Manuscript Click here to view linked References

COVID-19: In silico identification of potent α-ketoamide inhibitors targeting the main protease of the SARS-CoV-2

Mehdi Oubahmane [1], Ismail Hdoufane [1], Imane Bjij [1], Carola Jerves [2,3], Didier Villemin [4], and Driss Cherqaoui [1,*]

[1] Department of Chemistry, Faculty of Sciences Semlalia, BP, 2390, Marrakech, Morocco
[2] Departamento de Quimica, Faculdade de Ciencias da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto - Portugal.
[3] Facultad de Ciencias Químicas, Universidad de Cuenca. Cuenca, Ecuador
[4] 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

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 α-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 α-ketoamide-based compounds. In this regard, a set of 21 α-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}); α-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 (15th March, 2021) [1]. In the 21st century, SARS-CoV-2 is the 3rd *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 (PLpro, *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 (3CLpro)) 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.

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

- ✓ Establishing a reliable QSAR model using GA-MLR method that can be able to predict the SARS-CoV-2 M^{pro} inhibitory activity of α-ketoamides.
- ✓ Developing a pharmacophore fit model based on the structural features of the studied SARS-CoV-2 M^{pro} inhibitors.
- ✓ Proposing new drug candidates and screening for new hits through the investigation of the effect of active groups.
- ✓ Elucidating the dynamics of the complexes and the underlying inhibition mode of the proposed hits at the active pocket of the target protein and evaluating their pharmacological profiles.

2. Materials and methods

2.1. Data collection and molecular descriptor calculation

A dataset of 21 compounds of α-ketoamides derivatives with SARS-CoV inhibitory activity were collected from the published work by Zhang et al. (Table S1) [8]. The retrieved IC_{50} values were converted to their corresponding pIC_{50} ($-\log_{10}$ (IC_{50})) and used as a dependent variable. The ChemDraw 18.0 software was used to draw the chemical structures and their geometries were optimized using the AM1 method in the gas phase. This optimization was implemented in the Gaussian 09 program package [9]. Frequency analysis was used to verify the energy minima of the optimized samples.

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

from the initial computed pool was selected. Finally, four descriptors listed in Table S2 provided the best QSAR model.

2.2. 2D-QSAR Model Construction and Validation

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

(i). The dataset comprising the 21 α-ketoamide analogs was randomly divided, using the splitting option in the QSARINS software, into 70% training set (14 compounds) and 30% test set (7 compounds). The training set was employed for model construction and the test set for the external validation.

(ii). The GA-MLR 2D-QSAR model was constructed using default parameters of the QSARINS program. During the model development, the selected molecular descriptors were utilized to derive a simple and an informative GA-MLR model.

(iii). The confirmation of the model validity was proved by subjecting the established model to internal and external validations, Y-randomization, and model applicability domain (AD) analysis. Indeed, the statistical quality and robustness of the GA-MLR-based 2D-QSAR model was ensured on the basis of: (a) internal validation based on leave-one-out (LOO) and leave-many-out (LMO) procedure (i.e. cross-validation (CV)); (b) external validation; (c) Y-randomization; and (d) fulfillment of respective threshold values for the statistical metrics: the determination coefficient of the training set (R^2_{tr}) ≥ 0.6, the square of the CV correlation coefficient obtained from the leave-one-out (LOO) procedure (Q^2_{loo}) ≥ 0.5, the leave-many-out (LMO) Q^2_{lmo} ≥ 0.6, and the square of correlation coefficient obtained for the test set (R^2_{test}) ≥ 0.6. Finally, the root-mean square error (*RMSE*) and the mean absolute error (*MAE*) values should be close to zero. Any QSAR model that did not satisfy any of these criteria is therefore omitted.

2.3. Virtual Screening Database Preparation

Virtual screening (VS) is a computational method employed to screen available chemical databases for filtering molecules that are most likely to bind to a drug target by mapping them on generated chemoinformatic models [13]. The hit identification, using VS of compounds, is among the most popular computational techniques in drug design [14,15]. In this work, the screened database was prepared using different strategies and a total number of 713 molecules was generated. The first part was formed based on an initial pharmacophore model created by the ZincPharmer online server. Chemical features from the core structure of **13b** and **11r** compounds (Figure 1), possessing significant SARS-CoV-2 inhibitory activities, were employed to generate initial pharmacophore models (Figure S1) [16]. These models were applied to retrieve a total number of 197 compounds from the Zinc database [17]. The second part comprising 484 molecules was generated based on several modifications at the core structure of compounds **13b** and **11r** using the "Expand Generic structure" option implemented in the ChemDraw software. The modifications were made based on the optimization of P1', P2, and P3 substituents in compounds **13b** and **11r** (Figure 1). The P1 moiety was kept intact during the modifications due to its vital role in the interaction with the active pocket of the M^{pro}. The last part which contains 32 molecules was taken from a previously published series of α-ketoamide inhibitors [18].

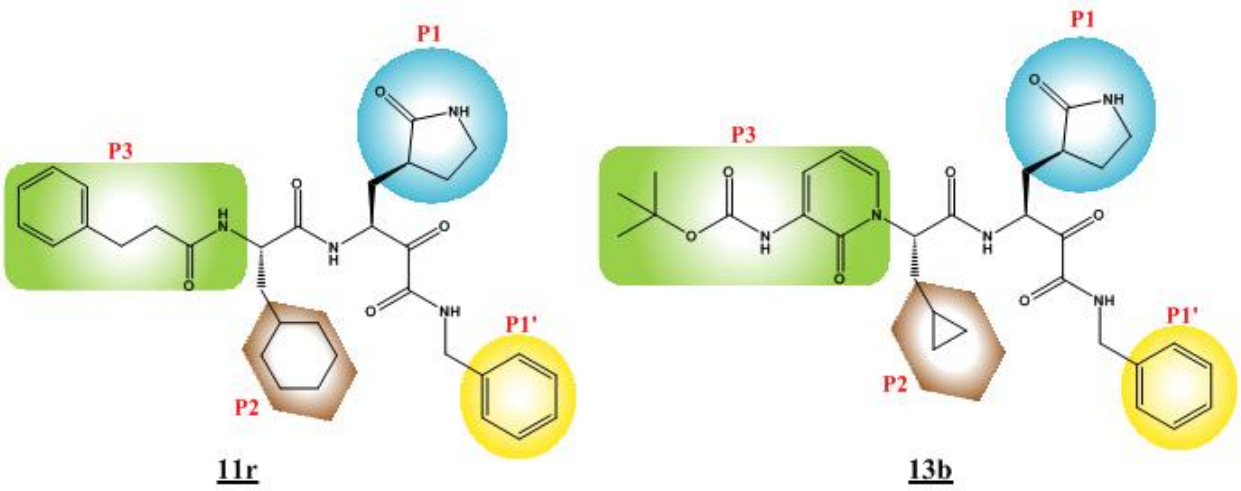


Figure 1. Chemical structures of α-ketoamide inhibitors **11r** and **13b**. Colored circles highlight the substituents where the modifications were made.

Herein, the identified molecules were subjected to two filters. Firstly, the above-mentioned GA-MLR model was employed in order to screen for compounds with the highest pIC_{50} values from the list of 713 generated ones. Secondly, the features fitting of each compound were

evaluated through the application of advanced pharmacophore fit model, facilitating the ranking of the screened molecules from the best GA-MLR QSAR model.

2.4. Ligand-based Pharmacophore model

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

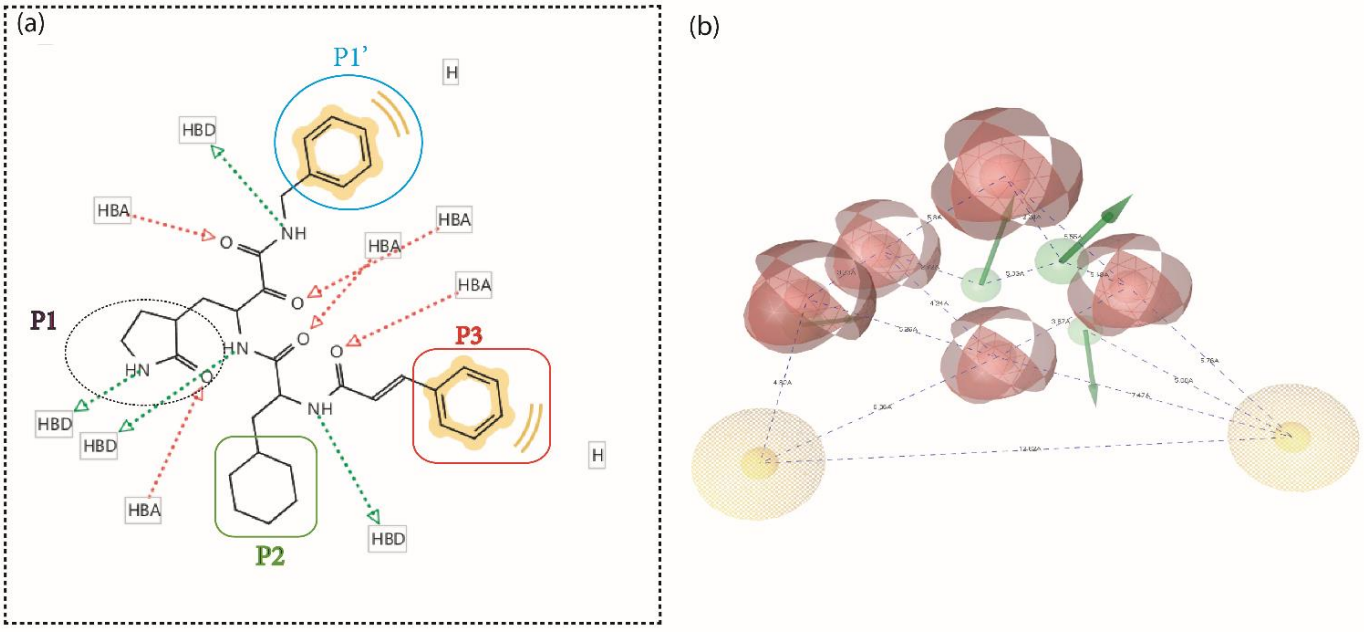


Figure 2. **a** | Ligand-based pharmacophore model generated by LigandScout software, showing different features including HBA, HBD, and Hydrophobic (H).

b | Pharmacophore model with distance constraints.

2.5. Molecular Docking of Selected Hits

Molecular docking analyses have been widely employed in medicinal chemistry research [22,23]. In the current work, molecular docking was employed to analyze the interactions and the conformational patterns of the selected hits within the active site of the M^{pro} of the SARS-CoV-2. The crystallographic protein of the M^{pro} was downloaded from the Protein Data Bank with the PDB code 6Y2F [16,24]. It is a monoclinic form of the complex resulting from the reaction between SARS-CoV-2 M^{pro} and the α-ketoamide inhibitor (**13b**). For the protein preparation, the co-crystallized ligand and the water molecules were removed and the hydrogens were added using Autodock Tools [25]. The selected hits were prepared and saved in pdbqt format using the same software. In the docking investigations, a box of 20 × 20 × 20 and a grid space of 0.375 Å were fixed. The center of the ligands was set using the following coordinates x = 10.88 Å, y = -0.25 Å and z = 20.75 Å. The mode number was fixed at 8. All the docking runs were performed by the Autodock Vina program [26].

2.6.Molecular Dynamics Simulations

Molecular dynamics (MD) simulation is an important approach in understanding the fundamental basis of the biological macromolecule structures, which can provide critical insights on their function and dynamics. Therefore, AMBER18 software package [27] was utilized to perform 160 ns of MD simulations. Five systems were simulated: complex**007 (**compound **007** in complex with M^{pro}), complex**329 (**compound **329** in complex with M^{pro}), complex**331 (**compound **331** in complex with M^{pro}), complex**13b** (the reference compound **13b** in complex with M^{pro}), and **APO** (the unbound form of M^{pro}). Prior to the MD simulations, the best-docked pose for each ligand was parameterized using Antechamber [28]. The TLeap module was used to ensemble the system, implementing the FF99SB force field to parameterize the protein. The general AMBER force field (GAFF) was employed to determine the partial charges of atoms in

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 $kcal.mol^{-1}.Å^{-2}$, 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 $kcal.mol^{-1}.Å^{-2}$. 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^{pro} 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:

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

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

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

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

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

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

The solvent accessible surface area (SASA) was used to compute the ΔG_{SA}; where γ and β are empirical constants for 0.00542 kcal.mol^{-1}.Å^{-2} and 0.92 kcal.mol^{-1}.Å^{-2}, respectively. Per-residue decomposition analysis was employed to calculate the affinity and to assess the stability of the studied compounds through the estimation of the different energy contributions of important residues at the active site.

2.8.Pharmacokinetic and Toxicity Predictions

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

3. Results and discussion

3.1. QSAR modeling

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

pIC50 = -17.0891 + 16.1547*(**GATS8i**) + 0.6723*(**NRS**) + 36.4881*(**G2p**) – 0.428*(**H8s**) (Equation 1)

N_{tr} = 14, R^2_{tr} = 0.91, $RMSE_{tr}$ = 0.19, Q^2_{LOO}= 0.78, Q^2_{lmo} = 0.75 R^2_{test}= 0.85, MAE_{test} = 0.26 , Q^2_{F1} = 0.82 , Q^2_{F2} = 0.81, Q^2_{F3} = 0.80 , CCC_{test} = 0.92, S = 0.24, and F = 26.75.

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

All these statistical metrics that have been calculated for the established model are associated with fitting (i.e. R^2_{tr} and $RMSE_{tr}$), internal (i.e. Q^2_{loo} and Q^2_{lmo}) and external validation (i.e. R^2_{test}, MAE_{test}, Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, and CCC_{test}) indicate the statistical reliability of the QSAR model. Indeed, these obtained parameters fulfill the recommended thresholds values and ensure the predictive power and stability of the derived GA-MLR model (Figure 3a). In addition, for a better validation of the built model, the AD assured by the leverage method and plotted as the Williams plot (Figure 3b), is used to assess the space of the AD of the created model. The

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

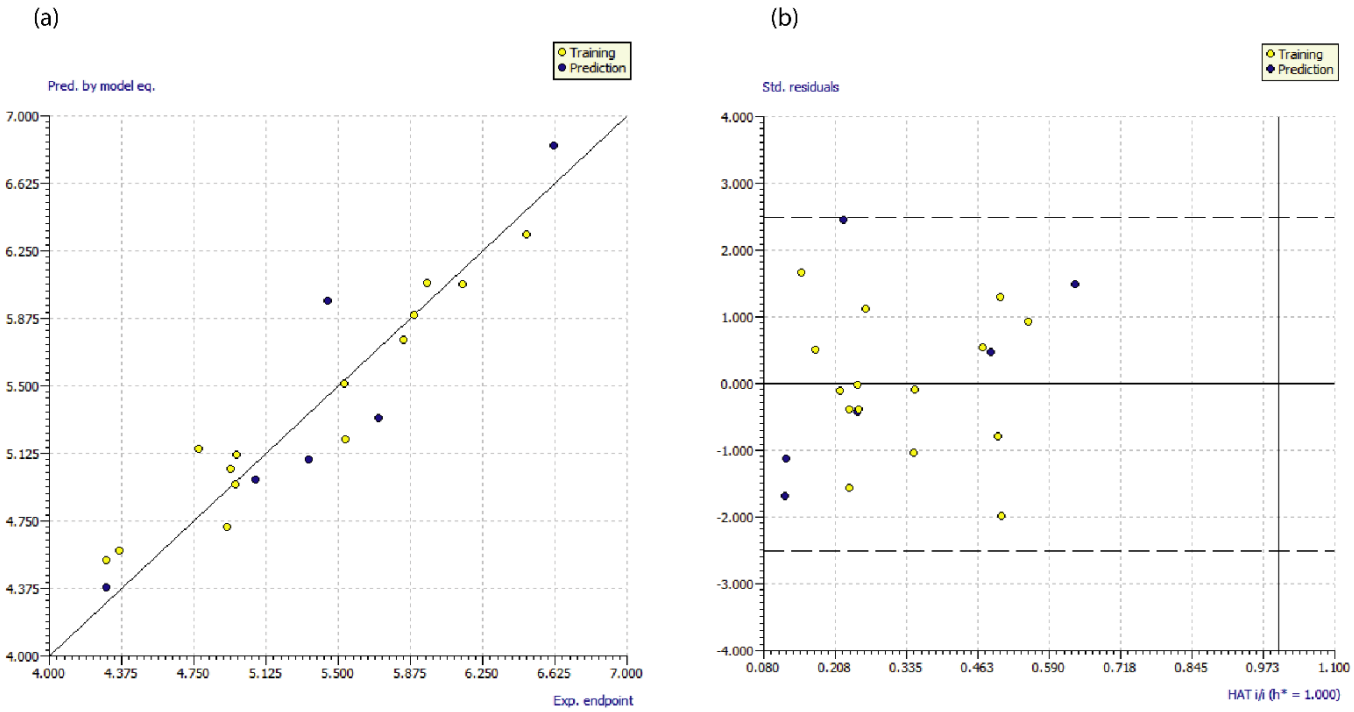
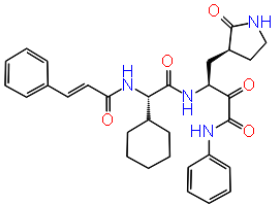
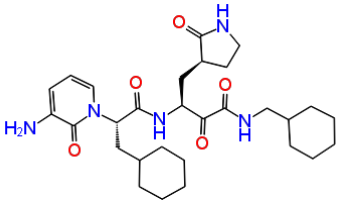
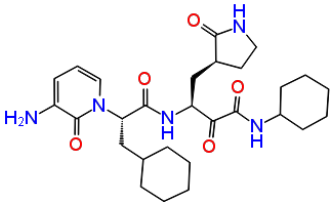
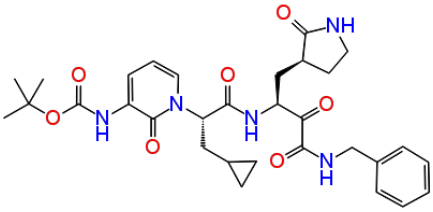


Figure 3. **a** | Plot of experimental vs. predicted pIC50 values; **b** | Williams plot.

3.2. Ligand-based Pharmacophore Modeling

Ligand-based pharmacophore models were built from 21 α-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_{50} using the GA-MLR model equation (Eq.1), their pharmacophore fit scores along with their binding affinity scores with the M^{pro} active site.

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

Hit ID	Chemical structure	pIC_{50} (Predicted)	Pharmacophore Fit score	Binding affinity (Kcal.mol^{-1})
007		6.0712	83.11	-7.9
329		6.2861	84.35	-7.2
331		6.4055	84.35	-8.2
13b		-	-	-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^{pro} 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^{-1}. It was found that the binding affinity score of the hit **329** (-7.2 kcal.mol^{-1}) was higher in comparison to the reference compound **13b** (-7.7 kcal.mol^{-1}). Unlike the hit compound **329**, the other two hits (**007** and **331**) showed lower binding affinity score (-7.9 and -8.2 kcal.mol^{-1}) towards the M^{pro} active site than that of the reference ligand (**13b**). The interaction of the amino acid residues of the M^{pro} 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

of the M^{pro}. The compound **13b** was stabilized by five hydrogen bonds and three hydrophobic bonds while interacting with the active site of the receptor protein, in which it forms two hydrogen bonds with the catalytic residue CYS145 and one hydrophobic interaction with the HIS41. The interaction with the compound **007** was found to be stabilized by forming two hydrogen bonds, one pi-sulfur interaction, and four hydrophobic interactions where one of them was seen with the catalytic residue HIS41. The compound **329** formed six hydrogen bonding interactions in which one hydrogen bonding interaction was seen with the catalytic residue CYS145, along with three hydrophobic interactions where two of them interact with HIS41 and CYS145 catalytic dyad. Besides, the compound **331** gets stabilized by six hydrogen binding interactions where two of them interact with the catalytic residues CYS145 and HIS41, and two out of three hydrophobic interactions were seen to be formed with the catalytic residues CYS145 and HIS41. Altogether, the analysis of non-covalent interactions between the selected three hit compounds and the M^{pro} shows that the selected compounds interact either with the two key catalytic residues, or with at least one of them (i.e. CYS145 or HIS41), and can thus serve as key protease inhibitors (Table S3).

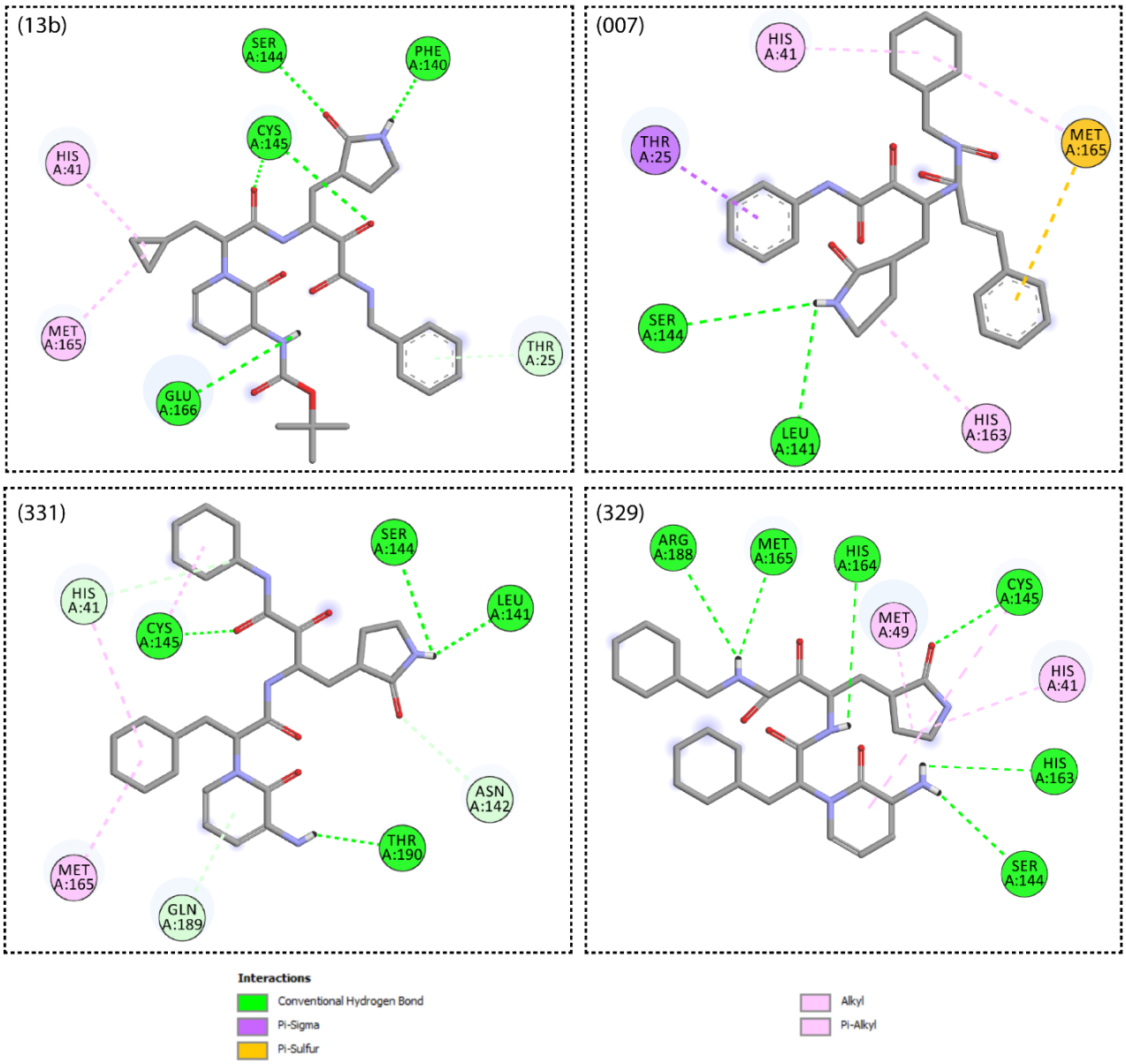


Figure 4. Interactions with key residues as exhibited by the hit compounds (**007**, **329** and **331**) and the reference ligand **13b** with the M^{pro} active site.

3.4. Molecular dynamics simulations

3.4.1. Conformational stability of the M^{pro}

The structural 3D stability of proteins could be estimated by measuring the overall atom deviations at the protein backbone. The Root-Mean-Square Deviation (RMSD) of the C-α atoms is a good parameter to estimate these variations. Great RMSD values can be an indicator of the expanded atom deviations at the protein backbone, also reduced RMSD values depict moderate

variations in backbone atoms. This can support the mapping of the protein dynamics and the equilibrium of the different structures. In this regard, throughout the MD simulations, the deviations of the amino-acids were investigated via RMSD calculations. The results were plotted as a function of the simulations time in Figure 5. As can be seen in the plots, all the complexes tend to converge around 40 ns and continue up to the end of the simulations showing deviations that differed across the simulations time. The binding of the compounds **007**, **329** and **331** to the M^{pro} was characterized by a decrease in deviation among the backbone atoms relating to the unbound conformation that displayed higher variations. The overall average values of RMSD of complex**007**, complex**329**, complex**331** and M^{pro} APO form are 1.88 Å, 1.95 Å, 1.94 Å and 2.08 Å, with maximum recorded values of 2.91 Å, 3.14 Å, 2.94 Å and 3.64 Å, respectively. In opposition to the impact of reduced RMSD upon binding of the previous compounds, the reference compound **13b** exhibited greater RMSD values with overall average value of 2.42 Å which is slightly higher than that disclosed by other systems. Although molecular binding to M^{pro} in general stabilizes the complexes with values that fall behind 3.5 Å and overall RMSD average values lower than 2.5 Å, a diverse impact on the stability of M^{pro} upon the binding of the compounds **007**, **329** and **331** was observed as compared to the binding of **13b** which could be attributed to the stronger interaction of the predicted compounds in the active site of M^{pro}.

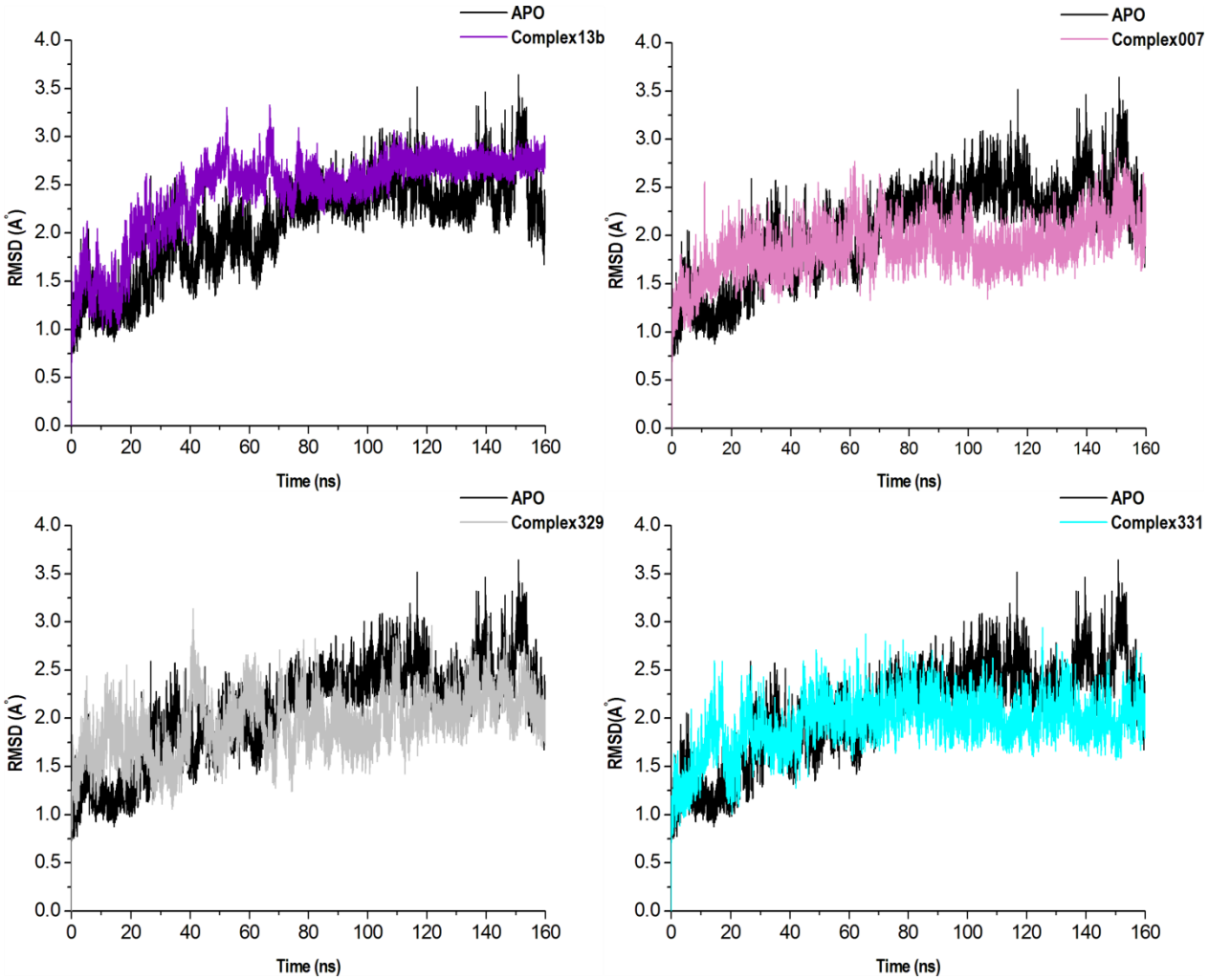


Figure 5. C-α backbone RMSD graphs of the three hits as well as the reference compound **13b** in complex with M^{pro} when compared to the unbound M^{pro}.

3.4.2. Structural fluctuations of the M^{pro}

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

residues, particularly in the catalytic region 140-192. Nevertheless, subtle differences were mainly observed outside this region. In the light of these observations, assessment of the secondary structure was performed.

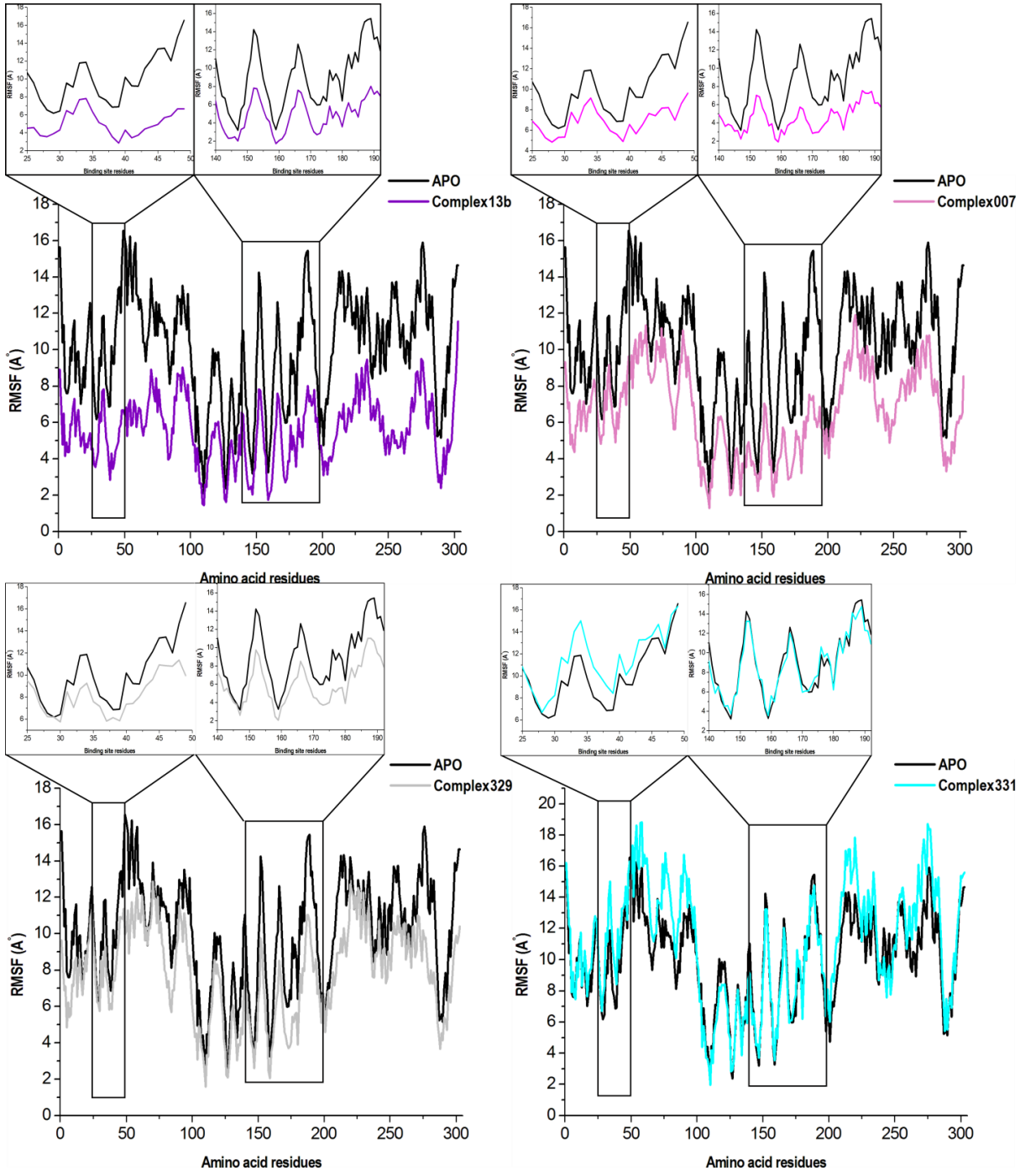


Figure 6. The RMSF plots of the three hits as well as the reference compound **13b** in complex with the M^{pro} in comparison with the unbound system.

3.4.3. Secondary structure analysis

As seen in Table 2, secondary structure analysis predicts a good amount of β strand (around 37% in all systems). The predicted percentage value of α helix was around 28% in all complexes except for the complex**331** which showed slightly higher percentage of α 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 β strand and α 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 %				
	α helix	**β strand**	**3_{10} helix**	**Beta Turn**	**Bend**
M^{pro} APO form	28.36	37.17	4.38	19.38	10.72
Complex**13b**	28.32	37.57	3.85	20.13	10.12
Complex**007**	28.64	37.37	4.78	18.64	10.57
Complex**329**	28.36	37.47	4.41	19.21	10.55
Complex**331**	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^{-1}) and complex**331** (-47.16 kcal.mol^{-1}) exhibited the most favorable binding when compared with complex**13b** and complex**007** (-35.01 and -31.51 kcal.mol^{-1}, 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^{pro}. The total binding free energies are highlighted in bold to distinguish them from the elements of the overall values.

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

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.

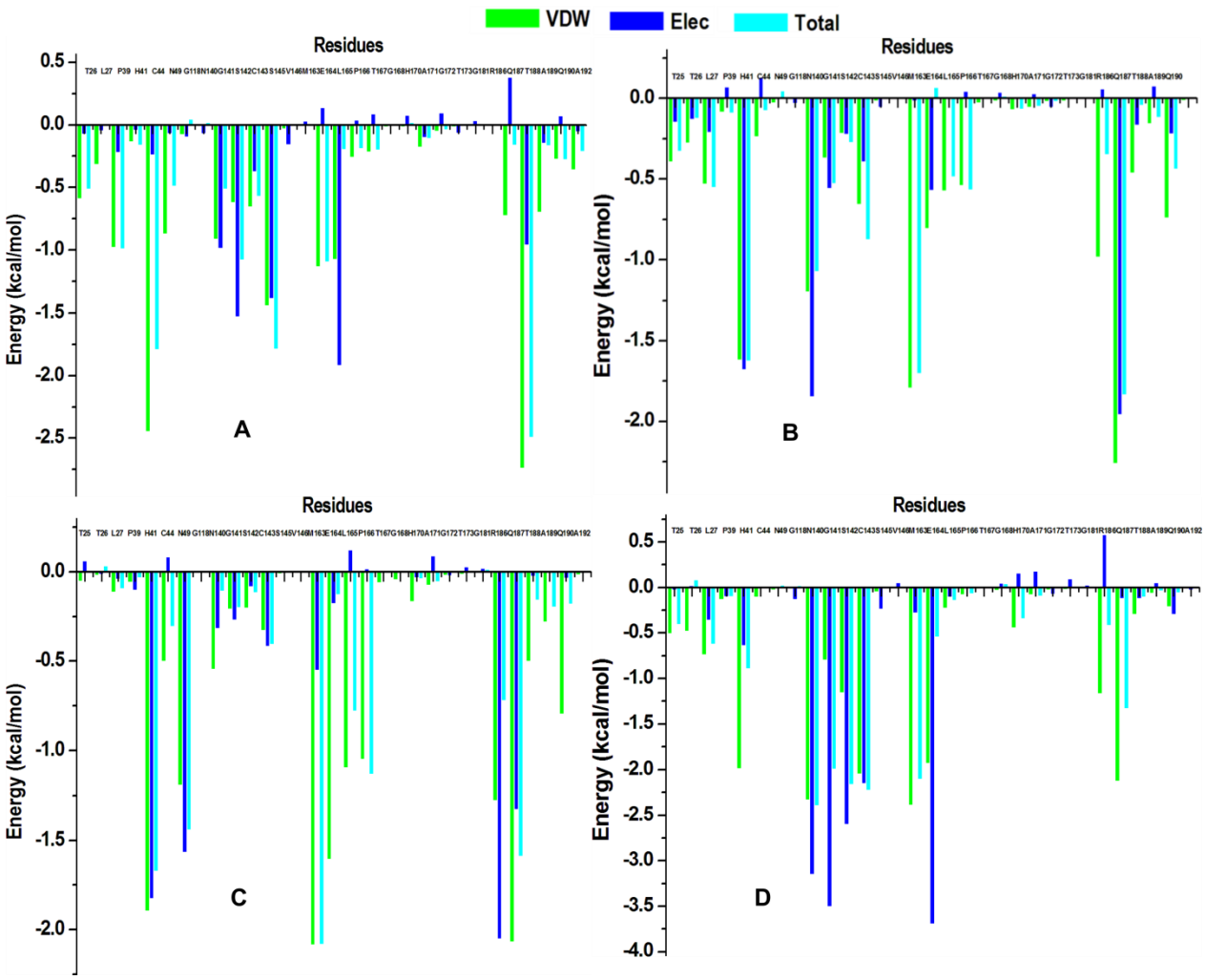


Figure 7. Per-residue decomposition analyses demonstrating the role of individual energy contributions of catalytic site residues to the stability and binding of the molecules **13b** (A), **007** (B), **329** (C) and **331** (D).

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

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

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

Toxicological properties	**007**	**329**	**331**	**13b**
Mutagenic	N	N	N	N
Tumorigenic	N	N	N	N
Irritant	N	N	N	N
Reproductive effect	N	N	N	N

N= No risk

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

Pharmacokinetic properties	**007**	**329**	**331**	**13b**
Molecular weight ($g.mol^{-1}$)	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 ($Å^2$)	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

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

4. Conclusion

The main protease (M^{pro}) has been an appealing target for discovering new SARS-CoV-2 replication inhibitors, as it is an important protein in post-transitional processing of replicase polyproteins. This study attempts to identify novel potent α-ketoamide based inhibitors targeting the M^{pro} of SARS-CoV-2. To accomplish this goal, QSAR and pharmacophore models were developed from the structural features of a series of α-ketoamide derivatives that had shown an inhibition effect against the M^{pro} of SARS-CoV. Three hits were retrieved by utilizing a constructed GA-MLR model and by applying a pharmacophore fit model. Moreover, Molecular dynamic simulations were used as a computational validation of the compounds. Overall, the three evaluated hits displayed important conformational and structural stability as well as

promising binding profiles. In addition, it has been shown that the suggested hit molecules are non-toxic and have acceptable pharmacological properties. The outcomes of this study clearly show that the three screened compounds may lead to potent anti-SARS-CoV-2 M^{pro} drug molecules where the binding details and the nature of activity were considered to be sufficient for blocking the active site of the M^{pro}. The ensemble of computational methods implemented herein allowed for the discovery of new drug candidates with potential inhibitory activity against the SARS-CoV-2 M^{pro}. These hit compounds can be further biologically evaluated for their potency and physiological toxicity.

Acknowledgment

The authors would like to thank Prof. Paola Gramatica for providing a free copy of the QSARINS Software.

Disclosure statement

All the authors declare no conflict of interest.

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.

[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, α-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.

[9] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G.

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

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

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

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

[13] A. Palmeira, E. Sousa, A. Köseler, R. Sabirli, T. Gören, İ. Türkçüer, Ö. Kurt, M.M. Pinto, M.H. Vasconcelos, Preliminary virtual screening studies to identify grp78 inhibitors which may interfere with sars-cov-2 infection, Pharmaceuticals. 13 (2020) 1–13. https://doi.org/10.3390/ph13060132.

[14] A. Jiménez-Alberto, R.M. Ribas-Aparicio, G. Aparicio-Ozores, J.A. Castelán-Vega, Virtual screening of approved drugs as potential SARS-CoV-2 main protease inhibitors,

Comput. Biol. Chem. 88 (2020) 107325. https://doi.org/10.1016/j.compbiolchem.2020.107325.

[15] S. Choudhary, Y.S. Malik, S. Tomar, Identification of SARS-CoV-2 Cell Entry Inhibitors by Drug Repurposing Using in silico Structure-Based Virtual Screening Approach, Front. Immunol. 11 (2020) 1664. https://doi.org/10.3389/fimmu.2020.01664.

[16] 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, Science (80-.). 3405 (2020) eabb3405. https://doi.org/10.1126/science.abb3405.

[17] J.J. Irwin, B.K. Shoichet, ZINC - A free database of commercially available compounds for virtual screening, J. Chem. Inf. Model. 45 (2005) 177–182. https://doi.org/10.1021/ci049714+.

[18] D. Zeng, Y. Ma, R. Zhang, Q. Nie, Z. Cui, Y. Wang, L. Shang, Z. Yin, Synthesis and structure–activity relationship of α-keto amides as enterovirus 71 3C protease inhibitors, Bioorganic Med. Chem. Lett. 26 (2016) 1762–1766. https://doi.org/10.1016/j.bmcl.2016.02.039.

[19] A. Vuorinen, R. Engeli, A. Meyer, F. Bachmann, U.J. Griesser, D. Schuster, A. Odermatt, Ligand-based pharmacophore modeling and virtual screening for the discovery of novel 17β-hydroxysteroid dehydrogenase 2 inhibitors, J. Med. Chem. 57 (2014) 5995–6007. https://doi.org/10.1021/jm5004914.

[20] W.R. Ferraz, R.A. Gomes, A.L. S Novaes, G.H. Goulart Trossini, Ligand and structure-based virtual screening applied to the SARS-CoV-2 main protease: an in silico repurposing study, Future Med. Chem. (2020) fmc-2020-0165. https://doi.org/10.4155/fmc-2020-0165.

[21] LigandScout, Version 4.3; Inte:Ligand GmbH, Clemesn-Maria-Hofbaurer-G. 6, 2344, Maria Enzersdorf, Austria. HYPERLINK, http://www.inteligand.com, (n.d.). http://www.inteligand.com.

[22] P.T. Mpiana, K. te N. Ngbolua, D.S.T. Tshibangu, J.T. Kilembe, B.Z. Gbolo, D.T. Mwanangombo, C.L. Inkoto, E.M. Lengbiye, C.M. Mbadiko, A. Matondo, G.N. Bongo,

D.D. Tshilanda, Identification of potential inhibitors of SARS-CoV-2 main protease from Aloe vera compounds: A molecular docking study, Chem. Phys. Lett. 754 (2020) 137751. https://doi.org/10.1016/j.cplett.2020.137751.

[23] R.K. Hussein, H.M. Elkhair, Molecular docking identification for the efficacy of some zinc complexes with chloroquine and hydroxychloroquine against main protease of COVID-19, J. Mol. Struct. 1231 (2021) 129979. https://doi.org/10.1016/j.molstruc.2021.129979.

[24] RCSB PDB - 6Y2F: Crystal structure (monoclinic form) of the complex resulting from the reaction between SARS-CoV-2 (2019-nCoV) main protease and tert-butyl (1-((S)-1-(((S)-4-(benzylamino)-3,4-dioxo-1-((S)-2-oxopyrrolidin-3-yl)butan-2-yl)amino)-3-cycloprop, (n.d.). https://www.rcsb.org/structure/6Y2F (accessed September 15, 2020).

[25] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791. https://doi.org/10.1002/jcc.21256.

[26] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2009) NA-NA. https://doi.org/10.1002/jcc.21334.

[27] T. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, AMBER Software, (n.d.) AMBER 2018, University of California, San Francisc.

[28] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Antechamber, An Accessory Software Package For Molecular Mechanical Calculations, J. Am. Chem. Soc. 222 (2001) U403.

[29] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins Struct. Funct. Bioinforma. 65 (2006) 712–725. https://doi.org/10.1002/prot.21123.

[30] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general Amber force field, J. Comput. Chem. 25 (2004) 1157–1174.

https://doi.org/10.1002/jcc.20035.

[31] M.F. Harrach, B. Drossel, Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity, J. Chem. Phys. 140 (2014) 174501. https://doi.org/10.1063/1.4872239.

[32] M.J. Harvey, G. Giupponi, G. De Fabritiis, ACEMD: Accelerating biomolecular dynamics in the microsecond time scale, J. Chem. Theory Comput. 5 (2009) 1632–1639. https://doi.org/10.1021/ct9000685.

[33] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

[34] Y. Lin, D. Pan, J. Li, L. Zhang, X. Shao, Application of Berendsen barostat in dissipative particle dynamics for nonequilibrium dynamic simulation, J. Chem. Phys. 146 (2017) 124108. https://doi.org/10.1063/1.4978807.

[35] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theory Comput. 9 (2013) 3084–3095. https://doi.org/10.1021/ct400341p.

[36] E. Seifert, OriginPro 9.1: Scientific Data Analysis and Graphing Software—Software Review, J. Chem. Inf. Model. 54 (2014) 1552–1552. https://doi.org/10.1021/ci500161d.

[37] K. Huang, S. Luo, Y. Cong, S. Zhong, J.Z.H. Zhang, L. Duan, An accurate free energy estimator: Based on MM/PBSA combined with interaction entropy for protein-ligand binding affinity, Nanoscale. 12 (2020) 10737–10750. https://doi.org/10.1039/c9nr10638c.

[38] 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.

[39] N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient, J. Chem. Inf. Model. 51 (2011) 2320–2335. https://doi.org/10.1021/ci200211n.

[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).