Natural Products-based drug design from Bahia Semi-Arid region against SARS-CoV-2 Mpro (3Clpro)

Rai Campos Silva[1], Humberto Fonseca de Freitas[2,3], Njogu M. Kimani[4], Samuel Silva da Rocha Pita[3], Cleydson Breno Rodrigues dos Santos[1,5*]

[1]Graduate Program on Medicinal Chemistry and Molecular modeling, Institute of Health science, Federal University of Pará. Augusto Corrêa, 01 - Guamá, Belém - PA, 66075-110, Brazil.

[2]Graduate Program on Pharmacy (PPGFAR-UFBA), Pharmacy College, Federal University of Bahia, Salvador, Bahia, Brazil

[3]Bioinformatics and Molecular Modeling Laboratory (LaBiMM), Federal University of Bahia, Av. Barão de Jeremoabo, 147, Pharmacy College, Térreo, Ondina, 40170-115, Salvador, BA, Brazil.

[4]Department of Physical Sciences, University of Embu, P. O. BOX 6-60100, Embu, Kenya.

[5]Laboratory of Modeling and Computational Chemistry, Department of Biological and Health Sciences, Federal University of Amapá, 68902-280 Macapá, AP, Brazil

* Corresponding author: breno@unifap.br; samuel.pita@ufba.br

ORCIDS ID's:

RCS https://orcid.org/0000-0003-1774-4164; camposchemistry@gmail.com

HFF https://orcid.org/0000-0003-3040-9694; humbarato@gmail.com

MNK https://orcid.org/0000-0002-5171-1940; njogu.mark@embuni.ac.ke

SSRP https://orcid.org/0000-0003-4053-8721; samuelrpita@gmail.com

CBS https://orcid.org/0000-0002-0271-335X; breno@unifap.br

Highlights:

Natural Products from Bahia Semi-Arid region (NatProDB, Brazil's biome) was studied.

In silico pharmacokinetic and toxicological (ADMET) properties selected compounds with potential antiviral activity against SARS-CoV-2 Mpro.
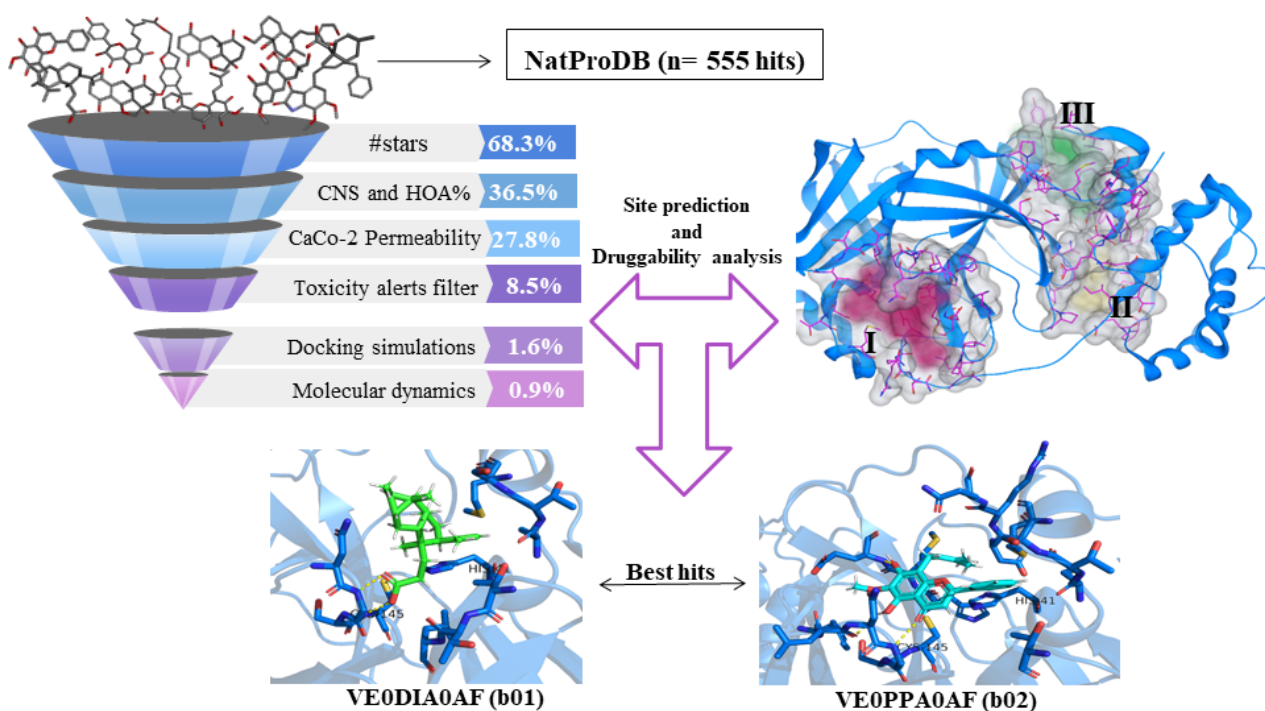
Site prediction and Druggability analysis of SARS-CoV-2 Mpro were combined with Molecular Docking Virtual Screening and Molecular Dynamics Simulations choosing Natural Products ligands.

This rational design approach elected VE0DIA0AF (b01) and VE0PPA0AF (b02) as potential inhibitors of SARS-CoV-2 Mpro (3Clpro).

Abstract

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), receives global attention due to the serious threat that affect public health. Since December 2019, its incidence affecting millions of patients and its rapid dissemination on a worldwide scale have led the searching for its treatment. To discover hit compounds that can be used alone or in combination with repositioned drugs, we initiated a structure-assisted drug design program including Virtual Screening, Docking, pharmacokinetic and toxicological analysis (ADMET) and Molecular Dynamics Simulation (MD) from Natural Products Database of the Bahia Semi-Arid region (NatProDB). We also aimed to identify novel scaffolds that target the SARS-CoV-2 Main Protease (Mpro, 3Clpro) since this protein plays a pivotal role in mediating viral replication and transcription, which makes it an attractive drug target for coronaviruses. Here, we selected 10 molecules that could be studied *in vitro* to lead discovery in response to this infectious diseases. The best SARS-CoV-2 Mpro complexes interactions revealed that some enzyme sites were accessed thereby confirming that this method can be employed as a suitable starting method for the identification of novel SARS-CoV-2 Mpro inhibitors. Two compounds (b01 and b02) suggest a better potential for interaction with SARS-CoV-2 main protease (Mpro) and could be further studied.

Graphical Abstract

Keywords:

1. Introduction

Although coronaviruses (CoVs) have been known since 1940 [1], reports of humans infections causing mild respiratory infections occurred in the 1960s. In December 2019 [2-6], a new viral respiratory disease initiated from Wuhan, China has been spread globally. This virus was named SARS-CoV-2, because its genomic RNA is about 82% identical to the coronavirus that causes Severe Acute Respiratory Syndrome (SARS-CoV). The disease caused by SARS-CoV-2 is called COVID-19 [4]. On March 11, the World Health Organization (WHO) declared this COVID-19 outbreak pandemic [3-5] and data indicated that the cumulative number of cases reported globally now exceeds 183 million and the number of deaths is almost 4 million (https://www.who.int/emergencies/diseases/novel-coronavirus-2019, accessed 09/06/2021).

Therefore, the rapid discovery of safe, effective, and broad-spectrum anti-COVID-19 drugs is urgent. Currently, there is no specific therapy for COVID-19 patients and treatment are limited [7], despite more than 80 clinical trials employing distinct drugs: chloroquine, hydroxychloroquine, arbidol, remdesivir, favipiravir, lopinavir, ritonavir, oseltamivir, methylprednisolone, bevacizumab [1, 8]. Due to its

relevance to viral replication, the Mpro enzyme, which plays a central role in mediating viral replication and transcription [4, 7, 9-11], was the main target investigated in this study and the compounds identified here will be evaluated in future assays.

Natural Products research and development (R&D) potentially plays a pivotal role in innovative drug discovery [12] like an important source of new drugs [12-14] and a great inspiration to generate compounds libraries [15]. Medicinal plants have been employed by humans for thousands of years to treat illnesses, health disorders [13] and identifiying some plants useful as antiviral drugs. Plants produce a high diversity of secondary metabolites with interesting biological activities and their natural products and natural-product-inspired agents have attracted significant attention because they have played an integral role in the treatment of many different conditions like viral infection, especially SARS-CoV-2 [13, 16-21].

Natural products (carolacton, homoharringtonine, emetine, and cepharanthine) and natural product-inspired small molecules (ivermectin, GS-5734, EIDD-2801, and ebselen) are potential anti-SARS-CoV-2 agents that have attracted significant attention due to their broad-spectrum antiviral activities [21]. However, they need to be subjected to detailed tests especially in animal models and if successful, in clinical trials [13].

A good starting point to find new antiviral natural products would be traditional medicinal plants, such as those from Asia, Africa or America that have been employed to treat infections [13]. Brazil hosts about 20% of all world biodiversity, being considered the country with the most endemic species worldwide [22]. The Natural Products Database of the Bahia Semi-Arid region (NatProDB, http://natprodb.uefs.br/) contains naturally occurring compounds with a wide chemical diversity. These compounds are yet to be fully explored for the discovery of bioactive molecules [23]. Thus, we searched for new molecular entities that could be potent inhibitors of SARS-CoV-2 Mpro in the NatProDB database.

Medicinal chemistry approaches have seen a great technological advancement over the years and have contributed immensely to the discovery of promising molecules. However, many compounds discovered by these techniques have shown unsatisfactory absorption, distribution, metabolism, excretion and toxicity (ADMET) properties in *in vivo* tests [24]. To overcome these pharmacokinetic and toxicological (PK & T) issues some *in silico* methods were developed and are freely available over

the web, namely Pred-Herg4.2 [25-26], http://labmol.com.br/predherg/; Aggregator Advisor [27], http://advisor.bkslab.org/; FAF-DRUGS [28], http://fafdrugs4.mti.univ-paris-diderot.fr/ and pKCSM [24], http://biosig.unimelb.edu.au/pkcsm/. Thus, in this study, we explore ADMET properties using two distinct tools QikProp™ and Derek™ intending to reduce the selection of false-positive ligands and obtain hits that can be subjected to *in vitro* tests against the SARS-CoV-2 Mpro enzyme.

For this reason, we propose to employ a structure-based Drug Design (SBDD) approach against the main sites on SARS-CoV-2 main protein (Mpro) through different computer-aided methods.

2. Materials and Methods:

Molecular Modelling calculations, estimation, and visualizations of SARS-CoV-2 Mpro binding affinities were carried out using SeeSAR version 9.2 software package (BioSolveIT GmbH) with HYDE visual affinity assessment [29-31] carried on a personal computer with 4 logical processors of an Intel® Core ™ i5-6200U, 2.3 GHz processor using Windows 10 Home operational system.

2.1 SARS-CoV-2 Mpro protein preparation:

The high-resolution X-ray crystallographic structure of SARS-CoV-2 Mpro (1.31Å) was obtained from the Protein Data Bank (PDB ID: 5R82, [32]) and was imported into the SeeSAR version 9.2 (BioSolveIT, Sankt Augustin, Germany, 2018, www.biosolveit.de/SeeSAR). The structures were then prepared using the Protein Editor tool as implemented in the BioSolveIT Suite. Briefly, raw PDB structure was processed by automatically assigning bond orders, adding hydrogens, adding missing side-chains, creating possible disulfide bridges, deleting waters, and generating hetero protonation states at pH 7.4. Residues with alternate positions were locked in the conformations with the highest average occupancy. Small ligand (6-(ethylamino) pyridine-3-carbonitrile - RZS) and dimethyl sulfoxide (DMS) originating from crystallization buffer were removed. The hydrogen-bonding networks were optimized automatically, by sampling water orientations and optimization of hydroxyls, Asn, Gln, and His residue states using Protein Editor.

*2.2 Site prediction and Druggability analysis of SARS-CoV-2 **Mpro***:

2.2.1   SeeSAR

The protein structure prepared as described above were exported as PDB files and inserted on the "Binding Site" tool using BioSolveIT Suite. The site predictions and druggability analysis were determined by applying the Difference of Gaussian (DoG) Site pocket finder [33-34]. DoGSite provides the functionality to detect potential binding pockets and subpockets of a protein of interest [34]. Subsequently, it analyzes the geometric and Physico-chemical properties of these pockets and estimates druggability with aid of a support vector machine (SVM) [34]. Thus, binding pockets and their druggability were extracted and analyzed employing the default parameters.

2.2.2   PockDrug

PockDrug predicts pocket druggability on both pockets guided and not guided by the ligand proximity [35] using different estimation methods based on 36 physicochemical and 16 geometrical descriptors to characterize each estimated pocket [36]. The prepared PDB file was uploaded to the PockDrug server (http://pockdrug.rpbs.univ-paris-diderot.fr/cgi-bin/index.py?page=Home) and assessed for binding sites, their 6 pocket descriptors [Volume of convex hull ($Å^3$); Kyte scale of residues' hydrophobicity [37]; Frequency of polar residues; Frequency of aromatic residues; Oxygen atoms frequency from Tyr residues in pocket [38]; Number of pocket residues] and corresponding average druggability probability and its associated standard deviation. For a probability greater than 0.5, pockets are considered as druggable according to the published data [35]. Local pocket properties, size, shape, and hydrophobicity were extracted for all identified sites and annotated to pocket numbers.

2.2.3   FTMap

FTMap identifies binding hot spots of macromolecules (regions of the surface with major contributions to the ligand-binding free energy) using a computational mapping approach [39]. It uses small organic molecules as probes to sampling the protein surface and it scores them using the interaction energy [40]. Protein sites that bind different probes indicate this region as binding hot spots [39]. The prepared PDB file was uploaded to the FTMap server (https://ftmap.bu.edu/) and examined for the number of probes per cluster found according to the published protocols [41-43].

Similar to Borrel and cols [35], in this study, the predicted pocket of interest from the entire set of candidate cavities detected was selected as the one that best overlaps the crystallographic ligand-binding site and presented the best druggability index on all three programs.

2.3 Molecular Docking on SARS-CoV-2 **Mpro**:

Docking experiments were performed on previously prepared SARS-CoV-2 **Mpro** structure with the SeeSAR version 9.2 through sampling their spatial coordinates on the best interaction sites which were defined by the DoG Site pocket finder [33-34] and druggability analysis from SeeSAR, PockDrug and FTMap. Then, the docking library was loaded, which includes a set of NatProDB ligands previously filtered from QikProp^{TM} and Derek^{TM}, plus **E64**, a well know cysteine protease inhibitor as a positive control.

Docking calculations were done for each compound generating 20 poses which one has their binding affinity estimated from µM into SARS-CoV-2 **Mpro** site. The best poses were selected based on their estimated affinity from the HYDE score function, while also considering the Torsion angle values to the binding conformation of protein-ligand [29, 33, 34, 44]. HYDE's empirical score function (equation (1)) rely on intrinsically balanced terms of atom specific desolvation, hydration and hydrogen bond (H-bond) (*i.e.*, free of weighting parameters) based on logP atomic increment system [45]. The binding affinity of a ligand described as the quotient of ΔG and the non-hydrogen numbers of atoms in the compound (equation (2)) [46].

$$\Delta G_{Hyde} = \Sigma_{atom\ i}\ [\Delta G^i_{Dehyd}{}^n + \Delta G^i_{h\text{-}bonds}] \tag{1}$$

$$LE = \Delta G / N \tag{2}$$

Where ΔG = -RT ln Ki, N = number of non-hydrogen atoms.

The selection of the best poses was based on their visual HYDE scores while also considering a statistics-based torsional analysis [44]. SeeSAR version 9.2 enables an interactive assessment of torsions and energies (in kJ/mol) including the desolvation (dehydration, −TΔS) and enthalpic (interaction, ΔH) contributions to binding for both protein and ligand. Furthermore, it also quantitatively reports the energy contributions for all heavy atoms (with a united atom approach for bounded H-atoms) and allows a semi-quantitative estimation of the thermodynamic profile for all tested compounds.

Analysis, comparisons, visualization of the predicted binding pose predictions and all images were done by educational Pymol version 2.4.1 [47].

2.4 SARS-CoV-2 ligand preparation:

The Natural Products Database of the Bahia Semi-Arid region (NatProDB, http://natprodb.uefs.br/) contains 555 molecular structures publicly available [23]. All these molecules were automatically employed to perform an ADMET property analysis by QikProp^{TM} and Derek^{TM}.

2.5 ADMET analysis by QikProp^{TM} and Derek^{TM}:

The predictions described below were made on a computer with 6 logical processors of an Intel® Core TM i5-9400, 4.10 GHz processor using Windows 10 Professional operational system.

QikProp program provides ranges to compare the physically significant descriptors and important ADMET properties of a specific compound with those of 95% of approved drugs. Predicted human oral absorption on 0 to 100% scale (HOA%), predicted apparent Caco-2 cell permeability in nm/sec (QPPCaco), predicted apparent Maden Darby Canine Kidney cell permeability (QPPMDCK), predicted brain/blood partition coefficient (QPlogBB), predicted octanol/water partition coefficient (QPlogPo/w) among many others. The program analyzes the chemical structure as fragments or a whole and supports predictions about the structure in three dimensions. In addition, flags 30 types of reactive functional groups that can cause false positives in high-throughput screening (HTS) assays [48-49].

Derek Nexus is an "expert rule-based" system for the prediction of toxicity [48-50] that uses a virtual base of molecules with reported toxicity alerts, for example, mutagenicity, genotoxicity, teratogenicity, carcinogenicity and hepatotoxicity for comparison with the potential toxic group present in each query molecule [51]. The program generates an alert based on literature evidence and describes the potential toxicity for the complete structure [52].

*2.6 Molecular Dynamics (MD) Simulations on SARS-CoV-2 **Mpro**-NatProDB complexes:*

The SARS-CoV-2 **Mpro** systems *apo*, *holo* and five complexed with better docked NatProDB compounds were submitted to Molecular Dynamic (MD) simulations using GROMACS 5.1.4 [53-59] available on National Center for High-Performance Computing in São Paulo (CENAPAD-SP) employing the following parameters: 1 atm, 310 K, pH 7.4, GROMOS54A7 force field updated [60], PME [61] for electrostatic treatment with cut-off = 1.0 nm in a dodecahedral box solvated with water model SPC/E [62] with periodic boundary conditions (PBC). Na^+ and Cl^- ions were added to maintain the physiological salt concentration (0.15M) and to neutralize the residual system charge at pH = 7.4. At first, the system was energy minimized (steepest descent /conjugate gradient) until forces reached $\leq$ 10 kJ.mol^{-1}.nm^{-1} followed by a pre-equilibrium simulation step (heavy atoms position restrained for 1 ns).

Finally, 100 ns of the unrestrained simulation was performed for all SARS-CoV-2 **Mpro** systems where atomic coordinates were recorded every 10 ps for later analyses in a NPT ensemble with V-rescale thermostat [63] and Berendsen barostat [64] with SETTLE [65] algorithm for solvent bonds a LINCS [66] for other bonds. The crystallographic ligand (RZS) and NatProDB compounds had their topology coordinates built in the Automated Topology Builder (ATB) version 3.0 server (http://compbio.biosci.uq.edu.au/atb/, [67-69].

The 3D-Root Mean Square Fluctuation (RMSF) analysis converted RMSF data to B-factor from *rmsf* module of GROMACS 5.1.4 [53-59]. The hydrogen bond analysis was made on GROMACS *hbond* module [70] and HbMap2Grace program [71] available on http://lmdm.biof.ufrj.br/software/hbmap2grace/index.html and Molecular Surface Area (Å^2) (Table 1) by SurfinMD program [72] accessible on http://lmdm.biof.ufrj.br/software/surfinmd/index.html.

2.7 Hydrogen bond capacity analysis

Hydrogen bonds have been calculated between the number of donors/acceptors and distance/angle among them and there are few discussions about this relationship in drug discovery. Hydrogen bond capacity was quantitatively described [73-74] through an empirical definition. Wang et al [75] described the hydrogen bond occurrence

probability from MD simulation results and based on their definition, we calculated the hydrogen bond capacity (Hbond$_{capac.}$) as defined by:

$$Hbond_{capac.} = \langle Hbond \rangle / \sum HBD, HBA \qquad (1)$$

Where $\langle Hbond \rangle$ is the average hydrogen bond number during the MD simulation, $\sum HBD, HBA$ is the sum of hydrogen bond donor (HBD) and acceptor (HBA) of the molecule.

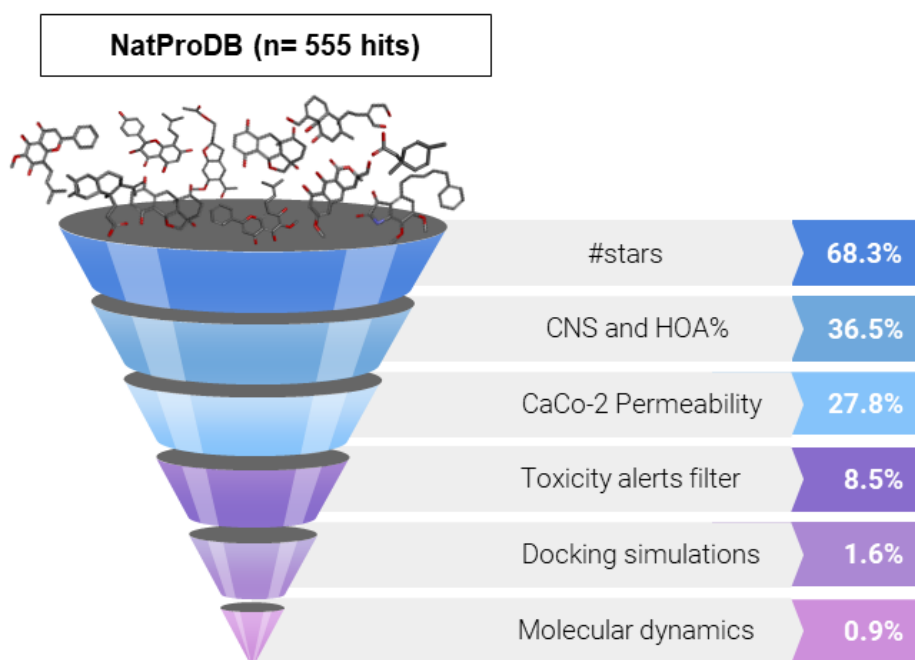All steps performed in this study are described in the Figure 1.

Figure 1: Virtual Screening steps applied in this study to identify new compounds from NatProdB with antiviral activity against SARS-CoV-2 Main protease (SARS-CoV-2 Mpro, PDB ID: 5R82). Pharmokinetic via **_Qikprop_** (#stars, CNS, HOA% and Caco-2 permeability); Toxicity alerts for chemical groups via **_DEREK_**; Docking simulations via **_GOLD_** program; and Molecular dynamics via **_GROMACS_**.

3. Results and Discussion:

In previous work, we analyzed the NatProDB compounds [76] interactions with a _Trypanosoma cruzi_ enzyme. Here we decided to analyze these data in further detail employing additional distinct ADMET _in silico_ methods: QikProp[TM] and Derek[TM] [77].

3.1 ADMET analysis by QikProp[TM]:

ADMET prediction was performed through the QikProp[TM] module. This tool predicts some relevant properties (e.g., water/gas log Ps, log S, log BB, overall CNS activity, Caco-2 and MDCK cell permeabilities; log Khsa for human serum albumin binding and log IC_{50} for HERG K^+-channel blockage) based on the full 3D molecular structure over twenty physical descriptors [78].

At first, we collected **555** molecules from the NatProDB and using the filter *#star* to calculate the number of properties that are outside the range for comparing a particular molecule's properties with those of 95% of known drugs, via QikProp[TM] [78-80]. Although it is acceptable to have at least 5 outside properties per hit, in this study, all hits with *#star* values ≥ 1 were eliminated following the **E64** (positive control), yielding **376** compounds.

Subsequently, we evaluated the predicted activity on the central nervous system (CNS) which ranges from -2 (inactive) to +2 (active) [78-80]. We define that only compounds with values < 0 were kept (**203** hits) since the antiviral action mechanism proposed in this study does not necessarily require an effect on the CNS.

Next, the human oral absorption [50] were predicted for remaining compounds using the values of a percentage ranging > 80% (high and desirable) to $\leq 25\%$ (low absorption) as standard criterion [78-80]. Aiming to search for good absorption potential, molecules with values $\geq 60\%$ were maintained since **E64** presented 14,41%, obtaining **201** candidates.

To assess the apparent permeability on Caco-2 cell membrane, Boehringer-Ingelheim scale (nm /s) was employed. Compounds with low permeability presented < 5 nm/s and high permeability showed values > 500 nm/s [78-80]. In this process, only molecules with values above 100 nm/s were considered and yielded in **153** molecules with major pharmacokinetics desirable properties for top-10 hits (Supplementary Table 1).

3.2 Toxicological analysis by Derek Nexus[TM]

For toxicity analysis predictions, we seek to identify and eliminate molecules that exhibited any toxicity alert for specific chemical groups, in the early stages of the process for developing prototype drugs, making it possible to optimize costs and time

during research. (carcinogenicity, mutagenicity, genotoxicity, skin sensitization, teratogenicity, nephrotoxicity, hepatotoxicity, irritation, respiratory sensitization and reproductive effects)[77].

Thus, the evaluation of chemical groups of the remaining **153** NatProDB compounds resulted in **47** candidates with low toxicological potential and adequate pharmacokinetic profiles for an oral route, *i. e.*: good intestinal absorption, high cellular permeability and low excretion. Therefore, promising antiviral candidate drugs.

The polar molecular surface (PSA) area is a property based on polar atoms surface that was shown to correlate with passive molecular transport through membranes, therefore, allows the prediction of transport properties of drugs [81-82]. Its desirable value ranges from 7.0 to 200.0 Å2 [78-80]. Molecule 6 had the highest PSA index of 107.35 $Å^2$, while compound 10 had the lowest value at 46.80 $Å^2$. This indicates good permeability of the selected molecules through the lipid membrane. It is noteworthy that a PSA value < 75 $Å^2$ has been associated with an increased risk of adverse effects when combined with high lipophilicity (logP greater than 4) and only 1 march that situation, which was excluded [81].

A drug bound to protein is inactive, and it is only the free drug that can bind with a receptor providing pharmacological action [83]. The free drug concentration is affected by binding on human serum albumin, therefore, this pharmacokinetic property was also predicted for the compounds [51, 84]. This was evaluated by its logarithmic constant and has recommended range from -1.5 to 1.5 [78-80].

Values outside the standard range may indicate non-enzymatic glycosylation of HSA, interference on drug binding potency and influence the distribution and excretion of molecules affecting the duration and intensity of the biologic effects [84-86]. All molecules selected showed satisfactory values and better than the **E64**, the smallest in the series, LogKHSA = -1.31 (Table S1).

Finally, the hERG encodes the inward rectifying voltage-gated potassium channel in the heart (IKr) which is associated with cardiac repolarization. Inhibition of the hERG current induces QT interval prolongation and could represent fatal ventricular arrhythmia called *Torsade de Pointes* [87]. In fact, recent studies pointed out that some medications such as hydroxychloroquine and azithromycin have facilitated the risk of fatal arrhythmias in patients with COVID-19 infection [88-90].

In our study, molecules that presented QPlogHERG values below -5 (**2** = -5.14; **4** = -5.53 and **5** = -6.48) will require greater attention and further *in vitro* analysis should be done to validate their pharmacological safety in the treatment against COVID-19.

3.3 Site prediction and Druggability analysis of SARS-CoV-2 Mpro:

We noted that all programs correctly identified the SARS-CoV-2 **Mpro** active site (Figure 2A-C) and additionally indicated that this site is druggable *e. g.*: druggability probability > 0.5 in PockDrug [35] and a strong primary hot spot with ≥ 16 probes clusters and hot spots were connected with a center-to-center distance < 8 Å in FTMap [92, 39]. These agree with all experimental SARS-CoV-2 **Mpro** structures [4, 7, 9, 93, 94] where distinct ligands were tested and crystallized.

After the protein preparation step, pockets predictions and druggability analysis of the SARS-CoV-2 **Mpro** were done using three different methods: Difference of Gaussian (DoG) [33, 91], PockDrug [35], FTMap [92]. the main pockets determined into this protein shown below (Figure 2)

Other sites were also signed as druggable in our analysis: SeeSAR (**II** and **III**-Figure 2A); PockDrug (green, yellow and grey sites- Figure 2B) and FTMap (yellow-Figure 2C). Since two SARS-CoV-2 **Mpro** sites are located on the same locations and were specified by distinct methods (**II**, yellow surface and yellow mesh; **III**, green surface and yellow mesh – Figure 2), we hypothesized that these two regions could be explored in further studies.
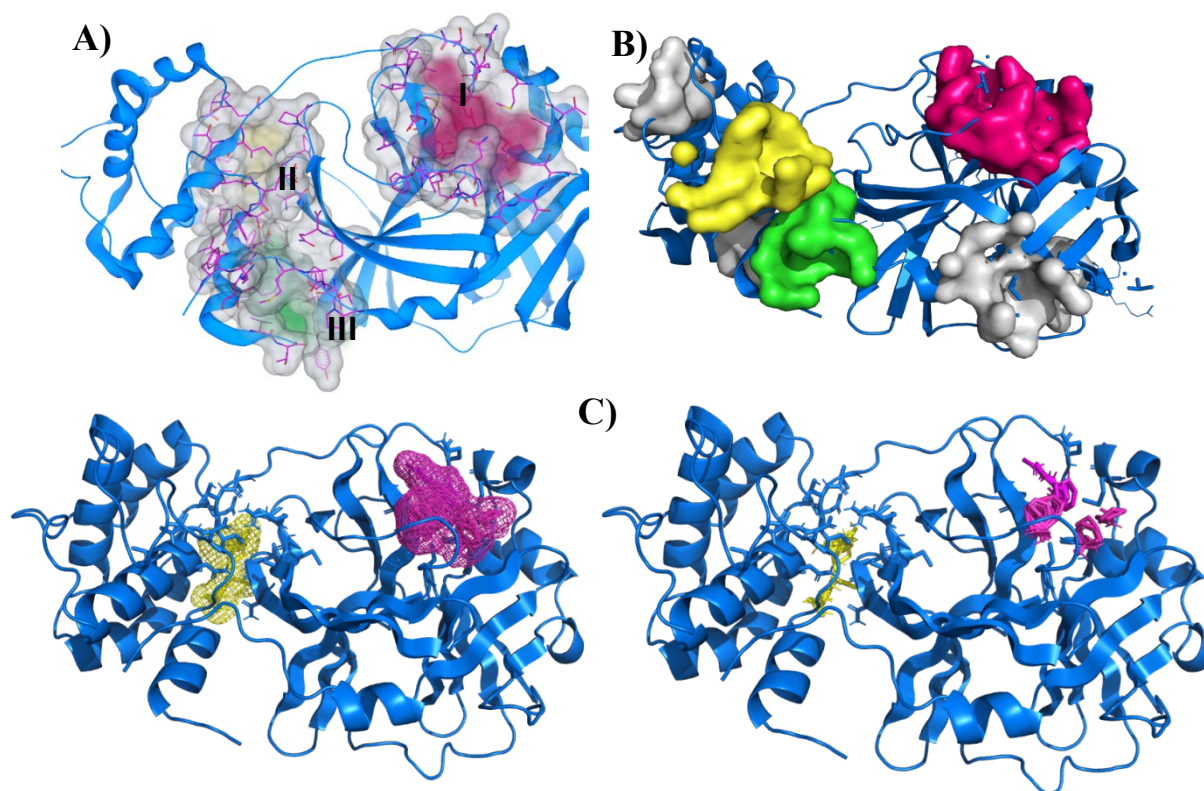
Figure 2: 3D structure of SARS-CoV-2 Main protease (SARS-CoV-2 **Mpro**, PDB ID: 5R82, blue cartoon). **A)** Binding sites (surface and pink sticks) found by SeeSAR version 9.2 (BioSolveIT GmbH) in the active site (**I**) is colored in pink (25 residues), site **II** (19 residues) in yellow and site **III** (15 residues) in green. This image was generated by SeeSAR program. **B)** Main pockets (color surfaces) predicted by PockDrug server [35]. The active site (**I**) is colored in pink (volume = 1,892 Å^3, druggability score = 0.63 ± 0.09), site **II** (volume = 643 Å^3, druggability score = 0.35 ± 0.08) in yellow and site **III** (volume = 760 Å^3, druggability score = 0.33 ± 0.03) in green. The remaining gray sites presented druggability score = 0.55 ± 0.10 (upper), 0.98 ± 0.01 (middle) and 0.61 ± 0.04 (lower). **C)** Main hot spots (mesh and sticks) calculated by FTMap server [92]. The active site is colored in pink and was classified as druggable (≥ 16 probes) [39]; the 2nd site (yellow) is borderline (13 ≤ probes < 16) and pink and yellow sticks show the location of the probe. The **B** and **C** images were generated by educational Pymol 2.4.1 [47].

The probes posed by FTMap reveals that the druggable site (magenta sticks, Figure 2C) accommodates polar acyclic compounds (ethanol, dimethylformamide and acetamide) and some rings (benzene, benzaldehyde and cyclohexane) and borderline site (yellow sticks, Figure 2C) lodges ethanol, methanamine, benzene and cyclohexane. This information is helpful for fragment-based drug design. In another further step, grey sites predicted by PockDrug (Figure 2B) which had higher druggability scores could also be explored.

3.4 Molecular Docking simulations

The docking calculations were done by SeeSAR version 9.2 (BioSolveIT GmbH) and focused on the active site of SARS-CoV-2 **Mpro** to better sampling the interaction energy (ΔG) between NatProDB ligands and this enzyme. The hydrogen bond and dehydration (HYDE) energy scoring function in SeeSAR provides a range of affinities, spanning an upper and lower limit.

We, therefore, used this descending range as an affinity parameter to compare distinct molecules and their interactions on Mpro active site. For the sake of clarity, we only show the 10 best docked compounds with higher estimated affinities (EA, µM) in Table 2 instead of 46 NatProDB and **E-64** molecules previously filtered from ADMET analysis (see preceding discussions).

SeeSAR version 9.2 (BioSolveIT GmbH) analyzes the estimated affinities (EA) energies and the best poses selection were based on HYDE scoring function [29, 30, 31] and torsional analysis [44]. The docked molecules at the SARS-CoV-2 **Mpro** active site were ordered by their EA on SARS-CoV-2 Mpro-NatProDB complexes and the "top 10" were selected for further analysis.

Comparing the docking results presented in Table 1, we noted that NatProDB compounds energy ranked on "top 10" were better ranked than the **Mpro** crystallographic ligand (RZS, 255th better energy). These results indicate a good interaction between NatProDB and SARS-CoV-2 Mpro residues, mainly on active site. Additionally, "top 10" NatProDB compounds presented better energy when compared to a positive control (**E-64**), a well-known cysteine protease inhibitor [98, 99] (562th better energy, Table 1).

The best-scored poses were selected by visual inspection and since they interacted better with **Mpro** active site residues, we decided to compare their interactions analyzing each complex (**Mpro**- RZS, E-64 and NatProDB) aiming to find some structural information that could help on drug discovery against COVID-19 (Figures S1-S4).

Table 1: Best docking results carried out by SeeSAR version 9.2 (BioSolveIT GmbH) on active site of SARS-CoV-2 Main protease (SARS-CoV-2 **Mpro**) using the Natural Products Database of Bahia Semi-Arid region (NatProDB), **E-64** (positive control) and **RZS** (crystallographic ligand, PDB ID: 5R82).

#	Compounds	Chemical Structure	Hydrogen Bond Donors (HBD)[1]	Hydrogen Bond Acceptors (HBA)	Lipinki's Rule[2] Violation	Molecular Surface Area (Å²)[3]	Solvent Accessible Surface (Å³)[4]	Estimated Affinity Range[5] (µM)	Inter clash type[6]
b01	VE0DIA0AF		1	2	1	134.92	478.2	3.29 – 326.4	+
b02	VE0PPA0AF		5	5	0	150.25	481.3	4.42 – 439.6	0

[1] The number of hydrogen bond donors (HBD), acceptors (HBA) and Lipinski's rule violation were determined from FAFF-Drugs server (https://fafdrugs4.rpbs.univ-paris-diderot.fr/) [28, 95, 96]

[2] See note above. Lipinski's rule was described here: [97]

[3] Molecular Surface Area were obtained from pkCSM server (http://biosig.unimelb.edu.au/pkcsm/) [24]

[4] The Solvent Accessible Surface (Å³) were calculated from MarvinSketch 18.8.0 Geometry plugin (http://www.chemaxon.com) employing the solvent radius 1.4Å and pH=7.4.

[5] Estimated Affinity (EA) ranges mark the lower and upper bound of the estimated affinities (±2 log-units).

[6] Inter clash type: atomic clashes between protein and docked ligand atoms predicted by SeeSAR9.2. Inter clash type are classified as + (no clashes), 0 (few clashes), - (many clashes).

b03	VE0ISA0AF	1	5	0	152.87	645.3	7.24 – 719.2	+
b04	VE0FKA0AF	2	5	0	150.34	576.2	17.25 – 1,714.3	+
b05	VE0FEA0SF	4	6	0	150.17	519.5	18.18 – 1,806.5	0
b06	VE0JDA0SF	3	5	0	121.72	355.9	20.82 – 2,068.7	0

ID	Compound								
b07	VE0ZDA0AF		1	5	0	151.74	652.2	41.3 – 4,103.0	+
b08	VE0NCA0SF		1	2	0	72.59	309.2	51.97 – 5,164.2	+
b09	VE0KJA0SI		3	8	0	148.67	599.8	52.40 – 5,206.5	0
b10	VE0NHA0SF		3	6	0	122.04	543.6	53.85 – 5,350.2	+
255 Control (RZS)			1	3	0	65.69	350.4	504.29 – 50,103.8	+
562 Positive control (E-64)			7	10	1	145.91	638.3	96,164.96 – 9,455,201.9	0

These hits occupied the active site in a varied manner and interacted with Mpro through hydrogen bonds with catalytic residues (His41 and Cys145) and made additional hydrophobic contacts as shown from docking results (Figures S1-S4). The main interactions were described by Protein-Ligand Interaction Profiler server (https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index) [100] at each pose (Table S2) where we noted that the 2 better docked NatProDB compounds (**VE0DIA0AF** and **VE0PPA0AF**) presented better contacts than the crystallographic ligand (**RZS**).

3.5 Molecular Dynamics (MD) Simulations

To understand the behavior of SARS-CoV-2 **Mpro** complexes, we analyzed them through Molecular Dynamics (MD) Simulations. These data allowed us to analyze the time influence on these interactions pattern. First, we analyze the stability of our eight systems: apo and complexes: holo (PDB ID: 5R82), **E-64** and the five better-ranked NatProDB compounds (Table S3). From our Root Mean Square Deviation (RMSD) data, we note that all systems equilibrated after 40ns (Figure S5 A) then we defined our productive phase into time intervals from 40-100ns for all simulations (Figure S5 B-C).

Comparing the Mpro structure fluctuations (3D RMSF) within all complexes we distinguished their overall protein folding stability, mainly at active site represented by catalytic residues (His41 and Cys145, Figure 3). These data correlate with Mpro crystallographic structure where active site residues are well solved in electron density maps from 2.1Å resolution [7]. Additionally, we analyzed their secondary structure stability by DSSP 3.1.4 module [101-103] installed on GROMACS 5.1.4 [53, 54, 56-59]. All Mpro complexes maintained their secondary structure stable during our simulation time (Figure S6).

Some regions presented higher fluctuation in Mpro structure, mostly at some loops and N-, C-terminals (Figure 3). SARS-CoV-2 Mpro crystal had a long loop region of domain II (residues 185 - 200) connecting domain III and this region is highly variable as pointed by superposition of 12 Mpros crystal structures [7]. Our MD data revealed that these loop on domain II fluctuated higher than other secondary protein structures corroborating with this experimental study. Previous results also cited the interface region among Mpro protomers composed by domain II residues as unstable [4, 7, 9] and these data were also visualized in our simulations (Figure 3).
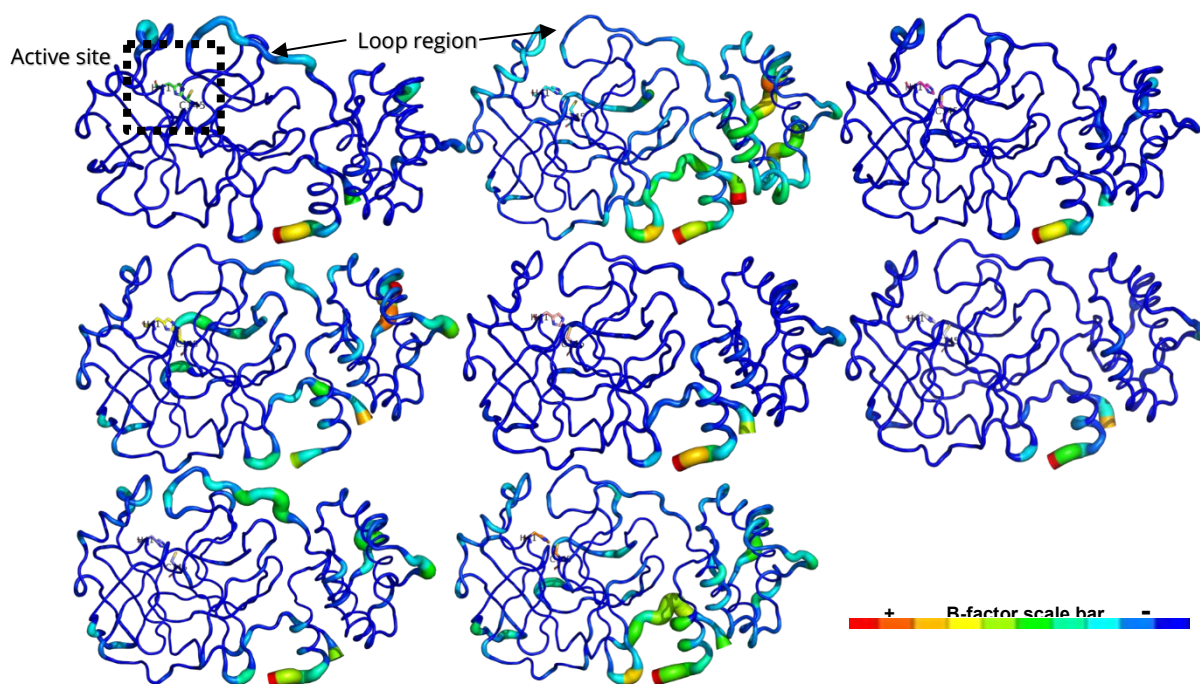
Figure 3: SARS-CoV-2 Mpro structures are shown in B-factor putty mode in PyMOL. The highest B-factor in each structure is colored in red and the lowest, in dark blue, as indicated by the B-factor scale bar. The thickness of the protein backbone also is proportional to the B-factors. The catalytic residues (His41 and Cys145) are displayed as sticks. This image was generated by educational Pymol 2.4.1 [43].

Many studies described Mpro subsites expanding from S6 to S1 with catalytic site located on S1 [4, 7, 9, 93]. Based on these results, we analyze the interaction pattern on SARS-CoV-2 Mpro of NatProDB derivatives compared with crystallographic ligand (**RZS**) and **E-64** . Additionally, we evaluated both the hydrogen bond pattern, through GROMACS *hbond* module [70] and HbMap2Grace program [71], and the molecular Surface Area by SurfinMD program [72].

The Molecular Dynamics (MD) data for hydrogen bond (H-bond) analysis shown that NatProDB compounds made more interactions than RZS and VE0FKA0AF and VE0FEA0SF were similar to **E-64**, which has the highest H-bond number (Table S3). Since these interaction patterns could favor their inhibitory behavior, we analyzed them in detail.

Initially, we defined and calculated the hydrogen bond capability for our ligands (Table S3), Hbond$_{capac.}$ during MD simulation. This measure could aid us to quantify how "tightly" the ligands interact with Mpro since Hbond$_{capac.} \geq 1$ means that each ligand's hydrogen bond atoms (donors and acceptors) interacted with protein residues.

Our results from MD productive phase correlates well with estimated binding energy obtained before from docking results (Table 1) *i.e.*, the ligands with best docking energies presented higher Hbond$_{capac.}$ which reveal to us that hydrogen bond is a favorable interaction on developing Mpro inhibitors. Additionally, we noted that our ligand accessed distinct sub-sites on SARS-CoV-2 Mpro interacting beyond the catalytic ones - S1 (Figure 4).

The H-bond pattern presented interactions with His41 and Cys145 (catalytic residues); Thr25 and 26 from S1'; Gly143 and Ser144 from canonical oxyanion role on S1 [7]; Glu166 from S1 and Gln189 and Thr190 from S4 [4, 7, 9, 94]. Since H-bond is considered as the "driving force" for Mpro inhibition, we could map and analyze them through our MD simulations.
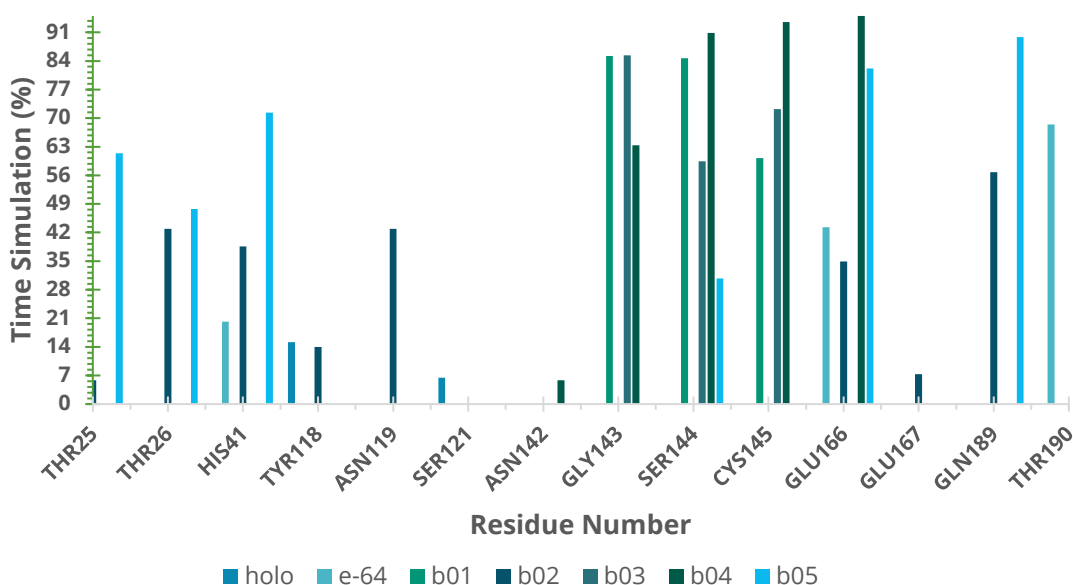

Figure 4: Hydrogen bond stability on SARS-CoV-2 Mpro complexes calculated by HbMap2Grace after 100ns of Molecular Dynamics simulation.

We also calculated the atomic contacts involving SARS-CoV-2 Mpro and NatProDB compounds (Figure 5). The contact surface area revealed additional interactions with apolar residues on the same sites described before [4, 7, 9]: S1' (Thr25, 26, Leu 27, Cys145), S2 (His41, Thr45, Ser46, Asp48, Met49), S4 (Met165, Glu166, Leu 167, Gln189) and S5 (Pro168).
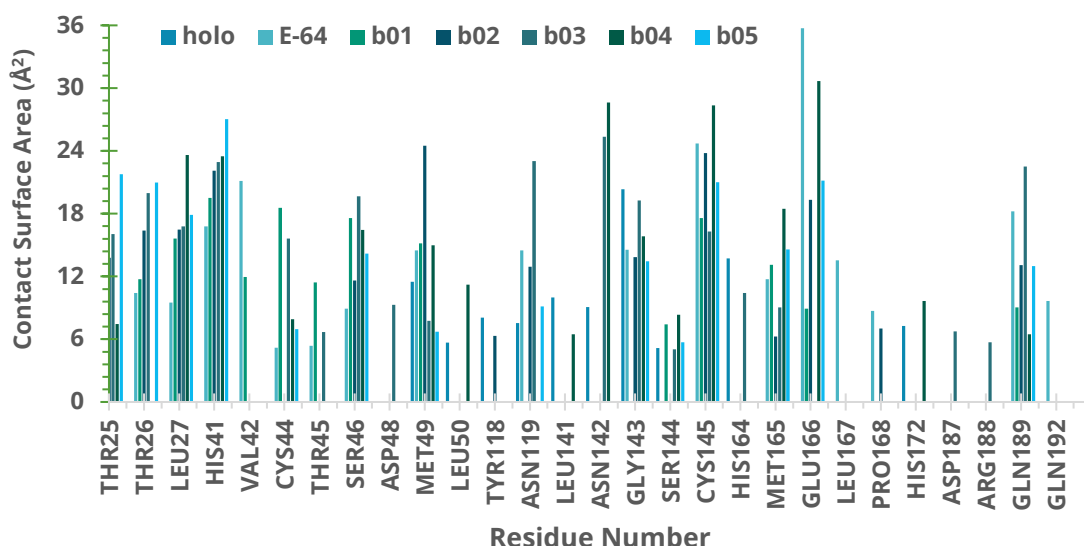
Figure 5: Surface Molecular Area (Å^2) of SARS-CoV-2 Mpro complexes calculated by SurfinMD after 100ns of Molecular Dynamics simulation.

Another measure to evaluate Mpro interaction with NatProDB ligands is their active site occupancy. Ideally, it is expected that potent inhibitors should present reversible action mechanisms and tight binding [104]. Since Mpro active site has a well-defined area (335.9 Å^2) and volume (364.1 Å^3) [105], we compare these results with our MD data. Fpocket program [106, 107] was employed to calculate the volume variation throughout the simulation. We noted that active site volume oscillates from 461.8 Å^3 of E-64 complexes to 262.9 Å^3 for VE0FKA0AF (Table S3) which agree well with previous data [4, 93, 105] and with the molecular area and volume of NatProDB ligands (Table 1). These results could point to us that SARS-CoV-2 Mpro and NatProDB ligands shown induced-fit mechanisms correlating with their dynamics and molecular interactions at the active site.

Based on interaction results of complexed after MD simulations, we assume that new Mpro sites were pointed: **i)** V42, C44, T45 and S46, near and ii) Y118 and N119 (Table S3). The former residues made hydrophobic interactions with alpha-ketoamide inhibitors described before [4], FDA approved drugs and medicinal plant compounds [105] and ZINC-15 docked compounds [108]. Our results (Figure 6) reinforce the importance of these non-polar interactions and opens up a new region (together with latter residues) for drug discovery development aiming to search for new Mpro inhibitors.
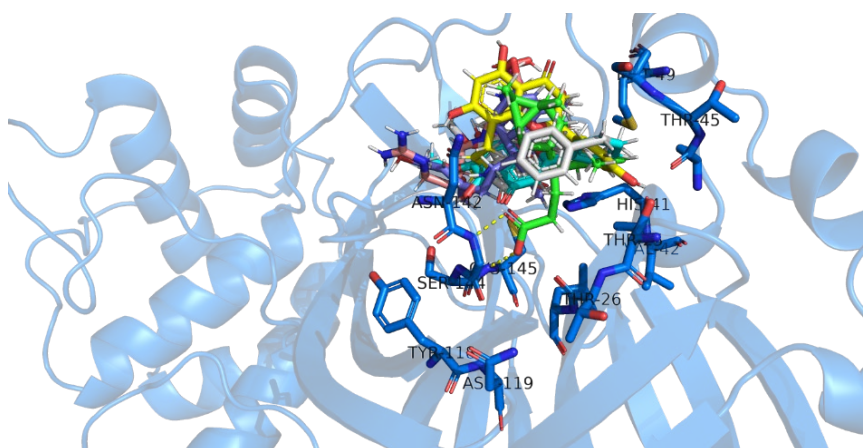
Figure 6: SARS-CoV-2 Mpro main residues that interact with best docked NatProDB compounds throughout the productive phase of Molecular Dynamics simulation. Mpro is shown as a cartoon (blue) with its main interacting residues labeled (blue sticks), **VE0DIA0AF** is shown in green sticks, **VE0PPA0AF** (cyan), **RZS** (magenta) and **E-64** (orange) and polar interactions is depicted as yellow dashed lines. This image was generated by educational Pymol 2.4.1 [43].

Additionally, we calculated the binding free energy of all Mpro complexes through MM-PBSA methods. The binding energy ($\Delta E_{binding}$) calculated by solvent accessible surface area shows that all compounds interacted favorable with Mpro (Table S3). Since these values are directly correlated to interacting protein residues, we decide to discriminate how amino acids presented better contacts with ligands. These residue decomposition energy analyses selected residues near the ligand (<5 Å) during the MD simulation and which participate actively in complex stabilization ($\Delta E_{binding} > \pm 5$ kJ/mol) as shown in Figure 7.
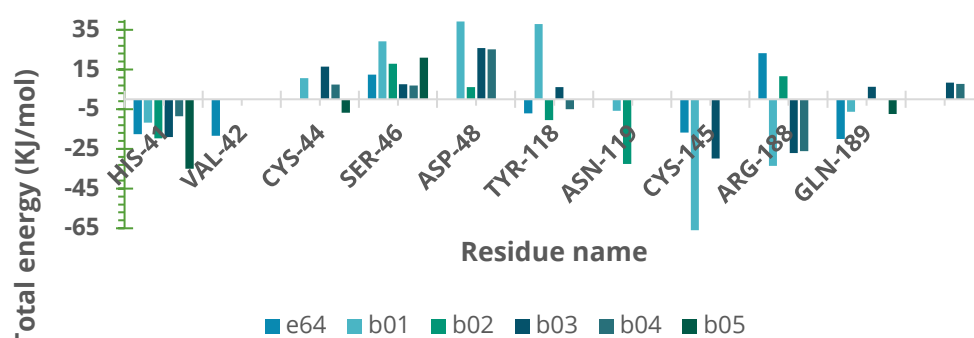

Figure 7: Residue contributions to the binding energy of SARS-CoV-2 Mpro complexes. The main residues with energy interaction ($\Delta E_{binding} > \pm 5$ kJ/mol) were highlighted. For Mpro sub-site definitions see the text.

We noted that ligand interaction with catalytic residues (His41 and Cys145) are highly favorable (negative values), as expected for reversible inhibitors. This behavior remains for two residues of the new binding site (Arg188 and Gln189), first described in this work. Other residues (Val42, Cys44, Ser46 and Glu48) do not interact favorably with ligands (positive values- Figure 6). This could be related to their position on Mpro site, *i.e.*, their side chains are pointed outwards instead of the active site cavity [4, 9].

4. Conclusion

In this study, some computer-aided drug discovery methods were sequentially employed against the main protease Mpro of SARS-CoV-2 (SARS-CoV-2 Mpro) aiming to select the most active molecules from the Natural Products Database of the Bahia Semi-Arid region (NatProDB, http://natprodb.uefs.br/). From "top 10" docked compounds some common scaffolds were selected (secotrachylobanoic acid, hexahydronaphthalen-1-yl, indolinedione, benzopyran-4-one) indicating that Mpro could privilege the hydrophobic chains and (poly) aromatic rings. The best interactions with SARS-CoV-2 Mpro reveals that some enzyme sites were accessed thereby confirming that this method can be employed as a suitable starting method for the identification of novel SARS-CoV-2 Mpro inhibitors.

The pharmacokinetic (ADMET) and toxicological analysis confirm that 9 of 10 best molecules does not have any issues and could be employed in subsequent *in vitro* assays. Finally, this study should point that sequential application of *in silico* methods (ADMET analysis, Docking and Molecular Dynamics) are valuable tools for searching the chemical space and selecting the best ligand structures that could be employed in experimental tests minimizing the costs of testing many compounds. We also hope that natural products from Bahia Semi-arid region choosen here could be safe and effective for treating SARS-CoV-2 infection.

REFERENCES:

1. Pillaiyar, T. et al. 2020. Recent discovery and development of inhibitors targeting coronaviruses. Drug Discovery Today. 25, 4 (2020),. DOI: 10.1016/j.drudis.2020.01.015

2. Guan, W. J. et al. 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. The new england journal o f medicine, 382, 18, (2020), 1708-1720. DOI: 10.1056/NEJMoa2002032

3. Wu, F. et al. 2020. A new coronavirus associated with human respiratory disease in China. Nature. 579, 7798 (2020) 265-269. DOI: 10.1038/s41586-020-2008-3

4. Zhang, L. et al. 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. (2020), eabb3405. DOI: 10.1126/science.abb3405

5. Zhou, P. et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 579, 7798 (2020), 270-273. DOI: 10.1038/s41586-020-2012-7

6. Zhu, N. et al. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019.The New England Journal of Medicine. 382, 8 (2020), 727-733. DOI: 10.1056/NEJMoa2001017

7. Jin, Z. et al. 2020. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. Nature, 582, (2020), 289-293. DOI: 10.1038/s41586-020-2223-y

8. Rosa, S. G. V. et al. 2020. Clinical trials on drug repositioning for COVID-19 treatment. Rev Panam Salud Publica. Brief communications. 44, n. march. DOI: 10.26633/RPSP.2020.40

9. Anand, K. et al. 2003. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. Science, 300, 5626, (2003), 1763-1767. DOI: 10.1126/science.1085658

10. Ziebuhr, J. Molecular biology of severe acute respiratory syndrome coronavirus. Current Opinion in Microbiology. 7, 4 (2004), 412. DOI: 10.1016/j.mib.2004.06.007

11. Ziebuhr, J. et al. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. Journal of General Virology. 81, Pt 4 (2000), 853-879. DOI: 10.1099/0022-1317-81-4-853

12. Thomford, N. E. et al. 2018. Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery. International Journal of Molecular Science, 19, 6 (2018), 1578. DOI: 10.3390/ijms19061578

13. Wink, M. 2012. Medicinal plants: a source of anti-parasitic secondary metabolites. Molecules. 17, 11 (2012), 12771-12791. DOI: 10.3390/molecules171112771

14. Newman, D. J. and Cragg, G. M. 2012. Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of Natural Products. 75, 3 (2012), 311-335. DOI: 10.1021/np200906s

15. Rodrigues, T. et al. 2016. Counting on natural products for drug design. Nature Chemistry. 8, 6 (2016), 531-541, 06 2016. DOI: 10.1038/nchem.2479

16. Fakhar, Z. et al. 2020. Anthocyanin derivatives as potent inhibitors of SARS-CoV-2 main protease: An in-silico perspective of therapeutic targets against COVID-19 pandemic. Journal of Biomolecular Structure and Dynamics. (2020), 1-13. DOI: 10.1080/07391102.2020.1801510

17. Mattio, L. M. et al. 2020. Natural and nature-inspired stilbenoids as antiviral agents. European Journal of Medicinal Chemistry. 202 (2020), 112541. DOI: 10.1016/j.ejmech.2020.112541

18. Newman, D. J. and Cragg, G. M. 2019. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. Journal of Natural Products. 83, 3 (2020) 770-803. DOI: 10.1021/acs.jnatprod.9b01285

19. Olubiyi, O. O. et al. 2020. High Throughput Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-CoV-2. Molecules.25, 14 (2020), 3193. doi: 10.3390/molecules25143193

20. Sillapachaiyaporn, C. and Chuchawankul, S. 2020. HIV-1 protease and reverse transcriptase inhibition by tiger milk mushroom Journal of Traditional and Complementary Medicine. 10, 4 (2020), 396-404. DOI: 10.1016/j.jtcme.2019.08.002

21. Wang, Z. and Yang, L. 2020. Turning the Tide: Natural Products and Natural-Product-Inspired Chemicals as Potential Counters to SARS-CoV-2 Infection. Frontiers in Pharmacology, 11 (2020), 1013. DOI: 10.3389/fphar.2020.01013

22. Kury, A. B. A. et al. 2006. Centro de Gestão e Estudos Estratégicos. Tecnologia, M. D. C. E.**:** 324, (2006).

23. Lucchese, A. M. and Vale, A. E. 2006. Plantas da caatinga: perfil botânico, fitoquímica e atividade biologica. (2006), 297. available at: http://natprodb.uefs.br/

24. Pires, D. E. V.; et al. 2015. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. Journal of Medicinal Chemistry. 58, 9 (2015), 4066-4072. DOI: 10.1021/acs.jmedchem.5b00104

25. Braga, R. C. et al. 2014. Tuning HERG out: antitarget QSAR models for drug development. Current Topics in Medicinal Chemistry, 14, 11, (2014), 1399-1415. DOI: 10.2174/1568026614666140506124442

26. Braga, R. C. et al. 2015. Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity. Molecular Informatics, 34, 10, (2015), 698-701. DOI: 10.1002/minf.201500040

27. Irwin, J. J. et al. 2015. An Aggregation Advisor for Ligand Discovery. Journal of Medicinal Chemistry, 58, 17, (2015), 7076-7087. DOI: 10.1021/acs.jmedchem.5b01105

28. Lagorce, D.; et al. 2008. FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC Bioinformatics, 9, 396, (2008). DOI: 10.1186/1471-2105-9-396

29. Reulecke, I.; et al. 2008. Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function. ChemMedChem. 3, 6 (2008), 885-897. DOI: 10.1002/cmdc.200700319.

30. Schneider, N. et al. 2012. Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. Journal of Computer-Aided Molecular Design. 26, 6 (2012), 701-723. DOI: 10.1007/s10822-011-9531-0

31. Schneider, N. et al. 2013. A consistent description of HYdrogen bond and DEhydration energies in protein-ligand complexes: methods behind the HYDE scoring function. Journal of Computer-Aided Molecular Design. 27, 1 (2013), 15-29. DOI: 10.1007/s10822-012-9626-2

32. Douangamath, A. et al. 2020. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. Nature Communications. 11, 1 (2020) 5047. DOI: 10.1038/s41467-020-18709-w

33. Volkamer, A. et al. 2010. Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. Journal of Chemical Information and Modeling. 50, 11 (2010), 2041-2052. DOI: 10.1021/ci100241y

34. Volkamer, A. et al. 2012. Combining Global and Local Measures for Structure-Based Druggability Predictions. Journal of Chemical Information and Modeling. 52, 2 (2012), 360-372. DOI: 10.1021/ci200454v

35. Hussein, H. A. et al. 2015. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Research, 43, W1, (2015), W436-442. DOI: 10.1093/nar/gkv462

36. Borrel, A. et al. 2015. PockDrug: A Model for Predicting Pocket Druggability That Overcomes Pocket Estimation Uncertainties. Journal of Chemical Information and Modeling, 55, 4, (2015), 882-895. DOI: 10.1021/ci5006004

37. Kyte, J.et al. 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biological, 157, 1, (1982), 105-132. DOI: 10.1016/0022-2836(82)90515-0

38. Milletti, F. and Vulpetti, A. 2010. Predicting polypharmacology by binding site similarity: from kinases to the protein universe. Journal of Chemical Information and Modeling. 50, 8 (2010), 1418-1431. DOI: 10.1021/ci1001263

39. Kozakov, D. et al. 2015. New Frontiers in Druggability. Journal Medicinal Chemistry, 58, 23, (2015), 9063-9088. DOI: 10.1021/acs.jmedchem.5b00586

40. Ngan, C. H. et al. 2012. FTMAP: extended protein mapping with user-selected probe molecules. Nucleic Acids Research. 40, Web Server issue (2012), W271-275. DOI: 10.1093/nar/gks441

41. Brenke, R. et al. 2009. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. Bioinformatics, 25, 5, (2009), 621-627. DOI: 10.1093/bioinformatics/btp036

42. Kozakov, D. et al. 2011. Structural conservation of druggable hot spots in protein-protein interfaces. Proc Natl Acad Sci U S A, 108, 33, (2011), 13528-13533. DOI: 10.1073/pnas.1101835108

43. Vajda, S. et al. 2018. Cryptic binding sites on proteins: definition, detection, and druggability. Current Opinion in Chemical Biology. 44, 06 (2018), 1-8. DOI: 10.1016/j.cbpa.2018.05.003

44. Schärfer, C.; et al. 2013. Torsion angle preferences in druglike chemical space: a comprehensive guide. Journal of Medicinal Chemistry. 56, 5 (2013), 2016-2028. DOI: 10.1021/jm3016816

45. Brethon, A. et al. 2017. New Caspase-1 inhibitor by scaffold hopping into bio-inspired 3D-fragment space. Bioorganic & Medicinal Chemistry Letters, 27, 24, (2017), 5373-5377. DOI: 10.1016/j.bmcl.2017.11.015

46. Hopkins, A. L. et al. 2004. Ligand efficiency: a useful metric for lead selection. Drug Discovery Today, 9, 10, (2004), 430-431. DOI: 10.1016/S1359-6446(04)03069-7

47. Schrödinger, L. The PyMOL Molecular Graphics System. Version 2.1.0. (2013).

48. Schrödinger. QikProp: Rapid ADME predictions of drug candidates. Versão 2017-2. 2017 ed. New York, NY: Schrödinger, LLC, 2017.

49. Hou, T. and Wang J. 2008. Structure-ADME relationship: still a long way to go? Expert Opinion on Drug Metabolism & Toxicolog. 4, 6 (2008) 759-70. DOI: 10.1517/17425255.4.6.759

50. Sutter, A. et al. 2013. Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regulatory Toxicology and Pharmacology. 67, 1 (2013), 39-52. DOI: 10.1016/j.yrtph.2013.05.001

51. Braggio, S. et al. 2014. Discovery Toxicology In Lead Optimisation. (The Handbook of Medicinal Chemistry: Principles and Practice (2014). DOI: 10.1039/9781782621836-00364

52. Greene, N. et al 1999. Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR and QSAR in Environmental Research, 10, 2-3, (1999), 299-314. DOI: 10.1080/10629369908039182

53. Abraham, M. J. et al. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1-2, (2015) 19-25. DOI: 10.1016/j.softx.2015.06.001

54. Berendsen, H. J. C.; et al. 1995. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. Computer Physics Communications, 91, 1-3, (1995) 43-56. DOI: 10.1016/0010-4655(95)00042-E

55. Hess, B. et al. 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. Journal of Chemical Theory and Computational, 4, 3, (2008), 435–447. DOI: 10.1021/ct700301q

56. Lindahl, E.; et al. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. Molecular modeling annual, 7, 8, (2001) 306-317. DOI: 10.1007/s008940100045

57. Pronk, S. et al. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 29, 7 (2013), 845-854. DOI: 10.1093/bioinformatics/btt055

58. Van der spoel, D. et al. 2005. GROMACS: fast, flexible, and free. Journal of Computational Chemistry. 26, 16 (2005), 1701-1718. DOI: 10.1002/jcc.20291.

59. Bekker, H., et al. 1993. GROMACS - A parallel computer for molecular-dynamics simulations. In RA. DeGroot, & J. Nadrchal (Eds.), Physics computing '92. (1993), 252-256. World Scientific Publishing.

60. Schmid, N. et al. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. European Biophysics Journal. 40, 7 (2011), 843-856, Jul 2011. DOI: 10.1007/s00249-011-0700-9

61. Darden, T. et al. 1993. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics. 98, 12, (1993) 10089-10092. DOI: 10.1063/1.464397

62. Berendsen, H. J. C. et al. 1987. The missing term in effective pair potentials. The Journal of Physical Chemistry, 91, 24, (1987), 6269-6271. DOI: 10.1021/j100308a038

63. Bussi, G. et al. 2007. Canonical sampling through velocity rescaling.The Journal of Chemistr Physics, 126, 1, (2007), 014101. DOI: 10.1063/1.2408420

64. Berendsen, H. J. C. et al. 1984. Molecular-Dynamics with Coupling to an External Bath. Journal of Chemical Physics, 81, 8, (1984) 3684-3690. DOI: 10.1063/1.448118

65. Miyamoto, S. and Kollman, P. A. Settle: 1992. An analytical version of the SHAKE and RATTLE algorithm for rigid water models. Journal of Computational Chemistry. 13, 8 (1992), 952-962. DOI: 10.1002/jcc.540130805

66. Berk, H. et al. 1997. LINCS: A linear constraint solver for molecular simulations. 18, 12 (1997), 1463-1472. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

67. Koziara, K. B. et al. 2014. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. Journal of

Computer-Aided Molecular Design, 28, 3, (2014), 221-233. DOI: 10.1007/s10822-014-9713-7

68. Malde, A. K. et al. 2011. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. Journal of Chemical Theory Computational, 7, 12, (2011), 4026-4037, Dec 2011. DOI: 10.1021/ct200196m

69. Stroet, M. et al. 2018. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. Journal of Chemistry Theory and Computational. 14, 11 (2018), 5834-5845. DOI: 10.1021/acs.jctc.8b00768

70. Van der spoel, D. et al. 2006.Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. The Journal of Physcal Chemistry B. 110, 9 (2006), 4393-4398. DOI: 10.1021/jp0572535

71. Gomes, D. E. B. et al. 2002. HbMap2Grace. Versão 1.0. (2002).

72. Gomes, D. E. B. et al. 2002. SurfinMD. (2002).

73. El Tayar, N. et al. 1991. Partitioning of solutes in different solvent systems: the contribution of hydrogen-bonding capacity and polarity. Journal of Pharmaceutical Sciences. 80, 6 (1991), 590-598. DOI: 10.1002/jps.2600800619

74. Seiler, P. 1974. Interconversion of lipophilicities from hydrocarbon/water systems into the octanol/water system.: European Journal of Medicinal Chemistry. 9, 5 (1974) 473-479.

75. Wang, T. et al. 2020. Role of hydrogen bond capacity of solvents in reactions of amines with CO. Journal of Environmental Science. 91, (2020), 271-278. DOI: 10.1016/j.jes.2020.01.019

76. Da paixão, V. G. and Pita, S. S. D. R. In silico identification and evaluation of new Trypanosoma cruzi trypanothione reductase (TcTR) inhibitors obtained from natural products database of the Bahia semi-arid region (NatProDB). Computational Biology and Chemistry, 79 (2019), 36-47. DOI: 10.1016/j.compbiolchem.2019.01.009

77. Macmillan et al. 2019. A defined approach for predicting skin sensitisation hazard and potency based on the guided integration of in silico, in chemico and

in vitro data using exclusion criteria, Regulatory Toxicology and Pharmacology. 101, (2019), 35-47. DOI: 10.1016/j.yrtph.2018.11.001

78. Silva, R. et al. 2019. Ligand- and structure-based virtual screening from 16-((diisobutylamino)methyl)-6α-hydroxyivouacapane-7β,17β-lactone a compound with potential anti-prostate cancer activity. Journal of the Serbian Chemical Society, 84, 2, (2019) 153-174. DOI: 10.2298/JSC180129047S

79. Duffy, E. M. and Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. Journal of the American Chemical Society. 122, 12 (2000), 2878-2888. DOI: 10.1021/ja993663t

80. Jorgensen, W. L. et al. 2000. Prediction of drug solubility from Monte Carlo simulations. Bioorganic & Medicinal Chemistry Letters, 10, 11, (2000), 1155-1158. DOI: 10.1016/S0960-894X(00)00172-4

81. Ertl, P. et al. 2000. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. Journal of Medicinal Chemistry. 43, 20 (2000), 3714-3717. DOI: 10.1021/jm000942e

82. Palm, K. et al. 1997. Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharmaceuticals Reserach. 14, 5 (1997) 568-571. DOI: 10.1023/A:1012188625088

83. Dasgupta, A. Chapter 4 - Monitoring Free Drug Concentration: Clinical Usefulness and Analytical Challenges. *In*: CLARKE, W. e DASGUPTA, A. (Ed.). Clinical Challenges in Therapeutic Drug Monitoring. San Diego: Elsevier, (2016), 71-100. DOI: 10.1016/B978-0-12-802025-8.00011-8

84. Hanai, T. et al. 2002. Prediction of human serum albumin–drug binding affinity without albumin, Analytica Chimica Acta, 454, 1, (2002), 101-108, 2002. DOI: 10.1016/S0003-2670(01)01515-X

85. Noctor, T. A. et al. 1992. Allosteric and competitive displacement of drugs from human serum albumin by octanoic acid, as revealed by high-performance liquid affinity chromatography, on a human serum albumin-based stationary phase.

Journal of Chromatography B: Biomedical Sciences and Applications. 577, 2 (1992), 305-15. DOI: 10.1016/0378-4347(92)80252-L

86. Urien, S. et al. 1993. Vinorelbine high-affinity binding to human platelets and lymphocytes: distribution in human blood. Cancer Chemotherapy and Pharmacology. 32, 3 (1993), 231-234, 1993. DOI: 10.1007/BF00685841

87. Haverkamp, W.et al. 2000. The potential for QT prolongation and proarrhythmia by non-antiarrhythmic drugs: clinical and regulatory implications. Report on a policy conference of the European Society of Cardiology. European Heart Journal, 21, 15, (2000), 1216-31. DOI: 10.1053/euhj.2000.2249

88. Gautret, P. et al. 2020. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. International Journal of Antimicrobial Agents, 56, 1, (2020). DOI: 10.1016/j.ijantimicag.2020.105949

89. Malviya, A. 2020. Ventricular arrhythmia risk due to chloroquine / hydroxychloroquine treatment for COVID-19: Should it be given. Indian Heart Journal. 72, 2 (2020), 131-132. DOI: 10.1016/j.ihj.2020.04.006

90. Saleh, M. et al 2020. Effect of chloroquine, hydroxychloroquine, and azithromycin on the corrected QT interval in patients with SARS-CoV-2 infection. Circulation: Arrhythmia and Electrophysiology. 13, 6 (2020), e008662. DOI: 10.1161/CIRCEP.120.008662

91. Volkamer, A. et al. 2012. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics. 28, 15 (2012), 2074-2075. DOI: 10.1093/bioinformatics/bts310

92. Kozakov, D. et al .2015. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc, 10, 5, (2015) 733-755. DOI: 10.1038/nprot.2015.043

93. Tan, J. et al. 2005. pH-dependent conformational flexibility of the SARS-CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. Journal of Molecular Biology. 354, 1 (2005), 25-40. DOI: 10.1016/j.jmb.2005.09.012

94. Yang, H. et al. 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS Biol. 3, 10 (2005), e324. DOI: 10.1371/journal.pbio.0030324

95. Lagorce, D. et al. 2017. Pan-assay interference compounds (PAINS) that may not be too painful for chemical biology projects. Drug Discovery Today, 22, 8, (2017), 1131-1133. DOI: 10.1016/j.drudis.2017.05.017

96. Lagorce, D. et al. 2015. FAF-Drugs3: a web server for compound property calculation and chemical library design. Nucleic acids research, 43, W1, (2015), W200-W207. DOI: 10.1093/nar/gkv353

97. Lipinski, C. A. et al. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 23, 1 (2001) 3-25. DOI: 10.1016/s0169-409x(00)00129-0

98. Barrett, A. J. et al. 1982. L-trans-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem Journal, 201, 1, (1982), 189-198. DOI: 10.1042/bj2010189

99. Matsumoto, K. et al. 1999. Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. Biopolymers, 51, 1, (1999), 99-107. DOI: 10.1002/(SICI)1097-0282(1999)51:1<99::AID-BIP11>3.0.CO;2-R

100. Salentin, S. et al. 2015. PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Research. 43, W1 (2015), W443-447. DOI: 10.1093/nar/gkv315

101. Kabsch, W. et al. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 22, 12, (1983), 2577-2637. DOI: 10.1002/bip.360221211

102. Sander, C. et al. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins. 9, 1 (1991), 56-68. DOI: 10.1002/prot.340090107

103.   Touw, W. G. et al. 2015. A series of PDB-related databanks for everyday needs. Nucleic Acids Research. 43, Database issue (2015), D364-368. DOI: 10.1093/nar/gku1028

104.   Patil, R. et al. 2010. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. PLoS One, 5, 8, (2010), e12029. DOI: 10.1371/journal.pone.0012029

105.   Sharma, P. et al. 2020. Identification of potential drug candidates to combat COVID-19: a structural study using the main protease (mpro) of SARS-CoV-2. Journal of Biomolecular Structure and Dynamics. (2020), 1-11. DOI: 10.1080/07391102.2020.1798286

106.   Le guilloux, V. et al. 2009. Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics, 10, 168, (2009). DOI: 10.1186/1471-2105-10-168

107.   Schmidtke, P. et al. 2010. fpocket: online tools for protein ensemble pocket detection and tracking. Nucleic Acids Research. 38, Web Server issue (2010) W582-589. DOI: 10.1093/nar/gkq383

108.   Ton, A. T. et al. 2020. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. Molecular Informatics. 39, 8 (2020), e2000028. DOI: 10.1002/minf.202000028

109.   Baker, N. A. et al. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences. 98, 18 (2001), 10037-10041. DOI: 10.1073/pnas.181342398

110.   Kumari R. and Kumar R. 2014. Open Source Drug Discovery Consortium, Lynn A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 28, 54(2014), 1951-62. DOI: 10.1021/ci500020m