Journal of Molecular Graphics and Modelling 140 (2025) 109100

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

# Combinatorial chemistry-driven In silico design and computational evaluation of covalent peptidomimetic SARS-CoV-2 main protease inhibitors via structure-based virtual screening and multivariate analysis

Alessia Bono [a], Gabriele La Monica [a], Federica Alamia [a], Francesco Mingoia [b], Annamaria Martorana [a], Antonino Lauria [a,c,*]

[a] Dipartimento di Scienze e Technologie Biologiche Chimiche e Farmaceutiche "STEBICEF" – University of Palermo, Viale delle Scienze – Ed. 17, 90128, Palermo, Italy
[b] Istituto per lo Studio dei Materiali Nanostrutturati - (ISMN)Consiglio Nazionale delle Ricerche (CNR), Via U. La Malfa 153, 90146, Palermo, Italy
[c] NBFC, National Biodiversity Future Center, Piazza Marina 61, Palermo, 90133, Italy

ARTICLE INFO

*Keywords:*
COVID-19
SARS-CoV-2
$M^{PRO}$
Covalent inhibitors
Electrophilic warhead
Catalytic cysteine

ABSTRACT

The COVID-19 pandemic has underscored the urgent need for specific pharmacological treatments beyond existing vaccines. One of the most attractive targets for antiviral therapies development is the SARS-CoV-2 Main Protease ($M^{PRO}$), a key enzyme in viral life. The lack of $M^{PRO}$ human homologs and its conservation rate among coronaviruses make this enzyme strategically important. Considering its mechanism of action, the catalytic cysteine residue ($Cys^{145}$) presents a prime target for covalent inhibition. Electrophilic warhead inhibitors, designed to react with this catalytic site, mimic the amide peptide bonds of the viral polyproteins, thereby facilitating their binding and subsequent inactivation of the enzyme. Their activity is further potentiated when incorporated into peptidomimetic structures. *In silico* approaches are gaining increasing importance in the search for effective COVID-19 treatments. In this view, this study focuses on developing an innovative *in silico* protocol for identifying anti-SARS-CoV-2 agents covalently targeting $M^{PRO}$. To this purpose, a combinatorial library of 450 peptidomimetic compounds with aldehydic warheads was generated, refined to 388 compounds through docking studies, and further evaluated for covalent binding capabilities, revealing that compounds **9**–**14**, **16**, and **20** exhibited significantly higher affinities compared to known inhibitors, even during a 200 ns dynamics simulation, thus affirming the validity of the adopted design strategy. Furthermore, this work takes the advantages of our in-house ligand-based tool, the Biotarget Predictor Tool, available in DRUDIT, integrated with both structure-based techniques and, interestingly, multivariate statistical analysis.

## 1. Introduction

The Coronavirus Disease-19 (COVID-19) global pandemic has been etiologically caused by the highly infectious Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) [1,2], a β-coronavirus (β-CoV) that infects both animals and humans, causing respiratory diseases [3]. As of today, the World Health Organization (WHO) has reported over 775 million cases globally, with a mortality rate of approximately 1 %. In response to the outbreak, the scientific community has recommended stringent social distancing measures and the rapid development of vaccines to contain the spread of the virus. Over 20 vaccines have successfully reached clinical trials, demonstrating their efficacy in preventing severe COVID-19 [4]. However, considering the high mutation rate characteristic of this virus family, the long-term efficacy and safety of these vaccines remain under debate. This highlights the urgent need for more specific pharmacological treatments to complement vaccines in both the prophylaxis and treatment of COVID-19 [5]. Moreover, there is the need to develop effective therapeutic strategies against SARS-CoV-2. Over the past four years, recombinant proteins, monoclonal antibodies, immunological treatments, and drug repurposing have emerged as effective approaches based on comparative studies among coronaviruses [6–35].

To date, with the goal of developing a selective antiviral drug against SARS-CoV-2, the research community worldwide evaluated several

* Corresponding author. Dipartimento di Scienze e Technologie Biologiche Chimiche e Farmaceutiche "STEBICEF" – University of Palermo, Viale delle Scienze – Ed. 17, 90128, Palermo, Italy
*E-mail address:* antonino.lauria@unipa.it (A. Lauria).

https://doi.org/10.1016/j.jmgm.2025.109100
Received 18 February 2025; Received in revised form 26 March 2025; Accepted 29 May 2025
Available online 1 June 2025

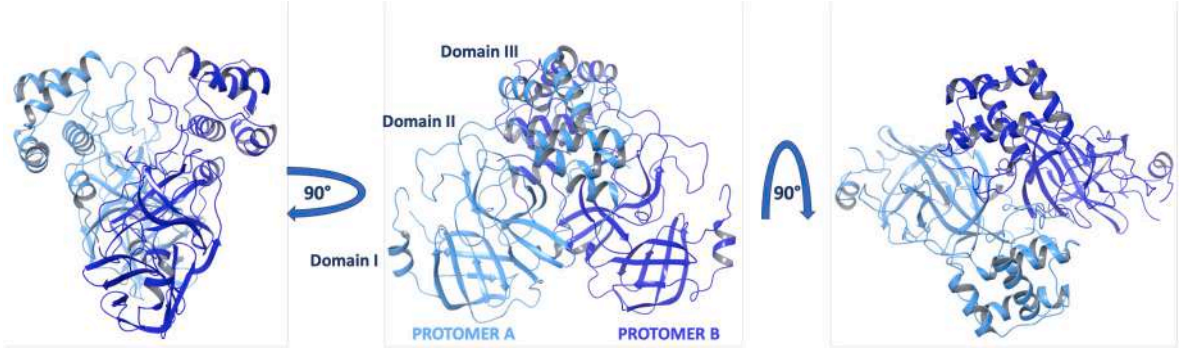



**Fig. 1.** X-Ray structure of dimeric SARS-CoV-2 $M^{PRO}$ (PDB code 6Y2F) [38].

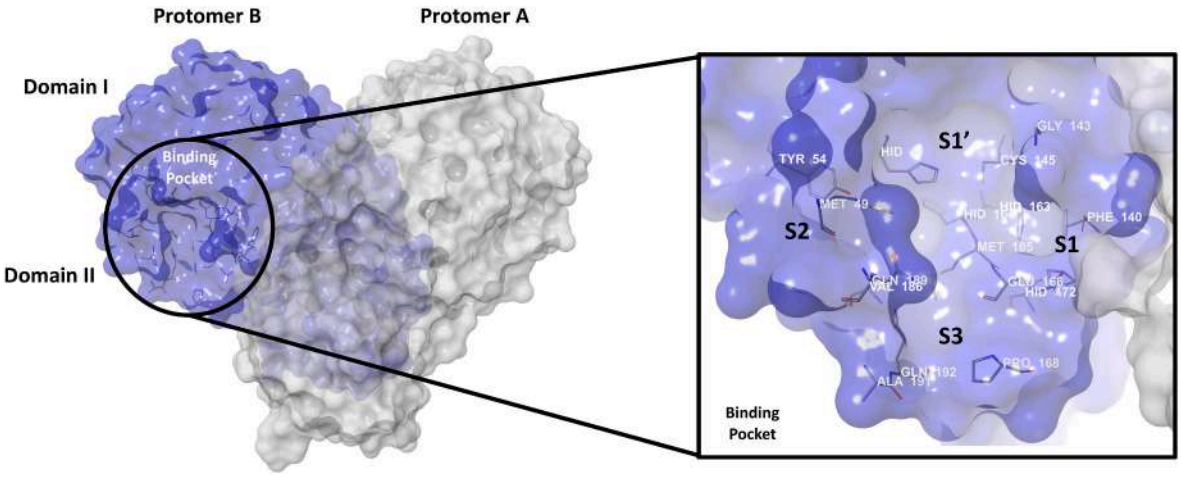



**Fig. 2.** X-Ray structure of dimeric SARS-CoV-2 $M^{PRO}$ and focus on the catalytic site, with the four region S1, S1′, S2 and S3/S4 highlighted (PDB code 6Y2F) [38].

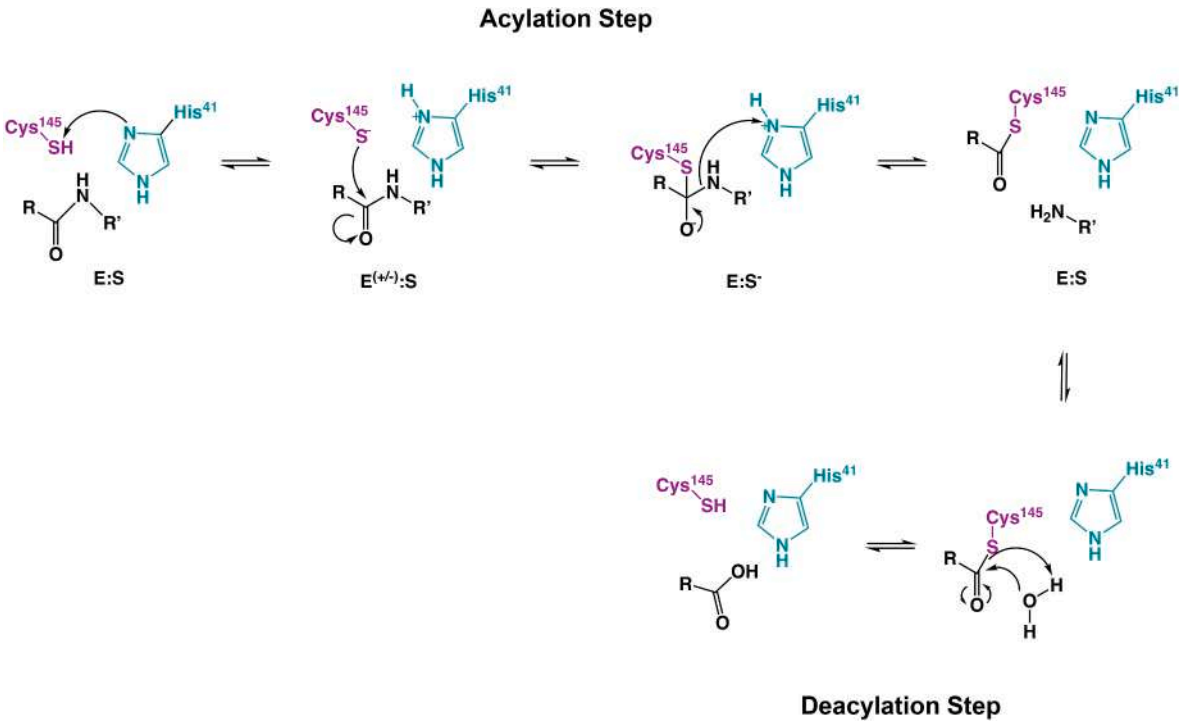
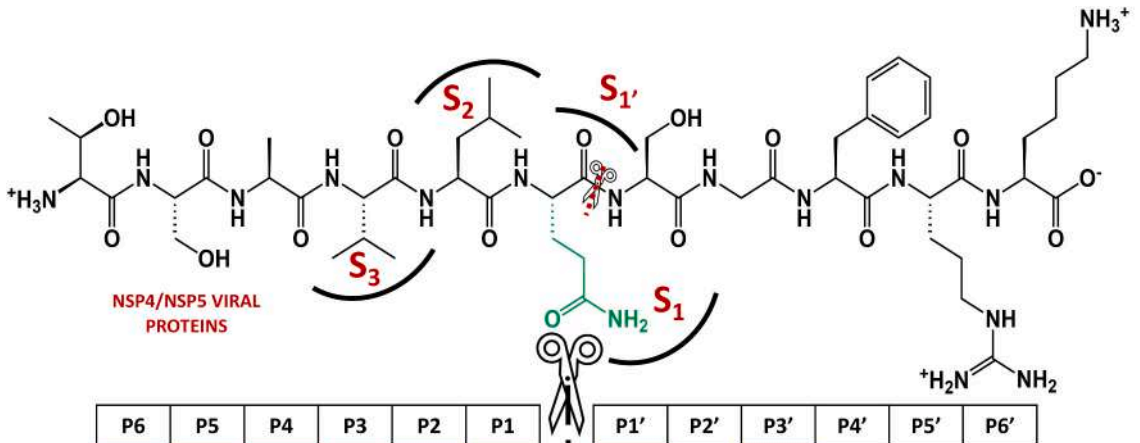



**Fig. 3.** SARS-CoV-2 $M^{PRO}$ mechanism of action processed by the catalytic dyad ($Cys^{145}$ and $His^{41}$, located in the S1′ subregion of the binding pocket). $Cys^{145}$ is highlighted in purple, while $His^{41}$ in aqua green. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)




| P6 | P5 | P4 | P3 | P2 | P1 | P1′ | P2′ | P3′ | P4′ | P5′ | P6′ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | S | A | V | L | Q | S | G | F | R | K | M | NSP4/NSP5 |
| S | G | V | T | F | Q | S | A | V | K | R | T | NSP5/NSP6 |
| K | V | A | T | V | Q | S | K | M | S | D | V | NSP6/NSP7 |
| N | R | A | T | L | Q | A | I | A | S | E | F | NSP7/NSP8 |
| S | A | V | K | L | Q | N | N | E | L | S | P | NSP8/NSP9 |
| A | T | V | R | L | Q | A | G | N | A | T | E | NSP9/NSP10 |
| R | E | P | M | L | Q | S | A | D | A | Q | S | NSP10/NSP11 |
| P | H | T | V | L | Q | A | V | G | A | C | V | NSP11/NSP12 NSP12/NSP13 |
| N | V | A | T | L | Q | A | E | N | V | T | G | NSP13/NSP14 |
| T | F | T | R | L | Q | S | L | E | N | V | A | NSP14/NSP15 |
| F | Y | P | K | L | Q | S | S | Q | A | W | Q | NSP15/NSP16 |

**Fig. 4.** 2D structures of the viral NSP4 and NSP5 proteins are depicted alongside the P6-P5-P4-P3-P2-P1-P1′-P2′-P3′-P4′-P5′-P6′ cleavage sequences for NSP4 to NSP16. The image highlights the recurrent glutamine residues at the P1 position, a critical feature across these sequences.

structural/non-structural viral proteins as possible SARS-CoV-2 druggable targets. Among these, the Main Protease ($M^{PRO}$) emerged as the most attractive target for drug design [36].

**$M^{PRO}$, a**lso known as **3C-Like Protease ($3CL^{PRO}$)**, is a crucial enzyme in the SARS-CoV-2 replication cycle, primarily involved in processing the viral polyproteins pp1a and pp1ab. By cleaving these polyproteins into functional Non-Structural Proteins (NSPs), $M^{PRO}$ enables the assembly of the Replication-Transcription Complex (RTC), which is essential for viral replication. $M^{PRO}$ is highly conserved among coronaviruses and functions as a homodimer, with its active site being a key target for antiviral drug development. The absence of similar proteases in human cells makes $M^{PRO}$ an ideal therapeutic target. Inhibition of $M^{PRO}$ has been shown to be a highly effective strategy in preventing the spread of SARS-CoV-2 and represents a cornerstone in the ongoing search for broad-spectrum antiviral therapies [36].

Structurally, $M^{PRO}$ protomers consist of 306 amino acids divided into 3 domains (I, II, and III): domain I (8–101) comprises 6 β-strands and an α-helix, while domain II (102–184) and domain III (201–303) include 6 β-strands and 5 α-helices respectively. All domains are connected by long loop regions (Fig. 1) [37].

Dimerization of the protein is a necessary step for the catalytic activity, because of in the monomeric state the active site pocket collapses and is not available for the binding with the substrate. The binding pocket is organized into four subsites, S1′, S1, S2, and S3/S4. In the region S1', between domains I and II, is located the catalytic site, characterized by the catalytic dyad ($Cys^{145}$ and $His^{41}$). During the hydrolysis of the peptide bond, $His^{41}$ activates the nucleophilic -SH of $Cys^{145}$ by deprotonation, with subsequent stabilization of the adduct by the so-called "oxyanion hole", formed by the $Gly^{143}$ and $Cys^{145}$ backbones. The S1 region, characterized by the side chains of $Phe^{140}$, $His^{163}$, $His^{164}$, $Glu^{166}$, and $His^{172}$, is highly specific for the glutamine residue [39]. The S2 consists of hydrophobic amino acids, such as $Met^{49}$, $Tyr^{54}$, $Met^{165}$, $Pro^{168}$, $Val^{186}$, and finally S3/S4, which are particularly exposed to the solvent, involve $Gln^{189}$, $Ala^{191}$, $Gln^{192}$, $Gly^{251}$ residues (Fig. 2).

The $M^{PRO}$ mechanism of action involves particularly the S1' region. The catalytic dyad ($His^{41}$/$Cys^{145}$) is pivotal for the enzyme as it drives the process of peptide bond cleavage in the viral polyprotein. In this context, the enzyme-substrate initial complex corresponds to the neutral $His^{41}$/$Cys^{145}$ dyad (**E:S** in Fig. 3). From this stable state, the imidazole group of $His^{41}$ polarizes and acts as an activator by removing a proton from the thiol group (SH) of $Cys^{145}$ to form a highly nucleophilic $CysS^-$/$HisH^+$ ion pair, that would react with the substrate (**$E^{(\pm)}$:S** in Fig. 3) [40]. The concomitant nucleophilic attack of $CysS^-$ on the carbonyl carbon atom of the peptide bond leads to a pseudo-stable intermediate (**E:$S^-$** in Fig. 3).

Then, the proton from the protonated $HisH^+$ is transferred to the N atom of the scissile peptide bond, forming an acyl-enzyme covalent intermediate (**E:S** in Fig. 3).

After this acylation step, the recovery of the enzyme in the following deacylation stage is assisted by a water molecule activated by the His, which performs the hydrolysis of the thioester intermediate. This reaction results in the production of a smaller viral protein fragment with a new amino-terminal end and the regeneration of the enzyme for subsequent catalytic cycles [41,42].

The strategic placement of $His^{41}$ and $Cys^{145}$ in the S1' region not only facilitates this reaction but also underscores the importance of this site in $M^{PRO}$ function. Disruptions in the S1' pocket could impede the enzyme ability to make timely and specific cuts in the polyprotein. An interesting approach has been provided by the covalent inhibition of the SARS-CoV-2 $M^{PRO}$ cysteinyl protease [43–45], where the catalytic sulfur of $Cys^{145}$ is

covalently trapped by an electrophilic moiety, a reactive warhead that mimics the amide peptide of the viral polyproteins.

According to the Schechter and Berger nomenclature, the peptide substrate residues recognized by $M^{PRO}$ are referred to, from N to C terminus, P3-P2-P1-P1′-P2′-P3′, with the hydrolysis taking place at the scissile bond between P1 and P1'. The multiple subsites S1′, S1, S2, and S3 are labeled as they can recognize residues in P1′, P1, P2, and P3, respectively. $M^{PRO}$ cleaves 11 of the 16 highly conserved recognition sites of the polyprotein pp1a/pp1ab, by specifically targeting the P1↓P1′ positions (where ↓ denotes the peptide bond cleavage location).

Naturally, $M^{PRO}$ recognizes diverse cleavage sequences but with a stringent requirement of glutamine at P1, which gets recognized by the S1 subsite in the active site of $M^{PRO}$. S1 subsite residues form an intricate network of interactions with the peptide substrate to ensure glutamine specificity at P1 substrate position [46]. P2 tolerates more hydrophobic amino acid with a clear preference for leucine, and P1′ is a non-conserved prime recognition site (a small amino acid, such as serine, glycine, or alanine) [47,48]. Fig. 4 shows the 2D structure of NSP4/NSP5 viral proteins and the P6-P5-P4-P3-P2-P1-P1′-P2′-P3′-P4′-P5′-P6′ sequences for NSP4-NSP16.

Since P4, P3, P2, P1, and P1′ residues interact with the substrate recognition sites (S3/S4–S1′), a recurrence in amino acids is appreciable in contrast with a higher variation rate for P2′, P3′, and P4'. In details, the analysis of amino acids in sequences highlights that residues as Ser and Ala are commonly found in P1', confirming that small aliphatic residues are favored at this position. The recurrent glutamine residue in P1 strongly suggests that Gln is critical for substrate binding and cleavage; a mutation of Gln would abolish the substrate activity completely. P2 position requires a large hydrophobic residue (mostly Leu, but also Phe and Val), indeed, hydrophobic interactions between the P2 residue and the S2 pocket of the enzyme are crucial for substrate recognition. As well as P2, P3 residues must be characterized by substituent groups capable of favorable hydrophobic interactions with the amino acids of S3/S4. The preferred amino acids in P3 are Val, Thr, Met, Lys, and Arg. Since no human host-cell proteases are known with this substrate specificity, $M^{PRO}$ inhibition could drive to blockage of viral replication [49], making itself the most attractive target for COVID-19 antiviral treatment.

Peptidomimetic inhibitors with an electrophilic warhead have emerged as a promising class of antiviral agents for targeting SARS-CoV-2 $M^{PRO}$. Indeed, their design leverages the enzyme substrate recognition and catalytic mechanism, effectively mimicking the natural substrates while introducing a reactive moiety at P1' position that covalently binds the $Cys^{145}$.

In the panorama of covalent inhibitors, aldehydes represent the most abundant class of small molecules and the most promising warhead in the design of new covalent inhibitors of SARS-CoV-2 $M^{PRO}$ [48]. Mechanistically, the carbonyl carbon is susceptible to the nucleophilic addition of the cysteine-SH, leading to the formation of a reversible hemithioacetal adduct. The high similarity between the latter and the intermediate formed by the natural substrate during the enzymatic catalytic cycle ensures high stability of the inhibitor−protein complex and a longer residence time [41].

In light of these considerations, this study employed an innovative *in silico* hybrid and hierarchical virtual screening approach to identify novel SARS-CoV-2 $M^{PRO}$ selective covalent inhibitors. Building on our preliminary findings [48], we gathered all the most recurrent pharmacophoric moieties from the P1, P2, and P3 positions and integrated them into a combinatorial analysis. This analysis facilitated the rational design of new small molecules featuring a peptidomimetic scaffold with an aldehydic electrophilic warhead. We leveraged structure-based studies, ligand-based studies, and multivariate statistical analysis to identify compounds with the potential to serve as new antiviral agents.

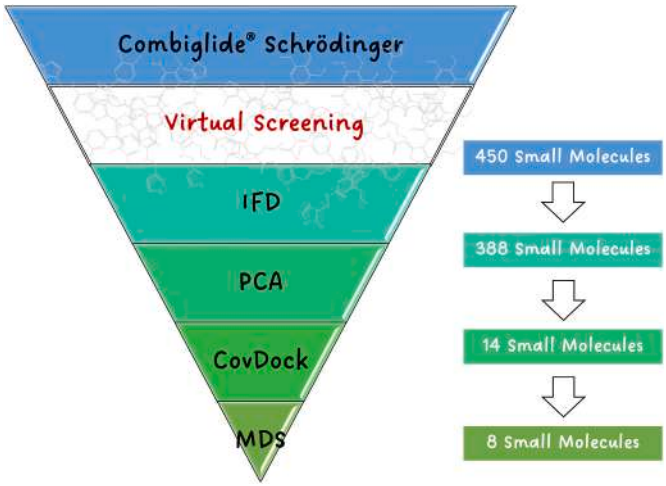



**Fig. 5.** Virtual Screening workflow, including database building, and the four key steps (IFD, PCA, CovDock, and MDS), for the identification of potential peptidomimetic covalent inhibitors with aldehydic warhead.

## 2. Results and discussion

In response to the growing interest in more potent and effective covalent inhibitors, this study proposes a rational design of peptidomimetic compounds, with aldehydic warhead, as promising SARS-CoV-2 $M^{PRO}$ covalent inhibitors.

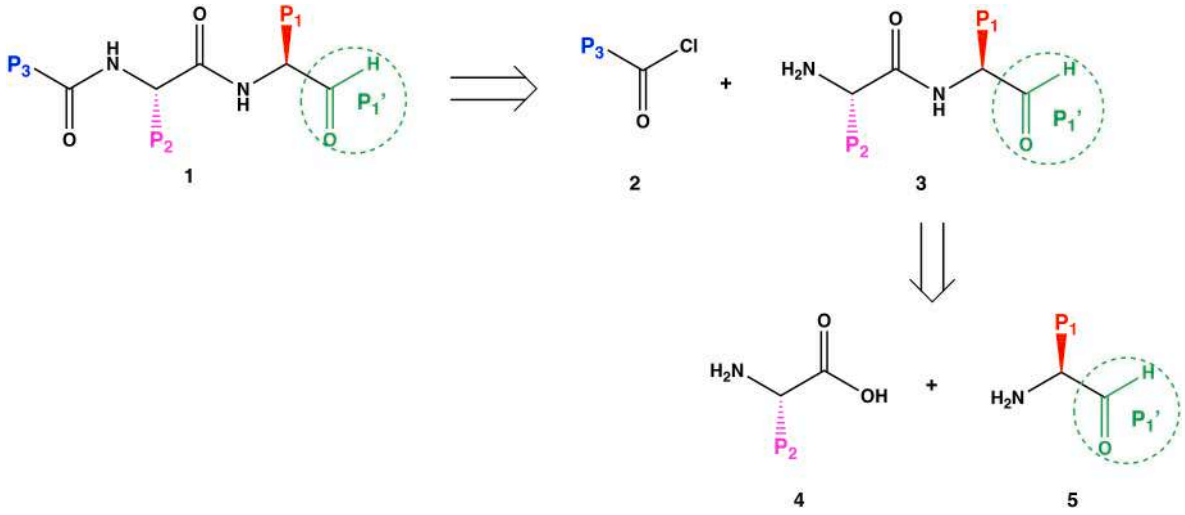
Our approach involved a systematic workflow, as shown in Fig. 5, structured in key phases to identify potential inhibitors. Initially, we identified the most recurrent and crucial fragments for each portion of the inhibitor (P1, P2, P3) [48]. Using a combinatorial approach (CombiGlide) integrated into the Maestro software suite, we constructed a database of 450 molecules, each designed with an aldehydic warhead at the P1' position to facilitate potential covalent interaction with the active site cysteine of $M^{PRO}$.

Subsequently, IFD studies were performed on these molecules within the $M^{PRO}$ binding site. Out of the 450 molecules, 388 were successfully docked, while 62 were excluded due to size incompatibility with the binding site. For the docked structures and a reference set of known $M^{PRO}$ inhibitors with established $IC_{50}$ values, we used DRUDIT and MOLDESTO to calculate molecular descriptors. Based on these descriptors, we developed a Principal Component Analysis (PCA) model, which produced a 2D plot where inhibitors with $IC_{50}$ values below 0.1 μM and those above 0.1 μM formed two distinct clusters. We then analyzed the distribution of the 388 molecules within the PCA plot to assess their positioning relative to the two clusters, providing insights into their potential inhibitory efficacy. Based on their favorable positioning in the PCA plot and alignment with the identified structural requirements, we selected 14 molecules for pharmacokinetic evaluation and detailed CovDock studies. The best 11 compounds were submitted to a preliminary Molecular Dynamics Simulation (MDS) over a simulation time of 50 ns, which led to select the final 8 compounds for longer MDS (200 ns).

Through this structured and methodical approach, we successfully identified promising new covalent peptidomimetic inhibitors featuring aldehydic warheads that demonstrate strong potential as SARS-CoV-2 $M^{PRO}$ inhibitors. These findings represent a significant contribution to the ongoing global effort in antiviral drug development and offer a robust framework for the future design of selective covalent inhibitors.

### 2.1. *In silico design of a new database of small molecules*

The initial phase of the research project focused on constructing a comprehensive database of peptidomimetic structures, each containing a reactive electrophilic aldehyde warhead. The database construction




**Scheme 1.** Retrosynthetic scheme of aldehyde-based peptidomimetic derivatives **1**. The electrophilic warhead in P1' is shown in green, while the P1−P3 fragments are depicted in red, purple, and blue, respectively.

process consisted of three sequential steps: designing a retrosynthetic scheme, preparing the reagents, and constructing the combinatorial library.

To efficiently plan the computational synthesis of aldehyde-based peptidomimetic derivatives **1**, a well-structured retrosynthetic scheme was developed. As outlined in Scheme 1, the synthesis of these derivatives follows a pathway designed solely for computational purposes; therefore, specific reaction conditions and experimental details are not included in this analysis. The computational scheme is developed to virtually generate the molecular files corresponding to derivatives **1**. Initially, the target molecules are formed via a nucleophilic acyl substitution reaction between synthons **2** and **3**, establishing the crucial acyl linkage that characterizes the peptidomimetic structure. Prior to this, synthon **3** is synthesized through a peptide coupling reaction between synthons **4** and **5**, assembling the necessary peptidic backbone. This retrosynthetic strategy ensures a systematic and coherent approach to the construction of the desired peptidomimetic derivatives, integrating all essential structural features to achieve the targeted biological activity.

Following the retrosynthetic analysis, the second step involves the utilization of the Reagent Preparation Panel. This crucial step ensures that the input files for the synthons (or reagents) are adequately prepared for use in the subsequent combinatorial library enumeration process. The Reagent Preparation Panel is designed to convert 2D molecular structures into optimized 3D structures, embedding all necessary information required to construct molecules for virtual screening.

The preparation begins with the conversion of 2D chemical structures into their corresponding 3D forms, performed by the Ligand Preparation task (LigPrep), which not only generates 3D conformers but also performs energy minimization to ensure that the structures are in a low-energy, physically realistic state. This step is vital as it reduces the likelihood of artifacts in the modeling process and ensures that the reagents are ready for subsequent computational operations.

The main task of the Reagent Preparation Panel is the selection of specific reagent types based on functional groups. For each reagent, the panel identifies the functional group within the molecule and pinpoints the specific bond that will be replaced or modified when the reagent is added to the core structure during the library enumeration process. This targeted approach ensures that the chemical transformations modeled *in silico* are both realistic and aligned with the intended synthetic strategy.

The output of the Reagent Preparation Panel consists of prepared reagent files, ready for integration into the combinatorial library. These files contain 3D structures of the reagents, along with information about the reactive sites. Notably, the process may yield multiple output structures for a single input reagent, reflecting the different energetically favorable conformations that the molecule can adopt. This conformational diversity is crucial for exploring a wide chemical space during virtual screening, as different conformers might interact differently with potential target sites.

The selection of P1, P2, and P3 fragments was guided by an analysis of common pharmacophoric features observed in known $M^{PRO}$ inhibitors. Specifically, this study builds upon a preliminary literature analysis [47,48]. Table 1 lists all the building blocks used in the combinatorial process. Specifically, three types of reagents were prepared: Amino acids (**6a-m**) for including P1 and P2 fragments, Acyl chlorides (**7a-j**), and Carbonyl chlorides (**8a-e**). By carefully preparing each reagent, this method ensures that all components of the combinatorial library are structurally sound and ready for subsequent steps in the computational database design workflow, ultimately enhancing the reliability and success of the virtual screening efforts.

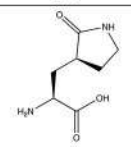
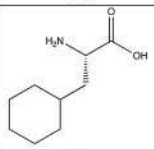
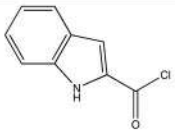
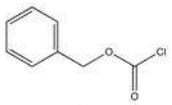
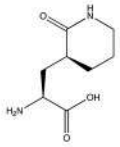
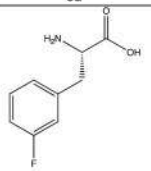
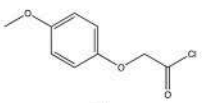
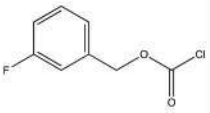
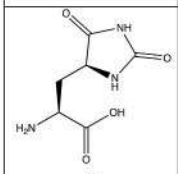
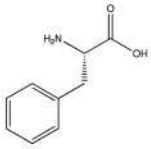
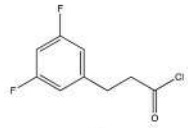
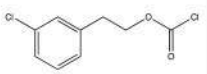
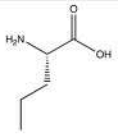
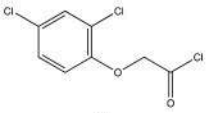
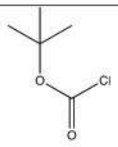
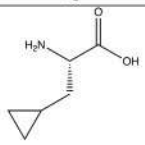
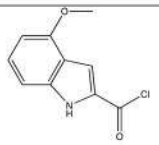
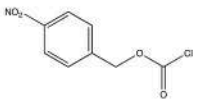
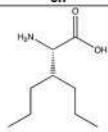
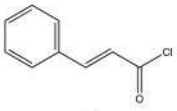
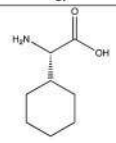
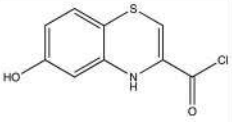
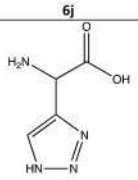
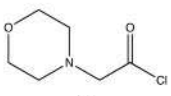
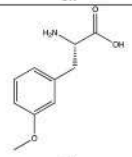
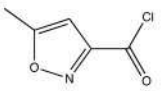
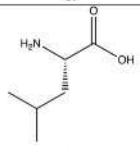
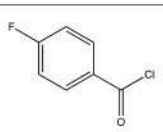
This preparation process must be repeated for each type of reagent intended for use in the combinatorial library enumeration. By carefully preparing each reagent, the method ensures that all components of the combinatorial library are structurally sound and ready for subsequent steps in the computational database design workflow, ultimately enhancing the reliability and success of the virtual screening efforts. Table 1 lists all the building blocks used in the combinatorial process.

Additionally, a core structure was prepared as an example molecule for the combinatorial process.

Once all the reagents were prepared, the third and final step in the library design was carried out using the Combinatorial Library Enumeration Panel to construct a library of compounds for subsequent virtual screening. This panel provides tools to develop a ligand library starting from a core structure to which molecular fragments are added at user-specified positions called “attachments”. Specifically, combinatorial library enumeration is performed by substituting fragments in the attachments of the core structure with building blocks from the reagent structures. Both the core and the reagents must have designated bonds to break (example in Fig. 6). Reagent fragments are then attached to the core at the site of the broken bond in a process referred to as "growing" and the bond being replaced is known as the "grow bond". These bonds are not necessarily the bonds broken and formed in the actual chemical reaction but serve to vary the substituents on a central structure.

At the end of this process, we obtained a database of 450 molecules with an aldehyde warhead for virtual screening. Supplementary Material S1, Tables S1 and S2 provide the SMILES representations for the fragments and the final obtained compounds.

**Table 1**
List of building blocks used in the combinatorial synthesis process. This includes amino acids (**6a-m**) used for P1 and P2 fragments, acyl chlorides (**7a-j**), and carbonyl chlorides (**8a-e**) employed for P3 fragments.

| P1 | P2 | P3 | P3 |
|---|---|---|---|
|  6a |  6d |  7a |  8a |
|  6b |  6e |  7b |  8b |
|  6c |  6f |  7c |  8c |
| |  6g |  7d |  8d |
| |  6h |  7e |  8e |
| |  6i |  7f | |
| |  6j |  7g | |
| |  6k |  7h | |
| |  6l |  7i | |
| |  6m |  7j | |

### 2.2. Induced Fit Docking studies

Following the construction of a database comprising 450 peptidomimetic small molecules with an aldehyde warhead, IFD studies were conducted as the first step of virtual screening to assess their potential compatibility with the active site of the SARS-CoV-2 $M^{PRO}$.

To perform this analysis, the crystallographic structure of the SARS-CoV-2 $M^{PRO}$ with PDB code 7VH8 [50] was selected. This X-ray structure was chosen for its high resolution and its co-crystallized inhibitor with a peptidomimetic core (*nirmatrelvir*, 2D structure in Fig. 7a), making it particularly relevant for the evaluation of similar compounds. We focused the docking grid on the SARS-CoV-2 $M^{PRO}$ binding pocket, including the four subsites S1', S1, S2, S3/S4 as described in the Materials and Methods section. Fig. 7b shows the 3D active binding site of SARS-CoV-2 $M^{PRO}$ in covalently bonding with *nirmatrelvir*, a second-generation orally available protease inhibitor currently in phase 3 clinical trials in combination with *ritonavir* (PAXLOVID®, see ClinicalTrials.gov identifier: NCT04960202).

The IFD studies led to the selection of 388 molecules from the initial 450, all of which exhibited IFD scores superior to that of the co-crystallized ligand. Supplementary Material S1, Table S3 shows IFD and docking scores for the selected 388 small molecules and *nirmatrelvir*. **Rigid docking was performed as a preliminary step in the IFD protocol**, allowing for an initial assessment of ligand binding poses before accounting for receptor flexibility. The final IFD scores, which include **both ligand flexibility and protein side-chain adaptations**, provide a more refined evaluation of binding interactions.

Table 2 presents ranges and averages of computed parameters (IFD and Docking scores) for the 388 optimal docked compounds, along with comparison scores for the co-crystalized and reference ligand *nirmatrelvir*.

The excluded 62 molecules were not docked by Maestro, because of structural incompatibility with the binding site of SARS-CoV-2 $M^{PRO}$. These molecules, likely due to their larger size or more rigid structural features, could not be successfully docked, suggesting that their steric or conformational properties rendered them unsuitable for fitting within the active site of $M^{PRO}$. This structural incompatibility highlights the importance of molecular size, shape, and flexibility when evaluating compounds for binding potential within a given target site. Consequently, these 62 molecules were excluded from further analysis, allowing the focus to shift to the remaining 388 compounds that exhibited favorable docking results and superior IFD scores compared to the co-crystallized ligand, *nirmatrelvir*. For all the 388 compounds we did not observe bad poses.

### 2.3. DRUDIT and Principal Component Analysis

To further evaluate the 388 molecules selected through IFD, a PCA was performed. This analysis involved calculating a set of molecular descriptors for a group of covalent inhibitors, which were previously identified through literature research and for which enzymatic inhibition assays against SARS-CoV-2 $M^{PRO}$ had been conducted. Supplementary Material S1, Table S4 shows SMILES and $IC_{50}$ for structures utilized for the PCA model building.

Specifically, the calculation of molecular descriptors was performed using the MOLecular DEScriptor Tool (MOLDESTO) software, from DRUDIT [51], a resource created by the research group to support drug discovery, which includes the developed Biotarget Predictor (BPT) and Antiproliferative Activity Predictor (APP) tools. MOLDESTO is our proprietary software, capable of generating over 1000 molecular descriptors (3D, 2D, and 1D) for each input structure [108].

The application of PCA to the structure *versus* Molecular Descriptors matrix (Supplementary Material S1, Matrix S1) revealed a total variance of 64.5 % explained by the first two principal components. The 2D plot (PC1 vs. PC2, Fig. 8) highlighted the clustering of the compounds based on their $IC_{50}$ values. Specifically, all molecules with an $IC_{50} < 0.1$ μM are represented in green, while all molecules with an $IC_{50} > 0.1$ μM are represented in red. The 0.1 μM cutoff was chosen to effectively distinguish between high- and low-affinity inhibitors. Inhibitors with $IC_{50}$ values below this threshold are generally considered potent binders. Indeed, while a more relaxed threshold (e.g., 1 μM) could allow the inclusion of additional candidates, a more stringent threshold could lead to the exclusion of potentially promising compounds. The 0.1 μM cutoff was selected to prioritize compounds with stronger predicted binding affinities for subsequent computational validation [52–54].

With the PCA model established, we then assessed the position of each of the 388 molecules in the plot relative to the red or green clusters. This evaluation involved determining whether each molecule fell into the cluster associated with $IC_{50} < 0.1$ μM (green) or the cluster with $IC_{50} > 0.1$ μM (red). By analyzing these distributions, we aimed to understand how the molecular descriptors influenced the molecules inhibitory activity against SARS-CoV-2 $M^{PRO}$.

In details, the selected small molecules were treated as described in the multi-step process below, shown in Fig. 9.

1 Calculation of Molecular Descriptors: for each of the 388 molecules, we recalculated the same set of molecular descriptors that were used to build the PCA model matrix (Supplementary Material S1, Matrix S2). This ensured that each molecule was represented consistently in terms of its chemical properties as well as molecules in the known inhibitors set;
2 Centroid Calculation: the centroids, or the central points, of the two clusters identified in the PCA plot (corresponding to the $IC_{50} < 0.1$ μM and $IC_{50} > 0.1$ μM groups) were calculated (Supplementary Material S1, Table S5). These centroids serve as reference points to evaluate the proximity of each molecule to the clusters;
3 Distance Calculation: for each molecule, we computed the distances from its position in the PCA plot to both centroids (Supplementary Material S1, Table S6). This allowed us to quantify how close each molecule was to the centers of the respective clusters;
4 Distance Ratio: the ratio between the distance from each molecule to the green centroid and the distance to the red centroid was computed (Supplementary Material S1, Table S7). This ratio provides insight into the relative proximity of each molecule to the two clusters: a smaller ratio indicates that the molecule is closer to the green cluster and farther from the red cluster;
5 Cutoff Value: a cutoff value of 0.3 was established for the ratio. Molecules with a ratio less than 0.3 are considered to be closer to the green cluster, indicating they are more similar to the high-potency inhibitors ($IC_{50} < 0.1$ μM) and relatively farther from the lower-potency inhibitors ($IC_{50} > 0.1$ μM).

By following these steps, we were able to systematically evaluate the distribution of each molecule in relation to the defined clusters and assess their potential classification based on their molecular descriptors and PCA positioning. This analysis enabled us to select 14 molecules (derivatives **9**–**22** in Fig. 10) for subsequent Covalent Docking studies.

### 2.4. Covalent docking studies

Computational insights into covalent docking are becoming increasingly crucial in addressing the challenges of selectivity and potency in covalent inhibitors. These inhibitors function through a dual mechanism: covalent bond formation between the ligand and the target protein, as well as stabilizing non-covalent interactions within the binding pocket. The CovDock protocol aids in identifying optimal covalent complexes using the well-validated Prime energy model, and calculates an apparent affinity score, which incorporates key elements of an effective covalent docking process.

For successful covalent docking, two critical conditions must be met: the pre-reactive form of the ligand must reside in the binding pocket long enough to facilitate the reaction between the ligand warhead and

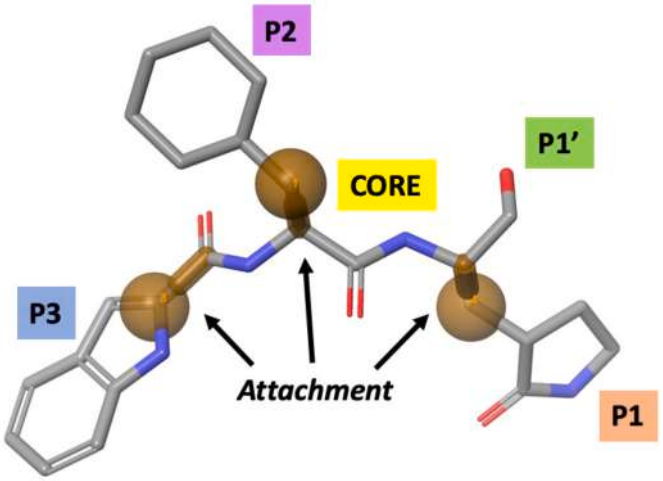



**Fig. 6.** Example of structure central core with the definition of the 3 attachments required for the process. The electrophilic warhead presents in P1′ is represented in green, while the fragments P1−P3 are shown in red, purple, and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the protein reactive residue, and the docking process must avoid unfavorable steric clashes and poor electrostatic interactions as the reaction progresses.

The distance between the active electrophilic warhead and the reactive thiol (-SH) group of $Cys^{145}$ was calculated for compounds **9**–**22** in their non-covalent complexes with SARS-CoV-2 $M^{PRO}$. The results, summarized in Table 3, reveal that all compounds exhibit a distance ranging from **3.37 to 3.85 Å**. This range suggests a spatial arrangement favorable for a potential nucleophilic attack by the thiol group on the electrophilic warhead, facilitating the formation of a covalent bond. This parameter is critical for assessing the covalent reactivity potential of these compounds, a key factor in the development of antiviral inhibitors [55–59]. All distances for compounds **9**–**22** are available in Fig. 11.

The CovDock protocol initiates by employing Glide docking on a receptor, temporarily replacing the reactive residue with alanine. Following this, the reactive residue is reinstated and sampled to form a covalent bond with the ligand, generating multiple binding poses. These covalent complexes are then minimized using the Prime energy model, selecting the top-scoring complexes. Furthermore, an apparent affinity score, derived from the Glide scores of both pre-reactive and post-reactive ligand conformations, is computed to estimate binding energies, providing an effective tool for virtual screening to assess potential inhibitors [60,61].

Compounds **9**–**22** were submitted to CovDock analysis to evaluate their capacity to covalently bind to the reactive $Cys^{145}$ within the aim, we choose a protein with **GC-373**, Fig. 12c, as co-crystalized compound

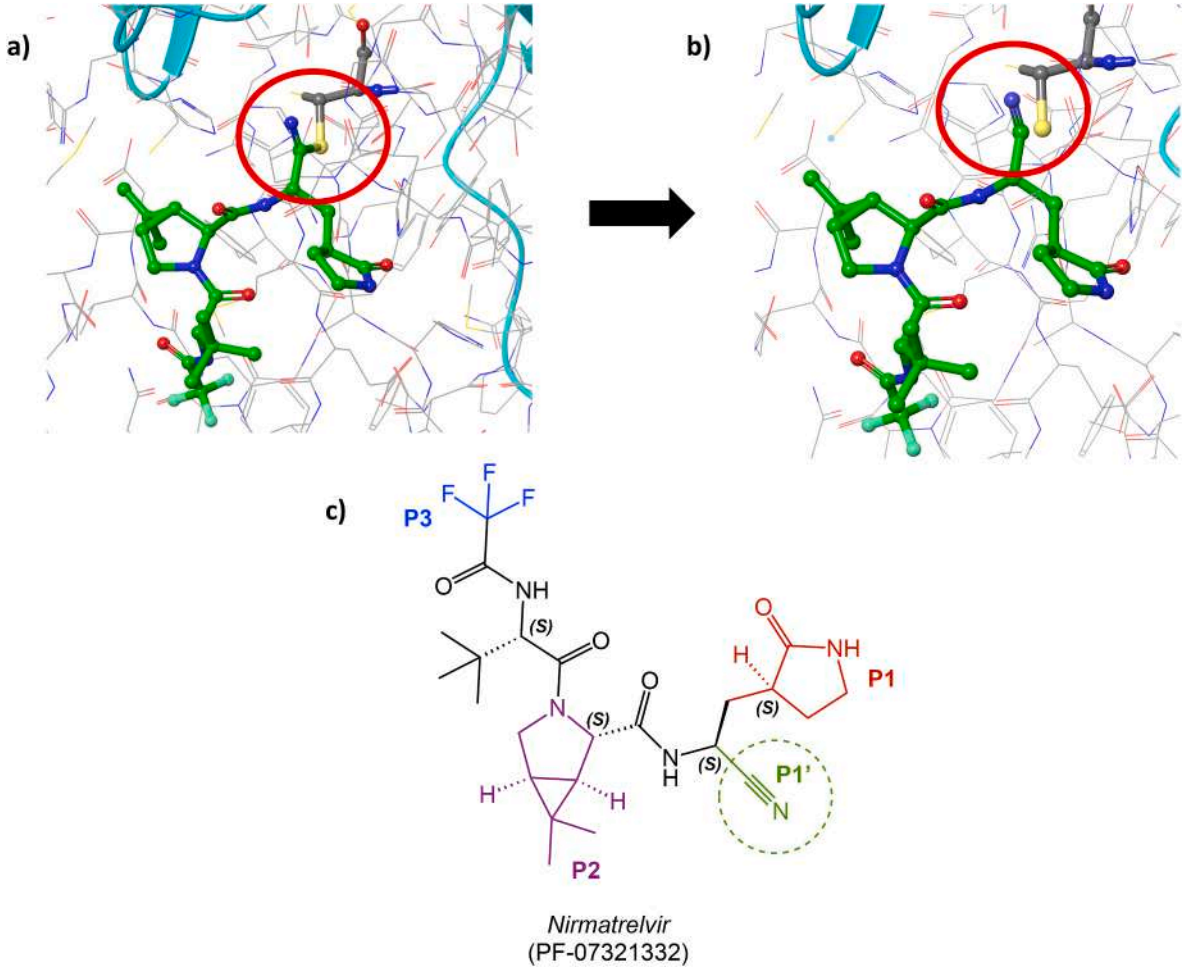



**Fig. 7.** (**a**) Representation of the covalent complex between SARS-CoV-2 $M^{PRO}$ and *nirmatrelvir*, highlighting the covalent bond between the $Cys^{145}$ residue and the ligand; (**b**) cleavage of the covalent bond between $Cys^{145}$ and *nirmatrelvir*, with the restoration of the thiol group (-SH) of the cysteine residue, simulating the unbound reactive state of $Cys^{145}$; (**c**) 2D structure of *nirmatrelvir*.

**Table 2**
Induced Fit Docking results for the selected 388 small molecules and *nirmatrelvir* against the SARS-CoV-2 $M^{PRO}$, showing ranges and averages for IFD and Docking scores.

| Target | PDB | Compound | IFD Score | Docking score |
|---|---|---|---|---|
| **SARS-CoV-2 $M^{PRO}$** | 7VH8 [51] | **Range** | −674.88 ÷ −680.75 | −7.85 ÷ −14.71 |
| | | **Average** | −676.86 | −10.77 |
| | | ***Nirmatrelvir*** | −673.14 | −10.17 |

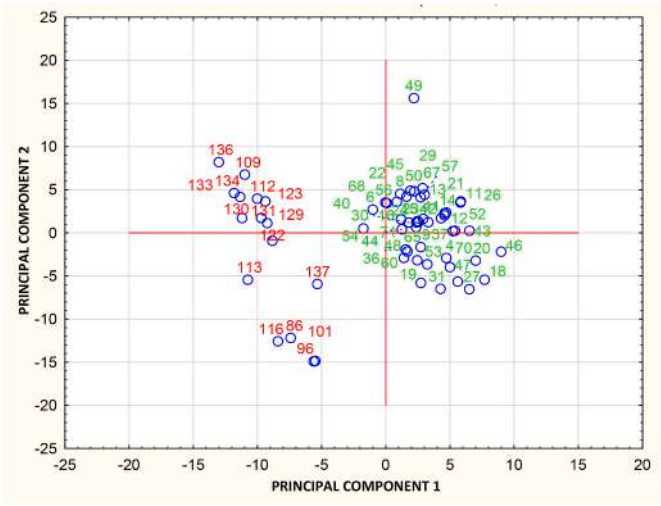



**Fig. 8.** Principal Component Analysis (PC1 *versus* PC2) applied to the molecular descriptors matrix of the selected known SARS-CoV-2 $M^{PRO}$ inhibitors.

(PDB code 6WTK [62]). In addition to generating the grid, the reaction type was specified by selecting the "Nucleophilic Addition to a Double Bond" option. Furthermore, the covalent bond between **GC-373** and $Cys^{145}$ was broken to restore the thiol group (SH) of the cysteine residue, allowing the system to simulate the unbound reactive state of $Cys^{145}$ before evaluating the binding potential of the tested compounds.

Fig. 12a illustrates the covalent $M^{PRO}$/**GC-373** complex, while Fig. 12b displays the broken covalent bond between $Cys^{145}$ and **GC-373**, highlighting the restoration of the cysteinyl -SH group.

It is important to note that compounds **18** and **22** were excluded during the CovDock calculations performed by Maestro. This exclusion could be attributed to various factors, such as the inability of these compounds to adopt a suitable geometry for covalent bond formation with $Cys^{145}$, or the failure to establish stable non-covalent interactions within the binding pocket. Additionally, steric clashes or unfavorable electrostatic interactions might have contributed to the software decision to reject these compounds as potential covalent binders.

Meanwhile, the docking scores and CovDock affinity values revealed that compound **11** exhibited the highest CovDock affinity (−9.609) and the lowest docking score (−10.213), indicating strong potential for covalent binding. Other compounds, such as **14** (−9.431), **13** (−9.230), and **16** (−8.858) also demonstrated notable affinity. In contrast, the reference compound **GC-373** showed a significantly lower affinity (−6.916) and a docking score of −6.335, suggesting that several of the identified compounds may outperform **GC-373** in binding to the target site. These results provide valuable insights into the potential of compounds **9**–**17** and **19**–**21** as effective covalent inhibitors. Docking score and CovDock affinity values for compounds **9**–**17** and **19**–**21** and the reference **GC-373** are shown in Table 4.

The analysis of 2D ligand interactions for derivatives **11**, **13**, **14**, and **16**, Figs. 13 and 14, revealed important details about their binding properties and interaction patterns with the SARS-CoV-2 $M^{PRO}$.

Once the covalent bond between the aldehydic warhead and the $Cys^{145}$ in the S1' pocket was established, the positioning of the P1, P2, and P3 moieties was examined.

The hydantoin moiety, which is recurrent in P1 of the top-ranking compounds, is capable of stabilizing hydrogen bonds with several amino acids in the S1 pocket, including $Phe^{140}$, $His^{163}$, $His^{164}$, $Glu^{166}$, $His^{172}$. Specifically, two crucial interactions were noted: the carboxylic backbone group of $Phe^{140}$ forms a hydrogen bond with the NH group of the hydantoin moiety ($Phe^{140}$-C=O—H-N-hydantoin), and the imidazole NH group of $His^{163}$ interacts with the Oxygen atom of the hydantoin moiety ($His^{163}$-N-H—O=C-hydantoin). These interactions suggest that the hydantoin moiety could mimic the Gln residue of natural substrates, providing a potential alternative to the extensively studied γ-lactam ring.

In the S2 pocket, which is mainly composed of hydrophobic amino acids such as $Met^{49}$, $Tyr^{54}$, $Met^{165}$, $Leu^{167}$, $Pro^{168}$, $Val^{186}$, $Asp^{187}$, $Arg^{188}$, there is considerable flexibility allowing for the binding of both small

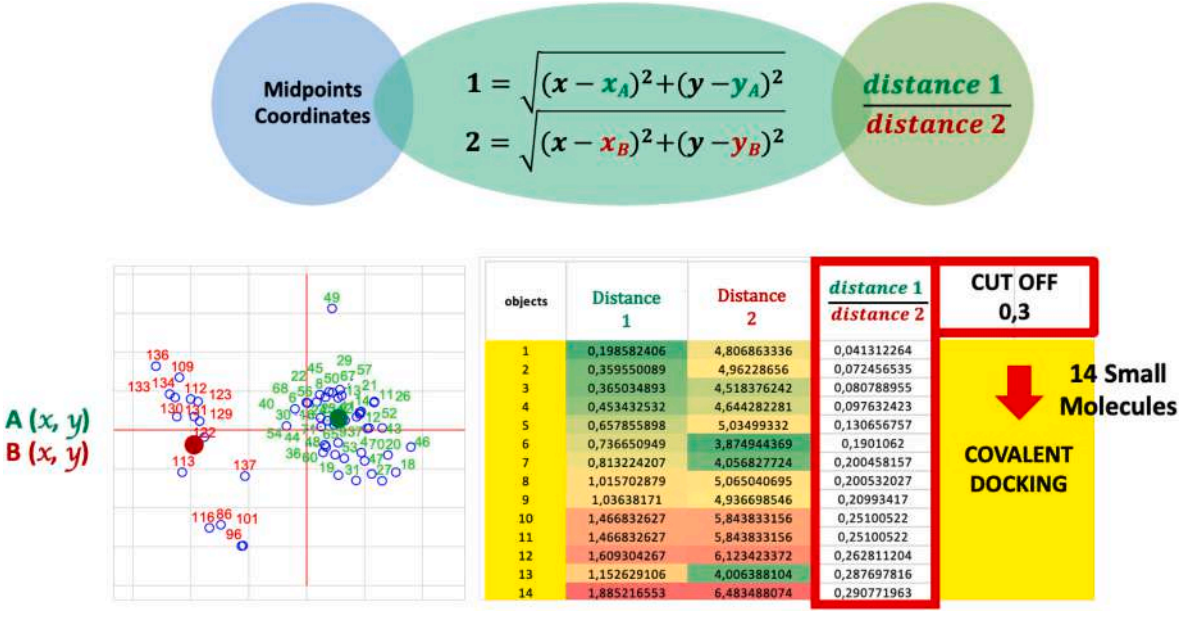



**Fig. 9.** Multi-step process for PCA data analysis.

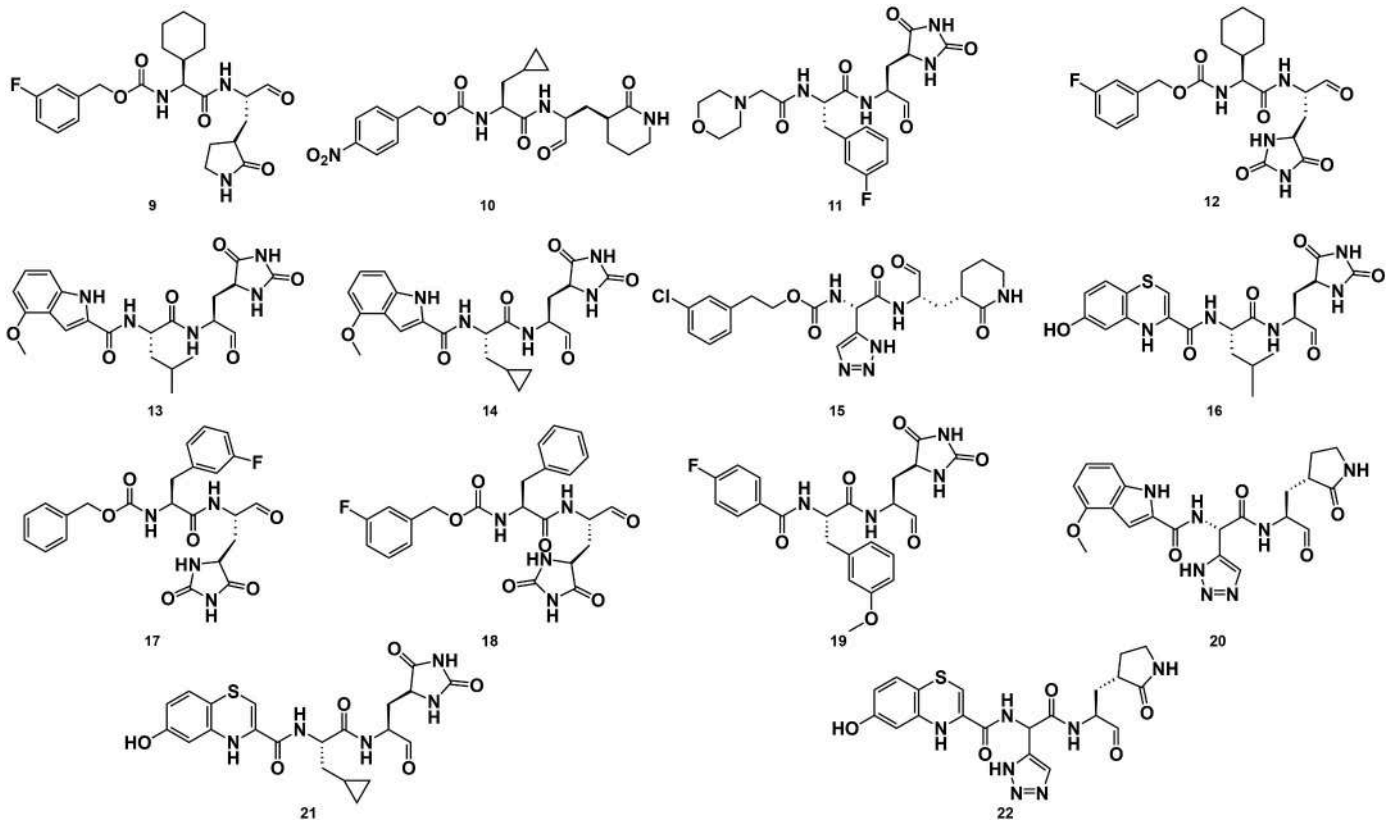


**Fig. 10.** 2D chemical structures of compounds **9**–**22**, screened through multivariate statistical analysis.

and bulky aromatic/alkyl groups. In compounds **11**, **13**, **14**, and **16**, cyclopropyl, isoleucine, and phenyl moieties are well-stabilized in the S2 subregion through hydrophobic interactions.

For interactions with the S3 subregion, compounds **13**, **14**, and **16** feature aromatic moieties in P3, while **11** a morpholine group, that interact with amino acids like Gln$^{189}$, Leu$^{167}$, Pro$^{168}$, Gly$^{251}$, and Asp$^{187}$. Overall, the identified compounds exhibit about 5–7 interactions with key residues in the binding pockets, in contrast to the reference compound **GC-373**, which shows only 2 interactions involving Glu$^{166}$ and Gly$^{143}$. This suggests that the new derivatives have a more extensive interaction profile, which may contribute to their enhanced binding efficacy.

Moreover, the analysis of the 3D binding poses for derivatives **9**–**14**, and **16**, as illustrated in Fig. 16, reveals a significant overlap among the poses. This overlap indicates a redundancy in the positioning of the key elements of these small molecules within the four sub-pockets of the binding site. The consistent positioning of these elements results in the formation of multiple interactions across the sub-pockets, suggesting a potentially similar binding mechanism or interaction pattern among these derivatives.

**Table 3**
Distances (in Å) between the electrophilic warhead of compounds **9**–**22** and the thiol group of Cys$^{145}$ in the non-covalent complexes with SARS-CoV-2 M$^{PRO}$.

| Molecule | Distance |
|---|---|
| **9** | 3.58 |
| **10** | 3.67 |
| **11** | 3.55 |
| **12** | 3.5 |
| **13** | 3.6 |
| **14** | 3.58 |
| **15** | 3.71 |
| **16** | 3.61 |
| **17** | 3.51 |
| **18** | 3.53 |
| **19** | 3.69 |
| **20** | 3.85 |
| **21** | 3.37 |
| **22** | 3.56 |

### 2.5. *ADMET properties prediction*

In drug discovery, the estimation of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) and drug-likeness parameters using *in silico* techniques is an invaluable aid to save both time and resources, increase the success rate, and consequently reduce the risks of failure in the preclinical and clinical phases [63]. To gain insight into the drug-likeness of our compounds, we decided to use the QikProp tool to calculate the #stars value [64]. This parameter indicates the number of property or descriptor values that fall outside the 95 % range of similar values computed for known drugs. Outlying descriptors and predicted properties are denoted with asterisks. A large number of stars suggests that a molecule is less drug-like than molecules with few stars. The following properties and descriptors are included in the determination of #stars: Molecular Weight, dipole, Ionization Potential (IP), Electron Affinity (EA), Solvent Accessible Surface Area (SASA), Hydrophobic Component of the SASA (FOSA), Hydrophilic Component of the SASA (FISA), π Component of the SASA (PISA), Weakly Polar Component of the SASA (WPSA), van der Waals Surface Area (PSA), Total Solvent-Accessible Volume in cubic angstroms (volume), Number of rotable bonds, Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution (donorHB), Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution (accptHB), Globularity descriptor (glob), Predicted polarizability in cubic angstroms (Qppolrz), Predicted hexadecane/gas partition

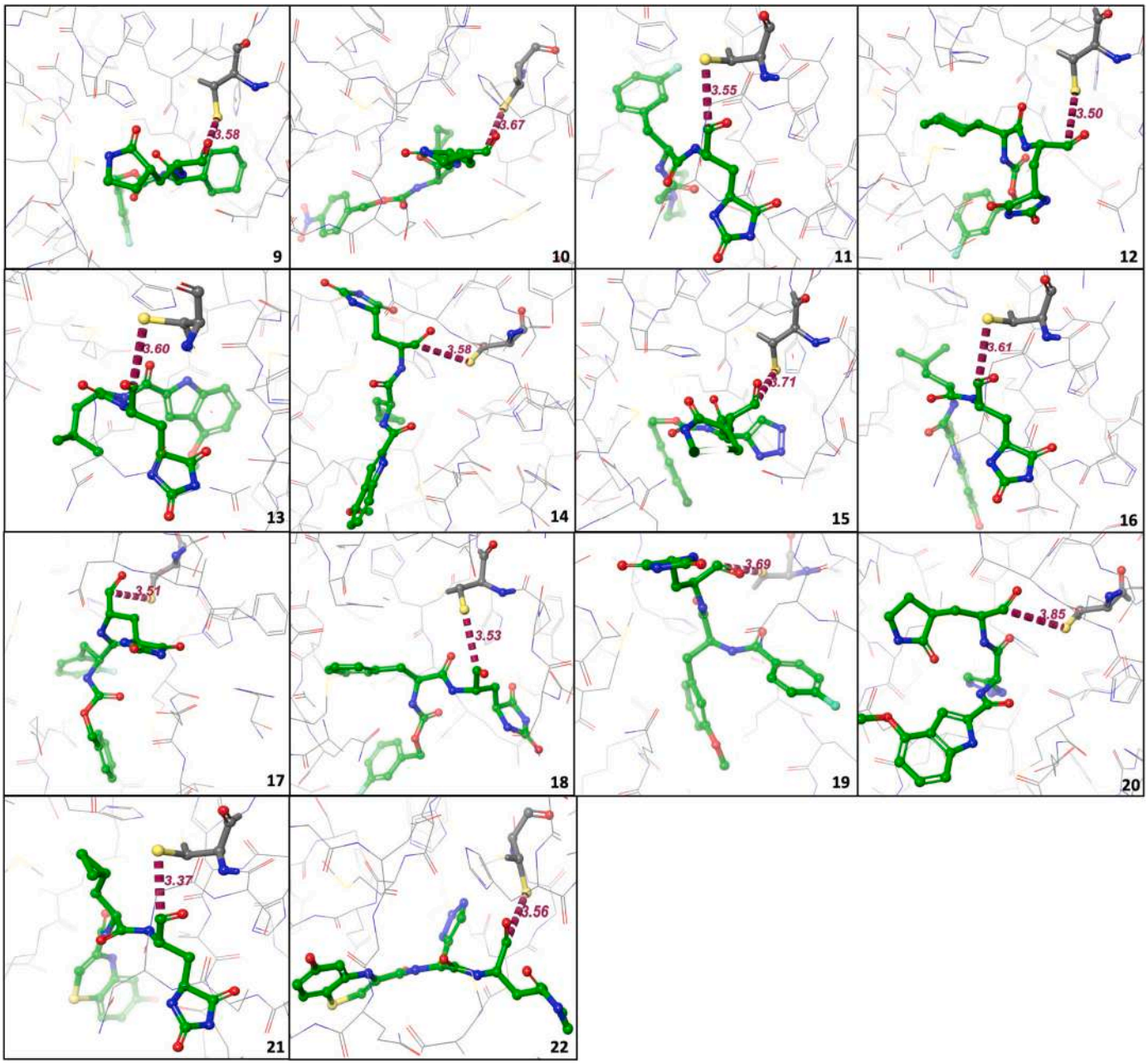



**Fig. 11.** Measured distances between the active electrophilic warhead and the reactive thiol (-SH) group of $Cys^{145}$ for compounds **9**–**22**.

coefficient (QPlogPC16), Predicted octanol/gas partition coefficient (QplogPoct), Predicted water/gas partition coefficient (QplogPw), Predicted octanol/water partition coefficient (QPlogPo/w), Predicted aqueous solubility (logS), Conformation-independent predicted aqueous solubility (QPLogKhsa), Predicted brain/blood partition coefficient (QplogBB), Number of likely metabolic reactions (metabol). For compounds **9**–**14**, **16**, **19**–**21**, *nirmatrelvir*, and GC-373 the obtained QikProp #stars value are reported in Table 5, while all the parameters, used for calculated #stars value, are reported in Supplementary Material S1, Table S8. #stars values, ranging from 0 to 4, indicate promising pharmacokinetic and physicochemical properties of our compounds, considering the recommended QikProp value $< 5$. For *nirmatrelvir* and GC-373 #stars value is 1 and 0, respectively. Table 5 also reports Predicted Central Nervous system activity (CNS), where a value of $-2$ indicates lack of activity, and Human Oral Absorption/Percent Human Oral Absorption. For these last parameters, for which 1, 2, and 3 stay down for low, medium, and high, our compounds showed great results, with high Percent Human Oral Absorption above 60 % for **9**, **12**, and **19**. As regard reference compounds, *nirmatrelvir* and GC-373, the CNS and Human Oral Absorption parameter are absolutely affinal with that of our compounds, while the Percent Human Oral Absorption is 65.684 % for *nirmatrelvir* and 53.522 % for GC-373, even lower than some compounds such as **9** and **12.**

In addition, the SwissADME tool [65] was used to consider a set of well consolidated parameters for searching bioactive compounds, such as PAINS filters, Lipinski's rules, Egan, and Muegge filters. The analysis of the data showed in Table 6 highlighted that compounds generally met the expectations in terms of bioactivity, with all registered 0 PAINS, as well as reference compounds *nirmatrelvir* and GC-373. Synthetic Accessibility for these compounds was evaluated, assigning scores from 1 (very easy) to 10 (very difficult). Compounds **13**, **14**, and **19** showed scores below 4, indicating a high likelihood of synthesis without significant difficulty. Compounds **9**–**12**, and **20** had scores lower or close to 4.30, suggesting a more challenging synthesis. Meanwhile, compounds

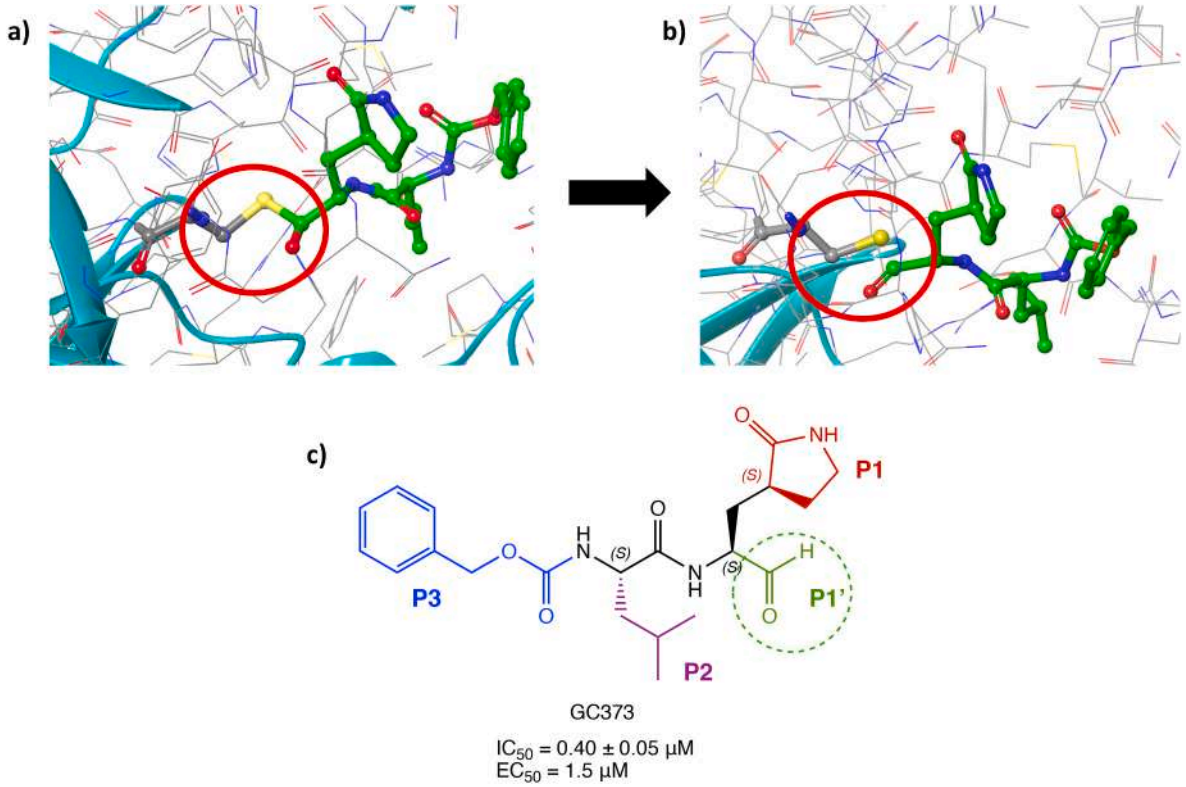



**Fig. 12.** (**a**) Representation of the covalent complex between SARS-CoV-2 $M^{PRO}$ and **GC-373**, highlighting the covalent bond between the $Cys^{145}$ residue and the ligand; (**b**) cleavage of the covalent bond between $Cys^{145}$ and **GC-373**, with the restoration of the thiol group (-SH) of the cysteine residue, simulating the unbound reactive state of $Cys^{145}$; (**c**) 2D structure of compound **GC-373**.

**Table 4**
*CovDock affinity* and *docking score* values for compounds **9**–**17**, **19**–**21**, and **GC-373**.

| Compound | CovDock Affinity | Docking Score |
|---|---|---|
| **11** | −9.609 | −10.213 |
| **14** | −9.431 | −9.531 |
| **13** | −9.230 | −8.970 |
| **16** | −8.858 | −9.746 |
| **12** | −8.849 | −9.770 |
| **10** | −8.608 | −9.566 |
| **9** | −8.520 | −9.333 |
| **21** | −8.454 | −8.727 |
| **20** | −8.156 | −8.016 |
| **19** | −7.933 | −7.905 |
| **GC-373** | −6.916 | −6.335 |
| **17** | −6.211 | −6.879 |
| **15** | −6.167 | −6.892 |

**16** and **21** exhibited the highest Synthetic Accessibility scores (4.94 and 4.86, respectively) revealing the greatest difficulty in synthesis. Synthetic Accessibility scores are available in Table 6.

Other parameters from SwissADME tool are reported in Supplementary Material S1, Table S9. Overall, results from QikProp and SwissADME tools highlight our compounds having great pharmacokinetic properties.

### 2.6. Molecular dynamics simulation

Molecular Dynamics Simulations were conducted to delve into the structural characteristics of compounds **9**–**14**, **16**, **19**–**21**, and **GC-373** in complex with $M^{PRO}$. This approach allowed us to explore the dynamic behavior of molecular systems, tracking the trajectories of individual atoms and molecules as they evolve over time.

In the first step, simulations were conducted for 50 ns for all compounds, providing a consistent timeframe to extract meaningful data. The Root Mean Square Deviation (RMSD) was computed for both the ligand and protein throughout the trajectory of simulations lasting 50 ns for each ligand-protein complex. This analysis aimed to evaluate the stability and convergence of the simulations by measuring the average change in displacement of the backbone from a reference frame at t = 0. According to the RMSD analysis, depicted in Fig. 17, almost all compounds assess a RMSD comparable to the reference ligand **GC-373**. The RMSD describes the structural similarity for all frames compared to the starting structure of the simulation (reference frame at t = 0). A stabilized RMSD value that only fluctuates around a mean value confirms the equilibration of the system and the convergence of the simulation. Fluctuations in the order of 1–3 Å are expected for small globular proteins, while larger fluctuations may indicate conformational changes.

In our simulations, both the protein and ligand are sufficiently equilibrated, and no conformational changes occur. Naturally, the protein side chains show higher RMSD values because they are more mobile compared to the protein backbone. The ligand RMSD aligned to the ligand itself shows the internal fluctuations of the ligand structure, which do not indicate any conformational changes here, particularly for compound **13**, **14**, and **16**, whit even better results than **GC-373**.

Only for compounds **19** and **21** we obtained bad results, with the RMSD reaching peaks of 10.5 and 12 Å, respectively. Therefore, we decided to exclude them for the next studies. Indeed, compounds **9**–**14**, **16**, and **20** were selected for further longer dynamics simulations, alongside the reference compound **GC-373**. The RMSD analysis was extended over a longer timescale to further verify the stability of the interactions and identify any structural variations not apparent in the initial 50 ns simulations. This extension allowed for observing the long-term dynamics of the selected compounds. Fig. 19 shows RMSDs for compounds **9**–**14**, **16**, and **20** over a simulation time of 200 ns.

The selected compounds demonstrated structural convergence and

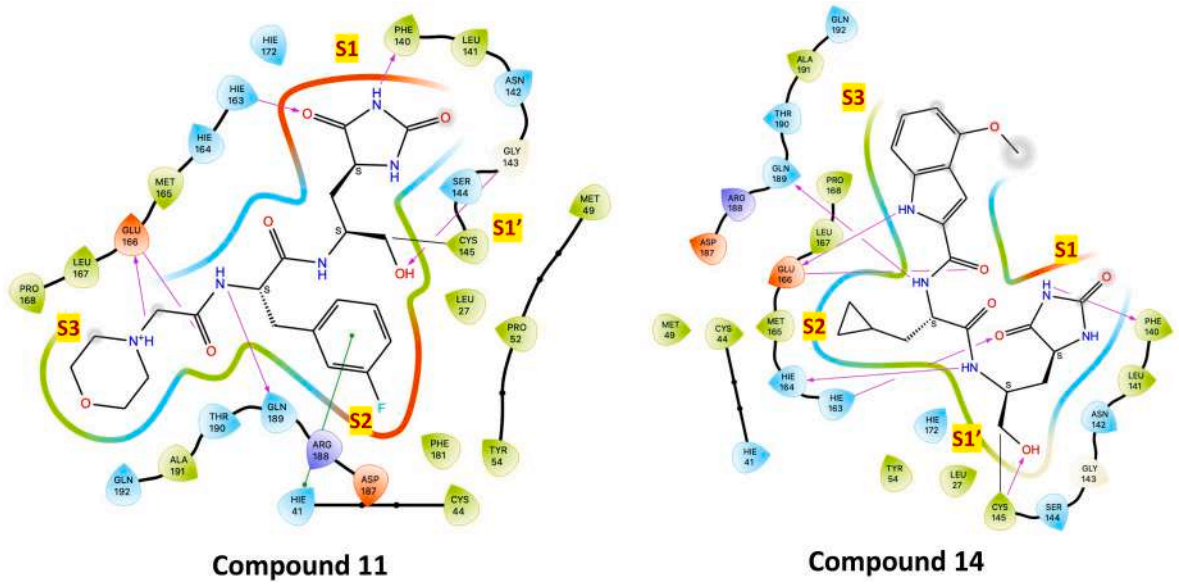



**Fig. 13.** 2D ligand interaction maps for compounds **11**, and **14**.

compatibility within the $M^{PRO}$ binding site. Their robust RMSD profiles indicate that they can maintain strong and stable interactions with $M^{PRO}$, like or better than the reference compound **GC-373**. This highlights the potential of compounds **9**–**14**, **16**, and **20** as covalent inhibitors and the importance of assessing them as therapeutic candidates in further in-depth studies. Subsequently, a range of structural parameters was computed for each complex, over 200 ns, to offer a comprehensive analysis of their molecular features. This included the calculation of the Ligand Root Mean Square Fluctuation (L-RMSF), and Protein Ligand contacts. Radius of Gyration (rGyr), which assesses the "extendedness" of the ligand and correlates with its principal moment of inertia, Intramolecular Hydrogen Bonds (intraHB), Molecular Surface Area (MolSA), Solvent Accessible Surface Area (SASA), and Polar Surface Area (PSA) were also analyzed and provided in Supplementary Material2, Figs. S1–9.

The L-RMSF serves as a crucial metric for tracking changes in the positions of ligand atoms during molecular dynamics simulations. During our simulations, L-RMSF analysis was conducted for compounds **9**–**14**, **16**, **20**, and **GC-373**. The reference time ($t_{ref}$) was set to the first frame, establishing it as the zero point in time. The L-RMSFs, illustrated in Fig. 20, offer insights into the local flexibility of individual atoms. The ligand is aligned on its reference frame and the RMSF is calculated for heavy atoms; 2D representation of the molecules is provided over each plot to show the atom number, name, and type.

Here, atoms with a large RMSF are the fluorine atoms in compounds **9**, **11**, and **12**, which are able to rotate freely around the C-C bonds. All other atoms have small RMSF values, ranging from 0 to 1.5 Å.

The dynamics of protein–ligand interactions in the complexes were analyzed using molecular dynamics simulations. Fig. 21 depicts plots of protein–ligand contacts and explains the interaction fraction of the protein residue with the ligand, which explains how a specific interaction is maintained between ligand and receptors complexes during the simulation time. In detail, Fig. 20a–i shows protein–ligand contacts for compounds **9**–**14**, **16**, **20**, and **GC-373** in complex with the SARS-CoV-2 $M^{PRO}$. Notably, all compounds showed highest interaction fractions with $Phe^{140}$, $Asn^{142}$, $Gly^{143}$, $Cys^{145}$, $His^{163}$, $His^{164}$, $Glu^{166}$, $Glu^{189}$, and $Glu^{192}$ throughout the simulation time, while, the reference compounds, **GC-373**, was able to maintain interactions in good fractions just with $Gly^{143}$, $Cys^{145}$, $Glu^{166}$, and $Glu^{189}$.

Moreover, the analysis of the 2D interaction maps (Fig. 22) reveals the intricate interaction network of compounds **9**–**14**, **16**, **20**, with high stabilization of the covalent $Cys^{145}$/C=O bond through hydrogen bonding.

The Molecular Mechanics with Generalized Born and Surface Area (MM-GBSA) analysis was conducted to evaluate the free binding energy between compounds **9**–**14**, **16**, **20**, and **GC-373** and the SARS-CoV-2 $M^{PRO}$. It uses an implicit solvent model to capture the effect of desolvating ligand and receptor. For this calculation, we use the VSGB water model to simulate solvation. Comparing the energy values (Table 7) for compounds **9**–**14**, **16**, and **20** with the **GC-373**/$M^{PRO}$ complex reveals insights into the relative binding affinities of the ligands. Indeed, all compounds showed lower energy than the reference ligand. In details, compounds **9**–**14**, **16**, **20** exhibits: Prime Coulomb Energy ranging from −10122.554 to −10103.552, Prime vdW Energy ranging from −1455.405 to −1441.063, Prime Solvent Energy ranging from −1051.173 to −10103.552, Prime Hbond Energy ranging from −102.241 to −101.868, and Prime Energy ranging from −13389.7 to −13360.6; while **GC-373** demonstrates Prime Coulomb Energy (−10103.535), Prime vdW Energy (−1439.000), Prime Solvent Energy (−1036.399), Prime Hbond Energy (−101.216), and Prime Energy (−13350.0), always lower than the analyzed compounds. Furthermore, it is noteworthy that the energy gap between the best-docked compound (compound **11**) and GC-373 is precisely 3.877 kcal/mol, while the Prime Energy gap between the lowest value (for compound **16**) and GC-373 is precisely 39.7 kcal/mol.

Therefore, compounds **9**–**14**, **16**, and **20** appear to keep being the best in all the studied simulations.

## 3. Materials and Methods

### 3.1. *In silico studies*

#### 3.1.1. *Ligand-based studies*

The DRUDIT web-service operates through four servers, each of which can perform more than ten jobs simultaneously, and several software modules implemented in C and JAVA running on MacOS Mojave.

*3.1.1.1. MOLDESTO: a new software for molecular descriptors calculation.* MOLDESTO, as we described in a previous work [51], is a software tool implemented in DRUDIT that represents the evolution of the expertise of the research group in the calculation/manipulation of molecular descriptors [158]. It is currently able to calculate more than 1000 1D, 2D, and 3D molecular descriptors for each input structure. The

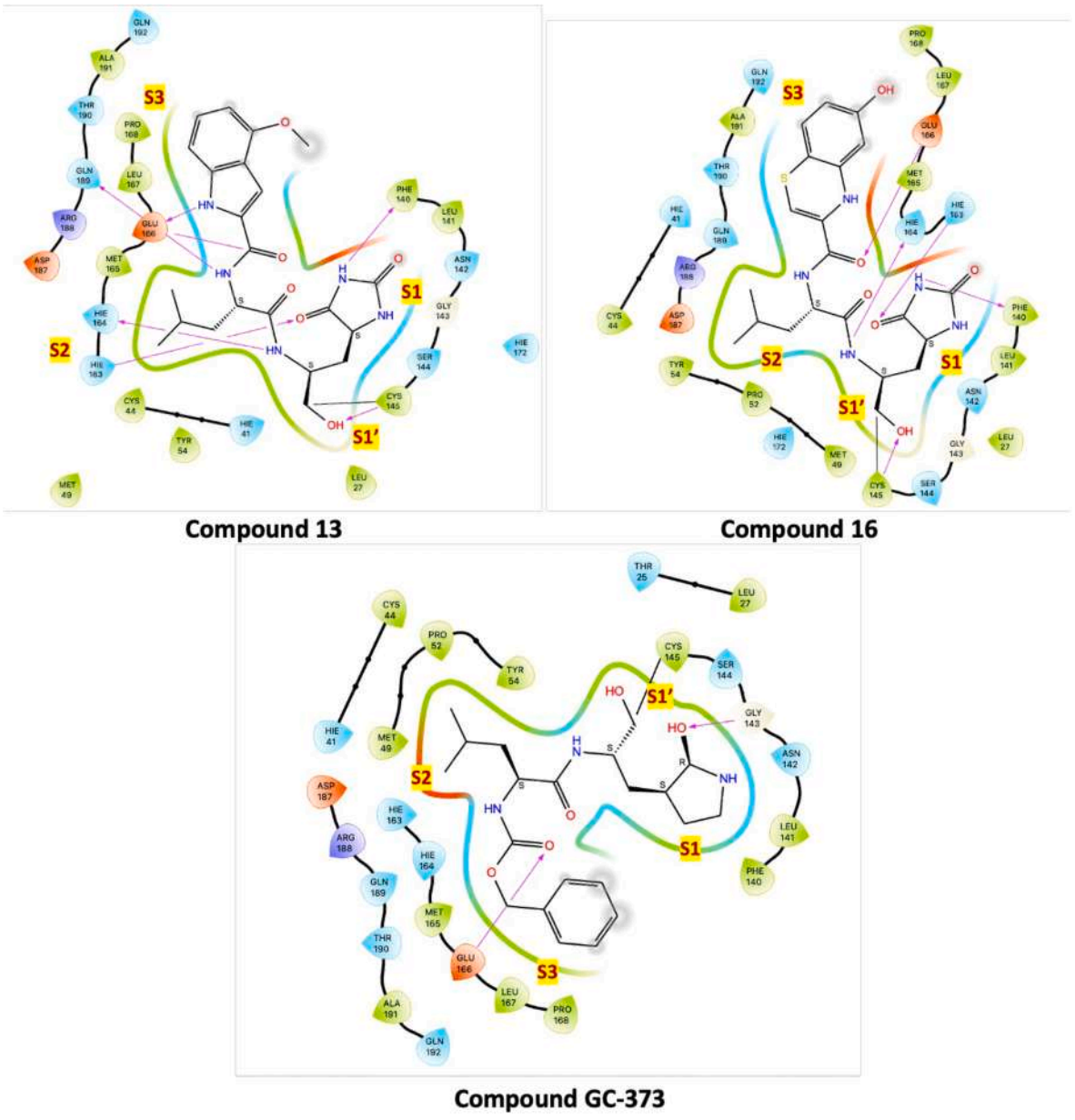



**Fig. 14.** 2D ligand interaction maps for compounds **13**, **14**, and **GC-373**.
The same trend is observed for compounds **9**, **10**, and **12**, as shown in the 2D ligand interaction maps reported in Fig. 15 while compounds **15**, **17**, **19**–**21** demonstrated a lower number of interactions within the SARS-CoV-2 $M^{PRO}$ binding site.

software is provided with a caching system to boost the calculation speed of previously submitted structures.

*3.1.1.2. Combiglide: design of a combinatorial library.* The combinatorial ligand library was generated using **CombiGlide** to explore potential chemical space efficiently. After the design of a retrosynthetic scheme, the following steps were involved in the library construction.

*3.1.1.3. Reagent Preparation Panel.* In this panel it is possible to prepare reagent files for use in combinatorial library enumeration. Reagent preparation ensures that the input structures are all-atom, 3D structures, and that that have the appropriate information stored with them to construct the molecules that are used in screening or are added to the library. Additionally, the main task in this panel is to select a reagent type (a functional group) and to identify the bond in the functional group that is replaced when the reagent is added to the core. A reagent preparation job for each reagent type was run.

*3.1.1.4. Combinatorial Library Enumeration Panel.* In this tab, the core-containing molecule was set up by importing a core-containing molecule and defining its attachment bonds.

*3.1.2. Structure-based studies*

The preparation process of ligands and protein-ligand complexes used for *in silico* studies has been performed as detailed below.

*3.1.2.1. Ligand Preparation.* The ligands for docking were prepared through the LigPrep tool, available in the Maestro Suite, Schrödinger software [66]. For each ligand, all possible tautomers and stereoisomers were generated for a pH of 7.0 ± 0.4, using default setting, through the Epik ionization method [67]. Consequently, the integrated Optimized Potentials for Liquid Simulations (OPLS) 2005 force field was used to minimize the energy status of the ligands [68].

*3.1.2.2. Protein Preparation.* The crystal structures of SARS-CoV-2 $M^{PRO}$

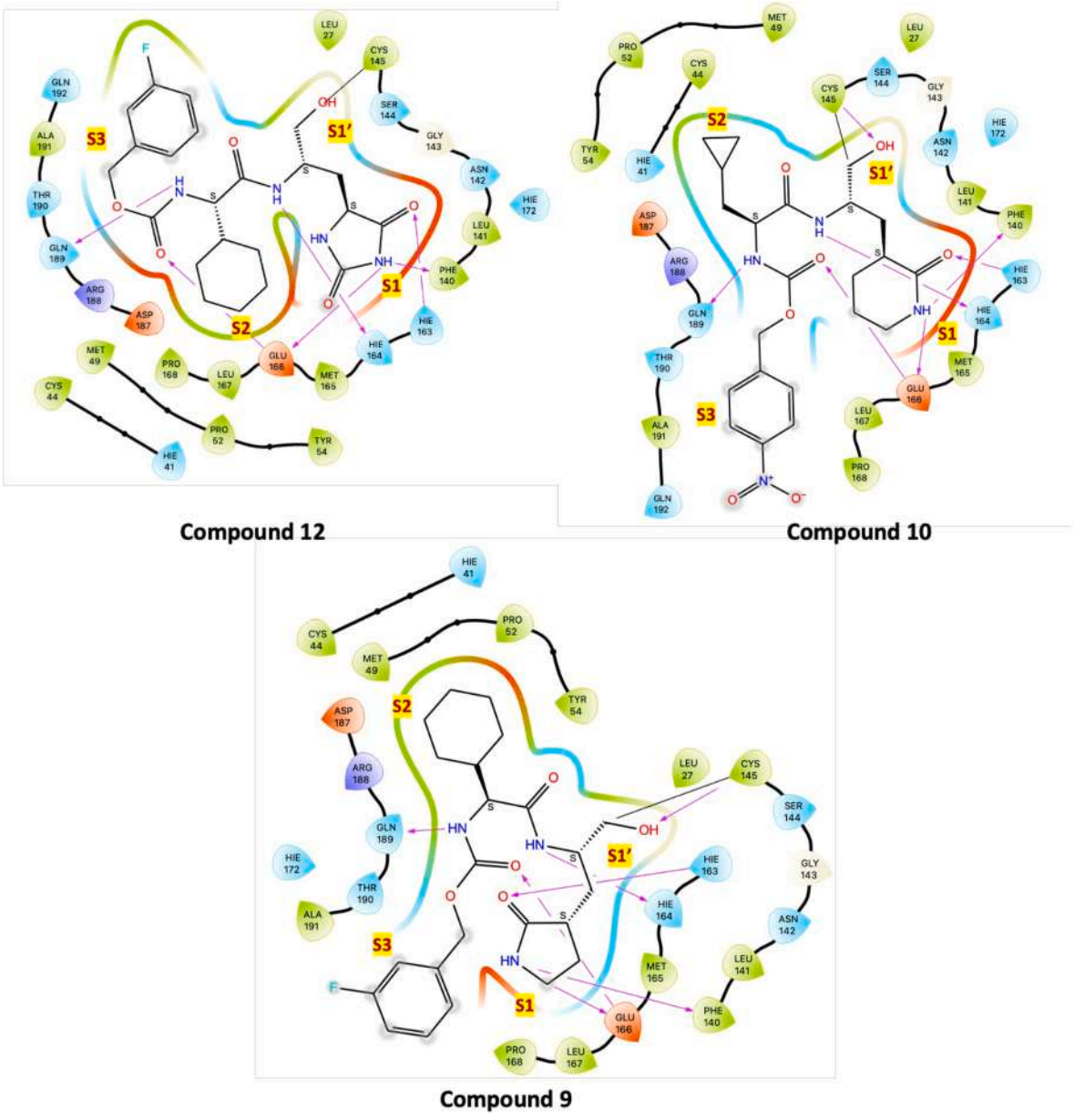



**Fig. 15.** 2D ligand interaction maps for compounds **9**, **10**, and **12**.

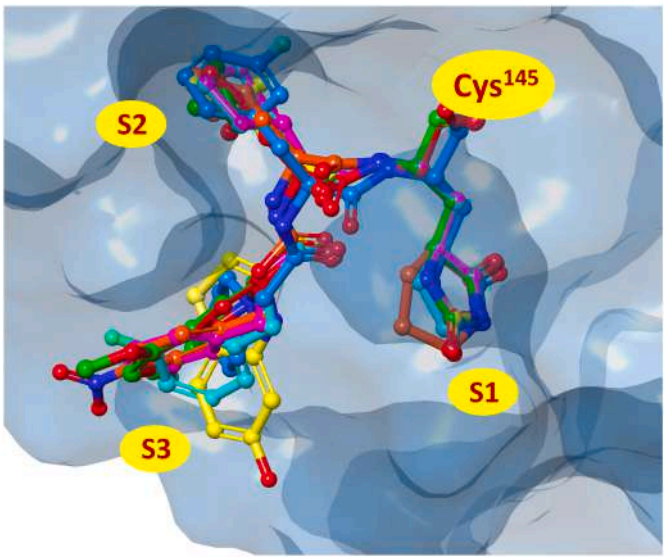



**Fig. 16.** 3D overlaps of compounds **9**–**14**, and **16** at the SARS-CoV-2 $M^{PRO}$ binding site.
Through a powerful blend of ligand- and structure-based approaches, supported by rigorous statistical analysis, this protocol has successfully identified new covalent peptidomimetic inhibitors with aldehydic warheads, demonstrating significant potential as SARS-CoV-2 $M^{PRO}$ inhibitors.

**Table 5**
Key predicted ADME properties calculated through QikProp tool.

| Compound | QikProp Stars | CNS | Human Oral Absorption | Percent Human Oral Absorption |
|---|---|---|---|---|
| **9** | 0 | −2 | 3 | 74.973 % |
| **10** | 1 | −2 | 2 | 34.858 % |
| **11** | 0 | −2 | 2 | 20.987 % |
| **12** | 0 | −2 | 3 | 66.295 % |
| **13** | 2 | −2 | 2 | 44.569 % |
| **14** | 2 | −2 | 2 | 39.569 % |
| **16** | 3 | −2 | 2 | 18.659 % |
| **19** | 0 | −2 | 2 | 61.704 % |
| **20** | 2 | −2 | 2 | 21.473 % |
| **21** | 4 | −2 | 2 | 17.108 % |
| ***Nirmatrelvir*** | 1 | −2 | 2 | 65.684 % |
| **GC-373** | 0 | −2 | 2 | 53.522 % |

(PDB codes 7VH8 [50] and 6WTK [62]) were downloaded from the Protein Data Bank [69,70]. A first breakup of the covalent bound between the cocrystal ligands and the $Cys^{145}$ was carried out. Successively, the proteins were prepared using the Protein Preparation Wizard, in the Schrödinger software, with the default setting [71]. In detail, bond orders were assigned, including Het group, hydrogen atoms were added, all water molecules were delated, and protonation of the heteroatom states were carried out using the Epik-tool (with the pH set at biologically relevant values, i.e., at 7.0 ± 0.4). The H-bond network was then optimized. The structure was finally subjected to a restrained energy minimization step (RMSD of the atom displacement for terminating the minimization was 0.3 Å), using the OPLS 2005 force field [68].

*3.1.2.3. Docking validation.* Molecular Docking studies were executed and scored by using the Glide module, available in the Schrödinger Suite program package. The receptor grids were obtained through assignment the original ligands (***nirmatrelvir*** and **GC-373** for PDB codes 7VH8 [50] and 6WTK [62], respectively) as the centroid of the grid boxes. Extra Precision (XP) mode, as scoring function, was used to dock the generated 3D conformers into the receptor model. The post-docking minimization step was performed with a total of 5 poses for each ligand conformer, and a maximum of 2 docking poses were generated per ligand conformer. The proposed docking procedure was able to re-dock the original ligands within the receptor-binding pockets with RMSD< 0.51 Å [68].

*3.1.2.4. Induced Fit Docking.* Induced Fit Docking simulation was performed using the IFD application, an accurate and robust Schrödinger technology that accounts for both ligand and receptor flexibility [72,73]. Schrödinger Induced Fit Docking validated protocol was applied by using SARS-CoV-2 $M^{PRO}$ protein from the PDB (PDB code 7VH8 [50]), previously refined by the Protein Preparation module. The IFD score (IFD score = 1.0 Glide Gscore + 0.05 Prime Energy), which includes protein–ligand interaction energy and system total energy, was calculated and used to rank the IFD poses. The more positive in modulus was the IFD score, the more favorable was the binding.

*3.1.2.5. Covalent docking.* Covalent Docking was carried out to dock a set of peptidomimetic ligands within the SARS-CoV-2 $M^{PRO}$ binding site (PDB code 6WTK [62]). Unlike classical docking, where interactions are mainly non-covalent, covalent docking requires specific modeling of the chemical reaction process occurring between a functional group of the ligand and a reactive residue on the protein.

The Prime Energy VSGB 2.0 (Variable-Scale Generalized Born) model was used for selecting the top covalent complexes, an advanced energy model designed for high-resolution protein structure modeling, capable of accurately minimizing complexes energy.

The covalent docking process was performed using Maestro Schrödinger, which incorporates an algorithm capable of handling both the pre-covalent phase (geometric alignment and orientation of the ligand in the active site) and the post-covalent phase (simulating the formation of the chemical bond).

The workflow adopted includes the following main steps.

- **Identification of the reactive residue:** the binding site and the reactive residue ($Cys^{145}$) were identified in the SARS-CoV-2 $M^{PRO}$ binding site (PDB code 6WTK [62]).

**Table 6**
Key predicted ADME properties calculated through SwissADME tool.

| Compound | Lipinski #violations | Ghose #violations | Egan #violations | Muegge #violations | Total #violations | PAINS #alerts | Synthetic Accessibility |
|---|---|---|---|---|---|---|---|
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 4.29 |
| **10** | 1 | 0 | 1 | 1 | 3 | 0 | 4.31 |
| **11** | 1 | 1 | 1 | 0 | 3 | 0 | 4.11 |
| **12** | 0 | 0 | 1 | 0 | 1 | 0 | 4.33 |
| **13** | 1 | 0 | 1 | 1 | 3 | 0 | 3.99 |
| **14** | 1 | 1 | 1 | 1 | 4 | 0 | 3.94 |
| **16** | 2 | 2 | 1 | 2 | 7 | 0 | 4.94 |
| **19** | 0 | 0 | 1 | 0 | 1 | 0 | 3.88 |
| **20** | 1 | 1 | 1 | 1 | 4 | 0 | 4.18 |
| **21** | 2 | 1 | 1 | 2 | 6 | 0 | 4.86 |
| ***Nirmatrelvir*** | 0 | 1 | 1 | 0 | 0 | 0 | |
| **GC-373** | 0 | 0 | 1 | 0 | 0 | 0 | |

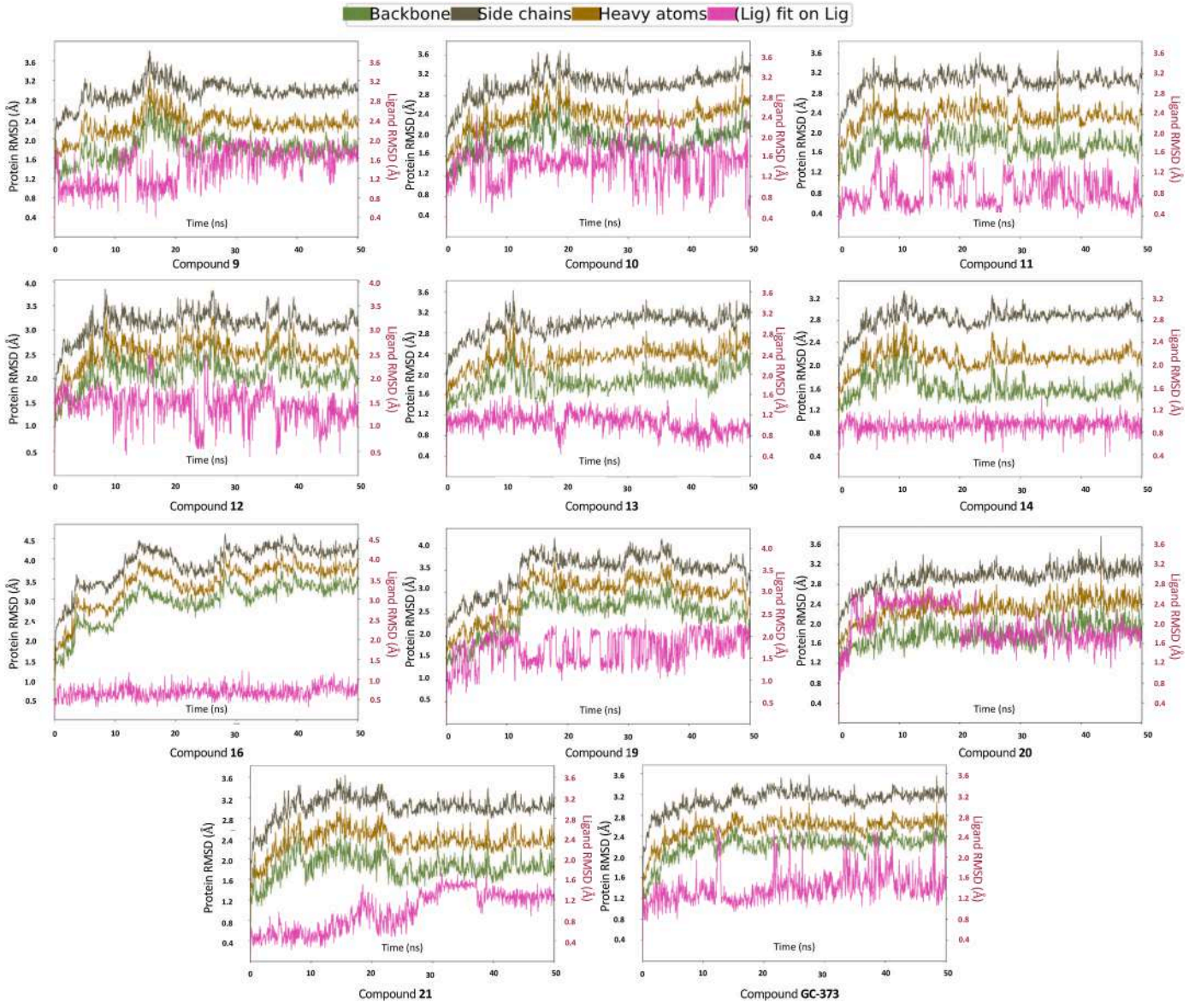



**Fig. 17.** Calculated RMSD for protein backbone, side chains, heavy atoms and ligand aligned to the ligand itself for compounds **9**–**14**, **16**, **19**–**21**, and **GC-373** during the simulation trajectory of 50ns.

The ligand RMSD aligned to the protein, in Fig. 18, shows the movements of the ligands relative to the protein. Ligands do not move significantly relative to the protein over the simulation time.

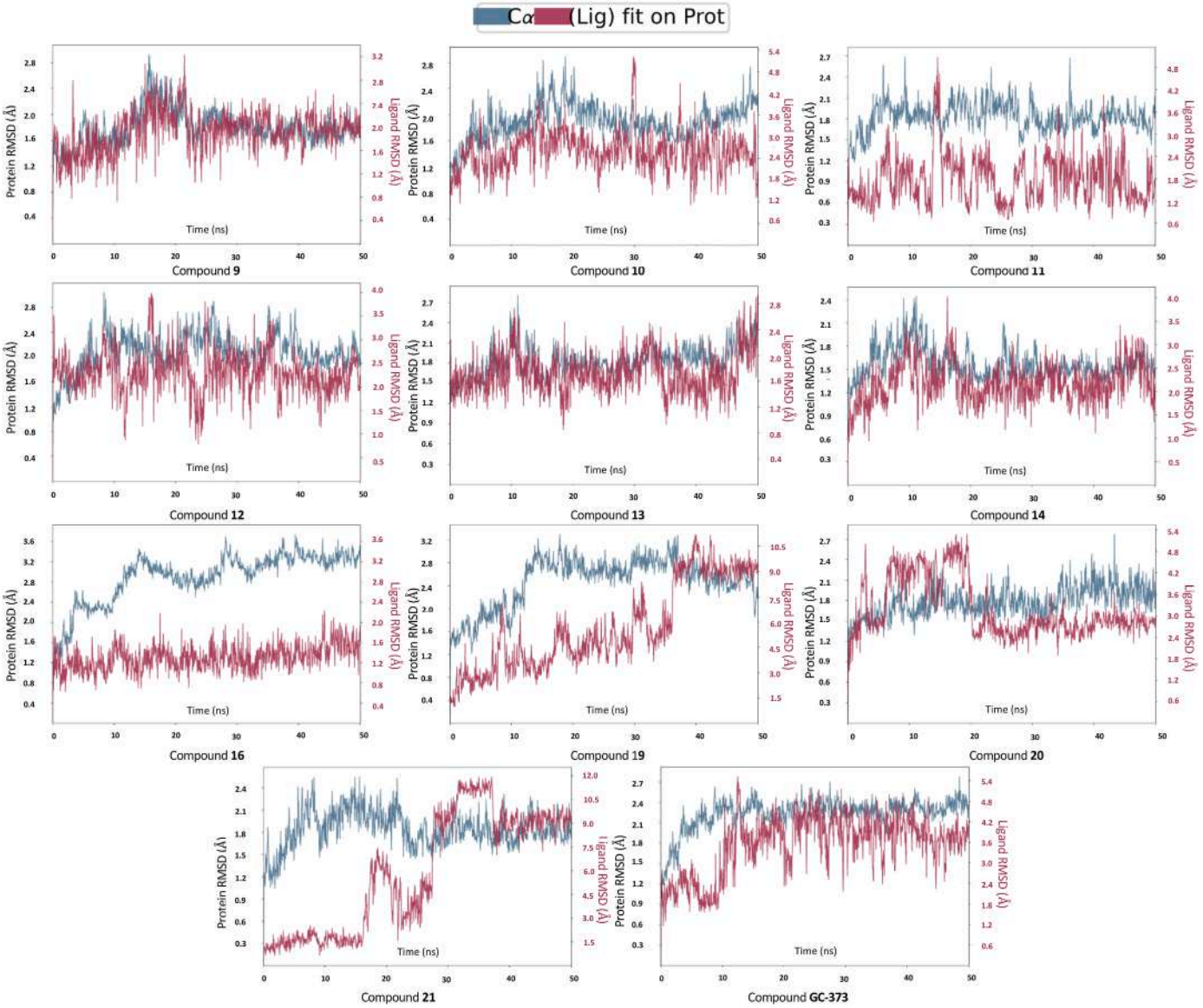



**Fig. 18.** Calculated RMSD for protein C-alphas and ligand aligned to the protein for compounds **9**–**14**, **16**, **19**–**21**, and **GC-373** during the simulation trajectory of 50ns.

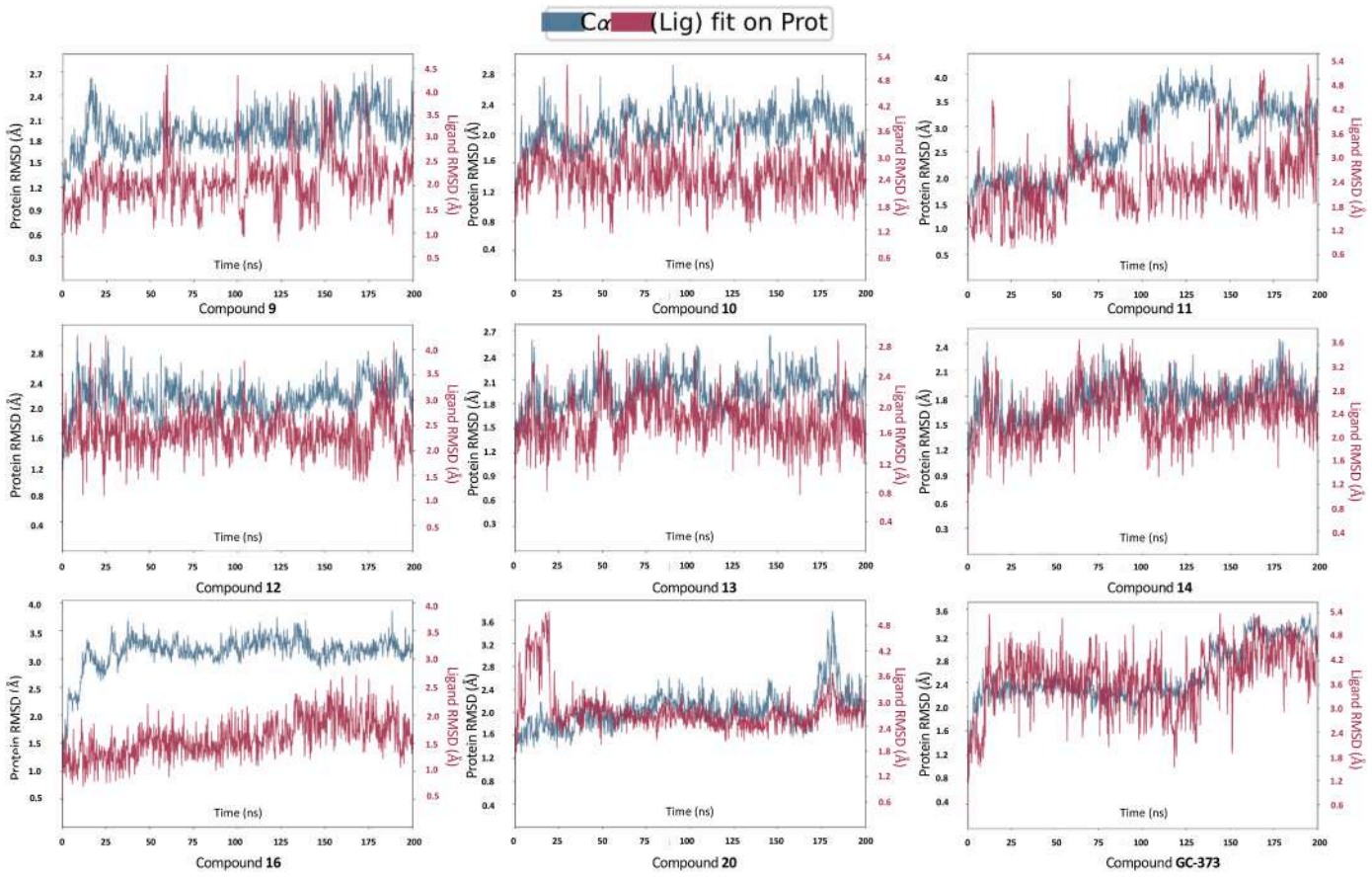



**Fig. 19.** Calculated RMSD for protein C-alphas and ligand aligned to the protein for compounds **9**–**14**, **16**, **20**, and **GC-373** during the simulation trajectory of 200ns.

- **Reaction Type selection:** the reaction type for the formation of covalent bond between ligand and receptor was selected. The reaction type defines the functional group on the ligand and the operations that must be performed on the structures to create the covalent bond (Nucleophilic Addition to a Double Bond).
- **Pre-covalent docking**: a preliminary docking was performed to optimally position the ligand in the binding site before covalent bond formation. The interaction energy was evaluated using an algorithm based on a scoring function that considers non-covalent interactions.
- **Covalent bond formation**: a chemical model was applied to simulate the formation of the covalent bond between the nucleophilic group of the reactive residue and the ligand.
- **Evaluation of the protein-ligand complex**: the stability of the covalent protein-ligand complex was assessed through a scoring function specific to covalent docking, accounting for both covalent bond energy and residual non-covalent interactions. Additionally, a short molecular dynamics simulation was performed to further validate and optimize the geometry of the final complex.

*3.1.2.6. Molecular dynamics simulation.* To assess the stability and binding affinity of compounds **9**–**14**, **16**, **20**, and **GC-373** in complex with the SARS-CoV-2 $M^{PRO}$, Molecular Dynamics Simulations (MDS) were conducted using the Desmond software. The simulations were performed under the constant-temperature–constant-pressure ensemble (NPT), allowing precise control over both temperature and pressure conditions. Pressure adjustments within the NPT ensemble were achieved by modifying the volume, while the unit cell vectors were allowed to change. Simulation parameters were configured with a system temperature of 300 K and a pressure of 1013.25 bar. The Nose-Hoover Chain thermostat and the Martyna-Tobias-Klein barostat were settled with a relaxation time of 1ps, and 2ps, respectively. Before commencing the production run, the systems underwent energy minimization for 1000 steps to establish a stable starting point. Subsequently, a preliminary production run of 50ns was conducted for compounds **9**–**14**, **16**, **20**, and **GC-373** in complex with the SARS-CoV-2 $M^{PRO}$. The simulation results were meticulously analyzed to observe the Root Mean Square Deviation (RMSD) of protein and ligands. A further longer simulation of 200 ns was performed, allowing to extract parameters like RMSD, Ligand Root Mean Square Fluctuation (L-RMSF), Protein Ligand contacts, Radius of Gyration (rGyr), Intramolecular Hydrogen Bonds (intraHB), Molecular Surface Area (MolSA), Solvent Accessible Surface Area (SASA), and Polar Surface Area (PSA). This comprehensive analysis offers crucial insights into the dynamic behavior of the complexes, providing valuable information about their structural integrity and the nature of molecular interactions over the simulation period.

*3.1.2.7. MM-GBSA analysis.* The Molecular Mechanics with Generalized Born and Surface Area (MM-GBSA) calculations helps in calculating binding free energies of the protein-ligand interaction, that includes interaction free energies such as van der waals, electrostatic, polar salvation, Hbond, and binding energies. In this study we adopted MM-GBSA in single step calculation method by using the VSGB as Solvation model, and OPLS-2005 as force field.

### 3.2. Multivariate statistical analysis

PCA, one of the most-widely used multivariate exploratory techniques, enables a drastic dimensionality reduction of an original raw data, transforming the original matrix to a new one, whose set of variables, termed as Principal Components (PCs), appear to be ordered with descending importance in terms of variance. Principal Components Analysis can be highly useful for data classification and pattern

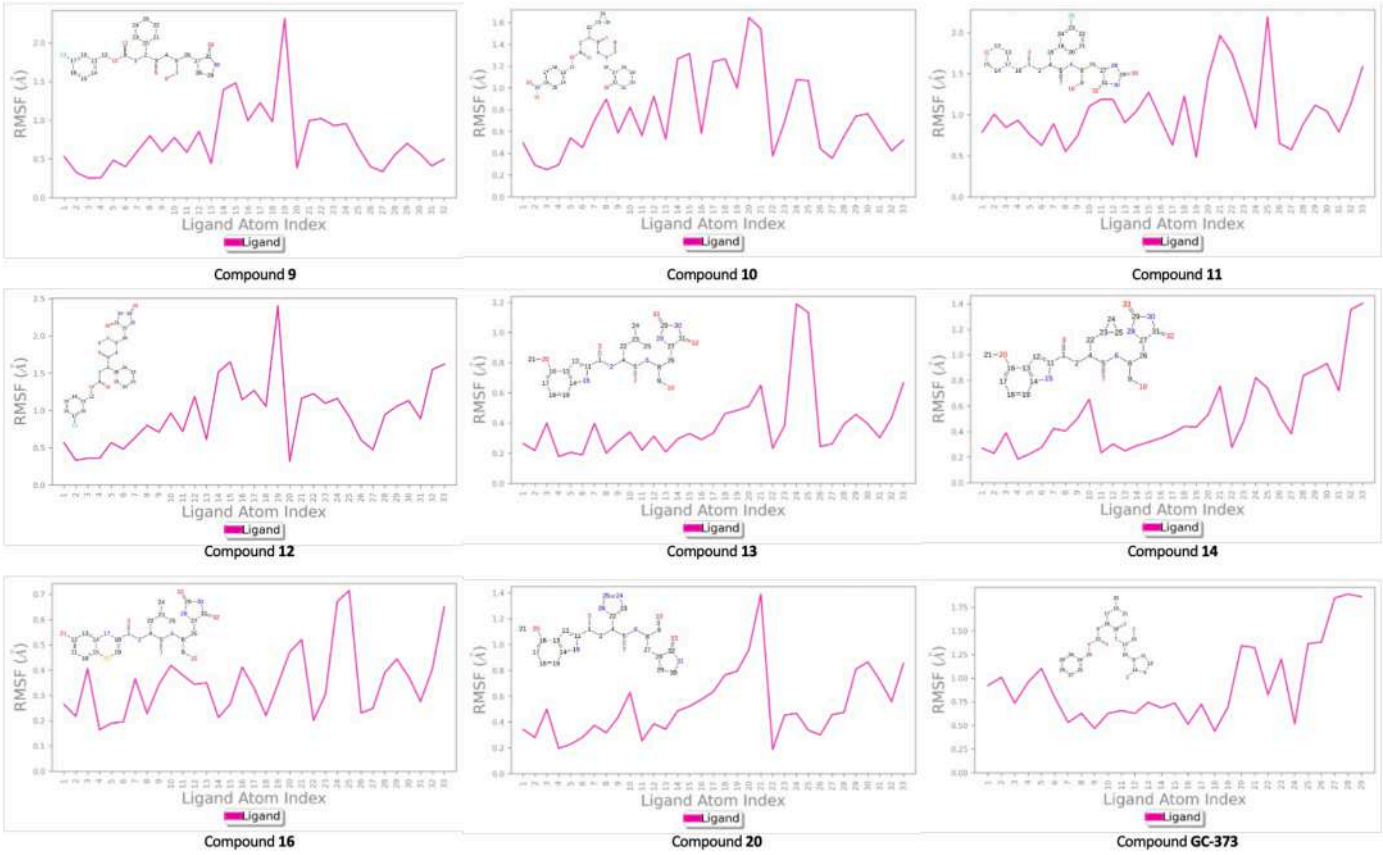



**Fig. 20.** Calculated L-RMSF for compounds **9**–**14**, **16**, **20**, and **GC-373** during the simulation trajectory of 200ns alongside 2D structures of related ligands.

recognition. In this work DRUDIT was used to obtain the original matrix of objects versus variables (Supplementary Material S1, Matrices S1-S2), and free version TIBCO Statistica® software was used to perform Principal Component Analysis.

## 4. Conclusions

In response to the unprecedented global health crisis posed by the COVID-19 pandemic, researchers worldwide focused on the need of effective antiviral treatments, with the SARS-CoV-2 $M^{PRO}$ emerging as a critical target for therapeutic intervention due to its essential role in the viral replication cycle. *In silico* simulations provide tools for qualitatively and quantitatively evaluating various treatments for specific diseases, leading to more practical and economical experiments. In the search for effective COVID-19 treatments, computational approaches have become essential in discovering and developing SARS-CoV-2 $M^{PRO}$ inhibitors.

In this context, the aim of this study is to develop and propose an innovative *in silico* protocol for the identification and optimization of anti-SARS-CoV-2 agents specifically targeting $M^{PRO}$ as covalent **peptidomimetic inhibitors with an aldehydic electrophilic warhead.** Peptidomimetic inhibitors are developed to closely mimic the structure of the natural viral polyproteins that are substrates of $M^{PRO}$, ensuring highly specific interactions. Utilizing the advanced CombiGlide Tool within the Maestro suite, a highly focused combinatorial library of 450 rationally designed peptidomimetic compounds with aldehydic warheads was generated. IFD studies targeted at the catalytic site of $M^{PRO}$ enabled the first refinement, narrowing the field to 388 compounds for subsequent multivariate statistical analysis. Of these, 14 standout compounds, identified through PCA, were meticulously evaluated using CovDock studies to assess their covalent binding capabilities. The results reveal significantly higher affinity compared to known inhibitors, affirming the validity of the adopted design strategy. Compounds, firstly analyzed for their pharmacokinetic properties, were explored in a dynamical point of view in a two steps Molecular Dynamics Simulation analysis. Notably, compound **9** exhibits zero violations across all the considered drug-likeness rules, including Lipinski, Ghose, Egan, Muegge, and PAINS, whereas both *Nirmatrelvir* and GC-373 fail to meet certain criteria (e.g., Ghose and Egan rules). This suggests that compound **9** may have better drug-likeness compared to these reference compounds. Furthermore, QikProp analysis reinforces this observation: compounds **9**, **12**, and **20** have no "stars", indicating that their properties fall within the desirable range for orally active drugs, whereas *Nirmatrelvir* has one "star", which may indicate a deviation from optimal properties.

Compounds with favorable ADME properties showed great stability, convergence, and compatibility within the $M^{PRO}$ binding site. Their robust RMSD profiles indicate that they can maintain strong and stable interactions with $M^{PRO}$, like or better than the reference compound GC-373.

The described protocol allowed to identify compounds **9**–**14**, **16**, and **20** as potential new SARS-CoV-2 $M^{PRO}$ inhibitors. This work aspires to contribute to the discovery and development of effective antiviral agents against COVID-19 and related coronaviruses.

**Supplementary Materials:** Table S1: SMILES representations for the fragments and the final obtained 300 compounds; Table S2: SMILES representations for the fragments and the final obtained 150 compounds; Table S3: IFD and docking scores for the selected 388 small molecules and ***nirmatrelvir***; Table S4: SMILES and $IC_{50}$ for structures utilized for the PCA model building; Table S5: cartesian coordinates for the calculation of the centroids, or the central points, of the two clusters identified in the PCA plot; Table S6: distance 1 and distance 2 for each input molecule; Table S7: calculated ratio between distance 1 and distance 2 for each molecule; Table S8: calculated parameters from QikProp for compounds **9**–**14**, **16**, and **19**–**21**; Table S9: calculated

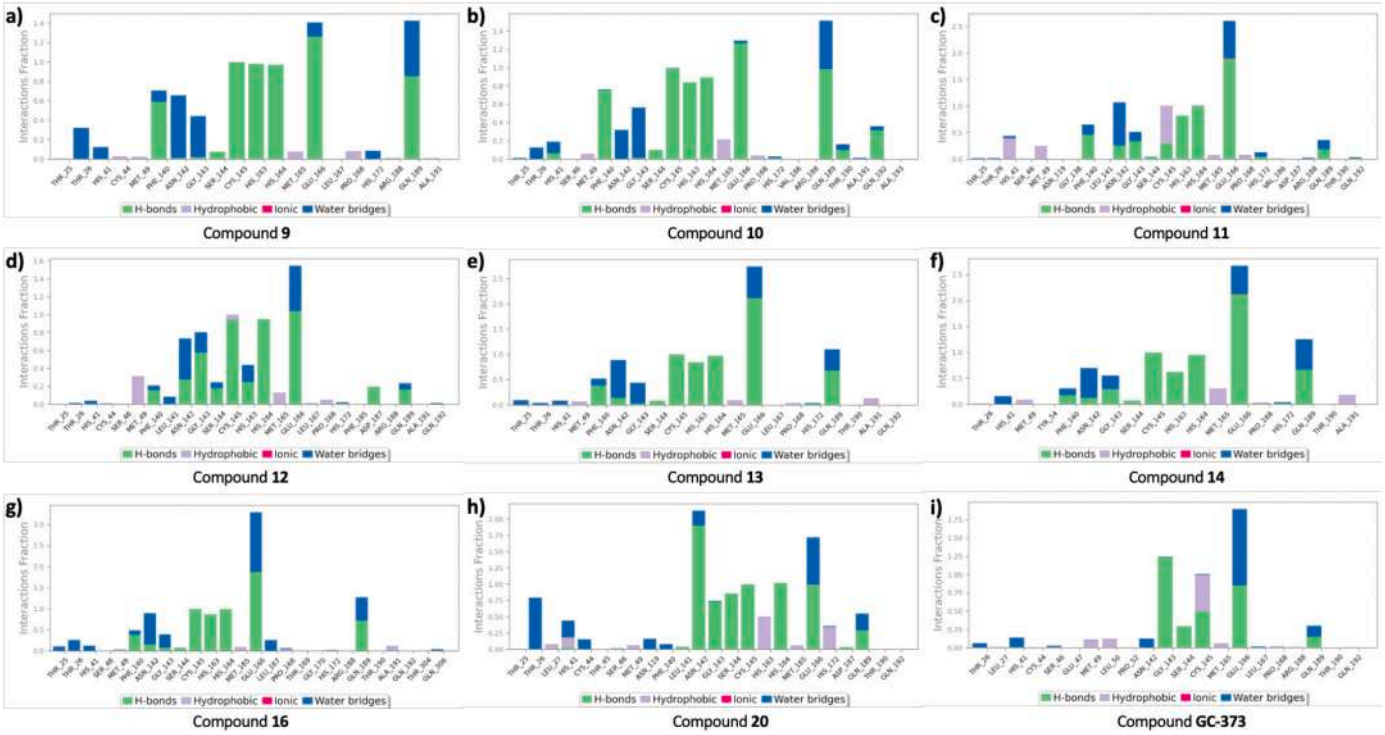



**Fig. 21. (a)** Protein–ligand interactions examination across the simulation time for **9**/M$^{PRO}$ complex; **(b)** protein–ligand interactions examination across the simulation time for **10**/M$^{PRO}$ complex; **(c)** protein–ligand interactions examination across the simulation time for **11**/M$^{PRO}$ complex; **(d)** protein–ligand interactions examination across the simulation time for **12**/M$^{PRO}$ complex; **(e)** protein–ligand interactions examination across the simulation time for **13**/M$^{PRO}$ complex; **(f)** protein–ligand interactions examination across the simulation time for **14**/M$^{PRO}$ complex; **(g)** protein–ligand interactions examination across the simulation time for **16**/M$^{PRO}$ complex; **(h)** protein–ligand interactions examination across the simulation time for **20**/M$^{PRO}$ complex; **(i)** protein–ligand interactions examination across the simulation time for **GC-373**/M$^{PRO}$ complex. Hydrogen bonds are labeled in green, hydrophobic interactions in light purple, ionic interactions in magenta, and water bridges in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

parameters from SwissADME for compounds **9**–**14**, **16**, and **19**–**21**; Matrix S1: Structure *versus* Molecular Descriptors matrix for known inhibitors; Matrix S2: Structure *versus* Molecular Descriptors matrix for 388 selected compounds; Fig. S1: In-depth structural analyses for complex **9**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S2: In-depth structural analyses for complex **10**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S3: In-depth structural analyses for complex **11**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S4: In-depth structural analyses for complex **12**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S5: In-depth structural analyses for complex **13**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S6: In-depth structural analyses for complex **14**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S7: In-depth structural analyses for complex **16**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S8: In-depth structural analyses for complex **20**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S9: In-depth structural analyses for complex **GC-373**/M$^{PRO}$: Calculated rGyr, intraHB, MolSA variations, SASA changes, and PSA dynamics over 200ns simulation; Fig. S10: Calculated L-RMSF for compound **9** during the simulation trajectory of 200ns; Fig. S11: Calculated L-RMSF for compound **10** during the simulation trajectory of 200ns; Fig. S12: Calculated L-RMSF for compound **11** during the simulation trajectory of 200ns; Fig. S13: Calculated L-RMSF for compound **12** during the simulation trajectory of 200ns; Fig. S14: Calculated L-RMSF for compound **13** during the simulation trajectory of 200ns; Fig. S15: Calculated L-RMSF for compound **14** during the simulation trajectory of 200ns; Fig. S16: Calculated L-RMSF for compound **16** during the simulation trajectory of 200ns; Fig. S17: Calculated L-RMSF for compound **20** during the simulation trajectory of 200ns; Fig. S18: Calculated L-RMSF for compound **GC-373** during the simulation trajectory of 200ns; Fig. S19: Protein–ligand interactions examination across the simulation time for **9**/M$^{PRO}$ complex; Fig. S20: Protein–ligand interactions examination across the simulation time for **10**/M$^{PRO}$ complex; Fig. S21: Protein–ligand interactions examination across the simulation time for **11**/M$^{PRO}$ complex.

## CRediT authorship contribution statement

**Alessia Bono:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gabriele La Monica:** Writing – original draft, Visualization, Validation, Formal analysis. **Federica Alamia:** Writing – original draft, Visualization, Validation. **Francesco Mingoia:** Writing – original draft, Visualization, Methodology. **Annamaria Martorana:** Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Antonino Lauria:** Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

## Informed consent statement:

Not applicable.

## Institutional review board statement:

Not applicable.

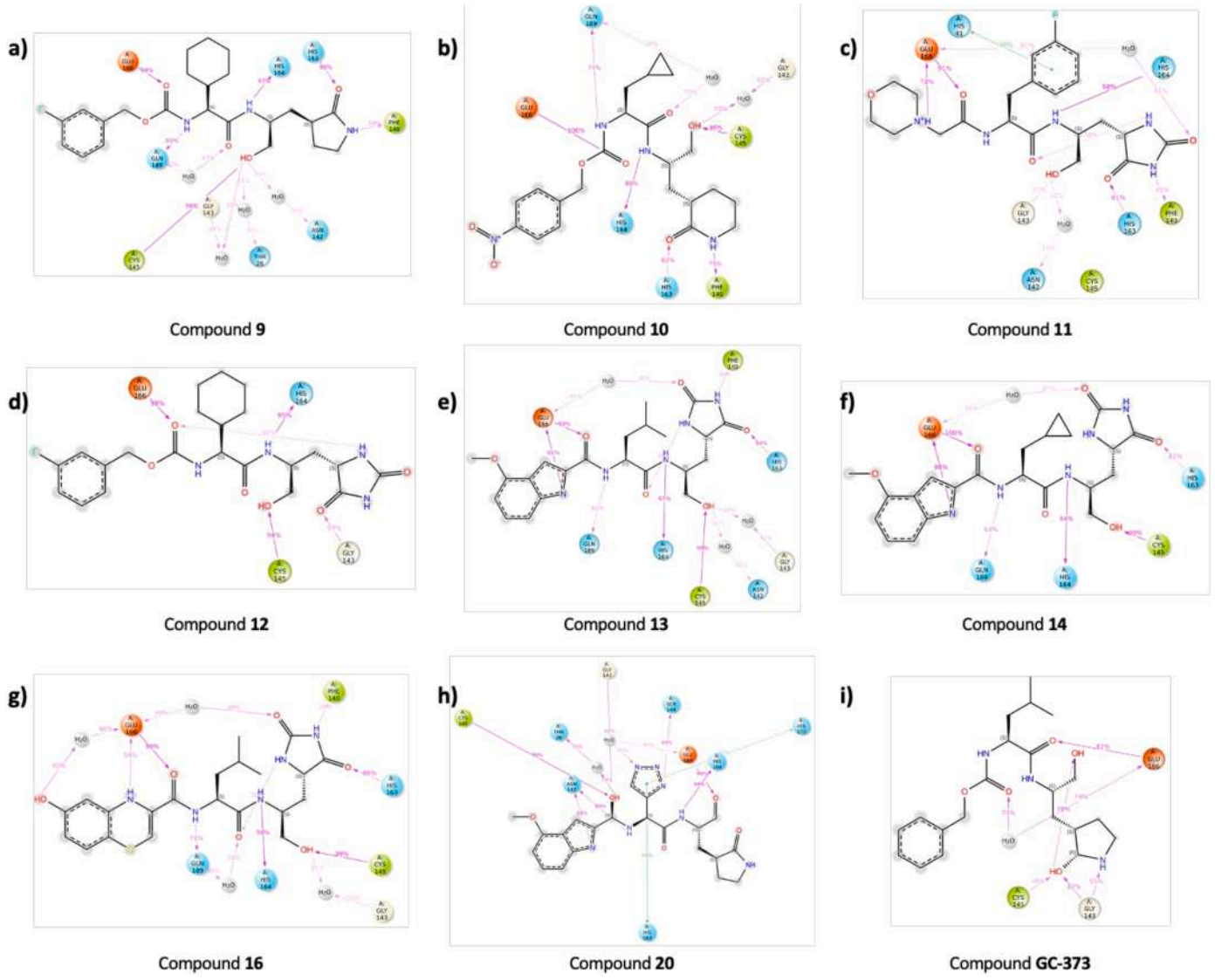



**Fig. 22.** **(a)** Detailed ligand atom interactions with the protein residues across the simulation time for **9**/$M^{PRO}$ complex; **(b)** detailed ligand atom interactions with the protein residues across the simulation time for **10**/$M^{PRO}$ complex; **(c)** detailed ligand atom interactions with the protein residues across the simulation time for **11**/$M^{PRO}$ complex; **(d)** detailed ligand atom interactions with the protein residues across the simulation time for **12**/$M^{PRO}$ complex; **(e)** detailed ligand atom interactions with the protein residues across the simulation time for **13**/$M^{PRO}$ complex; **(f)** detailed ligand atom interactions with the protein residues across the simulation time for **14**/$M^{PRO}$ complex; **(g)** detailed ligand atom interactions with the protein residues across the simulation time for **16**/$M^{PRO}$ complex; **(h)** detailed ligand atom interactions with the protein residues across the simulation time for **20**/$M^{PRO}$ complex; **(i)** detailed ligand atom interactions with the protein residues across the simulation time for **GC-373**/$M^{PRO}$ complex.

**Table 7**
Summary of MM-GBSA analysis: comparative binding energies of reference ligand **GC-373** and compounds **9**–**14**, **16**, and **20** in complex with the SARS-CoV-2 $M^{PRO}$.

| Complex | Docking score | Prime Coulomb | Prime vdW | Prime Solvent | Prime Hbond | Prime Energy |
|---|---|---|---|---|---|---|
| **9** | −9.333 | −10087.523 | −1441.063 | −1039.279 | −102.033 | −13360.6 |
| **10** | −9.566 | −10081.296 | −1451.701 | −1042.077 | −102.112 | −13363.9 |
| **11** | −10.213 | −10109.816 | −1443.346 | −1051.173 | −102.084 | −13384.5 |
| **12** | −9.770 | −10116.820 | −1442.974 | −1043.221 | −101.956 | −13383.7 |
| **13** | −8.970 | −10122.554 | −1450.794 | −1043.403 | −102.241 | −13389.2 |
| **14** | −9.531 | −10118.531 | −1448.834 | −1047.704 | −102.194 | −13374.1 |
| **16** | −9.746 | −10103.552 | −1455.405 | −1043.669 | −101.868 | −13389.7 |
| **20** | −8.016 | −10132.784 | −1445.776 | −1041.582 | −101.912 | −13382.0 |
| **GC-373** | −6.336 | −10103.535 | −1439.000 | −1036.399 | −101.216 | −13350.0 |

## Funding

This research was funded by SiciliAn MicronanOTecH Research and Innovation CEnter "SAMOTHRACE" (MUR, PNRR-M4C2,

ECS_00000022), spoke 3-Università degli Studi di Palermo "S2-COMMsMicro and Nanotechnologies for Smart & Sustainable Communities.

## Declaration of competing interest

I hereby declare that I (and my co-authors) have no conflict of interest to disclose related to this manuscript.

## Acknowledgments

The authors would like to thank also the "National Biodiversity Future Center" (identification code CN00000033, CUP.

B73C22000790001) on 'Biodiversity', financed under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment.

1.4 "Strengthening of research structures and creation of R&D 'national champions' on some Key Enabling Technologies"—Call for tender No. 3138 of December 16, 2021, rectified by Decree n.3175 of December 18, 2021 of Italian Ministry of University and Research funded by the European Union—NextGenerationEU.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2025.109100.

## Data availability

Data will be made available on request.

## References

[1] L. Wang, Y. Wang, D. Ye, Q. Liu, Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence, Int. J. Antimicrob. Agents 55 (2020) 105948, https://doi.org/10.1016/j.ijantimicag.2020.105948.

[2] D. Cucinotta, M. Vanelli, WHO declares COVID-19 a pandemic, Acta Biomed. 91 (2020) 157–160, https://doi.org/10.23750/abm.v91i1.9397.

[3] S. Krishnamoorthy, B. Swain, R.S. Verma, S.S. Gunthe, SARS-CoV, MERS-CoV, and 2019-nCoV viruses: an overview of origin, evolution, and genetic variations, Virusdisease (2020) 1–13, https://doi.org/10.1007/s13337-020-00632-9.

[4] L. Bian, F. Gao, J. Zhang, Q. He, Q. Mao, M. Xu, Z. Liang, Effects of SARS-CoV-2 variants on vaccine efficacy and response strategies, Expert Rev. Vaccines 20 (2021) 365–373, https://doi.org/10.1080/14760584.2021.1903879.

[5] A. Citarella, A. Scala, A. Piperno, N. Micale, SARS-CoV-2 $m^{pro}$: a potential target for peptidomimetics and small-molecule inhibitors, Biomolecules 11 (2021), https://doi.org/10.3390/biom11040607.

[6] H. Zhu, W. Du, M. Song, Q. Liu, A. Herrmann, Q. Huang, Spontaneous binding of potential COVID-19 drugs (Camostat and Nafamostat) to human serine protease TMPRSS2, Comput. Struct. Biotechnol. J. 19 (2021) 467–476, https://doi.org/10.1016/j.csbj.2020.12.035.

[7] X. Wang, R. Cao, H. Zhang, J. Liu, M. Xu, H. Hu, Y. Li, L. Zhao, W. Li, X. Sun, et al., The anti-influenza virus drug, arbidol is an efficient inhibitor of SARS-CoV-2 in vitro, Cell Discov 6 (2020) 28, https://doi.org/10.1038/s41421-020-0169-8.

[8] D.N. Juurlink, Safety considerations with chloroquine, hydroxychloroquine and azithromycin in the management of SARS-CoV-2 infection, CMAJ (Can. Med. Assoc. J.) 192 (2020) E450–E453, https://doi.org/10.1503/cmaj.200528.

[9] M. Costanzo, M.A.R. De Giglio, G.N. Roviello, SARS-CoV-2: recent reports on antiviral therapies based on lopinavir/ritonavir, darunavir/umifenovir, hydroxychloroquine, remdesivir, favipiravir and other drugs for the treatment of the new coronavirus, Curr. Med. Chem. 27 (2020) 4536–4541, https://doi.org/10.2174/0929867327666200416131117.

[10] H.L. Nguyen, N.Q. Thai, D.T. Truong, M.S. Li, Remdesivir strongly binds to both RNA-dependent RNA polymerase and main protease of SARS-CoV-2: evidence from molecular simulations, J. Phys. Chem. B 124 (2020) 11337–11348, https://doi.org/10.1021/acs.jpcb.0c07312.

[11] M. Hoffmann, S. Schroeder, H. Kleine-Weber, M.A. Müller, C. Drosten, S. Pöhlmann, Nafamostat mesylate blocks activation of SARS-CoV-2: new treatment option for COVID-19, Antimicrob. Agents Chemother. 64 (2020), https://doi.org/10.1128/AAC.00754-20.

[12] J.D. Gunst, N.B. Staerke, M.H. Pahus, L.H. Kristensen, J. Bodilsen, N. Lohse, L. S. Dalgaard, D. Brønnum, O. Fröbert, B. Hønge, et al., Efficacy of the TMPRSS2 inhibitor camostat mesilate in patients hospitalized with Covid-19-a double-blind randomized controlled trial, eClinicalMedicine 35 (2021) 100849, https://doi.org/10.1016/j.eclinm.2021.100849.

[13] N. Vankadari, Arbidol: a potential antiviral drug for the treatment of SARS-CoV-2 by blocking trimerization of the spike glycoprotein, Int. J. Antimicrob. Agents 56 (2020) 105998, https://doi.org/10.1016/j.ijantimicag.2020.105998.

[14] M. Plaze, D. Attali, A.C. Petit, M. Blatzer, E. Simon-Loriere, F. Vinckier, A. Cachia, F. Chrétien, R. Gaillard, Repurposing chlorpromazine to treat COVID-19: the reCoVery study, Encephale 46 (2020) 169–172, https://doi.org/10.1016/j.encep.2020.05.006.

[15] M. Plaze, D. Attali, M. Prot, A.C. Petit, M. Blatzer, F. Vinckier, L. Levillayer, J. Chiaravalli, F. Perin-Dureau, A. Cachia, et al., Inhibition of the replication of SARS-CoV-2 in human cells by the FDA-approved drug chlorpromazine, Int. J. Antimicrob. Agents 57 (2021) 106274, https://doi.org/10.1016/j.ijantimicag.2020.106274.

[16] PFIZER, Pfizer Initiates Phase 1 Study of Novel Oral Antiviral Therapeutic Agent against SARS-CoV-2, 2021.

[17] B. Halford, Pfizer Unveils its Oral SARS-CoV-2 Inhibitor, 2021.

[18] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, et al., Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, Nat. Commun. 11 (2020) 1620, https://doi.org/10.1038/s41467-020-15562-9.

[19] T. Mori, B.R. O'Keefe, R.C. Sowder, S. Bringans, R. Gardella, S. Berg, P. Cochran, J. A. Turpin, R.W. Buckheit, J.B. McMahon, et al., Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga Griffithsia sp, J. Biol. Chem. 280 (2005) 9345–9353, https://doi.org/10.1074/jbc.M411122200.

[20] Z.A. Shyr, K. Gorshkov, C.Z. Chen, W. Zheng, Drug discovery strategies for SARS-CoV-2, J. Pharmacol. Exp. Therapeut. 375 (2020) 127–138, https://doi.org/10.1124/jpet.120.000123.

[21] C.W. Day, R. Baric, S.X. Cai, M. Frieman, Y. Kumaki, J.D. Morrey, D.F. Smee, D. L. Barnard, A new mouse-adapted strain of SARS-CoV as a lethal model for evaluating antiviral agents in vitro and in vivo, Virology (New York, N. Y.) 395 (2009) 210–222, https://doi.org/10.1016/j.virol.2009.09.023.

[22] Y. Kumaki, M.K. Wandersee, A.J. Smith, Y. Zhou, G. Simmons, N.M. Nelson, K. W. Bailey, Z.G. Vest, J.K. Li, P.K. Chan, et al., Inhibition of severe acute respiratory syndrome coronavirus replication in a lethal SARS-CoV BALB/c mouse model by stinging nettle lectin, Urtica dioica agglutinin, Antivir. Res. 90 (2011) 22–32, https://doi.org/10.1016/j.antiviral.2011.02.003.

[23] F. Yu, R. Xiang, X. Deng, L. Wang, Z. Yu, S. Tian, R. Liang, Y. Li, T. Ying, S. Jiang, Receptor-binding domain-specific human neutralizing monoclonal antibodies against SARS-CoV and SARS-CoV-2, Signal Transduct. Targeted Ther. 5 (2020) 212, https://doi.org/10.1038/s41392-020-00318-0.

[24] X. Chi, X. Liu, C. Wang, X. Zhang, X. Li, J. Hou, L. Ren, Q. Jin, J. Wang, W. Yang, Humanized single domain antibodies neutralize SARS-CoV-2 by targeting the spike receptor binding domain, Nat. Commun. 11 (2020) 4528, https://doi.org/10.1038/s41467-020-18387-8.

[25] Y. Cao, B. Su, X. Guo, W. Sun, Y. Deng, L. Bao, Q. Zhu, X. Zhang, Y. Zheng, C. Geng, et al., Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells, Cell 182 (2020) 73–84.e16, https://doi.org/10.1016/j.cell.2020.05.025.

[26] W. Li, A. Drelich, D.R. Martinez, L. Gralinski, C. Chen, Z. Sun, X. Liu, D. Zhelev, L. Zhang, E.C. Peterson, et al., Potent neutralization of SARS-CoV-2 in vitro and in an animal model by a human monoclonal antibody, bioRxiv (2020), https://doi.org/10.1101/2020.05.13.093088.

[27] R. Shi, C. Shan, X. Duan, Z. Chen, P. Liu, J. Song, T. Song, X. Bi, C. Han, L. Wu, et al., A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2, Nature 584 (2020) 120–124, https://doi.org/10.1038/s41586-020-2381-y.

[28] Y. Wu, F. Wang, C. Shen, W. Peng, D. Li, C. Zhao, Z. Li, S. Li, Y. Bi, Y. Yang, et al., A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2, Sci. Technol. Humanit. 368 (2020) 1274–1278, https://doi.org/10.1126/science.abc2241.

[29] D. Wrapp, D. De Vlieger, K.S. Corbett, G.M. Torres, N. Wang, W. Van Breedam, K. Roose, L. van Schie, M. Hoffmann, S. Pöhlmann, et al., Structural basis for potent neutralization of betacoronaviruses by single-domain camelid antibodies, Cell 181 (2020) 1004–1015.e1015, https://doi.org/10.1016/j.cell.2020.04.031.

[30] P.J.M. Brouwer, T.G. Caniels, K. van der Straten, J.L. Snitselaar, Y. Aldon, S. Bangaru, J.L. Torres, N.M.A. Okba, M. Claireaux, G. Kerster, et al., Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability, Sci. Technol. Humanit. 369 (2020) 643–650, https://doi.org/10.1126/science.abc5902.

[31] W. Tai, X. Zhang, Y. He, S. Jiang, L. Du, Identification of SARS-CoV RBD-targeting monoclonal antibodies with cross-reactive or neutralizing activity against SARS-CoV-2, Antivir. Res. 179 (2020) 104820, https://doi.org/10.1016/j.antiviral.2020.104820.

[32] W. Wen, W. Su, H. Tang, W. Le, X. Zhang, Y. Zheng, X. Liu, L. Xie, J. Li, J. Ye, et al., Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing, Cell Discov 6 (2020) 31, https://doi.org/10.1038/s41421-020-0168-9.

[33] H.J. Lachmann, I. Kone-Paut, J.B. Kuemmerle-Deschner, K.S. Leslie, E. Hachulla, P. Quartier, X. Gitton, A. Widmer, N. Patel, P.N. Hawkins, et al., Use of canakinumab in the cryopyrin-associated periodic syndrome, N. Engl. J. Med. 360 (2009) 2416–2425, https://doi.org/10.1056/NEJMoa0810787.

[34] G. Gritti, F. Raimondi, D. Ripamonti, I. Riva, F. Landi, L. Alborghetti, M. Frigeni, M. Damiani, C. Mico, S. Fagiuoli, et al., Use of Siltuximab in Patients with COVID-19 Pneumonia Requiring Ventilatory Support, 2020.

[35] D.R. Burrage, S. Koushesh, N. Sofat, Immunomodulatory drugs in the management of SARS-CoV-2, Front. Immunol. 11 (2020) 1844, https://doi.org/10.3389/fimmu.2020.01844.
[36] S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, Bioorg Med Chem Lett 30 (2020) 127377, https://doi.org/10.1016/j.bmcl.2020.127377.
[37] E. Cho, M. Rosa, R. Anjum, S. Mehmood, M. Soban, M. Mujtaba, K. Bux, S.T. Moin, M. Tanweer, S. Dantu, et al., Dynamic profiling of β-coronavirus 3CL $m^{pro}$ protease ligand-binding sites, J. Chem. Inf. Model. 61 (2021) 3058–3073, https://doi.org/10.1021/acs.jcim.1c00449.
[38] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Sci. Technol. Humanit. 368 (2020) 409–412, https://doi.org/10.1126/science.abb3405.
[39] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, et al., Structure of $M^{pro}$ from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293, https://doi.org/10.1038/s41586-020-2223-y.
[40] K. Arafet, N. Serrano-Aparicio, A. Lodola, A.J. Mulholland, F.V. González, K. Świderek, V. Moliner, Mechanism of inhibition of SARS-CoV-2 $M^{pro}$ by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity, Chem. Sci. 12 (2020) 1433–1444, https://doi.org/10.1039/d0sc06195f.
[41] C.A. Ramos-Guzmán, J.J. Ruiz-Pernía, I. Tuñón, Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods, ACS Catal. 10 (2020) 12544–12554, https://doi.org/10.1021/acscatal.0c03420.
[42] K. Świderek, V. Moliner, Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 M, Chem. Sci. 11 (2020) 10626–10630, https://doi.org/10.1039/d0sc02823a.
[43] N. Shindo, A. Ojida, Recent progress in covalent warheads for in vivo targeting of endogenous proteins, Bioorg. Med. Chem. 47 (2021) 116386, https://doi.org/10.1016/j.bmc.2021.116386.
[44] Y. Liu, S. Lv, L. Peng, C. Xie, L. Gao, H. Sun, L. Lin, K. Ding, Z. Li, Development and application of novel electrophilic warheads in target identification and drug discovery, Biochem. Pharmacol. 190 (2021) 114636, https://doi.org/10.1016/j.bcp.2021.114636.
[45] M. Gehringer, S.A. Laufer, Emerging and Re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology, J. Med. Chem. 62 (2019) 5673–5724, https://doi.org/10.1021/acs.jmedchem.8b01153.
[46] T. Muramatsu, C. Takemoto, Y.T. Kim, H. Wang, W. Nishii, T. Terada, M. Shirouzu, S. Yokoyama, SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 12997–13002, https://doi.org/10.1073/pnas.1601327113.
[47] K. Fan, L. Ma, X. Han, H. Liang, P. Wei, Y. Liu, L. Lai, The substrate specificity of SARS coronavirus 3C-like proteinase, Biochem. Biophys. Res. Commun. 329 (2005) 934–940, https://doi.org/10.1016/j.bbrc.2005.02.061.
[48] G. La Monica, A. Bono, A. Lauria, A. Martorana, Targeting SARS-CoV-2 main protease for treatment of COVID-19: covalent inhibitors structure-activity relationship insights and evolution perspectives, J. Med. Chem. (2022), https://doi.org/10.1021/acs.jmedchem.2c01005.
[49] A.A.T. Naqvi, K. Fatima, T. Mohammad, U. Fatima, I.K. Singh, A. Singh, S.M. Atif, G. Hariprasad, G.M. Hasan, M.I. Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach, Biochim. Biophys. Acta Mol. Basis Dis. 1866 (2020) 165878, https://doi.org/10.1016/j.bbadis.2020.165878.
[50] Y. Zhao, C. Fang, Q. Zhang, R. Zhang, X. Zhao, Y. Duan, H. Wang, Y. Zhu, L. Feng, J. Zhao, et al., Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332, Protein Cell (2021), https://doi.org/10.1007/s13238-021-00883-2.
[51] A. Lauria, S. Mannino, C. Gentile, G. Mannino, A. Martorana, D. Peri, DRUDIT: web-based DRUgs DIscovery Tools to design small molecules as modulators of biological targets, Bioinformatics 36 (2020) 1562–1569, https://doi.org/10.1093/bioinformatics/btz783.
[52] T. Yu, C. Nantasenamat, S. Kachenton, N. Anuwongcharoen, T. Piacham, Cheminformatic analysis and machine learning modeling to investigate androgen receptor antagonists to combat prostate cancer, ACS Omega 8 (2023) 6729–6742, https://doi.org/10.1021/acsomega.2c07346.
[53] T. Yu, C. Nantasenamat, N. Anuwongcharoen, T. Piacham, Machine learning approaches to investigate the structure-activity relationship of angiotensin-converting enzyme inhibitors, ACS Omega 8 (2023) 43500–43510, https://doi.org/10.1021/acsomega.3c03225.
[54] S. Das, S.A. Amin, T. Jha, Insight into the structural requirement of aryl sulphonamide based gelatinases (MMP-2 and MMP-9) inhibitors - Part I: 2D-QSAR, 3D-QSAR topomer CoMFA and Naïve Bayes studies - first report of 3D-QSAR Topomer CoMFA analysis for MMP-9 inhibitors and jointly inhibitors of gelatinases together, SAR QSAR Environ. Res. 32 (2021) 655–687, https://doi.org/10.1080/1062936X.2021.1955414.
[55] G.S. Rieder, P.A. Nogara, F.B. Omage, T. Duarte, C.L. Dalla Corte, J.B.T. da Rocha, Computational analysis of the interactions between Ebselen and derivatives with the active site of the main protease from SARS-CoV-2, Comput. Biol. Chem. 107 (2023) 107956, https://doi.org/10.1016/j.compbiolchem.2023.107956.
[56] P.A. Nogara, F.B. Omage, G.R. Bolzan, C.P. Delgado, M. Aschner, L. Orian, J. B. Teixeira Rocha, In silico studies on the interaction between mpro and PLpro from SARS-CoV-2 and ebselen, its metabolites and derivatives, Mol Inform 40 (2021) e2100028, https://doi.org/10.1002/minf.202100028.
[57] A. Madabeni, P.A. Nogara, F.B. Omage, J.B.T. Rocha, L. Orian, Mechanistic insight into sars-cov-2 mpro inhibition by organoselenides: the ebselen case study, Appl. Sci. 11 (2021), https://doi.org/10.3390/app11146291.
[58] N. Kaiyawet, T. Rungrotmongkol, S. Hannongbua, Effect of halogen substitutions on dUMP to stability of thymidylate synthase/dUMP/mTHF ternary complex using molecular dynamics simulation, J. Chem. Inf. Model. 53 (2013) 1315–1323, https://doi.org/10.1021/ci400131y.
[59] Y. Zhang, D. Zhang, H. Tian, Y. Jiao, Z. Shi, T. Ran, H. Liu, S. Lu, A. Xu, X. Qiao, et al., Identification of covalent binding sites targeting cysteines based on computational approaches, Mol. Pharm. 13 (2016) 3106–3118, https://doi.org/10.1021/acs.molpharmaceut.6b00302.
[60] K. Zhu, K.W. Borrelli, J.R. Greenwood, T. Day, R. Abel, R.S. Farid, E. Harder, Docking covalent inhibitors: a parameter free approach to pose prediction and scoring, J. Chem. Inf. Model. 54 (2014) 1932–1940, https://doi.org/10.1021/ci500118s.
[61] D. Toledo Warshaviak, G. Golan, K.W. Borrelli, K. Zhu, O. Kalid, Structure-based virtual screening approach for discovery of covalently bound ligands, J. Chem. Inf. Model. 54 (2014) 1941–1950, https://doi.org/10.1021/ci500175r.
[62] W. Vuong, M.B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H.A. Saffran, R.T. McKay, M.J. van Belkum, M.A. Joyce, et al., Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication, Nat. Commun. 11 (2020) 4282, https://doi.org/10.1038/s41467-020-18096-2.
[63] J. Dulsat, B. López-Nieto, R. Estrada-Tejedor, J.I. Borrell, Evaluation of free online ADMET tools for academic or small biotech environments, Molecules (Basel) 28 (2023), https://doi.org/10.3390/molecules28020776.
[64] Schrödinger Release 2017-1 : QikProp, Schrödinger, LLC, New York, NY, 2017.
[65] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 42717, https://doi.org/10.1038/srep42717.
[66] Schrödinger Release 2017-2, LigPrep, Schrödinger, LLC, New York, NY, USA, 2017.
[67] Schrödinger Release 2017-2, Schrödinger Suite 2017-2 Protein Preparation Wizard, Epik, Schrödinger, LLC, New York, NY, USA, 2017.
[68] J.L. Banks, H.S. Beard, Y. Cao, A.E. Cho, W. Damm, R. Farid, A.K. Felts, T. A. Halgren, D.T. Mainz, J.R. Maple, et al., Integrated modeling program, applied chemical theory (IMPACT), J. Comput. Chem. 26 (2005) 1752–1780, https://doi.org/10.1002/jcc.20292.
[69] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.
[70] Rcsb PDB, Available online: www.rcsb.org. (Accessed 29 June 2022).
[71] G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, J. Comput. Aided Mol. Des. 27 (2013) 221–234, https://doi.org/10.1007/s10822-013-9644-8.
[72] W. Sherman, H.S. Beard, R. Farid, Use of an induced fit receptor structure in virtual screening, Chem. Biol. Drug Des. 67 (2006) 83–84, https://doi.org/10.1111/j.1747-0285.2005.00327.x.
[73] W. Sherman, T. Day, M.P. Jacobson, R.A. Friesner, R. Farid, Novel procedure for modeling ligand/receptor induced fit effects, J. Med. Chem. 49 (2006) 534–553, https://doi.org/10.1021/jm050540c.