Discovery of Natural Phenol Catechin as a Multitargeted Agent Against SARS-CoV-2 For the Plausible Therapy of COVID-19

Chandra Bhushan Mishra[1†], Preeti Pandey[2†], Ravi Datta Sharma [3†], Raj Kumar Mongre[1], Andrew. M.

Lynn[4], Rajendra Prasad [3,5], Raok Jeon[1]* and Amresh Prakash[5]*

[1]College of Pharmacy, Sookmyung Women's University, Hyochangwon gil-52, Yongsan-13 Gu, Seoul 140-742, Republic of Korea.
[2]Department of Chemistry & Biochemistry, University of Oklahoma, OK, USA.
[3]Amity Institute of Biotechnology, Amity University, Gurgaon, India.
[4]School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi-110067, India.
[5]Amity Institute of Integrative Sciences and Health, Amity University, Haryana, Gurgaon- 122413, India

[†]*authors contributed equally*

 Address for Correspondence

Amresh Prakash, PhD
Assistant Professor
Amity Institute of Integrative Sciences and Health (AIISH),
Amity University Haryana, Gurgaon-122413, India
Email: amreshprakash@jnu.ac.in; aprakash@ggn.amity.edu

Prof. Raok Jeon, Ph.D.
College of Pharmacy, Sookmyung Women's University,
Hyochangwon gil-52, Yongsan-13 Gu, Seoul 140-742,
Republic of Korea.
Email: rjeon@sookmyung.ac.kr

Abstract

The global pandemic crisis, COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has claimed the lives of millions of people across the world. Development and testing of anti-SARS-CoV-2 drugs or vaccines, are not turned to be realistic in the timeframe needed to combat this pandemic. Thus, rigorous efforts are still ongoing for the drug repurposing as a clinical treatment strategy to control COVID-19. Here we report a comprehensive computational approach to identify the multi-targeted drug molecules against the SARS-CoV-2 proteins, which are crucially involved in the viral-host interaction, replication of the virus inside the host, disease progression and transmission of coronavirus infection. Virtual screening of 72 FDA approved potential antiviral drugs against the target proteins: Spike (S) glycoprotein, human angiotensin-converting enzyme 2 (hACE2), 3-chymotrypsin-like cysteine protease (3CL^{pro}), Cathepsin L, Nucleocapsid protein, RNA-dependent RNA polymerase (RdRp) and non-structural protein 6 (NSP6) resulted in the selection of seven drugs which preferentially binds to the target proteins. Further, the molecular interactions determined by MD simulation, free energy landscape and the binding free energy estimation, using MM-PBSA revealed that among 72 drug molecules, catechin (flavan-3-ol) can effectively bind to 3CL^{pro}, Cathepsin L, RBD of S protein, NSP-6, and Nucleocapsid protein. It is more conveniently involved in key molecular interactions, showing binding free energy (ΔG_{bind}) in the range of -5.09 kcal/mol (Cathepsin L) to -26.09 kcal/mol (NSP6). At the binding pocket, catechin is majorly stabilized by the hydrophobic interactions, displays ΔE_{vdW} values -7.59 to -37.39 kcal/mol. Thus, the structural insights of better binding affinity and favourable molecular interaction of catechin towards multiple target proteins, signifies that catechin can be potentially explored as a multi-targeted agent in the rational design of effective therapies against COVID-19.

Keywords: SARS-CoV-2; COVID-19; multi-targeted drug; Catechin; free-energy landscape.

Introduction

There are different members of Coronaviridae family of virus which often cause mild, moderate to severe respiratory symptoms in humans. Recently, the novel coronavirus (2019-nCoV or SARS-CoV-2) appears to be the deadliest and highly contagious virus of this family. This mysterious virus emerged in Wuhan, China in December 2019 and, within a very short time, rattled the world's safety and ignited the sense of global panic. Although SARS-CoV-2 shares a high level of genetic similarity with SARS-CoV, the infection rate of SARS-CoV-2 is much higher than suspected. Declared as a pandemic in March 2020, the outbreak has already led to 0.5M deaths worldwide (WHO, 2020). The molecular and structural organization of the virus includes an envelope, non-segmented, positive-sense RNA which codes for several structural proteins such as spike (S) protein, envelope (E) protein, membrane (M) protein and the nucleocapsid (N) proteins and also 16 putative non-structural proteins (nsps, encoded by replicase complex (orf1ab)) (Chan et al., 2020; ul Qamar et al., 2020; Zhu et al., 2020). The genome consists of a 5′-untranslated region (UTR), 3′-UTR, and several unidentified non-structural open reading frames (Chan et al., 2020; Zhu et al., 2020). During the last couple of months, dozens of coronavirus vaccines have already been tested, approximately more than hundreds are either under development or in the preclinical investigation. However, the success of these efforts remains elusive. Therefore, the need of the hour is to identify novel and effective measures to control the pandemic (Cascella et al., 2020; WHO, 2020).

The major setback in identifying drugs/vaccines against COVID-19 remains the lack of knowledge about the mechanism of action of the virus, molecular targets and network of associated molecular interactions. Recently, several targets have already been discovered that interact with SARS-CoV-2 such as human angiotensin-converting enzyme 2 (hACE2), transmembrane protease serine 2 (TMPRSS2), phosphatidylinositol 3-phosphate 5-kinase (PIKfyve), two pore channel subtype 2 (TPC2), and cathepsin L (CTSL) (Bhowmik et al., 2020;

Hasan et al., 2020; Hoffmann et al., 2020; Ou et al., 2020a). At the same time, various research groups have identified the effective inhibitors against some of these targets such as main protease M^{pro} (3CL^{Pro}) (Alamri et al., 2020; Dai et al., 2020), envelope (E) protein (Gupta et al., 2020b), RNA dependent RNA polymerase (RdRp) (Yin et al., 2020) and Spike (S) protein (Barh et al., 2020; Tai et al., 2020).

Since the genes of RNA viruses (including SARS-CoV-2) are genetically variable (Cao et al., 2020; Lu et al., 2020; Phan, 2020), they can quickly accumulate genomic mutations through an error-prone viral reverse transcriptase, which advances their adaptation inside the human host. This further adds to the difficulty in designing active antiviral therapeutics against RNA viruses (Naik et al., 2020). Moreover, most of the antiviral drugs today are single target drugs designed against specific viral enzymes, which are essential for viral interaction, replication or invasion. Therefore, the high rate of mutations in these single viral drug targets has been main cause for reduced susceptibility of currently available antiviral drugs (Joshi et al., 2020).

Nevertheless, finding the compounds having efficacy for multiple molecular targets, remained a preferable approach in disorders caused by highly mutable pathogens (Ismail et al., 2016; Levitzki and Klein, 2019). Targeting different molecular targets with a single drug is always a preferable approach over combination therapy to avoid unwanted drug interactions (Kumari, Bhushan Mishra, and Tiwari, 2016; Naik et al., 2020). Additionally, the drugs designed for multiple protein targets are extensively preferable for the treatment of both infectious, inherited and complex diseases due to low treatment cost, less drug dosage and minimal side effect and drug-drug interactions (Battah et al., 2019; Gupta et al., 2020a; Kumari, Bhushan Mishra, and Tiwari, 2016; Mugumbate et al., 2017; Youdim, 2013). Therefore, with the immediate requirement of multi-targeted strategies (Naik et al., 2020; Shi et al., 2020) against the novel coronavirus SARS-CoV-2, with either biologically active drug-like molecules or approved drugs are in pressing priority. Recent advancement in the computational techniques have

proven their efficiency for identifying the potential drug candidates (Chodera et al., 2011; Ou et al., 2020b; Prakash and Luthra, 2012; Wang et al., 2019). Considering the improvement, reliability, and accuracy of computational methods, it has become a suitable choice to design structure-based drugs (Levitzki and Klein, 2019; Luthra, Kumar, and Prakash, 2009; Luthra et al., 2009; Williams-Noonan, Yuriev, and Chalmers, 2018). Keeping these facts in mind, we adopted a multi-target drug discovery approach to hit various druggable targets of SARS-CoV-2, which may appear highly beneficial to strike this highly mutated virus.

Thereby, we screened 72 FDA approved antiviral drugs against known targets of SARS-CoV-2. We have taken both human proteins as well as viral targets as a strategy. The targets were chosen which already have been published single drug interactions and defined PDB structure. Following 7 targets were identified: (1) The human angiotensin converting enzyme 2 (hACE2) interacting with the (2) transmembrane viral spike (S) glycoprotein at receptor-binding domain (RBD) that forms homotrimers protruding from the viral surface (Barh et al., 2020; Hoffmann et al., 2020; Ou et al., 2020a; Tai et al., 2020), (3) the highly immunogenic, antigenic and abundantly expressed viral nucleocapsid (N) protein, which plays essential roles in viral genome packaging by formation of helical ribonucleo proteins (Kang et al., 2020b), (4) main protease M ^{pro} (3CL ^{Pro}), an essential viral enzyme for processing the polyprotein complexes that are translated from the viral RNA (Zhang et al., 2020), (5) the human endosomal cysteine protease Cathepsin L required for viral entry (Liu et al., 2020; Ou et al., 2020a), (6) non-structural viral protein nsp-6 which dwells in the endoplasmic reticulum (ER) and has role in generation of autophagosomes (Benvenuto et al., 2020), (7) the nsp12, along with the two other cofactors nsp7-nsp8 as a complex, aiding increased RNA-dependent RNA polymerase (RdRp) template binding and processivity. Firstly, the virtual screening of 72 FDA approved antiviral drugs was performed against these targets to select a highly potent multi-targeted agent. The best binding poses of antiviral drugs with target proteins display a diverse range of binding affinities (6.23−12.43

kcal/mol). Among them, catechin emerged as a multi-targeted agent, effectively binding (range of 5.79 − 8.37 kcal/mol) with five target proteins: RBD, Cathepsin L, nucleocapsid protein, 3CLpro and NSP-6.

Catechin (flavan-3-ol) is a natural phenol and a major chemical component of sinecatechin, a first FDA approved herbal drug for the treatment of external genital warts caused by HPV infections. Topical ointment Veregen is a marketed medicine and it is a purified form of catechins, extracted from leaves of Chinese green tea which is comprising 80% catechins. It is well recognized for the anti-viral activity, anti-bacterial activity, anti-inflammatory and for the immunostimulatory actions (De Clercq and Li, 2016). Further, we have acquired several data contributing to the stable structural dynamics of the protein-ligand complex, including free-energy landscape, which validates the efficacy of these molecules as multi-targeted agents by using molecular dynamics (MD) simulation. The spatial stability of ligand molecules at the active site of protein was estimated in terms of binding free energy using MM/PBSA. These studies reveal that catechin occupies quite well the binding pocket of proteins, stabilized by energetically favourable molecular interactions during the simulation, which offers the prerequisite understanding for developing the effective drug molecules. Thus, identifying the catechin as a novel multi-targeted agent may provide the structural basis for the designing strategy of potential drug molecules targeting SARS-CoV-2 in the therapy of COVID-19.

Materials and Methods

Protein structure retrieval

The three-dimensional coordinates of protein structures were taken from the Protein Data Bank. The co-crystalized X-ray structure of SARS-CoV-2 spike receptor with ACE2 (PDB ID: 6M0J), 3-chymotrypsin-like (3CL) cysteine protease (PDB ID: 6M2N), Cathepsin L (PDB ID: 6F06), crystal structure of nucleocapsid protein (PDB ID: 6M3M), RNA-dependent RNA polymerase (PDB ID: 6M71), non-structural Protein 6 (NSP6) and Cryo–electron microscopy

structure of RNA-dependent RNA polymerase (RdRp) enzyme with remdesivir and nsp12-nsp7-nsp8 complex (PDB ID: 7BV2) were taken as the targets for the virtual screening of selected FDA approved antiviral agents as shown in supporting information **Figure S1**.

Virtual Screening and Molecular Docking

The virtual screening of FDA approved antiviral compounds against the selected SARS-CoV-2 proteins were performed using Glide, Schrodinger, LLC (Friesner et al., 2004; Halgren et al., 2004; Kumar et al., 2019). Glide involves three-step filtering methods, standard precision, extra precision and the selection of best docked compounds by integrating coulombic and van der Waals interaction energies and Glide scoring function. CORINA v2.64 software package (Sadowski, Gasteiger, and Klebe, 1994) was utilized to add the missing hydrogen atoms and optimize the sdf format structures of the ligands. The lowest energy three-dimensional structures of ligands were generated using Ligprep (Guillemette, Poitras, and Boulay, 1991). The ionization/tautomeric states of the selected compounds were taken care of by Epik and a maximum of up to 32 conformations was generated per ligand, using the Schrodinger protocol (Jorgensen, Maxwell, and Tirado-Rives, 1996; Jorgensen and Tirado-Rives, 2005; Shelley et al., 2007). The molecular interactions of docked complexes were analysed using PyMol and LigPlot.

Molecular dynamics (MD) simulation

MD simulations were carried out on the coordinates of protein-ligand complexes using GROMACS-2018.1 (Kutzner et al., 2019) with protein interactions approximated using CHARMM36 force field (Huang et al., 2017). The ligand parameters were generated utilizing CGenFF server (Vanommeslaeghe, Raman, and MacKerell, 2012). Each protein-ligand complex was placed in the center of a cubic simulation box with 10 Å distance to the edges

and solvated with TIP3P water molecules. The counterions (Na^+ Cl^-) 0.15 M added to neutralize the system. The periodic boundary condition was defined in x, y, and z directions (Darden, York, and Pedersen, 1993), and electrostatic interactions were evaluated using particle-Ewald summation, and a cut-off of 10 Å was used for calculation of vdW-interactions. The resulting systems were energy-minimized by steepest descent and conjugate gradient algorithms. Energy minimization was performed for 50000 steps. Equilibration was first performed for 500 ps in an NVT ensemble and subsequent 500 ps in an NPT ensemble. Temperature and pressure were set T=300 K and 1 bar, which was controlled by a Parrinello-Danadio-Bussi thermostat (Bussi, Donadio, and Parrinello, 2007) and Parrinello-Rahman pressure (Parrinello and Rahman, 1980), respectively. The integration step of 2 fs was used. Each system was simulated for 200 ns and the snapshots were saved every 10 ps for further analysis. All production runs were performed on CUDA enabled Tesla GPU machine (DELL T640 with V100 GPU), and OS Centos 7 (Prakash et al., 2018a; Singh et al., 2019).

MD analysis

The obtained MD trajectories were analyzed using GROMACS utilities. The structural order parameters that we measured are root-mean-square deviation (RMSD), the radius of gyration (Rg), solvent-accessible-surface-area (SASA), root-mean-square-fluctuation (RMSF), and hydrogen bond interactios. Hydrogen bonds were defined by a distance cut-off of 3.5 Å between the donor & acceptor atom and an angle cut-off of 30°. Similarly, a hydrophobic interaction was defined by the condition that the distance between two residues (i and j, with |i − j| > 3) is less than 4.5 Å. Principal component analysis performed using the projection of principal component (PCs), PC1, and PC2 along the native structure (Laberge and Yonetani, 2008; Prakash et al., 2018a) and gmx-sham utilized for the free energy landscape (Prakash et al., 2020; Prakash et al., 2018b).

Binding free energy estimation

The binding free energy of the protein-ligand complexes was evaluated using MM-PBSA (Molecular Mechanics-Poisson-Boltzmann Surface Area), which describes the structural, and molecular stability of the ligands in the active site of the protein (Batt et al., 2012; Prakash and Luthra, 2012; Sastry et al., 2013; Wang et al., 2019). The binding free energy of a protein-ligand complex ($\Delta G_{binding}$) can be written as,

$$\Delta G_{binding} = <G_{complex}> - <G_{receptor}> - <G_{ligand}>$$

where, $G_{complex}$ represents the free energy of the protein-ligand complex, $G_{receptor}$, the free energy of protein, G_{ligand} as the free energy of ligand and $<>$ represents the ensemble average.

Excluding the entropy term ($T\Delta S$), the above equation for the binding free energy can be approximately written as,

$$\Delta G_{binding} = \Delta E_{MM} + \Delta G_{solv}$$

where, ΔE_{MM} is the change in the average molecular mechanics interaction energy (gas-phase) upon ligand binding computed as the sum of the changes in the bonded and non-bonded (electrostatics and Van der Waals) interactions upon ligand binding ($\Delta E_{MM} = \Delta E_{bonded} + \Delta E_{electrostatics} + \Delta E_{vdw}$). ΔG_{solv} is the change in solvation free energy upon ligand binding. Further, ΔG_{solv} can be written as,

$$\Delta G_{solv} = \Delta G_{POL} + \Delta G_{NP}$$

where, ΔG_{POL} is the change in the polar part of the solvation free energy and ΔG_{NP} is the change in the non-polar part of the solvation free energy as a result of ligand binding to the proteins. Poisson-Boltzmann (PB) equation was used for the estimation of the polar part of the solvation free energy and the non-polar part was estimated with a surface area-based approach. Binding free energy ($\Delta G_{binding}$) for the protein-ligand complex was estimated using the MMPBSA.py script of the AMBER Tools (Wang et al., 2016). An ionic strength of 0.15 M and a solute

dielectric constant value of 2 was used for the PBSA calculations. Considering the convergence issues associated with the MM-PBSA calculation, only last 50 ns data was used.

Results and discussion

Targeting SARS-CoV-2 proteins for identifying multi-target agents

The ability of a chemical compound to work as a drug lies in its competency to bind efficiently to a druggable target. While there exist plenty of methods to evaluate the binding affinity of a ligand towards a target starting from very reliable and accurate alchemical free energy methods (computationally very costly) to less accurate docking methods, molecular docking and virtual docking remains the first choice to screen chemicals as they offer reasonable accuracy with modest computational efforts (Chodera et al., 2011; Williams-Noonan, Yuriev, and Chalmers, 2018). Considering this, we have first performed molecular docking study to screen the efficacy of 72 FDA approved drugs against various important druggable targets of SARS-CoV-2. Our docking study revealed that some of the FDA approved drugs have excellent interactions with particular target displaying satisfactory docking scores. Docking scores of the top 10 anti-viral drug molecules corresponding to each protein are enumerated in supporting information **Table S1**. Results indicate that among all, ritonavir, dolutegravir, tenofovir, tinofoviralafenamide, boceprevir, catechin, and zanamivir were found to be the most promising inhibitor against SARS-CoV-2 proteins: 3-chymotrypsin-like cysteine protease (3CL^pro), RNA-dependent RNA polymerase (RdRp), Angiotensin-converting enzyme 2 (ACE2), Cathepsin L, Non-Structural Protein 6 (NSP6), Nucleocapsid protein and receptor-binding domain (RBD) of Spike (S) protein, respectively. The molecular interactions of antiviral agents having the best dock score shown in supporting information **Figures S2 and S3.** Interestingly, the naturally derived polyhydroxy molecule catechin showed multi-targeted action against all seven targets; however, its promising binding capability (with the cut off range > 5.0 kcal/mol) was noticed against the five important targets engaged in invasion and survival of SARS-CoV-2 in to the human cells.

We found that catechin effectively binds to 3CL[pro], Cathepsin L, NSP-6, nucleocapsid protein, and RBD of S protein, showing the docking scores in the range of -5.79 to -8.34 kcal/mol.

Molecular interaction of catechin as multi-targeted agent

The structure-based virtual screening protocol bestowed catechin as the effective multi-targeted agent which effectively hits 3CL[pro], Cathepsin L, NSP-6, nucleocapsid protein, and RBD (**Figure 1 and** supporting information **Figure S3**). The SARS-CoV-2 protein, 3CL[pro] plays a critical role in the replication of the virus particles and is a potential target for anti-coronaviruses inhibitors screening. The active site of 3CL[pro] consists of Cys-His catalytic dyad (Cys145 and His41), which is highly conserved in the CoVs family, also referred to as the main protease, M[pro] (Muramatsu et al., 2016). The molecular docking result shows that catechin nicely fits in the active site of 3CL[pro] with the highest docking score -8.34 kcal/mol among all selected five target proteins. It displays H-bonds with Thr26, Met49, Arg188, and Gln189, whereas, Leu27, His41, and Leu58 are involved in hydrophobic interactions. Benzopyran moiety of catechin is oriented towards Met49, Arg188, and Gln189, noticeably, di-hydroxy phenyl occupied at the Cys145 and His41 catalytic dyad, crucial for ligand binding.

An excellent interaction of this compound was also seen with Cathepsin L, a crucial human protease that promotes SARS-CoV-2 entry by S protein activation (Jaimes, Millet, and Whittaker, 2020; Ou et al., 2020b). Indeed, catechin occupied the active site of this protease, stabilizing through hydrogen bond interactions as well as hydrophobic interaction. Phenolic hydroxy participate in hydrogen bonding with Ser216, and hydroxy at benzopyran moiety showed hydrogen bonding with Met161. Benzopyran ring was stabilized by hydrophobic interaction with Trp26, His163 and Ala214.

In the CoVs family, nucleocapsid acts as a multifunctional RNA-binding protein and plays an indispensable role in regulating viral RNA transcription/replication and the modulation of host cell metabolism (Tan et al., 2006). The recently solved N-terminal domain of SARS-CoV-2

nucleocapsid consists of antiparallel β-sheets at the core, protruding β-hairpin and short 3_{10}-helix (Kang et al., 2020a). The molecular interaction of catechin with nucleocapsid shows that phenolic hydroxy formed hydrogen bond interaction with Ala108, and phenyl moiety imparted hydrophobic interaction with Trp5, Ala109, and Ile110. Hydroxy group at benzopyran moiety was stabilized by hydrogen bond interaction with Asn28 and Ser31. Additionally, this benzopyran moiety was stabilized through hydrophobic interactions with His98 and Ile99. Another important CoVs protein, non-structural viral protein 6 (NSP6), plays an essential role in viral RNA synthesis by sequestering the membrane of endoplasmic reticulum of the host cell (Baliji et al., 2009). A strong affinity of catechin with NSP6 was also discerned. Indeed, catechin contoured well in the active site of NSP6, displaying satisfactory docking score (-6.68 kcal/mol). It was noticed that catechin forms tight interactions in the active site through three hydrogen bond interactions and several hydrophobic interactions. Hydroxy groups of catechin are found to be involved in hydrogen bonding with Leu239, Tyr242, and Asn232. Phenyl group, as well as benzopyran moiety, were well steadied by hydrophobic interactions with His62, Ala65, Tyr175 and Phe228. The configuration and orientation of catechin was also found to be favourable, showing convincing interactions with virus surface S glycoprotein which mediates entry into host cells, adhering at the host receptor (hACE2). Evidently, catechin also effectively inhibits the interaction of receptor-binding domain (RBD) of S protein with ACE2. Docking result indicates that the scaffold of catechin is well stabilized in the active site of RBD through four hydrogen bonding interactions. These hydrogen bonds are seen between hydroxy groups and active site amino acids such as Tyr453, Tyr495, Gly496 and Asn501. Additionally, hydrophobic interactions between catechin scaffold and amino acids: Arg403, Tyr453, Tyr495, Gly496 and Phe497 are also observed. Interestingly, catechin is also supported by π-π interactions, which can be noticed with Tyr453, Phe497 and Tyr495.

Thus, our molecular docking study evidently indicates that catechin strongly interacts with these five crucial targets associated with SARS-CoV-2, which clearly designates the multi-targeted action against SARS-CoV-2. However, considering the approximations made in molecular docking (lack of receptor flexibility and conformational entropy, lack of information about the number and free energy of water molecules in the binding site of the protein, etc.) to allow fast screening of chemicals, the dynamics of the protein-ligand interactions are overlooked and therefore, might not explain the stability of the ligands in the active site of the protein. In the cellular system, biomolecular interactions are dynamic in nature and the conformational flexibility is an intriguing property of proteins which triggers the biological functions and molecular recognitions (Amaral et al., 2017; Luthra, Kumar, and Prakash, 2009; Prakash and Luthra, 2012). A better understanding of protein-ligand interactions requires an accurate description of the spatial orientation of ligands at the active site of the protein, conformational dynamics which modulates the drug binding, interaction energy and molecular stability (Horoiwa et al., 2019; Lionta et al., 2014; Prakash et al., 2013). To understand the biomolecular interactions at atomic resolution, MD simulation is an efficient and well-established method which mimics the flexible nature of bio-molecules, protein conformational changes, protein-ligand interactions, structural perturbation and provide a more realistic picture with atomic details in reference to time (Koul et al., 2011; Prakash et al., 2018a; Wang et al., 2017). Thereby, to gain a deeper insight into the structural dynamics and stability of catechin binding with SARS-CoV-2 proteins, multiple MD simulations were performed for the period of 200 ns (Panda et al., 2020; Xu et al.). Additionally, the energetic contribution of binding pocket residues to accommodate the drug molecule, catechin is estimated using MM-PBSA (Pandey, Lynn, and Bandyopadhyay, 2017; Wang et al., 2019).

Conformational stability of protein-ligand complexes

We accessed the conformational stability of the protein-ligand complexes by measuring various structural order parameters like RMSD, R_g, SASA, RMSF, as shown in **Figure 2**. On comparing the Cα-RMSD of RBD, Cathepsin L and nucleocapsid protein complexed, we observed that catechin achieved stability in the active site of the protein very quickly (**Figure 2A**). These systems attained equilibrium in 0-5 ns and remained stable throughout the simulation time. The RMSD plot of 3CL^{pro} shows an initial rise in RMSD ~0.2 nm, which settles gradually, and a stable equilibrium can be seen up to 90 ns. We find a slight drop in the RMSD of ~0.1 nm around ~100 ns and the undisrupted trajectory is seen up to 200 ns, suggesting stable interaction of catechin in the binding pocket of Cathepsin L. The trajectory of NSP6 with catechin shows slightly large deviations in RMSD during 0-80 ns; thereafter, a gradual drop in RMSD can be seen which attains an equilibrium around ~120 ns. Notably, the stable conformational dynamics of NSP6-catechin is observed up to 200 ns. Thus, the shorter equilibration time taken by RBD, Cathepsin L, nucleocapsid and 3CL^{pro} to achieve a steady equilibrium suggests a better equilibrated and stabilized protein-ligand complex structure compared to NSP6. However, the stable trajectory of the NSP6-catechin complex during 120-200 ns signifies that the ligand is spatially well occupied and stabilized with the molecular interactions at the binding pocket of NSP6.

To further understand the structural stability of the protein-ligand complexes, we determined the compactness of protein structure by computing the radius of gyration (R_g). The R_g plots represented in **Figure 2B** show that the structural dynamics of RBD, Cathepsin L and nucleocapsid protein and 3CL^{pro} remain quite stable throughout the simulation time. The structural integrity of these four proteins was observed to be intact with the average R_g values, 1.83 ± 0.01 nm, 1.64 ± 0.01 nm, 1.50 ± 0.01 nm and 2.19 ± 0.01 nm, respectively. The slight deviations in the R_g plot of NSP6 can be seen during 0-80 ns; after that, the steady equilibrium

is noted till the end of simulation at 200 ns, which signifies the stable structural dynamics of NSP6-catechin complex with an average R_g value 2.23 ± 0.03 nm. The initial perturbation in the R_g trajectory may indicate the spatial adjustment of the ligand in the binding site of NSP6. Another important quantity that we measure and analyse to probe the conformational stability of the protein-ligand complex is the solvent-accessible surface area (SASA). The solvent environment around the protein plays a key role in maintaining the protein fold and govern the protein-ligand interaction processes, orientation and stability. Interestingly, we find that the SASA plots of all five protein-ligand complexes (RBD, Cathepsin L, nucleocapsid protein, 3CL^{pro} and NSP6) remain fairly equilibrated during the entire simulation period (0-200 ns) which provides clear evidence of the stable conformational dynamics of protein-ligand interactions (**Figure 2C**). The average values of structural order parameters, RMSD, R_g, and SASA are shown in supporting information **Table S2**.

Next, we investigated the binding stability of the catechin at the active site of the respected proteins by monitoring the time evolution plots of the average distance from the centre of binding pocket to the ligand, as shown in supporting information **Figure S4**. During the period of 200 ns simulation, the average distance of catechin from the binding site of all five proteins ranges between 0.33 and 0.40 nm. Although, the peaks of sharp drifts appeared transiently at ~80 ns and 140 ns for RBD and Nucleocapsid protein, respectively, the overall distance of catechin to the active site remains favourable for the stable molecular interaction. It is worth noting that the average distance plots of Cathepsin L, 3CL^{pro} and NSP-6 remain stable throughout simulation time. Thus, this analysis provides elegant evidence of the spatially well-fitted catechin orientation in binding sites of proteins.

We further performed RMSF analysis to evaluate the positional fluctuation of each amino acid around its average mean position (**Figure 2D**). This analysis provides a clue about the mobility of atomic fluctuations related to the structural stability of molecular interaction during the

simulation. Usually, the higher values of RMSF are often associated with loops or may be the terminal residues, whereas, the lower RMSF values indicate the rigid conformation of stable secondary structures of α-helices and β-sheets. The result indicates that all the complexes show equilibrium fluctuations, except NSP-6. The plot shows that the values of RMSF significantly vary for all residues of NSP-6 in comparison to the other proteins, RBD, Cathepsin L, Nucleocapsid protein and 3CL[pro], respectively. The RMSF plots of these proteins show an average atomic fluctuation <0.15 Å for amino acid residues, which belongs to the stable secondary structure and the regions which displayed high fluctuations represented the atomic flexibility of loops. The structure of NSP6 consists of a helical structure at the core, capped by antiparallel β-sheets and two small helices. The binding pocket is characterized by α-helix-2-4 and α-helix-6-7 and antiparallel β-sheets (β1 and β2) which is enclosed by hydrophobic loop (Phe235-Tyr242) connecting helices and β-sheets. The RMSF plot shows on average, high fluctuations for the residues belonging to longest loop (Val84-Leu110) connecting α-helix-3 and α-helix-4, loops (Gln257-Ser262 and Leu275-Pro282) connecting to small two helices (α8 and α9) with β-sheets, respectively. The average residual fluctuations observed reasonably lower for terminal residues of α-helix-2 (Phe42-Phe59), α-helix-3 (Lys63-Met86), α-helix-4 (Lys109-Arg129), α-helix-6 (Ala157-Thr196), α-helix-7(Tyr175-Tyr196) and N-terminal of connecting loop (Phe235-Tyr242) and β1 (Asp243-Val246) which are actively involved in molecular interactions with drug molecules. Further, we also established the conformational stability by analysing the secondary structural contents, which were observed intact during the simulation (supporting information **Figure S5**). The structural snapshots of protein-ligand interactions captured at time interval of 20 ns described in supporting information **Figure S6.** Thus, the overall results clearly indicate the stable conformational dynamics of the target proteins complexed with antiviral drug, catechin.

Another parameter, hydrogen bonds (H-bonds) interaction is one of the major players in governing the ligand stability at the active site of the protein. Thus, we further investigated the time evolution plots of H-bonds involved in the molecular interaction of catechin with proteins (**Figure 2E**). The average occupancy of H-bonds (donor and acceptor), excluding ionic interactions were examined during the last 50 ns of simulation time and are summarized in supporting information **Table S3**. We find the maximum occupancy of five H-bonds between the catechin and RBD; however, four H-bonds remain consistent up to ~50 ns. Out of these, two H-bonds vanished at ~60 ns, which are regained at ~100 ns. Nevertheless, three H-bonds observed stable during the last 50 ns of simulation. The result shows that catechin formed H-bond interaction with residues, Tyr451 (OH) with ligand (H7 and H8) and Asp442 (O) with ligand (H7). 3CL^{pro} shows six H-bond interactions with catechin which can be seen up to ~150 ns, but only three remains stable during the last 50 ns which are formed between the residues, Asp187 (O) – ligand (H7), His164 (HE2) – Ligand (O2) and Thr26 (O) – Ligand (H14). Cathepsin L form three H-bonds, which are observed consistent up to 170 ns. It shows maximum occupancy with Asp162 (OD2) – Ligand (H7), Asp160 (OD1) – Ligand (H8) and Asp160 (OD2) – Ligand (H8). Three H-bonds are formed between Nucleocapsid–catechin, which are observed with Arg46 (HN) – Ligand (O1), Asp56 (OD1) – Ligand (H7) and Thr44 (HG1) – Ligand (O2). NSP-6 shows the possibility of three-four H-bonds, however, two H-bonds between Thr238 (O) – Ligand (H14) and Thr130 (HG1) – Ligand (O1), having maximum occupancy during the last 50 ns of simulation. These results indicate that catechin is stabilized by an average of two-three H-bonds at the binding pocket of proteins.

Essential dynamics

Protein function is regulated by switching between various conformations. The modular nature of proteins to switch between various states is governed by the collective motion of protein,

which is intrinsic to many biological processes and plays a crucial role in the transmission of biological signals. For a protein to be functional, a reasonable amount of flexibility, as well as rigidity is required, specifically for the residues in the binding site. Essentially, a tighter interaction would restrict the motion of the protein, thereby not allowing it to sample some conformations required for its activity. Therefore, in order to understand the collective motion of protein occupied in the conformational space during the simulation, we applied the dimension reduction method, essential dynamics (ED) analysis by the projection of the first two principal components (PCs), PC1 and PC2. The PC1 and PC2 were calculated by diagonalizing the covariance matrix of eigenvectors to define the essential subspace in which most of the protein dynamics occur. The dynamic motion of proteins obtained through the projection of PC1 and PC2 are shown in **Figure 3**. It is apparent from these plots that the collective motion of proteins, RBD, Cathepsin L and Nucleocapsid protein is localized in a small conformational space in comparison to 3CL^{pro} and NSP-6, which revealed consistent results corresponding to the structural order analyses, RMSD, R_g and SASA as described in **Figure 2A-C**. The well-defined small clusters of RBD, Cathepsin L and Nucleocapsid protein, clearly indicate the reliability and stability of complex structure with catechin. The ED plot of 3CL^{pro} displays a slight increase in the conformational phase space, which can be seen along the PC2, which suggests that protein navigated the broad conformational space before achieving the ensemble of dynamically equilibrated state. Contrary to this, NSP-6 experiences a wide region of phase space. In fact, it explored a large conformational space in comparison to the other four proteins, which represent the overall higher flexibility of the protein. Thus, we observed a significantly compact structure of RBD, Cathepsin L and Nucleocapsid protein and 3CL^{pro} as compared to NSP-6, which may facilitate the vital interactions with catechin.

Free energy landscape

Free energy landscape (FEL) provides an accurate description of the minimum energy conformational ensembles of biomolecules, which is undoubtedly essential to understand the conformational transition underlying protein-ligand interactions (Boehr, Nussinov, and Wright, 2009). Thus, FEL plot is constructed using Boltzmann inversion ($F = -RT \ln P$), where P is the two-dimensional probability distribution of the first two PCs, PC1 and PC2, as reaction coordinates. **Figure 4** shows that the binding of catechin with proteins occurs through the minimum free energy pathway. The structural ensemble derived from FEL shows that the catechin bound complex with RBD navigated the broad conformational space, clustered in the different energy basins, distributed along the PC1. However, these energy minima separated through the low transition barrier < 2.0 kcal/mol indicates that with the small excursion, the ensemble states of RBD can easily move out from one energy basin to another. This may be the reason we observed a small and consolidated cluster of stable populations in ED analysis. The FEL plot of 3CL^{pro} shows the appearance of two distinct populations confined to two different energy basins, separated with high transition barrier >4.0 kcal/mol, which signifies the population of loosely and tightly ligand bound conformation of the protein. The conformational ensemble occupying the small energy basin represented the population of the equilibration phase, which readily achieved a stable equilibrium. These equilibrated ensembles of stable complex transverse to a broad and deep energy basin. The complex with Cathepsin L shows single but elongated energy minima, which depicts the heterogeneous population of different sub-states, but the very less transition barrier <1.0 kcal/mol between ensemble states suggested the stable conformation of protein-ligand complex confined to energy basin interplays between the subspace. Whereas, the rugged FEL with segmented small energy minima of Nucleocapsid protein suggested the population of loosely bound complex (Fu et al., 2019). The low transition barriers (~1.5 kcal/mol) between the small energy basins indicate a more prolonged equilibration phase of complex structure. During the progression of

simulation, the protein underwent structural modifications to accommodate ligand and adopted a stable conformation; thus, the equilibrated ensemble smoothly shifted to broad and deep energy minima. Remarkably, The FEL of NSP-6 shows that the stably bound conformation of the protein-ligand complex is widely populated to a single consolidated energy minimum, which provides the elegant evidence of interactions inducing the stable conformational transition of NSP-6–catechin complex. Thus, the comparison of FEL results indicates the different binding stability of catechin bound protein-ligand complexes.

Binding free energy and ligand-residue interaction decomposition

In order to understand the molecular interaction and stability associated with the binding of catechin to five different proteins of SARS-CoV-2, a detailed analysis of the binding free energy is executed through the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA). MM-PBSA provides the best prediction accuracy in terms of energy components of bonded, polar and non-polar solvation free energy, electrostatic and van der Waals interactions. In addition, it supplies the residue decomposition plot, which helps to probe the contribution of amino acid residues involved in the spatial interaction to stabilize ligands at the binding pocket of the protein. This analysis was performed on the fully converged trajectory of the last 50 ns with a solute dielectric constant value of 2 and an ionic strength of 0.15 M (supporting information **Figure S7**). Results show that catechin favourably binds to all five proteins, however, it shows a wide range of the total binding free energies (ΔG_{bind}), as enumerated in supporting information **Table S4.** As illustrated in **Figure 5**, catechin possesses highest binding affinity towards NSP-6 with a maximum value of ΔG_{bind} = -26.09 kcal/mol, whereas lowest towards Cathepsin L (ΔG_{bind} = -5.09 kcal/mol). Results show relatively more favourable contribution of van der Waals energies (ΔE_{vdW}) -37.39 kcal/mol and -24.76 kcal/mol for NSP-6 and 3CL^{pro}, as compared to Nucleocapsid protein (-16.34 kcal/mol) and RBD (-15.8

kcal/mol), respectively. Whereas, the less contribution of van der Waals interaction (-7.59 kcal/mol) is observed for Cathepsin L. Another binding energy component, electrostatic (ΔE_{eel}) energy, which describes ligand-protein interactions is a critical factor in determining the binding stability of ligand. The interaction of catechin with Cathepsin L shows the major contribution of electrostatic energy -29.61 kcal/mol, however, the lowest values of ΔE_{vdW} = -7.59 kcal/mol and ΔG_{bind} = -5.09 kcal/mol, signifies that the electrostatic energy contribute relatively less as compared to the other energies in the binding stability of catechin. Thus, the binding free energy analysis revealed that the binding of catechin at the active site of proteins are predominantly stabilized by hydrophobic interactions.

To further quantify the contribution of binding pocket residues to the molecular interaction of catechin with five different proteins, the free energy decomposition per residue was employed (**Figure 6**). The plot of free energy decomposition analysis shows that the active site residues, Agr346, Phe347, Leu441, Asp442, Lys444, Tyr449, Asn450 and Try451 energetically favour the binding stability of catechin to RBD. Remarkably, it is noted that Agr346 contributed the highest binding free energy, ΔE_{vdW} (-2.10 kcal/mol), ΔE_{eel} (-1.73 kcal/mol) and ΔG_{bind} (-2.10 kcal/mol) which indicated the favourable electrostatic and van der Waals interaction with catechin (**Figure 6A**). The protonated (-NH3$^+$) Agr346 shows the electrostatic interactions, whereas the side chain guanidinium (-C(NH$_2$)$_2$) facilitated the hydrophobic interaction with ligand. The binding interaction with 3CL^{pro} shows that the amino acid residues: Leu27, His47, Ser46, Met49, His164, Met165, Asp187 and Gln189, contributed the most to the total ΔG_{bind} (-16.98 kcal/mol). Although the van der Waals interaction primarily stabilizes the catechin at the binding pocket of 3CL^{pro}, the electrostatic interaction also contributes towards the observed stability by His164 (-3.80 kcal/mol) and Asp187 (-3.10 kcal/mol), respectively (**Figure 6B**). Indeed, the catechin is predominant stabilized in the binding pocket of Cathepsin L through the electrostatic interaction which is mostly contributed by residues: Asp71 (-11.31 kcal/mol),

Asp114 (-11.36 kcal/mol), Asp160 (-7.61 kcal/mol) and Asp162 (-7.89 kcal/mol), respectively (**Figure 6C**). **Figure 6D** showing the free energy decomposition plot of Nucleocapsid protein indicates the substantial contribution of amino acids, Thr44, Arg45, Arg46, Asp56, Arg60 and Tyr62 to energetically hold catechin at binding pocket. Surprisingly, it is noted that Arg45 contributed to both electrostatic (-3.07 kcal/mol) and van der Waals interaction (-1.88 kcal/mol), but Arg46 contributed only van der Waals interaction (-2.83 kcal/mol). This may be the reason we observed a moderate range of total binding energy (ΔG_{bind}) value -14.15 kcal/mol. The favourable binding of catechin with NSP-6 shows the significant contribution of residues, Lys61, His62, His64, Ala65, Asp133, Asp134, Phe228, Leu237, Thr238, Leu239 and Gln290 (**Figure 6E**). Interestingly, it is noted that Thr238 contributed the higher electrostatic energy (-5.0 kcal/mol), whereas the maximum van der Waals energy (-2.22 kcal/mol) contributed by Phe228. However, the binding pocket of NSP-6 mostly consist of hydrophobic residues; thus, we observed the major collective contribution of van der Waals energy for stabilizing the ligand interaction.

Conclusion

In summary, using molecular docking and classical molecular dynamics simulation, we have explored the possibility of 72 FDA approved antiviral drugs for their potential of being used as an effective therapeutic strategy to control SARS-CoV-2 infections. The virtual screening results showed that seven chemicals: ritonavir, dolutegravir, tenofovir, tinofoviralafenamide, boceprevir, catechin, and zanamivir could efficiently bind to SARS-CoV-2 proteins: 3CL^pro, RdRp, ACE2, Cathepsin L, NSP6, Nucleocapsid protein and RBD of Spike (S) protein, respectively. Of these, catechin has the potential to act as a multi-targeted agent, as it has the highest binding affinity towards the five crucial proteins of the virus: RBD, Cathepsin L, nucleocapsid protein, 3CLpro and NSP-6, which are essential for the invasion and infection of the host cell. Further, molecular dynamics simulation, free energy landscapes and binding free

estimation of catechin with the five target proteins explained the stable interactions of catechin with the critical residues in terms of occupancy of hydrogen bonds and residue contributions to the binding free energy. Thus, our investigation bestowed promising multitargeted agent catechin which can be explored as an effective therapeutic agent against SARS-CoV-2 virus to control the COVID-19 pandemic.

Conflict of Interest: Authors declare there is no conflict of interest.

References:

Alamri, M. A., Tahir ul Qamar, M., Mirza, M. U., Bhadane, R., Alqahtani, S. M., Muneer, I., Froeyen, M., and Salo-Ahen, O. M. (2020). Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro. *Journal of Biomolecular Structure and Dynamics*, 1-13.

Amaral, M., Kokh, D. B., Bomke, J., Wegener, A., Buchstaller, H. P., Eggenweiler, H. M., Matias, P., Sirrenberg, C., Wade, R. C., and Frech, M. (2017). Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nat Commun* **8**(1), 2276.

Baliji, S., Cammer, S. A., Sobral, B., and Baker, S. C. (2009). Detection of nonstructural protein 6 in murine coronavirus-infected cells and analysis of the transmembrane topology by using bioinformatics and molecular approaches. *J Virol* **83**(13), 6957-62.

Barh, D., Tiwari, S., Andrade, B. S., Giovanetti, M., Costa, E. A., Kumavath, R., Ghosh, P., Góes-Neto, A., Alcantara, L. C. J., and Azevedo, V. (2020). Potential chimeric peptides to block the SARS-CoV-2 spike receptor-binding domain. *F1000Research* **9**(576), 576.

Batt, S. M., Jabeen, T., Bhowruth, V., Quill, L., Lund, P. A., Eggeling, L., Alderwick, L. J., Futterer, K., and Besra, G. S. (2012). Structural basis of inhibition of Mycobacterium tuberculosis DprE1 by benzothiazinone inhibitors. *Proc Natl Acad Sci U S A* **109**(28), 11354-9.

Battah, B., Chemi, G., Butini, S., Campiani, G., Brogi, S., Delogu, G., and Gemma, S. (2019). A repurposing approach for uncovering the anti-tubercular activity of FDA-Approved drugs with potential multi-targeting profiles. *Molecules* **24**(23), 4373.

Benvenuto, D., Angeletti, S., Giovanetti, M., Bianchi, M., Pascarella, S., Cauda, R., Ciccozzi, M., and Cassone, A. (2020). Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *Journal of Infection*.

Bhowmik, D., Nandi, R., Jagadeesan, R., Kumar, N., Prakash, A., and Kumar, D. (2020). Identification of potential inhibitors against SARS-CoV-2 by targeting proteins responsible for envelope formation and virion assembly using docking based virtual screening, and pharmacokinetics approaches. *Infect Genet Evol* **84**, 104451.

Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**(11), 789-96.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**(1), 014101.

Cao, Y., Li, L., Feng, Z., Wan, S., Huang, P., Sun, X., Wen, F., Huang, X., Ning, G., and Wang, W. (2020). Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell discovery* **6**(1), 1-4.

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., and Di Napoli, R. (2020). Features, evaluation and treatment coronavirus (COVID-19). *In* "Statpearls [internet]". StatPearls Publishing.

Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., and Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging microbes & infections* **9**(1), 221-236.

Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., and Pande, V. S. (2011). Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol* **21**(2), 150-60.

Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., and Liu, F. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **368**(6497)**,** 1331-1335.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**(12)**,** 10089-10092.

De Clercq, E., and Li, G. (2016). Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* **29**(3)**,** 695-747.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**(7)**,** 1739-49.

Fu, H., Shao, X., Cai, W., and Chipot, C. (2019). Taming Rugged Free Energy Landscapes Using an Average Force. *Acc Chem Res* **52**(11)**,** 3254-3264.

Guillemette, G., Poitras, M., and Boulay, G. (1991). Two Ca2+ transport systems are distinguished on the basis of their Mg2+ dependency in a post-nuclear particulate fraction of bovine adrenal cortex. *Cell Calcium* **12**(1)**,** 51-60.

Gupta, A. K., Kumar, A., Rajput, A., Kaur, K., Dar, S. A., Thakur, A., Megha, K., and Kumar, M. (2020a). NipahVR: a resource of multi-targeted putative therapeutics and epitopes for the Nipah virus. *Database* **2020**.

Gupta, M. K., Vemula, S., Donde, R., Gouda, G., Behera, L., and Vadde, R. (2020b). In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *Journal of Biomolecular Structure and Dynamics***,** 1-11.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47**(7)**,** 1750-9.

Hasan, A., Paray, B. A., Hussain, A., Qadir, F. A., Attar, F., Aziz, F. M., Sharifi, M., Derakhshankhah, H., Rasti, B., and Mehrabi, M. (2020). A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *Journal of Biomolecular Structure and Dynamics***,** 1-9.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., and Nitsche, A. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.

Horoiwa, S., Yokoi, T., Masumoto, S., Minami, S., Ishizuka, C., Kishikawa, H., Ozaki, S., Kitsuda, S., Nakagawa, Y., and Miyagawa, H. (2019). Structure-based virtual screening for insect ecdysone receptor ligands using MM/PBSA. *Bioorg Med Chem* **27**(6)**,** 1065-1075.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmuller, H., and MacKerell, A. D., Jr. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**(1)**,** 71-73.

Ismail, H. M., Barton, V., Phanchana, M., Charoensutthivarakul, S., Wong, M. H., Hemingway, J., Biagini, G. A., O'Neill, P. M., and Ward, S. A. (2016). Artemisinin activity-based probes identify multiple molecular targets within the asexual stage of the malaria parasites Plasmodium falciparum 3D7. *Proc Natl Acad Sci U S A* **113**(8)**,** 2080-5.

Jaimes, J., Millet, J., and Whittaker, G. (2020). Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. *SSRN***,** 3581359.

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **118**(45)**,** 11225-11236.

Jorgensen, W. L., and Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A* **102**(19)**,** 6665-70.

Joshi, R. S., Jagdale, S. S., Bansode, S. B., Shankar, S. S., Tellis, M. B., Pandya, V. K., Chugh, A., Giri, A. P., and Kulkarni, M. J. (2020). Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. *Journal of Biomolecular Structure and Dynamics***,** 1-16.

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Chen, Q., Yan, Y., Zhang, C., Shan, H., and Chen, S. (2020a). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B*.

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., and Chen, Q. (2020b). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B*.

Koul, A., Arnoult, E., Lounis, N., Guillemont, J., and Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature* **469**(7331)**,** 483-90.

Kumar, N., Srivastava, R., Prakash, A., and Lynn, A. M. (2019). Structure-based virtual screening, molecular dynamics simulation and MM-PBSA toward identifying the inhibitors for two-component regulatory system protein NarL of Mycobacterium Tuberculosis. *J Biomol Struct Dyn***,** 1-15.

Kumari, S., Bhushan Mishra, C., and Tiwari, M. (2016). Polypharmacological drugs in the treatment of epilepsy: the comprehensive review of marketed and new emerging molecules. *Current pharmaceutical design* **22**(21)**,** 3212-3225.

Kutzner, C., Pall, S., Fechner, M., Esztermann, A., de Groot, B. L., and Grubmuller, H. (2019). More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J Comput Chem* **40**(27)**,** 2418-2431.

Laberge, M., and Yonetani, T. (2008). Molecular dynamics simulations of hemoglobin A in different states and bound to DPG: effector-linked perturbation of tertiary conformations and HbA concerted dynamics. *Biophys J* **94**(7)**,** 2737-51.

Levitzki, A., and Klein, S. (2019). My journey from tyrosine phosphorylation inhibitors to targeted immune therapy as strategies to combat cancer. *Proc Natl Acad Sci U S A* **116**(24)**,** 11579-11586.

Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem* **14**(16)**,** 1923-38.

Liu, T., Luo, S., Libby, P., and Shi, G.-P. (2020). Cathepsin L-selective inhibitors: A potentially promising treatment for COVID-19 patients. *Pharmacology & Therapeutics***,** 107587.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., and Zhu, N. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**(10224)**,** 565-574.

Luthra, P. M., Kumar, R., and Prakash, A. (2009). Demethoxycurcumin induces Bcl-2 mediated G2/M arrest and apoptosis in human glioma U87 cells. *Biochem Biophys Res Commun* **384**(4)**,** 420-5.

Luthra, P. M., Prakash, A., Barodia, S. K., Kumari, R., Mishra, C. B., and Kumar, J. B. (2009). In silico study of naphtha [1, 2-d] thiazol-2-amine with adenosine A 2A receptor and its role in antagonism of haloperidol-induced motor impairments in mice. *Neurosci Lett* **463**(3)**,** 215-8.

Mugumbate, G., Mendes, V., Blaszczyk, M., Sabbah, M., Papadatos, G., Lelievre, J., Ballell, L., Barros, D., Abell, C., and Blundell, T. L. (2017). Target identification of Mycobacterium

tuberculosis phenotypic hits using a concerted chemogenomic, biophysical, and structural approach. *Frontiers in pharmacology* **8,** 681.

Naik, B., Gupta, N., Ojha, R., Singh, S., Prajapati, V. K., and Prusty, D. (2020). High throughput virtual screening reveals SARS-CoV-2 multi-target binding natural compounds to lead instant therapy for COVID-19 treatment. *International Journal of Biological Macromolecules*.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., and Hu, J. (2020a). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature communications* **11**(1)**,** 1-12.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., and Qian, Z. (2020b). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* **11**(1)**,** 1620.

Panda, P. K., Arul, M. N., Patel, P., Verma, S. K., Luo, W., Rubahn, H. G., Mishra, Y. K., Suar, M., and Ahuja, R. (2020). Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci Adv* **6**(28)**,** eabb8097.

Pandey, P., Lynn, A. M., and Bandyopadhyay, P. (2017). Identification of inhibitors against alpha-Isopropylmalate Synthase of Mycobacterium tuberculosis using docking-MM/PBSA hybrid approach. *Bioinformation* **13**(5)**,** 144-148.

Parrinello, M., and Rahman, A. (1980). Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters* **45**(14)**,** 1196-1199.

Phan, T. (2020). Genetic diversity and evolution of SARS-CoV-2. *Infection, genetics and evolution* **81,** 104260.

Prakash, A., Dixit, G., Meena, N. K., Singh, R., Vishwakarma, P., Mishra, S., and Lynn, A. M. (2018a). Elucidation of stable intermediates in urea-induced unfolding pathway of human carbonic anhydrase IX. *J Biomol Struct Dyn* **36**(9)**,** 2391-2406.

Prakash, A., Kumar, K., Islam, A., Hassan, M. I., and Ahmad, F. (2013). Receptor Chemoprint Derived Pharmacophore Model for Development of CAIX Inhibitors. *Journal of Carcinogenesis & Mutagenesis* **5**(8)**,** 1-9.

Prakash, A., Kumar, V., Banerjee, A., Lynn, A. M., and Prasad, R. (2020). Structural heterogeneity in RNA recognition motif 2 (RRM2) of TAR DNA-binding protein 43 (TDP-43): clue to amyotrophic lateral sclerosis. *J Biomol Struct Dyn***,** 1-11.

Prakash, A., Kumar, V., Meena, N. K., and Lynn, Andrew M. (2018b). Elucidation of the structural stability and dynamics of heterogeneous intermediate ensembles in unfolding pathway of the N-terminal domain of TDP-43. *RSC Advances* **8**(35)**,** 19835-19845.

Prakash, A., and Luthra, P. M. (2012). Insilico study of the A(2A)R-D (2)R kinetics and interfacial contact surface for heteromerization. *Amino Acids* **43**(4)**,** 1451-64.

Sadowski, J., Gasteiger, J., and Klebe, G. (1994). Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *Journal of Chemical Information and Computer Sciences* **34**(4)**,** 1000-1008.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **27**(3)**,** 221-34.

Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007). Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* **21**(12)**,** 681-91.

Shi, Y., Zhang, X., Mu, K., Peng, C., Zhu, Z., Wang, X., Yang, Y., Xu, Z., and Zhu, W. (2020). D3Targets-2019-nCoV: a webserver for predicting drug targets and for multi-target and multi-site based virtual screening against COVID-19. *Acta Pharmaceutica Sinica B*.

Singh, R., Meena, N. K., Das, T., Sharma, R. D., Prakash, A., and Lynn, A. M. (2019). Delineating the conformational dynamics of intermediate structures on the unfolding pathway of beta-lactoglobulin in aqueous urea and dimethyl sulfoxide. *J Biomol Struct Dyn*, 1-10.

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., and Du, L. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & molecular immunology* **17**(6)**,** 613-620.

Tan, Y. W., Fang, S., Fan, H., Lescar, J., and Liu, D. X. (2006). Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells. *Nucleic Acids Res* **34**(17)**,** 4816-25.

ul Qamar, M. T., Alqahtani, S. M., Alamri, M. A., and Chen, L.-L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *Journal of pharmaceutical analysis*.

Vanommeslaeghe, K., Raman, E. P., and MacKerell, A. D., Jr. (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J Chem Inf Model* **52**(12)**,** 3155-68.

Wang, C., Greene, D., Xiao, L., Qi, R., and Luo, R. (2017). Recent Developments and Applications of the MMPBSA Method. *Front Mol Biosci* **4,** 87.

Wang, C., Nguyen, P. H., Pham, K., Huynh, D., Le, T. B., Wang, H., Ren, P., and Luo, R. (2016). Calculating protein-ligand binding affinities with MMPBSA: Method and error analysis. *J Comput Chem* **37**(27)**,** 2436-46.

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., and Hou, T. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev* **119**(16)**,** 9478-9508.

WHO, W. H. O. (2020). Coronavirus disease 2019 (COVID-19): situation report, 72.

Williams-Noonan, B. J., Yuriev, E., and Chalmers, D. K. (2018). Free Energy Methods in Drug Design: Prospects of "Alchemical Perturbation" in Medicinal Chemistry. *J Med Chem* **61**(3)**,** 638-649.

Xu, W. W., Huang, Z.-H., Liao, L., Zhang, Q.-H., Li, J.-Q., Zheng, C.-C., He, Y., Luo, T.-T., Wang, Y., Hu, H.-F., Zuo, Q., Chen, W.-Y., Yang, Q.-S., Zhao, J.-F., Qin, Y.-R., Xu, L.-Y., Li, E.-M., Liao, H.-X., Li, B., and He, Q.-Y. Direct Targeting of CREB1 with Imperatorin Inhibits TGFβ2-ERK Signaling to Suppress Esophageal Cancer Metastasis. *Advanced Science* **n/a**(n/a)**,** 2000925.

Yin, W., Mao, C., Luan, X., Shen, D.-D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., and Gao, M. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*.

Youdim, M. B. (2013). Multi target neuroprotective and neurorestorative anti-Parkinson and anti-Alzheimer drugs ladostigil and m30 derived from rasagiline. *Experimental neurobiology* **22**(1)**,** 1-10.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**(6489)**,** 409-412.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., and Lu, R. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*.
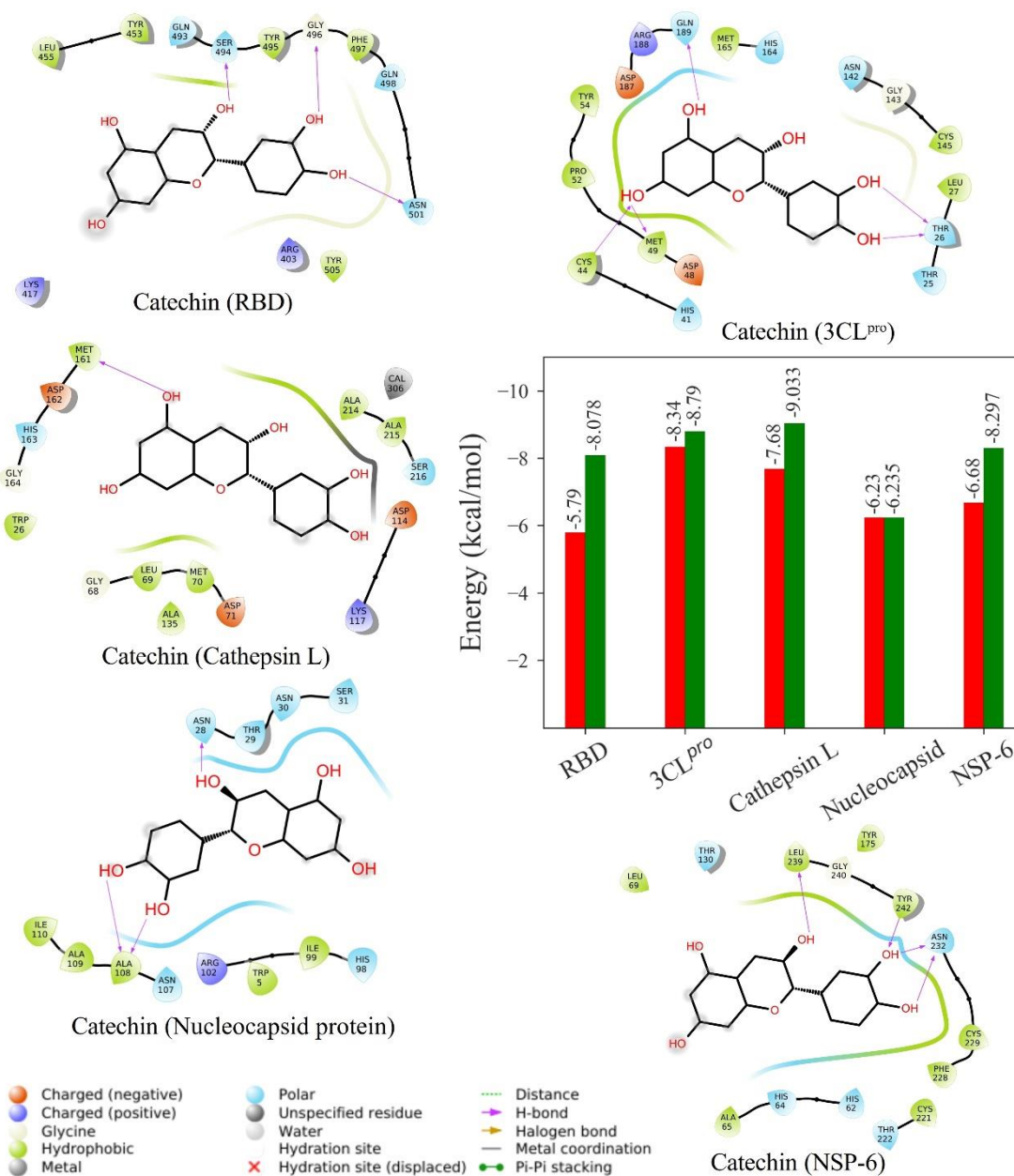
Figure 1. Molecular interactions of FDA approved antiviral drug, catechin at the binding pocket of potential target proteins: Receptor binding domain (RBD) of S protein, 3CL^pro, Cathepsin L, Nucleocapsid and non-structural protein 6 (NSP6), using LigPlot. The bar chart showing the target proteins specific antiviral drugs binding affinity (kcal/mol) obtained through virtual screening (represented by green bars) and the comparative binding affinity (kcal/mol) with catechin shown in red bars.
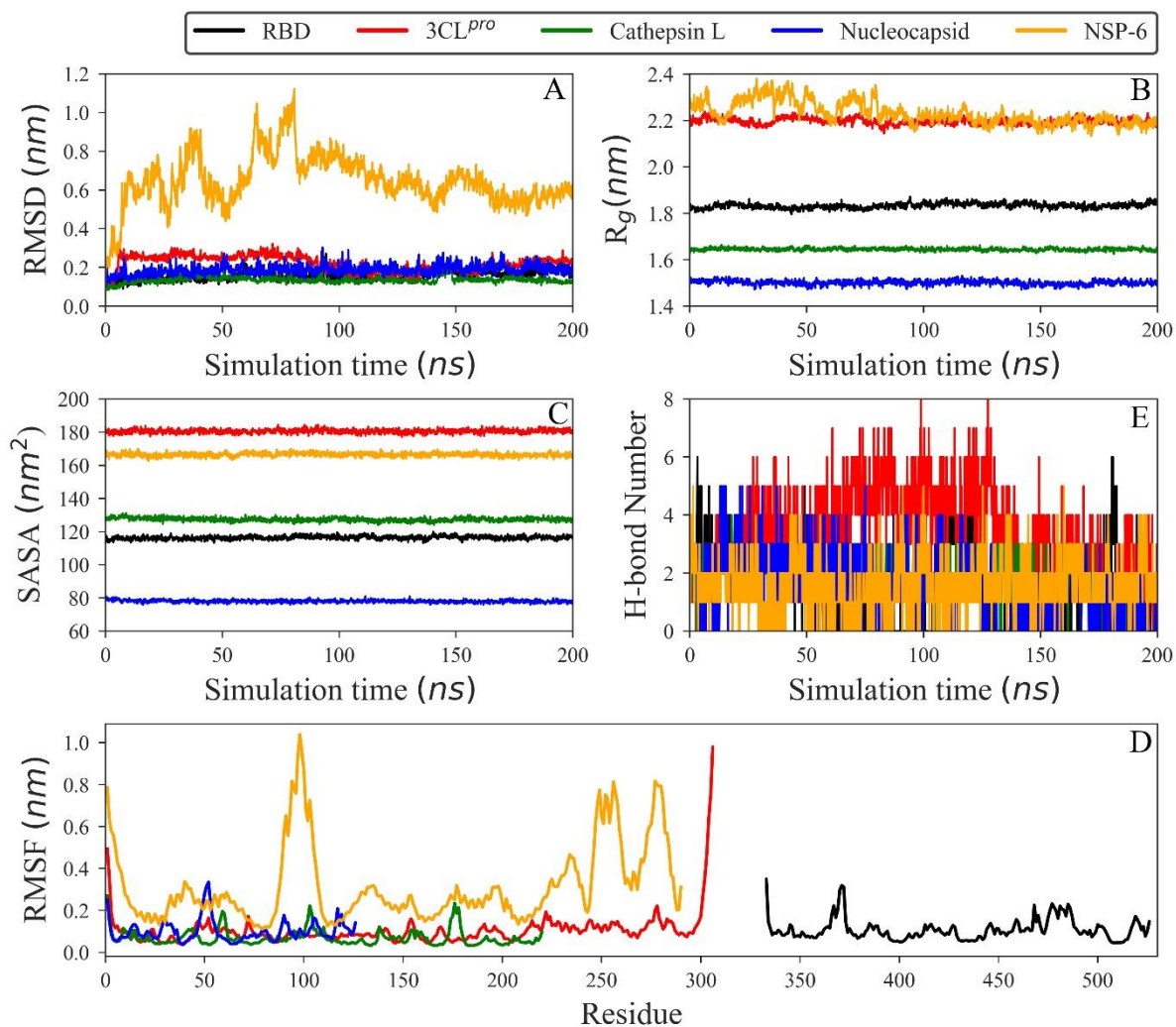
Figure 2. Time evolution plot of the structural order parameters of the antiviral drug, catechin docked complex with target proteins: RBD, 3CL^{pro}, Cathepsin L, Nucleocapsid and NSP6, (A) The root-mean square deviation (RMSD) of backbone C$^\alpha$- atoms (B) radius of gyration (R_g) (C) solvent accessible surface area (SASA) plots (D) root-mean square fluctuation showing the average fluctuation of amino acid residues and (E) the propensity of H-bonds interaction between the proteins ad catechin during the period of simulation (200 ns), at 300 K.
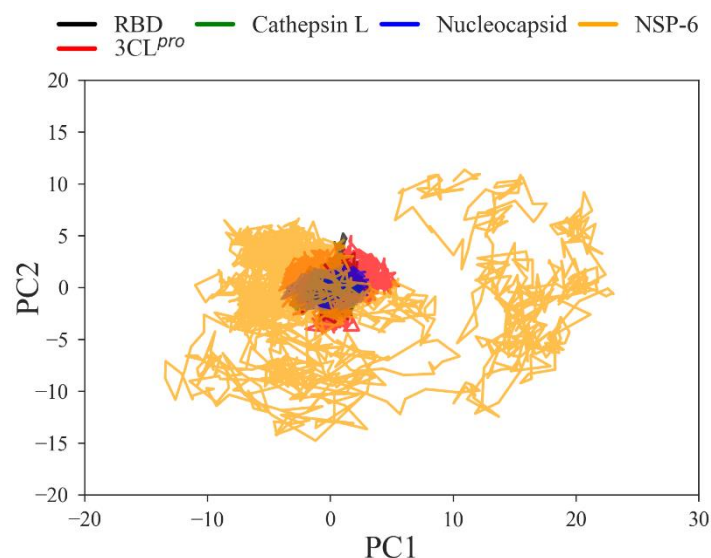
Figure 3. Essential dynamics plots. The conformational landscape of target proteins complexed with FDA approved antiviral drug, catechin. The projection of the collective motion of proteins in the essential subspace along the principal components (PCs), PC1 and PC2. The target proteins are represented with different color code as shown in schema.
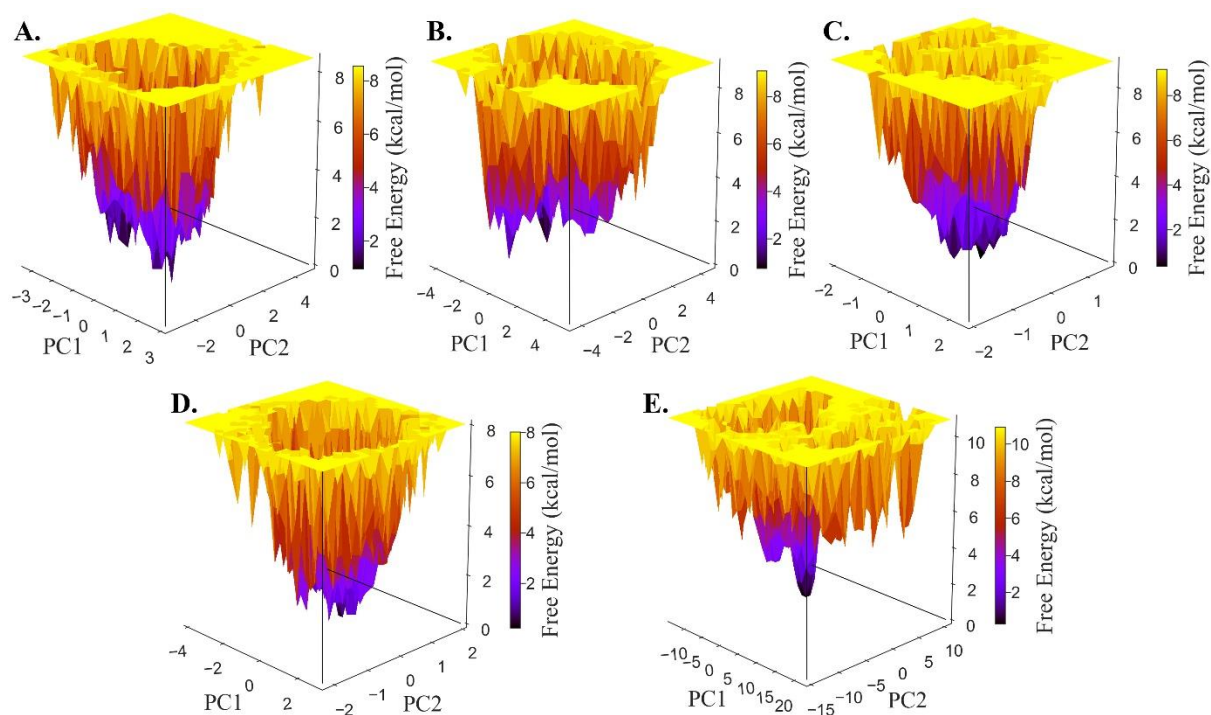

Figure 4. Free energy landscape (FEL) of target proteins complexed with catechin (A) RBD (B) 3CL^{pro} (C) Cathepsin L (D) Nucleocapsid and (E) NSP6. The free energy is given in kcal/mol and indicated by the color code, from lower to higher energy in the right panel.
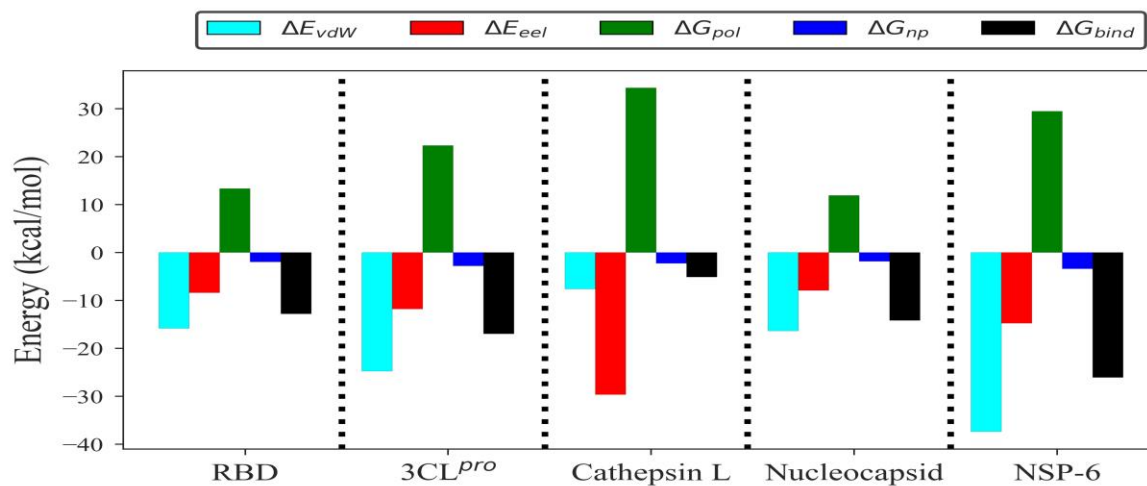
Figure 5. The binding free energy terms obtained from MM-PBSA calculations relative to the binding of catechin with five target proteins as labelled in plot and the color code for different energy components shown in schema.
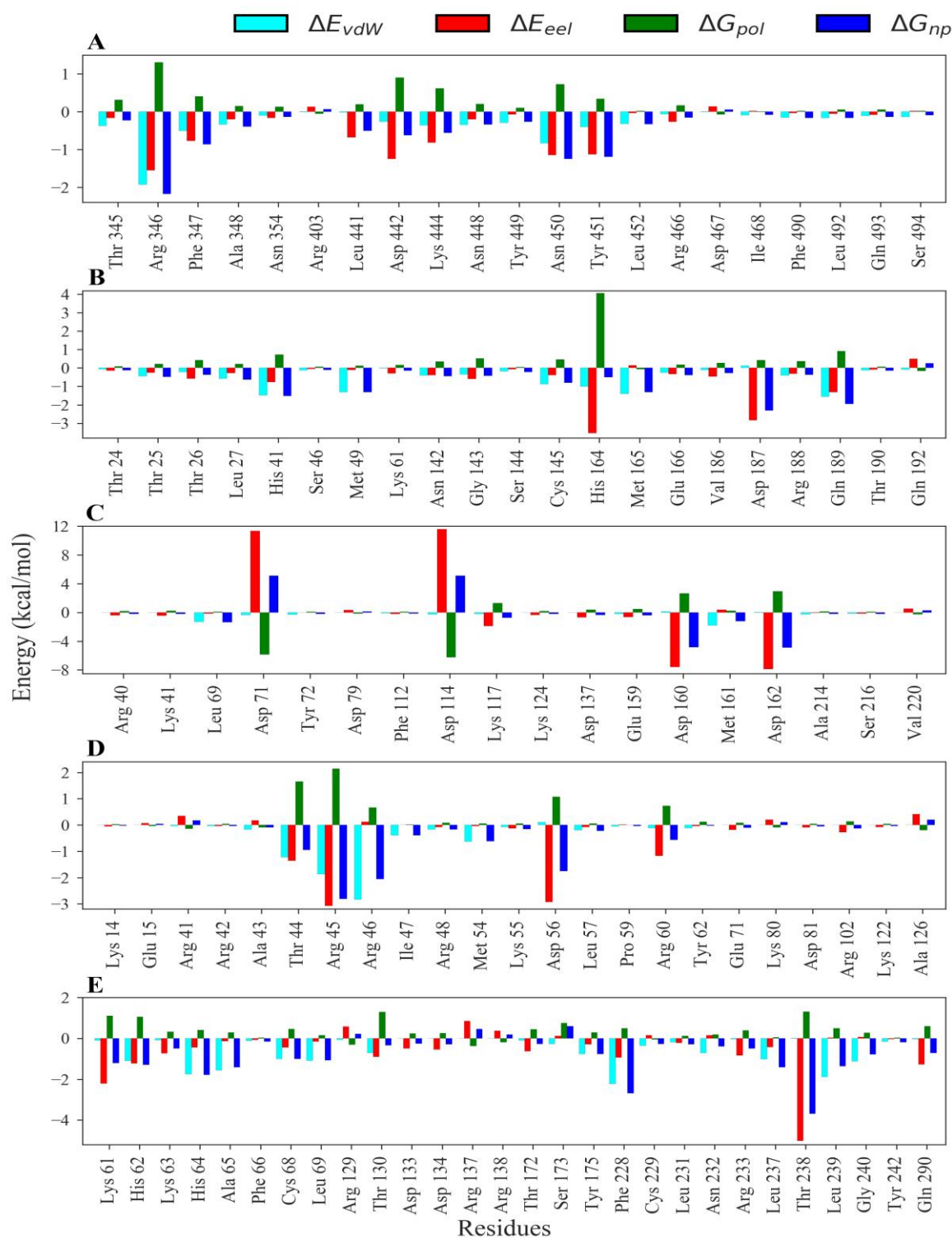
Figure 6. The residue decomposition plot (MM-PBSA), representing the binding energy contribution of active site residues of five target protein energetically stabilizing the catechin at binding pockets, (A) RBD (B) 3CL^{pro} (C) Cathepsin L (D) Nucleocapsid and (E) NSP6.