Modern Chemistry
2022; 10(2): 30-47
http://www.sciencepublishinggroup.com/j/mc
doi: 10.11648/j.mc.20221002.12
ISSN: 2329-1818 (Print); ISSN: 2329-180X (Online)

Design of New Inhibitors to Fight Against the 3CLpro Protease of SARS-CoV-2 (COVID-19)

Georges Stéphane Dembélé[1, 2], Mamadou Guy-Richard Koné[1, 2, *], Doh Soro[1], Fandia Konaté[1], Bibata Konaté[1, 2], Nahossé Ziao[1, 2]

[1]Laboratory of Thermodynamics and Physico-Chemistry of the Environment, Nangui Abrogoua University, Abidjan, Ivory Coast

[2]Ivorian Group for Research in Disease Modeling, Nangui Abrogoua University, Abidjan, Ivory Coast

Email address:

guyrichardkone@gmail.com (M. Guy-Richard K.)

[*]Corresponding author

To cite this article:

Georges Stéphane Dembélé, Mamadou Guy-Richard Koné, Doh Soro, Fandia Konaté, Bibata Konaté, Nahossé Ziao. Design of New Inhibitors to Fight Against the 3CLpro Protease of SARS-CoV-2 (COVID-19). *Modern Chemistry*. Vol. 10, No. 2, 2022, pp. 30-47.
doi: 10.11648/j.mc.20221002.12

Received: April 6, 2022; **Accepted**: April 23, 2022; **Published**: April 28, 2022

Abstract: The objective of this work is to design new compounds more active against SARS-CoV-2. This design study of new inhibitors on the main protease source of coronavirus (3CLpro) was conducted on ten molecules using Molecular Modeling techniques (Docking, QSAR, ADMET). Molecular docking between M5, M8 and M1 showing best, medium and low scores respectively. The active site residues revealed that the M5 ligand establishes more hydrogen bonds on all the ligands studied thus forming the most stable complex. Predicting the pIC_{50} of the molecules in the training set as a function of the variation in binding energy ($\Delta\Delta G$) to the pathogen, allowed us to develop a QSAR model accredited with very good statistical indicators $R^2 = 0.9137$; $S = 0.058$; $F = 52.942$. The applicability domain of the model obtained from the lever method shows that all compounds belong to the applicability domain. Moreover, the reliability of this model allowed the design of twenty (20) new potential molecules with theoretical inhibitory concentration potentials (pIC_{50}^{th}) values superior to those of the molecules in the database. Finally, the pharmacokinetic profile of the proposed molecules was confirmed by the satisfaction of the Lipinski and ADMET.

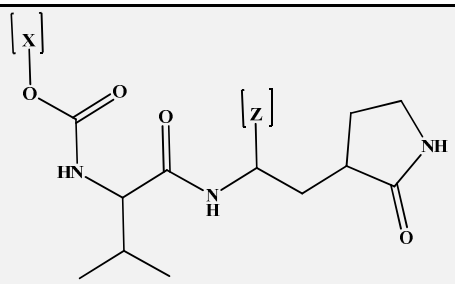
Keywords: SARS-CoV-2, QSAR, Docking, ADMET

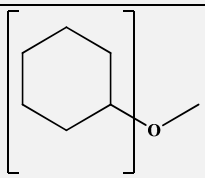
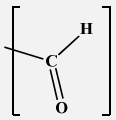
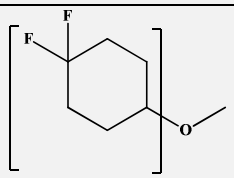
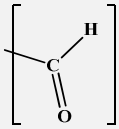
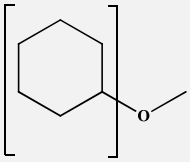
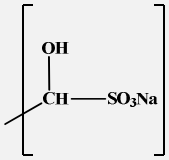
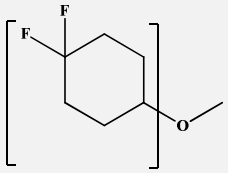
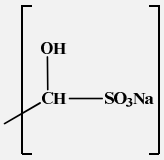
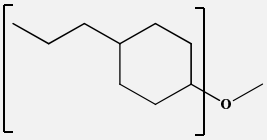
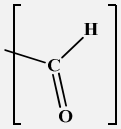
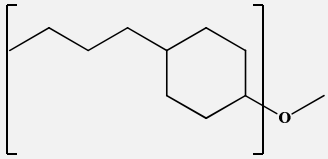
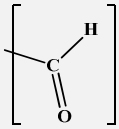
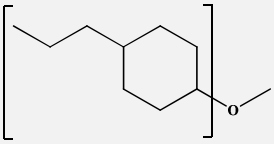
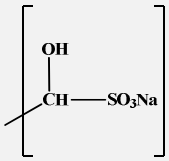
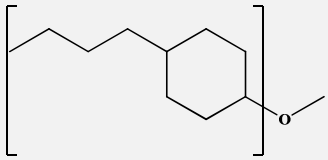
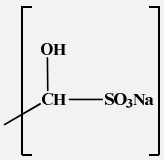
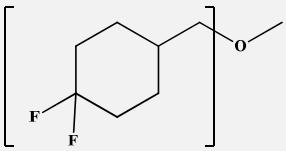
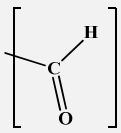
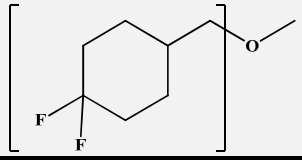
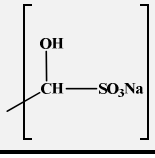
1. Introduction

The new coronavirus is an infectious disease caused by the coronavirus SARS-CoV-2 (severe acute respiratory syndrome-coronavirus-2). This disease can cause fever, coughing, breathing difficulties, headaches, in some cases diarrhea and vomiting in the infected person [1, 2]. In severe cases, the infected person dies. The first case of COVID-19 appeared in China precisely in Wuhan on November 17, 2019, and then spread around the world [3]. This virus has infected nearly 174,000,000 people causing over 3,000,000 deaths worldwide [4], This has resulted in school closures and a significant decline in economic activity. The year 2021 marked the launch of vaccination campaigns against COVID-19 throughout the world. However, there are already more than ten vaccines administered (Johnson & Johnson, Pfizer, Moderna, Astra Zeneca,...). Despite the existence of these vaccines, there is still an increase in infections and deaths due to virus mutations. It is therefore urgent to find molecules capable of blocking the mutations of this virus. Coronaviruses are frequent RNA viruses, of the Coronaviridae family, which are responsible for digestive and respiratory infections in humans and animals. They have two proteases [5]. The crystal structure of the major protease of SARS-CoV-2 is 3CLpro (3C-like proteinase) [6]. Its protein sequence is 96% identical to that of the 3CLpro proteases of other coronaviruses, which have been extensively studied [7, 8]. The 3CLpro protease is a potential viral drug target [9, 10] for coronavirus infections because of its essential role in processing polyproteins that are translated from viral RNA. Thus, Rathnayake et *al.* [11] in their work have shown that

a series of molecules (Table 1) show activity (IC_{50}) against the 3CL protease. The general objective of this work is to propose potential molecules that can inhibit the main protease of coronavirus (3CLpro).

Table 1. *2D representation and Inhibitory Concentration (IC_{50}) of 3CL pro inhibitors.*



CODES	[X]	[Z]	IC_{50} (µM)	CODES	[X]	[Z]	IC_{50} (µM)
M1			0.82	M5			0.17
M2			0.65	M6			0.20
M3			0.28	M7			0.43
M4			0.23	M8			0.41
M9			0.48	M10			0.45

2. Materials and Methods

2.1. Level of Computational Theory

In order to predict the antiCOVID-19 activity of the 3CLpro protease inhibitors quantum chemistry calculations were performed using Gaussian 09 [12]. The twenty (10) molecules used in this study have Inhibitory Concentration (IC) ranging from 0.17 to 0.82 µM. The median inhibitory concentration (IC50) is a measure of the effectiveness of a given compound in inhibiting a specific biological or biochemical function. Biological data are usually expressed as the opposite of the decimal-based logarithm of activity ($-\log_{10}(C)$) to obtain better mathematical values when structures are biologically active [13, 14]. The antiCOVID-19 activity will be expressed by the potential inhibitory concentration pIC defined by equation (1):

$$pIC_{50} = -\log_{10}(IC_{50} * 10^{-6}) \quad (1)$$

Where IC_{50}, the Inhibitory Concentration in µM.

The modeling was developed using the linear multiple regression (LMR) method that is implemented in Excel spreadsheets [15].

2.2. Selection and Preparation of Proteins

The therapeutic target studied in this work is the SARS-CoV-2 3CL protease. Its 3D structure was obtained in pdb format via the PDB database on the RCSB Protein Data Bank website "www.pdb.org". This one has a resolution of 1.70 Å,

has 2 monomers A and B and 4862 atoms. Each monomer contains a ligand (M10).

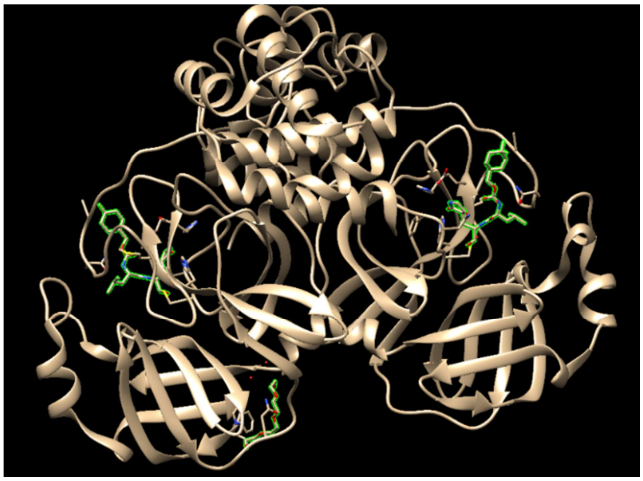


Figure 1. *Crystallographic structure of the 3CL protein of SARS-CoV-2 (PDB: 6xmk).*

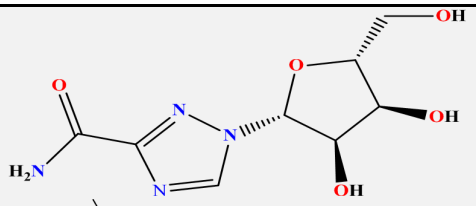
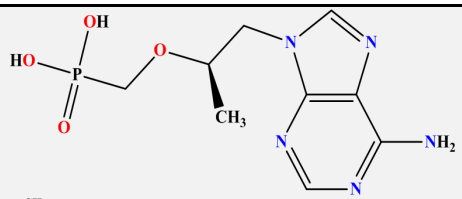
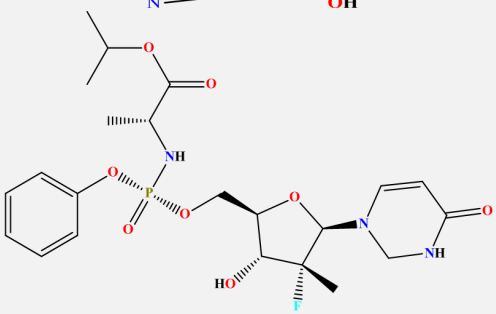
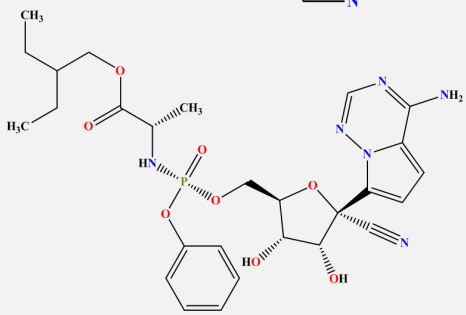
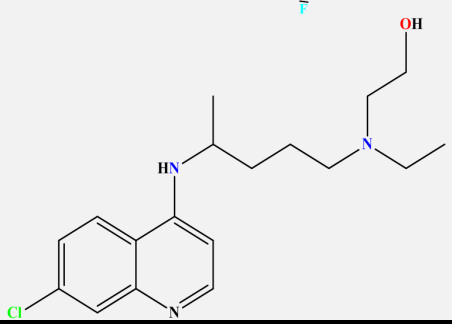
The preparation of the protein was performed using the DockPrep module included in the CHIMERA software [16]. The water molecules were removed, then the non-standard elements of the residues were repaired and finally some hydrogen molecules and charges were added. The Amber force field was used to assign the protein charges. The obtained protein was recorded in mol^2 format. Figure 2 shows the prepared protein.

Figure 2. *Crystallographic structure of 3CL pro after preparation and removal of endogenous ligands.*

2.3. Preparation of Ligands

In this work, we used a series composed of ten (10) molecules of 3CLpro protease inhibitors synthesized by Rathnayake et *al.* [11] as well as other compounds used in the treatment of COVID-19. These molecules include hydroxychloroquine, remdesivir, tenofovir, ribavirine, artemisinine [17-19]. These different molecules are represented in 2D in Table 2 below. All ligands were drawn in 3D using Gaussview 6.0 [20], then they were optimized from the CHIMERA software [16] in order to obtain the most stable conformation of the ligands. The output file was finally converted to Protein Data Bank (PDB) format.

Table 2. *Molecular structures of compounds used in the treatment of COVID-19.*

Ribavirin		Tenofovir r	
Sofosbuvir		Remdesivir	
Hydroxychloro quine			

2.4. Molecular Docking

Rathnayake *et al* [11] characterized a crystal structure of the main protease (3CLpro) of COVID-19 in complex with an inhibitor. For the molecular docking of the 6XMK protein, the cavities in which the inhibitor was located were defined as active sites.

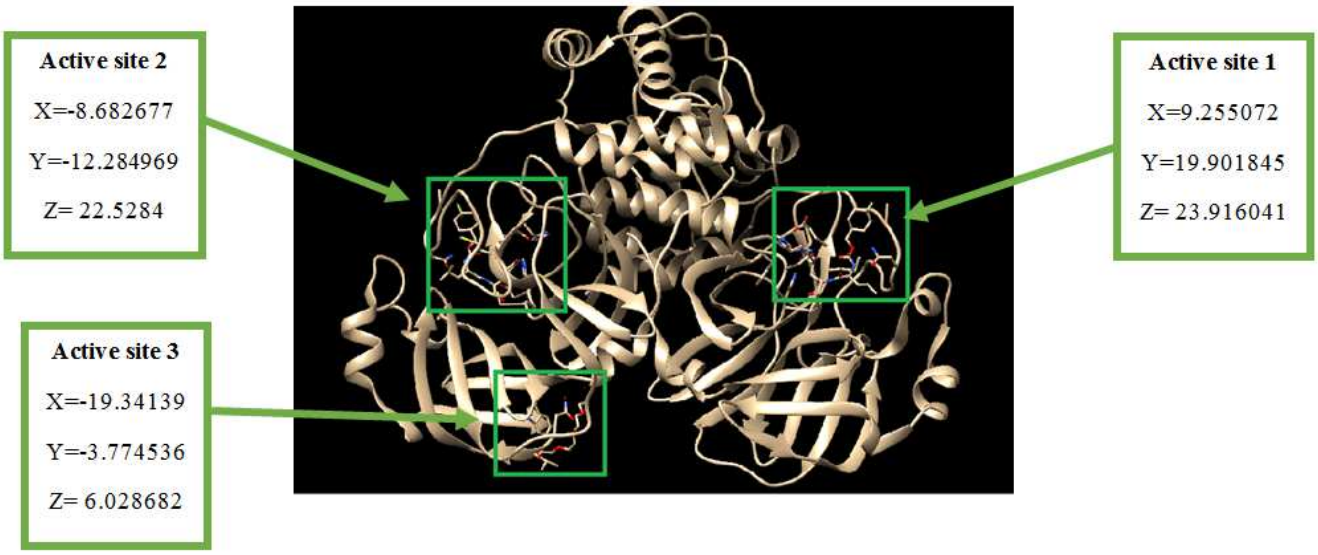


Figure 3. *The coordinates of the different active sites of the 6XMK protein.*

A virtual screening was performed using Autodock Vina and the best ligand/protein mode was identified based on the binding energy in the 3 active sites. Then, we selected the active site that best matches the experimental results.

2.5. Statistical Analysis

The statistical technique of simple linear regression (SLR) is used to study the relationship between a dependent variable (Property) and an independent variable (Descriptor). We note Y the real random variable to be explained (endogenous, dependent or response variable) and X the explanatory variable or fixed effect (exogenous). The model is based on the assumption that, on average, f(x) is an affine function of x. The writing of the model implicitly assumes a prior notion of causality in the sense that Y depends on X because the model is not symmetric.

$$Y = aX + b \quad (2)$$

2.6. Drug-Likeness

Druglikeness is a qualitative concept used in drug design to determine the efficacy of a drug candidate. It is estimated from the molecular structure even before the substance is synthesized and tested.

2.6.1. The Lipinski Rule (Rule of Five)

According to the following empirical principles, stated by Christopher Lipinski and grouped under the name of "rule of five", this rule is the most used for the identification of drug-like compounds, a substance will be better absorbed or penetrated if [21, 22]:

1) Its molecular weight is less than or equal to 500 Da.
2) It has less or 5 hydrogen bond donors HBD.
3) It has less or 10 hydrogen bond acceptors HBA.
4) Its log P value is less than or equal to 5.

The new molecular structures have been analyzed with the SWISSADME server [23] (http://www.swissadme.ch/) in order to verify whether or not the compounds respect the Lipinski rule (Rule of Five).

2.6.2. Prediction of Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET)

1) *Human Intestinal Absorption (HIA)* refers to the ability of the human intestine to absorb the drug. The higher the percentage of human intestinal absorption, the better the human intestine absorbs the drug (from 0~20% poor absorption; from 20~70% medium absorption, from 70 ~100% high absorption).
2) *Caco-2(nm/s) and MDCK (nm/s)* predicts the intestinal permeability of a compound on Caco-2(<4 poor permeability, between 4 ~70 medium permeability, >70 high permeability) and MDCK cells.
3) *PPB (Plasma Protein Binding)* refers to the degree to which drugs bind to proteins in the blood. The effectiveness of a drug can be affected by the degree to which it binds. The less bound a drug is, the more effectively it can cross cell membranes or diffuse. (<90 low binding, >90 high binding).
4) *BBB (Blood-Brain Barrier*) is the descriptor that indicates the ability of a compound to penetrate through the blood-brain barrier (BBB) and controls the passage of most compounds from the blood to the central nervous system (CNS). (<0.1 low absorption in the Central Nervous System (CNS), 0.1~2 medium

absorption in the CNS and >2 high absorption in the CNS).

5) *Cytochromes P450* are key enzymes involved in the metabolism of different endogenous or exogenous molecules. They exist under several iso-forms (CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4) but the most important are the last two. The prediction of the interaction of our best inhibitors with these iso-forms was also essential since the inhibition of these iso-enzymes is certainly one of the main causes of drug interactions leading to toxic or adverse effects [24].
6) *hERG (human Ether-à-go-go-Related Gene)* is a gene encoding a voltage-dependent potassium channel that draws potassium out of the cell. Blockage of this channel leads to fibrillations in cardiology which can result in cardiac arrest.
7) *AMES-Test (Salmonella typhimurium reverse Mutation Assay)* is a simple method to test the mutagenicity of a compound. It uses several strains of Salmonella typhimurium bacteria carrying mutations in genes involved in histidine synthesis, so that they require histidine for growth. This test consists in evaluating the capacity of a compound to induce a mutation allowing a return to growth on a medium without histidine.

These different parameters have been determined from the online server PreADMET [25] (https://preadmet.bmdrc.kr/).

3. Results and Discussion

3.1. Choice of the Active Site

In order to choose the appropriate active site for our work, we performed the docking of the ten (10) molecules in the three active sites of the protein. After docking the ligands and the protein in the three (3) active sites, the software proposed 8 possible bindings of each ligand in the active site. Each pose is associated with a binding energy (score). The best poses (smallest ΔG values), the variation in binding energy (ΔΔG) as well as the experimental activities are listed in Table 1. Note that the change in binding energy is calculated as the difference between the energy value of the molecule and the molecule with the smallest energy. Also, Figure 4 shows the evolution of the potential of the inhibitory concentration and the variation of the binding energy.

Table 3. *Binding energies (ΔG) and its variations (ΔΔG) of the ten (10) ligands in the three active sites.*

MOLECULES	*Active site 1*		*Active site 2*		*Active site 3*		pIC_{50}
	ΔG (kcal/mol)	*ΔΔG (kcal/mol)*	*ΔG (kcal/mol)*	*ΔΔG (kcal/mol)*	*ΔG (kcal/mol)*	*ΔΔG (kcal/mol)*	
M1	-5.9	1	-5.9	0.8	-6.6	0.9	6.086
M2	-5.9	1	-6.7	0	-7.3	0.2	6.187
M3	-6.1	0.8	-6	0.7	-7.5	0	6.553
M4	-6.5	0.4	-6.5	0.2	-6.6	0.9	6.638
M5	-6.9	0	-5.7	1	-7.2	0.3	6.77
M6	-6.8	0.1	-5.8	0.9	-6.6	0.9	6.699
M7	-6.4	0.5	-6.1	0.6	-7.1	0.4	6.367
M8	-6	0.9	-6.6	0.1	-7.2	0.3	6.387
M9	-5.8	1.1	-6.5	0.2	-6.8	0.7	6.319
M10	-6	0.9	-6.5	0.2	-7	0.5	6.347

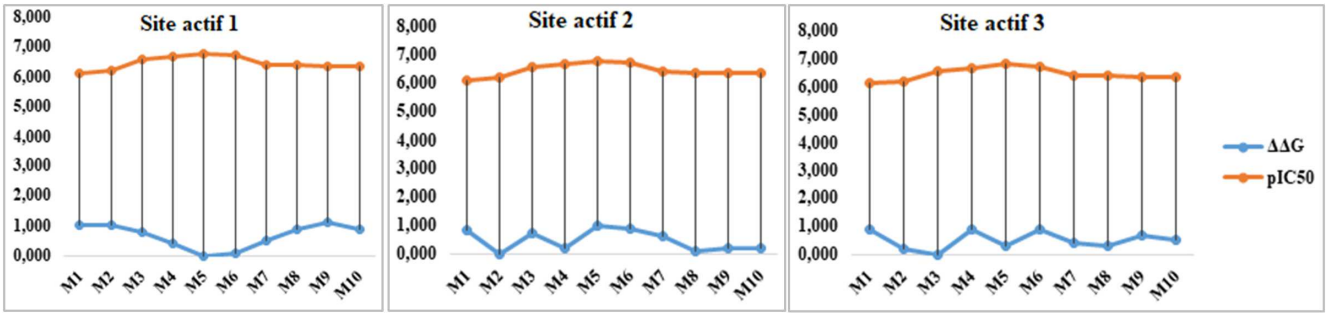


Figure 4. *Evolution of the change in binding energy (ΔΔG) and inhibitory concentration potential (pIC50) in the three active sites.*

Analysis of the table shows that all ligands have negative binding energy in all three active sites. This result reflects that the compounds have a good binding in the different sites. Also, we notice that the results of the binding energy of the ligands in the active site 1 agree more with the values of the experimental activities. This remark is supported by Figure 1, where a symmetrical evolution between the variation of the binding energy and the potential of the inhibitory concentration (pIC_{50}) of the molecules in site 1 is observed. From the above, the active site 1 is considered as the appropriate site for the continuation of our work.

3.2. Study of the Interactions Between Ligands and Proteins

3.2.1. Fixing Energies

In order to evaluate the affinity between the ligands and the protein, we performed a docking in the active site 1 of the protein using the 10 3CLpro inhibitors and some molecules

such as hydroxychloroquine, Remdesivir, Tenofovir, Ribavirin and artemisinin. The binding energies and the inhibitory concentration of these different ligands are presented in Table 4.

Table 4. *Energies of fixation (kcal/mol) of the ten (10) 3CLpro inhibitors studied and some reference molecules.*

MOLECULES	ΔG (kcal/mol)	IC_{50} (µM)
M1	-5.9	0.82
M2	-5.9	0.65
M3	-6.1	0.28
M4	-6.5	0.23
M5	-6.9	0.17
M6	-6.8	0.20
M7	-6.4	0.43
M8	-6	0.41
M9	-5.8	0.48
M10	-6	0.45
Hydroxychloroquine	-5.9	
Remdesivir	-6.1	
Tenofovir	-6.5	
Ribavirine	-6.5	
Artemisinine	-6.4	

Examination of the values in Table 4 shows that all the ten (10) 3CLpro inhibitors studied have low binding energies compared to the reference molecules. This result reflects that the ten (10) molecules inhibit SARS-CoV-2 protease better than the reference molecules. The sequences of evolution following the values of scores (ΔG) and then values of inhibition concentration (IC50) are shown below:

ΔG (kcal/mol): M5 >M6> M4> M7> M3> M8> M10> M1> M2> M10

IC_{50} (µM): M5 >M6> M4> M3> M8 >M7> M10> M9> M2> M1

The study of the visualization of the interactions between the active site residues and the ligands focused on the three ligands that are: M5, M8 and M1. Indeed, these three ligands present respectively the high, average and low scores. The same is true for the inhibitory concentrations IC_{50}.

3.2.2. Visualization of Interactions Between Active Site Residues and Ligands

Using the Discovery Studio Visualizer 2020 software, the interactions of the complexes were obtained. Figures 2, 3 and 4 give an illustration of the interactions of molecules M5, M8 and M1 and proteins. It should be noted that the molecule M5 is the compound with the best score, the molecule M8 is one of the compounds with a medium score and the molecule M1 is the compound with the lowest score. Various observations are made, including between hydrogen bond donors and acceptors, cationic and anionic groups performing ionic interactions, aromatic groups leading to π interactions and hydrophobic groups giving hydrophobic interactions. These interactions are defined as the pharmacophoric points. Pharmacophores are therefore sets of active atoms used in the design of drugs according to a suitable spatial arrangement [26, 27]. Pharmacophore approaches have been widely used in the drug discovery process, especially in the optimization phase [27]. A pharmacophore model can be established either by a "ligand-based" approach by superimposing active molecules and extracting the common physicochemical characteristics essential to their biological activity, or by the "structure-based" approach by searching for the interaction points between the target and the ligands [27].

The visualization of the different geometries of the complexes from the Discovery Studio Visualizer 2020 software allowed us to observe that:

1) The M5 ligand makes 7 hydrogen bonds with the residues HIS A: 38, HIS A: 160, CYS A: 142, SER A: 141, PHE A: 137, GLU A: 163 et GLN A: 186, 2 alkyl bonds with residues ALA A: 188 et PRO A: 165.
2) The M8 ligand makes 5 hydrogen bonds with the residues ASN A: 139, GLU A: 163, GLN A: 186, LEU A: 138 et GLY A: 140, 1 pi-cation bond with the residue HIS A: 38, 1 metal-acceptor bond with the residue HIS A: 161, 1 halogen bond (Fluorine) with the amino acid PRO A: 165.
3) The M1 ligand makes 1 hydrogen bond with the ASN A: 139, 1 pi-alkyl bond with the residue MET A: 162, 1 alkyl bond with the residue HIS A: 160.

The analysis of the different interactions between the residues of the active site of the protein and the ligands shows the different activities between complexes. Concerning the ligand M5 (more active), it has seven (7) hydrogen bonds compared to the compounds M8 and M1 which have respectively five (5) and one (1) hydrogen bond. It should be noted that hydrogen bonding and hydrophobic (alkyl) interactions are parameters that contribute to the stability of complexes. Thus the complex with the most active ligand (M5) is more stable than the other complexes. After the study of the interaction between the protein residues and ligands, the development of the mathematical model between the experimental biological activity and the variation of binding energy

(ΔΔGbinding) was invested. The variation of binding energy is a contribution of amino acid residues to the stabilization of a protein-ligand complex located at the interface of the active site [28].

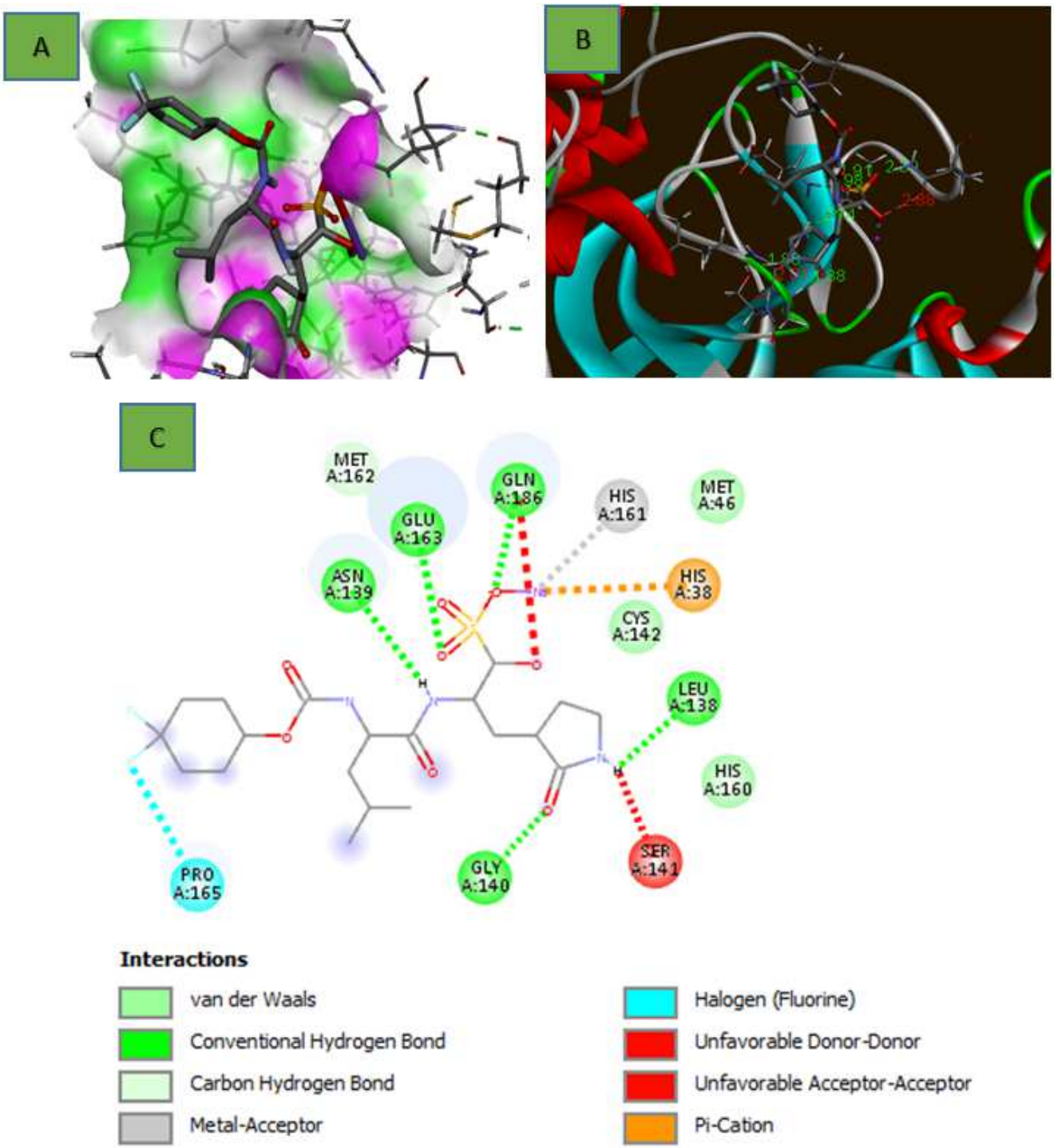


Figure 5. *3D (B) and 2D (C) interactions between the M8 molecule and the active site residues of the protein. (A) Surface of the hydrogen bonds formed between the active pocket and the M8 molecule.*

Table 5. *Variation of experimental binding energies and pIC50 of training and validation sets.*

MOLECULES	ΔG (kcal/mol)	ΔΔG (kcal/mol)	pIC_{50}
Training game			
M3	-6.1	0.8	6.553
M4	-6.5	0.4	6.638
M5	-6.9	0	6.77
M6	-6.8	0.1	6.699

MOLECULES	ΔG (kcal/mol)	ΔΔG (kcal/mol)	pIC_{50}
M8	-6	0.9	6.387
M9	-5.8	1.1	6.319
M10	-6	0.9	6.347
Validation Set			
M1	-5.9	1	6.086
M2	-5.9	1	6.187
M7	-6.4	0.5	6.367

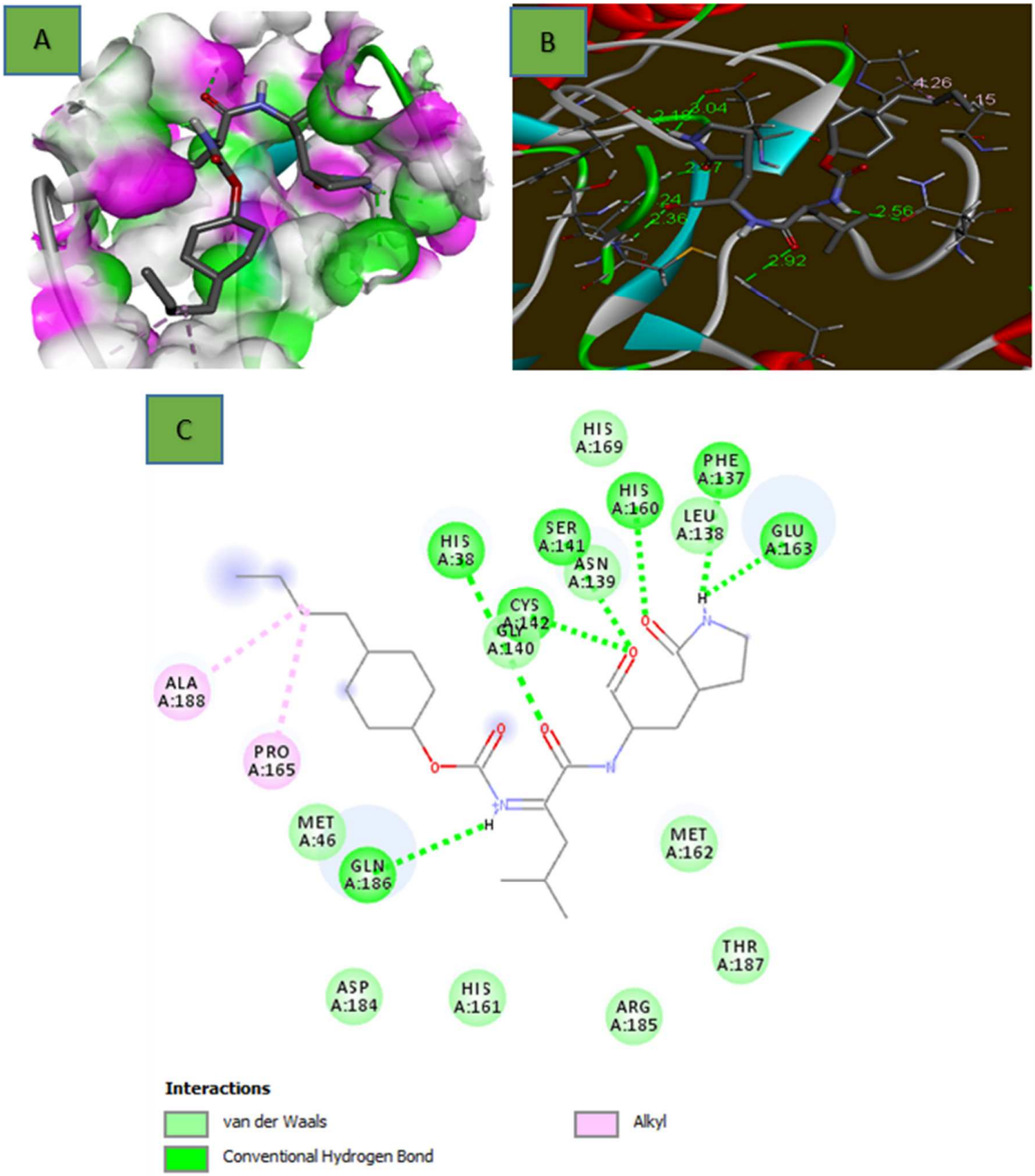


Figure 6. *3D (B) and 2D (C) interactions between the M5 molecule and the active site residues of the protein. (A) Surface of the hydrogen bonds formed between the active pocket and the M5 molecule.*

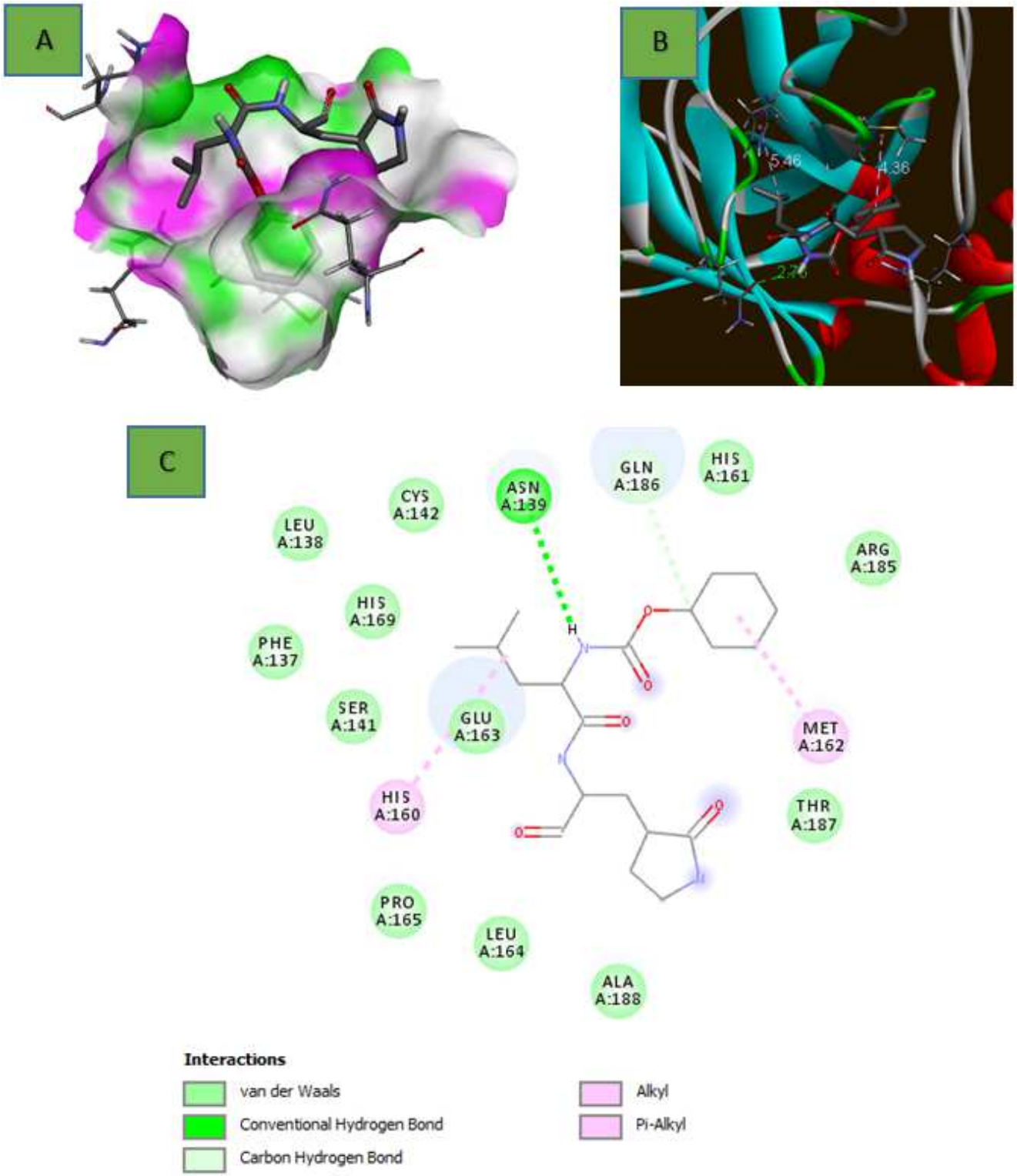


Figure 7. *3D (B) and 2D (C) interactions between the M1 molecule and the active site residues of the protein. (A) Surface of the hydrogen bonds formed between the active pocket and the M1 molecule.*

3.3. Development of the QSAR Model

This QSAR study was conducted using a series of ten (10) 3CLpro inhibitors synthesized and tested on *SARS-CoV 3CLpro* [11]. The molecules were split into two groups: seven (7) were used for the learning set and three (3) others for the validation set. The objective of this part of work is to model the antiCOVID-19 activity of the inhibitors from the variation of binding energy. The values of the binding energy variations ($\Delta\Delta G$) as well as those of the experimental biological activities of the molecules are recorded in Table 5.

In the equation of a model, the negative or positive sign of the coefficient of a descriptor reflects the change in the opposite or the same direction of the values of the explained variable (inhibitory potential, pIC50) with the values of the explanatory variable (descriptor ($\Delta\Delta G$)). Thus, the negative sign indicates that when the descriptor value is high, the inhibitory potential pIC50 decreases, whereas the positive sign reflects the opposite effect. The equation for the QSAR model is presented as follows:

$$pIC_{50} = -0.4006 * \Delta\Delta G_{binding} + 6.7707$$

The correlation curve between the change in binding energies and the experimental pIC_{50} of the 7 inhibitors in the training set is presented by the figure below.

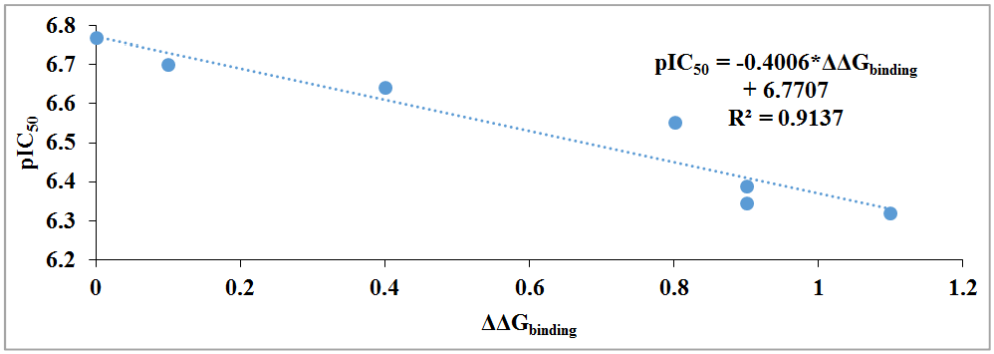


Figure 8. *Correlation curve of pIC_{50} as a function of $\Delta\Delta G_{binding}$.*

The negative sign of the coefficient of variation of the binding energy reflects that the anti COVID-19 activity will be enhanced for low values of this parameter.

The statistical indicators of the QSAR model are given in Table 6.

Table 6. *Statistical analysis report of the inhibitory potential IC_{50} of the QSAR model inhibitors.*

Number of observations N	7
Coefficient of determination R^2	0.9137
Standard deviation σ	0.058
Fischer test F	52.942
Cross-validation correlation coefficient Q^2_{CV}	0.9137
Confidence level α	> 95%

The value of the coefficient of determination R^2 which is 0.9137, shows that the estimated values of pIC50 contain 91.37% of the experimental values. The value of Fisher's test (F = 52.942) is relatively high compared to the critical value, from Fisher's table Fcr = 3.06 [29]. This value 52.942 of the Fisher test, shows that the error committed is less than what the model explains [29]. The standard deviation (σ = 0.058) tends to 0, expressing little variation in the predicted values from the experimental mean. For this model, the correlation coefficient of the cross-validation Q^2_{cv} is equal to $Q^2_{cv} = 0.9137$. This value, higher than 0.9, reflects a so-called excellent model according to Erikson et *al.* [30]. This model is acceptable because it is in agreement with the acceptance criteria of these authors $R^2 - Q^2_{cv}$=0.9137- 0.9137= 0.000 < 0.3. All these statistical indicators clearly show that the developed model explains the antiCOVID-19 activity in a statistically significant and satisfactory manner. The regression line of the model between the experimental and theoretical activities of the training set and the validation set is shown in Figure 8.

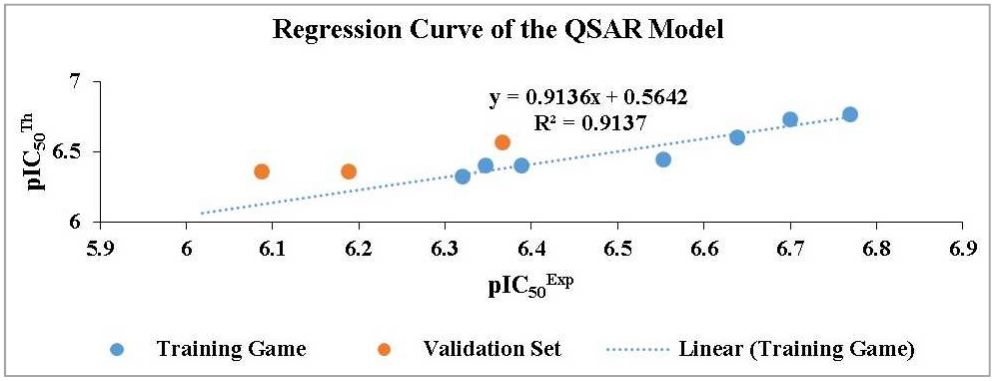


Figure 9. *Regression curve of the experimental and predicted values of the QSAR model.*

The analysis of the regression curve of the experimental and predicted values of the QSAR model shows that the points are around the regression line.

The low value of the standard error which is 0.058 attests

to the good similarity between the predicted and experimental values (Figure 9). This curve shows a similar evolution of the data of this model for the prediction of the inhibitory concentration IC_{50} of the inhibitors despite some deviations recorded (Figure 10).

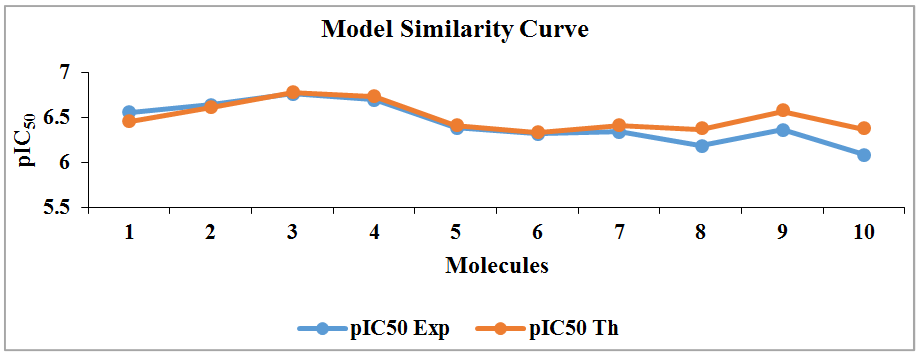


Figure 10. *Evolution curves of the experimental and predicted values of the QSAR model as a function of the variation of the binding energy.*

3.3.1. External Validation of the QSAR Model

The theoretical activities of the inhibitors of the "validation test" are obtained from the following equation:

$$pIC_{50} = -0.4006 * \Delta\Delta G_{binding} + 6.7707$$

For the molecules in the validation set, the theoretical activities will be compared to the experimental activities to judge this prediction. The prediction is good if the ratio $pIC_{50}^{Exp}/pIC_{50}^{Th}$ tends to 1.

Table 7. *Comparison between experimental and theoretical potential values of the validation set.*

MOLECULES	ΔΔG (kcal/mol)	pIC_{50}^{Exp}	pIC_{50}^{Th}	$pIC_{50}^{Exp}/pIC_{50}^{Th}$
M1	1	6.086	6.3701	0.9554
M2	1	6.187	6.3701	0.9713
M7	0.5	6.367	6.5704	0.9690

The different reports $pIC_{50}^{Exp}/pIC_{50}^{Th}$ obtained in Table 7 tend towards 1. The results of these ratios show that the docking model is reliable. The activities of new analogues can therefore be estimated following this model.

3.3.2. Scope of the QSAR Model

The domain of applicability of the model was defined using the standardized residuals technique based on the hii levers. The MINITAB software was used to calculate the levers of the molecules in the training set as well as those of the molecules in the test set. Figure 11 shows the plot of the standardized residues as a function of the hii levers of the compounds in both sets (learning and test). This graph allows to visualize the applicability domain of the model.

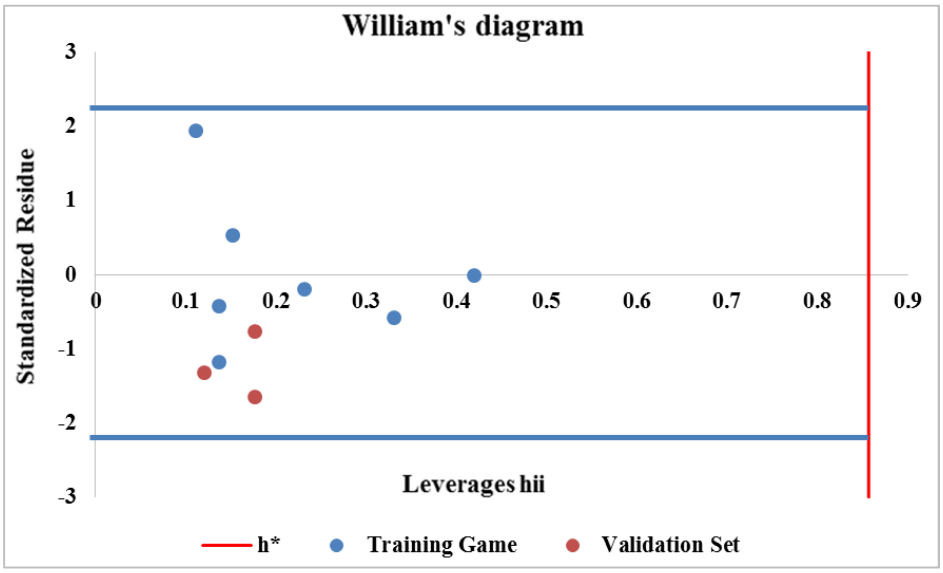


Figure 11. *Applicability of the QSAR model.*

The threshold value of the levers h* is 0.857 for the 7 compounds of the training set and the model descriptor. The extreme values of the standardized residuals are ±3 according to the "three sigma rule". These different values delimit the domain of applicability of the model as shown on the graph above. In this case, we notice that all the values of the molecules of the learning and validation set are lower than the threshold value of the levers h*, which reflects a good chemical space of our mathematical model.

3.4. Design of New SARS-CoV-2 Inhibitors 3CLpro

The compound M5 was taken as a starting structure to search from the swisssimilarity server. Indeed, this database proposed us a collection of 292 similar molecules with a similarity rate up to 95%. In order to improve the score of the starting inhibitor (-6.9 Kcal/mol), a series of molecular docking calculations were performed on the library of similar molecules. The results of this docking are presented in Table 8 and Figure 12.

Table 8. *Docking score in kcal/mol of approved drugs in the ZINC database with SARS-CoV-2 main protease (MPro) (PDB ID: 6XMK).*

Ligands	*ΔG (kcal/mol)*	Ligands	*ΔG (kcal/mol)*	Ligands	*ΔG (kcal/mol)*
ZINC00260582	-7	ZINC12330962	-7.9	ZINC40160151	-7.5
ZINC00260588	-7.4	ZINC12429337	-7.8	ZINC40489201	-6.9
ZINC00260594	-6.4	ZINC12429338	-7.9	ZINC40513622	-7.5
ZINC01092769	-8	ZINC12703103	-6.8	ZINC40513623	-6.5
ZINC01657627	-6.4	ZINC12703110	-7	ZINC40539203	-7.2
ZINC01659837	-7.6	ZINC12880574	-7.9	ZINC40539205	-6.6
ZINC01697993	-6.4	ZINC12942413	-6.6	ZINC40541241	-7
ZINC02172656	-7.1	ZINC12942419	-6.9	ZINC40541244	-6.7
ZINC02172662	-6.6	ZINC13718829	-7.4	ZINC42249104	-6.5
ZINC02172664	-6.7	ZINC14539351	-7.4	ZINC42249110	-6.6
ZINC02172665	-6.8	ZINC14731180	-7.3	ZINC42478542	-6.8
ZINC02172667	-7.2	ZINC14740684	-7.7	ZINC42478544	-6.6
ZINC02450405	-6.7	ZINC14746885	-7.6	ZINC43545870	-7.3
ZINC02572700	-6.4	ZINC14747170	-7.3	ZINC43763716	-7.2
ZINC02833134	-6	ZINC14754915	-7.6	ZINC44866384	-6.7
ZINC03010513	-7	ZINC14957156	-7.9	ZINC46055482	-7
ZINC03010515	-6.7	ZINC15785444	-6.6	ZINC47795046	-6.2
ZINC03010516	-7.1	ZINC15785448	-7	ZINC47795960	-7
ZINC03010519	-6.6	ZINC16478476	-7.2	ZINC47795961	-6.6
ZINC03010521	-6.1	ZINC19414675	-7	ZINC48009255	-7.6
ZINC03010523	-6.5	ZINC20780052	-6.8	ZINC48009256	-7.5
ZINC03010524	-6.2	ZINC20780055	-6.9	ZINC49879676	-6.5
ZINC03895728	-7	ZINC21970488	-8.1	ZINC49879677	-6.7
ZINC03895762	-7.4	ZINC21970492	-7.6	ZINC49928680	-6.6
ZINC03895763	-7.4	ZINC22987641	-6.4	ZINC50727728	-6.9
ZINC03895795	-7.8	ZINC23215110	-7.1	ZINC53504142	-7.1
ZINC03895796	-7.3	ZINC23457709	-6.9	ZINC55052666	-6.7
ZINC04023481	-7.6	ZINC25804320	-7.4	ZINC55459284	-6.6
ZINC04023484	-7.8	ZINC28197197	-6.9	ZINC57916448	-7.5
ZINC04029335	-7.3	ZINC28197201	-7.3	ZINC57995998	-7.2
ZINC04186266	-6.4	ZINC28622745	-7.6	ZINC58040698	-6.4
ZINC04186269	-6.6	ZINC28622747	-7.5	ZINC58044778	-7.2
ZINC04186272	-6.5	ZINC29572373	-6.2	ZINC58044794	-7.1
ZINC04186276	-6.7	ZINC29709309	-6.7	ZINC58151408	-7.9
ZINC04282968	-7.2	ZINC29709314	-7.3	ZINC58180677	-6.9
ZINC04350416	-6.3	ZINC30446225	-7	ZINC58180679	-7.1
ZINC04350420	-6.6	ZINC32911624	-6.7	ZINC58180681	-6.9
ZINC04350422	-6.7	ZINC32911687	-6.6	ZINC58264901	-6.7
ZINC04436010	-8	ZINC32911688	-6.1	ZINC58323989	-7.5
ZINC04778496	-7.1	ZINC32913889	-6.8	ZINC58323991	-7.6
ZINC04778497	-7.2	ZINC32990561	-6.4	ZINC58357359	-7.3
ZINC04778546	-7.5	ZINC32990562	-6.2	ZINC58357360	-6.9
ZINC04843016	-7.3	ZINC32999421	-6.7	ZINC58429395	-5.8
ZINC04943968	-6.9	ZINC32999422	-6.5	ZINC60323340	-7
ZINC04943969	-6.9	ZINC33185261	-7	ZINC61958002	-5.8
ZINC06357305	-6.7	ZINC33185263	-7.2	ZINC63656099	-8.2
ZINC06357307	-6.5	ZINC36390115	-6.8	ZINC64219574	-6.2
ZINC06727705	-7.1	ZINC36390116	-6.3	ZINC65237739	-7.1
ZINC06727708	-7.3	ZINC36390117	-6.7	ZINC65411677	-6
ZINC07808143	-7.1	ZINC36390118	-6.3	ZINC65428176	-6.3
ZINC08359415	-6.6	ZINC38819910	-6.9	ZINC65494509	-8.1
ZINC08844904	-7.1	ZINC39292200	-6.4	ZINC65591240	-6.8

Ligands	*ΔG (kcal/mol)*	Ligands	*ΔG (kcal/mol)*	Ligands	*ΔG (kcal/mol)*
ZINC11787762	-8	ZINC39292202	-6.3	ZINC65591241	-6.8
ZINC11787763	-7.7	ZINC40135089	-6.6	ZINC65608889	-7.2
ZINC11841690	-7.5	ZINC40148943	-8.3	ZINC65613124	-6.4
ZINC11872915	-7.3	ZINC40148944	-8.4	ZINC65613128	-6.5
ZINC11872916	-6.9	ZINC40148945	-9.7	ZINC65888263	-7.7
ZINC12056228	-7.9	ZINC40148946	-8.8	ZINC67021611	-7
ZINC12330959	-8.1	ZINC40160150	-7.6	ZINC67021613	-6.9

Table 8. *Continued.*

Ligands	Ligands	*ΔG (kcal/mol)*	Ligands	*ΔG (kcal/mol)*
ZINC00260582	ZINC67021615	-7.1	ZINC78869309	-6.9
ZINC00260588	ZINC67130804	-7.2	ZINC78879509	-6.5
ZINC00260594	ZINC67139617	-6.8	ZINC78934925	-7.3
ZINC01092769	ZINC67139644	-6.9	ZINC78934940	-7.4
ZINC01657627	ZINC67142242	-6.9	ZINC78970189	-7.4
ZINC01659837	ZINC67160435	-6.5	ZINC78992220	-7.5
ZINC01697993	ZINC67162584	-7	ZINC78992223	-6.9
ZINC02172656	ZINC67162585	-6.9	ZINC78992231	-7.5
ZINC02172662	ZINC67162586	-7.1	ZINC78992239	-6.9
ZINC02172664	ZINC67162587	-7.3	ZINC79032175	-6.9
ZINC02172665	ZINC67641493	-7.2	ZINC79032180	-6.5
ZINC02172667	ZINC67696163	-8.1	ZINC79056620	-7.2
ZINC02450405	ZINC67872385	-6.3	ZINC79056626	-7.1
ZINC02572700	ZINC69322591	-6.5	ZINC79481111	-7.2
ZINC02833134	ZINC69416513	-6.6	ZINC79481257	-6.9
ZINC03010513	ZINC69416516	-5.9	ZINC83253786.	-7.5
ZINC03010515	ZINC69452680	-6.2	ZINC83253796	-6.6
ZINC03010516	ZINC69638890	-7	ZINC84200225	-7.1
ZINC03010519	ZINC69663458	-6.9	ZINC84240782	-7.7
ZINC03010521	ZINC69663464	-7.1	ZINC85431150	-7.4
ZINC03010523	ZINC69702561	-6.4	ZINC87489344	-6.4
ZINC03010524	ZINC69845810	-6.3	ZINC89306084	-7.9
ZINC03895728	ZINC71868050	-6.6	ZINC89306088	-7.2
ZINC03895762	ZINC71920800	-6.7	ZINC89346194	-7.5
ZINC03895763	ZINC72004261	-6.4	ZINC89836608	-7.9
ZINC03895795	ZINC72072711	-5.8	ZINC89977400	-6.5
ZINC03895796	ZINC72233715	-6.9	ZINC90624800	-6.3
ZINC04023481	ZINC72233716	-7.1	ZINC91302651	-7.6
ZINC04023484	ZINC72294593	-6.9	ZINC91302690	-7.7
ZINC04029335	ZINC75093551	-6.1	ZINC91302974	-7.4
ZINC04186266	ZINC76329178	-7.5	ZINC91302975	-7.4
ZINC04186269	ZINC76722836	-6.5	ZINC91303121	-7.3
ZINC04186272	ZINC76722841	-6.9	ZINC91303284	-7.2
ZINC04186276	ZINC77288048	-6.4	ZINC91351586	-6.1
ZINC04282968	ZINC77322353	-6.8	ZINC91603571	-7.2
ZINC04350416	ZINC77322356	-7.3	ZINC91663118	-6.7
ZINC04350420	ZINC77378506	-7	ZINC91689947	-7.1
ZINC04350422	ZINC77378510	-7.4	ZINC91706387	-7
ZINC04436010	ZINC77387474	-6.9	ZINC91706389	-7
ZINC04778496	ZINC77387476	-6.8	ZINC95385948	-7.4
ZINC04778497	ZINC77471220	-7.1	ZINC95385949	-7.4
ZINC04778546	ZINC77471222	-6.5	ZINC95407889	-7.1
ZINC04843016	ZINC77489799	-6.7	ZINC95452056	-6.4
ZINC04943968	ZINC78598619	-6.3	ZINC95452057	-6.1
ZINC04943969	ZINC78598620	-6	ZINC95508797	-6.9
ZINC06357305	ZINC78601395	-6.4	ZINC95508798	-7
ZINC06357307	ZINC78713760	-6.6	ZINC95936773	-6.4
ZINC06727705	ZINC78713903	-7	ZINC95936774	-6.8
ZINC06727708	ZINC78713909	-6.9	ZINC95952890	-6.3
ZINC07808143	ZINC78713914	-6.6	ZINC95953000	-7.3
ZINC08359415	ZINC78713917	-6.3	ZINC96105481	-7.4
ZINC08844904	ZINC78727254	-7	ZINC96210017	-7
ZINC11787762	ZINC78727255	-6.8	ZINC97163082	-6.9
ZINC11787763	ZINC78727256	-6.4	ZINC97212428	-6.3
ZINC11841690	ZINC78727257	-6.6	ZINC97212429	-6
ZINC11872915	ZINC78822003	-6.7	ZINC98096468	-6.6

Ligands	Ligands	*ΔG (kcal/mol)*	Ligands	*ΔG (kcal/mol)*
ZINC11872916	ZINC78822030	-6.8		
ZINC12056228	ZINC78822097	-6.8		
ZINC12330959	ZINC78822104	-6.4		

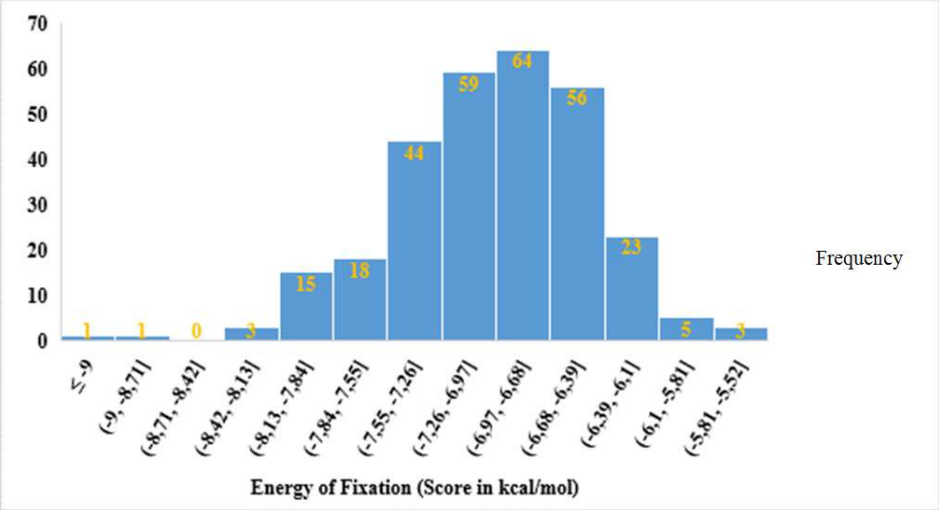


Figure 12. *Frequency distribution plot of the 292 docked compounds based on scores.*

The analysis of the results in Table 8 and Figure 12 showed that, out of the 292 molecules in the library, nearly 141 have a score greater than or equal to -6.9 kcal/mol. These results indicate that the 141 ligands are potentially more active than the starting molecules. In the following work, the 20 best ligands classified according to their score were used for the different studies.

Table 9. *Values of ΔG, ΔΔG, pIC_{50}^{pred} and hii of the top twenty (20) ligands.*

Ligands	*ΔG (kcal/mol)*	*ΔΔG (kcal/mol)*	pIC_{50}^{pred}	h_{ii}
ZINC40148945	-9.7	-2.8	7.892	0.629
ZINC40148946	-8.8	-1.9	7.532	0.165
ZINC40148944	-8.4	-1.5	7.372	0.066
ZINC40148943	-8.3	-1.4	7.332	0.052
ZINC63656099	-8.2	-1.3	7.291	0.041
ZINC12330959	-8.1	-1.2	7.251	0.035
ZINC21970488	-8.1	-1.2	7.251	0.035
ZINC65494509	-8.1	-1.2	7.251	0.035
ZINC67696163	-8.1	-1.2	7.251	0.035
ZINC01092769	-8	-1.1	7.211	0.033
ZINC04436010	-8	-1.1	7.211	0.033
ZINC11787762	-8	-1.1	7.211	0.033
ZINC12056228	-7.9	-1	7.171	0.036
ZINC12330962	-7.9	-1	7.171	0.036
ZINC12429338	-7.9	-1	7.171	0.036
ZINC12880574	-7.9	-1	7.171	0.036
ZINC14957156	-7.9	-1	7.171	0.036
ZINC58151408	-7.9	-1	7.171	0.036
ZINC89306084	-7.9	-1	7.171	0.036
ZINC89836608	-7.9	-1	7.171	0.036

Virtual screening of the chemical library similar to the M5 molecule shows the ligand ZINC40148945 as the best potential inhibitor of the main protease of SARS-CoV-2 with an interaction energy equal to -9.7 kcal/mol. Visualization of the interaction mode of this compound within the active site 1 of the protein reveals the presence of four (4) hydrogen bonds. The first one is formed between the OH group of the compound ZINC40148945 and the amine function of glycine (GLY140), separated by a distance of 2.41 Å. The second is found between the hydroxyl group of the inhibitor and the amine function of cysteine (CYS142), with a distance of 2.6 Å. The third is also observed between the hydroxyl group of the ligand and the amine function of leucine (LEU138) with a distance of 1.99 Å. As for the fourth, it takes place between the sp2 oxygen atom of the ester function of the compound and the amine function of glutamine (GLN186) with a distance of 2.37 Å. The protein-ligand complex ZINC40148945 is also stabilized by hydrophobic interactions involving residues: HIS38 and MET162 of the active site of SARS-CoV-2 (Figure 13).

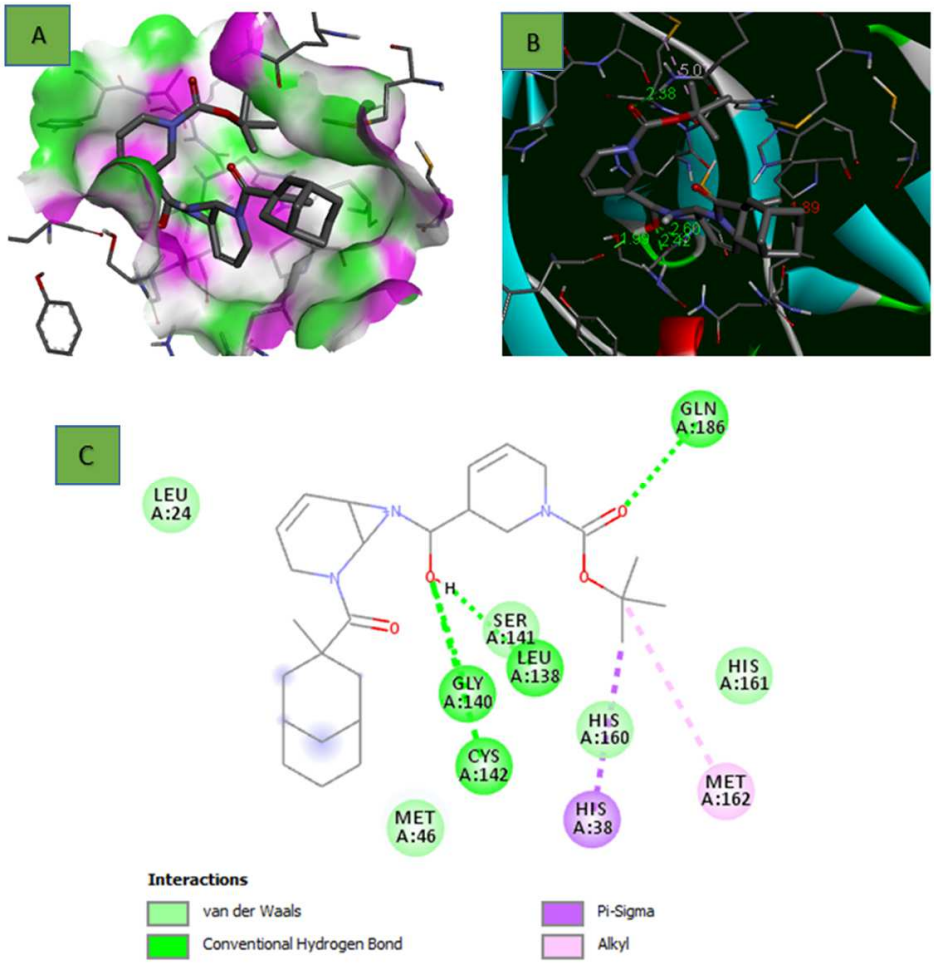


Figure 13. *3D (B) and 2D (C) interactions between the ZINC40148945 molecule and the active site residues of the protein. (A) Surface of the hydrogen bonds formed between the active pocket and the ZINC40148945 molecule.*

Table 10. *Lipinski parameters of new ligands with enhanced biological activities.*

Molecules	M (g/mol)	HBD	HBA	MlogP
Rule	**<500**	**<5**	**<10**	**<4.15**
ZINC40148945	473.65	1	4	3.06
ZINC40148946	473.65	1	4	3.06
ZINC40148944	473.65	1	4	3.06
ZINC40148943	473.65	1	4	3.06
ZINC63656099	341.4	0	5	0.67
ZINC12330959	469.57	0	5	2.55
ZINC21970488	433.58	2	4	2.45
ZINC65494509	341.4	0	5	0.67
ZINC67696163	340.42	3	5	-0.15
ZINC01092769	358.43	1	4	2.12
ZINC04436010	403.47	1	5	1.57
ZINC11787762	437.53	1	4	2.2
ZINC12056228	416.51	2	5	0.98
ZINC12330962	469.57	0	5	2.55
ZINC12429338	365.47	1	4	1.36
ZINC12880574	456.53	1	5	0.95
ZINC14957156	402.49	2	5	0.76
ZINC58151408	422.52	1	5	0.96
ZINC89306084	433.54	1	5	1.32
ZINC89836608	387.47	2	4	1.55

3.4.1. Druglikeness Prediction

For new molecules to be considered as drugs, they must comply with several rules. These molecules must be soluble, stable, and have pharmacokinetic and pharmacodynamic properties of absorption, distribution, metabolism, excretion and toxicity (ADMET). To verify these rules, we used the Lipinski rule predicted from the SWISSADME server [23] and ADMET prediction from the PreADMET server [25].

The Lipinski rule (Rule of five)

The quantities characterizing the Lipinski rule, namely the molar mass (M), the number of hydrogen donors (HBD), the number of hydrogen acceptors (HBA) and the lipophilicity (MlogP) have been determined. In order for a molecule to be administered as a drug, it must meet the Lipinski conditions. The values of these parameters are recorded in Table 10.

The analysis of the results of the table shows that all the molecules have values of molar mass lower than 500 g/mol, values of the number of hydrogen donors lower than 5 and values of the number of hydrogen acceptors lower than 10. This result implies that all the new molecules can easily pass by the cellular membrane. Also, we notice that these molecules have lipophilicity values lower than 4.15. This means that the compounds have a good solubility in water, a better gastric tolerance, an efficient elimination by the

kidneys and a good permeability through the cell membrane. In sum, the new biologically more active molecules respect Lipinski's rule, they are therefore orally administrable according to Lipinski.

3.4.2. Prediction of Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) of New Molecules

A good drug candidate must be rapidly and completely absorbed from the gastrointestinal tract, distributed specifically to its site of action in the body, metabolized in a manner that does not impair body functions, and eliminated appropriately without causing harm [31]. The prediction of pharmacological properties of absorption, distribution, metabolism, excretion and toxicity of the new molecules was performed using the PreADMET online server. We determined parameters such as, Human Intestinal Absorption (HIA), in vitro caco-2 cell permeability, in vitro MDCK (Mandin Darby Canine Kidney) cell permeability, plasma protein binding (PPB), blood-brain barrier penetration (BBB), cytochrome P450 enzyme inhibition (CYP2D6, CYP2C9, CYP2C19, CYP3A4), hERG (human Ether-à-go-Related Gene) inhibition, carcinogenicity and mutagenicity. These different parameters are listed in Table 6.

Table 11. *Prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) parameters of new compounds with improved activities.*

Molecules	Absorption			Distribution		Metabolism					
	HIA	Caco-2	MDCK	PPB	BBB	CYP2D6 Inhibition	CYP2D6 substrat	CYP2C19 Inhibition	CYP2C9 Inhibition	CYP3A4 Inhibition	CYP3A4 Substrat
ZINC40148945	96.693	23.244	0.048	82.174	0.181	No	No	No	No	No	Substrate
ZINC40148946	96.693	23.244	0.048	82.174	0.181	No	No	No	No	No	Substrate
ZINC40148944	96.693	23.244	0.048	82.174	0.181	No	No	No	No	No	Substrate
ZINC40148943	96.693	23.244	0.048	82.174	0.181	No	No	No	No	No	Substrate
ZINC63656099	93.789	23.59	12.217	57.955	0.308	No	No	No	No	No	Substrate
ZINC12330959	99.296	22.638	0.239	83.891	0.238	No	No	No	No	Yes	Substrate
ZINC21970488	93.416	21.486	0.125	85.549	0.432	No	No	No	No	No	Substrate
ZINC65494509	93.789	23.59	12.217	57.955	0.308	No	No	No	No	No	Substrate
ZINC67696163	76.246	14.94	25.205	10.66	0.039	Yes	Low	Yes	No	No	No
ZINC01092769	96.423	21.011	15.375	87.838	0.115	No	No	No	No	No	Low
ZINC04436010	96.584	21.781	6.197	75.205	0.013	No	No	No	No	No	Low
ZINC11787762	96.401	22.604	0.215	78.545	0.051	No	No	No	No	No	Low
ZINC12056228	93.977	22.294	6.684	38.812	0.022	Yes	Low	Yes	No	No	Substrate
ZINC12330962	99.296	22.638	0.239	83.891	0.238	No	No	No	No	Yes	Substrate
ZINC12429338	94.879	20.66	34.718	53.759	0.017	No	No	No	No	No	Low
ZINC12880574	95.89	21.119	2.181	62.852	0.088	No	No	No	No	No	Low
ZINC14957156	93.782	22.09	4.269	36.358	0.017	Yes	Low	Yes	No	No	Substrate
ZINC58151408	93.178	21.473	0.651	54.054	0.094	Yes	No	No	No	No	Substrate
ZINC89306084	96.157	30.169	0.371	65.604	0.085	No	No	No	No	No	Substrate
ZINC89836608	93.627	21.551	3.736	84.912	0.13	No	No	No	No	No	Low
Hydroxychloroquine	94.660	46.083	45.108	88.996	2.287	Yes	Substrate	No	No	No	Low

Table 11. *Continued.*

Molecules	Toxicity			
	hERG Inhibition	Carcinogenicity		Mutagenicity
		Mouse	Rat	(AMES-Test)
ZINC40148945	Low Risk	Negative	Positive	Non-Mutagenic
ZINC40148946	Low Risk	Negative	Positive	Non-Mutagenic
ZINC40148944	Low Risk	Negative	Positive	Non-Mutagenic
ZINC40148943	Low Risk	Negative	Positive	Non-Mutagenic
ZINC63656099	Low Risk	Negative	Positive	Mutagenic
ZINC12330959	Low Risk	Negative	Positive	Mutagenic
ZINC21970488	Low Risk	Negative	Positive	Non-Mutagenic
ZINC65494509	Low Risk	Negative	Positive	Mutagenic
ZINC67696163	Low Risk	Negative	Positive	Mutagenic
ZINC01092769	Medium risk	Negative	Négative	Mutagenic
ZINC04436010	Low Risk	Negative	Positive	Mutagenic
ZINC11787762	Medium risk	Negative	Positive	Mutagenic
ZINC12056228	Low Risk	Negative	Negative	Mutagenic
ZINC12330962	Low Risk	Negative	Positive	Mutagenic
ZINC12429338	Low Risk	Negative	Positive	Mutagenic
ZINC12880574	Ambiguous	Negative	Positive	Mutagenic
ZINC14957156	High risk	Negative	Negative	Mutagenic
ZINC58151408	Low Risk	Positive	Positive	Non-Mutagenic
ZINC89306084	Medium risk	Positive	Positive	Non-Mutagenic
ZINC89836608	High risk	Negative	Negative	Mutagenic
Hydroxychloroquine	Medium risk	Negative	Negative	Mutagenic

HIA (%) is the percentage of human intestinal absorption (from 0 ~ 20% poor absorption; from 20 ~ 70% medium absorption, from 70~100% high absorption). *Caco-2(nm/s) and MDCK (nm/s*) predicts the intestinal permeability of a compound on Caco-2(<4 poor permeability, between 4 ~70 medium permeability, >70 high permeability) and MDCK cells. *PPB (Plasma Protein Binding%)* predicts the degree of drug binding to proteins in blood (<90 low binding, >90 high binding*). BBB (Blood-Brain Barrier%*) predicts blood-brain barrier penetration (<0.1 low absorption in the Central Nervous System (CNS), 0.1~2 medium absorption in the CNS and >2 high absorption in the CNS*). P450 Cytochromes (CYP2D6, CYP2C19, CYP2C9 and CYP2A4)* are important in the oxidative metabolism of compounds. *hERG (human Ether-à - go - go-Related Gene)* is an ion (potassium) channel that moves potassium out of its cell. *AMES-Test (Salmonella typhimurium reverse Mutation Assay*) predicts the mutagenic potential of a molecule.

The analysis of the table shows us that at the level of absorption, the values of human intestinal absorption (HIA) are between 76.246 and 99.296%, these values are higher than 70%, which translates that the molecules have a strong absorption in the human intestine, therefore can be well assimilated by it. Moreover, the molecule ZINC40148945 has a higher HIA value than hydroxychloroquine. For the permeability on the Caco-2 cell the values of the compounds are between 14.94 and 30.169 nm/s, as for the permeability on the MDCK cell the values are between 0.048 and 34.718 nm/s, this result shows that the new compounds are permeable on the Caco-2 and MDCK cells at the distribution level, the values of the degree of plasma protein binding (PPB) are less than 90%, this reflects that the compounds have therefore a low binding on plasma proteins. For the values of the BBB, we notice that the compounds have either an average or a weak absorption at the level of the blood-brain barrier of the CNS. This criterion is much more important for compounds whose target is in the CNS, which is not the case here. In addition, this may protect the brain from a likely adverse effect of these compounds. For the metabolism process, the compound (ZINC40148945) with the highest docking score showed no effect on cytochrome P450 iso-form inhibition. However, it appears as a substrate for the CYP3A4 isoform. Inhibition of these enzymes is an important source of adverse drug interactions as changes in CYP enzyme activity can affect drug metabolism. Finally, for toxicity testing, the molecule ZINC40148945 has a low risk for hERG inhibition, so it is non-carcinogenic to mice and non-mutagenic by the AMES test. Nevertheless, it is carcinogenic for rats. These different results indicate that the new proposed molecules have good properties of absorption, distribution, metabolism and toxicity. These compounds can therefore be used as drugs.

4. Conclusion

In this work, the methods of Molecular Mechanics and Molecular Modeling were employed on ten (10) inhibitors in order to propose new molecular structures with improved anti-COVID-19 activities. This theoretical study was performed the Amber force field to minimize the protein energy and ligand structure. Molecular docking between the ligands and the main coronavirus protease (3CLpro) showed that compound M5 has the highest score. The different linkages involved in the interactions between the active site residues and the ligands identified compound M5 as the best ligand that leads to the formation of the most stable complex. The Quantitative Structure-Activity Relationship study conducted allowed the establishment of a mathematical model between the inhibitory potential (pIC50) of the 3CLpro inhibitors and the variation of the binding energy between the protein and the ligands. The statistical indicators of the model showed that it is acceptable, robust with a good predictive power ($R^2 = 0.9137$; $S = 0.058$, $F = 52.942$). The chemical space in which the model is applicable with lever threshold h*=0.667 covers all the molecules studied. The antiCOVID-19 activity of the inhibitors predicted in this work is reliable. The screening of the M5-like molecule library resulted in 292 molecules with better scores than the starting compound. Of the 292 ligands we selected the best 20. The ADMET study showed that the 20 compounds have good absorption, distribution, metabolism and toxicity properties and can be used as oral drugs. In perspectives, we plan to do a molecular dynamics simulation study.

References

[1] W. J. Guan, Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, H. Shan, C. L. Lei, D. Hui, B. Du, L. J. Li, G. Zeng, K. Y. Yuen, R. C. Chen, C. L. Tang, T. Wang, P. Y. Chen, J. Xiang, S. Y. Li, J. Wang et Z. Liang, «Clinical Characteristics of Coronavirus Disease 2019 in China,» *The New England Journal of Medicine,* vol. 382, n°%118, pp. 1708-1720, 2020.

[2] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang et Z. Peng, «Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan,» *JAMA,* vol. 323, n°%111, p. 1061–1069, 2020.

[3] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, «The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2,» *Nature Microbiology,* vol. 5, p. 536–544, 2020.

[4] M. Julie, «CORONAVIRUS WORLDWIDE WEDNESDAY, JUNE 9, 2021: NEW CASES AND DEATHS IN 24 HOURS,» 08 Juin 2021. [Online]. Available: https://www.sortiraparis.com/actualites/a-paris/articles/212134-coronavirus-dans-le-monde-mardi-8-juin-2021-nouveaux-cas-et-morts-en-24h. [Access on 2021 July 09].

[5] R. Hilgenfeld, «From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design.,» *FEBS J,* vol. 281, p. 4085–4096, 2014.

[6] L. Zhang, D. Lin et X. Sun, «Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors.,» *Science,* vol. 368, p. 409–412, 2020.

[7] K. Anand, J. Ziebuhr et P. Wadhwani, « Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.,» *Science,* vol. 300, p. 1763–1767, 2003.

[8] D. Juckel, J. Dubuisson et S. Belouzard, «Coronaviruses, uncertain enemies.,» *Med Sci,* vol. 36, p. 633–641, 2020.

[9] K. Anand, G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr et R. Hilgenfeld, «Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain,» *The EMBO Journal,* vol. 21, n°%113, p. 3213–3224, 2002.

[10] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. Gao, K. Anand, M. Bartlam, R. Hilgenfeld et Z. Rao, «The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor,» *Proceedings of the National Acamedy of Sciences of the United States of America,* vol. 100, n°%123, p. 13190–13195, 2003.

[11] A. Rathnayake, J. Zheng, Y. Kim, K. Perera, S. Mackin, D. Meyerholz, M. Kashipathy, K. Battaile, S. Lovell, S. Perlman, W. Groutas et K. Chang, «3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV-infected mice,» *Sci. Transl. Med,* vol. 12, n°%1557, pp. 1-16, August 2020.

[12] M. J. Frisch, G. W. Trucks, H. B. Schlegel et G. E. Scuseria, «Gaussian 09, Revision A.02,» Gaussian, Inc., Wallingford CT, 2009.

[13] S. Chaltterjee, A. Hadi et B. Price, «Regression Analysis by Examples,» *Wiley VCH: New York,* 2000.

[14] H. Phuong, «Synthesis and quantitative structure/activity relationship study (QSAR/2D) of Benzo[c]phenanthridine analogues,» France, 2007.

[15] M. Excel, 2016.

[16] E. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng et T. Ferrin, «UCSF Chimera--a visualization system for exploratory research and analysis,» *J. Comput. Chem.,* vol. 25, n°%113, pp. 1605-1612, 2004.

[17] J. Fantini et N. Yahi, «Structural and molecular modelling studies reveal a new mechanism of action of Chloroquine and hydroxychloroquine against SARS-CoV-2 infection,» *International Journal of Antimicrobial Agents,* vol. 55, n°%15, p. 105960, May 2020.

[18] A. Elfiky, «Ribavirin, Remdesivir, Sofosbuvir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study,» *Life Sci.,* vol. 253, p. 117592, 2020.

[19] A. Kumar, A. Kumar, G. Tiwari, R. Kumar et N. Misra, «In Silico Investigations on the Potential Inhibitors for COVID-19 Protease,» *arXiv: 2003.10642,* 2020.

[20] T. A. K. a. J. M. M. Roy Dennington, «GaussView, Version 6,» Semichem Inc. Shawnee Mission KS, 2016.

[21] C. Lipinski, F. Lombardo, B. Dominy et P. Feeney, «Experimental and computational approaches to estimating solubility and permeability in drug discovery and development parameters,» *Adv. Drug Deliv. Rév.,* vol. 1–3, n°%146, p. 3–26, mars 2001.

[22] C. Lipinski, «Lead and drug compounds: the rule of five revolution,» *Drug Discovery Today: Technologies,* vol. 1, n°%14, p. 337–341, décembre 2004.

[23] A. Daina, O. Michielin et V. Zoete, «SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.,» *Sci Rep,* vol. 7, p. 42717, 2017.

[24] « The metabolism of a drug,» 2021. [Online]. Available: http://udsmed.u-strasbg.fr/pharmaco/pdf/DCEM1_Pharmacologie_chapitre_5_Metabolisme_des_medic.pdf.. [Accès le 20 fevrier 2021].

[25] S. Lee, G. Chang, I. Lee, J. Chung, K. Sung et K. No, «PreADMET,» 2021. [Online]. Available: https://preadmet.bmdrc.kr/.

[26] F. W. Schueler, «Chemobiodynamics and drug design,» *Acad. Med.,* vol. 36, n°%13, p. 285–286, 1961.

[27] S.-Y. Yang, «Pharmacophore modeling and applications in drug discovery,» *Drug Discov. Today,* vol. 15, n°%111–12, p. 444–450, 2010.

[28] J. Wang, P. Morin, W. Wang et A. P. A. Kollman, «Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA.,» *J. Am. Chem. Soc.,* vol. 123, pp. 5221-5230, 2001.

[29] A. Fortuné, «Molecular Modeling Techniques applied to the Study and Optimization of Immunogenic Molecules and Modulators of Chemoresistance Drugs,» 2006.

[30] L. Eriksson, J. Jaworska, A. Worth, M. D. Cronin, R. M. Mc Dowell et P. Gramatica, «Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs,» *Environmental Health Perspectives,* vol. 111, n°%110, pp. 1361-1375, 2003.

[31] L. Stryer, J. M. Berg et J. L. Tymoczo, Drug development, 7e éd., vol. 36, Médecine Sciences, 2013, p. 1029–1054.