






In-silico homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain)

Phulen Sarma^{a*} , Nishant Shekhar^{a*}, Manisha Prajapat^{a*}, Pramod Avti^{b*} , Hardeep Kaur^a, Subodh Kumar^a, Sanjay Singh^{c#}, Harish Kumar^a, Ajay Prakash^a, Deba Prasad Dhibar^d and Bikash Medhi^a 

^aDepartment of Pharmacology, Postgraduate Institute of Medical Education and Research, Chandigarh, India; ^bDepartment of Biophysics, Postgraduate Institute of Medical Education and Research, Chandigarh, India; ^cBiomedical Informatics Centre, Postgraduate Institute of Medical Education and Research, Chandigarh, India; ^dInternal Medicine, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

The N terminal domain (NTD) of Nucleocapsid protein (N protein) of coronavirus (CoV) binds to the viral (+) sense RNA and results in CoV ribonucleoprotein (CoV RNP) complex, essential for the virus replication. In this study, the RNA-binding N terminal domain (NTD) of the N protein was targeted for the identification of possible inhibitors of RNA binding. Two NTD structures of N proteins were selected (2OFZ and 1SSK, 92% homology) for virtual screening of 56,079 compounds from Asinex and Maybridge library to identify top 15 hits for each of the targets based on 'docking score'. These top-hits were further screened for MM-GBSA binding free energy, pharmacokinetic properties (QikProp) and drug-likeness (SwissADME) and subjected to molecular dynamics (MD) studies. Two suitable binders (ZINC00003118440 and ZINC0000146942) against the target 2OFZ were identified. ZINC00003118440 is a theophylline derivative under the drug class 'bronchodilators' and further screening with approved bronchodilators was also studied to identify their ability to bind to the RNA binding region on the N protein. The other identified top hit is ZINC0000146942, which is a 3,4-dihydropyrimidone class molecule. Hence this study suggests two important class of compounds, theophylline and pyrimidone derivatives as possible inhibitors of RNA binding to the N terminal domain of N protein of coronavirus, thus opening new avenues for in vitro validations.

Abbreviations: CoV: Coronavirus; NTD: N terminal domain; MERS: Middle East Respiratory syndrome; SARS: Sudden acute respiratory syndrome; N Protein: Nucleocapsid protein; MD: Molecular dynamics; 2019-nCoV: 2019 novel coronavirus; COVID-19: Coronavirus disease 2019

ARTICLE HISTORY

Received 20 March 2020
Accepted 31 March 2020

KEYWORDS

Nucleocapsid protein; N terminal domain; drug design; RNA binding; 2019 novel corona virus; 2019-nCoV; SARS-CoV-2

Introduction

Coronaviruses are enveloped spherical or pleomorphic single stranded RNA viruses with characteristic bears club shaped glycoproteins on their surface (Prajapat et al., 2020). Seven strains of human CoVs are documented, which include 229E, NL63, OC43, HKU1, MERS-CoV, SARS-CoV and 2019-nCoV (Prajapat et al., 2020). The previous two Coronavirus epidemics, SARS-CoV and MERS-CoV, with 10% mortality for SARS-CoV and 37% for MERS-CoV affected ~10,000 individual lives (Huang et al., 2020). The reemergence and severe outbreak of Coronavirus, now as COVID-19 has caused global health emergency (Sarma et al., 2020). Originated from Wuhan, China, the virus has spread as far as Thailand, Japan, Korea, the USA, Vietnam, Singapore, India (Wang et al., 2020) and European nations. WHO has declared COVID-19 as a pandemic (*Coronavirus Disease (COVID-19)—Events as they happen, n.d.*). The 2019-nCoV is reported to be the seventh and


newest member of the Coronavirus, with human infections (Zhu et al., 2020).

The CoV has several conserved structural proteins e.g. the matrix protein (M protein), small envelop protein (E protein), trimeric spike (S) glycoprotein, nucleocapsid protein (N protein) and proteases (papain like and the main protease) (Prajapat et al., 2020). The nucleocapsid protein is typically located inside the virus and is one of the most abundant structural proteins in the Coronaviruses. The N protein binds with the viral RNA genome to form a virion core, which is vital for its replication and transcription. In addition, the N protein also takes an essential part in viral RNA synthesis (Chang et al., 2016). Recent studies suggest that N proteins of coronaviruses can be useful antiviral drug targets against infections caused by these viruses for its importance in the replication initiation machinery (Chenavas et al., 2013; Lin et al., 2014). One of the interesting features is that all CoV has the least variable structure of N protein (Chang et al.,

CONTACT Bikash Medhi  drbikashu@yahoo.com  Department of Pharmacology, Postgraduate Institute of Medical Education & Research, Research Block B, 4th Floor, Room no 4043, Chandigarh 160012, India

*Equal contribution.

[#]Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, India.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2020.1753580>.

© 2020 Informa UK Limited, trading as Taylor & Francis Group

2006, 2016). The N terminal domain (NTD) of the nucleocapsid protein (N protein) plays a significant role in binding to genomic and sub-genomic RNAs in MHV and IBV virions. The N protein of CoV plays a major role in the viral replication cycle by the formation of the ribonucleoprotein complex with the help of its interactions through the viral RNA and N terminal domain (NTD) of the N protein (Lin et al., 2014). It is due to this activity, the SARS-CoV N protein is suggested to be an RNA chaperone in a previous study and a stepping stone in viral genomic RNA replication.

Few studies are there, which tried to design a new drug targeting to disrupt the interactions between coronavirus N protein NTD and CoV-RNA. A molecule 'PJ34' designed targeting the N protein RNA binding domain of HCoV-OC43 N-NTD (PDB 4KXJ) is reported (Lin et al., 2014), however, the similarity between HCoV-OC43 N-NTD (PDB 4KXJ) and the 2019-nCoV N protein is only 46%. Here comes the need for designing new inhibitors against the RNA binding region of N protein to inhibit the binding of RNA and subsequent inhibition of viral replication in case of the 2019-nCoV.

Material and methods

The computational progressions were carried out on an Acer Predator Helios 300 laptop running on Linux Ubuntu OS 18.04.02 LTS, with Maestro release version 2019-3 from Schrödinger. GPU machine used to render molecular modeling and dynamic simulations of the mentioned molecules were Nvidia GeForce GTX 1660 Ti (6GB).

Selection of target PDB structure of nucleocapsid protein RNA binding domain

In this study we tend to design an inhibitor of RNA binding. Our target of interest was the NTD of nucleocapsid (N) protein. In the nucleocapsid (N) protein, our segment of interest was the RNA binding domain of the nucleocapsid protein. However as there was no structure available on the nucleocapsid protein of 2019-nCoV, the structures that were most close to the 2019-nCoV nucleocapsid protein N terminal domain (RNA binding domain) were 2OFZ (Saikatendu et al., 2007) and 1SSK (Huang et al., 2004) (92% sequence homology with the Wuhan sea food market pneumonia virus nucleocapsid phosphoprotein, accession number QHN73802). Although, both the structures represented the RBD of the N protein, none of the structures were bound to any reported inhibitor of RNA binding. Among the selected structures, PDB IDs, 1SSK was a NMR structure (resolution is not given) and the second target structure PDB ID 2OFZ is an ultrahigh resolution crystallized X-Ray structure (1.17 Å). The PDB IDs of the target structures (2OFZ and 1SSK) were retrieved from the RCSB database (PDB Database, n.d.).

Preparation of the protein

In order to prepare the structure of the protein, the protein preparation wizard tool of Maestro version 10.2 was used. The back chains and missing side of the protein were

included. On default settings and OPLS3 force field, the protein was minimized (Desmond, n.d.; Meng et al., 2011). Water molecules beyond 5.0 Å were removed to avoid hindrance in the binding region.

Ligand structure retrieval and preparation

36,750 compounds from the Maybridge library and 19,329 compounds of the Asinex library from the ZINC database were obtained and prepared in the ligprep module of Schrödinger. This step is crucial for processing input structure. The optimization was carried out by the OPLS 2005 force field, which produced the low-energy isomers with ready to dock poses (LigPrep, n.d.).

Determination of the RNA binding pocket

For the determination of the binding pocket, Sitemap was used. Picturized representation of the estimated binding pocket can be seen in Figure 1. The guided pocket coordinates were then docked with ribonucleoside 5'-monophosphates (AMP and UMP) to obtain an insight into the most active residues taking part in nucleotide binding (Lin et al., 2014). The imperative insight from the docking to be noted is the aberrant difference between the binding coordinates of 2OFZ and 1SSK, owing to the crystal and solution structure of the N protein of SARS-CoV, respectively. The spatial UMP-AMP docked pose RMSD for 2OFZ and 1SSK were found to be 3.74 Å and 3.85 Å, respectively. The pictorial and tabular illustration showing interacting residues can be seen in Table 1 and Figure 2.

Virtual screening

The glide tool of Maestro was used for the virtual screening purpose. A total of 56,079 compounds were screened through the three step VSW (virtual screen workflow) in a tandem HTVS, SP and XP progression. Enhanced sampling method and flexible docking were enabled in XP to screen hits for N protein NTD models (2OFZ and 1SSK) (Desmond, n.d.; Meng et al., 2011).

MM-GBSA: binding free energy calculation

Prime MM-GBSA was used to predict the free energy of binding between the receptor and the set of ligands. The binding free energy (ΔG_{bind}) was calculated using the default parameters of the Prime module in Maestro, Schrödinger, LLC, New York, NY, USA (Singh & Muthusamy, 2013). Prime uses a surface generalized Born (SGB) model employing a Gaussian surface instead of van der Waals surface for better representation of a solvent accessible surface area. MM/GBSA, provides better statistical correlations against experimental binding data than previous similar reported studies (Nain et al., 2019; Adasme-Carreño et al., 2014). The co-crystallized pose-viewer complexes were used as input files with implicit VSGB solvation model, OPLS3 force field and all other settings as default in Prime MM-GBSA (GlideScore/Docking

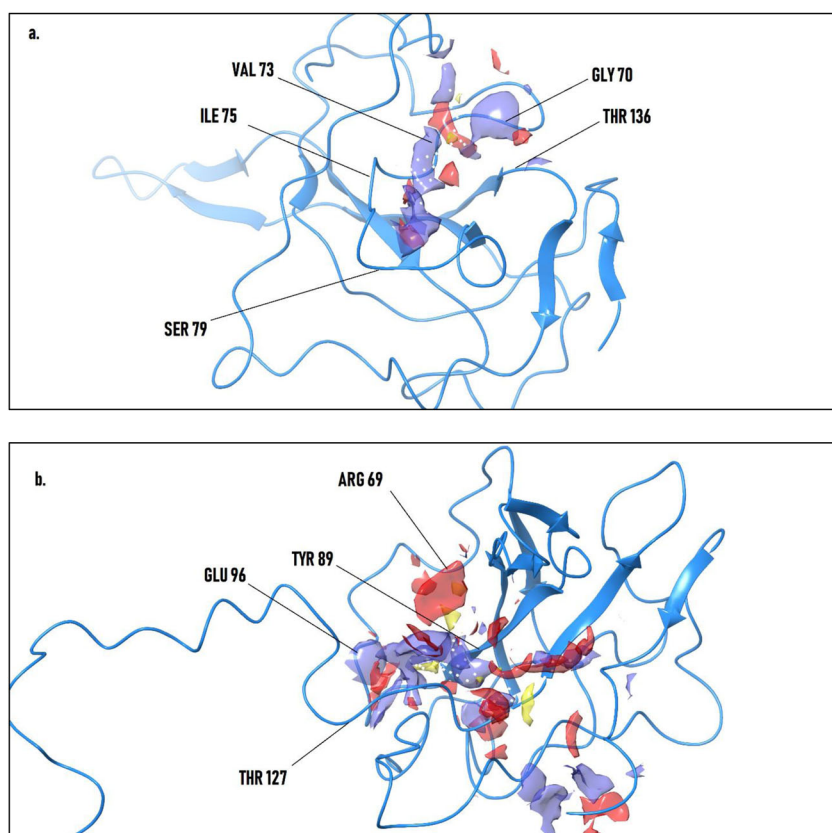


Figure 1. Sitemap derived binding pockets of 2OFZ (a) and 1SSK (b) at distinct regions with the hydrophobic region (blue bubbles), electrostatic regions (red bubbles) and lipophilic points (yellow bubble).

Table 1. Details of AMP and UMP binding to sitemap guided binding pocket of N protein NTD (2OFZ and 1SSK).

PDB id	Ligand	Docking score (GLIDE)	MM-GBSA	Active site residues
2OFZ (Sitemap)	AMP	-5.116	-34.722	Gly 70, Pro 74, Gln 84, Thr 136
2OFZ (Sitemap)	UMP	-4.473	-35.987	ASN 76, Gln 84, Thr 136
1SSK (Sitemap)	AMP	-5.028	-20.660	Arg 85, Tyr 87, Arg 127, Arg 155
1SSK (Sitemap)	UMP	-5.508	-22.266	Thr 27, Gly 125 Arg 70, Phe 88

Score doesn't correlate with my known activities. What is wrong? | Schrödinger, 2020, What do all the Prime MM-GBSA energy properties mean? | Schrödinger, 2020).

Pharmacokinetics and physicochemical properties and drug likeness of the ligands

The N protein NTD co-crystallized models (15 molecules for each 2OFZ and 1SSK) screened with the highest docking score were submitted for evaluation of the physicochemical and kinetic properties in QikProp module of Schrödinger. Drug-likeness of the hits were evaluated using values obtained from SwissADME (SwissADME, n.d.).

Molecular dynamics (MD) simulation

MD simulations can provide in-depth knowledge of the dynamic behavior of bio-molecules in a graphical simulation which can demonstrate a free-energy landscape that is close to the native protein state inside the body. Hence employing the docked complexes to MD simulations provide a more accurate interaction profile.

Molecular dynamics studies of the selected ligands were carried out against the chosen target structures. Module Desmond was used to carrying out MD simulations. In order to study the interactions of three targeted proteins with the ligands, the optimized potentials for the liquid simulations (OPLS)-2005 force field was employed. Firstly, a position restraint of 6000 ps was selected for the drug-target complex that allows water molecules in the system. Secondly, various frames were applied to minimize the complex upon which the force field is implied. Consequently, the root means square deviation (RMSD) for the C_{α} , Ligand root mean square fluctuation (RMSF), Ligand contacts were obtained from the data to monitor the stability of three protein in its dynamic form along the simulated trajectory (Desmond, n.d.).

Results and discussion

Virtual screening

We used two PDB structures for virtual screening (2OFZ and 1SSK). Fifteen hits of each were selected on the basis of docking score. The details of the screened molecules are shown in Table 2 (for 2OFZ) and Table 3 (for 1SSK).

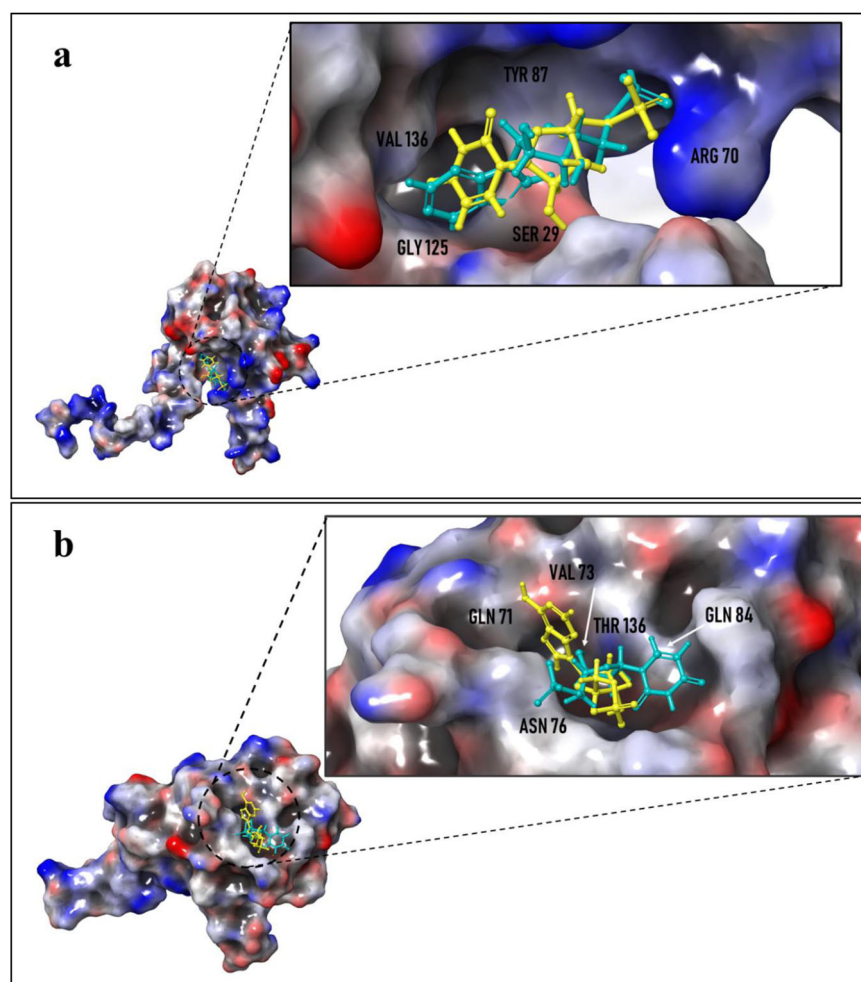


Figure 2. Potential energy surface structure with an elaborate display of binding sites of 1SSK (a) and 2OFZ (b) with in place docked poses of AMP (cyan) and UMP (yellow). The position of the interacting ligands can be noticed with faded regions indicating the non-polar clefts of CoV N protein NTD.

MM-GBSA: In the case of target 2OFZ, the top five ligands with the highest MM-GBSA score were ligand $1 > 7 > 3 > 11 > 9$. The highest MM-GBSA score was found to be -54.798 . In the case of the target 1SSK, Poor binding affinity was seen in the case of all ligands with the highest MMGBSA (ΔG_{bind}) found to be -44 . Ligands with the highest MMGBSA score were ligand no $8 > 5 > 12$.

Pharmacokinetics and physicochemical properties and drug likeness of the ligands and selection of hits for further molecular dynamic simulation studies

A log p value of 1–3 may indicate optimal physicochemical properties and a logP value more than five results in rapid turnover rate, low solubility and poor absorption. Log p not greater than 5 is a component of Lipinski's rule (20). In our study, as the target is intracellular (nucleocapsid protein), the lipophilic nature of the ligand will be very important and we kept the range of 1 to <5 .

In the case of 2OFZ, among the 15 molecules selected (2OFZ-ligand complexes), 8 ligands had log p value between 1 and <5 (ligands 1, 4, 5, 7, 8, 11, 14 and 15). Among the ligands with log p value in the range 1 to <5 , taking an arbitrary cutoff MM-GBSA score of -45 , we found that the

binding affinity was weak (-45) in case of ligands 4, 8, 14 and 15. Again ligand 14 showed strong HERG inhibition. So, on the basis of lipophilicity and binding affinity, the best performing ligands were ligand 1, 5, 7 and 11. The selected complexes were advanced for MD simulations on discrete basis of docking score (ligand no 1, 2 and 3), on the basis of lipophilicity (log p) and binding affinity (MMGBSA score) (ligands 1, 5, 7 and 11). So, we conducted molecular dynamics study of a total of 6 ligands (ligand 1, 2, 3, 5, 7 and 11).

In the case of 1SSK, among the 15 molecules selected (1SSK-ligand complexes), 9 ligands had log p value between 1 to <5 (ligands 2, 3, 4, 6, 7, 8, 9, 12 and 14). Among these, ligand no 6 showed strong potential for HERG inhibition and ligand 3 showed additional rule of three violations. The rest of the molecules can be arranged in the order $8 > 12 > 9 > 2 > 4 > 14 > 7$ on the basis of MM-GBSA. Top four molecules (8, 12, 9 and 2) were selected for MD simulation studies as the rest had poor MM-GBSA scores.

Molecular dynamics

Molecular dynamics simulation for the drug ligand complexes of Nucleocapsid protein N-terminal domain (PDB structure variants 2OFZ and 1SSK) were referred for the

Table 2. Virtual screen output for computationally potent hits for the binding site of 20FZ sorted on the basis of good complex interaction.

Log No	ZINC id	Docking score	MMGBSA (ΔG_{bind})	QPo/w	QPo/w	HERG	QPo/w	HERG	QPP	Caco	BBB permeab	QPo/w	QPo/w	logKhsa	P-gp inhibitor	% Human Oral Absorption	Rule Of Five	Rule Of Three	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	Leadlikeness
1	ZINC000012159837	-7.158	-54.798	1.039	-3.481	97.946	No	-0.149	Yes	68.664	0	No	No	No	No	0	0	No	No	No	No	No	1
2	ZINC000000141585	-6.865	-36.646	-1.087	-2.238	31.671	No	-0.918	No	47.441	0	No	No	No	No	0	0	No	No	No	No	No	1
3	ZINC000000318440	-6.728	-52.109	-0.107	-3.938	142.862	No	-0.651	No	64.886	0	No	No	No	No	0	0	No	No	No	No	No	1
4	ZINC000000134604	-6.387	-40.163	1.36	-4.179	474.366	Yes	-0.322	No	82.806	0	Yes	No	No	No	0	0	Yes	No	No	No	No	1
5	ZINC000012159732	-6.326	-45.194	1.466	-4.548	69.216	No	-0.032	Yes	68.466	0	No	No	No	No	0	0	No	No	No	No	No	1
6	ZINC0000003118446	-6.064	-44.878	0.449	-4.478	113.203	No	-0.445	No	66.336	0	No	No	No	No	0	0	No	No	No	No	No	0
7	ZINC000012159562	-5.964	-54.648	1.298	-4.903	95.894	No	-0.095	Yes	70.018	0	No	No	No	No	0	0	No	No	No	No	No	1
8	ZINC000012159729	-5.963	-40.331	1.483	-3.973	119.348	No	-0.082	Yes	72.802	0	No	No	No	No	0	0	No	No	No	No	No	1
9	ZINC000012157799	-6.008	-46.665	0.92	-3.331	27.983	No	-0.251	Yes	58.23	0	No	No	No	No	0	0	No	No	No	No	No	1
10	ZINC00000040052	-5.867	-45.259	0.898	-2.745	749.459	No	-0.414	No	83.654	0	No	No	No	No	0	0	No	No	No	No	No	1
11	ZINC000000146942	-5.856	-47.384	3.988	-4.397	815.316	Yes	0.709	No	100	0	No	No	No	No	0	0	No	No	No	No	No	0
12	ZINC000000135562	-5.856	-39.496	0.545	-4.208	284.361	No	-0.429	No	74.054	0	No	No	No	No	0	0	No	No	No	No	No	1
13	ZINC000012159395	-5.753	-44.640	0.602	-4.418	32.783	No	-0.199	Yes	57.596	0	No	No	No	No	0	0	No	No	No	No	No	1
14	ZINC000019285209	-5.668	-44.604	2.328	-5.534	376.55	No	0.093	Yes	86.679	0	No	No	No	No	0	0	No	No	No	No	No	0
15	ZINC000018202188	-5.632	-34.936	2.021	-4.04	536.499	No	0.003	No	87.633	0	No	No	No	No	0	0	No	No	No	No	No	1

QPo/w Po/w – Predicted octanol/gas partition coefficient (-2.0-6.5).

QPo/w HERG – Predicted IC50 value for the blockage of HERG K + channels (concern below -5).

QPPCaco – Predicted apparent Caco-2 cell permeability in nm/sec. (25 poor, > 500 great).

QPo/wBBB – Predicted brain/blood partition coefficient (-3.0-1.2).

QPo/wKhsa – Prediction of binding to human serum albumin. (-1.5-1.5).

For Leadlikeness and drug-likeness (Rule of five and three) – 0 = YES, 1 or above = NO.

For the sake of convenience of referring the molecule, we used the serial number used as is in-text.

Table 3. Virtual screen output for computationally potent hits for the binding site of 15SK sorted on the basis of good complex interaction.

Log No	ZINC id	Docking score	log Po/w	MM-GBSA (ΔG_{bind})	QPo/w	HERG	P	Caco	BBB permeab	logKhsa	P-gp inhibitor	% Human Oral Abs	Rule Of Five	Rule Of Three	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	Leadlikeness	
1	ZINC000003833811	-7.165	-3.174	-24.61015	-1.749	10.398	No	-6.223	Yes	13.602	1	No	No	No	No	No	No	No	No	1
2	ZINC000000156452	-6.813	1.64	-32.41165	-3.016	41.339	No	-3.959	No	65.478	0	No	No	No	No	No	No	No	No	0
3	ZINC000019924840	-6.725	3.109	-31.91928	-4.827	549.135	Yes	-2.978	Yes	94.186	0	Yes	Yes	Yes	Yes	Yes	Yes	No	No	0
4	ZINC000000147804	-6.622	1.723	-27.69639	-2.854	42.609	No	-3.103	No	66.197	0	No	No	No	No	No	No	No	No	0
5	ZINC000004325120	-6.459	0.766	-40.18159	-4.476	70.305	No	-4.1	No	64.486	0	No	No	No	No	Yes	Yes	No	No	0
6	ZINC000012410539	-6.215	1.516	-29.58703	-6.478	266.316	Yes	-5.081	Yes	79.229	0	Yes	Yes	Yes	Yes	No	No	Yes	Yes	1
7	ZINC000000141114	-6.183	1.598	-17.94346	-4.126	1519.174	Yes	-2.236	No	93.246	0	No	No	No	No	No	No	No	No	1
8	ZINC000000139516	-6.067	3.194	-44.13411	-4.312	1832.438	Yes	-0.988	Yes	100	0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	0
9	ZINC000000148164	-6.061	4.11	-34.84414	-4.762	1577.296	Yes	-1.346	Yes	100	0	Yes	Yes	Yes	Yes	Yes	Yes	No	No	1
10	ZINC000000146622	-6.006	0.612	-33.01976	-2.52	22.486	No	-4.627	No	54.725	0	No	No	No	No	No	No	No	No	0
11	ZINC000000135628	-5.99	0.406	-31.47537	-4.077	186.461	No	-4.331	No	69.961	0	No	No	No	No	No	No	No	No	0
12	ZINC000000158172	-5.806	3.497	-36.75966	-4.545	1679.706	Yes	-1.623	Yes	100	0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	0
13	ZINC000019520618	-5.736	5.171	-27.43031	-4.96	1713.442	No	-2.547	Yes	100	1	No	Yes	Yes	Yes	No	No	Yes	No	2
14	ZINC000000146742	-5.692	2.471	-23.93204	-4.773	1032.431	Yes	-1.31	Yes	95.358	0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	0
15	ZINC000003833811	-7.165	-3.174	-30.51850	-1.749	10.398	Yes	-6.223	Yes	13.602	1	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	1

QPo/w Po/w – Predicted octanol/gas partition coefficient (-2.0-6.5).

QPo/w HERG – Predicted IC50 value for the blockage of HERG K + channels (concern below -5).

QPPCaco – Predicted apparent Caco-2 cell permeability in nm/sec. (25 poor, > 500 great).

QPo/wBBB – Predicted brain/blood partition coefficient (-3.0-1.2).

QPo/wKhsa – Prediction of binding to human serum albumin (-1.5-1.5).

For Leadlikeness and drug-likeness (Rule of five and three) – 0 = YES, 1 or above = NO.

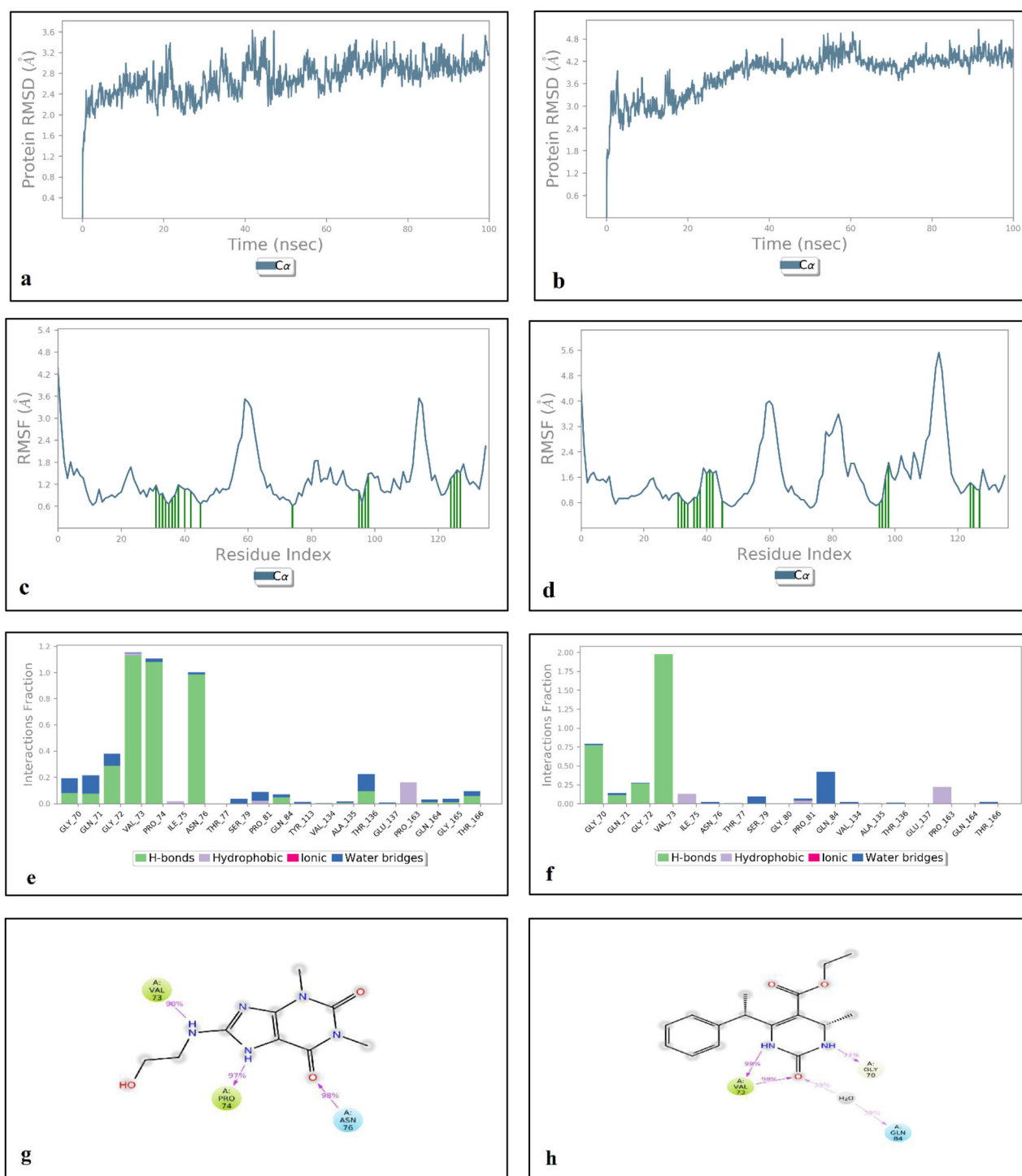


Figure 3. MD simulation interaction diagrams of 100 ns trajectory showing RMSD; Root mean square fluctuation of essential residue with viable ligand contacts; Residue-ligand interaction profile and interaction strength of the ligand in the binding cavity for SARS-CoV N protein RNA binding domain (2OFZ) with top hits ZINC00003118440 (a, c, e, g) and ZINC0000146942 (b, d, f, h), respectively in order.

dynamic profile assessment of N-protein interaction with the selected ligands (selected on the basis of MM-GBSA, Log P, HERG inhibition potential, rule of 5, rule of 3 and drug likeness rules). Considering the low structural constancy of the protein model itself, the 1SSK N-protein molecule simulation model was shed off from the picture as it was highly unstable with avg. RMSD(s) ranging between 10 and 12 Å (Supplementary Figure 1). On the contrary, selected molecules from N-protein (PDB id: 2OFZ) complexes revealed few

excellent binders. Lead molecules ZINC00003118440 and ZINC0000146942 emerged out to be high in interaction ratio and strength with N-protein active binding residues (Figure 3) having avg. RMSD of 2.8 Å and 4.0 Å. Protein side chain RMSF is positively stable in the two complexes with the local ligand-contact maxima at 1.8 Å and 2.2 Å for ZINC00003118440 and ZINC0000146942, respectively.

Interestingly, residues Gly 70, Val 73 and Pro 74 are the prime binding residues of N-protein that lie common with

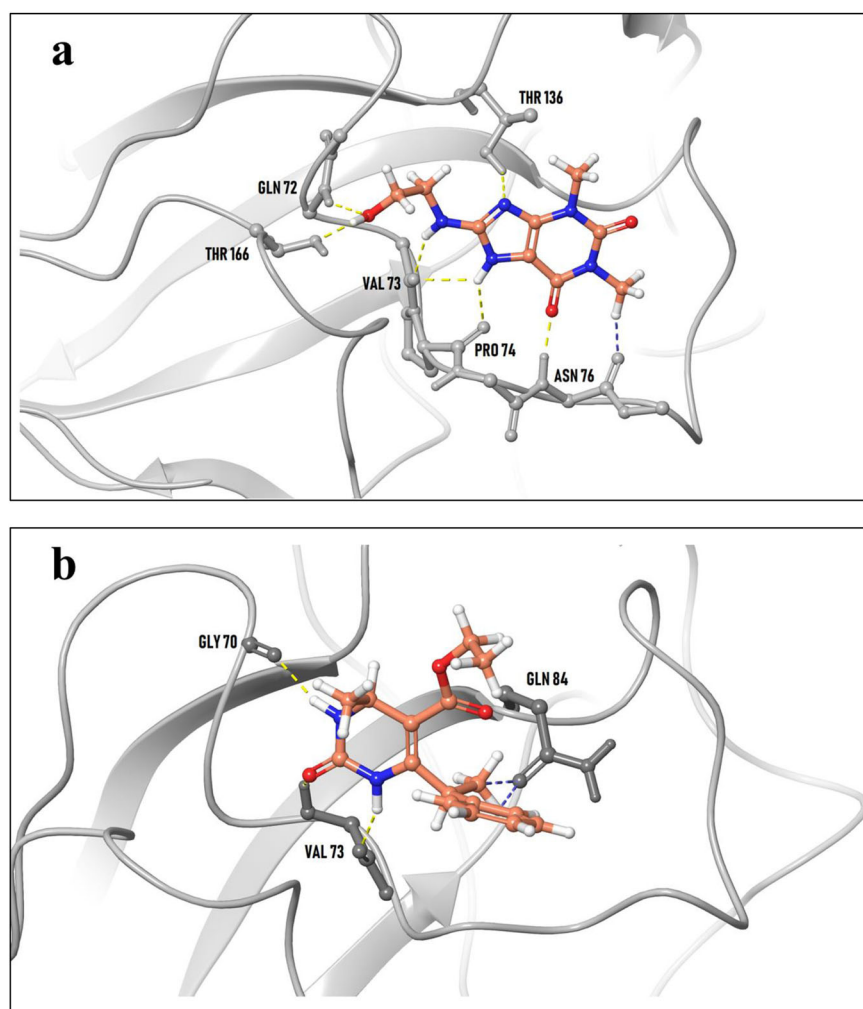


Figure 4. In place docking poses of most potent hits ZINC00003118440 (a) and ZINC0000146942 (b) with SARS-CoV N protein NTD (2OFZ).

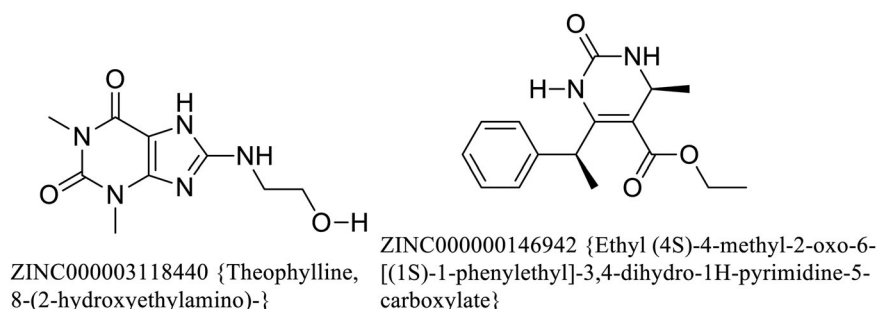


Figure 5. Structure of the two identified hits.

the top dynamically stable hit complexes and the nucleotide molecules AMP and UMP. The in-place docking poses of top hits ZINC00003118440 (left) and ZINC0000146942 (right) with SARS-CoV N protein NTD (2OFZ) is shown in Figure 4. The MD simulation profile for the ejected complexes can be found in supplementary figures 1 and 2 accompanying the movie file of the 2OFZ-ZINC00003118440 complex (Supplementary material Movie 1). One essential feature that compounds ZINC00003118440 displays in the molecular dynamics is its commitment to the binding site throughout

the trajectory, there was no leaving off the binding site accompanied by in-range protein fluctuation.

Identification of the top hits

The hit molecule ZINC000003118440 [8-(2-hydroxyethyl) aminophylline] is a synthetic derivative of the drug Theophylline (detailed structure showed in Figure 5). Theophylline is a bronchodilator of the Methylxanthine class. Few studies also reported the antiviral properties of Methylxanthines

Table 4. Binding affinities of approved Bronchodilator and Pyrimidone drug molecules with potential antiviral attributes to the RNA binding domain of 2OFZ.

Drug class	Drug name	Docking score	MM-GBSA
Beta-2-sympathomimetics	Salbutamol	-3.618	-32.862
	Terbutaline	-0.505	-54.513
	Bambuterol	-3.866	-31.47
	Salmeterol	-3.91	-33.54
	Formeterol	-1.336	-60.217
	Ephedrine	-3.009	-21.015
Methylxanthines	Theophylline	-3.763	-39.464
	Aminophylline	No significant docking	NA
	Choline theophyllinate	No significant docking	NA
	Hydroxyethyl theophylline	No significant docking	NA
	Theophylline ethanolate of piperazine	No significant docking	NA
	Doxophylline	No significant docking	NA
Anti-cholinergics	Ipratropium bromide	-3.669	-41.66
	Tiotropium Bromide	-3.055	-41.444

(Yamazaki and Tagaya, 1980). However, if we see specifically for theophylline or theophylline derivatives, only one study reported that theophylline possesses antiviral activity against the hepatitis B virus (Zheng et al., 2011) and no further data is available against other types of viruses.

In the management of COVID-19, the interim guidelines encourage the use of selective M1/M3 receptor blockers (anti-cholinergic agents) to reduce the secretion of lung glands, reduce spasm and associated wheeze and improve the respiratory function (Jin et al., 2020). As the best hit was a theophylline derivative, which is a bronchodilator, and bronchodilators are often used in the management of viral lung diseases, we further screened all approved bronchodilator drugs Tripathi (2013) for their binding to the N protein NTD (RNA binding domain). The detailed data is shown in Table 4. In addition to other activities, our study also demonstrated the binding affinity of the different bronchodilators against the NTD (RNA binding region) of the N protein of coronavirus. The binding affinity is in the order of Formeterol > Terbutaline > Ipratropium bromide > Tiotropium Bromide > Theophylline > salbutamol. One of the interesting facts is that formeterol also showed binding against 2019-nCoV PL protease (Zhavoronkov et al., 2020). A preliminary *in-vitro* assay may provide further details and may pave the way for a drug with dual action (antiviral and bronchodilator) and guide us on the choice of bronchodilators in case of COVID-19. However, toxicity and the presence of other comorbidity parameters will also govern the choice of bronchodilators.

The second compound ZINC000000146942 (Ethyl (4S)-4-methyl-2-oxo-6-[(1S)-1-phenylethyl]-3,4-dihydro-1H-pyrimidine-5-carboxylate) is a derivative of 3,4 Dihydropyrimidone. 3,4 Dihydropyrimidinones (DHPMs) are implicated in a wide range of biological activities. Pyrimidone derivatives are already being used against viral infections (Phucho et al., 2009; Wierenga et al., 1985; Skulnick et al., 1985; Sharma et al., 2014; Seley-Radtke et al., 2018). The pyrimidone nucleus is a core component of many of the anti-retroviral drugs (Sharma et al., 2014). The pyrimidone scaffold is a backbone of many of the approved anti-retrovirals e.g. Zidovudine, Didanosine and Zalcitabine (Seley-Radtke & Yates, 2018). Other pyrimidone derivatives 5-iododeoxyuridine and 5-iodo-2'-deoxyuridine are extensively used against viral infections (Sharma et al., 2014).

Thus it sounds likely to find a derivative in a screening study targeting SARS CoV.

Conclusion

We identified two potential hits ZINC000003118440 and ZINC000000146942 as inhibitors of RNA binding to NTD of N protein. The first hit is a theophylline derivative and the second hit is a pyrimidone derivative. Both the classes have reported to have antiviral effect against other viruses. However, this is the first time the binding potential of these two hits are reported against N protein NTD (RNA binding region) of SARS-CoV-2. Interestingly, the first compound was a theophylline derivative (commonly used bronchodilator). Hence, we screened all the approved bronchodilators against the N protein RNA binding site, which showed binding affinity (MMGBSA) in the order Formeterol > Terbutaline > Ipratropium bromide > Tiotropium Bromide > Theophylline > Salbutamol. These findings may help in the choice of bronchodilator in COVID-19. However these findings need *in-vitro* validation.

Acknowledgments

The authors thank Dr. Prajwal Nandekar and Mr. Vinod Devaraji of Schrodinger Corporation for their kind help and Mr. Nripendra Bhatta for his logistic help.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Phulen Sarma  <http://orcid.org/0000-0003-1926-7605>

Pramod Avti  <http://orcid.org/0000-0001-5603-4523>

Bikash Medhi  <http://orcid.org/0000-0002-4017-641X>

References

- Adasme-Carreño, F., Muñoz-Gutierrez, C., Caballero, J., & Alzate-Morales, J. H. (2014). Performance of the MM/GBSA scoring using a binding site hydrogen bond network-based frame selection: The protein kinase case. *Physical Chemistry Chemical Physics*, 16(27), 14047–14058. doi:10.1039/C4CP01378F

- Chang, C., Lo, S.-C., Wang, Y.-S., & Hou, M.-H. (2016). Recent insights into the development of therapeutics against coronavirus diseases by targeting N protein. *Drug Discovery Today*, 21(4), 562–572. doi:10.1016/j.drudis.2015.11.015
- Chang, C., Sue, S.-C., Yu, T., Hsieh, C.-M., Tsai, C.-K., Chiang, Y.-C., Lee, S., Hsiao, H., Wu, W.-J., Chang, W.-L., Lin, C.-H., & Huang, T. (2006). Modular organization of SARS coronavirus nucleocapsid protein. *Journal of Biomedical Science*, 13(1), 59–72. doi:10.1007/s11373-005-9035-9
- Chenavas, S., Crépin, T., Delmas, B., Ruigrok, R. W. H., & Slama-Schwok, A. (2013). Influenza virus nucleoprotein: Structure, RNA binding, oligomerization and antiviral drug target. *Future Microbiology*, 8(12), 1537–1545. doi:10.2217/fmb.13.128
- Coronavirus Disease (COVID-19)—Events as they happen. (n.d.). Retrieved March 16, 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
- Desmond, S. (n.d.). *Desmond* | Schrödinger. Retrieved from <https://www.schrodinger.com/desmond>
- Franco, R., Oñatibia-Astibia, A., & Martínez-Pinilla, E. (2013). Health Benefits of Methylxanthines in Cacao and Chocolate. *Nutrients*, 5(10), 4159–4173. <https://doi.org/10.3390/nu5104159>
- Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W., & Olejniczak, E. T. (2004). Structure of the N-Terminal RNA-Binding Domain of the SARS CoV Nucleocapsid Protein. *Biochemistry*, 43(20), 6059–6063. <https://doi.org/10.1021/bi036155b>
- GlideScore/Docking Score doesn't correlate with my known activities. What is wrong? | Schrödinger [Internet]. Retrieved February 21, 2020, from <https://www.schrodinger.com/kb/144>
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., ... Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223), 497–506. doi:10.1016/S0140-6736(20)30183-5
- Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W., & Olejniczak, E. T. (2004). Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry*, 43(20), 6059–6063. [10.2210/pdb1ssk/pdb](https://doi.org/10.2210/pdb1ssk/pdb)
- Jin, Y.-H., Cai, L., Cheng, Z.-S., Cheng, H., Deng, T., Fan, Y.-P., ... Wang, X.-H. (2020). A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Military Medical Research*, 7(1).
- LigPrep, S. (n.d.). *LigPrep* | Schrödinger. Retrieved from <https://www.schrodinger.com/ligprep>
- Lin, S. Y., Liu, C. L., Chang, Y. M., Zhao, J., Perlman, S., & Hou, M. H. (2014). Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target. *Journal of Medicinal Chemistry*, 57(6), 2247–2257. doi:10.1021/jm500089r
- Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer Aided-Drug Design*, 7(2), 146–157. doi:10.2174/157340911795677602
- Nain, Z., Sayed, S. B., Karim, M. M., Islam, M. A., & Adhikari, U. K. (2019). Energy-optimized pharmacophore coupled virtual screening in the discovery of quorum sensing inhibitors of LasR protein of *Pseudomonas aeruginosa*. Available from <https://www.preprints.org/manuscript/201910.0074/v1>
- PDB Database. (n.d.). *RCSB PDB: Homepage*. Retrieved January 13, 2020, from <https://www.rcsb.org/>
- Phucho, I. T., Nongpiur, A., Tumtin, S., Nongrum, R., Nongkhaw, R. L. (2009). Recent progress in the chemistry of dihydropyrimidinones. *Rasayan J Chem*, 2(3), 662–676.
- Prajapat, M., Sarma, P., Shekhar, N., Avti, P., Sinha, S., Kaur, H., Kumar, S., Bhattacharyya, A., Kumar, H., Bansal, S., & Medhi, B. (2020). Drug targets for corona virus: A systematic review. *Indian Journal of Pharmacology*, 52(1), 56–65. doi:10.4103/ijp.IJP_115_20
- Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., & Kuhn, P. (2007). Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *Journal of Virology*, 81(8), 3913–3921. doi:10.1128/JVI.02236-06
- Sarma, P., Prajapat, M., Avti, P., Kaur, H., Kumar, S., & Medhi, B. (2020). Therapeutic options for the treatment of 2019-novel coronavirus: An evidence-based approach. *Indian Journal of Pharmacology*, 52(1), 1–5. doi:10.4103/ijp.IJP_119_20
- Seley-Radtke, K. L., & Yates, M. K. (2018). The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Research*, 154, 66–86. doi:10.1016/j.antiviral.2018.04.004
- Singh, K. D., & Muthusamy, K. (2013). Molecular modeling, quantum polarized ligand docking and structure-based 3D-QSAR analysis of the imidazole series as dual AT(1) and ET(A) receptor antagonists. *Acta Pharmacologica Sinica*, 34(12), 1592–1606. [10.1038/aps.2013.129](https://doi.org/10.1038/aps.2013.129)
- Sharma, V., Chitranshi, N., & Agarwal, A. K. (2014). Significance and biological importance of pyrimidine in the microbial world. *International Journal of Medicinal Chemistry*, 2014, p. e202784. Available from <https://www.hindawi.com/journals/ijmc/2014/202784/>
- Skulnick, H. I., Weed, S. D., Eidson, E. E., Renis, H. E., Wierenga, W., & Stringfellow, D. A. (1985). Pyrimidinones. 1. 2-amino-5-halo-6-aryl-4(3H)-pyrimidinones. Interferon-inducing antiviral agents. *Journal of Medicinal Chemistry*, 28(12), 1864–1869. doi:10.1021/jm00150a018
- SwissADME. (n.d.). Retrieved March 18, 2020, from <http://www.swissadme.ch/>
- Tripathi, K. (2013). *Essentials of medical pharmacology* (17th ed.). JAYPEE.
- Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223), 470–473. doi:10.1016/S0140-6736(20)30185-9
- What do all the Prime MM-GBSA energy properties mean? | Schrödinger. (n.d.). Retrieved February 22, 2020, from <https://www.schrodinger.com/kb/1875>
- Wierenga, W. (1985). Antiviral and other bioactivities of pyrimidinones. *Pharmacology & Therapeutics*, 30(1), 67–89. doi:10.1016/0163-7258(85)90048-8
- Yamazaki, Z., & Tagaya, I. (1980). Antiviral Effects of Atropine and Caffeine. *Journal of General Virology*, 50(2), 429–431. <https://doi.org/10.1099/0022-1317-50-2-429>
- Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D. S., Polykovskiy, D., Shayakhmetov, R., Filimonov, A., Orekhov, P., Yan, Y., Popova, O., Vanhaelen, Q., Aliper, A., & Ivanenkov, Y. (2020). Potential 2019-nCoV 3C-like protease inhibitors designed using generative deep learning approaches. [10.26434/chemrxiv.11829102.v1](https://doi.org/10.26434/chemrxiv.11829102.v1)
- Zheng, Z., Li, J., Sun, J., Song, T., Wei, C., Zhang, Y., Rao, G., Chen, G., Li, D., Yang, G., Han, B., Wei, S., Cao, C., & Zhong, H. (2011). Inhibition of HBV replication by theophylline. *Antiviral Research*, 89(2), 149–155. doi:10.1016/j.antiviral.2010.12.004
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382(8), 727–733. doi:10.1056/NEJMoa2001017