ORIGINAL PAPER

Novel Molecules derived from 3-O-(6-galloylglucoside) inhibit Main Protease of SARS-CoV 2 In Silico

Haruna Isiyaku Umar[1,2] · Adeola Ajayi[1,2] · Ridwan Opeyemi Bello[2,3] · Hafsat Olateju Alabere[2,4] · Afees Akinbode Sanusi[5] · Olamide Olusegun Awolaja[1] · Mohammed Mansour Alshehri[6] · Prosper Obed Chukwuemeka[2,3]

Abstract

The ongoing pandemic caused by the severe acute respiratory syndrome 2 (SARS-CoV 2) has led to more than 168 million confirmed cases with 3.5 million deaths as at 28th May, 2021 across 218 countries. The virus has a cysteine protease called main protease (Mpro) which is significant to it life cycle, tagged as a suitable target for novel antivirals. In this computer-assisted study, we designed 100 novel molecules through an artificial neural network-driven platform called LigDream (https://playmolecule.org/LigDream/) using 3-O-(6-galloylglucoside) as parent molecule for design. Druglikeness screening of the molecules through five (5) different rules was carried out, followed by a virtual screening of those molecules without a single violation of the druglike rules using AutoDock Vina against Mpro. The in silico pharmacokinetic features were predicted and finally, quantum mechanics/molecular mechanics (QM/MM) study was carried out using Molecular Orbital Package 2016 (MOPAC2016) on the overall hit compound with controls to determine the stability and reactivity of the lead molecule. The findings showed that eight (8) novel molecules violated none of the druglikeness rules of which three (3) novel molecules (C33, C35 and C54) showed the utmost binding affinity of −8.3 kcal/mol against Mpro; C33 showed a good in silico pharmacokinetic features with acceptable level of stability and reactively better than our controls based on the quantum chemical descriptors analysis. However, there is an urgent need to carry out more research on these novel molecules for the fight against the disease.

Keywords 3-O-(6-galloylglucoside) · Artificial neural network-driven platform · LigDream · Main protease · Novel compounds · Quantum mechanics · SARS-CoV 2

✉ Haruna Isiyaku Umar
uhumar@futa.edu.ng; ariwajoye3@gmail.com

Adeola Ajayi
ajayideola21@gmail.com

Ridwan Opeyemi Bello
belloridwanopeyemi@gmail.com

Hafsat Olateju Alabere
alaberehafsat@gmail.com

Afees Akinbode Sanusi
sanusiafees@gmail.com

Olamide Olusegun Awolaja
molecools@gmail.com

Mohammed Mansour Alshehri
M.Mansouralshehri@outlook.sa

Prosper Obed Chukwuemeka
chukwuemekamcb154271@futa.edu.ng;
pochukwuemeka94@gmail.com

1 Department of Biochemistry, School of Life Sciences, Federal University of Technology, Along Owo-Ilesha Express way, P. M. B. 704, Akure, Ondo State, Nigeria

2 Computer-Aided Therapeutic Discovery and Design Group, FUTA, Akure, Nigeria

3 Department of Biotechnology, School of Life Sciences, Federal University of Technology, Akure, Nigeria

4 Department of Microbiology, School of Life Sciences, Federal University of Technology, Akure, Nigeria

5 Department of Chemistry, School of Physical Sciences, Federal University of Technology, Akure, Nigeria

6 Department of Pharmaceutical Care, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

Abbreviations

3O6G	3-O-(6-galloylglucoside)
ADMET	Adsorption, distribution, metabolism, excretion and toxicity
HOMO	Highest occupied molecular orbital
LUMO	Lowest unoccupied molecular orbital
Mpro	Main protease
PID	PubChem Identifier
PM7	Parametric method 7
SARS-CoV 2	Severe Acute Respiratory Syndrome 2

Introduction

Early 2020, the World Health Organization (WHO) acknowledged the most recent coronavirus (COVID-19) a pandemic and a globewide emergency that causes a global outbreak after its emergence from Wuhan, China in December of 2019 (Kamaz et al. 2020; Anjorin 2020). The ongoing pandemic has resulted in more than 168 million confirmed cases with 3.5 million deaths (as at 28th May, 2021) across 218 countries (Sakurai et al. 2021). The culprit virus that is devastating our world is the severe acute respiratory syndrome 2 (SARS-CoV 2), with a spherical shape, single-stranded, and positively sensed RNA virus. The virus has a genome length of approximately 30,000 base pairs consisting of about eleven open reading frames (ORFs) that encodes numerous proteins (structural and non-structural) that are of importance to the viral life cycle (Anjorin 2020; Alazmi and Motwalli 2020; Yoshimoto 2020). Once SARS-CoV 2 gains entry into the host cell, the viral RNA is liberated into the cytoplasm which results in the expression of replicase gene. The replicase gene accommodates two overlapping open reading frames, ORF1a/1b which covers two-third of the viral genome. The ORFs undergo expression to yield polyprotein 1a and 1ab. These polyproteins are biochemically sliced by two cysteine-rich proteinases known as 3-chymotrypsin-like protease or main protease (Mpro/3CLpro) and papain-like protease (PLpro) emancipating 16 non-structural proteins (nsp1 to 16).

3-Chymotrypsin-like protease (nsp5) is a three domain (I to III) protease that is of significance for the sustenance of the viral life cycle since its maturation leads to the production of nsp4 to nsp16 after self-cleaving from the precursor polyprotein (Dai et al. 2020; Jin et al. 2020). The cleft between domains I and II features a noncanonical Cys-His dyad that recognizes amino acids in substrates from the N terminus to C terminus (Dai et al. 2020; Jin et al. 2020). In it active form, is homodimeric with two protomers (A and B) (Jin et al. 2020). There are no homologs of Mpro present in humans, making it an ideal target for antiviral drug design. Additionally, the role it plays in the replication and proliferation of SARS-CoV 2 cannot be silenced. The process of

discovering and developing drugs is time-consuming and cost-effective, thus the need for the use of computer-aided therapeutic discovery (CATD) which embraces structure-based, system-based and ligand-based therapeutic design (Dai et al. 2020; Romano and Tatonetti 2019; Prieto-martı et al. 2019; Umar et al. 2021a). Today, the key roles played by computational methods in therapeutic design, discovery and development expedition cannot be overemphasized because of its numerous dimensional usage for assemblage of data, processing it before evaluation and interpretation (Dai et al. 2020; Romano and Tatonetti 2019; Prieto-martı et al. 2019; Umar et al. 2021a).

From our erstwhile computational study, we showed that 3-O-(6-galloylglucoside) serves as potential inhibitor of SARS-CoV 2's main protease. However, this compound was found to have poor druglike properties that might affect it down the drug discovery pipeline even though it showed a good binding score with main protease of SARS-CoV 2 (Umar et al. 2021a). In this current study, we designed 100 novel molecules through an artificial neural network-driven platform called LigDream (https://playmolecule.org/LigDream/) (Skalic et al. 2019) using 3-O-(6-galloylglucoside) as parent molecule for design. These new compounds were subjected to inflexible druglikeness screening to select those that could serve as oral drugs. Then a virtual screening was implemented to get the best hit compounds against Mpro. The hit compounds were examined for their pharmacokinetic properties in silico. Finally, quantum mechanics/molecular mechanics study was carried out on the overall hit compound with controls to determine the stability and reactivity of the lead molecule.

Materials and methods

Ligand selections

The ligands selected for this current in silico study are 3-O-(6-galloylglucoside) (seed molecule, PID: 73,157,749), remdesivir (control antiviral drug, PID: 121,304,016) and novel compounds (Fig S1 and Table S1). Three-dimensional conformers of the seed molecule and control drugs in structure data format (SDF) are sourced from a chemical repository server known as PubChem (https://pubchem.ncbi.nlm.nih.gov/compound/).

Generation of novel compounds from 3-O-(6-galloylglucoside)

A total of 100 novel compounds were generated using an artificial neural network-driven platform called LigDream module of playmolecules server (https://playmolecule.org/LigDream/) (Skalic et al. 2019). The SMILE string of

3-O-(6-galloylglucoside) was uploaded to the server and run to generate 100 new SMILE strings for different compounds. The platform uses two networks, auto-encoders and captioning networks that could differentially design several compounds (100) starting with a lone seed compound (3-O-(6-galloylglucoside)). Furthermore, these 100 novel compounds were filtered for druglikeness using five rules, viz Lipinski's (Lipinski et al. 2012), Veber's (Veber et al. 2002), Muegge's (Muegge et al. 2001), Egan's (Egan et al. 2000) with Ghose's et al. (1999); and bioavailability score via the SwissAdme server (http://www.swissadme.ch/index.php) (Daina et al. 2017). SMILE Strings of the novel molecules was uploaded onto the server and was run to evaluate their druglikeness. Those compounds which show no violation to the five rules and a bioavailability score of 0.55 and above were considered for the computer-aided molecular docking against main protease of SARS-CoV 2.

Selection and preparation of protein target

Three-dimensional (3D) structure of main protease (Mpro/nsp5) of SARS-CoV 2 (PDB ID: 6LU7) (Jin et al. 2020) was retrieved from the Protein Database (PDB) (http://www.pdb.org/pdb). Furthermore, the protein was free from all heteroatoms and consequently minimized using protein preparation and minimization tools in Cresset Flare© software, version 4.0 (https://www.cresset-group.com/flare/). The protein minimization was executed under the General Amber Force Field (GAFF), with gradient cutoff of 0.200 kcal/mol/A and iterations was set to 2000 iterations (Stroganov et al. 2011).

Protocol validation of virtual docking steps

Authentication of molecular docking step is required through step validation as done earlier (Umar et al. 2021a, 2021b) to corroborate its exactitude and consistency. Our intent is to replicate the binding posture of a re-docked ligand of a protein that was co-crystallised alongside it. Thus, the co-crystallised ligand (N-leucinamide) was detached from the Mpro and primed for re-docking using Cresset Flare software. The ligand was then re-docked back into Mpro's binding region using Auto Dock Vina integrated in Python Prescription (PyRx) (Trott and Olson 2010). The docked complex was aligned with the cognate crystal structure of Mpro bearing the co-crystalized ligand to acquire the root mean square deviation (RMSD) value in PyMOL.

Preparation of druglike novel compounds for virtual docking

The novel compounds that show druglikeness and good bioavailability score were sketched using MarvinSketch© (ver. 15.11.30) software and transformed into their best energetic and stable configurations using Merck molecular force field (MMFF94) (Halgren 1996) using Open babel integrated within Python Prescription (version 0.8).

Virtual docking

The molecular docking was achieved through flexible docking procedure (Trott and Olson, 2010) previously used by (Umar et al.2021b). PyRx 0.8, a suite integrated with Auto Dock Vina, was utilized for the molecular docking study. The specific target site for the receptor corresponding to the substrate-binding region was adjusted using the grid box with dimensions ($18.08 \times 26.45 \times 26.30$) Å, and the centre was attuned based on the site of substrate binding in the protein consisting of the following amino acids; Thr25, Thr26, His41, Cys44, Met49, Tyr54, Phe140, Leu141, Gly143, Cys145, Asn142, His163, His164, Met165, Ser144, Glu166, Pro168, His172, Val186, Asp187, Arg188, Gln189, Phe185, Thr190, and Gln192 (Dai et al. 2020; Jin et al. 2020; Umar et al. 2021a). The compounds with docking score similar to that of the Seed molecule and control drug at the end of the experiment, were subjected to molecular interaction analysis with the aid of PyMOL© Molecular Graphics (version 2.4, 2016, Shrodinger LLC) and LigPlot[+] (Laskowski and Swindells 2011).

In silico ADMET prediction

ADMET (Adsorption, Distribution, Metabolism, Excretion and Toxicity) is important to analyze the pharmacokinetics of the proposed molecule which could be used as a drug. ADMETSar server was used to predict the ADMET properties of the compounds with the best hits after molecular docking analysis (Cheng et al. 2012; Yang et al. 2018). SMILES of the ligands from PubChem (https://pubchem.ncbi.nlm.nih.gov/compound/) were uploaded onto the search bar of the servers and were predicted.

Quantum mechanics/molecular mechanics analysis

Quantum reactivity descriptors of a lead molecule, seed molecule and control drug was calculated using the Molecular Orbital Package (MOPAC2016). The computations were executed through the semi-empirical method Parametric Method 7 (Kishor and Bhoop 2013). The output file generated from the geometric optimization of the molecules was used to calculate quantum reactivity descriptors such as Highest Occupied Molecular Orbital (HUMO), Lowest Unoccupied Molecular Orbital (LUMO), Energy Gap, Chemical hardness and softness, Molecular surface electrostatic potential (MEP) and electronic energy.

Results and discussion

Generation of novel compounds and druglikeness screening

We used 3-O-(6-galloylglucoside) to produce 100 new compounds through LigDream platform (https://playmolecule.org/LigDream/). The significance of this platform to solve common drug discovery problems is tied to their potency to prevent over utilization of resources (Skalic et al. 2019). These new compounds show new scaffolds and functional groups which covers new site of chemical space that upholds lead-like characteristics (Supplementary Figure S1). The outcome of the inflexible druglikeness screening indicated that eight (8) new molecules scaled through without violating any of Lipinski's, Ghose's, Veber's, Egan's and Muegge's rules; also showing an Abbot bioavailability score of 0.55 (Table 1). Druglikeness is defined as a qualitative valuation that offer the chance for a compound to be an oral drug with reverence to bioavailability (Daina et al. 2017). In the early stage of drug discovery voyage, these valuations are routinely deployed to filter chemical libraries [in our case our chemical library is made of 100 novel molecules from 3-O-(6-galloylglucoside)] to expunge those molecules with properties that are discordant with an acceptable pharmacokinetics profile. This is an indication that the 8 novel molecules can serve as orally active drug.

New compounds bind to main protease of SARS CoV 2

We employed virtual docking to define potential binding and interactions between 8 new molecules, 3-O-(6-galloylglucoside) and remdesivir with SARS CoV 2's main protease. From our findings, 3O6G had a binding affinity of −8.4 kcal/mol while remdesivir displayed a binding affinity of −8.2 kcal/mol (Fig. 1 and Table 2). Also, the novel molecules (C1, C2, C17, C23, C30, C33, C35 and C54) showed good binding (Figs. 1, 2) although molecules C33, C35 and C54 showed the utmost binding affinity of -8.3 kcal/mol. The molecular interaction studies using PyMOL and LigPlot[+] of our compounds showed that they have interactions with amino acid residues domiciled within the binding domain for substrates in Mpro (Figs. 1, 2). Of note is the ability of our compounds to be able to interact with His41 and Cys145 either by hydrogen bonding or hydrophobic interaction. This interaction can be key to stopping the activity of Mpro for processing the replicase polyprotein and subsequently blocking the maturation of the virus as these two amino acid residues are key in interacting with the enzymes normal substrate.

The observation of 3-O-(6-galloylglucoside) binding to SARS-CoV 2 main protease from this current study is in agreement with the observation made in our previous study that looked at binding of gallic acid derivatives with five non-structural proteins (nsps) of SARS-CoV 2 (Umar et al. 2021a). (Das et al. 2020) have demonstrated that most molecules docked against SARS-CoV 2's main protease interacted with His41 and Cys145. Similarly, the in silico study of Umar et al. (2021) showed that Mangiferin, binded to Mpro by producing interactions with key amino acid residues significant in stopping the activity of Mpro especially the catalytic dyad residues, His41 and Cys145.

Our findings in this current in silico work through the molecular interaction fingerprints of the lead molecules are in tandem with those of previous studies (Kamaz et al. 2020; Jin et al. 2020; Putu et al. 2020; Zhao et al. 2020).

In silico ADMET prediction

For a drug to be effective, a potent small molecule should reach its target in the body in ample concentration, and remain there in its bioactive nature for a long period to exact its therapeutic influence. In line with this, it is pertinent to assess molecules early on in the drug development stage for their absorption, distribution, metabolism, excretion, and toxicity (ADMET) parameters as this effort will considerably reduce attrition. In our current in silico work, we assessed the ADMET-linked parameters of three best hit molecules (C33, C35 and C54), 3-O-(6-galloylglucoside) and remdesivir computationally using the ADMETSAR server. Our findings are presented in Table 3. The new compounds showed a good ADMET properties than our seed molecule (3O6G) and remdesivir especially C33. This feat could be linked to the earlier outcome from the inflexible druglikeness screening. Commendably, C33 shows none inhibition to some key liver enzymes (Cytochrome P450 isoforms), human-ether-a-go-go gene (i.e. might not cause prolonged QT interval), and P-glycoprotein transporter, in addition to non-carcinogenic and non-mutagenic aptitudes as predicted. Furthermore, C33 was predicted to be orally available and could be absorbed intestinally.

Quantum reactivity analysis

The quantum reactivity descriptors of our lead molecule (C33), 3-O-(6-galloylglucoside) and remdesivir were obtained from MOPAC2016 software. We were able to generate the energy of the highest occupied molecular orbital ($_E$HOMO) and the energy of the lowest unoccupied molecular orbital ($_E$LUMO), which describes the charge distribution of frontier molecular orbitals of our compounds (Fig. 3A–C). From this, the blue color portrays the negative periods, while the red color indicates the positive periods.

Table 1 Screening of 3-O-(6-galloylglucoside) and 100 new molecules derived using the LigDream online tool

Molecule	Formula	MW	TPSA	Lipinski #violations	Ghose #violations	Veber #violations	Egan #violations	Muegge #violations	Bioav. Score
3O6G	C20H20O14	484.36	243.9	2	2	1	1	3	0.11
1	**C18H18F4N4O5**	**446.35**	**105.78**	**0**	**0**	**0**	**0**	**0**	**0.55**
2	**C16H10F6N2O4**	**408.25**	**87.66**	**0**	**0**	**0**	**0**	**0**	**0.55**
3	C20H25N5O10	495.44	205.66	1	2	2	1	2	0.55
4	C15H19N5O10S	461.4	244.66	2	1	2	1	4	0.17
5	C19H20N2O11	452.37	204.97	2	1	2	1	3	0.11
6	C18H20N4O10S	484.44	224.59	1	1	2	1	2	0.11
7	C16H17N5O9S	455.4	228.25	1	1	2	1	2	0.11
8	C16H19N5O9S	457.42	220.41	1	0	2	1	2	0.11
9	C17H19N3O10S	457.41	201.73	1	0	2	1	2	0.11
10	C17H14F2N4O8	440.31	176.92	1	0	1	1	2	0.11
11	C17H13F2N3O9	441.3	195.38	2	0	1	1	3	0.11
12	C17H10F6N2O5	436.26	104.73	0	0	0	0	1	0.56
13	C17H12F3N3O8	443.29	175.15	1	0	1	1	2	0.11
14	C18H15F2N5O7	451.34	179.84	1	1	1	1	2	0.11
15	C17H15F2N5O8S	487.39	205.29	1	1	1	1	2	0.11
16	C18H16F2N4O7	438.34	159.97	1	0	1	1	2	0.11
17	**C17H11F5N2O5**	**418.27**	**104.73**	**0**	**0**	**0**	**0**	**0**	**0.56**
18	C17H10F6N2O5	436.26	115.73	0	0	0	0	1	0.56
19	C17H10F3N3O8	441.27	182.99	1	0	1	1	2	0.11
20	C17H15F3N2O8S	464.37	184.63	0	0	1	1	2	0.11
21	C17H13F2N3O8	425.3	159.27	1	0	1	1	2	0.11
22	C16H13F2N5O8	441.3	198.18	1	1	1	1	2	0.11
23	**C18H14F4N2O6**	**430.31**	**124.96**	**0**	**0**	**0**	**0**	**0**	**0.56**
24	**C17H11F5N2O5**	**418.27**	**115.73**	**0**	**0**	**0**	**0**	**0**	**0.56**
25	C18H13F5N2O6	448.3	124.96	0	0	0	0	1	0.56
26	C19H16F3N3O8	471.34	163.29	1	0	2	1	2	0.11
27	C17H10F6N2O6	452.26	124.96	0	0	0	0	1	0.56
28	C17H13F4N3O7S	479.36	173.27	0	0	1	1	2	0.11
29	C19H16F4N2O7	460.33	134.19	0	0	1	1	1	0.56
30	**C18H14F4N2O6**	**430.31**	**124.96**	**0**	**0**	**0**	**0**	**0**	**0.56**
31	C19H14F6N2O7	496.31	134.19	0	1	1	1	1	0.56
32	C17H11F5N2O6	434.27	124.96	0	0	0	0	1	0.56
33	**C19H14F5N3O5**	**459.32**	**121.96**	**0**	**0**	**0**	**0**	**0**	**0.56**
34	C19H16F2N2O8	438.34	143.42	0	0	1	1	0	0.56
35	**C19H13F5N2O5**	**444.31**	**106.94**	**0**	**0**	**0**	**0**	**0**	**0.56**
36	C20H17F4N3O7	487.36	146.05	0	1	2	1	1	0.55
37	C21H20F3N3O8	499.39	163.29	1	1	2	1	2	0.55
38	C18H14F4N2O7	446.31	145.19	0	0	1	1	1	0.56
39	C17H16F2N6O6	438.34	180.74	1	0	1	1	1	0.11
40	C17H13F2N3O9	441.3	179.5	1	0	1	1	2	0.11
41	C16H14F2N6O6	424.32	180.6	1	1	1	1	1	0.11
42	C18H13F5N2O7	464.3	145.19	0	0	1	1	1	0.56
43	C19H13F5N2O6	460.31	116.17	0	0	0	0	1	0.56
44	C18H14F4N2O7	446.31	145.19	0	0	1	1	1	0.56
45	C19H16F5N3O6	477.34	128.98	0	0	0	0	1	0.55
46	C20H22N4O9S	494.48	178.77	1	1	2	1	1	0.55
47	C17H15F2N3O7	411.31	147.08	0	0	1	1	0	0.56
48	C17H17F2N5O7	441.34	185.99	1	0	2	1	2	0.55

Table 1 (continued)

Molecule	Formula	MW	TPSA	Lipinski #violations	Ghose #violations	Veber #violations	Egan #violations	Muegge #violations	Bioav. Score
49	C17H8F8N2O4	456.24	84.5	1	1	0	1	1	0.55
50	C15H12F3N3O8S	451.33	183.27	1	0	1	1	2	0.11
51	C16H12F3N3O7S	447.34	173.27	0	0	1	1	2	0.55
52	C17H13F4N3O8S	495.36	182.5	1	1	2	1	2	0.55
53	C15H12F2N4O7S	430.34	186.16	1	0	1	1	2	0.55
54	**C20H17F4N3O6**	**471.36**	**126.73**	**0**	**0**	**0**	**0**	**0**	**0.55**
55	C17H15F4N3O7	449.31	150.12	0	0	2	1	2	0.55
56	C17H12F4N2O8	448.28	155.16	0	0	1	1	2	0.11
57	C17H13F4N3O7	447.29	160.95	0	0	1	1	2	0.55
58	C17H13F4N3O8	463.29	167.19	1	0	1	1	2	0.11
59	C20H17F5N4O6	504.36	130.09	1	1	0	0	1	0.55
60	C17H11F5N4O7	478.28	166.95	1	0	2	1	2	0.11
61	C17H12F5N5O6	477.3	172.74	1	0	2	1	2	0.55
62	C19H18F2N2O9S	488.42	171.08	1	1	1	1	2	0.55
63	C16H12F3N3O7S	447.34	173.27	0	0	1	1	2	0.55
64	C21H21F2N3O8	481.4	154.5	1	1	2	1	1	0.55
65	C20H19F2N3O8	467.38	166.28	1	0	2	1	1	0.55
66	C18H20N4O10S	484.44	224.85	1	1	2	1	2	0.55
67	C22H23F2N3O7	479.43	115.43	0	0	1	0	0	0.55
68	C20H19F2N3O8	467.38	152.29	1	0	2	1	1	0.55
69	C19H15F3N2O7	440.33	136.4	0	0	0	1	0	0.56
70	C16H13F2N5O7	425.3	177.95	1	0	1	1	2	0.11
71	C18H11F7N2O6	484.28	124.96	0	1	0	0	1	0.56
72	C15H11F5N4O5S	454.33	162.85	0	0	1	1	2	0.55
73	C20H16F6N2O5	478.34	88.1	0	0	0	0	1	0.55
74	C21H21N3O10	475.41	181.75	1	0	2	1	1	0.11
75	C21H23N3O9S	493.49	166.74	1	1	2	1	1	0.55
76	C15H19N5O9S	445.4	221.44	2	1	2	1	3	0.17
77	C17H10F7N3O4	453.27	110.52	0	0	0	0	1	0.55
78	C19H18F3N5O6	469.37	151.65	1	0	2	1	1	0.55
79	C17H12F3N3O8	443.29	175.15	1	0	1	1	2	0.11
80	C16H13F3N4O7	430.29	170.71	1	0	2	1	2	0.55
81	C17H13F2N3O9S	473.36	220.46	1	0	2	1	2	0.11
82	C19H15FN2O10	450.33	199.56	2	0	2	1	3	0.11
83	C19H17ClN4O8	464.81	165.18	1	0	2	1	1	0.55
84	C20H21N3O10S	495.46	186.97	1	1	2	1	1	0.55
85	C16H13F2N3O7S	429.35	183.16	0	0	2	1	1	0.11
86	C19H15F2N3O8	451.33	152.29	1	0	1	1	1	0.11
87	C17H11F5N2O6	434.27	124.96	0	0	0	0	1	0.55
88	C18H14F3N3O8	457.31	174.29	2	0	2	1	3	0.11
89	C19H17F4N5O5	471.36	135.1	0	0	0	1	0	0.56
90	C19H19FN4O8	450.37	171.29	1	0	1	1	1	0.11
91	C17H14FN3O9	423.31	195.38	2	0	1	1	3	0.11
92	C17H14F3N3O7S	461.37	173.27	0	0	1	1	2	0.11
93	C16H14F2N6O5	408.32	160.51	1	0	1	1	1	0.11
94	C18H13F2N3O7	421.31	143.06	0	0	1	1	0	0.55
95	C18H14F2N2O7	408.31	123.19	0	0	0	0	0	0.55
96	C20H19F3N2O8	472.37	145.63	0	0	2	1	1	0.56
97	C17H14F2N4O8	440.31	176.06	1	0	2	1	1	0.11

Table 1 (continued)

Molecule	Formula	MW	TPSA	Lipinski #violations	Ghose #violations	Veber #violations	Egan #violations	Muegge #violations	Bioav. Score
98	C18H20N4O10S	484.44	224.85	1	1	2	1	2	0.55
99	C20H20FN3O10	481.39	181.75	1	1	2	1	2	0.11
100	C17H11F7N4O4	468.28	136.54	0	0	0	1	1	0.55

3O6G = 3-O-(6-galloylglucoside); Those compounds in bold letters passed the stringent druglikeness screening; #violations = number of violations
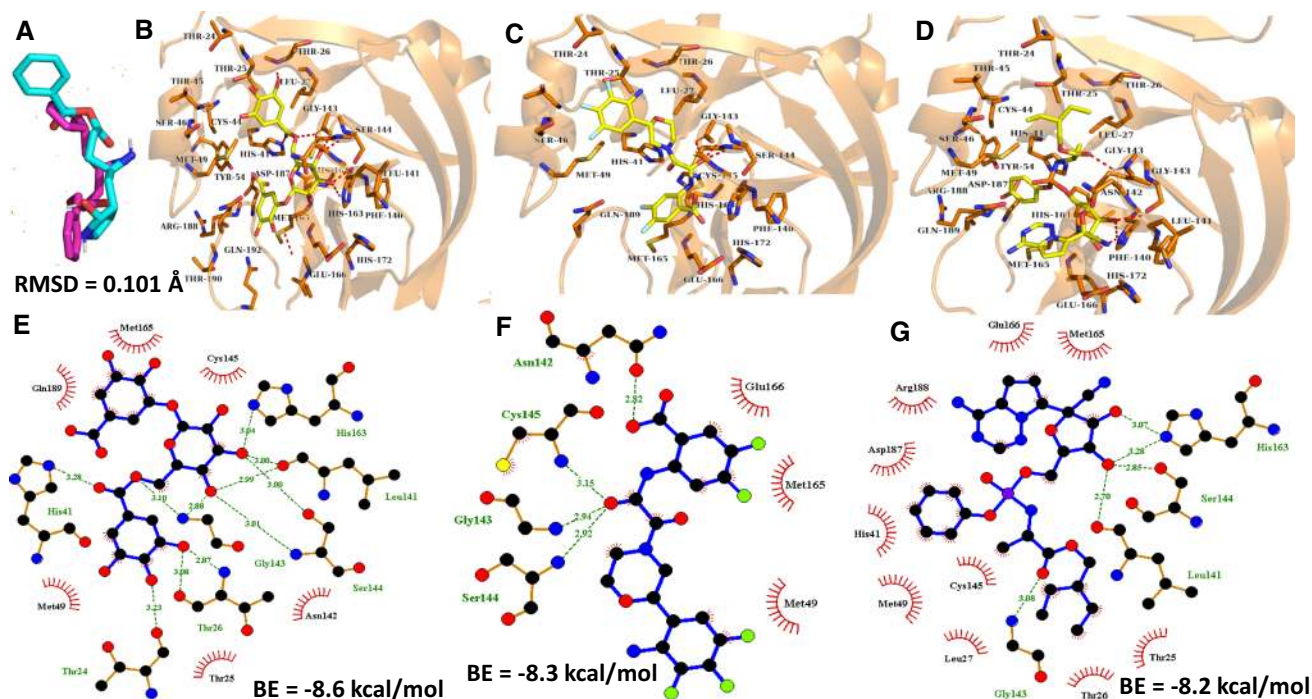
Fig. 1 Virtual Docking of C33, 3-O-(6-galloylglucoside) and remdesivir against Mpro of SARS CoV 2. **a** Validation of virtual docking steps. The superimposition of the re-docked (cyan) and native (pink) Ligands yields an RMSD value of 0.101 Å. The 2D and 3D molecular interactions of 3-O-(6-galloylglucoside) (**b**, **e**), C33 (**c**, **f**), and Remdesivir (**d**, **g**)

Table 2 Molecular interactions of lead ligands with the amino acid residues in the binding region of SARS-CoV 2's Main protease

| S/N	Compounds	Binding affinity (kcal/mol)	No. of residues involved in H-Bond	Residues involved in hydrogen bond (Distance, Å)	Residues involved in hydrophobic interaction
1	3-O-(6-galloylglucoside)	−8.6	7	Leu141 (3.00, 2.99), Ser144 (3.01, 3.00), His163 (3.04), Thr26 (3.08, 2.87), Thr24 (3.23), Gly143 (3.10, 2.88) and His41 (3.28)	Met49, Thr25, Asn142, Cys145, Met165 and Gln189
2	C33	−8.3	4	Ser144 (2.92), Gly143 (2.94), Cys145 (3.15) and Asn142 (2.82)	Asn142, Cys145, Met49, Met165 and Glu166
3	Remdesivir	−8.2	4	Gly143 (3.08), Leu141 (2.70), Ser144 (2.85) and His163 (3.28, 3.07)	Glu166, Met165, Arg188, Asp187, His41, Met49, Cys145, Leu27, Thr26 and Thr25
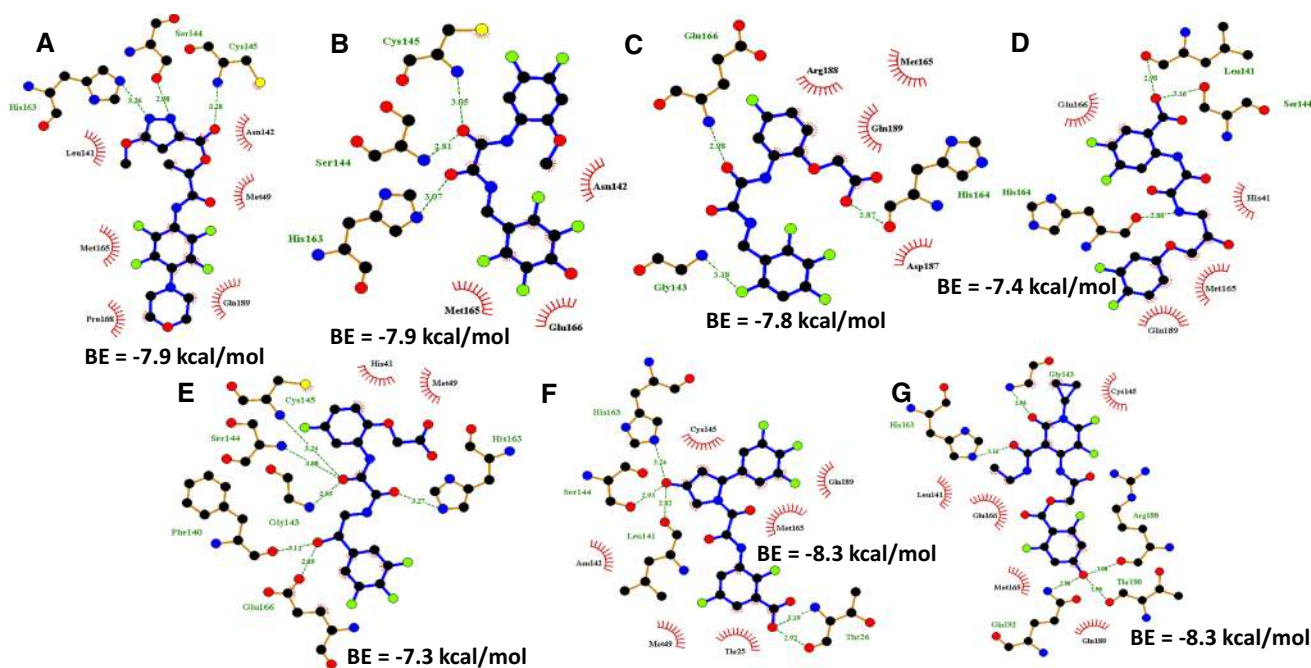
Fig. 2 2D molecular interaction plot of new compounds docked against Mpro. **a** C1, **b** C2, **c** C17, **d** C23, **e** C30, **f** C35 and **g** C54. Ligands are in blue sticks, amino acids interacting through hydrogen bond are in brown sticks, hydrogen bonds are represented in green dashes while hydrophobic interactions are presented as red curved spikes

Thus, we observe that the HOMO and LUMO are frequently located on the aromatic ring, sugar moiety (remdesivir), the amino-, floro-, carbonyl-, phospho- and methylene groups. Similarly, the values of HOMO (the capacity of a molecule to give electrons to an acceptor species is favored by high values) and LUMO (the ability of a molecule to take electrons is preferred by the low values) can determine the stability and reactivity of our compounds (Benbouguerra et al. 2021; Mubarik et al. 2021). C33 was seen to be able to donate electrons (HOMO = −8.898 eV) and also could take electrons (LUMO = −1.636 eV). The energy gap (ΔG) showed that the lead compound is reactive (−7.262 eV) than the seed molecule and control drug (Table 4) and moderately stable. This indicates that there is charge transfer between the atoms of C33. Other parameters termed the global reactivity descriptors such as chemical potential, electrophilicity index, and chemical hardness that is required to analyze the reactivity of the inhibitor molecules were also elucidated. In addition, descriptors such as ionization potential and electron affinity were also determined (Table 4).

The vertical ionization potentials Is and electron affinities as were obtained from the HOMO–LUMO values, which was used to define the global electrophilicity value ω (Table 4). It is a measure of the energy stabilization of the system (Siddiqui et al. 2012). Chemical hardness quantifies the resistance to change in the electron distribution in a collection of nuclei and electrons (Siddiqui et al. 2012; Kalaiarasi and Manivarman 2017). The calculated values of

the chemical potential, chemical hardness and global electrophilicity of C33 are − 5.267 eV, 3.631 eV and 3.820 eV, respectively. The small value of η for C33 indicates that it is relatively soft on the scale of hardness. By designation, the electrophilicity index is a degree of the susceptibility of chemical species to accept electrons (Siddiqui et al. 2012; Kalaiarasi and Manivarman 2017).

Molecular surface electrostatic potential

The active regions on our three compounds were elucidated by measuring their molecular surface electrostatic potentials (MEP) through quantum chemical calculations (Benbouguerra et al. 2021; Mubarik et al. 2021). This evaluation will locate the site of chemical reactivity of our molecules which is represented by the 3D maps at optimized geometry in Fig. 3D–F. The electrostatic potential difference is indicated by color gradients that is valuable in investigating the link between molecular structure and physicochemical property correlation of molecules, biomolecules and drugs inclusive. Based on the color gradients, the positive potentials corresponding to the nucleophilic reaction sites are depicted in blue color, while the negative potentials regions relate to the electrophilic reaction sites illustrated in yellow and red colors. However, the areas of zero potential are presented in green color (Fig. 3D–F). The variance that occurs in the electrostatic potential generated by our molecule is widely responsible for the binding of our molecule to Mpro binding

Table 3 In silico ADMET Profiling

Models	Predictions				
Absorption	33	35	54	3O6G	Remdesivir
Caco-2	−	−	−	−	−
Human Intestinal Absorption	+	−	+	−	+
Human oral bioavailability	+	+	−	−	−
Distribution					
Blood brain barrier	+	+	+	−	+
P-glycoprotein inhibitor	−	−	+	−	+
P-glycoprotein substrate	+	−	+	−	+
Subcellular localization	Mitochondria	Mitochondria	Nucleus	Mitochondria	Lysosomes
Metabolism					
CYP1A2 inhibition	−	−	−	−	−
CYP2C19 inhibition	−	−	−	−	−
CYP2C9 inhibition	−	−	−	−	−
CYP2C9 substrate	−	−	+	−	−
CYP2D6 inhibition	−	−	−	−	−
CYP2D6 substrate	−	−	−	−	−
CYP3A4 inhibition	−	−	−	−	−
CYP3A4 substrate	+	−	+	−	+
CYP inhibitory promiscuity	−	−	+	−	−
OATP1B1 inhibitor	+	+	+	+	+
OATP1B3 inhibitor	+	+	+	−	+
OATP2B1 inhibitor	−	−	−	−	−
OCT1 inhibitor	−	−	−	−	−
OCT2 inhibitor	−	−	−	−	−
MATE1 inhibitor	−	−	−	−	−
UGT catalyzed	−	−	+	+	−
Toxicity					
Hepatotoxicity	+	+	+	+	+
Human either-a-go-go inhibition	−	−	−	−	−
Ames mutagenesis	−	−	−	−	−
Acute oral toxicity (c)	III	III	III	III	III
Carcinogenicity (binary)	−	−	−	−	−

3O6G = 3-O-(6-galloylglucoside); + means Yes; − means No

domain, since the binding region cumulatively is expected to have opposing region of electrostatic potential. MEP of C33 clearly indicates the major negative potential sites cover the fluorine atoms of the phenyl and benzyl rings, and oxygen atoms of the morpholine ring, oxoacetamide chain and carboxylic group that were featured in yellow to red color, as they are the binding region for electrophilic attack. The hydrogen atoms of the amino and carboxylic groups bear the maximum level of positive potential while the rest of the compound seems to possess an almost neutral electrostatic potential (Fig. 4).

The IUPAC name of the lead molecule in our current computational study was generated using ChemSketch and MarvinSketch software and the query from both software returned similar name for the molecule which is 2-({[(2S)-2-(2-amino-3,4,5-trifluorophenyl)morpholin-4-yl](oxo)

acetyl}amino)-4,5-difluorobenzoic acid (See supplementary information).

Conclusion

We found out from this present study that eight (8) molecules from the 100 novel molecules generated from an artificial neural network-based platform showed the ability to serve as oral drugs. Also, C33, C35 and C54 out of these molecules displayed good binding affinity of −8.3 kcal/mol against Mpro. The ADMET profiling indicates that C33 was better than C35 and C54. Finally, the quantum chemical reactivity descriptors analysis showed that C33 was moderately stable and more reactive than the controls. However, there is an urgent need to carry out Molecular Dynamics Simulation (MDS) of at least 100 ns, synthesis of the lead
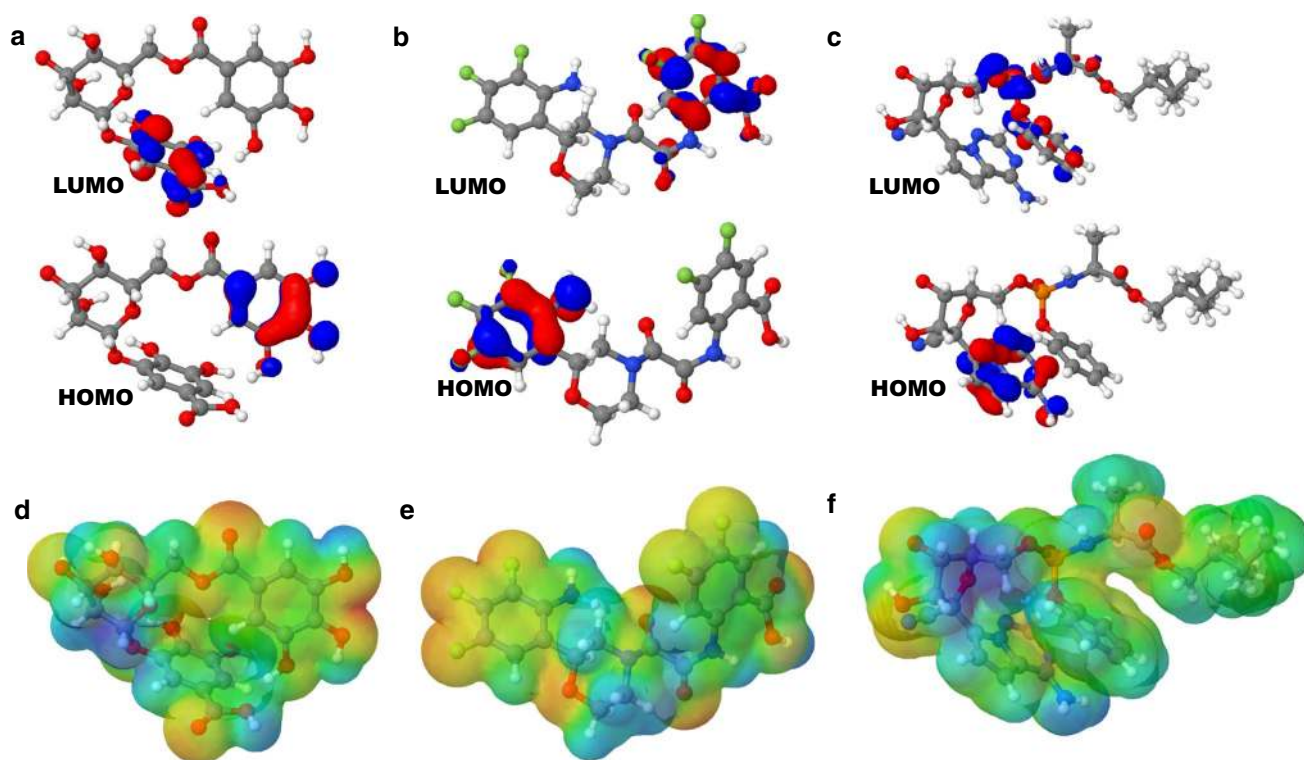
Fig. 3 Chemical reactivity descriptors based on HOMO–LUMO analysis of **a** 3-O-(6-galloylglucoside), **b** C33 and **c** Remdesivir. The molecular surface electrostatic potential (MEP) maps of **d** 3-O-(6-galloylglucoside), **e** C33 and **f** Remdesivir

Table 4 Calculated quantum reactivity descriptors of the lead, 3-O-(6-galloylglucoside) and Remdesivir using PM7 Hamiltonian method

SN	Quantum chemical property	Lead (C33)	3-O-(6-galloylglucoside)	Remdesivir
1	HOMO	− 8.898 eV	− 9.083 eV	− 8.728 eV
2	LUMO	− 1.636 eV	− 1.134 eV	− 0.497 eV
3	Energy Gap (ΔE)	− 7.262 eV	− 7.949 eV	− 8.231 eV
4	Ionization potential (I)	8.898 eV	9.083 eV	8.728 eV
5	Electron affinity (A)	1.636 eV	1.134 eV	0.497 eV
6	Chemical hardness (η)	3.631 eV	3.975 eV	4.116 eV
7	Chemical softness (ζ)	0.275 (eV)$^{-1}$	0.252 (eV)$^{-1}$	0.243 (eV)$^{-1}$
8	Electronegativity (χ)	5.267 eV	5.109 eV	4.613 eV
9	Chemical potential (μ)	− 5.267 eV	− 5.109 eV	− 4.613 eV
10	Electrophilicity index (ω)	3.820 eV	3.283 eV	2.585 eV
11	Dipole moment (M)	6.090 a.u	2.221 a.u	8.419 a.u

$\Delta E = HOMO\varepsilon - LUMO\varepsilon$, $I = -E_{HOMO}$, $A = -E_{LUMO}$, $\eta = (I - A)/2$, $\zeta = 1/\eta$, $\chi = (I + A)/2$, $\mu = -(I + A)/2$, $\omega = \mu^2/2\eta$

molecule, and follow-up with extensive experimental studies to ascertain its efficacy against this ongoing pandemic.

2-{2-[(2S)-2-(2-amino-3,4,5-trifluorophenyl)morpholin-4-yl]-2-oxoacetamido}-4,5-difluorobenzoic acid

Fig. 4 Structure of the lead molecule: 2-({[(2S)-2-(2-amino-3,4,5-tri-fluorophenyl)morpholin-4-yl](oxo)acetyl}amino)-4,5-difluorobenzoic acid

Declarations

Conflict of interest The authors declare no known competing financial interests or no personal relationships that could have appeared to influence the work reported in this paper.

References

Alazmi M, Motwalli O (2020) In silico virtual screening, characterization, docking and molecular dynamics studies of crucial SARS-CoV-2 proteins. J Biomol Struct Dyn [Internet] 0(0):1–11. https://doi.org/10.1080/07391102.2020.1803965

Anjorin AA (2020) The coronavirus disease 2019 (COVID-19) pandemic : a review and an update on cases in Africa. Acian Pacific J Trop Med 13:1–6

Benbouguerra K, Chafai N, Chafaa S, Islamtouahria Y, Tlidjane H (2021) New α-hydrazinophosphonic acid: synthesis, characterization, DFT study and in silico prediction of its potential inhibition of SARS-CoV-2 main protease. J Mol Struct [Internet] 130480. https://doi.org/10.1016/j.molstruc.2021.130480

Cheng F, Li W, Zhou Y, Jie S, Wu Z, Liu G et al (2012) admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. J Chem Inf Model 52:3099–3105

Dai W, Zhang B, Jiang X-M, Su H, Li J, Zhao Y et al (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science (80-) 368(6497):1331–5.

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug- likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:1–13

Das S, Sarmah S, Lyndem S, Roy AS (2020) An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. J Biomol Struct Dyn [Internet] 0(0):1–11. https://doi.org/10.1080/07391102.2020.1763201

Egan WJ, Merz KM, Baldwin JJ (2000) Prediction of drug absorption using multivariate statistics. J Med Chem 43:3867–3877

Ghose AK, Viswanadhan VN, Wendoloski JJ (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. J Comb Chem 1(1):55–68

Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17(5):490–519

Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y (2020) Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. Nature [Internet] 582(February):1–24. https://doi.org/10.1038/s41586-020-2223-y

Kalaiarasi N, Manivarman S (2017) AB initio (DFT) and vibrational studies of the synthesized heterocyclic compound 2-6-oxo-2- thi-oxotetrahydropyrimidin-41H-ylidene hydrazine carbothioamide. J New Dev Chem 1(2):100–110

Kamaz Z, Al-jassani MJ, Umar H (2020) Screening of common herbal medicines as promising direct inhibitors of Sars-Cov-2 in silico. Annu Res Rev Biol 35(8):53–67

Kishor A, Bhoop S (2013) Theoretical Studies of Vibrational Spectral modes and HOMO, LUMO Studies of Some Synthetic Organic Compounds. Res J Chem Sci 3(10):36–43

Laskowski RA, Swindells MB (2011) LigPlot + : multiple ligand À protein interaction diagrams for drug discovery. J Chem Inf Model 51:2778–2786

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 64(SUPPL.):4–17

Mubarik A, Rasool N, Hashmi MA, Mansha A, Zubair M, Shaik MR et al (2021) Computational study of structural, molecular orbitals. Opt Derivatives Cryst 11(211):1–14

Muegge I, Heald SL, Brittelli D (2001) Simple selection criteria for drug-like chemical matter. J Med Chem 44(12):1841–1846

Prieto-martı FD, López-López E, Juárez-Mercado KE, Medina-Franco JL. Computational Drug Design Methods—Current and Future Perspectives. In: In Silico Drug Design. 2019. p. 19–44.

Putu N, Laksmiani L, Putu L, Larasanty F, Agung A, Jaya G (2020) Active compounds activity from the medicinal plants against SARS-CoV-2 using in silico assay. Biomed Pharmacol J 13(2):873–881

Romano JD, Tatonetti NP (2019) Informatics and computational methods in natural product drug discovery : a review and perspectives. Front Genet 10(April):1–16

Sakurai Y, Myat MTN, Kurosaki Y, Sakura T, Inaoka DK, Fujine K et al (2021) 5-amino levulinic acid inhibits SARS-CoV-2 infection in vitro. Biochem Biophys Res Commun [Internet] 545:203–207. https://doi.org/10.1016/j.bbrc.2021.01.091

Schaduangrat N, Lampa S, Simeon S, Gleeson MP, Spjuth O, Nantasenamat C (2020) Towards reproducible computational drug discovery. J Cheminform [Internet] 12:1–30. https://doi.org/10.1186/s13321-020-0408-x

Siddiqui SA, Rasheed T, Faisal M, Pandey AK, Khan SB (2012) Electronic structure, nonlinear optical properties, and vibrational analysis of gemifloxacin by density functional theory. Spectrosc an Int J 27(3):185–206

Skalic M, Jimenez J, Sabbadin D, De FG (2019) Shape-based generative modeling for de Novo drug design. J Chem Inf Model 59:1205–1214

Stroganov OV, Novikov FN, Zeifman AA, Stroylov VS, Chilov GG (2011) TSAR, a new graph-theoretical approach to computational modeling of protein side-chain flexibility: modeling of ionization properties of proteins. Proteins Struct Funct Bioinform 79(9):2693–2710

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31(2):455–461

Umar HI, Josiah SS, Saliu TP, Jimoh TO, Ajayi A, Danjuma JB (2021) In-silico analysis of the inhibition of the SARS-CoV-2 main protease by some active compounds from selected African plants. J Taibah Univ Med Sci 16(2):162–76. https://doi.org/10.1016/j.jtumed.2020.12.005

Umar HI, Siraj B, Ajayi A, Jimoh TO, Chukwuemeka PO (2021a) Molecular docking studies of some selected gallic acid derivatives against five non-structural proteins of novel coronavirus. J Genet Eng Biotechnol 19:1–14

Umar HI, Olatunde I, Abegunde SM, Igbe FO, Siraj B (2021b) In silico molecular docking of bioactive molecules isolated from Raphia taedigera seed oil as potential anti-cancer agents targeting vascular endothelial growth factor receptor-2. Chem Africa 4(1):161–174

Veber DF, Johnson SR, Cheng H, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623

Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z et al (2018) admetSAR 2. 0 : web-service for prediction and optimization of chemical ADMET properties. Bioinformatics (August):1–2

Yoshimoto FK (2020) The proteins of severe acute respiratory syndrome coronavirus - 2 (SARS CoV - 2 or n - COV19), the cause of COVID - 19. Protein J [Internet] 39(3):198–216. https://doi.org/10.1007/s10930-020-09901-4

Zhao Y, Hu J, Song J, Zhao X, Shi Y, Jiang Y (2020) Exploration on Shufeng Jiedu capsule for treatment of COVID-19 based on network pharmacology and molecular docking. Chin Med 11:9–18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.