Identification of putative plant based antiviral compounds to fight against SARS-CoV-2 infection

Sujit Kumar Mishra[1], Rajguru Rajesh Raman Mishra[2], Saumya Dash[3], Jogeswar Panigrahi[4*]

[1]Laboratory of Natural products and Therapeutics, School of Applied Sciences, Centurion University of Technology and Management, Odisha, India

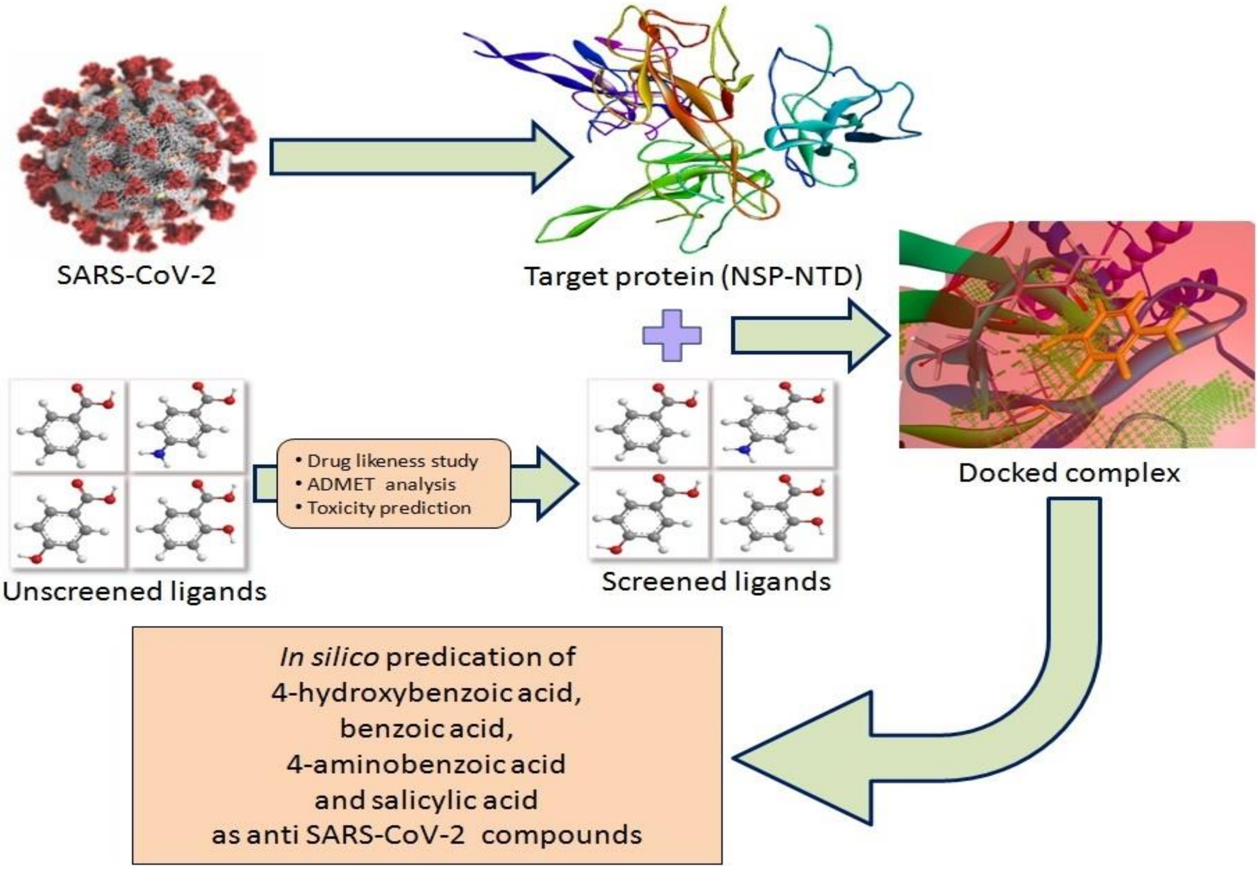
[2]Department of Chemistry, School of Applied Sciences, Centurion University of Technology and Management, Odisha, India

[3]Stem Cell Facility, All India Institute of Medical Sciences, New Delhi, India

[4]Department of Biosciences and Bioinformatics, Khallikote University, Berhampur, Odisha, India

*Corresponding author, E-mail: drjpanigrahi@gmail.com

Graphical abstract


SARS-CoV-2
Target protein (NSP-NTD)
Docked complex
Unscreened ligands
• Drug likeness study
• ADMET analysis
• Toxicity prediction
Screened ligands
In silico predication of
4-hydroxybenzoic acid,
benzoic acid,
4-aminobenzoic acid
and salicylic acid
as anti SARS-CoV-2 compounds

Highlights

- *In silico* analysis identified 14 natural compounds and their derivatives as druglikeness compounds.
- ADMET prediction study screened 10 compounds satisfying all the pharmacokinetics properties.
- Toxicity prediction study identified most of these compounds as non-degradable, non-mutagenic, and exhibiting mild to moderate occularirritancy and skin irritability, and low to moderate carcinogenicity.
- Molecular docking study identified four compounds (4-hydroxybenzoic acid, benzoic acid, 4-aminobenzoic acid and salicylic acid) as potential ligands inhibiting the target protein (NSP-NTD) of SARS-CoV-2.

Abstract

Background: This study aimed to examine the efficacy of some natural compounds and their derivatives in inhibiting the nucleocapsid protein N-terminal RNA binding domain (NSP-NTD), of SARS-CoV-2 virus by using the molecular doacking approach.

Methods: Physiochemical and drug likeness properties of the compounds were characterized by using SWISS ADME server tool. ADMET and TOPKAT modules of Discovery studio 4.0 were used for prediction of pharmacokinetic properties and toxicity of the compounds. Molecular docking of the ligands with the target protein (NSP-NTD) was carried out using the receptor-ligand interactions module of DS 4.0. The CDOCKER energy, CDOCKER interaction energy and binding energy of the interactions were calculated to identify the best interacting compounds.

Results: Four compounds including 4-hydroxybenzoic acid, benzoic acid, 4-aminobenzoic acid and salicylic acid have been predicted as effective compounds to inhibit the NSP-NTD (responsible for packing the viral RNA into the crown like capsid) *vis-à-vis* combat the SARS-Cov-2 virus infection.

Conclusions: In vitro and *in vivo* evaluation of these compounds against SARS-CoV-2 virus is required prior to assuring their potential roles in SARS-CoV-2 infection control.

Keywords: COVID-19, SARS-CoV-2, ADMET prediction, Molecular docking, Drug discovery

1 INTRODUCTION

The ongoing COVID-19 pandemic is a serious respiratory disease caused by the SARS- CoV-2 virus, belongs to the family *Coronaviridae* (Enjuanes et al., 2006). This family of virus primarily causes enzootic infections in animals like birds and mammals. However, during the last few decades, it has infected human and hence establish zoonotic infection (Schoeman and Fielding, 2019). Some viral strains of this virus family leads to severe acute respiratory syndrome coronavirus (SARS-CoV) disease during 2002-03 and infected around 8,000 people worldwide with a death rate of about 10% (Rota et al., 2003). During 2012-13, the middle east respiratory syndrome coronavirus (MERS-CoV) evolved and more than 1,700 people were infected, with a death rate of about 36% (Zaki et al., 2012). During 2013-14, porcine epidemic diarrhea corona virus (PEDV) again traumatized the United States with about 100% mortality rate in piglets (Mole, 2013). The world health organization reported more than 5.23 million confirmed cases with a death of more than 3, 38,000 people worldwide (as on 22nd May 2020) with a fatality rate of around 6.46%. The death statistic also reveals that the people who are above the age of 80 years are more vulnerable to death if infected with SARS-CoV-2.

In the present scenario, no drugs and vaccines has been developed to tackle this pandemic situation. To decline the infection and mortality rate, people have only option to maintain social distancing and strengthen immune system. However, for long run these aspects are not reliable.

At this outset, to combat this situation, it is essential to develop potential therapeutics against this virus. For drug discovery, one must have knowledge on the structural and molecular biology of the virus because the structural and molecular components of the virus could be targeted mostly for drug discovery. Structurally, the corona viruses are a group of large-enveloped positive-sense single-stranded RNA virus. It has four important structural proteins such as nucleocapsid protein, envelope protein, spike protein, and membrane protein (Li, 2016). The genome size of these viruses ranges from 27 to 32 kb and is packed inside a helical nucleocapsid formed by the nucleocapsid protein. Further, some important enzymes found in this virus, like SARS-CoV-2 main protease, are having some important implications in replication and virulence property of this virus.

Since time immemorial, natural products from plants, microbes and animals have been used in medicines to treat diseases. However, emergence of synthetic techniques during the nineteenth century led to the development of synthetic drugs as the pharmaceutical industries adapted the high-throughput chemistry-based synthesis approach in development of drugs. But, at par with the considerable effort, expected drug productivity target could not achieved in this direction. Many pharmaceutical industries were faced huge challenges in developing new products using this approach (Yuan et al., 2016). Further, the production of synthetic drug is more expensive and having serious side effects as compared to natural drugs derived from plants (Abiramasundari et al., 2011). During the last couple of years, more emphasis has been given to natural products based drug discovery (Ngo et al., 2013).

Natural products based development of new drugs has gaining much attention in the current time as these products have unique chemical diversity which results in diverse biological activities and drug-like characteristics **(**Yuan et al., 2016). Some plant based medicines like antihypertensive, anticancer, and antimigraine drugs are in current use and have benefited greatly to the human civilization (Joo, 2014; Newman et al., 2003). Considering their chemical diversity and unique mechanism of actions, the natural products will obviously play leading role in the coming years also to discover novel drugs for treating various critical human diseases (Galm and Shen, 2007). With this foregoing discussion, the present study aims to predict the therapeutic potential of some natural compounds and their derivatives towards inhibiting the nucleocapsid protein N-terminal RNA binding domain (NSP-NTD) of SARS-CoV-2 virus.

2 MATERIALS AND METHODS

2.1 Selection and preparation of target protein

Nucleocapsid protein RNA binding domain (PDB code: 6M3M) of SARS-CoV-2 virus was targeted to identify antiviral agent for SARS-CoV-2. The crystal structure of the protein was obtained from the RCSB-PDB (protein data bank) (http://www.rcsb.org). Further preparation and cleaning of the protein structure was carried out by using the Discovery Studio 4.0 (DS 4.0). Water molecules, cofactors, metal ions and any bonded molecules were removed from the 3D protein structure. Bond orders were assigned and hydrogens were added to heavy atoms. The protein structure was optimized and energy level was minimized by application of ‘prepare protein” protocol of DS 4.0. In order to clean the protein molecule, missing atoms and loops were added to it followed by assigning charges and fixing CHARMm force field. On the basis of the receptor cavity method, protein active sites were predicted. However, the suitable active site required for the interaction study between the ligand compounds and the protein was selected based on the amino acid residue situated in the binding pocket.

2.2 Ligand selection and preparation

A total of nineteen plant based chemical compounds and their derivatives including Ascorbic acid, 3-Glyceryl ascorbate, 4-Aminobenzoic acid, 4-Hydroxy benzoic acid, Apigenin, Benzoic acid, Citric acid, Gallic acid, L-Ascorbic acid 2-phosphate, Magnesium L-ascorbic acid-2-phosphate, Quercetin, Salicylic acid, Sodium ascorbate, Sodium ascorbyl phosphate, Sodium citrate, Syringic acid, Tricin, Vanillic acid, and Vitexin were selected based on their chemical diversity and antimicrobial property. The 2D structures of these compounds were either retrieved from PubChem database (https://pubchem.ncbi.nlm.nih.gov/) as SDF file or drawn using Biovia Draw (2019) and were incorporated into the DS 4.0 to generate the 3D structure in data source view (.DSV) file format. The ligand molecules were then prepared applying the “prepare ligand” protocol of DS 4.0 to add missing hydrogen bonds and the energy was also minimized using CHARMm force fields methods of energy minimization protocol of DS 4.0.

2.3 Elucidation of physiochemical and drug likeness properties of the compounds

Physiochemical properties of the compounds such as molecular weight, number of hydrogen bond acceptors, number of hydrogen bond donors, number of rotatable bonds, molar refractivity, lipophilicity (ALogP) and topological polar surface area (TPSA) were calculated using the SWISS ADME server tool (http://www.swissadme.ch/). Based on the physiochemical properties, drug like efficiency of the selected compounds were examined for first round screening using Lipinski, Ghose and Veber rules (Veber et al., 2002; Ghose et al., 1999; Lipinski et al., 1997). The compounds that satisfied either two of the Lipinski, Ghose and Veber rules were screened out for pharmacokinetic (ADMET) screening.

2.4 Prediction of ADMET properties of the compounds

To analyze the pharmacological potency, the compounds were subjected to *in silico* absorption, distribution, metabolism, excretion and toxicity (ADMET) prediction studies. The level of aqueous solubility, plasma protein binding (PPB), blood brain barrier (BBB), CYP2D6 binding, intestinal absorption (IA) and hepatotoxicity were evaluated. Additionally, AlogP98 and PSA_2D (2D polar surface area) were used for plotting a scatter plot of confidence ellipses. The compounds come outside the range of scatter plot of confidence ellipses were discarded from further studies. Variety of experimental data sources were used to derive the model to predict the ADMET properties, followed by archiving in the product documentation (Ponnan et al., 2013). The ADMET descriptors module quantitatively predict all the properties by a set of keys and offers detailed information of the PPB level, BBB level, absorption level, solubility level, hepatotoxicity, CYP2D6, and PSA_2D (polar surface area) values (**Table 2**). Based on these informations the pharmacological potency and toxicity were analyzed. The compounds with BBB values 2 or 3; absorption level 0 or 1; solubility level 3 or 4 were considered as optimal for a compound to be administered as drug (Rajendran et al., 2018; Ponnan et al., 2013). Further, hepatotoxicity probability and PPB level were also estimated. The hepatotoxicity probability of a drug greater than 0.5 was considered as toxic while less than 0.5 was considered as nontoxic (Rajendran et al., 2018). The PPB values were classified in two categories viz. “false-poorly bounded” and “true-highly bounded”.

2.5 Toxicity risk assessment of the compounds

In addition to ADMET studies, more detailed toxicity analyses were conducted for all screened compounds by using the toxicity prediction module ‘TOPKAT’ in DS 4.0 software. Parameters like aerobic biodegradability, ames mutagenicity, FDA carcinogenicity, occular irritancity, skin irritability, carcinogenic potency, max tolerated dose, and rat oral LD-50 were studied for detailed toxicity analysis. Compounds showing high carcinogenic and mutagenic potential with poor pharmacokinetic properties were excluded from further studies.

2.6 Molecular docking

Energy minimized ligands were docked with the target protein (NSP-NTD) using the “receptor-ligand interaction” module of DS 4.0. The docking was employed between selected ligand molecule and the identified target binding site of the protein by following ligand fit algorithm function. . Active sites of the receptor proteins were predicted using “Define and Edit Binding Site” protocol. The CDocker algorithm of “Dock lignad” menu was used for docking and the CHARMm force field algorithm was used to optimize the complex energies. All the parameters were set at their default values. The receptor-ligand interactions were visualized with the DS visualizer 4.0. CDcoker energy, CDocker interaction energy, and binding energy were calculated for all the ligand-receptor interactions. The alkyl bond, pi bond, hydrogen bond and vanderwall interactions between the atoms of ligand and the amino acids residues of the receptor molecule were analyzed and the interaction between the ligand and the receptor molecule was studied by estimating distance between interacting amino acid residues and the atoms of the ligands. The proximity of molecular dockings were analyzed on the basis of CDcoker energy, CDocker interaction energy, and binding energy. After each compound docked to the target protein (NSP-NTD), 10 best conformational poses were obtained. The compounds established binding poses with lowest binding energy, negative (-) CDOCKER energy, negative (-) CDOCKER interaction energy, and least energy difference between negative (-) CDOCKER energy and negative (-) CDOCKER interaction energy were chosen as good interacting compounds.

3 RESULTS AND DISCUSSION

3.1 Preparation of target proteins and ligands

For the development of novel therapeutics, the structure-based drug discovery has been emerged as a significant approach. Many recent studies have been targeting the spike proteins and viral proteases to combat the COVID-19 (Kang et al., 2020). However, the substantial outcome of these studies pertaining to the effective targeted therapeutic for the pandemic were reported to be negligible. Recent studies advocated that the nucleocapsid proteins could be an excellent target in corona viruses (CoVs) for drug discovery since they have important functions, such as RNA genomic packing, viral transcription and assembly, in the infectious cells (Kang et al., 2020). Based on this information the crystal structure of SARS-CoV-2 nucleocapsid protein N-terminal RNA binding domain with four polypeptide chains (A, B, C, D) was selected as the suitable target for this study and was retrieved from the RCSB PDB. The raw PDB protein structure usually not used in molecular docking studies directly as it contains unessential water molecules, heavy atoms, metal ions and cofactors. The PDB structure does not have the information about bond orders, topologies or formal atomic charges. Further, the X-ray structure analysis cannot distinguish between O and NH_2 ionization and hence the terminal amide groups may get misaligned and the tautomeric states may also unassigned. Therefore, the raw PDB protein structure should be prepared in an appropriate manner for molecular docking (Murugesan et al., 2014). Hence, the retrieved X-ray crystal structure was prepared and energy minimization was done to stabilize the structure for performing docking analysis. Six active sites were predicted from the receptor cavities of this target protein (NSP-NTD) using the Receptor-Ligand Interactions module of DS 4.0. Of these predicted active sites, the first active site was selected, based on its size, as the binding site for the study.

Based on earlier literatures nineteen natural compounds and their derivatives were selected and used as ligands for the present study. By using DS 4.0 3D structure of the compounds were prepared through compound cleaning, calculating 3D coordinates and generating possible isomers. Different conformations of the ligands were obtained after the preparation of ligands. 3-Glyceral ascorbate with eight, 4-Aminobenzoic acid with only one, 4-Hydroxybenzoic acid with two, Apigenin with seven, Ascorbic acid with eight, Citric acid with one, Benzoic acid with one, Gallic acid with one, L-ascorbic acid 2-phosphate with eight, Magnesium L-ascorbic acid-2-phosphate with eight, Quercetin with five, Salicylic acid with one, Sodium citrate with one, Sodium ascorbate with eight, Sodium ascorbyl phosphate with eight, Syringic acid with two, Tricin with seven, Vanillic acid with two, and vitexin with two hundred eighty eight conformations were predicted and these conformations of the suitable ligands were analyzed with the target protein (NSP-NTD) through molecular docking to obtain the top hit interacting poses.

3.2 Drug likeness, pharmacokinetic and toxicity profiling

In the process of drug development so many potent drugs fails either in clinical trial stage or in the later stag of drug discovery, due to poor drug likeliness and pharmacokinetic characteristics, and toxicity. In the present study, all the cleaned and optimized compounds were subjected to drug likeness, ADMET and detail toxicity prediction studies before molecular docking to make sure the toxicity risks and drug-likeness characteristics of these compounds.

3.2.1 Screening of drug likely compounds

Drug likeness properties of the compounds were assayed following the Lipinski's rule of five, Ghose and Veber rules of drug likeness (Veber et al., 2002; Ghose et al., 1999; Lipinski et al., 1997). For better results, the compounds those qualified either two of the Lipinski's rule of five, Ghose and Veber rules were screened as compounds with drug likeness properties. Accordingly five compounds out of 19 including citric acid, L-Ascorbic acid 2-phosphate, Magnesium L-ascorbic acid-2-phosphate, Sodium ascorbate, and Sodium citrate disqualified in these tests and discarded from further study (**Figure 1**; **Table 1**).

3.2.2 ADMET Analysis

The fourteen compounds satisfying the drug likeness properties were further used for predicting the pharmacokinetic parameters using the “ADMET descriptors” module of DS 4.0 (**Table 2**). In the drug discovery process the ADMET properties of the compounds have important role to play as these properties are mostly accountable for failure of drugs in approximately 60% of the clinical trials cases (Rajendran et al., 2018). The ADMET descriptors module in DS 4.0 calculates the AlogP98 and PSA (polar surface area), plasma protein binding level, hepatotoxicty, cytochrome P450 2D6 (CYP2D6) enzyme inhibition, aqueous solubility level, blood brain barrier penetration level, and intestinal absorption level of the drug like compounds. The value of AlogP98 determines the hydrophilicity of the compound where AlogP98>5 may indicates good absorption or permeability of the compound. PSA is another key parameter associated with bioavailability of the drug. Compounds with PSA<140 $Å^2$ can be absorbed passively and so have high oral bioavailability (Yang et al., 2020). The BBB level shows the amount of penetration of the drug into the central nervous system (CNS) after oral administration. An ideal drug should not penetrate the BBB level as it can cause side effects in the CNS. Thus, the drug compounds with BBB values 3 or 2 (low or medium) are usually considered as optimal for a drug to be administered (Rajendran et al., 2018). In the present study BBB predictions showed that all the compounds had low BBB permeability except one that had undefined permeability (**Table 2**). In the same line, absorption level and solubility level depicts the human intestinal absorption and drug likeliness of the compounds respectively after oral administration. The accepted values of the drug compounds should be either 0 (good absorption) or 1 (moderate absorption) for intestinal absorption and 3 (good) or 4 (optimal) for aqueous solubility. The present study predicted 10 compounds out of 14 compounds within the 99% confidence ellipse with good absorption. Regarding aqueous solubility, seven compounds were predicted with optimal solubility, two with good solubility and one with extremely high solubility (**Figure 2**; **Table 2**). The hepatotoxicty level of a drug compound can be based on its effect of causing dose dependent liver injuries and hence the drug toxicity is usually predicted based on the hepatotoxicity probability (Rajendran et al., 2018). In DS 4.0 the true element w.r.t. hepatotoxicity indicates toxic effect and false element indicates the non toxic effect of the compounds. In the present study, none of the compounds inhibited the CYP2D6 enzyme and no serious drug interaction toxicity caused in the liver. The PPB refers to the degree to which a drug binds to the blood proteins. The efficacy of the drug may be affected based on its degree to which it binds. With reference to PPB, all the 10 compounds had poor PPB activity except four that had strong PPB activity (**Table 2**). The molecules that are difficult for the intestine to absorb and easily penetrate the BBB were further discarded from further study by following the 95% and 99% confidence ellipses model formed from the two-dimensional polar surface area (PSA_2D) and the calculated value of AlogP98 (**Figure 2**). Four of the selected compounds did not pass the ADMET screening; leaving 10 candidate compounds for molecular docking study (**Table 4).** The detail ADMET scores of the compounds satisfying the drug likeness properties, validated through ADMET prediction analysis are given in **Table 2**. The relationship between the calculated value of AlogP98 and two-dimensional polar surface area (PSA_2D) for all the 14 compounds, with the HIA and BBB penetration model 95% and 99% confidence ellipses, were depicted in **Figure 2**.

3.2.3 Toxicity prediction

Detail toxicity prediction is also equally important for screening a potential drug. This was carried out by using the toxicity prediction (TOPKAT) module of DS 4.0, which computes a probable value of toxicity of the compounds from a quantitative structure-toxicity relationship (QSTR) equation. The product of a structure descriptors and its corresponding coefficient is the descriptors contribution to the probable toxicity. In this study majority of the compounds, out of the ten, found to be non-degradable, non-mutagenic, and exhibiting mild to moderate occular irritancity and skin irritability (**Table 3**). With respect to carcinogenicity of these compounds, some of them were predicted to be carcinogenic. However, the carcinogenic effects of these analyzed compounds with less carcinogenic potential may not be so harmful if the administration dose would be less than the carcinogenic potency (TD 50: the dose required to produce carcinogenic effect in 50% of animals) values (**Table 3**).

3.3 Molecular docking and analysis of protein-ligand interactions

Molecular docking has become an important computational approach for virtual screening of drug likely compounds. It paves a new way for rapid drug discovery as it evaluates the activity of a pretty good number of compounds against the target proteins and provides information about the candidate ligand-target protein interactions within a stipulated time period by cutting the cost of laboratory based screening. In this study, the candidate compounds 4-Aminobenzoic acid, 4-Hydroxybenzoic acid, Apigenin, Benzoic acid, Gallic acid, Quercetin, Salicylic acid, Syringic acid, Tricin, and Vanillic acid were docked with the nucleocapsid protein N-terminal RNA binding domain of SARS-CoV2 virus to find out the most favorable interacting complexes.

In the DS 4.0, the negative (–) CDOCKER energy, negative (–) CDOCKER interaction energy and the difference between the negative (–) CDOCKER and negative (–) CDOCKER interaction energies are the criterion to represents the ligand-protein interaction. The least difference between the negative (–) CDOCKER energy and negative (–) CDOCKER interaction energy indicates more favorable binding. The difference between the docking scores is quite dependent on the binding poses exhibited by the ligands. The top hit binding pose provides the least energy difference between negative (–) CDOCKER energy and negative (–) CDOCKER interaction energy. Considering the difference between the negative (-) CDOCKER energy and negative (-) CDOCKER interaction energy, molecular docking of the target protein (NSP-NTD) with the screened compounds was carried out which indicated that all the tested ligands were interactive with the target protein (**Table 4**).

To foresee the extent of interaction again the interacting complexes were utilized for calculating the binding energies as the binding energy of complex gives better understanding of binding affinity of the docked complex. It is known that the protein-ligand interactions occur spontaneously only when the free energy change in the interacting system is negative, and the difference in the free energy levels of complexed and unbound free states is relative to the binding energy [$\Delta E = E_{\text{complex}} - (E_{\text{enzyme}} + E_{\text{ligand}})$] and stability of the protein–ligand interaction (Afriza et al., 2018). The negative binding energy indicates the stability of the resulting complexes with receptor molecules, and this is an essential characteristic of effective drugs (Muthu et al., 2016). The present study advocates that out of the 10 analyzed compounds, four were found efficient to establish strong and stable interactions with the target protein (**Figure 3**; **Table 4).** Salicylic acid-target protein (NSP-NTD) interaction had the largest negative binding energy (-111.65) which was followed by the interactions of target protein (NSP-NTD) with Benzoic acid (-111.41), 4-hydroxybenzoic acid (-105.92) and 4-Aminobenzoic acid (-63.75). With emphasis to the difference in negative (–) CDOCKER energy and negative (–) CDOCKER interaction energy 4-Aminobenzoic acid had the least energy difference (1.45) followed by 4-hydroxybenzoic acid (1.97), Salicylic acid (2.16), and Benzoic acid (2.54). Thus, it can be stated that all these four ligands have more or less the potential to interact and inhibit the activity of the target protein (NSP-NTD).

Although binding energy, minus (-) CDOCKER energy and minus (-) CDOCKER interaction energy are informative for understanding the proximity of docked complexes, types of interactions such as hydrophobic interaction, hydrogen bonding, and electrostatic interactions with essential amino acid residues are indicative of docking of ligand in most favored conformations (Hariono et al., 2016). Findings in the present study shows that hydrogen bond, pi-pi interaction, pi-alkyl interaction and electrostatic interactions were mediated the complexes by different amino acid residues in every ligand–protein interaction (**Figure 3**). The binding site of the protein that are involved in different types of bonding such as hydrogen bond, hydrophobic, and electrostatic interactions with the ligands and also the amino acid residues in the binding pockets were different with respect to different ligands. 4-aminobenzoic-target protein (NSP-NTD) interaction was governed by hydrogen bond (2.54) and electrostatic interaction (1.93) at residue Lys B: 66, Pi-Pi interaction at Tyr A: 110 (4.18) residue and pi-alkyl interactions at Ala A: 51 (4.67) amino acid residue. Similarly, two hydrogen bonds were formed by Asn A: 49 (1.88) and Thr A: 50 (2.22) with 4-hydroxybenzoic acid. One pi-alkyl and another electrostatic interaction were also formed in the 4-hydroxybenzoic acid-protein complex by Ala A: 51 (5.26) and Lys B: 66 (2.95) respectively. No hydrogen bonds were imputed between benzoic acid and the protein binding site, whereas three interactions such as pi-pi, Pi-alkyl and an electrostatic interaction were formed by Tyr A: 110 (3.87), Ala A: 51 (4.88) and Lys B: 66 (2.26) amino acid residues respectively. Salicylic acid-protein interactions were supported by two hydrogen bonds at residues Arg A: 108 (2.46), and Lys B: 66 (1.93), and by three electrostatic interactions at residues Arg A: 108 (3.83), Arg A: 93 (5.04), and Lys B: 66 (5.42). The details about the intermolecular interaction results including the type of interactions, distance of interactions, involved amino acid residues and the receptor surfaces are illustrated in **Figure 3**.

4 CONCLUSIONS

The coronavirus disease (COVID-19) has emerged as a pandemic and affected the life and economy of every individual in many countries. Unavailability of specific drug against the causative agent of this disease (SARS-CoV-2 virus) created the situation more adverse over the globe. The present study evaluated 19 natural compounds and their derivatives with respect to their drug likeness, pharmacokinetic properties, and toxicity effects and finally molecular docking was carried out to identify the potent compounds with antiviral potency against SARS-CoV-2. In a nutshell, the study identified four compounds such as 4-hydroxybenzoic acid, benzoic acid, 4-aminobenzoic acid and salicylic acid as potential ligands that inhibit the target protein (NSP-NTD). These compounds may further be evaluated with respect to mutagenicity, carcinogenicity, toxicity and inhibitory activity through *in-vitro* and *in-vivo* studies to validate their potentiality to inhibit the target protein (NSP-NTD) of SARS-CoV-2 virus.

ACKNOWLEDGEMENTS

The authors acknowledge the Centurion University of Technology and Management, Odisha, India for the resources and facilities provided to carry out this work and thankful to Mr. Pradeep Sarangi (Regional Director, CUTM, Bolangir) for giving motivation to conduct the study.

CONFLICT OF INTEREST

The authors disclose that there is no potential conflict of interest.

REFERENCES

Abiramasundari, P.; Priya, V.; Jeyanthi, G. P.; Gayathri, D. (2011). Evaluation of the antibacterial activity of *Cocculus hirsutus. J Drugs Medicines*, **3**, 26-31.

Afriza, D.; Suriyah, W. H.; Ichwan, S. J. A. (2018). In silico analysis of molecular interactions between the anti-apoptotic protein survivin and dentatin, nordentatin, and quercetin. *Journal of Physics: Conference Series*, **1073**, 032001.

Enjuanes, L.; Almazan, F.; Sola, I.; Zuniga S. (2006). Biochemical aspects of coronavirus replication and virus-host interaction. *Ann. Rev. Microbiol.*, **60**, 211-230.

Galm, U.; Shen, B. (2007). Natural product drug discovery: The times have never been better. *Chem. Biol.*, **14**, 1098-1104.

Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of Combinatorial Chemistry*, **1**, 55-68.

Hariono, M.; Abdullah, N.; Damodaran, K. V.; Kamarulzaman, E. E.; Mohamed, N.; Hassan, S. S.; Shamsuddin, S.; Wahab, H. A. (2016). Potential new H1N1 neuraminidase inhibitors from ferulic acid and vanillin: molecular modelling, synthesis and in vitro assay. *Sci. Rep.*, **6**, 1-10.

Joo, Y. E. (2014) Natural product-derived drugs for the treatment of inflammatory bowel diseases. *Intest. Res.*, **12**, 103-109.

Kang, S.; Yang, M.; Hong, Z.; Zhang, L.; Huang, Z.; Chen, X.; et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites *Acta Pharmaceutica Sinica B*, 2020. https://doi.org/10.1016/j.apsb.2020.04.009

Li F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annual review of virology*, **3**, 237-261.

Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, **23**, 3-25.

Mole B. (2013). Deadly pig virus slips through US borders. *Nature*, **499**, 388.

Murugesan, D.; Ponnuswamy, R. D.; Gopalan, D. K. (2014). Molecular docking study of active phytocompounds from the methanolic leaf extract of *Vitex negundo* against cyclooxygenase-2. *Bangladesh Journal of Pharmacology*, **9**, 146-153.

Muthu, S.; Durairaj, B. (2016). Molecular docking studies on interaction of *Annona muricata* compounds with antiapoptotic proteins Bcl-2 and surviving. *Sky J Biochem Res*, **5**, 14-17.

Newman, D. J.; Cragg, G. M.; Snader, K. M. (2003). Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.*, **66**, 1022-1037.

Ngo, L. T.; Okogun, J.I.; Folk, W.R.(2013). 21st Century natural product research and drug development and traditional medicines. *Nat. Prod. Rep.*, **30**, 584-592.

Ponnan, P.; Gupta, S.; Chopra, M.; Tandon, R.; Baghel, A. S.; Gupta, G.; et al. (2013). 2D-QSAR, docking studies, and in silico ADMET prediction of polyphenolic acetates as substrates for protein acetyl transferase function of glutamine synthetase of *Mycobacterium tuberculosis*. *ISRN Structural Biology*, 2013.

Rajendran, B. K.; Suresh, M. X.; Bhaskaran, S. P.; Harshitha, Y.; Gaur, U.; Kwok, H. F. (2018). Pharmacoinformatic approach to explore the antidote potential of phytochemicals on bungarotoxin from Indian Krait, Bungarus caeruleus. *Computational and Structural Biotechnology Journal*, **16**, 450-461.

Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R.; et al. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 300, 1394-1399.

Schoeman, D.; Fielding, B. C. (2019). Coronavirus envelope protein: Current knowledge. *Virology Journal*, **16**, 69.

Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, **45**, 2615-2623.

Yang, Y.; Shi, C. Y.; Xie, J.; Dai, J. H.; He, S. L.; Tian, Y. (2020). Identification of potential dipeptidyl peptidase (DPP)-IV inhibitors among *Moringa oleifera* phytochemicals by virtual screening, molecular docking analysis, ADME/T-Based Prediction, and in vitro analyses. *Molecules*, **25**, 189.

Yuan, H.; Ma, Q.; Ye, L.; Piao, G. (2016). The traditional medicine and modern medicine from natural products. *Molecules*, **21**, 559.

Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.; Fouchier, R.A.M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.*, **367**, 1814-1820.

Table 1. Physiochemical and drug-likeness properties of the compounds

	Physiochemical properties							**Drug-likeness**			
Compound	**Mol. Wt. (g/mol)**	**Alog P**	**Rotat-able bonds**	**H bond acceptors**	**H bond donors**	**Molar refractivity**	**Oral bioavailability: TPSA (Å²)**	**Lipinski**	**Ghose**	**Veber**	**Bioavailabili-ty Score**
3-Glyceryl ascorbate	249.19	-3.237	6	8	4	49.44	139.51	Yes	No; 1 violation: WLOGP<-0.4	yes	0.56
4-Aminobenzoic acid	136.13	-0.761	1	2	1	35.86	66.15	Yes	No; 3 violations: MW<160, MR<40, #atoms<20	yes	0.56
4-Hydroxybenzoic acid	137.11	-0.257	1	3	1	33.48	60.36	Yes	No; 3 violations: MW<160, MR<40, #atoms<20	Yes	0.56
Apigenin	270.24	2.41	1	5	3	73.99	90.90	Yes	Yes	Yes	0.55
Ascorbic acid	175.12	-2.063	2	6	3	33.18	110.05	Yes	No; 3 violations: WLOGP<-0.4, MR<40, #atoms<20	Yes	0.56
Benzoic acid	121.11	-1.5	1	2	0	31.46	40.13	Yes	No; 3 violations: MW<160, MR<40, #atoms<20	Yes	0.56
Citric acid*	189.10	-5.74	5	7	1	31.64	140.62	Yes; 0 violation	No; 3 violations:	No; 1 violati	0.56

		1							WLOGP<-0.4, MR<40, #atoms<20	on: TPSA >140	
Gallic acid	249.19	-0.741	6	8	4	49.44	139.51	Yes; 0 violation	No; 1 violation: WLOGP<-0.4	Yes	0.56
L-Ascorbic acid 2-phosphate*	254.09	-3.978	4	9	3	42.56	169.22	Yes; 0 violation	No; 1 violation: WLOGP<-0.4	No; 1 violation: TPSA >140	0.11
Magnesium L-ascorbic acid-2-phosphate*	254.09	-3.978	4	9	3	42.56	169.22	Yes; 0 violation	No; 1 violation: WLOGP<-0.4	No; 1 violation: TPSA >140	0.11
Quercetin	300.22	0.224	1	7	3	74.25	137.02	Yes; 0 violation	Yes	Yes	0.56
Salicylic acid	137.11	-0.257	1	3	1	33.48	60.36	Yes; 0 violation	No; 3 violations: MW<160, MR<40, #atoms<20	Yes	0.56
Sodium ascorbate*	189.10	-5.741	5	7	1	31.64	140.62	Yes; 0 violation	No; 3 violations: WLOGP<-0.4, MR<40, #atoms<20	No; 1 violation: TPSA >140	0.56
Sodium ascorbyl phosphate	175.12	-2.063	2	6	3	33.18	110.05	Yes; 0 violation	No; 3 violations: WLOGP<-0.4, MR<40, #atoms<20	Yes	0.56

Sodium citrate[*]	254.09	-3.978	4	9	3	42.56	19.22	Yes; 0 violation	No; 1 violation: WLOGP<-0.4	No; 1 violation: TPSA >140	0.11
Syringic acid	197.16	-029	3	5	1	46.47	78.82	Yes; 0 violation	Yes	Yes	0.56
Tricin	330.29	2.377	3	7	3	86.97	109.36	Yes; 0 violation	Yes	Yes	0.55
Vanillic acid	167.14	-0.273	2	4	1	39.97	69.59	Yes; 0 violation	No; 2 violations: MR<40, #atoms<20	Yes	0.56
Vitexin	432.38	1.8	3	10	7	106.61	181.05	Yes; 1 violation: NH or OH>5	Yes	No; 1 violation: TPSA >140	0.55

[*] Not satisfying either two of the Lipinski, Ghose and Veber drug likeness rules

Table 2. *In silico* pharmacokinetic (ADMET) properties of the selected ligands

Compound	**PPB**	**Hepatotoxicity**	**CYP2D6 binding**	**[a]Aqueous Solubility**	**[b]BBB Penetration**	**[c]Intestinal Absorption**
3-Glyceryl ascorbate*	False (Poorly bounded)	False (Non-toxic)	False (Non-inhibitor)	5 (Too soluble)	4 (Undefined)	3 (Very low absorption)
4-Aminobenzoic acid	False (Poorly bounded)	True (Toxic)	False (Non-inhibitor)	4 (Optimal)	3 (Low)	0 (Good absorption)
4-Hydroxybenzoic acid	False (Poorly bounded)	True (Toxic)	False (Non-inhibitor)	4 (Optimal)	3 (Low)	0 (Good absorption)
Apigenin	True (Highly bounded)	True (Toxic)	False (Non-inhibitor)	3 (Good)	3 (Low)	0 (Good absorption)
Ascorbic acid*	False (Poorly bounded)	False (Non-toxic)	False (Non-inhibitor)	5 (Too soluble)	4 (Low)	2 (Low absorption)
Benzoic acid	True (Highly bounded)	True (Toxic)	False (Non-inhibitor)	4 (Optimal)	3 (Low)	0 (Good absorption)
Gallic acid	False (Poorly bounded)	True (Toxic)	False (Non-inhibitor)	5 (Too soluble)	3 (Low)	0 (Good absorption)
Quercetin	False (Poorly bounded)	True (Toxic)	False (Non-inhibitor)	4 (Optimal)	4 (Undefined)	0 (Good absorption)
Salicylic acid	False (Poorly bounded)	True (Toxic)	False (Non-inhibitor)	4 (Optimal)	3 (Low)	0 (Good absorption)
Sodium ascorbyl phosphate*	False (Poorly bounded)	False (Non-toxic)	False (Non-inhibitor)	5	4 (Undefined)	2 (Low absorption)
Syringic acid	True (Highly bounded)	True (Toxic)	False (Non-inhibitor)	4 (Optimal)	3 (Low)	0 (Good absorption)
Tricin	True (Highly bounded)	True (Toxic)	False (Non-inhibitor)	3 (Good)	3 (Low)	0 (Good absorption)
Vanillic acid	False (Poorly bounded)	False (Non-toxic)	False (Non-inhibitor)	4 (Optimal)	3 (Low)	0 (Good absorption)
Vitexin*	False (Poorly bounded)	True (Toxic)	False (Non-inhibitor)	3 (Good)	4 (Undefined)	3 (Very low absorption)

[a]**Aqueous solubility level**- 0: No (extremely low), 1: very low, 2: low, 3: good, 4: optimal, 5: No (Too soluble), 6: Warning: molecules with one or more unknown AlogP98 types; [b]**Blood brain barrier penetration (BBA) level**- 0: Very High, 1:High, 2: Medium, 3: Low, 4: Undefined, 5: Warning: molecules with one or more unknown AlogP calculation; [c]**Intestinal absorption level**- 0: Good absorption, 1: Moderate absorption, 2: Low absorption, 3: Very low absorption; *Compounds eliminate from molecular docking as not coming within the 95% and 99% confidence limit ellipses corresponding to the blood brain barrier (BBB) and intestinal absorption

Table 3. Detail toxicity risk assessment of the drug likeness compounds

Compound	**Aerobic biodegradability**	**Ames mutagenicity**	**FDA carcinogenicity (F:Female; M: Male)**	**Carcinogenic potency (TD 50 mouse); mg/kg body weight/day**	**Max tolerated dose rat g/kg body weight**	**Rat oral LD 50 (g/Kg body weight)**	**Occular irritancy**	**Skin irritability**
4-Aminobenzoic acid	Non-degradable	Mutagenic	Carcinogenic (F & M)	177	0.076	0.56	Moderate	Non-Irritant
4-Hydroxybenzoic acid	Non-degradable	Non-mutagenic	Carcinogenic (F & M)	516	0.232	0.67	Moderate	Non-Irritant
Apigenin	Non-degradable	Non-mutagenic	Carcinogenic (F & M)	75.8	0.541	0.75	Moderate	Non-Irritant
Benzoic acid	Degradable	Non-mutagenic	Carcinogenic (F & M)	121.11	0.064	1.28	Mild	Mild
Gallic acid	Non-degradable	Non-mutagenic	Carcinogenic (F & M)	291	0.605	0.74	Moderate	Non-Irritant
Quercetin	Non-degradable	Non-mutagenic	Non-carcinogenic (F & M)	113	0.295	0.49	Moderate	Non-Irritant
Salicylic acid	Non-degradable	Mutagenic	Non-carcinogenic (F) /Carcinogenic (M)	1.54	0.232	0.91	Moderate	Non-Irritant
Syringic acid	Degradable	Non-mutagenic	Non-carcinogenic (F) /Carcinogenic (M)	939	0.132	0.91	Mild	Mild
Tricin	Non-degradable	Non-mutagenic	Non-carcinogenic (F & M)	71.2	0.358	0.53	Mild	-
Vanillic acid	Degradable	Non-mutagenic	Non-carcinogenic (F & M)	1.48	0.125	1.18	Mild	Mild

Table 4. Molecular docking energies of SARS-CoV-2 nucleocapsid protein N-terminal RNA binding domain with the selected ligands

Ligand	**Minus (-) CDOCKER energy**	**Minus (-) CDOCKER INTERACTION energy**	**Difference**	**Binding energy**	**Ligand energy**	**Protein energy**	**Complex energy**
4-Aminobenzoic acid	28.7827	30.2343	1.4516	-63.7463	-11.7009	-9,523.37	-9,598.82
4-Hydroxybenzoic acid	29.9987	31.9704	1.9717	-105.92	-15.0794	-9,523.37	-9,644
Apigenin	24.2194	33.0373	8.8179	26294600	6.6932	-9,523	26,285,100
Benzoic acid	27.7849	30.3268	2.5419	-111.41	-3.0965	-9,523	-9,638
Gallic acid	39.6591	36.681	-2.9781	49,765.20	-37.5013	-9,523	40,204
Quercetin	65.8693	72.7389	6.8696	4692050	53.24	-9,523.37	4,682,580
Salicylic acid	26.7557	28.9212	2.1655	-111.646	3.853	-9,523.37	-9,631.16
Syringic acid	33.5135	40.2887	6.7752	194,955	-15.0322	-9,523.37	185,417.00
Tricin	30.1131	29.9892	-0.1239	135.40	-83.5634	-9,523.37	-9,471.54
Vanillic acid	31.5056	33.7383	2.2327	2,974.46	-21.1302	-9,523.37	-6,570

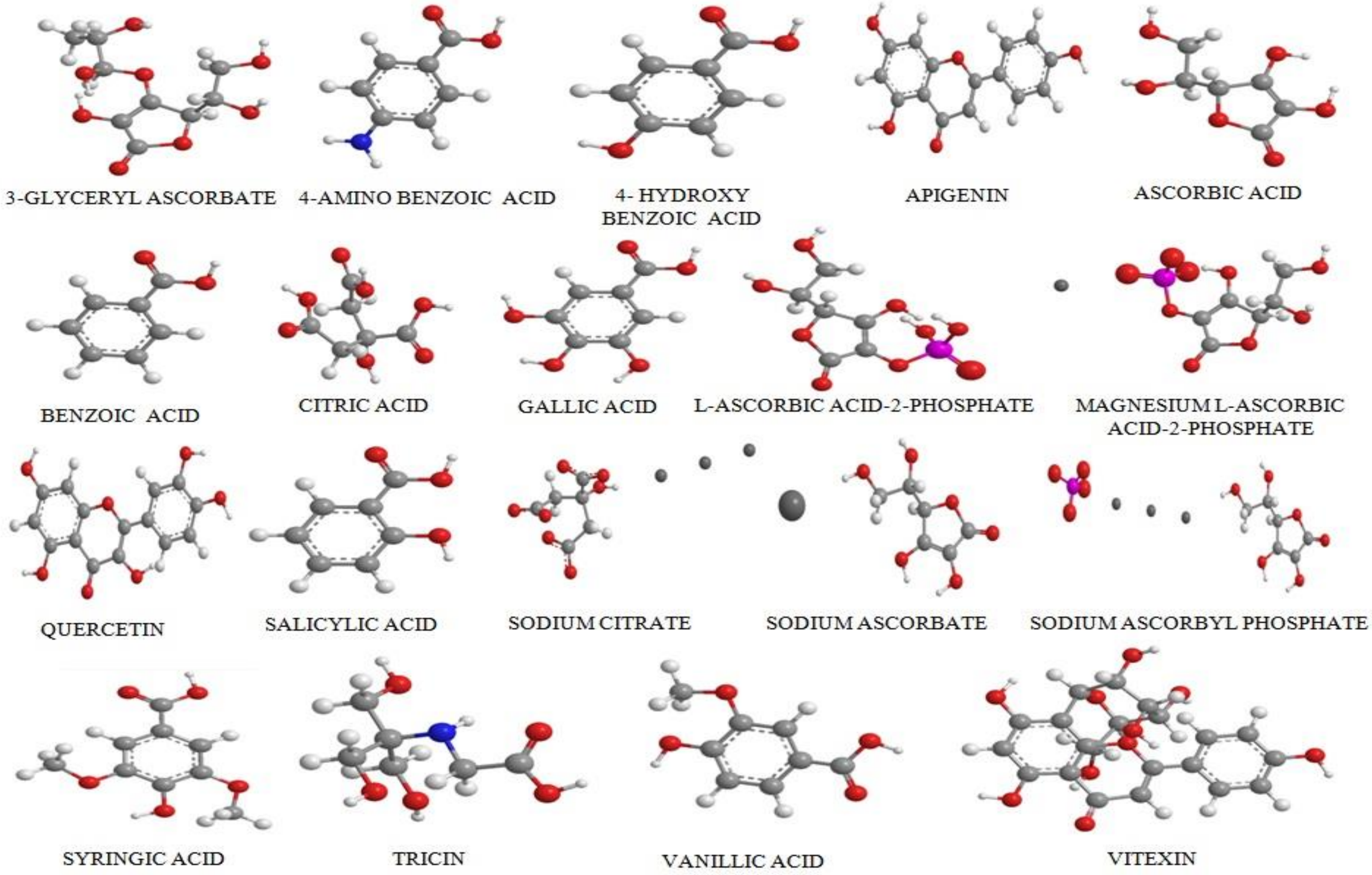

3-GLYCERYL ASCORBATE
4-AMINO BENZOIC ACID
4- HYDROXY BENZOIC ACID
APIGENIN
ASCORBIC ACID
BENZOIC ACID
CITRIC ACID
GALLIC ACID
L-ASCORBIC ACID-2-PHOSPHATE
MAGNESIUM L-ASCORBIC ACID-2-PHOSPHATE
QUERCETIN
SALICYLIC ACID
SODIUM CITRATE
SODIUM ASCORBATE
SODIUM ASCORBYL PHOSPHATE
SYRINGIC ACID
TRICIN
VANILLIC ACID
VITEXIN

Figure 1. 3D molecular structure of the natural compounds and their derivatives evaluated for their antiviral potential against SARS-CoV-2

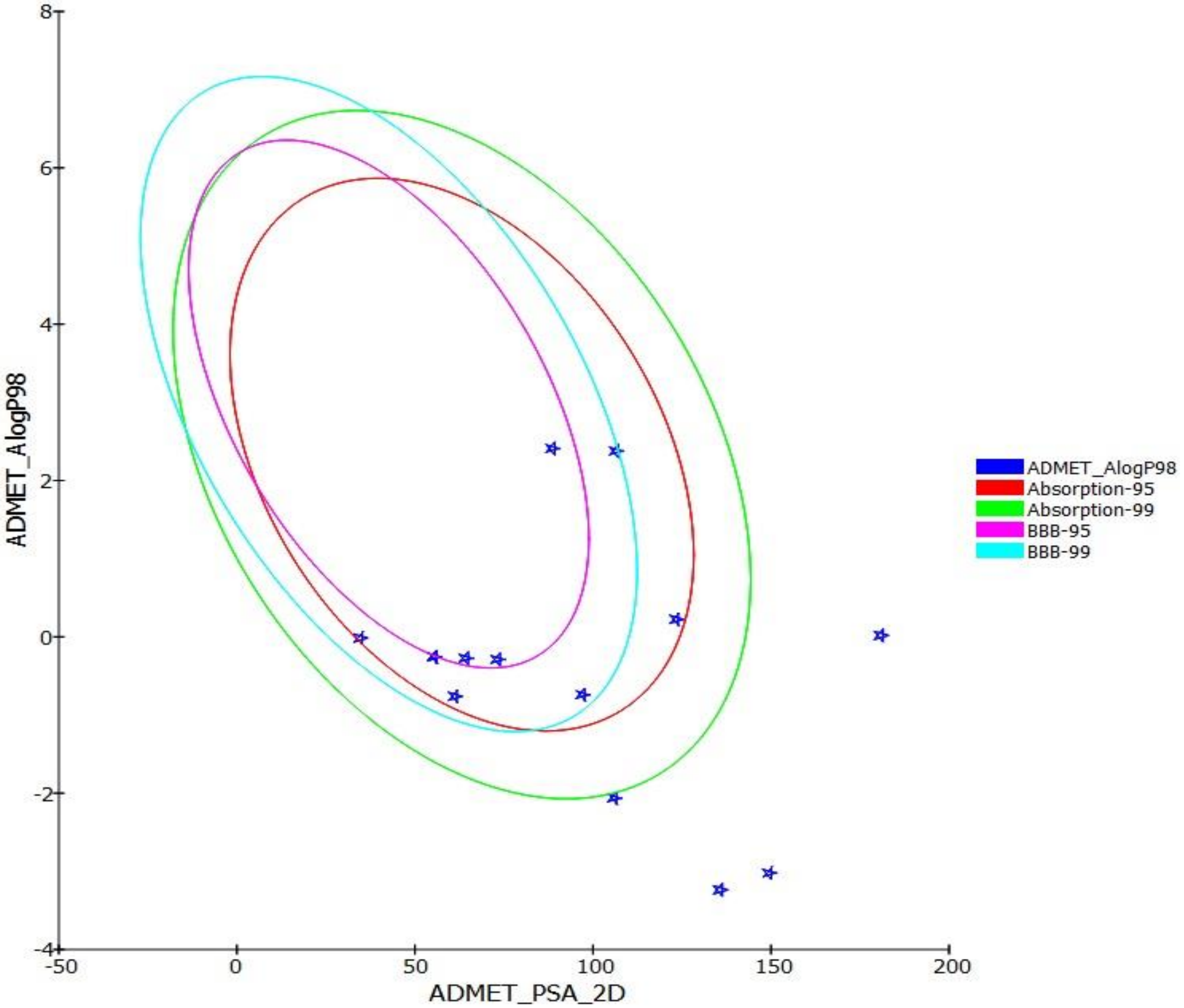
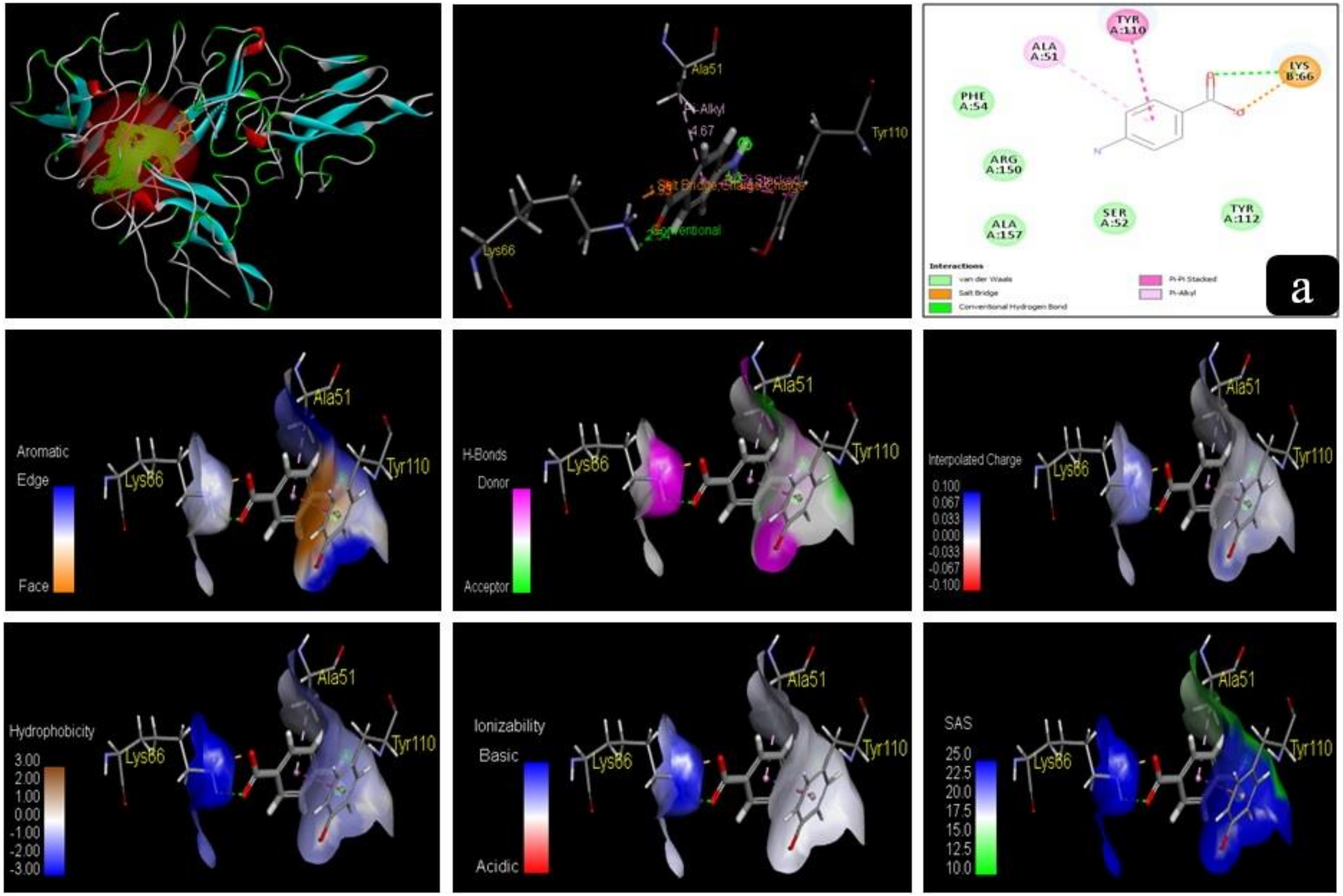
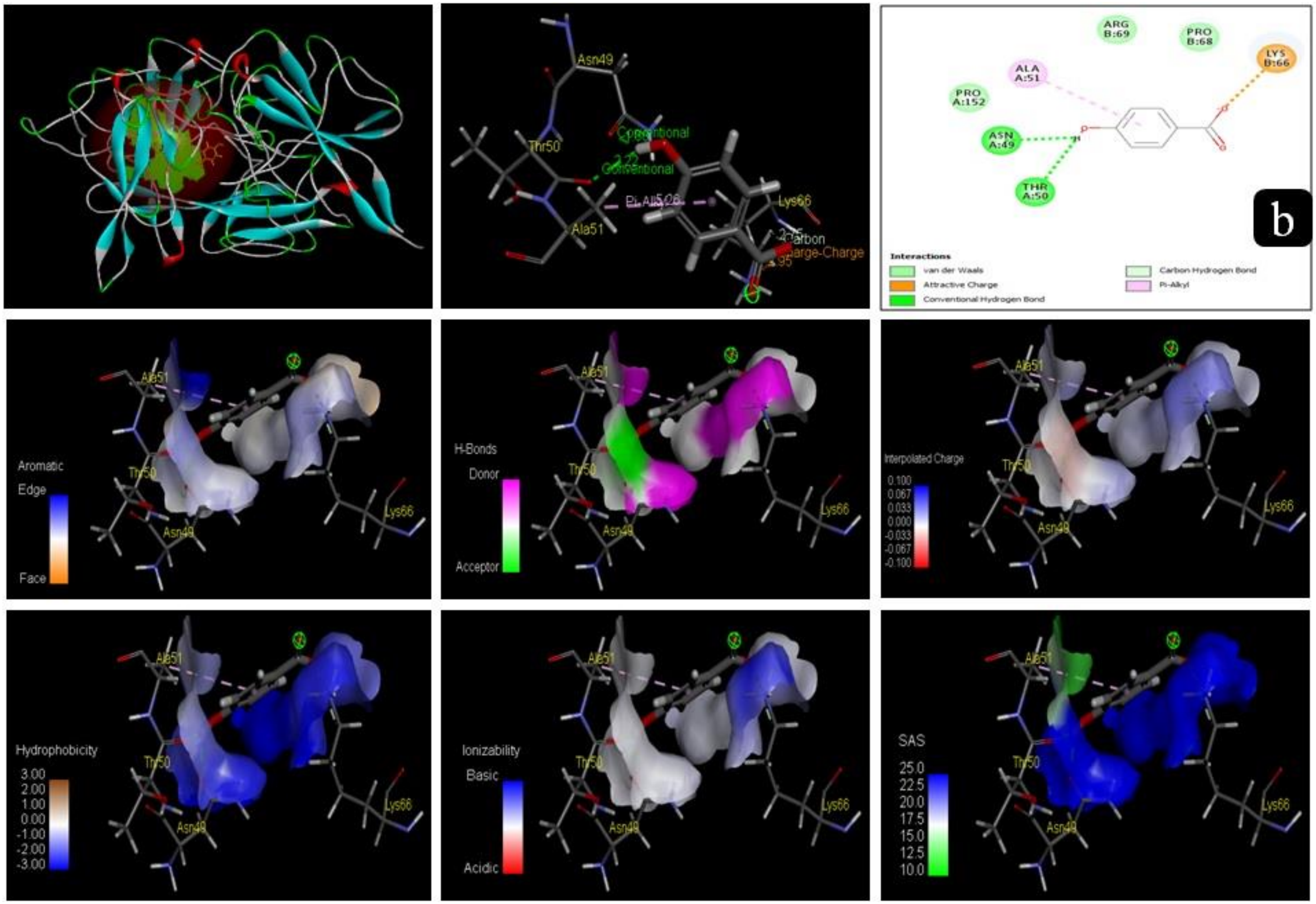
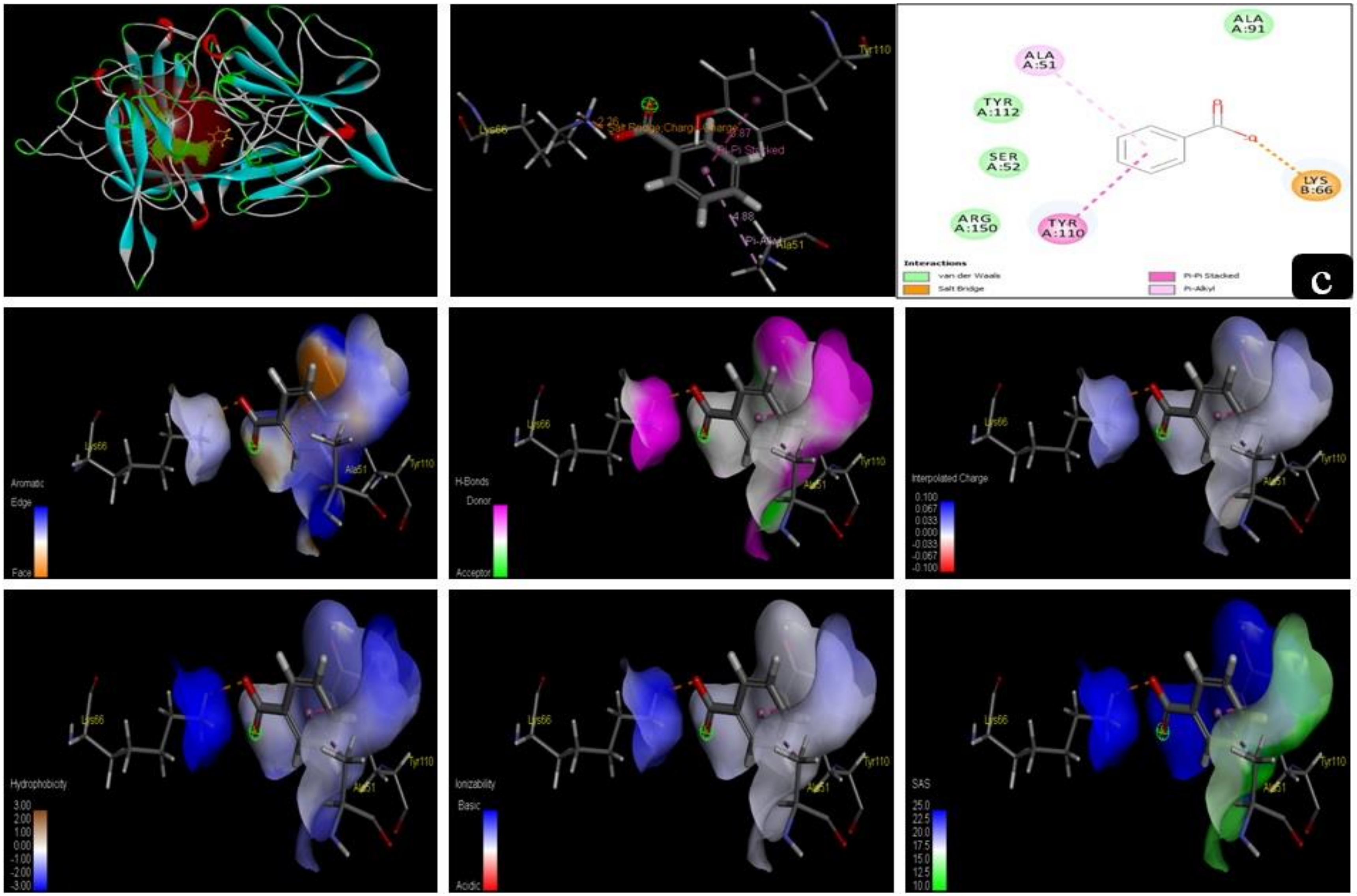


Figure 2. Plot of polar surface area (PSA_2D) versus ALogP for the selected ligands showing the 95% and 99% confidence limit ellipses corresponding to the blood brain barrier (BBB) and intestinal absorption



Ala51
Pi-Alkyl
4.67
Tyr110
Salt Bridge;Charge-Charge
Pi-Pi Stacked
Conventional
2.54
Lys66
TYR A:110
ALA A:51
LYS B:66
PHE A:54
ARG A:150
ALA A:157
SER A:52
TYR A:112
N
O
Interactions
van der Waals
Salt Bridge
Conventional Hydrogen Bond
Pi-Pi Stacked
Pi-Alkyl
a
Aromatic
Edge
Face
H-Bonds
Donor
Acceptor
Interpolated Charge
0.100
0.067
0.033
0.000
-0.033
-0.067
-0.100
Hydrophobicity
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
Ionizability
Basic
Acidic
SAS
25.0
22.5
20.0
17.5
15.0
12.5
10.0


Asn49
Thr50
Ala51
Lys66
Conventional
2.22
Conventional
Pi-Alkyl 5.26
Carbon
Charge-Charge
ARG B:69
PRO B:68
LYS B:66
ALA A:51
PRO A:152
ASN A:49
THR A:50
b
Interactions
van der Waals
Attractive Charge
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Alkyl
Aromatic
Edge
Face
H-Bonds
Donor
Acceptor
Interpolated Charge
0.100
0.067
0.033
0.000
-0.033
-0.067
-0.100
Hydrophobicity
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
Ionizability
Basic
Acidic
SAS
25.0
22.5
20.0
17.5
15.0
12.5
10.0


Tyr110
Lys66
2.26
Salt Bridge;Charge-Charge
3.87
Pi-Pi Stacked
4.88
Pi-Alkyl
Ala51
ALA A:91
ALA A:51
TYR A:112
SER A:52
ARG A:150
TYR A:110
LYS B:66
O
O
Interactions
van der Waals
Salt Bridge
Pi-Pi Stacked
Pi-Alkyl
c
Aromatic
Edge
Face
Lys66
Ala51
Tyr110
H-Bonds
Donor
Acceptor
Interpolated Charge
0.100
0.067
0.033
0.000
-0.033
-0.067
-0.100
Hydrophobicity
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
Ionizability
Basic
Acidic
SAS
25.0
22.5
20.0
17.5
15.0
12.5
10.0

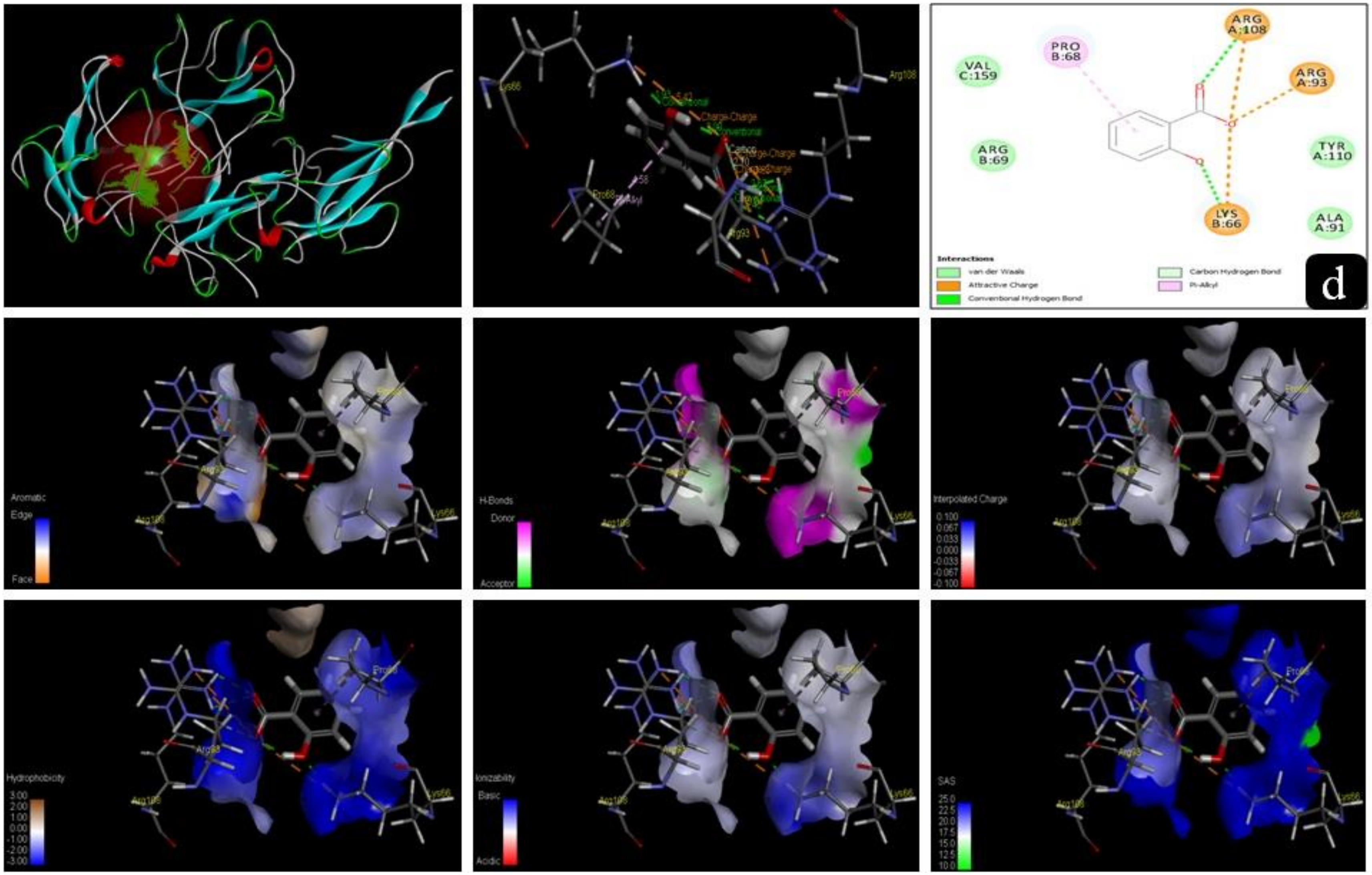

Lys66
Arg108
Pro68
Arg93
Charge-Charge
Conventional
Carbon
VAL C:159
PRO B:68
ARG A:108
ARG A:93
ARG B:69
TYR A:110
LYS B:66
ALA A:91
Interactions
van der Waals
Attractive Charge
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Alkyl
d
Aromatic
Edge
Face
H-Bonds
Donor
Acceptor
Interpolated Charge
0.100
0.067
0.033
0.000
-0.033
-0.067
-0.100
Hydrophobicity
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
Ionizability
Basic
Acidic
SAS
25.0
22.5
20.0
17.5
15.0
12.5
10.0

Figure 3. Receptor-ligand interaction of SARS-CoV-2 nucleocapsid protein N-terminal RNA binding domain with 4-aminobenzoic acid (a), 4-hydroxybenzoic acid (b), benzoic acid (c), and Salicylic acid (d)