In silico investigation of phenolic derivatives from Silybum marianum against SARS-CoV-2 proteins as a pharmacological drug repurposing strategies to mitigate the pandemic.

Yasmin Ahmad¹, Swaraj Mohanty¹, Soumya Lipsa Rath², and Poornima Sharma¹

¹DRDO Defence Institute of Physiology and Allied Sciences
²National Institute of Technology Warangal

Abstract

Background and Aim: The boutade of coronavirus disease-2019(COVID-19) has a striking impact on the worldwide healthcare system within a very short period of time. Availability of a large number of clinical data on SARS-CoV-2, conventional precautionary majors, and treatment strategies with the existing therapeutic antiviral drug molecules are also failing to control progression and disease transmission among the population. Hence, we implemented pharmacoinformatics approaches to facilitate the drug discovery process by repurposing naturally available therapeutic molecules as an effective intervention.

Experimental Procedure: The major phenolic derivatives of Silybum marianum(Milk thistle) have been identified and investigated for ADME(Absorption, Distribution, Metabolism and Excretion)/tox properties. Co-crystallized structure of three major proteins(i.e. Main protease, RNA binding domain of nucleocapsid phosphoprotein and Spike receptor binding domain) from SARS-CoV-2 investigated with molecular docking(MD) interaction with the phenolic compounds from milk thistle. Furthermore, based on ADMET and MD interaction a 100ns MD simulation was performed with silibinin molecule.

Outcomes: Being less toxic in ADME, a good MD interaction and stability of silibinin molecule across the MD simulation trajectories with targeted proteins explicate that silibinin molecule can be a promising drug candidate against the main protease and will be helpful to cease the enzymatic activity in viral replication and transcription.

Introduction:

In late 2019, the outbreak of COVID-19 completely changed the global healthcare scenario. This highly communicable disease created an emergency healthcare condition across the globe. The countries were also affected during this global pandemic due to poor hospital management conditions and economic crashes. Nevertheless, considering the urgency of the situation, companies worldwide worked towards generating vaccines and alternative therapeutics with a global objective of eradicating the disease. However, this objective could not be achieved completely due to emergence of new variants of the virus and the increasing drug resistance and escape strategies of the virus.

Developing a new drug or a vaccine is a tedious process, from drug discovery, in-vitro screening, preclinical and clinical trials. To gain FDA approval requires a minimum of 12 to 15 years, which is too long a time considering the mortality rate in COVID-19. An alternative strategy which can boost the process is to repurpose existing drugs. Repurposing FDA approved drugs drastically reduces the time and costs incurred in manufacturing. Major flavonolignans such as Silibinin, Isosilybin, Silychristin and Silidianin from Silybum marianum are most prominently known for its medicinal properties to cure different disease states of liver like chronic liver cirrhosis [1], fibrosis [2], necrosis [3, 4] and hepatocellular carcinoma [5, 6]. This compound is also well known for the anti-inflammatory, anti-oxidant, anti-carcinogenic and anti-mutagenic properties. Silimarin on the rat as an animal model demonstrated that it helps in prevention of
free radical generation of reactive oxygen species (ROS), triggers the antioxidant defense system, improves the anti-inflammatory response and activates the key genes that act as compensatory adaptive vascular response during the hypoxic condition [1, 7]. As silymarin is in the clinical trial pipeline and its less toxic properties is a unique characteristic which can ensure it as an effective drug molecule for different disease, we evaluated the efficacy of the molecule towards main protease, spike receptor binding domain and RNA binding domain of nucleocapsid phosphoprotein enzymes of SARS-CoV-2 [8]. The COVID-19 disease basically affects the endothelial lining of the lungs and the disease progression also relates to the physiological changes as in case of high-altitude hypoxia illness [9]. Hence, we want to check the possibility of Silibinin (a major constituent of silymarin) as an effective inhibitor.

The SARS-CoV-2 genome consists up 12 open reading frames (ORFs), 9 transcription regulatory sequence, 9 conserved leader sequences and 2 un-translated regions (UTRs) [10] (Figure 1). The multiple sequence alignment data of the SARS-CoV-1 & SARS-CoV-2 shows the sequence similarities and it is believed that 10-28 nucleotides in the 5’ UTR interacts with the non-structural protein 1 (Nsp1). Targeting the major viral protein to stop the progression of pathogenesis needs the clear understanding of the viral genome and its translational units that interacts with different host machinery starting from the viral entry till hijack of the cell immunity [11]. Among the four structural proteins of the SARS-CoV-2 virus we have targeted the binding sites of major proteins i.e. RNA binding domain of the nucleocapsid of phosphoprotein and the receptor binding domain of spike protein. From the virology aspects main proteases that plays a major role in viral proprotein maturation and assembly of other components to produce a whole virus after the translational event inside the host cell we also targeted the cysteine like protease i.e. main protease of the ORF1a region [8, 12]. Hence investigation to find the major inhibitory molecule that can actively bind and block the enzymatic activity of the major proteins of SARS-CoV-2 is the necessity to break the chain of the disease transmission from the pharmacological point of cure.

Materials and Methods:

Molecular Docking and ADME/tox Analysis:

The 3D structural data files (SDFs) of ligands (CID: 31553, 3085830, 441764 & 1982272) of Silybum marianum were taken from PubChem database (www.pubchem.ncbi.nlm.nih.gov) and checked for the pharmacokinetics and toxicity properties with ADMElab 2.0 (https://admetmesh.scbdd.com/). This ADMElab 2.0 tool considered to efficiently calculate and predict 17 physiochemical, 13 medicinal chemistry properties, 23 different ADME, 27 toxicity endpoints along with the 8 toxicophore rules.

The crystallographic structures of the proteins of SARS-CoV-2 were imported from protein data bank (PDB ID: 6LU7, 6VYO & 7BZ5). The ligands were transformed into PDBQT file format by using Open Babel (Version 3.1.1) tool [13]. The refinement of protein structures was done with AutoDock Vina (version 1.2.0) tool developed at The Scripps Research Institute [14]. The protein structures were imported and the water molecules present in crystal structure data were removed. Interacting small molecules were also removed from structure, missing residues and polar hydrogen molecules were appended in structure. The refined structures were subjected for energy minimization and then grid box with a dimension of X, Y & Z coordinates= 126.0 and spacing 0.375 Å was used for blind docking. The center of the grid was fixed as: center_x = -26.286, center_y = 12.608 & center_z = 58.965 for 6LU7, center_x = -14.253, center_y = 41.224 & center_z = 14.164 for 6VYO and center_x = -72.527, center_y = -29.929 & center_z = 11.449 for 7BZ5. Gasteiger charges were added to the protein before docking. All rotatable bonds of ligands were kept flexible while the protein remained rigid during the docking.

Molecular Dynamics Simulation:

Based on the strong docking interaction profile and the lesser toxicity for the silibinin compound in the ADMET results we focused to study the molecular evolution and the conformational changes of macromolecules (PDB ID: 6LU7, 6VYO & 7BZ5) with the ligand molecule during the 100ns molecular dynamics simulation. Molecular Dynamics (MD) Simulation was performed by using GROMACS (version 2020.4) MD simulation software package [15, 16]. Charmm36-Jul2020 force field parameters for protein was used for the study [17,
The topology and parameters for ligand molecules were generated using the CHARMM General Force Field Server [19]. The complexes were placed inside a water box filled with TIP3P water molecules and neutralizing ions. The box edge was kept at a distance of 1nm from the protein-ligand complex to avoid calculation artifacts. Initial minimization was performed using steepest descent and conjugate gradient algorithms followed by equilibration at an isothermal-isobaric ensemble. A time step of 2 fs was considered during the simulation and a trajectory of 100ns was generated to obtain data for analysis. Pymol [20, 21] and BIOVIA Discovery studio visualizer [22] were used to analyze structures generated during study.

**Results and Discussions:**

**Description of Silibilin & its derivatives:**

**ADME/Tox Properties:**

The absorption of any oral drug molecule takes place at the intestinal cells before releasing into the systematic circulation with active and passive diffusion process. Hence different in vitro models have been developed for the permeability study and oral bioavailability. Likewise, the distribution of the drug after administration has also been calculated along with the concentration to volume ratio, plasma protein binding along with the side effect in the central nervous system (CNS). The metabolism of any drug molecule occurs in liver by means of oxidative reaction and conjugative reaction, isozymes of cytochrome P450 family has been studied. Clearance of the drug molecule, its half-life and toxicity of the drug are also a crucial parameter to understand the pharmacokinetic of drug [23]. The effect of the chemical molecules can be well understood from the values (Table: 1) and the explanation of different properties are supplied as supplementary datasheet (Supplementary datasheet).

**Other Characterization:**

The Physiochemical and medicinal chemistry property of ligand molecules are properties satisfying which the molecules can be called as potent drug candidate and needs to further clinical and non-clinical studies to fall under drug development pipelines. The drug-likeness properties of the ligand molecules can better be understood from the bioavailability radar, and the parameters (like molecular weight, Van der walls volume, density, number of hydrogen bond acceptors, number of hydrogen bond donor, number of rotatable bonds, number of rings, number of atoms in biggest ring, number of heteroatoms, formal charge, number of rigid atoms, flexibility, number of stereocenters, topological polar surface area, logarithm of aqueous solubility value, logarithm of the n-octanol distribution co-efficient) were studied and represented (Table: 2) (Figure 2).

Silibilin as a potential drug molecule:

**Molecular Docking Results:**

The molecular docking of different proteins of SARS-CoV-2 with flavonoid compounds show good binding scores and interaction with the active residue of proteins with binding site of ligand molecules (Table 3). The analyzed from the docked complex with the help of PLIP [24] and BIOVIA discovery studio visualizer [22]. The 3D and 2D interactions of the complexes have been represented (Figure 3, 4 and 5).

**Bioactivity Score Analysis:**

The four lead molecules Silibinin, Isosilybin, Silychristin and Silidianin were subjected to bioactivity score analysis based on the parameters like GPCR, ICM, NRL and inhibitory enzymes (Protease and Kinase). The scores value greater than 0.00 considered to be highly active, values ranging from -0.50 to 0.00 were moderately active and scores less than -0.50 were inactive [25]. Based on the analysis score predicted Silibinin and Isosilybin were as effective due to its higher enzymatic activity inhibitor hence can be taken forward into the further drug development (Table 4).

**Stability of silibilin in protein pockets by MD simulations:**
Molecular dynamic simulation of 100ns were analysed by using different module of GROMACS with the pre-set algorithm and the results have been explained.

Protein characterization:

*Root Mean Square Deviation (RMSD):*

The root mean square deviation (RMSD) trajectory for simulation run for three macromolecules with silibinin molecule shows overall stability of protein-ligand complex during the binding with active site amino acid residues of the protein molecules. The protein molecules are globular in nature so the acceptable deviation ranges within a difference of 1-3 Å. From the C-α RMSD plot of the main protease (PDB ID: 6LU7) we can observe the trajectory lies within the range with very minimal deviation and after 80ns up to 100ns both the protein and the ligand RMSD shows a good binding affinity as in equilibrium and without any fluctuation. Likewise if we look into the trajectory of the nucleocapsid protein (PDB ID: 6VYO) we can clearly understand that the protein molecule is very stable and the fluctuations are within 2.5 Å and most importantly during the complete 100ns simulation, and the ligand molecular affinity shows a flattened trajectory from 78ns up to 97ns. And when we considered the RMSD trajectory for spike receptor binding domain (PDB ID: 7BZ5) with the ligand complex we have observed the protein molecule gradually came to equilibrium in fluctuation from 68ns up to 100ns and surprisingly the ligand fluctuation is within 0.5 Å throughout the 100ns simulation run (Figure 6a).

*Root Mean Square Fluctuation (RMSF):*

The root mean square fluctuation (RMSF) of protein and ligand molecules gives the idea of each residues present in molecular structure and this calculation generates a trajectory which denotes the flexibility of individual amino acid residues and atoms during the simulation process. If we observe the RMSF trajectory of the macromolecules we can clearly observe the flexibility of all 306 amino acid residues of main protease (6LU7), 128 amino acid residues of the nucleocapsid protein domain (6VYO) and 229 amino acid residues of the spike protein (7BZ5) (Figure 6b). The first 3 amino acid residues of the 6LU7 showed a high fluctuation of 6 Å but the fluctuation reduced to within a difference of range within 3 Å for other residues throughout the simulation run. In case of the 6VYO protein the fluctuation is very minimal and ranges in between 0.5 – 2.5 Å which indicates that there is no more stretching of the bonds formed between the active site residues of the complex. The trajectory of the 7BZ5 molecule also shown the minimal fluctuation for maximum number of residues during the simulation process.

*Principal Component Analysis (PCA):*

The protein-ligand interaction results a thousand of possible poses during the simulation process and it is very difficult to analyse each and every pose without a statistical tool. PCA is used as a mathematical tool to detect the correlation between a large set of datasets, its biological applications are to detect the flexible regions in a protein which hinder the equilibrium state of protein. This PCA can be helpful to integrate physical models of protein motions after removal of translational and rotational movements of atoms when interacted with drug molecule. The distribution of atoms during the simulation process has been represented in form of dot Cartesian trajectory coordinates from most eigenvector values. From our result interpretation we can clearly predict that during the interaction of drug molecule with nucleocapsid phosphoprotein and spike proteins shows very minimal changes whereas structure of main protease has been distorted with a large variation. The eigenvalues and eigenvectors of covariance matrix were diagonalized with first two principal components i.e. PC1 & PC2 and first 25 eigenvectors were considered (Figure 7). We can conclude the silibinin drug shows maximum effectiveness against the main protease (i.e. 6LU7) and can be used as a target against the maturation of viral accessory polyproteins inside the host cells and hence be helpful for the retardation of viral proliferation [26-28]. When we will observe the PCA plot for nucleocapsid phosphoprotein and spike glycoprotein we can state that residues of proteins shown less movement during simulation process and hence prefers to remain in natural state.

*Secondary Structure (SS):*
During the protein-ligand interaction, secondary structural changes occur in protein molecules along with evolution of time using DSSP command line tool of GROMACS. The changes in the elements like helix and beta-sheets were thoroughly observed for all the residues of the three proteins and ligand binding complexes (Figure 8). A very minute fluctuation has been observed in the turns, $\alpha$-helix and 3-10 helix of 6LU7(main protease). In case of 6VYO(nucleocapsid phosphoprotein) we can clearly visualize that the turns, $\beta$-sheets, $\beta$-turns, 3-10 helix and coils shows a very minimal alteration in natural structures. Similarly, if we consider the 7BZ5(spike protein) we have observed the distortion in secondary segment as compare to the natural structure in the turns, $\alpha$-helix, $\pi$-helix, 3-10 helix whereas $\beta$-sheets and $\beta$-turns shows the minimal fluctuations. Since the structural integrity of protein depends on backbones of protein and secondary structure of protein plays a vital role in protein folding and misfolding [29]. In case of the intervention of drug molecules the distortion of protein structure occurs and protein functions and bioactivity get disturbed [30, 31]. The in silico interaction of the silibinin molecule with the main protease and the spike protein shows promising results in distortion of the secondary structures at different backbone residues and hence can affect the protein conformational changes leading to protein functionality [32].

Ligand characterization:

**Root Mean Square Fluctuation(RMSF):** We have also analysed ligand flexibility of the molecules of silibinin compound during simulation run the trajectory for all atoms and observed deviation falls within the difference of maximum 3 Å for all three different molecular binding sets and represented below (Figure 9). The more the deviation of the atoms more is the flexibility of the atoms to bind with active site residues of the protein molecules leading to stronger molecular interaction among protein and ligand molecules [33]. It is being observed that all three proteins formed efficient binding interaction with silibinin molecules during 100ns MD simulation process and hence can be considered as a good inhibitory molecule [34].

**Radius of Gyration(Rg):**

The radius of gyration is defined as distribution of atoms in a protein molecule around its axis. During binding of a lead compound with protein the conformational changes in structure are observed as compactness of molecule get disturbed due to different binding forces. The lesser deviations in values from the central axis during simulation process more structural integrity of the molecules are preserved [35]. So the Rg plotted below shows a very less fluctuation within a difference in range of 0.5 - 1 Å (Figure 10a). This also predicts compactness of macromolecules is not getting disturbed when it is binding with silibinin inside the in silico environment [36].

**Solvent Accessible Surface Area(SASA):**

The protein-ligand binding is a solvent-substitution process in which the protein gets unfolded to provide surface area on the active site residues of the proteins when exposed to a solvent system. After molecular dynamics simulation it is highly necessary to compute the SASA which gives idea about how efficiently the computational system able to mimic the intracellular physiological environment [37]. The low fluctuation in the SASA value during the simulation process suggest the very good inhibitory action of the molecule against the protein targets [38, 39]. During the binding of the molecules with the ligand molecule parts of the proteins are buried inside the solvent and few portions are exposed to the solvent, the trajectory below represents that the fluctuation of the SASA values are within a difference range of 150 Å² and are the surface that are exposed into the solvent system (Figure 10b).

**Protein Ligand Interaction Profile:**

The molecular dynamics simulation different proteins of SARS-CoV-2 with silibinin compounds show good interaction during active pocket residue of proteins with binding site of ligand molecules (Table 5). The last frame complex during simulation process were analysed with the help of PLIP [24] and BIOVIA Discovery Studio Visualizer [22]. The 3D interaction of the complexes and the 2D interactions of complexes have been represented (Figure 11). We can conclude that among interaction of three macromolecules the maximum number of hydrogen bonding with bond orders less than 3.5 Å was seen between active site residues of main
protease and ligand molecule. Whereas number of hydrophobic interaction has been found between ligand and spike protein but since complete process is using water as a solvent system so we can elucidate that the hydrogen bonding to be the strongest interaction and more efficiently attachment of drug candidate to the target.

**Conclusions:**

It would be great aid to repurpose this medical emergency situation the surge in cases due to COVID-19 the worldwide researchers are striving to find an efficient drug candidate to alleviate the disease progression and treatments. These study outcomes can facilitate silibinin molecule as a promising molecule for binding to main protease and inhibit the formation of accessory viral assembly and replication of the SARS-CoV-2. And electrostatic interaction of silibinin molecule with spike protein of virus can impede interaction of ACE2 receptor of host by blocking receptor binding domain(RBD) of spike protein. Since COVID-19 shows critical pulmonary endothelial dysfunction by clogging atrial and venous fluid flow, so the reported anti-inflammatory and anti-thrombotic properties of silibinin molecule in treatment of hepatic diseases will be relaxing for complications during the disease progression. The reported study on size reduction and therapeutics intervention in animal model and its cell permeability data along with pharmacovigilance study are available as this drug is being used from many years in patients suffering from liver diseases. The candidate drug silibinin should be further tested into COVID-19 drug repurposing pipeline and if necessary the therapy may include other existing therapeutics based on the clinical condition of patient. We believe this silibinin as a lifesaving drug in case of COVID-19 management as a multitarget drug, without any harmful side effects and investment of huge financial cost, time and effort.

**Acknowledgement**

**Funding**

SM is a recipient of DRDO-SRF fellowship and PS is a recipient of DRDO-JRF fellowship. The authors have not received any special funding for this work apart from this institutional financial assistance.

**Conflict of interest**

Authors state no conflict of interests.

**Availability of data and materials**

The figure-1 has been adopted from respective author’s consent from a published article with the electronic permission, all other figures and tables have not been adopted from any available sources.

**Authors contribution**

YA conceptualized the study. SM performed the ADMET, molecular docking (MD) and wrote the entire manuscript with the necessary tables and sketched figures for convenient representation. SLR set the environment and carried out the molecular dynamics simulation(MDS) by using the computational resource facility of National Institute of Technology(NIT)-Warangal. SM and SLR equally contributed in this manuscript hence share the equal authorship. SM and PS checked the language and formatting of the manuscript. YA critically evaluated the manuscript.

**Ethics Approval**

Not applicable.

**Consent for Publication**

All authors agreed.

**Figure Legends:**
Figure 1: The genome organization of SARS-CoV-2 and its translational regions

Figure 2: The image represents the bioavailability radar for the ligand molecules Silibinin(a), Isosilybin(b), Silychristin(c) and Silidianin(d)

Figure 3: Interaction of main protease(6LU7) with different ligand molecules(Silibinin, Isosilybin, Silychristin & Silidianin) in 3D diagrammatic view(a) and 2D schematic view(b)

Figure 4: Interaction of RNA binding domain of nucleocapsid phosphoprotein(6VYO) with different ligand molecules(Silibinin, Isosilybin, Silychristin & Silidianin) in 3D diagrammatic view(a) and 2D schematic view(b)

Figure 5: Interaction of spike receptor binding domain(7BZ5) with different ligand molecules(Silibinin, Isosilybin, Silychristin & Silidianin) in 3D diagrammatic view(a) and 2D schematic view(b)

Figure 6: (a)Root mean square deviation(RMSD) trajectories of macromolecules(6LU7, 6VYO & 7BZ5)(BLACK) with silibinin molecule(RED) and (b)Root mean square fluctuation(RMSF) trajectories of macromolecules(6LU7, 6VYO & 7BZ5) during 100ns molecular dynamics simulation

Figure 7: PCA of different macromolecules(6LU7, 6VYO & 7BZ5) and eigen values of the covariance matrix during 100ns simulation process

Figure 8: Secondary structure of macromolecules(6LU7, 6VYO & 7BZ5) during 100ns MD simulation process

Figure 9: RMSF of silibinin ligand with different macromolecules(6LU7, 6VYO & 7BZ5) during 100ns MD simulation

Figure 10: Radius of Gyration(a) and solvent accessible surface area(b) of different macromolecules(6LU7, 6VYO & 7BZ5) during 100ns MD simulation process

Figure 11: Protein-ligand interaction profile(PLIP) of Silibinin with different macromolecules(6LU7, 6VYO & 7BZ5) in 3D(a) & 2D(b) schematic view representing different bonds and bond length after 100ns MD simulation last frame

References:


Hosted file

Figure legends for the article titled as: “In silico investigation of phenolic derivatives from *Silybum marianum* against SARS-CoV-2 proteins as a pharmacological drug repurposing strategies to mitigate the pandemic”
Graphical Abstract of the manuscript:

SARS-CoV-2 Virus

Machine learning approach can escalate the drug discovery process from available potential natural compounds

Pharmacoinformatics

Binds with host cell receptors in causing COVID-19 Disease

Phenolic derivatives from Silybum marianum

ADME/tox Property of Phenolic Derivatives from Milk thistle

Molecular Docking Analysis

100ns Molecular Dynamic Simulation Analysis
Figure 1: The genome organization of SARS-CoV-2 and its translational regions
Figure 2: The image represents the bioavailability radar for the ligand molecules Silibinin(a), Isosilybin(b), Silychristin(c) and Silidianin(d)
Figure 3: Interaction of main protease (6LU7) with different ligand molecules (Silibinin, Isosilybin, Silychristin & Silidianin) in 3D diagrammatic view (a) and 2D schematic view (b)
Figure 4: Interaction of RNA binding domain of nucleocapsid phosphoprotein (6VYO) with different ligand molecules (Silibinin, Isosilybin, Silychristin & Silidianin) in 3D diagrammatic view(a) and 2D schematic view(b)
Figure 5: Interaction of spike receptor binding domain(7BZ5) with different ligand molecules(Silibinin, Isosilybin, Silychristin & Silidianin) in 3D diagrammatic view(a) and 2D schematic view(b)
Figure 6: (a) Root mean square deviation (RMSD) trajectories of macromolecules (6LU7, 6VYO & 7BZ5) (BLACK) with silibinin molecule (RED) and (b) Root mean square fluctuation (RMSF) trajectories of macromolecules (6LU7, 6VYO & 7BZ5) during 100ns molecular dynamics simulation.

Figure 7: PCA of different macromolecules (6LU7, 6VYO & 7BZ5) and eigen values of the covariance matrix during 100ns simulation process.
Figure 8: Secondary structure of macromolecules (6LU7, 6VYO & 7BZ5) during 100ns MD simulation process.
Figure 9: RMSF of silibinin ligand with different macromolecules (6LU7, 6VYO & 7BZ5) during 100ns MD simulation.

Figure 10: Radius of Gyration (a) and solvent accessible surface area (b) of different macromolecules (6LU7, 6VYO & 7BZ5) during 100ns MD simulation process.
Figure 11: Protein-ligand interaction profile (PLIP) of Silibinin with different macromolecules (6LU7, 6VYO & 7BZ5) in 3D(a) & 2D(b) schematic view representing different bonds and bond length after 100ns MD simulation last frame.