Structure-Based Screening of Potential Triterpenoids for Anti-Viral activity against SARS-CoV-2 and related Coronaviruses

J. Vijaya Sundar*[,a] and Thiagarayaselvam Aarthy

[a] *Department of Chemistry, Indian Institute of Technology Madras, India*

Abstract

Triterpenoids possess valuable medicinal properties ranging from anti-microbial to anti-cancer. Some of them were known for their activity against HIV, Ebola and Influenza viruses. In this study, 108 triterpenoids were screened for its potential usage as anti-viral drug against SARS-CoV-2, MERS and SARS coronaviruses, using molecular docking calculation against their main proteases (M^{pro}) and ADME based drug-likeliness parameters. Among these triterpenoids, thirteen were found to possess ADME properties in optimal region and binding energies > -8.5 kcal mol^{-1} with M^{pro}. The thirteen triterpenoids are obacunone, citrusin, nomilin, epoxyazadiradione, secomahoganin, rutaevin, picrasin A, limonin, gedunin, glaucin B, isoobacunoic acid, 7-deacetylgedunin and andirobin. The H-bond and hydrophobic interactions for each of these triterpenoids with the important active site residues of M^{pro} (Cys145, His41, Glu166) were analyzed. The results revealed that citrus limonoids such as citrusin, obacunone, glaucin B, limonin and rutaevin as well as secomahoganin (found in meliaceae family) are potential inhibitors of M^{pro} of SARS-CoV-2 based on their non-covalent bonding to at least two of the active site amino acid residues. The order of effectiveness of these triterpenoids in inhibiting M^{pro} of SARS-CoV-2 was found to be citrusin > obacunone ≈ glaucin B > secomahoganin > limonin ≈ rutaevin > nomilin ≈ gedunin ≈ isoobacunoic acid > picrasin A > epoxyazadiradione > 7-deacetylgedunin ≈ andirobin, based on the number of active site residues involved in the bonding and the type of interaction. These triterpenoids also inhibit the active site of M^{pro} of SARS and MERS CoVs, however, in different order. The binding energy varies from -9.26 to -7.17 kcal mol^{-1} for MERS and from -10.51 to -7.38 kcal mol^{-1} for SARS coronaviruses, indicating active inhibition. Nevertheless, binding mode analyses showed that secomahoganin and nomilin binds directly to active site residues and potentially inhibits protease activity of MERS and SARS coronoviruses. Our study presents the potential triterpenoids as drug leads which needs further evaluation through cell culture studies and drug trials against coronaviruses.

AUTHOR INFORMATION

Corresponding Author

*****Email: vijayscien@gmail.com

Introduction

Coronavirus disease 2019 (COVID-19), first identified in Wuhan, China was found to spread worldwide in 2020. The etiological factor for COVID-19 is Novel coronavirus SARS-CoV-2, which causes respiratory illness with pneumonia like symptoms and acute respiratory distress syndrome (ARDS) in humans[1]. SARS-CoV-2 is an enveloped, single stranded RNA virus, which belongs to Coronaviridae family. So far, six human pathogenic betacoronaviruses (HCoV) belonging to this family other than SARS-CoV-2 were identified[1] . COVID-19 is the third epidemic outbreak caused by CoV, after Severe acute respiratory syndrome (SARS) due to SARS-CoV infections in 2002[2] and, Middle East respiratory syndrome (MERS) in 2012[3].

SARS-CoV-2 genome is comprised of 14 Open Reading Frames (ORF), which encodes for a total of 26 proteins that includes both structural and non-structural proteins[4],[5]. The four main structural proteins are spike glycoprotein (S), small envelope glycoprotein (E), membrane glycoprotein (M), and nucleocapsid protein (N). Non-structural proteins (nsps) are generated as two overlapping polyproteins, pp1a and pp1ab, which are processed into individual nsp1-16, performing many of the vital processes in virus and host cells[6] . To exemplify, nsp12 and nsp13 functions as RNA-dependent RNA polymerase (RDRP) and helicase involved in the replication of RNA, respectively[4]. Nsp3 and nsp5 of ORF 1a protease are papain-like cysteine protease (PL^{pro}) and 3-chymotrypsin-like cysteine protease ($3CL^{pro}$), also called main protease (M^{pro}), respectively. They are indispensable for the replication and transcription processes of coronavirus[6]. Therefore, inhibition of the protease affects the replication of the viruses in the host cell[7]. Moreover, the enzyme proteins exhibit considerable sequence conservation among different coronaviruses than structural proteins (S, E, M, N), [2] and structural proteins are unlikely to be considered for drug target. Taken together the crucial role played by main protease (M^{pro}), in the mediation of viral replication and absence of the enzyme homologues in human, M^{pro} is considered as a befitting drug target [3].

M^{pro} is active as a homodimer and is composed of two protomers. Each protomer has three-domain, domain I and II are found as antiparallel β-barrel and domain III is formed of five α-helices in the form of antiparallel globular cluster[8] . The enzyme has non-canonical Cys-His dyad in the active site, which is found between domains I and II. The catalytic site residues are highly conserved among SARS-CoV-2, SARS-CoV and MERS-CoV. So far, wide-spectrum inhibitors targeting M^{pro} of coronavirus based on its structure and substrate analogy have been synthesized and screened[9,10,11]. Further trials and optimization will help to discover a candidate of good clinical potential. Structure based rational design and development of drugs against the M^{pro} of coronavirus family will help in the preparedness against sudden zoonotic transmission of new coronavirus outbreak in the future. Due to an immediate need to alleviate COVID-19, repurposed anti-HIV drugs lopinavir and ritonavir (Kaletra®) in combination has been used as a M^{pro} protease target[12]. Similiarly, repurposed nucleotide analog prodrug remdesivir, an inhibitor of RDRP has already been in clinical trial for ebola virus infections, and it has also been approved for treatment of

the disease[13,1,14]. Though the nucleotide analog drugs work with low efficiency due to the exonuclease proof-reading enzyme, remdesivir was found to reduce the lung viral load significantly[13]. Unfortunately, adverse effects was observed in some patients during clinical trial and also the study revealed that no difference in time was observed for clinical improvement[15]. With various repurposed drug candidates used for the treatment of COVID-19, the pathogenic infections it causes cannot be reversed[13]. Apart from this, many synthetic drugs suffer from unknown, long term side effects. In this context, natural product based drugs are considered safe enough when compared to the synthetic ones. Understanding the natural product drug candidate helps in making new drugs such as natural product derivate (ND) and mimic of natural product (NM) etc[16].

Triterpenoids are class of natural products composed of six isoprene units and are characterized by tetracyclic or pentacyclic skeleton[17]. Tetracyclic triterpenes are of lanostane, dammarane, cycloartane, protostane, cucurbitane type and the pentacyclic scaffold includes oleanane, lupane, ursane and taraxastane type triterpenes[17]. Triterpenoids undergo ring cleavage in any of the ring system (A, B, C, D) and are called as seco triterpenoids. Plant triterpenoids exhibit numerous pharmacological and insecticidal activities[18]. Triterpenes such as oleanolic acid and betulinic acid act as viral-host cell fusion inhibitors, which have found to inhibit the entry of several viruses such as ebola, HIV and Influenza A [19]. Among different triterpenoids, limonoids are specific type called tetranortriterpenoids which possess a furan ring attached at C-17 [20]. These limonoids are mainly found in families such as meliaceae, and rutaceae [20],[21]. Limonoids such as limonin and nomilin found in Rutaceae family was found to be active against HIV-1 replication[22]. Another class of degraded triterpenes are quassinoids found in Simarouboidaea subfamily of Simaroubaceae[23]. Quassinoids were found to possess anti-HIV activity[24]. Synthetic analog of plant triterpenoids such as BMS-955176 and Bervimat has been used as anti-HIV drug, which is in the clinical trial phase[25,26].

In this study, triterpenoids from different plant sources, with diverse skeletal structure and functional groups has been selected and screened for its inhibitory activity of the active site of SARS-CoV-2 M^{pro} using molecular docking. The molecules that pass the selection criteria were further studied to understand the origin of M^{pro} protein-inhibitor interactions. The successful molecules were further tested for its inhibitory activity against M^{pro} of other viruses belonging to beta-coronavirus family, such as SARS-CoV and MERS-CoV.

Results and Discussion

i. Virtual screening of triterpenoids

Triterpenoids are highly potent molecules used by plants for diverse biological functions. The gene clustering responsible for the production of these molecules are highly evolved to defend against multiple pathogens. However, virtual screening of structurally diverse triterpenoids to target particular viral proteins was seldom carried out in viral drug discovery. Therefore, virtual screening of these potent molecules to target viral proteins is necessary to exploit their true potential. In this context, 108

triterpenoid compounds were taken for possible inhibitory activity against M^{pro}. The common names and corresponding structure in SMILES format are given in Table S2 in **Supporting Information**. All these compounds belong to diverse variety of medicinal plants such as *Azadirachta indica* (neem), *Citrus sp*. (lemon), *Ocimum sp*.(tulsi) etc. These compounds were screened by performing docking against the M^{pro} of SARS-CoV-2, obtained by homology modeling. A cutoff of -8.5 kcal mol^{-1} mean binding energy was set to obtain the potential candidates. After screening, 47 ligands were retrieved and subjected to further ADME screening criteria. Only the molecules that satisfies the drug likeliness rules were further studied. ADME properties such as molecular weight, partition coefficient (log $P_{O/W}$), topological polar surface area, number of hydrogen bond donors, number of hydrogen bond acceptors and solubility (log *S*) were important in oral availability and gastrointestinal absorption. In addition, properties including blood brain barrier, human enzymes inhibition and skin permeation have also been considered. These ADME values for all 47 candidates are provided in Table S3 and S4 of **Supporting Information**. Based on the above properties and drug likeliness rules, 13 compounds were further screened out. Those compounds were listed in **Table 1** along with their mean binding energies and ADME properties.

Table1. Mean binding energy between drug and M^{pro} along with ADME properties of successful triterpenoids.

Molecule	Mean B.E (kcal mol^{-1})	MW	Mlog*P*	N_A	N_D	TPSA	log *S*
Obacunone	-9.56	454.51	2.16	7	0	95.34	-4.83
Citrusin	-9.38	546.56	1.12	11	1	155.03	-4.11
Nomilin	-9.29	514.56	1.8	9	0	121.64	-4.77
Epoxy-azadiradione	-9.16	466.57	2.54	6	0	86.11	-5.76
Secomahoganin	-9.15	528.59	1.92	9	0	121.64	-5.83
Rutaevin	-9.09	486.51	0.67	9	1	124.8	-3.47
Picrasin A	-9.02	474.54	1.49	8	10	116.2	-3.85
Limonin	-8.86	470.51	1.45	8	0	104.57	-3.58
Gedunin	-8.84	482.57	2.56	7	0	95.34	-5.93
Glaucin B	-8.83	528.55	1.03	10	0	130.87	-4.16
Isoobacunoic acid	-8.83	472.53	1.45	8	1	115.57	-4.27
7-deacetylgedunin	-8.8	440.53	2.23	6	1	116.31	-5.2
Andirobin	-8.5	468.54	2.28	7	0	95.34	-5.02

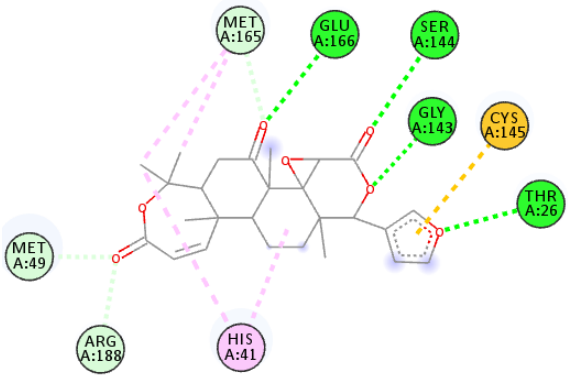
B.E – Binding energy, MW – molecular weight, N_A – number of acceptor atoms such as O and N, N_D – number of donor atoms such as N-H and O-H, TPSA – Topological Polar Surface Area, log S – logarithm of solubility product.

It should be noted from **Table 1** that log*S* values of epoxyazadiradione, secomahoganin, gedunin, deacetyl-gedunin and andirobin shows poorly soluble range and therefore their oral availability might be limited and requires modifications to drug structure to improve solubility. However, these values are predicted based on empirical calculations and therefore should be checked against experimental values. Therefore, these triterpenoids are still considered to have drug like properties against M^{pro} of multiple coronaviruses as it has prospects in future drug designing. Moreover, citrusin and secomahoganin have molecular weight more than 500 and does not satisfy Lipinski rule for drug likeliness. Nevertheless, their molecular weights are close to the limiting region and hence they are also considered for further characterization.

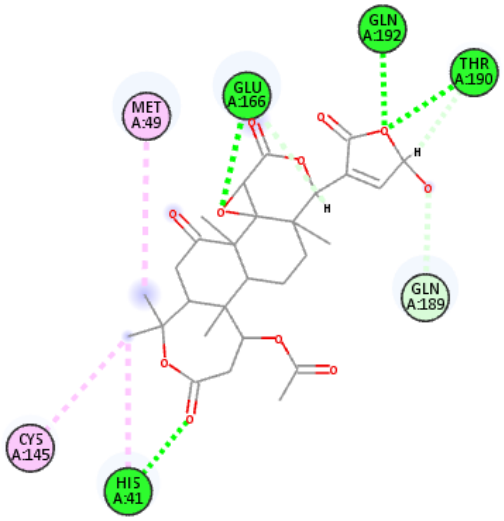
ii. Nature of M^{pro}*-triterpenoid drug interaction*

The triterpenoids with optimal binding energy and ADME properties were further examined for their binding modes with active sites. The residues involving in drug-protein interaction determines the ability of the drug to act as potential inhibitor. **Figure 1** illustrates the 2-D representation of triterpenoid interaction with the active site amino acid residues of M^{pro} of SARS-CoV-2. Apart from the catalytic dyad (His41 and Cys145), several other amino acids in the active site plays major role in the activity of M^{pro}. Both SARS-CoV and SARS-CoV-2 share similar conserved catalytic domain. Though, mutation studies of SARS-CoV-2 are still under progress, the same for SARS-CoV has been carried out. Mutations in Glu166, Arg188 and His163 have been shown to diminish the catalytic activity. Therefore, binding of the drugs to these residues apart from catalytic dyad are important in inhibiting M^{pro} activity.

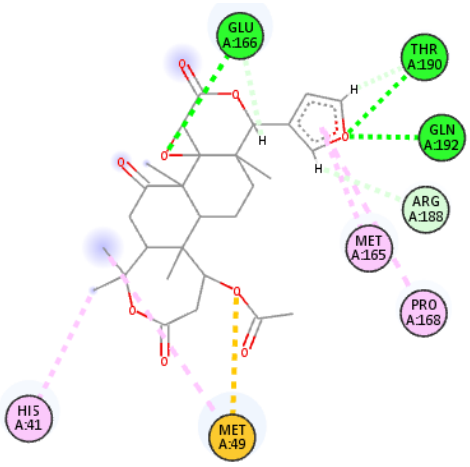
Obacunone, citrusin, nomilin and limonin are naturally occurring limonoids in citrus plants. The first three limonoids share similar structure with a lactide ring in their fused system. Their binding energies with M^{pro} are also similar (-9.56, -9.38 and -9.29 kcal mol^{-1}) despite having different binding modes and substitutions. Obacunone has no hydrogen donor and binds to amide N-H of Glu166, Ser144, Thr26 and Gly143 through hydrogen bonding. Cys145 and His41 interacts with Obacunone through π-S and π-alkyl interactions, respectively. Citrusin binds with the active site residues through H-bonding with amide N-H of Glu166 and Thr190, side-chain N-H of Gln192, imidazole N-H of His41. Nomilin interacts with active site of M^{pro} in the same fashion as that of Citrusin except His41. Limonin inhibits the active site of main protease with a binding energy of -8.86 kcal mol^{-1}. It forms H-bond with imidazole N-H of His41 and amide N-H of Gln189. All these citrus limonoids show excellent ADME properties with insignificant inhibitory effect on human enzymes, better skin permeation (log $k_p > 6.5$) and gastrointestinal absorption.



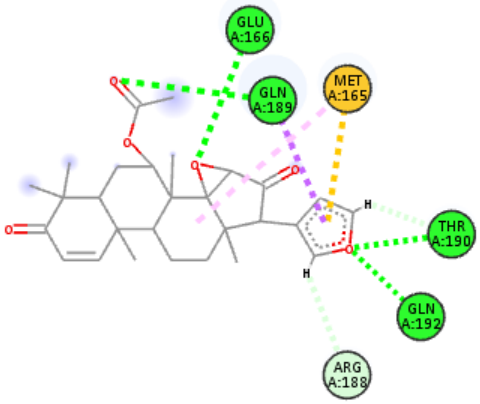
Obacunone



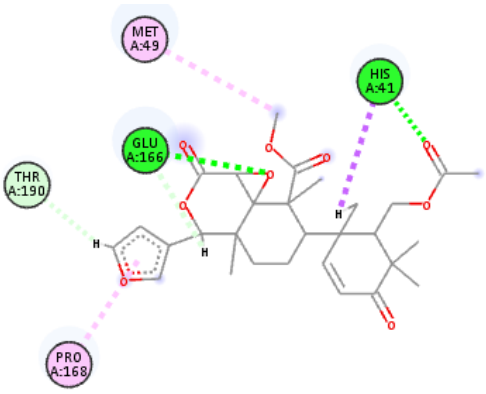
Citrusin



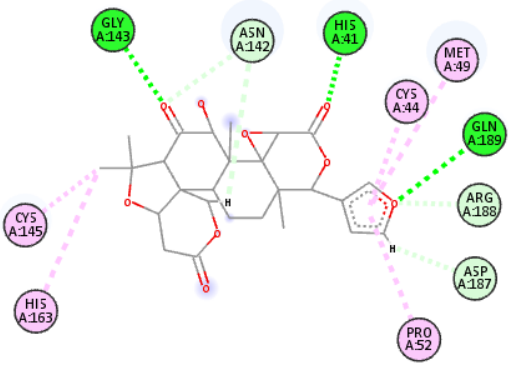
Nomilin



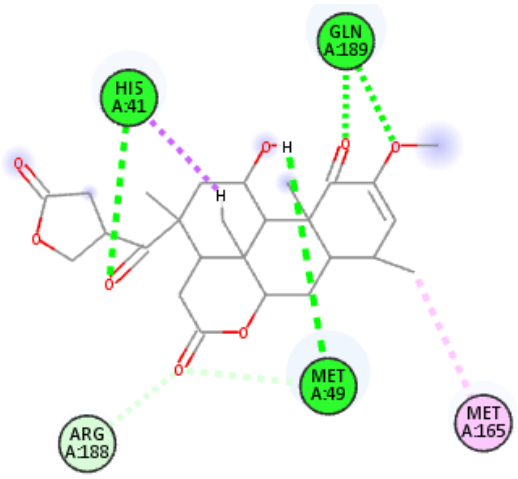
Epoxyazadiradione



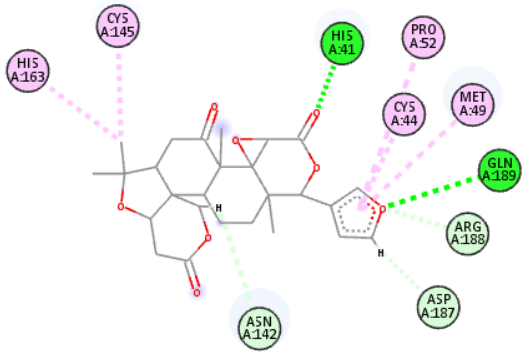
Secomahoganin



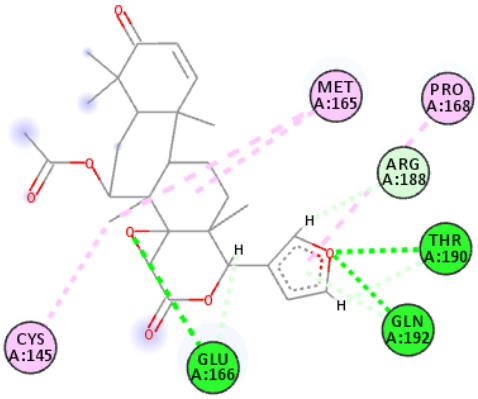
Rutaevin



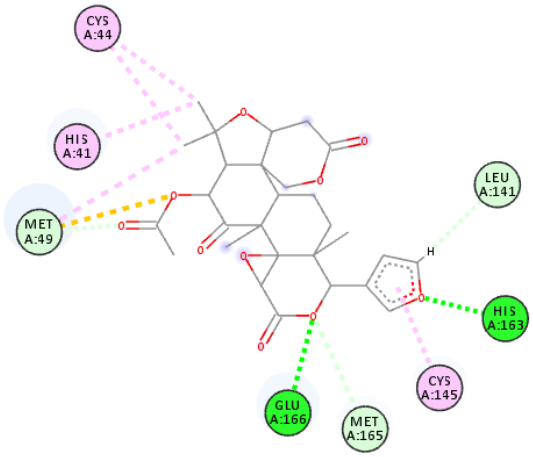
Picrasin A



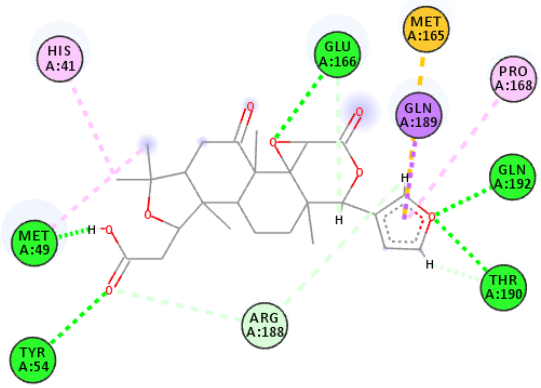
Limonin



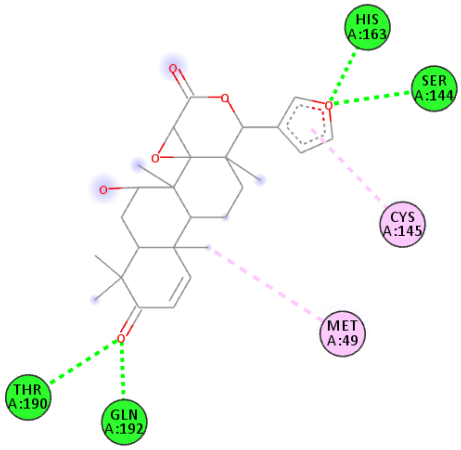
Gedunin



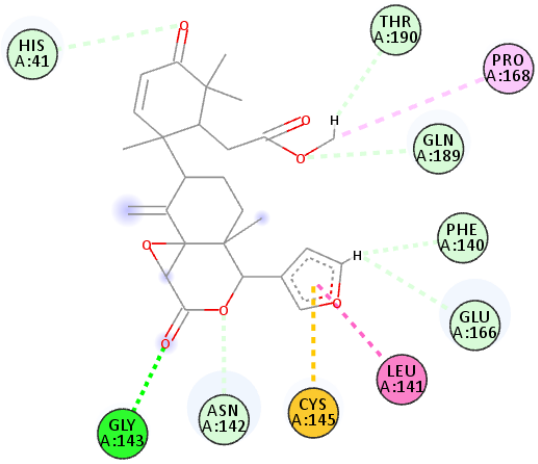
Glaucin B



Isoobacunoic acid



7-deacetylgedunin



Andirobin

Figure 1. 2D representation of the interaction between successful triterpenoid drug candidates and active site residues of M^{pro}. The drug is shown in lines and the interacting aminoacid residues with their numbers are shown inside circles. Each color of aminoacid residues and interaction markers indicates different types of interaction. Green represents conventional H-bonding, Yellow indicates π-SH interaction, Pink denotes π-amide interaction and rest of them represents weak van der Waals interaction.

Epoxyazadiradione, gedunin and 7-deacetylgedunin belong to Neem family (*Azadirachta indica*). They inhibit the active site of M^{pro} with mean binding energies of -9.16, -8.84 and -8.8 kcal mol^{-1} respectively. Epoxyazadiradione binds with the active site through H-bond interaction with amide N-H of Glu166, side-chain N-H of Gln189 and Gln192, amide N-H of Thr90. It also forms π-CH bond with Gln189 and S-π bond with Met165. Its interaction with His41 and Cys145 are negligible in the docked pose. Although gedunin and 7-deacetylgedunin shows similar structural features except an acetyl group, their binding mode with the active site of M^{pro} show larger differences. The furan ring of gedunin forms H-bond with amide N-H of Thr90 and side-chain N-H of Gln192. Its epoxy group interacts with N-H of Glu166 to form H-bond. An important note to consider here is that gedunin seldom interacts with His41 and Cys145 (catalytic dyad) except long-range van der Waals interaction. Similarly, 7-deacetylgedunin forms H-bond with imidazole N-H of His163, side-chain O-H of Ser144, amide N-H of Thr140, side-chain N-H of Gln192. It also has negligible interactions with the catalytic dyad apart from S-π interaction between Cys145 and furan ring of 7-deacetyl gedunin.

Secomahoganin (present in Meliaceae family plants) interacts with the active site of main protease through H-bond interactions with imidazole N-H of His41, amide N-H of Glu166 with mean binding energy of -9.15 kcal mol^{-1}. It also interacts with Met49, Thr190 and Pro168 through weak van der Waals forces. Rutaevin is found in Rutaceae family of plants and shows similar structural features as that of

limonin and binds with the active site in similar manner. The interaction energy for rutaevin is -9.09 kcal mol^{-1} and it results from H-bonds with amide N-H of Gly143, imidazole N-H of His41 and amide N-H of Gln189 and other non-bonded interactions with Cys145, His163, Cys44 and Met49. Picrasin A is a quassinoid and it is present in plants belongs to Simaroubaceae family. It forms strong H-bonds with imidazole N-H of His41, side-chain N-H of Gln189. It also binds with Met49 through O-H···S non-covalent bond. The binding energy of picrasin A with that of the active site of main protease is -9.02 kcal mol^{-1}. The triterpenoid glaucin B is also found in Rutaceae family and shows similar structure as that of limonin and rutaevin. However, its binding mode with that of M^{pro} is different when compared to the other two compounds. It forms H-bond with amide N-H of Glu166 and imidazole N-H of His163 and the binding energy with M^{pro} is -8.83 kcal mol^{-1}.

Isoobacunoic acid is also present in citrus family with less structural similarity to obacunone, citrusin and limonin. It interacts with the active site pocket with interaction energy of -8.83 kcal mol^{-1} and it has comparatively more H-bonds with the active site residues of M^{pro}. The carboxylic acid group interacts with phenyl O-H of Tyr54, amide C=O of Met49. Other groups such as side-chain N-H of Gln192, Thr190 and Glu166 also forms H-bond with isoobacunoic acid. andirobin is found in Meliaceae family and inhibits M^{pro} with binding energy -8.5 kcal mol^{-1}. It forms H-bonding interaction with amide N-H of Gly143. It also makes weak non-covalent interactions with Cys145 and Leu141.

Table 2 provides the screened triterpenoids and the active site residues interacting with them along with their interacting groups. The docked structures do not represent real bonding scenario. However, these predicted H-bonds provide an insight into the possible ways for triterpenoids to interact with active site residues. H-bonds dominate the overall energetics among other non-covalent interactions between drug and protein. It is also directional and re-orients ligand and binding site residues.

Table 2. Amino acid residues involved in H-bonding and hydrophobic interaction between triterpenoids and M^{pro} active site residues along with the chemical functionalities take part in H-bond corresponding to each amino acid. A, B, C, D represents the name of the ring in the limonoid and quassinoid nomenclature and subscript c indicates that the ring is cleaved.

	Amino acid residues		H-bond interacting groups	
	Hydrophobic	H-bond	Ligand	Amino acid
Obacunone	C145, H41, N188, M49	E166	A (>C=O)	Amide N–H
		G143	D (–O–)	Amide N–H
		S144	D (>C=O)	Amide N–H
		T26	Furan Ring O	Amide N–H

Citrusin	C145, M49, Q189	E166 Q192 T190 H41	D (epoxy O) Furan Ring O Furan Ring O A (lactide >C=O)	Amide N–H Side-chain N–H Amide N–H Imidazole N–H
Nomilin	N188, M165, P168, M49, H41	E166 Q192 T190	D (epoxy O) Furan Ring O Furan Ring O	Amide N–H Side-chain N–H Amide N–H
Epoxy-azadiradione	M165, N188	E166 Q189 T190 Q192	D (epoxy O) B (acetate >C=O) Furan Ring O Furan Ring O	Amide N–H Side-chain N–H Side-chain N–H Amide N–H
Secomahoganin	M49, T190, P168	E166 H41	D (epoxy O) A_c (>C=O)	Amide N–H Imidazole N–H
Rutaevin	C44, M49, P52, C145, H163	G143 H41 Q189	B (>C=O) D (>C=O) Furan Ring O	Amide N–H Imidazole N–H Amide N–H
Picrasin A	M165, N188	H41 Q189 M49	D (>C=O) A (methoxy O) D (O-H)	Imidazole N–H Side-chain N–H Side-chain S
Limonin	C145, H163, C44, M49, P52	H41 Q189	D (>C=O) Furan Ring O	Imidazole N–H Amide N–H
Gedunin	C145, M165, P168	T190 Q192 E166	Furan Ring O Furan Ring O D (epoxy O)	Amide N–H Side-chain N–H Amide N–H
Glaucin B	C145, H41, C44, M49	H163 E166	Furan Ring O D (–O–)	Imidazole N–H Amide N–H
Isoobacunoic-acid	H41, M165, Q189, P168	E166 Q192 T190 M49 Y54	D (epoxy O) Furan Ring O Furan Ring O A (–COO-H) A (–C(OH)=O)	Amide N–H Side-chain N–H Amide N–H Amide C=O Side-chain O-H

7-deacetyl-gedunin	C145, M49	H163 S144 Q192 T190	Furan Ring O Furan Ring O A (>C=O) A (>C=O)	Imidazole N–H Side-chain O–H Side-chain N–H Amide N–H
Andirobin	G163, C145, L141, P168	G143	D (>C=O)	Amide N–H

A closure look at **Table 1** indicates that most of the triterpenoids forms H-bond with amide N-H of peptide linkage, since they mostly contains H-bond acceptors. As already mentioned, only few of these residues actively take part in the proteolysis, among which His41, Cys145, Glu166 and His163 plays important role in protease activity. Only few triterpenoids make direct H-bonding interaction with His41. Citrusin, secomahoganin, rutaevin, picrasin A and limonin interacts with side-chain N-H group in imidazole ring of His41 through H-bonding. Citrusin and secomahoganin binds with His41 through >C=O in A ring and A-seco ring, respectively. Rutaevin, picrasin A and limonin interacts through >C=O D ring. Glu166 forms H-bond with obacunone, citrusin, nomilin, epoxyazadiradione, secomahoganin, glaucin B and isoobacunoic acid through amide N-H in peptide backbone. An interesting point to note is that only the epoxy/ether oxygen in ring D forms H-bond with active site residues, except obacunone where it is >C=O in ring A. The above observation indicates that both ring D and furan ring act as better anchoring point for H-bonding between triterpenoids and M^{pro}. Both citrusin and secomahoganin interacts with both His41 and Glu166 and serve as better inhibitors for M^{pro}.

Hydrophobic interaction between docked triterpenoids and active site of M^{pro} is as important as H-bonding interaction, as it improves overall binding energy between protein and ligand. However, caution must be taken in interpreting these interactions from molecular docking results due to the inadequacy in the definition of van der Waals interaction terms used in docking calculations. The screened triterpenoids have good long-range interactions with at least any one of the active site residues such as Cys145, His41 and Glu166 except epoxyazadiradione, secomahoganin and picrasin A. Considerably, obacunone and glaucin B forms H-bond with Glu166 and interacts with Cys145 and H41 through hydrophobic interaction. Citrusin forms H-bond with His41 and Glu166 and interacts with Cys145 through van der Waals forces. These three triterpenoids are most suitable as anti-viral drugs for the fight against CoV.

Considering both H-bond and hydrophobic interactions with active site residues, the order of effectiveness of thirteen triterpenoids has been found to be Citrusin > Obacunone ≈ Glaucin B > Secomahoganin > Limonin ≈ Rutaevin >

Nomilin ≈ Gedunin ≈ Isoobacunoic acid > Picrasin A > Epoxyazadiradione > 7-deacetylgedunin ≈ Andirobin.

iii. Inhibition against MERS and SARS coronaviruses

To understand the inhibitory effect of the screened triterpenoids against main protease of other coronaviruses, MERS and SARS viruses were taken and docked with triterpenoids. The binding energies were shown in **Table 3** along with their interacting residues.

Table 3. Mean binding energies and amino acid residues involving in hydrophilic and hydrophobic interactions between thirteen triterpenoids and active site region of main protease of both MERS and SARS CoV.

Compound	Mean Binding energy (kcal mol^{-1})		H-bonding residues		Hydrophobic Interaction residues	
	MERS	SARS	MERS	SARS	MERS	SARS
Obacunone	-8.17	-9.78	E169, Q195	-	H41, L49, M25, M168	N142, H163, C145, M165, R188, T190, Q189
Citrusin	-9.26	-10.51	E169, V193	Q192, N142, T190	H41, M25, K191	C145, H163, Q189, L141
Nomilin	-7.89	-8.70	E169, Q195	E166, N142, G143	H41, V193, M25, M168	M165, L167, D187, C145, H41, L27
Epoxy-azadiradione	-8.84	-7.85	E169, Q195	E166	M168	M49, H41, M165
Secomahoganin	-8.04	-7.56	H41, E169, Q195	T26, E166	M168, V193	M165
Rutaevin	-7.41	-8.87	H41, Q195	Q192, T190,	M168, E169,	P168, E166,

				N142, H163	L170, L49	Q189
Picrasin A	-8.05	-7.87	H41, Q167, G146	E166	Q192, M25	H41, Q189
Limonin	-8.21	-9.10	K191, Y54	Q192, T190, H163	C44, L49, M25, A171, M168	P168, E166, Q189
Gedunin	-8.85	-8.38	E169, Q195, Q192	C145, N142, S144	M168	M165
Glaucin B	-8.04	-8.77	C145, H41, Q169, Y54	N142, H163, T190	M168, A171	Q189, E166
Isoobacunoic-acid	-7.17	-7.38	Q192, K191, Y54	Q192, T190	M25, A46, L49, C44, M168, A171	H41, M165, C145, Q189
7-deacetyl-gedunin	-8.01	-9.45	E169	T190, H163, S144	C145, L144, M168, A171, V193	R188, Q192, M165
Andirobin	-7.46	-8.35	T26	Q192	M25, M168, E169	H163, C145, H41

Only few triterpenes bonds to the active site residues (Glu166, His41 and Cys145) through H-bonding and hydrophobic interactions. Among them, secomahoganin, rutaevin, picrasin A and glaucin B makes direct H-bond with His41 of M^{pro} of MERS CoV and acts as potent inhibitors. Also, obacunone, citrusin and nomilin binds to the active site through hydrophobic interactions and block the enzyme dynamics of protein binding.

Similarly, both nomilin and secomahoganin forms H-bond with Glu166 of M^{pro} of SARS CoV. Nomilin, epoxyazadiradione, picrasin A, isoobacunoic acid and andirobin inhibits the protease activity by making hydrophobic interactions with the active site of M^{pro} of SARS CoV. The interacting poses for these structures (MERS and SARS) are similar to that of SARS-CoV-2. From the above observation it could be understood that secomahoganin and nomilin are better at inhibiting both MERS and SARS CoV main proteases.

Conclusions

In this work, the efficacy of triterpenoids in acting as potential anti-viral drugs for coronaviruses (SARS-CoV-2, SARS, and MERS) were evaluated using molecular docking approach against the inhibition of main proteases of the same. About 108 triterpenoids were taken for this study and they were initially screened based on their binding energy (B.E > -8.5 kcal mol^{-1}) with the active site of M^{pro} of SARS-CoV-2. Further screening for optimal ADME properties and drug likeliness leads to thirteen successful triterpenoids. These docked structures were further examined to understand the molecular basis of active site inhibition through non-bonding interactions. It is observed that both H-bonding and hydrophobic interactions are necessary to block the entrance of active site from binding to their target proteins. Among other active site amino acid residues, Cys 145, His41 and Glu166 plays important role in proteolytic activity. Considering the ability of different triterpenoids in binding to multiple target residues, the potential of inhibition of thirteen triterpenoids for SARS-CoV-2 are Citrusin > Obacunone ≈ Glaucin B > Secomahoganin > Limonin ≈ Rutaevin > Nomilin ≈ Gedunin ≈ Isoobacunoic acid > Picrasin A > Epoxyazadiradione > 7-deacetylgedunin ≈ Andirobin. Here, the top triterpenoids with better inhibition property are citrus limonoids (citrusin, obacunone, glaucin B, limonin, rutaevin and nomilin), except secomahoganin.

The inhibition of SARS and MERS coronaviruses by these triterpenoids show larger differences in their binding pockets. Although, some of the triterpenoids have larger binding energy (citrusin, obacunone) with main proteases of these coronaviruses, their binding mode to Cys145, His41 and Glu166 are mostly hydrophobic and may limit their application as anti-virals against SARS and MERS coronaviruses. However, this inference is highly debatable, as it requires molecular dynamics or high-level QM/MM calculations to understand the effect of these citrus limonoids in interfering with protease activity. Yet, secomahoganin and nomilin inhibits both SARS and MERS main proteases with high interaction energy and optimal binding mode to active site residues.

Conclusively, citrus limonoids and secomahoganin shows better inhibition of the three coronaviruses under study. These natural products have better ADME properties and less toxicity as compared to synthetic drugs and covalent inhibitor drugs. Therefore, these triterpenoids can be further used for cell culture studies and clinical trials against coronaviruses.

Computational Methods

The three-dimensional structure of M^{pro} was constructed using homology modeling of genetic sequence obtained from NCBI (Ref.Seq: YP_009742612.1). The

templates used to construct the structure were obtained from Protein Data Bank (PDB ID: 6M03, 6W63 and 6LU7). One of them is in apo-form (6M03), the second one is non-covalently complexed to an inhibitor (6W63) and the final structure contains protein covalently bonded to inhibitor (6LU7). These structures pose slightly different configurations in active site region and therefore they were chosen for modeling M^{pro}. Modifications to their crystal structures were made by removing water, ligands and adding hydrogen wherever necessary. CHIMERA 1.14[27] was used to generate 3D structure based on the template of modified crystal structure of proteins. To understand the inhibitory effect of drugs against M^{pro} of other coronaviruses, MERS and SARS coronaviruses were also considered in this study. The crystal structures of main protease of MERS and SARS were taken from protein data bank with repository ID's 2YNA and 2GZ9 respectively. These structures were also prepared for docking using same procedure adopted for SARS-CoV-2.

The generated structure was further used for docking with triterpenoids. Most of the 3D structures of triterpenoids were generated using OPENBABEL[28] using the SMILES[29] obtained from PUBCHEM[30] or from CHEMDRAW 2D structures, wherever SMILES are not available. The triterpenoids were chosen based on various criteria such as anti-microbial and anti-viral properties as well as structural diversity in their skeletal ring. The stereo-centers of these drugs were checked against the 2D structures of the same.

Docking was performed using AUTODOCK 4.2[31] with gasteiger charges for all atoms (protein and ligand). The grid box was set around the active site of M^{pro} with dimensions 34×32×35 Å with its center at active site, which includes catalytic residues His41 and Cys145. The inhibitors were initially screened based on the average binding energy of clusters with at least more than 50% of total number of docking configurations. The screening criteria for binding energy were set to -8.5 kcal mol^{-1} and the successful ones were further studied for their drug likeliness. ADME (Absorption, Distribution, Metabolism and Excretion) properties, for the molecules with mean binding energy > -8.5 kcal mol^{-1}, were calculated with SwissADME[32] and Molinspiration[33] web services. Importantly, molecular weight (MW), log $P_{O/W}$, Topological Polar Surface Area (TPSA), number of hydrogen bond donors (N_D), number of hydrogen bond acceptors (N_A), Solubility (log *S*) were calculated and set as parameters for further screening. Only the molecules that satisfies drug-likeness rule (Lipinski[34], Ghose[35], Veber[36], Egan[37] and Muegge[38]) were further screened for their pharmacokinetic properties. GI absorption, Blood Brain Barrier (BBB)[39], enzyme inhibition activity, P-gp (P-glycoprotein) substrate activity and skin permeation[40] were also predicted using SwissADME.

The final set of molecules, that were screened through docking, ADME and pharmacokinetics, were studied for their nature of interaction and binding mode with M^{pro}. The visual characterization of H-bonding and hydrophobic interactions were carried out using BIOVIA Discovery Studio Visualizer[41]. Also, they were docked against MERS and SARS coronaviruses to identify their potential in inhibiting multiple main proteases with sequence and structural diversity.

Acknowledgment The author J. Vijaya Sundar acknowledges Indian Institute of Technology Madras (IITM), Chennai for granting Institute Post-Doctoral Fellowship (IPDF) and computational facility.

Supporting Information Available 3D representation of M^{pro} of SARS-CoV-2, MERS and SARS, SMILES format for triterpenoids along with their binding energies to M^{pro} of SARS-CoV-2 are provided. Selected ADME properties were also provided for successful triterpenoids. The mode of binding to active site of M^{pro} of the three coronaviruses were also shown.

References

(1) Guo, Y.-R.; Cao, Q.-D.; Hong, Z.-S.; Tan, Y.-Y.; Chen, S.-D.; Jin, H.-J.; Tan, K.-S.; Wang, D.-Y.; Yan, Y. The Origin, Transmission and Clinical Therapies on Coronavirus Disease 2019 (COVID-19) Outbreak – an Update on the Status. *Military Medical Research* **2020**, *7* (1), 11.

(2) Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R.; Icenogle, J. P.; Peñaranda, S.; Bankamp, B.; Maher, K.; Chen, M. hsin; et al. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* **2003**, *300* (5624), 1394–1399.

(3) Hilgenfeld, R. From SARS to MERS: Crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design. *The FEBS journal*. 2014, pp 4085–4096.

(4) Chan, J. F. W.; Kok, K. H.; Zhu, Z.; Chu, H.; To, K. K. W.; Yuan, S.; Yuen, K. Y. Genomic Characterization of the 2019 Novel Human-Pathogenic Coronavirus Isolated from a Patient with Atypical Pneumonia after Visiting Wuhan. *Emerging Microbes and Infections* **2020**, *9* (1), 221–236.

(5) Dawood, A. A. Mutated COVID-19 May Foretell a Great Risk for Mankind in the Future. *New Microbes and New Infections* **2020**, *35*, 100673.

(6) Chen, Y.; Liu, Q.; Guo, D. Emerging Coronaviruses: Genome Structure, Replication, and Pathogenesis. *Journal of Medical Virology*. 2020, pp 418–423.

(7) Astuti, I. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An Overview of Viral Structure and Host Response. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **2020**.

(8) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; et al. The Crystal Structures of Severe Acute Respiratory Syndrome Virus Main Protease and Its Complex with an Inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100* (23), 13190–13195.

(9) Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; et al. Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. *PLoS Biology* **2005**, *3* (10).

(10) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S. H. An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *Journal of Medicinal Chemistry*. 2016, pp 6595–6628.

(11) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. (3CL pro) Structure : Basis for Design of Anti-SARS Drugs. *Science* **2003**, *300* (June),

1763–1767.
(12) Nutho, B.; Mahalapbutr, P.; Hengphasatporn, K.; Pattaranggoon, N. C.; Simanon, N.; Shigeta, Y.; Hannongbua, S.; Rungrotmongkol, T. Why Are Lopinavir and Ritonavir Effective against the Newly Emerged Coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms. *Biochemistry* **2020**, *59* (18), 1769–1779.
(13) Martinez, M. A. Compounds with Therapeutic Potential against Novel Respiratory 2019 Coronavirus. *Antimicrobial Agents and Chemotherapy* **2020**, *64* (5), 1–7.
(14) Huang, J.; Song, W.; Huang, H.; Sun, Q. Pharmacological Therapeutics Targeting RNA-Dependent RNA Polymerase, Proteinase and Spike Protein: From Mechanistic Studies to Clinical Trials for COVID-19. *Journal of Clinical Medicine* **2020**, *9* (4), 1131.
(15) Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; et al. Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial. *The Lancet* **2020**, *395* (10236), 1569–1578.
(16) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*. 2016, pp 629–661.
(17) Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. Triterpene Biosynthesis in Plants. *Annual Review of Plant Biology* **2014**, *65* (1), 225–257.
(18) Dzubak, P.; Hajduch, M.; Vydra, D.; Hustova, A.; Kvasnica, M.; Biedermann, D.; Markova, L.; Urban, M.; Sarek, J. Pharmacological Activities of Natural Triterpenoids and Their Therapeutic Implications. *Natural Product Reports*. 2006, pp 394–411.
(19) Si, L.; Meng, K.; Tian, Z.; Sun, J.; Li, H.; Zhang, Z.; Soloveva, V.; Li, H.; Fu, G.; Xia, Q.; et al. Triterpenoids Manipulate a Broad Range of Virus-Host Fusion via Wrapping the HR2 Domain Prevalent in Viral Envelopes. *Science Advances* **2018**, *4* (11), 1–15.
(20) Roy, A.; Saraf, S. Limonoids: Overview of Significant Bioactive Triterpenes Distributed in Plants Kingdom. *Biological and Pharmaceutical Bulletin*. 2006, pp 191–201.
(21) Tan, Q. G.; Luo, X. D. Meliaceous Limonoids: Chemistry and Biological Activities. *Chemical Reviews*. 2011, pp 7437–7522.
(22) Battinelli, L., Mengoni, F., Lichtner, M., Mazzanti, G., Saija, A., Mastroianni, C. M., & Vullo, V. Effect of Limonin and Nomilin on HIV-1 Replication on Infected Human Mononuclear Cells. *Planta Medica* **2003**, *69* (10), 910–913.
(23) Chakraborty, D.; Pal, A. Quassinoids: Chemistry and Novel Detection Techniques. In *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*; Ramawat, K. G., Mérillon, J.-M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 3345–3366.
(24) Okano, M.; Fukamiya, N.; Tagahara, K.; Cosentino, M.; Lee, T. T.-Y.; Morris-Natschke, S.; Lee, K.-H. Anti-HIV Activity of Quassinoids. *Bioorganic & Medicinal Chemistry Letters* **1996**, *6* (6), 701–706.
(25) Morales-Ramirez, J.; Bogner, J. R.; Molina, J. M.; Lombaard, J.; Dicker, I. B.; Stock, D. A.; DeGrosky, M.; Gartland, M.; Dumitrescu, T. P.; Min, S.; et al. Safety, Efficacy, and Dose Response of the Maturation Inhibitor GSK3532795 (Formerly Known as BMS-955176) plus Tenofovir/Emtricitabine Once Daily in Treatment-Naive HIV-1-Infected Adults: Week 24 Primary Analysis from a Randomized Phase IIb Trial. *PLoS ONE* **2018**, *13* (10), 3–4.

(26) Wang, D.; Lu, W.; Li, F. Pharmacological Intervention of HIV-1 Maturation. *Acta Pharmaceutica Sinica B*. Elsevier 2015, pp 493–499.
(27) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* **2004**, *25* (13), 1605–1612.
(28) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *Journal of Cheminformatics* **2011**, *3* (1), 33.
(29) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *Journal of Chemical Information and Computer Sciences* **1988**, *28* (1), 31–36.
(30) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic acids research* **2019**, *47* (D1), D1102–D1109.
(31) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry* **2009**, *30* (16), 2785–2791.
(32) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Scientific Reports* **2017**, *7* (1), 42717.
(33) Molinspiration Cheminformatics. free web services https://www.molinspiration.com.
(34) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings1PII of Original Article: S0169-409X(96)00423-1. The Article Was Originally Published in Advanced Drug Delivery Reviews 23 (1997) 3. *Advanced Drug Delivery Reviews* **2001**, *46* (1), 3–26.
(35) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. 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* **1999**, *1* (1), 55–68.
(36) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *Journal of medicinal chemistry* **2002**, *45* (12), 2615–2623.
(37) Egan, W. J.; Merz, K. M.; Baldwin, J. J. Prediction of Drug Absorption Using Multivariate Statistics. *Journal of medicinal chemistry* **2000**, *43* (21), 3867–3877.
(38) Muegge, I.; Heald, S. L.; Brittelli, D. Simple Selection Criteria for Drug-like Chemical Matter. *Journal of medicinal chemistry* **2001**, *44* (12), 1841–1846.
(39) Daina, A.; Zoete, V. A Boiled-egg to Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **2016**, *11* (11), 1117.
(40) Potts, R. O.; Guy, R. H. Predicting Skin Permeability. *Pharmaceutical research* **1992**, *9* (5), 663–669.
(41) BIOVIA, Dassault Systèmes, Discovery Studio Visualizer V20.1, , San Diego: Dassault Systèmes, 2020.