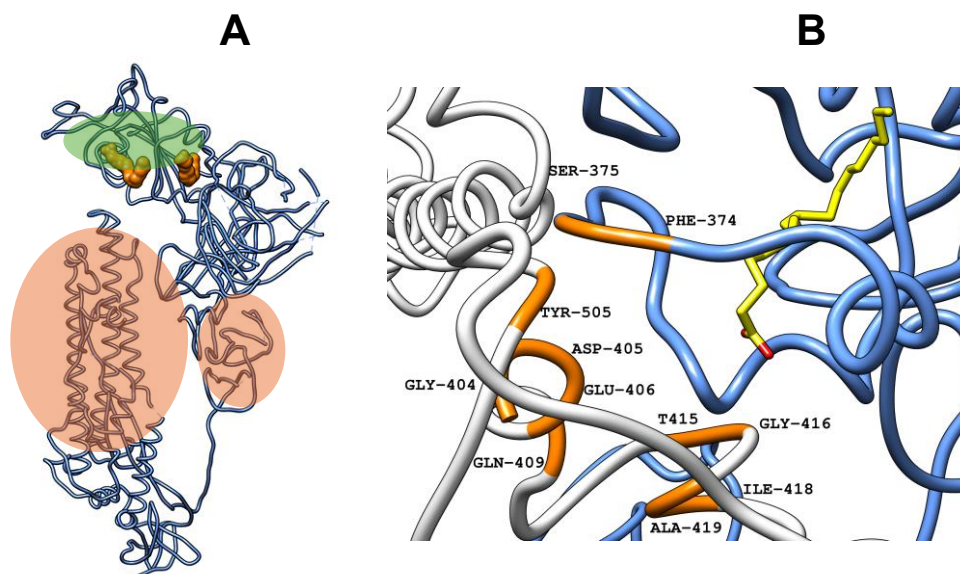


Figure 3 – A) Ribbon representation of a single chain in the trimeric S protein structure (PDB ID: 6ZB5). Two FABPs are represented by orange molecular surfaces. Residues highlighted with a green circle correspond to reduced fluctuation, while residues highlighted with red circles are linked to increased residue movement. B) A) Co-crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) and FABP in ribbon representation (blue and white ribbons corresponding to separate RBDs), with the stabilized residues nearby LA (orange ribbon) ³⁰. Abbreviations: S protein – Spike glycoprotein; PDB - Protein Data Bank; FABP - Fatty Acid Binding Pocket; LA - linoleic acid.

Additionally, the MOE Site Finder tool was applied to evaluate the stability and availability of FABP for ligand binding across the MD simulations ³¹. This tool detected possible pockets in the S protein, which were matched with the residues forming the FABP (Figure 4) ²¹.

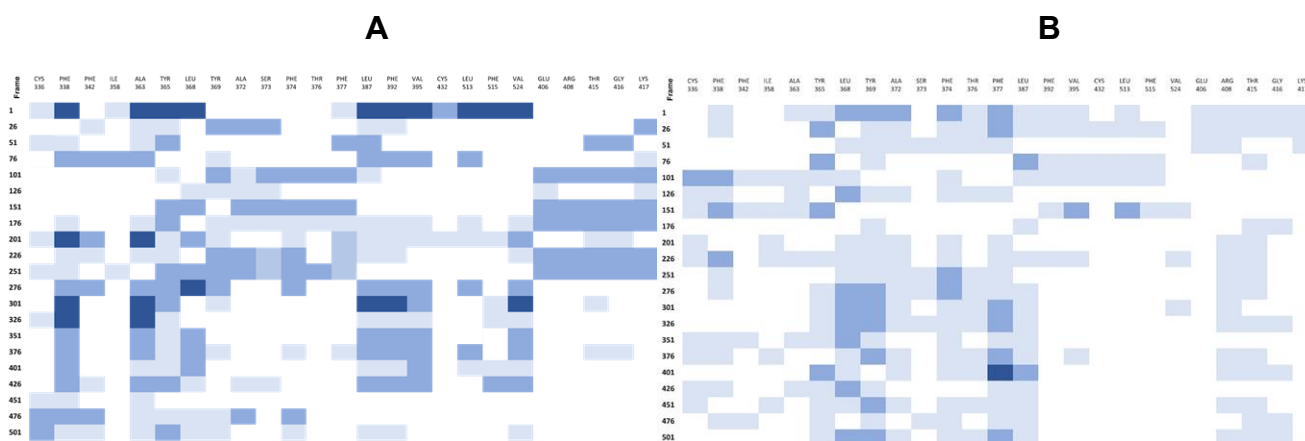


Figure 4 – Schematic representation of residues in pockets detected by Site Finder module in selected frames from MD of the closed conformation (A) and open (B). The color intensity corresponds to detection of each FABP residue in zero, one, two or three pocket for each frame ³¹. Abbreviations: MD – Molecular Dynamics; FABP - Fatty Acid Binding Pocket.

Overall, the site finder analysis revealed that the pocket is accessible across the MD in both the open and closed conformations of the S protein. However, in the closed conformation, the pockets identified in each frame are present in all three spike monomers and contain more FABP residues, suggesting that ligand binding is easier in this state. Despite this, the dynamic nature of the pocket shows available pockets even in open conformation, whose availability for potential binding varies across the simulation. Nevertheless, MOE Site Finder always detects a FABP in a spike monomer in every frame. These findings indicate that the FABP is available for binding, both in the open and closed conformation, and that binding stabilizes the S protein, particularly the RBD in the inactive conformation. Importantly, the hydrophilic area is only intermittently detected, although this is the area where the strong electrostatic interactions that stabilize the ligand occur, highlighted during the MD with residues Glu406, Arg408, Thr415, Gly416, Lys417. However, this analysis showed that the hydrophobic region is critical for initial binding, with residues Cys336, Phe338, Phe342, Ile358, Ala363, Tyr365, Leu368, Tyr369, Ala372, Ser373, Phe374, Thr376, Phe377, Leu387, Phe392, Val395, Cys432, Leu513, Phe515 maintained during the MD.

Hence, new binding scaffolds should promote more interactions in the hydrophobic area, perhaps exploring the aromatic nature of multiple phenylalanine residues that line the deeper areas of the pocket (Figure 5).

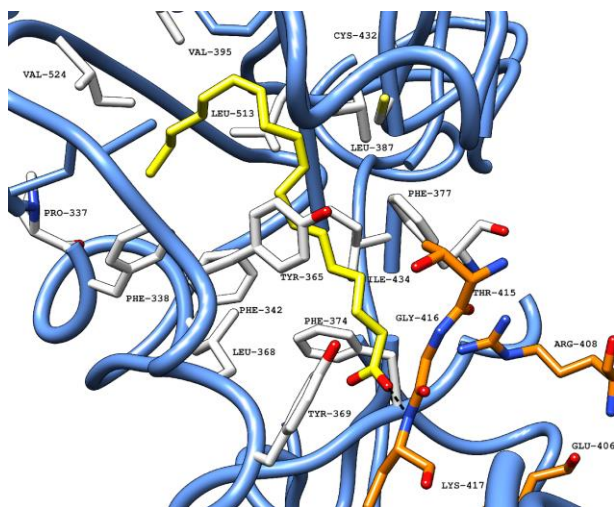


Figure 5 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in yellow) in the FABP in a ribbon representation (blue), located between two adjacent RBDs. The pocket is formed by the hydrophobic area (carbon atoms in white) and hydrophilic area (carbon atoms in orange). Black dashed lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein ³¹. Abbreviations: LA - Linoleic Acid; PDB - Protein Data Bank; RBD – Receptor Binding Domain.

Docking-Based Virtual Screening

In order to explore the FABP modulatory effects on the spike protein behavior, a docking based virtual screen was performed, to identify small molecules capable of binding, and therefore stabilize an inactive spike conformation. The crystal structure of LA bound to the S protein was used to screen the Enamine Screening Collection library of over 4,000,000 drug-like compounds ³².

The Glide high-throughput virtual screening tool (HTVS) ³³ was employed to virtually screen the database, against each FABP, with the top 50,000 molecules subsequently re-docked with Glide SP (standard precision).

To avoid potential bias introduced by any single docking program, the docking results (docking poses) were rescored with three scoring functions, Glide XP, CHEM_{PLP} (PLANTS) and OpenEye (ScorePose) ³⁴⁻³⁶. After applying an in-house optimized consensus scoring procedure, 3000 molecules for each binding site were chosen for visual inspection ³⁷. Combining the top molecules with best predicted interactions in the three pockets, along with analysis of drug-like properties, resulted in a final selection of 18 molecules (Figure 6), which

were purchased from Enamine and evaluated in binding assays and cell-based antiviral assays. As an example, the predicted binding for compound **1** is superimposed with LA in its binding site in Figure 7.

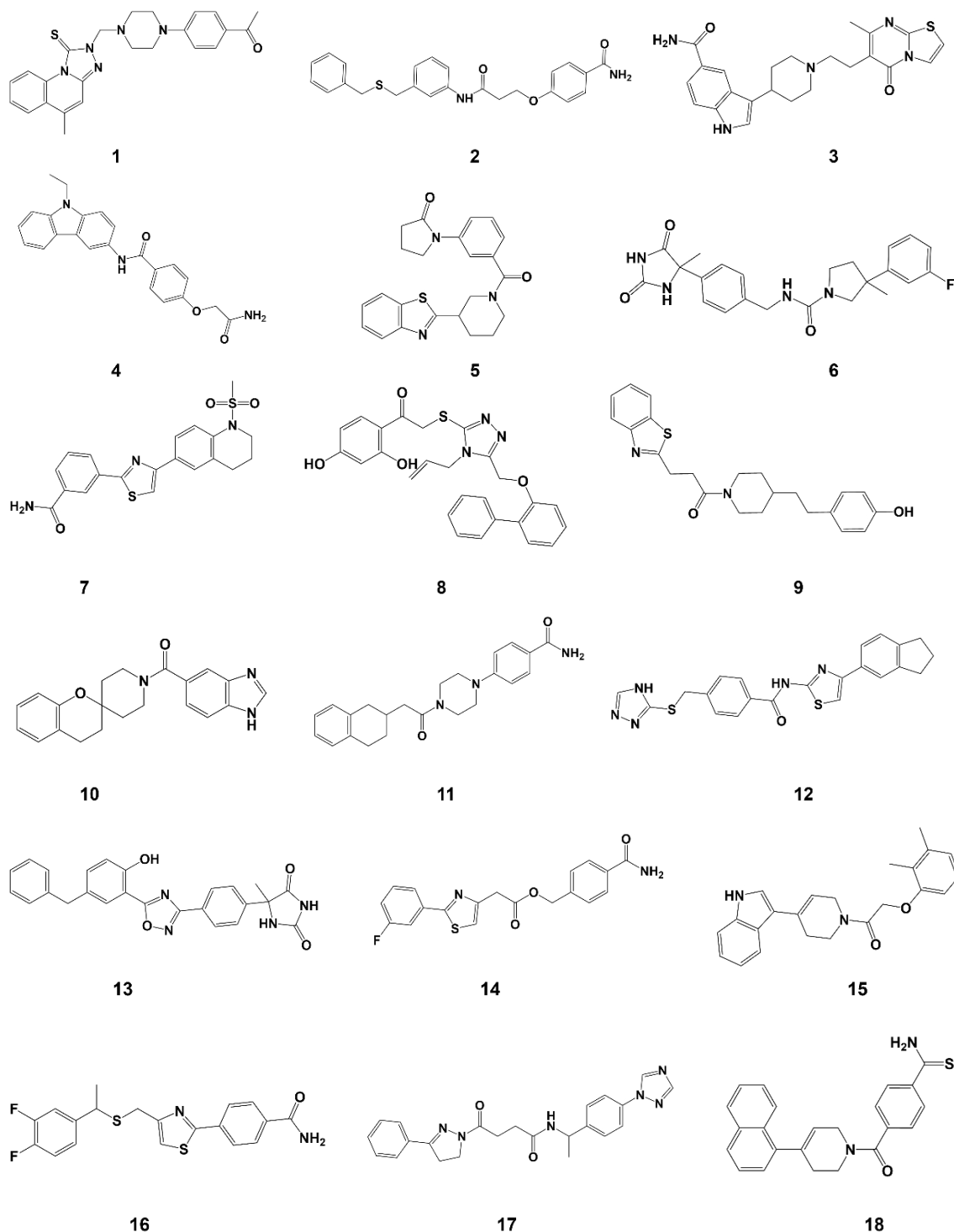


Figure 6 – Chemical structure of compounds selected after the structure-based virtual screening and purchased from Enamine.

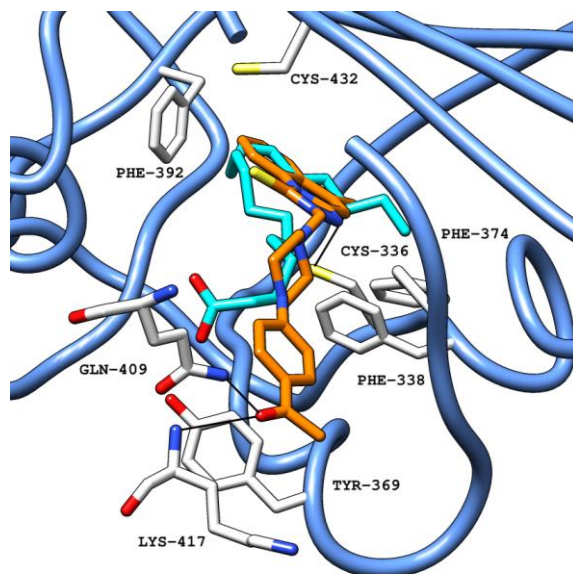


Figure 7 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in light blue) superimposed with docking pose for compound **1** (carbon atoms in orange) in the FABP in ribbon representation (blue), obtained with GlideSP. Black lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein ³⁰. Abbreviations: LA: Linoleic Acid; PDB - Protein Data Bank; FABP – Fatty Acid Binding Pocket.

Overall, the 18 selected compounds were predicted to achieve good pocket occupation and interactions with buried hydrophobic residues, along with multiple electrostatic and H-bond interactions with hydrophilic residues in the entrance. The main difference when compared with LA was the frequent presence of aromatic groups, against alkane chains in LA. The presence of aromatic features is shared with previously identified FABP bioactive molecules, such as lifitegrast, retinoids or SPC-14, with our molecular dynamics studies also pointing to possible stronger binding by exploring the hydrophobic area of the pocket. When compared with LA, compound **1** also extended further to the hydrophilic area and established direct contacts with the neighboring residues. Other selected compounds showed variability in pocket occupation, with varying focus in interactions with either hydrophobic or hydrophilic residues.

Biological assays

Binding assays

An ELISA-based inhibition assay was used to assess if the compounds selected inhibit the S-ACE2 interaction. The compounds were tested at 200 μ M and their activity was compared with positive and negative controls (vehicle – 2% dimethyl sulfoxide (DMSO)). LA was selected as a positive control due to its established inhibitory effect on the S-ACE2 interaction, showing 100% inhibition of the RBD-ACE2 binding at a concentration of 8.9 mM in in vitro assays.¹⁸ Additionally, palmitoylethanolamide (PEA) was also used as a positive control, as it has been confirmed to reduce RBD binding with ACE2 by ~50%¹⁷. At 200 μ M, six compounds had stronger inhibitory activity than PEA (17% inhibition), but none reached LA inhibition (84% inhibition) (Figure 8).

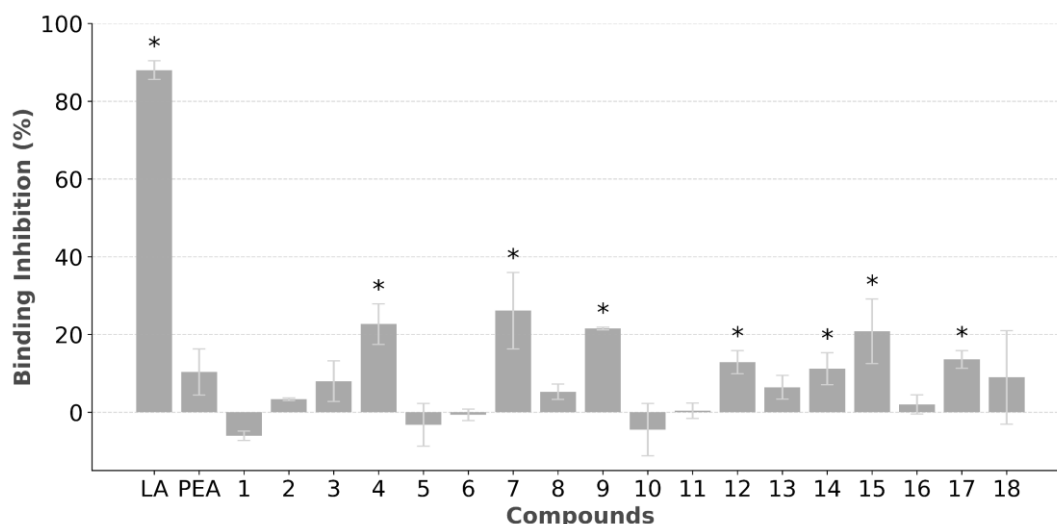


Figure 8 – Inhibitory activity of binding assays for the S-ACE2 interaction for the screened compounds in X-axis, tested at 100 μ M. Inhibition in the Y-axis determined as percentage of the vehicle-treated cells (2% DMSO). Bars represent the mean $\pm$ SEM from three experimental repeats. * Compounds with higher activity than PEA. Abbreviations: S-ACE2 – Spike glycoprotein-angiotensin converting enzyme; DMSO – Dimethyl Sulfoxide; LA – Linoleic Acid; PEA – Palmitoylethanolamide

The maximum detected inhibition was 28% for compound **4**, with compounds **7**, **9** and **15** reaching 20% inhibition of the interaction, while compounds **12** and **17** also surpassed PEA. When compared with linoleic acid at 200 μ M, reaching 84% inhibition of interaction derived signal, these compounds have shown a reduced ability to affect the interaction. However, since PEA can affect the virus-host interaction in virion-based assays, the screened compounds still hold the potential to replicate this effect.

Cell-based infection assays

The antiviral activity of the screened compounds was evaluated in Vero E6 cells in cell-based SARS-CoV-2 infection assays, with the original Wuhan strain. The compounds were evaluated for their ability to protect cells from virus-induced cytopathic effects (CPE), caused by SARS-CoV-2 damages to the host cell, and observed in Vero E6 cells by microscopy ³⁹. To quantitatively assess CPE, the CellTiter-Glo® methodology was used to measure the amount of ATP present in the medium, since CPE leads to ATP release and degradation ³⁹. The screened compounds were tested at multiple concentrations, with a 0.03 μ M-200 μ M range (Figure 9). The compound with highest activity also showed a dose response curve (Figure 10A). However, inhibition of CPE does not provide information on the number of infectious virions within a sample, hence a TCID₅₀ titration assay (median tissue culture infectious dose) was performed to assess the amount of replication-capable lytic virions released (Figure 10B).

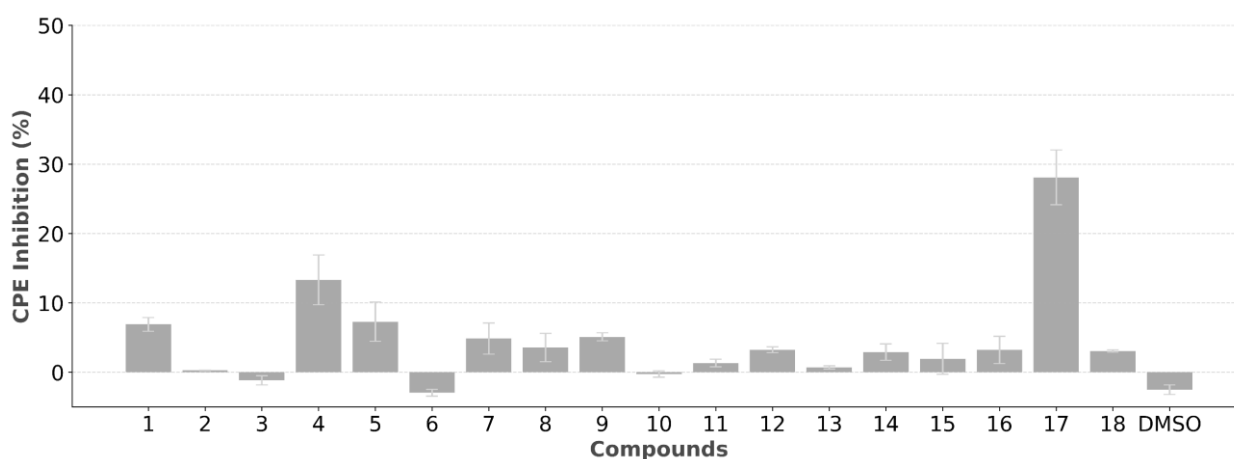


Figure 9 – Inhibitory activity of viral induced CPE by the screened compounds in X-axis, at 67 μ M, after 72h of infection, in Vero E6 cells, measured using the CellTiter-Glo® method. CPE inhibition in Y-axis determined by comparing cells exposed to compounds with cells infected in the absence of compounds. The bars represent the mean $\pm$ SEM from three experimental repeats. Abbreviations: CPE – cytopathic effect

Among the compounds screened in this assay, compound **17** showed the strongest antiviral activity at every concentration, with a maximum 28% reduction in viral induced CPE at 67 μ M and activity over 10% at 2.5 μ M. Overall, cytotoxicity was a limiting factor at higher concentrations. At 67 μ M compound **4** also surpassed 10%, while at 22 μ M, where cytotoxicity is less relevant, compounds **1** and **15** also achieved 10% inhibition.

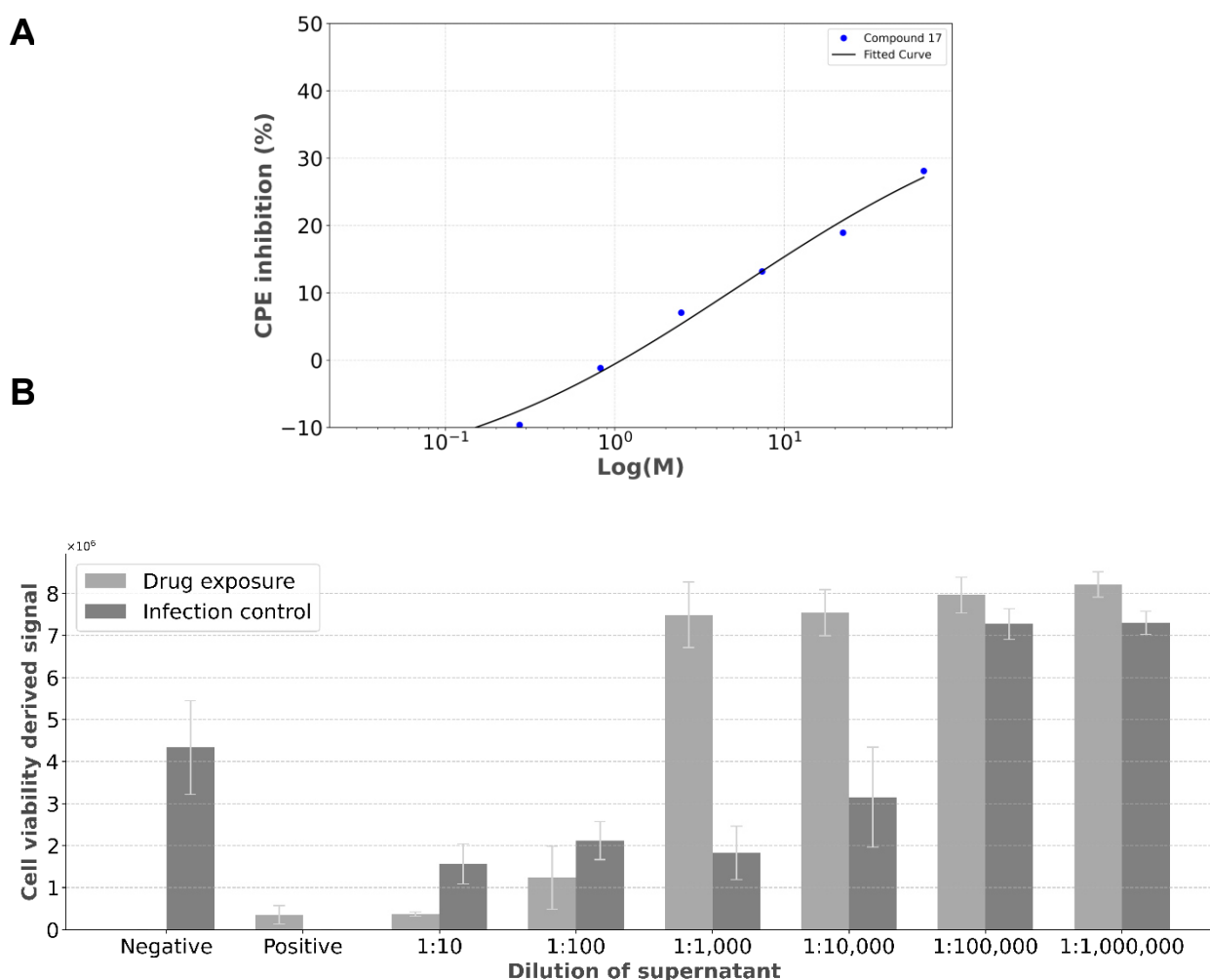


Figure 10 – A) Dose-response curve for compound **17** in Vero E6 cells, corresponding to the inhibitory activity of SARS-CoV-2 induced CPE in the Y-axis. The X-axis represents compound concentration expressed on a logarithmic scale B) Titration assay for compound **17** based on cell viability in the Y-Axis measured with CellTiter-Glo® method, for each dilution of viral supernatant in the X-axis. The bars represent the mean $\pm$ SEM from three experimental repeats. Abbreviations: CPE – cytopathic effect

A dose response curve was obtained for compound **17**, with strong cytotoxicity consistently observed at 200 μ M, that likely limits higher activity. The titration procedure was applied with the highest active concentration (67 μ M). While cell viability at early dilutions was similar in the presence and absence of compound **17**, the drug-exposed group showed cells with regular growth two ten-fold dilutions before the control. These results showed that the presence of compound **17** during the initial infection not only inhibited CPE but also resulted in a decrease

of 81% in the number of infectious virions released⁴⁰. Besides the quantifiable effect, the reduction in CPE for compound **17** was visible by microscopy, as healthy cells were observed. Overall, compound **17** was capable of partial inhibition of CPE during viral infection, while reducing the number of infecting new virions released, demonstrating ability to inhibit viral induced effects.

Predicted binding of antiviral hits

Four compounds showed an increased ability to inhibit viral induced CPE (compounds **1**, **4**, **5**, **17**), with their predicted binding pattern obtained from the docking studies. Following the observations for LA, extensive interactions are predicted in the hydrophilic portion of the pocket, although multiple aromatic rings extend toward the hydrophobic pocket, predicted to provide a better occupation and establish aromatic interactions (Figure 11).

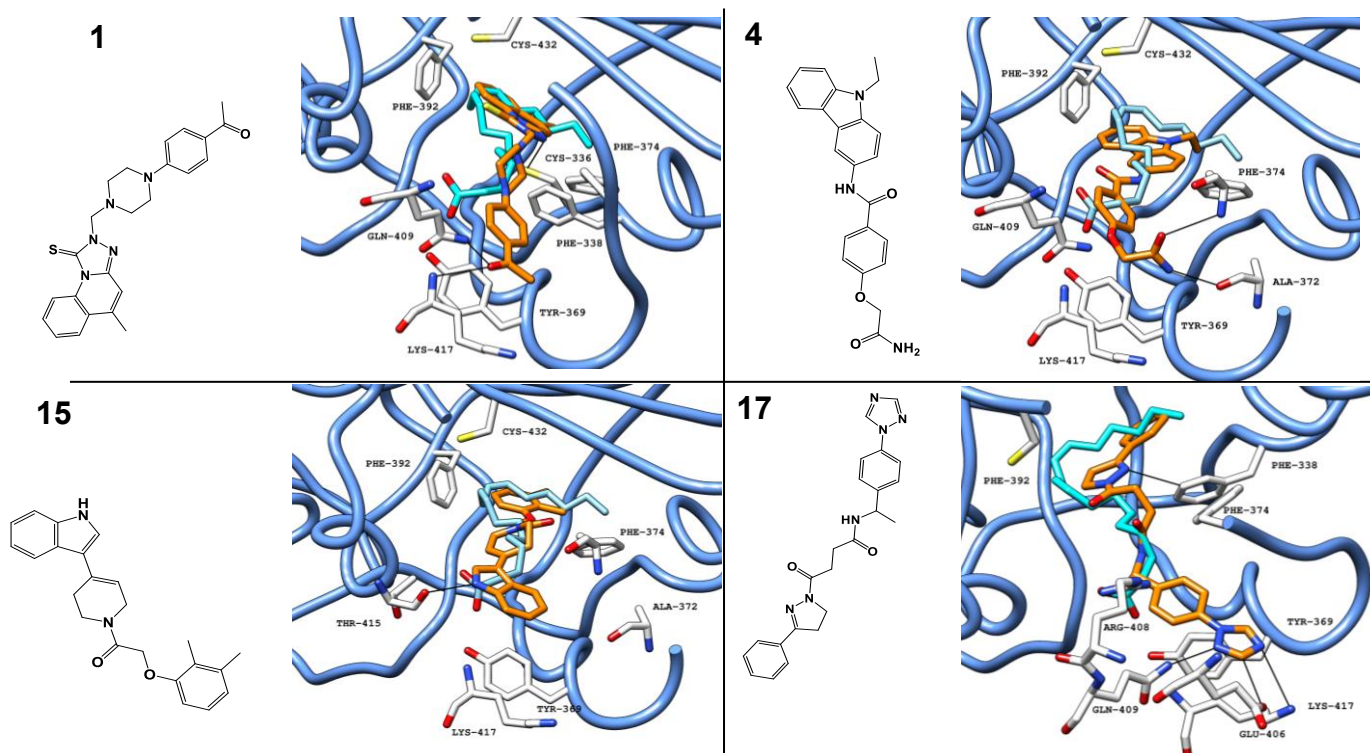


Figure 11 – Crystallized LA (PDB ID: 6ZB5) (carbon atoms in cyan) superimposed with LA in the FABP in ribbon representation (blue ribbon), along with predicted binding pose (obtained with GlideSP) for small molecule inhibitors with highest antiviral activity (compounds **1**, **4**, **15** and **17**), and with chemical structure. Black lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein³⁰. Abbreviations: LA – linoleic acid; FABP – Fatty Acid Binding Pocket.

In the four compounds with strongest activity, the aromatic rings are a recurring feature, with phenylalanine residues (Phe338, Phe374 and Phe392) consistently predicted to establish π -H and π -stacking interactions. On the other hand, regarding the hydrophilic entry of the pocket, compounds **1**, **4** and **17** have hydrophilic moieties (ketone, amide and triazole, respectively), while compound **15** exposes an indole moiety, and therefore is predicted to rely more upon hydrophobic interactions, with only one H-bond predicted. On the other hand, the ketone, amide and triazole groups are predicted to establish extensive interactions, with multiple residues in the pocket entrance (electrostatic interactions including H-bonds). Additionally, compounds **1** and **15** show a less efficient pocket occupation, even compared with LA, which extends more into the pocket.

Regarding the most active antiviral, compound **17**, the binding pattern can be compared to LA, having two main differences. Despite a similar extension in the hydrophobic area, compound **17** is predicted to achieve better pocket occupation due the aromatic nature of the scaffold, particularly in the vicinity of buried phenylalanine residues, allowing π -stacking and π -H interactions, with additional π -H interaction predicted with Cys432 and Tyr369 in the middle area of the pocket. In addition to a good occupation of the deeper pocket areas, compound **17** also extends towards functional groups in hydrophilic residues in the entrance of the pocket. The occupation of the anchoring entrance with a triazole group results in strong H-bond interactions with Glu406, Glu409 and Lys417, and additional π -H interactions with Arg408, which strongly stabilize the molecule in the pocket.

Cell viability

A preliminary cytotoxicity evaluation before cell infection for the 18 compounds was performed using Vero E6 cells, kidney epithelial cells originally isolated from African green monkey (*Chlorocebus* sp.)⁴¹. This is a commonly used cell model for coronavirus infection as it highly expresses ACE2, the functional receptor recognized by the S protein and shows characteristic CPE⁴². In the virtual screening selection protocol, predicted toxicity was an important consideration, with ADME and PAINS (pan-assay interference compounds) analysis included to exclude potential toxic compounds (Supplementary Data file – Table 1)³⁸.

In addition to the test compounds, cells were treated with DMSO, since it is capable of altering cell membrane permeability and selectivity, justifying its use as a control for cytotoxicity⁴³.

In the preliminary assay control, cell growth for most compounds was observed in the initial stage, while a stabilization in cell viability is generally observed, with DMSO exposed cells showing strong cytotoxicity. Overall, a concentration over 100 μ M could be used in infection assays, to allow maximum inhibitory effect without high levels of toxicity (Supplementary Data File – Figure 1). During viral infection assays, concentration causing 50% cytotoxicity (CC₅₀) was determined for all screened compounds (Supplementary Materials - Figure 2), with CC₅₀

values ranging from 59 μM to 220 μM , and strong cytotoxicity only at high concentrations. The most active compound (**17**) showed a CC_{50} of 90 μM . The preliminary evaluation of cytotoxicity was, therefore, critical to guarantee adequate separation between activity and cytotoxicity at testing concentrations.

Overall, from an initial virtual screening study targeted at the S protein, compound **17** was identified as the most active compound, reducing viral induced CPE by around 30%, along with a reduction in viral particles released. It also showed approximately 20% ability to reduce the S protein-ACE2 interaction in binding assays. Therefore, it is likely that the activity shown in the infection assays is, at least in part, derived from an effect in the FABP, stabilizing the inactive conformation of the RBD, which becomes incapable of recognizing target receptors, recognize and infect new cells.

Analog identification

Starting from the compound with highest activity, with compound **17** showing micromolar activity, a few commercial analogues were selected by similarity for activity confirmation and the study of structure-activity relationships. The twenty selected analogues were divided in three main groups: 1) substitution of the terminal triazole by apolar groups, such as tertbutyl, phenyl or methyl; 2) substitution of the terminal triazole by functional polar entities such as tetrazole, thiazole and amide; 3) substitution of linker between the aromatic groups, extension or reduction in overall size, and new scaffolds (Figure 12). These changes were intended to explore the interactions with the hydrophilic anchor in the pocket, improving pocket occupation and exploring new interactions, as well as improving drug-like properties.

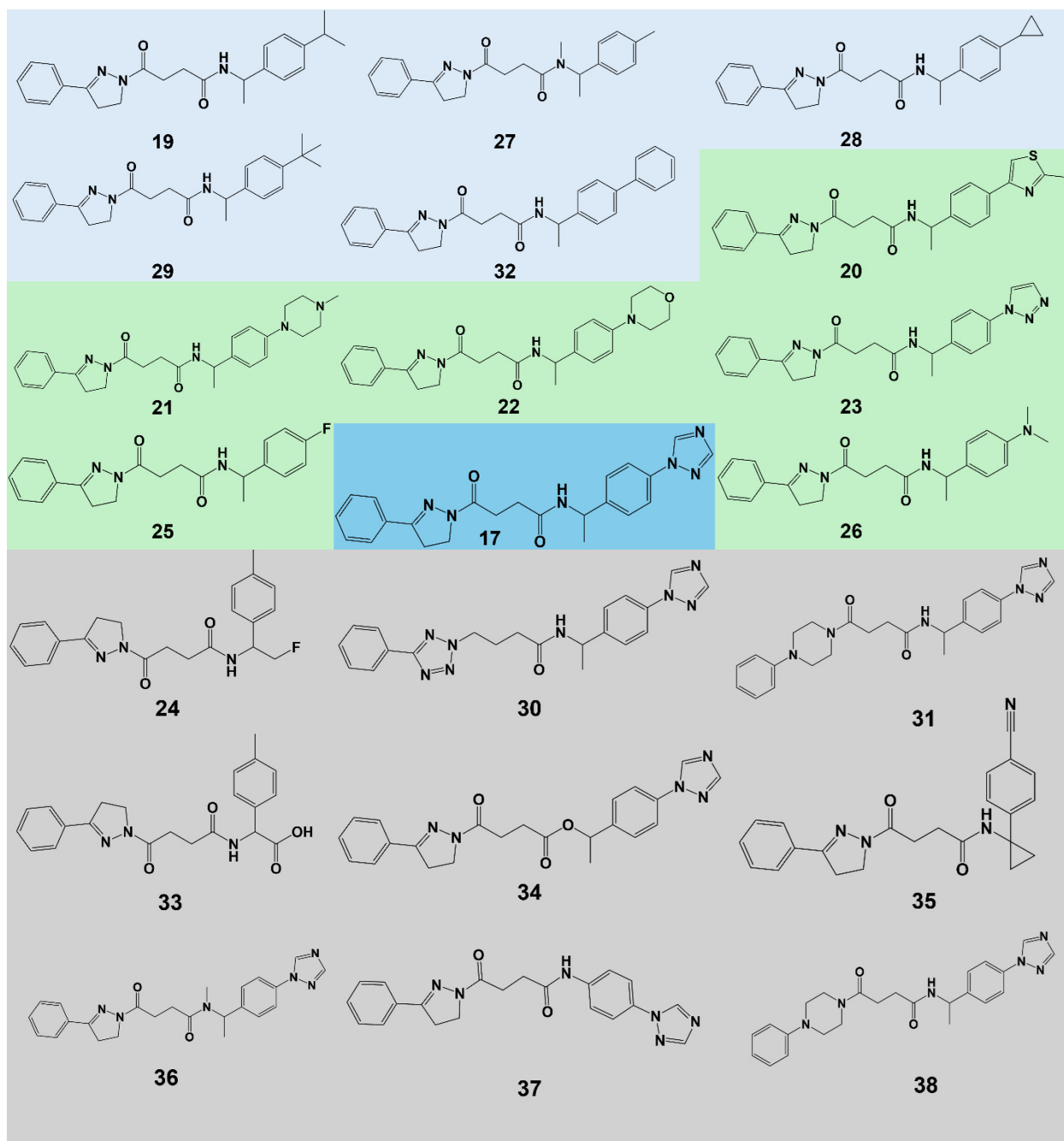


Figure 12 – Chemical structure of analogues compounds to the strongest antiviral identified, divided in three categories and the original bioactive compound, with hydrophobic substitutions (light blue), hydrophilic substitutions (green), other changes (grey), original compound (blue).

Biological evaluation of the hit analogues

The twenty structural analogues selected (compounds **19** to **38**) and the parent compound (compound **17**), used as a positive control, were evaluated for their ability to inhibit the viral induced CPE and viral replication. The compounds were tested at concentrations ranging from 0.15 μM to 100 μM . Since compound **17** was able to reduce CPE evident by microscopy, only compounds with the ability to visibly reduce CPE were subject to quantitative testing, upon confirmation of positive and negative controls. Upon optical confirmation of CPE, ten compounds were subject to quantitative evaluation of activity and toxicity, with CPE inhibition at 33 μM showing maximum activity levels (Figure 13A), while cytotoxicity was limited with CC_{50} values ranging from 37 μM to 243 μM and 50 % cytotoxicity not detected at 100 μM (Supplementary Data – Figure 3). The best analogue, compound **36**, exhibited a maximum of 91% inhibition of CPE at 33 μM , and a dose response curve (Figure 13B).

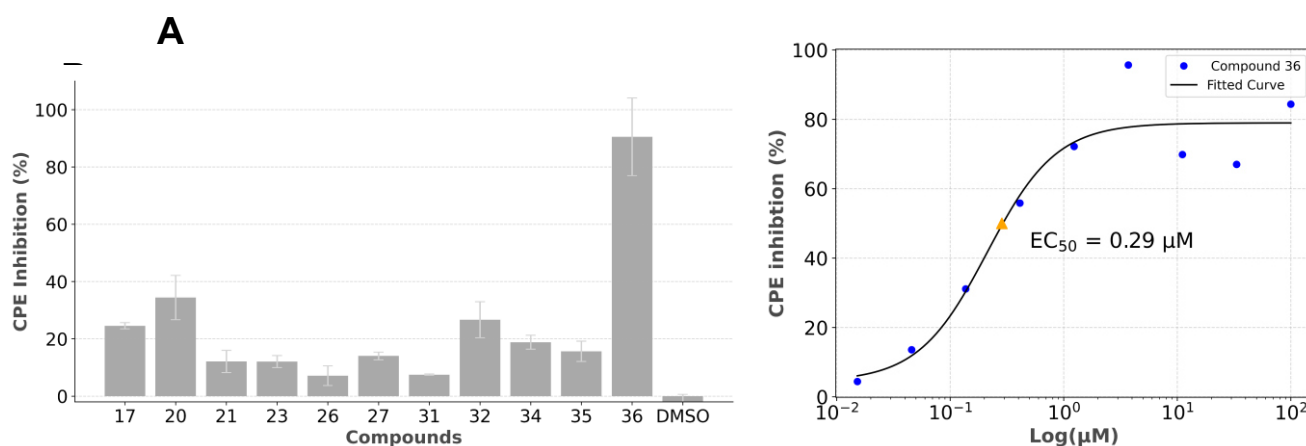


Figure 13 – A) Inhibitory activity of viral induced CPE by analogue compounds in the X-axis at 33 μM , after 72h of infection, in Vero E6 cells, measured using the CellTiter-Glo® method. The Y-axis represents CPE inhibition, determined by comparing cells exposed to compounds with cells infected in the absence of compounds. The bars represent the mean $\pm$ SEM from three experimental repeats. B) Dose-response curve for compound **36** in Vero E6 cells, corresponding to the inhibitory activity of SARS-CoV-2 induced CPE in the Y-axis. The X-axis represents compound concentration expressed on a logarithmic scale. Abbreviations: CPE – cytopathic effect, DMSO - Dimethyl sulfoxide, EC_{50} - half maximal effective concentration.

From the analogues that induced a visible reduction by microscopy in viral induced CPE, most had their activity confirmed in the infection assays, with nine showing at least 10% inhibition at 33 μM , and compound **36** showing 91% inhibition with an EC_{50} of 0.29 μM . When compared with compound **17** (23%), compound **20** (34%), compound **32** (27%) and compound **36** (91%) achieved stronger inhibition of viral induced CPE. Overall, cytotoxicity was limited, not

reaching 50% reduction in cell viability in all compounds, apart from compound **27** and **34**.

Structure-activity analysis

The analogues selected for the optimization stage of this study introduced multiple structural modifications aimed at increasing activity and establishing a preliminary structure activity relationship. For the hydrophobic substitution analogues, introduction of an aliphatic chain in the hydrophilic anchor is associated with loss of activity, with interactions in this area severely reduced (Figure 14).

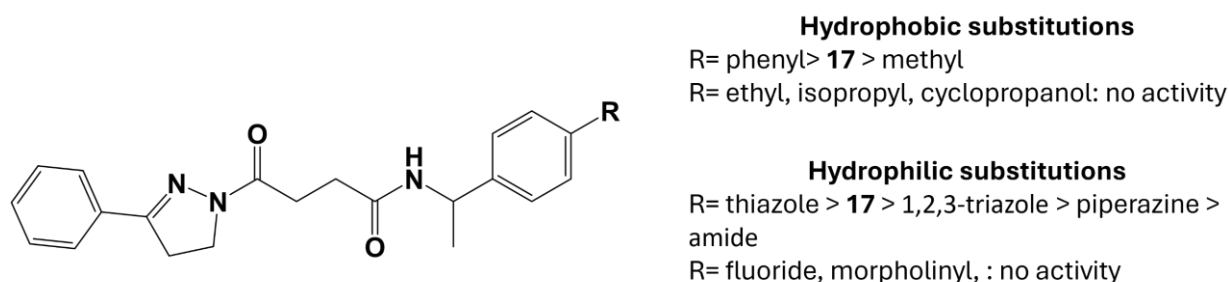


Figure 14 – Quantitative/qualitative structure activity relationship concerning the first and second group of analogues.

The aromatic nature of the triazole is confirmed as critical since compound **32** (phenyl) resulted in increased activity, while the importance of interaction in the hydrophobic area of the pocket is highlighted by the activity maintained in compound **27** (methyl). Regarding the hydrophilic substitutions (Figure 15), the increased activity of the methylthiazole substitution (compound **20**), could be due to increased aromatic and hydrophilic interactions, or potentially by an increase in LogP (2.36 (compound **17**) and 3.91 (compound **20**)), which, given the hydrophobic nature of the FABP, might benefit initial binding and therefore activity.

On the third set of analogues, the strongest antiviral analogue (compound **36**) represented a significant improvement when compared with compound **17**, although they only differ in the methylation of a linking secondary amine group, resulting in a tertiary amine (Figure 15). The nitrile group in compound **35** might also be an interesting terminal group since some activity is retained.

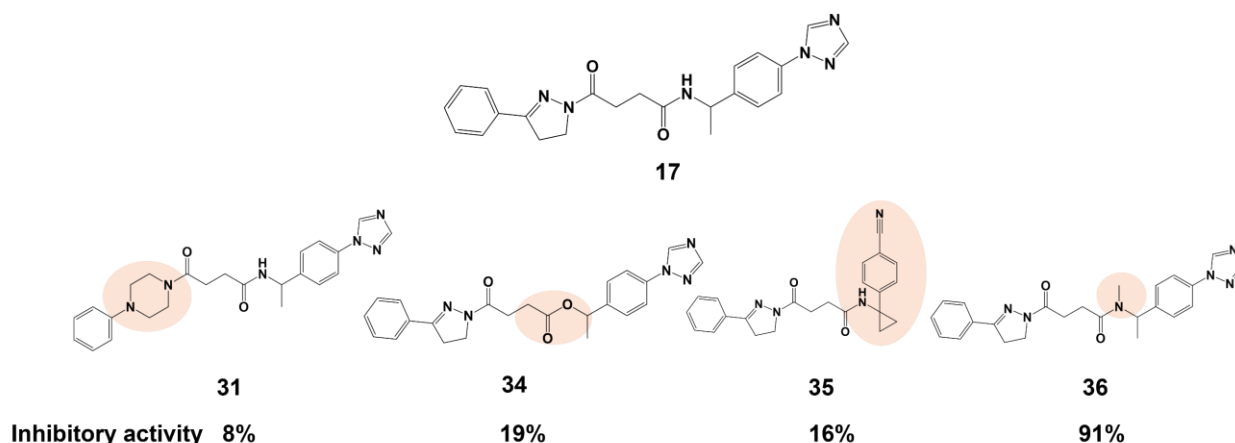


Figure 15 – Chemical structure of analogues compounds from the third category that resulted in compounds with inhibitory activity.

Although compound **36** loses the ability to be a hydrogen donor for H-bonds, the ability to be a receptor for H-bonds, such as with Tyr365 and Tyr369, is reinforced, which might result in a stronger anchoring effect of this portion of the molecule (Figure 16). Additionally, the increase in LogP (2.36 (compound **17**) to 2.62 (compound **36**)) might also improve initial interaction with the target.

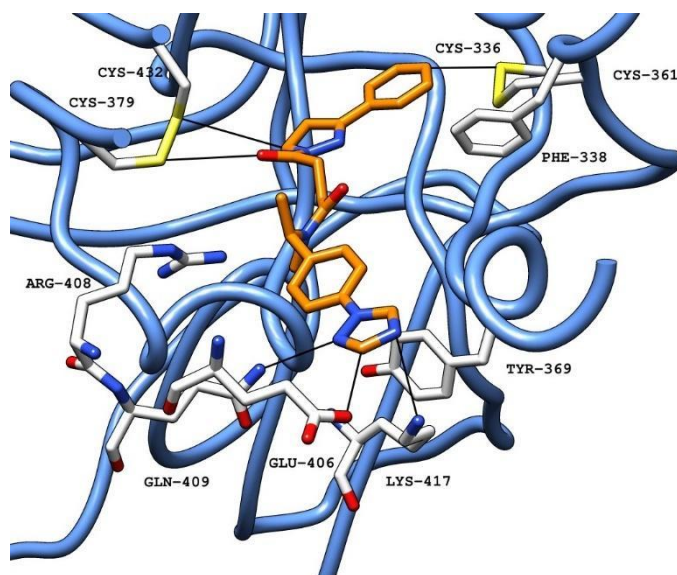


Figure 16 – Predicted binding pose for compound **36** (carbon atoms in orange) obtained with GlideSP, in the FABP in ribbon representation (blue) (PDB ID: 6ZB5). Black lines represent polar interactions (hydrogen bonds and electrostatic interactions) between the ligand and amino acid residues in the protein³⁰. Abbreviations: FABP – Fatty Acid Binding Pocket, PDB – Protein Data Bank

Activity against SARS-CoV-2 variants of concern

In addition to the Wuhan SARS-CoV-2 strain, compound **36** was also tested against the SARS-CoV-2 Delta and Omicron variants of concern^{10, 23}. Since the FABP is not under selective pressure from antibodies and vaccine induced immune response, it is expected that active compounds maintain the ability to inhibit cell induced CPE and affect the virus life cycle²². Therefore, compound **36** was tested against SARS-CoV-2 Delta and Omicron variants. The results with the Delta variant followed a similar pattern to the Wuhan variant, with a maximum inhibition detected at 11 μM , while a dose response is present towards lower dilutions (Figure 17). Compound **36** showed a maximum inhibition of CPE by 78% at 33 μM and an EC_{50} of 5.77 μM .

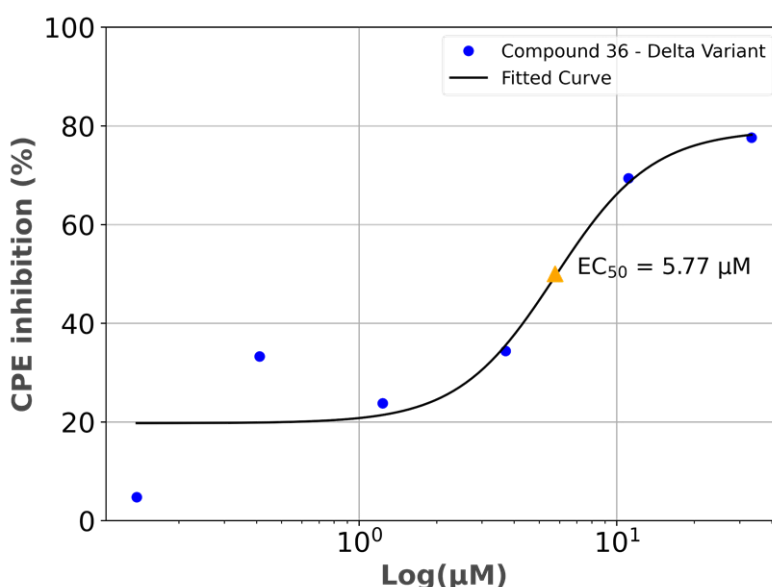


Figure 17 – Dose-response curve for compound 36 in Vero E6 cells, corresponding to the inhibitory activity of SARS-CoV-2 – Delta Variant induced CPE in the Y-axis. The X-axis represents compound concentration expressed on a logarithmic scale to better visualize the dose-response relationship. Abbreviations: CPE – cytopathic effect

An important deviation from expected behavior with the Delta strain was detected, when compared with the Wuhan strain. Infected cells did not show CPE after 96h and only after an additional 72 h period could CPE be observed. Further assays performed with the Omicron strain did not show CPE even after this extended period (Supplementary data - Figure 4). There seems to be a link between viral evolution from the initial strains (Wuhan and Alpha) to later strains (Delta and Omicron), that leads to reduced CPE⁴⁴. Cell entry by SARS-CoV-2 Wuhan strain can proceed by two ways, S-protein initiated cell fusion after a proteolysis by Transmembrane Protease Serine 2 (TMPRSS2), or alternatively by endocytosis, with the

acidic pH activating cathepsin for the proteolysis step ⁴⁵. The Omicron variant has a strong preference for the endocytosis pathway, whereas the Wuhan and Delta strains preferentially initiate entry by the TMPRSS2 cleaved S- protein. Viral induced CPE is dependent on infection through adjacent cells, which requires TMPRSS2 mediated activation. This change is also associated with a variation in cell tropism: early SARS-CoV-2 strains target lower respiratory tract, rich in TMPRSS2, while Omicron targets upper respiratory tract, rich with cathepsin, which justifies the lack of observed CPE for the omicron variant ^{39, 45}.

Compound **36** maintained activity in the heavily mutated Delta variant, showing conservation of the FABP, which is also conserved in other human CoVs, particularly in the other highly pathogenic CoVs: SARS-CoV and MERS-CoV ²³. Therefore, it would be important to perform further assessment of pan-coronavirus activity, as compound **36** holds the potential of broad-spectrum activity. Overall, the activity shown by this molecule could be highly relevant not only to the current pandemic, but also to potential future-emerging new CoVs.

RT-ddPCR

SARS-CoV-2 infected cells were used for antiviral assays, with the resulting viral induced CPE analyzed. To confirm and quantify that the pathogen causing this effect indeed was SARS-CoV-2, PCR analyses were performed on the supernatant retrieved from infected cells. To this end, we applied reverse transcriptase (RT) digital droplet (dd)PCR ⁴⁶. For this PCR, the target samples are compartmentalized in thousands of droplets together with the PCR reagents, making multiple reactions that occur simultaneously ^{46, 47}. Due to this, absolute quantification is possible ⁴⁷. The supernatant samples from infected cells with SARS-CoV-2 Wuhan strain and treated with compound **36** were retrieved and analysed by RT-ddPCR (Table 1).

Table 1 – ddPCR results, with detected viral genomic copies for SARS-CoV-2 infected cells treated with compound **36** compared to non-treated and infected cells.

Compound	Copies/20 μ L
36	16.9
Positive control	146.6

The number of viral genome copies present in the supernatant after infection was reduced by 89% at 33 μ M. Additionally, TCID₅₀ was also determined for compound **36** (), with a reduction of 90% of replication-capable lytic virions. Hence both independent methodologies point to a similar reduction in infectious virions released during infection, with compound **36** capable of significantly affecting the virus lifecycle.

Table 2 – Titration results for infection with SARS-CoV-2 Wuhan strain exposed to compound 36.

TCID ₅₀ /mL	Drug	No drug	PFU/mL	Drug	No drug
Compound 36	3.6	4.6		2.49E+07	1.25E+08

Conclusions

The FABP shows great promise in modulation S-protein behavior and, consequently, SARS-CoV-2 infection, due to its essential role in cell recognition and entry. However, none of the currently approved SARS-CoV-2 treatments targeting the FABP. Given the conservation of the pocket across highly pathogenic human coronaviruses, it holds the potential for pan-coronavirus activity, which is crucial for addressing both current and future coronavirus outbreaks.²³.

In this work, the FABP effects on the S-protein behavior were explored using molecular dynamics simulations, which guided a docking-based virtual screening of a commercial library of small-molecule, drug-like compounds. This approach led to the identification of potential inhibitors, which were purchased and evaluated in direct binding and cell-based antiviral assays. Among these, one compound demonstrated significant inhibition of virus-induced effects, likely by binding to the FABP, thus validating the computational approach. This compound served as the starting point for evaluating twenty structural analogues, leading to the identification of a sub-micromolar inhibitor of SARS-CoV-2 with low toxicity. These findings might provide a promising starting point for further optimization and drug development efforts against SARS-CoV-2.

Materials and methods

Molecular modelling studies

Molecular Dynamics Simulations

Molecular dynamics and molecular docking studies supporting the virtual screening study were performed on an Asus WS X299 PRO Intel i9-10980XE CPU @ 3.00 GHz × 36 running Ubuntu 18.04 (graphic card: GeForce RTX 2080 Ti). Molecular Operating Environment (MOE) 2022.02³¹ and Maestro (Schrödinger Release 2020-2)³³ were used as molecular modelling software. The crystal structure of the S protein in the open and closed conformations, along with LA in the S protein were downloaded from the Protein Data Bank (PDB) (<http://www.rcsb.org/> (accessed on 15 May 2023); PDB code 7DF3, 6VYB and 6ZB5, respectively)^{3, 21}. Missing loops in structures were constructed through homology modelling. A cubic water box with a 10 Å buffer distance was used for the solvation system between each box side and the protein atoms, with four sodium atoms used to neutralize the system. Before the MD simulation, the system was pre-equilibrated using a default relaxation routine implemented in Desmond. Three 100 ns MD simulations were performed, during which the equation of motion was integrated using a 2 fs time step in the NPT ensemble with temperature (300 K) and pressure (1 atm) constant. All other parameters were set using the Desmond default values. Data was collected every 1.2 ps (energy) and every 200 ps (trajectory). Each individual system was simulated in triplicate with a separate random seed as a starting point. Visualization of the S protein and MD trajectory analyses were carried out using Maestro. RMSD, RMSF, secondary structure and protein–protein interactions analyses were performed using the Simulation Event Analysis tool and the Simulation Interaction Diagram of Desmond.

Site finder

The Site Finder application in MOE was used to find potential ligand binding pockets in generated S protein structures from frames in the molecular dynamics generated trajectory files. A fifteen-frame interval was applied, corresponding to a 3 ns, and were submitted to the Site Finder application, in PDB format. Site Finder produced a list of potential binding sites, including interacting residues and size of the predicted pocket. Each pocket identified was visually analyzed and pockets similar to the FABP were selected. The procedure was repeated in both open and closed conformations, and for each of the three FABP, in each S protein structure.

Virtual screening

The Enamine library of commercially available drug candidates was screened against the fatty acid binding pocket using PDB ID 6ZB5 crystal structure (<http://www.rcsb.org/> (accessed on 15 May 2023); PDB code 6ZB5)²¹. The structures of the compounds analyzed were built in MOE2019.10, saved in .sdf format, and prepared using the Maestro LigPrep tool by energy minimizing the structures (OPLS_2005 force field) and generating possible ionization states at pH 7 ± 2, tautomers, all possible stereoisomers per ligand, and low-energy ring conformers. The protein was pre-processed with the MOE Protein Preparation tool, and the resulting protein–ligand complex was saved in .mae format and prepared using the Schrödinger Protein Preparation Wizard by assigning bond orders, adding hydrogens, and performing a restrained energy minimization of the added hydrogens using the OPLS_2005 force field. Additionally, the protein was also saved in .oedu format and .mol2 format to be used with scoring software ScorePose (OpenEye) and Plants, respectively^{35, 36}. The Glide high-throughput virtual screening tool (HTVS) was used to virtually screen the commercial database against the binding site, followed by a docking procedure with Glide SP³³. A 15 Å docking grid was prepared using the co-crystallized LA as the centroid, in parallel for the three FABP in the S protein. The library was docked on the active sites using the Glide HTVS docking algorithm¹⁶, keeping the default parameters, setting to three the number of output poses per input ligand to include in the solution. The top 50 000 compounds scored by Glide HTVS were selected for a new docking procedure with Glide SP (standard precision) docking algorithm, setting to three the number of output poses per input ligand to include in the solution and performing a post-docking minimization of each of the poses kept³³. The output poses were saved as mol2 files. Docking poses obtained were then rescored (maintaining the identified pose) using Glide XP, CHEMPLP (PLANTS) and OpenEye (ScorePose) scoring functions³⁴⁻³⁶. Using a single docking program and scoring function might introduce potential bias, which justified the use of three programs for rescoring. The value of each scoring functions for each docking pose were then combined (consensus score) and only docking poses falling in the top 25% of the score value range for all the four scoring functions were selected for visual inspection in the three FABP. The docking results were visually inspected in MOE 2022.02. The docking poses of the compounds obtained from the visual inspection were evaluated considering the following criteria: ability of a compound to adequately occupy the fatty acid binding site (similar to LA); interactions predicted between compound and protein residues defining the site. Given that LA has been confirmed as a ligand and has shown antiviral activity in experimental assays, docked molecules were superimposed with a crystallographic structure of LA bound to the S protein (PDB accession code 6ZB5). In the next step, the set of molecules from each pocket was combined, with only molecules capable of good predicted interaction in all pockets selected for further stages. Finally, the set of molecules identified for experimental validation

was reduced to 20, by applying the Lipinski rule of five (selecting for good medicinal chemistry properties) and the SWISS-ADME webtool, to screen compound potential for toxicity (PAINS and BREAK analysis) ³⁸.

Source of Small Molecules

All the compounds in this study, both initial screened set and analogue compounds were purchase from Enamine, Ltd (Kyiv, Ukraine). The library used for the virtual screening study was the Enamine Screening Collection ³². Additionally, the analogue compounds were selected based on a similarity search on the Enamine REAL database, through which it was purchased ⁴⁸. Molecular formulas (SMILES), molecular weight, PAINS and BRENK analysis of tested compounds are reported in the Supplementary Data file (Table 1).

Biological assays

Binding assays

An Inhibitor Screening Assay Kit was used to screen inhibitors of the S-ACE2 interaction (BPS Bioscience Catalog # 78012) ⁴⁹. The kit includes the S protein in its native trimeric conformation from the Wuhan strain, providing the best physiologically relevant model for this interaction ⁵⁰. The assay kit also contains Biotinylated-ACE2, Streptavidin-HRP, and assay buffers. The assay procedure was performed as follows: SARS-CoV-2 S protein was first coated onto a 96-well plate. Following this, Biotinylated-ACE2 was incubated with the S protein on the plate, and Streptavidin-HRP was added to the plate. The interaction between Biotinylated-ACE2 and SARS-CoV-2 S protein was then detected using a colorimetric substrate. The resulting colour change was quantified by measuring the absorbance using a UV/Vis microplate reader. Compounds were dissolved in DMSO and diluted until a testing concentration of 200 μ M, with each compound tested in triplicate. Finally, one negative control (vehicle – 2% DMSO) and two positive controls were used, LA and PEA.

Cell culture

Vero E6 cells used for cytotoxicity and antiviral assays were cultivated in Dulbecco's Modified Eagle Medium (DMEM) with 10% FCS, 1% penicillin G/streptomycin (P/S), and grown in standard culture conditions, namely in a humidified incubator at 37° and 5% CO₂. Cells were evaluated daily, and the culture medium was changed whenever necessary, with cell passaging was performed when desirable confluences of 70–80% were observed.

Cytotoxicity – Presto Blue™ Viability Assay

The Presto Blue™ assay was performed to determine the cytocompatibility between the cellular system and test compounds. This assay is based on a ready to use, commercially available water-soluble preparation and allows a live-cell evaluation. The resazurin solution was used to assess cell viability, based on the mitochondrial metabolization of this substance solution. Viable cells reduce the phenoxazine dye (resazurin), which results in colour modification from blue to reddish over time that can be not only directly observed but also quantitatively measured by UV-VIS spectrophotometry, functioning as a cell viability indicator. Vero E6 cells were seeded over a 96-well plate and maintained in incubation overnight (standard culture medium, 37°C, 5% CO₂ environment and 80% humidified atmosphere). To perform the Presto Blue™ assay, the culture medium was removed from each well at every time-point (24, 28, 96 hours) and replaced by complete medium with 10% (v/v) of 10 Presto Blue™ cell viability reagent (Invitrogen, A13262, Thermo Fisher Scientific, Waltham, MA USA). To perform the analysis, cells were incubated for 60 minutes standard conditions to allow metabolization of the reagent. The supernatant medium was then collected and transferred to a 96-well plate and absorbance was read at 570nm and 595 nm in a Multiskan™ FC Microplate Photometer (Thermo Scientific™, 51119000, Thermo Fisher Scientific, Waltham, MA USA). Afterwards, wells were washed with Dulbecco's phosphate-buffered saline solution (DPBS, Gibco, 14190169) until the Presto Blue sediments were removed. Then, fresh culture medium was added to each well, according to the timepoint specifications. Regular growth medium was used until the first timepoint (24h), when it was replaced with DMEM (10% FCS, 1% P/S) supplemented with the test compound (determination of acute cytotoxicity). At the second timepoint (28 h) the medium was replaced with DMEM (2% FCS, 1% P/S) supplemented with test compounds, then left for 72h until the last timepoint (96h) (determination of acute cytotoxicity). For the Presto Blue assessment, both control group and test compounds were considered, and for each group, blank wells (without cell seeding), were included. The wavelength for excitation is 570 nm and for emission 595 nm. For that reason, the value obtained at 595 nm was subtracted from the value obtained for 570 nm (normalized value), for each well. In addition, the corrected absorbance for each experimental well, only considering seeded wells, was obtained by the subtraction of blank wells average from the normalized values, of the respective sample group. The absorbance values were measured in triplicates. Data was further processed and normalized to the mean of the gold standard group and presented in a ratio between the 24h timepoint, and both 28h and 92h timepoints, representing variation against initial cell viability as a baseline. Statistical analysis was performed with a One-Way Anova, with Dunnett's test Post-Hoc.

SARS-CoV-2 antiviral assay - Virus infection

Vero E6 cells were seeded into 96-well plates at a density of 40,000 cells per well in 100 μ L of DMEM supplemented with 10% FCS and 1% P/S. Blank and vehicle controls were included. The cells were incubated overnight at 37°C with 5% CO₂ to reach approximately 70-80% confluence. Serial dilutions of the test compounds were prepared in DMEM (2% FCS, 1% P/S), with nine three-fold dilutions ranging from 200 μ M to 0.03 μ M. The cells and compound solutions were then transferred to a Biosafety Level 3 laboratory for infection with SARS-CoV-2. The growth medium in each well was replaced with test compound-containing solutions immediately before infection. The top half of each plate received an additional mock DMEM (2% FCS, 1% P/S), while the bottom half received virus-containing solution. Hence, cell infection was performed in two sets: 1) Vero E6 cells treated with test compound and infected with SARS-CoV-2, and 2) Vero E6 cells treated with test compound only, allowing a direct comparison of cell viability in presence or absence of virus. Cell viability was compared with mock-treated samples treated with vehicle only (0.5% DMSO [vol/vol]). Two controls were used (with an additional blank without cells to measure background signal): a positive control where cells are infected without test compound and a negative control with non-infected cells treated with DMSO only (vehicle). The cells were infected with 0.05 MOI (Multiplicity of infection) of SARS-CoV-2 strain INMI1 P4. The virus was obtained from EVAg⁵¹. Additionally, the antiviral assays were also performed with other variants of concern, particularly the SARS-CoV-2 Delta and SARS-CoV-2 Omicron obtained through private collaboration. The mock-treated and infected cells were incubated at 37°C with 5% CO₂ for 72-96 hours, until a cytopathic effect (CPE) was observed in all control wells. Upon CPE detection, some of the supernatant was harvested and frozen for downstream analyses. Cells and remaining SN were frozen for viability assays.

SARS-CoV-2 antiviral assay - Cytotoxicity and cell viability

Cell viability was determined using CellTiter-Glo® (Promega), that quantifies the amount of ATP present in the sample provided. ATP is a measure of metabolically active cells, allowing the determination of the number of viable cells. Cells/SN that were treated with different concentrations of compounds and infected were thawed after freezing at -80 °C and equilibrated to room temperature. In a white bottom plate (PerkinElmer® - 1/2 Area ViewPlate) 10 μ L of lysed cells/medium were mixed with 10 μ L of CellTiter-Glo® Reagent (Promega). The plate was then mixed for two minutes in an orbital shaker to induce cell lysis and allowed to incubate at room temperature for 10 minutes to stabilize the luminescence signal. The luminescence signal was then measured with a plate reader (VICTOR® Nivo™, Perkin Elmer). Antiviral activity is measured by the ability to reduce the viral effects, which reduce the number of viable cells and cause CPE. As a result, the ATP levels detected are also reduced, along

with the luminescence signal. This was expressed as percentage of inhibitory effect of viral induced reduction of signal compared against untreated virus-infected positive control cells (100% CPE).

$$\text{Inhibitory activity} = \frac{[(\text{Cells} + \text{virus} + \text{antiviral}) - (\text{Cells} + \text{virus})]}{[(\text{Cells}) - (\text{Cells} + \text{virus})]}$$

When 50% inhibition of viral induced reduction of CPE is detected, the antiviral activity is expressed by half maximal effective concentration (EC_{50}), the concentration of compound that achieves 50% of viral effect.

TCID₅₀ determination – virus titration

To determine the 50% tissue culture infectious dose (TCID₅₀), a standard virus titration assay was performed. As previously, Vero E6 cells were seeded in a 96-well plate at a density of 40,000 cells per well and incubated overnight at 37°C with 5% CO₂. Six serial tenfold dilutions of virus supernatant were prepared in DMEM (2% FCS, 1% P/S). The supernatant used corresponds to the compound concentration with the highest activity and reduced cytotoxicity. Each dilution was added to six wells, starting with the lowest dilution. The plates were incubated at 37°C with 5% CO₂ for 96 hours, during which CPE was monitored daily. Following the incubation period, each well was scored for the presence or absence of CPE. TCID₅₀ can then be calculated using the Reed-Muench method, which involves determining the dilution at which 50% of the wells show CPE. This is done by plotting the number of positive wells against the dilution factor and finding the dilution at which there is a 50% probability of infection. Alternatively, cell viability can be determined with CellTiter-Glo® and the virus titre causing 50% reduction in CPE can be determined.

RT-ddPCR

RT-ddPCR was employed to quantify SARS-CoV-2 with high precision and sensitivity. The RT-ddPCR reactions were prepared using One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad®), with target-specific primers and probes (sequence, Forward Primer: ACAGGTACGTTAATAGTTAATAGCGT; Reverse Primer: ATATTGCAGCAGTACGCACACA; Probe Sequence: ACACTAGCCATCCTTACTGCGCTTCG)⁵². Each 20 µL reaction mixture contained 5 µL of the Supermix, 2 µL of reverse transcriptase (RT), 1 µL of primers/probes mix, and 11 µL of inactivated viral sample from the dilution with highest activity for each compound. Droplets were generated using the Bio RAD QX200 Droplet Generator™. Briefly, the reaction mixture was loaded into the sample wells of a DG8™ Cartridge, along with

ddPCR™ Droplet Reader Oil for probes in the oil wells and covered by DG8™ Gaskets. The cartridge was then placed into the QX200™ Droplet Generator. After droplet generation, the emulsified PCR reactions were transferred to a 96- well PCR plate and sealed with a foil seal using a PX1 PCR Plate Sealer. PCR amplification was carried out using a thermal cycler with the following conditions: initial cDNA synthesis from RNA template by RT at 50°C for 15 minutes, followed by DNA polymerase activation at 95°C for 2 minutes, and finally by 40 cycles of 95°C for 15 seconds and 60°C for 30 seconds. The droplets were then read using a QX200 Droplet Digital™ System, which counted the number of fluorescent-positive and -negative droplets to determine the absolute quantity of target molecules using Poisson distribution analysis. Data analysis was performed using the manufacturer's software, for absolute quantification of the SARS-CoV-2 genome copies per 20 µL of the initial diluted sample.

List of abbreviations

ACE2 - Angiotensin converting enzyme 2
ADME - Absorption, Distribution, Metabolism, Excretion
CC₅₀ - concentration 50% cytotoxic
COVID-19 - Coronavirus disease 19
CPE - cytopathic effects
DMEM - Dulbecco's Modified Eagle Medium
DMSO - Dimethylsulfoxide
FABP - Fatty acid Binding Pocket
FCS - Fetal calf serum
LA - Linoleic acid
MD - molecular dynamics
MERS-CoV – Middle Eastern Respiratory Syndrome Coronavirus
P/S - penicillin G/streptomycin
PEA - Palmitoylethanolamide
RBD - Receptor binding domain
RMSD - Root Mean Square Deviation
RMSF - root mean square fluctuation
RT- ddPCR Digital droplet PCR
S protein - Glycosylated spike protein
SARS-CoV-2 - Severe Acute Respiratory Syndrome Coronavirus 2
SARS-CoV - Severe Acute Respiratory Syndrome Coronavirus
TCID₅₀ -50% tissue culture infectious dose
TMPRSS2 - Transmembrane protease serine 2

Data and Software Availability

Not Applicable

Conflicts of Interest:

The authors declare no conflict of interest.

Funding and Acknowledgments

L.Q.-R. would like to acknowledge Fundação para a Ciência e para a Tecnologia for the grant “2020.10230.BD” under the program “DOCTORATES 4 COVID-19” and EEA Grants/Norway Grants for the grant “FBR_OC52_53”. R.A acknowledges the Centro de Estudos de Ciência Animal (CECA), Instituto de Ciências, Tecnologias e Agroambiente (ICETA), Porto University (UP), and Fundação para a Ciência e Tecnologia (FCT) for the funding and availability of all

technical, structural, and human resources necessary for the development of this work. His participation in this project was supported through project UIDB/00211/2020 merged by FCT/MCTES through national funds and through project 2022.04501.PTDC (Olfabionerve — Olfactory Mucosa Mesenchymal StemCells and Biomaterials Promoting Peripheral Nerve Regeneration).

Supporting information

Supplementary File contains chemical structures of all compounds reported , Molecular Weight, calculated LogP, H-bond accepting groups, H-bond donating groups, PAINS and BRENK analysis. Additionally, cytotoxicity results and SARS-CoV-2 Omicron infection cell viability are also reported here.

References

- (1) WHO. *Coronavirus Disease (COVID-19) Dashboard*. WHO 2020. <https://covid19.who.int/> (accessed 2021/09/04).
- (2) Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **2016**, 3 (1), 237-261. DOI: 10.1146/annurev-virology-110615-042301.
- (3) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veessler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, 181 (2), 281-292.e286. DOI: 10.1016/j.cell.2020.02.058.
- (4) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, 581 (7807), 221-224. DOI: 10.1038/s41586-020-2179-y.
- (5) Fung, T. S.; Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annual Review of Microbiology* **2019**, 73 (1), 529-557. DOI: 10.1146/annurev-micro-020518-115759.
- (6) Woo, P. C.; Lau, S. K.; Huang, Y.; Yuen, K. Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* **2009**, 234 (10), 1117-1127. DOI: 10.3181/0903-mr-94.
- (7) Perlman, S. Another Decade, Another Coronavirus. *New England Journal of Medicine* **2020**, 382 (8), 760-762. DOI: 10.1056/NEJMe2001126.
- (8) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K. Y.; et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, 181 (4), 894-904.e899. DOI: 10.1016/j.cell.2020.03.045.
- (9) Wang, M. Y.; Zhao, R.; Gao, L. J.; Gao, X. F.; Wang, D. P.; Cao, J. M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front Cell Infect Microbiol* **2020**, 10, 587269. DOI: 10.3389/fcimb.2020.587269.
- (10) Harvey, W. T.; Carabelli, A. M.; Jackson, B.; Gupta, R. K.; Thomson, E. C.; Harrison, E. M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S. J.; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **2021**, 19 (7), 409-424. DOI: 10.1038/s41579-021-00573-0.
- (11) Queirós-Reis, L.; Gomes da Silva, P.; Gonçalves, J.; Brancale, A.; Bassetto, M.; Mesquita, J. R. SARS-CoV-2 Virus-Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *International Journal of Molecular Sciences* **2021**, 22 (19), 10836.
- (12) Vogel, A. B.; Kanevsky, I.; Che, Y.; Swanson, K. A.; Muik, A.; Vormehr, M.; Kranz, L. M.; Walzer, K. C.; Hein, S.; Güler, A.; et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* **2021**. DOI: 10.1038/s41586-021-03275-y.

- (13) Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J.-H.; Portnoff, A. D.; et al. Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* **2020**, *370* (6520), 1089-1094. DOI: 10.1126/science.abe1502.
- (14) National Institutes of Health. *Antiviral Agents, Including Antibody Products*. 2024. <https://www.covid19treatmentguidelines.nih.gov/therapies/antivirals-including-antibody-products/summary-recommendations/#:~:text=Remdesivir%20and%20ritonavir%2Dboosted%20nirmatrelvir,the%20treatment%20of%20COVID%2D19>. (accessed 2024/06/27).
- (15) Staufer, O.; Gupta, K.; Hernandez Bücher, J. E.; Kohler, F.; Sigl, C.; Singh, G.; Vasileiou, K.; Yagüe Relimpio, A.; Macher, M.; Fabritz, S.; et al. Synthetic virions reveal fatty acid-coupled adaptive immunogenicity of SARS-CoV-2 spike glycoprotein. *Nature Communications* **2022**, *13* (1), 868. DOI: 10.1038/s41467-022-28446-x.
- (16) Wang, Q.; Meng, F.; Xie, Y.; Wang, W.; Meng, Y.; Li, L.; Liu, T.; Qi, J.; Ni, X.; Zheng, S.; et al. In Silico Discovery of Small Molecule Modulators Targeting the Achilles' Heel of SARS-CoV-2 Spike Protein. *ACS Cent Sci* **2023**, *9* (2), 252-265. DOI: 10.1021/acscentsci.2c01190.
- (17) Fonnesu, R.; Thunuguntla, V.; Veeramachaneni, G. K.; Bondili, J. S.; La Rocca, V.; Filipponi, C.; Spezia, P. G.; Sidoti, M.; Plicanti, E.; Quaranta, P.; et al. Palmitoylethanolamide (PEA) Inhibits SARS-CoV-2 Entry by Interacting with S Protein and ACE-2 Receptor. *Viruses* **2022**, *14* (5). DOI: 10.3390/v14051080.
- (18) Goc, A.; Niedzwiecki, A.; Rath, M. Polyunsaturated ω -3 fatty acids inhibit ACE2-controlled SARS-CoV-2 binding and cellular entry. *Sci Rep* **2021**, *11* (1), 5207. DOI: 10.1038/s41598-021-84850-1.
- (19) Xiao, T.; Lu, J.; Zhang, J.; Johnson, R. I.; McKay, L. G. A.; Storm, N.; Lavine, C. L.; Peng, H.; Cai, Y.; Rits-Volloch, S.; et al. A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent in vitro. *bioRxiv* **2020**. DOI: 10.1101/2020.09.18.301952.
- (20) Xia, S.; Yan, L.; Xu, W.; Agrawal, A. S.; Algaissi, A.; Tseng, C.-T. K.; Wang, Q.; Du, L.; Tan, W.; Wilson, I. A.; et al. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Science advances* **2019**, *5* (4), eaav4580-eaav4580. DOI: 10.1126/sciadv.aav4580 PubMed.
- (21) Toelzer, C.; Gupta, K.; Yadav, S. K. N.; Borucu, U.; Davidson, A. D.; Kavanagh Williamson, M.; Shoemark, D. K.; Garzoni, F.; Staufer, O.; Milligan, R.; et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **2020**, *370* (6517), 725-730. DOI: 10.1126/science.abd3255.
- (22) Queirós-Reis, L.; Mesquita, J. R.; Brancale, A.; Bassetto, M. Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein – Confirmed and Potential Ligands. *Journal*

of *Chemical Information and Modeling* **2023**, 63 (23), 7282-7298. DOI: 10.1021/acs.jcim.3c00803.

(23) Toelzer, C.; Gupta, K.; Yadav, S. K. N.; Hodgson, L.; Williamson, M. K.; Buzas, D.; Borucu, U.; Powers, K.; Stenner, R.; Vasileiou, K.; et al. The free fatty acid binding pocket is a conserved hallmark in pathogenic coronavirus spike proteins from SARS-CoV to Omicron. *Science Advances* **2022**, 8 (47), eadc9179. DOI: doi:10.1126/sciadv.adc9179.

(24) Day, C. J.; Bailly, B.; Guillon, P.; Dirr, L.; Jen, F. E.-C.; Spillings, B. L.; Mak, J.; Itzstein, M. v.; Haselhorst, T.; Jennings, M. P. Multidisciplinary Approaches Identify Compounds that Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2–ACE2 Receptor Interactions. *mBio* **2021**, 12 (2), 10.1128/mbio.03681-03620. DOI: doi:10.1128/mbio.03681-20.

(25) Shoemark, D. K.; Colenso, C. K.; Toelzer, C.; Gupta, K.; Sessions, R. B.; Davidson, A. D.; Berger, I.; Schaffitzel, C.; Spencer, J.; Mulholland, A. J. Molecular Simulations suggest Vitamins, Retinoids and Steroids as Ligands of the Free Fatty Acid Pocket of the SARS-CoV-2 Spike Protein*. *Angew Chem Int Ed Engl* **2021**, 60 (13), 7098-7110. DOI: 10.1002/anie.202015639.

(26) Carino, A.; Moraca, F.; Fiorillo, B.; Marchianò, S.; Sepe, V.; Biagioli, M.; Finamore, C.; Bozza, S.; Francisci, D.; Distrutti, E.; et al. Hijacking SARS-CoV-2/ACE2 Receptor Interaction by Natural and Semi-synthetic Steroidal Agents Acting on Functional Pockets on the Receptor Binding Domain. *Frontiers in Chemistry* **2020**, 8, Original Research. DOI: 10.3389/fchem.2020.572885.

(27) Berman, H. M.; Westbrook, J. D.; Feng, Z.; Gilliland, G. L.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Protein Data Bank. **2000**, N/a (N/a), N/a..

(28) Xu, C.; Wang, Y.; Liu, C.; Zhang, C.; Han, W.; Hong, X.; Wang, Y.; Hong, Q.; Wang, S.; Zhao, Q.; et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances* **2021**, 7 (1), eabe5575. DOI: doi:10.1126/sciadv.abe5575.

(29) Arora, P.; Pöhlmann, S.; Hoffmann, M. Mutation D614G increases SARS-CoV-2 transmission. *Signal Transduction and Targeted Therapy* **2021**, 6 (1), 101. DOI: 10.1038/s41392-021-00502-w.

(30) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, 25 (13), 1605-1612. DOI: 10.1002/jcc.20084.

(31) Chemical Computing Group, I. *Molecular Operating Environment (MOE 2019.10)*. Chemical Computing Group, Inc., <http://www.chemcomp.com>. Accessed: 2023/05/15

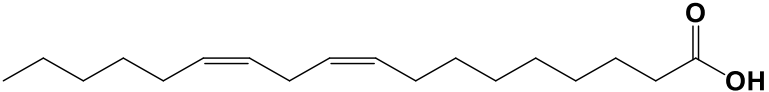
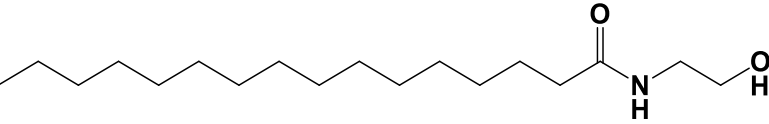
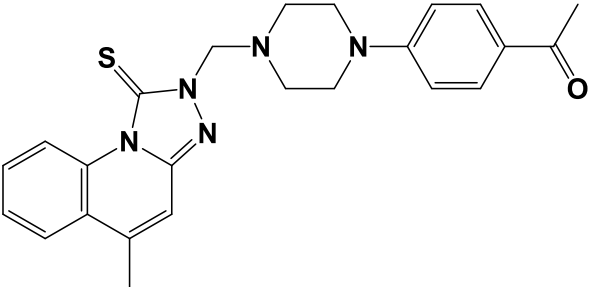
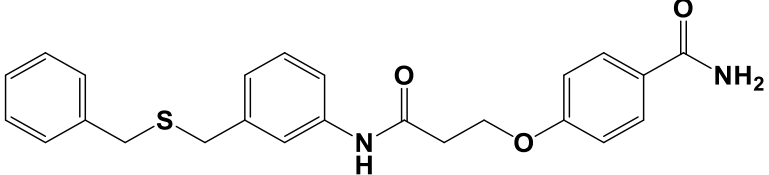
(32) Enamine Ltd. *Screening Collection*. <https://enamine.net/compound-collections/screening-collection>. 2022/02/08

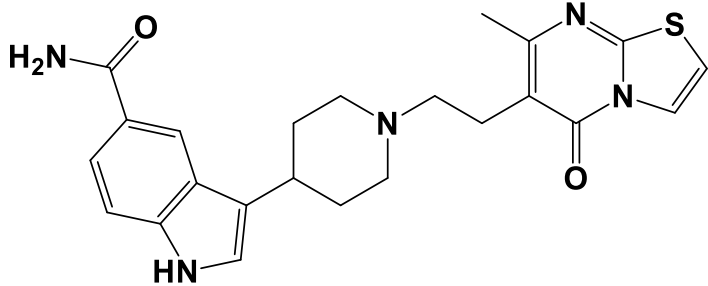
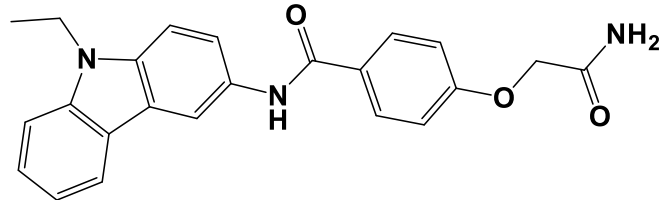
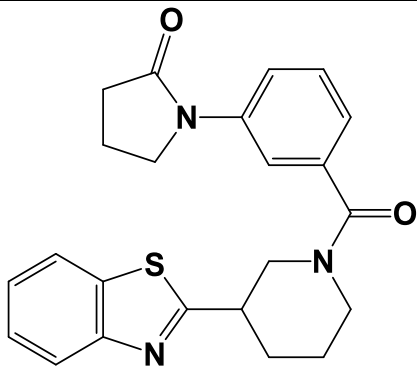
- (33) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, *47* (7), 1739-1749. DOI: 10.1021/jm0306430.
- (34) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *Journal of Medicinal Chemistry* **2006**, *49* (21), 6177-6196. DOI: 10.1021/jm051256o.
- (35) Korb, O.; Stützle, T.; Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* **2009**, *49* (1), 84-96. DOI: 10.1021/ci800298z.
- (36) OpenEye. OEDOCKING 4.3.1.0. *Cadence Molecular Sciences, Inc.*
- (37) Ferla, S.; Netzler, N. E.; Ferla, S.; Veronese, S.; Tuipulotu, D. E.; Guccione, S.; Brancale, A.; White, P. A.; Bassetto, M. In silico screening for human norovirus antivirals reveals a novel non-nucleoside inhibitor of the viral polymerase. *Sci Rep* **2018**, *8* (1), 4129. DOI: 10.1038/s41598-018-22303-y.
- (38) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **2017**, *7* (1), 42717. DOI: 10.1038/srep42717.
- (39) Zhu, N.; Wang, W.; Liu, Z.; Liang, C.; Wang, W.; Ye, F.; Huang, B.; Zhao, L.; Wang, H.; Zhou, W.; et al. Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. *Nature Communications* **2020**, *11* (1), 3910. DOI: 10.1038/s41467-020-17796-z.
- (40) Lei, C.; Yang, J.; Hu, J.; Sun, X. On the Calculation of TCID₅₀ for Quantitation of Virus Infectivity. *Viol Sin* **2021**, *36* (1), 141-144. DOI: 10.1007/s12250-020-00230-5.
- (41) Chen, D. Y.; Turcinovic, J.; Feng, S.; Kenney, D. J.; Chin, C. V.; Choudhary, M. C.; Conway, H. L.; Semaan, M.; Close, B. J.; Tavares, A. H.; et al. Cell culture systems for isolation of SARS-CoV-2 clinical isolates and generation of recombinant virus. *iScience* **2023**, *26* (5), 106634. DOI: 10.1016/j.isci.2023.106634.
- (42) Ogando, N. S.; Dalebout, T. J.; Zevenhoven-Dobbe, J. C.; Limpens, R.; van der Meer, Y.; Caly, L.; Druce, J.; de Vries, J. J. C.; Kikkert, M.; Bárcena, M.; et al. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J Gen Virol* **2020**, *101* (9), 925-940. DOI: 10.1099/jgv.0.001453.
- (43) Gironi, B.; Kahveci, Z.; McGill, B.; Lechner, B. D.; Pagliara, S.; Metz, J.; Morresi, A.; Palombo, F.; Sassi, P.; Petrov, P. G. Effect of DMSO on the Mechanical and Structural Properties of Model and Biological Membranes. *Biophys J* **2020**, *119* (2), 274-286. DOI: 10.1016/j.bpj.2020.05.037.

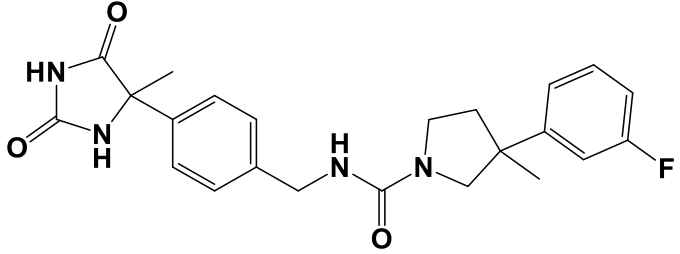
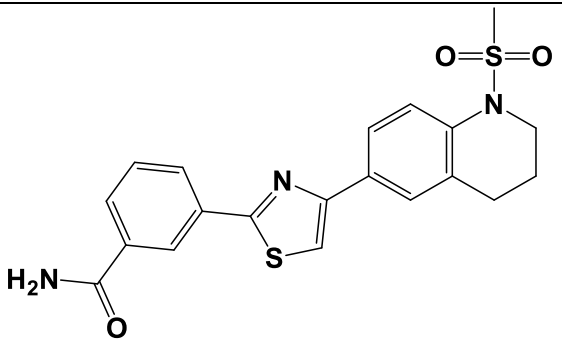
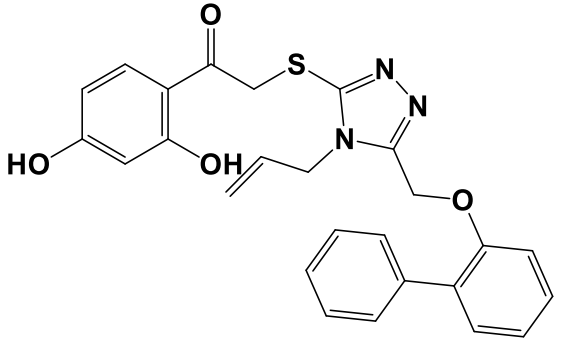
- (44) Saito, A.; Irie, T.; Suzuki, R.; Maemura, T.; Nasser, H.; Uriu, K.; Kosugi, Y.; Shirakawa, K.; Sadamasu, K.; Kimura, I.; et al. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* **2022**, *602* (7896), 300-306. DOI: 10.1038/s41586-021-04266-9.
- (45) Willett, B. J.; Grove, J.; MacLean, O. A.; Wilkie, C.; De Lorenzo, G.; Furnon, W.; Cantoni, D.; Scott, S.; Logan, N.; Ashraf, S.; et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nature Microbiology* **2022**, *7* (8), 1161-1179. DOI: 10.1038/s41564-022-01143-7.
- (46) Kojabad, A. A.; Farzanehpour, M.; Galeh, H. E. G.; Dorostkar, R.; Jafarpour, A.; Bolandian, M.; Nodooshan, M. M. Droplet digital PCR of viral DNA/RNA, current progress, challenges, and future perspectives. *J Med Virol* **2021**, *93* (7), 4182-4197. DOI: 10.1002/jmv.26846.
- (47) Taylor, S. C.; Laperriere, G.; Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Scientific Reports* **2017**, *7* (1), 2409. DOI: 10.1038/s41598-017-02217-x.
- (48) Enamine Ltd. *REAL database*. Available from: <https://enamine.net/compound-collections/real-compounds/real-database>. Accessed 2024/06/27.
- (49) Bioscience, B. *SARS-CoV-1 Spike Trimer (S1+S2) ACE2 Inhibitor Screening Colorimetric Assay Kit (Catalog No. 78012)*. <https://bpsbioscience.com/sars-cov-1-spike-trimer-s1-s2-ace2-inhibitor-screening-colorimetric-assay-kit-78012> Accessed: 2022/09/22.
- (50) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260-1263. DOI: 10.1126/science.abb2507.
- (51) European Virus Archive Global (EVAg). **2024**.
- (52) Corman, V. M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M. L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* **2020**, *25* (3). DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.

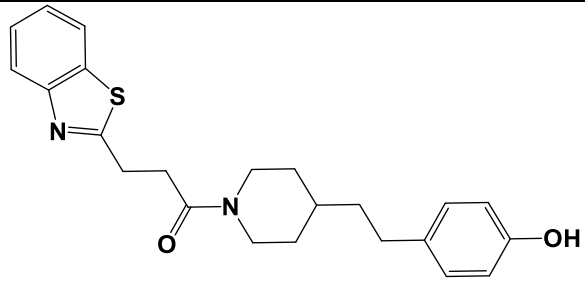
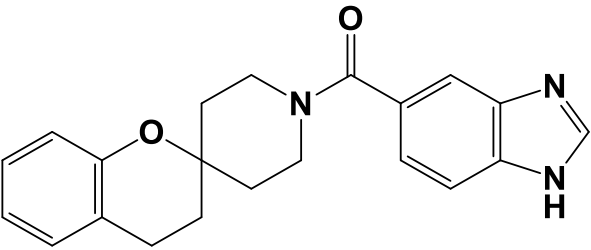
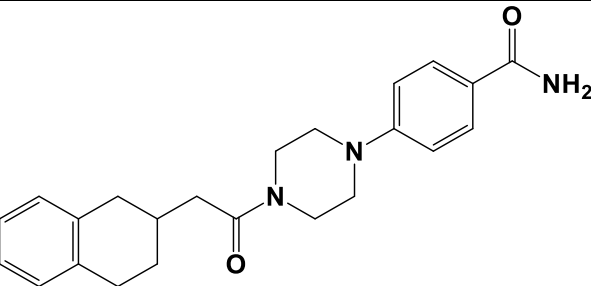
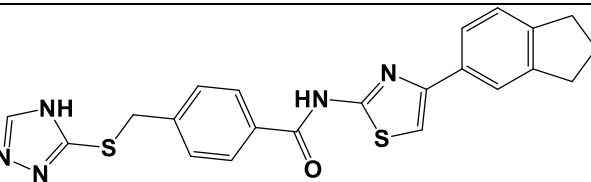
Supplementary Material

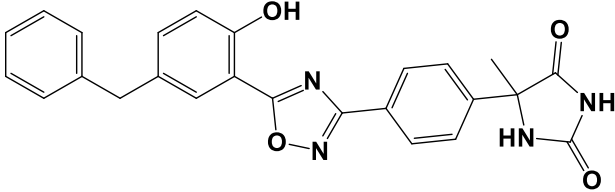
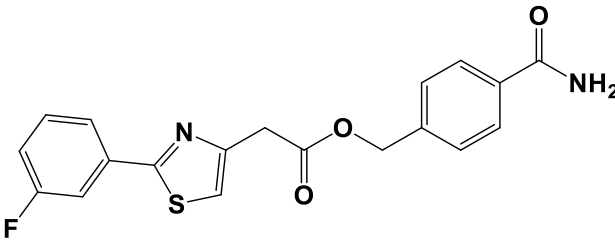
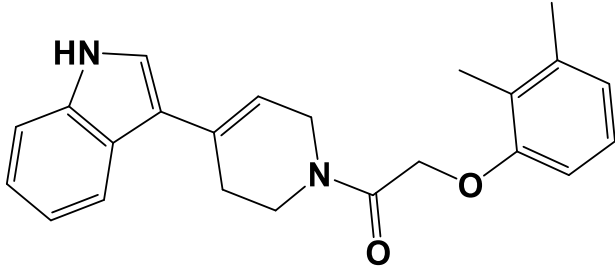
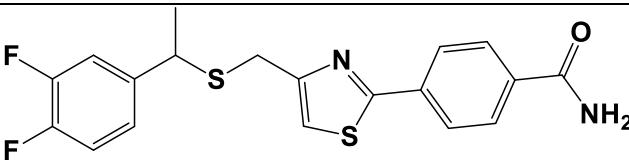
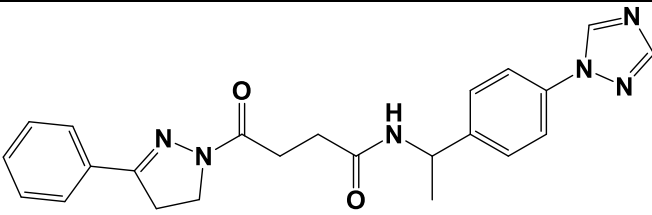
Table S1 – Chemical structures of compounds reported, including molecular Weight, calculated LogP, H-bond accepting groups, H-bond donating groups, as well as PAINS and BRENK analysis.

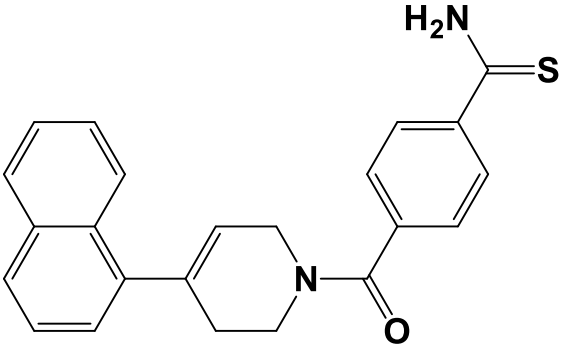
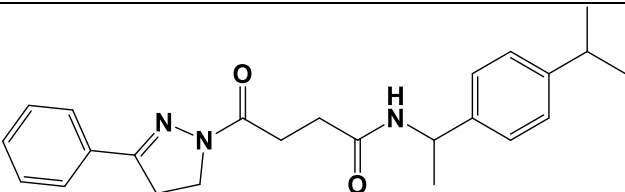
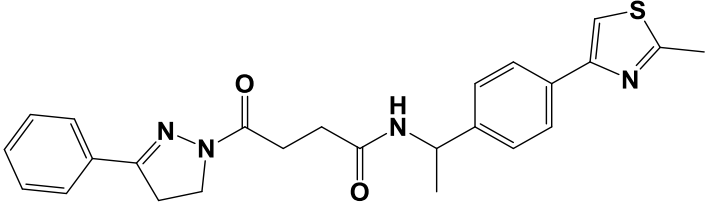
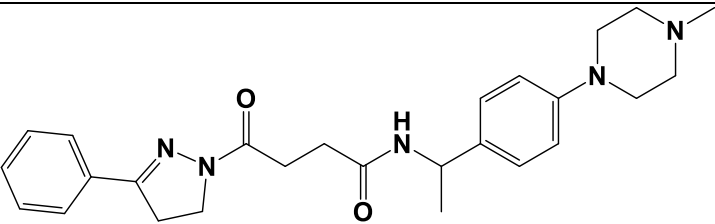
Molecule	SMILES	MW	PAINS	BRENK	H-bond acceptors	H-bond donors	LogP
Linoleic acid		280.45	0	0	2	1	5.88
PEA		299.49	0	0	2	2	4.76
1		431.55	0	1	3	0	3.55
2		420.52	0	0	3	2	3.74

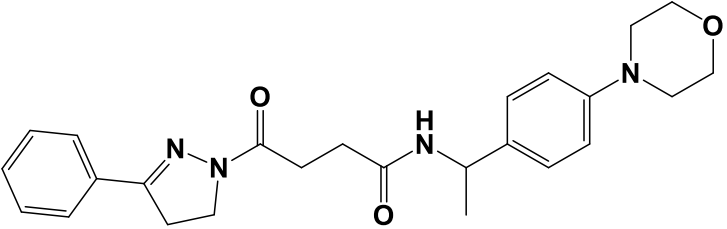
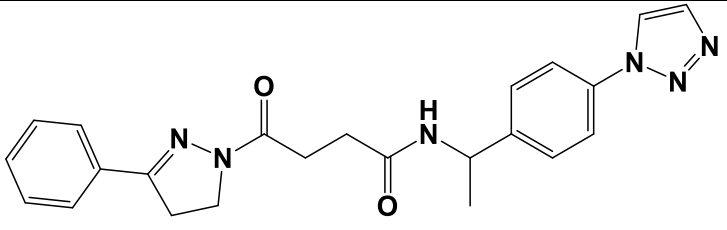
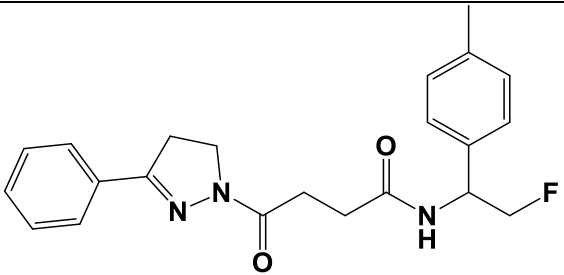
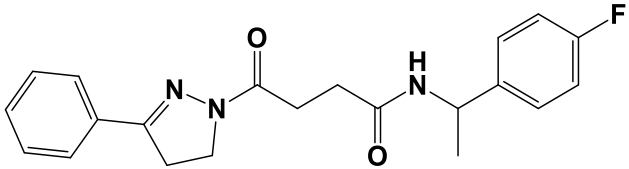
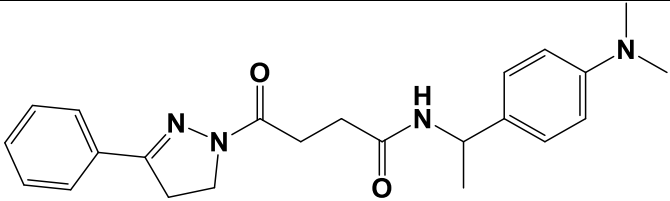
3		435.54	0	0	4	2	2.86
4		387.43	0	0	3	2	3.03
5		405.51	0	0	3	0	3.72

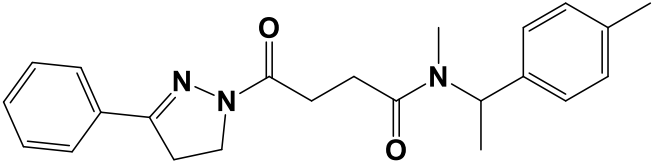
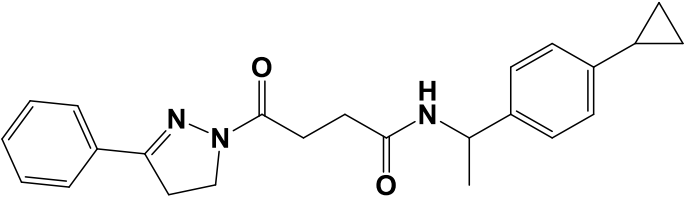
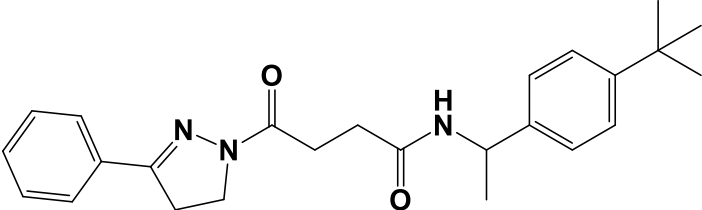
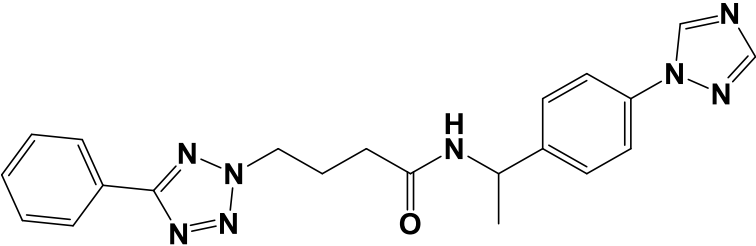
6		424.47	0	1	4	3	2.65
7		413.51	0	0	4	1	2.91
8		473.54	0	1	6	2	4.16

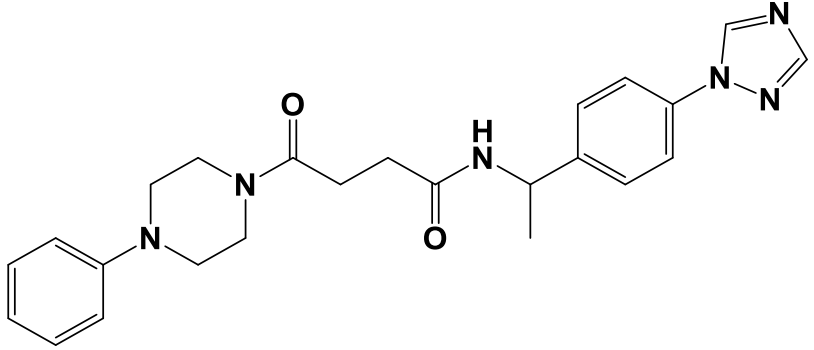
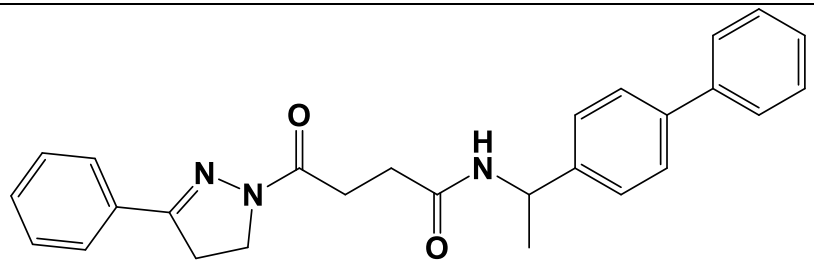
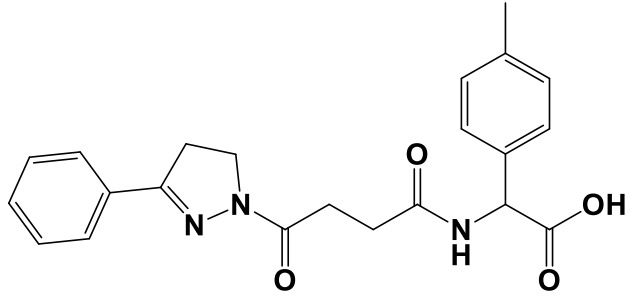
9		394.53	0	0	3	1	4.43
10		347.41	0	0	3	1	3.11
11		377.48	0	0	2	1	2.67
12		433.55	0	0	4	2	4.12

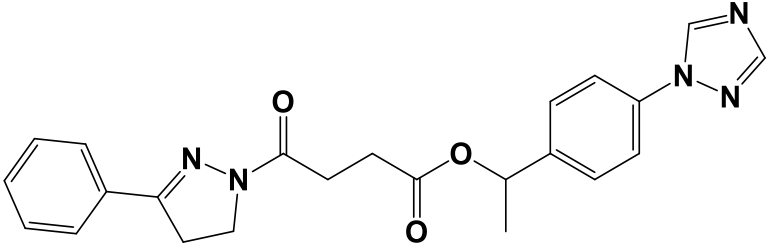
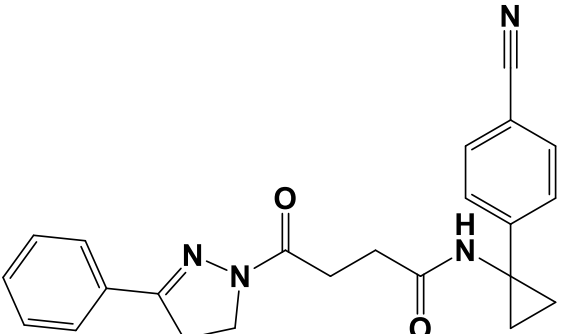
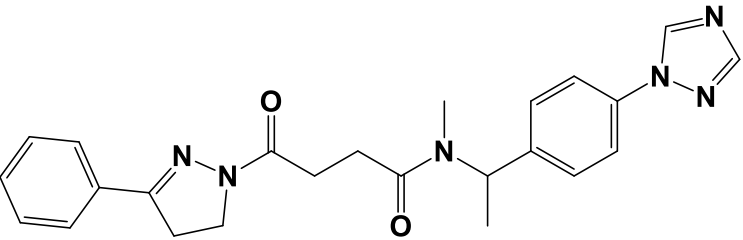
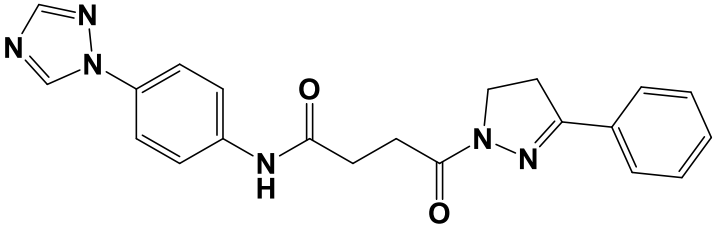
13		440.45	0	1	6	3	3.39
14		370.4	0	0	5	1	3.38
15		360.45	0	0	2	1	3.89
16		390.47	0	0	4	1	4.57
17		416.48	0	0	5	1	2.36

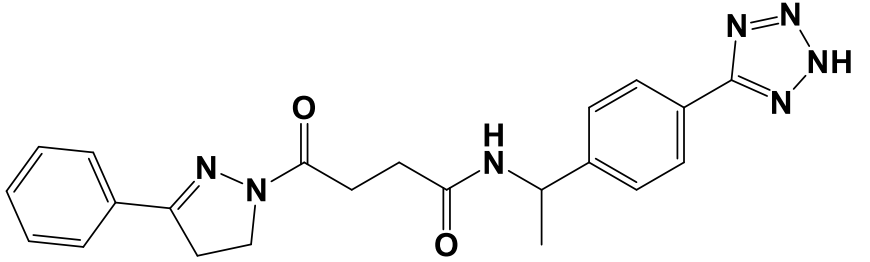
18		372.48	0	1	1	1	4.02
19		391.51	0	0	3	1	3.79
20		446.56	0	0	4	1	3.91
21		447.57	2	0	4	1	2.6

22		434.53	2	0	4	1	2.66
23		416.48	0	0	5	1	2.29
24		381.44	0	0	4	1	3.24
25		367.42	0	0	4	1	3.17
26		392.49	2	0	3	1	2.88

27		377.48	0	0	3	0	3.34
28		389.49	0	0	3	1	3.63
29		405.53	0	0	3	1	4.06
30		402.45	0	0	6	1	2.44

31		432.52	0	0	4	1	2.19
32		425.52	0	0	3	1	4.15
33		393.44	0	0	5	2	2.44

34		417.46	0	0	6	0	2.83
35		386.45	0	0	4	1	2.73
36		430.5	0	0	5	0	2.62
37		388.42	0	0	5	1	1.99

38		417.46	0	0	6	2	2.14
----	--	--------	---	---	---	---	------

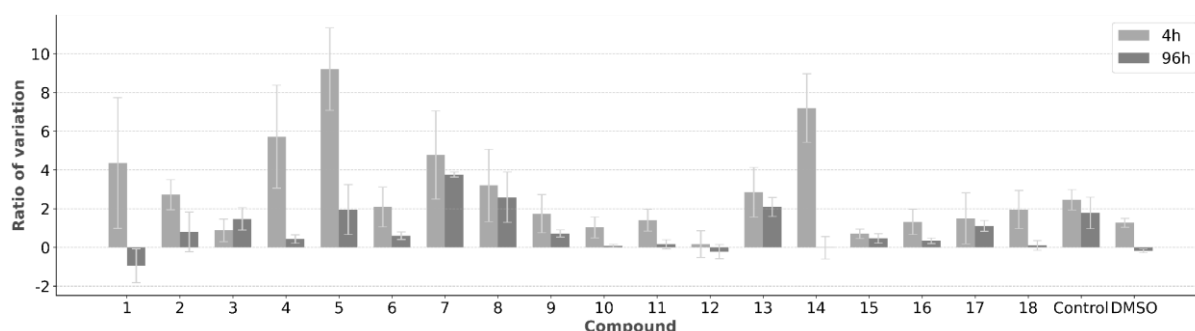


Figure S1 – Preliminary cytotoxicity evaluation performed on Vero E6 cells, with the screened compounds, a growth control and a cytotoxicity control. The initial growth at 0h is a baseline against which growth at 4h and 96h is compared, as observed in the experimental protocol for viral infection in the proceeding cells assays.

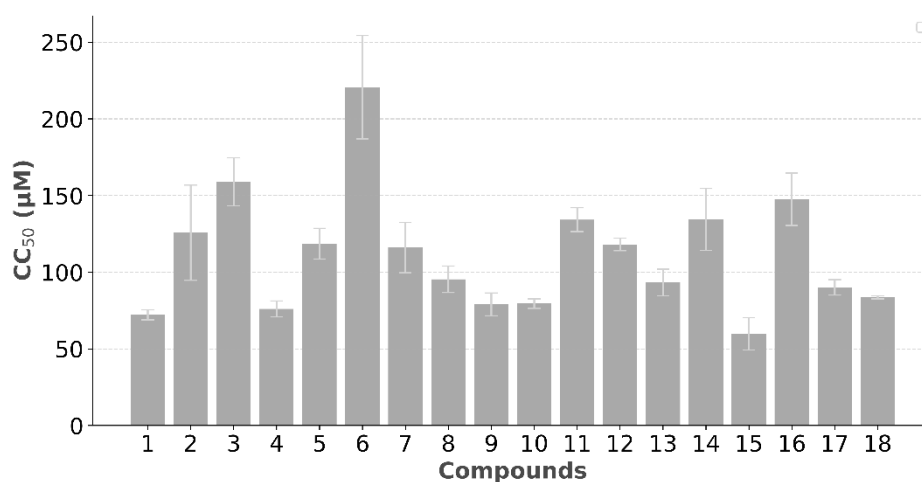


Figure S2 – Cytotoxicity of screened compounds (X-axis) in Vero E6 cells, measured by the CellTiter-Glo® method. Cytotoxicity is represented by CC₅₀ in the Y-axis, the concentration of test compounds required to reduce cell viability by 50%. The bars represent the mean ± SEM from three experimental repeats. Abbreviations: CC₅₀ – cytotoxic concentration 50%.

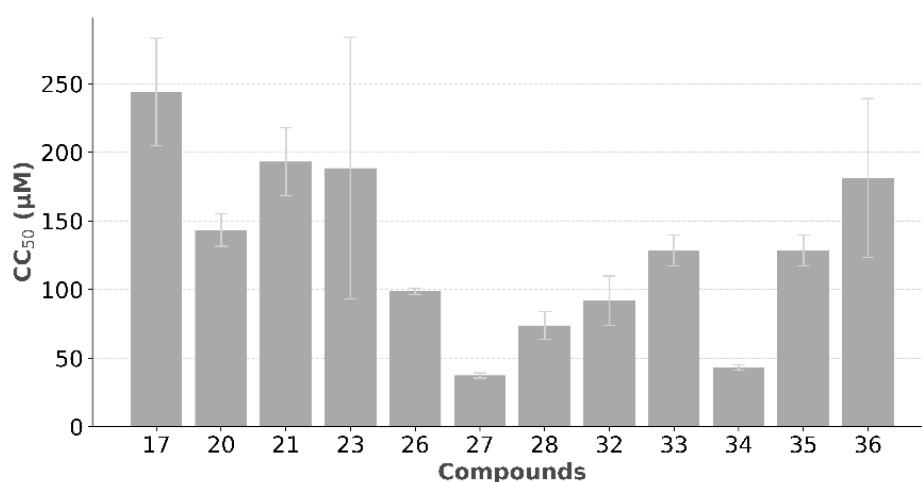


Figure S3 – Cytotoxicity of screened compounds (X-axis) in Vero E6 cells, measured by the CellTiter-Glo® method. Cytotoxicity is represented by CC_{50} in the Y-axis, the concentration of test compounds required to reduce cell viability by 50%. The bars represent the mean $\pm$ SEM from three experimental repeats. Abbreviations: CC_{50} – cytotoxic concentration 50%.

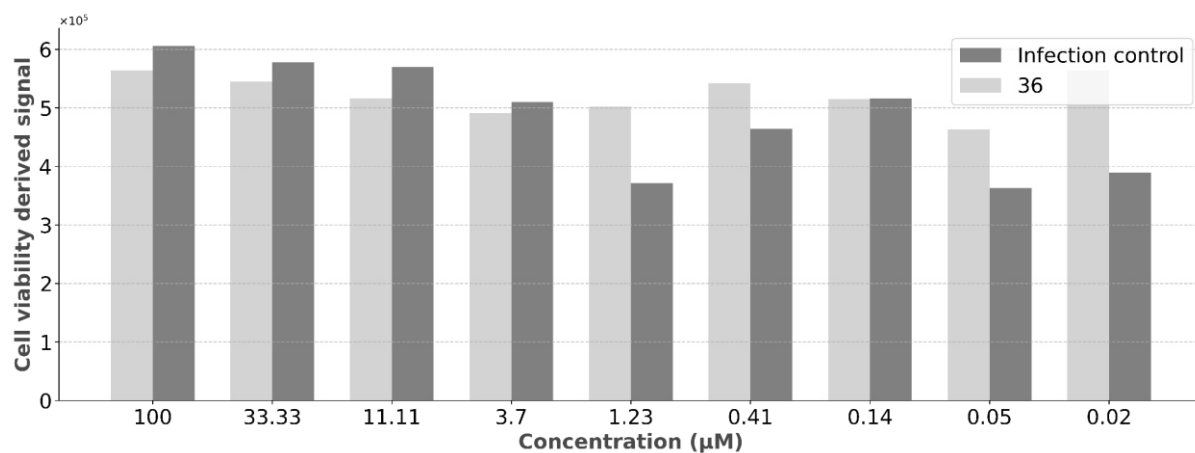


Figure S4 – Infection of Vero E6 cells with SARS-CoV-2 Omicron Variant at multiple concentrations. Cell viability was evaluated using the CellTiter-Glo® method.



Chapter IV

Discussion

“The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.”

Stephen Hawking

Chapter IV – Discussion

4.1 General discussion

In the early stages of the COVID-19 pandemic, the need for therapeutic interventions led to a focus on repurposing existing drugs. At the same time, a global race to develop vaccines marked a significant turning point, dramatically reducing severe cases and deaths. One major focus of research was the S protein, a critical protein in SARS-CoV-2 responsible for cell recognition, infection ability, and cell tropism [1]. This focus proved effective, as S-targeted vaccines demonstrated efficacy, with doses administered worldwide and leading to reduced hospitalization rates, further establishing the S protein as a prime target for ongoing therapeutic efforts.

4.1.1. Exploration of S Protein and identification of pharmacological targets

4.1.1.1 Exploration of S protein

This project set out to explore the S glycoprotein as a determinant for viral infection, aiming to study its behaviour and identify new bioactive scaffolds. However, in early 2020, the structure and behaviour were still partly unknown, although the overall behaviour was comparable to previous CoV outbreaks, MERS-CoV and particularly SARS-CoV [2]. Past CoV infections pointed to potential receptors for interaction, such as ACE2 in SARS-CoV or DPP-IV in MERS-CoV, and to the existence of extensive conformational changes elicited by cell recognition [2]. Importantly, previous research on CoVs, due to structural similarities, maintained wide applications for SARS-CoV-2. The rapid determination of the S protein structure by Cryo-EM, achieved within two weeks of genome identification, exemplified the application of previous knowledge to a swift response during the pandemic [1-3]. During the first year of the pandemic, along with the focus of many viral proteins, many authors also extensively studied the S protein, and analysing this information would be the basis to identify a suitable target for drug discovery [1, 4-8].

In this initial stage, the first step was to clarify the virus-host interaction and the conformational changes across the protein, leading to the fusion and cell infection [8]. The complex formed in cell recognition, due to the interaction between ACE2 and the RBD, affects affinity and is critical for infectivity, with consequences in the epidemiology. This is therefore a natural start point to affect S protein activity [8]. Directly inhibiting the S protein-ACE2 interaction has proved successful, particularly as this is the mechanism behind most approved vaccines and immune system neutralization of the virus, by which antibodies compete to bind the RBD active conformation, and block interaction with ACE2 [9-11].

However, despite the importance of the RBD-ACE2 interaction as a target to affect cell recognition, the S protein has additional functions in budding and cell fusion, mainly through conformational changes [5,6].

The S protein requires two activation steps, a cleavage step at the multibasic S1/S2 site and a second cleavage at the S2' site. The initial step is performed by furin and promotes a stark increase in spike trimers with one RBD in the open conformation, which primes the S protein for cell recognition [6].

Once this cleavage event is performed, the S protein becomes highly sensitive to the presence of ACE2, as binding with the RBD has an allosteric effect that destabilizes the S1/S2 interaction. Particularly in the first RBD-ACE2 binding event, a significant conformational change is observed, with a rigid body rotation and movement in the center of mass of the S protein, that exposes unbound RBDs [8]. As these become available for interaction, the new S conformation is stabilized due to the high affinity with ACE2. The conformational changes and consecutive binding to ACE2 continually weakens the S1–S2 interaction, eventually releasing the S2 subunit, that becomes free to perform helical rearrangements required for viral–host membrane fusion and the post fusion conformation. When compared with previous CoVs, two features contribute to increased infectivity in SARS-CoV-2, the multibasic S1/S2 site recognized by furin, which enables a more efficient cleavage and activation of the S protein, and a particularly low energy barrier to the final conformational change, facilitating the fusion event [8].

However, membrane fusion cannot proceed without the second cleavage event, performed by TMPRSS2 or cathepsin L, that exposes the fusion peptide in the S2 subunit [12]. Cleavage by TMPRSS2 is performed at the cell surface, which enables viral fusion to adjacent cells, while cathepsin L is mainly relevant for endosomal entry.

In SARS-CoV-2, each RBD can move independently between an exposed state and a buried state, with this ratio influenced by multiple factors, such as the cleavage event and pH [5, 13, 14]. In the presence of ACE2 and at pH 7.4, 84% of S protein trimers bind at least two molecules of ACE2, whereas at pH 5.5 this is reduced to 69% [13,14]. However, in the absence of ACE2 at pH 5.5, no S protein trimers showed more than one RBD in the conformation available for interaction, which underscores the effect of ACE2 in stabilizing the open conformation. At low lysosomal pH (pH 4.0 and 4.5), even after cleavage is performed, all RBDs are in the down conformation. As acidity increases, at pH 5.5, a region in subunit S2 near the fusion peptide, rich in aspartic acid residues, undergoes a pH-dependent switch, twisting the orientation of disulphide bonds in the structure and causing RBD movement. The domain switch, caused by the pH in endosomal entry, includes aspartic acid residues at position 830, 839, 843, and 848. The affected areas include the Heptad repeat 1 (HR1) (in the

S2 subunit) and the NTD in the adjacent S1 subunit, locking the RBDs in the down position, representing a strategy of immune evasion that SARS-CoV-2 uses to avoid neutralization by RBD-up-recognizing Abs [8].

Further viral mutations during the pandemic period, particularly with mutation D614G, increase the flexibility of the S protein, facilitating the cleavage event, and reduce even more the energy barrier for S1/S2 separation due to the weakened interactions established by the side chain of glycine, in comparison with aspartic acid, increasing SARS-CoV-2 infectivity [7, 8].

4.1.1.2 Identification of target sites

The first potential target emerging from this in-depth analysis was the RBD, and indeed affecting the interaction with ACE2 has shown inhibitory activity, with competing antibodies achieving virus neutralization. On the other hand, stabilization of the prefusion conformation, with the RBDs in the down conformation yielded effective vaccines. However, as the portion of viral genome most affected by mutations, the S protein has adapted to escape antibody neutralization, and therefore any drugs binding in this area would also be highly susceptible to development of resistance [10, 11, 15-17].

While the first step for cell infection is recognition of target, blocking fusion with the host cell can also severely disrupt viral infection. This step is mainly performed by the S2 subunit, containing the fusion peptide, heptad repeat 1, central helix, connector domain, heptad repeat 2 (HR2), transmembrane domain, and cytoplasmic tail [18]. The interaction between HR1 and HR2 forms a six-helix bundle, that brings viral and cellular membranes together for fusion, resulting in viral entry into host cells [18]. The interaction of HR1 and HR2 has been explored, with a peptide based fusion inhibitor (CP1 was identified in 2003) during the SARS-CoV outbreak, derived from the HR2 domain, blocking the interaction and consequently the fusion event and cell entry [19, 20]. During the SARS-CoV-2 pandemic, EK1 was designed with the same rational, with inhibitory activity in SARS-CoV-2 pseudo virions, including multiple SARS-CoV-2 variants [19, 20]. However, being a peptide, it intrinsically has significant limitations, particularly in regard to pharmacokinetic properties, with oral administration limited, and a short half-life, due to digestion enzymes and proteolytic enzymes, respectively. Additionally, specific storing conditions are required due to low stability. However, this is a target with a strong potential, given that it can effectively neutralize the virus ability to replicate early in the lifecycle [21]. Still considering the S2 as a possible target, a small-molecule solution appeared during drug repurposing efforts, umifenovir, an antiviral designed for Influenza viruses, identified as an effective antiviral against SARS-CoV-2 (currently approved for Influenza infections in Russia and China) [22]. Later studies identified the mechanism of action: it acts as a fusion inhibitor by binding the S2 subunit of SARS-CoV-2, blocking membrane fusion,

which leads to lysosomal degradation of the virus, while showing micromolar *in vitro* activity against SARS-CoV-2, along with other human CoVs [22]. However, further clinical trials showed little benefit in the clinical outcome of hospitalized patients with mild/moderate COVID-19 [23, 24]. Nevertheless, the potential for pan-CoV activity and low evolutionary pressure (conserved among CoVs), makes it a good pharmacological target [22, 25].

The final target site considered was the Fatty Acid Binding Pocket (FABP) [26]. This pocket was identified with a ligand, linoleic acid (LA), showing a stabilization of the S protein in an inactive conformation, with the majority of RBDs buried and inaccessible for interaction, and a stark change in S protein dynamics [26]. This is a mechanism reminiscent to the effect of reduced pH, that hides the RBDs. Similarly, S protein availability for interaction becomes limited, cell recognition is inefficient, and cell infection cannot proceed [7]. The FABP is located between two RBD interfaces, which results in three pockets per spike trimer, and is conserved in highly pathogenic CoVs, while still partly identified in other human CoVs, with a strong impact in the virus lifecycle [26, 27]. Biological assays have shown bioactivity for LA and other polyunsaturated fatty acids (linolenic acid and arachidonic acid), while in human cells, LA supplementation synergizes with the COVID-19 drug remdesivir, suppressing SARS-CoV-2 replication [26]. Furthermore, epidemiological studies have shown a link between consumption of ω -6 and ω -9 fatty acids with reduced mortality during the pandemic [28, 29].

The direct inhibition of the S-ACE2 interaction, either targeting the human receptor or the S protein, was excluded due to the expected emergence of resistance. Previous strategies targeting this area, particularly antibodies and soluble ACE2 show bioactivity but place a large evolutionary pressure in the RBD, in an area already favoured for mutations in new viral variants [11]. Fusion inhibitors provided a potential target for pan CoV activity, with both approaches here identified focusing on the HR1-HR2 bundle of the S2 subunit. Since this approach blocks the conformational change into the post fusion conformation that elicits cell fusion, it can have a strong effect in the virus lifecycle by blocking cell infection [19]. Furthermore, umifenovir is a small molecule, showing that this might be a good target for structure-based drug design. However, the precise binding site is unclear and neither crystallographic nor Cryo-EM structures accurately representing this binding event have been determined, limiting further research. Finally, the FABP is a well-defined pocket in the RBD, with confirmed bioactivity and available crystallographic structures, required for accurate computer-based studies. Additional *in-vitro* assays have revealed that multiple fatty acids, both poly- and monounsaturated, along with retinoids and vitamins (A and K), can affect S protein binding and cell infection [27, 29, 30]. Different compounds and fatty acids have varying activity levels, showing a relationship between structural features and activity levels, and a

potential for activity improvement [29]. Therefore, the FABP was selected as the focus of the remaining research in this study.

4.1.2 Hit molecule identification and optimization

The main objective at this stage was to explore the FABP as a drug target, and to identify new bioactive scaffolds suitable for drug development, by mimicking the biological effect of LA binding. Although previous research has focused on the FABP, most confirmed bioactive compounds have limitations, such as fatty acids, displaying inadequate properties, and therefore the aim is to identify new scaffolds more suitable for translational or clinical applications.

4.1.2.1 Fatty acid binding pocket - confirmed ligands

Since the identification of the FABP in late 2020, subsequent research has substantially advanced the understanding of this site, including the discovery of new ligands. These insights have provided critical information regarding the binding dynamics.

By systematically analysing the binding modes of confirmed ligands—identifying favoured functional groups, scaffolds, and essential interactions—this approach aims to support the identification of new molecules with optimized binding properties.

The presence of fatty acids in the FABP, particularly LA, influences conformational rearrangements of the spike protein, impacting its infectivity and immunogenicity. Beyond the direct stabilization of the RBDs that affects ACE2 recognition, LA binding may induce distal effects within the S protein, modulating the behaviour of key residues involved in the major conformational changes preceding cell fusion. This effect is likely linked to the role of LA in the equilibrium between active and inactive RBD conformations, similar to pH-induced shifts, by stabilizing the inactive conformation and favouring a closed, less accessible RBD state [31]. Additionally, the FABP exhibits a highly conserved structure not only among SARS-CoV-2 variants of concern but also in related coronaviruses infecting intermediate hosts such as bats and pangolins. This conservation underscores its functional importance and potential as a therapeutic target [31].

Despite the experimental effect in the virus-host interaction detected for multiple compounds, only two have been confirmed to directly bind the FABP and exert a biological activity, lifitegrast and LA, while a third compound, SPC-14, has been co-crystalized in the S protein, but lacks biological assays to confirm its bioactivity [31]. When present in the medium, SPC-14 results in 48% of S protein population appearing in the closed conformation, with the remaining only showing one RBD exposed for interaction with ACE2 [31]. However, multiple additional compounds have been predicted to act in the pocket and show bioactivity in either binding

assays or reduced cell infection detected in cell-based assays [31]. Among these, compounds with a steroid backbone, namely dexamethasone, have shown reduced activity in cell-based assays, being used in the treatment of COVID-19, with a clear benefit of critically ill patients, although this might also be due to its activity as a corticosteroid and the wide anti-inflammatory action [32].

Among the confirmed ligands, common binding patterns include the ability to interact with hydrophobic residues buried in the pocket and with polar functional groups of amino acid residues that define the pocket entrance. The first confirmed ligand with large ring systems was lifitegrast, showing that long, flexible carbon chains typical of fatty acids are not an essential requirement for binding. This observation was later confirmed with SPC-14 [31]. Alternatively, bioactive compounds also show the presence of only up to two aromatic rings, attached to a more flexible backbone, as observed with LA, retinoids, and vitamin K [31]. Carbon chains and aromatic rings are predicted to have the ability to accommodate in the pocket and establish different nonbonded interactions. Particularly, the presence of aromatic rings is likely essential so that π interactions can be established with the numerous phenylalanine residues, buried, and spread across the pocket. Scaffolds capable of interacting with other hydrophobic amino acids that coat the deeper portions of the site are also favoured, as is the case for retinoids [32]. The carboxylic acid group appears to be an important feature to interact with the hydrophilic residues, but other functionalities, such as the sulfoxide in lifitegrast, also show the potential to make direct contacts with electrophilic residues in the vicinity. However, in the SPC-14 co-crystallized structure, the ligand is deeply buried in the pocket, and too far to directly interact with the entrance residues, which suggests that these interactions in the entrance are not a requirement for binding. Despite the structural insights, biological activity for SPC-14 is still lacking experimental validation of the inhibitory effect of Spike-ACE2 interaction or antiviral cell-based assays [31].

4.1.2.2 Preliminary virtual screening

The exploration of the FAPB started with a VS study aimed at identifying small molecules with therapeutic potential. A library of over 350,000 drug like molecules, the SPECS library [33], was screened against the co-crystallographic structure of the FAPB and LA, using the LA molecules as centroids in the SARS-CoV-2 Wuhan S protein. The ranking of molecules from the docking study followed a consensus scoring procedure, resulting in 26 compounds submitted to experimental testing. The validation step focused on two features: inhibitory activity of the SARS-CoV-2-ACE2 interaction and cytotoxicity evaluation, to ensure minimal toxic effects. The ability to block the virus-host interaction was tested using an ELISA-based inhibition assay, with two controls, LA and palmitoylethanolamide (PEA), while cytotoxicity was

evaluated in L929 murine cells. Among the 26 compounds selected, seven have shown the ability to affect the interaction more than PEA (14%), with a maximum of 34%, while none surpassed LA with 58% inhibitory activity.

The inhibitory activity observed for LA justified the use as a positive control, and despite the weaker effects observed for the screened compounds, these could also affect the S protein behaviour, through the FABP, as some still achieved higher inhibition than PEA.

Regarding a structural comparison between the screened compounds and LA, the presence of cyclic and aromatic rings in compounds is a clear divergence from the carbon chains. These large substituents can potentially interact with regions of the pocket rich in phenylalanine residues, by establishing π based interactions. As for the entrance of the pocket, terminal hydrophilic groups are present in active compounds and are predicted to establish H-bonds in this area. However, unlike LA, among the compounds with the highest activity, only one has a carboxylic acid. Thus, the carboxylic terminal group is not required for adequate interaction, as other terminal groups also show activity.

The FABP is formed between two adjacent RBDs, resulting in three pockets within each S protein trimer. Given the structural similarity among S protein monomers, it is reasonable to assume that the properties are consistent across pockets. This assumption, was however, also tested during this thesis, while the ability of compounds to effectively engage with all three pockets was a critical selection criterion. The binding poses by the most active compound showed consistency in predicted adopted conformations across all three pockets, forming identical hydrogen bonds and occupying similar regions. This consistency validates the hypothesis regarding pocket similarity.

This VS study identified several compounds that partially inhibit the virus-host interaction, with low toxicity, an effect that, at least partially, might be derived from an effect in the FABP, stabilizing the inactive conformation and reducing binding between the S protein and ACE2. This variation in activity for different compounds provided proof-of-concept that this target is druggable and can be modulated by small molecules. Furthermore, these results reinforce the reliability of computational approaches, such as VS. From an initial library of approximately 350,000 compounds, the experimental validation identified four molecules capable of inhibiting the virus-host interaction by over 25%. These findings highlight the precision of VS in narrowing down a vast chemical space to a few biologically active candidates.

To deepen our understanding of the observed effect and explore its potential applications more comprehensively, a revised VS approach was undertaken. This new procedure incorporated an in-depth study of the FABP, along with a larger and more diverse chemical library, aiming to identify additional compounds that might exhibit similar or enhanced activity.

4.1.2.3 S protein dynamics and FABP availability

To better understand fatty acid binding within the FABP, MD simulations were performed using the co-crystallized LA ligand. These simulations explored the modulatory effects of the FABP on the S protein, focusing on key binding interactions in this site. The full-length 3D structure of the spike protein was analysed in both the presence and absence of LA. The open conformation of the S protein exhibited the highest stability overall, while the presence of LA stabilized the closed conformation, achieving an intermediate level of stabilization. Throughout the MD simulations, the pocket remained accessible in both open and closed conformations. However, in the closed state, pockets detected in each frame included more FABP residues, suggesting ligand binding is more favourable in this conformation. Notably, the hydrophilic area of the pocket, critical for electrostatic interactions that stabilize the ligand, was only intermittently detected as part of the pocket. This indicates that while ligand binding is possible in both conformations, the closed state may offer a more consistent site for interaction.

Analysis of interaction patterns for confirmed ligands, combined with molecular dynamics evaluation of pocket availability, suggests that hydrophobic interactions within the FABP play a pivotal role in initial ligand binding. These interactions might even be more critical for binding than the hydrophilic interactions at the pocket entrance. Interestingly, the hydrophilic anchor appears to have a limited role during initial binding, further emphasizing the importance of the hydrophobic core.

These findings provide valuable guidance for new, potent, and broad-spectrum compounds. By targeting the FABP with strong hydrophobic interactions, such compounds could effectively disrupt cell recognition processes. This knowledge would directly inform the expanded VS study aimed at identifying novel bioactive scaffolds with enhanced binding affinities for the FABP.

4.1.2.4 Hit identification

The crystal structure of LA bound to the S protein was used to screen an expanded library of over 4,000,000 chemically diverse commercially available compounds, the Enamine Screening Collection library [34]. This broadened chemical variety aims to enhance the screening's ability to identify unique binding interactions and potential novel scaffolds. Similarly to the previous VS procedure, the ranking of molecules followed a consensus rescoring procedure. The best predicted affinity and analysis of drug-like properties resulted in a final selection of 18 compounds. An ELISA-based inhibition assay was applied to assess

the ability to affect the S-ACE2 interaction, while cell-based assays evaluated antiviral activity, by quantitatively assessing cytopathic effect.

During the docking procedure, each compound was assessed for its potential affinity to the FABP using five distinct scoring functions, selected to complement one another and address various parts of the docking problem. For initial pose determination, Glide HTVS was employed to efficiently evaluate the library (due to a reduced the number of intermediate conformations), with the top ranked compounds undergoing redocking with by Glide SP. This docking program refines ligand-receptor binding predictions while maintaining a balance between speed and accuracy, with these docking methods ensuring the identification of key molecular interactions, enabling a computationally efficient screening of large compound libraries [35]. However, Glide SP has innate inaccuracies, particularly a large number of false positives due to exceptions to physical chemistry principles [35, 36]. Therefore, to improve quality when ranking compounds for a final selection, a rescoring step is performed, complemented by high precision scoring functions. Glide XP is particularly useful at this stage, being designed to lend more weight to the physical aspects of the model, and a strong ability to detect false-positive poses with more accuracy in binding predictions, mainly focusing on the physical interactions (electrostatic and van der Waals forces). However, it is less accurate regarding hydrophobicity, intrinsically difficult for physics-based systems, since this property is entropic in nature [36, 37]. To solve this inaccuracy, PLANTS offers the piecewise linear potential (PLP) approach, making it effective at scoring flexible ligands and with increased accuracy for hydrophobic interactions, since the strong physical aspect of GLIDE XP can lead to low affinity in good ligands [36]. Native ligand geometries have been found to exhibit high strain energies, which are severely penalized by physic-based systems. Given that these functions are only used to perform rescoring, they are limited by the rigid-receptor framework [36]. In this context, the PLANTS scoring function is particularly efficient at dealing with ligand flexibility [38]. Finally, ScorePose, an empirical scoring approach, emphasizes ligand-receptor complementarity by focusing on shape and electrostatics, making it a more balanced scoring system [39].

Regarding the binding assays, the maximum detected inhibition was 28%, with six compounds surpassing PEA activity. In the cell based viral infection assays, **17** (Figure 8) was identified as the most active compound, reducing viral induced CPE by around 30%, along with a reduction in viral particles released.

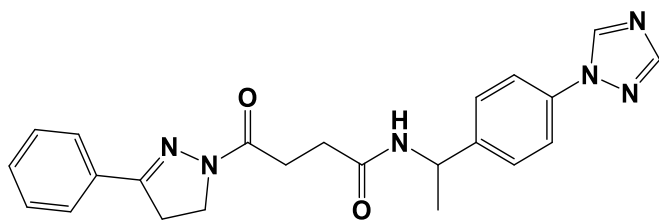


Figure 1 - Chemical structure of the strongest antiviral identified in the second virtual screening study (**17**).

The predicted binding pattern for **17** can be compared to LA, with increased predicted interactions due to the aromatic nature of the scaffold. This is particularly significant in the vicinity of phenylalanine and tyrosine residues in the pocket, with multiple π -stacking and π -H interactions. On the hydrophilic terminal end of the molecule, **17** extends towards the entrance of the pocket with a triazole group, as opposed to a carboxylic acid in LA, that is predicted to result in H-bond interactions with multiple hydrophilic residues, which stabilize the molecule in the pocket and the adjacent RBD. As for the remaining compounds, the frequent presence of aromatic groups is observed, in line with previous predictions. The presence of these aromatic features, as their frequency, is validated by confirmed FABP bioactive molecules, such as lifitegrast and SPC-14, while the molecular dynamics studies also pointed to possible stronger binding by exploring the hydrophobic area of the pocket. **17** follows the previous established rules interaction in the FABP, confirming the trend for new ligands: deviation from the alkane chains in fatty acids, towards an increased number of aromatic ring systems. Given that antiviral activity for this compound was confirmed experimentally and predicted computationally, along with inhibition of the virus-host interaction in binding assays, it is likely that the activity shown in the viral infection assays is at least partially due to an effect on the FABP, stabilizing the inactive conformation of the S protein.

These findings not only validate the initial computational predictions but also highlights the pocket's potential as a druggable target.

To build on these insights and refine the compound's therapeutic potential, the next phase of the study focused on hit optimization.

4.1.2.5 Hit optimization

A systematic approach was undertaken to explore structural variations in the selected compounds. This included selecting structural analogues to probe specific interactions within the FABP binding site, with the aim of identifying critical contacts that contribute to binding affinity.

The selection of structural analogues was aimed at potentially increasing antiviral activity of the identified scaffolds and allow a preliminary study of SAR, that could guide the design of

new more potent compounds. Twenty structural analogues were selected, targeting hydrophobic and hydrophilic substitutions in the terminal triazole ring and linkers between the multiple ring systems in **17**. These compounds were submitted to new antiviral assays, with CPE as the indicator of antiviral activity. From the initial selection, nine compounds showed at least 10% inhibition of CPE, while activity for **17** was confirmed (23%) and three compounds showed increased stronger inhibition of viral induced CPE: **20** (34%), **32** (27%) and **36** (91%). The best analogue **36** achieved an EC_{50} in the nanomolar range (EC_{50} =290 nM), along with reduced cytotoxicity (CC_{50} =181 μ M). Additional assays were performed to evaluate the release of virions, with **36** demonstrating a ~90% reduction in both release of replication-capable lytic virions and genome copies in infection samples, measured by reverse transcriptase digital droplet PCR (RT-ddPCR).

The structural analogues identified from **17** demonstrated a broad range of effects on CPE reduction, varying from 0% to 90%, providing a foundation for a preliminary SAR analysis. This variation offered valuable insights into the binding pocket's characteristics and key molecular interactions (Figure 9).

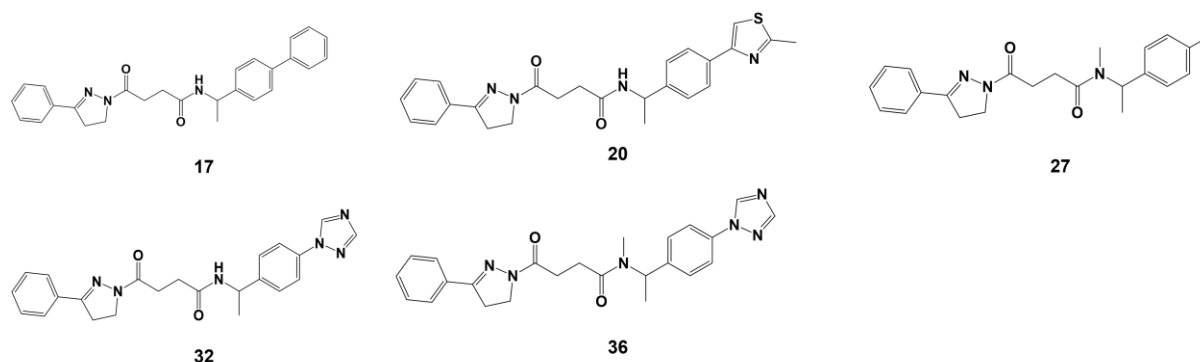


Figure 2 - Chemical structure of the most active structural analogues to **17**

As previously suggested by the confirmed ligands of the FABP, the molecular dynamics simulations and pocket availability, **27** (methyl substitution of triazole – 14% activity) highlights the hydrophobic interactions as critical, since the methyl group is incapable of aromatic, electrostatic or H-bond interactions in the pocket entrance. However, it still retains some activity even in the absence of a terminal group. On the other hand, the presence of an aromatic terminal ring observed for **32** (phenyl substitution the triazole – 27% activity), that results in increased activity against the parent compound, underlines the importance of π interactions in this area, that can be further explored by heterocyclic rings with increased activity such as **20** (thiazole substitution - 34% activity). In addition to the hydrophilic interaction with amino acids in the adjacent RBD, aromatic groups in this area might enable interactions with tyrosine and phenylalanine residues, which results in higher binding affinity.

Indeed, interactions in this area appear to be behind the increased activity in **36**, with the tertiary amine that might result in a stronger anchoring effect, leading to a stable interaction in the FABP and stabilization of the S protein, due to increased ability to be an H-bond acceptor. However, given the previous considerations regarding hydrophobic interactions as essential for initial binding, an increase in hydrophobicity (LogP) might also play an important role, as this is increased in both **20** and **36** (**17**: 2.36, **20**: 3.91, **36**:2.62).

When compared with currently approved medicines against COVID-19, the inhibition of CPE in Vero E6 cells for **36** (EC_{50} =290 nM) is superior when compared with Remdesivir (EC_{50} =770 nM) [40], similar to molnupiravir (EC_{50} =300 nM) [41] and weaker than Nirmatrelvir (EC_{50} =74.5 nM) [42]. Therefore, the detected activity is comparable with other already approved drugs against SARS-CoV-2, which might clear the way for further optimization in potential new biologically relevant experiments. However, these results should be interpreted with caution. While Vero E6 cells are highly susceptible and permissive to SARS-CoV-2, being widely used in CoV research due to the high viral titres of SARS-CoV-2 and characteristic CPE (being well suited to antiviral screening), they also show limitations [41]. Since these cells are derived from a non-human primate (African green monkey), polymorphisms to the human proteome might limit analysis, given the overall absence of lung-specific or host immune effects, since they lack the capacity to produce IFN-I following infection, which likely plays a role in supporting robust viral replication [44]. They promote rapid phenotypic adaptation, with furin cleavage mutations frequently obtained and have variations in metabolism when compared with human cell lines, although this is particularly relevant for drugs designed to undergo metabolic activation (such as prodrugs), which is not the expected mechanism of action [43, 45]. Additionally, later variants, particularly Omicron have shown limited results in Vero E6 cells, due to attenuated fusogenicity and pathogenicity of these later variants, which reduces CPE [45, 46]. This effect was indeed detected in the antiviral assays with the Delta and Omicron variant. For the Delta variant, the infection period was extended by 72h, to obtain similar results when compared with the Wuhan original strain (EC_{50} =0.29 μ M and EC_{50} =5.77, respectively). However, with the Omicron variant, the extension in infection time did not originate CPE. Nevertheless, these results, do not exclude activity in these strains and, in fact, taking into account that the S protein of Delta variant is heavily mutated, the activity detected suggests that this target is conserved, even in mutated variants and that this scaffold holds the potential for broad anti-CoV activity [47]. Regarding the Omicron variant, it has become clear that CPE analysis may not yield the expected results, requiring different methodologies or cell types to adequately assess antiviral activity. Although the FABP has been mainly unaffected by mutations, it is still unclear if the exploration of this target would exert an evolutionary pressure leading to quick emergence of resistance. However, the

suggested physiological function of the FABP, modulating the exposure of immunodominant sites, points to a potential decrease in viral fitness from mutations in this area. Linoleic acid and arachidonic acid are essential eicosanoid precursors, and along with tissue inflammatory regulators, show a stark increase in concentration during COVID-19 infection and lung inflammation [48]. While at basal serum concentrations antibody binding to the S protein is possible, under elevated “inflammatory” concentrations the neutralization activity is reduced due to activation of the FABP, in a suggested immune escape mechanism by stabilization of the inactive conformation of the RBDs. This mechanism enables viral cell infection and replication in early stages of infection, while suppressing immunogenicity once the host immune response is activated, leading to increased viral titers, due of reduced virus clearance [27].

4.2 Future perspectives

The discovery of a sub-micromolar anti-SARS-CoV-2 agent opens a wide range of possibilities for future research and therapeutic development, primarily focusing in confirming binding to the expected target, along with complementary experimental assays for continued pre-clinical development. While it is expected that the compounds in this study act through the FABP this has not been experimentally confirmed, which could be performed by an NMR or crystallographic study. This would not only confirm binding but also clarify binding dynamics and stabilizing interactions, allowing further optimization of activity. Mechanistic studies can also elucidate how these compounds act, as with time-of-addition assays, that could confirm an effect in the initial infection stage of the virus lifecycle (during the infection stage). Additional binding assays, such as Surface Plasmon Resonance, could also confirm binding to the target protein [49]. Further pre-clinical development should, therefore, focus on different cell lines, particularly from human origin and the respiratory tract, which would allow a precise determination of activity and toxicity. Evaluating the pharmacokinetic properties of these compounds, including their absorption, metabolism, and excretion, is critical to advance their development as viable antiviral agents. New analogues could be obtained to modulate activity and toxicity, guided by experiments in different cell types that more adequately represent human replication kinetics, entry pathways, off-target effects, immune capabilities and other signalling pathways.

References

1. Wrapp, D., et al., 2020, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 367(6483): p. 1260-1263.

2. Wu, F., et al., 2020, A new coronavirus associated with human respiratory disease in China. *Nature*. 579(7798): p. 265-269.
3. Scudellari, M., 2020, The sprint to solve coronavirus protein structures — and disarm them. *Nature*. 581: p. 252-255.
4. Walls, A.C., et al., 2020, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 181(2): p. 281-292.e6.
5. Lan, J., et al., 2020, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 581(7807): p. 215-220.
6. Wang, Q., et al., 2020, Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*. 181(4): p. 894-904.e9.
7. Benton, D.J., et al., 2020, Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature*. 588(7837): p. 327-330.
8. Queirós-Reis, L., et al., 2021, SARS-CoV-2 Virus–Host Interaction: Currently Available Structures and Implications of Variant Emergence on Infectivity and Immune Response. *International Journal of Molecular Sciences*. 22(19): p. 10836.
9. Li, G., et al., 2023, Therapeutic strategies for COVID-19: progress and lessons learned. *Nature Reviews Drug Discovery*. 22(6): p. 449-475.
10. Xue, S., et al., 2024, Mutations in the SARS-CoV-2 spike receptor binding domain and their delicate balance between ACE2 affinity and antibody evasion. *Protein & Cell*. 15(6): p. 403-418.
11. Chakraborty, C., et al., 2022, A Detailed Overview of Immune Escape, Antibody Escape, Partial Vaccine Escape of SARS-CoV-2 and Their Emerging Variants With Escape Mutations. *Frontiers in Immunology*. 13.
12. Papa, G., et al., 2021, Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *PLOS Pathogens*. 17(1): p. e1009246.
13. Zhou, T., et al., 2020, Cryo-EM Structures Delineate a pH-Dependent Switch that Mediates Endosomal Positioning of SARS-CoV-2 Spike Receptor-Binding Domains. *bioRxiv*. p. 2020.07.04.187989.
14. Zhou, T., et al., 2020, A pH-dependent switch mediates conformational masking of SARS-CoV-2 spike. *bioRxiv*. p. 2020.07.04.187989.
15. Cox, M., et al., 2023, SARS-CoV-2 variant evasion of monoclonal antibodies based on *in vitro* studies. *Nat Rev Microbiol*. 21(2): p. 112-124.
16. Troyano-Hernández, P., R. Reinoso, and Á. Holguín, 2021, Evolution of SARS-CoV-2 Envelope, Membrane, Nucleocapsid, and Spike Structural Proteins from the Beginning of the Pandemic to September 2020: A Global and Regional Approach by Epidemiological Week. *Viruses*. 13(2): p. 243.

17. Lamb, K.D., et al., 2024, Mutational signature dynamics indicate SARS-CoV-2's evolutionary capacity is driven by host antiviral molecules. *PLOS Computational Biology*. 20(1): p. e1011795.
18. Wang, M.Y., et al., 2020, SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front Cell Infect Microbiol*. 10: p. 587269.
19. Xia, S., et al., 2019, A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Science advances*. 5(4): p. eaav4580-eaav4580.
20. Xia, S., et al., 2020, Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res*. 30(4): p. 343-355.
21. Cooper, B.M., et al., 2021, Peptides as a platform for targeted therapeutics for cancer: peptide-drug conjugates (PDCs). *Chem Soc Rev*. 50(3): p. 1480-1494.
22. Leneva, I., et al., 2021, Antiviral Activity of Umifenovir *In Vitro* against a Broad Spectrum of Coronaviruses, Including the Novel SARS-CoV-2 Virus. *Viruses*. 13(8).
23. Li, Y., et al., 2020, Efficacy and Safety of Lopinavir/Ritonavir or Arbidol in Adult Patients with Mild/Moderate COVID-19: An Exploratory Randomized Controlled Trial. *Med*. 1(1): p. 105-113.e4.
24. Amani, B., et al., 2021, Efficacy and safety of arbidol (umifenovir) in patients with COVID-19: A systematic review and meta-analysis. *Immun Inflamm Dis*. 9(4): p. 1197-1208.
25. Guo, L., et al., 2023, Targetable elements in SARS-CoV-2 S2 subunit for the design of pan-coronavirus fusion inhibitors and vaccines. *Signal Transduction and Targeted Therapy*. 8(1): p. 197.
26. Toelzer, C., et al., 2020, Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science*. 370(6517): p. 725-730.
27. Stauffer, O., et al., 2022, Synthetic virions reveal fatty acid-coupled adaptive immunogenicity of SARS-CoV-2 spike glycoprotein. *Nature Communications*. 13(1): p. 868.
28. Weill, P., et al., 2020, May omega-3 fatty acid dietary supplementation help reduce severe complications in Covid-19 patients? *Biochimie*. 179: p. 275-280.
29. Goc, A., A. Niedzwiecki, and M. Rath, 2021, Polyunsaturated ω -3 fatty acids inhibit ACE2-controlled SARS-CoV-2 binding and cellular entry. *Scientific Reports*. 11(1): p. 5207.
30. Shoemark, D.K., et al., 2021, Molecular Simulations suggest Vitamins, Retinoids and Steroids as Ligands of the Free Fatty Acid Pocket of the SARS-CoV-2 Spike Protein*. *Angew Chem Int Ed Engl*. 60(13): p. 7098-7110.
31. Queirós-Reis, L., et al., 2023, Exploring the Fatty Acid Binding Pocket in the SARS-CoV-2 Spike Protein – Confirmed and Potential Ligands. *Journal of Chemical Information and Modeling*. 63(23): p. 7282-7298.

32. Romanou, V., et al., 2021, Dexamethasone in the Treatment of COVID-19: Primus Inter Pares? J Pers Med. 11(6).
33. SPECS Library. Available online: <https://www.specs.net/> (accessed on 2024/11/21).
34. Enamine Screening Collection. Available online: <https://enamine.net/compoundcollections/screening-collection> (accessed on 2024/11/21).
35. Friesner, R.A., et al., 2004, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. Journal of Medicinal Chemistry. 47(7): p. 1739-1749.
36. Friesner, R.A., et al., 2006, Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. Journal of Medicinal Chemistry. 49(21): p. 6177-6196.
37. The Handbook of Medicinal Chemistry: Principles and Practice, ed. S.E. Ward and A. Davis. 2023: The Royal Society of Chemistry.
38. Korb, O., T. Stütze, and T.E. Exner, 2009, Empirical Scoring Functions for Advanced Protein–Ligand Docking with PLANTS. Journal of Chemical Information and Modeling. 49(1): p. 84-96.
39. OpenEye, OEDOCKING 4.3.1.0. Cadence Molecular Sciences, Inc.
40. Wang, M., et al., 2020, Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. Cell Research. 30(3): p. 269-271.
41. Sheahan, T.P., et al., 2020, An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. Science Translational Medicine. 12(541): p. eabb5883.
42. Owen, D.R., et al., 2021, An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19. Science. 374(6575): p. 1586-1593.
43. Pires De Souza, G.A., et al., 2022, Choosing a cellular model to study SARS-CoV-2. Frontiers in Cellular and Infection Microbiology. 12.
44. Lokugamage, K.G., et al., 2020, Type I Interferon Susceptibility Distinguishes SARS-CoV-2 from SARS-CoV. J Virol. 94(23).
45. Chen, D.Y., et al., 2023, Cell culture systems for isolation of SARS-CoV-2 clinical isolates and generation of recombinant virus. iScience. 26(5): p. 106634.
46. Suzuki, R., et al., 2022, Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. Nature. 603(7902): p. 700-705.
47. Toelzer, C., et al., 2022, The free fatty acid–binding pocket is a conserved hallmark in pathogenic β -coronavirus spike proteins from SARS-CoV to Omicron. Science Advances. 8(47): p. eadc9179.

48. Irún, P., et al., 2023, Serum lipid mediator profiles in COVID-19 patients and lung disease severity: a pilot study. *Sci Rep.* 13(1): p. 6497.
49. Day, C.J., et al., 2021, Multidisciplinary Approaches Identify Compounds that Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2–ACE2 Receptor Interactions. *mBio.* 12(2): p. e03681-20.



Chapter V

Conclusions

“I am among those who think that science has great beauty”

Marie Curie

5 Chapter V - Conclusions

Summarizing the research carried out in this thesis, the overall results indicate the following general conclusion:

The S protein of SARS-CoV-2 plays a central role in virus-host interactions and remains a critical target for antiviral drug discovery. Through computational methodologies, an antiviral compound with sub-micromolar activity against SARS-CoV-2 was successfully identified. The discovery of a novel scaffold offers a promising foundation for further optimization and drug development against current and emerging CoVs.

5.1 Exploration of S protein structure and behaviour

The S protein is the central determinant of SARS-CoV-2 recognition and infection, playing a pivotal role in viral entry into host cells. Despite structural similarities to S proteins from earlier CoV outbreaks, its functional behaviour is uniquely complex. Effective drug development targeting this protein requires a comprehensive understanding of its conformational changes and the transition to the post-fusion state that facilitates cell infection. Initial recognition is driven by the high-affinity interaction between the RBD and ACE2, a hallmark of SARS-CoV-2 infectivity and a key factor in its rapid global spread. Compared to prior CoV outbreaks, the introduction of the multibasic cleavage site markedly enhances activation by furin, a feature that synergizes with increased receptor binding affinity to amplify viral entry efficiency. However, the origins of this multibasic cleavage site remain unclear, though its contribution to viral infectivity and pathogenicity is profound. Following cleavage at the S1/S2 boundary, the spike protein becomes increasingly responsive to ACE2 binding, triggering an allosteric cascade that destabilizes the S1/S2 interface. This destabilization facilitates S2 release and subsequent conformational changes essential for membrane fusion. Notably, the transition to the post-fusion conformation occurs with a remarkably low energy barrier in SARS-CoV-2, further distinguishing its infectivity. These synergistic factors—enhanced receptor binding, efficient activation by host proteases, and low-energy fusion initiation—culminate in an exceptionally infectious virus capable of sustaining pandemic-level transmission. Adaptations in the spike protein since the first confirmed cases, such as mutations increasing affinity, flexibility and cleavage efficiency, and promoting RBD availability, have further lowered the energy barrier for fusion events. These changes have amplified infectivity, reinforcing the virus's capacity for widespread human transmission.

The S protein offers multiple potential drug targets, each with distinct challenges. Initially, the RBD emerged as a promising target due to its critical role in mediating viral entry via ACE2 binding. However, this region is highly mutable, enabling the virus to escape antibody neutralization. Consequently, drugs targeting the RBD or directly inhibiting the S-ACE2 interaction are prone to resistance, making them less viable for long-term therapeutic strategies. An alternative approach involves targeting the fusion process, a crucial step for viral infection. Fusion inhibitors, particularly those targeting the HR1-HR2 bundle in the S2 subunit, could offer broad-spectrum activity across coronaviruses. However, this strategy is hindered by the lack of structural clarity; neither crystallographic nor Cryo-EM structures have captured this binding event in sufficient detail, limiting the ability to design effective inhibitors. The FABP near the RBDs represents well-defined target. Structural studies and *invitro* assays have demonstrated that multiple fatty acids, including polyunsaturated and monounsaturated types, along with retinoids and vitamins A and K, influence S protein stability and cell infection. The availability of high-resolution crystallographic structures for this pocket enables precise computational modelling and drug design. Given its critical role in modulating the conformational dynamics of the spike protein, the FABP represents a compelling target for developing novel antiviral agents.

5.2 Hit molecule identification and *in vitro* validation

The FABP emerged as an attractive target for the development of innovative antiviral therapies, and its exploration began with a virtual screening study. This initial screening aimed to identify compounds capable of binding in the FABP and mimicking LA binding effects. This would result in the stabilization of the S protein in its closed conformation, thereby preventing interaction between the RBDs and the ACE2 receptor. In the screened compounds, large substituent groups were positioned within FABP, contrasting with the alkane chains observed LA, potentially promoting favourable aromatic interactions. While several promising compounds were identified based on their binding potential, their activity in biological assays was limited, pointing to the need for further refinement.

Building on these early results, additional insights from research, particularly regarding known FABP ligand,s were incorporated, highlighting the structural features that contribute to binding affinity. Compounds with aromatic and cyclic groups were quickly confirmed to bind the pocket and have a biological effect, as opposed to the long carbon chains, typical of fatty acids like LA. These bulkier groups can interact with hydrophobic regions within the pocket, particularly with phenylalanine and tyrosine residues, while the terminal hydrophilic groups can still form H-bonds at the pocket entrance. Interestingly, while fatty acid carboxylate groups play a role in binding, they were not strictly necessary for effective interaction. This understanding helped

shift the focus to molecules with non-carboxylate terminal groups, broadening the potential pool of candidate molecules.

Further MD simulations revealed that hydrophobic interactions within the pocket are pivotal during initial binding, likely playing a more significant role than the hydrophilic interactions at the pocket's entrance, although the stabilization of the hydrophilic residues in the adjacent RBD is likely the responsible for the wider activity in the protein. These findings refined the criteria to select potent inhibitors. Following this, an expanded computational screening was performed, with the insights obtained from ligand analysis.

The most promising compound identified through these efforts demonstrated significant inhibition of SARS-CoV-2 in cell-based assays, with activity likely stemming from its ability to stabilize the inactive conformation of the spike protein, reducing binding to ACE2. This result provided strong validation for the computational screening approach, as well as evidence that targeting the FABP could effectively disrupt the virus-host interaction. To advance the development of potent inhibitors, twenty structural analogues of the most bioactive compound were obtained and evaluated. Among these, one analogue displayed submicromolar potency against the Wuhan strain and micromolar activity against the Delta variant, showing its sustained antiviral potential and highlighting an optimizing step in the compound's therapeutic efficacy.

In comparison with currently approved COVID-19 therapeutics, the identified compound showed favourable results, with comparable efficacy in Vero E6 cells, although a slight variation in potency between the Wuhan and delta variants was found. These promising results suggest that the compound could be developed as a novel antiviral agent with potential for both SARS-CoV-2 and other emerging coronaviruses.

The insights gained through this research not only highlight the FABP as a viable therapeutic target for SARS-CoV-2 but also establish a framework for optimizing small-molecule inhibitors against viral entry. Just as prior CoV research was instrumental in accelerating the discovery of inhibitors for SARS-CoV-2, the findings from this work may play a pivotal role in addressing future coronavirus outbreaks. This foundation offers the potential to guide the development of novel antiviral therapies, providing both an alternative to existing treatments while contributing to global preparedness for emerging viral threats.

5.3 Closing remarks

This thesis provides key insights into the role of the FABP in SARS-CoV-2 S protein function, offering a better understanding of its potential as a druggable target for antiviral therapy. Although the role of FABP in stabilizing an inactive S protein conformation is still under active investigation, this study contributes to that ongoing exploration, providing further evidence of its involvement in modulating the S protein's interaction with ACE2 and influencing viral entry

mechanism. The conserved nature of this target across CoVs positions the FABP as a promising focus for therapeutic strategies in future outbreaks. This study demonstrates the effective application of computational methods to identify potential therapeutic compounds targeting a rationally defined biological target. Several of the identified compounds exhibited micromolar and sub-micromolar activity. Nonetheless, further investigations are necessary to elucidate their precise mechanisms of action and to confirm whether the FABP within the S protein serves as their molecular target. Future research should aim to further optimize the binding interactions with FABP and explore the broader antiviral effects of these compounds, particularly their ability to inhibit viral replication across different SARS-CoV-2 variants. In vivo and clinical studies will be essential to assess their full potential. Further structural biology studies are needed to confirm the interactions between the identified scaffolds and FABP. Additionally, evaluating the pharmacokinetic properties of these compounds is essential to advance their development as antiviral agents. Addressing these gaps will facilitate the development of effective therapies for SARS-CoV-2 and future emerging coronaviruses.

5.4 Personal remarks

The project described in this thesis unfolded during a time of rapid evolution, in a field marked by frequent breakthroughs in information, theories, and experimental techniques. This dynamic environment invited constant reassessment and reformulation of objectives and near-term goals to align with the latest developments. Even before the project's formal initiation, some of the original goals had already become partially outdated. Keeping pace with global research and ensuring efforts did not overlap with other teams remained a constant challenge. Despite these challenges, the context offered unique learning opportunities. The focused attention of the scientific community, particularly from virology experts, created numerous collaborative possibilities. The core partnership with supervisors from Swansea and Cardiff Universities provided essential guidance and support, and their expertise was essential to the design and execution of the entire project. A six-month scientific mission was performed in Swansea University, which enabled the contact with multidisciplinary teams, with extensive expertise in chemistry, medicinal chemistry and antiviral research, particularly in computer aided techniques, which significantly contributed to the success of this work

Midway through the project, a significant restructuring occurred to include cell-based assays, particularly viral infection assays, requiring access to BSL-3 laboratories for live virus replication. This expansion was made possible by funding from the EEA Norway grants, which supported extensive research activities and two scientific missions. However, even this subproject was subject to changes in timeline and experimental setup. To overcome this, a new collaboration with Oslo University Hospital was initiated, further expanding the existing partnership with Swansea and Cardiff Universities. This collaboration enabled access to BSL-

3 facilities, allowing for critical live-virus infection assays, which were key to evaluate the antiviral activity of computationally identified compounds, significantly enhancing the scope and outcomes of this research. These collaborations not only enabled the practical execution of this work but also provided exceptional learning opportunities by integrating expertise from leading researchers in the field.

While COVID-19 is no longer classified as a Public Health Emergency of International Concern, and largely faded from public discourse, it is pertinent to reflect on the relevance of this research and its findings. The disease now exhibits more predictable patterns of circulation and is managed through proven therapies and vaccination efforts, but its potential to cause significant health impacts, particularly among vulnerable populations, remains. This transition is common with other endemic diseases, such as influenza or HIV, that still pose challenges despite widespread immunity or established treatments. Furthermore, lessons learned from SARS-CoV and MERS-CoV played a critical role in the rapid pace of SARS-CoV-2 research, with numerous examples of CoV-specific knowledge adapted to SARS-CoV-2. Given that SARS-CoV-2 represents the third highly pathogenic CoV outbreak in recent history, the knowledge and tools developed during this research not only retain their applicability to SARS-CoV-2 but also might serve as a foundation for addressing newly emerging CoVs.



INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

