



STUDYING THE LIMITS OF PROTEIN-LIGAND DOCKING TO PREDICT NEW NON-COVALENT INHIBITORS FOR THE SARSCOV-2 MAIN PROTEASE

Guillem Macip Sancho

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



UNIVERSITAT
ROVIRA I VIRGILI

Studying the limits of protein-ligand docking to predict new non-covalent inhibitors for the SARS-CoV-2 main protease

GUILLEM MACIP SANCHO

DOCTORAL THESIS

2023

UNIVERSITAT ROVIRA I VIRILI

STUDYING THE LIMITS OF PROTEIN-LIGAND DOCKING TO PREDICT NEW NON-COVALENT INHIBITORS FOR THE SARSCOV-2 MAIN PROTEASE

Guillem Macip Sancho

Guillem Macip Sancho

**Studying the limits of protein-ligand docking to
predict new non-covalent inhibitors for the SARS-
CoV-2 main protease**

Doctoral Thesis

Supervised by

Dr. Gerard Pujadas Anguiano and Dr. Santiago Garcia Vallvé

Cheminformatics & Nutrition Research Group
Biochemistry & Biotechnology Department



UNIVERSITAT ROVIRA I VIRGILI

Tarragona 2023

UNIVERSITAT ROVIRA I VIRGILI

STUDYING THE LIMITS OF PROTEIN-LIGAND DOCKING TO PREDICT NEW NON-COVALENT INHIBITORS FOR THE SARSCOV-2 MAIN PROTEASE

Guillem Macip Sancho



UNIVERSITAT
ROVIRA I VIRGILI

WE STATE that the present study, entitled “Studying the limits of protein-ligand docking to predict new non-covalent inhibitors for the SARS-CoV-2 main protease”, presented by Guillem Macip Sancho for the award of the degree of Doctor, has been carried out under our supervision at the Biochemistry and Biotechnology Department of this university.

FEM CONSTAR que aquest treball, titulat “Studying the limits of protein-ligand docking to predict new non-covalent inhibitors for the SARS-CoV-2 main protease”, que presenta Guillem Macip Sancho per a l'obtenció del títol de Doctor, ha estat realitzat sota la nostra direcció al Departament de Bioquímica i Biotecnologia d'aquesta universitat.

HACEMOS CONSTAR que el presente trabajo, titulado “Studying the limits of protein-ligand docking to predict new non-covalent inhibitors for the SARS-CoV-2 main protease”, que presenta Guillem Macip Sancho para la obtención del título de Doctor, ha sido realizado bajo nuestra dirección en el Departamento de Bioquímica i Biotecnologia de esta universidad.

Tarragona, April 10th, 2023

Doctoral Thesis Supervisors

Dr. Gerard Pujadas Anguiano

Dr. Santiago Garcia Vallvé

UNIVERSITAT ROVIRA I VIRGILI

STUDYING THE LIMITS OF PROTEIN-LIGAND DOCKING TO PREDICT NEW NON-COVALENT INHIBITORS FOR THE SARSCOV-2 MAIN PROTEASE

Guillem Macip Sancho

Table of Contents

Summary.....	9
Introduction	14
COVID 19 as a disease and its causative agent SARS-CoV-2	16
Social Impact.....	16
Viral Genome.....	16
Viral Infection	17
Structural proteins	17
M-pro as target for the treatment of COVID 19	19
Enzyme active site and subsites: Catalytic mechanism.....	20
The importance of the water molecules	21
Residue protonation on the binding site	22
Efforts against M-pro.....	23
Protein-ligand docking	23
State of the art	23
Protein-ligand docking limitations	27
Validation of the docking results.....	28
References.....	32
Hypotheses and objectives	38
Results	42
Manuscript 1	45
A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet?	
Manuscript 2	103
Haste makes waste: A critical review of docking- based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition	
Manuscript 3	151
Are the protein-ligand docking programs good enough to predict experimental poses of non-covalently inhibitors bound to SARS-CoV-2 main protease?	
Discussion	188
Conclusions	194
Publication List	198

Summary

The global pandemic known as COVID-19 (COronaVirus Disease 2019) had a profound impact on healthcare systems around the world, straining countries and societies in ways not seen since globalization. SARS-CoV-2, a positive sense single-stranded RNA virus from the Betacoronavirus genus, is the causative agent of COVID-19, and it has been thoroughly investigated by the scientific community since December 2019. The SARS-CoV-2 main protease (M-pro) is one of the more frequently investigated viral targets. Our research group devised a virtual screening methodology based on protein-ligand docking during the start of the COVID-19 pandemic to identify new SARS-CoV-2 M-pro inhibitors among approved drugs. When we compared our results to those of other suggested SARS-CoV-2 M-pro inhibitors, we discovered that numerous articles predicted different compounds, albeit utilizing the same starting dataset and approach. In a short period of time, there were a large number of proposed M-pro inhibitors and structures of non-covalent complexes between SARS-CoV-2 M-pro, which supplied the breeding ground for this thesis. The current PhD thesis examines in depth the limitations of protein-ligand docking in predicting SARS-CoV-2 M-pro inhibitors.

Resum

La pandèmia global coneguda com a COVID-19 (COronaVirus Disease 2019) va tenir un profund impacte en els sistemes sanitaris d'arreu del món, tensionant els països i les societats d'una manera que no s'havia vist des de la globalització. El SARS-CoV-2, un virus d'ARN monocatenari de sentit positiu del gènere Betacoronavirus, és l'agent causant de COVID-19, i ha estat investigat a fons per la comunitat científica des de desembre de 2019. La proteasa principal del SARS-CoV-2 (M-pro) és una de les dianes virals més investigades. El nostre grup de recerca va idear una metodologia de cribratge virtual basada en el docking proteïna-ligand durant l'inici de la pandèmia COVID-19 per identificar nous inhibidors del SARS-CoV-2 M-pro entre els fàrmacs aprovats. Quan vam comparar els nostres resultats amb els d'altres inhibidors suggerits de la M-pro del SARS-CoV-2 vam descobrir que nombrosos articles predien diferents compostos, tot i que utilitzaven el mateix conjunt de dades i enfocament inicial. En un curt període de temps, es va generar un gran nombre d'estructures de complexos no covalents entre la SARS-CoV-2 M-pro i diferents inhibidors, cosa que va propiciar el caldo de cultiu per a aquesta tesi. La actual tesi doctoral examina en profunditat les limitacions del docking proteïna-ligand a l'hora de predir els inhibidors de la SARS-CoV-2 M-pro.

Resumen

La pandemia global conocida como COVID-19 (COronaVirus Disease 2019) tuvo un profundo impacto en los sistemas sanitarios de todo el mundo, tensionando a los países y sociedades de una forma que no se había visto desde la globalización. El SARS-CoV-2, un virus de ARN monocatenario de sentido positivo del género Betacoronavirus, es el agente causante de COVID-19, y ha sido investigado a fondo por la comunidad científica desde diciembre de 2019. La proteasa principal del SARS-CoV-2 (M-pro) es una de las dianas virales más investigadas. Nuestro grupo de investigación ideó una metodología de cribado virtual basada en el docking proteína-ligando durante el inicio de la pandemia COVID-19 para identificar nuevos inhibidores del SARS-CoV-2 M-pro entre los fármacos aprobados. Cuando comparamos nuestros resultados con los de otros SARS sugeridos. -Inhibidores del CoV-2 M-pro, descubrimos que numerosos artículos predicen distintos compuestos, aunque utilizaban el mismo conjunto de datos y enfoque inicial. En un corto período de tiempo, se generó un gran número de estructuras de complejos no covalentes entre la SARS-CoV-2 M-pro y diferentes inhibidores, cosa que proporcionó el caldo de cultivo para esta tesis. La tesis doctoral actual examina en profundidad las limitaciones del docking proteína-ligando para predecir los inhibidores del SARS-CoV-2 M-pro.

Introduction

Introduction

COVID 19 as a disease and its causative agent SARS-CoV-2

Social Impact

Even though the global alarm has disappeared, the disease COVID-19 (COronaVirus Disease 2019) is still affecting 228 countries and territories ¹ accumulating more than 6,7 million deaths. The earlier stages of the pandemic affected quite deeply the healthcare systems all around the world, making some of them even collapse. In the lack of a vaccine or effective medications, nations had been forced to use quarantines and other health strategies such as social distancing to mitigate the spread of the virus ². Quarantines, however, come with a significant financial cost since when workers are forced to stay at home, production stops. Longer quarantines either delay the peak (if they are applied relatively early) or dampen the peak, lowering the overall number of infected individuals and deaths (if they are imposed later in the pandemic), but they come with higher economic costs ³. Numerous restrictions worldwide ended up having a detrimental effect on the mental faculties on all levels of society. In many cases, this social isolation developed into loneliness, which in turn can turn into depression or anxiety, among other health complications like coronary heart disease and stroke and other chronic diseases ⁴.

The pandemic's effects on research may last longer than previously thought. The amount of time that scientists are devoting to their research has nearly reached pre-pandemic levels, and the majority of indicators based on publications only show slight declines. However, despite returning to the workforce, scientists have been much less likely to start new research projects. This aspect of impact seems to be rather uniform across disciplines and disproportionately affects female scientists and those who are raising young children ⁵.

Viral Genome

The coronaviruses have the biggest RNA virus genome known to date. They are an enclosed virus with a non-segmented, single-stranded, positive-sense RNA genome of about 32 kilobases. Coronaviruses are members of the nidovirales order, the coronaviridae family, and the coronavirinae subfamily. The Coronavirinae subfamily has four genera: alphacoronavirus, betacoronavirus, deltacoronavirus, and gammacoronavirus. The SARS-CoV-2 strain is categorized within the betacoronavirus genus according to genome sequence analysis. Since the coronavirus genome is known to have a 5' cap and a 3' poly (A) tail, it serves as an mRNA for the translation of the replicase polyproteins necessary for viral replication once the virus has infected the host cell ^{6,7}.

Approximately 29 to 30 kB in size, the SARS-CoV-2 RNA genome is structured as methyl-capped-5'UTR-ORF1a/b-S-ORF3-E-M-ORF6-ORF7a/b-ORF8-N/ORF9b-ORF10-3'UTR-poly-A tail. The hemagglutinin esterase gene present in other betacoronaviruses is absent from the genome ⁸⁻¹⁰.

Figure 1 showcases the general genomic structure of these viruses.

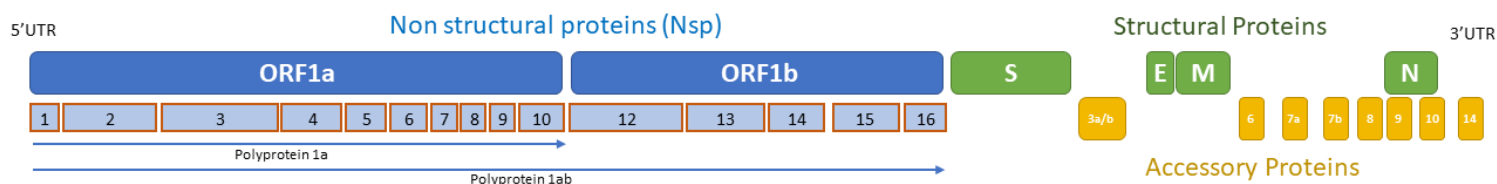


Figure 1. Genomic structure of SARS-CoV-2.

Introduction

Two thirds of the genome (ORF1a and ORF1b) are translated into two large overlapping polyproteins (pp1a and pp1ab), the second occurring thanks to a programmed frame-shift.^{7,11,12} The encoded polyproteins contain the proteases PLpro and M-pro, which help cleave 16 non-structural proteins (nsp1-nsp16) from polyprotein pp1ab. These nsp include various replication proteins such RdRp, RNA helicase, and exoribonuclease (ExoN). The four structural proteins: the envelope (E), nucleocapsid (N), membrane (M) and spike (S) as well as auxiliary proteins (ORF3a to ORF9b) are housed in the final third of the viral genome^{7,11,12}.

Viral Infection

Bat-derived coronavirus (CoV) seldom infect humans, therefore an intermediate host is usually necessary for the coronavirus to spread to humans^{6,11}. The majority of wild pangolins tested positive for coronaviruses similar to SARS-CoV-2. As a result, it was discovered that the Pangolin-CoV spike protein differed only slightly from the SARS-CoV-2 spike protein in terms of one amino acid. Thus, before spreading to humans, SARS-CoV-2 may have developed via a viral recombination between Pangolin-CoV and Bat-nCoV^{6,7,11}.

SARS-CoV-2 spreads through respiratory droplets expelled by an infected person, which means that coughing and sneezing can cause the virus to become airborne and infect others who are not already sick. Because they house so many infected people, hospitals are recognized as one of the sources of secondary SARS-CoV-2 transmission as well as by direct contact with polluted surfaces^{6,11}. Asymptomatic or slightly symptomatic patients make up 30–60% of all SARS-CoV-2 patients⁷.

All of the SARS-like CoVs display a typical replication and translation mechanism upon infection in the host cell. The first stage of viral infection, the binding of human coronaviruses (HCoV) to host-cell receptors, determines the severity of the infection and the pathogenesis¹¹.

The HCoVs infection process starts with the viral particle's attachment to the host cell surface via the spike (S) protein, which is composed of S1 and S2 domains. When the S1 domain attaches to a host-cell receptor, specifically the angiotensin-converting enzyme 2 (ACE2) for SARS-CoV-2, the viral membrane fuses with the host cell membrane. As a result, host cell proteases cleave the spike glycoprotein, allowing the S2 domain to connect and allows the virus to enter the cell^{7,9,11–13}.

Two distinct pathways have been identified for viral entrance into cells after attachment to the host-cell surface. The first is via an endosome that is mediated by clathrin-dependent and -independent endocytosis mediated via a cleavage by the host transmembrane serine protease 2 (TMPRSS2) or transmembrane serine protease 11D (TMPRSS11D) on the cell surface. The RNA is then released into the cytoplasm^{7,11,12}.

Structural proteins

There are four main viral structural proteins in SARS-CoV-2: S, E, M, and N. **Table 1 with the general ideas for each protein.**

Table 1. Main proteins of SARS-CoV-2

STRUCTURAL PROTEINS	
<i>SPIKE</i>	Attaches to the host-cell receptor
<i>ENVELOPE</i>	Formation of the viroporin (virus assembly, release and virulence)
<i>MEMBRANE</i>	Viral morphogenesis, interferon suppression and virion shape
<i>NUCLEOCAPSID</i>	Adaptive immunity and stabilization of the RNA into helical form
NON-STRUCTURAL PROTEINS	
<i>NSP1</i>	Disrupts protein synthesis in host cell
<i>NSP2</i>	Interacts with signals involved in host cell cycle
<i>NSP3</i>	Scaffold for other NSP. Viral replication transcription complex
<i>NSP4</i>	Assembly of the viral double-membrane vesicles
<i>NSP5</i>	3CLPro/M-pro. Protease activity
<i>NSP6</i>	Induction of autophagosomes
<i>NSP7</i>	Primer synthesis and rna replication
<i>NSP8</i>	
<i>NSP9</i>	Involved in the virulence
<i>NSP10</i>	Scaffold for other NSP. mRNA cap methylation complex
<i>NSP11</i>	Intrinsically disordered protein. Unknown
<i>NSP12</i>	RNA-dependent RNA polymerase, mRNA capping
<i>NSP13</i>	helicase activity during RNA replication
<i>NSP14</i>	Proof reading and methyltransferase during mRNA capping
<i>NSP15</i>	Endoribonuclease that cleaves RNA at polyuridylylate sites
<i>NSP16</i>	Methyltransferase during mRNA capping

S is a trimeric class I fusion protein. The N-terminal domain (NTD) and receptor binding domain (RBD) of the S1 subunit, which are the targets of neutralizing antibodies, are highly concentrated with important immunogenic epitopes in the S protein. The major domains of the S2 subunit are two heptad repeats (HR1 and HR2), a central helix (CH), fusion peptide (FP) and a connector domain (CD) which connects the spike to the virus membrane^{10,13}. The envelope protein (E) which is a transmembrane protein with an ion channel activity, most likely contributes to the formation of a viroporin, a component required for virus assembly and release as well as a potential indicator of virulence. The lipid envelope's highly expressed structural protein known as the membrane protein (M) plays a crucial role in viral morphogenesis and interferon suppression as well as virion shape. Last but not least, the nucleocapsid protein (N) is a crucial target of adaptive immunity and stabilizes the RNA genome in a helical structure. Additionally, it functions as an interferon antagonist^{7,10–12,14}. Apart from the structural proteins, other important proteins are Nsp1, 3, 5, 9, 10, 14 and 16.

Nsp1 is produced from the cleavage of the N-terminal polyprotein precursors, pp1a and pp1ab, via proteolysis by viral papain-like protease (PLpro). Nsp1 disrupts the synthesis of proteins in the host cell^{9,10}.

Nsp3 is the largest protein encoded by the coronavirus. It is a multi-domain protein and acts as a scaffold that interacts with the other Nsps and the host proteins to form the viral replication-

Introduction

transcription complex, anchoring them to the membrane. It has a PLpro domain that cleaves polyproteins pp1a and pp1ab to free Nsp3^{9,10}.

Nsp5, also known as the main protease (M-pro) or 3C-like protease (3CLpro) is a cysteine protease. The sequence of M-pro from SARS-CoV-2 is 96% similar to that of SARS-CoV¹⁵. It produces 12 functional proteins by cleaving the viral polyproteins pp1a and pp1ab at 11 different locations with the cleavage sequences LQ(S/A/G). M-Pro is a desirable target for the development of anti-CoV medications since inhibitors targeting this enzyme would prevent viral replication^{9,10}.

Nsp9 is a single-stranded RNA-binding protein. Nsp9 is thought to contribute to the virus's virulence. The SARS-CoV-2 Nsp9 and the SARS-CoV ortholog share 97% of their sequence similarity. According to the structural data, disrupting the dimer interface of Nsp9 may be interesting to create treatments for coronavirus illnesses^{9,10}.

Nsp10 is a tiny, single-domain protein that shares 99% of its sequence from the SARS-CoV equivalent. Together with Nsp14 (exonuclease and N7-methyltransferase) and Nsp16 (2'-O-methyltransferase), Nsp10 functions as a scaffold protein to assemble the mRNA cap methylation complex. It's interesting to note that the activity of things complex is targeted by a Nsp10 derived peptide both in vivo and in vitro^{9,10}.

Nsp14 is a bifunctional protein with an N-terminal exonuclease (ExoN) domain that is thought to play a role in proofreading and a C-terminal guanine-N7 methyl transferase (N7-MTase) domain involved in the methylation of viral RNA cap. It has been hypothesized that Nsp14's proofreading function is what drives the evolution and upkeep of coronaviruses' expansive genomes. By preventing host immune response-induced degradation of viral mRNA, Nsp14's mRNA capping function is crucial for the stability of viral mRNA^{9,10}.

Nsp16 is a m7GpppA (Cap-1)-specific, S-adenosyl methionine (SAM) dependent, nucleoside 2'-O-methyltransferase which catalyzes the 5'-methyl capping of viral mRNA. The 5'-methyl capping prevents host 5'-exoribonucleases from degrading viral RNA and works to prevent the activation of an innate immune response. There is a ten-fold decrease in the synthesis of viral RNA when SARS-CoV Nsp16 is genetically disrupted which makes this an attractive target. Nsp16 from SARS-CoV-2 and SARS-CoV are 93.3% identical^{9,10}.

M-pro as target for the treatment of COVID 19

As previously stated, inhibiting the protein resulting from Nsp5, meaning the M-pro would halt the spread of viruses. Thus, during the recent years many approaches against COVID-19 have been focused on this specific protein.

The SARS-CoV-2 M-pro's performs as an active homodimer with roughly C2 symmetry, similar to other CoV M-pros. There are three domains on each subunit (I, II, and III) (**Figure 2**). The domains I and II adopt a chymotrypsin- six-stranded antiparallel β -barrel fold (residues 10-99 and 100-182, respectively). The substrate-binding site is formed by a large gap between domains I and II, and it contains the N-terminal finger (domain I, residues 1-7), which is crucial for catalysis^{9,16,17}. Five antiparallel α -helices make up the C-terminal globular domain III (residues 198-303), which participates in dimerization via an intermolecular salt-bridge interaction between E290 and R4. The loop (residues 183-198) that connects domains II and III has two different conformations; in some, it takes on a very extended conformation, while in others, residues 186-190 form a short helix. Because the N-finger of each subunit interacts with E166 of the other subunit to shape the S1 pocket of the substrate-binding site, dimerization of the

enzyme is required for greater catalytic activity. Due to a substitution (T285A) in domain III, the dimer of SARS-CoV-2 M-pro is tighter and closer than that of SARS-CoV M-pro, increasing catalytic efficiency. Although each subunit contributes mostly by itself to its own substrate binding site, the two subunits switch N-termini to maintain their S1 pockets^{9,16,17}. The superposition of 12 M-pro's crystal structures reveals that the most variable regions are the

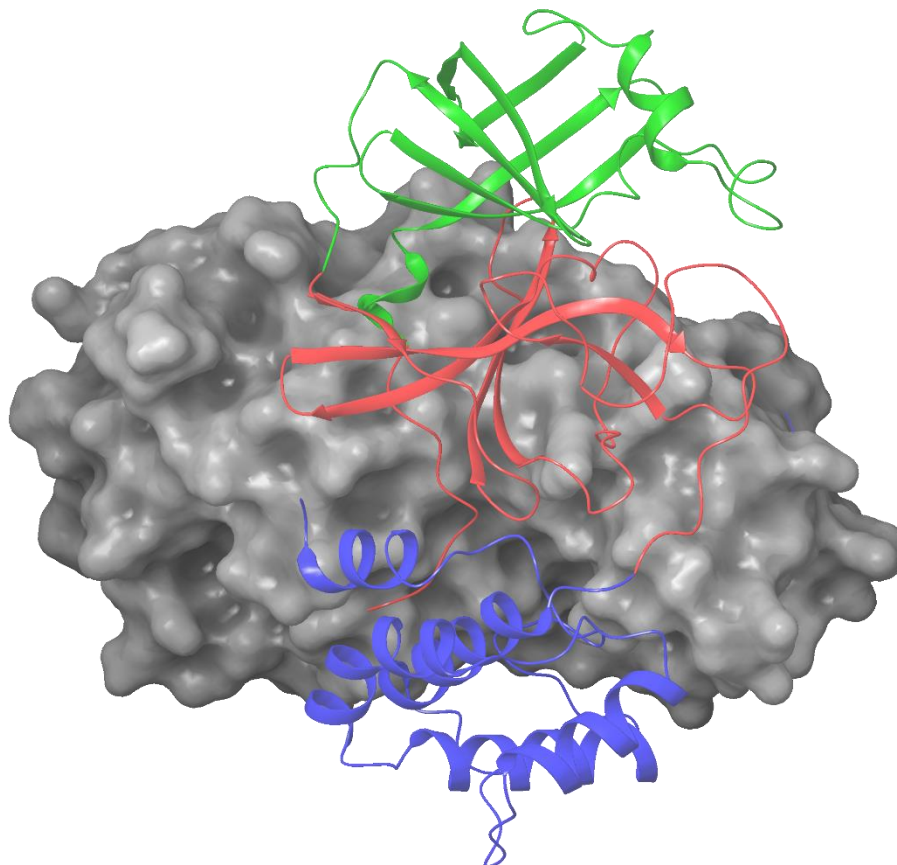


Figure 2. Homodimer of SARS-CoV-2 M-pro. PDB ID:6WQF. One subunit is represented with their molecular surface, while the other shows Domain I(green), II (red), III (blue).

helical domain III and surface loops, and that the substrate-binding pocket (located in a cleft between domain I and domain II) is highly conserved across all coronaviruses; this suggests that antiviral inhibitors targeting this pocket should have broad antiviral activity against coronaviruses¹⁶.

Enzyme active site and subsites: Catalytic mechanism

His41 and Cys145 make up the catalytic dyad at the substrate-binding site, which is located in the middle of the cleft between domain I and II and a conserved water molecule forms a hydrogen bond with His41 to facilitate catalysis (**Figure 3**)^{15,17,18}. Other important subsites at the M-pro binding site, as in any protease, are S3, S2, S1, and S1', which are occupied by the P3, P2, P1, and P1' residues of its peptidic substrate (where the point of peptide cleavage is at the peptide bond that binds residue P1 with residue P1'). Met165, Leu167, Gln189, Thr190, and Gln192 form S3; Met49, Tyr54, His164, Asp187, and Arg188 form S2; Ser1 (from the other protomer), Phe140, Leu141, Asn142, His163, and Glu166 make S1'; and His41, Gly143, Ser144, and Cys145 form S1'. Other relevant residues at the M-pro binding site include Thr24, Thr25, Pro168, His172, Phe185, and Ala191^{15,17,18}.

Introduction

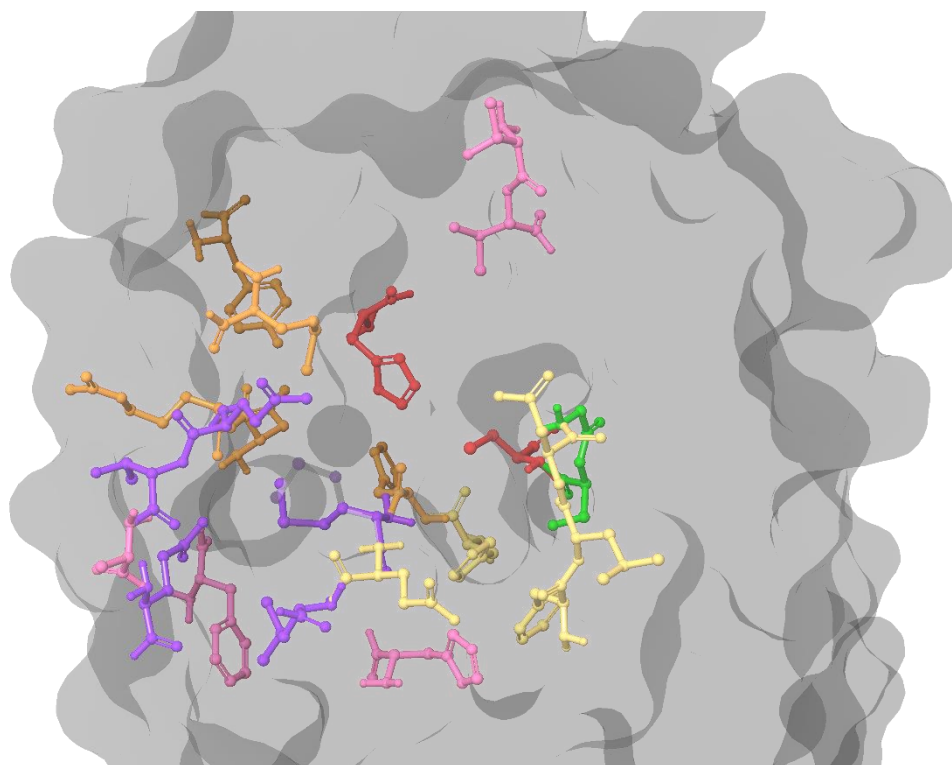


Figure 3. Important residues inside the binding site of SARS-CoV-2 M-pro. In red the catalytic dyad (His41, Cys145), in yellow the S1, in green the S1' (which also contains the catalytic dyad), in orange the S2, in purple the S2' and in pink other important residues.

The substrate adopts an extended configuration from P6 to P1 regions, with the M-pro residues and its backbone forming an antiparallel-sheet. A conserved glutamine residue at the P1 position of the substrate forms two powerful hydrogen bonds with the side chain atom of His163 and the main chain carbonyl oxygen of Phe140^{9,15,16,19}. With certain side chains filling the S2-S4 M-Pro pockets, the P2 to P4 residues bind to the enzyme in an extended beta-sheet shape. Through van der Waals interactions, the substrate residues at the P5 and P6 sites interact with the enzyme. Only short residues, Ser, Gly, or Ala, can fit in the P1' position on the C-terminal side of the substrate's S1' subsite. A lengthy side chain, like Lys, can fit in the S2' subsite's small yet deep P2' location. The side chain in the P3' position is accessible to solvents^{9,15,16,19}.

The importance of the water molecules

When a protein is in solution, it is enveloped by water molecules that have different properties from those away from it. Extensive networks of water molecules are found on protein surfaces, as shown by cryogenic X-ray data and molecular dynamics simulations. The majority of water molecules in close proximity to a protein are weakly bound to it, remain mobile, have brief interaction times, and are difficult to observe using conventional X-ray crystallography. Others are unmistakably securely attached to the protein surface, especially in clefts like ligand binding sites (**Figure 4**), despite the fact that some water molecules seen in crystal structures are by-products of the analysis²⁰. In protein crystal structures, water molecules frequently occupy conserved locations in structurally related proteins. Water molecules usually become crucial components of a protein's binding site, since they can mediate interactions between ligands and proteins and form hydrogen-bonded networks that stabilize the complex. Thus, the presence of water molecules has a significant impact on the accuracy of ligand-protein docking predictions

²⁰.

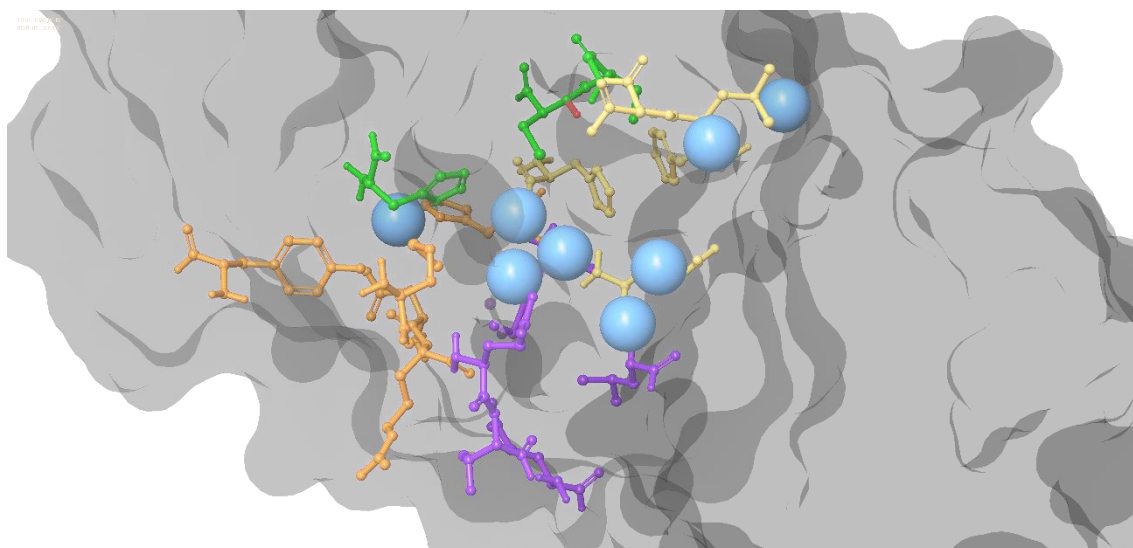


Figure 4. Water molecules inside the binding site of M-pro (PDB ID: 6WQF). The water molecules will be to either be displaced or reinforce the interaction between the ligand and the binding site.

Specifically for M-pro, the amino acid residues His41, Cys117, Glu166, and Arg298 (which are in the catalytic dyad, dimerization site, substrate-binding site, and allosteric site, respectively) present a distorted structure of water molecules close to the interface, according to Venugopal PP and coworkers²¹.

Longer hydrogen bond lifetimes indicate that the protein binding sites are stable. The binding sites of the M-pro homodimer are hidden due to the interaction between the two subunits. The binding sites of the dimer contain fewer water molecules than those of the monomer. As a result, the dimer has a shorter water-water hydrogen bond lifetime than the monomer. Thus, it can be inferred that hydrogen bonds between the water molecules serve to stabilize the protein's binding sites^{17,21}.

Another indicative of the water importance is the irreversible symmetry-breaking event that undergoes the M-pro triggered by water molecules. This causes a binding site to become inactive while the other remains active²². Depending on the content of water, the N-finger of a subunit interacts with the binding site: 1) strongly and thus opening a path for water to enter, consequently collapsing the S1 binding site; or 2) with a single hydrogen bond that allows hydrophobic interactions between His163 and Phe140 to stabilize the binding site. The subunits end up displaying different conformations of the binding pocket and consequently each subunit presents a different collection of long-lived water molecules inside the binding site. Furthermore, four water molecules play a crucial structural role for the whole subunit (two of them located between domain I and II and the others are present inside II and III respectively)²².

Residue protonation on the binding site

The combination of histidine protonation states in or near the catalytic site can have a significant impact on M-pro's structural stability. As previously stated, M-pro displays a His41–Cys145 catalytic dyad, but on top of that other histidines are found near the binding site: His64, His80, His163, His164 and His172¹⁸. Neutron crystallographic structures have demonstrated that histidines are able to form hydrogens bonds with amide nitrogen from the backbone, resulting in low pKas and thus remain in a neutral deprotonated state. Changes in the enzymatic activity of M-pro have been attributed to protonation and deprotonation of several of these histidines

Introduction

¹⁸. It has been suggested that His172 protonated state occurs because it acts as a hydrogen bond donator and that His41 acts as a proton shuttle during the catalysis.

Several studies have specifically targeted the protonation states of His41, Cys145 and His164^{23,24}. While they did not reach a consensus, as the mechanistic studies determined that the catalytic residues likely remain in a neutral state, the MD simulations found increased distances between both in the charged state. Nonetheless, Cys145 deprotonation coupled with protonation of His164, could not be ruled out¹⁸. According to bibliography, Cys145, His41, and His163 form interactions with substrates and inhibitors¹⁸.

Of the histidines present near the binding site, His64 is directly in contact with the solvent and His80 forms a hydrogen bond with the side chain of Asn63. Due to their close proximity, Cys145, His41, and His164 were assumed to have interdependent protonation states, whereas His163 and His172 were regarded as having independent protonation states from the other residues¹⁸. Interactions with water are critical for His41 because it interacts with both His164 and Asp187 via a water molecule, which serves as the third component of the catalysis mechanism¹⁸. In crystal structure conformations, the protonated His41/deprotonated Cys145 state of the catalytic dyad is structurally unstable but previous computational studies have proposed that this state exists as a transient reaction intermediate^{24,25}.

Efforts against M-pro

Numerous efforts have been made to create and produce SARS-CoV-2 M-pro inhibitors (see Manuscript 1). Specifically, CADD (Computer-Aided Drug Discovery) methods have been used to find potential many candidates. Unfortunately, most of these studies did not adhere to industry best practices, which call for knowledge gained through a variety of meticulously designed benchmarking studies and meticulous experimental validation²⁶

Among the different crystallographic structures of SARS-CoV-2 M-pro in the Protein Data Bank (PDB), one of the first was that of 6LU7. This structure has been used in protein-ligand docking to predict many SARS-CoV-2 M-pro inhibitors.

Protein-ligand docking

State of the art

Molecular docking is a method for computing interactions between macromolecules (receptors) and small molecules (ligand). They try to mirror the lock-and-key method of drug action to estimate the experimental binding pose and affinity of a ligand within the binding site of the target protein²⁷⁻²⁹. Since docking methods are quick enough to scan through millions of compounds, they are frequently used in structure-based virtual screenings (VS) on large molecular libraries. Docking programs rely on scoring functions to assess protein-ligand interactions³⁰.

The binding free energy of a protein and its ligand determines their binding affinity. Extensive sampling of complex conformations and detailed handling of the aqueous solution environment, such as free energy perturbation (FEP) or thermodynamic integration (TI), are required for accurate prediction of binding free energy. The drawback of these techniques is that they are ill suited for large scale VS due to their computational cost. Due to their scoring functions, which calculate the protein-ligand binding free energy based on a single protein-ligand complex structure, molecular dockings provide an alternative³⁰.

Docking programs usually involve two steps, the docking itself, which tries to assess the correct structure of the small molecule in relation to the receptor and the scoring function. Scoring functions are able to rank the generated poses and select the most favourable one according to their calculations (best-scored pose), which represents the binding affinity of the compound.³⁰

Scoring functions are usually classified in the following groups: (i) physics-based approaches, (ii) statistical potentials based on knowledge, (iii) empirical scoring functions, and (iv) machine-learning.

- Physics-based scoring functions are frequently based on molecular physics like Van der Waals interactions (Lennard-Jones potential), electrostatic interactions (coulombic potential), and desolvation energies. Both experimental data and ab initio quantum mechanical computations can be used to determine these terms. Solvation and entropy factors are typically oversimplified or disregarded in physics-based scoring algorithms due to the processing cost. This type of scoring function is used by programs such as GoldScore, DOCK, and early versions of AutoDock^{30–32}.
- The statistical potentials are obtained from experimentally obtained protein-ligand structures. The inverse Boltzmann distribution is utilized to produce these potentials by using the frequency of specific interactions from multiple protein-ligand complexes. Using a high number of protein-ligand atom-pairwise terms, this method approximates intricate and difficult-to-characterize physical interactions. As a result, there is no immediate physical meaning for the scoring function. Examples of knowledge-based scoring functions are DrugScore, ITScore, and PMF^{30,33}.
- Empirical scoring functions use a collection of weighted scoring parameters to characterize the binding affinity of protein-ligand complexes. These score terms could include VDW descriptors, electrostatics, hydrogen bonds, hydrophobic, desolvation, entropy, among others. The descriptor weights are determined by fitting experimental binding affinity data of protein-ligand complexes using linear regression. Both physics-based and knowledge-based scoring functions are used in empirical scoring functions. Empirical scoring functions, like physics-based scoring functions, use physically relevant meaning. Similarly to knowledge-based scoring functions, the contribution (weight) of each phrase is learned from the training data. Because of the limits provided by the physical terms, empirical scoring systems are less prone to overfitting than knowledge-based scoring functions and provide insight into which has a bigger role in the interactions. ChemScore, GlideScore, X-Score and Autodock Vina are some of the examples^{30,34–37}.
- A collection of approaches known as machine learning (ML) scoring functions makes use of ML algorithms to identify patterns in training data and learn the functional form of the binding affinity. The intermolecular interactions that are challenging to explicitly model can be implicitly captured by ML scoring functions without the use of a predetermined functional form. Recent years have seen a noticeable increase in the ability of ML scoring systems to predict binding affinities. The previous three types are usually referred as classical scoring functions. They share the linear nature of their force field of interactor descriptors. However, by applying ML techniques like Support Vector Machine (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGB), Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Graph Neural Network(GNN), ML scoring functions may adopt far more complex functional forms. Despite these successes, the use of ML modeling in drug development, particularly for

Introduction

deep learning, is still in its infancy. Deep learning techniques are frequently criticized for being uninterpretable, a "black-box," and easily overtrained^{30,38}.

There are many prominent protein-ligand docking systems now on the market, and their number has been rising over the past few years. The subsequent programs are ones the most used.

Autodock Vina

One of the most popular open-source docking programs is Autodock Vina, whose scoring function consists of a ligand torsion count term and five empirical interaction terms (two gaussian terms, a repulsion term, a hydrogen bond term, and a hydrophobic term)³⁰.

AutoDock Vina use a Monte-Carlo iterated local search approach, which includes sampling, scoring, and optimization iterations. Initially, a particular ligand's initial random conformation, which is represented by its location, orientation, and torsion, is sampled (POT). After then, a disturbance randomly modifies its position, orientation, or torsion. The binding position between a ligand and a protein is then used to evaluate the affinity. In AutoDock Vina, the scoring function that sums the intramolecular energy and the intermolecular energy (ligand-receptor) is used to determine the binding affinity (ligand–ligand). Additionally, a Broyden-Fletcher-Goldfarb-Shanno (BFGS) technique that takes into account the gradients of the scoring function is used to optimise the conformation. By using these gradients, the ligand can be guided to a better conformation. The docking score differential between the optimised conformation and the starting conformation is used to create a metropolis acceptance rule to determine whether or not the optimised conformation may be accepted. The acceptable conformation will be saved as the starting conformation and progressively improved in the following rounds³⁹.

Virtual screening with AutoDock Vina on huge databases is very slow. Therefore, an acceleration of AutoDock Vina has become a central problem in current virtual screens of drug hits from huge compound databases. Till now, there have been several attempts for the acceleration of AutoDock Vina in large virtual screens^{40–42}. Due to the seriality of the AutoDock Vina approach, the usual way to accelerate the process is the allocation of resources and tasks, ie. the VirtualFlow platform. The popularity of AutoDock Vina and its flexibility in contemporary drug development will be severely constrained by the high resource requirements and high access threshold of users. The wide adoption of AutoDock Vina in massive virtual screening for contemporary drug discovery will be facilitated by a reduction in the computational resource investment and user access threshold. For instance, AutoDock-GPU and Vina-GPU 2.0 offers an OpenCL implementation of AutoDock4 that takes advantage of both parallel GPU and CPU architectures^{39,43}.

Recently, Lin F9, a linear empirical scoring function inspired by Vina's scoring function, was created to enhance scoring performance and get around some of Vina's drawbacks by adding new empirical terms like metal-ligand interactions and mid-range interactions. Lin F9 outperformed Vina in scoring accuracy when trained on a small but high-quality protein-ligand dataset for binding affinity prediction³⁰.

Vina's source code has been changed and extended by research teams worldwide, which has improved the search algorithm (QuickVina2⁴⁴), made the user interface more comprehensible by enabling modification of scoring terms through the user interface (Smina⁸), and enhanced the scoring function for carbohydrate docking (Vina-Carb⁴⁵) and halogen bonds (VinaXB⁴⁶) as well as ranking and scoring (Vinardo⁴⁷).

Despite these positive advancements, there are still a number of AutoDock Suite techniques that are not compatible with Vina since they were developed exclusively for the AutoDock 4 (AD4) scoring function or software. AutoDock 4 provides a wide range of search techniques and a scoring mechanism based on a linear regression analysis, the Assisted Model Building with Energy Refinement (AMBER) force field, and a vast collection of various protein-ligand complexes with known inhibition constants⁴⁸. AD4 contributed to the widespread use of the Autodock suite, but it was later surpassed by Vina. The ineffective search performance of the AD4 software limits the utility of such customised solutions. Depending on the search complexity, AD4 can be up to 100 times slower than Vina⁴⁹.

This changes with the most recent version of the program. Simple and complex docking simulations may be designed and carried out more easily with the help of AutoDock Vina 1.2.0. The updated version offers Python bindings, making it simpler to script for complex applications like virtual screening. In order to speed up high-throughput virtual screens, it also incorporated batch processing and simultaneous multiple-ligand docking against a single target structure⁴⁹. Additionally, the hydrated docking protocol was created to simulate the water molecules that are directly engaged in the interaction between ligands and receptors. The method, which is based on docking ligands explicitly hydrated with spherical water molecules, enhances predictions of ligand pose⁴⁹.

Glide

It is widely acknowledged that the Schrödinger's Glide docking suite, when it includes additional components accounting for solvation and repulsion contributions in addition to other chemical forces in the formation of ligand-protein complexes, provides better predictions of binding affinities, higher binding accuracy and consistency, and more reliable differentiation between active and inactive compounds³⁶. Glide approximates a thorough systematic search of the conformational, orientational, and positional space of the docked ligand, in contrast to existing approaches for docking ligands to the hard 3D structure of a known protein receptor. In this search, torsionally flexible energy optimization on an OPLS-AA nonbonded potential grid is applied to a small number of the surviving candidate poses after an initial rough positioning and scoring phase that drastically reduces the search space. A Monte Carlo sample of position conformation is used to further hone the top choices. A model energy function that incorporates empirical and force-field-based components is used to choose the best docked position³⁶. The Glide application contains a covalent docking panel, high throughput (HTVS), standard precision (SP), and extra precision (XP) noncovalent docking features. The covalent docking virtual screening procedure used by the Glide is noteworthy since it has been shown to be both quick and extremely accurate in covalent docking experiments^{37,50}.

The empirically based ChemScore formula⁵¹ serves as the basis for Glide scoring. Two different GlideScore variants are utilised: GlideScore SP, which is used by SP and HTVS, and GlideScore XP, which is used by X. Although these functions have similar nomenclature, they were created with different goals in mind. GlideScore SP is a "softer," "more forgiving" tool that excels at locating ligands with a fair tendency to bind, even when the Glide pose has substantial flaws. The scoring function is the same in both HTVS and SP docking, but in HTVS, the number of intermediate conformations along the docking funnel is decreased, and the final torsional refinement and sampling is less complete. GlideScore XP imposes harsh penalties on positions that defy well-established physical chemistry rules, such the need that charged and highly polar groups be appropriately exposed to solvent^{36,37}.

Introduction

FRED

FRED uses a docking algorithm based on Gaussian function⁵² and makes use of a precalculated conformation database⁵³. Run time is decreased by producing conformers before starting FRED (since conformation generation is done independently, it is run once per given ligand database). The tool OMEGA is utilized in the OpenEye workflow to generate ligand conformers before docking. OMEGA identifies and counts rotatable bonds and flexible rings using torsion and ring libraries. Conformers that exhibit internal conflicts or significant strain are eliminated. Using their root-mean-square deviation (RMSD) as a basis, the conformers are clustered and only the centers are kept. Additionally, if the number of generated conformers is greater than a predetermined maximum, only the lowest strain energy are kept⁵⁴. FRED also allows for internal consensus docking. This entails docking each compound in the database to several targets and standardizing a specific compound's score on a given target against a range of scores for that molecule on the various targets⁵³.

FRED docks molecules by employing a thorough search algorithm that performs systematic rotations and translations of each ligand conformer within the active site at a predetermined resolution/precision. Improbable poses are filtered out during the thorough search, and those that remain are rated. Following a thorough search, systematic solid body optimization is applied to the top 100 poses (a local search that is more detailed than the global one). The best pose is then used against other ligands in the screening database to properly score the ligand itself. During the docking procedure, both the conformers of the ligand and the protein are maintained rigid. The docking process is actually flexible with respect to the ligand since several conformers of each ligand are docked into the site⁵⁴. During the docking process, FRED makes use of scoring functions three times: during the thorough search, during optimization, and during the final scoring. Shapegauss, PLP, Chemgauss, Chemscore, Screenscore, Chemical Gaussian Overlay (CGO), and Chemical Gaussian Tanimoto (CGT) are the scoring algorithms available in FRED⁵⁴. FRED can use the bound ligand to guide docking (although in its default mode it is a pure docking program). This strategy, which blurs the distinction between ligand- and structure-based approaches, is called hybrid docking⁵⁴.

Protein-ligand docking limitations

Although molecular docking programs can typically create ligand poses that are comparable to the crystallographic conformation for a variety of targets, protein-ligand docking has a number of limitations. One of these is that most docking approaches adopt a rigid protein approximation for practical reasons, in which the ligand is viewed as flexible but the protein conformation is limited^{55,56}. Protein flexibility can be studied using methods such as ensemble docking, which operates on a group of protein conformations⁵⁷. The inability to find bioactive ligands for binding sites that are not explicitly docked against or implicitly represented in the training set is a fundamental flaw in both structure and ligand-based VS methods. Ligands may very well (and frequently do) bind to alternate binding pockets in the experimental validation⁵⁸. Another issue is that, contrary to expectations, the docked poses and docking scores might be significantly dependent on the ligand input structure, so slight changes in the ligand input conformation can result in large variances in the docked poses. One of the fundamental disadvantages of docking software scoring systems is that they do not always provide a helpful prediction of ligand binding affinity⁵⁹. As a result, in most docking applications, the best energy score is not a reliable criterion for picking the ideal solution⁶⁰. Docking scores should not be utilized as a measure or threshold for compound activity for these reasons, but rather to expand the library by removing compounds that do not fit into the binding site. It should also be emphasized that docking scores from different docking programs cannot be directly compared because the scores are affected

by the force fields and procedures utilized by each program. Another drawback of molecular docking is how it handles water molecules that are present in the binding site during the docking process⁵⁵ which can lead to inaccuracies when determining the probable interaction between the ligand and the target. The effectiveness of a docking program to distinguish between decoys and active molecules is typically highly dependent on the protein structure utilized as well as the similarity between the screened ligands and the co-crystallized ones. The use of docking in drug development is however restricted to biological targets for which crystal structures are known. This latter barrier has been overcome using a variety of strategies. For instance, creating homology models from structural templates with highly homologous sequences is a common way to get around the lack of 3D structures⁵⁹.

Docking programs can be used with other technologies to overcome their limitations. Protein flexibility can be investigated using standard molecular dynamics or binding free energy estimations such as MM-GBSA and MM-PBSA⁶¹. These approaches are typically used to evaluate compounds identified by molecular docking because they are not suited for screening huge libraries of compounds^{59,61}. However, the success of these methods is dependent on the target structure⁶², and they cannot be used to predict actual drug candidates in drug design without experimental validation due to their poor accuracy⁶³. The fact that these post-docking refining and rescoring techniques typically result in higher success rates in virtual screening campaigns and better correlation with experimental data makes them extremely valuable for the drug discovery process⁵⁹.

As was previously mentioned, docking frequently cannot give a reliable approximation of the significance of the solvation penalties associated with the binding. Subsequently, a number of complementary computational techniques have been created to recognize and examine the water molecules in the vicinity of the protein-ligand complex in order to determine their contribution to binding⁵⁵. The fundamental drawback of these approaches is their reliance on the protein shape employed for computation. It is currently unclear if these methodologies can be used in large virtual screening campaigns, despite the fact that they are now more widely used and have shown value in explaining lead chemical structure-activity connections⁵⁵.

Although being created and utilized initially as a stand-alone technique, docking is today primarily used in conjunction with other computational techniques inside integrated workflows. This makes it possible to get beyond some of the most important inherent restrictions that molecular docking has⁵⁹. The use of mixed techniques typically leads to higher prediction performances and enables for better exploitation of data from various sources. Docking has in fact been used in applications of integrated workflows to help with various drug development activities. For instance, to enhance prediction performances, docking has been combined with ligand-based, molecular dynamics, binding free energy calculations, and artificial intelligence techniques⁵⁹.

Validation of the docking results

Generally speaking, when a docking program has generated a possible structure for the small molecule, commonly known as a binding pose, approaches are used to evaluate its performance.

In silico validation

The bioactivity of docking hits must be confirmed both *in silico* and *in vitro* or *in vivo* to demonstrate their accuracy. Applying the docking protocol simultaneously to a set of active compounds and a set of inactive or decoy compounds is a common way to carry out an *in silico* validation^{58,64}.

Introduction

Compounds that have been shown to exhibit a high activity toward the desired target are known as **active** compounds. Although there is no set amount of activity over which a chemical is deemed active, compounds are frequently regarded as actives if their IC50, Ki, or EC50 activity values fall in the micromolar or nanomolar range^{58,64}. Compounds that have been reported to exhibit little to no action (or none at all) toward the desired target are known as **inactive** compounds. The use of hit compounds that are structurally related to **inactives** is discouraged since they are thought to have a higher likelihood of being inactive^{58,64}. **Decoy** compounds are substances that mimic active substances but whose activity against the target of interest has not been reported; as a result, it is assumed that they are inactive⁶⁵. Due to the small number of null results that have been reported in the literature and the resulting paucity of information on inactive compounds, decoys are typically utilized in validation processes in place of inactives^{58,64}.

Once the active and inactive/decoy groups have been specified, after the docking, each compound will fall into the confusion matrix (**Table 2**), which contains True positives (actives predicted as such), True negatives (inactive predicted as such), False positives (inactive predicted as active) or False negatives (actives predicted as inactive). They can be used to calculate a series of statistical measures for each step. These include true positive rate, true negative rate, positive predictive value, negative predictive value, false negative rate, false positive rate, accuracy, etc. among others.

A scoring function's ability to rank known active above many inactive compounds in the library is normally estimated by the screening performance. To measure the screening effectiveness of a scoring function, many assessment measures are available, such as the enrichment factor (EF), area under the curve (AUC), and receiver operating characteristics (ROC) curve. Of note, the EF is one of the most used. The EF is a measurement of how much the sample has been enriched with actives following the application of a certain filter or series of filters. The ratio of the actives after and before the docking is used to calculate it⁶⁴. The AUC-ROC approach is used to evaluate the classifier's performance in assessing true binder identification. ROC projects the false positive rates (FPR, also known as specificity) and true positive rates (TPR, also known as recall or sensitivity) onto a curve whose area under the curve (AUC) spans from 0 to 1, with 1 denoting perfect selection and 0.5 representing a random-level selection^{30,66}.

Table 2. Confusion Matrix of a binary classifier.

PREDICTED	AUTHENTIC	
	POSITIVE	NEGATIVE
POSITIVE	True positives (TP)	False positives (FP)
NEGATIVE	False Negatives (FN)	True negatives (TN)

These measures help us evaluate the performance of each step and thus prioritize. For example, during the initial stages of a VS it would be helpful to maximize the true negative rate, since inactives usually outnumber actives. Model parameters should be changed to provide a more desirable outcome⁶⁴. Overall, these criteria enable objective evaluation of the model's multiple elements and user-preference-based parameter adjustment to provide the most appropriate model for each circumstance.

Re-docking and cross docking

Finding the native binding pose among the others generated, is the aim of binding pose identification. A trustworthy scoring algorithm should be able to rank the native binding pose as the top. A typical method for verifying the docking step is also re-docking, which involves taking

a ligand out of a protein-ligand complex and re-docking after converting it into a 2D structure to see if a docking software can accurately anticipate its original and 3D binding position. A frequently used assessment measure is RMSD between the top docking position and the experimentally determined ligand pose (**Figure 5**). The RMSD between the docking pose and the crystallographic one, provides an estimate of the capacity of the docking algorithm to reproduce the co-crystallized binding pose of the ligand. RMSD readings under 2.0 Å are typically regarded as good, although readings near zero are desirable⁵⁶. The RMSD measure for binding posture prediction has been widely utilized in the area since it is straightforward and simple to apply^{30,67,68}.

It is advised to try cross docking in addition to re-docking to try and predict a ligand's binding pose by using a different PDB structure. This makes it possible to assess the potential impact of protein flexibility or, in the event that there are many protein structures, to choose the most advantageous one. Re-docking and cross docking, however, cannot adequately validate a VS process as it only involves the docking step.

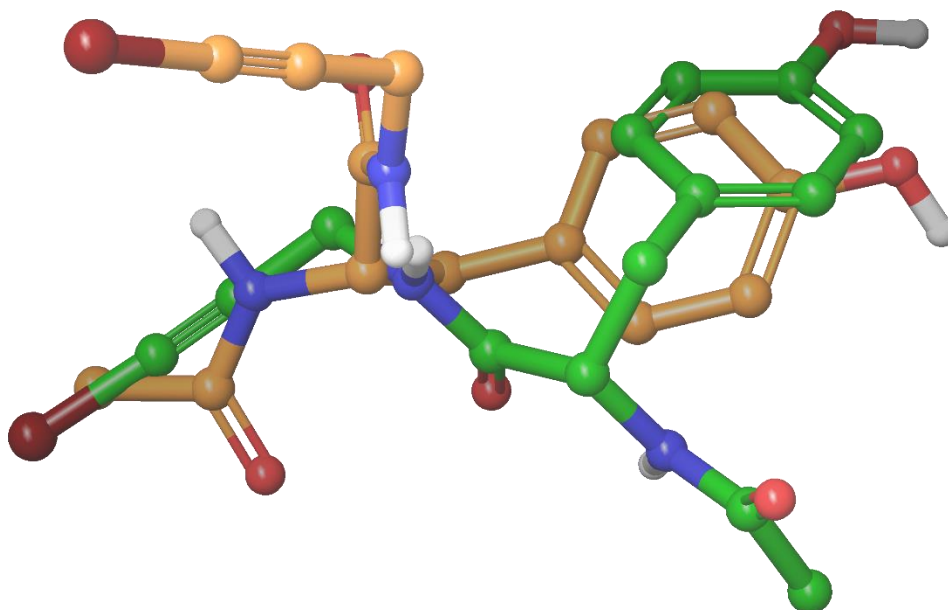


Figure 5. Comparison between the experimentally obtained pose of the ligand (green) with one predicted using a docking program (orange). The RMSD between them is 4.78

In vitro

As previously mentioned, docking hits are compounds which are expected to be active on the target of interest, but this has to be demonstrated. The activity of the compound for the target of interest can be tested in the laboratory and this is usually achieved by using a commercial enzymatic activity kit or by developing an in-house method for the detection of enzymatic activity and comparing the activities of the target with and without the presence of the compound that was obtained as a hit.

The in vitro activity of docking hits for the target of interest is a key step in order to validate a workflow. To minimize the expensive in vitro assays, it is essential to choose compounds that differ from one another. This can be done by sorting the hit compounds into several clusters

Introduction

based on their structural similarities⁶⁴. As a result, it is possible to: (a) choose hit compounds that are structurally diverse and belong to different clusters than known actives, making them more novel and interesting; (b) avoid choosing compounds that cluster with inactives and are therefore more likely to be inactive; and (c) choose hit compounds that are members of different clusters. There are various clustering techniques, including hierarchical clustering and k-means clustering, and the choice of approach is typically impacted by the features of the dataset and the limits of each technique⁶⁴.

Predicting the binding affinity for a specific protein-ligand complex is the goal of binding affinity prediction. Certain scoring methods cannot provide scores that can directly match to experimental binding data. The Pearson correlation value between projected scores and experimental binding data on benchmark test sets is therefore a commonly used criteria for predicting affinity. An alternate criterion is the Spearman ranking correlation coefficient which determines the correlation between the two ranking sets after first ranking the expected and experimental scores^{30,69}.

Another mean to better select which predicted actives to test is the use of machine learning. Training a model with active and inactive compounds^{58,64} against SARS-CoV-2 M-pro and using it after the docking would allow us predict the activity and filter the results. Contrary to traditional similarity approaches, which are based on comparing the compounds, giving equal weight to all parts of the molecule, machine-learning methods avoid this issue because they are constructed from multiple active molecules, and fingerprint bits of each compound relate to the bioactivity separately. As a result, machine learning can distinguish the fingerprint bits that are crucial for bioactivity, their main advantage^{58,64}.

References

1. Worldometer COVID-19 Coronavirus pandemic at <https://www.worldometers.info/coronavirus/> (accessed on April 9, 2023).
2. Whitelaw, S., Mamas, M. A., Topol, E. & van Spall, H. G. C. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit Health* **2**, e435-e440 (2020).
3. Forslid, R. & Herzing, M. Assessing the consequences of quarantines during a pandemic. *Eur J Health Econ.* **22**, 1115-1128 (2021).
4. Sayin Kasar, K. & Karaman, E. Life in lockdown: Social isolation, loneliness and quality of life in the elderly during the COVID-19 pandemic: A scoping review. *Geriatr Nurs.* **42**, 1222-1229 (2021).
5. Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R. & Wang, D. Potentially long-lasting effects of the pandemic on scientists. *Nat Commun.* **12**, 6188 (2021).
6. Sharma, A., Ahmad Farouk, I. & Lal, S. K. COVID-19: A Review on the Novel Coronavirus Disease Evolution, Transmission, Detection, Control and Prevention. *Viruses* **13**, 202 (2021).
7. To, K. K. W. *et al.* Lessons learned 1 year after SARS-CoV-2 emergence leading to COVID-19 pandemic. *Emerg Microbes Infect.* **10**, 507-535 (2021).
8. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. *J Chem Inf Model.* **53**, 1893-1904 (2013).
9. Arya, R. *et al.* Structural insights into SARS-CoV-2 proteins. *J Mol Biol.* **433**, 166725 (2021).
10. Kadam, S. B., Sukhramani, G. S., Bishnoi, P., Pable, A. A. & Barvkar, V. T. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights. *J Basic Microbiol.* **61**, 180-202 (2021).
11. Kirtipal, N., Bharadwaj, S. & Kang, S. G. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infect Genet Evol.* **85**, 104502 (2020).
12. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol.* **23**, 3-20 (2022).
13. Mistry, P. *et al.* SARS-CoV-2 Variants, Vaccines, and Host Immunity. *Front Immunol.* **12**, 809244 (2021).
14. Ochani, R. K. *et al.* COVID-19 pandemic: from origins to outcomes. A comprehensive review of viral pathogenesis, clinical manifestations, diagnostic evaluation, and management. *Infez Med.* **29**, 20-36 (2021).
15. Gimeno, A. *et al.* Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition. *Int J Mol Sci.* **21**, 3793 (2020).
16. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293 (2020).

Introduction

17. Sobhia, M. E., Ghosh, K., Sivangula, S., Kumar, S. & Singh, H. Identification of potential SARS-CoV-2 Mpro inhibitors integrating molecular docking and water thermodynamics. *J Biomol Struct Dyn.* **40**, 5079-5089 (2020).
18. Pavlova, A. *et al.* Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *Chem Sci.* **12**, 1513-1527 (2021).
19. Panchariya, L. *et al.* Zinc²⁺ ion inhibits SARS-CoV-2 main protease and viral replication in vitro. *Chem Commun (Camb).* **57**, 10083–10086 (2021).
20. Roberts, B. C. & Mancera, R. L. Ligand-protein docking with water molecules. *J Chem Inf Model.* **48**, 397–408 (2008).
21. Venugopal, P. P., Singh, O. & Chakraborty, D. Understanding the role of water on temperature-dependent structural modifications of SARS CoV-2 main protease binding sites. *J Mol Liq.* **363**, 119867 (2022).
22. Ansari, N., Rizzi, V., Carloni, P. & Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. *J Am Chem Soc.* **143**, 12930–12934 (2021).
23. Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* **43**, 4568–4574 (2004).
24. Paasche, A. *et al.* Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry* **53**, 5930–5946 (2014).
25. Poater, A. Michael Acceptors Tuned by the Pivotal Aromaticity of Histidine to Block COVID-19 Activity. *J Phys Chem Lett.* **11**, 6262–6265 (2020).
26. Gentile, F., Oprea, T. I., Tropsha, A. & Cherkasov, A. Surely you are joking, Mr Docking! *Chem Soc Rev.* **52**, 872-878 (2023).
27. Ferreira, L. G., dos Santos, R. N., Oliva, G. & Andricopulo, A. D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **20**, 13384-13421 (2015).
28. De Ruyck, J., Brysbaert, G., Blossey, R. & Lensink, M. F. Molecular docking as a popular tool in drug design, an in silico travel. *Adv Appl Bioinform Chem.* **9**, 1-11 (2016).
29. Torres, P. H. M., Sodero, A. C. R., Jofily, P. & Silva-Jr, F. P. Key Topics in Molecular Docking for Drug Design. *Int J Mol Sci.* **20**, 4574 (2019).
30. Yang, C., Chen, E. A. & Zhang, Y. Protein–Ligand Docking in the Machine-Learning Era. *Molecules* **27**, 4568 (2022).
31. Allen, W. J. *et al.* DOCK 6: Impact of New Features and Current Docking Performance. *J Comput Chem.* **36**, 1132-1156 (2015).
32. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J Mol Recognit.* **9**, 1–5 (1996).
33. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol.* **295**, 337–356 (2000).

34. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem.* **31**, 455-461 (2010).
35. Wang, R., Lai, L. & Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des.* **16**, 11–26 (2002).
36. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem.* **47**, 1739–1749 (2004).
37. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem.* **47**, 1750–1759 (2004).
38. Ain, Q. U., Aleksandrova, A., Roessler, F. D. & Ballester, P. J. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip Rev Comput Mol Sci.* **5**, 405-424 (2015).
39. Tang, S. *et al.* Accelerating AutoDock Vina with GPUs. *Molecules* **27**, 3041 (2022).
40. Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663-668 (2020).
41. Li, H., Leung, K. S. & Wong, M. H. Idock: A multithreaded virtual screening tool for flexible ligand docking. 2012 IEEE Symposium on Computational Intelligence and Computational Biology (CIBCB), San Diego, CA, USA. 77–84 (2012)
42. Jaghoori, M. M., Bleijlevens, B. & Olabarriaga, S. D. 1001 Ways to run AutoDock Vina for virtual screening. *J Comput Aided Mol Des* **30**, 237-249 (2016).
43. Ding, J. *et al.* Vina-GPU 2.0: Further Accelerating AutoDock Vina and Its Derivatives with Graphics Processing Units. *J Chem Inf Model.* In Press (2023)
44. Alhossary, A., Handoko, S. D., Mu, Y. & Kwok, C. K. Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics* **31**, 2214–2216 (2015).
45. Nivedha, A. K., Thieker, D. F., Makeneni, S., Hu, H. & Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J Chem Theory Comput.* **12**, 892–901 (2016).
46. Koebel, M. R., Schmadeke, G., Posner, R. G. & Sirimulla, S. AutoDock VinaXB: implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina. *J Cheminform.* **8**, 27 (2016).
47. Quiroga, R. & Villarreal, M. A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS One* **11**, e0155183 (2016).
48. Vieira, T. F. & Sousa, S. F. Comparing AutoDock and Vina in Ligand/Decoy Discrimination for Virtual Screening. *Appl Sci.* **9**, 4538 (2019).
49. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* **61**, 3891–3898 (2021).

Introduction

50. Rants'o, T. A., Johan van der Westhuizen, C. & van Zyl, R. L. Optimization of covalent docking for organophosphates interaction with Anopheles acetylcholinesterase. *J Mol Graph Model*. **110**, 108054 (2022).
51. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. v. & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*. **11**, 425–445 (1997).
52. Jug, G., Anderluh, M. & Tomašič, T. Comparative evaluation of several docking tools for docking small molecule ligands to DC-SIGN. *J Mol Model*. **21**, 1–12 (2015).
53. McGaughey, G. B. *et al.* Comparison of topological, shape, and docking methods in virtual screening. *J Chem Inf Model*. **47**, 1504–1519 (2007).
54. McGann, M. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model*. **51**, 578–596 (2011).
55. Pantsar, T. & Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **23**, 1899 (2018).
56. Macip, G. *et al.* Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Med Res Rev*. **42**, 744-769 (2022).
57. Amaro, R. E. *et al.* Ensemble Docking in Drug Discovery. *Biophys J*. **114**, 2271-2278 (2018).
58. Scior, T. *et al.* Recognizing pitfalls in virtual screening: a critical review. *J Chem Inf Model*. **52**, 867–881 (2012).
59. Pinzi, L. & Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int J Mol Sci*. **20**, 4331 (2019).
60. Ramírez, D. & Caballero, J. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?. *Molecules* **23**, 1038 (2018).
61. Menin, S., Pavan, M., Salmaso, V., Sturlese, M. & Moro, S. Thermal Titration Molecular Dynamics (TTMD): Not Your Usual Post-Docking Refinement. *Int J Mol Sci*. **24**, 3596 (2023).
62. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*. **10**, 449-461 (2015).
63. Wang, E. *et al.* End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev*. **119**, 9478–9508 (2019).
64. Gimeno, A. *et al.* The Light and Dark Sides of Virtual Screening: What Is There to Know?. *Int J Mol Sci*. **20**, 1375 (2019).
65. Cereto-Massagué, A. *et al.* DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* **28**, 1661–1662 (2012).
66. Sunseri, J. & Koes, D. R. Virtual Screening with Gnina 1.0. *Molecules* **26**, 7369 (2021).
67. Brozell, S. R. *et al.* Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des*. **26**, 749-773 (2012).

68. Liebeschuetz, J. W., Cole, J. C. & Korb, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J Comput Aided Mol Des.* **26**, 737–748 (2012).
69. Forli, S. *et al.* Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc.* **11**, 905-919 (2016).

Introduction

Hypotheses and objectives

Introduction

At the beginning of the COVID-19 pandemic, our research group developed a virtual screening protocol based on protein-ligand docking to suggest new SARS-CoV-2 M-pro inhibitors among approved drugs¹⁵. When we compared our results with other suggested SARS-CoV-2 M-pro inhibitors, we found that several manuscripts predicted different compounds, even when using a similar starting dataset and methodology. The high number of proposed M-pro inhibitors and structures of non-covalent complexes between SARS-CoV-2 M-pro and different inhibitors available in a short span of time, provided the breeding ground for this thesis.

The main objective of this thesis is to analyse in detail the **limits** of protein-ligand docking to predict SARS-CoV-2 M-pro inhibitors.

To achieve this goal, other important objectives are:

- To **compile** a non-redundant set of SARS-CoV-2 M-pro **inhibitors**. This set of inhibitors would include only those compounds whose inhibitory capacity, mainly expressed as the IC₅₀ value, against M-pro from SARS-CoV-2 has been determined.
- To assess whether docking scores obtained from known SARS-CoV-2 M-pro **inhibitors** **correlate** with their **activity**.
- To **review** several **VS protocols**, based on protein-ligand docking, that predict drugs as SARS-CoV-2 M-pro inhibitors.
- To test the **re-docking** and **cross docking** capabilities of the two most widely used docking programs (Autodock Vina and Glide) when trying to **reproduce** experimentally known SARS-CoV-2 M-pro inhibitors.
- To analyse the impact of **water molecules** in the **binding site** of SARS-CoV-2 M-pro during the docking.
- To evaluate the **influence** of the protonation of key **histidines** in the binding site of SARS-CoV-2 M-pro.
- To find the **best conditions** for predicting by docking the crystallized pose of a set of SARS-CoV-2 inhibitors complexed with the M-pro.

Hypotheses and objectives

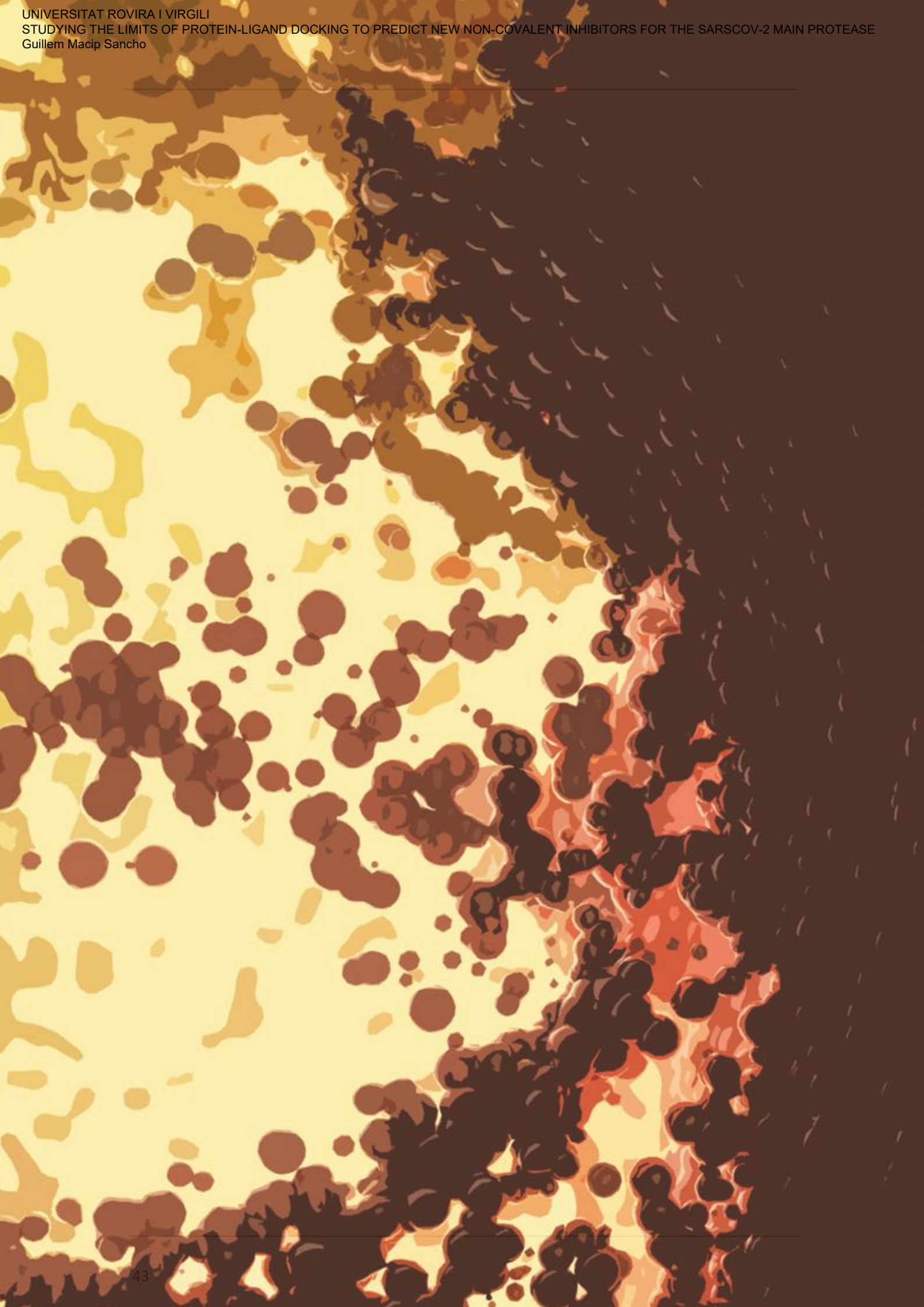
Al començament de la pandèmia COVID-19, el nostre grup de recerca va desenvolupar un protocol de cribratge virtual basat en docking de proteïna-l·ligand per suggerir nous inhibidors de la M-pro del SARS-CoV-2 entre fàrmacs aprovats. Quan vam comparar els nostres resultats amb altres inhibidors de la M-pro del SARS-CoV-2 suggerits, vam trobar que diversos manuscrits prediuen diferents compostos, fins i tot quan s'utilitzaven un conjunt de dades i una metodologia de partida similars. L'elevat nombre d'inhibidors M-pro proposats i estructures de complexos no covalents entre la M-pro del SARS-CoV-2 i diferents inhibidors disponibles en un curt període de temps, va proporcionar el caldo de cultiu per a aquesta tesi.

L'objectiu principal d'aquesta tesi és analitzar amb detall els límits del docking proteïna-l·ligand per predir els inhibidors de la M-pro del SARS-CoV-2.

Per aconseguir aquesta meta, altres objectius importants són:

- **Compilar** un conjunt no redundant d'**inhibidors** de la M-pro del SARS-CoV-2. Aquest conjunt d'inhibidors inclouria només aquells compostos la capacitat inhibidora dels quals, expressada principalment com a valor IC50, contra la M-pro del SARS-CoV-2 s'ha determinat experimentalment.
- Avaluar si el docking score obtinguts dels inhibidors coneguts de la M-pro del SARS-CoV-2 es **correlacionen** amb la seva activitat.
- **Revisar** diversos **protocols VS**, basats en docking de proteïna-l·ligand, que prediuen fàrmacs com a inhibidors de la M-pro del SARS-CoV-2.
- Per provar les capacitats de **re-docking** i **cross docking** dels dos programes d'acoblament més utilitzats (Autodock Vina i Glide) quan s'intenta **reproduir** inhibidors de la M-pro del SARS-CoV-2 coneguts experimentalment.
- Analitzar l'impacte de les **molècules d'aigua** al **binding site** de la M-pro del SARS-CoV-2 durant el docking.
- Avaluar la influència de la protonació d'**histidines** clau en el lloc d'unió de la M-pro del SARS-CoV-2.
- Trobar les **millors condicions** per predir mitjançant docking de la pose cristal·litzada d'un conjunt d'inhibidors del SARS-CoV-2 complexats amb la M-pro.

Results



A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet?

Guillem Macip, Pol Garcia-Segura, Júlia Mestres-Truyol, Bryan Saldivar-Espinoza, Gerard Pujadas* and Santiago Garcia-Vallvé*

Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Campus Sescelades, Universitat Rovira i Virgili, 43007 Tarragona, Catalonia, Spain;
guillem.macip@gmail.com (G.M.); polgarse2@gmail.com (P.G.-S.); juliamt110@gmail.com (J.M.-T.); bsaldvar.emc2@gmail.com (B.S.-E.)

* Correspondence: gerard.pujadas@gmail.com (G.P.); santi.garcia-vallve@urv.cat (S.G.-V.)



Review

A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet?

Guillem Macip , Pol Garcia-Segura , Júlia Mestres-Truyol, Bryan Saldivar-Espinoza, Gerard Pujadas * and Santiago Garcia-Vallvé *

Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Campus Sescelades, Universitat Rovira i Virgili, 43007 Tarragona, Catalonia, Spain; guillem.macip@gmail.com (G.M.); polgarse2@gmail.com (P.G.-S.); juliamt110@gmail.com (J.M.-T.); bsaldivar.emc2@gmail.com (B.S.-E.)

* Correspondence: gerard.pujadas@gmail.com (G.P.); santi.garcia-vallve@urv.cat (S.G.-V.)

Abstract: In this review, we collected 1765 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) M-pro inhibitors from the bibliography and other sources, such as the COVID Moonshot project and the ChEMBL database. This set of inhibitors includes only those compounds whose inhibitory capacity, mainly expressed as the half-maximal inhibitory concentration (IC₅₀) value, against M-pro from SARS-CoV-2 has been determined. Several covalent warheads are used to treat covalent and non-covalent inhibitors separately. Chemical space, the variation of the IC₅₀ inhibitory activity when measured by different methods or laboratories, and the influence of 1,4-dithiothreitol (DTT) are discussed. When available, we have collected the values of inhibition of viral replication measured with a cellular antiviral assay and expressed as half maximal effective concentration (EC₅₀) values, and their possible relationship to inhibitory potency against M-pro is analyzed. Finally, the most potent covalent and non-covalent inhibitors that simultaneously inhibit the SARS-CoV-2 M-pro and the virus replication in vitro are discussed.

Keywords: COVID-19; M-pro inhibitors; 3CL-pro inhibitors; computational chemistry; protease inhibitors; virtual screening



Citation: Macip, G.; Garcia-Segura, P.; Mestres-Truyol, J.; Saldivar-Espinoza, B.; Pujadas, G.; Garcia-Vallvé, S. A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet? *Int. J. Mol. Sci.* **2022**, *23*, 259. <https://doi.org/10.3390/ijms23010259>

Academic Editors: Francesco Caruso and Miriam Rossi

Received: 3 December 2021

Accepted: 25 December 2021

Published: 27 December 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/>).

1. Introduction

Since the onset of the COVID-19 pandemic, the scientific community has focused on studying the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus that causes the disease and on developing therapies and vaccines, several of which have been developed in record time. In the pharmacological field, no drugs have yet been definitively approved to inhibit the replication of SARS-CoV-2 and stop the development of COVID-19. Several targets are being studied, including the main protease (M-pro), which plays an essential role in virus replication [1]. This protease and the papain-like protease cleave the pp1a and pp1ab polyproteins to produce several nonstructural proteins, including M-pro itself, required for virus replication and transcription [1]. The high conservation of M-pro among related viruses, the importance of M-pro in the replication of the virus and the fact that M-pro only exists in coronaviruses and not in humans makes it an attractive target for the development of antiviral drugs [2–4]. SARS-CoV-2 M-pro has 306 amino acids that form three domains (I, II and III) [4]. The M-pro binding site is located between domains I and II, and domain III is involved in dimerization, which is essential for M-pro activity [4]. Similar to the M-pro from SARS-CoV-1 and other coronaviruses, SARS-CoV-2 M-pro has two catalytic amino acids, His41 and Cys145 (Figure 1). A catalytic water molecule is also important and makes a strong hydrogen bond with His41 [5]. Although some allosteric binding sites have been identified for the SARS-CoV-2 M-pro [6–9], most of the inhibitors crystallized within the M-pro bind to the active site [10]. One strategy used to find SARS-CoV-2 M-pro inhibitors, especially at the beginning of the pandemic, was

drug repositioning [2,11–13]. This strategy is based on looking for drugs approved for one disease (therefore, its safety and possible adverse effects are known) that can be used to treat another—in this case, COVID-19. One of the most widely used computational tools for repositioning drugs, or looking for compounds with new activities, is what is known as protein-ligand docking. This tool predicts whether a particular molecule can bind (and, if it can, how) to a particular target (for example, the SARS-CoV-2 M-pro [14]). However, protein-ligand docking has several limitations, such as the consideration of the protein as a rigid body and the lack of confidence in the ability of scoring functions to give accurate binding energies [15,16]. In addition, the flexibility of the SARS-CoV-2 M-pro makes it a challenging target for small-molecule inhibitor design [17]. Using two different SARS-CoV-2 M-pro structures and five protein-ligand docking methods, we have recently shown that docking scores or the Gibbs free energy (ΔG) calculated with an MM-GBSA method [18] do not correlate with bioactivity [19], probably because of the inability of common docking programs to correctly reproduce the binding modes of SARS-CoV-2 M-pro inhibitors [20]. This reinforces the idea that it is essential to validate the results obtained by protein-ligand docking or any other computational tool, especially when analyzing SARS-CoV-2 M-pro inhibitors [19,21–23]. The results of protein-ligand docking can be computationally validated by re-docking, cross-docking and applying the same protocol to a set of known active compounds and a set of decoy or inactive compounds [19]. Protein-ligand docking is expected to discriminate decoys from active compounds. If docking scores are used to rank the potency of a set of compounds, it must first be demonstrated that there is a correlation between docking scores and activity or potency, for example, expressed as IC_{50} values [19]. However, the best way to validate the predictions of protein-ligand docking is to experimentally test the predicted bioactivity of selected hits.

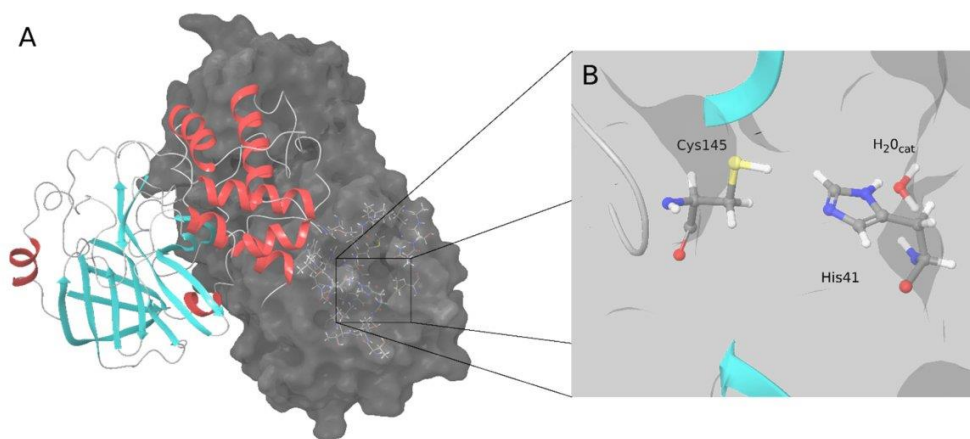


Figure 1. SARS-CoV-2 M-pro structure. (A) Biological assembly of the M-pro in its dimeric form. The left protomer is shown in cartoon representation, colored by protein secondary structure, and the right protomer is displayed as a surface. (B) Detailed snapshot of the catalytic water, Cys145 and His41.

Since the beginning of the COVID-19 pandemic, developing SARS-CoV-2 M-pro inhibitors has been an active area of research. However, it did not have to start from scratch. Previous research about protease inhibitors, especially from SARS-CoV and MERS-CoV, proved to be useful [24,25]. Known inhibitors of proteases from HIV and Hepatitis C virus, in addition to calpain and caspase-3 inhibitors, were systematically analyzed to test their capacity to inhibit the SARS-CoV-2 M-pro [26]. Compounds developed against the M-pro of other coronaviruses were also tested, and some were found to be potent SARS-CoV-2 M-pro inhibitors [24,27,28]. The complete genome sequence of SARS-CoV-2 [29] and the first crystallized structure of the SARS-CoV-2 M-pro [4] were two important

milestones in the development of new SARS-CoV-2 M-pro inhibitors. The article describing the first crystallized SARS-CoV-2 M-pro structure (the 6LU7 structure) also presented the first SARS-CoV-2 M-pro inhibitors [4]. These first inhibitors included the N3 compound, which had previously been developed as a protease inhibitor for multiple coronaviruses, including SARS-CoV and MERS-CoV, approved drugs (such as disulfiram and carmo-fur) and preclinical or clinical-trial drug candidates (ebselen, shikonin, tideglusib, PX-12 and TDZD-8) [4]. Since then, thousands of compounds have been suggested as SARS-CoV-2 M-pro inhibitors through computational methods such as protein-ligand docking, high-throughput screening experiments, computer-aided design and synthesis of new compounds. Several articles have reviewed the SARS-CoV-2 M-pro inhibitors discovered to date [25,28,30–38]. In this review, we collected 1765 SARS-CoV-2 M-pro inhibitors from the bibliography and other sources such as the COVID Moonshot project [39,40] and the ChEMBL release 29 database [41]. This set of inhibitors includes only those compounds whose inhibitory capacity, mainly expressed as the half-maximal inhibitory concentration (IC_{50}) value, against M-pro from SARS-CoV-2 has been determined. We did not include compounds predicted only by docking or other computational tools. As we have discussed previously, to avoid false positives and false negatives, the results of automated protein-ligand docking should not be used as the only evidence to predict SARS-CoV-2 M-pro inhibitors [19]. Several covalent warheads are used to treat covalent and non-covalent inhibitors separately. Chemical space, the variation of the IC_{50} inhibitory activity when measured by different methods or laboratories and the influence of 1,4-dithiothreitol (DTT) are discussed. When available, virus replication inhibition values, measured with a cellular antiviral assay, were collected, and their relationship with the inhibitory potency against M-pro is shown. Finally, the most potent inhibitors that simultaneously inhibit the SARS-CoV-2 M-pro and the virus replication in vitro are discussed.

2. SARS-CoV-2 M-Pro Inhibitors

Table 1 shows the origin of the non-redundant set of 1765 SARS-CoV-2 M-pro inhibitors collected (see supplementary file S1). This set of inhibitors includes only those compounds whose inhibitory capacity, mainly expressed as the IC_{50} value, against M-pro from SARS-CoV-2 has been determined. A total of 758 compounds were extracted from peer-reviewed articles published between January 2020 and August 2021. When multiple IC_{50} values were found for the same compound, the mean value was calculated. From a set of 1037 M-pro inhibitors with an IC_{50} value downloaded from the COVID Moonshot project [39,40] on 1st October 2021, the compounds that had already been collected from the bibliographic search were discarded. In the end, 999 compounds were collected from COVID Moonshot. The IC_{50} values of these compounds were estimated as the mean value of the IC_{50} values from two biochemical assays: a fluorescence-based assay and a Rapid-Fire Mass Spectrometry assay. Finally, 8 compounds were collected from the ChEMBL database [41], which contained more SARS-CoV-2 M-pro inhibitors, but most of them had already been collected from the bibliography. The SMILES of the 1765 SARS-CoV-2 M-pro inhibitors were standardized with the Standardizer 21.15.0 program from ChemAxon (<http://www.chemaxon.com>, accessed on 4 September 2021). The pIC_{50} values of the SARS-CoV-2 M-pro inhibitors collected range from 2.5 to 9.0 (Table 1). Putative covalent inhibitors were identified by the presence of typical covalent warheads (Table 2). When one of these warheads is in the appropriate position within the M-pro binding site, it can form a covalent bond, usually with the catalytic Cys145 [25]. There are twice as many non-covalent inhibitors as putative covalent inhibitors (Table 1), although pIC_{50} values are highest in some putative covalent inhibitors (Table 1 and Figure 2). However, conventional IC_{50} measurements are of limited value for characterizing the potency of irreversible covalent inhibitors, because incubation for different periods of time would provide different IC_{50} values [42]. Other parameters, such as molecular weight, LogP, number of hydrogen bond donors and hydrogen bond acceptors were similar between the covalent and non-covalent sets (see Supplementary Figure S1).

Table 1. Number of SARS-CoV-2 M-pro inhibitors collected.

SARS-CoV-2 M-Pro Inhibitor Set	Number of Compounds (Covalent/Non-Covalent) ¹	pIC ₅₀ Range	pIC ₅₀ Range Covalent	pIC ₅₀ Range Non-Covalent
From the bibliography	758 (346/412)	2.5–9.0	3.4–9.0	2.5–8.3
From COVID Moonshot	999 (205/794)	4.0–7.8	4.0–7.8	4.0–7.4
From ChEMBL	8 (1/7)	5.4–6.1	5.4	5.5–6.1
All	1765 (552/1213)	2.5–9.0	3.4–9.0	2.5–8.3

¹ Putative covalent and non-covalent inhibitors were identified by the presence or absence of the covalent warheads shown in Table 2.

Table 2. Covalent warheads that can be used to identify putative covalent inhibitors. It shows the SMARTS that can be used to identify each warhead and some examples of SARS-CoV-2 inhibitors that contain each warhead. These covalent warheads were used to identify putative covalent inhibitors among the known SARS-CoV-2 M-pro inhibitors.

Warhead	SMARTS	Examples
Acrylamide	[C;H2:1]=[C;H1]C(N)=O	CVD-0004255
Chloroacetamide	Cl[C;H2:1]C(N)=O	BFC204
Vinylsulfonamide	NS(=O)([C;H1]=[C;H2:1])=O	
Nitrile	N#[C:1]-[*]	Isavuconazole
Michael acceptors	C=!@CC=[O,S]	Cinanserin, MPI2, MPI9, N3
Alpha-ketoamide	C(=O)(C=O)N	Boceprevir, narlaprevir, telaprevir, UAWJ248
Aldehyde	[CX3H1](=O)	GC373, MI-05, MI-06, MI-09, MI-11, MI-13, MI-14, MI-21, MI-23, MI-28
Bisulfite adduct of aldehyde	C(O)S(=[OX1])([O])(=[OX1])	GC376
Urea carbonyl	[NX3][CX3](=[OX1])([NX3,nX3])	Carmofur
Bis(dialkylaminethiocarbonyl)disulfide	[CX3](=[SX1])SS[CX3](=[SX1])	Disulfiram
Carbamoylsulfanyl	[NX3,nX3][C,c](=[OX1])([SX2,sx2])	Tideglusib
Disulfide	[SX2][SX2]	PX-12
Hydroxymethyl ketone	[CX3H0](=[OX1])[CH2][OH]	PF-00835231
Alkoxyethyl ketone	[CX3H0](=[OX1])[CH2][OX2H0]	2683066-41-1, 2683066-42-2, 2683066-47-7
Acyloxyethyl ketone	[CX3H0](=[OX1])[CH2][OX2H0][CX3H0](=O)	2683066-41-1, 2683066-42-2, 2683066-47-7
Fluoro, Chloro-methyl ketone	[CX3H0](=[OX1])[CH2][Cl,F]	Z-AVLD-FMK
Ebselen related	[Se]n(c=O)	Ebselen

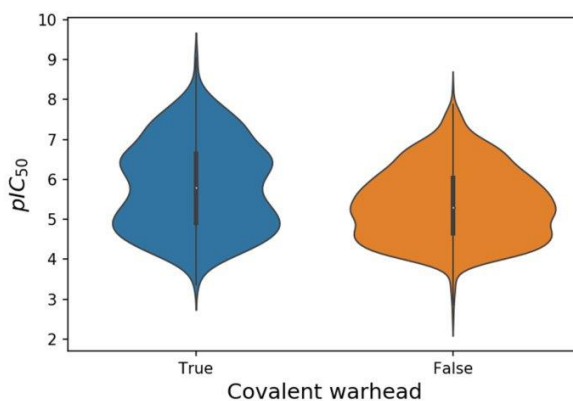


Figure 2. Violin plots of the pIC₅₀ values from 552 putative covalent and 1213 non-covalent SARS-CoV-2 M-pro inhibitors.

Figure 3 shows the t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of the chemical space of the set of SARS-CoV-2 M-pro inhibitors extracted from the bibliography. In this representation, more similar compounds are closer together. Peptidomimetic compounds, such as alpha-acyloxymethylketones, telaprevir, boceprevir, GC373 and their derivatives, which mimic natural peptide substrates, are closer together at the top left of the figure. Other clusters of compounds represent derivative compounds that have been synthesized from a lead compound to increase its bioactivity. Thus, derivatives from perampanel, ML300, ML188, ebsulfur, ebselen and myricetin form well-defined clusters. Perampanel derivatives are an example of a very successful increase in activity. Perampanel was first predicted as a SARS-CoV-2 M-pro inhibitor by consensus docking [2]. This prediction was confirmed by Jorgensen and coworkers, although perampanel showed only an approximate IC_{50} of 100–250 μM [43]. The same authors also optimized this compound and synthesized several derivative compounds [44–46]. Some of these perampanel derivatives have IC_{50} values in the low nanomolar range and are some of the most potent non-covalent SARS-CoV-2 M-pro inhibitors found to date.

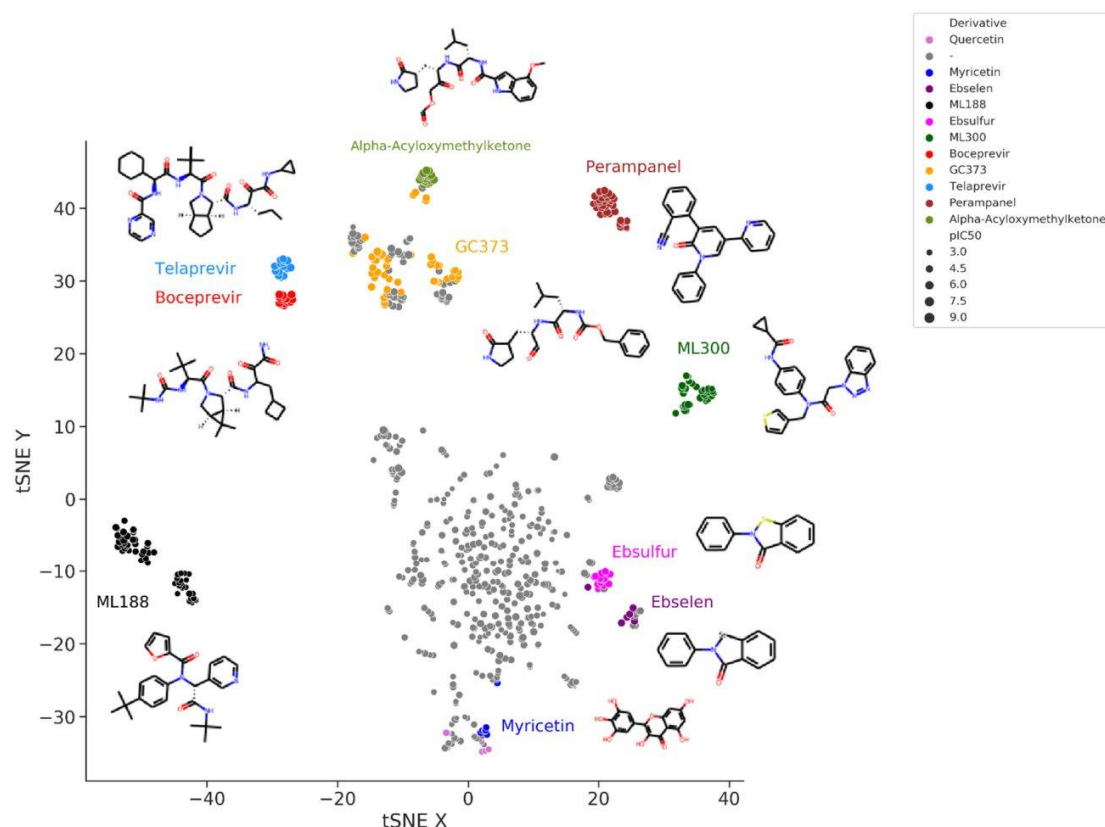


Figure 3. t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of the chemical space of a set of SARS-CoV-2 M-pro inhibitors extracted from the bibliography. Embedding is based on the 2048-bit Morgan fingerprint. Markers are colored according to several manually attributed chemotypes.

ML300 and ML188 are non-covalent inhibitors that were developed against the M-pro from SARS-CoV-1 [47,48]. Both compounds have been used to obtain more potent SARS-CoV-2 M-pro inhibitors that can inhibit SARS-CoV-2 replication in infected cells [49,50]. Boceprevir and telaprevir are approved protease inhibitors for treating hepatitis caused by the hepatitis C virus. Both compounds have been identified several times as covalent

inhibitors of the SARS-CoV-2 M-pro [43,51–58]. New bicyclopipeline derivatives have been designed and synthesized from them both [59]. All compounds inhibited SARS-CoV-2 M-pro in vitro, with IC_{50} values ranging from 7.6 to 748.5 nM [59]. In addition, two of them, MI-09 and MI-30, showed excellent antiviral activity in a cell-based assay and significantly reduced lung viral loads and lesions in a transgenic mouse model of SARS-CoV-2 infection [59]. GC376 is a covalent M-pro inhibitor that was developed as an inhibitor of the main protease of the feline coronavirus FCoV [60] that also showed activity against the M-pro from MERS and SARS-CoV viruses [61]. Its IC_{50} activity against SARS-CoV-2 M-pro ranges between 0.026 and 0.89 μ M [51,52,61–67]. GC376 is a prodrug, and its bisulphite adduct is converted to an aldehyde to form GC373. This aldehyde forms a covalent interaction with the catalytic Cys145 of the SARS-CoV-2 M-pro [61]. Several GC373 and GC376 derivative compounds have been designed and assayed [63,68,69]. Some of them, such as UAWJ248 [70], are more potent than GC376. A group of peptidomimetic compounds with an α -acyloxymethyl ketone warhead designed to form an irreversible covalent bond with Cys145 showed IC_{50} values against the SARS-CoV-2 M-pro in the nM range [71]. They also inhibited SARS-CoV-2 replication and presented low cytotoxicity and good stability [71]. Ebselen is a covalent inhibitor of the SARS-CoV-2 M-pro, although its specificity has been questioned [72,73]. Several derivative compounds of ebselen and its sulfur derivative ebsulfur have been analyzed [74,75]. Some of the derivative compounds displayed more potent M-pro inhibition than ebselen and ebsulfur [74,75]. However, the promiscuous behavior of ebselen and ebsulfur and their lack of cellular antiviral activity [74,75] may also be applied to their derivatives. Myricetin is a flavonoid that acts as a non-peptidomimetic and covalent inhibitor of SARS-CoV-2 [76,77]. Its covalent behavior was unexpected and caused by the pyrogallol moiety that formed a covalent bond with Cys145 [76]. Myricetin and its derivatives inhibit SARS-CoV-2 M-pro and SARS-CoV-2 replication in cells [76–79], and form a cluster at the bottom of Figure 3, near quercetin and other flavonoids.

2.1. Activity Assays for Identifying SARS-CoV-2 M-Pro Inhibitors

Several methods have been developed or adapted for quantifying the potency of the SARS-CoV-2 M-pro inhibitors [73,80–89]. These methods demonstrate the mechanism of action of antiviral drugs and do not require cells infected with SARS-CoV-2 or a laboratory with biosafety level 3 containment facilities. When combined with high-throughput sample processing and analysis, hundreds or thousands of compounds can be screened. These methods use a marked substrate, usually a peptide derivative, and when the M-pro is present, the substrate is cut, which induces the emission of a signal, usually fluorescence. In the presence of an M-pro inhibitor, signal intensity is reduced, and the potency of the inhibitor can be quantified. Some methods consist of an in vitro screen and use a purified M-pro protein, but they do not account for cell permeability, metabolism or cytotoxicity and cannot be used to accurately predict the cellular activity of M-pro inhibitors [73,83]. The M-pro protein can be expressed by transforming a bacterium with a plasmid encoding the SARS-CoV-2 M-pro. For purification purposes, an M-pro modified with the addition of specific residues, known as a tag, to the N- or C-terminus of the protein can be used. However, the tag may interfere with the binding of M-pro to its ligands [90]. M-pro requires a native N-terminus to form the active dimer [70,91], so a C-terminal His-tag has been used. However, this C-terminal His-tag can lower the binding affinity of a given ligand [90]. To overcome the limitations of the in vitro screens, cell-based assays have been developed [73,80,83,84,87,88]. In these assays, the cells express the M-pro and the reporter used and can differentiate cytotoxicity from true M-pro inhibition [83]. However, these assays often require a biosafety level 2 laboratory. To complement the activity assays, a thermal shift binding assay can be performed. This assay is based on the thermal stabilization of a protein when it binds to a protein. A ΔT_m shift of up to 18 °C has been observed for some SARS-CoV-2 inhibitors [51,70,73,92,93]. However, the thermal shift

assay might not be suitable for analyzing non-covalent M-pro inhibitors [90]. For covalent inhibitors, the binding can be confirmed by native mass spectrometry.

In all assays for identifying SARS-CoV-2 M-pro inhibitors, negative and positive controls are needed. As a positive control, a known SARS-CoV-2 M-pro inhibitor is used. Some of the inhibitors most commonly used as positive controls are GC376, boceprevir, ebselen, disulfiram and telaprevir. After the inhibition of the SARS-CoV-2 M-pro has been detected, a dose-response assay can be used to calculate the IC_{50} or its derivative pIC_{50} . Lower values of IC_{50} and higher values of pIC_{50} represent more potent inhibitors. Comparisons between IC_{50} values obtained with different methods and by different laboratories must be made with great care. Furthermore, irreversible covalent M-pro inhibitors cannot be unambiguously ranked for potency using IC_{50} values [94]. To show the differences in the pIC_{50} estimations for the same compound, we identified the compounds with the highest number of pIC_{50} values from our dataset of SARS-CoV-2 M-pro inhibitors. Figure 4 shows the variation in the pIC_{50} values of the five most evaluated compounds, GC376, boceprevir, ebselen, disulfiram and telaprevir. These compounds are usually used as positive controls, and multiple pIC_{50} values for each compound have been calculated. The variation in the pIC_{50} values for three of these five compounds can be higher than two pIC_{50} units. This means that the estimations of the IC_{50} values of a compound could differ by a factor of 100 between two different laboratories or methods.

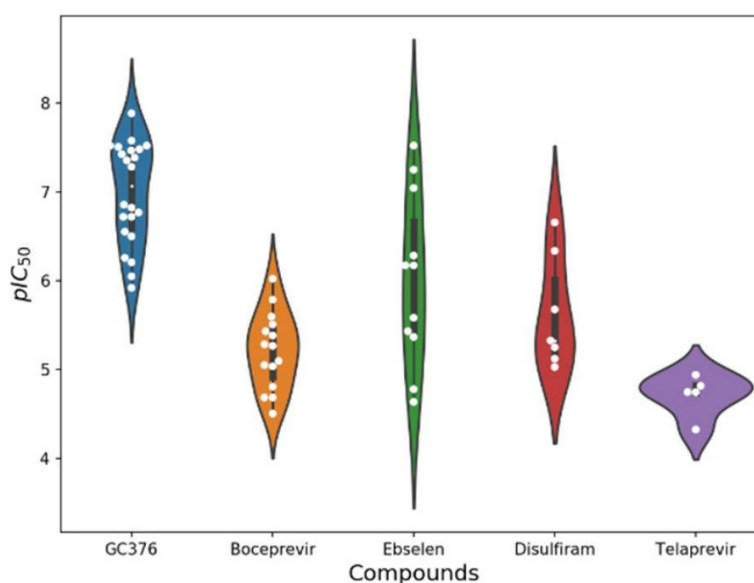


Figure 4. Comparison of multiple pIC_{50} values calculated by different laboratories. The pIC_{50} values of the five most evaluated compounds, GC376, boceprevir, ebselen, disulfiram and telaprevir, are shown as white points to highlight that different laboratories and methods can estimate different pIC_{50} values.

The inhibitory activity achieved in enzymatic assays is sensitive to the method and conditions used [73,95]. Compounds such as indinavir, lopinavir, nelfinavir, saquinavir, and tipranavir have shown no M-pro inhibitory activity at some of the concentrations analyzed [54,55,62,96]. Other compounds, such as candesartan, chloroquine, dipyridamole, montelukast and oxytetracycline, did not inhibit M-pro in four different assays tested [73]. The presence of DTT has been reported to affect the inhibitory activity of M-pro covalent inhibitors, as it maintains M-pro in a reduced state and eliminates non-specific thiol-reactive compounds [51,91]. Thus, if the inhibitory effect of an M-pro inhibitor is eliminated or greatly reduced by the presence of DTT, then the inhibition is non-specific. Therefore, the enzymatic inhibition potency of cysteine protease inhibitors measured in the absence of

DTT should not be used to predict cellular antiviral activity [72]. Some of the SARS-CoV-2 M-pro inhibitors whose inhibitory activity is eliminated or reduced by the presence of DTT are carmofur, disulfiram, ebselen, tideglusib, shikonin, PX-12 [72,73,97], and zinc pyrithione [77]. The results of a thermal shift-binding assay in the presence of DTT showed that some of these compounds (i.e., disulfiram, ebselen, tideglusib, shikonin and PX-12) did not bind to SARS-CoV-2 M-pro [72]. This means that these compounds are promiscuous cysteine inhibitors that are not specific for M-pro [72,73]. Figure 5 shows the effect of 1mM of DTT on the M-pro inhibition in a group of 246 SARS-CoV-2 M-pro inhibitors [77]. A total of 156 of these compounds showed a relative reduction in SARS-CoV-2 M-pro inhibition of more than 30% and can be considered as “DTT sensitive” [77].

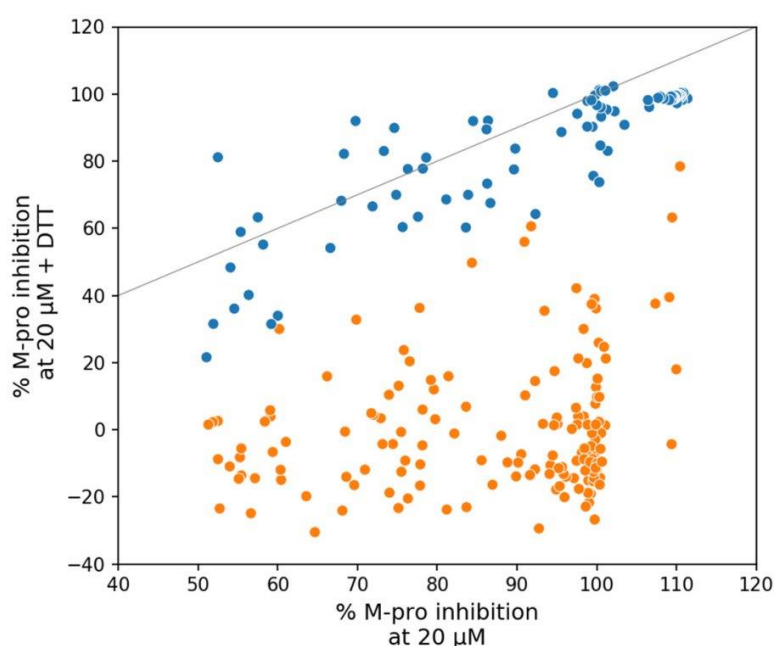


Figure 5. Effect of 1mM of DTT on M-pro inhibition in a group of 246 SARS-CoV-2 M-pro inhibitors [77]. A gray line represents the points where the presence of DTT does not affect the % of inhibition. Compounds far below the gray line are DTT-sensitive inhibitors and are colored orange. Compounds near the gray line are DTT insensitive inhibitors and are colored blue. The R^2 value of the insensitive compounds is 0.75 ($p = 1.1 \times 10^{-27}$).

2.2. Activity Assays for Identifying Molecules That Inhibit SARS-CoV-2 Replication

In addition to testing the inhibition of SARS-CoV-2 M-pro in vitro or in a cell-based assay, the ability of a compound to inhibit the SARS-CoV-2 replication also needs to be tested. An antiviral assay, using cells infected with SARS-CoV-2, is the gold standard assay, although it requires a biosafety level 3 laboratory. The Vero E6 cell line is one of the most common cell lines used for this analysis, but other cell lines have also been used. However, it has been reported that Vero cells express high levels of some efflux transporters, which may mask the true activity of some compounds [93]. A dose-response assay can be used to calculate a half maximal effective concentration (EC_{50}) value, defined as the concentration of the compound that reduces the viral-induced cytopathic effect by 50% (with respect to the virus control). However, the cytotoxicity of the compounds needs to be estimated to discount that the observed effect is not due to the toxicity of the compounds. For this purpose, the half-maximal cytotoxic concentration (CC_{50}) is usually used. Not all the compounds that inhibit the SARS-CoV-2 M-pro in vitro can inhibit the SARS-CoV-2 replication. For example, carmofur, disulfiram, ebselen, PX-2, shikonin and tideglusib

cannot inhibit SARS-CoV-2 replication in cell cultures [64,72,73]. It has been suggested that the possible antiviral activity of lopinavir and nelfinavir is due to their cytotoxicity [83,98]. The antiviral activity of other compounds against SARS-CoV-2, measured as an EC_{50} value, is slightly more than 10 times their M-pro inhibitory activity, measured as an IC_{50} value [54]. One possible explanation why compounds with high inhibitory activity in an in vitro M-pro assay show little or no activity in a cell assay is their low lipophilicity and the resulting poor cell membrane permeability. Figure 6 compares the pIC_{50} values of some SARS-CoV-2 M-pro inhibitors with their anti SARS-CoV-2 activities, measured as pEC_{50} values. To avoid comparisons between values from different laboratories, data from five articles [49,59,69,99,100] that calculate the pIC_{50} and pEC_{50} for a set of compounds are shown independently. In all but one case, the pEC_{50} values are lower than the pIC_{50} values, showing that the potency of a compound to inhibit SARS-CoV-2 replication in cells cannot always be inferred from the potency to inhibit the M-pro determined in vitro.

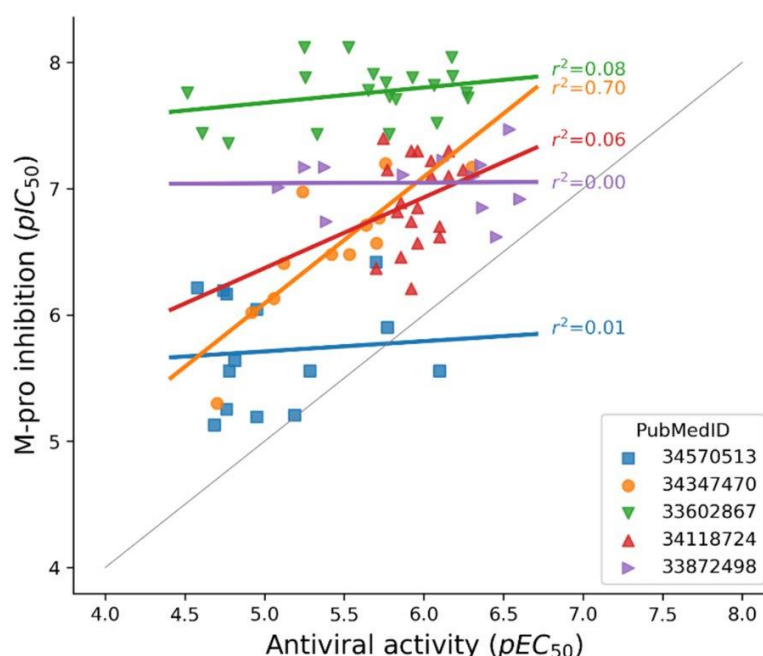


Figure 6. Comparison between the pIC_{50} values of some SARS-CoV-2 M-pro inhibitors and their antiviral activities, measured as pEC_{50} values. To avoid comparisons between values from different laboratories, data from five articles that calculate the pIC_{50} and pEC_{50} for a set of compounds are shown independently. The PubMedID of the articles is indicated. A gray line represents the diagonal where pIC_{50} and pEC_{50} values are equal.

2.3. Most Potent Covalent and Non-Covalent Inhibitors of SARS-CoV-2 M-Pro

Figure 7 and Table 3 show the 15 most potent non-covalent inhibitors of the SARS-CoV-2 M-pro. Only M-pro inhibitors that have pIC_{50} values and can inhibit SARS-CoV-2 replication in a cellular antiviral assay were included. The compound with the CAS number 339096-59-2, also named compound M3 [101], was obtained from the optimization of a previous SARS-CoV-2 M-pro inhibitor, named compound 13 [102]. Compound 339096-59-2 inhibited the SARS-CoV-2 M-pro in vitro with an IC_{50} of 13 nM and displayed anti-SARS-CoV-2 activity in Vero-E6 cells infected with SARS-CoV-2 [101]. The EC_{50} value was 16 nM [101]. This compound also inhibited in vitro human TMPRSS2 and furin enzymes, which are required for viral entry [101]. Compounds with the CAS number 2603242-35-7, 2603242-41-5, 2679814-93-6, 2679814-92-5, 2679814-91-4, 2679814-96-9, 2603242-04-0 are perampanel derivatives that can inhibit the SARS-CoV-2 M-pro in vitro, with IC_{50} values

between 0.128 and 0.018 μM [44,46]. The most potent compounds in this set, 2603242-35-7 and 2603242-41-5, showed antiviral activity in a viral plaque assay but lacked antiviral activity in a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay [44] (Table 3). In addition, both compounds showed the highest cytotoxicity [44]. More promising are the results of the other peramppanel derivatives. In particular, 2679814-93-6 showed lower values of EC_{50} and negligible cytotoxicity in Vero E6 cells [44]. Compounds 2694063-46-0, CCF0058981, 2694063-44-8 and 2694063-65-3 are derived from ML300 with IC_{50} values against SARS-CoV-2 M-pro of 0.063, 0.068, 0.106 and 0.171 μM , respectively [49]. CCF0058981 showed the highest anti-SARS-CoV-2 activity in infected Vero E6 cells, with an EC_{50} of 497 nM, and low cytotoxicity ($\text{CC}_{50} > 50 \mu\text{M}$) [49]. The compound with the CAS number 81418-42-0 is a juglone derivative that exhibited an IC_{50} value of 72 nM against the SARS-CoV-2 M-pro and that effectively suppressed the replication of SARS-CoV-2 in Vero E6 cells [103]. Compound 392732-12-6 (MCULE-7013373725-0) is a SARS-CoV-2 M-pro inhibitor, with an IC_{50} value of 0.11 μM that also showed significant inhibition activity against SARS-CoV-2 replication ($\text{EC}_{50} = 0.11 \mu\text{M}$) [102]. Interestingly, this compound also inhibited the SARS-CoV-2 papain protease (PL-pro) and human furin protease [102]. The dual activity against the viral and host proteases and dual activity against SARS-CoV-2 M-pro and PL-pro are very interesting. However, more studies about the cytotoxicity of this compound are needed [102]. Its selectivity index (i.e., the ratio between the antiviral activity and cytotoxicity) has room for improvement [102]. PET-UNK-29afea89-2 is a 3-aminopyridine compound from the COVID Moonshot project submitted in October 2020. It showed a high SARS-CoV-2 inhibition and anti-SARS-CoV-2 activity, with IC_{50} and EC_{50} values of 0.129 and 0.244 μM , respectively [40]. However, it was revealed to be a metabolically unstable compound [40]. The attempts to improve the stability of the compound slightly decreased compound potency but significantly increased metabolic stability and oral bioavailability [40].

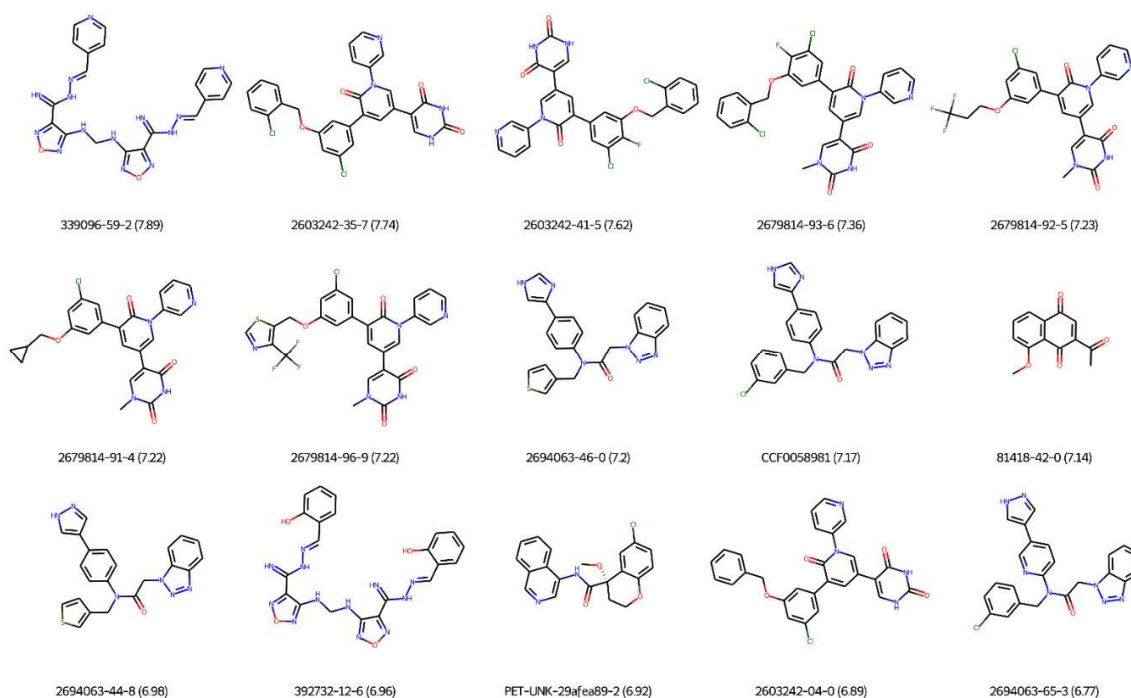


Figure 7. The 15 most potent non-covalent inhibitors of the SARS-CoV-2 M-pro. Only M-pro inhibitors with pIC_{50} values that can inhibit SARS-CoV-2 replication in a cellular antiviral assay are shown. pIC_{50} values are shown in parentheses.

Table 3. Data from the 15 most potent non-covalent inhibitors of the SARS-CoV-2 M-pro.

Compound	IC ₅₀ μ M (pIC ₅₀)	EC ₅₀ μ M	CC ₅₀ μ M	Reference
339096-59-2 (M3)	0.013 (7.89)	0.016 ^a	Higher than the EC ₅₀ by 2.3-fold	[101]
2603242-35-7 (21)	0.018 (7.74)	11.30 ^{a,b}	1.7 ^a	[44]
2603242-41-5 (23; 18)	0.024 (7.62)	0.84 ^{a,b}	1.15 ^a	[44,46]
2679814-93-6 (19)	0.044 (7.36)	0.08 ^a	>32.5 ^a	[46]
2679814-92-5 (17)	0.059 (7.23)	0.82 ^a	>100 ^a	[46]
2679814-91-4 (16)	0.061 (7.22)	1.20 ^a	82 ^a	[46]
2679814-96-9 (21)	0.061 (7.22)	1.08 ^a	>100 ^a	[46]
2694063-46-0 (21)	0.063 (7.20)	1.74 ^a	-	[49]
CCF0058981 (41)	0.068 (7.17)	0.50 ^a	>50 ^a	[49]
81418-42-0 (15)	0.072 (7.14)	4.55 ^a	viability ^a >90% at conc. $\leq$ 20 μ M	[103]
2694063-44-8 (19)	0.106 (6.98)	5.76 ^a	-	[49]
392732-12-6 (13)	0.11 (6.96)	0.11 ^a	0.41 ^c	[102]
PET-UNK-29afea89-2	0.129 (6.92)	0.244 ^d	lack of toxicity ^d	[40]
2603242-04-0 (14)	0.128 (6.89)	3.20 ^a	12.3 ^a	[44]
2694063-65-3 (40)	0.171 (6.77)	1.91 ^a	-	[49]

^a In Vero E6 cells. ^b Lacked antiviral activity in an MTT assay. ^c In mammalian cells. ^d In Calu-3 cell line.

Some of the covalent inhibitors of the SARS-CoV-2 M-pro are more potent than the non-covalent inhibitors. Table 4 and Figure 8 show the 15 most potent covalent inhibitors of the SARS-CoV-2 M-pro. Only M-pro inhibitors that have pIC₅₀ values and can inhibit SARS-CoV-2 replication in a cellular antiviral assay were included. Z-AVLD-FMK is a peptidomimetic fluoromethylketone (FMK) inhibitor that was improved from a caspase inhibitor to mimic the natural substrates of the SARS-CoV-2 M-pro [97]. It has the highest inhibitory potency of all SARS-CoV-2 M-pro inhibitors, with an IC₅₀ value of 0.9 nM [97]. It also inhibited the SARS-CoV-2 replication in Vero E6 cells, with an EC₅₀ value of 66 μ M [97]. Although Z-AVLD-FMK showed no toxicity at the concentrations assayed, FMKs may present host cell toxicity [97]. Z-AVLD-FMK could be used as a starting point for developing effective antiviral drugs against SARS-CoV-2 [97]. Compounds 2683066-42-2, 2683066-41-1 and 2683066-47-7 are peptidomimetic with an alpha-acyloxymethylketone warhead and a six-membered lactam glutamine mimic [71]. They are potent SARS-CoV-2 M-pro inhibitors with IC₅₀ values of 1, 6.4 and 14 nM, respectively [71]. They inhibited SARS-CoV-2 replication in Vero E6 cells and showed low cytotoxicity and good plasma and glutathione stability [71]. Compound 2683066-42-2 also displayed selectivity for SARS-CoV-2 M-pro over several cathepsins [71,93]. More advanced compounds based on alpha-acyloxymethylketone should improve metabolic stability [71]. PF-00835231 is a peptidomimetic compound with a hydroxymethylketone warhead that was a development candidate for SARS-CoV-1, but the end of the SARS-CoV-1 outbreak suspended its development [104]. It is one of the most potent SARS-CoV-2 M-pro inhibitors with an IC₅₀ value between 5.8 and 8 nM [104–106] and shows high selectivity over human proteases [93]. It exhibited potent activity against SARS-CoV-2 and other coronaviruses as a single agent [93,104]. Furthermore, PF-00835231 has synergistic activity in combination with remdesivir [93]. Pfizer has started a clinical trial of PF-07304814 (also known as lufotrelvir), a prodrug that is metabolized to PF-00835231 [93]. However, PF-07304814 must be administered by intravenous infusion, and other orally active candidates, such as PF-07321332, are better for clinical development [38]. PF-07321332 has recently been described as a SARS-CoV-2 M-pro inhibitor with in vitro pan-human coronavirus antiviral activity, excellent off-target selectivity and in vivo safety profiles [107]. It forms a reversible covalent bond with the M-pro Cys145 through a nitrile substituent [107]. Furthermore, it

has demonstrated oral activity in a mouse-adapted SARS-CoV-2 model and has achieved oral plasma concentrations exceeding the in vitro antiviral cell potency in a phase I clinical trial in healthy human participants [107]. Currently, PF-07321332 is in phase 3 trials, and it could be the first approved M-pro inhibitor to be used to treat SARS-CoV-2. MI-21, MI-23, MI-28, MI-13, MI-14, MI-05, MI-11, MI-06 and MI-09 are derivatives of boceprevir or telaprevir with IC_{50} values against the SARS-CoV-2 M-pro ranging between 7.6 and 15.2 nM [59] (Table 4). All of these compounds are aldehyde-based inhibitors and showed anti-SARS-CoV-2 activity in Vero E6 cells with EC_{50} values ranging from 0.66 to 5.63 μ M [59]. MI-30 showed an even better EC_{50} value of 0.54 nM, although its IC_{50} value (17.2 nM) is slightly higher than the previously described boceprevir or telaprevir derivatives [59]. The two compounds that showed the best anti-SARS-CoV-2 activity, MI-09 and MI-30, also improved SARS-CoV-2 induced lung lesions in a transgenic mouse model of SARS-CoV-2 infection and displayed good pharmacokinetic properties in rats [59]. UAWJ248 is a GC376 analog with an alpha-ketoamide warhead that binds irreversibly to the Cys145 from SARS-CoV-2 M-pro [70]. This compound was designed on the basis of the x-ray crystal structure of SARS-CoV-2 Mpro with GC-376 (with PDB code 6WTT), to satisfy the side-chain preferences of S1', S2, S3, and S4 M-pro pockets [70]. UAWJ248 inhibited the SARS-CoV-2 M-pro with an IC_{50} value of 12 nM [108]. However, its anti-SARS-CoV-2 potency was lower, showing an EC_{50} of 20.49 μ M against infected Vero E6 cells (Table 4) [70].

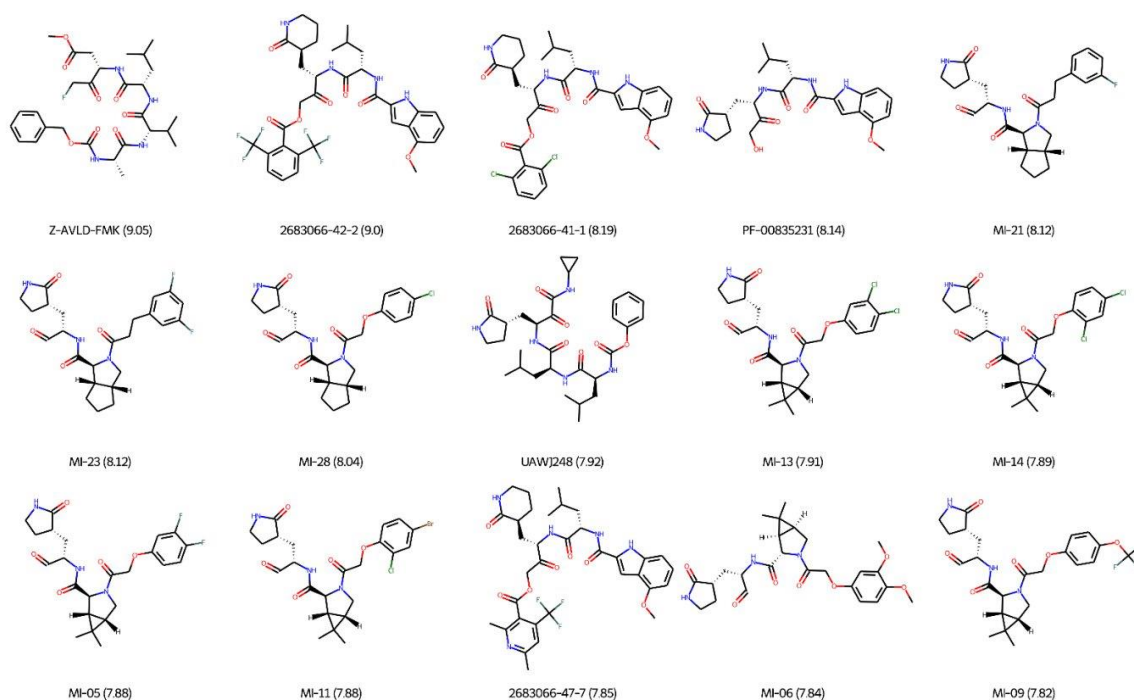


Figure 8. The 15 most potent covalent inhibitors of the SARS-CoV-2 M-pro. Only M-pro inhibitors with pIC_{50} values that can inhibit the SARS-CoV-2 replication in a cellular antiviral assay are shown. pIC_{50} values are shown in parentheses.

Table 4. Data from the 15 most potent covalent inhibitors of the SARS-CoV-2 M-pro.

Compound	IC ₅₀ μ M (pIC ₅₀)	EC ₅₀ μ M	CC ₅₀ μ M	Reference
Z-AVLD-FMK	0.0009 (9.05)	66.0 ^a	-	[97]
2683066-42-2 (15h)	0.001 (9.0)	0.16 ^a	>200	[71]
2683066-41-1 (15g)	0.0064 (8.19)	0.52 ^a	>200	[71]
PF-00835231 (4)	0.0057–0.008 0.0072 ^b (8.14)	88.9 ^a	>100 ^a	[93,104–106]
MI-21	0.0076 (8.12)	2.97 ^a	>500 ^a	[59]
MI-23	0.0076 (8.12)	5.63 ^a	>500 ^a	[59]
MI-28	0.0092 (8.04)	0.67 ^a	>500 ^a	[59]
UAWJ248	0.012 (7.92)	20.49 ^a	>250 ^a	[70]
MI-13	0.0124 (7.91)	2.08	>500 ^a	[59]
MI-14	0.0130 (7.89)	0.66 ^a	>500 ^a	[59]
MI-05	0.0132 (7.88)	5.57 ^a	>500 ^a	[59]
MI-11	0.0133 (7.88)	1.18 ^a	>500 ^a	[59]
2683066-47-7 (15m)	0.014 (7.85)	0.47 ^a	>200	[71]
MI-06	0.0145 (7.84)	1.73 ^a	>500 ^a	[59]
MI-09	0.0152 (7.82)	0.86 ^a	>500 ^a	[59]

^a In Vero E6 cells. ^b Mean value of multiple IC₅₀.

3. Conclusions

Since the beginning of the COVID-19 pandemic, the scientific community has tried to find a drug to inhibit the SARS-CoV-2 life cycle. The SARS-CoV-2 M-pro enzyme has been extensively studied, and its inhibitors are promising effective drugs for fighting against SARS-CoV-2. The first attempts to discover SARS-CoV-2 M-pro inhibitors used previously developed protease inhibitors or tried to repurpose drugs from other diseases. Neither attempt was entirely effective, due to, among other things, the flexibility of the SARS-CoV-2 M-pro and the inability of protein-docking methods to predict the binding modes and potency of SARS-CoV-2 M-pro inhibitors. Several *in vitro* and *in cellulo* (using live cells) methods have been developed to measure the inhibitory potency of a compound against the SARS-CoV-2 M-pro. *In vitro* methods need to express and purify SARS-CoV-2 M-pro, so some tags are sometimes added. However, especially if they are located at the N-terminus, these tags can interfere with the binding of M-pro to its ligands. The activity values obtained by different laboratories or with different methods or conditions must be compared with great care. The presence of DTT has been reported to affect the inhibitory activity of covalent M-pro inhibitors. If the inhibitory effect of an M-pro inhibitor is eliminated or greatly reduced by the presence of DTT, the inhibition is not specific. Therefore, the potency of inhibition measured in the absence of DTT should not be used by itself. The potency of a compound to inhibit SARS-CoV-2 replication in cells cannot always be inferred from the potency to inhibit M-pro, determined *in vitro*. An antiviral assay that uses cells infected with SARS-CoV-2 provides a better estimate of the potency of a compound to inhibit virus replication. However, if it is to be ruled out that the toxicity of the compounds is responsible for the antiviral activity, the cytotoxicity of the compounds needs to be determined.

In this review, we collected 1765 SARS-CoV-2 M-pro inhibitors. This set of compounds could be useful to validate a virtual screening procedure. The search for common covalent warheads identifies putative covalent inhibitors. Although we have not yet hit the bullseye and no drug has yet been approved to inhibit M-pro, we may be close. Improving derivatives of a leading compound has proven to be a very successful strategy for finding potent SARS-CoV-2 M-pro inhibitors. Some derivative compounds designed in less than two years since the start of the COVID-19 pandemic represent an important step toward the

development of new anti-SARS-CoV-2 drugs. Currently, there are several compounds with low nanomolar IC₅₀ values against SARS-CoV-2 M-pro and high anti-SARS-CoV-2 efficacy in cell models, with values comparable to those of the FDA-approved RNA polymerase inhibitor remdesivir. We hope that a SARS-CoV-2 M-pro inhibitor will be approved soon, so we can add a new tool to fight against SARS-CoV-2 or future coronavirus pandemics.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms23010259/s1>.

Author Contributions: G.P. and S.G.-V. conceived and designed the review. P.G.-S. and J.M.-T. collected the data. G.M. and B.S.-E. prepared the figures. G.M. and S.G.-V. wrote the first draft. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 713679 and from the Universitat Rovira i Virgili (grant 2019PFR-URV-B2-69).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated and analyzed during this study are included in this published article and its supplementary information files or are available from the corresponding author on reasonable request.

Acknowledgments: We thank ChemAxon for kindly providing us with an academic license to use their programs.

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

References

1. Ziebuhr, J.; Herold, J.; Siddell, S.G. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J. Virol.* **1995**, *69*, 4331–4338. [\[CrossRef\]](#)
2. Gimeno, A.; Mestres-Truyol, J.; Ojeda-Montes, M.J.; Macip, G.; Saldivar-Espinoza, B.; Cereto-Massagué, A.; Pujadas, G.; Garcia-Vallvé, S. Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition. *Int. J. Mol. Sci.* **2020**, *21*, 3793. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J.R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763–1767. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of M^{Pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [\[CrossRef\]](#)
5. Kneller, D.W.; Phillips, G.; O'Neill, H.M.; Jedrzejczak, R.; Stols, L.; Langan, P.; Joachimiak, A.; Coates, L.; Kovalevsky, A. Structural plasticity of SARS-CoV-2 3CL M^{Pro} active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020**, *11*, 3202. [\[CrossRef\]](#)
6. El-Baba, T.J.; Lutomski, C.A.; Kantsadi, A.L.; Malla, T.R.; John, T.; Mikhailov, V.; Bolla, J.R.; Schofield, C.J.; Zitzmann, N.; Vakonakis, I.; et al. Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 23544–23548. [\[CrossRef\]](#)
7. Günther, S.; Reinke, P.Y.A.; Fernández-García, Y.; Lieske, J.; Lane, T.J.; Ginn, H.M.; Koua, F.H.M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; et al. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372*, 642–646. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Du, R.; Cooper, L.; Chen, Z.; Lee, H.; Rong, L.; Cui, Q. Discovery of chebulagic acid and punicalagin as novel allosteric inhibitors of SARS-CoV-2 3CLpro. *Antivir. Res.* **2021**, *190*, 105075. [\[CrossRef\]](#)
9. Eberle, R.J.; Olivier, D.S.; Amaral, M.S.; Gering, I.; Willbold, D.; Arni, R.K.; Coronado, M.A. The Repurposed Drugs Suramin and Quinacrine Cooperatively Inhibit SARS-CoV-2 3CLpro In Vitro. *Viruses* **2021**, *13*, 873. [\[CrossRef\]](#)
10. Nguyen, D.D.; Gao, K.; Chen, J.; Wang, R.; Wei, G.-W. Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning. *Chem. Sci.* **2020**, *11*, 12036–12046. [\[CrossRef\]](#)
11. Liu, Z.; Fang, H.; Reagan, K.; Xu, X.; Mendrick, D.L.; Slikker, W.; Tong, W. In silico drug repositioning: What we need to know. *Drug Discov. Today* **2013**, *18*, 110–115. [\[CrossRef\]](#)
12. Chakraborti, S.; Bheemireddy, S.; Srinivasan, N. Repurposing drugs against the main protease of SARS-CoV-2: Mechanism-based insights supported by available laboratory and clinical data. *Mol. Omics* **2020**, *16*, 474–491. [\[CrossRef\]](#)
13. Wang, X.; Guan, Y. COVID-19 drug repurposing: A review of computational screening methods, clinical trials, and protein interaction assays. *Med. Res. Rev.* **2021**, *41*, 5–28. [\[CrossRef\]](#) [\[PubMed\]](#)

14. Gimeno, A.; Ojeda-Montes, M.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int. J. Mol. Sci.* **2019**, *20*, 1375. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Warren, G.L.; Andrews, C.W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Senger, S.; et al. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931. [\[CrossRef\]](#)
16. Painsar, T.; Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **2018**, *23*, 1899. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Bzówka, M.; Mitusińska, K.; Raczynska, A.; Samol, A.; Tuszyński, J.A.; Góra, A. Structural and Evolutionary Analysis Indicate That the SARS-CoV-2 M^{Pro} Is a Challenging Target for Small-Molecule Inhibitor Design. *Int. J. Mol. Sci.* **2020**, *21*, 3099. [\[CrossRef\]](#)
18. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [\[CrossRef\]](#)
19. Macip, G.; Garcia-Segura, P.; Mestres-Truyol, J.; Saldivar-Espinoza, B.; Ojeda-Montes, M.J.; Gimeno, A.; Cereto-Massagué, A.; Garcia-Vallvé, S.; Pujadas, G. Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Med. Res. Rev.* **2021**. [\[CrossRef\]](#)
20. Zev, S.; Raz, K.; Schwartz, R.; Tarabeh, R.; Gupta, P.K.; Major, D.T. Benchmarking the Ability of Common Docking Programs to Correctly Reproduce and Score Binding Modes in SARS-CoV-2 Protease M^{Pro}. *J. Chem. Inf. Model.* **2021**, *61*, 2957–2966. [\[CrossRef\]](#)
21. Dotolo, S.; Marabotti, A.; Facchiano, A.; Tagliaferri, R. A review on drug repurposing applicable to COVID-19. *Brief. Bioinform.* **2021**, *22*, 726–741. [\[CrossRef\]](#)
22. Bellera, C.L.; Llanos, M.; Gantner, M.E.; Rodriguez, S.; Gavernet, L.; Comini, M.; Talevi, A. Can drug repurposing strategies be the solution to the COVID-19 crisis? *Expert Opin. Drug Discov.* **2021**, *16*, 605–612. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Llanos, M.A.; Gantner, M.E.; Rodriguez, S.; Alberca, L.N.; Bellera, C.L.; Talevi, A.; Gavernet, L. Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: The Main Protease (M^{Pro}) Case Study. *J. Chem. Inf. Model.* **2021**, *61*, 3758–3770. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Amin, S.A.; Banerjee, S.; Gayen, S.; Jha, T. Protease targeted COVID-19 drug discovery: What we have learned from the past SARS-CoV inhibitors? *Eur. J. Med. Chem.* **2021**, *215*, 113294. [\[CrossRef\]](#)
25. Yang, H.; Yang, J. A review of the latest research on M^{Pro} targeting SARS-COV inhibitors. *RSC Med. Chem.* **2021**, *12*, 1026–1036. [\[CrossRef\]](#)
26. Cui, W.; Yang, K.; Yang, H. Recent Progress in the Drug Development Targeting SARS-CoV-2 Main Protease as Treatment for COVID-19. *Front. Mol. Biosci.* **2020**, *7*, 616341. [\[CrossRef\]](#)
27. Pillaiyar, T.; Wendt, L.L.; Manickam, M.; Easwaran, M. The recent outbreaks of human coronaviruses: A medicinal chemistry perspective. *Med. Res. Rev.* **2021**, *41*, 72–135. [\[CrossRef\]](#)
28. Banerjee, R.; Perera, L.; Tillekeratne, L.M.V. Potential SARS-CoV-2 main protease inhibitors. *Drug Discov. Today* **2021**, *26*, 804–816. [\[CrossRef\]](#) [\[PubMed\]](#)
29. 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. [\[CrossRef\]](#)
30. Gao, K.; Wang, R.; Chen, J.; Tepe, J.J.; Huang, F.; Wei, G.-W. Perspectives on SARS-CoV-2 Main Protease Inhibitors. *J. Med. Chem.* **2021**, *64*, 16922–16955. [\[CrossRef\]](#)
31. Xu, J.; Xue, Y.; Zhou, R.; Shi, P.; Li, H.; Zhou, J. Drug repurposing approach to combating coronavirus: Potential drugs and drug targets. *Med. Res. Rev.* **2021**, *41*, 1375–1426. [\[CrossRef\]](#)
32. Cannalire, R.; Cerchia, C.; Beccari, A.R.; Di Leva, F.S.; Summa, V. Targeting SARS-CoV-2 Proteases and Polymerase for COVID-19 Treatment: State of the Art and Future Opportunities. *J. Med. Chem.* **2020**. [\[CrossRef\]](#)
33. Chen, C.; Yu, X.; Kuo, C.; Min, J.; Chen, S.; Ma, L.; Liu, K.; Guo, R. Overview of antiviral drug candidates targeting coronaviral 3C-like main proteases. *FEBS J.* **2021**, *288*, 5089–5121. [\[CrossRef\]](#)
34. Xiong, M.; Su, H.; Zhao, W.; Xie, H.; Shao, Q.; Xu, Y. What coronavirus 3C-like protease tells us: From structure, substrate selectivity, to inhibitor design. *Med. Res. Rev.* **2021**, *41*, 1965–1998. [\[CrossRef\]](#)
35. Mengist, H.M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021**, *9*, 622898. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Citarella, A.; Scala, A.; Piperno, A.; Micale, N. SARS-CoV-2 M^{Pro}: A Potential Target for Peptidomimetics and Small-Molecule Inhibitors. *Biomolecules* **2021**, *11*, 607. [\[CrossRef\]](#)
37. Chia, C.S.B.; Xu, W.; Shuyi Ng, P. A Patent Review on SARS Coronavirus Main Protease (3CL pro) Inhibitors. *ChemMedChem* **2021**. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Vandyck, K.; Deval, J. Considerations for the discovery and development of 3-chymotrypsin-like cysteine protease inhibitors targeting SARS-CoV-2 infection. *Curr. Opin. Virol.* **2021**, *49*, 36–40. [\[CrossRef\]](#) [\[PubMed\]](#)
39. COVID Moonshot. Available online: <https://covid.postera.ai/covid> (accessed on 1 October 2021).
40. The COVID Moonshot Consortium; Chodera, J.; Lee, A.; London, N.; von Delft, F. Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitors. *ChemRxiv* **2021**. [\[CrossRef\]](#)
41. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [\[CrossRef\]](#)
42. Singh, J.; Petter, R.C.; Baillie, T.A.; Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317. [\[CrossRef\]](#)

43. Ghahremanpour, M.M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J.A.; Zhang, C.-H.; Cabeza de Vaca, I.; Liosi, M.-E.; Anderson, K.S.; Jorgensen, W.L. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **2020**, *11*, 2526–2533. [\[CrossRef\]](#)
44. Zhang, C.-H.; Stone, E.A.; Deshmukh, M.; Ippolito, J.A.; Ghahremanpour, M.M.; Tirado-Rives, J.; Spasov, K.A.; Zhang, S.; Takeo, Y.; Kudalkar, S.N.; et al. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021**, *7*, 467–475. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Deshmukh, M.G.; Ippolito, J.A.; Zhang, C.-H.; Stone, E.A.; Reilly, R.A.; Miller, S.J.; Jorgensen, W.L.; Anderson, K.S. Structure-guided design of a perampanel-derived pharmacophore targeting the SARS-CoV-2 main protease. *Structure* **2021**, *29*, 823–833.e5. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Zhang, C.-H.; Spasov, K.A.; Reilly, R.A.; Hollander, K.; Stone, E.A.; Ippolito, J.A.; Liosi, M.-E.; Deshmukh, M.G.; Tirado-Rives, J.; Zhang, S.; et al. Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med. Chem. Lett.* **2021**, *12*, 1325–1332. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Jacobs, J.; Grum-Tokars, V.; Zhou, Y.; Turlington, M.; Saldanha, S.A.; Chase, P.; Eggler, A.; Dawson, E.S.; Baez-Santos, Y.M.; Tomar, S.; et al. Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL pro. *J. Med. Chem.* **2013**, *56*, 534–546. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Turlington, M.; Chun, A.; Tomar, S.; Eggler, A.; Grum-Tokars, V.; Jacobs, J.; Daniels, J.S.; Dawson, E.; Saldanha, A.; Chase, P.; et al. Discovery of N-(benzo[1,2,3]triazol-1-yl)-N-(benzyl)acetamido)phenyl) carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: Identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 6172–6177. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Han, S.H.; Goins, C.M.; Arya, T.; Shin, W.-J.; Maw, J.; Hooper, A.; Sonawane, D.P.; Porter, M.R.; Bannister, B.E.; Crouch, R.D.; et al. Structure-Based Optimization of ML300-Derived, Noncovalent Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus 3CL Protease (SARS-CoV-2 3CL pro). *J. Med. Chem.* **2021**. [\[CrossRef\]](#)
50. Kitamura, N.; Sacco, M.D.; Ma, C.; Hu, Y.; Townsend, J.A.; Meng, X.; Zhang, F.; Zhang, X.; Ba, M.; Szeto, T.; et al. Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *J. Med. Chem.* **2021**. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Ma, C.; Sacco, M.D.; Hurst, B.; Townsend, J.A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M.T.; Chen, Y.; et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020**, *30*, 678–692. [\[CrossRef\]](#)
52. Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q.; Wu, Y.; Zhao, C.; Sun, H.; Huang, B.; Niu, P.; et al. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020**, *11*, 4417. [\[CrossRef\]](#)
53. Kneller, D.W.; Galanie, S.; Phillips, G.; O'Neill, H.M.; Coates, L.; Kovalevsky, A. Malleability of the SARS-CoV-2 3CL M^{pro} Active-Site Cavity Facilitates Binding of Clinical Antivirals. *Structure* **2020**, *28*, 1313–1320.e3. [\[CrossRef\]](#)
54. Pathak, N.; Chen, Y.-T.; Hsu, Y.-C.; Hsu, N.-Y.; Kuo, C.-J.; Tsai, H.P.; Kang, J.-J.; Huang, C.-H.; Chang, S.-Y.; Chang, Y.-H.; et al. Uncovering Flexible Active Site Conformations of SARS-CoV-2 3CL Proteases through Protease Pharmacophore Clusters and COVID-19 Drug Repurposing. *ACS Nano* **2021**, *15*, 857–872. [\[CrossRef\]](#)
55. Jan, J.-T.; Cheng, T.-J.R.; Juang, Y.-P.; Ma, H.-H.; Wu, Y.-T.; Yang, W.-B.; Cheng, C.-W.; Chen, X.; Chou, T.-H.; Shie, J.-J.; et al. Identification of existing pharmaceuticals and herbal medicines as inhibitors of SARS-CoV-2 infection. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021579118. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Mody, V.; Ho, J.; Wills, S.; Mawri, A.; Lawson, L.; Ebert, M.C.C.J.C.; Fortin, G.M.; Rayalam, S.; Taval, S. Identification of 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents. *Commun. Biol.* **2021**, *4*, 93. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Baker, J.D.; Uhrich, R.L.; Kraemer, G.C.; Love, J.E.; Kraemer, B.C. A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. *PLoS ONE* **2021**, *16*, e0245962. [\[CrossRef\]](#)
58. Manandhar, A.; Blass, B.E.; Colussi, D.J.; Almi, I.; Abou-Gharbia, M.; Klein, M.L.; Elokely, K.M. Targeting SARS-CoV-2 M3CLpro by HCV NS3/4a Inhibitors: In Silico Modeling and In Vitro Screening. *J. Chem. Inf. Model.* **2021**, *61*, 1020–1032. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Qiao, J.; Li, Y.-S.; Zeng, R.; Liu, F.-L.; Luo, R.-H.; Huang, C.; Wang, Y.-F.; Zhang, J.; Quan, B.; Shen, C.; et al. SARS-CoV-2 M^{pro} inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371*, 1374–1378. [\[CrossRef\]](#)
60. Kim, Y.; Shivanna, V.; Narayanan, S.; Prior, A.M.; Weerasekara, S.; Hua, D.H.; Kankanamalage, A.C.G.; Groutas, W.C.; Chang, K.-O. Broad-spectrum inhibitors against 3C-like proteases of feline coronaviruses and feline caliciviruses. *J. Virol.* **2015**, *89*, 4942–4950. [\[CrossRef\]](#)
61. Vuong, W.; Khan, M.B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H.A.; McKay, R.T.; van Belkum, M.J.; Joyce, M.A.; et al. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **2020**, *11*, 4282. [\[CrossRef\]](#)
62. Hung, H.; Ke, Y.; Huang, S.Y.; Huang, P.-N.; Kung, Y.; Chang, T.-Y.; Yen, K.; Peng, T.-T.; Chang, S.-E.; Huang, C.-T.; et al. Discovery of M Protease Inhibitors Encoded by SARS-CoV-2. *Antimicrob. Agents Chemother.* **2020**, *64*, e00872-20. [\[CrossRef\]](#)

63. Rathnayake, A.D.; Zheng, J.; Kim, Y.; Perera, K.D.; Mackin, S.; Meyerholz, D.K.; Kashipathy, M.M.; Battaile, K.P.; Lovell, S.; Perlman, S.; et al. 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV-infected mice. *Sci. Transl. Med.* **2020**, *12*, eabc5332. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Gurard-Levin, Z.A.; Liu, C.; Jekle, A.; Jaisinghani, R.; Ren, S.; Vandyck, K.; Jochmans, D.; Leyssen, P.; Neyts, J.; Blatt, L.M.; et al. Evaluation of SARS-CoV-2 3C-like protease inhibitors using self-assembled monolayer desorption ionization mass spectrometry. *Antivir. Res.* **2020**, *182*, 104924. [\[CrossRef\]](#)
65. Wang, Y.-C.; Yang, W.-H.; Yang, C.-S.; Hou, M.-H.; Tsai, C.-L.; Chou, Y.-Z.; Hung, M.-C.; Chen, Y. Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum anti-coronaviral drug. *Am. J. Cancer Res.* **2020**, *10*, 2535–2545.
66. Zhu, W.; Xu, M.; Chen, C.Z.; Guo, H.; Shen, M.; Hu, X.; Shinn, P.; Klumpp-Thomas, C.; Michael, S.G.; Zheng, W. Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative High-Throughput Screening. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 1008–1016. [\[CrossRef\]](#)
67. Iketani, S.; Forouhar, F.; Liu, H.; Hong, S.J.; Lin, F.-Y.; Nair, M.S.; Zask, A.; Huang, Y.; Xing, L.; Stockwell, B.R.; et al. Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors. *Nat. Commun.* **2021**, *12*, 2016. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Yang, K.S.; Ma, X.R.; Ma, Y.; Alugubelli, Y.R.; Scott, D.A.; Vatansever, E.C.; Drelich, A.K.; Sankaran, B.; Geng, Z.Z.; Blankenship, L.R.; et al. A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors. *ChemMedChem* **2021**, *16*, 942–948. [\[CrossRef\]](#)
69. Vuong, W.; Fischer, C.; Khan, M.B.; van Belkum, M.J.; Lamer, T.; Willoughby, K.D.; Lu, J.; Arutyunova, E.; Joyce, M.A.; Saffran, H.A.; et al. Improved SARS-CoV-2 M^{Pro} inhibitors based on feline antiviral drug GC376: Structural enhancements, increased solubility, and micellar studies. *Eur. J. Med. Chem.* **2021**, *222*, 113584. [\[CrossRef\]](#)
70. Sacco, M.D.; Ma, C.; Lagarias, P.; Gao, A.; Townsend, J.A.; Meng, X.; Dube, P.; Zhang, X.; Hu, Y.; Kitamura, N.; et al. Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M^{Pro} and cathepsin L. *Sci. Adv.* **2020**, *6*, eabe0751. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Bai, B.; Belovodskiy, A.; Hena, M.; Kandadai, A.S.; Joyce, M.A.; Saffran, H.A.; Shields, J.A.; Khan, M.B.; Arutyunova, E.; Lu, J.; et al. Peptidomimetic α -Acylloxymethylketone Warheads with Six-Membered Lactam P1 Glutamine Mimic: SARS-CoV-2 3CL Protease Inhibition, Coronavirus Antiviral Activity, and in Vitro Biological Stability. *J. Med. Chem.* **2021**. [\[CrossRef\]](#)
72. Ma, C.; Hu, Y.; Townsend, J.A.; Lagarias, P.I.; Marty, M.T.; Kolocouris, A.; Wang, J. Ebselen, Disulfiram, Carmofur, PX-12, Tideglusib, and Shikonin Are Nonspecific Promiscuous SARS-CoV-2 Main Protease Inhibitors. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 1265–1277. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Ma, C.; Tan, H.; Choza, J.; Wang, Y.; Wang, J. Validation and invalidation of SARS-CoV-2 main protease inhibitors using the Flip-GFP and Protease-Glo luciferase assays. *Acta Pharm. Sin. B* **2021**. [\[CrossRef\]](#)
74. Sun, L.-Y.; Chen, C.; Su, J.; Li, J.-Q.; Jiang, Z.; Gao, H.; Chigan, J.-Z.; Ding, H.-H.; Zhai, L.; Yang, K.-W. Ebsulfur and Ebselen as highly potent scaffolds for the development of potential SARS-CoV-2 antivirals. *Bioorg. Chem.* **2021**, *112*, 104889. [\[CrossRef\]](#)
75. Ampornnanai, K.; Meng, X.; Shang, W.; Jin, Z.; Rogers, M.; Zhao, Y.; Rao, Z.; Liu, Z.-J.; Yang, H.; Zhang, L.; et al. Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives. *Nat. Commun.* **2021**, *12*, 3061. [\[CrossRef\]](#)
76. Su, H.; Yao, S.; Zhao, W.; Zhang, Y.; Liu, J.; Shao, Q.; Wang, Q.; Li, M.; Xie, H.; Shang, W.; et al. Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease. *Nat. Commun.* **2021**, *12*, 3623. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Kuzikov, M.; Costanzi, E.; Reinshagen, J.; Esposito, F.; Vangeel, L.; Wolf, M.; Ellinger, B.; Claussen, C.; Geisslinger, G.; Corona, A.; et al. Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen. *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 1096–1110. [\[CrossRef\]](#)
78. Liu, H.; Ye, F.; Sun, Q.; Liang, H.; Li, C.; Li, S.; Lu, R.; Huang, B.; Tan, W.; Lai, L. Scutellaria baicalensis extract and baicalin inhibit replication of SARS-CoV-2 and its 3C-like protease in vitro. *J. Enzyme Inhib. Med. Chem.* **2021**, *36*, 497–503. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Nguyen, T.T.H.; Jung, J.-H.; Kim, M.-K.; Lim, S.; Choi, J.-M.; Chung, B.; Kim, D.-W.; Kim, D. The Inhibitory Effects of Plant Derivate Polyphenols on the Main Protease of SARS Coronavirus 2 and Their Structure–Activity Relationship. *Molecules* **2021**, *26*, 1924. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Froggatt, H.M.; Heaton, B.E.; Heaton, N.S. Development of a Fluorescence-Based, High-Throughput SARS-CoV-2 3CLpro Reporter Assay. *J. Virol.* **2020**, *94*, e01265–20. [\[CrossRef\]](#)
81. Franko, N.; Teixeira, A.P.; Xue, S.; Charpin-El Hamri, G.; Fussenegger, M. Design of modular autoproteolytic gene switches responsive to anti-coronavirus drug candidates. *Nat. Commun.* **2021**, *12*, 6786. [\[CrossRef\]](#)
82. Coelho, C.; Gallo, G.; Campos, C.B.; Hardy, L.; Würtele, M. Biochemical screening for SARS-CoV-2 main protease inhibitors. *PLoS ONE* **2020**, *15*, e0240079. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Rawson, J.M.O.; Duchon, A.; Nikolaitchik, O.A.; Pathak, V.K.; Hu, W.-S. Development of a Cell-Based Luciferase Complementation Assay for Identification of SARS-CoV-2 3CLpro Inhibitors. *Viruses* **2021**, *13*, 173. [\[CrossRef\]](#) [\[PubMed\]](#)
84. O'Brien, A.; Chen, D.-Y.; Hackbart, M.; Close, B.J.; O'Brien, T.E.; Saeed, M.; Baker, S.C. Detecting SARS-CoV-2 3CLpro expression and activity using a polyclonal antiserum and a luciferase-based biosensor. *Virology* **2021**, *556*, 73–78. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Dražić, T.; Kühl, N.; Leuthold, M.M.; Behnam, M.A.M.; Klein, C.D. Efficiency Improvements and Discovery of New Substrates for a SARS-CoV-2 Main Protease FRET Assay. *SLAS Discov. Adv. Sci. Drug Discov.* **2021**, *26*, 1189–1199. [\[CrossRef\]](#)
86. Ihssen, J.; Faccio, G.; Yao, C.; Sirec, T.; Spitz, U. Fluorogenic in vitro activity assay for the main protease M^{Pro} from SARS-CoV-2 and its adaptation to the identification of inhibitors. *STAR Protoc.* **2021**, *2*, 100793. [\[CrossRef\]](#)

87. Mathieu, C.; Touret, F.; Jacquemin, C.; Janin, Y.L.; Nougairède, A.; Brailly, M.; Mazelier, M.; Décimo, D.; Vasseur, V.; Hans, A.; et al. A Bioluminescent 3CLPro Activity Assay to Monitor SARS-CoV-2 Replication and Identify Inhibitors. *Viruses* **2021**, *13*, 1814. [\[CrossRef\]](#)
88. Dey-Rao, R.; Smith, G.R.; Timilsina, U.; Falls, Z.; Samudrala, R.; Stavrou, S.; Melendy, T. A fluorescence-based, gain-of-signal, live cell system to evaluate SARS-CoV-2 main protease inhibition. *Antiviral Res.* **2021**, *195*, 105183. [\[CrossRef\]](#)
89. Alalam, H.; Sigurdardóttir, S.; Bourgard, C.; Tiukova, I.; King, R.D.; Grøtli, M.; Sunnerhagen, P. A Genetic Trap in Yeast for Inhibitors of SARS-CoV-2 Main Protease. *mSystems* **2021**, *6*, e0108721. [\[CrossRef\]](#)
90. Li, Z.; Lin, Y.; Huang, Y.-Y.; Liu, R.; Zhan, C.-G.; Wang, X.; Luo, H.-B. Reply to Ma and Wang: Reliability of various in vitro activity assays on SARS-CoV-2 main protease inhibitors. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2024937118. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Ma, C.; Wang, J. Dipyridamole, chloroquine, montelukast sodium, candesartan, oxytetracycline, and atazanavir are not SARS-CoV-2 main protease inhibitors. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2024420118. [\[CrossRef\]](#)
92. Xia, Z.; Sacco, M.; Hu, Y.; Ma, C.; Meng, X.; Zhang, F.; Szeto, T.; Xiang, Y.; Chen, Y.; Wang, J. Rational Design of Hybrid SARS-CoV-2 Main Protease Inhibitors Guided by the Superimposed Cocrystal Structures with the Peptidomimetic Inhibitors GC-376, Telaprevir, and Boceprevir. *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 1408–1421. [\[CrossRef\]](#)
93. Boras, B.; Jones, R.M.; Anson, B.J.; Arenson, D.; Aschenbrenner, L.; Bakowski, M.A.; Beutler, N.; Binder, J.; Chen, E.; Eng, H.; et al. Preclinical characterization of an intravenous coronavirus 3CL protease inhibitor for the potential treatment of COVID19. *Nat. Commun.* **2021**, *12*, 6055. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Gossen, J.; Albani, S.; Hanke, A.; Joseph, B.P.; Bergh, C.; Kuzikov, M.; Costanzi, E.; Manelfi, C.; Storici, P.; Gribbon, P.; et al. A Blueprint for High Affinity SARS-CoV-2 M^{Pro} Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics. *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 1079–1095. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Behnam, M.A.M.; Klein, C.D. Inhibitor potency and assay conditions: A case study on SARS-CoV-2 main protease. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2106095118. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Mahdi, M.; Mótyán, J.A.; Szojka, Z.I.; Golda, M.; Miczi, M.; Tózsér, J. Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2's main protease. *Virol. J.* **2020**, *17*, 190. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Milligan, J.C.; Zeisner, T.U.; Papageorgiou, G.; Joshi, D.; Soudy, C.; Ulferts, R.; Wu, M.; Lim, C.T.; Tan, K.W.; Weissmann, F.; et al. Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp5 main protease. *Biochem. J.* **2021**, *478*, 2499–2515. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Hattori, S.; Higshi-Kuwata, N.; Raghavaiah, J.; Das, D.; Bulut, H.; Davis, D.A.; Takamatsu, Y.; Matsuda, K.; Takamune, N.; Kishimoto, N.; et al. GRL-0920, an Indole Chloropyridinyl Ester, Completely Blocks SARS-CoV-2 Infection. *mBio* **2020**, *11*, e01833-20. [\[CrossRef\]](#)
99. Huff, S.; Kummetha, I.R.; Tiwari, S.K.; Huante, M.B.; Clark, A.E.; Wang, S.; Bray, W.; Smith, D.; Carlin, A.F.; Endsley, M.; et al. Discovery and Mechanism of SARS-CoV-2 Main Protease Inhibitors. *J. Med. Chem.* **2021**. [\[CrossRef\]](#)
100. Dai, W.; Jochmans, D.; Xie, H.; Yang, H.; Li, J.; Su, H.; Chang, D.; Wang, J.; Peng, J.; Zhu, L.; et al. Design, Synthesis, and Biological Evaluation of Peptidomimetic Aldehydes as Broad-Spectrum Inhibitors against Enterovirus and SARS-CoV-2. *J. Med. Chem.* **2021**. [\[CrossRef\]](#)
101. Hamdy, R.; Fayed, B.; Mostafa, A.; Shama, N.M.A.; Mahmoud, S.H.; Mehta, C.H.; Nayak, Y.; Soliman, S.S.M. Iterated Virtual Screening-Assisted Antiviral and Enzyme Inhibition Assays Reveal the Discovery of Novel Promising Anti-SARS-CoV-2 with Dual Activity. *Int. J. Mol. Sci.* **2021**, *22*, 9057. [\[CrossRef\]](#)
102. Elseginy, S.A.; Fayed, B.; Hamdy, R.; Mahrous, N.; Mostafa, A.; Almehtdi, A.M.; Soliman, S.S.M. Promising anti-SARS-CoV-2 drugs by effective dual targeting against the viral and host proteases. *Bioorg. Med. Chem. Lett.* **2021**, *43*, 128099. [\[CrossRef\]](#)
103. Cui, J.; Jia, J. Discovery of juglone and its derivatives as potent SARS-CoV-2 main proteinase inhibitors. *Eur. J. Med. Chem.* **2021**, *225*, 113789. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Hoffman, R.L.; Kania, R.S.; Brothers, M.A.; Davies, J.F.; Ferre, R.A.; Gajiwala, K.S.; He, M.; Hogan, R.J.; Kozminski, K.; Li, L.Y.; et al. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63*, 12725–12747. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Liu, C.; Boland, S.; Scholle, M.D.; Bardiot, D.; Marchand, A.; Chaltin, P.; Blatt, L.M.; Beigelman, L.; Symons, J.A.; Raboisson, P.; et al. Dual inhibition of SARS-CoV-2 and human rhinovirus with protease inhibitors in clinical development. *Antivir. Res.* **2021**, *187*, 105020. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Vankadara, S.; Wong, Y.X.; Liu, B.; See, Y.Y.; Tan, L.H.; Tan, Q.W.; Wang, G.; Karuna, R.; Guo, X.; Tan, S.T.; et al. A head-to-head comparison of the inhibitory activities of 15 peptidomimetic SARS-CoV-2 3CLpro inhibitors. *Bioorg. Med. Chem. Lett.* **2021**, *48*, 128263. [\[CrossRef\]](#)
107. Owen, D.R.; Allerton, C.M.N.; Anderson, A.S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R.D.; Carlo, A.; Coffman, K.J.; et al. An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, *374*, eabl4784. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Sacco, M.D.; Ma, C.; Lagarias, P.; Gao, A.; Townsend, J.A.; Meng, X.; Dube, P.; Zhang, X.; Hu, Y.; Kitamura, N.; et al. Structure and inhibition of the SARS-CoV-2 main protease reveals strategy for developing dual inhibitors against M^{Pro} and cathepsin L. *bioRxiv Prepr. Serv. Biol.* **2020**, *10*, 2020-07. [\[CrossRef\]](#)

Supplementary material for

A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet?

Guillem Macip, Pol Garcia-Segura, Júlia Mestres-Truyol, Bryan Saldivar-Espinoza, Gerard Pujadas* and Santiago Garcia-Vallvé*

Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Campus Sescelades, Universitat Rovira i Virgili, 43007 Tarragona, Catalonia, Spain;
guillem.macip@gmail.com (G.M.); polgarse2@gmail.com (P.G.-S.); juliamt110@gmail.com (J.M.-T.); bsaldivar.emc2@gmail.com (B.S.-E.)

* Correspondence: gerard.pujadas@gmail.com (G.P.); santi.garcia-vallve@urv.cat (S.G.-V.)

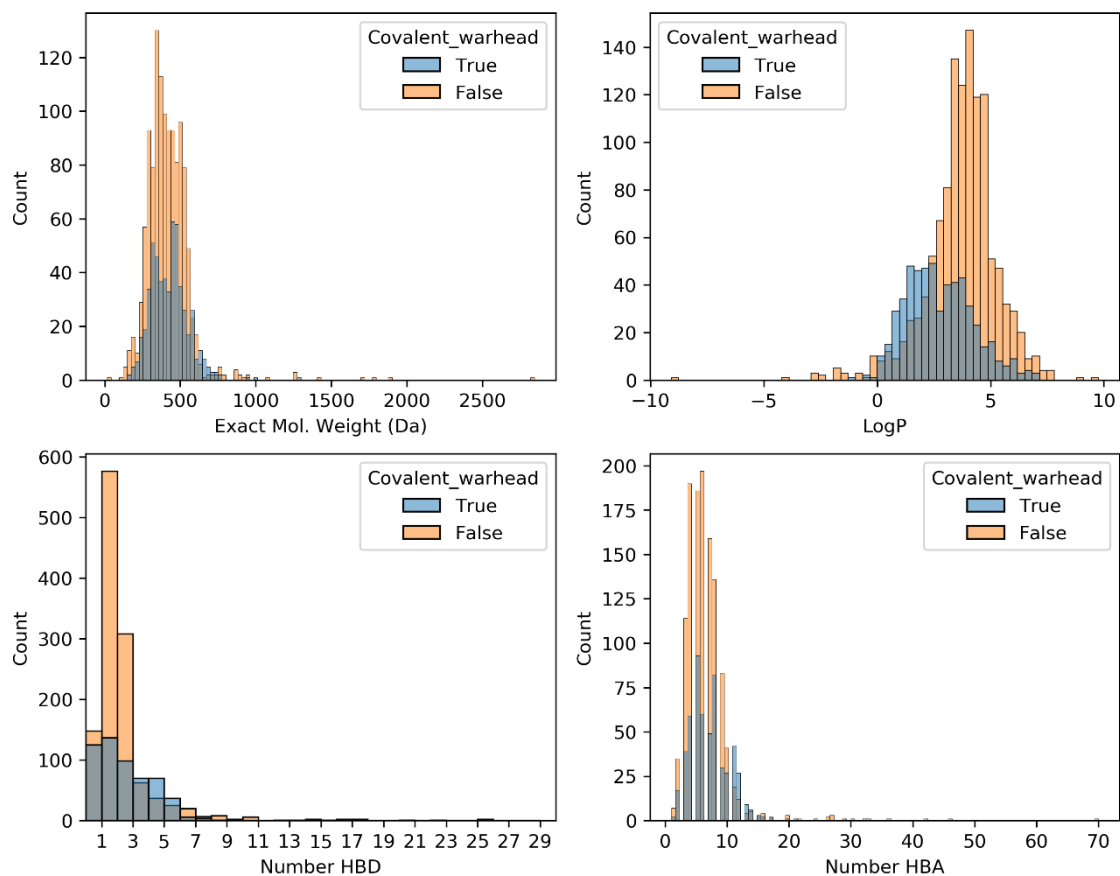
SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Molecular weight, LogP, number of hydrogen bond donors and hydrogen bond acceptors between the 1765 M-pro inhibitors when comparing the covalent and non-covalent groups.

SUPPLEMENTARY TABLES

Table S1. Collection of 1765 SARS-CoV-2 M-pro inhibitors which could be useful to validate a VS. It also contains their International Chemical Identifier (Inchikey) as well as their pIC_{50} (the negative log of the IC_{50} value when converted to molar).

Manuscript 1



Compound	InchiKey	pIC50
1815600-11-3	BBQRBOIMSKMFFO-LTGZKZEYSA-N	4.738
184904-82-3	NTNYHMYSGPGCQ-UHFFFAOYSA-N	7.347
184904-84-5	VDNGDPGHZKPFHH-UHFFFAOYSA-N	7.276
2103278-86-8	HPKJGHVHQWJOOT-ZJOUEHCJSA-N	7.486
2411412-50-3	QTGORSKTHIBLJT-FIVJMQNTSA-N	6.745
2412965-58-1	JYNPCEUBSGNIIR-VJBMBRPKSA-N	5.622
2412965-59-2	NLVRHQFXQFSBQK-XWGVYQGASA-N	6.174
2413716-71-7	HCRBVFBQANEDDM-CUWPLCDZSA-N	7.521
2477631-41-5	JAZGKZKQQSCVMM-ZOBUZTSGSA-N	6.086
2477631-44-8	PHUZMTSQHVZDOX-DZKIICNBSA-N	6.367
2477631-46-0	COXVPYKZDDKVRF-ULQDDVLXSA-N	6.319
2477631-49-3	ORKNIPZHPCWIDI-AOTTWXCTSA-N	6.187
2477631-55-1	MODFCHUDBQIAH-QAYVNIIMSA-N	6.387
2477631-59-5	BHZBRFONZANPNK-ZYHFAYPJSA-N	6.347
2477734-17-9	BLRXSZOWSREMKN-DHRHBVOQSA-N	6.553
2477734-19-1	FOLCUFKJHSQMEL-BIXPGCQOSA-N	6.770
2477734-23-7	VCQYBHSNDWQLAS-RLGDLGOESA-N	6.638
2477734-27-1	WKNCKKOHQQLNCY-ULUMWSQISA-N	6.699
2488719-73-7	SKGYHEXXWSDNGH-UHFFFAOYSA-N	7.328
2488719-74-8	MIUFWBYRBXTIDZ-UHFFFAOYSA-N	4.604
2488719-75-9	RCSXHMPDXUMRL-UHFFFAOYSA-N	4.407
674786-33-5	GIPLQLQEWXCSIX-UHFFFAOYSA-N	4.975
69471-29-0	VYIBOCHDJVESIU-UHFFFAOYSA-N	5.046
898264-59-0	PWBQMVUCCVWDEJ-UCWBBBRQSA-N	6.821
898542-39-7	DKYBERVCUALHHG-UHFFFAOYSA-N	4.379
898542-48-8	UPFRLUKDHMINPH-UHFFFAOYSA-N	4.991
898542-49-9	HRVIHQDQVDQVYBE-UHFFFAOYSA-N	4.750
898542-52-4	MIUBNBDMUMQCLE-UHFFFAOYSA-N	4.810
Agaric acid	HZLCGUXUOFWCCN-UHFFFAOYSA-N	4.628
Andro-NBD	HQRLZOFUJXQDOA-TXNPGWHDSA-N	5.554
Andrographolide	BOJKULTULYSRAS-OTESTREVSA-N	4.822
Atazanavir	AXRYRYVKAWYZBR-GASGPIRDSA-N	4.467
Baicalein	FXNFHKRTJBSTCS-UHFFFAOYSA-N	4.156
Baicalin	DAIGSUHCWOZEQG-ZFORQUDYSA-N	4.348
Bedaquiline	QUIJNHUBAXPXS-XLJNKUFUSA-N	4.728
Bendamustine	YTKUWDBFDASYHO-UHFFFAOYSA-N	4.585
Bepridil	UIEATEWHFDRYRU-UHFFFAOYSA-N	4.143
Beta-Lapachone	QZPQTZZNNJUOLS-UHFFFAOYSA-N	4.875
Boceprevir	LHHCSNFAOIFYRV-DOVBMPENSA-N	5.027
Bronopol	LVDKZNITIUWNER-UHFFFAOYSA-N	5.618
Calpain inhibitor I	FMYKJLXRRQTBOR-BZSNMDCSA-N	5.162
Calpain inhibitor II	RJWLAIMXRBUMH-ULQDDVLXSA-N	5.448
Calpain inhibitor XII	PLVWMBFPIAQRHK-FGZHOGPDSA-N	5.613
Calpeptin	PGGUOGKHUUUWAF-ROUUACIISA-N	5.186
Candesartan Cilextil	GHOSNRCGJFBJIB-UHFFFAOYSA-N	4.455
Carmofur	AOCCBINRVIKJHY-UHFFFAOYSA-N	5.094
CAY-10581	WKHJQIQDQXGUOD-UWJYYQICSA-N	5.036
Chloranil	UGNWTBMOAKPKBL-UHFFFAOYSA-N	5.686
Chloroquine	WHTVZRBWZFKQO-UHFFFAOYSA-N	5.409
Cinanserin	RSUVYMGADVXGOU-BUHFOSPRSA-N	3.903
Clemastine	YNNUSGIPVFPVBX-NHCUHLMSSA-N	3.304
Cyanidin-3-O-glucoside	RKWHWFONKJEUFE-GQUPQBGVSA-O	4.186
DA-3003-1	BMKPDVDQDJQWBPD-UHFFFAOYSA-N	5.580

Manuscript 1

Compound	InchiKey	pIC50
Darunavir	CJBHJOAVZSMMDJ-HEXNFIEUSA-N	4.442
Dipyridamole	IZEKFCXSFNUWAM-UHFFFAOYSA-N	6.222
Disulfiram	AUZONCFQVSMFAP-UHFFFAOYSA-N	5.368
Duloxetine	ZEUITGRIYCTCEM-KRWDZBQOSA-N	2.516
Ebastine	MJJALKDDGIKVBE-UHFFFAOYSA-N	4.244
Ebselen	DYEFUKCXAQOFHX-UHFFFAOYSA-N	5.323
Efonidipine	NSVFSAJIGAJDMR-UHFFFAOYSA-N	4.415
Emedastine	KBUZBQVCBVDWKX-UHFFFAOYSA-N	4.086
Epigallocatechin gallate	WMBWREPUVVBILR-WIYYLYMNSA-N	4.411
Evans blue	LIRUWPZKVZZTHC-UHFFFAOYSA-N	5.335
Fascaplysin	WYQIPCUPNMRAP-UHFFFAOYSA-O	5.002
GC373	QYENXTYKACLCGO-FQECFTEESA-N	6.656
GC376	BSPZFJDYQHDZNR-HTCLRFROSA-N	6.641
GW-0742	HWVNEWGKWGRSRK-UHFFFAOYSA-N	4.652
Hematoporphyrin	DWZDGMDDUHAQA-UHFFFAOYSA-N	5.409
Hydroxocobalamin	BAOLGUBNOUWZAV-RQJMMSBLSA-N	5.483
Hydroxychloroquine	XXSMGPRMXLTPCZ-UHFFFAOYSA-N	5.538
Hypericin	BTXNYTINYBABQR-UHFFFAOYSA-N	4.197
Indinavir	CBVCZFGXHXORBI-PXQQMZJSSA-N	4.366
Isavuconazole	DDFOUSQFMRYUQK-RCDICMHDSA-N	3.359
Itraconazole	VHVPQPYKVGDNFY-UHFFFAOYSA-N	3.955
Lercanidipine	ZDXUKAKRHYTAKV-UHFFFAOYSA-N	4.790
Leupeptin	GDBQQVLCIARPGH-ULQDDVLXSA-N	4.036
LLL-12	CQBHSRLUQDYPBU-UHFFFAOYSA-N	5.007
Lopinavir	KJHKTHWMRKYKJE-SUGCFTRWSA-N	3.444
MAC-5576	PAVJYVZIJTAHT-UHFFFAOYSA-N	7.092
Manidipine	ANEBWFXPVPTET-UHFFFAOYSA-N	5.119
Masitinib	WJEOLQLKVOPQFV-UHFFFAOYSA-N	5.602
Metixene	MJFJKKXQDNNUJF-UHFFFAOYSA-N	3.197
MG-115	QEJRGURBLQWEOU-FKBYEOEOSA-N	5.101
MG-132	TZYWCYJVHRLUCT-VABKMULXSA-N	5.296
MK 0893	DNTVJEMGHBIUMW-IBGZPJMESA-N	4.614
Montelukast	UCHDWCPVSPXUMX-TZIWL TJVSA-N	5.137
1065478-52-5	HXAHMXYAYHWWRI-ZCTWNQIISA-N	6.987
890648-84-7	RUHYGYHBVKWREW-CMOCDZPBSA-N	8.071
2481133-53-1	WOAWOVDMMGJUQG-QORCZRPOSA-N	7.824
2481133-56-4	HWYRKZDETMXFE-QORCZRPOSA-N	7.481
856242-41-6	GTBAQTKMASRCMG-FFYZIMEISA-N	7.222
2481133-61-1	DFRWJNPTXOHWGQ-FFXRMZKPSA-N	7.328
856242-63-2	YFUZVOLEKQFVCQ-FFXRMZKPSA-N	7.244
2481133-66-6	BLKJSUVYKJMKNN-ZHRPPACSA-N	7.252
N3	IDBWWEGDLCFCTD-VNEMRZQUSA-N	7.523
Narlaprevir	RICZEKVVNZFTNZ-LFGITCQGSA-N	5.166
Nelfinavir	QAGYKUNXZXKMR-HKWSIXNMSA-N	3.754
NSC 95397	MAASHDQFQDDECQ-UHFFFAOYSA-N	4.746
Omeprazole	SUBDBMMJDZJVOS-UHFFFAOYSA-N	3.818
Oritavancin diphosphate	VHFGEVPHAGQPI-MYYQHNLBSA-N	4.617
Oxiconazole	QRJJEGAJXVEBNE-HKOYGPOVSA-N	4.004
Oxytetracycline	LUIPXMIQGDWPOS-NUORWIKGSA-N	4.818
Penta-o-galloyl-beta-D-glucose hydrate	QJYNZEYHSMRWBK-NIKIMHBISA-N	4.804
PF-00835231	QDIMHKWNHMDVJB-WBAXXEDZSA-N	8.145
Phenylmercuric acetate	XEBWQGVWTUSTLN-UHFFFAOYSA-M	6.688
Pimozide	YVUQSNJEYSNKRX-UHFFFAOYSA-N	4.377

Compound	InchiKey	pIC50
Protoporphyrin IX	HGXFBVDWWKPUKZ-UHFFFAOYSA-N	4.638
PSI	TYFTWYMXUWCOOB-YICLCXKGSA-N	4.984
PX-12	BPBPYQWMFCTCNG-UHFFFAOYSA-N	5.126
Rimonabant	JZCPYUJPEARBJL-UHFFFAOYSA-N	4.071
Ritonavir	NCDNCNXCDXHOMX-XGKFQTDJSA-N	4.863
Roxatidine acetate	SMTZFNFIKUPEJC-UHFFFAOYSA-N	4.692
Rupintrivir	CAYJBRBGZBCZKO-BHGBQCOSSA-N	4.167
Saquinavir	QWAXKHKRTORLEM-UGJKXSETSA-N	3.655
Sepantronium bromide	OTSOHRUMBRSHZ-UHFFFAOYSA-N	4.866
Sertaconazole	JLGKQTAYUIMGRK-UHFFFAOYSA-N	4.119
Shikonin	NEZONWMXZKDMKF-SNVBAGLBSA-N	4.969
Simeprevir	JTZZSQYMACOLNN-VDWJNHBNSA-N	4.623
SP 100030	GHEKQBWHQPMXSH-UHFFFAOYSA-N	4.540
STR2104	LAMQVIQMVKWXOC-UHFFFAOYSA-N	4.071
SU 16f	APYYTEJNOZQZNA-UHFFFAOYSA-N	4.603
Suramin	FIAFUQMPZJWCLV-UHFFFAOYSA-N	5.194
Tannic acid	LRBQNJMCXXYIU-PPKXGCFTSA-N	5.088
TBB	OMZYUVOATZSGJY-UHFFFAOYSA-N	4.688
Teicoplanin	BJNLLBUOHPVGFT-VPMXZCLFSA-N	5.824
Telaprevir	BBAWEDCPNXPBQM-GDEBMAJSA-N	4.657
Theaflavin	SPQUOZYCYRJYJU-ZKSIBHASSA-N	4.825
Thimerosal	HXQVQGWHFERNKMS-UHFFFAOYSA-M	6.516
Tideglusib	PMJIHLSCWIDGMD-UHFFFAOYSA-N	5.851
Tipranavir	MYXACKZCYFKMTE-JETRDCLMSA-N	3.984
Trihexyphenidyl	HWHLPGVTWGOCJO-UHFFFAOYSA-N	3.432
UAWJ246	OFKZKNOAEKYJPI-MFUMQWNRSA-N	7.347
UAWJ247	JKMYXNKXJWIHOA-AKQSQHNNSA-N	7.377
UAWJ248	COHKXTNMOWXZEZ-HCUBDLJSSA-N	7.921
Valganciclovir	WPVFJKSGQUFQAP-GKAPJAKFSA-N	4.778
Vanitolidide	WQYRHRANNNRDIA-UHFFFAOYSA-N	5.529
Vitamin B12	BAOLGUBNOUWZAV-MNFZBJFUSA-N	4.744
Walrycin B	XRVMPWWKLKLPB-UHFFFAOYSA-N	5.876
Z-DEVD-FMK	GBJVAVGBSGRRKN-JYBCORGSA-N	5.609
Z-FA-FMK	ASXVEBPZMSPHB-KPMSDPLLSA-N	4.905
Zopiclone	GBBSUAFBMRNDJC-UHFFFAOYSA-N	3.457
Cobicistat	ZCIGNRJZKPOIKD-CQXVEOKZSA-N	5.171
Cangrelor	PAEBIVWUMLRPSK-IDTAVKCVSA-N	3.046
Denufosol	FPNPSEMJLALQSA-MIYUEGBISA-N	2.863
Bisindolylmaleimide IX	DSXXEELGXBCYNQ-UHFFFAOYSA-N	3.944
TL-3	BJJPNOGMLLUCER-KUTQPOQPSA-N	4.602
QZ2	AGUIQRSXBJPXDU-UHFFFAOYSA-N	5.000
QZ4	MPYJETPFMNPJCJO-UHFFFAOYSA-N	5.187
Auranofin	SFOZKJGZNOBSHF-UHFFFAOYSA-N	5.139
TPCK	MQUQNUAYKLCRME-INIZCTEOSA-N	5.959
IPA-3	RFAXLXKIAKIUDT-UHFFFAOYSA-N	6.538
5-thioguanine	WYWHKKSPHMUBEB-UHFFFAOYSA-N	4.870
89971-88-0	GPCAFSOAVIHODN-SVFBPWRDSA-N	6.046
56942-10-0	DMTDXMMYVHJDJE-NVKBOPVSA-N	5.678
Phenethyl isothiocyanate	IZJDOKYDEWTZSO-UHFFFAOYSA-N	4.994
Scutellarein	JVXZRQGOGOXCEC-UHFFFAOYSA-N	5.237
Dihydromyricetin	KJXSIXMJHKAJOD-LSDHHAIUSA-N	4.486
Quercetagenin	ZVOLCUVKHLEPEV-UHFFFAOYSA-N	4.136
Myricetin	IKMDFBPHZNJCSN-UHFFFAOYSA-N	4.933

Manuscript 1

Compound	InchiKey	pIC50
Ethacrynic acid	AVOLMBLBETYQH-XUHFFFAOYSA-N	5.097
Ivermectin	AZSNMRSAGSSBNP-XPNPUAGNSA-N	4.667
Micafungin	PIEUQSKUWLMALL-YABMTYFHSA-N	4.322
Paritaprevir	UAUIUKWPKRJZJV-QPLHLKROSA-N	4.230
Ombitasvir	PIDFDZJZLOTZTM-KHVQSSSXSA-N	4.122
ML188	JXGIYKRRPGCLFV-JOCHJYFZSA-N	4.930
GRL-1720	JROSKNTUYSHPGX-UHFFFAOYSA-N	6.495
Aurothioglucose	JUSMHIGDXPKSID-GASJEMHNSA-N	5.123
PYR-41	ARGIPZKQJGFSGQ-LCYFTJDESA-N	4.760
Ciluprevir	PJZPDFUUXKKDNB-KNINVKUSA-N	4.683
Tolcapone	MIQPIUSUKVNLNT-UHFFFAOYSA-N	5.102
Levothyroxine	XUIIKFGFIJCVMT-LBPRGKRZSA-N	4.717
GRL-0496	BOSZJNSICHFHMA-UHFFFAOYSA-N	5.723
Quercetin	REFJWTPEDVJJIY-UHFFFAOYSA-N	4.132
Luteolin	IQPNAANSBPBGFQ-UHFFFAOYSA-N	3.879
Asunaprevir	XRWSZZJLZRKHHD-WVWJVSJSA-N	4.111
MI-01	HIKUMVQVEVEKEC-ZONLJJFTSA-N	6.344
MI-02	SVAHUSOLVZWUTO-VATUJKLESA-N	7.283
MI-03	PTPKDVBZJQYYAA-OMBLBEMTSA-N	7.783
MI-04	XEPIMNAFLOXIEZ-KIFLDFMFSA-N	7.733
MI-05	YAZLWQBKWOGOIO-ZUFXTUFISA-N	7.879
MI-06	XWPVHYBRIIHPSA-MQGCXBLESA-N	7.839
MI-07	DYCMIBQYGBOWHW-PILWDZAYSA-N	7.364
MI-08	ITLIZQKXLAJTCO-XAHGZOFNSA-N	7.429
MI-09	RBWMYPKAPIYTJQ-VMBFOHBNSA-N	7.818
MI-10	ZNLCZOIWTNELKD-ZUFXTUFISA-N	7.294
MI-11	AKYGJRDGGYTORB-ZUHDPNGNSA-N	7.876
MI-12	UJTKRTAWEQKHLE-PILWDZAYSA-N	7.721
MI-13	WUFNYTJVPNYFBX-ZUFXTUFISA-N	7.907
MI-14	BDFIPRXCQILSGM-ZUHDPNGNSA-N	7.886
MI-15	QGGYUARATLXVFL-HILJTLORSA-N	6.126
MI-16	LWCOBWFCZRXIAT-SOADLSRISA-N	6.815
MI-17	CFLPDEMDXINBKA-OHTAKOAQSA-N	6.525
MI-18	PPIGFETZIJRNZ-WOPHEGALSA-N	6.279
MI-19	YMZJOZDEQYMIOV-KBCMKWCFSA-N	6.709
MI-20	GCESTRYMIZNLSQH-JTLWTYSLSA-N	6.426
MI-21	FIOKKXQKMFASF-DFPMAUPISA-N	8.119
MI-22	CYRUHAHDZRPFIU-NSBVLCQTSA-N	7.759
MI-23	ZUAZJPLYVHKFN-NAFADOBPSA-N	8.119
MI-24	NXGBAEZSUURIHI-RNDDGGPQSA-N	6.422
MI-25	IDEWJDIVGHDEFX-HVUMFHIUSA-N	7.441
MI-26	KDVUTHOMPYFSDW-IQHRTIQUSA-N	7.161
MI-27	LXKAJLKRKVRCL-GBLYOYEHSAN	7.029
MI-28	PTTIGZHWNGKYAO-BPNDGJTPSA-N	8.036
MI-29	QRFWUTDNKUVKKL-ZEAPZOAHSA-N	7.427
MI-30	HKJKOHDJUKCNRG-ZEAPZOAHSA-N	7.764
MI-31	LQMJJXKADORALLN-ICWLNENESA-N	7.523
MI-32	SBNACRDYLQAFAN-KJDZKPBBSA-N	7.706
2603241-56-9	DUXWXGQOMNHTOZ-UHFFFAOYSA-N	5.000
2603241-59-2	XOPZNMANGWUPLN-UHFFFAOYSA-N	5.195
2603241-62-7	KXGLBZMRWKVPSY-UHFFFAOYSA-N	5.396
2603241-67-2	KFGXONJZGIXSGH-UHFFFAOYSA-N	6.854
2603241-70-7	QRCHFFNDYYKHQM-UHFFFAOYSA-N	6.328

Compound	InchiKey	pIC50
2603241-75-2	PUDFTLRNHLQJHN-UHFFFAOYSA-N	6.553
2603241-79-6	VMIGFUOVZDKJCW-UHFFFAOYSA-N	6.292
2603241-88-7	UMETVGKNHTXVQR-UHFFFAOYSA-N	5.921
2603241-91-2	GJCBTWHGKCNAB-UHFFFAOYSA-N	6.921
2603241-96-7	ZSDMEQDHYASRTO-UHFFFAOYSA-N	6.602
2603241-99-0	XEQQANRELYKWAL-UHFFFAOYSA-N	6.721
2603242-04-0	IVPUTCRGJCQUQU-UHFFFAOYSA-N	6.893
2603242-09-5	JVSFPWJWCIVVJO-UHFFFAOYSA-N	6.959
2603242-14-2	CXKRCAINPQDWER-UHFFFAOYSA-N	7.000
2603242-17-5	RTUXHKBGIXAWJQ-UHFFFAOYSA-N	6.959
2603242-22-2	JNZBKQUMTMLNQO-UHFFFAOYSA-N	7.620
2603242-25-5	DGQRJEVFDWWDCX-UHFFFAOYSA-N	7.432
2603242-30-2	BUTFYHOATLSTQT-UHFFFAOYSA-N	7.444
2603242-35-7	NFXZAFZCOMISJH-UHFFFAOYSA-N	7.745
2603242-38-0	JZJVGWWXTLJZIK-UHFFFAOYSA-N	7.444
2603242-41-5	FBYZCSUBKCAJAK-UHFFFAOYSA-N	7.620
2603242-46-0	MAQWOCWKZDBXGP-UHFFFAOYSA-N	7.432
2603242-51-7	BHQBUBHCKLKLKLD-UHFFFAOYSA-N	7.602
2603242-56-2	UIXRRGFIAPQEF-UHFFFAOYSA-N	6.770
2603242-58-4	GHBWVWMKFOCWLG-UHFFFAOYSA-N	6.921
JCP474	HBWFNALLDLXEL-ADUOGZPQSA-N	6.018
JCP400	LVUKWBPECUZOET-NSHDSACASA-N	5.759
BFC220	OXKRHUJDYNUUBQ-UHFFFAOYSA-N	5.558
BFC204	OJQCBOLVUZALMV-UHFFFAOYSA-N	5.509
BFC210	GRQUYYHAXFLTMJ-UHFFFAOYSA-N	5.478
JCP403	NSAPQHIHCHMAMY-INIZCTEOSA-N	5.524
JCP543	GBIVZALZNKBBEK-BFRYPLHYSA-N	6.310
Ginkgolic acid	YXHVCZZLWZYHSA-FPLPWBNLISA-N	5.582
Anacardic acid	ADFQWBGTDJIESE-UHFFFAOYSA-N	5.859
923691-87-6	LZJFXMUXSZGUNF-UHFFFAOYSA-N	5.437
1184668-18-5	WPOQXWCCIJIAH-UHFFFAOYSA-N	5.078
799837-78-8	OBXJAFXYJIYLRG-UHFFFAOYSA-N	5.328
924084-26-4	ZCRFRMSPOHWZEJ-UHFFFAOYSA-N	4.426
1311313-65-1	ZHVMATQYQXBYRA-UHFFFAOYSA-N	4.260
796106-54-2	ZBZGUXCOHGXIIRI-UHFFFAOYSA-N	4.333
882223-89-4	GJSOCNYKVHSCX-UHFFFAOYSA-N	4.092
2637531-57-6	KKIJAAVRSUCNW-UHFFFAOYSA-N	4.740
3156-42-1	CVWPMONBGNBYCJ-UHFFFAOYSA-N	4.376
3179-10-0	JKQUXSHVQGBODD-UHFFFAOYSA-N	4.323
6316-70-7	PEPYBEJQAIGLRE-UHFFFAOYSA-N	4.338
15795-14-9	NWHBBZILZXZIBO-UHFFFAOYSA-N	4.836
897296-66-1	DQICACAYOYNJIG-UHFFFAOYSA-N	3.848
1043922-53-7	FTJHEEIVOBMMMC-SFHVURJKSA-N	5.660
1043922-54-8	GZBOGFDHJAMENH-HNNXBMFYSA-N	5.324
1043922-56-0	MEESNWIPVLGWFL-VWLOTQADSA-N	5.271
1043922-57-1	SENKHALBSUKUMS-QFIPXVFZSA-N	5.123
1043922-55-9	GAEREYBMNZIBLC-DEOSSOPVSA-N	6.495
1259945-86-2	LRGPAVMYPWWORG-KRWDZBQOSA-N	4.593
1259945-88-4	MWBPTVRPPPPZPG-SFHVURJKSA-N	4.747
1259945-93-1	KFTMZNIURYIOKI-KRWDZBQOSA-N	4.399
1259945-81-7	JESRGSWARNNMLP-HNNXBMFYSA-N	5.120
1259946-02-5	AUPZRAAKTGPDLG-KRWDZBQOSA-N	4.836
1259946-06-9	VQKRNTTUZDCHOZ-SFHVURJKSA-N	5.256

Manuscript 1

Compound	InchiKey	pIC50
1417900-19-6	GNDGQHDQDHABCO-LBPRGKRZSA-N	4.830
1627094-89-6	JRHFOUWUVMCHX-UHFFFAOYSA-N	5.182
1627094-88-5	LHJKRUYOLRIDGD-UHFFFAOYSA-N	4.000
57389-70-5	QAAYQUWLDDDJFC-UHFFFAOYSA-N	4.611
Gü3609	SKSFVXRDFVHXJJ-INIZCTEOSA-N	5.839
Gü3608	HUGYLJWKXGABNS-IBGZPJMESA-N	5.633
Gü3635	HIDSPZDVVVECFJ-PZJWPPBQSA-N	6.111
Gü3641	BYRKEJAEIYZMBW-LADGPHEKSA-N	6.627
Gü3634	PFTKOJFIYWUUPD-WPFOTENUSA-N	7.428
Gü3619	JGNSK MJHNUAKGY-OHSXHVKISA-N	7.423
585571-89-7	FDUHUAOVUUJXMZ-UHFFFAOYSA-N	5.796
PZB00410075	VYUGMEGJNINBO-UHFFFAOYSA-N	4.648
936371-64-1	LJTINUKMKJKLBT-UHFFFAOYSA-N	7.607
1025906-65-3	YWKPSFBPBXOGRP-UHFFFAOYSA-N	4.664
1026559-41-0	TYLOEMFGLQBHPE-UHFFFAOYSA-N	5.714
1087243-12-6	OPNCFNLNJUPFKN-UHFFFAOYSA-N	6.232
1087243-18-2	FDPMNGSSFLBBSI-UHFFFAOYSA-N	7.260
PZB10620022	WGTDRRECQYDFW-UHFFFAOYSA-N	7.466
PZB10620019	LXYZUFZEQSUDBI-UHFFFAOYSA-N	6.644
PZB10620065A	KNFSNAVOLGYYP-UHFFFAOYSA-N	6.451
936371-56-1	QHBWMDVARYUYJW-UHFFFAOYSA-N	8.267
219928-80-0	MNUHXGJUJTWVLU-UHFFFAOYSA-N	7.583
PZB10620023	IQLXSGBYBLCDBO-UHFFFAOYSA-N	7.633
VS12	LXOLOTHZSBNZFS-UHFFFAOYSA-N	5.724
Sitagliptin	MFFMDFZMYVKS-SECBINFHSA-N	5.160
Melittin	VDXZNPDIRNWWCW-JFTDCZMZSA-N	5.789
53250-52-5	UQRUMZDBXRBEPM-UHFFFAOYSA-N	6.991
Sennoside B	QFJIQUOMKJNYLI-VZZBSESQSA-N	6.983
Naringenin	FTVWIRXFELQLPI-UHFFFAOYSA-N	4.125
578-74-5	KMOUJOKENFFTPU-QNDFHXLGSA-N	7.131
Acteoside	FBKJMQYURKNSU-ZLSOWSIRSA-N	5.640
Puerarin	HKEAFJYKMMKDOR-VPRICQMDSA-N	4.310
Daidzein	ZQSIJRDFPHDXIC-UHFFFAOYSA-N	4.252
Astragalin	JPUKWEQWGBDDQB-QSOFNFLRSA-N	3.845
Vitexin	SGEWCQFRYRRZDC-VPRICQMDSA-N	3.745
Resveratrol	LUKBXSAWLPMMSZ-OWOJBTEDSA-N	3.987
Caffeic acid	QAIPRVGONGVQAS-DUXPYHPUSA-N	3.706
Rutin	IKGXIBQEEMLURG-NVPNHPEKSA-N	4.495
Genkwanin	JPMYFOBNRRGFNO-UHFFFAOYSA-N	4.974
Isorhamnetin	IZQSVBODKVDZ-UHFFFAOYSA-N	4.500
Apigenin	KZNIFHPLKGYRTM-UHFFFAOYSA-N	4.357
Sciadopitysin	YCXRBCHEOFVYEN-UHFFFAOYSA-N	5.963
Ginkgetin	AIFCFBUSLAEIBR-UHFFFAOYSA-N	5.526
Isoginkgetin	HUOQMAOYXQFIDQ-UHFFFAOYSA-N	5.633
Amentoflavone	YUSWMAULDZXHPY-UHFFFAOYSA-N	4.120
Bilobetin	IWEIJEPIYMAGTH-UHFFFAOYSA-N	4.951
Ginkgolic acid C17:1	MBYNDKVOZOAIS-FPLPWBNSA-N	5.924
Ginkgolic acid C13:0	VEPUCZUJLKAVNM-UHFFFAOYSA-N	5.447
1359941-93-7	KSCRVOKQPYZBHZ-IXPOFIJOSA-N	7.027
898264-58-9	WDMSSWWUPIMEGNZ-UHCVVMIDSA-N	6.544
2204222-43-3	PMAJEYHZNBPFI-ZJGAYNASAN-N	7.187
217798-51-1	INJSPYUOTZNFSE-UHFFFAOYSA-N	6.187
106363-74-0	ODMXJEXAQZKNMV-UHFFFAOYSA-N	6.387

Compound	InchiKey	pIC50
106363-71-7	FSGQFSVVMZXOLX-UHFFFAOYSA-N	6.456
153871-75-1	FOEPTYDLLXMPLJ-UHFFFAOYSA-N	6.886
1034707-14-6	PLADVUNGKXPXM-X-UHFFFAOYSA-N	6.824
81743-92-2	JWLXOHUMYJGZPC-UHFFFAOYSA-N	6.585
2237228-66-7	LGLOPWIQZGBWRF-UHFFFAOYSA-N	6.620
2641582-47-8	MWOBNUZBAFCVJM-UHFFFAOYSA-N	6.481
106966-77-2	QZJDTYZIHCOMMO-UHFFFAOYSA-N	7.131
2634-33-5	DMSMPAJRVJJAGA-UHFFFAOYSA-N	6.180
2527-66-4	RDWXSJJCICPOOKO-UHFFFAOYSA-N	6.108
4299-05-02	CIZCWSJDCWBINO-UHFFFAOYSA-N	6.137
4299-09-06	BKCDKWJRHGJSRF-UHFFFAOYSA-N	6.387
180801-33-6	JZMXCENGODLQED-UHFFFAOYSA-N	6.319
35159-81-0	BUQJUCVPPUJQBJ-UHFFFAOYSA-N	6.201
443286-41-7	PRLIOOFJTHOJPZ-UHFFFAOYSA-N	6.886
2514-36-5	GTTHESNNWLOVLD-UHFFFAOYSA-N	6.770
2514-30-9	KRXMYBAZKJBAB-UHFFFAOYSA-N	6.757
2620-91-9	XZZLYYYXSWYYKL-UHFFFAOYSA-N	6.409
2304826-28-4	SCOBQUKCPDHPB-UHFFFAOYSA-N	6.959
1909243-65-7	CPXPGFUBQYOVQL-UHFFFAOYSA-N	6.387
Ebsulfur	KDMQSUUXKLZVNU-UHFFFAOYSA-N	6.626
4322-93-4	ZIFJJRRCTKAIGB-UHFFFAOYSA-N	6.444
788092-78-4	KWLOKUMNWAEMHY-UHFFFAOYSA-N	6.638
51584-03-3	GYVKPAVYPMIITO-UHFFFAOYSA-N	6.041
1601403-03-5	IRBBDJYBNHNLKO-UHFFFAOYSA-N	6.420
Zafirlukast	YEEZWCHGZLNKEK-UHFFFAOYSA-N	4.331
Lomustine	GQYIWUULTXOXAJ-UHFFFAOYSA-N	5.232
Closantel sodium	JMPFSEBWVLAJKM-UHFFFAOYSA-N	5.276
Alexidine	LFVVPNBFBUSHL-UHFFFAOYSA-N	4.643
Vitamin K3	MJVAVZPDRWSRRC-UHFFFAOYSA-N	5.091
1,4-Naphthoquinone	FRASJONUBLZVQX-UHFFFAOYSA-N	6.398
1010-60-2	CCTJHVLTAJTPBV-UHFFFAOYSA-N	5.717
Juglone	KQPYUDDGWXQXHS-UHFFFAOYSA-N	5.109
475-38-7	RQNVIKXOOKXAJQ-UHFFFAOYSA-N	6.310
Plumbagin	VCMMXZQDRFWYSE-UHFFFAOYSA-N	5.036
14918-69-5	KFAVXUQFBGHURZ-UHFFFAOYSA-N	5.815
Chebulagic acid	HGJXAVROWQLCTP-YABCKIEDSA-N	5.041
Punicalagin	ZJVUMAFASBFUBG-UYMKNUMKSA-N	5.335
MR6-7-2	UWQXYXVYDDJKSY-UHFFFAOYSA-N	6.440
MR6-17-1	ZFXLHLJHFSDXFG-UHFFFAOYSA-N	6.154
MR6-18-4	OQWCODYRVBTZGK-UHFFFAOYSA-N	6.462
MR6-26-2	DDEAVXZRFSPDI-UHFFFAOYSA-N	6.331
MR6-31-2	XZWKXIFLMYRVTK-UHFFFAOYSA-N	6.084
LASSBio-1945	RTNFALLGIDGVDR-UHFFFAOYSA-N	4.797
Z-VAD-FMK	MIFGOLAMNLSLGH-QOKNQOGYSA-N	5.958
Z-IETD-FMK	PHLCQASLWHYEMX-DEKIMQJDSA-N	5.933
330453-41-3	WNJONLIYGJXNFC-UHFFFAOYSA-N	4.721
383896-47-7	GGDRMKGZRIVKSM-UHFFFAOYSA-N	4.420
2648754-66-7	HAHPXCSLUABGFD-BZSNMDCSA-N	7.301
1681060-26-3	YDHLZYPCVKECFG-ACRUOGEOSA-N	6.215
1065478-18-3	JKMYXNKXJWIHOA-ACRUOGEOSA-N	6.783
2648754-67-8	CZMUQIYXGOYLKC-SZMVWBNQSA-N	6.886
2648754-68-9	AFYLRCTUYTWBEO-WDSOQIARSA-N	6.824
890648-76-7	QCRLRBSKQJIUTG-MPGHIAIKSA-N	7.222

Manuscript 1

Compound	InchiKey	pIC50
2648754-69-0	BZJBCSRHPATGV-WNHJNPCNSA-N	6.569
2648754-70-3	BZJBCSRHPATGV-BVBHFADKSA-N	6.456
2648754-71-4	LWPZKXWMHWKVCB-SZMVWBNQSA-N	6.854
2648754-72-5	OPFBFTVBHSYZLW-WDSOQIARSA-N	6.620
2648754-73-6	CQURQDKUTCKDIG-MPGHIAIKSA-N	7.301
2648754-74-7	KAXIWAYDOIRHKD-QZHFEQFPISA-N	6.367
2648754-76-9	FHMWCUZBRKDEDN-HTCLRFROSA-N	7.097
2648754-78-1	VGTSDAFHDXWTEF-XAXASBLLSA-N	7.155
2648754-80-5	DRWFTJQFNJPAAY-YQASGENESA-N	7.155
2648754-82-7	FPXMZZWNZPMKPB-PIJSLZJWSA-N	7.097
2648754-84-9	PCJKUFVSLOJHBP-XAXASBLLSA-N	7.398
2648754-86-1	XDBBVZVFGANLDV-PNXRDEECSA-N	7.301
HZZRMKZNUIJMBM-UHFFFAOYSA-N	HZZRMKZNUIJMBM-UHFFFAOYSA-N	5.292
ZSSAWRSYSQEDJU-UHFFFAOYSA-N	ZSSAWRSYSQEDJU-UHFFFAOYSA-N	4.620
PXYZVGLGISWIRT-UHFFFAOYSA-N	PXYZVGLGISWIRT-UHFFFAOYSA-N	4.959
GSCFTMXOXFVPAI-UHFFFAOYSA-N	GSCFTMXOXFVPAI-UHFFFAOYSA-N	5.523
YSWKVKNPZIMIS-UHFFFAOYSA-N	YSWKVKNPZIMIS-UHFFFAOYSA-N	4.886
ZBKPEDGUBSTOIK-UHFFFAOYSA-N	ZBKPEDGUBSTOIK-UHFFFAOYSA-N	5.481
HFPUFXJMLZLTRB-UHFFFAOYSA-N	HFPUFXJMLZLTRB-UHFFFAOYSA-N	5.658
Dalcetrapib	YZQLWPMZQVHJED-UHFFFAOYSA-N	4.842
ML300	OCKMJWOFMCSMEX-UHFFFAOYSA-N	4.976
1473419-56-5	BRLBTOJQLXGOHO-UHFFFAOYSA-N	4.739
1473419-57-6	LHBIUOYHJIXRST-UHFFFAOYSA-N	6.022
2694063-36-8	BDMCAWVDUSVNBO-UHFFFAOYSA-N	4.724
2694063-37-9	AMNRXIOJAADJRF-UHFFFAOYSA-N	6.357
2694063-38-0	UTWHGTFFVPLUKB-UHFFFAOYSA-N	4.798
2694063-39-1	PAOYPYMKBNVXRO-UHFFFAOYSA-N	5.793
2694063-40-4	WMGPUGILOFJULP-UHFFFAOYSA-N	5.057
1473419-59-8	ZFYLRJRHNLKAHPV-UHFFFAOYSA-N	5.854
2694063-41-5	ICGCVLDLCRLNGJ-UHFFFAOYSA-N	5.680
1473419-58-7	NGIZLVDUIAGJQR-UHFFFAOYSA-N	5.295
2694063-42-6	AZBYGSYNZGNIEI-UHFFFAOYSA-N	6.036
2694063-43-7	MHMNSHNWWVMPDA-UHFFFAOYSA-N	4.691
2694063-44-8	PYVQFTGANJRGEF-UHFFFAOYSA-N	6.959
2694063-45-9	IIHIUQOLONBIHR-UHFFFAOYSA-N	5.020
2694063-46-0	XZJQABWECIWHRU-UHFFFAOYSA-N	7.222
2694063-47-1	OQGTZTSFRMUFBA-UHFFFAOYSA-N	5.335
2694063-48-2	CIABYILRZWPLFX-UHFFFAOYSA-N	5.600
2694063-49-3	DLTPUSCBKNOVLV-UHFFFAOYSA-N	6.081
2694063-50-6	MUXXBMHXURKPHX-UHFFFAOYSA-N	5.936
2694063-51-7	YPAJVLNWSXODGF-UHFFFAOYSA-N	4.638
2694063-52-8	SDKAWNLSLJAJOGF-UHFFFAOYSA-N	5.799
2694063-53-9	QLYAVYHRKCZVAB-UHFFFAOYSA-N	6.409
2694063-54-0	LBKWIIALDNORK-UHFFFAOYSA-N	5.069
2694063-56-2	UHMIVSNGFNZGBS-UHFFFAOYSA-N	6.924
2694063-57-3	FXOXREWVQYLPFR-UHFFFAOYSA-N	6.754
2694063-58-4	YYPWVCNJTBELIN-UHFFFAOYSA-N	6.506
2694063-59-5	AZRNSMNPERAERM-UHFFFAOYSA-N	6.268
2694063-60-8	OMFBUEFMMHUWOX-UHFFFAOYSA-N	6.127
2694063-61-9	JNKQPWXEHPYJAU-UHFFFAOYSA-N	6.569
2694063-62-0	ZUSJSFVCQUBXFQ-UHFFFAOYSA-N	6.484
2694063-63-1	ZJWZLVITQIYDKL-UHFFFAOYSA-N	6.706
2694063-64-2	WCLFFPWYAFAVEX-UHFFFAOYSA-N	6.478

Compound	InchiKey	pIC50
2694063-65-3	GITOTVPGLSPTSF-UHFFFAOYSA-N	6.767
CCF0058981	NQTRFDLUODRJJC-UHFFFAOYSA-N	7.167
2694063-66-4	LLFIKGDQQRLTFJ-UHFFFAOYSA-N	5.896
2694063-67-5	HQKYYKUPSYGXIW-UHFFFAOYSA-N	5.463
Rottlerin	DEZFNHCVIZBHBI-ZHACJKMWSA-N	4.432
M-1805	YMGIJUKMVDMGRY-VGOFMYFVSA-N	4.000
Pectolinarin	DUXQKCCCELUXOE-UHFFFAOYSA-N	3.330
1197726-57-0	XSBPSLHXSBBPQO-UHFFFAOYSA-N	6.097
1788623-85-7	FGELQLYSYCCROF-UHFFFAOYSA-N	6.137
941105-74-4	OPVFOSHQEHUQNB-UHFFFAOYSA-N	5.688
871308-04-2	WRWLTWLDXNKDO-UHFFFAOYSA-N	6.119
2109487-17-2	COZHWXLQRALJBE-UHFFFAOYSA-N	6.161
1197793-62-6	GRZJWHXVWUOSAS-UHFFFAOYSA-N	5.764
1417699-67-2	JXGIYKRRPGCLFV-UHFFFAOYSA-N	5.152
2596275-00-0	BFMIBWJNRQASNC-UHFFFAOYSA-N	5.270
2596275-04-4	OMEKWEAKEKVEHF-UHFFFAOYSA-N	5.413
2596275-10-2	VNOUKGOBWQZPOV-UHFFFAOYSA-N	5.600
2498659-44-0	OXLNXQQCDCGFLH-UHFFFAOYSA-N	5.622
2596275-16-8	USOVOWGAZPCBIJ-UHFFFAOYSA-N	5.107
2596275-19-1	JGOVLTVDNDLBEO-UHFFFAOYSA-N	5.126
2596275-25-9	BYWXQKBXORVZIJ-UHFFFAOYSA-N	5.090
2596275-30-6	GPFAPGKAYCRHKV-UHFFFAOYSA-N	5.056
2596275-33-9	BROLJTKQRJMMHO-LWAJAJQLZSA-N	5.244
2596275-38-4	FIUMZUXZKZLSW-LWAJAJQLZSA-N	5.706
2596275-41-9	QVUDAPUVFJLICX-LWAJAJQLZSA-N	4.660
2596275-43-1	UEDFBTZVUPSJIJ-QASNKKAYSAN	5.263
2596275-46-4	POPWPHIGQFWRDK-IHOKFDBFSA-N	6.292
2596275-49-7	IJUSQHGCLSCLBK-QDLLFRBVSA-N	5.595
2596275-53-3	NJXPUSRCRCOWEQ-QDLLFRBVSA-N	6.252
2596275-55-5	WVXFUFQTNVWHME-XYXHBKGXSA-N	6.027
2596275-59-9	DBSGFTPGMRLXJV-RBQQCVMASA-N	5.742
2596275-64-6	IUUHHCBBPQRRQS-ACXKHFGCSA-N	4.455
2596275-66-8	RYGJIKURQDMMQZ-YUDQIZAISAN	6.700
2596275-74-8	RZTKCADKVFEBSR-RBQQCVMASA-N	6.018
2596275-76-0	CIRLSXXITHQIMD-RBQQCVMASA-N	5.607
2596275-80-6	OOWIYZOIKWPXEB-RBQQCVMASA-N	5.304
1417700-04-9	WLZURSBQLFFPCE-UHFFFAOYSA-N	5.307
2596275-87-3	AFFQDEHJIFFRBV-UHFFFAOYSA-N	4.918
2596275-90-8	GODDDBMSKPRAAE-UHFFFAOYSA-N	5.670
2596275-93-1	LYHYGWZUSDXIHQ-UHFFFAOYSA-N	5.387
2596275-96-4	LLIAUWAYLJLCJT-UHFFFAOYSA-N	5.138
2596275-99-7	OKHRVKARLOVBRJ-UHFFFAOYSA-N	5.572
2596276-02-5	UKLYROLISOJUEM-UHFFFAOYSA-N	5.235
2596276-05-8	DFVOOMGEUDYTE-RBQQCVMASA-N	6.092
2596276-09-2	MEWSAFKJOGMCKZ-QDLLFRBVSA-N	6.174
2596276-13-8	LERCEIGOROJLAY-RBQQCVMASA-N	6.553
2596276-15-0	NWPFRNAKGQTGOK-UHFFFAOYSA-N	5.830
MYVUXRRMNMASTR-UHFFFAOYSA-N	MYVUXRRMNMASTR-UHFFFAOYSA-N	5.979
Quinacrine	GPKJTRJOBQKQK-UHFFFAOYSA-N	5.108
Bacitracin	MNJKVJAYSVAQLU-RNHDWVCBSA-N	5.869
Zinc Pyrithione	FGVVTMRZYROCTH-UHFFFAOYSA-N	6.014
Z-YVAD-FMK	MVPQJUFFTWWKBT-LBDWYMBGSA-N	5.090
Z-WEHD-FMK	NLZNSSWGRVBWIX-KRCBVYEFSA-N	4.771

Manuscript 1

Compound	InchiKey	pIC50
Z-AVLD-FMK	AFQSECJYXLRBY-FVLCUVFVSA-N	9.046
Z-SAVLD-FMK	IRAIPRIZVRGSTD-OPTWNZHXSA-N	7.770
Z-ASAVLD-FMK	ISHHUUQOINBYKF-GHXSILMLSA-N	6.585
Diclazuril	ZSZFUDFOPOMEET-GFCCVEGCSA-N	4.538
Oprea1_409650	UOUXILZUBDIWQU-UHFFFAOYSA-N	4.000
ALP-POS-ddb41b15-4	APWZRPKMUTUINS-UHFFFAOYSA-N	5.389
ALP-POS-ddb41b15-1	IMGNMZGBRMQRDL-UHFFFAOYSA-N	4.753
ALP-POS-ddb41b15-3	PBTDXWZFLBQGLA-UHFFFAOYSA-N	4.238
1489258-19-6	CISWBUKGCQINLS-AWNIVKPZSA-N	6.222
Ciprofloxacin	MYSWGUAQZAJ5OK-UHFFFAOYSA-N	5.290
Remdesivir	RWWYLEGWBNMMLJ-YSOARWBDSA-N	5.996
Phycocyanobilin	NNMALANKTSRILL-UHFFFAOYSA-N	4.149
Tigecycline	LQQYICBWGHYRRN-QCERJBSOSA-N	4.688
Tetracycline	YRKNPFVJGRPARO-XWDZYTQJSA-N	4.682
FAD(Na)2	VWWQXMAJTJZDQX-UYBVJOGSSA-N	4.372
QYENXTYKACLCGO-UIWACHPISA-N	QYENXTYKACLCGO-UIWACHPISA-N	6.770
BSPJDKCMFIPBAW-ITQFNVDMSA-M	BSPZFJDYQHDZNR-HVQOYDIZSA-N	6.745
WIKBMBOAVOAE-SPMFUAFBSA-M	BYNZDGFDRVJWKF-YRVKCN5QSA-N	6.149
LLFRIYQRGIGTCQ-JJVIDZKBSA-M	KVZZRDKCLBWPQO-PSQUDNLJSA-N	6.097
VXWHJWYFXCLWID-UJSXZRH5SA-N	VXWHJWYFXCLWID-UJSXZRH5SA-N	5.883
QYENXTYKACLCGO-IUIHOACCSA-N	QYENXTYKACLCGO-IUIHOACCSA-N	6.796
BSPJDKCMFIPBAW-KYVYQFSESA-M	BSPZFJDYQHDZNR-HJGFZKFC5A-N	6.721
LLFRIYQRGIGTCQ-QUYBAAFTSA-M	KVZZRDKCLBWPQO-RYSMJKKJSA-N	6.167
QYENXTYKACLCGO-FOAIQHMTSA-N	QYENXTYKACLCGO-FOAIQHMTSA-N	6.620
BSPJDKCMFIPBAW-NDDIRMSCSA-M	BSPZFJDYQHDZNR-RPQLVXMLSA-N	6.678
LLFRIYQRGIGTCQ-ZRBRQOCOSA-M	KVZZRDKCLBWPQO-UVELGDQISA-N	6.114
Purpurogallin	WDGFFVCWBZVLCE-UHFFFAOYSA-N	5.339
CCG-50014	QUIIYITNGOFEI-UHFFFAOYSA-N	6.824
PD119507	YDEUKNRKEYICTH-UHFFFAOYSA-N	5.688
Thioguanosine	OTDJAMXESTUWLO-UHFFFAOYSA-N	5.967
NSC228155	ICCFXXDUYSPKOL-UHFFFAOYSA-N	7.000
MLS0315771	GRGDWLDSVPJEJU-UHFFFAOYSA-N	7.699
86-93-1	GGZHVNZHFYCSEV-UHFFFAOYSA-N	5.627
CHEMBL1339149	SHVMSYDYOYCGDZ-UHFFFAOYSA-N	7.699
ML311	NGAPBLRRJ5KIRT-UHFFFAOYSA-N	6.824
CL-17107	BOJLJKMUGYYKCZ-UHFFFAOYSA-N	6.409
Phenacyl Chloride	IMACFCSSMIZSPP-UHFFFAOYSA-N	6.721
PD120911	AWQPMEGCCWPTRI-UHFFFAOYSA-N	6.268
PD119792	ZGZGUAKCJRZXRV-UHFFFAOYSA-N	5.910
TDZD-8	JDSJDASOXWCHPN-UHFFFAOYSA-N	6.854
SKF-38393	JUDKOGFHZYDMDF-UHFFFAOYSA-N	5.553
Benzothiazyl disulfide	AFZSMODLJJC5PP-UHFFFAOYSA-N	6.796
PD096194	QSVDRGTXPDNZQZ-UHFFFAOYSA-N	5.690
HC-056456	RUQGCDMXFBOTMW-UHFFFAOYSA-N	4.970
HQ-415	ZUOOFZJZQYFZQW-UHFFFAOYSA-N	6.569
Benserazide	BNQDCRGUHNALGH-UHFFFAOYSA-N	6.854
BN82002	GOKYHQGR1XMNE-UHFFFAOYSA-N	5.216
ML025	KTPSEIKHZBDCBJ-UHFFFAOYSA-N	5.375
THIRAM	KUAZQDVKQLNFPE-UHFFFAOYSA-N	7.301
BI-78D3	QFRLDZGQE5CCJZ-UHFFFAOYSA-N	6.523
Rabeprazole (sodium)	YREYEV1YC5EVJK-UHFFFAOYSA-N	6.131
LDN-57444	OPQRFP5LZZPCCH-LTGZKZE5SA-N	4.748
INCA-6	GCHPUOHXXCNSQL-UHFFFAOYSA-N	6.620

Compound	InchiKey	pIC50
PD121351	FXPVOLWQNNGFHA-UHFFFAOYSA-N	5.496
SU3327	NQQBNZBOOHVQP-UHFFFAOYSA-N	7.523
SCH-202676	FFUBTEITUNMMOK-UHFFFAOYSA-N	6.009
SENNOSIDE A	QFJIQUOMKJNYLI-QSXPVFSBSA-N	5.799
pyrrolidine-dithiocarbamate	VSWDORGPiHIGNW-UHFFFAOYSA-N	6.678
PYRITHIONE ZINC	YBBJKCMCRQZMA-UHFFFAOYSA-N	5.609
eseroline(-)	HKGWQUVGHPDEBZ-OLZOCXBDSA-N	5.701
erbstatin-analog	BQCNSTFWSKOWMA-GORDUTHDSA-N	4.886
PROFLAVINE HEMISULFATE	WDVSHHCDHLJJJR-UHFFFAOYSA-N	6.081
Delavirdine (mesylate)	WHBIGIKBNXZKFE-UHFFFAOYSA-N	5.398
MIRA-1	YXEWPGYLMHXLPS-UHFFFAOYSA-N	6.523
MERBROMIN	PVBQYTFCVWZSJ-K-UHFFFAOYSA-N	6.538
JAK3-inhibitor-V	FOTCGZPFSUWZBN-UHFFFAOYSA-N	5.907
silver-sulfadiazine	SEEPANYCNGTZFQ-UHFFFAOYSA-N	6.125
Tetramethylthiuram monosulfide	REQPQFUJGGOFQL-UHFFFAOYSA-N	6.027
N-oxydiethylenebenzothiazole-2-sulfenamide	MHKLKWCYGIBEQF-UHFFFAOYSA-N	5.292
NT157	NIPUPOUEGOSAAO-OWOJBTEDSA-N	6.301
(-)-Epicatechin	PFTAWBLQPZVEMU-UKRRQHHQSA-N	4.772
dihydroxidine	BGOQGUHWXBGXJW-RHSMWYFYSA-N	5.959
Tasuldine	HMCTXMOKMWELFJ-UHFFFAOYSA-N	5.044
oxyfedrine	GDYUVHBMFVMBAF-LIRRHJRJNSA-N	5.629
Oleuropein	RFWGABANNQMhMZ-DXNYSgJVSA-N	5.471
Felbinac ethyl	NPPJLSILDPVHCM-UHFFFAOYSA-N	6.699
BAY-11-7082	DOEWDSDBFrhVAP-KRXBUXKQSA-N	6.086
BAY-11-7085	VHKZGNPOHPFPER-ONNFQVAWSA-N	6.076
riodoxol	XKFZYVWWXCHIX-UHFFFAOYSA-N	6.481
Dexrazoxane hydrochloride	BMKDZUISNHGIBY-ZETCQYMhSA-N	5.205
PR-619	ZXOBLNBVNROVLC-UHFFFAOYSA-N	6.387
Ro-106-9920	JFSXSNSCPNFCDM-UHFFFAOYSA-N	5.951
omega-(4-Iodophenyl)pentadecanoic_acid	YAONEUNUMVOKNQ-UHFFFAOYSA-N	5.648
EOSIN	DBZJJPROPLPSM-N-UHFFFAOYSA-N	5.378
CISPLATIN	QGZKDVFNNGYKY-UHFFFAOYSA-N	5.418
Bonaphthone	MXWZRRPNVLCHMY-UHFFFAOYSA-N	7.000
TAS-103 (dihydrochloride)	ROWSTIYZUWEOMM-UHFFFAOYSA-N	5.524
MNS (3;4-Methylenedioxy-alphanitrostyrene)	KFLWBZPSJQPRDD-ONEGZZNKSA-N	4.826
SB-747651A	MBCJUIJWPYUEBX-UHFFFAOYSA-N	5.684
Satraplatin	PAFZNILMFXTMIY-UHFFFAOYSA-N	5.243
Oridonin (Isodonol)	VPZCKRKZFRCZMX-QRTGZHHDSA-N	5.590
PFK-015	UJJUKZPBUMCSJZ-BQYQJAHWSA-N	5.188
XL228	ALKJNCZNEOTEMP-UHFFFAOYSA-N	5.827
carboxypyridine-disulfide	GSASOFRDSIKDSN-UHFFFAOYSA-N	6.620
Dixanthogen	FVIGODVHAVLZOO-UHFFFAOYSA-N	5.224
Chicago Sky Blue	BGULJNMSWJQIEK-UHFFFAOYSA-N	5.057
CARBIDOPA	TZFNLOMSOLWIDK-UHFFFAOYSA-N	5.450
caracemide	JURAJLFHWXNPHG-UHFFFAOYSA-N	5.876
TRi-1	PRKWNRUVAZZHPH-UHFFFAOYSA-N	5.297
Ambazone	ANZIOUQAFBXNHU-UHFFFAOYSA-N	5.032
exifone	XEDWWPGWIXPVRQ-UHFFFAOYSA-N	5.359
SB-268262	YXHLPJQQPOMPGE-UHFFFAOYSA-N	5.728
TENATOPRAZOLE	ZBFDAUIVDSSISP-UHFFFAOYSA-N	6.013
ditiocarb-sodium-trihydrate	LMBWSYZSUOEYSN-UHFFFAOYSA-N	4.989
U-0124	GXGXSVFKGDVBMD-UHFFFAOYSA-N	5.210
Salvianolic_acid B	SNKFFCBZYFGCQN-PDVBOLEISA-N	5.466

Manuscript 1

Compound	InchiKey	pIC50
Brilliant Green	HXCILVUBKWANLN-UHFFFAOYSA-N	4.937
Razoxane	BMKDZUISNHGIBY-UHFFFAOYSA-N	5.496
Parthenolide	KTEXNACQROZXEV-SLXBATTESA-N	4.998
Sodium Danshensu	PAFLSMZLRSPALU-UHFFFAOYSA-N	5.243
compound 3 [PMID: 23849879]	BIEYFOHSWAIQDF-UHFFFAOYSA-N	5.295
gallium-triquinolin-8-olate	MCJGNVYPOGVAJF-UHFFFAOYSA-N	5.188
JK-184	ROYXIPOUVGDTAO-UHFFFAOYSA-N	5.297
STRYCHNINE SULFATE	QMGVPVSNSZLJIA-OFYMUHMXSA-N	4.984
thioguanosine	OTDJAMXESTUWLO-UUOKFMHZA-N	5.200
PFK-158	IAJOMYABKVAZCN-AATRIKPKSA-N	5.116
ZM-39923	JSASWRWALCMOQP-UHFFFAOYSA-N	6.237
Myricitrin (Myricitrine)	DCYOADKBABEMIQ-FLCVNNLFSA-N	5.437
Temoporfin	KSXLETNRCMCRAQ-UHFFFAOYSA-N	5.184
APOMORPHINE HYDROCHLORIDE	VMWNQDUVQKEIOC-CYBMUJFWSA-N	6.284
sulforaphane	SUVVMJBTUFCVSAD-UHFFFAOYSA-N	5.261
Hematoxylin (Hydroxybrazilin)	WZUVPKPBWHMQCE-XJKSGUPXSA-N	6.658
CGH2466	BEPGKLOHQTUHX-UHFFFAOYSA-N	5.270
Captan	LDVVMCZRFWMZSG-UHFFFAOYSA-N	6.174
BX-795	VAVXGGRQQJZYBL-UHFFFAOYSA-N	5.060
CEFACLOX	QYIYFLOTGYLRGG-GPCCPHFNSA-N	4.755
3-Bromopyruvic_acid	CHNJCELMWMIDSC-UHFFFAOYSA-N	4.939
ozolinone	NQFBZYYUAFJYNS-FPLPWBNSA-N	4.934
MKC-3946	IVQVBMWPWPTSNO-UHFFFAOYSA-N	5.876
GNE-0877	ZPPUMAMZIMPJGP-UHFFFAOYSA-N	5.022
CBS; N-Cyclohexyl-2-benzothiazolesulfenamide	DEQZTKGFXNUBJL-UHFFFAOYSA-N	6.114
MIANSERIN HYDROCHLORIDE	UEQUQVLFPOEMF-UHFFFAOYSA-N	4.717
lofepramine	SAPNXPWPAUFUJU-UHFFFAOYSA-N	4.942
WAY-308264	ZFKWFHMKVZEDEN-UHFFFAOYSA-N	5.499
heclin	SPTWXRJNCFIDRQ-ZHACJKMWSA-N	5.305
PD119466	KJFHJZMOFADPOP-UHFFFAOYSA-N	5.058
NSC697923	GAUHWPWCXOLCZ-UHFFFAOYSA-N	4.836
Ubiquitin Isopeptidase Inhibitor I	SSAMIOWVUGBIJK-KAVGSPWWSA-N	5.410
LEVODOPA	WTDQRQBEARUVNC-UHFFFAOYSA-N	5.255
PD 150606	DJCVSFWGKYHMKH-YVMONPNESA-N	4.693
Aldoxorubicin	OBMJQRLIQQTJLR-USGQOSEYSA-N	6.237
NIFEDIPINE	HYIMSNHJOBLJNT-UHFFFAOYSA-N	5.182
2376907-80-9	AGSOCSLLVKCMJG-HOFKKMOUSA-N	4.148
2376907-85-4	POVPOLRRHQMDKJ-WNJJXGMVSA-N	4.000
2380146-82-5	KNPIJQSFBVXHNH-UHFFFAOYSA-N	4.287
2380146-80-3	XFIVYEBVEADCBM-QABMSTFYSA-N	4.281
2380147-10-2	QZOWNHGXNWECDR-JSOSNVBQSA-N	4.524
2380147-06-6	SDJUQUXATYBXDP-KAYWLYCHSA-N	4.110
2380147-02-2	MFPMIEOKNMGOQZ-MYVCOICNSA-N	4.583
2380146-91-6	SDADTGNLMAGZPQ-NJHZRGNWSA-N	4.684
2679814-83-4	PPVFDMTQNZQFKN-UHFFFAOYSA-N	6.484
2679814-84-5	YMXUIKGOSFGGAJ-UHFFFAOYSA-N	7.081
2679814-85-6	CVJBJRQQYWUAPJ-UHFFFAOYSA-N	7.071
2679814-86-7	KDFIKGUUIGLMGS-UHFFFAOYSA-N	6.883
2679814-87-8	GNSONAJPJOLHV-UHFFFAOYSA-N	7.569
2679814-88-9	RJHJFSODWKKAIK-UHFFFAOYSA-N	7.377
2679814-89-0	BZOQLNKQJLJRJD-UHFFFAOYSA-N	6.979
2679814-90-3	LXBUHWWQNTWTSJL-UHFFFAOYSA-N	7.420
2679814-91-4	WQPXNFAUKKEFHC-UHFFFAOYSA-N	7.215

Compound	InchiKey	pIC50
2679814-92-5	UJBVSKKPZZUOLN-UHFFFAOYSA-N	7.229
2679814-93-6	KIJKUSPNZALBLD-UHFFFAOYSA-N	7.357
2679814-94-7	ZVTHINWYSDBXSJ-UHFFFAOYSA-N	6.886
2679814-96-9	CMYSAEAJENJTSK-UHFFFAOYSA-N	7.215
CDD-1714	GKHVCTHOJIZMSY-UHFFFAOYSA-N	7.111
CDD-1713	LDHIKLLVVUEYPE-UHFFFAOYSA-N	7.438
CDD-1976	IXTRHVKNNCAPDT-UHFFFAOYSA-N	7.264
CDD-1777	UBXULBPPTHWXTK-UHFFFAOYSA-N	5.618
Vanicoside A	RNXDQKHOEWIRNH-FBMBONOOSA-N	4.636
Vanicoside B	ALSDWGAQQGXOHC-PWYSLETCSA-N	4.361
2683066-29-5	SEPKTDPAPERLPET-IHPCNDPISA-N	5.824
2683066-30-8	RCHQWDODKNPHAO-SNRMKQJ TSA-N	7.710
2683066-34-2	HLSIDWWRICRQTA-NYVOZVTQSA-N	7.193
2683066-36-4	XQNMBIMRDZGLC-NYVOZVTQSA-N	7.660
2683066-37-5	DNUOOXQKGWPVRL-YYWHXJBOSA-N	7.699
2683066-38-6	JDBHNHGTZQNIEF-VJBM BRPKSA-N	7.903
QOTVXHVQJYHEFL-OYDLWJJNSA-N	QOTVXHVQJYHEFL-OYDLWJJNSA-N	7.064
2683066-40-0	MSRIXUCOJKCCJJ-TUSQITKMSA-N	7.197
2683066-41-1	QHAHSNXJBQCKEG-NWVWQQAFSA-N	8.194
2683066-42-2	PKJWLVLG HATRZ-NWVWQQAFSA-N	9.000
2683066-43-3	PFYSTXWNQVFKTC-IGKWTDBASA-N	6.901
2683066-44-4	KIRIGTQPZIKIAF-IGKWTDBASA-N	8.398
2683066-45-5	DHBIOTHYZUQBSL-OYDLWJJNSA-N	7.483
2683066-46-6	DVFM RMSQBLLRFB-HRNNMHK YSA-N	7.721
2683066-47-7	ZOSDANWQBTXOLM-OPXMRZJ TSA-N	7.854
2683066-48-8	JNFGOKUCIAOPPF-TUSQITKMSA-N	7.656
2683066-49-9	NALFXGDWJKEEGP-OYDLWJJNSA-N	7.146
2683066-35-3	NJKJFKJKSOABDF-IGKWTDBASA-N	7.770
161511-45-1	VMKPCENEHPWJT-DOPYIHRPSA-N	7.523
16280-27-6	BDZXSHDKBKYQKJ-UHFFFAOYSA-N	6.523
RATYOANKPNDVGM-UHFFFAOYSA-N	RATYOANKPNDVGM-UHFFFAOYSA-N	6.131
URMMIFIBZDAXOX-UHFFFAOYSA-N	URMMIFIBZDAXOX-UHFFFAOYSA-N	5.717
DVHDLRUJXOVRRY-UHFFFAOYSA-N	DVHDLRUJXOVRRY-UHFFFAOYSA-N	5.611
351010-72-5	OEYMPTGVGRACKD-JKSUJKDBSA-N	6.585
TZHYGMWJMVYHIL-UHFFFAOYSA-N	TZHYGMWJMVYHIL-UHFFFAOYSA-N	5.179
RSHHAZIWMJLBFQ-UHFFFAOYSA-N	RSHHAZIWMJLBFQ-UHFFFAOYSA-N	5.504
ZYTPLZKDUYQTPE-RRPNLBNLSA-N	ZYTPLZKDUYQTPE-RRPNLBNLSA-N	5.735
UAWJ9-36-1	XXWDDQJFYNJUCB-LREPDPJSA-N	7.292
UAWJ9-36-3	YVCYQGVRUNWVOI-WTOUUDQASA-N	7.268
95636-28-5	ZSQWAAHEKUTDPH-UHFFFAOYSA-N	6.656
75445-60-2	QPPBYEMNWPF RJ D-UHFFFAOYSA-N	6.793
67513-19-3	MHTRQSAKBZLLJM-UHFFFAOYSA-N	6.886
81418-42-0	QASXCIRROULNNX-UHFFFAOYSA-N	7.143
339096-59-2	RKERUHFQUMGLTH-UHFFFAOYSA-N	7.886
392732-12-6	NDDNJ OIBDROFPF-UHFFFAOYSA-N	6.959
1173776-71-0	CWAHMAWZJYIHRF-UHFFFAOYSA-N	6.347
302-79-4	SHGAZHPCJJPHSC-YCNIQYBTSA-N	4.607
2103278-26-6	SIAOKSJKWWLBHX-FHWLQOOXSA-N	7.108
2640359-56-2	HYHVR SBWXVDHDE-ACRUOGEOSA-N	7.229
2103278-37-9	HQBGJISGMNPOIV-BJLQDIEVSA-N	7.013
2103278-39-1	GMDMXLGBCBVPKJ-UFYCRDLUSA-N	7.097
2103278-67-5	TVJOQSLBWQGF KP-IPJJNNNSSA-N	7.187
2103278-60-8	FCEUKDIRLWHKPK-JTAQYXEDSA-N	6.854

Manuscript 1

Compound	InchiKey	pIC50
2103278-66-4	JKGAELNCICKZCD-LEDOBFOHSA-N	6.620
2103278-23-3	BLUNMHAXHLUNSH-HSQYWUDLSA-N	6.921
2103278-25-5	BRFDPJYHCJVJEP-ZJOUEHCJSA-N	7.469
2103278-24-4	GXGOVLGROSYUMC-XWUOBKMESA-N	7.174
2103278-27-7	AQHVFHFFDFXWGMJ-AAAXIEHDTSA-N	7.167
2640359-57-3	SFTZNDKMPHPRCR-HZLPDVBGSA-N	7.174
2640359-58-4	MNLRGPSJVOWSQR-ZCUSHOGSSA-N	6.745
CVD-0004255	SDIRCAZHIVHPY-UHFFFAOYSA-N	5.514
CVD-0004253	XBUJTNSECKROFV-UHFFFAOYSA-N	5.189
2629207-60-7	PHGKEUOLMGCCSU-UHFFFAOYSA-N	5.206
2629208-30-4	BNAGYCBNPOVAEF-UHFFFAOYSA-N	5.042
2629202-05-5	KCFLXQWWKKQMLY-UHFFFAOYSA-N	5.039
2629203-22-9	KBBXUVMNJAZEQI-UHFFFAOYSA-N	4.935
2629208-46-2	URCFWMRSFHGMSN-UHFFFAOYSA-N	4.780
2629203-91-2	NZIKWPPLRKAMJI-UHFFFAOYSA-N	4.502
CVD-0004229	QULOSEKXAYFLQG-UHFFFAOYSA-N	4.697
2629201-98-3	OHLKSHLIFVTZEX-UHFFFAOYSA-N	4.718
2629205-37-2	GUONEMIASAKEAJ-UHFFFAOYSA-N	4.696
CVD-0004210	XSPNJGFIWWEIKX-UHFFFAOYSA-N	4.647
CVD-0004248	VDNVZBAVOFLYGM-UHFFFAOYSA-N	4.544
CVD-0004236	XDFOKSZMVIQLAW-UHFFFAOYSA-N	4.508
CVD-0004251	FFFOTIFNUFRVNW-UHFFFAOYSA-N	4.371
CVD-0004237	NLPVUZFBIFDWRG-UHFFFAOYSA-N	4.409
2418650-36-7	FIGUSFMOHPETJW-UHFFFAOYSA-N	4.385
CVD-0004234	INVGZNWCNPHPRU-UHFFFAOYSA-N	4.302
CVD-0004232	NUOMPYHHXJKZMD-UHFFFAOYSA-N	4.253
2418649-19-9	PSLBTNLSDUNTPD-UHFFFAOYSA-N	4.282
CVD-0004262	SYWJSTOAQQLUHU-UHFFFAOYSA-N	4.183
2418649-92-8	IGHGFRRJIOKIDB-UHFFFAOYSA-N	4.178
CVD-0004424	CZGNSKRUKWZCCU-UHFFFAOYSA-N	4.172
CVD-0004239	NYYAATOUFFMYLO-UHFFFAOYSA-N	4.043
2418649-76-8	VFOVROUJSUVTRD-UHFFFAOYSA-N	4.010
Alpha-rubromycin	LQNGOIZVRFNQLO-UHFFFAOYSA-N	5.268
Cromoglicic acid	IMZMKUWMOSJXDT-UHFFFAOYSA-N	5.959
YHWWOAJOMPQXED-LEDIMMKISA-N	YHWWOAJOMPQXED-LEDIMMKISA-N	6.745
RCZBVOKMICIXNE-CRGUEAJTSA-M	RCZBVOKMICIXNE-CRGUEAJTSA-M	6.770
YHWWOAJOMPQXED-OXESIOJZSA-N	YHWWOAJOMPQXED-OXESIOJZSA-N	6.538
RCZBVOKMICIXNE-ZTBSZZGISA-M	RCZBVOKMICIXNE-ZTBSZZGISA-M	6.538
XMVDTWUKZMEFKE-FSPHOAMCSA-N	XMVDTWUKZMEFKE-FSPHOAMCSA-N	6.509
OHUZHLDAXYGKCN-SEWBHGOTSA-M	OHUZHLDAXYGKCN-SEWBHGOTSA-M	6.585
XMVDTWUKZMEFKE-GSCWJDCWSA-N	XMVDTWUKZMEFKE-GSCWJDCWSA-N	6.553
OHUZHLDAXYGKCN-JXBRSJJTSA-M	OHUZHLDAXYGKCN-JXBRSJJTSA-M	6.523
RPHGJUFZINUIBX-SWBMUHMYSAN	RPHGJUFZINUIBX-SWBMUHMYSAN	6.585
WZJYVJHLXCYWV-FV-OAOPQBDDSA-M	WZJYVJHLXCYWV-FV-OAOPQBDDSA-M	6.553
SUXXLPVLPGMTJN-SWBMUHMYSAN	SUXXLPVLPGMTJN-SWBMUHMYSAN	6.854
CDEICDIJDMTMS-OAOPQBDDSA-M	CDEICDIJDMTMS-OAOPQBDDSA-M	7.000
RZISQWRNJACRJT-OLGWUGKESA-N	RZISQWRNJACRJT-OLGWUGKESA-N	5.721
PVUSLPJJBDAJMSM-CYWORZNESA-M	PVUSLPJJBDAJMSM-CYWORZNESA-M	5.767
XEJUYBJCYVOVBW-YDUAIELLSA-N	XEJUYBJCYVOVBW-YDUAIELLSA-N	6.409
MUGMTSGNWGIKNT-KRHYPISQSA-M	MUGMTSGNWGIKNT-KRHYPISQSA-M	6.432
UYCGANVIMDDJSZ-UEUFYOKWSA-N	UYCGANVIMDDJSZ-UEUFYOKWSA-N	6.569
MADFDBAXBFAOMQ-ZYBIIAARSA-M	MADFDBAXBFAOMQ-ZYBIIAARSA-M	6.523
UYCGANVIMDDJSZ-LDWJCNMMSA-N	UYCGANVIMDDJSZ-LDWJCNMMSA-N	6.509

Compound	InchiKey	pIC50
MADFDBAXBFAOMQ-INYFNYJMSA-M	MADFDBAXBFAOMQ-INYFNYJMSA-N	6.509
LZAUUVQWXLSHGJJ-KRNMPOHKSA-N	LZAUUVQWXLSHGJJ-KRNMPOHKSA-N	6.745
UYPRMIJZKBTNNB-LWCMEJDOSA-M	UYPRMIJZKBTNNB-LWCMEJDOSA-M	6.824
LZAUUVQWXLSHGJJ-RPEAECGFSA-N	LZAUUVQWXLSHGJJ-RPEAECGFSA-N	6.538
UYPRMIJZKBTNNB-IWDBCACLSA-M	UYPRMIJZKBTNNB-IWDBCACLSA-M	6.569
YBRZKVAPDVWTRT-CSOMKMDOSA-N	YBRZKVAPDVWTRT-CSOMKMDOSA-N	6.569
PTOBNGNAWDQBAE-QQSQGYMKSA-M	PTOBNGNAWDQBAE-QQSQGYMKSA-M	6.602
HCMUAXJAEVVGND-ITLNLTKASA-N	HCMUAXJAEVVGND-ITLNLTKASA-N	6.086
OXFCTAAAUQBQAO-FKLBIKJRSA-M	OXFCTAAAUQBQAO-FKLBIKJRSA-M	5.987
HCMUAXJAEVVGND-WKZSNBNESA-N	HCMUAXJAEVVGND-WKZSNBNESA-N	6.131
OXFCTAAAUQBQAO-AGLQCOFBSA-M	OXFCTAAAUQBQAO-AGLQCOFBSA-M	6.108
Blonanserine	XVGOZDAJGBALKS-UHFFFAOYSA-N	4.606
Emodin	RHMXXJGYXNZAPX-UHFFFAOYSA-N	4.290
Ugonin J	VCPZVNMTXCALPA-INIZCTEOSA-N	6.027
Kaempferol	IYRMWMYZSQPKC-UHFFFAOYSA-N	4.664
Isokaempferide	VJJZJBUCDWKPLC-UHFFFAOYSA-N	4.395
LOR-NOR-30067bb9-18	JVRRUXPWYUPVTN-UHFFFAOYSA-N	5.421
LON-WEI-babf2c61-11	NQZPCTWJLLTANT-UHFFFAOYSA-N	5.321
MAR-UNI-c84db004-17	OEEOHKZVBKYMBA-UHFFFAOYSA-N	5.075
KAD-UNI-8a629cb0-21	SWYGVGJOJGQZBB-MHZLTWQESA-N	5.033
MAT-POS-43fe65f4-1	OIUXWYDUFACEFE-HNNXBMFYSA-N	4.991
LON-WEI-babf2c61-10	IIIMAJHNXHUPPG-UHFFFAOYSA-N	4.964
LOR-NOR-30067bb9-6	DJJXAFOCUYLUSE-UHFFFAOYSA-N	4.906
LOR-NOR-30067bb9-10	GAPUGZRYCWGUBE-UHFFFAOYSA-N	4.845
MED-COV-4280ac29-5	BGHJNYWPPPBXGZ-UHFFFAOYSA-N	4.818
WAR-XCH-b72a1bbc-11	NMONTLDKYFRAMZ-UHFFFAOYSA-N	4.802
LOR-NEU-c8f11034-7	BCWHJYKXPWMEAT-UHFFFAOYSA-N	4.777
LOR-NOR-30067bb9-2	PLPOPFVKSKBFJP-UHFFFAOYSA-N	4.762
LON-WEI-babf2c61-8	BNYQFEKMSAJYCY-UHFFFAOYSA-N	4.761
TAM-UNI-d1c3dd9f-18	KNNAVLOTACJOHO-UHFFFAOYSA-N	4.757
DAN-LON-a5fc619e-10	UJJWABRALJJPLY-IBGZPJMESA-N	4.725
ALP-UNI-3496895b-8	WROWCVDJFRRWBJ-NDEPHWFRSA-N	4.713
BAR-COM-0f94fc3d-41	YPZKHKYPLKRQED-UHFFFAOYSA-N	4.700
LON-WEI-babf2c61-4	SCTRRRLFFBAYPJ-UHFFFAOYSA-N	4.582
BEN-BAS-baaab58b-1	PBJNXFMTQKJPBJ-UHFFFAOYSA-N	4.564
WAR-XCH-b0339bbe-3	NJDGZUKDLATBMO-UHFFFAOYSA-N	4.527
LOR-NEU-c8f11034-3	MBVCESWADCIXJN-UHFFFAOYSA-N	4.519
WAR-XCH-b6889685-26	ZENPIZMBLVEDLM-UHFFFAOYSA-N	4.407
LOR-NOR-30067bb9-15	KYUDISFMFVEROE-UHFFFAOYSA-N	4.373
LOR-NOR-30067bb9-8	UTYQTGGBFCQOTL-UHFFFAOYSA-N	4.373
LOR-NEU-c8f11034-1	OCVKSIBWTJCXPV-UHFFFAOYSA-N	4.365
MAT-POS-02ae579f-2	WXWBKLXWXKNUBW-UHFFFAOYSA-N	4.351
LOR-NOR-30067bb9-13	CNUWGOYZCFOACA-UHFFFAOYSA-N	4.345
TAM-UNI-d1c3dd9f-12	HBVIEMZUPVVQQP-UHFFFAOYSA-N	4.298
JAN-GHE-83b26c96-18	GSEMEGPVJJVGNC-UHFFFAOYSA-N	4.238
MAR-UNI-c84db004-16	YAMFGQLZCQRUIX-UHFFFAOYSA-N	4.235
ALP-POS-79636100-2	MAJXFJOPQNQUFY-UHFFFAOYSA-N	4.207
ALP-POS-90c18d1d-1	GWXIPZCYRYXJTF-LJQANCHMSA-N	4.171
MAT-POS-b5746674-78	VVKODFKULNXOAE-UHFFFAOYSA-N	4.164
JAN-GHE-83b26c96-13	MUVSGYUDBWMCJG-UHFFFAOYSA-N	4.161
MAT-POS-ea426761-45	TVPQPADWXWREOA-UHFFFAOYSA-N	4.107
MAT-POS-b5746674-102	YVXKUWQUDAHOHS-UHFFFAOYSA-N	4.083
MAT-POS-b5746674-108	CYCSTOLONAFFHI-UHFFFAOYSA-N	4.035

Manuscript 1

Compound	InchiKey	pIC50
MAT-POS-2492181e-4	NBORDABFEIFAKS-UHFFFAOYSA-N	4.018
MAT-POS-2492181e-6	XPXRFBRETSHZLT-UHFFFAOYSA-N	4.008
STE-KUL-d79e3d6a-3	LRSOQVGVAVGVKR-OXQOHEQNESA-N	7.682
JOH-MSK-46727e7b-1	BSPZFJDYQHDZNR-ASPXRTSYSAN	7.555
MAT-POS-e194df51-1	YAYDSAMHSQCKIM-OAQYLSRUSAN	7.434
MAT-POS-e69ad64a-2	XLVAPPYYTCUGRF-UHFFFAOYSA-N	7.392
STE-KUL-d79e3d6a-2	DKOIBTIRJIPLGM-IBZIQKAZSAN	7.357
MAT-POS-f9802937-8	AXGVWFZIMFSNCN-CYBMUJFWSAN	7.324
EDJ-MED-1981ceba-4	JMLGBOOPVUWCHB-UHFFFAOYSA-N	7.295
MAT-POS-4223bc15-11	DHYOVGSVGPDDR-X-UHFFFAOYSA-N	7.288
EDJ-MED-1981ceba-3	BOBUOSCSQDEHLU-UHFFFAOYSA-N	7.226
MAT-POS-4223bc15-12	GJPHCXZDDOVJHH-UHFFFAOYSA-N	7.223
MAT-POS-ec6d90b7-6	RQXHXECUYIQDP-ZDUSSCGKSA-N	7.219
EDJ-MED-611d11e7-7	RMPFEPAZBNRHOP-NRFANRHFSA-N	7.213
MAT-POS-dc2604c4-1	YAYDSAMHSQCKIM-UHFFFAOYSA-N	7.192
MAT-POS-af1eef35-2	GLLFHKJANJSNEG-UHFFFAOYSA-N	7.189
MAT-POS-4223bc15-23	BNHZFFZRNBJNJB-UHFFFAOYSA-N	7.159
PET-UNK-bcc8fd08-1	FRAUGLSYTUOAF-ZNCSYKOHSA-N	6.918
PET-UNK-29afea89-2	FRAUGLSYTUOAF-FQEVSTJZSAN	6.918
EDJ-MED-1981ceba-2	DAHOAXWCUDTPIS-UHFFFAOYSA-N	7.051
PET-UNK-bcc8fd08-2	PAPXBUNGDDTTOJ-ZNCSYKOHSA-N	7.014
EDJ-MED-611d11e7-4	PAPXBUNGDDTTOJ-FQEVSTJZSAN	7.014
MAT-POS-a13804f0-3	NNQIVNFMEXUPCH-CYBMUJFWSAN	7.010
MAT-POS-f2460aef-1	WPJAOHFPXGCVSV-MHZLTWQESAN	6.691
MAT-POS-53907a1c-3	XZGGGAUTQFFNSC-UHFFFAOYSA-N	6.391
MAT-POS-96f51285-3	KZEBYUBBYZGMBN-UHFFFAOYSA-N	6.970
ALP-POS-02c6a514-44	URCIXQCRKXZPV-UHFFFAOYSA-N	6.951
MAT-POS-4223bc15-25	IDQZBINPDWFWFE-UHFFFAOYSA-N	6.945
EDJ-MED-1981ceba-5	WLYDPFRGDOGKP-UHFFFAOYSA-N	6.937
MAT-POS-96f51285-1	RMPFEPAZBNRHOP-UHFFFAOYSA-N	6.927
MAT-POS-090737b9-1	AMHRDRHRXYOOCU-UHFFFAOYSA-N	6.914
MAT-POS-993cdc78-1	FBLHAFJQTKYFLC-HXUWFJFHSA-N	6.905
MAT-POS-96f51285-6	RQXHXECUYIQDP-UHFFFAOYSA-N	6.883
MAT-POS-4223bc15-14	OVJUBIHYMLYNOD-UHFFFAOYSA-N	6.869
MAT-POS-61f37a1a-2	BEPOGYHLKHCDLE-JHRSTNGLSAN	6.864
EDJ-MED-1981ceba-1	TYKRJZJUJBHCQ-UHFFFAOYSA-N	6.862
MAT-POS-4223bc15-2	POHNQKSSAWVPGD-UHFFFAOYSA-N	6.857
MAT-POS-24589f88-3	QOWQBZLRTFIIMG-GOSISDBHSA-N	6.850
EDG-MED-10fcb19e-1	OANVIGJLLAVHQQ-UHFFFAOYSA-N	6.837
MAT-POS-4223bc15-1	ZTUNSYBLVPAFKR-UHFFFAOYSA-N	6.828
MAT-POS-ec6d90b7-5	NUXDKRDLDGPBW-LBPRGKRZSAN	6.823
MAT-POS-4223bc15-8	GMQADXIAHJPSAZ-UHFFFAOYSA-N	6.818
ALP-UNI-3735e77e-4	AXGVWFZIMFSNCN-UHFFFAOYSA-N	6.814
MAT-POS-4223bc15-40	LZTSPUJFXOCHAI-UHFFFAOYSA-N	6.809
EDJ-MED-e4b030d8-11	SGXOKACJPFKOBH-HXPMCKFVSA-N	6.796
MAT-POS-4223bc15-39	QYCQHFUQBZCSRT-UHFFFAOYSA-N	6.776
MAT-POS-dc2604c4-3	GXLIAZMEDWVSBX-UHFFFAOYSA-N	6.771
BEN-DND-c852c98b-4	XUFSEMPSIPKIN-UHFFFAOYSA-N	6.755
RAL-THA-05e671eb-10	NNQIVNFMEXUPCH-UHFFFAOYSA-N	6.541
EDG-MED-0e5afe9d-3	FRAUGLSYTUOAF-UHFFFAOYSA-N	6.680
MAT-POS-b3e365b9-1	QRUNNMAOPKEFMJ-OAHLLOKOSAN	6.666
MAT-POS-932d1078-3	RMXNRMBUTBCRQX-FQEVSTJZSAN	6.719
ALP-POS-fe871b40-16	CBVDIOCBIRWMPB-UHFFFAOYSA-N	6.708

Compound	InchiKey	pIC50
ALP-POS-3b848b35-2	LMTQPYQFHVEEMK-UHFFFAOYSA-N	6.748
MAT-POS-4223bc15-33	ZAXAQOGTEJZPRU-UHFFFAOYSA-N	6.691
ALP-POS-869ac754-1	ATZNGVMHQNZVPP-UHFFFAOYSA-N	6.598
MAT-POS-78e1d523-1	UTAUZKDGXZMBJU-UHFFFAOYSA-N	6.685
MAT-POS-dd3ad2b5-4	QOWQBZLRTFIIMG-UHFFFAOYSA-N	6.683
MIC-UNK-91acba05-1	TVJGSLPAQMOAIJ-UHFFFAOYSA-N	6.682
MAT-POS-4223bc15-6	QWMBVZGCWHTXFR-UHFFFAOYSA-N	6.675
MAT-POS-4223bc15-4	WDNZEVDIXFNUTD-UHFFFAOYSA-N	6.672
MAT-POS-4223bc15-30	AVYNUDKWZTW DST-UHFFFAOYSA-N	6.663
LON-WEI-4d77710c-76	FJTJSQTVHIUWGE-UHFFFAOYSA-N	6.736
RAL-THA-6e4c80cf-1	XFYXDYCVHVXIQS-UHFFFAOYSA-N	6.650
MAT-POS-4223bc15-41	NABYXCUJYBHZGW-UHFFFAOYSA-N	6.650
EDJ-MED-37aac4bd-4	RMXNRMBUTBCRQX-UHFFFAOYSA-N	6.648
MAT-POS-f9802937-7	FBODVAAXKKPILK-UHFFFAOYSA-N	6.539
MAT-POS-14f31916-1	CGIVTVFISUXBPO-UHFFFAOYSA-N	6.642
EDJ-MED-92e193ae-1	TUZBEGQKTWHIIP-OAHLLOKOSA-N	6.690
EDJ-MED-e4b030d8-2	GRKLXICDGD XNRS-SJKOYZFVSA-N	6.458
MAT-POS-4223bc15-5	NDQJXHMH SBCNTF-UHFFFAOYSA-N	6.634
MAT-POS-0c8fa4a7-1	DGPUYQMHP I IWJQ-UHFFFAOYSA-N	6.862
MAT-POS-ec6d90b7-7	VPVQGE OZXDR IEP-LBPRGKRZSA-N	6.631
RAL-THA-6e4c80cf-2	FBECVEGEKPHQNB-UHFFFAOYSA-N	6.626
ALP-UNI-8d415491-2	PZWQFDUSZJLSSR-UHFFFAOYSA-N	6.621
ALP-POS-fe871b40-2	FIGCWDOCORWPMS-UHFFFAOYSA-N	6.616
STE-KUL-d79e3d6a-1	BLJMV SBNQKASLS-MTWWTJSRSA-N	6.610
EDG-MED-5d232de5-5	NIZLDWOVHDTV KM-INIZCTEOSA-N	6.611
MAT-POS-4223bc15-7	VAEAAYNBAZNSCY-UHFFFAOYSA-N	6.609
MAT-POS-a5ef2d74-2	PAPXBUNGDDTTOJ-UHFFFAOYSA-N	6.609
ALP-UNI-8d415491-6	NUXDKRDLDDGPBW-UHFFFAOYSA-N	6.605
MAT-POS-dd3ad2b5-3	PVUJOYSOOMMEIC-UHFFFAOYSA-N	6.602
MAT-POS-64942dd0-3	NUTKWGCEX IJKRZ-UHFFFAOYSA-N	6.589
ALP-POS-477dc5b7-2	TUZBEGQKTWHIIP-UHFFFAOYSA-N	6.365
EDJ-MED-c314995a-1	DDMCNIBUQPRTBH-UHFFFAOYSA-N	6.296
MAT-POS-23a8a11a-1	OQQVTWZLQH BDCJ-UHFFFAOYSA-N	6.732
ALP-UNI-3735e77e-2	UETIOCMJDODDEB-UHFFFAOYSA-N	6.551
MAT-POS-4223bc15-31	JJFXISRP NUSPLB-UHFFFAOYSA-N	6.572
MAT-POS-dd3ad2b5-5	JODGLUXNHMXYIP-UHFFFAOYSA-N	6.565
MAT-POS-4223bc15-15	SNOKHRLNRHDJFL-UHFFFAOYSA-N	6.563
MAT-POS-4223bc15-18	GYXPARZCBIHWKB-UHFFFAOYSA-N	6.560
ALP-POS-b0bc6a46-5	FXMUCWQWZQMNSJ-UHFFFAOYSA-N	6.550
LON-WEI-4d77710c-8	KTMLBSZTILEAGY-UHFFFAOYSA-N	6.691
EDJ-MED-e4b030d8-13	CPQDCUNLQZNEBN-HXUWFJFHSA-N	6.795
PET-UNK-1901c25b-1	NWMNITATBGAXEW-UHFFFAOYSA-N	6.312
MAT-POS-6c284e65-1	UJFDJUBIAPUKQV-UHFFFAOYSA-N	6.508
MIC-UNK-91acba05-3	PFMVVUXLJ S OIKF-UHFFFAOYSA-N	6.541
MAT-POS-3ccb8ef6-1	SYSIGBRVBHDHN-QGZVFWFLSA-N	6.540
MAT-POS-f2460aef-3	YFVNUYLPSADEEV-UHFFFAOYSA-N	6.257
MAT-POS-61f37a1a-13	JHHRPPWGEXJGQX-UHFFFAOYSA-N	6.513
EDJ-MED-d08626de-3	GPTYSZBZQA EHFH-DXUWFJFHSA-N	6.424
MAT-POS-4223bc15-3	ZZQVZXHXOZIEHC-UHFFFAOYSA-N	6.508
MAT-POS-96f51285-5	VPVQGE OZXDR IEP-UHFFFAOYSA-N	6.503
MAT-POS-f2460aef-4	FLTATSNCGZWHJB-UHFFFAOYSA-N	6.395
EDG-MED-5d232de5-1	YVSNGUCAHVBF FA-CQSZACIVSA-N	6.534
ALP-POS-477dc5b7-1	QTESWALCLGKUKC-UHFFFAOYSA-N	6.514

Manuscript 1

Compound	InchiKey	pIC50
MAT-POS-e10a589d-1	VBWGHNNNSBLMGQY-UHFFFAOYSA-N	6.486
EDJ-MED-ce467fd5-1	NOAFNXVSUVZFJU-UHFFFAOYSA-N	6.482
MIC-UNK-50cce87d-3	VPQQORJHUJYTNP-UHFFFAOYSA-N	6.480
MAT-POS-61f37a1a-12	OKUSBOLTCQCKRF-UHFFFAOYSA-N	6.475
MAT-POS-86c60949-2	KLKGDNQJTAVAQU-UHFFFAOYSA-N	6.452
EDG-MED-5d232de5-4	OYAZCEPXBNBUAT-KRWDZBQOSA-N	6.436
VLA-UCB-1dbca3b4-15	QRUNNMAOPKEFMJ-UHFFFAOYSA-N	6.408
RAL-THA-2d450e86-6	FVAXWBDMWOKGB-UHFFFAOYSA-N	6.459
RAL-THA-2d450e86-14	LJRAPVIJDYXPLI-UHFFFAOYSA-N	6.389
MAT-POS-dd3ad2b5-2	VTQUGRCHBZGAGI-UHFFFAOYSA-N	6.417
ALP-POS-fe871b40-11	FBHAFJQTKYFLC-UHFFFAOYSA-N	6.406
MAT-POS-4223bc15-37	RKZPHVXSZMGJBL-UHFFFAOYSA-N	6.401
LON-WEI-4d77710c-2	CIMCKIQLVFYXTP-UHFFFAOYSA-N	6.702
ALP-POS-02c6a514-32	XSIUYZLFELGGKB-UHFFFAOYSA-N	6.395
MAT-POS-96f51285-2	WQGNBNAAPSMTNB-UHFFFAOYSA-N	6.372
JIN-POS-6dc588a4-22	HWRAFWHTDZDTRB-UHFFFAOYSA-N	6.372
MAT-POS-7ddaf7de-2	YZZABCCJOKORAD-UHFFFAOYSA-N	6.369
RAL-THA-2d450e86-17	RIBHDGSWGQEONR-UHFFFAOYSA-N	6.381
EDG-MED-ba1ac7b9-9	VZSKCSDUDDDIHW-UHFFFAOYSA-N	6.360
MIC-UNK-91acba05-6	NIZLDWOVHDTVKM-UHFFFAOYSA-N	6.299
LON-WEI-4d77710c-75	RKKBKFEJZACNKO-UHFFFAOYSA-N	6.268
MIC-UNK-45817b9b-1	WGORDSNJHZEQJN-UHFFFAOYSA-N	6.329
MAT-POS-a13804f0-4	RZBYEPRBZPLACH-XAGPSQNTSA-N	6.346
DAV-CRI-14a23e73-1	QXFHJCBKHQAQDM-UHFFFAOYSA-N	6.306
ALP-UNI-dbb9503d-1	DLQGJQRILXLRFE-UHFFFAOYSA-N	6.330
MAT-POS-3b97339c-2	RPJBIMKATYTDKW-LJQANCHMSA-N	6.325
MAT-POS-c74bc7b3-1	LNNHTGQQOQWKNR-HXUWFJFHSA-N	6.320
BEN-BAS-c2bc0d80-6	QQLVOMOIQPMUFC-UHFFFAOYSA-N	6.313
MAT-POS-4223bc15-17	ITZFRGSKFSHGCA-UHFFFAOYSA-N	6.312
ALP-POS-477dc5b7-5	SVWQDFXKODKDNW-UHFFFAOYSA-N	6.311
ALP-POS-6d04362c-2	DJHGLSHYLFLFFG-UHFFFAOYSA-N	6.356
DAR-DIA-5ff57136-9	VUFXBEOBILVVHI-GFOMBABLSA-N	6.252
EDG-MED-ba1ac7b9-11	PJTWJPJYCTTZBKW-UHFFFAOYSA-N	6.299
LON-WEI-ff7b210a-4	SNABBCSDVVWSGP-UHFFFAOYSA-N	6.191
LON-WEI-9739a092-1	DZQXYCUOULSTTG-UHFFFAOYSA-N	6.297
ALP-UNI-8d415491-4	OMEXTCWEXYMPTB-UHFFFAOYSA-N	6.283
DAR-DIA-6a508060-2	YWNGLZMXNZHQGK-UHFFFAOYSA-N	6.282
ALP-UNI-8d415491-5	LVPPKRIRMRVSJZ-UHFFFAOYSA-N	6.279
RUB-POS-1325a9ea-4	BSUXNZPXAUGDQM-UHFFFAOYSA-N	6.270
MED-COV-4280ac29-30	GZBIFDCOYSBXDA-UHFFFAOYSA-N	6.321
EDJ-MED-d08626de-5	GPTYSZBZQAEHFD-UHFFFAOYSA-N	6.196
EDJ-MED-12c115cc-1	VNMUCGFGSYBMK-OAHLLOKOSA-N	6.410
ALP-POS-02c6a514-50	FSZSSCCKOPNHHS-UHFFFAOYSA-N	6.254
BEN-BAS-c2bc0d80-7	JBTXEVNTEFVYSG-UHFFFAOYSA-N	6.250
RAL-THA-4aa06b95-7	NFLFPVHJTJCDNE-UHFFFAOYSA-N	6.127
RAL-THA-2d450e86-16	RXQWWUHGRIJKJHB-UHFFFAOYSA-N	6.172
MAT-POS-14f31916-3	CNRNRSYYOIMTNE-UHFFFAOYSA-N	6.244
MAT-POS-e6dd326d-6	XCYPEVNRWFVHLA-UHFFFAOYSA-N	6.242
ALP-POS-ce760d3f-1	CBPDYJATDXRIOM-UHFFFAOYSA-N	6.256
EDJ-MED-611d11e7-12	AFRPJKUVYRVGEC-LLVKDONJSA-N	6.236
EDG-MED-10fcb19e-3	VXDSZNRKIKUNBB-UHFFFAOYSA-N	6.226
JAG-UCB-a3ef7265-20	MMSFMHOFPDOWCH-UHFFFAOYSA-N	6.223
MAT-POS-e6dd326d-13	FHOISZYJCVIOOD-UHFFFAOYSA-N	6.220

Compound	InchiKey	pIC50
EDJ-MED-9e38fd34-1	FZPBWXZGXKTLK-UHFFFAOYSA-N	6.215
ALP-POS-ced8ea4d-12	SLIIUFPUFZWQJW-UHFFFAOYSA-N	6.215
EDG-MED-ba1ac7b9-10	DOHYITWAEMWMBS-UHFFFAOYSA-N	6.210
EDG-MED-ba1ac7b9-21	BEMREGHUUYXDNTR-UHFFFAOYSA-N	6.208
BEN-DND-93268d01-8	YHXAYSKGSZLMHI-UHFFFAOYSA-N	6.195
LON-WEI-9739a092-6	KIVBNAMDYXZGMT-UHFFFAOYSA-N	6.195
ALP-POS-e6e0c683-3	OTUDMKAJBQXUAD-UHFFFAOYSA-N	6.192
LON-WEI-4d77710c-73	SPAGBNWGOKCXAI-UHFFFAOYSA-N	6.390
MAT-POS-fc9ede84-1	YJHLFRCHQPNWFN-UHFFFAOYSA-N	6.186
ALP-POS-d0876c20-1	BUYQKSMLRBDWSQ-UHFFFAOYSA-N	6.185
ALP-POS-c3a96089-4	LTNIKFLJOUNFAI-UHFFFAOYSA-N	6.235
MAT-POS-8a69d52e-2	GRKLXICDGDGNRS-PXAZEXFGSA-N	5.872
LON-WEI-9739a092-10	BHQBDYRJFBGVOT-UHFFFAOYSA-N	6.167
VLA-UNK-cf7facf1-1	FSSOHAWMSBZQQQ-HXUWFJFHSA-N	6.174
MAT-POS-fa06b69f-6	BCSVXUUBZJSQEE-BENRWUELSA-N	6.154
JOH-UNI-0e1753c1-5	RXNNWOBJBKPHPT-INIZCTEOSA-N	6.156
PET-UNK-dd44aeb6-1	GEKAMRLLPCKGPA-UHFFFAOYSA-N	6.135
BRU-THA-92256091-17	CXQMRNJWEULDQY-UHFFFAOYSA-N	6.001
TAT-ENA-80bfd3e5-4	YYAXBTAUFRRRPP-UHFFFAOYSA-N	6.147
ADA-UCB-6c2cb422-1	VFGOVUIYBMUVAD-UHFFFAOYSA-N	6.057
MAT-POS-8695a11f-3	YJEUVMCEEMSKGZ-IBGZPJMESA-N	6.132
EDJ-MED-93390d0c-2	NSVYYDJNANPBPR-UHFFFAOYSA-N	6.128
EDJ-MED-fa7708b3-3	BIWPJTCDWZBOFO-UHFFFAOYSA-N	6.125
MAT-POS-8695a11f-2	FLVRFAVTGZINX-UHFFFAOYSA-N	6.121
MAT-POS-fce787c2-3	MTMXPPBZDNZMGK-UHFFFAOYSA-N	6.117
MAT-POS-14f31916-2	WSVBQTGCXSVXQG-UHFFFAOYSA-N	6.119
MAT-POS-f7918075-2	VNWMUCGFGSYBMK-UHFFFAOYSA-N	6.205
ALP-POS-696356e4-1	YVSNGUCAHVBFFA-UHFFFAOYSA-N	6.222
EDJ-MED-b7309adf-1	CXWLJEFNFIICLC-UHFFFAOYSA-N	6.114
RAL-THA-8416115c-5	ALSINHANKODPSI-UHFFFAOYSA-N	6.012
EDJ-MED-fa7708b3-2	UUESBJLAWFRMEQ-UHFFFAOYSA-N	6.106
DAN-LON-a5fc619e-3	RXNNWOBJBKPHPT-MRXNPFEDSA-N	6.197
ALP-POS-02c6a514-49	WUBAALIYHKNSBF-UHFFFAOYSA-N	6.097
EDJ-MED-8c98ee63-2	ZITSXTKDCXOESQ-UHFFFAOYSA-N	6.095
MAT-POS-fce787c2-8	PDHWWHKIMZBYRZ-UHFFFAOYSA-N	6.191
RAL-THA-8416115c-6	BSIBGMRLHQEJNZ-UHFFFAOYSA-N	6.057
VLA-UCB-50c39ae8-7	AMHRDRHRXYOOCQ-QGZVFWFLSA-N	6.104
MAT-POS-a5ef2d74-3	FXDATWAVSATAMS-UHFFFAOYSA-N	6.080
ALP-POS-88a7a97e-24	SZPJNXOUFFTIPT-OTRRWNHKSAN	6.077
EDJ-MED-b7309adf-2	NYAALPIGCNLRV-UHFFFAOYSA-N	6.077
MAT-POS-11b63608-1	NKEZIGFFPYIIL-UHFFFAOYSA-N	5.955
MAT-POS-e6dd326d-8	DWQRAXAUIMRQLB-UHFFFAOYSA-N	6.060
MAT-POS-de59a476-4	NBMVMUFILVGST-UHFFFAOYSA-N	6.035
MAT-POS-3b536971-1	YEJFDEVIYATGAM-UHFFFAOYSA-N	6.390
LON-WEI-4d77710c-67	ZEETYMQMBUWUTD-UHFFFAOYSA-N	6.276
JIN-POS-6dc588a4-13	WFMSCOMUABQZQN-UHFFFAOYSA-N	6.051
ALP-POS-95f71980-23	CEFARIWHARMJGO-UHFFFAOYSA-N	6.050
MAT-POS-4223bc15-24	RWZLDLVVBNSPHP-UHFFFAOYSA-N	6.048
EDJ-MED-00c1612e-1	RPTKBCKDMLWOT-UHFFFAOYSA-N	6.153
MAT-POS-61f37a1a-5	PCXPKBXCXYEGMW-UHFFFAOYSA-N	6.039
MAT-POS-4223bc15-38	XDQFHRKSDLPFIX-UHFFFAOYSA-N	6.037
LON-WEI-4d77710c-11	SEGJTGSMYFNGA-UHFFFAOYSA-N	6.252
ALP-POS-966f8da6-2	OYAZCEPXBNUAT-UHFFFAOYSA-N	6.042

Manuscript 1

Compound	InchiKey	pIC50
EDJ-MED-d1555997-1	JIOVCUKELUZZKX-UHFFFAOYSA-N	6.022
BEN-DND-c852c98b-5	ULOCCHXCKCKCRD-UHFFFAOYSA-N	5.903
TAT-ENA-80bfd3e5-1	MUWIOILLWYOCEU-UHFFFAOYSA-N	6.011
ALP-UNI-8d415491-1	SAWVWQKOOIRWCC-UHFFFAOYSA-N	6.010
EDG-MED-971238d3-1	QFIUTQBIZOZBNY-IBGZPJMESA-N	5.943
ALP-POS-ce760d3f-3	FJCJCAOYJHVYOI-ZDUSSCGKSA-N	6.059
RAL-THA-2d450e86-30	LPKOUSTWJUXVBF-UHFFFAOYSA-N	6.001
RAL-THA-4aa06b95-1	RKWNGNYSSGAJBH-UHFFFAOYSA-N	6.088
PET-UNK-dd44aeb6-2	ZLUZOFQKSMEEQC-UHFFFAOYSA-N	6.026
MAT-POS-14f31916-5	QJXRDKHIZVLHEP-UHFFFAOYSA-N	5.985
ALP-UNI-8d415491-3	RQFXRLLVVUPDMA-UHFFFAOYSA-N	5.984
RAL-THA-8416115c-1	DTQXKFGSEIHQQB-UHFFFAOYSA-N	5.890
MAT-POS-4223bc15-36	IDJINHQEGPIDBX-UHFFFAOYSA-N	5.972
MAT-POS-de59a476-2	HQIMBBKRSDUGNW-UHFFFAOYSA-N	5.992
MAT-POS-9db1e783-2	JEYBJKIBGGOUDS-UHFFFAOYSA-N	5.947
RAL-THA-2d450e86-8	VOYXLGCTTFJWHV-UHFFFAOYSA-N	5.987
MAT-POS-a5ef2d74-1	AFRPJKUVYRVGEC-UHFFFAOYSA-N	5.959
ERI-UCB-ce40166b-17	SNQHLYWSRJDLGK-UHFFFAOYSA-N	5.943
ALP-POS-95f71980-21	PBNGMKZNAWWFOY-UHFFFAOYSA-N	5.954
MAT-POS-24589f88-5	DFXXSAQHCVREDF-UHFFFAOYSA-N	5.952
ALP-UNI-b33a865d-1	GZEFUJZPAJUBC-UHFFFAOYSA-N	5.881
MAT-POS-8d5af1ef-1	NJMXMRVMIOEVON-UHFFFAOYSA-N	5.950
MAT-POS-fb82b63d-1	RTYGDBGOBVKIGM-UHFFFAOYSA-N	5.948
CHO-MSK-00c5269a-2	RVLKWCGLMRBJIF-UHFFFAOYSA-N	5.596
MAT-POS-4223bc15-16	HSICGNDPAFFAHN-UHFFFAOYSA-N	5.943
TAT-ENA-80bfd3e5-7	XMGGTELXPXRGJKE-UHFFFAOYSA-N	5.880
PET-UNK-c9c1e0d8-3	WATOKUBUEBNRTN-UHFFFAOYSA-N	6.039
ALP-POS-d9b6304e-1	SOAAHEUGOZFXMC-UHFFFAOYSA-N	5.935
MAT-POS-ea426761-9	PIBCKVSMRKHTLH-UKTHLTGXSA-N	5.613
ALP-POS-305f6ec3-14	NZCJAOIFBGIMMM-UHFFFAOYSA-N	5.932
NIR-THE-c331be7a-2	KKJBJMPUNLISGA-UHFFFAOYSA-N	5.886
MAT-POS-3b92565d-1	UFUGOWGYSILVPI-UHFFFAOYSA-N	5.895
ALP-POS-d0876c20-21	JIEOITNBUFLYIT-UHFFFAOYSA-N	5.917
PET-UNK-c9c1e0d8-4	RNKBROMGPBKNFK-UHFFFAOYSA-N	5.980
ALP-POS-e0fe77e5-13	GMEVLWIDIJTOMX-UHFFFAOYSA-N	5.889
MAT-POS-f9802937-3	KPAUUUFZYGHRE-HSZRJFAPSA-N	5.804
MIC-UNK-50cce87d-1	BLQARRAYSPMQLM-UHFFFAOYSA-N	5.884
ALP-POS-a2344586-3	DJWSIIHOQOUZJL-UHFFFAOYSA-N	5.882
VLA-UCB-29506327-1	FSSOHAWMSBZQQQ-UHFFFAOYSA-N	6.160
MAT-POS-9ff17035-3	BIKFVJZSUZPAE-UHFFFAOYSA-N	5.852
MAT-POS-24589f88-16	PPTMIVBWSXGWRF-UHFFFAOYSA-N	5.870
MAT-POS-fce787c2-2	UEJJRTXZTCTAIK-UHFFFAOYSA-N	5.982
MAT-POS-61f37a1a-14	HERXWNGKPDDPHP-UHFFFAOYSA-N	5.859
LON-WEI-9739a092-3	KDGKLIKOAXSJEW-UHFFFAOYSA-N	5.855
ALP-POS-82da25a3-1	PLYUGCGALYOYCH-UHFFFAOYSA-N	5.855
LON-WEI-4d77710c-10	YXJAEFWDSKWRJT-UHFFFAOYSA-N	6.074
RAL-THA-8416115c-13	OGUJARWRDDBIAJ-UHFFFAOYSA-N	5.903
ROB-IMP-e811baff-1	VEPQAUCCEYXBOH-UHFFFAOYSA-N	5.331
MAT-POS-7ddaf7de-3	PUBHXVXXNHSPKG-UHFFFAOYSA-N	5.842
ALP-UNI-3735e77e-1	CJNMPNBDTPWBLB-UHFFFAOYSA-N	5.840
VLA-UCB-05e51b3f-10	AMHRDRHRXYOOC-D-KRWDZBQOSA-N	5.831
MAT-POS-bbbbc21a-3	LHNWKYSEHJROHH-UHFFFAOYSA-N	6.074
JIN-POS-6dc588a4-11	XWEXHTMLQJSLGI-UHFFFAOYSA-N	5.833

Compound	InchiKey	pIC50
MAT-POS-86c60949-3	PNJIAFGGPCUFSE-UHFFFAOYSA-N	5.832
EDJ-MED-2f867453-1	GTVALCQWCWDFRA-UHFFFAOYSA-N	5.831
MAT-POS-fce787c2-6	RSRGRXAYHCGNSK-UHFFFAOYSA-N	5.827
RUB-POS-1325a9ea-14	YMBXBLIFBSFTRM-UHFFFAOYSA-N	5.825
LON-WEI-9739a092-9	FPNVOOOMGSKXIQ-UHFFFAOYSA-N	5.823
MAT-POS-4223bc15-22	TYTJKOJLSROVHC-UHFFFAOYSA-N	5.810
MAT-POS-96f51285-4	FZPAOKYJXGPGTG-UHFFFAOYSA-N	5.809
PET-UNK-bb7fe78-2	MWFXNZNKYUINCG-UHFFFAOYSA-N	5.809
MAK-UNK-7c9d1431-9	TYOQEDJEURTNRB-UHFFFAOYSA-N	5.808
MAT-POS-ea426761-13	UUMLXNZZKFQMMR-UHFFFAOYSA-N	6.122
LON-WEI-ff7b210a-5	IGPUJHRBDQDWSZ-UHFFFAOYSA-N	5.889
MAT-POS-54c4bf04-3	MFUBKAXERGMRJX-UHFFFAOYSA-N	5.937
ALP-POS-c59291d4-2	TWIVXCFEBRGEKY-UHFFFAOYSA-N	5.345
MAT-POS-ee51dedd-2	MBZYSHRESYQQNC-UHFFFAOYSA-N	6.051
KAD-UNI-80f122c8-3	MTHFEAGIPLVPM-UHFFFAOYSA-N	5.778
MAT-POS-8a69d52e-7	SGXOKACJPFKOBBCWTRNNRKSAN	5.752
RAL-THA-2d450e86-26	HVJFVWJFZCGHBBH-UHFFFAOYSA-N	5.766
MAT-POS-d8472c4f-2	JISJOVVMRJRBIW-UTCGOKLKSA-N	5.766
MAT-POS-24589f88-13	YEYRSZKXZCWDA-UHFFFAOYSA-N	5.762
MAT-POS-1f3f1a6f-4	FZQAFCWOUHYFCA-UHFFFAOYSA-N	5.758
ALP-POS-02c6a514-52	XGMHIXNFUOIGEJ-UHFFFAOYSA-N	5.750
LON-WEI-4d77710c-19	PNOHRUSRGMQJPS-CQSZACIVSA-N	5.843
VLA-UNK-82501c2c-2	ZIQTWXWOFHJVRA-UHFFFAOYSA-N	5.748
ALP-POS-d2866bdf-1	UEXLZKRRYKGPRD-UHFFFAOYSA-N	5.508
PET-UNK-bb7fe78-1	AJZWQKNTFPYK-UHFFFAOYSA-N	5.745
MAT-POS-f9802937-5	WFJMXJRCQHLHHH-JOCHJYFZSA-N	5.741
ALP-POS-9c80c481-1	ZVIGUBNCYLIRDY-UHFFFAOYSA-N	5.739
RAL-THA-2d450e86-12	BVFDAXUESDGGQS-UHFFFAOYSA-N	5.738
MAK-UNK-8be7dca9-1	WJCHRCGMUWDDOY-UHFFFAOYSA-N	5.737
VLA-UNK-82501c2c-1	IOZNTHKFTJIBSE-UHFFFAOYSA-N	5.734
RAL-THA-2d450e86-10	IMZAGSRTOVZPED-UHFFFAOYSA-N	5.732
MAT-POS-916a2c5a-1	AQTIYCHYBRWPPP-UHFFFAOYSA-N	5.871
PAU-UNI-8cdd41c7-1	QTUTYCLSGSFLSK-UHFFFAOYSA-N	6.023
JIN-POS-6dc588a4-14	CPHDKSOPZHQTJB-UHFFFAOYSA-N	5.722
EDJ-MED-bbe8e3da-2	UPODGWBKTLFDGB-UHFFFAOYSA-N	5.716
ALP-POS-b0bc6a46-10	BZADTLDBHNWZMX-UHFFFAOYSA-N	5.715
EDJ-MED-4c7486ba-9	DUXVLLPXVHPYTC-DQEYMECFSA-N	5.713
MAT-POS-24589f88-9	QEHFSTJYVNDTE-UHFFFAOYSA-N	5.711
BEN-DND-f2e727cd-5	SYSIIGBRVBHDHN-UHFFFAOYSA-N	5.631
SIM-SYN-f15aaa3a-1	LJSDPVAMDBJJSB-UHFFFAOYSA-N	6.059
MAT-POS-9ff17035-2	YPXGSTHCACVCPX-UHFFFAOYSA-N	5.704
MAT-POS-500f4700-2	GFVYBCHBOACIPT-UHFFFAOYSA-N	5.703
LEE-CAM-7ab9b158-3	YOMQHTOPSVGZNU-UHFFFAOYSA-N	5.702
EDJ-MED-4c7486ba-8	QLLUEKGKVUQPEV-DEOSSOPVSA-N	5.701
ALP-POS-5bb456a5-8	GXDHPHUFVWZLL-OLVCOQPYSA-N	5.701
NIR-THE-c331be7a-6	QVSGONMDHLKFLP-UHFFFAOYSA-N	6.038
BEN-DND-f2e727cd-6	VDQMWTFIKRGKRL-UHFFFAOYSA-N	5.697
MAT-POS-fce787c2-4	CKFWJQDYTZGUBQ-UHFFFAOYSA-N	5.696
EDJ-MED-6af13d92-3	NXFNSWGPJUOQPY-UHFFFAOYSA-N	5.688
RAL-THA-2d450e86-9	KJHYOZXTJLNKLS-UHFFFAOYSA-N	5.693
ALP-POS-d0876c20-5	AHCYNPHSHHAJHU-UHFFFAOYSA-N	5.687
LON-WEI-4d77710c-32	NQUCPOLKPQGXRI-UHFFFAOYSA-N	5.776
ALP-UNI-dbb9503d-2	YZGBUNKDXWNRHI-UHFFFAOYSA-N	5.682

Manuscript 1

Compound	InchiKey	pIC50
RUB-POS-1325a9ea-12	WTWQCPAXWFJLQF-UHFFFAOYSA-N	5.677
EDG-MED-ba1ac7b9-18	SPPHIOOHXLCTTK-UHFFFAOYSA-N	5.675
MAT-POS-173a45da-1	LUZOTWUPXCLDCX-UHFFFAOYSA-N	5.833
MAT-POS-d8472c4f-1	YJEUVMCEEMSKGZ-UHFFFAOYSA-N	5.673
ERI-UCB-a0b0dbcb-2	PYWVPVQQQANJLAT-UHFFFAOYSA-N	5.662
DAN-LON-a5fc619e-8	VSKDLORAYNFUSV-UHFFFAOYSA-N	5.924
MAT-POS-61f37a1a-1	XINQFXWLSISPMW-UHFFFAOYSA-N	5.664
ALP-POS-02c6a514-55	VWUCIEGKSAHXHU-UHFFFAOYSA-N	5.660
LON-WEI-9739a092-2	VTZQVNQSEIIUST-UHFFFAOYSA-N	5.659
MED-COV-4280ac29-37	PMYFITCYHWIAOC-UHFFFAOYSA-N	5.637
MAT-POS-24589f88-14	UJULDYYGTLAQFU-UHFFFAOYSA-N	5.655
LON-WEI-4d77710c-64	LCYWJQISYSPYAO-UHFFFAOYSA-N	5.684
MAT-POS-4223bc15-28	BDYIWAGQQMTVBW-UHFFFAOYSA-N	5.645
ALP-POS-a2344586-1	ZXQKARKVPSFSQW-UHFFFAOYSA-N	5.642
EDJ-MED-cf4b0d25-3	PUPNRPVHHRWKJX-XJDOXCRVSA-N	5.641
CHO-MSK-00c5269a-1	XEPQWAGFLLJHCF-UHFFFAOYSA-N	5.557
ALP-POS-88a7a97e-39	KRWOLCCLEWFBJG-UHFFFAOYSA-N	5.629
EDJ-MED-ee07cf00-18	VYUFPKYULHKDHH-UHFFFAOYSA-N	5.629
MAT-POS-8293a91a-2	RZBYEPRBZPLACH-YOQGVNIDSA-N	5.624
ALP-POS-88a7a97e-6	GQZXCXWKICAHQU-BGNYOVEWSA-N	5.623
MAT-POS-a3f7f96a-1	WLZJUTGEBNKHNR-UHFFFAOYSA-N	5.621
LON-WEI-4d77710c-74	YQWYXJYITQVIMI-UHFFFAOYSA-N	5.909
MAT-POS-24589f88-12	JCBMGTCFYFNELO-UHFFFAOYSA-N	5.614
ALP-POS-88a7a97e-16	IXPAUIYIETYLJJ-UHFFFAOYSA-N	5.611
DAR-DIA-0587064e-15	ZRSJGEZCFQSFAS-UHFFFAOYSA-N	5.604
MAT-POS-61f37a1a-11	MSSUYVMWFRRWKQ-UHFFFAOYSA-N	5.600
EDG-MED-ba1ac7b9-19	MQGPZHKWCMCBLL-UHFFFAOYSA-N	5.598
LON-WEI-2e27a2e5-1	WPJAOHFPXGCVSV-UHFFFAOYSA-N	5.970
MED-COV-4280ac29-15	ZMKLMLZZCDRNGW-UHFFFAOYSA-N	5.930
RAL-THA-8416115c-9	YXVMSWGLJCUMES-UHFFFAOYSA-N	5.595
MAT-POS-281d2ee9-1	ONBZJNTURWVQJU-UHFFFAOYSA-N	5.581
PET-UNK-bb7ffe78-3	CXVHATHIVLZSIH-UHFFFAOYSA-N	5.577
ALP-POS-88a7a97e-40	SZPJNXOUFFTIPR-UQVBRKOYSA-N	5.576
MIH-UNI-e573136b-5	CMTWIJXHGNMPL-UHFFFAOYSA-N	5.822
MAT-POS-e6dd326d-5	JJNUJZJHUUXGBO-UHFFFAOYSA-N	5.565
EDJ-MED-827e7cb4-1	CJNQWBCIUCYVDE-UHFFFAOYSA-N	5.562
DAR-DIA-fb20be43-7	ZEGKFIPTKWEFJK-UHFFFAOYSA-N	5.676
BEN-DND-c852c98b-1	MSLHVPCGUUCQAT-UHFFFAOYSA-N	5.561
EDJ-MED-ee07cf00-14	KLOPSAKEOLUBSS-UHFFFAOYSA-N	5.559
ALP-POS-88a7a97e-30	OTRBVEXVUIJRMP-UHFFFAOYSA-N	5.558
EDJ-MED-cf4b0d25-2	MSMKVLKCBNTHPS-NFJZVGLOSA-N	5.558
ALP-POS-5bb456a5-1	RZBKTFCAVJXIAX-UHFFFAOYSA-N	5.555
ALP-POS-88a7a97e-17	NZZHCOXGXARYLP-UHFFFAOYSA-N	5.553
MAT-POS-24589f88-11	GYQTZXUETADLMD-UHFFFAOYSA-N	5.551
TAT-ENA-80bfd3e5-32	GSFHOSYSNDVKNB-UHFFFAOYSA-N	5.466
PAU-WEI-5b8d5051-1	SDIRCAZHIVHPY-SANMLTNESA-N	5.535
MAT-POS-e9e99895-13	CDOWQSLSBXGJJG-UHFFFAOYSA-N	5.543
EDG-MED-70ae9412-1	BSTRMBRUYSEKM-UHFFFAOYSA-N	5.541
ALP-POS-d0876c20-17	XKEYBZZRFYMNFKC-UHFFFAOYSA-N	5.534
MAT-POS-5d65ec79-1	KPAUUUFZYGHRE-UHFFFAOYSA-N	5.534
ALP-POS-95f71980-26	OXMQDLBJPKYXDF-UHFFFAOYSA-N	5.530
EDG-MED-5d232de5-7	RNKBROMGPBKNFK-QGZVFWFLSA-N	5.529
MAT-POS-fb82b63d-3	NHDKKXCWRNZVNJ-UHFFFAOYSA-N	5.522

Compound	InchiKey	pIC50
EDJ-MED-ee07cf00-12	HZRXTJZXTSNCRY-UHFFFAOYSA-N	5.518
MAT-POS-f39f51fd-1	WKVJSMOQHXYAGM-UHFFFAOYSA-N	5.518
JOR-UNI-2fc98d0b-12	GCADPMUYKNHGSM-UHFFFAOYSA-N	5.498
MAT-POS-61f37a1a-10	DIFKEHQIIVEHOK-UHFFFAOYSA-N	5.510
BEN-DND-4f474d93-1	KCXBRLJFVUQUIR-CQSZACIVSA-N	5.509
DAR-DIA-0587064e-10	UQGLKJIMTUFOWG-UHFFFAOYSA-N	5.503
ALP-POS-c0c213c9-6	PYAHERHDUWOPJL-UHFFFAOYSA-N	5.499
LON-WEI-9739a092-5	AWCBIARJMKDDOD-UHFFFAOYSA-N	5.493
DAR-DIA-fb20be43-4	LEGWKCQAECEARJ-UHFFFAOYSA-N	5.639
MAT-POS-932d1078-4	MSNYXPXEMNEZGI-OAQYLSRUSA-N	5.492
TAT-ENA-80bfd3e5-36	HATLBPWOVVIHEI-UHFFFAOYSA-N	5.324
LON-WEI-4d77710c-70	PLYACWUKKBQQNP-UHFFFAOYSA-N	5.965
MAR-UCB-f313ec4d-2	SWRAJWUUCVFLBR-UHFFFAOYSA-N	5.364
MAT-POS-51833a24-1	IXUNYIMCNPQMGN-UHFFFAOYSA-N	5.487
LON-WEI-8f408cad-2	KNPLACWNGKQKQV-UHFFFAOYSA-N	5.532
ALP-POS-869ac754-2	DKWFAWGDLEBRQAX-UHFFFAOYSA-N	5.480
MAT-POS-5d65ec79-2	WFJMXJRCQHLHHH-UHFFFAOYSA-N	5.477
MAT-POS-e6dd326d-9	POYIPMIDTAGBFC-UHFFFAOYSA-N	5.476
MAT-POS-e9e99895-10	IMEKXLUOCJLETQ-UHFFFAOYSA-N	5.472
MAT-POS-7ddaf7de-4	PPXCCJYLJDOHSR-UHFFFAOYSA-N	5.471
MAT-POS-1f3f1a6f-1	WYMRTHFDSBYEJI-UHFFFAOYSA-N	5.471
ALP-POS-88a7a97e-13	LUYNSZZGDGRFLY-UHFFFAOYSA-N	5.470
PET-UNK-3c72d439-1	WBAYRCXMIZFJCT-UHFFFAOYSA-N	5.206
MAT-POS-4223bc15-26	BNNAPGIWIHPHRX-UHFFFAOYSA-N	5.467
MAT-POS-5369c344-7	UVCQZEJUCAAOCO-UHFFFAOYSA-N	5.462
MAT-POS-61f37a1a-4	OHDIYBCJUFVPVX-UHFFFAOYSA-N	5.458
NIR-THE-0d6461ce-8	HSLMOVJLEHUPFA-UHFFFAOYSA-N	5.677
ALP-POS-a3de0cb1-1	RGFAECRTGVORCK-UHFFFAOYSA-N	5.437
TRY-UNI-2eddb1ff-7	JCZDGTYSQGAPDN-UHFFFAOYSA-N	5.735
CHO-MSK-00c5269a-5	QVNRSGLDLIUOBK-UHFFFAOYSA-N	5.329
ALP-POS-88a7a97e-25	ZOYZTJHOEXZPAY-UHFFFAOYSA-N	5.438
LON-WEI-8f408cad-4	KOZXPZQYHAJRLQ-UHFFFAOYSA-N	5.565
EDJ-MED-e4b030d8-8	NKJSOSASDLIWEZ-OAHLLOKOSA-N	5.435
CHO-MSK-00c5269a-4	ZZVHWWZEQBNVHQ-UHFFFAOYSA-N	5.427
ALP-UNI-8e43a71e-16	IXXQKRILSLINRQ-UHFFFAOYSA-N	5.429
NAU-LAT-64f4b287-4	KBINLPLGLVLGVKA-UHFFFAOYSA-N	5.588
EDG-MED-ba1ac7b9-23	OFDISXHJSXTNLS-NUGJNUBVSAN	5.422
MAT-POS-6344a35d-1	ZJXCXDXGMIYGBJ-UHFFFAOYSA-N	5.710
KAD-UNI-80f122c8-2	XZDUBNODKZYWIO-UHFFFAOYSA-N	5.422
STU-CHA-6cae54d3-1	PPJDCJHMJYNHCW-UHFFFAOYSA-N	5.338
ERI-UCB-a0b0dbcb-4	NKHQYQSOEVSMTN-UHFFFAOYSA-N	5.334
ALP-UNI-c3ef0aba-3	BQABIPVMWKLWCS-UHFFFAOYSA-N	5.420
ALP-POS-88a7a97e-46	QAHQNZYUKISDPQ-UHFFFAOYSA-N	5.415
MAT-POS-e6dd326d-4	CEZQFZCAZILLII-UHFFFAOYSA-N	5.412
ALP-POS-88a7a97e-9	ICMQPSAZFLYNCL-UHFFFAOYSA-N	5.409
LON-WEI-9739a092-8	QWAHTESKWRFBSE-UHFFFAOYSA-N	5.408
ALP-POS-b0bc6a46-12	DJAHYHWLWJBXSK-UHFFFAOYSA-N	5.404
ALP-POS-88a7a97e-15	IQOFZCHTTCKOMV-UHFFFAOYSA-N	5.399
EDJ-MED-37aac4bd-2	MSNYXPXEMNEZGI-UHFFFAOYSA-N	5.074
MAT-POS-1f3f1a6f-3	GOKFUHGNQXJOGJ-UHFFFAOYSA-N	5.392
MAR-TRE-6a44bbf2-24	QJBKADADTCEHIA-UHFFFAOYSA-N	5.391
ALP-POS-88a7a97e-34	GIFDPZIHATTTCJ-UHFFFAOYSA-N	5.390
MAT-POS-e9e99895-3	BPWKRQZQRMWVCS-UHFFFAOYSA-N	5.388

Manuscript 1

Compound	InchiKey	pIC50
MAT-POS-e194df51-2	YAYDSAMHSQCKIM-NRFANRHFSA-N	5.386
GIA-UNK-20b63697-6	JDEVBATVXNYFFP-UHFFFAOYSA-N	5.788
LON-WEI-4d77710c-65	YPFACJFVYXDVFY-UHFFFAOYSA-N	6.083
MAT-POS-24589f88-1	RMPFPEAZBNRHOP-OAQYLSRUSA-N	5.378
TRY-UNI-714a760b-3	YFDTZFBVQPPSTPK-UHFFFAOYSA-N	5.230
ALP-POS-95f71980-4	GUZHPIDHMSDRPJ-UHFFFAOYSA-N	5.371
MAT-POS-f7918075-5	VQNYZMIKTVJRNS-UHFFFAOYSA-N	5.520
MAT-POS-1f3f1a6f-2	XCNSOFZGYUQECM-UHFFFAOYSA-N	5.369
ALP-POS-c59291d4-4	RMLYDSUPMGBCJP-UHFFFAOYSA-N	5.046
MAT-POS-5f1400cf-2	FLVRAXDXEKIINQ-UHFFFAOYSA-N	5.361
MAT-POS-9db1e783-3	VUFXBEOBILVVHI-BLGFXRMMMSA-N	5.361
MAT-POS-8293a91a-6	RZBYEPRBZPLACH-NTZARQNWWSA-N	5.348
ALP-POS-91609ae9-3	XMGGURAYXACURV-UHFFFAOYSA-N	5.439
EDJ-MED-50011917-1	PPBRUMHDHDMFJJ-UHFFFAOYSA-N	5.347
ERI-UCB-a0b0dbcb-12	AHSCMCZAWZBVOI-UHFFFAOYSA-N	5.340
DAR-DIA-fb20be43-6	KYSQSSMYYPVSQL-UHFFFAOYSA-N	5.636
MAT-POS-e9e99895-7	OCCECNEIZISYOAC-UHFFFAOYSA-N	5.333
LON-WEI-4d77710c-24	WHDLSUWGWXZWHR-UHFFFAOYSA-N	5.451
ALP-UNI-8e43a71e-2	ZRSVJHDPDCAAQO-UHFFFAOYSA-N	5.332
MAT-POS-3b536971-2	JBDonKQUEWDGLZ-UHFFFAOYSA-N	5.532
FRA-DIA-c7e803f4-1	DSHIHRHPWOORRQ-UHFFFAOYSA-N	5.297
MAT-POS-fce787c2-5	TXAXJHHNYJHGRH-UHFFFAOYSA-N	5.323
MED-COV-4280ac29-33	BXMVYJZFZGBLMT-UHFFFAOYSA-N	5.602
ALP-POS-305f6ec3-49	PTRDOWFMZUDTGK-UHFFFAOYSA-N	5.321
RAL-THA-2d450e86-13	VGTLCIFBGULFDX-UHFFFAOYSA-N	5.318
EDJ-MED-ee07cf00-4	SAQODVFGAOFELU-UHFFFAOYSA-N	5.318
MAT-POS-e9e99895-8	HZCWDQDHBUPETN-UHFFFAOYSA-N	5.317
ALP-POS-87c86d55-1	BZTBYCRNDCQCTD-UHFFFAOYSA-N	5.079
MAT-POS-afb6844f-4	ZCLSQNWMJQLIFL-UHFFFAOYSA-N	5.306
MAT-POS-b3e365b9-3	KJMNRTLNVIILOB-LBPRGKRZSA-N	5.250
RAL-THA-2d450e86-7	XYWOJIRBYMBSKX-UHFFFAOYSA-N	5.304
MAT-POS-a13804f0-1	AKMIAYKEPASZKQ-UHFFFAOYSA-N	5.302
EDG-MED-70ae9412-2	XTRUZCCFIYMEIT-UHFFFAOYSA-N	5.301
EDG-MED-971238d3-5	ILSGMAUMNXSNQB-FQEVSTJZSA-N	5.299
PET-UNK-372a2f82-1	OLHLDKCWBHUASQ-UHFFFAOYSA-N	5.162
ALP-POS-6d04362c-1	KOZCSUYCMLWBNN-UHFFFAOYSA-N	5.551
EDG-MED-ba1ac7b9-17	YVHJBYYDDXUZZBQ-UHFFFAOYSA-N	5.296
ALP-POS-88a7a97e-27	JGWMHGXGUQNOYSO-UHFFFAOYSA-N	5.288
MAT-POS-199e2e7c-1	NJHYHXQHRGWKMO-UHFFFAOYSA-N	5.310
ALP-POS-88a7a97e-36	RHFQHFVDVXQNSIS-UHFFFAOYSA-N	5.279
ALP-POS-02c6a514-46	BQPQIXCOIPUANU-UHFFFAOYSA-N	5.277
RAL-THA-1d44ff04-8	IROCFXWUODMOSA-UHFFFAOYSA-N	5.276
MAT-POS-7ddaf7de-1	JNXYXEFKJGHSML-UHFFFAOYSA-N	5.274
ALP-POS-88a7a97e-37	SZPJNXOUFFTIPR-NQXVELOOSA-N	5.274
ALP-POS-02c6a514-9	FXQLAAYABUFWKJ-UHFFFAOYSA-N	5.273
MAT-POS-06036648-5	NZEBZGMXAYXWMP-UHFFFAOYSA-N	5.125
ALP-POS-a0a4abd7-1	GPXJYKOZZUYXCY-UHFFFAOYSA-N	5.271
ALP-POS-88a7a97e-41	OJISJWVOUHNWKS-UHFFFAOYSA-N	5.270
MAT-POS-5f1400cf-1	UXQXQCWWBHBNOA-UHFFFAOYSA-N	5.270
EDJ-MED-cf4b0d25-5	YWWMTUUYUDUMGZ-UHFFFAOYSA-N	5.267
ALP-POS-88a7a97e-26	PJYUEHDFDXBPM-UHFFFAOYSA-N	5.266
MAR-UCB-f313ec4d-6	UIQBSVLMOPGDAQ-UHFFFAOYSA-N	5.202
ALP-POS-02c6a514-17	YHHPYQQITSJHCM-UHFFFAOYSA-N	5.255

Compound	InchiKey	pIC50
DAR-DIA-0587064e-12	IYYXIIGOPZZEPN-UHFFFAOYSA-N	5.249
LON-WEI-4d77710c-66	YBQZYGXKIKXKH-UHFFFAOYSA-N	5.763
MED-COV-4280ac29-31	OXJKQVHSPFERIM-UHFFFAOYSA-N	5.643
EDJ-MED-611d11e7-6	VPVQGEQZXDRIEP-GFCCVEGCSA-N	5.239
MAT-POS-e9e99895-2	LWLQXUAOFURQPB-UHFFFAOYSA-N	5.237
STE-KUL-d79e3d6a-5	OFFBWMVPMMERSJ-GDDXUJBHSA-N	5.237
MAT-POS-bfb445d4-2	OYXJREXTFZDGLT-UHFFFAOYSA-N	4.872
MAT-POS-f2460aef-5	BOYAMDIEACWWDH-UHFFFAOYSA-N	5.275
TAT-ENA-80bfd3e5-45	DKNPWKVKVOUSDJK-UHFFFAOYSA-N	5.234
ALP-POS-88a7a97e-49	MDDQRNGTLRONLR-UHFFFAOYSA-N	5.231
TAT-ENA-80bfd3e5-42	KGOAKHKFHXXKAMA-UHFFFAOYSA-N	5.229
EDJ-MED-827e7cb4-7	JTSDHELQJXQKHL-UHFFFAOYSA-N	5.227
MAT-POS-61f37a1a-7	FLZOKXCUXZXZOB-UHFFFAOYSA-N	5.224
JOH-UNI-ea1df7a8-1	LAMPSQTVGNGTCW-UHFFFAOYSA-N	5.328
TAT-ENA-80bfd3e5-37	YIHQTDMLDQZILW-UHFFFAOYSA-N	5.218
EDG-MED-ba1ac7b9-16	PRAMQJXVZJSNCS-UHFFFAOYSA-N	5.214
BEN-DND-c852c98b-10	MLMSDXNATONIOS-UHFFFAOYSA-N	5.210
EDG-MED-ba1ac7b9-15	ARURGDWYZCELGB-UHFFFAOYSA-N	5.209
LON-WEI-8f408cad-5	CNYRPLHEARDSQA-UHFFFAOYSA-N	5.344
ALP-POS-88a7a97e-48	IOKHMQAFXRLQEE-UHFFFAOYSA-N	5.205
ALP-POS-88a7a97e-19	UTGFQINFHOBBJI-UHFFFAOYSA-N	5.204
LON-WEI-4d77710c-69	NVDDPKBABKGJNB-UHFFFAOYSA-N	5.932
ALP-POS-ced8ea4d-32	FRXMHZHXACZFFE-UHFFFAOYSA-N	5.199
ALP-POS-3fc1724e-1	GOIVLPNYJWWFAT-UHFFFAOYSA-N	4.988
EDG-MED-ba1ac7b9-13	QQWBTWVCIPHEIH-BPGSNBDMISA-N	5.194
MAT-POS-24589f88-8	PCXJSSFZCKFKCU-UHFFFAOYSA-N	5.194
MAT-POS-5369c344-5	GYZZUYOBWQICSG-UHFFFAOYSA-N	5.192
PET-UNK-8922bd3c-1	OCDIEQRRHRVJBQ-UHFFFAOYSA-N	5.166
ALP-POS-6d04362c-3	HRRQJRBCBDNWBP-UHFFFAOYSA-N	5.192
MAT-POS-edabd372-2	SYVPDBBNVDUWBZ-UHFFFAOYSA-N	5.179
ALP-POS-5bb456a5-7	GXDPHPUFBVWZLL-UHFFFAOYSA-N	5.175
AGN-NEW-891393a6-1	SQQGVRHLDXZSJR-UHFFFAOYSA-N	5.497
ALP-POS-88a7a97e-23	SZPJNXOUFFTIPR-ZQLHYOEWSA-N	5.166
ALP-POS-88a7a97e-43	BWIYLCNBORKVOV-UHFFFAOYSA-N	5.160
RAL-THA-2d450e86-31	RLXPIIDJMBKGLO-UHFFFAOYSA-N	5.158
MAT-POS-fce787c2-1	MEEAZNUAAIDWKR-UHFFFAOYSA-N	5.157
MAT-POS-bb423b95-2	AFQAFSYTYCJOEN-UHFFFAOYSA-N	5.180
ALP-UNI-c3ef0aba-1	KHRYPQZVLWTRHW-UHFFFAOYSA-N	5.153
ALP-POS-6d96567b-2	UBRPTXKQWBAFHO-UHFFFAOYSA-N	5.146
ALP-POS-95b75b4d-5	TVYREULMZIPGLM-UHFFFAOYSA-N	5.138
ALP-POS-88a7a97e-5	FRSFZWYUOPENKL-UHFFFAOYSA-N	5.144
EDJ-MED-ee07cf00-6	KSYXKDYSTCOWMM-UHFFFAOYSA-N	5.143
MAT-POS-916a2c5a-2	IOKPKVSTRDCKNE-UHFFFAOYSA-N	5.301
BEN-DND-76ad4ac9-13	KHFBRRKYEIYOTB-UHFFFAOYSA-N	5.140
MAT-POS-14ad9fe9-1	CVTLEGSZWVPMPP-UHFFFAOYSA-N	5.266
ALP-UNI-8e43a71e-10	SWYGVGJOJGQZBB-UHFFFAOYSA-N	5.138
LON-WEI-4d77710c-61	STJUKLCGPTUUSK-UHFFFAOYSA-N	5.659
MAT-POS-e9e99895-6	CDPYVHPXDFRJSH-UHFFFAOYSA-N	5.135
EDJ-MED-cf4b0d25-1	GLYKCQXPAQOKNP-UHFFFAOYSA-N	5.135
BEN-DND-4f474d93-2	FJCJCAOYJHVYOI-CYBMUJFWSA-N	5.127
BEN-DND-7e92b6ca-12	DLURFVIIIGNLAAB-UHFFFAOYSA-N	4.914
MED-COV-4280ac29-13	UIWDEDSTMCQHLE-UHFFFAOYSA-N	5.397
BRU-CON-c4e3408a-1	KJMNRTLNVIILOB-UHFFFAOYSA-N	5.109

Manuscript 1

Compound	InchiKey	pIC50
ALP-POS-88a7a97e-29	CIMVMIMVXWHOCA-UHFFFAOYSA-N	5.102
MAT-POS-bb423b95-1	ITQMQTJYJVFFAW-UHFFFAOYSA-N	5.265
EDG-MED-ba1ac7b9-12	PWLHKSOQBQSNIM-UHFFFAOYSA-N	5.098
PET-UNK-8df914d1-2	GRCPKASYWFJZEJ-UHFFFAOYSA-N	5.095
BEN-DND-f2e727cd-4	LANOLLUVYHFGJP-UHFFFAOYSA-N	5.093
ALP-POS-c3a96089-1	WGIVEHFTGVGGES-UHFFFAOYSA-N	4.979
ALP-POS-88a7a97e-38	IYSFAGLTQKWOOB-UHFFFAOYSA-N	5.087
JAN-GHE-5a013bed-2	YHLCUCKEOZVUDL-UHFFFAOYSA-N	5.225
ALP-POS-ced8ea4d-16	PJHNCDMUUHAWTR-UHFFFAOYSA-N	5.079
ALP-POS-95f71980-25	VLQUGEUPDVOABA-UHFFFAOYSA-N	5.075
ALP-POS-d0876c20-19	HJRRRJUDFMAQRX-UHFFFAOYSA-N	5.074
MAT-POS-4223bc15-21	MVXPUDNKNQUUDE-UHFFFAOYSA-N	5.074
JIN-POS-6dc588a4-21	QUYQRVGZEMEFBI-UHFFFAOYSA-N	5.074
MAT-POS-f42f3716-1	XLUOUCSKIFQLV-UHFFFAOYSA-N	5.311
MAT-POS-1f3f1a6f-5	QYSQMAWAWFDUIX-UHFFFAOYSA-N	5.073
ALP-POS-88a7a97e-31	YSTGZYMKNLTUESF-UHFFFAOYSA-N	5.073
MAT-POS-916a2c5a-3	WKFNJCXVHWMQRU-UHFFFAOYSA-N	5.249
MAT-POS-e9e99895-5	VNZZPEPBUTUXBE-UHFFFAOYSA-N	5.070
MAT-POS-24589f88-4	MNEMEPBVUPTVMK-UHFFFAOYSA-N	5.069
MAT-POS-24589f88-15	OACUIVYIKIVPHQ-UHFFFAOYSA-N	5.068
JAN-GHE-5a013bed-3	VAYRFIMVASFMKQ-UHFFFAOYSA-N	5.063
BEN-DND-7e92b6ca-10	CARZEXKCGVUISS-UHFFFAOYSA-N	4.892
ALP-POS-88a7a97e-3	UMJMEZWWUYWJL-UHFFFAOYSA-N	5.056
BEN-DND-76ad4ac9-9	DEPIHNLTFXPDAF-UHFFFAOYSA-N	5.055
MAT-POS-e9e99895-11	MDXYQARUSPNBMR-UHFFFAOYSA-N	5.054
MAT-POS-f42f3716-5	ZPHJPHDQVDZROB-UHFFFAOYSA-N	4.932
KAD-UNI-80f122c8-1	PPXPBVMYMHCBZQ-UHFFFAOYSA-N	5.043
EDJ-MED-cf4b0d25-4	TWJHVIHWLJPSHA-UHFFFAOYSA-N	5.040
ALP-POS-5bb456a5-3	OLQFESLXZBSXOR-UHFFFAOYSA-N	5.037
ERI-UCB-d6de1f3c-2	OMBIOAUKLYASJI-UHFFFAOYSA-N	5.147
NIR-THE-c331be7a-1	UCMXTNBXERXONW-UHFFFAOYSA-N	4.989
EDG-MED-ba1ac7b9-22	PTXZFELDLUFZII-UHFFFAOYSA-N	5.033
UNK-CYC-68f84b31-67	DWQPUNLINDZEJ-UHFFFAOYSA-N	5.032
EDJ-MED-ee07cf00-16	PFQHGBCEXNBSTD-UHFFFAOYSA-N	5.030
PET-UNK-c9c1e0d8-2	CLLYBOMGFSNDSO-UHFFFAOYSA-N	5.028
ALP-POS-42f3de95-3	DJKLEAQONGAKOU-UHFFFAOYSA-N	5.024
MAT-POS-a3f7f96a-8	MDXGMRQDZMDNOI-UHFFFAOYSA-N	5.024
ALP-POS-88a7a97e-32	DVVNJLDZAKWTBH-UHFFFAOYSA-N	5.021
MAT-POS-61f37a1a-8	JTRAWPKSEVSNTB-UHFFFAOYSA-N	5.019
EDG-MED-ba1ac7b9-20	MOLYEONJAYVSLN-UHFFFAOYSA-N	5.007
ALP-POS-64a710fa-1	COUGPWQMAYMVGH-UHFFFAOYSA-N	5.263
ALP-POS-88a7a97e-44	XKKFENBTDQAABL-UHFFFAOYSA-N	5.003
ALP-UNI-8e43a71e-8	WROWCVDJFRRWBJ-UHFFFAOYSA-N	5.001
EDJ-MED-ee07cf00-15	GSGQFGIQQLCNS-UHFFFAOYSA-N	4.989
MAT-POS-e9e99895-4	JYNWFFMHMBDISO-UHFFFAOYSA-N	4.985
ALP-POS-b0bc6a46-24	GDVJEWWRXFVOOF-UHFFFAOYSA-N	4.983
ALP-POS-305f6ec3-11	QAXXYFLLJOHGPE-UHFFFAOYSA-N	4.979
MAT-POS-45b13633-2	HBVMSYLLJFZUYAS-UHFFFAOYSA-N	4.979
DUN-NEW-f8ce3686-22	IGYKULZPTHKSHF-UHFFFAOYSA-N	5.228
EDJ-MED-6af13d92-1	FVKXXNFROYGAEF-UHFFFAOYSA-N	5.279
ALP-POS-88a7a97e-33	YMXHHTCLQZWEII-UHFFFAOYSA-N	4.968
ALP-POS-02c6a514-53	BEBXTRSDRKUMEW-UHFFFAOYSA-N	4.961
RAL-THA-eb6cb89c-2	FBTMEABCOKXEU-UHFFFAOYSA-N	4.955

Compound	InchiKey	pIC50
ALP-POS-f13221e1-2	SXTDMNODKPPWQE-UHFFFAOYSA-N	5.086
ALP-POS-02c6a514-56	XKSATFPPIVWXRL-UHFFFAOYSA-N	4.950
WIL-MOD-03b86a88-6	NLTJDKXVUTASZ-UHFFFAOYSA-N	4.894
EDJ-MED-ee07cf00-1	BZVKRMJCTWUAMA-UHFFFAOYSA-N	4.874
JIN-POS-6dc588a4-9	AGRSKVFVSLDNFF-UHFFFAOYSA-N	4.945
EDJ-MED-8c98ee63-1	ICBOLVNQRIKQMG-UHFFFAOYSA-N	4.796
BEN-DND-f2e727cd-1	ZILZTRUBCRJCMW-UHFFFAOYSA-N	4.944
MAT-POS-a3f7f96a-9	YTRHZBYARFZPTB-UHFFFAOYSA-N	4.938
MAT-POS-45b13633-3	CLOIMUFARJKWGD-UHFFFAOYSA-N	4.938
TRY-UNI-2eddb1ff-6	FOACOGDXINRMHQ-UHFFFAOYSA-N	5.000
ALP-POS-6d04362c-4	PQVLYZJUGXIBNY-UHFFFAOYSA-N	4.803
MAT-POS-a3f7f96a-7	XUPGSFXHGWNLQS-UHFFFAOYSA-N	4.933
MAT-POS-24589f88-7	KBLBPXJQWCZLSF-UHFFFAOYSA-N	4.930
LON-WEI-9739a092-4	GXAWOVNPQXJWJZ-UHFFFAOYSA-N	4.927
EDJ-MED-976da9a6-1	MGLFKWBXQLXODC-UHFFFAOYSA-N	4.966
LON-WEI-9739a092-7	UEGYGPZUHJQJWRF-UHFFFAOYSA-N	4.924
EDJ-MED-611d11e7-5	NUXDKRDLDGDPBW-GFCCVEGCSA-N	4.919
ALP-POS-b0bc6a46-19	TYPFQXXQFLKXLN-UHFFFAOYSA-N	4.918
MAT-POS-24589f88-10	ROVYWUSZIYZDMG-UHFFFAOYSA-N	4.917
MAT-POS-61f37a1a-6	DTYVPXGUZYPHHQ-UHFFFAOYSA-N	4.917
STU-CHA-7b3b91d5-1	LSMNUGUNNUAPSA-MRXNPFEDSA-N	4.806
BEN-DND-76ad4ac9-10	SZRVCMBDBFMEQS-UHFFFAOYSA-N	4.916
RAL-THA-2d450e86-33	RMUFMFDOQHTPNN-UHFFFAOYSA-N	4.912
MAT-POS-e9e99895-9	MIIMNSJUHLVLA-UHFFFAOYSA-N	4.911
MAT-POS-24589f88-2	QOWQBZLRTFIIMG-SFHVURJKSA-N	4.909
MAT-POS-24589f88-6	NOSTYJJMTBEVMJ-UHFFFAOYSA-N	4.901
MAT-POS-f42f3716-3	PNMKEGJHDLIGFG-UHFFFAOYSA-N	4.909
MAT-POS-2db6411e-2	AWAQBATWGYJNHH-UHFFFAOYSA-N	5.347
MAT-POS-2bb0cf2b-2	BKYVVDKZFPPWFC-UHFFFAOYSA-N	4.885
ALP-POS-95f71980-24	YGHHBDDFTHZIFG-UHFFFAOYSA-N	4.880
RAL-THA-2d450e86-23	HVJOKHNKBZMQSO-UHFFFAOYSA-N	4.880
MAT-POS-f7918075-8	NXBRVTFHRPPFLN-UHFFFAOYSA-N	4.988
EDJ-MED-ee07cf00-10	DKSLTJIIHBAOKF-UHFFFAOYSA-N	4.865
LON-WEI-0a73fcb8-7	XLONNNGKPLHXDC-UHFFFAOYSA-N	4.521
DUN-NEW-f8ce3686-23	GNIFKHXEAFAXDF-UHFFFAOYSA-N	4.963
RAL-THA-2d450e86-1	IBEFUXAVGLACIF-UHFFFAOYSA-N	4.854
JAN-GHE-5a013bed-1	AJOFHXMHBPDFHQ-UHFFFAOYSA-N	4.492
WAR-XCH-79d12f6e-6	YKYWKHQEHNAJT-UHFFFAOYSA-N	5.318
MAT-POS-61f37a1a-3	ADLBIVOVUUDWEX-UHFFFAOYSA-N	4.845
MAT-POS-61f37a1a-9	CDMJXMMOQORACL-UHFFFAOYSA-N	4.844
MAT-POS-ea426761-53	KPUXXANPHIWRFP-UHFFFAOYSA-N	4.843
ALP-POS-88a7a97e-20	LDXVDPOFAXGNFU-UHFFFAOYSA-N	4.842
ALP-POS-42f3de95-2	JRQPEUUGSKDFGE-UHFFFAOYSA-N	4.840
EDG-MED-ba1ac7b9-27	KYYYVADSEVGPIQY-UHFFFAOYSA-N	4.838
RAL-THA-6b94ceba-13	SJZQDLPQMTZPX-UHFFFAOYSA-N	4.859
BEN-DND-7e92b6ca-11	HWQXSFIOFYVQC-UHFFFAOYSA-N	4.426
MAK-UNK-af83ef51-6	JXLGZKTTWILFHL-UHFFFAOYSA-N	5.090
MAT-POS-044491d2-6	YPFFNUZDAIMEIX-UHFFFAOYSA-N	4.587
MAT-POS-54c4bf04-4	VXCZIEOLFXENOB-UHFFFAOYSA-N	4.831
STE-KUL-d79e3d6a-4	JIIINDHMQARXKOK-GZNUUUAQSA-N	4.819
ALP-POS-ced8ea4d-67	HJGBGEHHJAYHCI-UHFFFAOYSA-N	4.811
MAT-POS-f2460aef-2	WPJAOHFPXGCVSV-HHHXNRCGSA-N	4.823
BEN-DND-76ad4ac9-11	DQIHRAOMUCROJG-UHFFFAOYSA-N	4.806

Manuscript 1

Compound	InchiKey	pIC50
RAL-THA-e002e396-1	LVFVBKQULJRERK-UHFFFAOYSA-N	4.806
MAT-POS-e10a589d-4	YIOYEPDFEQAXPP-UHFFFAOYSA-N	5.173
ERI-UCB-d6de1f3c-1	BJJZCSKOWCBJFK-UHFFFAOYSA-N	4.909
ALP-POS-305f6ec3-52	SOEFAJVCFRAYCA-UHFFFAOYSA-N	4.798
PET-UNK-7374c256-2	NAUKDTRKBNSCDS-UHFFFAOYSA-N	4.420
WIL-MOD-03b86a88-4	BYFAHRYDVSZUMV-UHFFFAOYSA-N	4.806
TRY-UNI-714a760b-19	MDGSUONMRGWOAQ-UHFFFAOYSA-N	4.643
MAT-POS-590ac91e-41	YFXMEQJREOQWSW-UHFFFAOYSA-N	4.673
BEN-DND-7e92b6ca-14	YZERBHLVPUBWNC-UHFFFAOYSA-N	4.530
RAL-THA-6b94ceba-12	QSEQGWMIJUKJZ-UHFFFAOYSA-N	4.636
ALP-POS-b0bc6a46-17	SLSSRDLDJYCON-UHFFFAOYSA-N	4.779
MAT-POS-932d1078-5	MSNYXPXEMNEZGI-NRFANRHFSA-N	4.772
EDG-MED-5d232de5-3	OYAZCEPXBNUAT-QGZVFWFLSA-N	4.765
BEN-DND-76ad4ac9-14	VEKHPVRMLNEAJV-UHFFFAOYSA-N	4.764
JAN-GHE-83b26c96-8	QJMYJFPUHDEEBX-UHFFFAOYSA-N	5.011
DUN-NEW-f8ce3686-24	LNZPSYUTCDEHI-UHFFFAOYSA-N	4.850
ALP-POS-305f6ec3-46	UICSECGSYFZTTK-UHFFFAOYSA-N	4.754
SAM-UNK-2684b532-12	LTODPBLUNCYZHX-UHFFFAOYSA-N	4.741
WAR-XCH-b0339bbe-19	ITTXFESNHRYVTJ-UHFFFAOYSA-N	4.784
ALP-POS-02c6a514-45	XRDWVUPYWPTTRA-UHFFFAOYSA-N	4.740
EDJ-MED-ee07cf00-9	PUFNGCQTRCODIU-UHFFFAOYSA-N	4.598
EDJ-MED-ee07cf00-5	VRHSXYQSHKEFBU-UHFFFAOYSA-N	4.728
ALP-POS-305f6ec3-38	HPDVWFGGHGBPBO-UHFFFAOYSA-N	4.727
WIL-MOD-03b86a88-2	ZTNNOENPPVCJEI-UHFFFAOYSA-N	4.482
MAT-POS-ee51dedd-3	OAOTYZBHACZHLK-UHFFFAOYSA-N	4.715
MAT-POS-f42f3716-7	OLNZSLRRSLAEOL-UHFFFAOYSA-N	4.755
MAT-POS-fce787c2-11	SWOKAOBZNNNGSGO-UHFFFAOYSA-N	4.714
MAT-POS-8293a91a-7	RZBYEPRBZPLACH-NHIDMIJISA-N	4.713
MAT-POS-8e4737f4-1	CPQDCUNLQZNEBN-FQEVSTJZSA-N	4.665
MAT-POS-bb423b95-5	MXJNXNCRYRDVPRY-UHFFFAOYSA-N	4.644
ALP-POS-305f6ec3-37	LSXHQZYVYKSHHK-UHFFFAOYSA-N	4.710
EDJ-MED-ee07cf00-17	UWWQOIKCCVRMFA-UHFFFAOYSA-N	4.710
STU-CHA-83dacf15-1	NZJQITRTVLSFEK-UHFFFAOYSA-N	4.607
MAT-POS-ee51dedd-1	IGUSJGIDMUBABY-UHFFFAOYSA-N	4.809
MAT-POS-044491d2-1	SXRYZSBUMPNSOT-UHFFFAOYSA-N	4.558
ALP-POS-6d04362c-5	QMSHHMMACMYJRL-UHFFFAOYSA-N	4.722
EDJ-MED-e4b030d8-3	GRKLXICDGDGNRS-YVEFUNNKSA-N	4.520
BEN-DND-76ad4ac9-12	NUNBVNSEDQNJIZ-UHFFFAOYSA-N	4.695
MAK-UNK-af83ef51-2	VPPOHECJIPUOJL-UHFFFAOYSA-N	4.780
MAT-POS-f9802937-4	KPAUUUFZYGHRE-QHCPKHFHSA-N	4.689
VLA-UCB-34f3ed0c-11	FSSOHAWMSBZQQQ-FQEVSTJZSA-N	4.688
JAN-GHE-83b26c96-12	ZTLMWKAKNVXEJW-UHFFFAOYSA-N	4.742
CHO-MSK-6e55470f-7	KOXYBCASPNTQIV-UHFFFAOYSA-N	4.780
EDG-MED-0e5afe9d-1	YNHUBGVMTXFOMC-UHFFFAOYSA-N	4.685
GIA-UNK-20b63697-2	DWCGNWLFDFOFSP-UHFFFAOYSA-N	5.228
MAT-POS-f9802937-2	YLSYWDUPJWRNFZ-UHFFFAOYSA-N	4.683
MAT-POS-c7771779-1	WTGNNXIIBVYQGH-UHFFFAOYSA-N	4.683
ALP-POS-f13221e1-3	YNKVVMYWRZOLAS-UHFFFAOYSA-N	4.698
MAT-POS-450cb4f9-1	QXMZFFYSZUSZDT-UHFFFAOYSA-N	4.680
AAR-POS-d2a4d1df-27	VXKRRLAIQCSGHI-UHFFFAOYSA-N	4.677
EDJ-MED-ee07cf00-8	WRNJYSYBCQEAPV-UHFFFAOYSA-N	4.641
ALP-UNI-8e43a71e-7	VFLCJUJQOPYLJP-UHFFFAOYSA-N	4.672
MAT-POS-281d2ee9-3	ZSBKUCQNXRTRNY-UHFFFAOYSA-N	4.670

Compound	InchiKey	pIC50
JIN-POS-6dc588a4-3	LRJBZROURSRBNZ-UHFFFAOYSA-N	4.667
MAT-POS-932d1078-1	QJTMCPJHBMXIA-UHFFFAOYSA-N	4.666
EDJ-MED-50fe53e8-1	ATNZGZHLHMEMFV-UHFFFAOYSA-N	4.665
MAT-POS-ea426761-52	MJHCWEXYAIJKMH-UHFFFAOYSA-N	4.663
WIL-MOD-03b86a88-5	NJLSSHJXUZPKHO-UHFFFAOYSA-N	4.662
RUB-POS-1325a9ea-13	UBADBBNAESQMGM-UHFFFAOYSA-N	4.661
BRU-THA-92256091-16	CRWDSMFFLKVOTQ-UHFFFAOYSA-N	4.655
ALP-POS-ced8ea4d-19	KWZOVLTJZSARM-UHFFFAOYSA-N	4.654
BRU-THA-92256091-80	REJFBYXBUOSADI-UHFFFAOYSA-N	4.644
MAT-POS-f42f3716-8	WLFQCYKMUSGMFP-NDENLUEZSA-N	4.729
EDG-MED-0da5ad92-20	YMFZNLWSSHHIRN-UHFFFAOYSA-N	4.640
JIN-POS-6dc588a4-6	QOGVQOIEARJILE-UHFFFAOYSA-N	4.640
MAT-POS-044491d2-2	QHVGFWBWFFNMBS-UHFFFAOYSA-N	4.614
MAT-POS-ea426761-49	CDXLHNKOWAMLPC-UHFFFAOYSA-N	4.637
MAT-POS-ec6d90b7-4	PAPXBUNGDDTTOJ-HXUWFJFHSA-N	4.632
RAL-THA-2d450e86-19	JTKPUZPWCDULOI-UHFFFAOYSA-N	4.628
RAL-THA-2d450e86-2	ACXVUTMBJYJIOO-UHFFFAOYSA-N	4.625
MAT-POS-1e5f28a7-1	LOFUGDOYUSJHEA-UHFFFAOYSA-N	4.696
EDJ-MED-ee07cf00-3	XDVVVMGUPNDBE-UHFFFAOYSA-N	4.624
ALP-POS-ced8ea4d-20	FLUFIIUXSOKYNJ-UHFFFAOYSA-N	4.624
CHO-MSK-6e55470f-5	ISRHBBIYTNRSMU-UHFFFAOYSA-N	4.618
PET-UNK-b1ef24dc-1	WLHXPTOEEMUXEC-UHFFFAOYSA-N	4.621
BEN-DND-7e92b6ca-16	ONQQLMKVMVPYKH-UHFFFAOYSA-N	4.620
CHR-SOS-7098f804-10	CNVXVFXVGMPROS-UHFFFAOYSA-N	4.616
ALP-POS-305f6ec3-41	BSHWEVACTVWMCC-UHFFFAOYSA-N	4.616
MAT-POS-a3f7f96a-6	QLQAFEAWFXQMMH-UHFFFAOYSA-N	4.612
ALP-POS-ced8ea4d-30	GCBOAYQZGNACKE-UHFFFAOYSA-N	4.611
MAT-POS-86c60949-1	CXWLJEFNFIICLC-HNNXBMFYSA-N	4.610
MAT-POS-53907a1c-2	BYXFIRCJOPCCSU-UHFFFAOYSA-N	4.610
TRY-UNI-714a760b-6	ZROFBWFSOGSOQG-UHFFFAOYSA-N	4.733
RAL-THA-2d450e86-37	HOEDDRRLZQIJMK-UHFFFAOYSA-N	4.607
EDG-MED-0da5ad92-7	BNLQGVZVZNCOKF-UHFFFAOYSA-N	4.705
RAL-THA-2d450e86-32	OPHPFMNEMZRFFW-UHFFFAOYSA-N	4.605
MAT-POS-5369c344-6	LUZCWVXVLDWXRAN-UHFFFAOYSA-N	4.603
RAL-MED-2de63afb-14	SFHIIQYDZORTHQ-UHFFFAOYSA-N	4.603
ALP-POS-87c86d55-2	MBOCIJFRGXBBRS-UHFFFAOYSA-N	4.470
ALP-POS-88a7a97e-4	LUNCSISTTZQMTN-UHFFFAOYSA-N	4.597
MAT-POS-044491d2-3	ROXLXWPALVDCDC-UHFFFAOYSA-N	4.651
ALP-POS-ce760d3f-4	CJFRFTLFKMMUSPS-HNNXBMFYSA-N	4.590
ALP-POS-e980f4ea-32	OFVJDCHKAIRYNL-UHFFFAOYSA-N	4.588
JAG-UCB-52b62a6f-19	NCVYFRLPLSKVCP-UHFFFAOYSA-N	4.297
ALP-POS-305f6ec3-1	IGEMYIKWFUZTSN-UHFFFAOYSA-N	4.582
TRY-UNI-714a760b-18	WNGGAWMEODUSED-UHFFFAOYSA-N	4.735
ALP-POS-568d9d36-1	VVABZIZCFLMRHI-UHFFFAOYSA-N	5.070
DAR-DIA-23aa0b97-19	MNTWRMFSZRVDHS-UHFFFAOYSA-N	4.704
EDJ-MED-12c115cc-2	VNWMUCGFGSYBMK-HNNXBMFYSA-N	4.570
EDJ-MED-611d11e7-8	RQXHXECUYIQDP-CYBMUJFWSA-N	4.565
ALP-POS-ced8ea4d-35	JTFZSOHLZFKYHF-UHFFFAOYSA-N	4.564
ERI-UCB-a0b0dbcb-11	VCHAPWLUGINYTM-UHFFFAOYSA-N	4.563
MAT-POS-916a2c5a-4	NZVVMCRLJDCU-UHFFFAOYSA-N	4.772
MAK-UNK-7c9d1431-17	KADYJZKMMKSOLG-UHFFFAOYSA-N	4.560
GIA-UNK-a79af1bc-1	SAVXKRAOMBRMTM-UHFFFAOYSA-N	4.635
LON-WEI-4d77710c-51	YMWAYXNOWZLQDU-UHFFFAOYSA-N	4.558

Manuscript 1

Compound	InchiKey	pIC50
ALP-POS-ced8ea4d-11	BUKGGEYRDKYROU-UHFFFAOYSA-N	4.558
ALP-POS-95f71980-28	ANIVJGSJFINECR-UHFFFAOYSA-N	4.555
DAV-CRI-3edb475e-6	NBYHXZHPJRULHZ-UHFFFAOYSA-N	4.616
MAT-POS-d8472c4f-4	RDLMQTIEQKQNMJ-UHFFFAOYSA-N	4.551
ALP-POS-95f71980-3	SXTGLRCYPWOPRX-UHFFFAOYSA-N	4.546
EDJ-MED-ee742daa-4	SUENFCMPEYZUDH-UHFFFAOYSA-N	4.540
ALP-POS-305f6ec3-2	PQHUGDUYNBFUPL-UHFFFAOYSA-N	4.536
ALP-POS-88a7a97e-10	QGXXCVWFZDKXPT-UHFFFAOYSA-N	4.534
RAL-THA-2d450e86-38	KNVLVTWBEZVOQM-UHFFFAOYSA-N	4.530
LON-WEI-4d77710c-44	BJENSXWOGUUEKX-UHFFFAOYSA-N	4.528
MED-COV-4280ac29-25	YAXUQSCFTRMAAW-UHFFFAOYSA-N	4.494
EDJ-MED-49816e9b-1	VKQLRXLEZSWDDW-UHFFFAOYSA-N	4.552
RAL-THA-2d450e86-11	HKGRIOGEDMBBDB-UHFFFAOYSA-N	4.516
ALP-POS-b0bc6a46-22	FRJIZPGLFJUFBY-UHFFFAOYSA-N	4.516
MAT-POS-78e1d523-2	WDFPFZJIPSBMOZ-UHFFFAOYSA-N	4.513
RAL-THA-58fba2bc-1	BHCFZKZVZCXPAP-UHFFFAOYSA-N	4.503
MAT-POS-f42f3716-4	HHEGLDWZMHOEHA-UHFFFAOYSA-N	4.543
MAT-POS-f42f3716-6	KWOZOQHUQWBGBH-UHFFFAOYSA-N	4.373
MAT-POS-afb6844f-1	BKFTUQOQWCYOPU-UHFFFAOYSA-N	4.490
ALP-POS-95f71980-14	ISYQGYABPRUSER-UHFFFAOYSA-N	4.489
ALP-POS-cd485364-2	KFDIFJWRZNKQEO-UHFFFAOYSA-N	4.377
BEN-DND-7e92b6ca-13	MNPRGBHSSXXPMF-UHFFFAOYSA-N	4.485
MAT-POS-3ccb8ef6-2	SYSIIGBRVBHDHN-KRWDZBQOSA-N	4.482
MAT-POS-590ac91e-40	VYMXSDTVXIFDGZ-UHFFFAOYSA-N	4.482
ALP-POS-95f71980-33	WFOKUGVDOYWGSG-UHFFFAOYSA-N	4.478
ALP-POS-d0876c20-4	ZXGJRTJTMHZAPS-UHFFFAOYSA-N	4.476
WAR-XCH-79d12f6e-4	HUWSBYFJZNFUCU-UHFFFAOYSA-N	4.604
ERI-UCB-a0b0dbcb-13	CRMCDYBMJIZOOL-UHFFFAOYSA-N	4.466
MAT-POS-932d1078-2	RMXNRMBUTBCRQX-HXUWFJFHSA-N	4.466
RAL-MED-2de63afb-2	BBXRTGNKDKIMPV-UHFFFAOYSA-N	4.464
MAT-POS-bb423b95-7	XFUDGCDECOIROA-UHFFFAOYSA-N	4.390
RAL-THA-2d450e86-39	AKWMESZDJUMPL-UHFFFAOYSA-N	4.456
BRU-THA-92256091-86	IAYQGBDEHLWTGX-UHFFFAOYSA-N	4.450
BAR-COM-4e090d3a-6	ZKNOWBIAEIEKLM-UHFFFAOYSA-N	4.447
ALP-POS-ced8ea4d-37	CVMGGVMKMCBKDJ-UHFFFAOYSA-N	4.439
DAR-DIA-0587064e-19	SFYBYVPPWUOOQP-UHFFFAOYSA-N	4.437
MAT-POS-044491d2-7	UELOTNVORDVOMK-UHFFFAOYSA-N	4.460
MAT-POS-c9973a83-1	WXCJKPMYIYFEDR-UHFFFAOYSA-N	4.397
RAL-THA-6b94ceba-2	CGZHAQKIMHZAPY-UHFFFAOYSA-N	4.261
NEH-REV-107bcf72-5	LQSMXXIHSSFBKQ-UHFFFAOYSA-N	4.730
ALP-POS-02c6a514-34	JZSZCXIRJJBCSS-UHFFFAOYSA-N	4.426
EDG-MED-ba1ac7b9-28	QEVMDZIWEFGMAM-UHFFFAOYSA-N	4.425
AAR-POS-8a4e0f60-7	AUTHSSSKYCIITDF-UHFFFAOYSA-N	4.422
JIN-POS-6dc588a4-20	ZGJBVFVWERWPHL-UHFFFAOYSA-N	4.420
ERI-UCB-ce40166b-6	CHVVOHNIGAZFET-UHFFFAOYSA-N	4.564
EDJ-MED-d08626de-1	FRAUGLSYTUOAAF-HXUWFJFHSA-N	4.414
ALP-POS-c0c213c9-1	NIAFAWJPFVWDCR-UHFFFAOYSA-N	4.452
CHR-SOS-7098f804-14	INMXZXUMJHQYMO-UHFFFAOYSA-N	4.408
ALP-POS-95f71980-11	UKBNJSPBXCXYHOB-UHFFFAOYSA-N	4.406
ALP-POS-a3de0cb1-16	OFMUKPPIQPQGGE-UHFFFAOYSA-N	4.404
EDG-MED-5d232de5-8	RNKBROMGPKBNFK-KRWDZBQOSA-N	4.404
ALP-POS-305f6ec3-59	ACRBJLNCDELWOB-UHFFFAOYSA-N	4.401
TAM-UNI-d1c3dd9f-5	WTWQIUHSHKVLGJ-UHFFFAOYSA-N	4.406

Compound	InchiKey	pIC50
MAT-POS-afd4d4fd-2	YZQKBSHQXMHNSJ-UHFFFAOYSA-N	4.391
ALP-POS-477dc5b7-3	CPOMOUZWGWPYOA-UHFFFAOYSA-N	4.209
ALP-POS-0c2c77e1-1	VSOZJTYJJHNDPD-UHFFFAOYSA-N	4.494
ALP-POS-0c0a3b93-1	HKFCOEPSJJKAKE-UHFFFAOYSA-N	4.388
BRU-THA-92256091-20	NYRIRZYSBXGAPF-UHFFFAOYSA-N	4.386
RAL-THA-6b94ceba-5	XDVRTJWPDGZLAS-UHFFFAOYSA-N	4.320
JAG-UCB-a3ef7265-3	BWTWLKMAKBVNDU-GFCCVEGCSA-N	4.554
ALP-POS-b0bc6a46-16	JZXMWXQMFWCKGG-UHFFFAOYSA-N	4.366
MAT-POS-0bc33984-1	BDDAUIXRVHLSPG-UHFFFAOYSA-N	4.359
BEN-DND-03406596-8	FSSOUDWKRDPMFL-UHFFFAOYSA-N	4.357
ALP-POS-305f6ec3-21	UWZUUTBLGHYWCV-UHFFFAOYSA-N	4.356
LON-WEI-4d77710c-39	KPNLTDFHLPMBTA-UHFFFAOYSA-N	4.352
EDJ-MED-e66f2cc5-1	JNMYIISRJKNRPB-UHFFFAOYSA-N	4.620
ALP-POS-95f71980-35	YRGKHXSBDLPGV-UHFFFAOYSA-N	4.346
JAN-GHE-5a013bed-11	HFAVJGNWJPFAGR-UHFFFAOYSA-N	4.344
TRY-UNI-2eddb1ff-2	OCRRKPZGCOGESB-UHFFFAOYSA-N	4.333
ALP-POS-2da19ca7-2	GXDPHPUFBVWZLL-JKDDQTOVSA-N	4.332
EDG-MED-971238d3-4	RPJBIMKATYTDKW-IBGZPJMESA-N	4.323
BRU-THA-92256091-77	PNHPGDNJETUHJB-UHFFFAOYSA-N	4.321
ALP-POS-ced8ea4d-8	ZWIRFTZOXMUTSY-UHFFFAOYSA-N	4.319
EDJ-MED-3c65e9ce-4	BZNRCHXVMKBNNF-UHFFFAOYSA-N	4.308
MAT-POS-f9802937-6	WFJMXJRCQHLLHHH-QFIPXVFZSA-N	4.304
MAT-POS-b3e365b9-2	QRUNNMAOPKEFMJ-HNNXBMFYSA-N	4.267
ALP-POS-ced8ea4d-27	IAUUHGWQXDVIAS-UHFFFAOYSA-N	4.299
ALP-POS-a3de0cb1-13	LKAZBEZWQPCSAW-UHFFFAOYSA-N	4.190
ALP-POS-95f71980-2	CJAMVKLPJGKUJT-UHFFFAOYSA-N	4.292
ALP-POS-95f71980-16	AOJLRYIZMACLCL-UHFFFAOYSA-N	4.290
RAL-THA-6b94ceba-7	FLSUUSNRZPAOBF-UHFFFAOYSA-N	4.323
MAT-POS-8a69d52e-3	GRKLXICDGDGNRS-SJCJKPOMSA-N	4.283
MED-COV-4280ac29-12	YDDPBPJVBZUEOJ-UHFFFAOYSA-N	4.282
RAL-THA-2d450e86-34	CZLYCQXGUITFRL-UHFFFAOYSA-N	4.279
JOH-UNI-0e1753c1-2	FRETUPOIWATUEC-UHFFFAOYSA-N	4.276
EDJ-MED-e58735b6-4	BRBVXYZNRUYZKX-UHFFFAOYSA-N	4.274
NAU-LAT-64f4b287-5	ZWQIUHDNTGABHU-UHFFFAOYSA-N	4.272
EDG-MED-0da5ad92-2	JXVSAHJMMLDNGG-UHFFFAOYSA-N	4.270
ALP-POS-8b8a49e1-6	ZNDORTNMEWZIKY-UHFFFAOYSA-N	4.270
ALP-POS-95f71980-27	OJVLUTWBCIOPLX-UHFFFAOYSA-N	4.269
DAR-DIA-23aa0b97-8	AXUCLULHXCDGDW-UHFFFAOYSA-N	4.265
EDG-MED-5d232de5-6	NIZLDWOVHDTVKM-MRXNPFEDSA-N	4.264
MIC-UNK-66895286-3	BYGPKLOEYHRCEP-UHFFFAOYSA-N	4.263
ALP-POS-b0bc6a46-20	AGRSSHJIIRXABU-UHFFFAOYSA-N	4.261
TRY-UNI-714a760b-22	QZIZMNNUMXCPSV-UHFFFAOYSA-N	4.302
ALP-POS-b0bc6a46-23	HIOYWZXSFGAGAK-UHFFFAOYSA-N	4.250
DAR-DIA-23aa0b97-20	XBRJTBLYIAYTOZ-UHFFFAOYSA-N	4.293
RAL-THA-2d450e86-3	RLHWHJQBWCGSCF-UHFFFAOYSA-N	4.234
BEN-DND-93268d01-6	RBRNVPIASJICKN-UHFFFAOYSA-N	4.233
BAR-COM-0f94fc3d-48	SHDRXRMTUQTxFE-UHFFFAOYSA-N	4.316
ALP-POS-ced8ea4d-24	NDFPLCAJWCYRNK-UHFFFAOYSA-N	4.218
EDJ-MED-d08626de-4	GPTYSZBZQAEHFD-FQEVSTJZSA-N	4.184
ALP-POS-95f71980-15	CVTSVTWEHBDKFW-UHFFFAOYSA-N	4.214
TRY-UNI-714a760b-20	OTFRLGRHXMKEAF-UHFFFAOYSA-N	4.285
JAG-UCB-52b62a6f-10	OGUSAQXXDSEGCM-UHFFFAOYSA-N	4.242
ALP-POS-b0bc6a46-14	FANONCNPAKTPAJ-UHFFFAOYSA-N	4.203

Manuscript 1

Compound	InchiKey	pIC50
EDG-MED-0e5afe9d-2	CAAKDEVMDLUSTN-UHFFFAOYSA-N	4.202
ALP-POS-88a7a97e-42	VDHCIAQEXRFNGV-UHFFFAOYSA-N	4.202
BRU-THA-92256091-23	AVSHDFLDPKNUPO-UHFFFAOYSA-N	4.202
ALP-POS-ced8ea4d-1	WIOAFGQREWPHPA-UHFFFAOYSA-N	4.197
DAR-DIA-23aa0b97-11	XMKDHSWNVVTQII-UHFFFAOYSA-N	4.197
TRY-UNI-714a760b-12	SXMQMOJIKMECGG-UHFFFAOYSA-N	4.266
MAT-POS-6da3605a-2	PZYJRCQIDAUQHX-UHFFFAOYSA-N	4.185
JIN-POS-6dc588a4-24	OWVCFIMAQDIPNS-UHFFFAOYSA-N	4.181
ALP-POS-ced8ea4d-5	CYRCUXIYRIXMSY-UHFFFAOYSA-N	4.180
MIC-UNK-08cd9c58-1	ACBAQCKSRSEYGV-UHFFFAOYSA-N	4.205
ALP-POS-95f71980-36	YWCBSNXXKGUOFZ-UHFFFAOYSA-N	4.169
NAU-LAT-8502cac5-8	UOUSLMAPSOMURZ-UHFFFAOYSA-N	4.367
RUB-POS-1325a9ea-20	SZILMDJJNXTTF-UHFFFAOYSA-N	4.161
ALP-POS-ced8ea4d-63	JAMSUCBBAYAIEO-UHFFFAOYSA-N	4.160
MAT-POS-993cdc78-2	FBLHAFJQTKYFLC-FQEVSTJZSA-N	4.160
MAT-POS-b5c94171-1	BILHTENSLKPFTQ-UHFFFAOYSA-N	4.158
JAG-UCB-52b62a6f-16	AYODLXBPDLEXFB-UHFFFAOYSA-N	4.153
LON-WEI-8f408cad-1	ZLKNMNXMYFJRNP-UHFFFAOYSA-N	4.153
CHO-MSK-6e55470f-22	JSUYYWXFTLGULD-UHFFFAOYSA-N	4.503
MAT-POS-500f4700-1	LBYWTRXCIGSDSO-UHFFFAOYSA-N	4.145
ALP-POS-6747fa38-1	SIFMSWJDYLFPPS-UHFFFAOYSA-N	4.142
RAL-THA-2d450e86-4	KEALMVHBRJFONR-UHFFFAOYSA-N	4.139
ROD-UFR-86d8d6a3-1	YAHFBZSYGMQIY-UHFFFAOYSA-N	4.138
ALP-POS-305f6ec3-55	ZZGRQMVYMHMUSG-UHFFFAOYSA-N	4.130
EDJ-MED-c8e7a002-4	YOMCSEGWKKANHS-UHFFFAOYSA-N	4.101
ALP-POS-ced8ea4d-9	DNUADCRMUGXTIU-UHFFFAOYSA-N	4.120
JIN-POS-6dc588a4-10	JNBAFDRVDQGPHJ-UHFFFAOYSA-N	4.119
ALP-POS-95f71980-6	GJUVFAGHKMUSIR-UHFFFAOYSA-N	4.117
ALP-POS-95f71980-10	JWGDZWVYLJNDNS-UHFFFAOYSA-N	4.116
ALP-POS-305f6ec3-33	SOLHUAFAHQXZDX-UHFFFAOYSA-N	4.113
TRY-UNI-2eddb1ff-8	NGODSAZXSLLARZ-UHFFFAOYSA-N	4.084
ALP-POS-ced8ea4d-18	VMOASLBZNBXIPK-UHFFFAOYSA-N	4.110
ALP-POS-ce760d3f-8	SHKWJYBWRXZSMY-UHFFFAOYSA-N	4.444
RUT-UNI-630c5802-20	DQSJMGMFBNRMNT-UHFFFAOYSA-N	4.105
ALP-POS-b0bc6a46-28	YHRBBKBYBCPWDW-UHFFFAOYSA-N	4.099
MAT-POS-ec6d90b7-1	FECHZMWPYJRLGC-FQEVSTJZSA-N	4.092
JAG-UCB-cedd89ab-1	ZTWAZISELGQBHC-UHFFFAOYSA-N	4.088
MAT-POS-c478e61b-1	BJMMDGNCVMQTHW-UHFFFAOYSA-N	4.086
MAT-POS-2bb0cf2b-1	ILSGMAUMNXSNQB-HXUWFJFHSA-N	4.085
ALP-POS-88a7a97e-47	RERDNFCIYRPOFQ-UHFFFAOYSA-N	4.083
MAT-POS-3b92565d-6	CBLOHDRIMKSMKB-UHFFFAOYSA-N	4.079
ALP-POS-ced8ea4d-36	XMFIQIFKBRISLKT-UHFFFAOYSA-N	4.076
EDJ-MED-49816e9b-4	ADQNLTVCLDBKV-UHFFFAOYSA-N	4.118
MAT-POS-9ff17035-1	PTPWDISECPKTAU-UHFFFAOYSA-N	4.066
LON-POS-586a7c45-1	SDIRCAZHIVHPY-AREMUKBSSA-N	4.081
ALP-POS-902f21bb-2	UJHVTPSDBZZEES-UHFFFAOYSA-N	4.040
BEN-DND-61647d40-13	SWHLHIOFZDCMKY-UHFFFAOYSA-N	4.060
RAL-MED-2de63afb-1	OCMQJJQEJQFRH-UHFFFAOYSA-N	4.054
ALP-POS-ced8ea4d-28	NGNMIIYBBLGTTG-UHFFFAOYSA-N	4.050
MAT-POS-8695a11f-1	YJEUVMCEEMSKGZ-LJQANCHMSA-N	4.043
EDG-MED-5d232de5-2	YVSNGUCAHVBFBA-AWEZLNQCLSA-N	4.036
ALP-POS-305f6ec3-26	VSVUWJVEHSQHLX-UHFFFAOYSA-N	4.036
BRU-THA-92256091-19	HIKVJPICOQZEEB-UHFFFAOYSA-N	4.028

Compound	InchiKey	pIC50
PET-UNK-8df914d1-4	DVAZCJFDVZFHLG-UHFFFAOYSA-N	4.027
JOH-UNI-0e1753c1-1	WUYDAFIWBDPMJ-UHFFFAOYSA-N	4.027
ALP-POS-d0876c20-15	VXERXFBZBNHAGQ-UHFFFAOYSA-N	4.023
ALP-POS-95b75b4d-1	MLLQAKCWQMHQOK-UHFFFAOYSA-N	4.018
MAR-UCB-f313ec4d-3	DSMFMWCNKBKXKKQ-UHFFFAOYSA-N	4.042
JAN-GHE-83b26c96-22	QOWQEKIGADCHLL-UHFFFAOYSA-N	4.312
MAT-POS-c74bc7b3-2	LNNHTGQQOQWKNR-FQEVSTJZSA-N	4.010
TRY-UNI-714a760b-24	KUGXNNCPAFBCHD-UHFFFAOYSA-N	4.156
EDJ-MED-e4b030d8-12	IMWRDWGDTVKNBA-AWEZLNQCLSA-N	5.449
CHO-MSK-6e55470f-14	YKFSTTQUNHQEOP-UHFFFAOYSA-N	5.131
LON-WEI-babf2c61-14	JAKBSRIOBQUWTD-UHFFFAOYSA-N	4.977
LOR-NOR-c954e7ad-1	XRSFAJNWGRERK-UHFFFAOYSA-N	4.744
MAT-POS-f42f3716-2	VYNTZCCGGDYLMZ-UHFFFAOYSA-N	4.709
VLA-UCB-00f2c2b3-7	CYQCEAJKJVOFMS-UHFFFAOYSA-N	4.627
LOR-NOR-c954e7ad-3	MWXFKVGXECELTR-UHFFFAOYSA-N	4.615
LOR-NOR-c954e7ad-6	KRWGZAGTABTYNO-UHFFFAOYSA-N	4.570
BEN-DND-61647d40-12	YKPILQYKTOXPJC-UHFFFAOYSA-N	4.531
ALP-POS-ddb41b15-8	KWLXUMUUNZANDS-UHFFFAOYSA-N	4.504
JAG-UCB-1d922829-3	KJTYEBJEAMAHXL-UHFFFAOYSA-N	4.492
MAT-POS-8a69d52e-6	SGXOKACJPFKOBBLVVKFPJESA-N	4.454
MAT-POS-8e4737f4-2	IMWRDWGDTVKNBA-CQSZACIVSA-N	4.423
JAG-UCB-1d922829-1	CAHXXSIYLMZEH-UHFFFAOYSA-N	4.361
EDG-MED-0da5ad92-5	YUJIYWCEKJSKBQ-UHFFFAOYSA-N	4.325
JAG-UCB-c61058a9-45	STEFORABHLOGBH-UHFFFAOYSA-N	4.319
CHO-MSK-6e55470f-19	ZUVWWKYSRRNDCK-UHFFFAOYSA-N	4.312
EDJ-MED-c8e7a002-3	URUZYZWVKAJJDE-UHFFFAOYSA-N	4.284
MAT-POS-0c8fa4a7-4	WJZPCYNHBAKRE-UHFFFAOYSA-N	4.279
ALP-POS-e2fddb0f-2	SAIXPJGIXYRJA-UHFFFAOYSA-N	4.274
JAG-UCB-1d922829-2	AWSCEIJUQVXEJ-UHFFFAOYSA-N	4.263
LOR-NOR-c954e7ad-2	UOVBGVWCWYOO-UHFFFAOYSA-N	4.240
ALP-POS-6495d03e-6	PRHCZWPUCDCJFV-UHFFFAOYSA-N	4.238
LON-WEI-b2874fec-23	NZZKWOCHZOWVQW-OAQYLSRUSA-N	4.238
MAT-POS-8fd29122-2	YSSSROBDYWKWLX-UHFFFAOYSA-N	4.230
LON-WEI-b2874fec-25	GKKKXDGPOXUFAL-LJQANCHMSA-N	4.225
LON-WEI-b2874fec-24	HMNQGYVEXVJOLE-HXUWFJFHSA-N	4.222
LON-WEI-b2874fec-16	BHVNAOSKIGSHPN-HXUWFJFHSA-N	4.187
JAG-UCB-52b62a6f-15	KEQGEYSNGXRBCH-UHFFFAOYSA-N	4.130
MAT-POS-8a69d52e-5	SGXOKACJPFKOBBLVKBTVSA-N	4.099
EDJ-MED-c8e7a002-13	FHQURCUGGWIOO-UHFFFAOYSA-N	4.087
EDJ-MED-c8e7a002-5	UBMUDQXFJYVGAS-UHFFFAOYSA-N	4.086
LON-WEI-f8a636e2-1	DDLJCTBSXJGSQA-LJQANCHMSA-N	4.086
JOH-UNI-f51e3bbc-1	KQTVGHXFNNMNF-UHFFFAOYSA-N	4.077
AAR-POS-5507155c-1	KNQFDOLIXOOIGX-UHFFFAOYSA-N	4.077
EDJ-MED-c8e7a002-9	QUDPTQZFACCPRK-UHFFFAOYSA-N	4.053
EDJ-MED-c8e7a002-10	HLCOLJGMRBZZHP-UHFFFAOYSA-N	4.039
PET-UNK-e44fd04-1	PICHRHZICCXQLR-UHFFFAOYSA-N	4.028
ALP-POS-cd485364-1	NSVLSSWSODGTMA-UHFFFAOYSA-N	4.026
LOR-NEU-c8f11034-5	DELLOEULSHGYCG-UHFFFAOYSA-N	4.915
LOR-NOR-30067bb9-11	PGAYPKUTCQWJU-UHFFFAOYSA-N	7.769
JOH-UNI-44664832-1	PXBKXVGKPBXJN-UHFFFAOYSA-M	6.931
MAT-POS-ea426761-10	CHHCUIYTYGXNN-UHFFFAOYSA-N	6.699
JOH-UNI-c7afdb96-2	MZCNLYUDKUARK-UHFFFAOYSA-N	6.632
LOR-NEU-c8f11034-2	GWIDRHBTFZIOZ-UHFFFAOYSA-N	6.314

Manuscript 1

Compound	InchiKey	pIC50
ALP-POS-05819dc4-1	YMIBPTVNIIGSMV-UHFFFAOYSA-N	6.160
LOR-NOR-30067bb9-7	NFAJNAYSRGOCAG-UHFFFAOYSA-N	5.910
JOH-UNI-44664832-2	AUJRCFUBUPVWSZ-XTZHGVARSA-M	5.906
ALP-POS-ce760d3f-2	KCXBRLJFVUQUIR-AWEZNQCLSA-N	5.889
LON-WEI-4d77710c-71	QHPBANPKTYSJRV-UHFFFAOYSA-N	5.843
EDJ-MED-6af13d92-2	HLMOMNRZAYBAJZ-UHFFFAOYSA-N	5.738
EDJ-MED-a364e151-1	KVBRDRRWOWIOL-UHFFFAOYSA-N	5.735
LOR-NOR-30067bb9-5	SDDDFQWUUCROW-UHFFFAOYSA-N	5.668
CHEMBL1977761	XSIOKTWDEOJMGG-UHFFFAOYSA-O	5.710
CHEMBL324842	DOUMFZQKYFQNTF-WUTVXBCWSA-N	5.410
CHEMBL1188627	FFUBTEITUNMMOK-NXVVXOECSA-N	6.010
CHEMBL518800	HOCBJBNQIQQGT-LJQANCHMSA-N	6.080
CHEMBL52	HCZKYJDFEPMADG-UHFFFAOYSA-N	5.590
CHEMBL111654	YXIWHUQXZSMYRE-UHFFFAOYSA-N	5.760
CHEMBL1201073	ACTRVOBWPAIOHC-XIXRPRMCSA-N	5.760
CHEMBL1201236	TZFNLOMSOLWIDK-JTQLQIEISA-N	5.450

Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition

Guillem Macip¹ | Pol Garcia-Segura¹ | Júlia Mestres-Truyol¹ | Bryan Saldivar-Espinoza¹ | María José Ojeda-Montes² | Aleix Gimeno³ | Adrià Cereto-Massagué⁴ | Santiago Garcia-Vallvé^{1,5} | Gerard Pujadas^{1,5}

¹Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Research group in Cheminformatics & Nutrition, Tarragona, Tarragona, Spain

²Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

³Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

⁴EURECAT Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Spain

⁵EURECAT, TECNIO, CEICS, Avinguda Universitat 1, Reus, Spain



Received: 24 March 2021 | Revised: 30 June 2021 | Accepted: 11 October 2021

DOI: 10.1002/med.21862

REVIEW ARTICLE

Medicinal Research Reviews WILEY

Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition

Guillem Macip¹ | Pol Garcia-Segura¹ |
Júlia Mestres-Truyol¹ | Bryan Saldivar-Espinoza¹ |
María José Ojeda-Montes² | Aleix Gimeno³ |
Adrià Cereto-Massagué⁴ | Santiago Garcia-Vallvé^{1,5} |
Gerard Pujadas^{1,5}

¹Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Research group in Cheminformatics & Nutrition, Tarragona, Tarragona, Spain

²Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

³Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

⁴EURECAT Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Spain

⁵EURECAT, TECNIO, CEICS, Avinguda Universitat 1, Reus, Spain

Correspondence

Santiago Garcia-Vallvé and Gerard Pujadas, Departament de Bioquímica i Biotecnologia, Research group in Cheminformatics & Nutrition, Campus de Sescelades, Universitat

Abstract

This review makes a critical evaluation of 61 peer-reviewed manuscripts that use a docking step in a virtual screening (VS) protocol to predict SARS-CoV-2 M-pro (M-pro) inhibitors in approved or investigational drugs. Various manuscripts predict different compounds, even when they use a similar initial dataset and methodology, and most of them do not validate their methodology or results. In addition, a set of known 150 SARS-CoV-2 M-pro inhibitors extracted from the literature and a second set of 81 M-pro inhibitors and 113 inactive compounds obtained from the COVID Moonshot project were used to evaluate the reliability of using docking scores as feasible predictors of the potency of a SARS-CoV-2 M-pro inhibitor. Using two SARS-CoV-2 M-pro structures and five protein-ligand docking programs, we proved that the correlation between the pIC₅₀ and docking scores is not good. Neither

Guillem Macip, Pol Garcia-Segura, Júlia Mestres-Truyol contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Medicinal Research Reviews* published by Wiley Periodicals LLC.

Rovira i Virgili, c/Marcel·li domingo, 1,
Tarragona 43007, Catalonia, Spain.
Email: santi.garcia-vallve@urv.cat and
gerard.pujadas@gmail.com

Funding information

Universitat Rovira i Virgili,
Grant/Award Number: 2019PFR-URV-B2-69;
European Union's Horizon 2020 research and
innovation program under the Marie
Sklodowska-Curie grant agreement,
Grant/Award Number: 713679

was any correlation found between the pIC_{50} and the ΔG calculated with an MM-GBSA method. When a group of experimentally known inactive compounds was added, neither the docking scores or the ΔG were able to distinguish between compounds with or without M-pro experimental inhibitory activity. Performances improved when covalent and noncovalent inhibitors were treated separately, but were not good enough to fully support using a docking score as a cutoff value for selecting new putative M-pro inhibitors or predicting the relative bioactivity of a compound by comparison with a reference compound. The two sets of known SARS-CoV-2 M-pro inhibitors presented here could be used for validating future VS protocols which aim to predict M-pro inhibitors.

KEYWORDS

drug repurposing, molecular docking, M-pro inhibitors, SARS-CoV-2 3C protease, virtual screening

1 | INTRODUCTION

It has been more than a year since the onset of the coronavirus disease 2019 (COVID-19) global pandemic, which has affected 219 countries and territories around the world. According to the Johns Hopkins Coronavirus map tracker,¹ there have been more than 123 million cases worldwide. During the last year, more than 100 million people have recovered from the disease while over 2.7 million have died.² The pandemic has had a detrimental effect on healthcare systems worldwide and has affected almost all aspects of life. When the World Health Organization declared the COVID-19 outbreak a global emergency, governments started to enforce border shut-downs, travel restrictions, quarantines, curfews and other restrictions, which in turn fuelled an impending economic crisis and recession.³ The scientific community reacted to the COVID-19 global pandemic in an outstandingly rapid and effective way, and thousands of articles were published in the first few months. Between January and May 2020, roughly 100 new articles on COVID-19 were added every day to PubMed,⁴ with a median time from submission to acceptance of only 6 days.⁵ Although the nature of the COVID-19 emergency warranted accelerated publishing, measures need to be taken to safeguard the integrity of the scientific evidence.⁵ There is concern among some scientists that poor quality research prevents an effective evidence-based response.^{6,7}

Coronaviruses (CoVs) are a group of viruses containing a linear, positive sense single-stranded RNA genome.⁸ Their genome size varies, but is larger than that of other types of virus. The genome is housed within an enveloping lipid bilayer in which three structural proteins—membrane (M), envelope (E) and spike (S)—are anchored. The E and M proteins are responsible for the shape of the virus, while the S protein mediates receptor attachment and viral and host cell membrane fusion.⁸ The nucleocapsid (N) protein binds to the viral RNA and forms a ribonucleoprotein that is packaged in the virus envelope. In addition to these four structural proteins, the CoV genome also codes for several nonstructural proteins, synthesized as two very large polyproteins that are cleaved by two or three virus proteases, and several auxiliary proteins. CoVs are classified within four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*. To date, seven strains of *Alphacoronavirus* and *Betacoronavirus* that infect

humans have been identified.⁹ But our main concern here lies within the *Betacoronavirus* genus which contains the severe acute respiratory syndrome (SARS) viruses. SARS-like viruses emerged as human infectious diseases in 2002, and since then have caused several epidemics and outbreaks. Before COVID-19, outbreaks of SARS in 2003¹⁰ and Middle East Respiratory Syndrome (MERS) in 2012¹¹ had already been reported as serious diseases caused by *Betacoronaviruses*. The causative agent of COVID-19 is SARS-CoV-2.¹² This virus has the capacity to affect multiple organs as well as the central nervous system and sometimes causes respiratory problems which might end up having fatal consequences. The SARS-CoV-2 genome consists of around 30,000 nucleotides and contains 12 functional open reading frames (ORFs): from 5' to 3', ORF1a, ORF1b, S, ORF3a, E, M, ORF6, ORF7a, ORF7b, ORF8, N, and ORF10.^{13,14} Two-thirds of the SARS-CoV-2 genome, located at the 5' genomic region, codify the ORF1a and ORF1b genes, which encode the polyproteins pp1a and pp1ab. Polyprotein pp1ab contains the pp1a protein and its synthesis requires a -1 ribosomal frameshift.¹⁵ Both polyproteins are further cleaved by the main protease (M-pro, also known as 3CLpro) and papain-like (PLpro) protease to produce 16 proteins that include both proteases, an RNA-dependent RNA polymerase, a helicase, two nucleases and a methyltransferase, among other proteins.¹⁵

Due to the lack of target-specific drugs for treating COVID-19, several strategies are being investigated. Among these strategies, several clinical trials are being developed to: (i) inhibit the RNA-dependent RNA polymerase,^{16,17} (ii) inhibit the viral M-pro or PLpro proteases,^{18–21} (iii) block the fusion with the host cells,^{22–24} (iv) enhance the innate immune system,²⁵ (v) attenuate the inflammatory response,²⁶ (vi) control the symptoms,²⁷ and (vii) create pathogen-specific artificial antigen-presenting cells.²⁸ Given its importance in the replication of SARS-CoV-2, M-pro has become a key target for the development of antiviral drugs.²⁹ M-pro proteins from SARS-CoV-2 and SARS-CoV show a high sequence (i.e., >96%) and structural similarity (i.e., an root mean-squared deviation [RMSD] of 0.44 Å between the Cα of both structures). SARS-CoV-2 M-pro has a length of 306 amino acids that form three different domains (domains I, II, and III).³⁰ The substrate-binding site is located at a cleft between domain I (residues 8–101) and domain II (residues 102–184), which share a common fold consisting of antiparallel β-barrel structures (Figure 1). Domain III (residues 201–303) consists of an antiparallel globular cluster of five alpha-helices

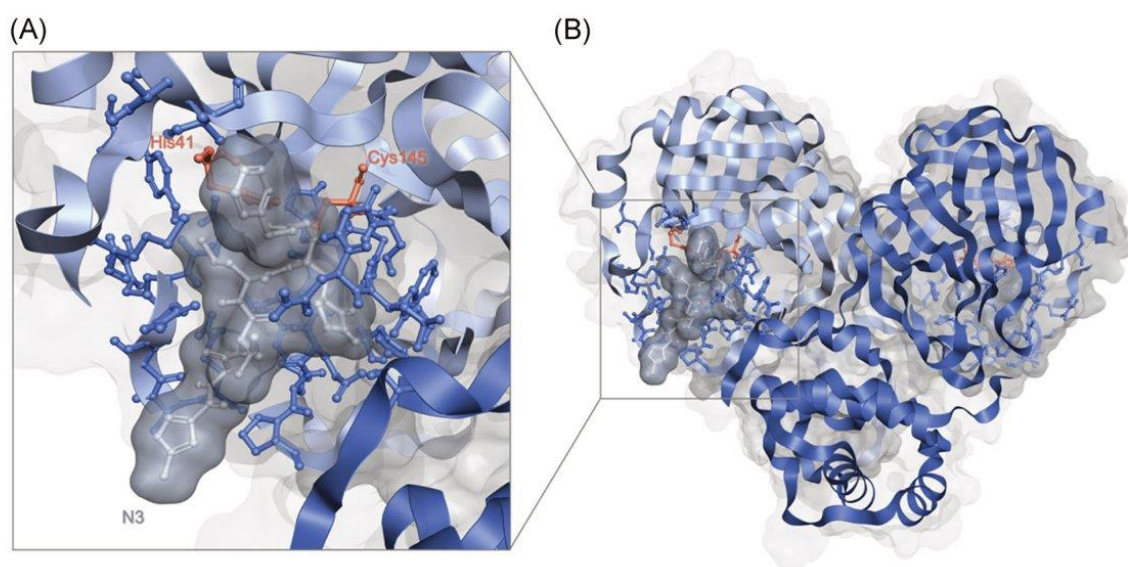


FIGURE 1 SARS-CoV-2 M-Pro homodimer (PDBid 6LU7). Panel (B) shows an overview of the main structural features of M-Pro and panel (A) displays in detail one of its binding pockets where the co-crystallized ligand N3, in grey, fits. Each of the two protomers is a different tone of blue. In both panels, the residues that form the binding pocket (according to Gimeno et al.²⁹) are also highlighted in ball and stick format and the residues of the catalytic dyad (His41 and Cys145) are colored orange and labeled. This figure has been generated using the Flare software.^{32–34} [Color figure can be viewed at wileyonlinelibrary.com]

and is involved in M-pro dimerization, which is essential for M-pro activity.³⁰ Unlike the conventional Ser(Cys)-His-Asp(Glu) triad found in other chymotrypsin-like enzymes, SARS-CoV-2 M-pro has a catalytic dyad formed by His41 and Cys145 (Figure 1) that catalyzes the hydrolysis of the peptide bond at highly specific sites of a polypeptide chain through a common nucleophilic-type reaction. It was suggested that a water molecule might complete the catalytic triad by mediating crucial interactions between His41 and other important conserved residues, such as His164 and Asp187.³¹ While noncovalent small inhibitors have the ability to interact with different residues that lie within the binding pocket of SARS-CoV-2 M-pro, covalent inhibitors are found to mainly interact with the catalytic cysteine residue after an initial noncovalent mechanistic step that ensures the warhead is correctly oriented.²¹ Thus, it is of great importance to consider covalent and noncovalent inhibitors as different, because their binding mode and mode of action are quite distinct. Most of the characterized SARS-CoV-2 M-pro inhibitors were originally designed for other pathogens, such as SARS-CoV, HIV and Hepatitis C viruses, which shows the importance of drug repurposing strategies.

Drug repurposing, also known as drug repositioning, is a well-known strategy in drug development, that is used to identify new medical uses for already approved or investigational drugs. This has enormous advantages compared to traditional drug development from scratch. Briefly, drug repurposing shortens time-consuming steps in drug development, as safety assessment and testing are usually completed, while it is also known to substantially diminish the investment needed. This, together with the decreased risk of failure, makes drug repurposing an interesting approach when time presses, such as in a global pandemic like COVID-19.³⁵ Given the high throughput of computational approaches, they are one of the mainstays of drug repurposing. Specifically, the most popular structure-based *in silico* approach, protein-ligand molecular docking, has been extensively used not only in drug discovery but also in drug repurposing.^{36,37} Molecular docking is a receptor-based VS method, that can be used to assess the possible binding poses of the ligands (e.g., a library of approved drugs) to a target (e.g., SARS-CoV-2 M-pro). Thus, molecular docking makes it possible to interrogate thousands of drugs against the protein target at a relatively low computational cost, so it is a popular tool in computational and structural biology.

Several articles have reviewed the SARS-CoV-2 M-pro inhibitors discovered to date.^{20,21,38–43} In this review, we focus on the computational methods, especially protein-ligand docking, which have been used to discover new SARS-CoV-2 M-pro inhibitors. We check 61 peer-reviewed manuscripts that use at least one docking step in a virtual screening (VS) protocol to predict SARS-CoV-2 M-pro inhibitors in approved or investigational drugs. We put special emphasis on the validation of the docking process and whether the use of docking scores or other estimates of binding affinity are a reliable way to select new putative inhibitors of SARS-CoV-2 M-pro. For this purpose, we have taken a set of 150 M-pro inhibitors with known experimental bioactivity from the literature and a second set of 81 M-pro known inhibitors and 113 inactive compounds from the COVID Moonshot project to evaluate the reliability of using docking scores as feasible predictors of the potency of a SARS-CoV-2 M-pro inhibitor.

2 | RECENT VIRTUAL SCREENINGS EFFORTS THAT USE PROTEIN-LIGAND DOCKING TO PREDICT NEW SARS-COV-2 M-PRO INHIBITORS

Computational approaches can contribute to drug discovery by reducing the cost and time of drug development and speeding up the analyses of the interactions between a target and candidate drugs.³⁷ Of all the VS methods, and specifically in drug repurposing, molecular docking is an attractive choice because it can predict the preferred orientation of a drug at the binding site of a target.⁴⁴ Molecular docking methods use two independent steps: pose generation and scoring.⁴⁵ The first step refers to the methods used to create different ligand poses within the binding site of the protein. The second, scoring, is required to rank the different poses and quantitatively estimate the pose quality.⁴⁵ Scoring functions attempt to estimate the binding affinity between the ligand and the protein.⁴⁵

We collected 61 articles that use a VS protocol based on protein-ligand docking to predict drugs as SARS-CoV-2 M-pro inhibitors. These articles were published in peer-reviewed journals between the beginning of the COVID-19

pandemic in early February 2020 and November 2020. It should be pointed out that preprints were not included. Table 1 outlines the most relevant information of these 61 articles. Further details and the drugs predicted in each can be found in Supporting Information File 1. As shown in Table 1, most of these articles used an initial database of approved drugs (e.g., the Drugbank database). A quarter of the articles also included a group of antiviral molecules, such as HIV or Hepatitis C protease inhibitors. The first crystallized structure of the SARS-CoV-2 M-pro protein, the 6LU7 structure released on February 5, 2020³⁰ and crystallized with the covalent inhibitor N3 was largely used. Before this structure was available, eight articles used a structural model of the M-pro constructed using the homologous protein from SARS-CoV as a template. Subsequently, many other SARS-CoV-2 M-pro structures were released and several articles used the M-pro structure, such as 6W63, complexed with a noncovalent inhibitor. Most of the VS protocols reviewed here used AutoDock Vina or other docking methodologies available at the Schrödinger Drug Discovery suite (i.e., Glide HTVS, Glide SP and Glide XP; see Table 1). The majority of these VS protocols used only one docking program, although some of them used two or three. Generally, protocols using multiple docking programs are performed in a stepwise manner; i.e., the programs are used in a funnel approach to reduce the number of hits by using the different prediction algorithms and/or punctuation functions, which is indeed what happens when the increasingly exhaustive methodologies of the Schrödinger package Glide (Glide HTVS, SP, and XP) are used. Nonetheless, there is also room to use different programs in a parallel manner and then evaluate the equivalent poses predicted by each of them, as in our last contribution.²⁹ Usually, the docking score of the best pose of a reference compound (normally, a known or expected inhibitor) is generally used as a cutoff value for selecting new putative M-pro inhibitors. If the best docking score of a compound is more negative than the docking score of the reference compound, it is usually assumed that the compound will be more active than the reference one. When a reference compound is not used, the most negative docking score values tend to be used as a cutoff to select a group of hit compounds and analyze them in further detail. The predicted interactions between the substrate-binding site of the M-pro protein and each selected compound are usually analyzed. In addition, some reviewed articles use molecular dynamics (MD) to evaluate the binding dynamics of the group of selected hits to the M-pro structure. The molecular mechanics generalized Born surface area (MM-GBSA) or the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) algorithms were generally used to estimate the binding free energies of the putative inhibitors.

The best way to validate the methodology and results of a VS protocol is to experimentally prove the bioactivity of some selected hits. Only two of the 61 articles reviewed confirmed their computational predictions *in vitro*.^{29,46} When this is not possible, the methodology should be subject to a computational validation.³⁷ A common strategy for validating the docking step in some of the 61 articles was re-docking or self-docking. Re-docking consists of removing the ligand from a protein-ligand complex to dock it back and evaluate if a docking program can predict its initial binding pose. The RMSD between the best docking pose and the crystallographic one, or the lowest RMSD value of all the predicted poses, gives an estimation of the ability of the docking program to reproduce the co-crystallized binding pose of the ligand. RMSD values under 2.0 Å are usually considered to be acceptable, but values close to zero are ideal. Although 18 of the 61 VS protocols analyzed used re-docking to estimate the docking accuracy (see Table 1 and Supporting Information File 1), only two^{47,48} converted the extracted ligand to SMILES format and then back to 3D. This is done to lose the binding coordinates of the ligand and evaluate the docking accuracy from scratch, similar to how the compounds are predicted. Because docking calculations are very sensitive to the starting geometries of ligands,⁴⁹ it may also be beneficial to re-run the re-docking with different starting conformations each time.⁵⁰ In addition to re-docking, it is also recommended to use cross-docking to try to predict the binding pose of a ligand using a different PDB structure. In this way, the possible effect of protein flexibility can be evaluated or the best protein structure identified when several are available. However, re-docking and cross-docking cannot suitably validate an entire VS process or each of its steps. In addition, if docking scores are used to rank the bioactivity of the predicted compounds, it needs to be shown that they correlate with the bioactivity. Without a proper validation, the predictions obtained are speculative and dubious. A VS protocol can be computationally validated by applying the same VS protocol to a set of active compounds and to a set of inactive or decoy compounds (when a group of inactive compounds is not available). The more accurate the methodology is, the more

TABLE 1 Outlines of the 61 articles that predict drugs to be SARS-CoV-2 M-pro inhibitors

Compounds database	Counts	% articles
FDA-approved	27	44.26
Approved	36	59.02
Drugbank	21	34.43
Selleckchem	7	11.48
ZINC	12	19.67
PubChem	5	8.20
ChEMBL	6	9.84
Antiviral	15	24.59
Others	3	4.92
PDB structure		
6LU7	39	63.93
6W63	6	9.84
6M03	4	6.56
Model	5	8.20
Others	8	13.11
Docking software		
AutoDock Vina	26	42.62
Glide HTVS	13	21.31
Glide SP	13	21.31
Glide XP	14	22.95
GOLD	4	6.56
Others	13	21.31
N° of docking programs		
1	37	60.66
2	12	19.67
3	12	19.67
Molecular dynamics		
Yes	35	57.38
No	26	42.62
Validation method		
Actives enrichment in silico	3	4.92
In vitro	2	3.28
Re-docking	18	29.51

(Continues)

TABLE 1 (Continued)

Compounds database	Counts	% articles
No	41	67.21
Total of articles ^a	61	

Note: Information is shown about the initial database, the PDB structure, the protein-ligand docking method, whether they use molecular dynamics or not, and the validation method used (if any).

Abbreviations: FDA, Food and Drug Administration.

^a% Of articles does not necessarily add 100% because some articles may be included in more than one group.

true positives (i.e., active compounds that pass the VS procedure or VS step) and true negatives (i.e., inactive molecules that do not pass the VS procedure or VS step) there will be. To quantify the performance of the entire methodology or each VS step, several metrics can be used. Of these, the most popular are the area under a ROC curve and the Enrichment Factor (EF), which quantify the extent to which a step of the VS can be enriched with active compounds. Only three of the 61 articles reviewed used a group of known M-pro inhibitors to confirm that their VS procedure was able to enrich the predicted compounds with these inhibitors.^{29,51,52} In one of the articles⁵² 12 M-pro inhibitors, 10 weak M-pro inhibitors and 114 DUDE-generated decoys were screened to calculate the EF for each step of the VS and to choose the optimal scoring function for the final hit selection. Although some of the 61 articles used re-docking as a validation step, more than the 67% did not apply a validation step (Table 1). This lack of validation of the structure-based approaches to drug repurposing that target M-pro and other SARS-CoV-2 proteins has been described elsewhere.^{53,54} In a survey of the literature before 26 August, Bellera and coworkers⁵⁴ reviewed 96 articles that reported the use of molecular docking for identifying potential compounds against SARS-CoV-2 M-pro. None of the 96 articles reported any level of experimental validation and 40.6% of the articles did not perform any kind of in silico validation.⁵⁴ We agree with Bellera and coworkers that “the preceding figures are surprising, considering the known fact that the performance of docking studies is highly dependent on the system under study, and the high false-positive rate associated with docking-based VS”.⁵⁴

An interesting approach used in some of the articles reviewed^{48,52,55–57} was to add a pharmacophore-based procedure before the docking step to enrich the database with some characteristics of the reported M-pro inhibitors or M-pro co-crystallized ligands, or to consider only the compounds that can engage in specific interactions with the M-pro active site. This strategy makes it possible to identify which of the initial number of compounds are most likely to be inhibitors, even though it requires prior knowledge of the interactions between the receptor and its ligands. Thus, it cannot be applied in early attempts, when little information is available, but it becomes a more powerful tool as more information arises.

Figure 2 shows that the 61 VS protocols analyzed predict very few M-pro inhibitors in common. It could be argued that this lack of agreement about the potential inhibitors is due to differences in the initial set of compounds analyzed and the methods used. However, Figure 3 shows that when different protocols using similar initial datasets are compared, such as FDA-approved drugs (Figure 3A) or approved and investigational drugs (Figure 3B), very few compounds are predicted by two protocols even if the same PDB structure and docking program have been used (Figure 3). Some of the compounds most commonly predicted as inhibitors of SARS-CoV-2 M-pro in several of the 61 VS protocols reviewed are: the covalent inhibitors of the HIV-protease ritonavir, saquinavir, indinavir, lopinavir and nelfinavir; other antiviral compounds, such as the hepatitis C protease inhibitors simeprevir and paritaprevir; and the HIV integrase inhibitor raltegravir. All of these compounds were predicted by more than five different articles (Figure 2). As these antiviral compounds were initially used in clinical trials to combat the SARS-CoV-2 pandemic, they are also often selected to be included among the final hits of the VS protocols analyzed. Ritonavir inhibits cytochrome P450-3A4, and is often simultaneously used with lopinavir or other HIV or hepatitis C protease inhibitors since it increases their half-life. Although ritonavir was found to inhibit the

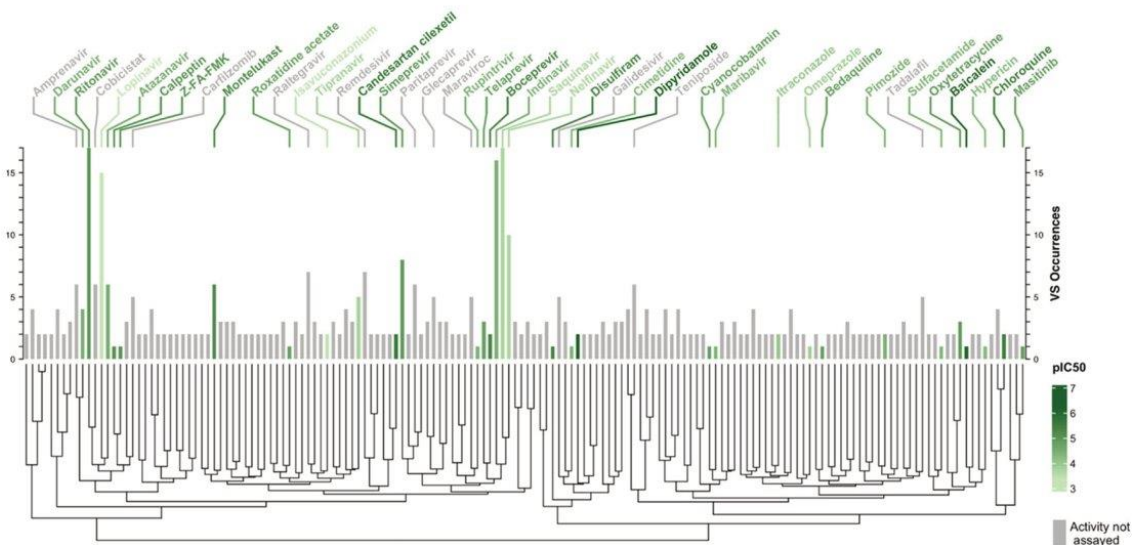


FIGURE 2 Dendrogram and bar plot of the compounds that are most predicted to be inhibitors of M-pro in several VS. The dendrogram clusters the compounds in terms of their structural pairwise similarity established by the Tanimoto coefficients calculated from the ECFP6 circular fingerprint from each compound. The bar plot represents the number of times that each compound has been predicted as a putative M-pro inhibitor in the reviewed articles (VS occurrences). The in vitro activity, when available, is expressed as pIC₅₀ values and represented in a green scale. The only compounds displayed are those that are predicted in two or more different VS or for which experimental pIC₅₀ is available. The names of the compounds with known activity and those with five or more VS occurrences are displayed. VS, virtual screening [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

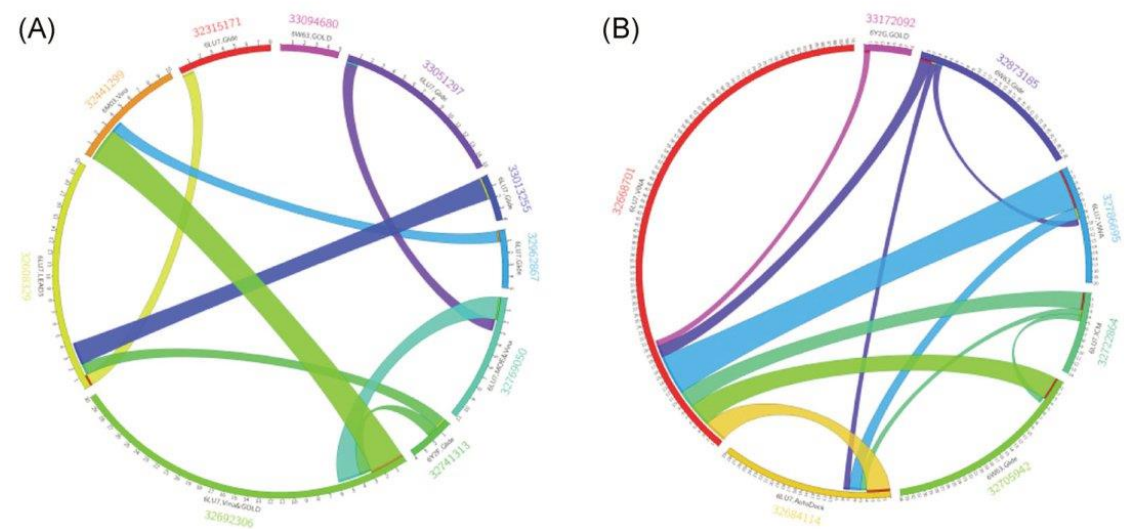


FIGURE 3 Circos plots showing the number of shared compounds among the predicted hits of different VS studies which use similar initial datasets. Panel (A) shows the number of compounds that are predicted simultaneously by two different articles that use a group of FDA-approved drugs (between 1500 and 2800 drugs) as the initial dataset. Panel (B) shows the compounds predicted simultaneously by two articles that use a group of approved and investigational drugs (between 8500 and 13,563 drugs). The information shown is the PubMed ID of each article, the PDB structure and docking program used, the number of compounds predicted by each article and the number of common compounds. FDA, Food and Drug Administration; VS, virtual screening [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

10981128, 2022, 2, Downloaded from <https://onlinelibrary.wiley.com/doi/10.1002/mr.21862> by Readcube (Labtiva Inc.), Wiley Online Library on [13/10/2022]. See the Terms and Conditions (<https://onlinelibrary.wiley.com/terms-and-conditions>) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SARS-CoV-2 M-pro in cell culture,⁵⁸ high concentrations are required to achieve significant inhibition, so it will be of limited clinical relevance.⁵⁸ The results of several clinical trials that have evaluated the role of lopinavir/ritonavir combination therapy in patients with COVID-19 have shown little or no improvement.^{59–61} Remdesivir was also predicted as an M-pro inhibitor by 7 of the 61 articles. This compound was the first compound approved by the FDA to treat patients with severe symptoms of COVID-19. However, remdesivir is an adenosine nucleoside triphosphate analog that inhibits the SARS-CoV-2 RNA-dependent RNA polymerase and induces an irreversible chain terminator. So, it is surprising that it is predicted to be a SARS-CoV-2 M-pro inhibitor. There is no experimental evidence to suggest that remdesivir binds to SARS-CoV-2 M-pro.⁶² In addition, remdesivir is a prodrug, converted to its active form by host enzymes, and this was often not taken into account in the VS protocols analyzed. To the best of our knowledge, only one study has docked remdesivir and its active form to SARS-CoV-2 M-pro and suggested that both molecules have a comparable affinity against SARS-CoV-2 M-pro.⁶² More experiments are required to precisely determine if, in addition to inhibiting the RNA-dependent RNA polymerase of SARS-CoV-2, remdesivir's mechanism of action includes the inhibition of M-pro.

Figure 2 also shows that the M-pro inhibitors predicted by most articles are not the most active inhibitors. The IC_{50} values of lopinavir, saquinavir, indinavir and nelfinavir have been estimated to be 486,⁶³ between 31.4 and 411,^{58,63} 43.1,⁴⁶ and 234 μM ,⁶³ respectively. In other studies, lopinavir, indinavir, nelfinavir, tipranavir and other HIV protease inhibitors have not shown any inhibitory activity between 20⁶⁴ and 200 μM .⁵⁸ It has been suggested that the reduction in viral activity shown by nelfinavir and lopinavir in cell cultures is due to the inherent cytostatic and/or cytotoxic effects of these compounds.⁶⁵ Dipyridamole,^{46,66} baicalein,⁶⁶ candesartan cilexetil^{46,67} (predicted in 2, 1, and 2 studies, respectively) have the lowest IC_{50} values of the most predicted M-pro inhibitors, but in the three cases, their respective IC_{50} of 0.6, 0.94 and 2.78 μM ^{46,68} are a long way from the most potent SARS-CoV-2 M-pro inhibitors, such as PF-00835231, MI-21 and MI-23, which have an IC_{50} in the nanomolar range.^{69,70} As depicted in the dendrogram in Figure 2, in structural terms the most predicted compounds in VS studies vary considerably, probably due to the high flexibility induced by ligand binding and to the size of the M-pro's binding site.^{31,71} Also, relatively similar compounds (e.g., ritonavir and lopinavir) can have a quite different pIC_{50} , which makes M-pro a challenging target in docking and SAR studies, as a wide variety of scaffolds, activity and selectivity cliffs might have to be taken into account. To discover new therapeutics for COVID-19 it would be useful to make a comprehensive and systematic assessment of the chemical space of known SARS-CoV-2 inhibitors.⁷² Although this approach opens up an interesting line in SARS research for the discovery of new SARS-CoV-2 M-pro inhibitors, it falls outside the scope of the present work and, for the sake of brevity, no further discussion will be conducted.

In the following sections we describe a group of known SARS-CoV-2 M-pro inhibitors and inactive compounds, and show how their bioactivity can be used to validate a protein-ligand docking step.

3 | KNOWN SARS-COV-2 M-PRO INHIBITORS

Figure S1 and Supporting Information File 2 show the 2D structure, bioactivity and other details of 150 known SARS-CoV-2 M-pro inhibitors extracted from the literature. This set of inhibitors includes only those compounds whose inhibitory capacity, mainly expressed as the IC_{50} value, against M-pro from SARS-CoV-2 has been determined in vitro. The bioactivities of these compounds have been obtained since the second trimester of 2020 and show how quickly the scientific community has responded to the challenge of the COVID-19 pandemic, which appeared abruptly at the beginning of 2020. Most of these 150 compounds are found in other reviews that describe recent progress made in searching for SARS-CoV-2 M-pro inhibitors.^{20,21,38–43} Among these M-pro inhibitors there are SARS-CoV M-pro inhibitors such as the compound MAC-5576; HIV-protease inhibitors such as atazanavir, darunavir, indinavir, lopinavir, nelfinavir, ritonavir and saquinavir; Hepatitis C protease inhibitors such as boceprevir, narlaprevir, simeprevir and telaprevir; Caspase-3 inhibitors such as Z-DEVD-FMK and Z-FA-FMK; Calpain and other protease inhibitors; and inhibitors obtained from high-throughput screening experiments. Most

of the HIV, SARS-CoV or hepatitis C protease inhibitors can make a covalent bond with the catalytic Cys145 of SARS-CoV-2 M-pro. Among the 150 SARS-CoV-2 M-pro inhibitors reviewed here, there are compounds with moderate or low inhibitory activity; compounds, such as carmofur, ebselen, disulfiram, PX-2, shikonin and tideglusib, whose inhibition potencies decreased in the presence of reducing agents, such as DTT, and are considered to be nonspecific inhibitors^{73,74}; compounds such as shikonin, which quench the signal used to evaluate its activity and therefore may represent a false positive of the analysis⁷³; and compounds such as the HIV-protease inhibitors lopinavir, indinavir, nelfinavir saquinavir, and tipranavir, which have not shown M-pro inhibitory activity in some concentrations analyzed.^{58,64,75,76} To test the hypothesis that inhibiting the enzymatic activity of M-pro will inhibit SARS-CoV-2 replication, a cellular antiviral assay is usually performed and an EC₅₀ value calculated. This EC₅₀ value is defined as the concentration of the compound that reduces the viral-induced cytopathic effect by 50% of with respect to the virus control. Several SARS-CoV-2 M-pro inhibitors have EC₅₀ values that are slightly more than 10 times their IC₅₀ values.⁷⁶ Some of the 150 known SARS-CoV-2 M-pro inhibitors, such as carmofur, darunavir, disulfiram, ebselen, MG-132, PX-2, shikonin and tideglusib, cannot inhibit SARS-CoV-2 replication in cell cultures or do inhibit it due to their cytotoxicity.^{73,74,77–79} This shows that not all the compounds that inhibit the SARS-CoV-2 M-pro in vitro will be useful in vivo.

Boceprevir, calpain inhibitors I, II, and XII, carmofur, cinanserin, disulfiram, ebselen, GC376, N3, naltrexone, PX-12, shikonin, telaprevir and tideglusib are some of the covalent inhibitors described for the SARS-CoV-2 M-pro (see Figure S1 and Supporting Information File 2). The role of covalent inhibitors in drug discovery is controversial. Some of the risks they pose are toxicity, immune-mediated drug hypersensitivity, nonspecificity, and hyperreactivity that might lead to further drug-induced toxicity (e.g., hepatotoxicity, mutagenicity, or carcinogenicity).⁸⁰ However, three potential benefits are that the high biochemical efficiency may lead to lower doses and reduced off-target effects, lower drug resistance and fewer cases of decreased potency due to competing endogenous substrates.⁸⁰ Reversible or irreversible covalent inhibitors that have some kind of specificity for a target or group of targets must first be recognized by the target binding site to place a reactive functional group, usually referred to as a warhead, in an appropriate position to form a covalent bond between the inhibitor and its target.^{21,81} The formation of the covalent bond depends not only on the correct structural complementarity between the protein and the inhibitor, but also on the appropriate chemical reactivity of the inhibitor and the protein environment that stabilizes the covalent complex.⁸² The protonation state of catalytic Cys145 and His41 and His172 is crucial for the activity variation caused by pH changes and the M-pro catalytic mechanism^{83–85} and may be crucial to the action of covalent inhibitors. The majority of SARS-CoV-2 covalent inhibitors are large substrate mimetics with an electrophilic warhead that covalently binds with the catalytic Cys145 sulfur atom.²¹ But other more nucleophilic cysteines, such as Cys44 located near the substrate binding pocket, are additional sites for covalent linkage.⁸⁵ Table 2 shows such covalent warheads, as acrylamide, chloroacetamide, vinylsulfonamide, nitrile, a Michael acceptor group, alpha-ketoamide, aldehyde, disulfide and several methyl ketone groups, which have been used here to identify putative M-pro covalent inhibitors. Whereas some covalent warheads, such as acrylamide and chloromethyl ketones, form irreversible bonds with their biological targets, others, such as nitriles or electron-deficient ketones, have been found to form a reversible covalent bond.^{21,80} For irreversible covalent inhibitors, if a sufficient period of time is allowed, inhibition is expected to be complete.⁸⁶ Thus, conventional IC₅₀ measurements are of limited value for characterizing the potency of irreversible inhibitors because incubation for a different period of time would provide a different IC₅₀ value.⁸⁶ Of the 150 SARS-CoV-2 M-pro inhibitors reviewed here, 73 present a covalent warhead (Supporting Information File 2) and are plausibly covalent inhibitors.

Among the most potent inhibitors of the SARS-CoV-2 M-pro, the compounds PF-00835231, MPI3, UAWJ248, MPI4 and MPI5 have a pIC₅₀ higher than 7.5, and an IC₅₀ in the nanomolar range. PF-00835231 is a hydroxymethyl ketone covalent inhibitor selected as a development candidate for SARS-CoV, the clinical advancement of which was suspended in 2003 due to the end of the SARS-CoV outbreak.⁶⁹ It is the most potent SARS-CoV-2 M-pro inhibitor known to date, with an IC₅₀ of 0.00027 μ M.⁶⁹ It is a highly selective pan-coronavirus M-pro inhibitor that can also inhibit SARS-CoV-2 replication in cells.⁶⁹ Preclinical research is ongoing in the use of

TABLE 2 Covalent warheads that can be used to identify putative covalent inhibitors. It shows the SMARTS that can be used to identify each warhead and some examples of SARS-CoV-2 inhibitors that contain each warhead

Warhead	SMARTS	Examples
acrylamide	<chem>[C;H2:1]=[C;H1]C(N)=O</chem>	
chloroacetamide	<chem>Cl[C;H2:1]C(N)=O</chem>	
vinylsulfonamide	<chem>NS(=O)([C;H1]=[C;H2:1])=O</chem>	
nitrile	<chem>N#[C:1]-[*]</chem>	Cimetidine, isavuconazole
Michael_acceptors	<chem>C=!@CC=[O,S]</chem>	Cinanserin, MPI2, MPI9, N3
alpha-ketoamide	<chem>C(=O)(C=O)N</chem>	Boceprevir, narlaprevir, telaprevir
aldehyde	<chem>[CX3H1](=O)</chem>	Leupeptin, 2413716-71-7
bisulfite_adduct_of_aldehyde	<chem>C(O)S(=[OX1])([O])(=[OX1])</chem>	GC376
urea_carbonyl	<chem>[NX3][CX3](=[OX1])([NX3,nX3])</chem>	Carmofur
bis(dialkylaminethiocarbonyl)disulfide	<chem>[CX3](=[SX1])SS[CX3](=[SX1])</chem>	Disulfiram
carbamoylsulfanyl	<chem>[NX3,nX3][C,c](=[OX1])([SX2,sx2])</chem>	Tideglusib
disulfide	<chem>[SX2][SX2]</chem>	PX-12
Hydroxymethyl_ketone	<chem>[CX3H0](=[OX1])[CH2][OH]</chem>	PF-00835231
Alkoxyethyl_ketone	<chem>[CX3H0](=[OX1])[CH2][OX2H0]</chem>	GW-0742, 2434601-94-0
Acyloxyethyl_ketone	<chem>[CX3H0](=[OX1])[CH2][OX2H0] [CX3H0](=O)</chem>	2434601-94-0, 2434601-96-2
Fluoro, Chloro-methyl_ketone	<chem>[CX3H0](=[OX1])[CH2][Cl,F]</chem>	Z-DEVD-FMK, Z-FA-FMK

Note: These covalent warheads were used to identify putative covalent inhibitors among the known SARS-CoV-2 M-pro inhibitors (Figures S1 and S2) and SARS-CoV-2 M-pro inactive compounds (Figure S3).

PF-00835231 as a possible treatment for COVID-19.⁶⁹ Pfizer has initiated a clinical trial with PF-07304814, a prodrug which metabolizes into PF-00835231, against the M-pros of SARS-CoV-2 and other coronaviruses. GC376 is one of the SARS-CoV-2 M-pro covalent inhibitors that is most commonly used as a reference compound or positive control. Its IC₅₀ activity ranges between 0.03 and 0.89 μM^{64,73,74,78,87-93} and this activity is consistent both in the absence and presence of the reducing agent DTT.⁷⁴ GC376 was developed as an inhibitor of the main protease of the feline coronavirus FCoV, and it also showed activity against the M-pro from MERS and SARS-CoV viruses.⁸⁸ GC376 is a prodrug and its bisulphite adduct is converted to an aldehyde that has been proved to form a covalent interaction with the catalytic Cys145 of the SARS-CoV-2 M-pro.⁸⁸ Several GC376 analogs or derived compounds have shown greater inhibitory activity than GC376.²¹ MPI3, MPI4, and MPI5 are reversible covalent inhibitors derived from the GC376 inhibitor.⁹¹ MPI3, MPI4, and MPI5 display an IC₅₀ value of 0.0085, 0.015, and 0.033 μM, respectively.⁹¹ Although MPI3 is ranked second in potency as a SARS-CoV-2 M-pro inhibitor, its inhibition of SARS-CoV-2 replication is much lower than other compounds, which may be explained by the compound's low cell stability.⁹¹ UAWJ248 is a GC376 analog with an alpha-ketoamide reactive warhead that binds irreversibly to Cys145.⁹⁴ The IC₅₀ of UAWJ248 is 0.012 μM, 3.2-fold more potent than GC376.⁹⁴

Another set of SARS-CoV-2 M-pro inhibitors and inactive compounds can be obtained from the COVID Moonshot project^{95,96} set up to develop M-pro inhibitors. It focused on turning early fragment-screening results into potent compounds.⁹⁶ Drug-designers the world over have proposed more than 10,000 M-pro inhibitors, and over 1000 compounds have been either ordered or synthesized and tested for activity against SARS-CoV-2 M-pro

using two complementary biochemical assays: a fluorescence based assay and a RapidFire Mass Spectrometry assay.⁹⁶ The Fluorescence M-pro inhibition assay consists of a screening in duplicate at 20 and 50 μ M. Compounds with a value greater than 50% M-pro inhibition at 50 μ M are confirmed by a dose response assay and the value of IC_{50} is calculated.⁹⁶ For the RapidFire Mass Spectrometry assay, compounds are triaged by testing the % of inhibition at a final concentration of 20 and 50 μ M, and IC_{50} values are calculated from dose response curves.⁹⁶ From the COVID Moonshot activity data available on the August 18, 2020, we selected those compounds that have a pIC_{50} difference between the two activity assays of less than two units. In addition, further quality control steps were carried out: (i) a compound was not selected if the percentage of inhibition at 20 μ M was higher than the percentage at 50 μ M or (ii) if the percentage of inhibition at 50 μ M and the IC_{50} of a compound were both simultaneously higher than 50. Finally, compounds with undefined chirality were discarded. Figure S2 and Supporting Information File 3 show the final list of the 81 SARS-CoV-2 M-pro inhibitors selected from the COVID Moonshot activity data. The pIC_{50} values of these 81 compounds are between 4.14 and 7.28, and 36 of them contain at least one of the covalent warheads defined in Table 2 (Supporting Information File 3) and are putative covalent inhibitors of the SARS-CoV-2 M-pro.

The activity data available from the COVID Moonshot initiative makes it possible to select a group of inactive compounds against the SARS-CoV-2 M-pro. Although some articles have reported inactive compounds against SARS-CoV-2 M-pro, such as the DDP-4 inhibitors linagliptin, other gliptins and structural analogues,⁹⁷ this kind of information is not usually available. From the COVID Moonshot activity data available on 18 August 2020, we selected the compounds with a percentage of inhibition less than 2% at 50 μ M in the two different assays simultaneously (fluorescence and RapidFire). Compounds with undefined chirality were discarded. Thus, a group of 113 inactive compounds against SARS-CoV-2 were selected (see Figure S3 and Supporting Information File 4). Of these 113 inactive compounds, 29 present at least one of the covalent warheads from Table 2.

The set of 81 M-pro inhibitors from the COVID Moonshot project and the 150 M-pro inhibitors extracted from the literature and presented above could be used to validate a VS procedure and/or a docking step that aims to predict new SARS-CoV-2 M-pro inhibitors. The COVID Moonshot set of active compounds is less diverse than the set taken from the literature, but their inhibitory activities were estimated using the same procedure and laboratory, and all the inhibitors were confirmed by two different assays (fluorescence and RapidFire). In the section below, these sets of compounds are used to check whether protein-ligand docking methods can rank these compounds in order of their potency and if they can discriminate between active and inactive compounds.

4 | DOCKING METHODOLOGY NEEDS TO BE VALIDATED BEFORE USE

Although molecular docking programs can usually generate ligand poses that are similar to the crystallographic conformation for several targets,⁹⁸ protein-ligand docking has several limitations. One of these is that, for practical reasons, most docking methods use a rigid protein approximation, where the ligand is treated as flexible, but the protein conformation is restricted.^{45,99} Some methods, such as ensemble docking which takes place in an ensemble of protein conformations,¹⁰⁰ enable protein flexibility to be analysed. Another limitation is that, contrary to what is expected, the poses and docking scores obtained by docking can be heavily dependent on the ligand input structure,⁴⁹ so small changes in the ligand input conformation can lead to large differences in the docked poses.⁴⁹ One of the major limitations of the scoring functions used by docking programs is that they do not always make a useful prediction of ligand binding affinity.^{36,98} Thus, the best energy score is not a reliable criterion for selecting the best solution in common docking applications.¹⁰¹ For these reasons, docking scores should not be used as a measure or threshold for compound activity, but rather to enrich the library by discarding those compounds that do not fit inside the binding site.³⁷ It should also be noted that the docking scores of different docking programs cannot be directly compared because the scores are dependent on the force fields and protocols used for each program. Another limitation of molecular docking is how they deal with water molecules present in the binding site

during the docking process^{45,102} which can lead to errors when trying to ascertain the possible interaction between the ligand and the target. Finally, the ability of a docking program to separate decoys from active compounds is usually strongly dependent on the protein structure used and the similarity between the screened ligands and the co-crystallized ones.³⁶

We used the aforementioned sets of 150 M-pro inhibitors extracted from the literature and 81 M-pro inhibitors from the COVID Moonshot project to check whether docking scores can be used to predict the potency of a SARS-CoV-2 M-pro inhibitor. To test this hypothesis we performed a protein-ligand docking of both sets of compounds with two different 3D structures of the SARS-CoV-2 M-pro protein, the 6LU7³⁰ and 6WQF,³¹ and three docking programs, i.e., FRED,¹⁰³ Glide¹⁰⁴ and Autodock Vina.¹⁰⁵ For Glide, we used three distinct docking methodologies: Glide HTVS, Glide SP and Glide XP. So in total we considered 5 different docking methods, which appeared to be the most common in the reviewed articles (see Table 1). Briefly, the three methodological approaches from the Schrödinger package Glide differ from each one. Essentially, Glide HTVS and Glide SP use the same docking algorithm, but their main difference is that Glide HTVS minimizes the number of intermediate conformations during the docking procedure and the accuracy of the final torsional refinement and sampling.¹⁰⁶ On the other hand, Glide XP uses a more exhaustive and sophisticated scoring function and sampling process, and is stricter in terms of ligand-receptor shape complementarity.¹⁰⁷

We used the 6LU7 structure³⁰ because it is the most used structure in molecular docking studies of the SARS-CoV-2 M-pro (see Table 1). The 6WQF structure of unliganded SARS-CoV-2 M-pro was also used as it was suggested that this structure may provide more relevant information at physiological temperatures, which could be helpful in protein-ligand molecular docking studies.³¹ Both M-pro structures (6LU7 and 6WQF) were obtained from the Protein Data Bank and were superimposed using the alpha carbons (C α) as reference points. The 6LU7 ligand (N3) was removed from the structure before the protein was prepared. Protein preparation was performed with different tools, depending on the docking software used. For Glide, the structures were prepared with the Protein Preparation Wizard¹⁰⁸ with the following settings: (a) N and C termini were capped; (b) hydrogens were added after original hydrogens had been removed; (c) water molecules were removed; (d) Epik¹⁰⁹ was used to generate tautomers and protonation states with neutral pH; (e) H-bond assignment was optimized with PROPKA¹¹⁰; and (f) the structure was minimized with the default force field. The grid for docking was generated using Glide. The center coordinates corresponded to the centroid of N3 (-10.36, 12.46, 68.7) and the box size was set to 35 × 35 × 35 Å. In the case of FRED, the utility MakeReceptor¹⁰³ was used to (a) define a box, (b) set a shape potential and (c) define the inner and outer contours of the grid based on those used to create the grid for Glide. The preparations needed for AutoDock Vina were performed with AutoDockTools¹¹¹ which consisted of removing all waters and adding polar hydrogens. The grid was based on the same coordinates and size as the previous programs, and the M-pro inhibitors were converted into isomeric SMILES format. Only achiral molecules or molecules with a fully defined chirality were used. Before the docking procedure with Autodock Vina and the three methods of Glide, Omega¹¹² was used to generate the 3D conformation with the lowest energy and Ligprep¹¹³ was used to generate all the possible protonation states in the pH range 7.2 ± 1.0 and the default number of tautomers. The docking program FRED had its own pipeline that included (i) *fixpka*, *tautomer* and *molcharge* from the QUACPAC program¹¹⁴ which set the ionization states, enumerated the tautomeric forms and assigned atomic partial charges, and (ii) Omega¹¹² which generated one conformation per compound and discarded unspecified chiralities.

The most negative docking scores obtained for each compound and each protein-ligand methodology were correlated with the inhibitory potency, measured as pIC₅₀ values. Figure 4 and Figure S4 show that for both datasets, the correlations were far from satisfactory. The correlations improved if covalent and noncovalent inhibitors were treated separately (see Figure 4 and Figure S4). For the 6LU7 structure, and considering solely the covalent inhibitors extracted from the literature, correlations were negative, as one would expect because the more negative the docking score, the higher pIC₅₀ value (Figure 4A–E). The best correlation had a Pearson's *r* coefficient of -0.335 (*p* value 0.003) for the covalent inhibitors when AutoDock Vina was used (Figure 4E). However, the Pearson correlation coefficients for the noncovalent inhibitors were positive or around zero (Figure 4A–E). For the

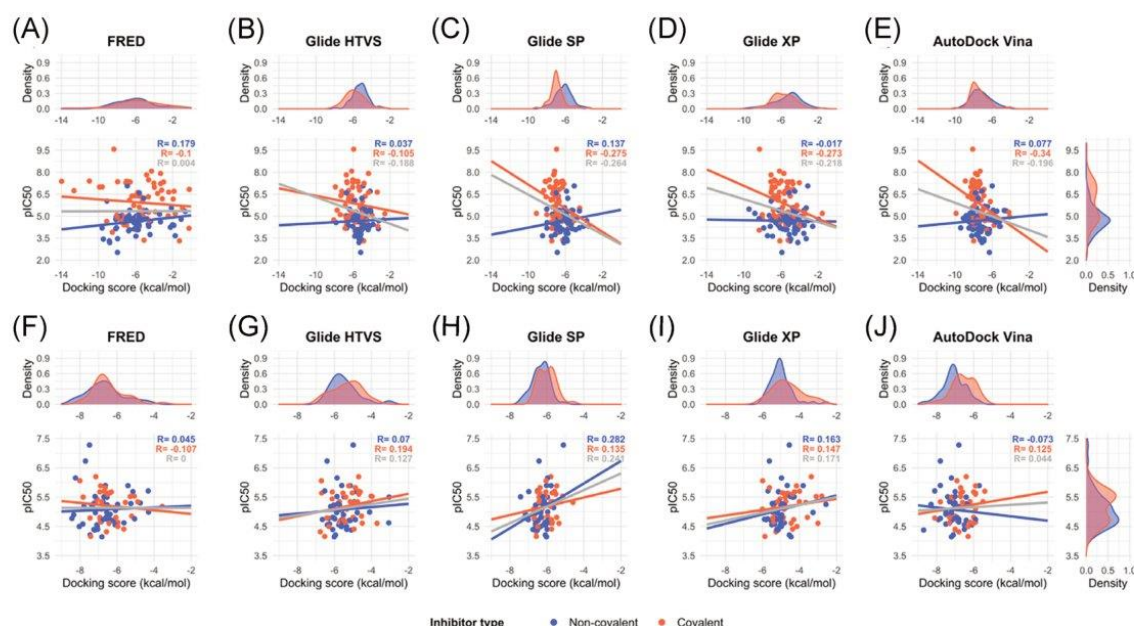


FIGURE 4 Correlation between the highest docking scores from five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) and the experimental inhibitory activity, measured as pIC_{50} , of a group of 150 M-pro inhibitors extracted from the literature (A–E) and 81 M-pro inhibitors from the COVID Moonshot project (F–J). The protein from the 6LU7 PDB structure was used for docking. Data has been adjusted to a linear regression and the correlation coefficient is displayed in gray. Covalent and noncovalent inhibitors are also represented separately in orange and blue, respectively [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

set of 81 M-pro inhibitors from the COVID MoonShot project (Figure 4F–J), most of the correlations were unexpectedly found to be positive. Conclusions were similar when the 6WQF structure was used (Figure S4). This lack of correlation between the bioactivities and docking scores for this target was previously reported^{115,116} and questions the reliability of using a docking score as a cutoff value for selecting new putative M-pro inhibitors or predicting the relative bioactivity of a compound by comparison with reference compounds or even with other predicted compounds.

To overcome the limitations of docking programs, they can be combined with other methods. Standard molecular dynamics or binding free energy estimations, such as MM-GBSA and MM-PBSA enable the protein flexibility to be explored. MM-GBSA or MM-PBSA are popular methods for estimating ligand-binding affinities with a modest computational effort.¹¹⁷ Their results are considered to be more accurate than those of most docking scoring functions¹¹⁸ and they have been widely used in virtual screening protocols to find M-pro inhibitors (see Supporting Information File 1). Because these methods are not suitable for screening large libraries of compounds,³⁶ they are usually used to analyze compounds selected by molecular docking. The best docking poses of a selected group of hits can be the input of the MM-GBSA or MM-PBSA algorithms which can be used to estimate their ligand-binding affinities, expressed as a ΔG value. The combination of molecular docking and MM-GBSA or MM-PBSA re-scoring has proven to be a promising strategy for identifying the correct binding poses and correctly ranking the binding affinities of a series of ligands.¹¹⁸ However, the success of these methods depends on the target structure¹¹⁷ and they cannot be used to predict true drug candidates in drug design without experimental verification because their accuracy is relatively limited.¹¹⁸ To assess how successfully these methods predict the bioactivity ranking of the two sets of known M-pro inhibitors presented above, we used the MM-GBSA minimization available at Prime¹¹⁹ to re-score the best docking pose obtained with the five docking methods and two SARS-CoV-2 M-pro structures, and to estimate the ΔG value of ligand-binding affinity for each compound. Figure 5 and Figure S5 show the correlation

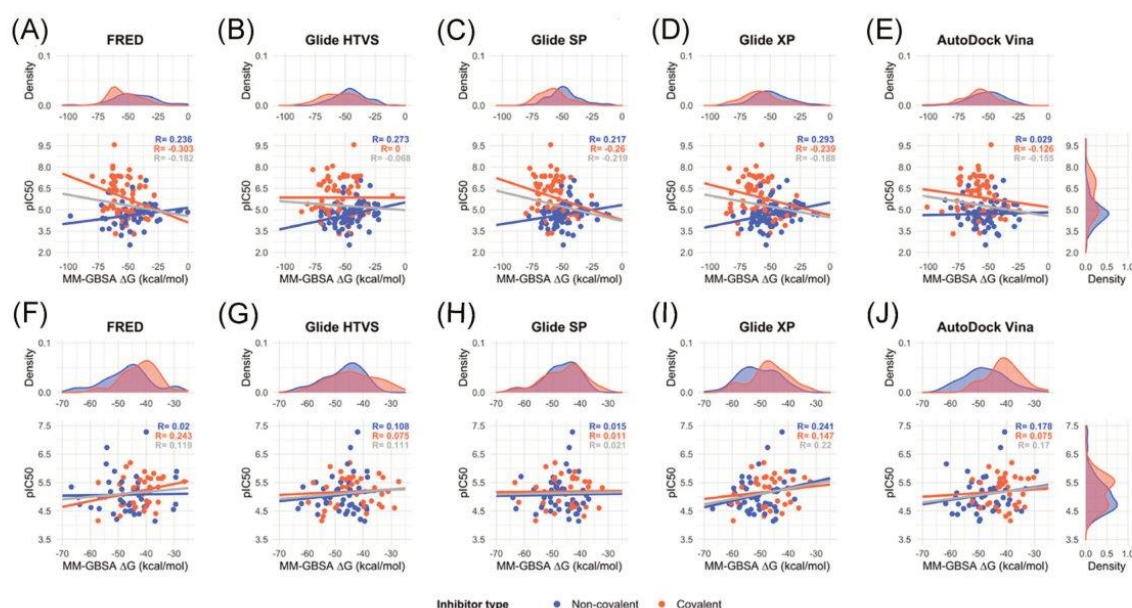


FIGURE 5 Correlation between the binding energy, calculated as the ΔG of the MM-GBSA method, and the experimental inhibitory activity, measured as pIC_{50} , of a group of 150 M-pro inhibitors extracted from the literature (A–E) and 81 M-pro inhibitors from the COVID Moonshot project (F–J). The protein from the 6LU7 PDB structure was used for docking. The poses with the highest docking scores from each of the five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP, and AutoDock Vina) were used as the input to the MM-GBSA program to estimate the binding energy. Data has been adjusted to a linear regression and the correlation coefficient is displayed in gray. Covalent and noncovalent inhibitors are also represented separately in orange and blue, respectively. MM-GBSA, molecular mechanics generalized Born surface area [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/med.21862)]

between the pIC_{50} and the estimated ΔG values. A negative correlation is expected, because the higher the pIC_{50} of a compound is, the lower the ΔG value. Figure 5 and Figure S5 show that most correlations were positive, even though noncovalent and covalent inhibitors were treated separately. Moderate negative correlations were obtained for covalent inhibitors from the bibliography (Figure 5A–E and Figure S5A–E). The best Pearson r coefficient (with a value of -0.303) was obtained with the covalent inhibitors from the bibliography docked to the 6LU7 structure with the FRED program (Figure 5A). Thus, for the inhibitors and M-pro structures assayed here, the MM-GBSA algorithm was not able to correct the lack of correlation obtained with the docking programs.

Although the docking scores or ΔG values estimated with the MM-GBSA algorithm of known SARS-CoV-2 M-pro inhibitors do not correlate with their potency, we wanted to find out whether these parameters made it possible to discriminate between SARS-CoV-2 M-pro inhibitors and compounds with no SARS-CoV-2 M-pro inhibitory activity. For this reason, we added a group of inactive compounds to the set of COVID Moonshot inhibitors, that is to say, compounds with no activity against M-pro in the two assays used in the COVID Moonshot project. As inactive compounds do not have IC_{50} values, we used ROC curves to test whether M-pro inhibitors could be distinguished from inactive compounds by either the docking score or ΔG estimated with MM-GBSA. Again, we used the five aforementioned docking methods and the two SARS-CoV-2 M-pro structures. ROC curves are used to assess the performance of a classification method. They represent true positive versus false positive rates and the area under the ROC curve (AUC) measures the performance across all possible classification thresholds. AUC values range from 0 to 1, and represent the probability that a random positive example is positioned before a random negative example, when they are sorted by decreasing docking scores or ΔG values. A random two-class classification has an AUC value of 0.5. Figure 6 and Figure S6 show the ROC curves and AUC values for the COVID

Moonshot M-pro inhibitors and inactive compounds presented above when using the two M-pro structures. Most of the AUC values obtained were near to 0.5 (Figure 6 and Figure S6), which means that M-pro inhibitors generally cannot be distinguished from inactive compounds when using docking scores or ΔG values estimated by the MM-GBSA method. In some cases, the predictions improved if noncovalent and covalent inhibitors were treated separately (Figure 6 and Figure S6). The best AUC obtained was 0.744 when the 6WQF M-pro structure and the FRED docking poses of the noncovalent inhibitors were used to estimate the ΔG values with the MM-GBSA

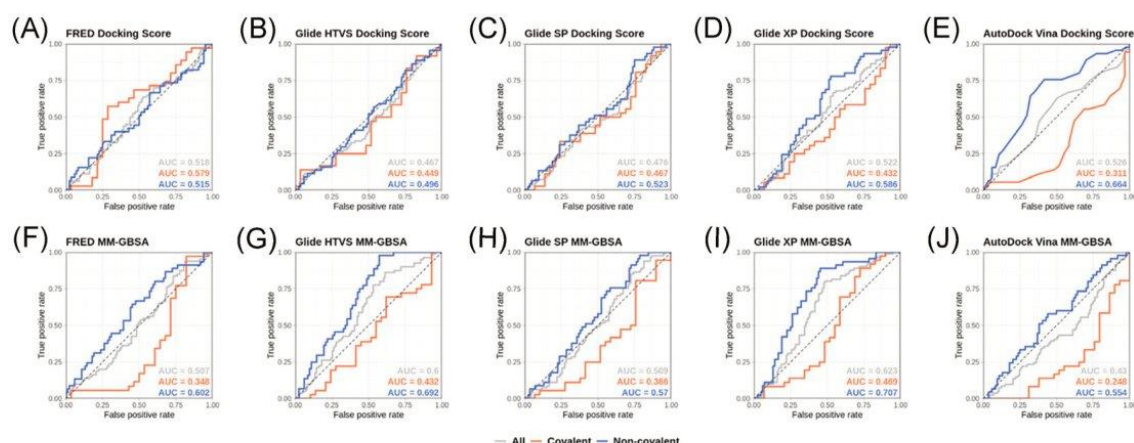


FIGURE 6 ROC curves obtained from the docking scores of five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) (A–E) and the ΔG calculated with the MM-GBSA program (F–J) of a group of known SARS-CoV-2 M-pro inhibitors and compounds with no activity against M-pro from the COVID Moonshot initiative. The protein from the 6LU7 PDB structure was used for docking. Covalent, noncovalent and the whole set of inhibitors are represented separately. The area under the curve (AUC) values are displayed for each group. MM-GBSA, molecular mechanics generalized Born surface area [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/mr.21862)]

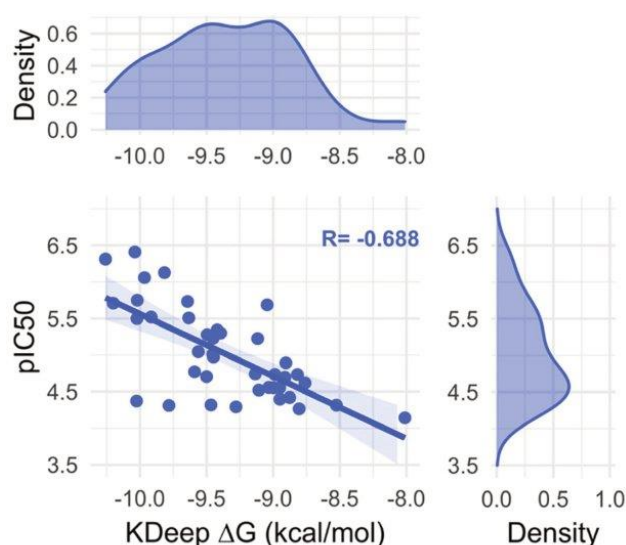


FIGURE 7 Correlation between the pIC_{50} and the ΔG values estimated with the KDeep server from the crystallized protein-ligand complexes of a group of 40 noncovalent SARS-CoV-2 M-pro inhibitors that were co-crystallized with the SARS-CoV-2 M-pro protein [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/mr.21862)]

method (Figure S6F). For the 6LU7 structure, the best AUC was obtained with the Glide XP docking poses of the noncovalent inhibitors and ΔG values estimated with the MM-GBSA method (Figure 6I).

To test if there is a correlation between the pIC_{50} of a group of SARS-CoV-2 M-pro inhibitors and their ΔG values without using docking, from the COVID Moonshot activity data we selected a group of 40 noncovalent inhibitors that were co-crystallized with the SARS-CoV-2 M-pro (Supporting Information File 5). The 3-D structures of the protein-ligand complex of these compounds, available from Fragalysis,¹²⁰ were used to estimate the protein-ligand affinity, expressed as ΔG values, using the K_{Deep} server.¹²¹ In this server, the structure of the protein and the ligand is represented using 8 pharmacophoric-like features (i.e., hydrophobic, aromatic, hydrogen-bond donor and acceptor, positive and negative ionizable, metallic and total excluded volume) and protein-ligand affinity is predicted using a Deep Convolutional Neural Network pre-trained with the PDBbind v.2016 database.¹²¹ Figure 7 shows that the correlation between the pIC_{50} and ΔG values is good, with a Pearson correlation coefficient of -0.688 (p value $9.32E-07$). However, this correlation decreased when the ligands were transformed to SMILES format to lose the binding coordinates of the ligand and were docked with the five docking programs used above (Figure 8 and Figure S7). The best correlation, with a Pearson coefficient value of -0.359 (p value 0.023), was obtained with the FRED docking program, a long way from the correlation of -0.688 obtained with the crystallographic poses. This shows that for the two SARS-CoV-2 M-pro structures and the five docking methods used here, these methodologies were not able to predict the crystallographic pose as the best docking pose, as it has recently been shown.¹²² Again this reinforces the idea that it is essential that docking results be validated before use.

5 | OUTLOOK

Although the SARS-CoV-2 pandemic needed an accelerated process of drug development, measures are required to safeguard this process. Dotolo and coworkers put this idea into words with the proverb “haste makes waste”, and remarked that “what is true for popular wisdom is doubly true for scientific research”.⁵³ Overconfident statements and conclusions that are not fully supported by experimental evidence are frequently found in articles that use protein-ligand docking to find SARS-CoV-2 M-pro inhibitors.⁵⁴ The methods used and the results obtained by computational approaches must be validated so that these methods are not devalued and efforts focus on the most promising compounds. If docking scores are used to rank or select a set of compounds or to compare the potency of two inhibitor compounds, it must first be shown that there is a correlation between docking scores and activity or potency, for example expressed as the IC_{50} . If there is no such correlation, there is no reason to apply this

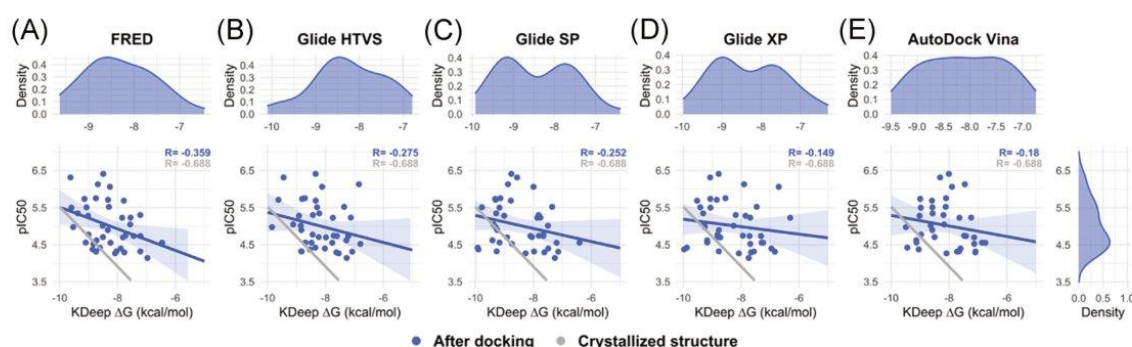


FIGURE 8 Correlation between the pIC_{50} and the ΔG values estimated with the K_{Deep} server from the best docking poses from five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) (A–E) and the 6LU7 M-pro structure of a group of 40 noncovalent SARS-CoV-2 M-pro inhibitors that were co-crystallized with the SARS-CoV-2 M-pro protein [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

methodology. A VS protocol can be computationally validated by applying the same VS protocol to a set of both active and inactive compounds. Therefore, the methodology is expected to be able to selectively identify the known active compounds from the validation set. The use of inactive compounds for this validation is crucial because they act as a negative control. Of course, the validation of a VS protocol with active and inactive compounds requires prior knowledge of the target, so it is not applicable if information is scarce, as is the case when targets are new or poorly understood. However, it must be used as an efficient validation tool when more information is available. The set of SARS-CoV-2 M-pro inhibitors and inactive compounds reviewed here could be helpful in validating future screenings.

The flexibility and plasticity of the SARS-CoV-2 M-pro make it a challenging target for small-molecule inhibitor design.⁷¹ Thus, techniques that have been applied to other targets are not always effective. When using any methodology, such as protein-ligand docking, it is important to consider not only its well-known advantages, but also its weaknesses. Binding free energies solely based on docking poses are, in general, insufficiently reliable.¹¹⁶ Here, in two large groups of SARS-CoV-2 M-pro inhibitors, using two different SARS-CoV-2 M-pro structures and five protein-ligand docking methods we have shown that there is not a good correlation between the bioactivity and such affinity scores, as the docking scores or the ΔG calculated with the MM-GBSA method. In addition, these affinity scores cannot distinguish between compounds with or without M-pro inhibitory activity, especially if covalent and noncovalent inhibitors are analyzed together. This is not to suggest that protein-ligand docking is useless. On the contrary, it is indeed an invaluable tool for reducing the initial compounds in a virtual screening when used in the initials steps, for plausible pose prediction and for semiquantitative assessment of the inhibitory power of a ligand.¹¹⁶ There are some successful examples of the use of protein-ligand docking as an initial step to identify new SARS-CoV-2 M-pro inhibitors,^{46,63,123–125} but none of them used protein-ligand docking alone. Consensus docking, i.e., the use of different protein-ligand docking programs, has proved to be also helpful.^{29,123} When designing a new VS procedure to find SARS-CoV-2 M-pro inhibitors it should be considered if either covalent or noncovalent inhibitors (or both) are expected. The covalent warheads provided herein can be used to differentiate between potentially covalent and noncovalent M-pro inhibitors and to treat both sets separately. Covalent docking^{126,127} is an alternative when the aim is to find covalent inhibitors, and it has been used successfully to find M-pro inhibitors.^{128,129} In addition to docking, other computational methods have been used to identify new SARS-CoV-2 M-pro inhibitors. These methods include pharmacophores^{76,130,131}; combined quantum mechanics/molecular mechanics (QM/MM) methods^{132,133}; a Docking Consensus Approach and Common Hits Approach that combine molecular dynamics simulations with molecular docking and pharmacophores, respectively¹³¹; interactive molecular dynamics simulation in virtual reality¹³⁴; Quantitative Structure-activity Relationship models,¹¹⁵ and binding free energies calculations.^{82,116,135} Among the various methods to calculate protein-ligand binding affinities, Free Energy Perturbation (FEP) has been used successfully to predict new M-pro inhibitors.^{46,124,136} Advances in computer power, enhanced sampling algorithms, and increase of force field accuracy has led to newer FEP implementations that has enabled accurate and reliable calculations of protein-ligand binding free energies.¹³⁷ Only when combined with other methods, should the results of automated protein-ligand docking be considered worthy for further experimental testing.

ACKNOWLEDGMENTS

We thank OpenEye Scientific Software and Cresset for kindly providing us with a software bursary to use their programs. This manuscript has been edited by the English language service of our university. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 713679 and from the Universitat Rovira i Virgili (grant 2019PFR-URV-B2-69).

ORCID

Guillem Macip  <http://orcid.org/0000-0003-4016-5558>
 Pol Garcia-Segura  <http://orcid.org/0000-0002-0169-824X>
 Júlia Mestres-Truyol  <http://orcid.org/0000-0002-8552-5281>
 Bryan Saldivar-Espinoza  <http://orcid.org/0000-0002-9667-2818>
 María José Ojeda-Montes  <http://orcid.org/0000-0002-9290-1935>
 Aleix Gimeno  <http://orcid.org/0000-0002-7654-6689>
 Adrià Cereto-Massagué  <http://orcid.org/0000-0002-9319-6377>
 Santiago Garcia-Vallvé  <http://orcid.org/0000-0002-0348-7497>
 Gerard Pujadas  <http://orcid.org/0000-0003-2598-8089>

REFERENCES

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1
- Worldometers Coronavirus Update (Live). Accessed March 22, 2021. <https://www.worldometers.info/coronavirus/>
- Nicola M, Alsafi Z, Sohrabi C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. *Int J Surg*. 2020;78:185-193. doi:10.1016/j.ijssu.2020.04.018
- Odono A, Galea S, Stuckler D, Signorelli C. University Vita-Salute San Raffaele COVID-19 literature monitoring working group. The first 10 000 COVID-19 papers in perspective: are we publishing what we should be publishing? *Eur J Public Health*. 2020;30(5):849-850. doi:10.1093/eurpub/ckaa170
- Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. *Nat Hum Behav*. 2020;4(7):666-669. doi:10.1038/s41562-020-0911-0
- Guha R, Willighagen E, Zdrzil B, Jeliakova N. What is the role of cheminformatics in a pandemic? *J Cheminform*. 2021;13:16. doi:10.1186/s13321-021-00491-6
- Glasziou PP, Sanders S, Hoffmann T. Waste in covid-19 research. *Br Med J*. 2020;369:m1847. doi:10.1136/bmj.m1847
- Masters PS. The molecular biology of coronaviruses. *Adv Virus Res*. 2006;66:193-292. doi:10.1016/S0065-3527(06)66005-3
- Mishra SK, Tripathi T. One year update on the COVID-19 pandemic: where are we now? *Acta Trop*. 2021;214:105778. doi:10.1016/j.actatropica.2020.105778
- Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*. 2003;348(20):1953-1966. doi:10.1056/NEJMoa030781
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367(19):1814-1820. doi:10.1056/NEJMoa1211721
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5(4):536-544. doi:10.1038/s41564-020-0695-z
- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3
- Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. *Cell*. 2020;181(4):914-921. doi:10.1016/j.cell.2020.04.011.e10.
- Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol*. 2020;92(4):418-423. doi:10.1002/jmv.25681
- Wang M, Cao R, Zhang L, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*. 2020;30(3):269-271. doi:10.1038/s41422-020-0282-0
- Furuta Y, Komeno T, Nakamura T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc Jpn Acad Ser B Phys Biol Sci*. 2017;93(7):449-463. doi:10.2183/pjab.93.027
- Amin SA, Banerjee S, Gayen S, Jha T. Protease targeted COVID-19 drug discovery: what we have learned from the past SARS-CoV inhibitors? *Eur J Med Chem*. 2021;215:113294. doi:10.1016/j.ejmech.2021.113294
- Petushkova AI, Zamyatnin AA. Papain-like proteases as coronaviral drug targets: current inhibitors, opportunities, and limitations. *Pharmaceuticals*. 2020;13(10):277. doi:10.3390/ph13100277
- Banerjee R, Perera L, Tillekeratne LMV. Potential SARS-CoV-2 main protease inhibitors. *Drug Discov Today*. 2021;26(3):804-816. doi:10.1016/j.drudis.2020.12.005
- Xiong M, Su H, Zhao W, Xie H, Shao Q, Xu Y. What coronavirus 3C-like protease tells us: from structure, substrate selectivity, to inhibitor design. *Med Res Rev*. 2021;41:1965-1998. doi:10.1002/med.21783

22. Monteil V, Kwon H, Prado P, et al. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell*. 2020;181(4):905-913.e7. doi:10.1016/j.cell.2020.04.004
23. Gautret P, Lagier J-C, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents*. 2020;56(1):105949. doi:10.1016/j.ijantimicag.2020.105949
24. Nojomi M, Yassin Z, Keyvani H, et al. Effect of Arbidol (Umifenovir) on COVID-19: a randomized controlled trial. *BMC Infect Dis*. 2020;20(1):954. doi:10.1186/s12879-020-05698-w
25. Lee WS, Wheatley AK, Kent SJ, DeKosky BJ. Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nat Microbiol*. 2020;5(10):1185-1191. doi:10.1038/s41564-020-00789-5
26. Zhou G, Zhao Q. Perspectives on therapeutic neutralizing antibodies against the Novel Coronavirus SARS-CoV-2. *Int J Biol Sci*. 2020;16(10):1718-1723. doi:10.7150/ijbs.45123
27. Stasi C, Fallani S, Voller F, Silvestri C. Treatment for COVID-19: an overview. *Eur J Pharmacol*. 2020;889:173644. doi:10.1016/j.ejphar.2020.173644
28. Tu Y-F, Chien C-S, Yarmishyn AA, et al. A Review of SARS-CoV-2 and the ongoing clinical trials. *Int J Mol Sci*. 2020;21(7):2657. doi:10.3390/ijms21072657
29. Gimeno A, Mestres-Truyol J, Ojeda-Montes MJ, et al. Prediction of novel inhibitors of the main protease (M-pro) of SARS-CoV-2 through consensus docking and drug reposition. *Int J Mol Sci*. 2020;21(11):3793. doi:10.3390/ijms21113793
30. Jin Z, Du X, Xu Y, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*. 2020;582(7811):289-293. doi:10.1038/s41586-020-2223-y
31. Kneller DW, Phillips G, O'Neill HM, et al. Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat Commun*. 2020;11(1):3202. doi:10.1038/s41467-020-16954-7
32. Flare, version 4.0.2., Cresset®, Litlington, Cambridgeshire, UK.
33. Cheeseright T, Mackey M, Rose S, Vinter A. Molecular field extrema as descriptors of biological activity: definition and validation. *J Chem Inf Model*. 2006;46(2):665-676. doi:10.1021/ci050357s
34. Bauer MR, Mackey MD. Electrostatic complementarity as a fast and effective tool to optimize binding and selectivity of protein-ligand complexes. *J Med Chem*. 2019;62(6):3036-3050. doi:10.1021/acs.jmedchem.8b01925
35. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18(1):41-58. doi:10.1038/nrd.2018.168
36. Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. *Int J Mol Sci*. 2019;20(18):4331. doi:10.3390/ijms20184331
37. Gimeno A, Ojeda-Montes MJ, Tomás-Hernández S, et al. The light and dark sides of virtual screening: what is there to know? *Int J Mol Sci*. 2019;20(6):1375. doi:10.3390/ijms20061375
38. Cui W, Yang K, Yang H. Recent progress in the drug development targeting SARS-CoV-2 main protease as treatment for COVID-19. *Front Mol Biosci*. 2020;7:616341. doi:10.3389/fmolb.2020.616341
39. Pillaiyar T, Wendt LL, Manickam M, Easwaran M. The recent outbreaks of human coronaviruses: a medicinal chemistry perspective. *Med Res Rev*. 2021;41(1):72-135. doi:10.1002/med.21724
40. Wang X, Guan Y. COVID-19 drug repurposing: a review of computational screening methods, clinical trials, and protein interaction assays. *Med Res Rev*. 2021;41(1):5-28. doi:10.1002/med.21728
41. Cannalire R, Cerchia C, Beccari AR, Di Leva FS, Summa V. Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: state of the art and future opportunities. *J Med Chem*. 2020. <https://pubs.acs.org/doi/10.1021/acs.jmedchem.0c01140>
42. Xu J, Xue Y, Zhou R, Shi P, Li H, Zhou J. Drug repurposing approach to combating coronavirus: potential drugs and drug targets. *Med Res Rev*. 2021;41(3):1375-1426. doi:10.1002/med.21763
43. Chen CC, Yu X, Kuo CJ, et al. Overview of antiviral drug candidates targeting coronaviral 3C-like main proteases. *FEBS J*. 2021;288(17):5089-5121. doi:10.1111/febs.15696
44. Lengauer T, Rarey M. Computational methods for biomolecular docking. *Curr Opin Struct Biol*. 1996;6(3):402-406. doi:10.1016/s0959-440x(96)80061-3
45. Pansar T, Poso A. Binding affinity via docking: fact and fiction. *Molecules*. 2018;23(8):1899. doi:10.3390/molecules23081899
46. Li Z, Li X, Huang Y-Y, et al. Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc Natl Acad Sci USA*. 2020;117(44):27381-27387. doi:10.1073/pnas.2010470117
47. Fiorucci D, Milletti E, Orofino F, Brizzi A, Mugnaini C, Corelli F. Computational drug repurposing for the identification of SARS-CoV-2 main protease inhibitors. *J Biomol Struct Dyn*. 2020;39(16):5121-6248. doi:10.1080/07391102.2020.1796805

48. Kanhed AM, Patel DV, Teli DM, Patel NR, Chhabria MT, Yadav MR. Identification of potential Mpro inhibitors for the treatment of COVID-19 by using systematic virtual screening approach. *Mol Divers*. 2021;25(1):383-401. doi:10.1007/s11030-020-10130-1
49. Feher M, Williams CI. Effect of input differences on the results of docking calculations. *J Chem Inf Model*. 2009;49(7):1704-1714. doi:10.1021/ci9000629
50. Mohan V, Gibbs A, Cummings M, Jaeger E, DesJarlais R. Docking: successes and Challenges. *Curr Pharm Des*. 2005;11(3):323-333. doi:10.2174/1381612053382106
51. Meyer-Almes F-J. Repurposing approved drugs as potential inhibitors of 3CL-protease of SARS-CoV-2: virtual screening and structure based drug design. *Comput Biol Chem*. 2020;88:107351. doi:10.1016/j.compbiolchem.2020.107351
52. Mandour YM, Zlotos DP, Alaraby Salem M. A multi-stage virtual screening of FDA-approved drugs reveals potential inhibitors of SARS-CoV-2 main protease. *J Biomol Struct Dyn*. 2020;1-12. doi:10.1080/07391102.2020.1837680
53. Dotolo S, Marabotti A, Facchiano A, Tagliaferri R. A review on drug repurposing applicable to COVID-19. *Brief Bioinform*. 2020;22(September):726-741. bbaa288 doi:10.1093/bib/bbaa288
54. Bellera CL, Llanos M, Gantner ME, et al. Can drug repurposing strategies be the solution to the COVID-19 crisis? *Expert Opin Drug Discov*. 2021;16(6):605-612. doi:10.1080/17460441.2021.1863943
55. Arun KG, Sharanya CS, Abhithaj J, Francis D, Sadasivan C. Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening, molecular docking and molecular dynamics with main protease as the target. *J Biomol Struct Dyn*. 2020;39:1-12. doi:10.1080/07391102.2020.1779819
56. Mitra K, Ghanta P, Acharya S, Chakrapani G, Ramaiah B, Doble M. Dual inhibitors of SARS-CoV-2 proteases: pharmacophore and molecular dynamics based drug repositioning and phytochemical leads. *J Biomol Struct Dyn*. 2020;39(16):6324-6337. doi:10.1080/07391102.2020.1796802
57. J A, Francis D, C SS, K GA, C S, Variyar EJ. Repurposing simeprevir, calpain inhibitor IV and a cathepsin F inhibitor against SARS-CoV-2 and insights into their interactions with Mpro. *J Biomol Struct Dyn*. 2020;1-12. doi:10.1080/07391102.2020.1813200
58. Mahdi M, Mótán JA, Szojka ZI, Golda M, Miczi M, Tózsér J. Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2's main protease. *Virol J*. 2020;17:190. doi:10.1186/s12985-020-01457-0
59. Li Y, Xie Z, Lin W, et al. Efficacy and safety of lopinavir/ritonavir or arbidol in adult patients with mild/moderate COVID-19: an exploratory randomized controlled trial. *Med*. 2020;1(1):105-113.e4. doi:10.1016/j.medj.2020.04.001
60. Cao B, Wang Y, Wen D, et al. A trial of lopinavir-ritonavir in adults hospitalized with severe covid-19. *N Engl J Med*. 2020;382(19):1787-1799. doi:10.1056/NEJMoa2001282
61. RECOVERY Collaborative G. Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet*. 2020;396(10259):1345-1352. doi:10.1016/S0140-6736(20)32013-4
62. Chakraborti S, Bheemireddy S, Srinivasan N. Repurposing drugs against the main protease of SARS-CoV-2: mechanism-based insights supported by available laboratory and clinical data. *Mol Omi*. 2020;16(5):474-491. doi:10.1039/d0mo00057d
63. Vatansever EC, Yang KS, Drelich AK, et al. Bepridil is potent against SARS-CoV-2 in vitro. *Proc Natl Acad Sci USA*. 2021;118(10):e2012201118. doi:10.1073/pnas.2012201118
64. Hung HC, Ke YY, Huang SY, et al. Discovery of M protease inhibitors encoded by SARS-CoV-2. *Antimicrob Agents Chemother*. 2020;64(9):e00872-20. doi:10.1128/AAC.00872-20
65. Hattori S, Higshi-Kuwata N, Raghavaiah J, et al. GRL-0920, an Indole Chloropyridinyl Ester, Completely Blocks SARS-CoV-2 Infection. Weiss LM, ed. *MBio*. 2020;11(4):e01833-20. doi:10.1128/mBio.01833-20
66. Mittal L, Kumari A, Srivastava M, Singh M, Asthana S. Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. *J Biomol Struct Dyn*. 2020;39(10):1-19. doi:10.1080/07391102.2020.1768151
67. Cavasotto CN, Di Filippo JL. In silico Drug Repurposing for COVID-19: targeting SARS-CoV-2 Proteins through Docking and Consensus Ranking. *Mol Inform*. 2021;40(1):e2000115. doi:10.1002/minf.202000115
68. Su HX, Yao S, Zhao WF, et al. Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol Sin*. 2020;41(9):1167-1177. doi:10.1038/s41401-020-0483-6
69. Hoffman RL, Kania RS, Brothers MA, et al. Discovery of Ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. *J Med Chem*. 2020;63(21):12725-12747. doi:10.1021/acs.jmedchem.0c01063
70. Qiao J, Li Y-S, Zeng R, et al. SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science*. 2021;371(6536):1374-1378. doi:10.1126/science.abf1611

71. Bzówka M, Mitusińska K, Raczynska A, Samol A, Tuszyński JA, Góra A. Structural and evolutionary analysis indicate that the SARS-CoV-2 Mpro is a challenging target for small-molecule inhibitor design. *Int J Mol Sci*. 2020;21(9):3099. doi:10.3390/ijms21093099
72. Kumar A, Loharch S, Kumar S, Ringe RP, Parkesh R. Exploiting cheminformatic and machine learning to navigate the available chemical space of potential small molecule inhibitors of SARS-CoV-2. *Comput Struct Biotechnol J*. 2021;19:424-438. doi:10.1016/j.csbj.2020.12.028
73. Gurard-Levin ZA, Liu C, Jekle A, et al. Evaluation of SARS-CoV-2 3C-like protease inhibitors using self-assembled monolayer desorption ionization mass spectrometry. *Antiviral Res*. 2020;182:104924. doi:10.1016/j.antiviral.2020.104924
74. Ma C, Hu Y, Townsend JA, et al. Ebselen, disulfiram, carmofur, PX-12, tideglusib, and shikonin are nonspecific promiscuous SARS-CoV-2 main protease inhibitors. *ACS Pharmacol Transl Sci*. 2020;3(6):1265-1277. doi:10.1021/acspstsci.0c00130
75. Jan J-T, Cheng TR, Juang Y-P, et al. Identification of existing pharmaceuticals and herbal medicines as inhibitors of SARS-CoV-2 infection. *Proc Natl Acad Sci USA*. 2021;118(5):e2021579118. doi:10.1073/pnas.2021579118
76. Pathak N, Chen Y-T, Hsu Y-C, et al. Uncovering flexible active site conformations of SARS-CoV-2 3CL proteases through protease pharmacophore clusters and COVID-19 drug repurposing. *ACS Nano*. 2021;15(1):857-872. doi:10.1021/acsnano.0c07383
77. Bhimraj A, Morgan RL, Shumaker AH, et al. Infectious diseases society of america guidelines on the treatment and management of patients with coronavirus disease 2019 (COVID-19). *Clin Infect Dis*. 2020. ciaa478 doi:10.1093/cid/ciaa478
78. Ma C, Sacco MD, Hurst B, et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res*. 2020;30(8):678-692. doi:10.1038/s41422-020-0356-z
79. Rawson JMO, Duchon A, Nikolaitchik OA, Pathak VK, Hu W-S. Development of a cell-based luciferase complementation assay for identification of SARS-CoV-2 3CLpro inhibitors. *Viruses*. 2021;13(2):173. doi:10.3390/v13020173
80. Bauer RA. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov Today*. 2015;20(9):1061-1073. doi:10.1016/j.drudis.2015.05.005
81. Lonsdale R, Burgess J, Colclough N, et al. Expanding the armory: predicting and tuning covalent warhead reactivity. *J Chem Inf Model*. 2017;57(12):3124-3137. doi:10.1021/acs.jcim.7b00553
82. Mondal D, Warshel A. Exploring the mechanism of covalent inhibition: simulating the binding free energy of α -ketoamide inhibitors of the main protease of SARS-CoV-2. *Biochemistry*. 2020;59(48):4601-4608. doi:10.1021/acs.biochem.0c00782
83. Kneller DW, Phillips G, O'Neill HM, et al. Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: insights into enzyme mechanism and drug design. *IUCrJ*. 2020;7(Pt 6):1028-1035. doi:10.1107/S2052252520012634
84. Kneller DW, Phillips G, Weiss KL, et al. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J Biol Chem*. 2020;295(50):17365-17373. doi:10.1074/jbc.AC120.016154
85. Verma N, Henderson JA, Shen J. Proton-coupled conformational activation of SARS coronavirus main proteases and opportunity for designing small-molecule broad-spectrum targeted covalent inhibitors. *J Am Chem Soc*. 2020;142(52):21883-21890. doi:10.1021/jacs.0c10770
86. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. *Nat Rev Drug Discov*. 2011;10(4):307-317. doi:10.1038/nrd3410
87. Rathnayake AD, Zheng J, Kim Y, et al. 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV-infected mice. *Sci Transl Med*. 2020;12(557):eabc5332. doi:10.1126/scitranslmed.abc5332
88. Vuong W, Khan MB, Fischer C, et al. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat Commun*. 2020;11:4282. doi:10.1038/s41467-020-18096-2
89. Fu L, Ye F, Feng Y, et al. Both boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat Commun*. 2020;11:4417. doi:10.1038/s41467-020-18233-x
90. Wang Y-C, Yang W-H, Yang C-S, et al. Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum anti-coronaviral drug. *Am J Cancer Res*. 2020;10(8):2535-2545.
91. Yang KS, Ma XR, Ma Y, et al. A quick route to multiple highly potent SARS-CoV-2 main protease inhibitors. *ChemMedChem*. 2021;16(6):942-948. doi:10.1002/cmdc.202000924
92. Zhu W, Xu M, Chen CZ, et al. Identification of SARS-CoV-2 3CL protease inhibitors by a quantitative high-throughput screening. *ACS Pharmacol Transl Sci*. 2020;3(5):1008-1016. doi:10.1021/acspstsci.0c00108
93. Iketani S, Forouhar F, Liu H, et al. Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors. *Nat Commun*. 2021;12(1):2016. doi:10.1038/s41467-021-22362-2

94. Sacco MD, Ma C, Lagarias P, et al. Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M pro and cathepsin L. *Sci Adv.* 2020;6(50):eabe0751. doi:10.1126/sciadv.abe0751
95. COVID Moonshot. Accessed March 22, 2021. <https://covid.postera.ai/covid>
96. Achdout H, Aimon A, Bar-David E, et al. COVID moonshot: open science discovery of SARS-CoV-2 main protease inhibitors by combining crowdsourcing, high-throughput experiments, computational simulations, and machine learning. *bioRxiv.* 2020;2020 doi:10.1101/2020.10.29.339317
97. Nar H, Schnapp G, Hücke O, Hardman TC, Klein T. Action of dipeptidyl peptidase-4 inhibitors on SARS-CoV-2 main protease. *ChemMedChem.* 2021;16(9):1425-1426. doi:10.1002/cmdc.202000921
98. Warren GL, Andrews CW, Capelli A-M, et al. A critical assessment of docking programs and scoring functions. *J Med Chem.* 2006;49(20):5912-5931. doi:10.1021/jm050362n
99. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem.* 2004;47(1):45-55. doi:10.1021/jm030209y
100. Amaro RE, Baudry J, Chodera J, et al. Ensemble docking in drug discovery. *Biophys J.* 2018;114(10):2271-2278. doi:10.1016/j.bpj.2018.02.038
101. Ramírez D, Caballero J. Is it reliable to take the molecular docking top scoring position as the best solution without considering available structural data? *Molecules.* 2018;23(5):1038. doi:10.3390/molecules23051038
102. Sethi A, Joshi K, Sasikala K, Alvala M. Molecular docking in modern drug discovery: principles and recent applications. In: Gaitonde V, Karmakar P, Trivedi A, eds. *Drug Discovery and Development - New Advances.* IntechOpen 2020. doi:10.5772/intechopen.85991
103. OEDOCKING 3.4.0.2: OpenEye Scientific Software, Santa Fe, NM, USA, 2019.
104. Schrödinger Release 2020-3. *Glide, Schrödinger.* New York, NY: LLC, 2020.
105. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31(2):455-461. doi:10.1002/jcc.21334
106. Friesner RA, Banks JL, Murphy RB, 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. doi:10.1021/jm0306430
107. Friesner RA, Murphy RB, Repasky MP, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem.* 2006;49(21):6177-6196. doi:10.1021/jm051256o
108. Schrödinger Release 2020-3. *Protein Preparation Wizard, Schrödinger.* New York, NY, USA: LLC; 2020.
109. Schrödinger Release 2020-3. *Epik, Schrödinger.* New York, NY, USA: LLC; 2020.
110. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput.* 2011;7(2):525-537. doi:10.1021/ct100578z
111. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem.* 2009;30(16):2785-2791. doi:10.1002/jcc.21256
112. OMEGA 3.1.1.2: OpenEye Scientific Software, Santa Fe, NM, USA, 2019.
113. Schrödinger Release 2020-3. *LigPrep, Schrödinger.* New York, NY: LLC; 2020.
114. QUACPAC 2.0.2.2: OpenEye Scientific Software, Santa Fe, NM, USA, 2019.
115. Alves VM, Bobrowski T, Melo-Filho CC, et al. QSAR modeling of SARS-CoV M pro inhibitors identifies sufugolix, cenicriviroc, proglumetacin, and other drugs as candidates for repurposing against SARS-CoV-2. *Mol Inform.* 2021; 40(1):2000113. doi:10.1002/minf.202000113
116. Macchiagodena M, Pagliai M, Karrenbrock M, Guarnieri G, Iannone F, Procacci P. Virtual double-system single-box: a nonequilibrium alchemical technique for absolute binding free energy calculations: application to ligands of the SARS-CoV-2 main protease. *J Chem Theory Comput.* 2020;16(11):7160-7172. doi:10.1021/acs.jctc.0c00634
117. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov.* 2015;10(5):449-461. doi:10.1517/17460441.2015.1032936
118. Wang E, Sun H, Wang J, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev.* 2019;119(16):9478-9508. doi:10.1021/acs.chemrev.9b00055
119. Schrödinger Release 2020-3. *Prime, Schrödinger.* New York, NY, USA: LLC; 2020.
120. Mpro: Fragalysis. Accessed March 22, 2021. <https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro>
121. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J Chem Inf Model.* 2018;58(2):287-296. doi:10.1021/acs.jcim.7b00650
122. Zev S, Raz K, Schwartz R, Tarabeh R, Gupta PK, Major DT. Benchmarking the ability of common docking programs to correctly reproduce and score binding modes in SARS-CoV-2 protease Mpro. *J Chem Inf Model.* 2021;61(6): 2957-2966. doi:10.1021/acs.jcim.1c00263
123. Ghahremanpour MM, Tirado-Rives J, Deshmukh M, et al. Identification of 14 known drugs as inhibitors of the main protease of SARS-CoV-2. *ACS Med Chem Lett.* 2020;11(12):2526-2533. doi:10.1021/acsmedchemlett.0c00521

124. Zhang C-H, Stone EA, Deshmukh M, et al. Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations. *ACS Cent Sci*. 2021;7(3):467-475. doi:10.1021/acscentsci.1c00039
125. Gupta A, Rani C, Pant P, et al. Structure-based virtual screening and biochemical validation to discover a potential inhibitor of the SARS-CoV-2 main protease. *ACS Omega*. 2020;5(51):33151-33161. doi:10.1021/acsomega.0c04808
126. Scarpino A, Ferenczy GG, Keserű GM. Comparative evaluation of covalent docking tools. *J Chem Inf Model*. 2018;58(7):1441-1458. doi:10.1021/acs.jcim.8b00228
127. Kumalo H, Bhakat S, Soliman M. Theory and applications of covalent docking in drug discovery: merits and pitfalls. *Molecules*. 2015;20(2):1984-2000. doi:10.3390/molecules20021984
128. Shitrit A, Zaidman D, Kalid O, et al. Conserved interactions required for inhibition of the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Sci Rep*. 2020;10:20808. doi:10.1038/s41598-020-77794-5
129. Amendola G, Ettari R, Previti S, et al. Lead discovery of SARS-CoV-2 main protease inhibitors through covalent docking-based virtual screening. *J Chem Inf Model*. 2021;61(4):2062-2073. doi:10.1021/acs.jcim.1c00184
130. Kouznetsova VL, Huang DZ, Tsigelny IF. Potential SARS-CoV-2 protease Mpro inhibitors: Repurposing FDA-approved drugs. *Phys Biol*. 2021;18(2):025001. doi:10.1088/1478-3975/abcb66
131. Battisti V, Wieder O, Garon A, Seidel T, Urban E, Langer T. A computational approach to identify potential novel inhibitors against the coronavirus SARS-CoV-2. *Mol Inform*. 2020;39(10):e2000090. doi:10.1002/minf.202000090
132. Arafet K, Serrano-Aparicio N, Lodola A, et al. Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem Sci*. 2021;12(4):1433-1444. doi:10.1039/d0sc06195f
133. Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I. Multiscale simulations of SARS-CoV-2 3CL protease inhibition with aldehyde derivatives. Role of protein and inhibitor conformational changes in the reaction mechanism. *ACS Catal*. 2021;11(7):4157-4168. doi:10.1021/acscatal.0c05522
134. Deeks HM, Walters RK, Barnoud J, Glowacki DR, Mulholland AJ. Interactive molecular dynamics in virtual reality is an effective tool for flexible substrate and inhibitor docking to the SARS-CoV-2 main protease. *J Chem Inf Model*. 2020;60(12):5803-5814. doi:10.1021/acs.jcim.0c01030
135. Jukič M, Janežič D, Bren U. Ensemble docking coupled to linear interaction energy calculations for identification of coronavirus main protease (3CLpro) Non-Covalent Small-Molecule Inhibitors. *Molecules*. 2020;25(24):5808. doi:10.3390/molecules25245808
136. Ngo ST, Tam NM, Pham MQ, Nguyen TH. Benchmark of popular free energy approaches revealing the inhibitors binding to SARS-CoV-2 Mpro. *J Chem Inf Model*. 2021;61(5):2302-2312. doi:10.1021/acs.jcim.1c00159
137. Wang L, Chambers J, Abel R. Protein-ligand binding free energy calculations with FEP. In: Bonomi M, Camilloni C, eds. *Biomolecular Simulations. Methods in Molecular Biology*. Vol 2022. Humana; 2019:201-232. doi:10.1007/978-1-4939-9608-7_9

AUTHOR BIOGRAPHIES

Guillem Macip Sancho is a PhD student at the Cheminformatics & Nutrition Research Group of the URV currently focused in finding inhibitors of SARS-CoV-2 main protease (Mpro). His academic studies previous to this position were a Bachelor Degree in Biochemistry by the Autonomous University of Barcelona (UAB) and a Master's Degree in Bioinformatics and Biostatistics by the Open University of Catalonia—University of Barcelona (UOC-UB). As a student trainee in the Department of Biophysics at the UAB he has worked in the analysis of *Mycoplasma genitalium* membrane proteome in order to find therapeutic targets against its infection.

Pol Garcia-Segura is an undergraduate student of a double BSc. in Biochemistry and Molecular Biology and Biotechnology at Rovira i Virgili University (URV). He has been working as a trainee student in the Cheminformatics & Nutrition Research Group at URV for a couple of years, performing protein-ligand docking simulations, genomic sequences analysis and data analysis.

Júlia Mestres-Truyol is an undergraduate student of a double BSc. in Biochemistry and Molecular Biology and Biotechnology at Rovira i Virgili University (URV). She has been part of the Cheminformatics & Nutrition

Research Group at URV for two years as a trainee student, where she has worked on protein-ligand molecular docking, lead optimization and drug development as well as data analysis.

Bryan P. Saldivar-Espinoza has a bachelor degree in telecommunications engineering and he is a PhD student at the Cheminformatics & Nutrition Research Group of the URV. He has worked applying Machine learning on medical images for tuberculosis and anemia diagnosis. Similarly, he has applied machine learning methods to text mining in a public policy context and process optimization in a business operation framework.

María José Ojeda-Montes holds a Bsc. in Biotechnology and a Msc. in Nutrition and Metabolism, both by Universitat Rovira i Virgili (URV), and a Msc. in Biostatistics and Bioinformatics by Universitat Oberta de Catalunya (UOC). She obtained the PhD in 2017 in the Cheminformatics & Nutrition Research Group of the URV. Her PhD thesis was focused on the improvement of virtual screening performance using ligand-based and receptor-based methods with the aim of identifying bioactive inhibitors for interesting targets in the medicinal chemistry field, such as DPP-IV involved in Type 2 Diabetes. Then, she was working in Intelligent Pharma company as a computational chemist consultant for different pharma clients and European projects. María José has joined the Molecular Modeling Group of the Swiss Institute Bioinformatics (SIB) as a Postdoctoral Computational Chemist in September 2020 developing and providing methods for computer-aided drug design, which are applied to academic and industrial drug discovery.

Aleix Gimeno acquired a bachelor in Biochemistry and Molecular Biology at the university Universitat Rovira i Virgili (URV, Tarragona) in 2013. In 2013, he was also an exchange student of the Erasmus program at the Institute of Immunology from the Universitetet i Oslo (Norway). In 2014, he acquired a master of Bioinformatics at the Universitat Autònoma de Barcelona. He was awarded with the CSA Trust Grant in 2015. He obtained the PhD in 2017 at the Cheminformatics & Nutrition Research Group of the URV. Currently, he is a member of the Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), at The Barcelona Institute of Science and Technology.

Adrià Cereto-Massagué acquired a bachelor in Biotechnology by the Universitat Rovira i Virgili (URV, Tarragona) in 2011, a MSc in Nutrition and Metabolism in 2012, and PhD in Cheminformatics in 2017 by the same university. Currently, he is the lead bioinformatician of the Center for Omic Sciences (COS) in Reus (Catalonia, Spain), where he works on data analysis and integration from different omics sources. In his beginning as a researcher, he was introduced to the fields of cheminformatics, structure-activity relationship (SAR) and machine learning, in which he developed tools aid in the process of virtual screening and target fishing.

Santiago Garcia-Vallvé received his Bachelor degree in Chemistry in 1995 and received his PhD in Biochemistry in 1999 at the Universitat Rovira i Virgili (URV) in Tarragona, Spain. Since 2000 he is member of the Department of Biochemistry and Biotechnology at URV, where in 2003 he won a permanent position. From 1995 to 2009 his research included the development of several databases, servers and tools in the fields of bioinformatics, sequence analysis and molecular evolution for, among others, the prediction of horizontally transferred genes, highly expressed genes and phylogenetic analysis. In 2001, he was a visitor of the Computational Genomics Group at the European Bioinformatics Institute (EBI) in Cambridge, UK. At present, he is member of the Cheminformatics & Nutrition Research Group at URV. His current research interests include the use of cheminformatics tools to detect new bioactive compounds for a given target and the development of new tools to improve virtual screening performance and predict the bioactivity of a compound. He has been involved in the development of several virtual screening procedures for predicting anti-inflammatory (i.e., IKK-2 inhibitors) or anti-diabetic (i.e., DPP-IV inhibitors and PPAR-gamma partial agonists) compounds and tools for target

fishing. In summer 2016, he was a visitor at the Bender Group at the Department of Chemistry, Cambridge University, UK. Since 2020, part of the research of the Cheminformatics & Nutrition Research Group is focused on SARS-CoV-2 and finding inhibitors of the main protease (Mpro).

Gerard Pujadas is graduate in Chemistry by the Universitat de Barcelona (1992) and PhD in Chemistry by the Universitat Rovira i Virgili (URV, Tarragona) in 1998. In his beginning as a researcher, he was introduced to the fields of Bioinformatics and protein structure analysis and its relationship with sequence evolution. In 2001 he made a post-doctoral stay at Prof. Richard Haser's lab at the Institut de Biologie et Chimie des Protéines (Lyon, France). After this post-doctoral stay, he returned to the URV where in September 2002 he won a permanent position at the Biochemistry and Biotechnology Department. Since then, his main contributions to science has been the development of: (1) several virtual screening workflows for predicting molecules that act as anti-inflammatories (i.e., IKK-2 inhibitors) or anti-diabetics (i.e., DPP-IV inhibitors and PPAR-gamma partial agonists); (2) tools for validating virtual screening workflows (i.e., DecoyFinder); and (3) tools for validating the coordinates of drug-target complexes (i.e., VHELIBS). Since April 2014, he leads the research group on "Cheminformatics & Nutrition" at the URV that aims to: (1) develop potent antidiabetic compounds from lead molecules identified with our virtual screening workflows; and (2) develop new tools and databases for improving virtual screening workflows performance.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Macip G, Garcia-Segura P, Mestres-Truyol J, et al. Haste makes waste: a critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Med Res Rev.* 2022;42:744-769. doi:10.1002/med.21862

Supplementary material for

Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition

Guillem Macip¹ | Pol Garcia-Segura¹ | Júlia Mestres-Truyol¹ | Bryan Saldivar-Espinoza¹ | María José Ojeda-Montes² | Aleix Gimeno³ | Adrià Cereto-Massagué⁴ | Santiago Garcia-Vallvé^{1,5} | Gerard Pujadas^{1,5}

¹Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Research group in Cheminformatics & Nutrition, Tarragona, Tarragona, Spain

²Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

³Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

⁴EURECAT Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Spain

⁵EURECAT, TECNIO, CEICS, Avinguda Universitat 1, Reus, Spain

Supplementary figure 1. 2D structure of the 150 known SARS-CoV-2 M-pro inhibitors extracted from the bibliography. Covalent warheads are highlighted in red. 2D structures were drawn from the SMILES using the RDKit library.

Supplementary figure 2. 2D structure of the 81 SARS-CoV-2 M-pro inhibitors from the COVID Moonshot project used for validation. Covalent warheads are highlighted in red. 2D structures were drawn from the SMILES using the RDKit library.

Supplementary figure 3. 2D structure of the 113 compounds from the COVID Moonshot project that are inactive against SARS-CoV-2 M-pro and which were used for validation. Covalent warheads are highlighted in red. 2D structures were drawn from the SMILES using the RDKit library.

Supplementary figure 4. Correlation between the highest docking scores from five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) and the experimental inhibitory activity, measured as pIC_{50} , of a group of 150 M-pro inhibitors extracted from the literature (A-E) and 81 M-pro inhibitors from the COVID Moonshot project (F-J). The protein from the 6WQF PDB structure was used for docking. Data has been adjusted to a linear regression and the correlation coefficient is displayed in gray. Covalent and non-covalent inhibitors are also represented separately in orange and blue, respectively.

Supplementary figure 5. Correlation between the binding energy, calculated as the ΔG of the MM-GBSA method, and the experimental inhibitory activity, measured as pIC_{50} , of a group of 150 M-pro inhibitors extracted from the literature (A-E) and 81 M-pro inhibitors from the COVID Moonshot project (F-J). The protein from the 6WQF PDB structure was used for docking. The poses with the highest docking scores from each of the five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) were used as the input to the MM-GBSA program to estimate the binding energy. Data has been adjusted to a linear regression and the correlation coefficient is displayed in gray. Covalent and non-covalent inhibitors are also represented separately in orange and blue, respectively.

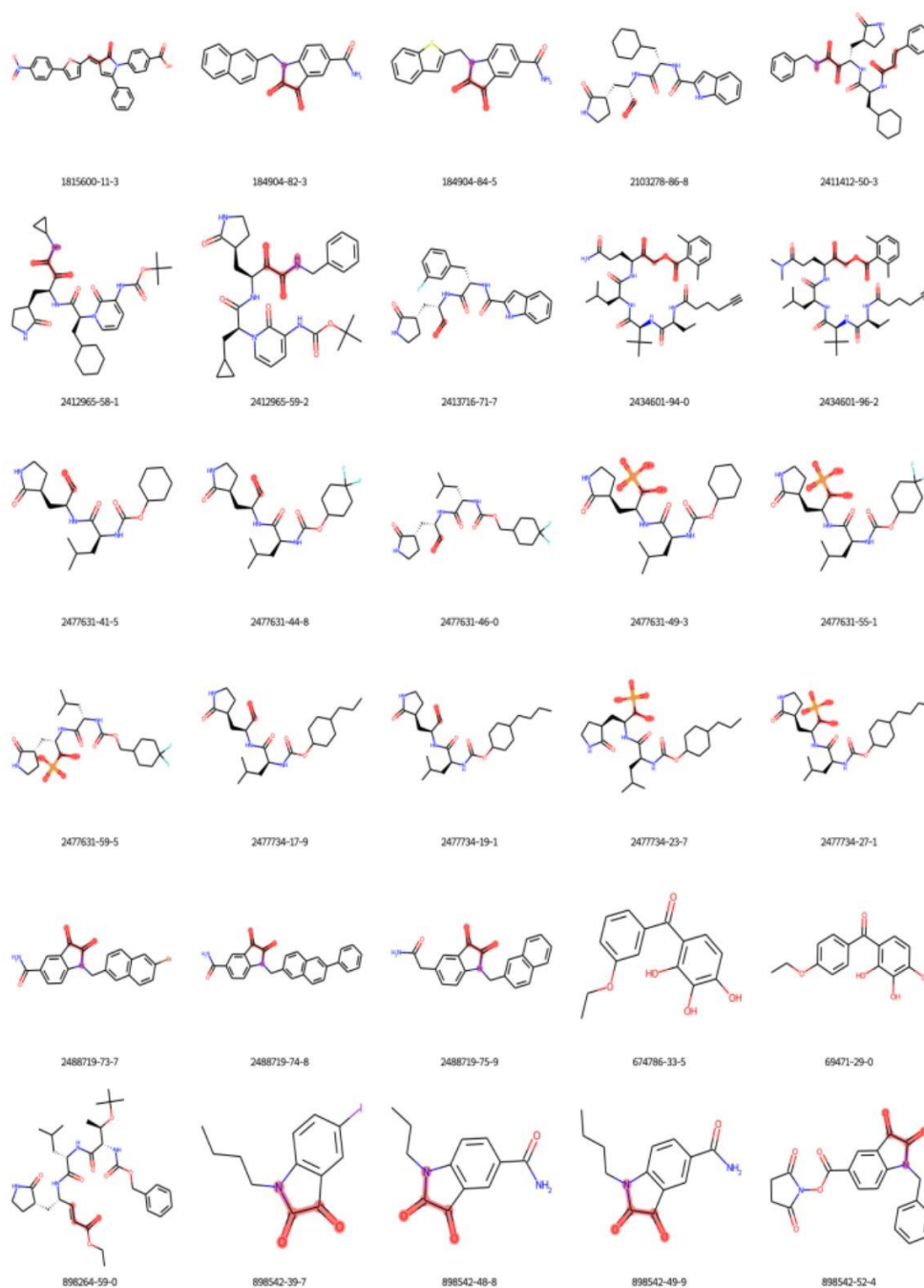
Supplementary figure 6. ROC curves obtained from the docking scores of five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) (A-E) and the ΔG calculated with the MM-GBSA program (F-J) of a group of known SARS-CoV-2 M-pro inhibitors and compounds with no activity against M-pro from the COVID Moonshot initiative. The protein from the 6WQF PDB structure was used for docking. Covalent, non-covalent and the whole set of inhibitors are represented separately. The area under the curve (AUC) values are displayed for each group.

Supplementary figure 7. Correlation between the pIC_{50} and the ΔG values estimated with the KDeep server from the best docking poses from five different docking methods (FRED, Glide HTVS, Glide SP, Glide XP and AutoDock Vina) (A-E) and the 6WQF M-pro structure of a group of 40 non-covalent SARS-CoV-2 M-pro inhibitors that were co-crystallized with the SARS-CoV-2 M-pro protein.

Macip et al.

2

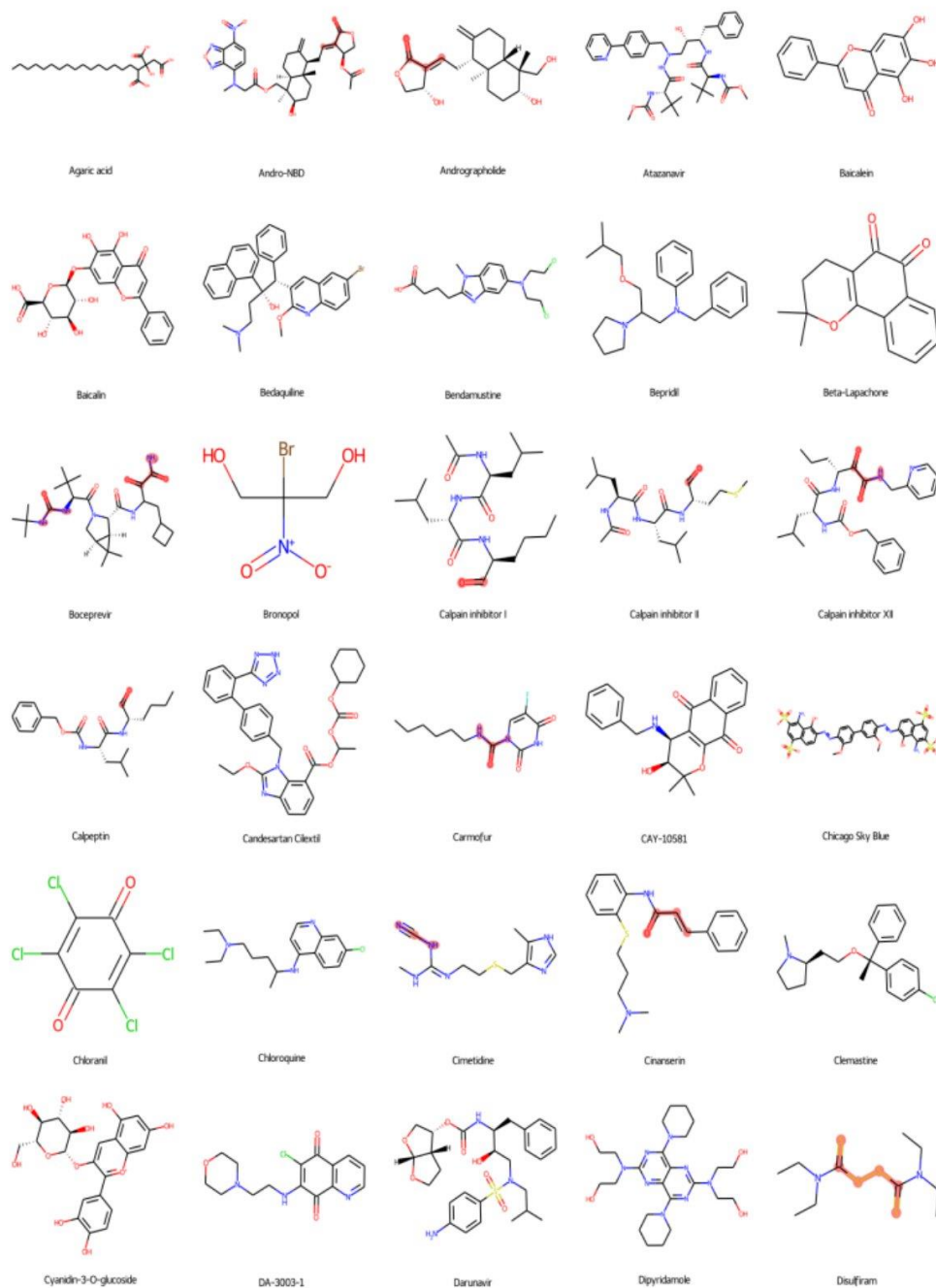
Figure S1



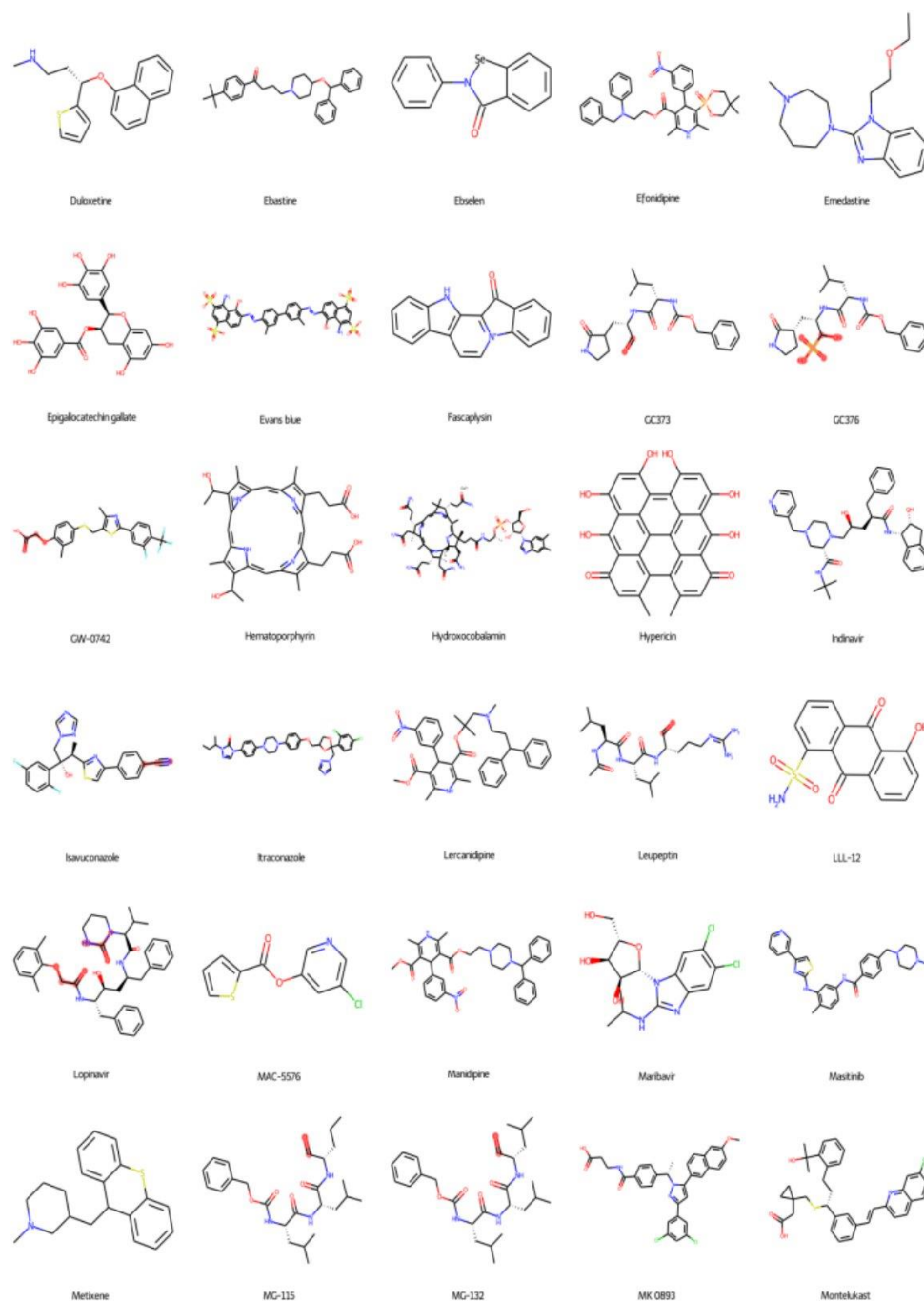
Page 1 of 5

Macip et al.

3

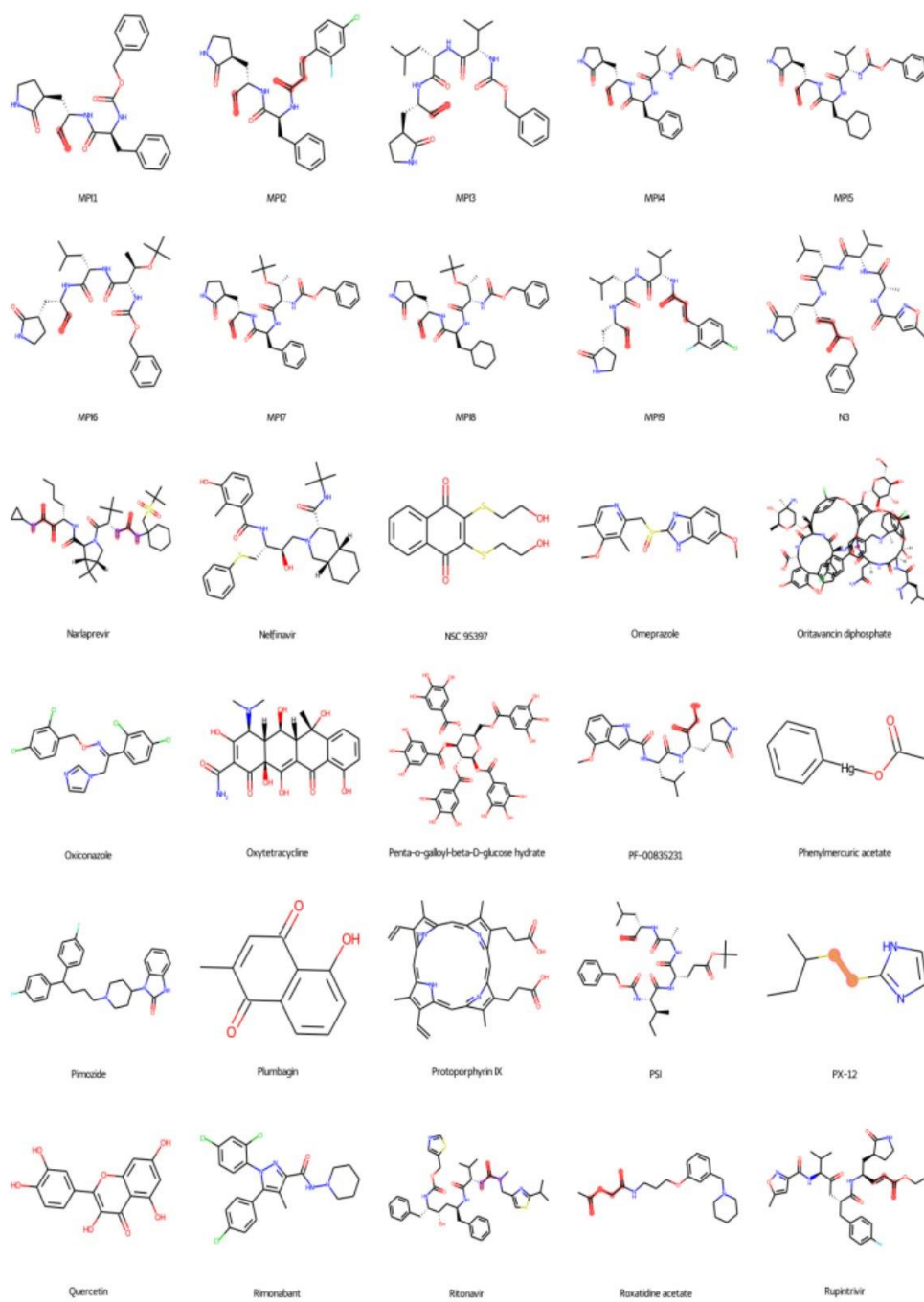


Page 2 of 5

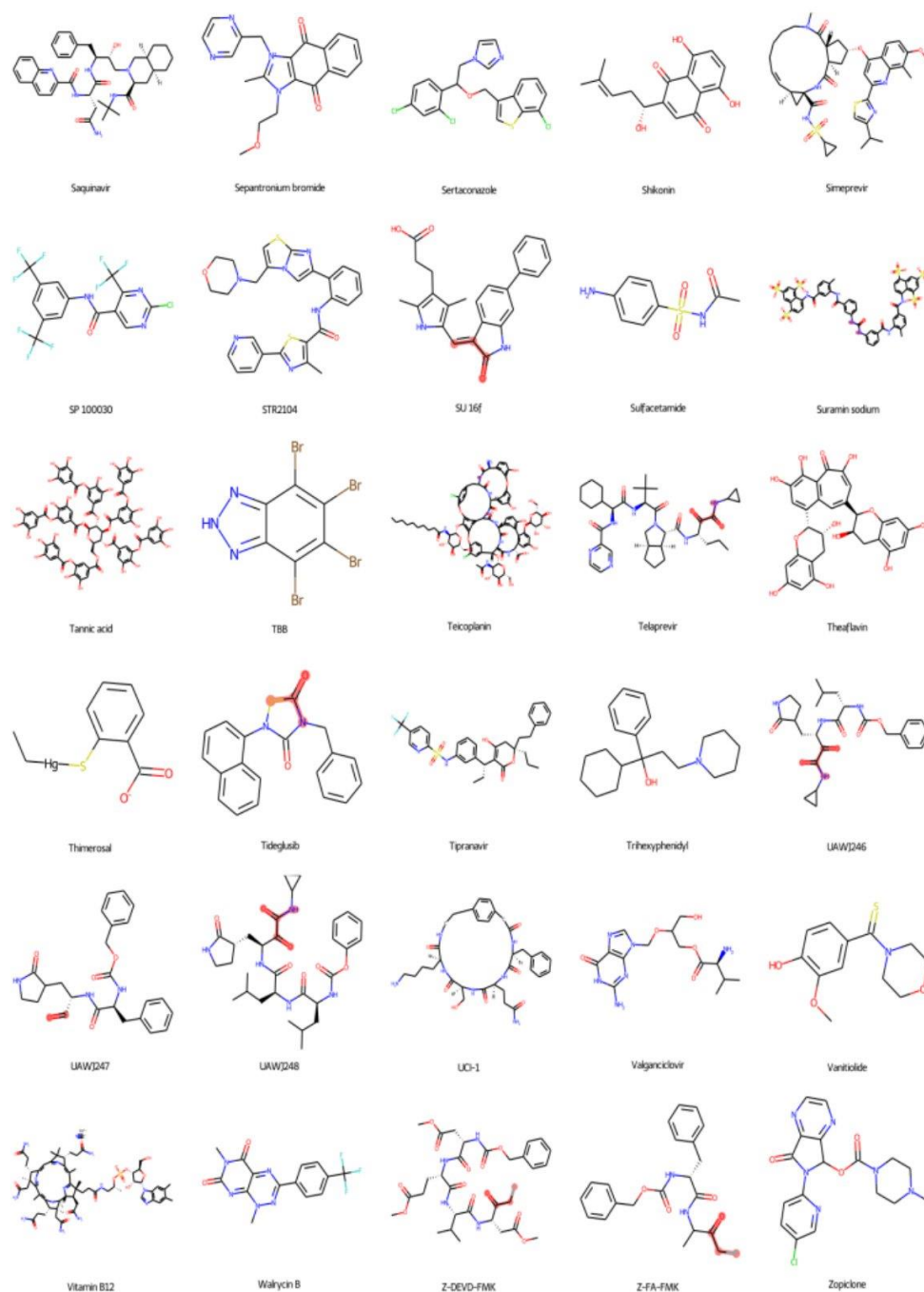


Macip et al.

5



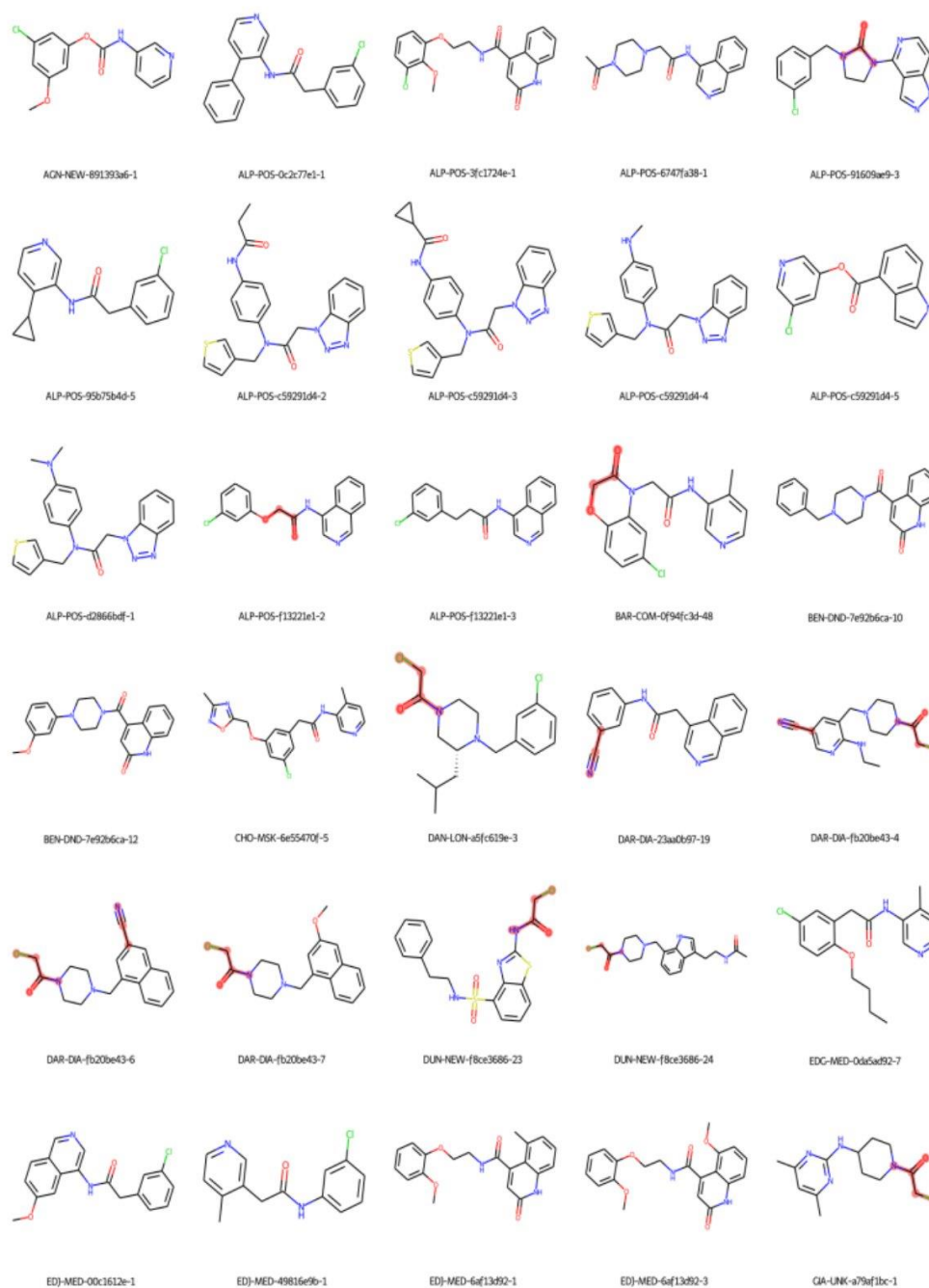
Page 4 of 5



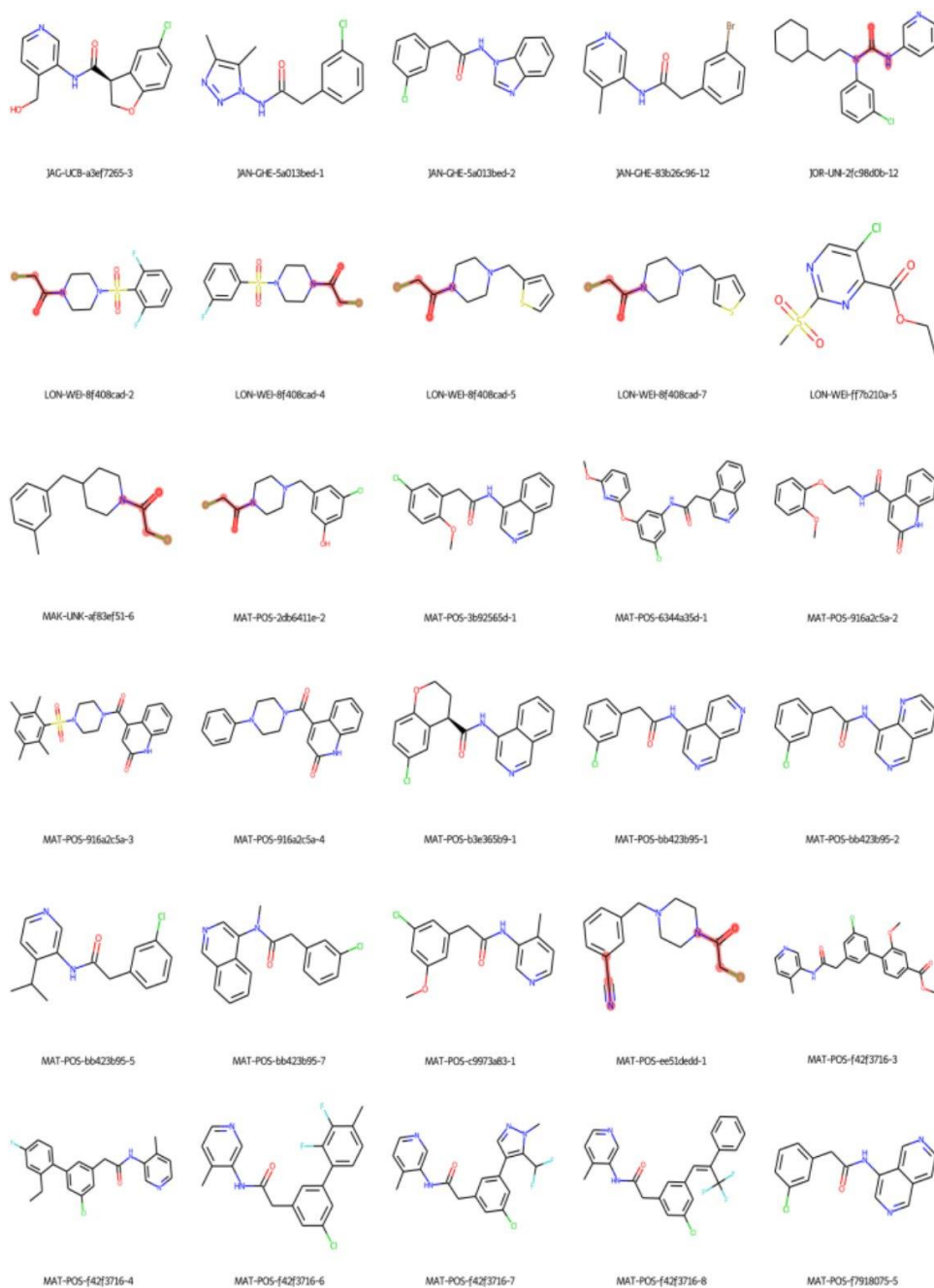
Macip et al.

7

Figure S2

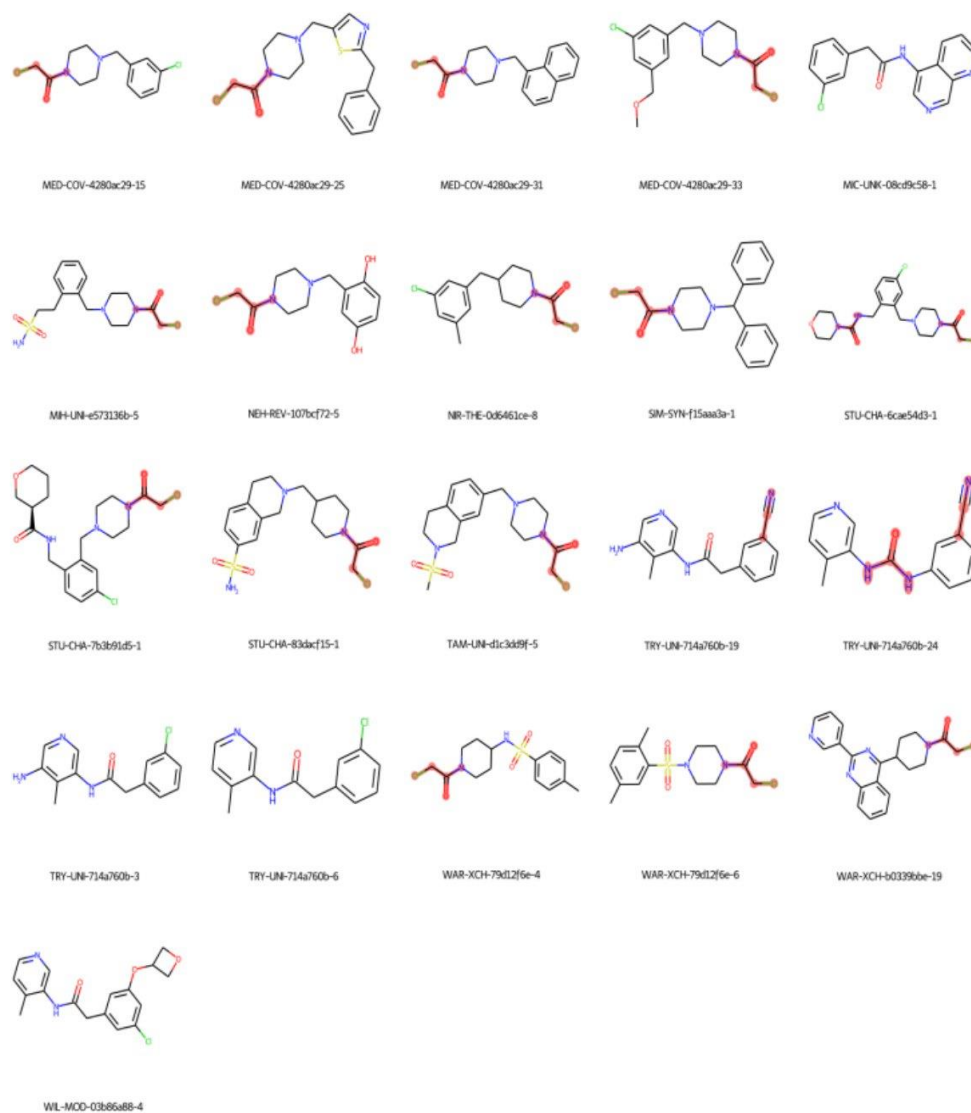


Page 1 of 3



Macip et al.

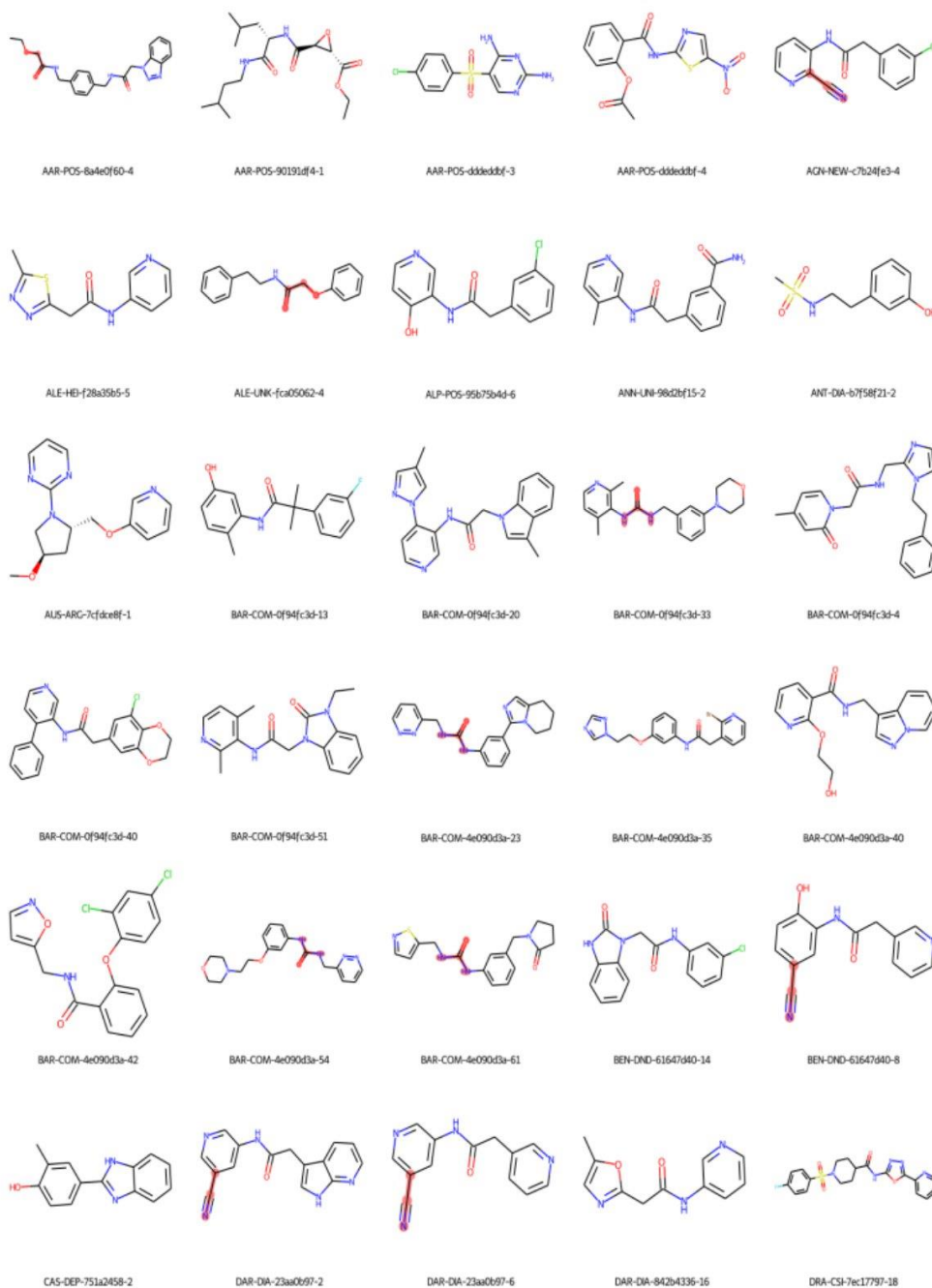
9



Macip et al.

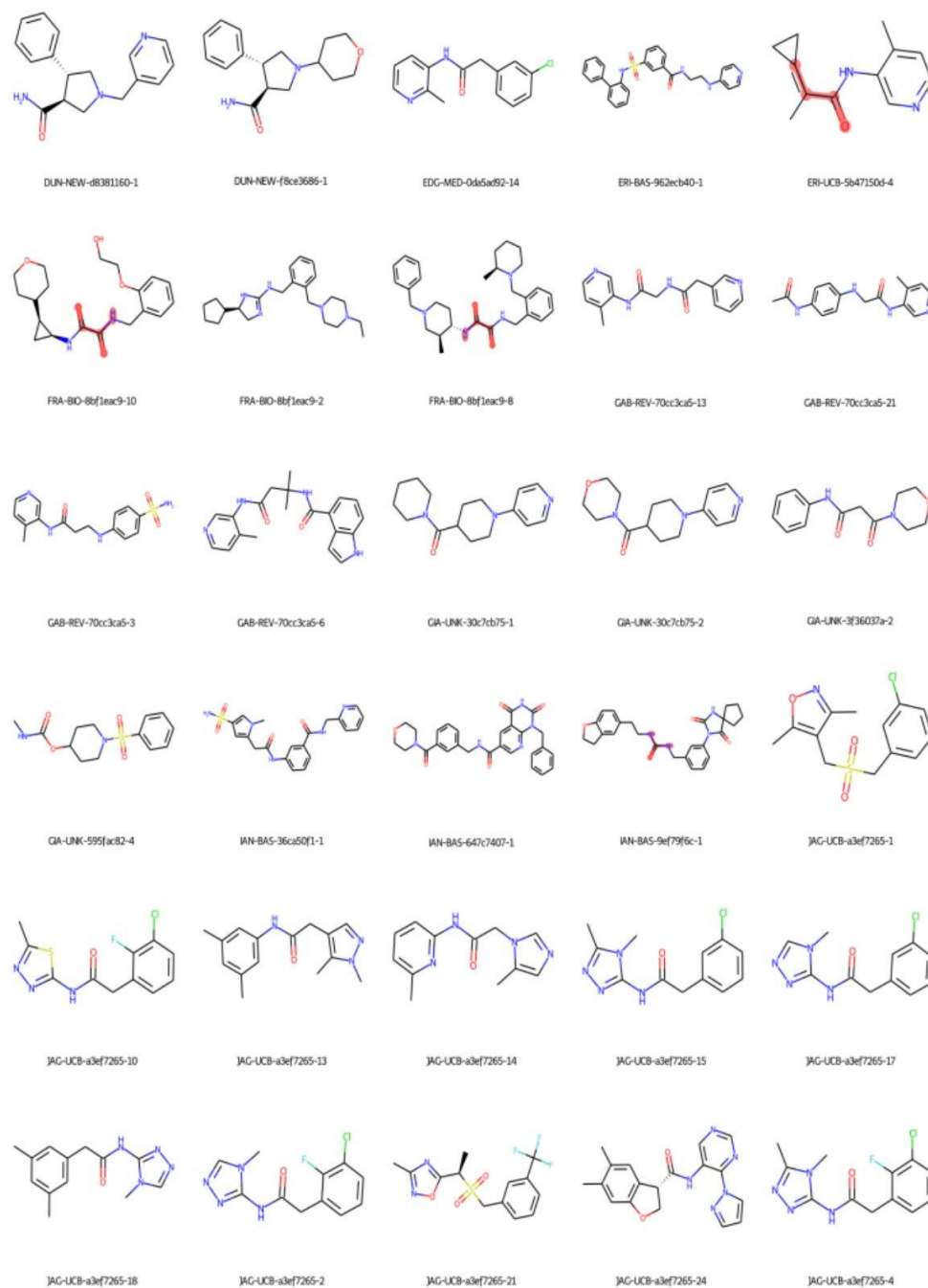
10

Figure S3

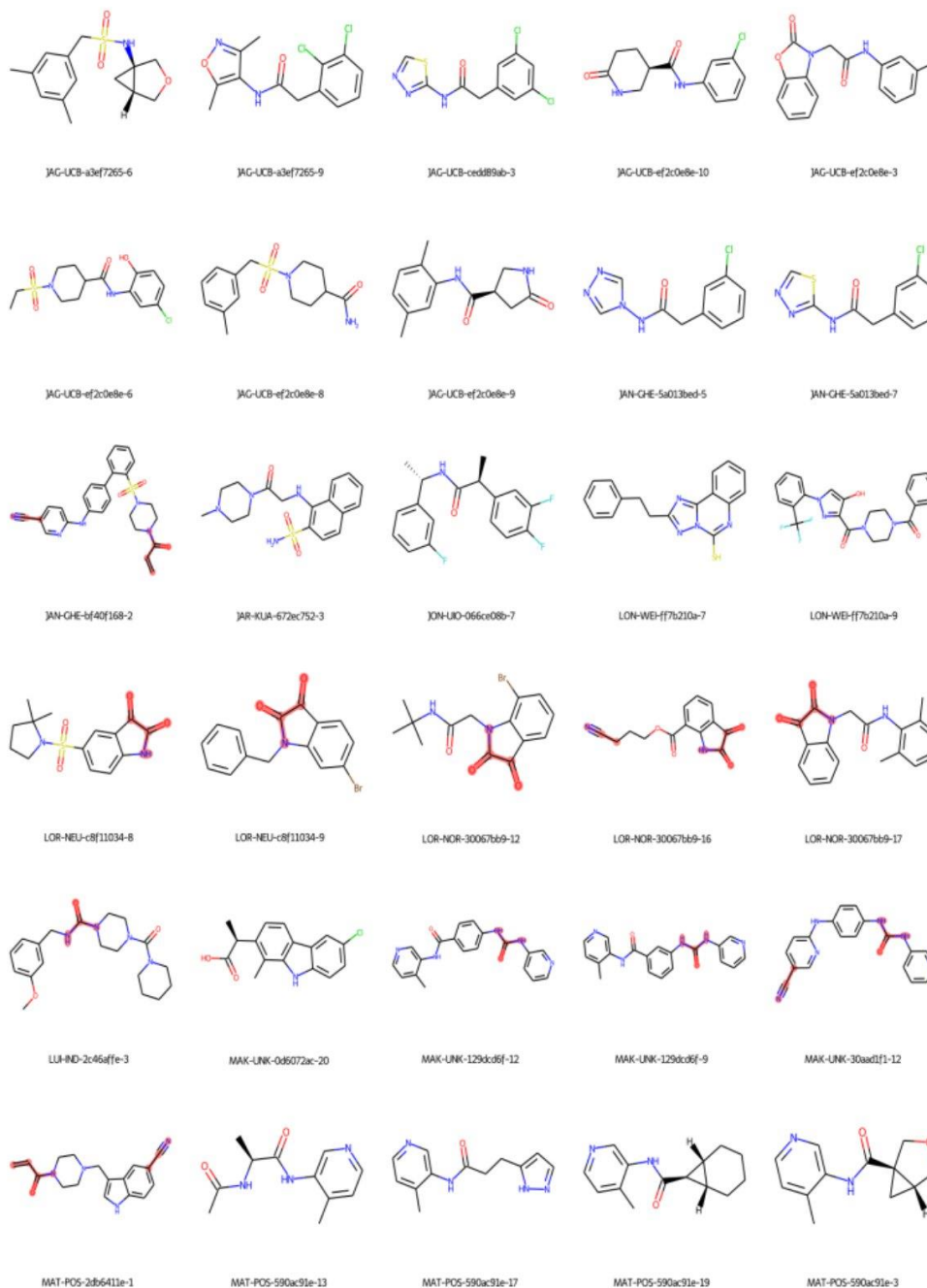


Macip et al.

11



Page 2 of 4



Macip et al.

13

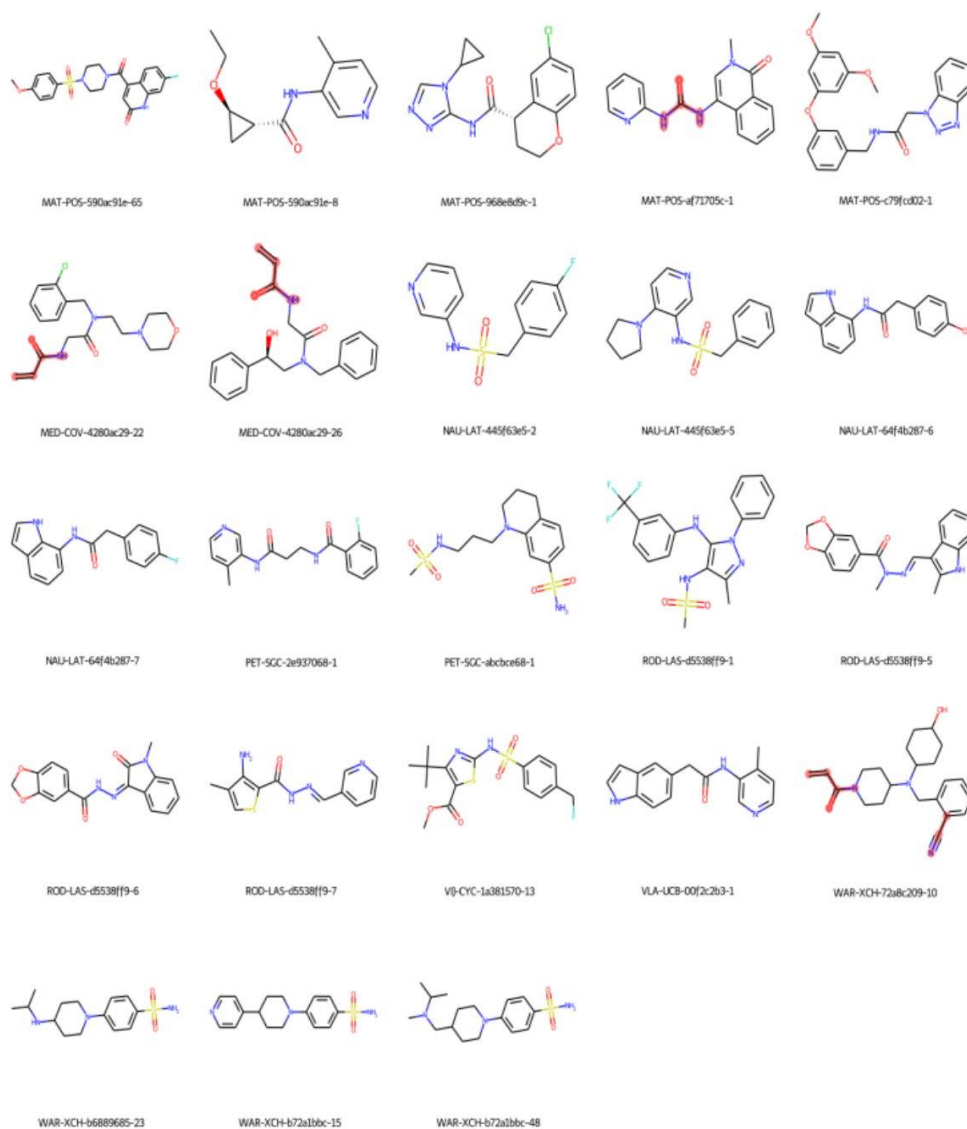


Figure S4

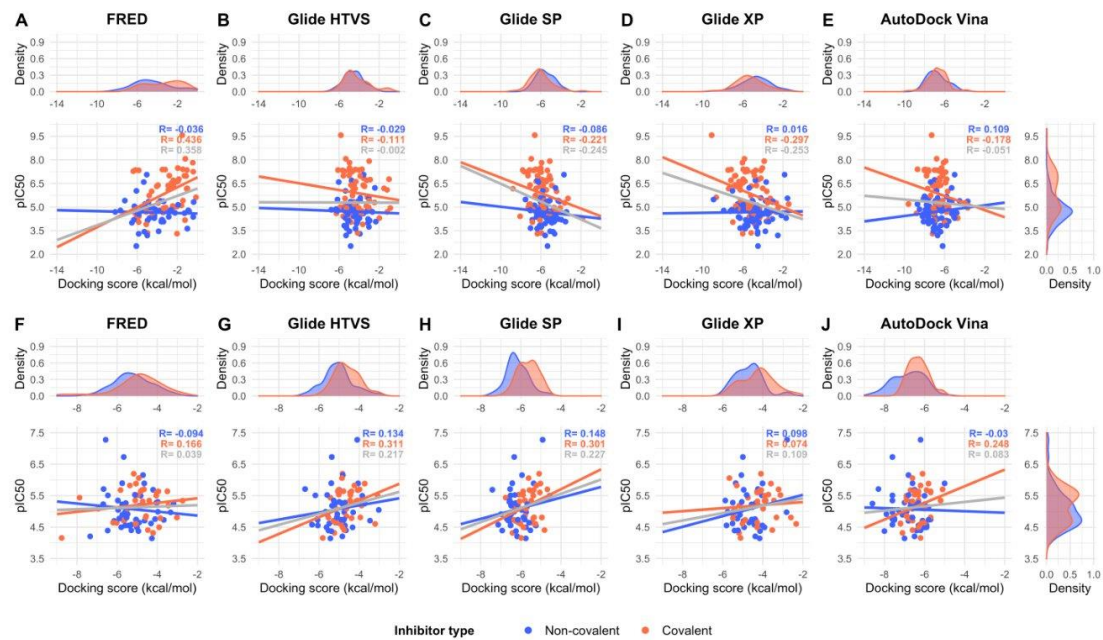


Figure S5

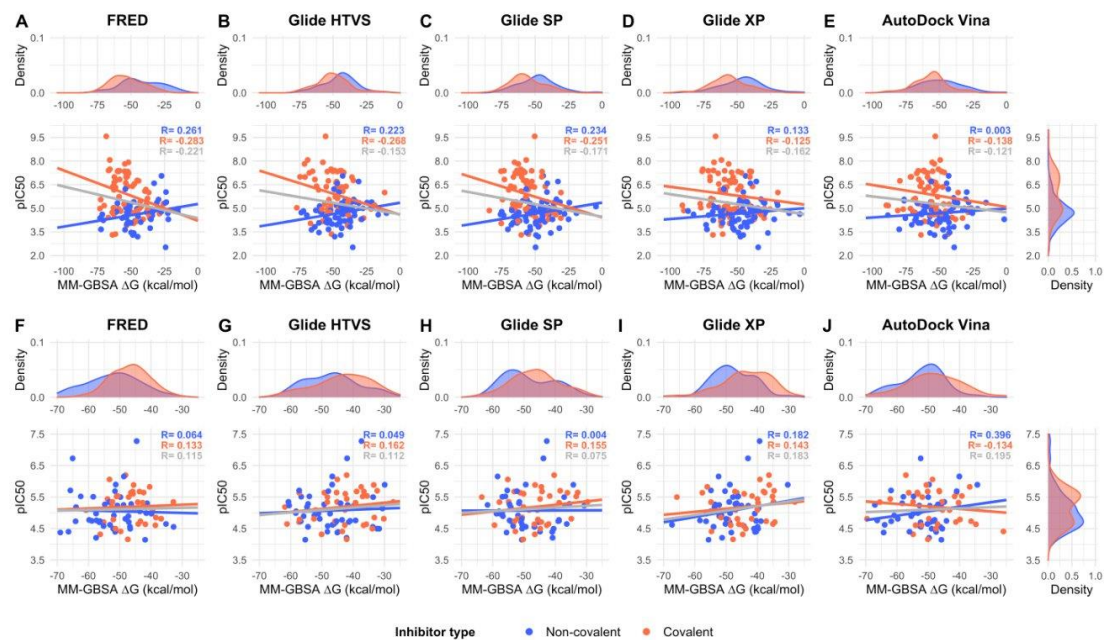


Figure S6

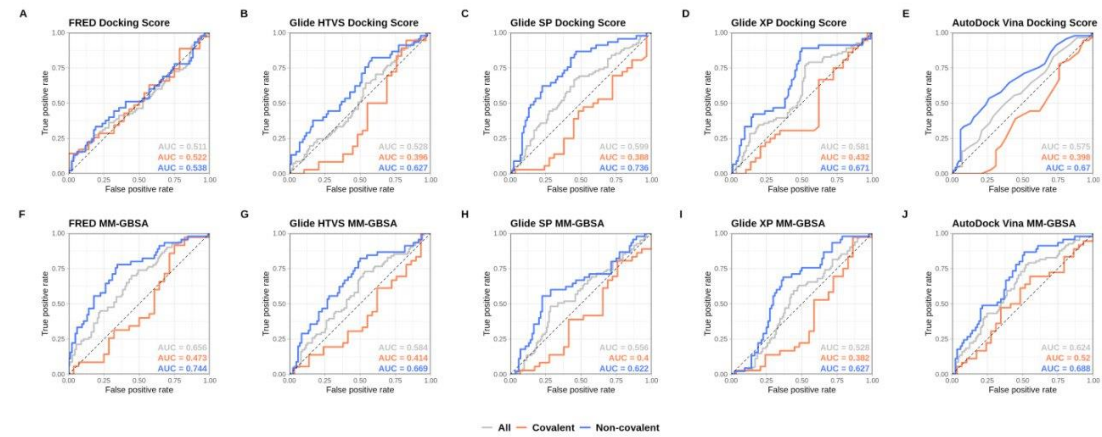
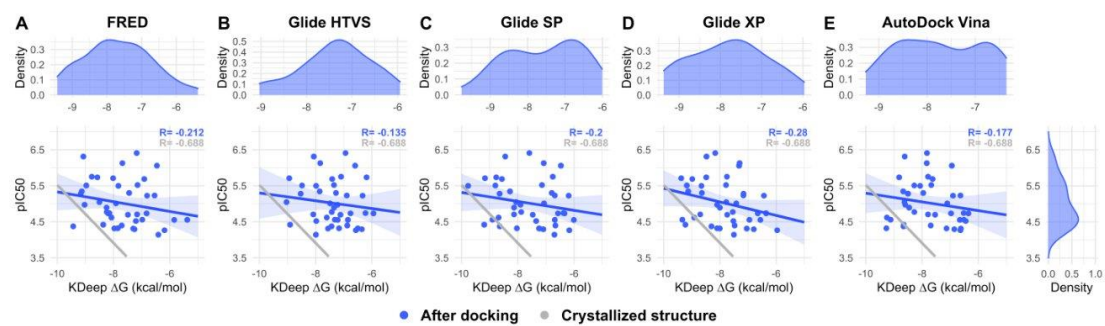


Figure S7



Manuscript 2

Are the protein-ligand docking programs good enough to predict experimental poses of non-covalently inhibitors bound to SARS-CoV-2 main protease?

Guillem Macip¹, Júlia Vilalta¹, Júlia Mestres-Truyol¹, Pol Garcia-Segura¹, Bryan Saldivar-Espinoza¹, Adrià Cereto-Massagué², Santiago Garcia-Vallvé^{1,*}, Gerard Pujadas^{1,*}.

¹Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Research group in Cheminformatics & Nutrition, Tarragona, Catalonia, Spain.

²EURECAT Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Catalonia, Spain.

*Corresponding authors

ABSTRACT

Even with the successful implementation of vaccines, antiviral treatment against SARS-CoV-2 is still necessary. In this regard, its main (M-pro) or 3C-like protease remains one of the best targets for the treatment of COVID-19 and numerous virtual screening (VS) studies have targeted this viral enzyme to find small molecules that inhibit its proteolytic action. Most of these VS have been performed using AutoDock Vina or one of the three methods available for Glide (i.e., HTVS, SP or XP), either independently or as part of a VS workflow. In addition, in the Protein Data Bank (PDB) there are several complexes between M-pro and different non-covalent inhibitors. Therefore, all these 3D complexes are an excellent benchmark for testing the ability of docking programs to predict their structure. Here we analyze whether different docking programs, using different target structures and various target preparations, can predict the experimental pose of these complexes. The overall performance of the docking programs is low. Depending on the conditions, the docking programs reproduce the experimental pose for 48-81% of the inhibitors in a re-docking experiment. The cross-docking results show that the accuracy of the predictions depends more on the PDB file used in the docking than on the docking program itself. In addition, they also show the importance of the correct preparation of the 3D structures of the targets before using them for docking. Finally, we evaluate the results of the cross docking against the histidine protonation of key residues and the similarities of the compounds used.

1. INTRODUCTION

More than three years have passed since the start of the global pandemic known as COVID-19 (COronaVirus Disease 2019). Since then, almost 677 million people worldwide have been infected and from those, 6.7 million resulted in deceasements.¹ The pandemic deeply affected the healthcare systems worldwide, straining and weakening until some of them collapsed. Even now, with 13.286 million vaccines administered, around 5.9 million people get infected every 28-days globally.²

The causative agent of COVID-19 is SARS-CoV-2, a positive sense single-stranded RNA virus from the *Betacoronavirus* genus, and since December 2019, it has been earnestly studied by all the scientific community.³ Apart from the spike protein, which is the target of currently approved vaccines, one of the viral targets more widely studied is the SARS-CoV-2 main protease (M-pro).⁴ This protease, also known as 3-chymotrypsin-like cysteine protease (3CL-Pro), is involved in the first enzymatic step, cleaving the pp1a and pp1ab polyproteins to produce several nonstructural proteins that results in the viral replication and transcription of the virus^{4,5} and thus is a key target for the development of antiviral drugs.^{4,6}

M-pro activity depends on its homodimeric quaternary structure⁷ and each subunit has 306 residues, constituting three domains (domains I and II are antiparallel six-stranded β -barrel structure, while domain III consists of five α -helices; see Figure 1). The active site of M-pro, which is located between domains I and II, differs from other chymotrypsin-like enzymes in that the usual catalytic triad (Ser/Cys-His-Asp/Glu) seems to lack a third component because it is formed by the dyad His41 and Cys145 (see Figure 1). Because of that fact, it has been suggested that a water molecule might fill the role of the Asp/Glu residue in the canonical catalytic triad by mediating interactions between His41 and other conserved residues, like His164 and Asp187.⁸

Numerous studies are trying to inhibit the activity of the M-pro with covalent and non-covalent inhibitors.^{4,9–12} While covalent inhibitors are directly bonded to the catalytic cysteine residue, non-covalent inhibitors achieve their effect through a more distributed network of interactions with different binding site residues.¹³ Hence their modes of action are unambiguously different.¹⁴ Some of the compounds outlined as SARS-CoV-2 M-pro inhibitors were initially conceived for other pathogens, such as HIV (Lopinavir) and Hepatitis C (Remdesivir and Ribavirina) viruses, which stresses the importance of drug repurposing strategies.^{15–17}

Of the *in silico* approaches used in drug discovery, protein-ligand molecular docking is the most popular and has been broadly used, usually in order to get rid of non-suitable compounds or to validate virtual screening (VS) workflows previous to *in vitro* studies.^{18,19} Molecular docking allows to probe thousands of compounds against the protein target at a relatively low computational cost. The docking software integrates a search algorithm to explore different conformations and orientations of the ligand/compound into a predefined binding site (*i.e.*, docked poses) with a subsequent scoring function that estimates the binding energy associated with each generated pose.²⁰ Specifically regarding M-pro, molecular docking is currently controversial since the great amount of *in silico* “findings” have not been translated to *in vitro* results.^{21,22} Hence it is imperative to check if docking programs struggle to predict the correct binding mode or the docking score simply fails to properly punctuate the correct/best pose.²³ It

is also important to note that most molecular docking studies simply delete all water and as it has been said before, water might have an important role inside the binding site of M-pro²⁴.

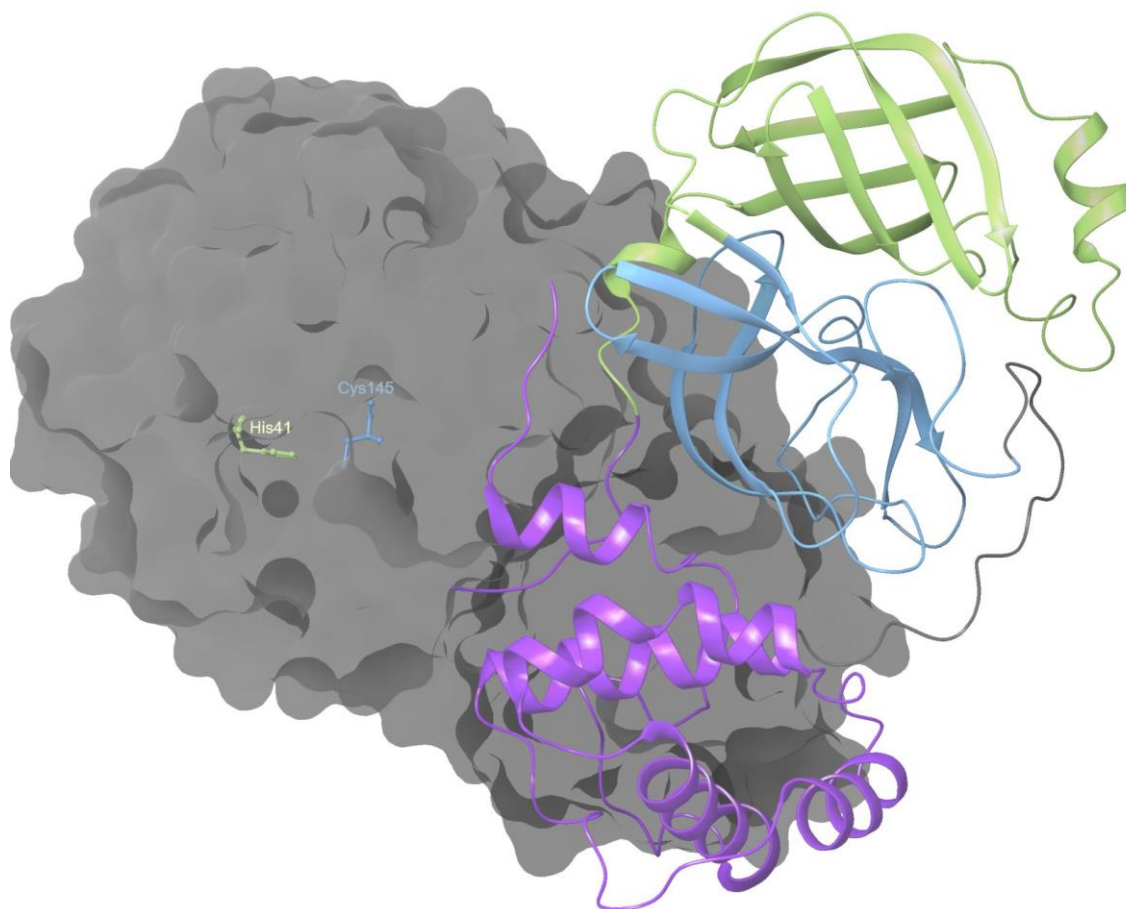


Figure 1. Representation of the M-pro homodimeric structure (molecular surface on the left subunit and backbone on the right one). The backbone representation shows the limits of each of the 3 domains (Domain I in green for residues from 8 to 101; Domain II in blue for residues from 102 to 184; and Domain III in purple for residues from 201 to 303). The surface representation shows the catalytic dyad in green (His41) and blue (Cys145)

The most widely used programs in VS studies for finding SARS-CoV-2 M-pro inhibitors, and by a great margin, are Glide and AutoDock Vina.²¹ The former has been updated and improved over the years with each Schrodinger release, while AutoDock Vina had not had any major release until recently when they launched AutoDock Vina 1.2.²⁵

Glide offers the full range of speed vs. accuracy options with its three methods: HTVS (high-throughput virtual screening), SP (standard precision) and XP (extra precision). HTVS and SP docking use the same scoring function, but HTVS reduces the number of intermediate conformations and the thoroughness of the final torsional refinement and sampling.^{26,27} Glide XP does more extensive sampling than SP and it starts with SP sampling before beginning its own anchor-and-grow procedure. XP also employs a more sophisticated scoring function that weeds out false positives that SP lets through.²⁸

AutoDock Vina (so-called Vina) is the most widely used open-source program for molecular protein-ligand docking.²⁹ AutoDock Vina uses a non-deterministic algorithm and by its nature as an open-source has had many different programs derived from it (e.g., smina, VirtualFlow).^{30,31} It uses an iterated local search global optimizer algorithm, which consists in a success of steps of mutation and optimization, each according to the Metropolis criterion; and a Broyden-Fletcher-Goldfarb-Shanno method for the local optimization.²⁹ Recently AutoDock Vina 1.2 was released which allows, among other things, to hydrate the ligand with explicit bridging water molecules prior to docking.²⁵

Due to the increasingly high number of SARS-CoV-2 M-pro structures in the Protein Data Bank (PDB) that are complexed with non-covalent inhibitors, it is possible to use them to create a benchmark for analyzing the performance of the most commonly used protein-ligand docking programs.^{5,23,32–34} Even though these studies showed accurate docking results, there are some points to discuss. First, some authors perform the validation by considering PDB structures with covalently bound and noncovalently bound inhibitors indiscriminately which may lead to an inferior performance of the protein-ligand docking programs under evaluation, since they have been specifically designed for non-covalent binding.^{23,34} Second, only one of these studies analyzed the influence on the docked poses of the water molecules found at the M-pro binding site.³³ Third, one of these studies generates a different number of poses depending on the protein-ligand docking program that it is used and, therefore, the comparison between them can suffer from bias.⁵ Fourth, a re-docking (*i.e.*, comparing the docked versus the experimental pose for one inhibitor at the protein coordinates from its corresponding PDB file) and a cross docking (*i.e.*, comparing the docked versus the experimental poses for all the inhibitors at a different PDB file than the one where they are co-crystallized) analysis is only performed in one of these benchmark studies.³² Fifth, the effect of the different protonation states for critical residues are only considered in two of these studies.^{32,34}

Therefore, the aim of this current study is to make a thorough evaluation of the performance of the four most widely used docking programs/methodologies (Glide SP, Glide XP, Glide HTVS and AutoDock Vina) in both re-docking and cross docking experiments in order to analyze to which extend they are able to predict the experimental poses of the non-covalent inhibitors that are bound to SARS-CoV-2 M-pro in the 3D structure of a set of experimental complexes obtained from the PDB. Furthermore, we analyze the influence of the binding site environment to assess the docking methodology capable of better predicting the structure of these complexes by using binding sites with or without water molecules and with different protonation states for critical residues.

2. MAIN CHARACTERISTICS AND PREPARATION CONDITIONS OF THE NON-COVALENT M-PRO/INHIBITOR COMPLEXES USED IN THIS STUDY.

The 47 PDB files from SARS-CoV-2 M-pro complexes with non-covalent inhibitors that were used in this study were obtained from the May 8th 2021 update of the Protein Data Bank (PDB) and superposed with Maestro (Schrödinger Release 2021-1: Maestro; Schrödinger, LLC.: New York, NY, USA, 2021). Apart from 6m2n, which included two homodimers (6m2n-1 and 6m2n-2) and were considered as independent, the asymmetric unit of the remaining 46 M-pro complexes included only one. Therefore, the number of complexes studied in this study is 48. Figure 2 shows how the non-covalent inhibitors are structurally related according to PubChem's fingerprints (their 2D structure is shown in Figure S1). In general, the inhibitor's structure is quite varied and only compounds at the next sets of PDB files share part of their 2D structure: **(1)** 5r83,

5rgk, 5rgu, 5rgv, 5rgw, 5rgx, 5rgy, 5rgz, 5rh0, 5rh1, 5rh2, 5rh3 and 5rh8; and **(2)** 7l10, 7l11, 7l12 and 7l14. On the other hand, 6w63 and 6w79 share the same ligand.

Considering that the reliability of the binding site and ligand coordinates in these 48 SARS-CoV-2 M-pro non-covalent complexes is crucial for a proper evaluation of the predictive success of the different protein-ligand docking programs/methods that are being tested, the fitting of their coordinates to the electron density map (EDM) was analyzed with VHELIBS.³⁵ Then, according to this analysis, the binding site for 9 out of 48 complexes (*i.e.*, 5r80, 5r84, 5reb, 5rez, 5rh8, 7a1u, 7ans, 7ap6, 7aqe) shows coordinates that are qualified as “bad” by VHELIBS and therefore, this can affect the results of the protein-ligand docking performed with these protein structures. In a similar way, the non-covalent inhibitors for 5 out of the 48 complexes (*i.e.*, 5rf2, 5rfe, 5rgi, 5rgk, 7aqe) show coordinates that VHELIBS labels also as “bad” and, therefore comparing docked poses with them may lead to an underestimation of docking results.

Crystallographic water molecules within 5.0 Å from any inhibitor atom were also analyzed in order to find which ones are spatially equivalent (*i.e.*, within 1.0 Å of another water molecule in a different complex) across the 48 SARS-CoV-2 M-pro complexes (see Tables S1 and S2). This will make it possible to assess the influence of the solvent molecules on the protein-ligand docking results by comparing three different situations in the binding site: **(1)** *dry* (*i.e.*, no water molecules are considered at the binding site); **(2)** *conserved* (only those water molecules conserved at least in 15 out of 48 complexes are kept); and **(3)** *water* (all waters present at the binding site in the original PDB file are kept).

Finally, all the superposed complexes have undergone a preparation prior to being used in protein-ligand docking with the Protein Preparation Wizard v2021-1³⁶ by using the default setup except for the next settings: **(a)** N and C termini were capped; and **(b)** hydrogens were added after original hydrogens had been removed. For AutoDock Vina docking, the proteins were then converted to the specific format (PDBQT) with the *Prepare_receptors_rigid.sh* script provided by the self-program.

3. RE-DOCKING OF NON-COVALENT M-PRO/INHIBITOR COMPLEXES WITH KNOWN 3D STRUCTURE.

The ability of the 4 most widely used docking programs/methodologies (Glide HTVS, Glide SP, Glide XP and AutoDock Vina) to reproduce the 48 experimental complexes in the three previously mentioned binding site situations (*i.e.*, *dry*, *conserved* and *water*) was analyzed by performing a re-docking experiment (*i.e.*, removing the inhibitor from the corresponding PDB file and then dock it back).

Here, to not bias the results, special care was taken to avoid that the input ligand conformation did not bear any relation to the experimental pose that is being predicted. This was achieved by **(1)** initially converting the experimental 3D poses from the inhibitors to 1D SMILES with Maestro, **(2)** then using Omega v3.1.1.2³⁷ to generate the 3D conformation with the lowest energy, and **(3)** finally generating all the possible protonation states in the pH range 7.2 ± 1.0 and the default number of tautomers with Ligprep v2020-3³⁸. The resulting 3D conformers of the ligands were then used as input for all the corresponding re-docking experiments.

Grid preparation was different for Glide and for AutoDock Vina. For Glide, the grid for each binding site situation was generated by using the XGlide panel with default options. Here the size of the outer box size was set to 35 x 35 x 35 Å while the inner box was set to 10 x 10 x 10 Å. The grid center corresponded to the centroid of the N3 ligand at 6LU7 (−10.36, 12.46, 68.7).⁷

For AutoDock Vina, the spatial location of the grid was coincident with the inner box for Glide and the grid preparation was performed with AutoDock Vina 1.2.0 with default options.²⁵ Here, instead of considering the *conserved* and *water* situations, the Meeko python package was used for the automated addition of explicit water molecules directly on the ligand.³⁹

Finally, all the resulting grids and the 46 inhibitors found in the 48 complexes (see **Figure S1**) were used by either Glide or AutoDock Vina for the corresponding re-docking experiments by using default options except for the number of docked poses that were kept for each protein-ligand docking that was set to 20. Then, each docked pose was compared with its corresponding experimental pose and the root mean-squared deviation (RMSD) was calculated to quantify the ability of the docking program to predict the correct inhibitor binding.

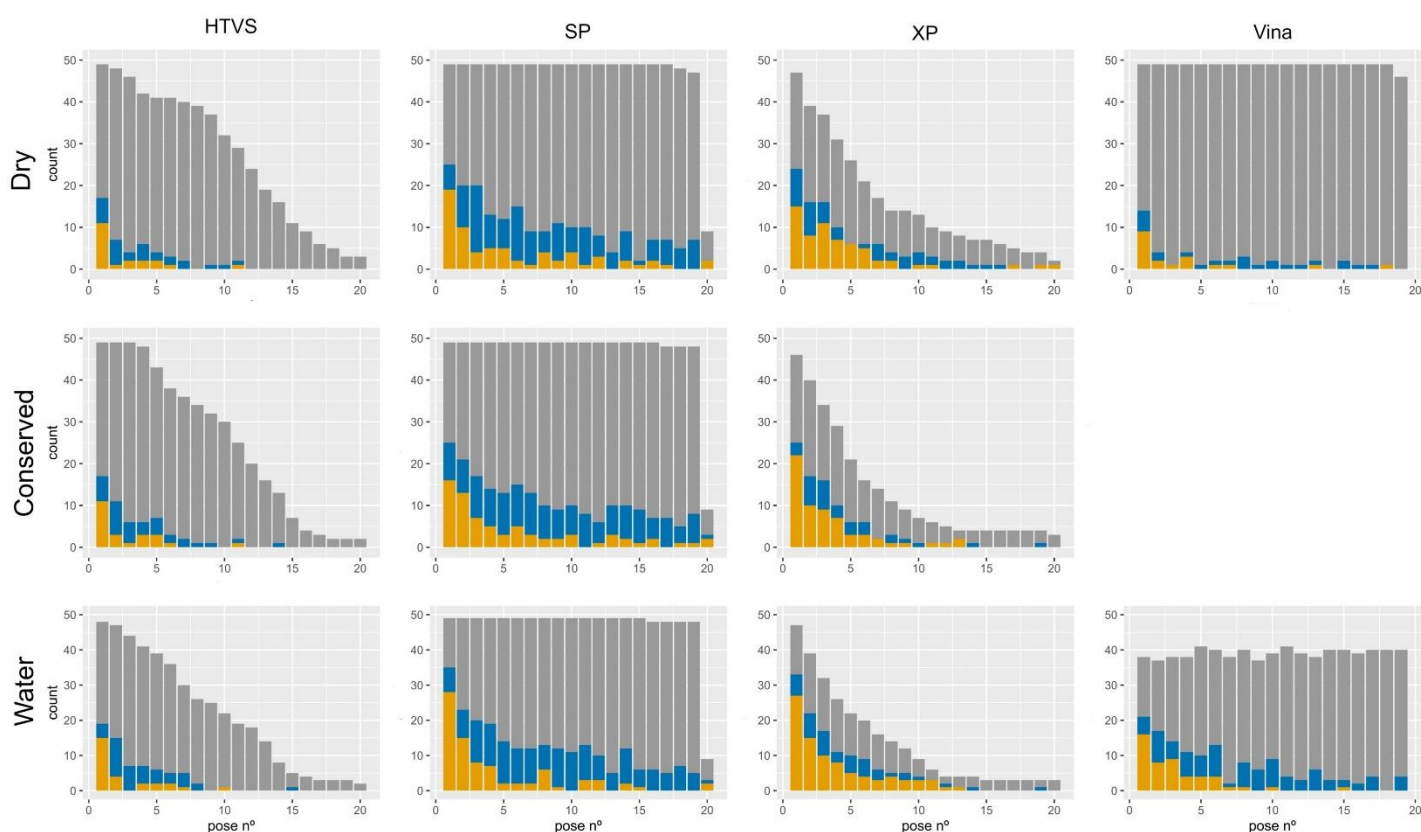


Figure 3. Re-docking success evaluation by comparing the RMSD of the first 20 docked poses of each co-crystallized inhibitor with its experimental pose. The performance of the three protein-ligand docking methodologies that can be run with Glide (i.e., HTVS, SP and XP) is evaluated in three different binding site environments [where dry corresponds to deleting all binding site water molecules before docking; conserved corresponds to consider during the protein-ligand docking only those water molecules that are found at least in 15 complexes (see Table S2); and water corresponds to consider during the protein-ligand docking all the binding site water molecules (see Table S1)]. In the case of AutoDock Vina, the re-docking success is evaluated in two different binding site environments (where dry corresponds to deleting all binding site water molecules before docking whereas and water corresponds to consider during the protein-ligand docking the explicit water molecules generated by the Meeko python package). Bar colors at the histogram group the results according to different RMSD ranges.

Figure 3 shows how each one of the 20 docked poses per ligand compares with the corresponding co-crystallized pose. The resulting RMSD values are then grouped according to different ranges to better assess the impact of the scoring function on the reliability of the predicted poses. As expected, the first docked pose obtained by the different methodologies/programs (and in the different situations analyzed at the binding site) is, in general, the closest to the experimental pose (*i.e.*, when the rank of a pose decreases, the probability that this pose is similar to the co-crystallized inhibitor also decreases). Figure 3 also shows that for the first docked pose, there is nearly no difference in the number of poses with an RMSD ≤ 2.0 Å found by either Glide methodology in a *dry* or *conserved* binding site environment (*i.e.*, 17 in HVTs, 25 in SP and from 24 to 25 for XP if the *dry* and *conserved* situations are compared). This situation completely changes when all waters within 5.0 Å of any inhibitor atom (see Table S1) are considered as integral part of the binding site during the protein-ligand docking because in that case the number of poses with an RMSD ≤ 2.0 Å increases for each Glide protocol (*i.e.*, from 17 to 19 in HVTs, from 25 to 35 in SP and from 24 to 33 for XP if the *dry* and *water* situations are compared). For AutoDock Vina, the use of explicit water molecules also increases the reliability of the first docked pose in the same RMSD range (from 14 to 21 if the *dry* and *water* situations are compared).

	HTVS	SP	XP	Vina
<i>Dry</i>	23/18/6/2 5r80, 5r83 , 5r84 , 5re4 , <i>5re9</i> , 5reh , <i>5rfe</i> , 5rg1 , 5rgi , 5rgk , 5rgx , 5rgz , 5rh1 , 5rh8 , 6w63 , 6w79 , <i>7alu</i> , 7ap6 , <i>7aqe</i> , 7kx5 , 710d , 7112 , 7114	14/6/6/2 <i>5re9</i> , 5reh , <i>5rfe</i> , 5rg1 , <i>5rgh</i> , 5rgk , <i>5rh0</i> , <i>5rh8</i> , <i>6w63</i> , <i>6w79</i> , <i>7alu</i> , 7ap6 , <i>7aqe</i> , 7kx5	15/5/6/2 <i>5r7y</i> , <i>5r80</i> , 5re4 , <i>5re9</i> , <i>5reh</i> , <i>5rf1</i> , <i>5rfe</i> , <i>5rg1</i> , <i>5rgh</i> , 5rgk , <i>5rgv</i> , <i>5rh0</i> , <i>5rh8</i> , 7ap6 , 710d	16/9/6/2 5r7y , <i>5r7z</i> , <i>5r82</i> , 5re9 , <i>5reb</i> , 5reh , <i>5rez</i> , 5rfe , <i>5rg1</i> , 5rgk , 5rgv , <i>5rgx</i> , <i>5rgz</i> , 7ap6 , 7aqe , 710d
<i>Conserved</i>	24/18/8/2 <i>5r7y</i> , <i>5r82</i> , 5r83 , 5r84 , 5re4 , <i>5re9</i> , 5reh , <i>5rfe</i> , 5rg1 , 5rgi , 5rgk , 5rgx , 5rgz , <i>5rh0</i> , 5rh1 , 5rh8 , 6w63 , 6w79 , <i>7alu</i> , 7ap6 , 7kx5 , 710d , 7112 , 7114	13/6/8/2 <i>5re4</i> , <i>5re9</i> , 5reh , 5rg1 , <i>5rgh</i> , 5rgk , <i>5rgx</i> , <i>5rh8</i> , <i>6w63</i> , 7ap6 , 7aqe , 7kx5 , 710d	15/5/8/2 <i>5r7y</i> , <i>5r81</i> , <i>5r83</i> , 5re4 , <i>5re9</i> , <i>5reb</i> , <i>5rg1</i> , <i>5rgh</i> , 5rgk , <i>5rgv</i> , <i>5rgx</i> , <i>5rh0</i> , <i>5rh8</i> , 7ap6 , 710d	
<i>Water</i>	22/18/4/2 <i>5r82</i> , 5r83 , 5r84 , 5re4 , 5reh , 5rg1 , 5rgi , 5rgk , 5rgx , 5rgz , <i>5rh0</i> , 5rh1 , 5rh8 , 6w63 , 6w79 , 7ap6 , <i>7aqe</i> , 7kx5 , 710d , 7111 , 7112 , 7114	8/6/4/2 5reh , 5rg1 , 5rgk , <i>6w79</i> , 7ap6 , 7aqe , 7kx5 , 710d	8/5/4/2 5re4 , <i>5reh</i> , 5rgk , <i>5rgz</i> , <i>5rh0</i> , <i>6w79</i> , 7ap6 , 710d	16/9/4/2 5r7y , <i>5r80</i> , <i>5r81</i> , 5re9 , 5reh , <i>5rf1</i> , <i>5rf2</i> , 5rfe , 5rgk , 5rgv , <i>5rh8</i> , <i>5rhd</i> , <i>7ans</i> , 7ap6 , 7aqe , 710d

Table 1. List of PDB codes for the ligands for which no pose below 2.0 Å was found in the re-docking experiment under the different combinations of binding site environment and docking programs/methodologies used in the present study (a maximum of 20 poses per ligand were evaluated). PDB codes in bold did not obtain a single pose for the given program/methodology, regardless of the treatment of the binding site environment. PDB codes in italic did not obtain a single pose for the given treatment of the binding site environment regardless of the program/methodology used. Underlined PDB codes did not obtain a single pose in any treatment of the binding site environment and program/methodology used. The numbers summarize how many ligands are found in the above situations in each cell.

Despite the in-depth analysis in the previous paragraph, there remain some important questions about the performance of the four programs/methodologies that cannot be answered by **Figure 3**. These are **(1)** are the experimental poses reproduced in the 48 complexes at an $\text{RMSD} \leq 2.0 \text{ \AA}$ by any of the combinations of binding site environments and programs/methodologies? and **(2)** how many docked poses are needed to find a solution at $\text{RMSD} \leq 2.0 \text{ \AA}$ from the experimental one? **Figure 4** tries to answer both questions by no longer considering all the consequent low-scored docked poses after the first occurrence of a RMSD under $2.0 \text{ \AA}$, for the comparisons with the experimental pose of one re-docked inhibitor. Interestingly, **Figure 4** shows, irrespective of the combinations of binding site environments and programs/methodologies, that: **(1)** the first pose is, overwhelmingly, the one with most RMSD under $2.0 \text{ \AA}$; **(2)** in general, no new poses at $\text{RMSD} \leq 2.0 \text{ \AA}$ from the experimental one are found after the 5th ranked one; and **(3)** there are ligands for which no docked pose is found with a RMSD under $2.0 \text{ \AA}$.

This later result is quite surprisingly if it considered that **Figure 4** is evaluating the re-docking performance and, therefore, it merits further analysis. Therefore, **Table 1** details which are the ligands for which no re-docked pose is found with a RMSD under $2.0 \text{ \AA}$.

Considering this the main conclusion of the re-docking part of the experiment are the following: **(1)** depending on the program and conditions used, docking programs are only able to reproduce the experimental pose for 50-83% of the non-covalent inhibitors; **(2)** depending on the docking program and the conditions used, the first pose only corresponds to the experimental pose in 37-76% of the inhibitors (whereas the poses from 2 to 5 and the ones from 6 to 20 corresponds to 10-56% and 7-31%, respectively); **(3)** for some of the non-covalent inhibitors, none of the docking programs or setups used is able to reproduce the experimental pose; **(4)** if the water molecules at the binding site are considered as an integral part of the active center during docking, it is more likely that the first pose will correspond to the experimental pose; **(5)** if only those water molecules that are more conserved among the structures of the different experimental complexes are considered as an integral part of the active center, then the results obtained are not always better than if no water is considered; and **(6)** if the options that require less user intervention are used (*i.e.*, Hydrated docking of the AutoDock Vina or no water present at the active center), then the first pose coincides with 39-53% of the experimental poses.

Manuscript 3

Generally speaking, Glide SP and XP boast the better results whereas Vina barges into the equation only with the hydrated docking. That being said, the main inconvenience of Hydrated Vina is that it cannot hydrate all inputted inhibitors and fail with ligands from 5r7y, 5r80, 5r81 5rf1, 5rhd and 7ans.

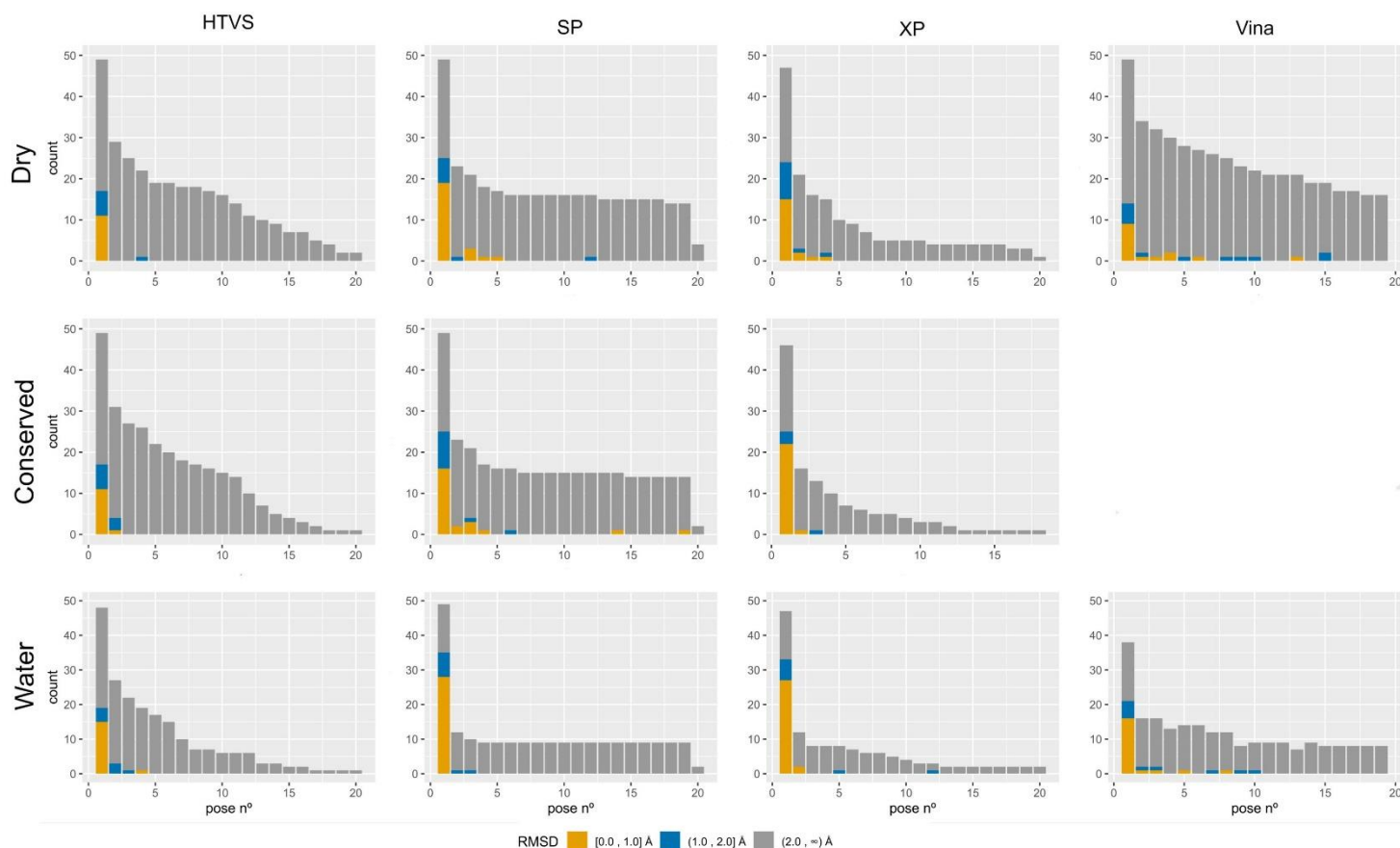


Figure 4. Equivalent to Figure 3 but, after a re-docked pose for an inhibitor is found with $RMSD \leq 2.0 \text{ \AA}$, all its lowest scored poses are no longer considered in the comparisons.

4. CROSS DOCKING OF NON-COVALENT M-PRO/INHIBITOR COMPLEXES WITH KNOWN 3D STRUCTURE.

Usually, cross-docking studies are quite similar to the situation of a protein-ligand docking step in a VS study (*i.e.*, a target structure is used to dock a library of putative ligands from that target). Therefore, cross docking is a good way to evaluate how docking will behave during a VS.

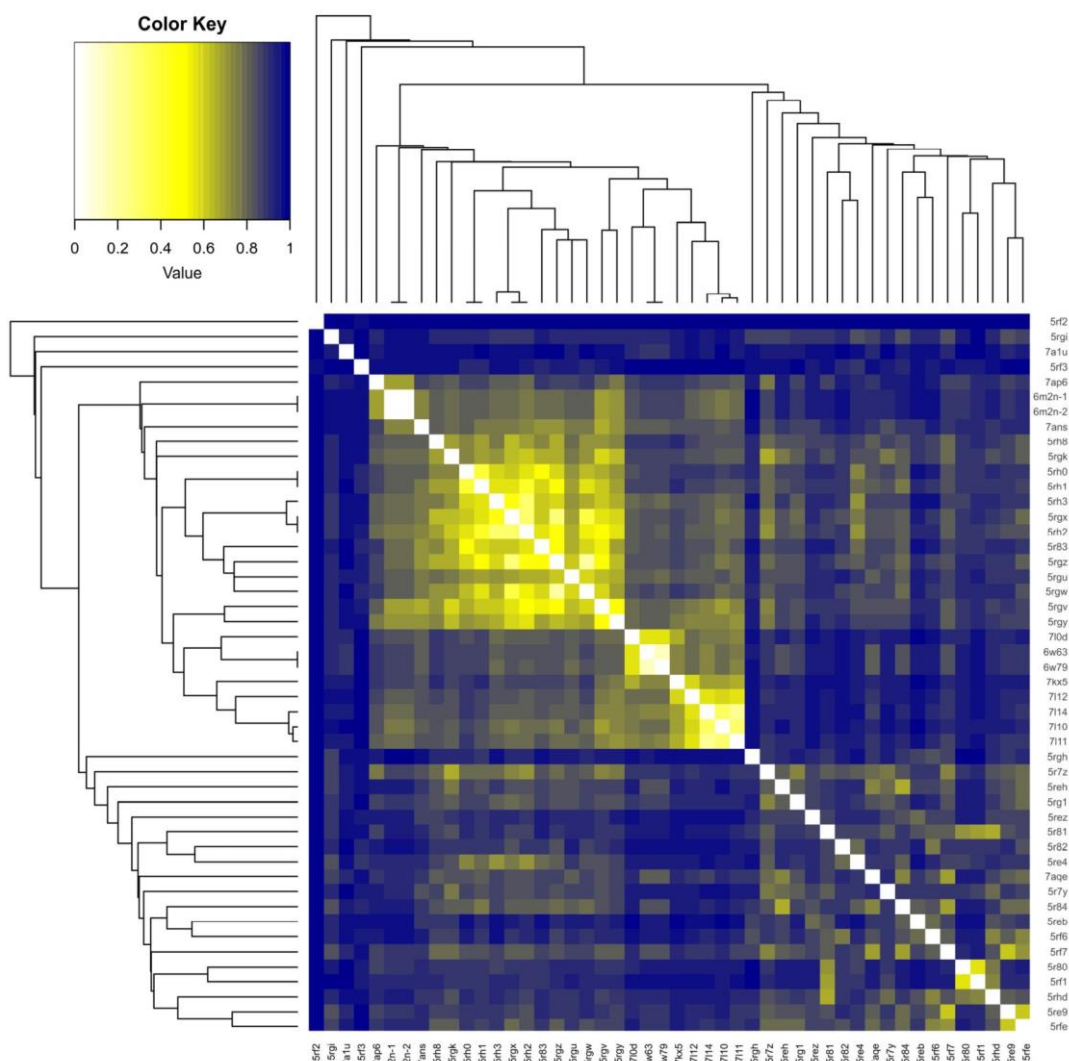


Figure 2. Structural relationships between the 2D structures of all the non-covalently bound inhibitors from all the 48 complexes analyzed in the current study. Each inhibitor is identified by the code of the PDB file to which it belongs. Inhibitors are clustered by using PubChem's fingerprints. In the heatmap, similarity increases from blue (totally different compounds) to white (very similar or the same compounds).

Manuscript 3

Here we try to find if any combination of: **(1)** docking program; **(2)** PDB structure; and **(3)** target preparation is better than the other in predicting the experimental pose when only the first docked pose is considered. For that we perform docking with the dry states of all the structures plus the hydrated docking of AutoDock Vina which hydrates the ligand and then docks it in a dry binding site. Additionally, we try to answer if we can make consensus docking with these programs. **Figure 5** shows the cross-docking results obtained for Glide XP whereas **Figures S2, S3, S4 and S5** provides the results for HVTs, SP, AutoDock Vina dry and AutoDock Vina hydrated, respectively. Comparison between **Figures 5, S2, S3, S4 and S5** shows that, when considering diversity in ligand chemical structure (a key point during any VS scenario), Glide XP provides the best cross-docking results followed by Glide SP, Glide HVTs, AutoDock Vina hydrated and AutoDock Vina dry. Thus, **Figure 5** shows that for Glide XP apparently the best PDB structures in cross docking are 5rgw (8/3), 5rh2 (8/3) and 5rh3 (7/3) (with 11, 11 and 10 docked poses within 2.0 Å, respectively of the corresponding experimental ones) but this seems to be biased by the similarity between ligand 2D structures (see **Figure 2**).

In addition, our analysis shows how the 6lu7 structure (used as a target in most of the non-covalent M-pro inhibitor virtual screening studies)²¹, does not seem to be the most suitable structure for these tasks because it only finds 4 docked poses with an RMSD $\leq$ 2.0 Å to the corresponding experimental poses (Glide XP data). At this point, the low performance of dry or hydrated AutoDock Vina with a single docked pose in the (1.0-2.0] Å range for 6lu7 is very remarkable because this program has been used by about 40% and this PDB structure by 60% of the VS studies looking for non-covalent M-pro inhibitors.²¹ This suggests that all these VS have classified as false negatives a large quantity of compounds that could be potentially used for COVID-19 treatment.

Figure 5 also shows that: **(1)** there are structures (*i.e.*, 6m2n-1, 5rgk, 5re9; Glide XP data) that, when used for cross-docking, do not find any docked poses within 2.0 Å of the corresponding experimental poses although their binding site coordinates fit well their corresponding EDM; and **(2)** there are some non-covalent inhibitors (*i.e.*, those found in 6m2n-1, 5rgk, 5rgv, 5rgy, 7kx5, 5reh, 5rf7, 5r80, 5rf1, 5re9, 5rfe; Glide XP data) that, independently of the target structure used during the protein-ligand docking, do not find any docking pose within 2.0 Å of the corresponding experimental pose (and only the non-covalent inhibitors at 5rgk and 5rfe do not fit the EDM according to VHELIBS).

Generally speaking, structures and ligands that the evaluation with VHELIBS deemed “bad” have worse results than the others, not scoring a single pose at 2.0 or even within the generous 2.7 angstroms. This exemplifies the necessity for the correct interpretation of experimental data and good coordinates versus electron density map fit.

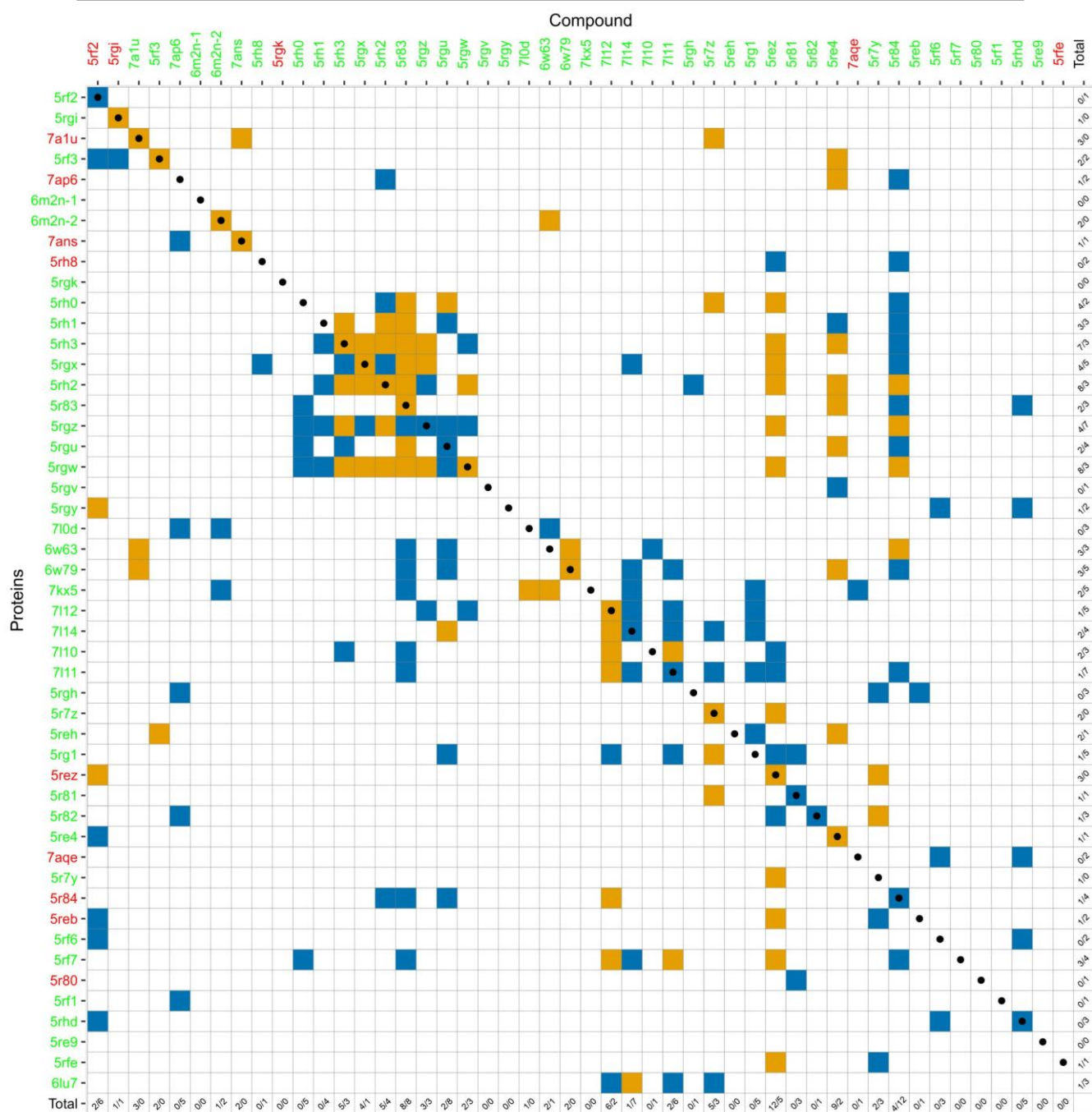


Figure 5. Cross-docking (off-diagonal) and re-docking (diagonal) results obtained with Glide XP for the top-one docked pose in comparison with the corresponding experimental pose. PDB codes in columns and in rows identify, respectively, the non-covalent inhibitors and the protein structure found at the corresponding PDB files. Colors at the PDB labels show whether the corresponding ligand or binding site coordinates fit (green) or do not fit (in red) to the electron density map according to VHELIBS. The colors of the cells mirror the RMSD ranges from Figures 3 and 4 (being orange an $\text{RMSD} \leq 1.0 \text{ \AA}$, blue between 1.0 and $2.0 \text{ \AA}$ and white $> 2.0 \text{ \AA}$). The dotted line corresponds to the re-docking results. Cross-docking results for 6LU7 (by large, the most used SARS-CoV-2 M-pro structure in the literature for docking studies) are also shown in the last line. Data are shown in the same order as in Figure 2 to allow easier correlation of 2D chemical similarity with cross-docking and re-docking results. The numbers in the last column/row are the corresponding counts of the different RMSD ranges in increasing range order (i.e., number of yellow and blue cells).

Given our interest in consensus docking we compare the best “blind” docking method of the two distinct programs (Glide and AutoDock Vina) which in this case are Glide XP and AutoDock Vina hydrated, agreeing with the results seen on the re-docking. Taking into account the best structure from Glide XP re-docking (**Figure 5**, 5rgw) it is observed that the two programs simultaneously predict that the first docking pose of 9 of the non-covalent inhibitors is within 2.0 angstroms of the corresponding experimental pose. Is of note that these 9 poses for AutoDock Vina Hydrated are the only within 2.0 angstroms that the program predicts for the given structure. **Figure 6** shows each ligand of these ligands for both programs plus the experimental pose.

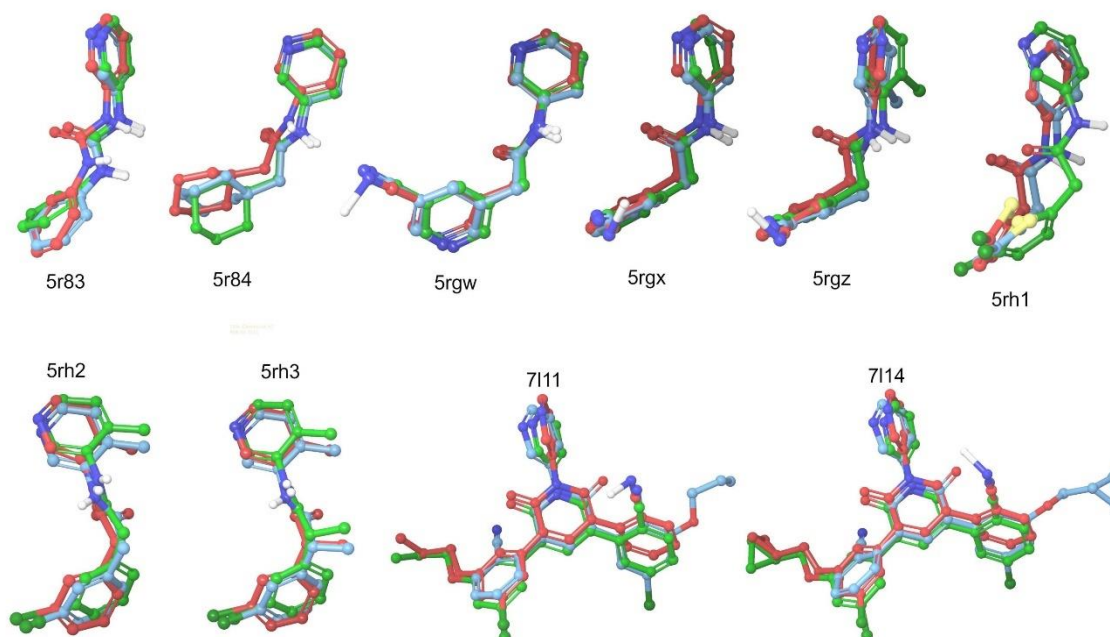


Figure 6. Coincidental results for the structure 5rgw from Glide XP and AutoDock Vina hydrated. In green the experimental pose, in blue the one obtained from XP and in red the one from Vina Hydrated.

5. HOW DOES THE PROTONATION OF ACTIVE SITE HISTIDINES INFLUENCE THE RESULTS OF CROSS-DOCKING?

To better understand the disparity in docking results across different PDB structures, we also analyzed the histidines of the M-pro binding site.^{40–42} His41, His163, His164 and His172 are key residues during the intermolecular interaction between M-pro and non-covalent inhibitors.^{43–45}

Table 2 shows how these four histidine residues are protonated in all the protein structures used in this paper prior to any protein-ligand docking. Of the 4, His172 always shows the same protonation state in all structures and therefore does not play a role in the disparity of results.

With the other three histidines, we tried to find some kind of pattern that allowed us to decide which histidine protonation combination to use during docking to obtain better results. **Figures 7 and 8** shows the histidine protonation of the best and worst performing structures in our dataset. While the best are indeed our top performers, from the worst performing we subtracted those that presented a bad fit with the electron density map within their binding site when evaluated with VHELBS.

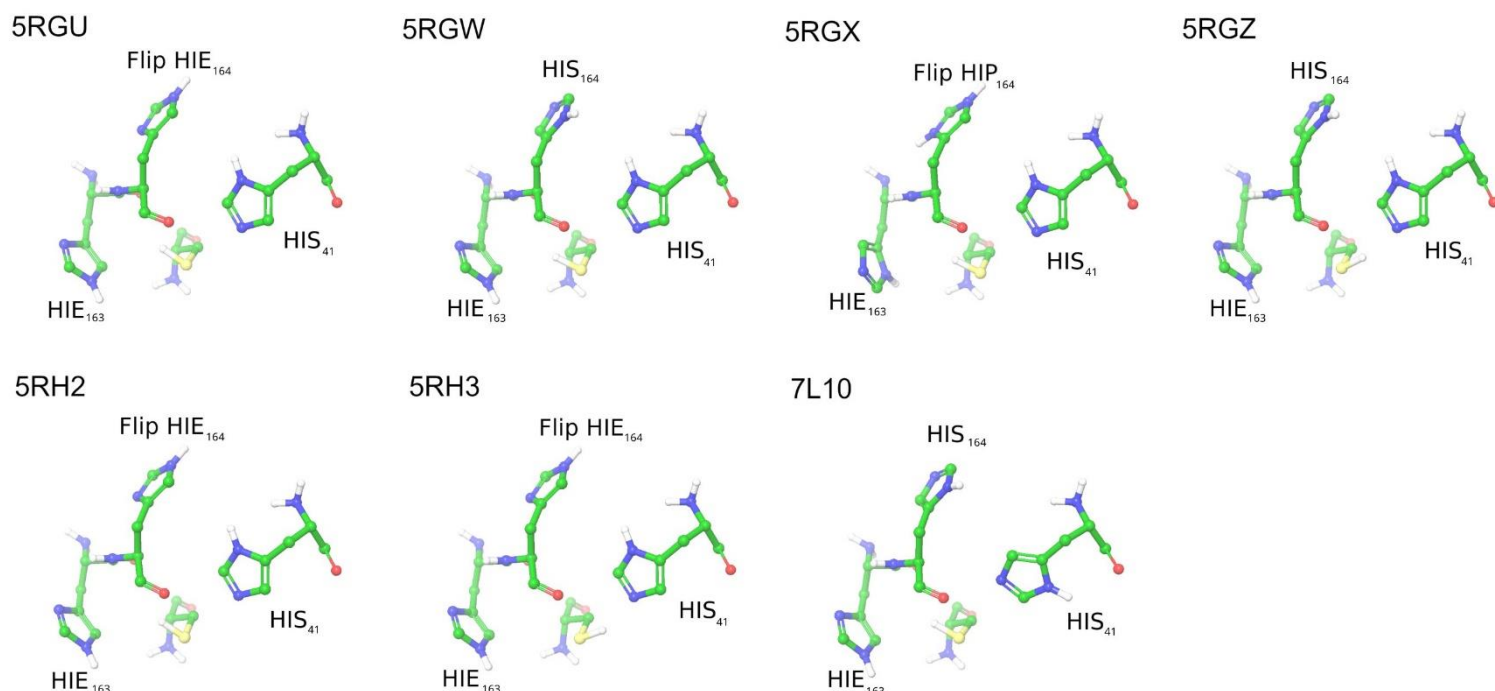


Figure 7 and 8. Set formed by His₄₁, His₁₆₃, His₁₆₄, and Cys₁₄₅ in the binding site of the structures that give the “bests” results on the cross-docking evaluation.

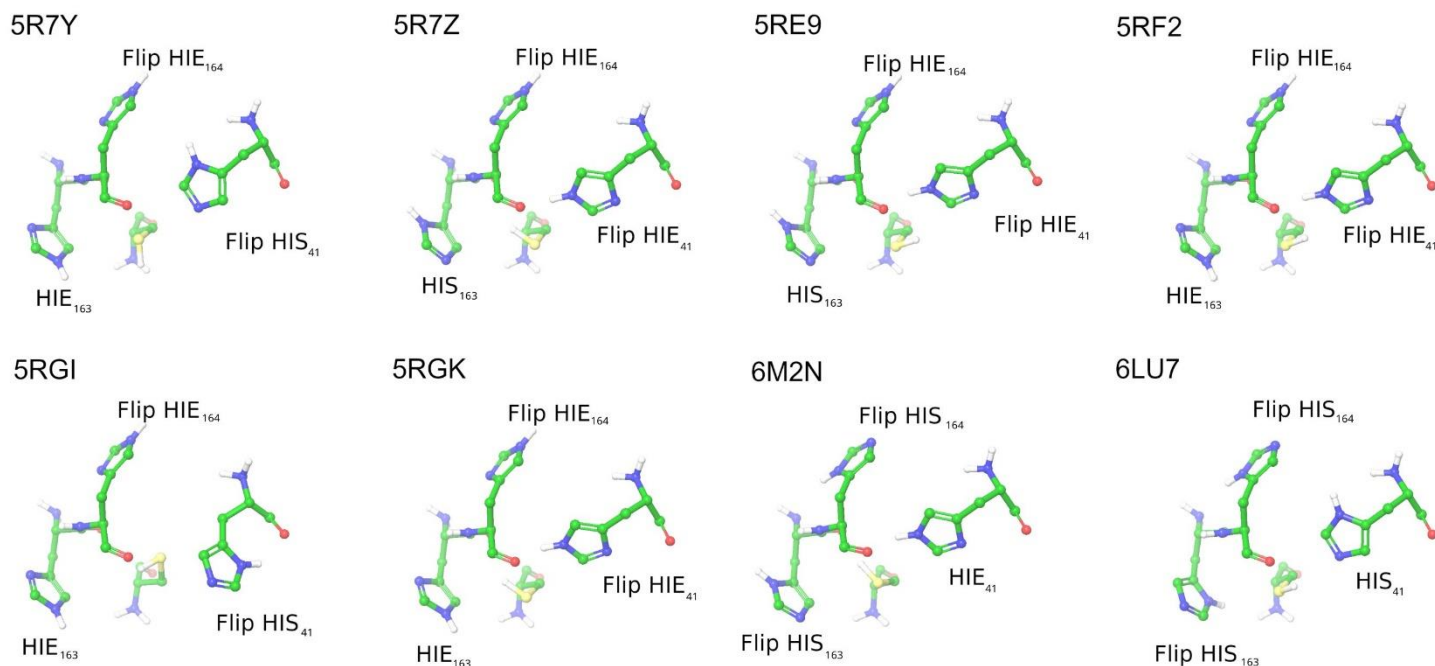


Figure 8. Set formed by His₄₁, His₁₆₃, His₁₆₄, and Cys₁₄₅ in the binding site of the structures that give the “worsts” results on the cross-docking evaluation.

PDB ID	His41	His163	His164	His172
5r7y	Flip HIS	HIE	Flip HIE	HIE
5r7z	Flip HIE	HIS	Flip HIE	HIE
5r80	Flip HIP	Flip HIS	Flip HIE	HIE
5r81	Flip HIE	HIS	Flip HIE	HIE
5r82	Flip HIE	Flip HIS	Flip HIS	HIE
5r83	Flip HIE	HIE	Flip HIP	HIE
5r84	HIS	HIE	Flip HIE	HIE
5re4	Flip HIE	HIE	Flip HIS	HIE
5re9	Flip HIE	HIS	Flip HIE	HIE
5reb	Flip HIE	HIS	Flip HIE	HIE
5reh	Flip HIE	HIE	HIS	HIE
5rez	Flip HIE	HIS	Flip HIE	HIE
5rf1	Flip HIE	Flip HIS	Flip HIE	HIE
5rf2	Flip HIE	HIE	Flip HIE	HIE
5rf3	Flip HIE	HIE	Flip HIP	HIE
5rf6	HIE	HIE	HIS	HIE
5rf7	Flip HIE	HIS	HIS	HIE
5rfe	Flip HIE	HIS	HIS	HIE
5rg1	HIE	HIS	Flip HIP	HIE
5rgh	Flip HIE	HIS	Flip HIS	HIE
5rgi	Flip HIS	HIE	Flip HIE	HIE
5rgk	Flip HIE	HIE	Flip HIE	HIE
5rgu	HIS	HIE	Flip HIE	HIE
5rgv	Flip HIE	HIS	Flip HIE	HIE
5rgw	HIS	HIE	HIS	HIE
5rgx	HIS	Flip HIS	Flip HIP	HIE

PDB ID	His41	His163	His164	His172
5rgy	Flip HIE	HIS	Flip HIP	HIE
5rgz	HIS	HIE	HIS	HIE
5rh0	Flip HIE	HIS	Flip HIE	HIE
5rh1	Flip HIE	HIE	HIE	HIE
5rh2	HIS	HIE	Flip HIE	HIE
5rh3	HIS	HIE	Flip HIE	HIE
5rh8	Flip HIE	HIS	Flip HIE	HIE
5rhd	Flip HIE	HIS	Flip HIE	HIE
6lu7	HIE	Flip HIS	Flip HIS	HIE
6m2n-1	HIS	Flip HIS	Flip HIS	HIE
6m2n-2	HIE	HIE	Flip HIE	HIE
6w63	HIE	HIE	HIS	HIE
6w79	HIE	HIE	HIS	HIE
7a1u	HIS	HIE	HIS	HIE
7ans	HIS	HIE	Flip HIE	HIE
7ap6	HIE	HIE	Flip HIE	HIE
7aqe	HIE	Flip HIE	Flip HIP	HIE
7kx5	HIE	HIE	HIE	HIE
7l0d	HIS	HIE	HIS	HIE
7l10	HIE	HIE	HIS	HIE
7l11	Flip HIE	HIE	HIS	HIE
7l12	HIE	HIE	HIS	HIE
7l14	HIS	Flip HIS	Flip HIS	HIE
5rgy	Flip HIE	HIS	Flip HIP	HIE
5rgz	HIS	HIE	HIS	HIE
5rh0	Flip HIE	HIS	Flip HIE	HIE

Table 2. Binding site histidine protonation of each structure after performing the Protein Preparation Wizard prior to any docking. Nomenclature used for histidine protonation: HIP (+1 charged, both δ - and ϵ -nitrogens protonated), HIS (neutral, δ -nitrogen protonated), and HIE (neutral, ϵ -nitrogen protonated).

We speculated that His41 should be in HIS while His163 should be in HIE (**note 1**) in order to guide the ligand properly, while His134 is too variable and hence should not influence too much in the results. For the sake of completion, we modified two structures: one with “good” results (5rgw) and one with “bad” ones (6lu7) and performed cross docking with the same conditions as before. Specifically, the modification was performed on the His163 (HIE, Flip HIS respectively) and mirroring the original state of the other (5rgw to Flip HIS, 6lu7 to HIE). The results can be seen on **Table 2** under the original ones of each structure, marked by the -m suffix and shows that with a single His change we can guide the structures to perform better or worse within our dataset.

After careful evaluation, the reason of this higher performance is so because their ligands are similar to each other and thus by necessity their histidine states are similar (see **Figure 2 and 5**). With that fact, we can no longer say that there is a “better” histidine state than the others, but that the histidine protonation of the binding site guides the docking results towards specific families of ligands.

6. CONCLUSIONS

The fight against the SARS-CoV 2 continues and many ongoing studies rely on the correct functioning of the docking program they use. In this study we evaluate how Glide HTVS, SP, XP, AutoDock Vina and AutoDock Vina Hydrated perform under different circumstances.

Re-docking studies are considered the easiest situation a docking program can confront and the situation the program should work best, its peak performance. Depending on conditions (dry, conserved and waters), the dockings programs evaluated here can reproduce 50 to 80% of the experimental poses, but only if taken into account at least the first 5. Using only the 1st lowers the results to 37-71% which is the pose the program tells you it is its best one. Glide SP and XP are the ones that come on top, performing similar to one another in all cases. Comparing all three groups we can see that using conserved waters might not necessarily improve the docking results, thus unless you have a good indication that a water should be there in all cases it might be better to perform a dry docking.

Evaluating which ligands each program reproduces correctly we found that some experimental poses are simply beyond the capabilities of these programs, be it because the electron density of the binding site or the ligand are not properly represented by the coordinates or other causes inherent to the programs themselves. Specifically in the case of AutoDock Vina Hydrated the caveat here seems to be the hydration step that needs to be performed, which does not allow some ligands to continue. The docking for this program reaches over 80% performance for those that have overcome the hydration step.

When considering a cross-docking scenario, Glide XP achieves better results than other programs, but SP does not fall behind. Since Glide XP requires a lot of computing power and SP has similar results requiring less, SP should be taken into consideration when performing docking at large scales. Since they have the same docking algorithm and only the sampling and scoring are different using both and then trying to perform consensus docking would not be significant but AutoDock Vina Hydrated does not fall behind the two too much and it might be interesting in the case given the high performance it boasts after the hydration step discards those it cannot hydrate.

Finally, the protonation of the histidines inside the binding site have a pivotal role in which families of ligands perform a successful docking and reproduce the experimental pose. Further testing all possible histidine protonations in different structures might shine some more light into the regard.

REFERENCES

1. Worldometer COVID-19 Coronavirus Pandemic at <https://www.worldometers.info/coronavirus/> (accessed on April 9, 2023).
 2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) at <https://coronavirus.jhu.edu/map.html> (accessed on April 9, 2023).
 3. Sadeghalvad, M. *et al.* Recent developments in SARS-CoV-2 vaccines: A systematic review of the current studies. *Rev Med Virol.* **33**, e2359 (2023).
 4. Macip, G. *et al.* A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet?. *Int J Mol Sci.* **23**, 259 (2021).
 5. Manelfi, C. *et al.* Combining Different Docking Engines and Consensus Strategies to Design and Validate Optimized Virtual Screening Protocols for the SARS-CoV-2 3CL Protease. *Molecules* **26**, 797 (2021).
 6. Gimeno, A. *et al.* Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition. *Int J Mol Sci.* **21**, 3793 (2020).
 7. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
 8. Kneller, D. W. *et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat Commun.* **11**, 7–12 (2020).
 9. Hu, Q. *et al.* The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19. *MedComm (Beijing).* **3**, 1–27 (2022).
 10. Antonopoulou, I., Sapountzaki, E., Rova, U. & Christakopoulos, P. Inhibition of the main protease of SARS-CoV-2 (Mpro) by repurposing/designing drug-like substances and utilizing nature's toolbox of bioactive compounds. *Comput Struct Biotechnol J.* **20**, 1306–1344 (2022).
 11. Yan, F. & Gao, F. An overview of potential inhibitors targeting non-structural proteins 3 (PLpro and Mac1) and 5 (3CLpro/Mpro) of SARS-CoV-2. *Comput Struct Biotechnol J.* **19**, 4868–4883 (2021).
 12. Gao, K. *et al.* Perspectives on SARS-CoV-2 Main Protease Inhibitors. *J Med Chem* **64**, 16922-16955 (2021).
 13. Xiong, M. *et al.* What coronavirus 3C-like protease tells us: From structure, substrate selectivity, to inhibitor design. *Med Res Rev.* **41**, 1965-1998 (2021).
 14. Awoonor-Williams, E. & Abu-Saleh, A. A. A. Covalent and non-covalent binding free energy calculations for peptidomimetic inhibitors of SARS-CoV-2 main protease. *Phys Chem Chem Phys.* **23**, 6746–6757 (2021).
 15. Martinez, M. A. Clinical Trials of Repurposed Antivirals for SARS-CoV-2. *Antimicrob Agents Chemother.* **64**, e01101-e01120 (2020).
 16. Kumar, S. *et al.* Drug repurposing against SARS-CoV-2 using computational approaches. *Drug Discov Today.* **27**, 2015–2027 (2022).
 17. Martinez, M. A. Efficacy of repurposed antiviral drugs: Lessons from COVID-19. *Drug Discov Today.* **27**, 1954–1960 (2022).
-

18. Gimeno, A. *et al.* The Light and Dark Sides of Virtual Screening: What Is There to Know?. *Int J Mol Sci.* **20**, 1375 (2019).
19. Pinzi, L. & Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int J Mol Sci.* **20**, 4331 (2019).
20. Pujadas, G. *et al.* Protein-ligand Docking: A Review of Recent Advances and Future Perspectives. *Curr Pharm Anal.* **4**, 1–19 (2008).
21. Macip, G. *et al.* Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Med Res Rev.* **42**, 744–769 (2021).
22. Bellera, C. L. *et al.* Can drug repurposing strategies be the solution to the COVID-19 crisis?. *Expert Opin Drug Discov.* **16**, 605–612 (2021).
23. Zev, S. *et al.* Benchmarking the Ability of Common Docking Programs to Correctly Reproduce and Score Binding Modes in SARS-CoV-2 Protease Mpro. *J Chem Inf Model.* **61**, 2957–2966 (2021).
24. Kumar, P. & Mohanty, D. Development of a Novel Pharmacophore Model Guided by the Ensemble of Waters and Small Molecule Fragments Bound to SARS-CoV-2 Main Protease. *Mol Inform.* **41**, e2100178 (2022).
25. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* **61**, 3891–3898 (2021).
26. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem.* **47**, 1739–1749 (2004).
27. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem.* **47**, 1750–1759 (2004).
28. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem.* **49**, 6177–6196 (2006).
29. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* **31**, 455–461 (2010).
30. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model.* **53**, 1893–1904 (2013).
31. Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).
32. Llanos, M. A. *et al.* Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study. *J Chem Inf Model.* **61**, 3758–3770 (2021).
33. Bassani, D., Pavan, M., Bolcato, G., Sturlese, M. & Moro, S. Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro. *Pharmaceuticals (Basel)* **15**, 180 (2022).

34. Ngo, S. T., Tam, N. M., Pham, M. Q. & Nguyen, T. H. Benchmark of Popular Free Energy Approaches Revealing the Inhibitors Binding to SARS-CoV-2 Mpro. *J Chem Inf Model.* **61**, 2302–2312 (2021).
35. Cereto-Massagué, A. *et al.* The good, the bad and the dubious: VHELBS, a validation helper for ligands and binding sites. *J Cheminform.* **5**, 36 (2013).
36. Schrödinger Release 2021-1; Schrödinger, LLC. : N. Y. N. U. 2021. Protein Preparation Wizard.
37. OpenEye Scientific Software; OpenEye Scientific Software: Santa Fe, NM. U. 2019. OMEGA 3.1.1.2.
38. Schrödinger Release 2020-3; Schrödinger, LLC. : N. Y. N. U. 2020. LigPrep.
39. GitHub - forlilab/Meeko: Interfacing RDKit and AutoDock. <https://github.com/forlilab/Meeko>.
40. Ten Brink, T. & Exner, T. E. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. *J Chem Inf Model.* **49**, 1535–1546 (2009).
41. Park, M.-S., Gao, C. & Stern, H. A. Estimating binding affinities by docking/scoring methods using variable protonation states. *Proteins* **79**, 304–314 (2011).
42. Santos, V. C., Campos, A. C. B., Waldner, B. J., Liedl, K. R. & Ferreira, R. S. Impact of different protonation states on virtual screening performance against cruzain. *Chem Biol Drug Des.* **99**, 703–716 (2022).
43. Pavlova, A. *et al.* Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *Chem Sci.* **12**, 1513–1527 (2021).
44. Kneller, D. W. *et al.* Direct Observation of Protonation State Modulation in SARS-CoV-2 Main Protease upon Inhibitor Binding with Neutron Crystallography. *J Med Chem.* **64**, 4991-5000 (2021).
45. Ono, J., Koshimizu, U., Fukunishi, Y. & Nakai, H. Multiple protonation states in ligand-free SARS-CoV-2 main protease revealed by large-scale quantum molecular dynamics simulations. *Chem Phys Lett.* **794**, 139489 (2022).

Supplementary material for

Are the protein-ligand docking programs good enough to predict experimental poses of non-covalently inhibitors bound to SARS-CoV-2 main protease?

Guillem Macip¹, Júlia Vilalta¹, Júlia Mestres-Truyol¹, Pol Garcia-Segura¹, Bryan Saldivar-Espinoza¹, Adrià Cereto-Massagué², Santiago Garcia-Vallvé^{1,*}, Gerard Pujadas^{1,*}.

¹Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Research group in Cheminformatics & Nutrition, Tarragona, Catalonia, Spain.

²EURECAT Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Catalonia, Spain.

*Corresponding authors

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Two-dimensional structure of the inhibitors non-covalently bound to each of the 48 complexes analyzed in the present study. Each inhibitor is identified by the code of the PDB file to which it belongs.

Figure S2. Equivalent to Figure 5 but for Glide HVTS results.

Figure S3. Equivalent to Figure 5 but for Glide SP results.

Figure S4. Equivalent to Figure 5 but for AutoDock Vina (dry binding site) results.

Figure S5. Equivalent to Figure 5 but for AutoDock Vina (hydrated ligand) results.

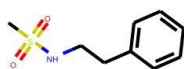
SUPPLEMENTARY TABLES

Table S1. Labels for water molecules located within 5.0 Å of any inhibitor atom in the 48 non-covalent complexes considered in this study. Bold and underlined labels correspond to water molecules that do not fit to the electron density map and, therefore, have been removed prior to any protein-ligand docking. A total of 298 water molecules have reliable coordinates and are located within 5.0 Å of any inhibitor atom. Some structures also contain a DMS (*i.e.*, 5r82, 5re9, 5rf6, 5rgh, 5rgy, 5rhd, 7a1u, 7ans) or IMD (*i.e.*, 7a1u) molecule at the binding site.

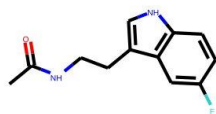
Table S2. Water molecules at equivalent locations in the different non-covalent complexes. The labels of water molecules that are within 5.0 Å of any inhibitor atom and that are in equivalent spatial coordinates in different complexes are shown in the same column (208 water molecules in total). The remaining 90 water molecules (see Table S1) are exclusives of only one of the complexes. The orange columns are equivalent water molecules found in at least 30% of the 48 complexes under study.

Table S3. Overview of all the programs of each group. Name, Total poses generated, % of the total poses below 2.0Å, Total poses under 1.0Å, Total poses between 1 and 2.0Å, Total failed poses, Total compounds that have any poses under 2.0Å, Total compounds that do not have any poses under 2.0Å. % of poses under 2.0Å in the first pose, Total compounds that have the 1s pose under 2.0Å

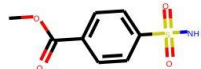
Table S4. Results of each of the 48 structures used separated by groups and docking program. Each cell contains two numbers. The first is the number of poses under 2 Å while the second is the total number of poses generated by the program. The columns Good Ligand and Good Binding Site refer to the evaluation performed with VHELBS to validate the coordinates of the ligand and binding site respectively so as they fit their electron density map.



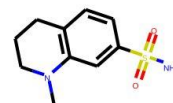
5r7y



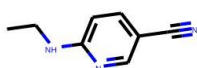
5r7z



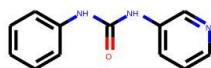
5r80



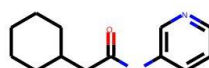
5r81



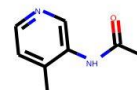
5r82



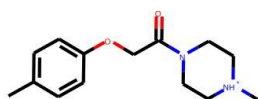
5r83



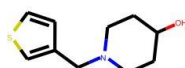
5r84



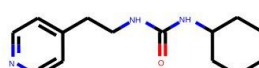
5re4



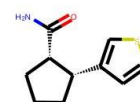
5re9



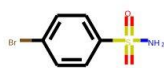
5reb



5reh



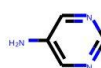
5rez



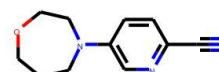
5rf1



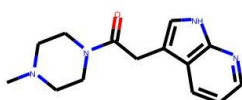
5rf2



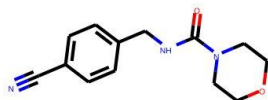
5rf3



5rf6



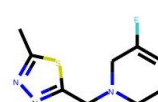
5rf7



5rfe



5rg1



5rgh



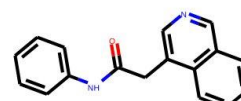
5rgi



5rgk



5rgu



5rgv



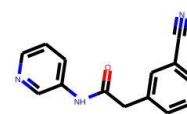
5rgw



5rgx



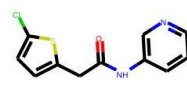
5rgy



5rgz



5rh0



5rh1



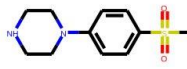
5rh2



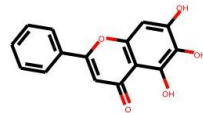
5rh3



5rh8



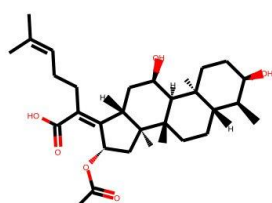
5rhd



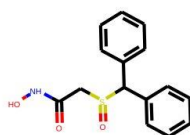
6m2n-1/6m2n-2



6w63/6w79



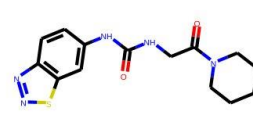
7a1u



7ans



7ap6



7aqe



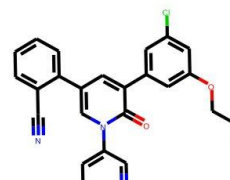
7kx5



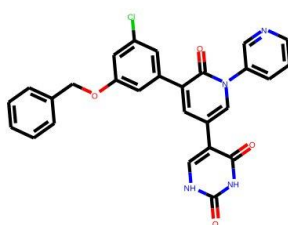
7l0d



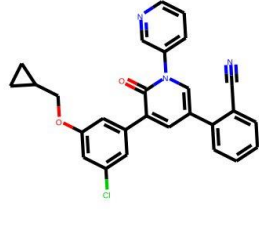
7l10



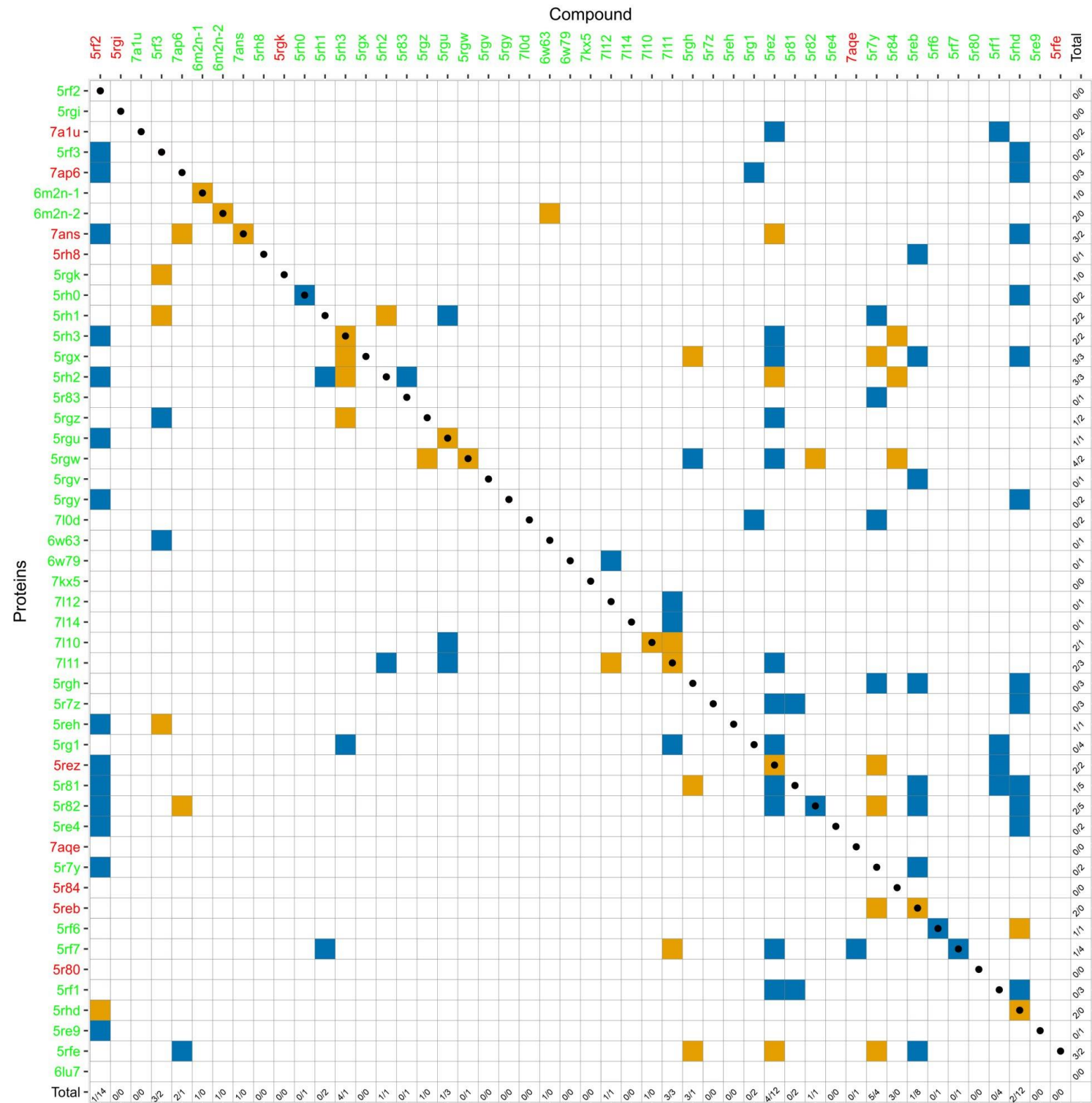
7l11

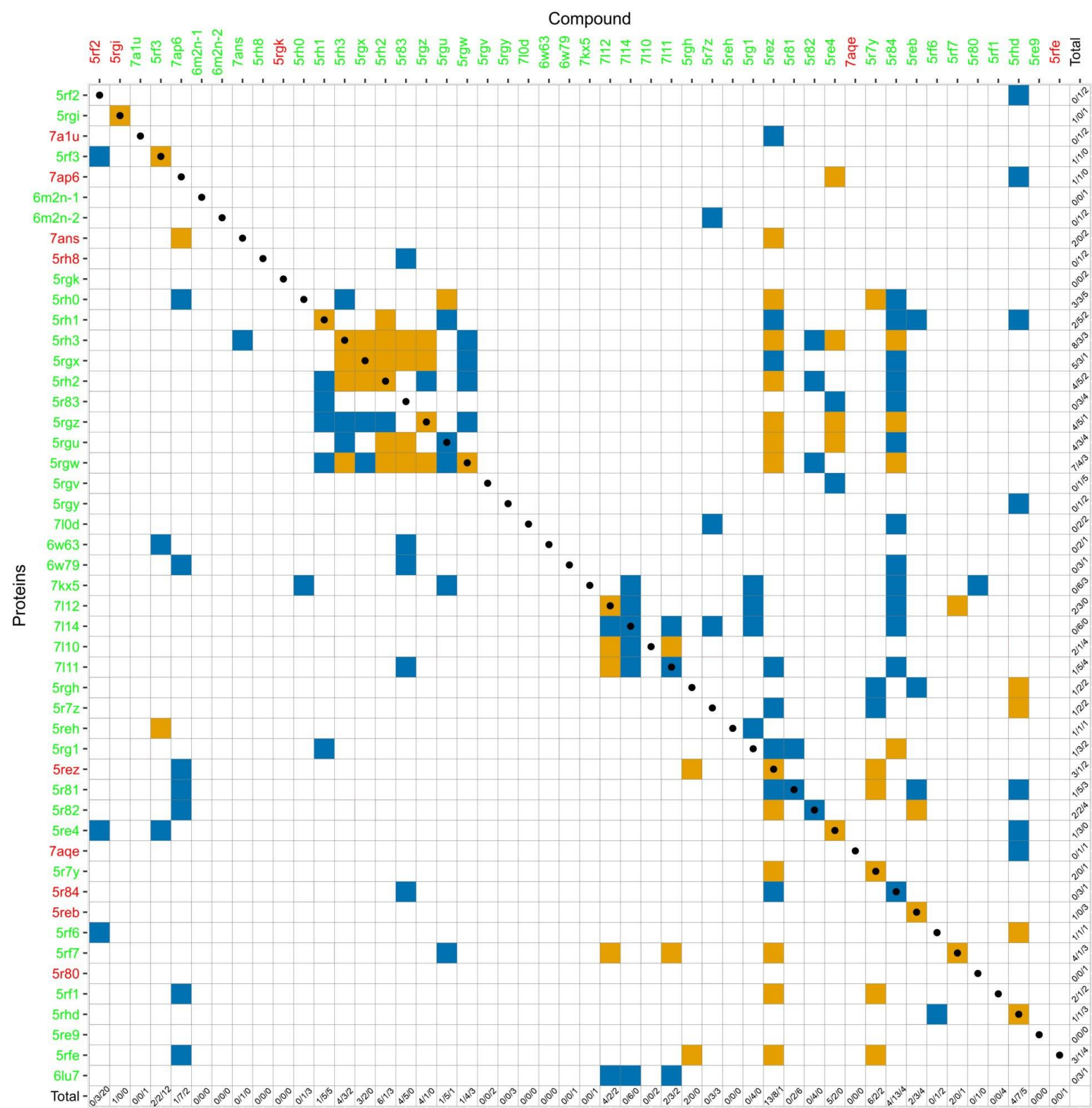


7l12

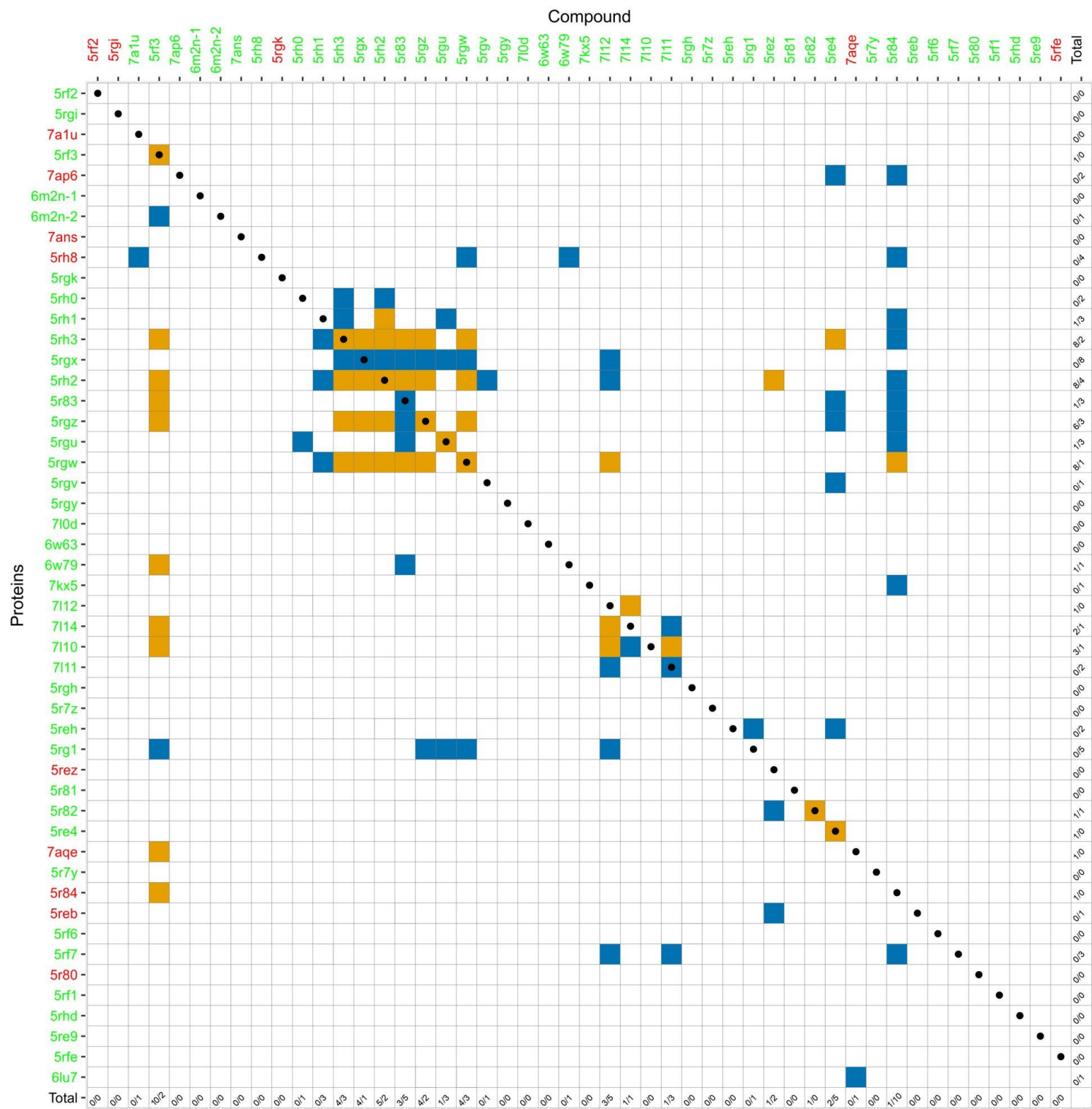


7l14









Manuscript 3

PDB code	Labels for water molecules located within 5.0 Å of any inhibitor atom
5r7y	1281A, 1319A, 1322A
5r7z	580A, 596A, 601A , 697A, 751A
5r80	518A, 542A, 676A
5r81	1173A, 1259A, 1280A, 1298A, 1422A
5r82	1171A, 1189A, 1250A, 1284A, 1407A, 1410
5r83	547A, 595A, 642A, 808A
5r84	1118A, 1313A
5re4	506A, 617A, 640A, 642A, 719A, 786A , 820A
5re9	734A, 749A, 751A, 802A, 841A, 842A
5reb	669A, 689A, 749A, 760A, 798A
5reh	501A, 622A, 667A, 681A, 686A, 695A, 772A
5rez	537A, 565A, 661A, 670A, 723A, 755A, 775A, 818A
5rf1	519A, 608A, 710A
5rf2	595A, 710A, 722A, 815A
5rf3	509A, 701A, 724A, 735A, 745A
5rf6	506A, 515A, 536, 619A, 827A
5rf7	504A, 507A, 512A, 519A, 595A, 619A, 647A, 704A, 765A
5rfe	504A, 514A, 585A, 600A, 615A, 665A, 692A, 715A, 719A , 778A, 797A
5rg1	531A, 632A, 637A, 669A, 728A, 789A
5rgh	623A, 804A
5rgi	516A, 529A , 638A, 683A, 726A(A), 726A(B) , 791A
5rgk	607A, 642A, 721A, 793A
5rgu	1103A, 1212A, 1244A
5rgv	1166A, 1169A, 1233A, 1266A, 1271A
5rgw	1103A, 1107A, 1173A, 1260A
5rgx	1103A, 1116A, 1152A, 1183A, 1264A, 1275A
5rgy	1241A, 1253A, 1261A, 1288A, 1289A
5rgz	1126A, 1161A, 1165A, 1270A
5rh0	1170A, 1189A, 1220A
5rh1	1232A, 1249A, 1255A
5rh2	1136A, 1160A, 1200A, 1232A, 1254A, 1258A, 1269A, 1344A, 1362A, 1365A
5rh3	1118A, 1185A, 1189A, 1240A, 1271A, 1284A, 1299A, 1323A, 1368A, 1373A
5rh8	555A, 673A, 681A , 711A, 714A, 726A, 740A
5rhd	502A, 547A, 594A, 617A, 759A, 814A, 817A
6m2n-1	532A
6m2n-2	532B
6w63	515A, 518A, 546A, 576A, 594A, 600A, 606A, 618A, 650A, 671A
6w79	510A, 545A, 571A, 573A, 623A, 638A, 666A, 705A, 744A, 750A, 767A, 775A, 793A , 836A, 892A, 913A, 918A, 945A
7a1u	508A, 517A, 628A, 644A, 707A, 709A, 714A, 732A, 747A
7ans	592A, 596A, 616A, 648A, 743A, 774A
7ap6	502A, 602A, 626A, 640A, 649A, 654A, 739A
7aqe	610A, 628A, 634A, 646A, 661A, 676A
7kx5	No water
7l0d	524A, 535A, 549A, 577A, 592A
7l10	553A, 564A, 574A, 579A, 601A, 604A, 635A, 669A, 671A, 731A
7l11	510A, 524A, 596A, 609A, 662A, 688A, 692A, 701A, 712A, 717A, 724A, 734A
7l12	514A, 530A, 544A, 559A, 597A, 620A, 621A, 663A, 667A, 689A , 718A, 815A
7l14	508A, 512A, 524A, 526A, 541B, 631A, 659A, 699A, 716A, 736A, 765A, 771A, 772A, 775A, 776A, 802A

PDB code	Water molecules at equivalent locations in the different non-covalent complexes														
5r7y						1281 A				1322A					
5r7z						751A	697A	580A	596A						
5r80						518A	676A	542A							
5r81						1259A	1298 A	1173A				1280 A			
5r82						1171A	1284 A		1407 A	1189 A		1410	1250 A		
5r83			595A		642A							808A		547A	
5r84		1313 A			1118 A										
5re4			642A	640 A	617A			506A	820A						
5re9				734 A			751A		802A				749A	841A	
5reb							689A				749A		669A	760A	
5reh			686A		695A		681A	622A	501A						
5rez								661A	537A		723A	755A			
5rf1						608A	710A		519A						
5rf2													710A		
5rf3	509A		745A		724A		701A						735A		
5rf6	506A							536	515A	827A					619 A
5rf7	512A		704A		619A	647A			507A					595A	
5rfe	504A		715A				600A	585A	514A		615A	778A	692A		
5rg1		531A		632 A	669A					789A	637A				
5rgh										804A	623A				
5rgi	516A		726A(A)					683A							638 A
5rgk					642A		607A			793A			721A		
5rgu	1103 A							1212 A							
5rgv					1169 A	1233A								1166A	
5rgw	1103 A							1260 A	1107A		1173 A				
5rgx		1116 A	1264A			1152A					1183 A	1275A			
5rgy								1253 A							
5rgz		1126 A						1165 A			1161 A	1270A			
5rh0					1189 A	1220A		1170 A							
5rh1		1249 A			1232 A	1255A									
5rh2			1232A		1269 A	1258A			1136A	1362 A	1160 A	1254A		1200A	
5rh3	1189 A		1284A			1271A		1271 A	1118A	1373 A	1185 A	1323A	1368 A	1240A	
5rh8	555A			714 A	681A					711A					
5rhd								547A		814A		594A		759A	617 A
6m2n-1															
6m2n-2															
6w63	515A		546A	518 A						671A	576A				594 A
6w79	510A		892A	571 A				573A	775A	545A	623A		913A		750 A
7a1u					508A	709A	732A	707A							644 A

Manuscript 3

7ans					743A 648A						596A				616 A		
7ap6	502A		649A		626A	739A	602A				654A				640 A		
7aqe		610A			634A		661A				676A		628A		646 A		
7kx5																	
7l0d	535A			524 A								592A			577 A		
7l10	669A		635A	671 A				574A			604A				601 A		
7l11				524 A		688A		596A							609 A		
7l12	544A		597A	621 A		663A			559A		620A		718A		530 A		
7l14			699A	512 A		765A		631A							716 A		
Waters per cluster	15	6	16	11	16	18	17	16	15	13	15	10	8	9	8	11	4
% in 48 complexes	31.25	12.5	33.3	22.9	33.3	37.5	35.4	33.3	31.25	27.0	31.25	20.8	16.6	18.75	16.6	22.9	8.3

PDB code	Not in a cluster											
5r7y	1319A											
5r7z	601A											
5r80												
5r81	1422A-Subcluster?											
5r82												
5r83												
5r84												
5re4	719A	786A										
5re9	842A											
5reb	798A											
5reh	667A	772A										
5rez	565A	670A	775A	818A								
5rf1												
5rf2	595A	722A	815A									
5rf3												
5rf6												
5rf7	504A	519A	765A-Subcluster?									
5rfe	665A	719A	797A									
5rg1	728A											
5rgh	623A	DMS405A										
5rgi	529A	791A										
5rgk												
5rgu	1244A											
5rgv	1266A	1271A										
5rgw												
5rgx	1103A											
5rgy	1241A	1261A	1288A	1289A								
5rgz												
5rh0												
5rh1												
5rh2	- 1344A-Subcluster?	1365A										
5rh3	1299A											
5rh8	673A	726A	740A									
5rhd	502A	817A	DMS405A									
6m2n-1	532A											
6m2n-2	532B											
6w63	600A	606A	618A	650A								
6w79	638A	666A	705A	744A	767A	793A	836A	918A	945A			
7a1u	628A	714A	747A									
7ans	592A	774A										
7ap6												
7aqe												
7kx5												
7l0d	549A											
7l10	553A	564A	579A	731A								
7l11	510A	662A	692A	701A	712A	717A	724A	734A				
7l12	514A	667A	689A	815A								
7l14	508A	524A	526A	541B	659A	736A	771A	772A	775A	776A	802A	
Waters with no clusters	99											

	Total poses	% Hits	<1	<2	>2.0	Compounds found	Not found	% Hits 1s pose	Compounds 1st pose
Dry									
Glide_htvs	540	8.70	20	27	493	25	23	35.42%	17
Glide_sp	937	21.88	68	137	732	34	14	52.08%	25
Glide_xp	321	33.64	61	47	213	33	15	50.00%	24
Vina	928	4.63	19	24	885	32	16	29.17%	14
Conserved									
Glide_htvs	502	11.35	23	34	445	24	24	35.42%	17
Glide_sp	937	23.48	72	148	717	35	13	52.08%	25
Glide_xp	269	34.94	62	32	175	33	15	52.08%	25
Water									
Glide_htvs	437	15.56	27	41	369	26	22	39.58%	19
Glide_sp	936	25.96	82	161	693	40	8	72.92%	35
Glide_xp	273	47.62	87	43	143	40	8	68.75%	33
Vina	743	18.84	49	91	603	32	16	43.75%	21

PDB ID	VHELIBS	Dry				Conserved				Water			
		HTVS	SP	XP	Vina	HTVS	SP	XP	Vina	HTVS	SP	XP	Vina
5r7y	√/√	1/12	3/19	0/19	0/19	0/11	6/19	0/3	NA/NA	4/13	14/19	6/7	NA/NA
5r7z	√/√	2/14	3/19	1/1	0/19	2/11	1/19	1/7	2/19	1/10	1/19	1/3	2/19
5r80	√/x	0/3	6/19	0/1	2/19	2/5	7/19	1/4	NA/NA	2/5	7/19	1/4	NA/NA
5r81	√/√	6/14	9/19	1/1	4/19	7/10	9/19	NA/NA	NA/NA	5/13	10/19	1/1	NA/NA
5r82	√/√	2/13	4/19	4/5	0/19	0/6	5/19	1/1	4/19	0/2	7/19	4/4	4/19
5r83	√/√	0/16	3/19	1/4	1/18	0/14	5/19	0/4	9/19	0/6	7/19	1/2	9/19
5r84	√/x	0/3	14/19	14/16	2/19	0/5	16/19	9/9	9/19	0/1	17/19	10/10	9/19
5re4	√/√	0/9	2/19	NA/NA	2/19	0/11	0/19	NA/NA	6/19	0/9	2/19	NA/NA	6/19
5re9	√/√	0/20	0/20	0/17	0/19	0/20	0/20	0/9	0/18	3/12	3/20	2/20	0/18
5reb	√/x	7/13	9/19	1/4	0/19	5/13	10/19	0/2	2/19	7/13	17/19	5/5	2/19
5reh	√/√	0/11	0/19	0/7	0/19	0/4	0/19	1/5	0/18	0/7	0/19	0/10	0/18
5rez	√/x	8/12	11/19	1/1	0/19	6/14	11/19	3/3	4/15	9/14	16/19	3/3	4/15
5rf1	√/√	2/8	9/19	NA/NA	2/19	3/4	8/19	2/2	NA/NA	3/4	8/19	2/2	NA/NA
5rf2	x/√	7/7	18/18	10/10	3/19	5/5	15/16	8/8	0/19	7/7	13/19	3/3	0/19
5rf3	√/√	1/11	5/19	2/3	4/19	1/5	5/19	1/1	19/19	1/3	2/15	1/1	19/19
5rf6	√/√	2/10	12/19	3/3	5/19	5/8	10/19	4/5	8/19	7/8	16/19	3/3	8/19
5rf7	√/√	7/15	8/20	4/9	1/19	12/15	18/20	2/2	4/16	6/13	2/20	1/1	4/16
5rfe	x/√	0/10	0/19	0/6	0/19	0/12	2/19	5/7	0/19	3/9	2/19	3/3	0/19
5rg1	√/√	0/12	0/19	0/2	0/19	0/12	0/19	0/3	2/19	0/12	0/19	1/6	2/19
5rgh	√/√	2/12	0/19	0/7	1/19	2/12	0/19	0/4	2/19	1/13	3/19	2/7	2/19
5rgi	x/√	0/3	3/19	2/10	1/19	0/7	2/19	4/5	3/19	0/7	2/19	3/5	3/19
5rgk	x/√	0/16	0/19	0/19	0/19	0/11	0/19	0/4	0/16	0/3	0/19	0/8	0/16
5rgu	√/√	3/15	4/19	4/5	1/19	1/14	3/19	2/2	16/18	2/15	3/19	1/1	16/18
5rgv	√/√	1/18	3/20	0/11	0/19	1/14	2/20	0/11	NA/NA	4/13	2/20	4/6	NA/NA
5rgw	√/√	1/16	2/19	1/1	1/19	1/14	4/19	4/4	6/19	1/14	4/19	4/4	6/19
5rgx	√/√	0/14	2/19	2/5	0/19	0/11	0/19	NA/NA	6/19	0/12	1/19	2/2	6/19
5rgy	√/√	1/11	7/19	3/5	1/19	3/13	7/19	3/4	11/19	3/10	2/19	6/6	11/19
5rgz	√/√	0/11	2/19	1/1	0/19	0/13	3/19	1/1	7/19	0/12	3/19	NA/NA	7/19
5rh0	√/√	2/9	0/19	0/4	1/19	0/5	3/19	0/1	4/19	0/5	3/19	0/1	4/19
5rh1	√/√	0/14	8/19	1/6	4/19	0/9	8/19	1/5	7/19	0/6	8/19	2/8	7/19
5rh2	√/√	2/9	6/19	1/1	1/19	2/10	4/19	2/2	5/19	1/11	1/19	1/2	5/19
5rh3	√/√	2/9	5/19	1/1	1/19	2/10	4/19	2/2	6/19	6/6	7/19	1/1	6/19
5rh8	√/x	0/13	0/19	0/4	2/19	0/8	0/19	0/6	0/7	0/6	18/19	10/10	0/7
5rhd	√/√	3/3	16/19	2/2	4/19	4/4	17/19	5/5	NA/NA	2/2	14/19	2/2	NA/NA
6m2n	√/√	3/14	8/20	6/7	1/19	3/14	8/20	6/7	6/19	4/20	7/20	8/11	6/19
6m2n	√/√	5/20	9/20	5/6	3/19	5/20	9/20	5/6	10/19	3/19	9/20	8/9	10/19
6w63	√/√	0/2	0/20	12/20	1/19	0/6	0/20	8/20	2/15	0/6	1/20	14/14	2/15
6w79	√/√	0/6	0/20	0/20	1/19	0/12	1/20	1/20	2/19	0/3	0/20	0/20	2/19
6w79	√/√	0/8	0/20	2/13	2/19	0/17	1/20	2/20	3/13	0/7	0/20	0/20	3/13
7a1u	√/x	0/1	0/19	12/15	1/19	0/3	2/19	4/4	1/14	3/5	19/19	11/11	1/14
7ans	√/x	1/10	8/19	4/4	3/19	4/10	18/19	1/1	NA/NA	4/10	13/19	2/2	NA/NA
7ap6	√/x	0/17	0/19	0/5	0/19	0/16	0/19	0/8	0/19	0/16	0/19	0/9	0/19
7aqe	x/x	0/18	0/19	2/12	0/19	1/15	0/19	2/19	0/19	0/14	0/19	2/3	0/19
7kx5	√/√	0/2	0/17	7/10	2/18	0/4	0/19	2/4	3/18	0/4	0/19	2/4	3/18
7l0d	√/√	0/4	1/19	0/6	0/19	0/4	0/19	0/12	0/18	0/2	0/19	0/6	0/18
7l10	√/√	3/12	16/19	3/3	2/18	2/9	16/19	3/3	7/18	2/9	11/19	1/1	7/18
7l11	√/√	2/9	17/19	3/3	1/19	2/7	17/19	3/3	4/19	NA/NA	11/19	1/1	4/19
7l12	√/√	0/20	12/20	3/3	1/19	0/15	7/20	8/10	4/8	0/20	2/20	3/9	4/8
7l14	√/√	0/11	16/19	3/3	2/19	0/10	19/19	1/1	4/18	0/6	8/19	2/2	4/18

Manuscript 3

Discussion

Discussion

In this present thesis we focused on the evaluation and validation of several protein docking programs when used as a part of a VS for the treatment of SARS-CoV-2, specifically targeted to its M-pro. Additionally we also provide a collection of 1765 M-pro SARS-CoV-2 inhibitors to help verify a virtual screening method (**Manuscript 1**). This set is expected to be helpful as it contains many inhibitors that could be used to improve derivatives of lead compounds.

As the pandemic started, a myriad of peer-reviewed articles began to be published with the aim of predicting SARS-CoV-2 M-pro inhibitors among approved or investigational drugs. In order to evaluate them, an exhaustive comparison was done between them (**Manuscript 2**). We found that very few peer-reviewed articles, even with similar methodologies and datasets, predicted the same inhibitors. Although the SARS-CoV-2 pandemic demanded an accelerated drug development process, safeguards are required to protect this process, both to save time and to obtain meaningful results. The methodologies employed and the results obtained by computational approaches must be validated in order for these kinds of approaches not to be devalued and efforts to be directed toward the most promising compounds. Furthermore, the flexibility and plasticity of the SARS-CoV-2 M-pro make it a challenging target for small-molecule inhibitor design.

In **Manuscript 2**, we collected 61 peer-reviewed articles that employ a protein-ligand docking-based VS methodology to predict compounds that will act as SARS-CoV2 M-pro inhibitors. With the aim to reevaluate with up to date information, we analysed 245 peer-reviewed articles collected from November 2020 to December 2022 (see **Annex**). In contrast to the start of the pandemic, only 29 (11,6%) presented raw docking scores as the last step of their VS lacking any kind of internal validation. This marks a promising but still somewhat discouraging landscape as of at least June 2022, SARS-CoV-2 M-pro docking results were published without further validation as possible inhibitors. The other 221 (88,4%) had at least another step in their VS. Molecular dynamics, binding free energy calculations and machine learning are one of the most commons. The most surprising finding of the analysis was that 104 (41,6%) either tried their hits in vitro or extracted them from the bibliography to justify the outcomes of their VS protocols. This is, of course, the ideal situation for all VS protocols because it demonstrates the reliability of their predictions by providing a tangible result. In terms of docking validation methods, 89 (35,6%) use re-docking and only 6 (2,4%) use cross docking, while 21 (8,4%) use active enrichment in silico. Ten percent (10%) of the that use active/decoys back up their claims with IC50 results. Even if we consider all previous numbers as individual peer-reviewed articles (which we cannot because some overlap and use multiple validation methods), less than 47% of the examined do not directly validate their docking step.

Since most of the previous studies did not have a validation step, a key part of a virtual screening, we describe in **Manuscript 2** how a group of known SARS-CoV-2 M-pro inhibitors and inactive compounds can be used to validate the protein-ligand docking step. For this purpose, we extracted from the earlier stages of **Manuscript 1** (as they were written in tandem) a set of 150 inhibitors, and a second set of 81 inhibitors and 113 inactive compounds from the COVID Moonshot project¹. Using both sets, we performed a protein-ligand docking with two different SARS-CoV-2 M-pro structures, the 6LU7 and 6WQF², and three docking programs, i.e, FRED³, Glide^{4,5} and Autodock Vina^{6,7}. For Glide, we used three distinct docking methodologies: Glide HTVS, Glide SP and Glide XP. So in total we considered 5 different docking methods, which appeared to be the most common in the reviewed studies. As seen in the introduction, to overcome the limitations of docking programs and correct them, they can be combined with other methods such as MM-GBSA to enable the protein flexibility to be explored^{8,9}. The

Discussion

desirable outcome, a negative correlation between either the docking score or the ΔG value from MM-GBSA with the experimental pIC50 would signify a way to directly validate the outcome of a docking result. The more negative the docking score or ΔG value, the higher pIC50 value. The correlations between docking score/ ΔG value and activity were unsatisfactory even when it improved if covalent and noncovalent inhibitors were treated separately. Most correlations were found to be positive, showing a clear limitation of both the docking programs and the MM-GBSA. Additionally, these methods are unable to discriminate M-pro inhibitors from inactive drugs as the area under the ROC curve (AUC) values were close to 0.5.

The KDeep server¹⁰ proved useful for predicting the protein-ligand affinity (measured as ΔG values) of a group of 40 non-covalent inhibitors co-crystallized with M-pro SARS-CoV-2 and relating ΔG values to pIC50 (**Manuscript 2**). The correlation between the pIC50 and ΔG values was good, with a Pearson correlation coefficient of -0.688 (p-value 9.32E-07). However, this correlation decreased when the ligands were transformed to SMILES format to lose the binding coordinates of the ligand and were docked with the five docking programs used above.

Continuing with the previous endeavour and focusing on non-covalent inhibitors, we decided to perform a benchmark of the different ligand-docking programs (**Manuscript 3**). Since the docking score did not correlate with the activity of the inhibitors, maybe the docking programs are simply unable to reproduce the correct pose of the complex with the M-pro, thus a way to compare with experimental poses is needed. Given the increasingly high number of SARS-CoV-2 M-pro structures in the PDB that are complexed with non-covalent inhibitors, we decided to work with them. This could also allow us to compare each structure, specifically the binding site, and find the best possible one for future VS of non-covalent inhibitors, as the most commonly used (6LU7) has been so primarily because it was one of the first.

To analyse the performance of the four most widely used docking programs/methodologies (Glide SP, Glide XP, Glide HTVS and AutoDock Vina) we use both re-docking and cross docking experiments (**Manuscript 3**) on 48 non-covalent M-pro/ligand complexes from the PDB. When performing the re-docking experiments, we took into account the following three situations: **a)** considering the waters present in the structure as part of the binding site (water), **b)** removing the water molecules (dry), and **c)** keeping only the conserved water molecules across structures (at least 30%). Considering all the different situations, Glide SP and XP obtain the better results whereas Vina competes only with the hydrated docking.

In cross docking studies, several ligand-protein complexes are aligned and then split to create two libraries: one with the proteins and one with the ligands. Each component of the protein library is then used to dock the ligand library and thus it is a good way to evaluate how docking will behave during a virtual screening. Using the same 48 complexes “dry” for each of the four programs and hydrated docking of AutoDock Vina (as it hydrates the ligand itself) we try to find if a combination of: (a) docking program; (b) PDB structure; and (c) target preparation is better than the other in predicting the experimental pose.

The “best” PDB structures in cross docking are 5rgx, 5rh2 and 5rgw (with 10, 11 and 11 poses within 2.0 Å, respectively, of the corresponding experimental poses). In order to better understand why some structures performed differently than others, we also analysed the binding site and ligands with a validation helper for ligands and binding sites (VHELIBS), which checks how ligand and binding site coordinates fit to the electron density map. Generally speaking, structures and ligands that the evaluation with VHELIBS deemed “bad” have worse results than the others, not scoring a single pose at 2.0 angstroms.

In order to better understand the disparity in the docking results in all the different structures from the PDB, we also analysed the histidines present in the binding site (**Manuscript 3**). We speculated that His41 should be in HIS (neutral, δ -nitrogen protonated) while His163 should be in HIE (neutral, ϵ -nitrogen protonated) to guide the ligand properly, while His134 is too variable and hence should not influence too much in the results. After careful evaluation, the reason of this higher performance (5rgx, 5rh2 and 5rgw) occurs because their ligands are similar to each other and thus by necessity their histidine states are similar. With that fact, we can no longer say that there is a “better” histidine state than the others, but that the histidine protonation of the binding site guides the docking results towards specific families of ligands.

REFERENCES

1. The COVID Moonshot Consortium et al. Open Science Discovery of Potent Non-Covalent SARS-CoV-2 Main Protease Inhibitors. *bioRxiv* 2020.10.29.339317
2. Kneller, D. W. *et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat Commun.* **11**, 3202 (2020).
3. McGann, M. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model.* **51**, 578–596 (2011)
4. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem.* **47**, 1739–1749 (2004).
5. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47**, 1750–1759 (2004).
6. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem.* **31**, 455–461 (2010).
7. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* **61**, 3891–3898 (2021).
8. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov.* **10**, 449–461 (2015).
9. Wang, E. *et al.* End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev.* **119**, 9478–9508 (2019).
10. Jiménez, J., Škalič, M., Martínez-Rosell, G. & de Fabritiis, G. KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J Chem Inf Model.* **58**, 287–296 (2018).

Conclusions

Conclusions

In this thesis, the following conclusions were drawn:

- We have collected 1765 SARS-CoV-2 M-pro inhibitors which could be useful to validate a VS. It only includes compounds whose inhibitory potential against M-pro from SARS-CoV-2 has been established experimentally.

From the analysis of 61 peer-reviewed articles that employ a protein-ligand docking-based VS methodology to predict compounds that will act as SARS-CoV2 M-pro inhibitors:

- Most VS protocols, based on protein-ligand docking, that predict drugs as SARS-CoV-2 M-pro inhibitors have not been properly validated.

From the analysis of 150 SARS-CoV-2 M-pro inhibitors:

- The activity of known SARS-CoV-2 M-pro inhibitors do not correlate with their docking scores nor the ΔG calculated using the MM-GBSA method. The docking score cannot be used to define or predict the activity of a given compound.
- The docking score and nor the ΔG calculated using the MM-GBSA method are unable to distinguish actives from inactive compounds.
- The results improve when covalent and noncovalent inhibitors were treated separately.

From the analysis of the performance of the four most widely used docking programs:

- The most popular docking programs (Autodock Vina and Glide) can reproduce, when at least the first 5 poses are taken into account, up to 50%-80% of a set of 47 experimentally known SARS-CoV-2 M-pro inhibitor poses in a re-docking study. For the 1st pose the results are 37-71% accurate depending on the conditions (water, dry, conserved). The experimental poses of some ligands cannot be reproduced via re-docking.
- In cross docking experiments Glide XP and SP are the best performers in all cases but Autodock Vina Hydrated docking works quite well if the compounds can be hydrated as it reaches over 80% performance when removing the compounds unable to be hydrated.
- 6LU7 is not the best structure for these VS because it only discovers four docked positions with an $\text{RMSD} \leq 2.0 \text{ \AA}$. This indicates that many compounds that might have been used to treat COVID-19 have been classified as false negatives.
- The impact of water molecules in the binding site of SARS-CoV-2 M-pro during the docking is quite clear but using conserved waters might not improve the docking results.
- The influence of the protonation of key histidines in the binding site of SARS-CoV-2 M-pro is evident, but there is no overall “best” histidine protonation state. However, it guides the docking towards specific families of ligands.

Conclusions

En aquesta tesi s'han extret les següents conclusions:

- Hem recollit 1765 inhibidors de la M-pro del SARS-CoV-2 que podrien ser útils per validar un VS. Només s'inclouen compostos el potencial inhibidor dels quals s'ha establert experimentalment contra la M-pro del SARS-CoV-2.

A partir de l'anàlisi de 61 articles peer-reviewed que utilitzen una metodologia VS basada en docking de proteïna-ligand per predir compostos que actuaran com a inhibidors del SARS-CoV2 M-pro:

- La majoria dels protocols VS basats en docking de proteïna-ligand, que prediuen fàrmacs com a inhibidors de la M-pro del SARS-CoV-2 no s'han validat correctament.

A partir de l'anàlisi de 150 inhibidors de la M-pro:

- L'activitat dels inhibidors coneguts de la M-pro del SARS-CoV-2 no es correlaciona amb el seu docking score ni amb el ΔG calculat mitjançant el mètode MM-GBSA. La puntuació d'acoblament no es pot utilitzar per definir o predir l'activitat d'un determinat compost.
- El docking score ni el ΔG calculat mitjançant el mètode MM-GBSA no poden distingir els compostos actius dels inactius.
- Els resultats milloren quan els inhibidors covalents i no covalents es tracten per separat.

A partir de l'anàlisi del rendiment dels quatre programes de docking més utilitzats:

- Els programes de docking més utilitzats (Autodock Vina i Glide) poden reproduir, quan es tenen en compte almenys les 5 primeres poses, fins a un 50%-80% d'un conjunt de 47 inhibidors experimentalment coneguts de la M-pro del SARS-CoV-2 en un estudi de re-docking. Per a la 1a pose, els resultats tenen una precisió del 37-71% segons les condicions (water, dry, conserved). Les poses experimentals d'alguns lligands no es poden reproduir mitjançant re-docking.
- En els experiments de cross docking, Glide XP i SP són els millors en tots els casos, però Autodock Vina Hydrated funciona força bé si els compostos es poden hidratar, ja que arriba al 80% de rendiment quan s'eliminen els compostos que no es poden hidratar.
- 6LU7 no és la millor estructura per a aquests treballs perquè només descobreix quatre poses amb un $\text{RMSD} \leq 2,0 \text{ \AA}$. Això indica que molts compostos que podrien haver estat utilitzats per tractar la COVID-19 s'han classificat com a falsos negatius.
- L'impacte de les molècules d'aigua al binding site de la M-pro del SARS-CoV-2 durant el docking és força clar, però l'ús d'aigües conservades pot no millorar els resultats del docking.
- L'influència de la protonació de les histidines clau al lloc d'unió de SARS-CoV-2 M-pro és evident, però no hi ha un estat de protonació de la histidina "millor" general. Tanmateix, guia l'acoblament cap a famílies específiques de lligands.

Publication List

Annex

Annex

1. Gimeno, Aleix, Júlia Mestres-Truyol, María José Ojeda-Montes, **Guillem Macip**, Bryan Saldivar-Espinoza, Adrià Cereto-Massagué, Gerard Pujadas, and Santiago Garcia-Vallvé. 2020. 'Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition'. *International Journal of Molecular Sciences* 21(11):3793. doi: 10.3390/ijms21113793.
2. **Macip, Guillem**, Pol Garcia-Segura, Júlia Mestres-Truyol, Bryan Saldivar-Espinoza, María José Ojeda-Montes, Aleix Gimeno, Adrià Cereto-Massagué, Santiago Garcia-Vallvé, and Gerard Pujadas. 2021. 'Haste Makes Waste: A Critical Review of Docking-Based Virtual Screening in Drug Repurposing for SARS-CoV-2 Main Protease (M-pro) Inhibition'. *Medicinal Research Reviews* 42(2):744–69. doi: 10.1002/med.21862.
3. **Macip, Guillem**, Pol Garcia-segura, Júlia Mestres-truyol, Bryan Saldivar-espinoza, Gerard Pujadas, and Santiago Garcia-Vallvé. 2022. 'A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet?' *International Journal of Molecular Sciences* 23(1):259. doi: 10.3390/IJMS23010259/S1.
4. **Macip, Guillem**, Júlia Vilalta, Júlia Mestres-Truyol, Pol Garcia-Segura, Bryan Saldivar-Espinoza, Adrià Cereto-Massagué, Santiago Garcia-Vallvé, Gerard Pujadas. Are the protein-ligand docking programs good enough to predict experimental poses of non-covalently inhibitors bound to SARS-CoV-2 main protease?
In preparation
5. Saldivar-Espinoza, Bryan, **Guillem Macip**, Pol Garcia-Segura, Júlia Mestres-Truyol, Pere Puigbò, Adrià Cereto-Massagué, Gerard Pujadas, and Santiago Garcia-Vallve. 2022. 'Prediction of Recurrent Mutations in SARS-CoV-2 Using Artificial Neural Networks'. *International Journal of Molecular Sciences* 23(23):14683. doi: 10.3390/IJMS232314683/S1.
6. Saldivar-Espinoza, Bryan, **Guillem Macip**, Gerard Pujadas, and Santiago Garcia-Vallve. 2022. 'Could Nucleocapsid Be a Next-Generation COVID-19 Vaccine Candidate?' *International Journal of Infectious Diseases* 125:231–32. doi: 10.1016/j.ijid.2022.11.002.

ANNEX

Annex

ANNEX

245 articles analysed in the discussion.

Abdallah, H. M. <i>et al.</i> Repurposing of Some Natural Product Isolates as SARS-COV-2 Main Protease Inhibitors via In Vitro Cell Free and Cell-Based Antiviral Assessments and Molecular Modeling Approaches. <i>Pharmaceuticals</i> 14 , 213 (2021).
Abdelaal Ahmed Mahmoud M. Alkhatip, A. <i>et al.</i> Metal-bound methisazone; novel drugs targeting prophylaxis and treatment of sars-cov-2, a molecular docking study. <i>Int. J. Mol. Sci.</i> 22 , 1–13 (2021).
Abdelmohsen, U. R. <i>et al.</i> Natural coumarins as potential anti-SARS-CoV-2 agents supported by docking analysis. <i>RSC Adv.</i> 11 , 16970–16979 (2021).
Abdul-Hammed, M. <i>et al.</i> Target-based drug discovery, ADMET profiling and bioactivity studies of antibiotics as potential inhibitors of SARS-CoV-2 main protease (Mpro). <i>VirusDisease</i> 32 , 642–656 (2021).
Abdusalam, A. A. A. & Murugaiyah, V. Identification of Potential Inhibitors of 3CL Protease of SARS-CoV-2 From ZINC Database by Molecular Docking-Based Virtual Screening. <i>Front. Mol. Biosci.</i> 7 , 603037 (2020).
Abel, R. <i>et al.</i> Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. <i>Front. Chem.</i> 8 , 590263 (2020).
Acharya, A. <i>et al.</i> Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. <i>J. Chem. Inf. Model.</i> 60 , 5832–5852 (2020).
Adelusi, T. I. <i>et al.</i> Dietary polyphenols mitigate SARS-CoV-2 main protease (Mpro)–Molecular dynamics, molecular mechanics, and density functional theory investigations. <i>J. Mol. Struct.</i> 1250 , 131879 (2022).
Adem, Ş. <i>et al.</i> Caffeic acid derivatives (CAFDs) as inhibitors of SARS-CoV-2: CAFDs-based functional foods as a potential alternative approach to combat COVID-19. <i>Phytomedicine</i> 85 , 153310 (2021).
Aghaee, E., Ghodrati, M. & Ghasemi, J. B. In silico exploration of novel protease inhibitors against coronavirus 2019 (COVID-19). <i>Informatics Med. Unlocked</i> 23 , 100516 (2021).
Ahmad, S., Waheed, Y., Ismail, S., Najmi, M. H. & Ansari, J. K. Rational design of potent anti-COVID-19 main protease drugs: An extensive multi-spectrum in silico approach. <i>J. Mol. Liq.</i> 330 , 115636 (2021).
Ahmad, S. <i>et al.</i> Fragment-based in silico design of SARS-CoV-2 main protease inhibitors. <i>Chem. Biol. Drug Des.</i> 98 , 604–619 (2021).
Ahmed, M. Z. <i>et al.</i> Aminoglycosides as potential inhibitors of SARS-CoV-2 main protease: an in silico drug repurposing study on FDA-approved antiviral and anti-infection agents. <i>J. Infect. Public Health</i> 14 , 611–619 (2021).
Alanazi, K. M., Farah, M. A. & Hor, Y.-Y. Multi-Targeted Approaches and Drug Repurposing Reveal Possible SARS-CoV-2 Inhibitors. <i>Vaccines</i> 10 , 24 (2021).
Alharthi, N. M. <i>et al.</i> Assessment of pharmacists' perception toward continuing education. <i>J. Adv. Pharm. Technol. Res.</i> 12 , 368–372 (2021).

Annex

Almeida, J. S. F. D. <i>et al.</i> Searching for potential drugs against SARS-CoV-2 through virtual screening on several molecular targets. <i>J. Biomol. Struct. Dyn.</i> 40 , 5229–5242 (2022).
Alnajjar, R., Mostafa, A., Kandeil, A. & Al-Karmalawy, A. A. Molecular docking, molecular dynamics, and in vitro studies reveal the potential of angiotensin II receptor blockers to inhibit the COVID-19 main protease. <i>Heliyon</i> 6 , e05641 (2020).
Al-Rabia, M. W. <i>et al.</i> Repurposing of sitagliptin-melittin optimized nanoformula against sars-cov-2: Antiviral screening and molecular docking studies. <i>Pharmaceutics</i> 13 , 1–21 (2021).
Alves, V. M. <i>et al.</i> QSAR Modeling of SARS-CoV M pro Inhibitors Identifies Sufugolix, Cenicriviroc, Proglumetacin, and other Drugs as Candidates for Repurposing against SARS-CoV-2. <i>Mol. Inform.</i> 40 , 2000113 (2021).
Anowar Hosen, M. <i>et al.</i> Synthesis, antimicrobial, molecular docking and molecular dynamics studies of lauroyl thymidine analogs against SARS-CoV-2: POM study and identification of the pharmacophore sites. <i>Bioorg. Chem.</i> 125 , 105850 (2022).
Augustin, T. L., Hajbabaie, R., Harper, M. T. & Rahman, T. Novel Small-Molecule Scaffolds as Candidates against the SARS Coronavirus 2 Main Protease: A Fragment-Guided in Silico Approach. <i>Molecules</i> 25 , 5501 (2020).
Avti, P. <i>et al.</i> Computational basis of SARS-CoV 2 main protease inhibition: an insight from molecular dynamics simulation based findings. <i>J. Biomol. Struct. Dyn.</i> 40 , 8894–8904 (2022).
Azad, I. <i>et al.</i> Identification of Severe Acute Respiratory Syndrome Coronavirus-2 inhibitors through in silico structure-based virtual screening and molecular interaction studies. <i>J. Mol. Recognit.</i> 34 , e2918 (2021).
Azeem, M., Mustafa, G. & Mahrosh, H. S. Virtual screening of phytochemicals by targeting multiple proteins of severe acute respiratory syndrome coronavirus 2: Molecular docking and molecular dynamics simulation studies. <i>Int. J. Immunopathol. Pharmacol.</i> 36 , 039463202211427 (2022).
Bahadur Gurung, A., Ajmal Ali, M., Al-Hemaid, F., El-Zaidy, M. & Lee, J. In silico analyses of major active constituents of fingerroot (<i>Boesenbergia rotunda</i>) unveils inhibitory activities against SARS-CoV-2 main protease enzyme. <i>Saudi J. Biol. Sci.</i> 29 , 65–74 (2022).
Bajrai, L. H., Faizo, A. A., Alkhaldy, A. A., Dwivedi, V. D. & Azhar, E. I. Repositioning of anti-dengue compounds against SARS-CoV-2 as viral polyprotein processing inhibitor. <i>PLoS One</i> 17 , e0277328 (2022).
Balogun, T. A. <i>et al.</i> Discovery of putative inhibitors against main drivers of SARS-CoV-2 infection: Insight from quantum mechanical evaluation and molecular modeling. <i>Front. Chem.</i> 10 , 964446 (2022).
Banerjee, S. <i>et al.</i> Drug Repurposing to Identify Nilotinib as a Potential SARS-CoV-2 Main Protease Inhibitor: Insights from a Computational and in Vitro Study. <i>J. Chem. Inf. Model.</i> 61 , 5469–5483 (2021).
Batool, F. <i>et al.</i> Synthetic flavonoids as potential antiviral agents against SARS-CoV-2 main protease. <i>J. Biomol. Struct. Dyn.</i> 40 , 3777–3788 (2022).

Beck, B. R., Shin, B., Choi, Y., Park, S. & Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. <i>Comput. Struct. Biotechnol. J.</i> 18 , 784–790 (2020).
Bello, M., Martínez-Muñoz, A. & Balbuena-Rebolledo, I. Identification of saquinavir as a potent inhibitor of dimeric SARS-CoV2 main protease through MM/GBSA. <i>J. Mol. Model.</i> 26 , 340 (2020).
Bepari, A. K. & Reza, H. M. Identification of a novel inhibitor of SARS-CoV-2 3CL-PRO through virtual screening and molecular dynamics simulation. <i>PeerJ</i> 9 , e11261 (2021).
Bernardi, M., Ghaani, M. R. & Bayazeid, O. Phenylethanoid glycosides as a possible COVID-19 protease inhibitor: a virtual screening approach. <i>J. Mol. Model.</i> 27 , 341 (2021).
Bharadwaj, S. <i>et al.</i> Exploration of natural compounds with anti-SARS-CoV-2 activity via inhibition of SARS-CoV-2 Mpro. <i>Brief. Bioinform.</i> 22 , 1361–1377 (2021).
Bhowmick, S. <i>et al.</i> Structure-based identification of SARS-CoV-2 main protease inhibitors from anti-viral specific chemical libraries: an exhaustive computational screening approach. <i>Mol. Divers.</i> 25 , 1979–1997 (2021).
Bibi, S. <i>et al.</i> Virtual screening and molecular dynamics simulation analysis of Forsythoside A as a plant-derived inhibitor of SARS-CoV-2 3CLpro. <i>Saudi Pharm. J.</i> 30 , 979–1002 (2022).
Borkotoky, S., Banerjee, M., Modi, G. P. & Dubey, V. K. Identification of high affinity and low molecular alternatives of boceprevir against SARS-CoV-2 main protease: A virtual screening approach. <i>Chem. Phys. Lett.</i> 770 , 138446 (2021).
Bray, S. <i>et al.</i> Galaxy workflows for fragment-based virtual screening: a case study on the SARS-CoV-2 main protease. <i>J. Cheminform.</i> 14 , 22 (2022).
Bu, Q. <i>et al.</i> Mililatensols A–C, New Records of Sarsolenane and Capnosane Diterpenes from Soft Coral Sarcophyton mililatensis. <i>Mar. Drugs</i> 20 , 566 (2022).
Bung, N., Krishnan, S. R., Bulusu, G. & Roy, A. De novo design of new chemical entities for SARS-CoV-2 using artificial intelligence. <i>Future Med. Chem.</i> 13 , 575–585 (2021).
Burch, M. L., Getachew, R., Osman, N., Febbraio, M. A. & Little, P. J. Thrombin-mediated proteoglycan synthesis utilizes both protein-tyrosine kinase and serine/threonine kinase receptor transactivation in vascular smooth muscle cells. <i>J. Biol. Chem.</i> 288 , 7410–7419 (2013).
Cayona, R. & Creencia, E. Phytochemicals of Euphorbia hirta L. and Their Inhibitory Potential Against SARS-CoV-2 Main Protease. <i>Front. Mol. Biosci.</i> 8 , 801401 (2021).
Chen, Y. W., Yiu, C.-P. B. & Wong, K.-Y. Virtual screening on the web for drug repurposing: a primer. <i>J. Biol. Methods</i> 8 , e148 (2021).
Cheng, A. J. T. <i>et al.</i> In silico evaluation of Philippine Natural Products against SARS-CoV-2 Main Protease. <i>J. Mol. Model.</i> 28 , 345 (2022).
Choe, J., Har Yong, P. & Xiang Ng, Z. The Efficacy of Traditional Medicinal Plants in Modulating the Main Protease of SARS-CoV-2 and Cytokine Storm. <i>Chem. Biodivers.</i> 19 , e202200655 (2022).

Annex

Chowdhury, K. H. <i>et al.</i> Drug repurposing approach against novel coronavirus disease (COVID-19) through virtual screening targeting SARS-CoV-2 main protease. <i>Biology (Basel)</i> . 10 , 1–14 (2021).
Clyde, A. <i>et al.</i> High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor. <i>J. Chem. Inf. Model.</i> 62 , 116–128 (2022).
Cooper, K. <i>et al.</i> Novel Development of Predictive Feature Fingerprints to Identify Chemistry-Based Features for the Effective Drug Design of SARS-CoV-2 Target Antagonists and Inhibitors Using Machine Learning. <i>ACS omega</i> 6 , 4857–4877 (2021).
Costa, A. S., Martins, J. P. A. & de Melo, E. B. SMILES-based 2D-QSAR and similarity search for identification of potential new scaffolds for development of SARS-CoV-2 MPRO inhibitors. <i>Struct. Chem.</i> 33 , 1691–1706 (2022).
Cristina Soares Rodrigues, G. <i>et al.</i> Send Orders for Reprints to reprints@benthamscience.net Ligand and Structure-based Virtual Screening of Lamiaceae Diterpenes with Potential Activity against a Novel Coronavirus (2019-nCoV). <i>Curr. Top. Med. Chem.</i> 20 , 2126–2145 (2020).
Daoud, I., Mesli, F., Melkemi, N., Ghalem, S. & Salah, T. Discovery of potential SARS-CoV 3CL protease inhibitors from approved antiviral drugs using: virtual screening, molecular docking, pharmacophore mapping evaluation and dynamics simulation. <i>J. Biomol. Struct. Dyn.</i> 40 , 12574–12591 (2022).
Debnath, P., Bhaumik, S., Sen, D., Muttineni, R. K. & Debnath, S. Identification of SARS-CoV-2 Main Protease Inhibitors Using Structure Based Virtual Screening and Molecular Dynamics Simulation of DrugBank Database. <i>ChemistrySelect</i> 6 , 4991–5013 (2021).
Deeks, H. M., Walters, R. K., Barnoud, J., Glowacki, D. R. & Mulholland, A. J. Interactive Molecular Dynamics in Virtual Reality Is an Effective Tool for Flexible Substrate and Inhibitor Docking to the SARS-CoV-2 Main Protease. <i>J. Chem. Inf. Model.</i> 60 , 5803–5814 (2020).
Djokovic, N. <i>et al.</i> An Integrative in silico Drug Repurposing Approach for Identification of Potential Inhibitors of SARS-CoV-2 Main Protease. <i>Mol. Inform.</i> 40 , 2000187 (2021).
Dong, J. <i>et al.</i> Ligand-based discovery of coronavirus main protease inhibitors using MACAW molecular embeddings. <i>J. Enzyme Inhib. Med. Chem.</i> 38 , 24–35 (2023).
Duran, N. <i>et al.</i> New chalcone derivatives as effective against SARS-CoV-2 agent. <i>Int. J. Clin. Pract.</i> 75 , e14846 (2021).
Durdagi, S. <i>et al.</i> Screening of Clinically Approved and Investigation Drugs as Potential Inhibitors of SARS-CoV-2: A Combined in silico and in vitro Study. <i>Mol. Inform.</i> 41 , 2100062 (2022).
El Hassab, M. A. <i>et al.</i> Computational Prediction of the Potential Target of SARS-CoV-2 Inhibitor Plitidepsin via Molecular Docking, Dynamic Simulations and MM-PBSA Calculations. <i>Chem. Biodivers.</i> 19 , e202100719 (2022).
El-Ashrey, M. K., Bakr, R. O., Fayed, M. A. A., Refaey, R. H. & Nissan, Y. M. Pharmacophore based virtual screening for natural product database revealed possible inhibitors for SARS-COV-2 main protease. <i>Virology</i> 570 , 18–28 (2022).

El-Demerdash, A. <i>et al.</i> Comprehensive virtual screening of the antiviral potentialities of marine polycyclic guanidine alkaloids against sars-cov-2 (Covid-19). <i>Biomolecules</i> 11 , 1–25 (2021).
Elgohary, A. M. <i>et al.</i> Investigating the structure-activity relationship of marine polycyclic batzelladine alkaloids as promising inhibitors for SARS-CoV-2 main protease (Mpro). <i>Comput. Biol. Med.</i> 147 , 105738 (2022).
Elkadeed, E. B. <i>et al.</i> A Multistage In Silico Study of Natural Potential Inhibitors Targeting SARS-CoV-2 Main Protease. <i>Int. J. Mol. Sci.</i> 23 , 8407 (2022).
Emon, N. U. <i>et al.</i> Virtual screenings of the bioactive constituents of tea, prickly chaff, catechu, lemon, black pepper, and synthetic compounds with the main protease (Mpro) and human angiotensin-converting enzyme 2 (ACE 2) of SARS-CoV-2. <i>Futur. J. Pharm. Sci.</i> 7 , 121 (2021).
Erlina, L. <i>et al.</i> Virtual screening of Indonesian herbal compounds as COVID-19 supportive therapy: machine learning and pharmacophore modeling approaches. <i>BMC Complement. Med. Ther.</i> 22 , 207 (2022).
Eskandari, V. Repurposing the natural compounds as potential therapeutic agents for COVID-19 based on the molecular docking study of the main protease and the receptor-binding domain of spike protein. <i>J. Mol. Model.</i> 28 , 153 (2022).
Fakhar, Z., Khan, S., AlOmar, S. Y., Alkhuriji, A. & Ahmad, A. ABBV-744 as a potential inhibitor of SARS-CoV-2 main protease enzyme against COVID-19. <i>Sci. Rep.</i> 11 , 234 (2021).
Farhat, N. & Khan, A. U. Repurposing drug molecule against SARS-Cov-2 (COVID-19) through molecular docking and dynamics: a quick approach to pick FDA-approved drugs. <i>J. Mol. Model.</i> 27 , 312 (2021).
Federico, L. B. <i>et al.</i> Identification of known drugs as potential SARS-CoV-2 Mpro inhibitors using ligand- And structure-based virtual screening. <i>Future Med. Chem.</i> 13 , 1353–1366 (2021).
Franco, L. S., Maia, R. C. & Barreiro, E. J. Identification of LASSBio-1945 as an inhibitor of SARS-CoV-2 main protease (MPRO) throughin silicoscreening supported by molecular docking and a fragment-based pharmacophore model. <i>RSC Med. Chem.</i> 12 , 110–119 (2021).
Freidel, M. R. & Armen, R. S. Mapping major SARS-CoV-2 drug targets and assessment of druggability using computational fragment screening: Identification of an allosteric small-molecule binding site on the Nsp13 helicase. <i>PLoS One</i> 16 , e0246181 (2021).
Gaudêncio, S. P. & Pereira, F. A Computer-Aided Drug Design Approach to Predict Marine Drug-Like Leads for SARS-CoV-2 Main Protease Inhibition. <i>Mar. Drugs</i> 18 , 633 (2020).
Gentile, F. <i>et al.</i> Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus Deep Docking of 40 billion small molecules. <i>Chem. Sci.</i> 12 , 15960–15974 (2021).
Ghahremanpour, M. M. <i>et al.</i> Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. <i>ACS Med. Chem. Lett.</i> 11 , 2526–2533 (2020).
Ghosh, A., Chakraborty, M., Chandra, A. & Alam, M. P. Structure-activity relationship (SAR) and molecular dynamics study of withaferin-A fragment derivatives as potential

Annex

therapeutic lead against main protease (Mpro) of SARS-CoV-2. <i>J. Mol. Model.</i> 27 , 97 (2021).
Glaab, E., Manoharan, G. B. & Abankwa, D. Pharmacophore Model for SARS-CoV-2 3CLpro Small-Molecule Inhibitors and in Vitro Experimental Validation of Computationally Screened Inhibitors. <i>J. Chem. Inf. Model.</i> 61 , 4082–4096 (2021).
Gogoi, N. <i>et al.</i> Computational guided identification of a citrus flavonoid as potential inhibitor of SARS-CoV-2 main protease. <i>Mol. Divers.</i> 25 , 1745–1759 (2021).
Gorgulla, C. <i>et al.</i> A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening. <i>iScience</i> 24 , 102021 (2021).
Gossen, J. <i>et al.</i> A Blueprint for High Affinity SARS-CoV-2 Mpro Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics. <i>ACS Pharmacol. Transl. Sci.</i> 4 , 1079–1095 (2021).
Goyzueta-Mamani, L. D., Barazorda-Ccahuana, H. L., Mena-Ulecia, K. & Chávez-Fumagalli, M. A. Antiviral Activity of Metabolites from Peruvian Plants against SARS-CoV-2: An In Silico Approach. <i>Molecules</i> 26 , 3882 (2021).
Guedes, I. A. <i>et al.</i> Drug design and repurposing with DockThor-VS web server focusing on SARS-CoV-2 therapeutic targets and their non-synonym variants. <i>Sci. Rep.</i> 11 , 5543 (2021).
Guo, S. <i>et al.</i> Discovery of novel inhibitors against main protease (Mpro) of SARS-CoV-2 via virtual screening and biochemical evaluation. <i>Bioorg. Chem.</i> 110 , 104767 (2021).
Gupta, A. & Zhou, H. X. Profiling SARS-CoV-2 Main Protease (MPRO) Binding to Repurposed Drugs Using Molecular Dynamics Simulations in Classical and Neural Network-Trained Force Fields. <i>ACS Comb. Sci.</i> 22 , 826–832 (2020).
Gupta, A. <i>et al.</i> Structure-Based Virtual Screening and Biochemical Validation to Discover a Potential Inhibitor of the SARS-CoV-2 Main Protease. <i>ACS Omega</i> 5 , 33151–33161 (2020).
Gupta, A. <i>et al.</i> Exploration of Novel Lichen Compounds as Inhibitors of SARS-CoV-2 Mpro: Ligand-Based Design, Molecular Dynamics, and ADMET Analyses. <i>Appl. Biochem. Biotechnol.</i> 194 , 6386–6406 (2022).
Gupta, Y. <i>et al.</i> Antiviral evaluation of hydroxyethylamine analogs: Inhibitors of SARS-CoV-2 main protease (3CLpro), a virtual screening and simulation approach. <i>Bioorg. Med. Chem.</i> 47 , 116393 (2021).
Gupta, Y. <i>et al.</i> Bisindolylmaleimide IX: A novel anti-SARS-CoV2 agent targeting viral main protease 3CLpro demonstrated by virtual screening pipeline and in-vitro validation assays. <i>Methods</i> 195 , 57–71 (2021).
Gyebi, G. A. <i>et al.</i> Structure-based virtual screening suggests inhibitors of 3-Chymotrypsin-Like Protease of SARS-CoV-2 from <i>Vernonia amygdalina</i> and <i>Ocimum gratissimum</i> . <i>Comput. Biol. Med.</i> 136 , 104671 (2021).
Hagar, M., Ahmed, H. A., Aljohani, G. & Alhaddad, O. A. Investigation of Some Antiviral N-Heterocycles as COVID 19 Drug: Molecular Docking and DFT Calculations. <i>Int. J. Mol. Sci.</i> 21 , 3922 (2020).

Haider, Z. <i>et al.</i> In-silico pharmacophoric and molecular docking-based drug discovery against the Main Protease (Mpro) of SARS-CoV-2, a causative agent COVID-19. <i>Pak. J. Pharm. Sci.</i> 33 , 2697–2705 (2020).
Hakami, A. R., Bakheit, A. H., Almehezia, A. A. & Ghazwani, M. Y. Selection of SARS-CoV-2 main protease inhibitor using structure-based virtual screening. <i>Future Med. Chem.</i> 14 , 61–79 (2022).
Halimi, M. & Bararpour, P. Natural inhibitors of SARS-CoV-2 main protease: structure based pharmacophore modeling, molecular docking and molecular dynamic simulation studies. <i>J. Mol. Model.</i> 28 , 279 (2022).
Hamdy, R. <i>et al.</i> Iterated Virtual Screening-Assisted Antiviral and Enzyme Inhibition Assays Reveal the Discovery of Novel Promising Anti-SARS-CoV-2 with Dual Activity. <i>Int. J. Mol. Sci.</i> 22 , 9057 (2021).
Hariyono, P., Dwistuti, R., Yusuf, M., Salin, N. H. & Hariono, M. 2-Phenoxyacetamide derivatives as SARS-CoV-2 main protease inhibitor: In silico studies. <i>Results Chem.</i> 4 , 100263 (2022).
Hartini, Y. <i>et al.</i> Biflavonoid as potential 3-chymotrypsin-like protease (3CLpro) inhibitor of SARS-Coronavirus. <i>Results Chem.</i> 3 , 100087 (2021).
Hassam, M. <i>et al.</i> Identification of potent compounds against SARs-CoV-2: An in-silico based drug searching against Mpro. <i>Comput. Biol. Med.</i> 151 , 106284 (2022).
Hayek-Orduz, Y. <i>et al.</i> Novel covalent and non-covalent complex-based pharmacophore models of SARS-CoV-2 main protease (Mpro) elucidated by microsecond MD simulations. <i>Sci. Rep.</i> 12 , 14030 (2022).
Hengphasatporn, K. <i>et al.</i> Halogenated Baicalein as a Promising Antiviral Agent toward SARS-CoV-2 Main Protease. <i>J. Chem. Inf. Model.</i> 62 , 1498–1509 (2022).
Hernández González, J. E., Eberle, R. J., Willbold, D. & Coronado, M. A. A Computer-Aided Approach for the Discovery of D-Peptides as Inhibitors of SARS-CoV-2 Main Protease. <i>Front. Mol. Biosci.</i> 8 , 816166 (2021).
Holanda, V. N. <i>et al.</i> Identification of 1,2,3-triazole-phthalimide derivatives as potential drugs against COVID-19: a virtual screening, docking and molecular dynamic study. <i>J. Biomol. Struct. Dyn.</i> 40 , 5462–5480 (2022).
Hosseini, M., Chen, W., Xiao, D. & Wang, C. Computational molecular docking and virtual screening revealed promising SARS-CoV-2 drugs. <i>Precis. Clin. Med.</i> 4 , 1–16 (2021).
Ibezim, A. <i>et al.</i> Structure-based virtual screening and molecular dynamics simulation studies to discover new SARS-CoV-2 main protease inhibitors. <i>Sci. African</i> 14 , e00970 (2021).
Ibrahim, M. A. A. <i>et al.</i> In Silico Mining of Terpenes from Red-Sea Invertebrates for SARS-CoV-2 Main Protease (Mpro) Inhibitors. <i>Molecules</i> 26 , 2082 (2021).
Ibrahim, M. A. A. <i>et al.</i> In Silico Evaluation of Prospective Anti-COVID-19 Drug Candidates as Potential SARS-CoV-2 Main Protease Inhibitors. <i>Protein J.</i> 40 , 296–309 (2021).
Issa, S. S., Sokornova, S. V., Zhidkin, R. R. & Matveeva, T. V. The Main Protease of SARS-CoV-2 as a Target for Phytochemicals against Coronavirus. <i>Plants</i> 11 , 1862 (2022).

Annex

Jade, D. <i>et al.</i> Virtual high throughput screening: Potential inhibitors for SARS-CoV-2 PLPRO and 3CLPRO proteases. <i>Eur. J. Pharmacol.</i> 901 , 174082 (2021).
Jairajpuri, D. S. <i>et al.</i> Identification of natural compounds as potent inhibitors of SARS-CoV-2 main protease using combined docking and molecular dynamics simulations. <i>Saudi J. Biol. Sci.</i> 28 , 2423–2431 (2021).
Jamir, E. <i>et al.</i> Applying polypharmacology approach for drug repurposing for SARS-CoV2. <i>J. Chem. Sci.</i> 134 , 57 (2022).
Jawarkar, R. D. <i>et al.</i> QSAR based virtual screening derived identification of a novel hit as a SARS CoV-229E 3CLpro Inhibitor: GA-MLR QSAR modeling supported by molecular Docking, molecular dynamics simulation and MMGBSA calculation approaches. <i>Arab. J. Chem.</i> 15 , 103499 (2022).
Jena, A. B. <i>et al.</i> A Computational Insight on the Inhibitory Potential of 8-Hydroxydihydrosanguinarine (8-HDS), a Pyridone Containing Analog of Sanguinarine, against SARS CoV2. <i>Chem. Biodivers.</i> 19 , e202200266 (2022).
Jiménez-Avalos, G. <i>et al.</i> Comprehensive virtual screening of 4.8 k flavonoids reveals novel insights into allosteric inhibition of SARS-CoV-2 MPRO. <i>Sci. Rep.</i> 11 , 15452 (2021).
Johnson, T. O. <i>et al.</i> A Computational Approach to Elucidate the Interactions of Chemicals From <i>Artemisia annua</i> Targeted Toward SARS-CoV-2 Main Protease Inhibition for COVID-19 Treatment. <i>Front. Med.</i> 9 , 907583 (2022).
Jukič, M., Janežič, D. & Bren, U. Ensemble Docking Coupled to Linear Interaction Energy Calculations for Identification of Coronavirus Main Protease (3CLpro) Non-Covalent Small-Molecule Inhibitors. <i>Molecules</i> 25 , 5808 (2020).
Jukič, M., Kores, K., Janežič, D. & Bren, U. Repurposing of Drugs for SARS-CoV-2 Using Inverse Docking Fingerprints. <i>Front. Chem.</i> 9 , 757826 (2021).
Kaliyamurthi, S. <i>et al.</i> Structure-Based Virtual Screening Reveals Ibrutinib and Zanubrutinib as Potential Repurposed Drugs against COVID-19. <i>Int. J. Mol. Sci.</i> 22 , 7071 (2021).
Kang, N., Heo, S.-Y., Cha, S.-H., Ahn, G. & Heo, S.-J. In Silico Virtual Screening of Marine Aldehyde Derivatives from Seaweeds against SARS-CoV-2. <i>Mar. Drugs</i> 20 , 399 (2022).
Kar, B., Dehury, B., Singh, M. K., Pati, S. & Bhattacharya, D. Identification of phytocompounds as newer antiviral drugs against COVID-19 through molecular docking and simulation based study. <i>J. Mol. Graph. Model.</i> 114 , 108192 (2022).
Khalidan, A. <i>et al.</i> Prediction of potential inhibitors of SARS-CoV-2 using 3D-QSAR, molecular docking modeling and ADMET properties. <i>Heliyon</i> 7 , e06603 (2021).
Khan, J. <i>et al.</i> Identification of potential phytochemicals from Citrus Limon against main protease of SARS-CoV-2: molecular docking, molecular dynamic simulations and quantum computations. <i>J. Biomol. Struct. Dyn.</i> 40 , 10741–10752 (2022).
Khan, M. T. <i>et al.</i> Marine natural compounds as potents inhibitors against the main protease of SARS-CoV-2—a molecular dynamic study. <i>J. Biomol. Struct. Dyn.</i> 39 , 3627–3637 (2021).
Khan, S. <i>et al.</i> Structure-based identification of potential SARS-CoV-2 main protease inhibitors. <i>J. Biomol. Struct. Dyn.</i> 40 , 3595–3608 (2022).

Kralj, S., Jukić, M. & Bren, U. Commercial SARS-CoV-2 Targeted, Protease Inhibitor Focused and Protein–Protein Interaction Inhibitor Focused Molecular Libraries for Virtual Screening and Drug Design. <i>Int. J. Mol. Sci.</i> 23 , 393 (2021).
Kumar, A. <i>et al.</i> Inhibition Potencies of Phytochemicals Derived from Sesame Against SARS-CoV-2 Main Protease: A Molecular Docking and Simulation Study. <i>Front. Chem.</i> 9 , 744376 (2021).
Kumar, K. A. <i>et al.</i> Multifunctional inhibitors of SARS-CoV-2 by MM/PBSA, essential dynamics, and molecular dynamic investigations. <i>J. Mol. Graph. Model.</i> 107 , 107969 (2021).
Kumar, R., Kumar, V. & Lee, K. W. A computational drug repurposing approach in identifying the cephalosporin antibiotic and anti-hepatitis C drug derivatives for COVID-19 treatment. <i>Comput. Biol. Med.</i> 130 , 104186 (2021).
Kumar, S. <i>et al.</i> A multi-targeted approach to identify potential flavonoids against three targets in the SARS-CoV-2 life cycle. <i>Comput. Biol. Med.</i> 142 , 105231 (2022).
Kumar, V., Kancharla, S. & Jena, M. K. In silico virtual screening-based study of nutraceuticals predicts the therapeutic potentials of folic acid and its derivatives against COVID-19. <i>VirusDisease</i> 32 , 29–37 (2021).
Kumari, M. & Subbarao, N. Deep learning model for virtual screening of novel 3C-like protease enzyme inhibitors against SARS coronavirus diseases. <i>Comput. Biol. Med.</i> 132 , 104317 (2021).
Kushari, S. <i>et al.</i> An integrated computational approach towards the screening of active plant metabolites as potential inhibitors of SARS-CoV-2: an overview. <i>Struct. Chem.</i> 1–32 (2022).
Kwofie, S. K. <i>et al.</i> Cheminformatics-Based Identification of Potential Novel Anti-SARS-CoV-2 Natural Compounds of African Origin. <i>Molecules</i> 26 , 406 (2021).
Lee, J.-Y. <i>et al.</i> Identification of non-covalent 3C-like protease inhibitors against severe acute respiratory syndrome coronavirus-2 via virtual screening of a Korean compound library. <i>Bioorg. Med. Chem. Lett.</i> 42 , 128067 (2021).
Li, S. G. <i>et al.</i> An Enhanced Hybrid Screening Approach to Identify Potent Inhibitors for the SARS-CoV-2 Main Protease From the NCI Compound Library. <i>Front. Chem.</i> 10 , 816576 (2022).
Liang, H., Zhao, L., Gong, X., Hu, M. & Wang, H. Virtual screening FDA approved drugs against multiple targets of SARS-CoV-2. <i>Clin. Transl. Sci.</i> 14 , 1123–1132 (2021).
Lipiński, P. F. J., Zaborniak, J., Garnuszek, P. & Szurmak, P. Virtual screening for small molecular non-covalent binders of the SARS-CoV-2 main protease. <i>Arch. Med. Sci.</i> 17 , 838–842 (2021).
Liu, J. <i>et al.</i> Discovery of genistein derivatives as potential SARS-CoV-2 main protease inhibitors by virtual screening, molecular dynamics simulations and ADMET analysis. <i>Front. Pharmacol.</i> 13 , 961154 (2022).
Liu, W. S. <i>et al.</i> Screening potential FDA-approved inhibitors of the SARS-CoV-2 major protease 3CLpro through high-throughput virtual screening and molecular dynamics simulation. <i>Aging (Albany. NY)</i> . 13 , 6258–6272 (2021).

Annex

Luttens, A. <i>et al.</i> Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. <i>J. Am. Chem. Soc.</i> 144 , 2905–2920 (2022).
Maghsoudi, S. <i>et al.</i> A review on computer-aided chemogenomics and drug repositioning for rational COVID-19 drug discovery. <i>Chem. Biol. Drug Des.</i> 100 , 699–721 (2022).
Mahgoub, R. E. <i>et al.</i> The Discovery of Small Allosteric and Active Site Inhibitors of the SARS-CoV-2 Main Protease via Structure-Based Virtual Screening and Biological Evaluation. <i>Molecules</i> 27 , 6710 (2022).
Mahmud, S. <i>et al.</i> Virtual screening and molecular dynamics simulation study of plant-derived compounds to identify potential inhibitors of main protease from SARS-CoV-2. <i>Brief. Bioinform.</i> 22 , 1402–1414 (2021).
Manelfi, C. <i>et al.</i> Combining Different Docking Engines and Consensus Strategies to Design and Validate Optimized Virtual Screening Protocols for the SARS-CoV-2 3CL Protease. <i>Molecules</i> 26 , 797 (2021).
Manish, M., Mishra, S., Anand, A. & Subbarao, N. Computational molecular interaction between SARS-CoV-2 main protease and theaflavin digallate using free energy perturbation and molecular dynamics. <i>Comput. Biol. Med.</i> 150 , 106125 (2022).
Masand, V. H., Sk, M. F., Kar, P., Rastija, V. & Zaki, M. E. A. Identification of Food Compounds as inhibitors of SARS-CoV-2 main protease using molecular docking and molecular dynamics simulations. <i>Chemom. Intell. Lab. Syst.</i> 217 , 104394 (2021).
Mehmood, A., Nawab, S., Wang, Y., Chandra Kaushik, A. & Wei, D.-Q. Discovering potent inhibitors against the Mpro of the SARS-CoV-2. A medicinal chemistry approach. <i>Comput. Biol. Med.</i> 143 , 105235 (2022).
Mekni, N., Coronello, C., Langer, T., Rosa, M. De & Perricone, U. Support Vector Machine as a Supervised Learning for the Prioritization of Novel Potential SARS-CoV-2 Main Protease Inhibitors. <i>Int. J. Mol. Sci.</i> 22 , 7714 (2021).
Mercorelli, B. <i>et al.</i> Discovery of novel SARS-CoV-2 inhibitors targeting the main protease Mpro by virtual screenings and hit optimization. <i>Antiviral Res.</i> 204 , 105350 (2022).
Miandad, K. <i>et al.</i> Virtual Screening of Artemisia annua Phytochemicals as Potential Inhibitors of SARS-CoV-2 Main Protease Enzyme. <i>Molecules</i> 27 , 8103 (2022).
Mishra, A., Mulpuru, V. & Mishra, N. Identification of SARS-CoV-2 inhibitors through phylogenetics and drug repurposing. <i>Struct. Chem.</i> 33 , 1789–1797 (2022).
Mishra, C. B. <i>et al.</i> Identifying the natural polyphenol catechin as a multi-targeted agent against SARS-CoV-2 for the plausible therapy of COVID-19: An integrated computational approach. <i>Brief. Bioinform.</i> 22 , 1346–1360 (2021).
Mishra, S. S. <i>et al.</i> Computational investigation of potential inhibitors of novel coronavirus 2019 through structure-based virtual screening, molecular dynamics and density functional theory studies. <i>J. Biomol. Struct. Dyn.</i> 39 , 4449–4461 (2021).
Mithilesh, S., Raghunandan, D. & Suresh, P. K. In-Silico Identification of Natural Compounds from Traditional Medicine as Potential Drug Leads against SARS-CoV-2 Through Virtual Screening. <i>Proc. Natl. Acad. Sci. India Sect. B - Biol. Sci.</i> 92 , 81–87 (2022).

Mohammed, A. O., Abo-Idrees, M. I., Makki, A. A., Ibraheem, W. & Alzain, A. A. Drug repurposing against main protease and RNA-dependent RNA polymerase of SARS-CoV-2 using molecular docking, MM-GBSA calculations and molecular dynamics. <i>Struct. Chem.</i> 33 , 1553–1567 (2022).
Mohan, A. <i>et al.</i> Structure-based virtual screening, in silico docking, ADME properties prediction and molecular dynamics studies for the identification of potential inhibitors against SARS-CoV-2 Mpro. <i>Mol. Divers.</i> 26 , 1645–1661 (2022).
Mohan, S., Dharani, J., Natarajan, R. & Nagarajan, A. Molecular docking and identification of G-protein-coupled receptor 120 (GPR120) agonists as SARS COVID-19 MPro inhibitors. <i>J. Genet. Eng. Biotechnol.</i> 20 , 108 (2022).
Mohseni, M. <i>et al.</i> Indole alkaloids as potential candidates against COVID-19: an in silico study. <i>J. Mol. Model.</i> 28 , 144 (2022).
Morena, F., Argentati, C., Tortorella, I., Emiliani, C. & Martino, S. De novo ssRNA Aptamers against the SARS-CoV-2 Main Protease: In Silico Design and Molecular Dynamics Simulation. <i>Int. J. Mol. Sci.</i> 22 , 6874 (2021).
Muhammad, I. <i>et al.</i> Screening of potent phytochemical inhibitors against SARS-CoV-2 protease and its two Asian mutants. <i>Comput. Biol. Med.</i> 133 , 104362 (2021).
Muhtar, E., Wang, M. & Zhu, H. In silico discovery of SARS-CoV-2 main protease inhibitors from the carboline and quinoline database. <i>Future Virol.</i> 16 , 507–518 (2021).
Muteeb, G., Alshoaibi, A., Aatif, M., Rehman, M. T. & Qayyum, M. Z. Screening marine algae metabolites as high-affinity inhibitors of SARS-CoV-2 main protease (3CLpro): an in silico analysis to identify novel drug candidates to combat COVID-19 pandemic. <i>Appl. Biol. Chem.</i> 63 , 79 (2020).
Naasani, I. COMPARE Analysis, a Bioinformatic Approach to Accelerate Drug Repurposing against Covid-19 and Other Emerging Epidemics. <i>SLAS Discov.</i> 26 , 345–351 (2021).
Nabati, F. <i>et al.</i> Virtual screening based on the structure of more than 105 compounds against four key proteins of SARS-CoV-2: MPro, SRBD, RdRp, and PLpro. <i>Informatics Med. Unlocked</i> 35 , 101134 (2022).
Nada, H., Elkamhawy, A. & Lee, K. Identification of 1H-purine-2,6-dione derivative as a potential SARS-CoV-2 main protease inhibitor: molecular docking, dynamic simulations, and energy calculations. <i>PeerJ</i> 10 , e14120 (2022).
Nallusamy, S. <i>et al.</i> Exploring Phytochemicals of Traditional Medicinal Plants Exhibiting Inhibitory Activity Against Main Protease, Spike Glycoprotein, RNA-dependent RNA Polymerase and Non-Structural Proteins of SARS-CoV-2 Through Virtual Screening. <i>Front. Pharmacol.</i> 12 , 667704 (2021).
Nand, M. <i>et al.</i> Virtual screening of anti-HIV1 compounds against SARS-CoV-2: machine learning modeling, chemoinformatics and molecular dynamics simulation based analysis. <i>Sci. Rep.</i> 10 , 20397 (2020).
Nath, V., Rohini, A. & Kumar, V. Identification of Mpro inhibitors of SARS-CoV-2 using structure based computational drug repurposing. <i>Biocatal. Agric. Biotechnol.</i> 37 , 102178 (2021).

Annex

Nayarisseri, A. <i>et al.</i> Shape-based Machine Learning Models for the Potential Novel COVID-19 Protease Inhibitors Assisted by Molecular Dynamics Simulation. <i>Curr. Top. Med. Chem.</i> 20 , 2146–2167 (2020).
Nazir, M. <i>et al.</i> Isolation of Thioinosine and Butenolides from a Terrestrial Actinomycetes sp. GSCW-51 and Their in Silico Studies for Potential against SARS-CoV-2. <i>Chem. Biodivers.</i> 19 , e202100843 (2022).
Negru, P. A. <i>et al.</i> Virtual screening of substances used in the treatment of SARS-CoV-2 infection and analysis of compounds with known action on structurally similar proteins from other viruses. <i>Biomed. Pharmacother.</i> 153 , 113432 (2022).
Novak, J., Rimac, H., Kandagalla, S., Grishina, M. A. & Potemkin, V. A. Can natural products stop the SARS-CoV-2 virus? A docking and molecular dynamics study of a natural product database. <i>Future Med. Chem.</i> 13 , 363–378 (2021).
Ochoa, R., Palacio-Rodriguez, K., Clemente, C. M. & Adler, N. S. dockECR: Open consensus docking and ranking protocol for virtual screening of small molecules. <i>J. Mol. Graph. Model.</i> 109 , 108023 (2021).
Oh, E. <i>et al.</i> (+)-Usnic acid and its salts, inhibitors of SARS-CoV-2, identified by using in silico methods and in vitro assay. <i>Sci. Rep.</i> 12 , 13118 (2022).
Oktaç, L. <i>et al.</i> Binary-qsar guided virtual screening of fda approved drugs and compounds in clinical investigation against sars-cov-2 main protease. <i>Turkish J. Biol.</i> 45 , 459–468 (2021).
Omer, S. E. <i>et al.</i> Drug repurposing for SARS-CoV-2 main protease: Molecular docking and molecular dynamics investigations. <i>Biochem. Biophys. Reports</i> 29 , 101225 (2022).
Onyango, H., Odhiambo, P., Angwenyi, D. & Okoth, P. In Silico Identification of New Anti-SARS-CoV-2 Main Protease (Mpro) Molecules with Pharmacokinetic Properties from Natural Sources Using Molecular Dynamics (MD) Simulations and Hierarchical Virtual Screening. <i>J. Trop. Med.</i> 2022 , 1–22 (2022).
Ortega, J. T., Serrano, M. L., Pujol, F. H. & Rangel, H. R. Unrevealing sequence and structural features of novel coronavirus using in silico approaches: The main protease as molecular target. <i>EXCLI J.</i> 19 , 400–409 (2020).
Pathak, N. <i>et al.</i> Uncovering Flexible Active Site Conformations of SARS-CoV-2 3CL Proteases through Protease Pharmacophore Clusters and COVID-19 Drug Repurposing. <i>ACS Nano</i> 15 , 857–872 (2021).
Pathak, Y., Mishra, A., Choudhir, G., Kumar, A. & Tripathi, V. Rifampicin and Letemovir as potential repurposed drug candidate for COVID-19 treatment: insights from an in-silico study. <i>Pharmacol. Reports</i> 73 , 926–938 (2021).
Paul, D., Basu, D. & Ghosh Dastidar, S. Multi-conformation representation of Mpro identifies promising candidates for drug repurposing against COVID-19. <i>J. Mol. Model.</i> 27 , 128 (2021).
Portilla-Martínez, A. <i>et al.</i> In silico evaluation of flavonoids as potential inhibitors of SARS-CoV-2 main nonstructural proteins (Nsps)—amentoflavone as a multitarget candidate. <i>J. Mol. Model.</i> 28 , 404 (2022).

Quimque, M. T. J. <i>et al.</i> Virtual screening-driven drug discovery of SARS-CoV2 enzyme inhibitors targeting viral attachment, replication, post-translational modification and host immunity evasion infection mechanisms. <i>J. Biomol. Struct. Dyn.</i> 39 , 4316–4333 (2021).
Rahman, M. M. <i>et al.</i> In silico investigation and potential therapeutic approaches of natural products for COVID-19: Computer-aided drug design perspective. <i>Front. Cell. Infect. Microbiol.</i> 12 , 929430 (2022).
Rahman, M. R., Banik, A., Chowdhury, I. M., Sajib, E. H. & Sarkar, S. Identification of potential antivirals against SARS-CoV-2 using virtual screening method. <i>Informatics Med. Unlocked</i> 23 , 100531 (2021).
Raj, R. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. <i>Biochem. Biophys. Reports</i> 25 , 100847 (2021).
Raj, V., Lee, J.-H., Shim, J.-J. & Lee, J. Antiviral activities of 4H-chromen-4-one scaffold-containing flavonoids against SARS-CoV-2 using computational and in vitro approaches. <i>J. Mol. Liq.</i> 353 , 118775 (2022).
Raj, V. <i>et al.</i> Assessment of antiviral potencies of cannabinoids against SARS-CoV-2 using computational and in vitro approaches. <i>Int. J. Biol. Macromol.</i> 168 , 474–485 (2021).
Rajendran, M. <i>et al.</i> In silico screening and molecular dynamics of phytochemicals from Indian cuisine against SARS-CoV-2 MPro. <i>J. Biomol. Struct. Dyn.</i> 40 , 3155–3169 (2022).
Rajpoot, S., Alagumuthu, M. & Baig, M. S. Dual targeting of 3CLpro and PLpro of SARS-CoV-2: A novel structure-based design approach to treat COVID-19. <i>Curr. Res. Struct. Biol.</i> 3 , 9–18 (2021).
Ramesh, V. <i>et al.</i> Current Update of Phytotherapeutic Agents in the Treatment of COVID-19: In-Silico Based Virtual Screening Approach for the Development of Antiviral Drug. <i>Front. Biosci.</i> 27 , 123 (2022).
Rashdan, H. R. M. & Abdelmonsef, A. H. In silico study to identify novel potential thiadiazole-based molecules as anti-Covid-19 candidates by hierarchical virtual screening and molecular dynamics simulations. <i>Struct. Chem.</i> 33 , 1727–1739 (2022).
Rauf, A. <i>et al.</i> Docking-based virtual screening and identification of potential COVID-19 main protease inhibitors from brown algae. <i>South African J. Bot.</i> 143 , 428–434 (2021).
Ray, A. K. <i>et al.</i> Repurposing of FDA-approved drugs as potential inhibitors of the SARS-CoV-2 main protease: Molecular insights into improved therapeutic discovery. <i>Comput. Biol. Med.</i> 142 , 105183 (2022).
Refaey, R. H., El-Ashrey, M. K. & Nissan, Y. M. Repurposing of renin inhibitors as SARS-COV-2 main protease inhibitors: A computational study. <i>Virology</i> 554 , 48–54 (2021).
Rehman, M. F. <i>ur et al.</i> Effectiveness of Natural Antioxidants against SARS-CoV-2? Insights from the In-Silico World. <i>Antibiotics</i> 10 , 1011 (2021).
Reiner, Ž. <i>et al.</i> Statins and the Covid-19 main protease: In silico evidence on direct interaction. <i>Arch. Med. Sci.</i> 16 , 490–496 (2020).
Reyaz, S., Tasneem, A., Rai, G. P. & Bairagya, H. R. Investigation of structural analogs of hydroxychloroquine for SARS-CoV-2 main protease (Mpro): A computational drug discovery study. <i>J. Mol. Graph. Model.</i> 109 , 108021 (2021).

Annex

Roy, R., Sk, M. F., Tanwar, O. & Kar, P. Computational studies indicated the effectiveness of human metabolites against SARS-Cov-2 main protease. <i>Mol. Divers.</i> 1–16 (2022).
Rubio-Martínez, J. <i>et al.</i> Discovery of Diverse Natural Products as Inhibitors of SARS-CoV-2 MproProtease through Virtual Screening. <i>J. Chem. Inf. Model.</i> 61 , 6094–6106 (2021).
Saakre, M., Mathew, D. & Ravisankar, V. Perspectives on plant flavonoid quercetin-based drugs for novel SARS-CoV-2. <i>Beni-Suef Univ. J. Basic Appl. Sci.</i> 10 , 21 (2021).
Said, M. A., Albohy, A., Abdelrahman, M. A. & Ibrahim, H. S. Importance of glutamine 189 flexibility in SARS-CoV-2 main protease: Lesson learned from in silico virtual screening of ChEMBL database and molecular dynamics. <i>Eur. J. Pharm. Sci.</i> 160 , 105744 (2021).
Sanachai, K. <i>et al.</i> Identification of repurposing therapeutics toward SARS-CoV-2 main protease by virtual screening. <i>PLoS One</i> 17 , e0269563 (2022).
Santos, L. H. <i>et al.</i> Structure-Based Identification of Naphthoquinones and Derivatives as Novel Inhibitors of Main Protease Mproand Papain-like Protease PLproof SARS-CoV-2. <i>J. Chem. Inf. Model.</i> 62 , 6553–6573 (2022).
Scior, T., Abdallah, H. H., Mustafa, S. F. Z., Guevara-García, J. A. & Rehder, D. Are vanadium complexes druggable against the main protease Mpro of SARS-CoV-2? – A computational approach. <i>Inorganica Chim. Acta</i> 519 , 120287 (2021).
Selvaraj, C. <i>et al.</i> Microsecond MD Simulation and Multiple-Conformation Virtual Screening to Identify Potential Anti-COVID-19 Inhibitors Against SARS-CoV-2 Main Protease. <i>Front. Chem.</i> 8 , 595273 (2020).
Semenov, V. A. & Krivdin, L. B. Combined Computational NMR and Molecular Docking Scrutiny of Potential Natural SARS-CoV-2 M pro Inhibitors. <i>J. Phys. Chem. B</i> 126 , 2173–2187 (2022).
Sen, D. <i>et al.</i> Identification of potential edible mushroom as SARS-CoV-2 main protease inhibitor using rational drug designing approach. <i>Sci. Rep.</i> 12 , 1503 (2022).
Sen, D., Debnath, P., Debnath, B., Bhaumik, S. & Debnath, S. Identification of potential inhibitors of SARS-CoV-2 main protease and spike receptor from 10 important spices through structure-based virtual screening and molecular dynamic study. <i>J. Biomol. Struct. Dyn.</i> 40 , 941–962 (2022).
Hisham Shady, N. <i>et al.</i> Sterols and Triterpenes: Antiviral Potential Supported by In-Silico Analysis. <i>Plants (Basel, Switzerland)</i> 10 , 41 (2020).
Shahhamzehei, N., Abdelfatah, S. & Efferth, T. In Silico and In Vitro Identification of Pan-Coronaviral Main Protease Inhibitors from a Large Natural Product Library. <i>Pharmaceuticals</i> 15 , 308 (2022).
Sharma, T. <i>et al.</i> Screening of drug databank against WT and mutant main protease of SARS-CoV-2: Towards finding potential compound for repurposing against COVID-19. <i>Saudi J. Biol. Sci.</i> 28 , 3152–3159 (2021).
Sisakht, M., Mahmoodzadeh, A. & Darabian, M. Plant-derived chemicals as potential inhibitors of SARS-CoV-2 main protease (6LU7), a virtual screening study. <i>Phyther. Res.</i> 35 , 3262–3274 (2021).

Siva Kumar, B., Anuragh, S., Kammala, A. K. & Ilango, K. Computer Aided Drug Design Approach to Screen Phytoconstituents of Adhatoda vasica as Potential Inhibitors of SARS-CoV-2 Main Protease Enzyme. <i>Life</i> 12 , 315 (2022).
Sobhia, M. E. <i>et al.</i> Rapid structure-based identification of potential SARS-CoV-2 main protease inhibitors. <i>Future Med. Chem.</i> 13 , 1435–1450 (2021).
Srivastava, A. <i>et al.</i> Exploring nature's bounty: identification of Withania somnifera as a promising source of therapeutic agents against COVID-19 by virtual screening and in silico evaluation. <i>J. Biomol. Struct. Dyn.</i> 40 , 1858–1908 (2022).
Stefaniu, A., Pirvu, L., Albu, B. & Pintilie, L. Molecular Docking Study on Several Benzoic Acid Derivatives against SARS-CoV-2. <i>Molecules</i> 25 , 5828 (2020).
Swain, S. S. <i>et al.</i> Integrated bioinformatics–cheminformatics approach toward locating pseudo-potential antiviral marine alkaloids against SARS-CoV-2-Mpro. <i>Proteins Struct. Funct. Bioinforma.</i> 90 , 1617–1633 (2022).
Tahir ul Qamar, M. <i>et al.</i> Target-Specific Machine Learning Scoring Function Improved Structure-Based Virtual Screening Performance for SARS-CoV-2 Drugs Development. <i>Int. J. Mol. Sci.</i> 23 , 11003 (2022).
Talluri, S. Molecular Docking and Virtual Screening Based Prediction of Drugs for COVID-19. <i>Comb. Chem. High Throughput Screen.</i> 24 , 716–728 (2020).
Tejera, E., Munteanu, C. R., López-Cortés, A., Cabrera-Andrade, A. & Pérez-Castillo, Y. Drugs Repurposing Using QSAR, Docking and Molecular Dynamics for Possible Inhibitors of the SARS-CoV-2 Mpro Protease. <i>Molecules</i> 25 , 5172 (2020).
Trezza, A., Mugnaini, C., Corelli, F., Santucci, A. & Spiga, O. In Silico Multi-Target Approach Revealed Potential Lead Compounds as Scaffold for the Synthesis of Chemical Analogues Targeting SARS-CoV-2. <i>Biology (Basel)</i> . 11 , 465 (2022).
Tripathi, N. <i>et al.</i> Virtual screening and molecular simulation study of natural products database for lead identification of novel coronavirus main protease inhibitors. <i>J. Biomol. Struct. Dyn.</i> 40 , 3655–3667 (2022).
Umar, H. I. <i>et al.</i> Novel Molecules derived from 3-O-(6-galloylglucoside) inhibit Main Protease of SARS-CoV 2 In Silico. <i>Chem. Pap.</i> 76 , 785–796 (2022).
Uniyal, A., Mahapatra, M. K., Tiwari, V., Sandhir, R. & Kumar, R. Targeting SARS-CoV-2 main protease: structure based virtual screening, in silico ADMET studies and molecular dynamics simulation for identification of potential inhibitors. <i>J. Biomol. Struct. Dyn.</i> 40 , 3609–3625 (2022).
Venugopal, P. P. & Chakraborty, D. Molecular mechanism of inhibition of COVID-19 main protease by β -adrenoceptor agonists and adenosine deaminase inhibitors using in silico methods. <i>J. Biomol. Struct. Dyn.</i> 40 , 5112–5127 (2022).
Verma, A. K. & Aggarwal, R. Repurposing potential of FDA-approved and investigational drugs for COVID-19 targeting SARS-CoV-2 spike and main protease and validation by machine learning algorithm. <i>Chem. Biol. Drug Des.</i> 97 , 836–853 (2021).
Verma, S., Patel, C. N. & Chandra, M. Identification of novel inhibitors of SARS-CoV-2 main protease (Mpro) from Withania sp. by molecular docking and molecular dynamics simulation. <i>J. Comput. Chem.</i> 42 , 1861–1872 (2021).

Annex

Wadanambi, P. M., Jayathilaka, N. & Seneviratne, K. N. A Computational Study of Carbazole Alkaloids from <i>Murraya koenigii</i> as Potential SARS-CoV-2 Main Protease Inhibitors. <i>Appl. Biochem. Biotechnol.</i> 195 , 573–596 (2023).
Wang, L. <i>et al.</i> A pocket-based 3D molecule generative model fueled by experimental electron density. <i>Sci. Rep.</i> 12 , 15100 (2022).
Wang, S., Sun, Q., Xu, Y., Pei, J. & Lai, L. A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2. <i>Brief. Bioinform.</i> 22 , bbab211 (2021).
Xu, J., Gao, L., Liang, H. & Chen, S. In silico screening of potential anti-COVID-19 bioactive natural constituents from food sources by molecular docking. <i>Nutrition</i> 82 , 111049 (2021).
Xu, T. <i>et al.</i> Efficient Identification of Anti-SARS-CoV-2 Compounds Using Chemical Structure- and Biological Activity-Based Modeling. <i>J. Med. Chem.</i> 65 , 4590–4599 (2022).
Yang, F. <i>et al.</i> Signaling repurposable drug combinations against COVID-19 by developing the heterogeneous deep herb-graph method. <i>Brief. Bioinform.</i> 23 , bbac124 (2022).
Yang, J. <i>et al.</i> Structure-Based Discovery of Novel Nonpeptide Inhibitors Targeting SARS-CoV-2 Mpro. <i>J. Chem. Inf. Model.</i> 61 , 3917–3926 (2021).
Yang, X., Xing, X., Liu, Y. & Zheng, Y. Screening of potential inhibitors targeting the main protease structure of SARS-CoV-2 via molecular docking. <i>Front. Pharmacol.</i> 13 , 962863 (2022).
Yu, W. <i>et al.</i> Computational Simulation of HIV Protease Inhibitors to the Main Protease (Mpro) of SARS-CoV-2: Implications for COVID-19 Drugs Design. <i>Molecules</i> 26 , 7385 (2021).
Yuce, M. <i>et al.</i> Repurposing of FDA-approved drugs against active site and potential allosteric drug-binding sites of COVID-19 main protease. <i>Proteins Struct. Funct. Bioinforma.</i> 89 , 1425–1441 (2021).
Zaki, M. E. A. <i>et al.</i> Identification of Anti-SARS-CoV-2 Compounds from Food Using QSAR-Based Virtual Screening, Molecular Docking, and Molecular Dynamics Simulation Analysis. <i>Pharmaceuticals</i> 14 , 357 (2021).
Zare, S., Pirhadi, S., El Seedi, H. R. & Jassbi, A. R. Anti-COVID-19 and antidiabetic activities of new oleanane and ursane-type triterpenoids from <i>Salvia grossheimii</i> : an in-silico approach. <i>J. Recept. Signal Transduct.</i> 42 , 540–548 (2022).
Zhang, L., Howland, M., Hilgenfeld, R., Anderson, M. O. & Eagon, S. Identification of non-covalent SARS-CoV-2 main protease inhibitors by a virtual screen of commercially available drug-like compounds. <i>Bioorg. Med. Chem. Lett.</i> 41 , 127990 (2021).
Zhang, X. <i>et al.</i> A network pharmacology based approach for predicting active ingredients and potential mechanism of lianhuapingwen capsule in treating covid-19. <i>Int. J. Med. Sci.</i> 18 , 1866–1876 (2021).
Zhang, Y. <i>et al.</i> Cotton flower metabolites inhibit SARS-CoV-2 main protease. <i>FEBS Open Bio</i> 12 , 1886–1895 (2022).

Zhang, Z. jia <i>et al.</i> Active constituents and mechanisms of Respiratory Detox Shot, a traditional Chinese medicine prescription, for COVID-19 control and prevention: Network-molecular docking-LC–MSE analysis. <i>J. Integr. Med.</i> 18 , 229–241 (2020).
Zhao, J. <i>et al.</i> Exploration of SARS-CoV-2 3CLproInhibitors by Virtual Screening Methods, FRET Detection, and CPE Assay. <i>J. Chem. Inf. Model.</i> 61 , 5763–5773 (2021).
Zhao, L. <i>et al.</i> Multi-target mechanisms against coronaviruses of constituents from Chinese Dagang Tea revealed by experimental and docking studies. <i>J. Ethnopharmacol.</i> 297 , 115528 (2022).
Zhao, X., Liu, R., Miao, Z., Ye, N. & Lu, W. Retracted: A Study of Potential SARS-CoV-2 Antiviral Drugs and Preliminary Research of Their Molecular Mechanism, Based on Anti-SARS-CoV Drug Screening and Molecular Dynamics Simulation. <i>J. Comput. Biol.</i> 27 , e1699–e1713 (2020).
Zígolo, M. A., Goytia, M. R., Poma, H. R., Rajal, V. B. & Irazusta, V. P. Virtual screening of plant-derived compounds against SARS-CoV-2 viral proteins using computational tools. <i>Sci. Total Environ.</i> 781 , 146400 (2021).
Zrieq, R. <i>et al.</i> Tomatidine and Patchouli Alcohol as Inhibitors of SARS-CoV-2 Enzymes (3CLpro, PLpro and NSP15) by Molecular Docking and Molecular Dynamics Simulations. <i>Int. J. Mol. Sci.</i> 22 , 10693 (2021).

UNIVERSITAT ROVIRA I VIRGILI

STUDYING THE LIMITS OF PROTEIN-LIGAND DOCKING TO PREDICT NEW NON-COVALENT INHIBITORS FOR THE SARSCOV-2 MAIN PROTEASE

Guillem Macip Sancho



UNIVERSITAT
ROVIRA I VIRGILI