



**HAL**  
open science

# COVID-19: In silico identification of potent $\alpha$ -ketoamide inhibitors targeting the main protease of the SARS-CoV-2

Mehdi Oubahmane, Ismail Hdoufane, Imane Bjj, Carola Jerves, Didier Villemin, Driss Cherqaoui

## ► To cite this version:

Mehdi Oubahmane, Ismail Hdoufane, Imane Bjj, Carola Jerves, Didier Villemin, et al.. COVID-19: In silico identification of potent  $\alpha$ -ketoamide inhibitors targeting the main protease of the SARS-CoV-2. Journal of Molecular Structure, 2021, 1244, pp.130897. 10.1016/j.molstruc.2021.130897. hal-03264711

**HAL Id: hal-03264711**

**<https://normandie-univ.hal.science/hal-03264711v1>**

Submitted on 18 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## COVID-19: In silico identification of potent $\alpha$ -ketoamide inhibitors targeting the main protease of the SARS-CoV-2

Mehdi Oubahmane <sup>1</sup>, Ismail Hdoufane <sup>1</sup>, Imane Bjjj <sup>1</sup>, Carola Jerves <sup>2,3</sup>, Didier Villemin <sup>4</sup>, and Driss Cherqaoui <sup>1,\*</sup>

<sup>1</sup> Department of Chemistry, Faculty of Sciences Semlalia, BP, 2390, Marrakech, Morocco

<sup>2</sup> Departamento de Quimica, Faculdade de Ciencias da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto - Portugal.

<sup>3</sup> Facultad de Ciencias Químicas, Universidad de Cuenca. Cuenca, Ecuador

<sup>4</sup> Ecole Nationale Supérieure d'Ingénieurs (ENSICAEN) Laboratoire de Chimie Moléculaire et Thioorganique. UMR 6507 CNRS, INC3M, FR3038, Labex EMC3, Labex SynOrg ENSICAEN & Université de Caen, France

\* Corresponding author: [cherqaoui@uca.ma](mailto:cherqaoui@uca.ma)

### ABSTRACT

The COVID-19 has been creating a global crisis, causing countless deaths and unbearable panic. Despite the progress made in the development of the vaccine, there is an urge need for the discovery of antivirals that may better work at different stages of SARS-CoV-2 reproduction. The main protease ( $M^{pro}$ ) of the SARS-CoV-2 is a crucial therapeutic target due to its critical function in virus replication. The  $\alpha$ -ketoamide derivatives represent an important class of inhibitors against the  $M^{pro}$  of the SARS-CoV. While there is 99% sequence similarity between SARS-CoV and SARS-CoV-2 main proteases, anti-SARS-CoV compounds may have a huge demonstration's prospect of their effectiveness against the SARS-CoV-2. In this study, we applied various computational approaches to investigate the inhibition potency of novel designed  $\alpha$ -ketoamide-based compounds. In this regard, a set of 21  $\alpha$ -ketoamides was employed to construct a QSAR model, using the genetic algorithm-multiple linear regression (GA-MLR), as well as a pharmacophore fit model. Based on the GA-MLR model, 713 new designed molecules were reduced to 150 promising hits, which were later subject to the established pharmacophore fit model. Among the 150 compounds, the best selected compounds (3 hits) with greater pharmacophore fit score were further studied *via* molecular docking, molecular dynamic simulations along with the Absorption, distribution, metabolism, excretion, and toxicity (ADMET) analysis. Our approach revealed that the three hit compounds could serve as potential inhibitors against the SARS-CoV-2  $M^{pro}$  target.

**Keywords:** SARS-CoV-2; Main protease ( $M^{pro}$ );  $\alpha$ -ketoamide; QSAR; Pharmacophore modeling; Molecular docking; Molecular dynamics simulations.

## 1. Introduction

The spread of novel coronavirus disease COVID-19, caused by serious acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has had significant morbidity, mortality, and social and economic disruption throughout the globe. This disease has occurred firstly in Wuhan, Hubei Province, China, where a pneumonia of an unknown cause was detected in December 2019. As of now, SARS-CoV-2 has spread over to almost all parts of the world (213 nations) with over 120 million confirmed cases and over 2.67 million confirmed deaths worldwide at the time of writing (15<sup>th</sup> March, 2021) [1]. In the 21<sup>st</sup> century, SARS-CoV-2 is the 3<sup>rd</sup> *Coronaviridae* family member after the SARS-CoV in 2002 and the Middle East respiratory syndrome coronavirus MERS-CoV in 2012 that have infected 8422 (mortality rate of about 10%) and 1700 people (mortality rate of about 36%), respectively [2,3].

Like other coronaviruses, SARS-CoV-2 genome is around 30 kb in size and its genomic organization followed the gene characteristic order to known CoVs, [5'-replicase (*rep*), spike (*S*), envelope (*E*), membrane (*M*), and nucleocapsid (*N*)-3'] [4]. The 5' terminal more than two-thirds of the genome contains two long open reading frames, ORF1a and ORF1b which are translated into two polyproteins *pp1a* and *pp1b* to encode 16 non-structural proteins (*nsp1-nsp16*) which form the viral replicase-transcriptase complex (RTC). On the other hand, the 3' third of the genome contains the remaining ORFs that encode the 4 viral structural proteins (*S*, *M*, *E* and *N*) as well as the 9 auxillary proteins (*ORF3a*, *ORF3b*, *ORF6*, *ORF7a*, *ORF7b*, *ORF8*, *ORF9a*, *ORF9b* and *ORF10* genes) [5]. The RTC consists of multiple enzymes, the main protease ( $M^{pro}$ , *nsp5*) and the papain-like protease (PL<sup>pro</sup>, *nsp3*) participate in the cleavage of the polyproteins to produce *nsp2-nsp16* involved in the RTC [6].

The  $M^{pro}$  is essential for viral replication and maturation. It has been paid attention to the  $M^{pro}$  (i.e. 3C-like protease (3CL<sup>pro</sup>)) as an attractive target for anti-COVID-19 drug discovery and development [7]. The  $M^{pro}$  monomer is made up of three domains: domain I, II, and III with amino acid residues 8-101, 102-184, and 201-306, respectively. The catalytic dyad composed of CYS145 and HIS41, where is located at the cleft between domains I and II, is reported to initiate the activation through dimerization process mechanism [6]. Thus, it may therefore be logical to block the catalytic site in order to inhibit the main protease function.

1  
2  
3  
4 As the SARS-CoV-2 genome has over 80% identity to SARS-CoV (about 99% sequence  
5 similarity for their M<sup>pro</sup>), previously reported 21  $\alpha$ -ketoamides SARS-CoV M<sup>pro</sup> inhibitors may  
6 have significant efficacy to show against the SARS-CoV-2 [3]. In this view, various  
7 computational approaches including QSAR modeling, pharmacophore modeling, molecular  
8 docking, molecular dynamic (MD) simulations, and Absorption, distribution, metabolism,  
9 excretion, and toxicity (ADMET) test were used to investigate the possible inhibitory activity of  
10 novel designed  $\alpha$ -ketoamide-based compounds. Indeed, the aim of this work consists of:

- 11 ✓ Establishing a reliable QSAR model using GA-MLR method that can be able to predict  
12 the SARS-CoV-2 M<sup>pro</sup> inhibitory activity of  $\alpha$ -ketoamides.
- 13 ✓ Developing a pharmacophore fit model based on the structural features of the studied  
14 SARS-CoV-2 M<sup>pro</sup> inhibitors.
- 15 ✓ Proposing new drug candidates and screening for new hits through the investigation of  
16 the effect of active groups.
- 17 ✓ Elucidating the dynamics of the complexes and the underlying inhibition mode of the  
18 proposed hits at the active pocket of the target protein and evaluating their  
19 pharmacological profiles.

## 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 **2. Materials and methods**

### 35 36 *2.1. Data collection and molecular descriptor calculation*

37  
38 A dataset of 21 compounds of  $\alpha$ -ketoamides derivatives with SARS-CoV inhibitory activity were  
39 collected from the published work by Zhang et al. (Table S1) [8]. The retrieved IC<sub>50</sub> values were  
40 converted to their corresponding pIC<sub>50</sub> (-log<sub>10</sub> (IC<sub>50</sub>)) and used as a dependent variable. The  
41 ChemDraw 18.0 software was used to draw the chemical structures and their geometries were  
42 optimized using the AM1 method in the gas phase. This optimization was implemented in the  
43 Gaussian 09 program package [9]. Frequency analysis was used to verify the energy minima of  
44 the optimized samples.

45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
Molecular descriptor values (0D-3D) of the aforementioned compounds were calculated with the  
OCHEM server (it is a free implementation environment, available online <https://ochem.eu>) [10].  
The variables (nearly 5000) were collected and pre-filtered by eliminating constant and nearly  
constant value (>80%). Pairs of a highly correlation coefficient (> 0.95) were pruned, avoiding  
development of any biased models. The process continued till a set of the relevant 10 descriptors

1  
2  
3  
4 from the initial computed pool was selected. Finally, four descriptors listed in [Table S2](#) provided  
5 the best QSAR model.  
6

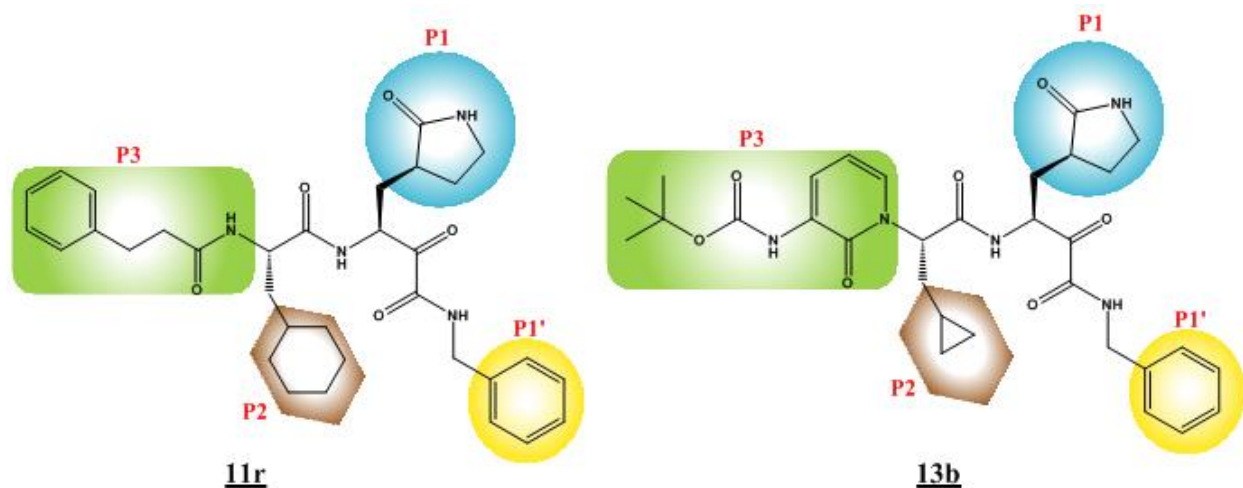
## 7 8 **2.2. 2D-QSAR Model Construction and Validation** 9

10 The statistically robust 2D-QSAR model based on the genetic algorithm-multiple linear  
11 regression (GA-MLR) has been established utilizing the QSARINS software [11,12]. This model  
12 was subjected to various thorough statistical validations as per the Organization for Economic  
13 Co-operation and Development (OECD) guidelines. The general procedure for developing the  
14 QSAR model is composed of three main sessions:  
15  
16  
17

- 18 (i). The dataset comprising the 21  $\alpha$ -ketoamide analogs was randomly divided, using the  
19 splitting option in the QSARINS software, into 70% training set (14 compounds) and 30%  
20 test set (7 compounds). The training set was employed for model construction and the test  
21 set for the external validation.  
22  
23  
24  
25  
26
- 27 (ii). The GA-MLR 2D-QSAR model was constructed using default parameters of the QSARINS  
28 program. During the model development, the selected molecular descriptors were utilized to  
29 derive a simple and an informative GA-MLR model.  
30  
31  
32
- 33 (iii). The confirmation of the model validity was proved by subjecting the established model to  
34 internal and external validations, Y-randomization, and model applicability domain (AD)  
35 analysis. Indeed, the statistical quality and robustness of the GA-MLR-based 2D-QSAR  
36 model was ensured on the basis of: (a) internal validation based on leave-one-out (LOO) and  
37 leave-many-out (LMO) procedure (i.e. cross-validation (CV)); (b) external validation;  
38 (c) Y-randomization; and (d) fulfillment of respective threshold values for the statistical  
39 metrics: the determination coefficient of the training set ( $R_{tr}^2 \geq 0.6$ ), the square of the CV  
40 correlation coefficient obtained from the leave-one-out (LOO) procedure ( $Q_{loo}^2 \geq 0.5$ ), the  
41 leave-many-out (LMO)  $Q_{lmo}^2 \geq 0.6$ , and the square of correlation coefficient obtained for the  
42 test set ( $R_{test}^2 \geq 0.6$ ). Finally, the root-mean square error (RMSE) and the mean absolute  
43 error (MAE) values should be close to zero. Any QSAR model that did not satisfy any of  
44 these criteria is therefore omitted.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

## 55 56 **2.3. Virtual Screening Database Preparation** 57 58 59 60 61 62 63 64 65

1  
2  
3  
4 Virtual screening (VS) is a computational method employed to screen available chemical  
5 databases for filtering molecules that are most likely to bind to a drug target by mapping them on  
6 generated chemoinformatic models [13]. The hit identification, using VS of compounds, is  
7 among the most popular computational techniques in drug design [14,15]. In this work, the  
8 screened database was prepared using different strategies and a total number of 713 molecules  
9 was generated. The first part was formed based on an initial pharmacophore model created by the  
10 ZincPharmer online server. Chemical features from the core structure of **13b** and **11r** compounds  
11 (Figure 1), possessing significant SARS-CoV-2 inhibitory activities, were employed to generate  
12 initial pharmacophore models (Figure S1) [16]. These models were applied to retrieve a total  
13 number of 197 compounds from the Zinc database [17]. The second part comprising 484  
14 molecules was generated based on several modifications at the core structure of compounds **13b**  
15 and **11r** using the “Expand Generic structure” option implemented in the ChemDraw software.  
16 The modifications were made based on the optimization of P1', P2, and P3 substituents in  
17 compounds **13b** and **11r** (Figure 1). The P1 moiety was kept intact during the modifications due  
18 to its vital role in the interaction with the active pocket of the M<sup>Pro</sup>. The last part which contains  
19 32 molecules was taken from a previously published series of  $\alpha$ -ketoamide inhibitors [18].  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33



51  
52 **Figure 1.** Chemical structures of  $\alpha$ -ketoamide inhibitors **11r** and **13b**. Colored circles highlight  
53 the substituents where the modifications were made.  
54

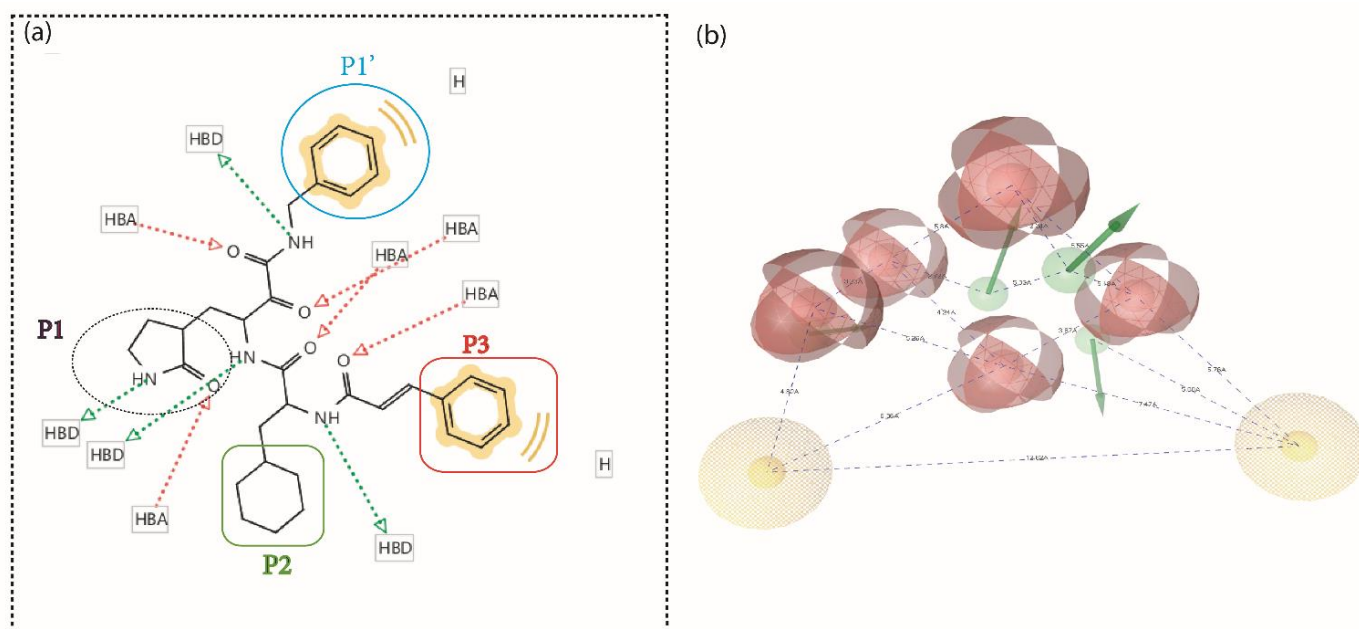
55  
56 Herein, the identified molecules were subjected to two filters. Firstly, the above-mentioned  
57 GA-MLR model was employed in order to screen for compounds with the highest pIC<sub>50</sub> values  
58 from the list of 713 generated ones. Secondly, the features fitting of each compound were  
59  
60  
61  
62  
63  
64  
65



1  
2  
3  
4 evaluated through the application of advanced pharmacophore fit model, facilitating the ranking  
5 of the screened molecules from the best GA-MLR QSAR model.  
6  
7  
8  
9

#### 10 **2.4. Ligand-based Pharmacophore model**

11  
12  
13 Ligand-based pharmacophore approach has become an important tool in the identification and  
14 extraction of the main chemical features from a set of active compounds [19,20]. Over the 713  
15 generated compounds that were subject to the established GA-MLR model, the best 150  
16 molecules were selected for advanced pharmacophore fit models. The pharmacophore models  
17 were generated based on the initial 21  $\alpha$ -ketoamide analogs using LigandScout software [21]. On  
18 investigation, it was observed that a combination of three distinct chemical features of the  
19 molecules in the training set, including hydrogen-bond acceptor (HBA), hydrogen-bond donor  
20 (HBD), and hydrophobic interaction (H), was found to effectively assess all critical chemical  
21 features (Figure 2). These chemical characteristics were generated in order to remain close to the  
22 common core structure of the studied  $\alpha$ -ketoamide analogs and later were employed to screen for  
23 new hit molecules. Finally, 3 hits (**007**, **329** and **331**) with the higher pharmacophore fit score  
24 were subjected to molecular docking and MD simulations analysis followed by the ADMET  
25 study to confirm their pharmacokinetic profiles.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37



1  
2  
3  
4 **Figure 2. a** | Ligand-based pharmacophore model generated by LigandScout software, showing  
5 different features including HBA, HBD, and Hydrophobic (H).  
6  
7

8 **b** | Pharmacophore model with distance constraints.  
9

## 10 11 12 **2.5. Molecular Docking of Selected Hits** 13

14  
15 Molecular docking analyses have been widely employed in medicinal chemistry research  
16 [22,23]. In the current work, molecular docking was employed to analyze the interactions and the  
17 conformational patterns of the selected hits within the active site of the M<sup>Pro</sup> of the SARS-CoV-2.  
18 The crystallographic protein of the M<sup>Pro</sup> was downloaded from the Protein Data Bank with the  
19 PDB code 6Y2F [16,24]. It is a monoclinic form of the complex resulting from the reaction  
20 between SARS-CoV-2 M<sup>Pro</sup> and the  $\alpha$ -ketoamide inhibitor (**13b**). For the protein preparation, the  
21 co-crystallized ligand and the water molecules were removed and the hydrogens were added  
22 using Autodock Tools [25]. The selected hits were prepared and saved in pdbqt format using the  
23 same software. In the docking investigations, a box of 20  $\times$  20  $\times$  20 and a grid space of 0.375 Å  
24 were fixed. The center of the ligands was set using the following coordinates x = 10.88 Å, y = -  
25 0.25 Å and z = 20.75 Å. The mode number was fixed at 8. All the docking runs were performed  
26 by the Autodock Vina program [26].  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 40 **2.6. Molecular Dynamics Simulations** 41

42 Molecular dynamics (MD) simulation is an important approach in understanding the  
43 fundamental basis of the biological macromolecule structures, which can provide critical insights  
44 on their function and dynamics. Therefore, AMBER18 software package [27] was utilized to  
45 perform 160 ns of MD simulations. Five systems were simulated: complex**007** (compound **007** in  
46 complex with M<sup>Pro</sup>), complex**329** (compound **329** in complex with M<sup>Pro</sup>), complex**331**  
47 (compound **331** in complex with M<sup>Pro</sup>), complex**13b** (the reference compound **13b** in complex  
48 with M<sup>Pro</sup>), and **APO** (the unbound form of M<sup>Pro</sup>). Prior to the MD simulations, the best-docked  
49 pose for each ligand was parameterized using Antechamber [28]. The TLeap module was used to  
50 ensemble the system, implementing the FF99SB force field to parameterize the protein. The  
51 general AMBER force field (GAFF) was employed to determine the partial charges of atoms in  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



the ligand. Neutralizing counterions and missing hydrogens were included [29,30]. All the systems were placed inside a TIP3P water box of a distance of 10 Å [31]. The partial Mesh Ewald (PME) [32] method was used to account for the long-range electrostatic forces using a cutoff of 12 Å, and the SHAKE algorithm [33] was used to constrain all the hydrogen atom bonds. The studied systems were then minimized using double phase method. First, 2500 steps of steepest descent minimization were applied with restraint conditions of 500 kcal.mol<sup>-1</sup>.Å<sup>-2</sup>, considering the solute molecule. This was followed by a conjugate gradient minimization process in the absence of all restraints; thus, the overall system was relaxed. The systems were progressively heated at a constant volume and at a constant pressure from 0 to 300 K using a harmonic restraint of 10 kcal.mol<sup>-1</sup>.Å<sup>-2</sup>. In order to equilibrate the system, weak restraints were performed during 1000 ps at a temperature of 300 K. Also, the Berendsen barostat was used to maintain the system pressure at 1 bar [34]. Total of 160 ns of MD simulations were performed on all the systems. The resulting coordinates and trajectories were analyzed using the integrated CPPTRAJ and PTRAJ modules of AMBER18 [35]. The attained data were plotted using Origin software [36].

## 2.7. Calculations of the binding free energy

The binding interactions of the compounds **13b**, **007**, **329** and **331** with the M<sup>PTO</sup> were estimated through the calculation of binding free energies using the method of Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PBSA) [37]. The binding free energy is mathematically represented as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}}$$

$$\Delta G_{\text{bind}} = \Delta E_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S$$

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdw}} + \Delta E_{\text{ele}}$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$$

$$\Delta G_{\text{SA}} = \gamma \text{SASA} + \beta$$

where  $\Delta S$  and  $T$  denote the total entropy of the solute and the temperature, respectively. However,  $\Delta E_{\text{gas}}$ ,  $\Delta E_{\text{int}}$ ,  $\Delta E_{\text{ele}}$  and  $\Delta E_{\text{vdw}}$  designate gas-phase energy, internal energy, electrostatic and van der Waals interactions, respectively.  $\Delta G_{\text{sol}}$  is the solvation free energy which can be separated into polar solvation free energy ( $\Delta G_{\text{pb}}$ ) and non-polar solvation free energy ( $\Delta G_{\text{SA}}$ ).

1  
2  
3  
4 The solvent accessible surface area (SASA) was used to compute the  $\Delta G_{SA}$ ; where  $\gamma$  and  $\beta$  are  
5 empirical constants for  $0.00542 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$  and  $0.92 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ , respectively. Per-residue  
6 decomposition analysis was employed to calculate the affinity and to assess the stability of the  
7 studied compounds through the estimation of the different energy contributions of important  
8 residues at the active site.  
9

## 14 **2.8. Pharmacokinetic and Toxicity Predictions**

16 The ADMET (i.e., Absorption (A), Distribution (D), Metabolism (M), Excretion (E), and  
17 Toxicity) predictions of the selected hit compounds were explored through the AdmetSAR and  
18 the Osiris property explorer servers to evaluate their drug likeness properties [38]. These  
19 properties explain the disposition of drugs inside an organism and consequently, impacts their  
20 pharmacological activity.  
21  
22  
23  
24  
25

## 26 **3. Results and discussion**

### 28 **3.1. QSAR modeling**

30 In the present work, a GA-MLR model was developed using 21  $\alpha$ -ketoamide derivatives reported  
31 as SARS-CoV inhibitors with a defined endpoint ( $IC_{50}$ ). The best-established model from the  
32 four selected molecular descriptors with its statistical parameters are shown below:  
33  
34

$$36 \text{pIC}_{50} = -17.0891 + 16.1547 * (\mathbf{GATS8i}) + 0.6723 * (\mathbf{NRS}) + 36.4881 * (\mathbf{G2p}) - 0.428 * (\mathbf{H8s})$$

37 (Equation 1)  
38

$$39 N_{tr} = 14, R_{tr}^2 = 0.91, RMSE_{tr} = 0.19, Q_{LOO}^2 = 0.78, Q_{lmo}^2 = 0.75, R_{test}^2 = 0.85,$$

40  $MAE_{test} = 0.26, Q_{F1}^2 = 0.82, Q_{F2}^2 = 0.81, Q_{F3}^2 = 0.80, CCC_{test} = 0.92, S = 0.24, \text{ and } F = 26.75.$   
41  
42

43 In this equation,  $N_{tr}$  is the number of training samples,  $CCC$  is the concordance correlation  
44 coefficient [39].  $Q_{F1}^2$ ,  $Q_{F2}^2$  and  $Q_{F3}^2$  are external validation criteria [40],  $S$  is the standard  
45 deviation,  $F$  is Fischer-ratio between the variances of calculated and observed activities  
46  
47  
48

49 All these statistical metrics that have been calculated for the established model are associated  
50 with fitting (i.e.  $R_{tr}^2$  and  $RMSE_{tr}$ ), internal (i.e.  $Q_{LOO}^2$  and  $Q_{lmo}^2$ ) and external validation (i.e.  $R_{test}^2$ ,  
51  $MAE_{test}$ ,  $Q_{F1}^2$ ,  $Q_{F2}^2$ ,  $Q_{F3}^2$ , and  $CCC_{test}$ ) indicate the statistical reliability of the QSAR model.  
52  
53 Indeed, these obtained parameters fulfill the recommended thresholds values and ensure the  
54 predictive power and stability of the derived GA-MLR model (Figure 3a). In addition, for a  
55 better validation of the built model, the AD assured by the leverage method and plotted as the  
56 Williams plot (Figure 3b), is used to assess the space of the AD of the created model. The  
57  
58  
59  
60  
61  
62  
63  
64  
65

warning line for the  $X$  outlier ( $h^*$ ) is 1. The dashed lines show the cutoff value of  $\pm 3$  standard deviation (s.d.). From the Williams plot, all compounds are within the scope of the AD.

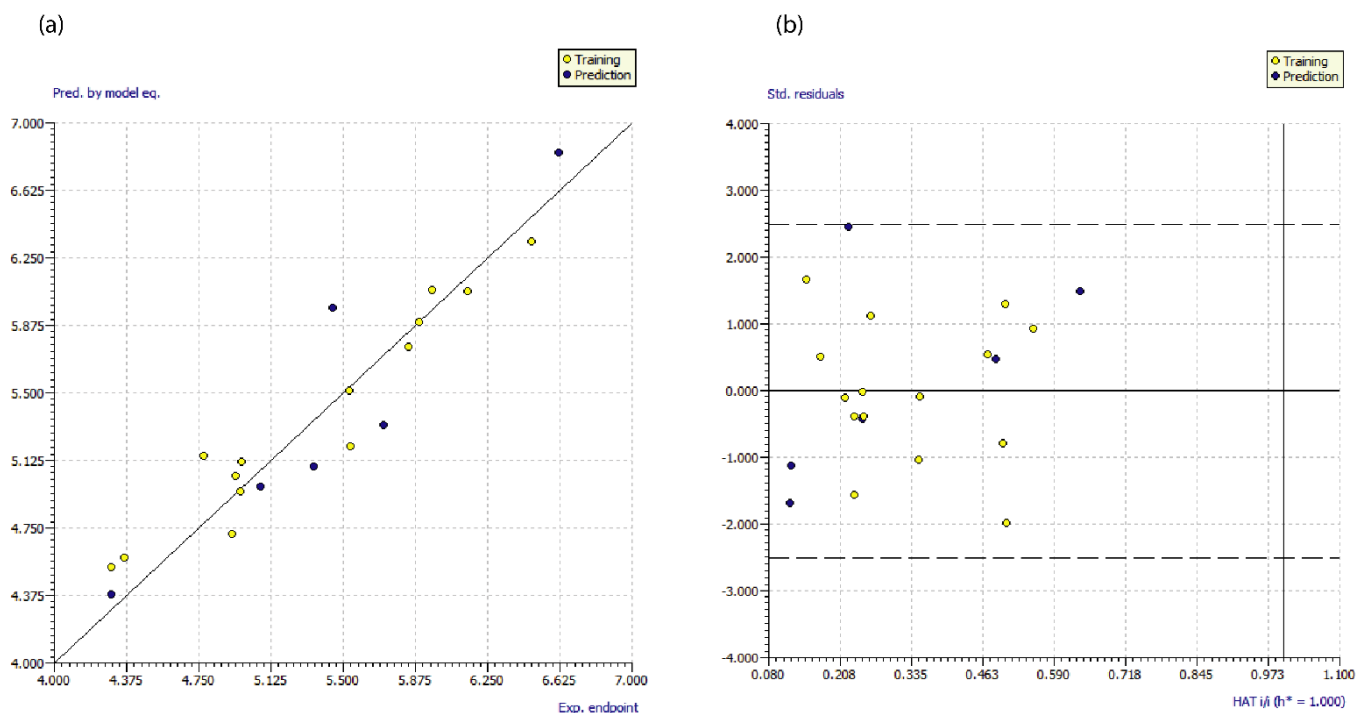
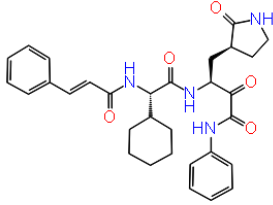
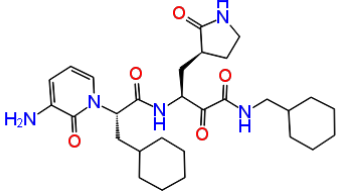
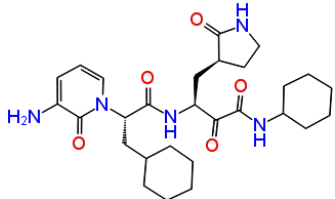
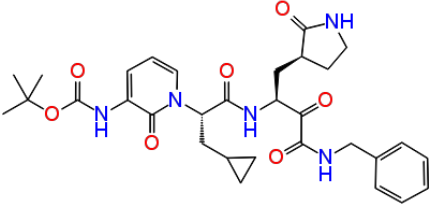


Figure 3. **a** | Plot of experimental vs. predicted pIC<sub>50</sub> values; **b** | Williams plot.

### 3.2. Ligand-based Pharmacophore Modeling

Ligand-based pharmacophore models were built from 21  $\alpha$ -ketoamide inhibitors. Molecular features involving hydrophobic interactions, HBA, and HBD were selected while generating the pharmacophore models. The best pharmacophore model was employed to select the best hits from the top 150 compounds. These hits were further subjected to molecular docking analysis to confirm their ability to be used as SARS-CoV-2 inhibitors. Table 1 shows the chemical structure of the three selected hits that match the different pharmacophore features illustrated in the Figure 2, their predicted pIC<sub>50</sub> using the GA-MLR model equation (Eq.1), their pharmacophore fit scores along with their binding affinity scores with the M<sup>PTO</sup> active site.

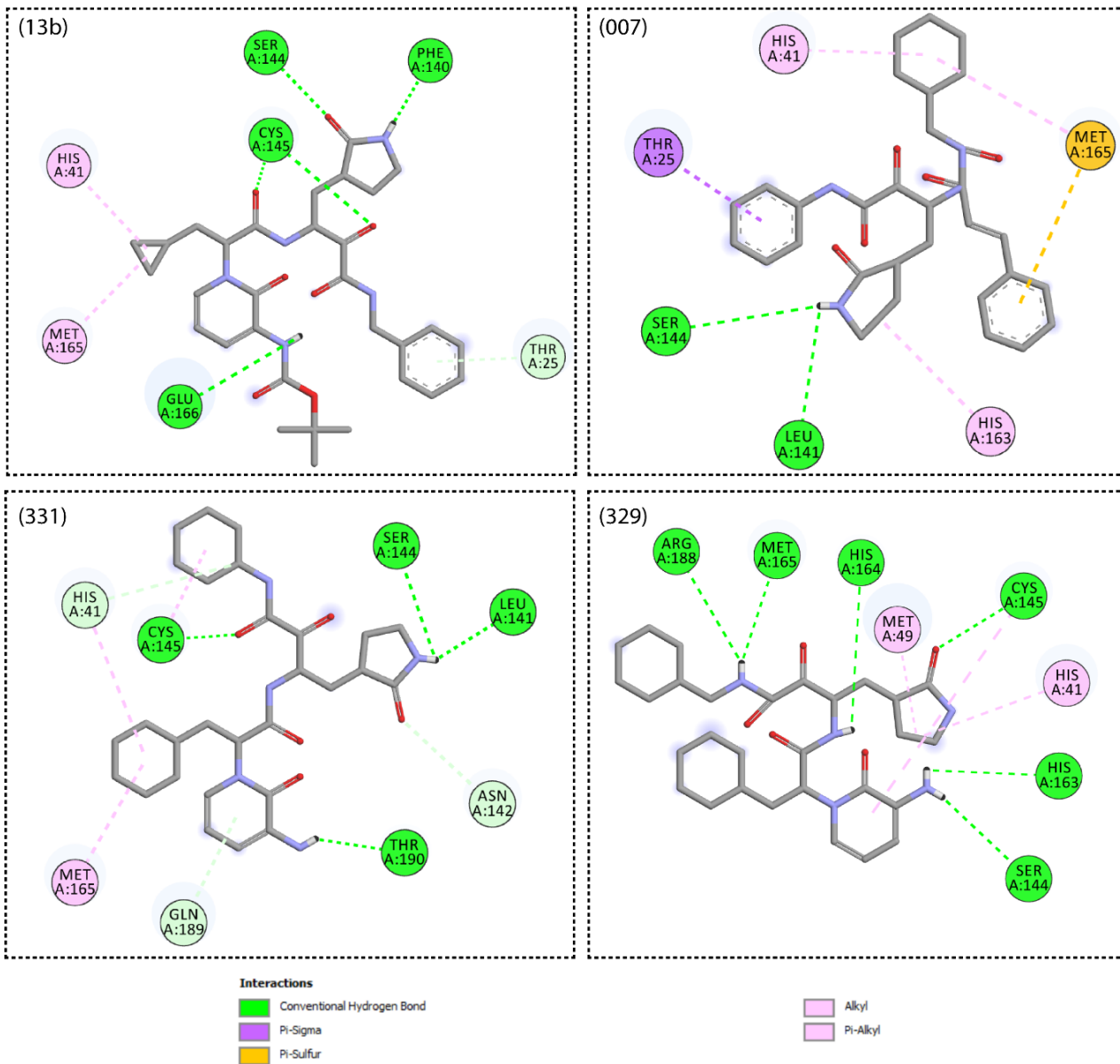
Table 1. Three hit compounds that well passed the filtration procedure.

Hit ID	Chemical structure	pIC <sub>50</sub> (Predicted)	Pharmacophore Fit score	Binding affinity (Kcal.mol <sup>-1</sup> )
<b>007</b>		6.0712	83.11	-7.9
<b>329</b>		6.2861	84.35	-7.2
<b>331</b>		6.4055	84.35	-8.2
<b>13b</b>		-	-	-7.7

### 3.3. Molecular Docking Analysis

The docking result and binding affinity estimation of the selected hit compounds and the reference re-docked inhibitor (**13b**) are shown in the Table 1. The interaction details with the active pocket amino-acids of SARS-CoV-2 M<sup>Pro</sup> are shown in Figure 4. The detailed amino acid residues are represented in Table S3 and the Figure S2 illustrates different hydrophobic and aromatic binding modes of **13b**, **007**, **329** and **331** ligands. The maximum range of energy differences between the studied ligands was 1 kcal.mol<sup>-1</sup>. It was found that the binding affinity score of the hit **329** (-7.2 kcal.mol<sup>-1</sup>) was higher in comparison to the reference compound **13b** (-7.7 kcal.mol<sup>-1</sup>). Unlike the hit compound **329**, the other two hits (**007** and **331**) showed lower binding affinity score (-7.9 and -8.2 kcal.mol<sup>-1</sup>) towards the M<sup>Pro</sup> active site than that of the reference ligand (**13b**). The interaction of the amino acid residues of the M<sup>Pro</sup> target with these three compounds were carefully analyzed using the discovery studio visualizer program [41]. It was evinced that all compounds efficiently interacted with different residues of domain I and II

1  
2  
3  
4 of the M<sup>pro</sup>. The compound **13b** was stabilized by five hydrogen bonds and three hydrophobic  
5 bonds while interacting with the active site of the receptor protein, in which it forms two  
6 hydrogen bonds with the catalytic residue CYS145 and one hydrophobic interaction with the  
7 HIS41. The interaction with the compound **007** was found to be stabilized by forming two  
8 hydrogen bonds, one pi-sulfur interaction, and four hydrophobic interactions where one of them  
9 was seen with the catalytic residue HIS41. The compound **329** formed six hydrogen bonding  
10 interactions in which one hydrogen bonding interaction was seen with the catalytic residue  
11 CYS145, along with three hydrophobic interactions where two of them interact with HIS41 and  
12 CYS145 catalytic dyad. Besides, the compound **331** gets stabilized by six hydrogen binding  
13 interactions where two of them interact with the catalytic residues CYS145 and HIS41, and two  
14 out of three hydrophobic interactions were seen to be formed with the catalytic residues CYS145  
15 and HIS41. Altogether, the analysis of non-covalent interactions between the selected three hit  
16 compounds and the M<sup>pro</sup> shows that the selected compounds interact either with the two key  
17 catalytic residues, or with at least one of them (i.e. CYS145 or HIS41), and can thus serve as key  
18 protease inhibitors ([Table S3](#)).  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



45 **Figure 4.** Interactions with key residues as exhibited by the hit compounds (**007**, **329** and **331**)  
46 and the reference ligand **13b** with the M<sup>PRO</sup> active site.  
47

### 48 **3.4. Molecular dynamics simulations**

#### 49 **3.4.1. Conformational stability of the M<sup>PRO</sup>**

50  
51  
52  
53  
54 The structural 3D stability of proteins could be estimated by measuring the overall atom  
55 deviations at the protein backbone. The Root-Mean-Square Deviation (RMSD) of the C- $\alpha$  atoms  
56 is a good parameter to estimate these variations. Great RMSD values can be an indicator of the  
57 expanded atom deviations at the protein backbone, also reduced RMSD values depict moderate  
58  
59  
60  
61  
62  
63  
64  
65



1  
2  
3  
4 variations in backbone atoms. This can support the mapping of the protein dynamics and the  
5 equilibrium of the different structures. In this regard, throughout the MD simulations, the  
6 deviations of the amino-acids were investigated via RMSD calculations. The results were plotted  
7 as a function of the simulations time in [Figure 5](#). As can be seen in the plots, all the complexes  
8 tend to converge around 40 ns and continue up to the end of the simulations showing deviations  
9 that differed across the simulations time. The binding of the compounds **007**, **329** and **331** to the  
10 M<sup>pro</sup> was characterized by a decrease in deviation among the backbone atoms relating to the  
11 unbound conformation that displayed higher variations. The overall average values of RMSD of  
12 complex**007**, complex**329**, complex**331** and M<sup>pro</sup> APO form are 1.88 Å, 1.95 Å, 1.94 Å and 2.08  
13 Å, with maximum recorded values of 2.91 Å, 3.14 Å, 2.94 Å and 3.64 Å, respectively. In  
14 opposition to the impact of reduced RMSD upon binding of the previous compounds, the  
15 reference compound **13b** exhibited greater RMSD values with overall average value of 2.42 Å  
16 which is slightly higher than that disclosed by other systems. Although molecular binding to M<sup>pro</sup>  
17 in general stabilizes the complexes with values that fall behind 3.5 Å and overall RMSD average  
18 values lower than 2.5 Å, a diverse impact on the stability of M<sup>pro</sup> upon the binding of the  
19 compounds **007**, **329** and **331** was observed as compared to the binding of **13b** which could be  
20 attributed to the stronger interaction of the predicted compounds in the active site of M<sup>pro</sup>.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

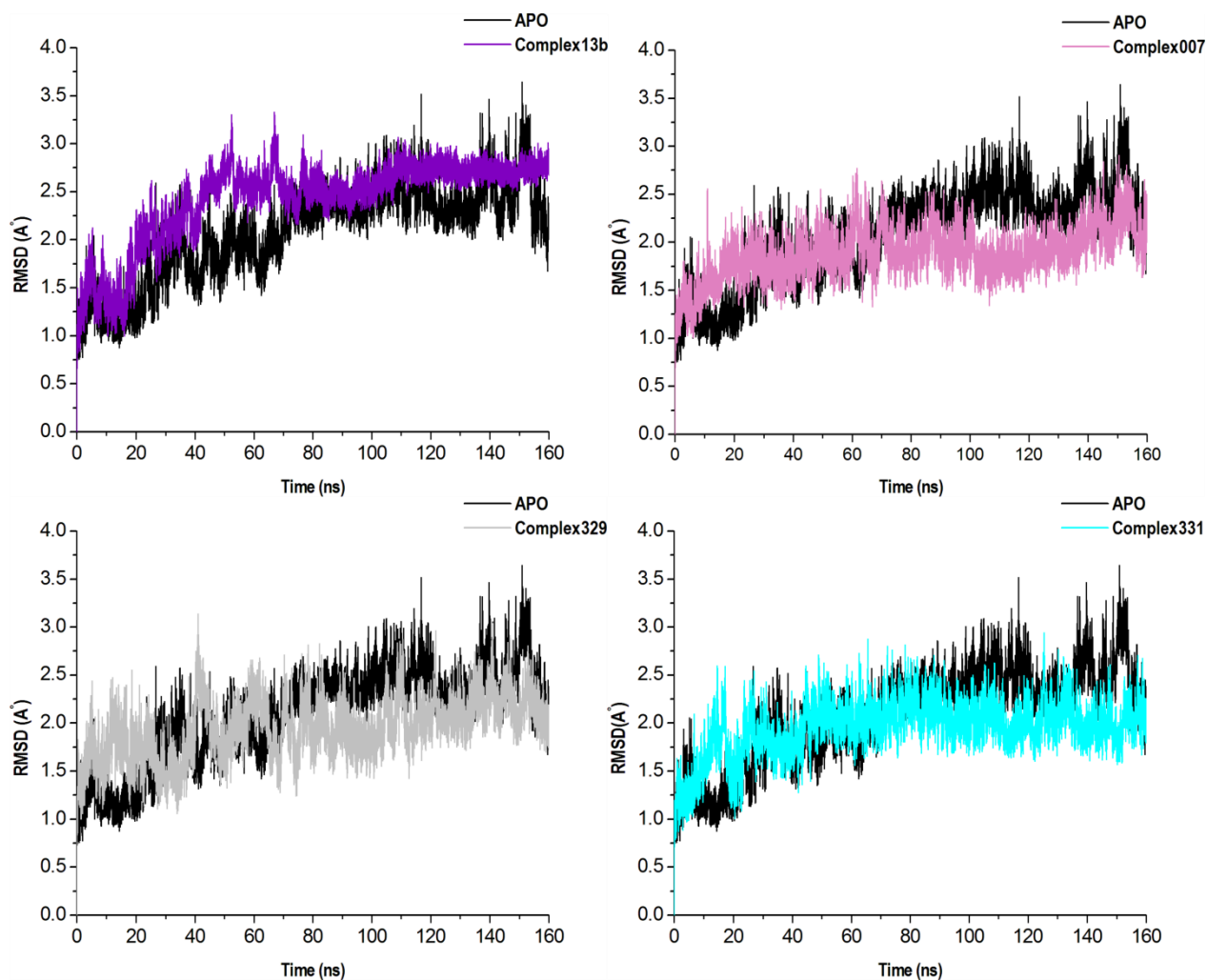
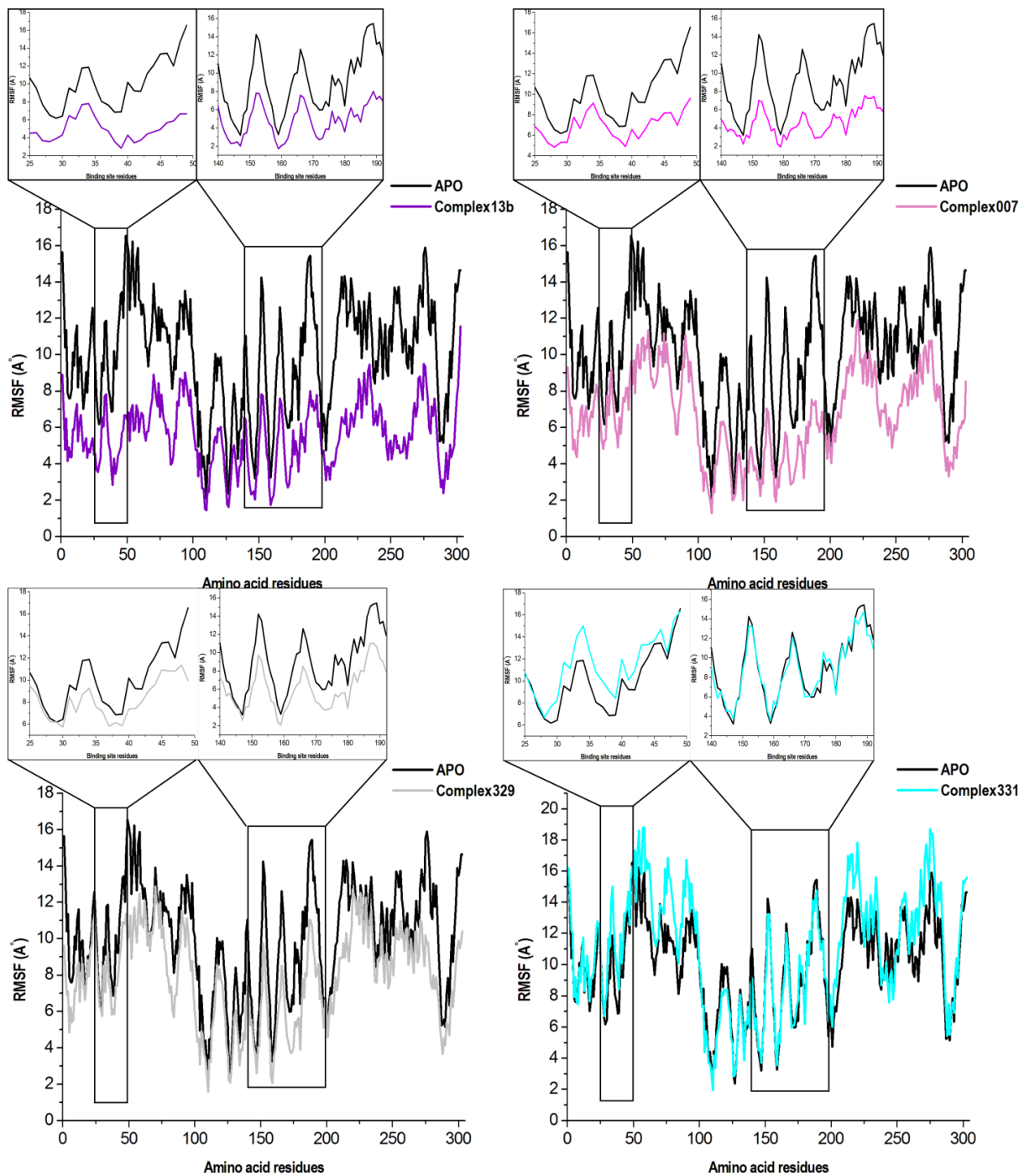


Figure 5. C- $\alpha$  backbone RMSD graphs of the three hits as well as the reference compound **13b** in complex with M<sup>pro</sup> when compared to the unbound M<sup>pro</sup>.

#### 3.4.2. Structural fluctuations of the M<sup>pro</sup>

The root mean square fluctuation (RMSF) is a valuable indicator of the structural behavior of the protein. The values depicted from RMSF elucidate the fluctuation of each amino acid residue as they interact with the ligand throughout the trajectory, which offers better insights into protein features. The RMSF values of each of the four complexes were compared to the RMSF of the unbound system (Figure 6). Based on the RMSF results, it was evident that the complexes **007**, **13b** and **329** exhibited less fluctuation than the unbound system especially in the catalytic region (residues 25-49, 118, 140-192) elucidating the attenuation of structural fluctuation in the presence of these ligands. The complex**331** displayed the greatest similarity to the unbound system in terms of structural changes. It was observed that the fluctuations occur with similar

1  
2  
3  
4 residues, particularly in the catalytic region 140-192. Nevertheless, subtle differences were  
5  
6 mainly observed outside this region. In the light of these observations, assessment of the  
7  
8 secondary structure was performed.  
9



10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57 **Figure 6.** The RMSF plots of the three hits as well as the reference compound **13b** in complex  
58 with the M<sup>pro</sup> in comparison with the unbound system.  
59  
60  
61  
62  
63  
64  
65

### 3.4.3. Secondary structure analysis

As seen in [Table 2](#), secondary structure analysis predicts a good amount of  $\beta$  strand (around 37% in all systems). The predicted percentage value of  $\alpha$  helix was around 28% in all complexes except for the complex**331** which showed slightly higher percentage of  $\alpha$  helix with value of 29.37%. However, other elements displayed modest contribution to the secondary structure. The detailed information is shown in [Table 2](#). The results revealed that  $\beta$  strand and  $\alpha$  helix dominated among secondary structure components. However, no significant change was seen in the secondary structure of the  $M^{pro}$  upon binding of the compounds. The secondary structure analysis described herein provides useful conformational insights on  $M^{pro}$  which might prompt to the design/discovery of selective SARS-CoV-2  $M^{pro}$  inhibitors. We assume that the application of rational approaches may contribute to the development of effective and safe antiviral agents against  $M^{pro}$  enzyme.

[Table 2](#): Contribution of different elements to the secondary structure of the  $M^{pro}$  enzyme (in %).

Percentage of Protein Secondary Structure %					
	$\alpha$ helix	$\beta$ strand	$3_{10}$ helix	Beta Turn	Bend
<b><math>M^{pro}</math> APO form</b>	28.36	37.17	4.38	19.38	10.72
<b>Complex<b>13b</b></b>	28.32	37.57	3.85	20.13	10.12
<b>Complex<b>007</b></b>	28.64	37.37	4.78	18.64	10.57
<b>Complex<b>329</b></b>	28.36	37.47	4.41	19.21	10.55
<b>Complex<b>331</b></b>	29.37	37.46	3.33	20.20	9.63

### 3.4.4. Binding free energy profiles of the selected hits

The ligand-binding thermodynamic energy is a significant parameter that contributes to the total binding free energy of the protein-ligand complex, surmounting the stabilizing forces of an inhibitor in the active site. Hence, the stability of the system throughout the simulation. To determine the basis for possible inhibition against the  $M^{pro}$  target, we computed the binding free energy and per-residue decomposition of each hit molecule at the active site. [Table 3](#) sums up the

binding free energy of the complexes, considering Van der Waals, solvation, and electrostatic energies. Also, the individual energy contributions of catalytic site residues are presented in the per-residue decomposition analyses in [Figure 7](#). In the light of these findings, the binding energies of complex**329** (-38.60 kcal.mol<sup>-1</sup>) and complex**331** (-47.16 kcal.mol<sup>-1</sup>) exhibited the most favorable binding when compared with complex**13b** and complex**007** (-35.01 and -31.51 kcal.mol<sup>-1</sup>, respectively).

**Table 3:** MMPBSA-based binding free energy profiles of the reference compound **13b** and compounds **007**, **329** and **331** at the binding pocket of the M<sup>Pro</sup>. The total binding free energies are highlighted in bold to distinguish them from the elements of the overall values.

Energy Components (kcal.mol <sup>-1</sup> )					
Complex	$\Delta E_{vdW}$	$\Delta E_{elec}$	$\Delta G_{gas}$	$\Delta G_{solv}$	$\Delta G_{bind}$
Complex <b>13b</b>	-49.66 ± 6.18	-22.09 ± 8.83	-71.75 ± 13.1	36.74 ± 6.91	<b>-35.01 ± 7.22</b>
Complex <b>007</b>	-43.07 ± 8.00	-17.86 ± 9.35	-60.93 ± 15.7	29.42 ± 9.09	<b>-31.51 ± 7.51</b>
Complex <b>329</b>	-51.50 ± 5.67	-17.21 ± 6.98	-68.72 ± 9.77	30.11 ± 5.51	<b>-38.60 ± 6.71</b>
Complex <b>331</b>	-57.47 ± 6.69	-35.04 ± 12.22	-92.51 ± 12.56	45.34 ± 8.06	<b>-47.16 ± 6.26</b>

The binding energy observed in complex**331** was associated to higher van der Waals and electrostatic energy between the amino acid residues ASN140, GLY141, SER142 and CYS143 and the compound **331**. However, the optimal binding energy noticed in complex**329** was due to increased electrostatic and van der Waals interactions between the residues HIS41, ASN49, MET163, GLN187 and the compound. Residues THR167 and GLY168 contributed minor interactions in all four complexes; this may have been attributed to the position of the amino-acids at the shallow region of the binding site. Although the assessed binding free energies are not absolute values compared to the experimental ones, they are still reliable considering the residue interactions within the binding site area.

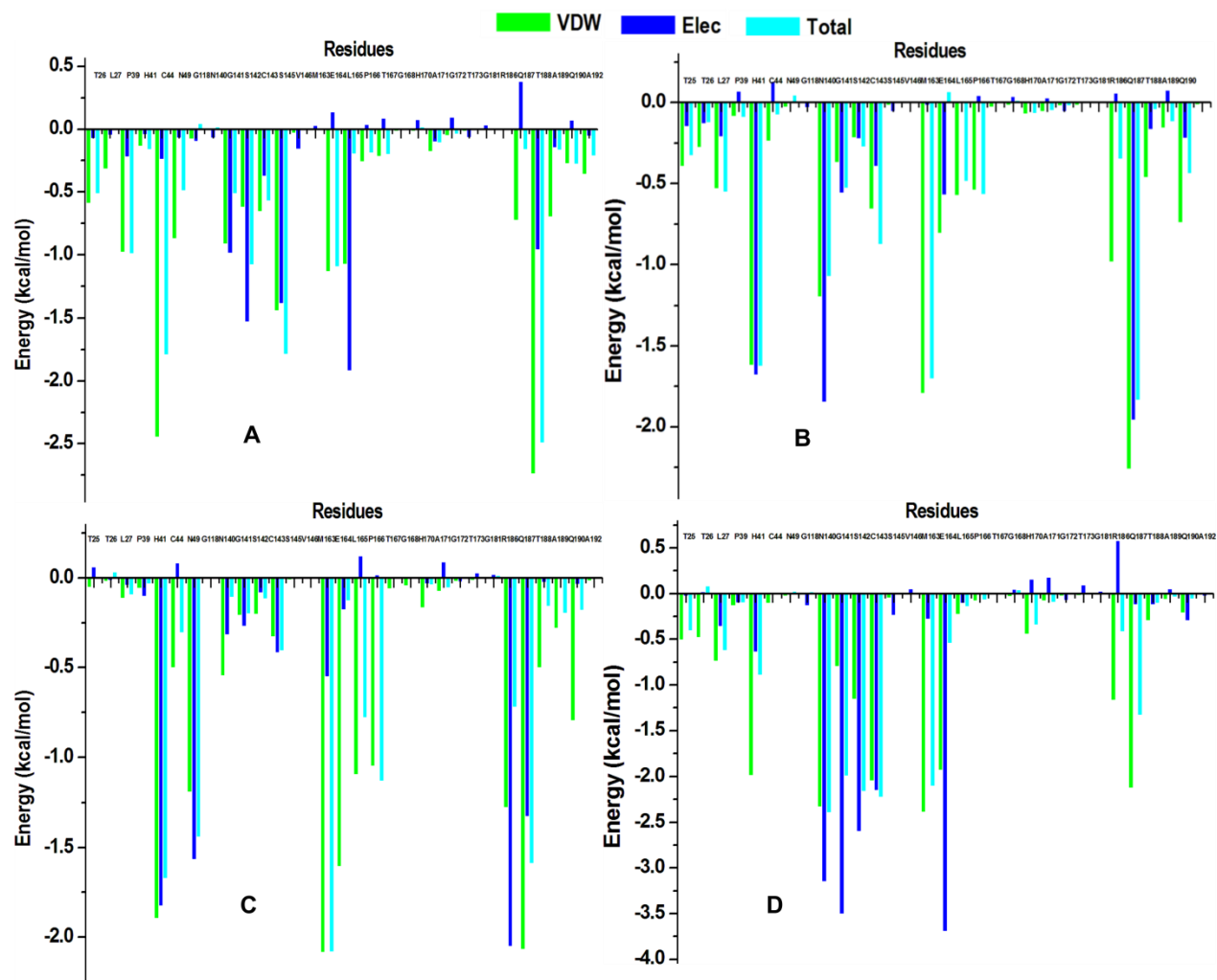


Figure 7. Per-residue decomposition analyses demonstrating the role of individual energy contributions of catalytic site residues to the stability and binding of the molecules **13b** (A), **007** (B), **329** (C) and **331** (D).

### 3.5. Pharmacokinetic and Toxicity Predictions

To be an auspicious drug candidate against biological target, several pharmacokinetic propriety studies must be prioritized at early drug discovery phases (before the clinical trial stage). Not only would this increase the overall quality of drug candidates, but also their success probability, in order to shorten the phase of drug discovery.

The generated result of the selected hits from the ADMET filtering analyses are represented in Tables 4 and 5. The selected hit compounds have no risks of mutagenic, tumorigenic, irritant, or reproductive effect profiles. All three compounds have better penetration score through the



1  
2  
3  
4 blood-brain barrier (BBB) than the **13b** compound. While the absorption from the intestinal tract  
5 upon oral administration test turns out to be slightly better for the **007** compound.  
6  
7

8 **Table 4.** Toxicological properties of the selected compounds and the **13b** inhibitor assessed  
9 through AdmetSAR and Osiris property explorer.  
10  
11

Toxicological properties	<b>007</b>	<b>329</b>	<b>331</b>	<b>13b</b>
Mutagenic	N	N	N	N
Tumorigenic	N	N	N	N
Irritant	N	N	N	N
Reproductive effect	N	N	N	N

12  
13  
14  
15  
16  
17  
18  
19  
20  
21 N= No risk

22 **Table 5.** Pharmacokinetic and ADME properties of the selected compounds and the **13b** inhibitor  
23 assessed through the AdmetSAR and the Osiris property explorer.  
24  
25

Pharmacokinetic properties	<b>007</b>	<b>329</b>	<b>331</b>	<b>13b</b>
Molecular weight (g.mol <sup>-1</sup> )	544.65	541.69	527.66	579.65
cLog P	2.35	0.41	0.22	1.1
Solubility	-5.32	-4.47	-4.41	-4.84
TPSA (Å <sup>2</sup> )	133.47	150.7	150.7	163.01
HBA	9	10	10	12
HBD	4	4	4	4
BBB	0.96	0.96	0.96	0.93
HIA	0.96	0.93	0.93	0.94

26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 TPSA: total polar surface area; HIA: human intestinal absorption.

## 42 **4. Conclusion**

43  
44  
45 The main protease (M<sup>pro</sup>) has been an appealing target for discovering new SARS-CoV-2  
46 replication inhibitors, as it is an important protein in post-translational processing of replicase  
47 polyproteins. This study attempts to identify novel potent  $\alpha$ -ketoamide based inhibitors targeting  
48 the M<sup>pro</sup> of SARS-CoV-2. To accomplish this goal, QSAR and pharmacophore models were  
49 developed from the structural features of a series of  $\alpha$ -ketoamide derivatives that had shown an  
50 inhibition effect against the M<sup>pro</sup> of SARS-CoV. Three hits were retrieved by utilizing a  
51 constructed GA-MLR model and by applying a pharmacophore fit model. Moreover, Molecular  
52 dynamic simulations were used as a computational validation of the compounds. Overall, the  
53 three evaluated hits displayed important conformational and structural stability as well as  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 promising binding profiles. In addition, it has been shown that the suggested hit molecules are  
5 non-toxic and have acceptable pharmacological properties. The outcomes of this study clearly  
6 show that the three screened compounds may lead to potent anti-SARS-CoV-2 M<sup>pro</sup> drug  
7 molecules where the binding details and the nature of activity were considered to be sufficient  
8 for blocking the active site of the M<sup>pro</sup>. The ensemble of computational methods implemented  
9 herein allowed for the discovery of new drug candidates with potential inhibitory activity against  
10 the SARS-CoV-2 M<sup>pro</sup>. These hit compounds can be further biologically evaluated for their  
11 potency and physiological toxicity.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

## 22 **Acknowledgment**

23  
24 The authors would like to thank Prof. Paola Gramatica for providing a free copy of the  
25 QSARINS Software.  
26  
27

## 28 **Disclosure statement**

29  
30 All the authors declare no conflict of interest.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## References

- [1] Worldometer, Coronavirus Cases, Worldometer. 368 (2020) 1–22. <https://doi.org/10.1101/2020.01.23.20018549V2>.
- [2] D. Fisman, C. Rivers, E. Lofgren, M.S. Majumder, Estimation of MERS-Coronavirus Reproductive Number and Case Fatality Rate for the Spring 2014 Saudi Arabia Outbreak: Insights from Publicly Available Data, *PLoS Curr.* 6 (2014). <https://doi.org/10.1371/currents.outbreaks.98d2f8f3382d84f390736cd5f5fe133c>.
- [3] E. Petersen, M. Koopmans, U. Go, D.H. Hamer, N. Petrosillo, F. Castelli, M. Storgaard, S. Al Khalili, L. Simonsen, Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics, *Lancet Infect. Dis.* 20 (2020) e238–e244. [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9).
- [4] D. Kim, J.Y. Lee, J.S. Yang, J.W. Kim, V.N. Kim, H. Chang, The Architecture of SARS-CoV-2 Transcriptome, *Cell.* 181 (2020) 914–921.e10. <https://doi.org/10.1016/j.cell.2020.04.011>.
- [5] F.K. Yoshimoto, The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19, *Protein J.* 39 (2020) 198–216. <https://doi.org/10.1007/s10930-020-09901-4>.
- [6] 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>.
- [7] A.M. Lesk, A.S. Konagurthu, L. Allison, M. Garcia de la Banda, P.J. Stuckey, D. Abramson, Computer modeling of a potential agent against SARS-Cov-2 (COVID-19) protease., *Proteins.* 88 (2020) 1557–1558. <https://doi.org/10.1002/prot.25980>.
- [8] L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E.J. Snijder, H. Liu, R. Hilgenfeld,  $\alpha$ -Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment, *J. Med. Chem.* (2020) [acs.jmedchem.9b01828](https://doi.org/10.1021/acs.jmedchem.9b01828). <https://doi.org/10.1021/acs.jmedchem.9b01828>.
- [9] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G.

1  
2  
3  
4 Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, H.P.H. X.  
5  
6 Li, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota,  
7  
8 R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T.  
9  
10 Vreven, J.A.M. Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N.  
11  
12 Kudin, V.N. Staroverov, R. Kobayashi, Normand, J.R. Rendell, K.A.J.C. Burant, S.S.  
13  
14 Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V.  
15  
16 Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R.  
17  
18 Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A.  
19  
20 Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D.D.O. Farkas, J.B. Foresman, J.V.  
21  
22 Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Wallingford CT, (2009). <http://gaussian.com/>.

23 [10] I. Sushko, S. Novotarskyi, R. Körner, A.K. Pandey, M. Rupp, W. Teetz, S. Brandmaier,  
24  
25 A. Abdelaziz, V. V. Prokopenko, V.Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P.  
26  
27 Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I.I. Baskin, V.A. Palyulin, E. V.  
28  
29 Radchenko, W.J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-De-  
30  
31 Sousa, Q.Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko, I. V.  
32  
33 Tetko, Online chemical modeling environment (OCHEM): Web platform for data storage,  
34  
35 model development and publishing of chemical information, *J. Comput. Aided. Mol. Des.*  
36  
37 *25* (2011) 533–554. <https://doi.org/10.1007/s10822-011-9440-2>.

38 [11] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: A new software  
39  
40 for the development, analysis, and validation of QSAR MLR models, *J. Comput. Chem.*  
41  
42 *34* (2013) 2121–2132. <https://doi.org/10.1002/jcc.23361>.

43  
44 [12] P. Gramatica, S. Cassani, N. Chirico, QSARINS-chem: Insubria datasets and new  
45  
46 QSAR/QSPR models for environmental pollutants in QSARINS, *J. Comput. Chem.* *35*  
47  
48 (2014) 1036–1044. <https://doi.org/10.1002/jcc.23576>.

49  
50 [13] A. Palmeira, E. Sousa, A. Kösele, R. Sabirli, T. Gören, İ. Türkçüer, Ö. Kurt, M.M. Pinto,  
51  
52 M.H. Vasconcelos, Preliminary virtual screening studies to identify grp78 inhibitors  
53  
54 which may interfere with sars-cov-2 infection, *Pharmaceuticals*. *13* (2020) 1–13.  
55  
56 <https://doi.org/10.3390/ph13060132>.

57  
58 [14] A. Jiménez-Alberto, R.M. Ribas-Aparicio, G. Aparicio-Ozores, J.A. Castelán-Vega,  
59  
60 Virtual screening of approved drugs as potential SARS-CoV-2 main protease inhibitors,  
61  
62  
63  
64  
65

- 1  
2  
3  
4 Comput. Biol. Chem. 88 (2020) 107325.  
5  
6 <https://doi.org/10.1016/j.compbiolchem.2020.107325>.  
7  
8  
9 [15] S. Choudhary, Y.S. Malik, S. Tomar, Identification of SARS-CoV-2 Cell Entry Inhibitors  
10 by Drug Repurposing Using in silico Structure-Based Virtual Screening Approach, *Front.*  
11 *Immunol.* 11 (2020) 1664. <https://doi.org/10.3389/fimmu.2020.01664>.  
12  
13  
14 [16] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R.  
15 Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of  
16 improved  $\alpha$ -ketoamide inhibitors, *Science* (80-. ). 3405 (2020) eabb3405.  
17  
18 <https://doi.org/10.1126/science.abb3405>.  
19  
20  
21  
22 [17] J.J. Irwin, B.K. Shoichet, ZINC - A free database of commercially available compounds  
23 for virtual screening, *J. Chem. Inf. Model.* 45 (2005) 177–182.  
24  
25 <https://doi.org/10.1021/ci049714+>.  
26  
27  
28 [18] D. Zeng, Y. Ma, R. Zhang, Q. Nie, Z. Cui, Y. Wang, L. Shang, Z. Yin, Synthesis and  
29 structure–activity relationship of  $\alpha$ -keto amides as enterovirus 71 3C protease inhibitors,  
30 *Bioorganic Med. Chem. Lett.* 26 (2016) 1762–1766.  
31  
32 <https://doi.org/10.1016/j.bmcl.2016.02.039>.  
33  
34  
35  
36 [19] A. Vuorinen, R. Engeli, A. Meyer, F. Bachmann, U.J. Griesser, D. Schuster, A. Odermatt,  
37 Ligand-based pharmacophore modeling and virtual screening for the discovery of novel  
38  $17\beta$ -hydroxysteroid dehydrogenase 2 inhibitors, *J. Med. Chem.* 57 (2014) 5995–6007.  
39  
40 <https://doi.org/10.1021/jm5004914>.  
41  
42  
43  
44 [20] W.R. Ferraz, R.A. Gomes, A.L. S Novaes, G.H. Goulart Trossini, Ligand and structure-  
45 based virtual screening applied to the SARS-CoV-2 main protease: an in silico  
46 repurposing study, *Future Med. Chem.* (2020) fmc-2020-0165.  
47  
48 <https://doi.org/10.4155/fmc-2020-0165>.  
49  
50  
51  
52 [21] LigandScout, Version 4.3; Inte:Ligand GmbH, Clemesn-Maria-Hofbauer-G. 6, 2344,  
53 Maria Enzersdorf, Austria. HYPERLINK, <http://www.inteligand.com>, (n.d.).  
54  
55 <http://www.inteligand.com>.  
56  
57  
58 [22] P.T. Mpiana, K. te N. Ngbolua, D.S.T. Tshibangu, J.T. Kilembe, B.Z. Gbolo, D.T.  
59 Mwanangombo, C.L. Inkoto, E.M. Lengbiye, C.M. Mbadiko, A. Matondo, G.N. Bongo,  
60  
61  
62  
63  
64  
65

- 1  
2  
3  
4 D.D. Tshilanda, Identification of potential inhibitors of SARS-CoV-2 main protease from  
5 Aloe vera compounds: A molecular docking study, *Chem. Phys. Lett.* 754 (2020) 137751.  
6 <https://doi.org/10.1016/j.cplett.2020.137751>.  
7  
8  
9
- [23] R.K. Hussein, H.M. Elkhair, Molecular docking identification for the efficacy of some  
10 zinc complexes with chloroquine and hydroxychloroquine against main protease of  
11 COVID-19, *J. Mol. Struct.* 1231 (2021) 129979.  
12 <https://doi.org/10.1016/j.molstruc.2021.129979>.  
13  
14  
15  
16  
17
- [24] RCSB PDB - 6Y2F: Crystal structure (monoclinic form) of the complex resulting from the  
18 reaction between SARS-CoV-2 (2019-nCoV) main protease and tert-butyl (1-((S)-1-(((S)-  
19 4-(benzylamino)-3,4-dioxo-1-((S)-2-oxopyrrolidin-3-yl)butan-2-yl)amino)-3-cycloprop,  
20 (n.d.). <https://www.rcsb.org/structure/6Y2F> (accessed September 15, 2020).  
21  
22  
23  
24  
25
- [25] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J.  
26 Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor  
27 flexibility, *J. Comput. Chem.* 30 (2009) 2785–2791. <https://doi.org/10.1002/jcc.21256>.  
28  
29  
30  
31
- [26] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with  
32 a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31  
33 (2009) NA-NA. <https://doi.org/10.1002/jcc.21334>.  
34  
35  
36  
37
- [27] T. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III,  
38 V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W.  
39 Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny,  
40 AMBER Software, (n.d.) AMBER 2018, University of California, San Francisc.  
41  
42  
43  
44  
45
- [28] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Antechamber, An Accessory Software  
46 Package For Molecular Mechanical Calculations, *J. Am. Chem. Soc.* 222 (2001) U403.  
47  
48  
49
- [29] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of  
50 multiple Amber force fields and development of improved protein backbone parameters,  
51 *Proteins Struct. Funct. Bioinforma.* 65 (2006) 712–725.  
52 <https://doi.org/10.1002/prot.21123>.  
53  
54  
55  
56
- [30] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing  
57 of a general Amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174.  
58  
59  
60  
61  
62  
63  
64  
65



- 1  
2  
3  
4 <https://doi.org/10.1002/jcc.20035>.
- 5  
6  
7 [31] M.F. Harrach, B. Drossel, Structure and dynamics of TIP3P, TIP4P, and TIP5P water near  
8 smooth and atomistic walls of different hydroaffinity, *J. Chem. Phys.* 140 (2014) 174501.  
9 <https://doi.org/10.1063/1.4872239>.
- 10  
11  
12 [32] M.J. Harvey, G. Giupponi, G. De Fabritiis, ACEMD: Accelerating biomolecular dynamics  
13 in the microsecond time scale, *J. Chem. Theory Comput.* 5 (2009) 1632–1639.  
14 <https://doi.org/10.1021/ct9000685>.
- 15  
16  
17 [33] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian  
18 equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J.*  
19 *Comput. Phys.* 23 (1977) 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- 20  
21  
22 [34] Y. Lin, D. Pan, J. Li, L. Zhang, X. Shao, Application of Berendsen barostat in dissipative  
23 particle dynamics for nonequilibrium dynamic simulation, *J. Chem. Phys.* 146 (2017)  
24 124108. <https://doi.org/10.1063/1.4978807>.
- 25  
26  
27 [35] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for processing and analysis of  
28 molecular dynamics trajectory data, *J. Chem. Theory Comput.* 9 (2013) 3084–3095.  
29 <https://doi.org/10.1021/ct400341p>.
- 30  
31  
32 [36] E. Seifert, OriginPro 9.1: Scientific Data Analysis and Graphing Software—Software  
33 Review, *J. Chem. Inf. Model.* 54 (2014) 1552–1552. <https://doi.org/10.1021/ci500161d>.
- 34  
35  
36 [37] K. Huang, S. Luo, Y. Cong, S. Zhong, J.Z.H. Zhang, L. Duan, An accurate free energy  
37 estimator: Based on MM/PBSA combined with interaction entropy for protein-ligand  
38 binding affinity, *Nanoscale.* 12 (2020) 10737–10750. <https://doi.org/10.1039/c9nr10638c>.
- 39  
40  
41 [38] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate  
42 pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules,  
43 *Sci. Rep.* 7 (2017) 42717. <https://doi.org/10.1038/srep42717>.
- 44  
45  
46 [39] N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models: How To Evaluate  
47 It? Comparison of Different Validation Criteria and Proposal of Using the Concordance  
48 Correlation Coefficient, *J. Chem. Inf. Model.* 51 (2011) 2320–2335.  
49 <https://doi.org/10.1021/ci200211n>.
- 50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

- [40] N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection, *J. Chem. Inf. Model.* 52 (2012) 2044–2058. <https://doi.org/10.1021/ci300084j>.
- [41] B.D. Systèmes, Discovery Studio Modeling Environment, San Diego Dassault Systèmes. (n.d.). <http://accelrys.com/products/collaborative-science/biovia-discovery-studio/> (accessed April 15, 2018).