Virtual repurposing of Ursodeoxycholate and Chenodeoxycholate as lead candidates against SARS-Cov2-Envelope protein: A molecular dynamics investigation

Short running title: Effect of Ursodeoxycholate and Chenodeoxycholate on SARS-Cov2-E

Authors:
Reena Yadav¹#; Ph.D. student
Chinmayee Choudhury¹#; Ph.D.
Yashwant Kumar²; MD, DNB
Alka Bhatia¹*; MD

¹Department of Experimental Medicine and Biotechnology, and ²Department of Immunopathology, Postgraduate Institute of Medical Education and Research, Chandigarh, India

# These authors contributed equally to this work

*Correspondence
Dr. Alka Bhatia
Professor
Department of Experimental Medicine & Biotechnology,
Research Block B, PGIMER, Chandigarh
E-mail: alkabhatia@ymail.com
Phone no. 0172-2755271
Fax: 0091-172-2744401

Author contribution
CC and RY have performed the calculations, analysed the results and written the manuscript, YK and AB have contributed to conceptualization, designing and drafting of the final version of the manuscript.
Abstract

Drug repurposing is an apt choice to combat the currently prevailing global threat of COVID-19, caused by SARS-Cov2 in absence of any specific medication/vaccine. The present work attempts to computationally evaluate binding affinities and effect of two widely used surfactant drugs i.e. chenodeoxycholate (CDC) and ursodeoxycholate (UDC) with the envelope protein of SARS-Cov2 (SARS-Cov2-E) using homology modelling, molecular docking and molecular dynamics simulations. A good quality homo-pentameric structure of SARS-Cov2-E was modelled from its homologue with more than 90% sequence identity followed by symmetric docking. The pentameric structure was embedded in a DPPC membrane and subsequently energy minimized. The minimized structure was used for blind molecular docking of CDC and UDC to obtain the best scoring SARS-Cov2-E–CDC/UDC complexes, which were subjected to 230ns molecular dynamics simulations in triplicates in DPPC membrane environment. Comparative analyses of structural and enthalpic properties and molecular interaction profiles from the MD trajectories revealed that, both CDC and UDC could stably bind to SARS-Cov2-E through H-bonds, water-bridges and hydrophobic contacts in the transmembraneresidues. T30 was observed to be a key residue for CDC/UDC binding. The polar functional groups of the bound CDC/UDC facilitated entry of a large number of water molecules into the channel and affected the H-bonding pattern between adjacent monomeric chains, loosening the compact transmembrane region of SARS-Cov2-E. These observations suggest the potential of CDC/UDC as repurposed candidates to hinder the survival of SARS-Cov2 by disrupting the structure of SARS-Cov2-E and facilitate entry of solvents/polar inhibitors inside the viral cell.
Keywords

Chenodeoxycholate; Molecular docking; Molecular Dynamics simulations; SARS-Cov2-E;
Ursodeoxycholate; drug repurposing
Abbreviations

ACE2: Angiotensin converting enzyme 2; CDC: Chenodeoxycholate; SARS-Cov2: Severe Acute Respiratory Syndrome Corona Virus 2; UDC: Ursodeoxycholate;
1 INTRODUCTION

Coronavirus disease 2019 (COVID-19) is the currently prevailing pandemic which is caused by Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV2) and has emerged as a global threat. SARS-CoV2 is the recent active member of the β-coronaviruses class which emerged from Wuhan, China in December 2019 and spread to approximately 216 countries within no time. It was isolated and sequenced in January 2020 and found that SARS-CoV2 show 79.5% and 96% genetic similarity to SARS-CoV and bat SARS-CoV respectively (Chen, Liu et al. 2020). Molecular level studies have explained that SARS-CoV2 like SARS-CoV binds to cells of respiratory mucosa through the angiotensin converting enzyme 2 (ACE 2) receptors in the human body. The virus is enveloped and possesses single stranded positive sense RNA. The RNA genome is of 30 kB in size and is involved in encoding viral replicases associated with genome synthesis and other sub-genomic mRNAs. The viral RNA encodes structural proteins like Spike glycoprotein, Envelope protein, Membrane and Nucleoprotein and the non-structural proteins for example chymotrypsin like main protease (Khailany, Safdar et al. 2020). Although many candidate vaccines and potential drugs with probable efficacy against virus are in process of development, till date no specific treatment is available for treatment. The current treatment being used is purely supportive and symptomatic with advancements being incorporated day by day.

Drug discovery, is a laborious, time consuming and expensive process as it takes around 10-15 years which emphasize upon the importance of screening of potential drugs. Drug repurposing, is the recently used strategy for investigating new uses outside the original medication scope of already approved drugs. This strategy has grabbed greater attention due to its various advantages over the conventional drug development like low failure risk(as the repurposed drug is already safe in preclinical models and human trials) and lesser time and cost involvement(Pushpakom, Iorio et al. 2019). In case of SARS-CoV2 also the first line of
treatment are the repurposed drugs. Based on past experiences with other SARS, antiviral drugs like ribavirin and lopinavir-ritonavir are being used (Caly, Druce et al. 2020). An FDA approved anti-parasitic, Ivermectin has also been shown to inhibit SARS-Cov2 in vitro. Another antiviral drug remdesivir which have been shown to be effective in both patients and in vitro conditions is also emerging as a potent treatment for COVID-19(Cascella, Rajnik et al. 2020). Other repurposed drugs currently used in SARS-Cov2 treatment involve hydroxychloroquine, chloroquine, favipiravir and nitazoxanide (Wang, Jiang et al. 2019; Cai, Yang et al. 2020; Yao, Ye et al. 2020).

Computational methods are being popularly used to screen molecules and understand drug-target interactions as cheaper and faster alternatives of high throughput screenings (Choudhury, Priyakumar et al. 2015; Choudhury, Priyakumar et al. 2016; Choudhury 2020). Membrane and membrane associated proteins of SARS-Cov2 are attractive drug targets for anti-Covid-19 drug discovery. In this scenario, considering common bile salts, which act as solubilizers and emulsifiers for cholesterol and other water insoluble compounds in intestines (Roda, Cerré et al. 1995) as inhibitors of the SARS-Cov2 membrane associated proteins is interesting in its own right. These salts have a chemical structure composed of steroid nucleus, hydroxyl groups and carboxyl group which makes them amphiphilic (Scheme 1). Due to their structure, they act as natural biosurfactants/detergents having micellar properties (Ninomiya, Matsuoka et al. 2003). Ursodeoxycholate (UDC) and chenodeoxycholate (CDC) are the drugs recommended for the treatment of gallstones as they are reported to reduce cholesterol saturation of bile (Hirota, Chijiiwa et al. 1992). The tetracyclic steroid nucleus has been recognized as one of the five most common unique ‘privileged’ molecular scaffolds, occurring in the natural product space because of its capability to bind to multiple proteins (Choudhury, Deva Priyakumar et al. 2016). CDC works by dissolving cholesterol molecules in mixed micelles. Clinically, UDC is as effective as CDC despite its poor micellar solubility
of cholesterol as the former is believed to solubilize cholesterol as liquid crystals (vesicles). They are also used as transporters or carriers of physically complexed or chemically conjugated drugs due to their drug absorption enhancer characteristics. They increase the drug bioavailability either by increasing solubility and dissolution rate of drugs or by increasing the membrane fluidity and permeability (Pavlović, Golocočorbin-Kon et al. 2018). Few recent communications have also pointed towards the anti-inflammatory, immunomodulatory and anti-apoptotic effects of UDC to tack the cytokine storm syndrome which is strongly associated with severe tissue damage and fatal outcomes of COVID-19 (Abdulrab, Al-Maweri et al.; Subramanian, Iles et al. 2020). In addition derivatives of UDC and obeticholic acid have also been shown to prevent virus entry by binding to the receptor binding domain of spike glycoprotein of SARS-CoV2 (Carino, Moraca et al. 2020). Considering the presence of a privileged natural product scaffold in CDC and UDC and their surfactant and absorption enhancer properties, the effect of these two on the SARS-CoV2-E is investigated here using computational methods.

Scheme 1. 2D structure of CDC and UDC

SARS-CoV2-E shares more than 90% sequence identity with the envelope of SARS-CoV, which is a short, integral membrane protein crucial for viral assembly, release of virions and pathogenesis of the virus (Corse and Machamer 2003; Teoh, Siu et al. 2010; Ruch and Machamer 2011). The hydrophobic region of transmembrane domain contains amphipathic α-helix that oligomerizes to form an ion conductive pore in membrane. Synthetic
peptides corresponding to full length or N-terminal of SARS-CoV envelope have shown that it forms cation selective ion channels in planar lipid bilayers (Wilson, McKinlay et al. 2004). The effects of UDC and CDC on the overall structural dynamics and solvent permeability of SARS-Cov2-E has been computationally investigated in the study and the outcome points towards the capability of UDC and CDC to act as emerging leads against SARS-Cov2-E.

2 METHODOLOGY

2.1 Homology modelling, refinement and structure validation

The experimental structure for the full length SARS-Cov2-E is not available except for a recently reported NMR solution structure of only the transmembrane region (residue 8-38). So the SWISS-MODEL server (Waterhouse, Bertoni et al. 2018) was used to model the 3D structure of the full length protein from the FASTA sequence based on the homologous template structure of SARS-Cov envelope protein (2mm4.pdb), which shares more than 90% sequence similarity with SARS-Cov2-E. The generated model was further submitted for structure quality assessment in SWISS-MODEL, ERRAT (Colovos and Yeates 1993) and PROCHECK (PDBsum) servers (Laskowski, Chistyakov et al. 2005). Based on these assessments, the disordered regions of the proteins were refined by iteratively performing steepest descent of energy minimizations and quality assessments. The final good quality monomeric model was used to build a pentameric form, as the template SARS-Cov-E forms a pentameric quaternary structure. The SymmDock server (Schneidman-Duhovny, Inbar et al. 2005) was used to generate a pentameric structure from the modelled monomer of SARS-Cov2-E. The server returned 20 best pentameric structures, out of which the second-best model was selected for further work based on the lowest RMSD (3.44 Å) with the available three-dimensional pentameric structure of SARS-Cov-E pentameric structure (5x29.pdb).

2.2 Blind Docking with the pre-processed Comparative Model of SARS-Cov2-E
The modelled pentameric structure of SARS-Cov2-E was subjected to pre-docking preparation using Protein Preparation Wizard (PPW) module (PPW; Epik, Schrödinger, LLC, New York, NY, 2019) of Schrödinger software package, version 2019-2. The structure was pre-processed by adding missing hydrogens and assigning appropriate bond orders structure. The protonation states of the polar residues were optimized using PROPKA and all the newly added hydrogen atoms were minimized impref to avoid steric clashes. The resultant pre-processed structure was further used for preparation of grids followed by docking and molecular dynamics (MD) simulations. Before proceeding for docking, the pentameric model was put inside a predefined dipalmitoyl phosphatidylcholine (DPPC) membrane (at 325K) taking residues 17-37 as the trans membrane atoms using the system builder tool of Desmond (Desmond Research, 2018). The membrane embedded protein was then solvated with TIP3P water and 10000 steps of steepest descent energy minimization was performed on the system to remove steric clashes and generate a low energy conformer of the protein in a membrane environment. The SARS-Cov2-E protein was then extracted from the membrane and used for docking. ‘Receptor Grid Generation’ module of Schrödinger was utilized to define interaction grids for molecular docking keeping the centroid of the whole protein as the grid centre. The structures of UDC and CDC were prepared using LigPrep (LigPrep, Schrödinger, LLC, New York, NY, 2019), where their ionization states at pH 7.0 (± 2.0) were generated using Epikionizer, their energies were minimized and 5 best conformers for each of the ligands were generated. The size of the interaction grid was fixed to 16 Å for inner box and 20Å as outer box so that the whole protein was covered. UDC and CDC were docked to the interaction grid using the Glide XP module of Schrödinger software package and 10 best poses were generated for each ligand conformer. OPLS_2005 force field (Shivakumar, Williams et al. 2010) was used for docking with all default parameters. Two protein-ligand complexes taking best energy poses of both the ligands with SARS-Cov2-E were generated.
for further study. The resultant complexes of the molecules with SARS-CoV2-E were further submitted for binding energy estimation, where Molecular Mechanics-Generalized BornSurfaceArea (MM/GBSA) based binding free energy ($\Delta G_{\text{bind}}$) was computed for the complexes and MD simulations.

2.3 Molecular Dynamics Simulations

MD simulations were executed on the complexes of SARS-CoV2-E with UDC and CDC and also the SARS-CoV2-E model without any ligand, using the Desmond MD simulation package (release 2018) of Schrodinger (Desmond Research, 2018). The OPLS_2005 force field was employed for the protein-ligand complexes. All the three model systems were first put inside a predefined DPPC membrane (at 325K) taking residues 17-37 as the transmembrane atoms using the system builder tool of Desmond as described in the previous section. Then, the membrane bound complexes were solvated in a cubical water box (TIP3P water model) keeping 10 Å buffer space in x, y and z dimensions. Each system was neutralized by adding appropriate counter ions and an ionic concentration of 0.15 M was maintained by adding Na+ and Cl− ions. The systems were minimized with 5000 steepest descent steps followed by gradual heating from 0 to 300 K, under NVT ensemble. The systems were thermally relaxed before the production run using Nose-Hoover Chain thermostat method for 5 ns and 5 ns of pressure relaxation with Martyna-Tobias-Klein barostat method. The production run was performed in two phases. Initially 30 ns of production run was carried out under NPT ensemble each system using a cut-off distance of 12 Å for non-bonded interactions. Coordinates were saved at each 30 ps to generate trajectories of 1000 frames each. MMGBSA binding energies of CDC and UDC were calculated from 30 snapshots collected at equal intervals from each trajectory and two lowest energy complexes were obtained from each trajectory, which are denoted as SARS-Cov2-
E+CDC (2), SARS-Cov2-E+CDC (3), SARS-Cov2-E+UDC (2) and SARS-Cov2-E+UDC (3).

Figure 1. Overall Workflow of the study
These four complexes were subjected to 100 ns of final phase MD simulations (in DPPC membrane, with similar parameters described above) to confirm the stable binding of CDC/UDC with SARS-CoV2-E. 100 ns of simulations were performed on the lowest energy snapshot of SARS-CoV2-E (out of the 30 snapshots collected at equal intervals from the 30 ns trajectory), which was denoted as SARS-Cov2-E (2). Thus, the CDC/UDC bound systems were simulated for a total of 230 ns each, while SARS-CoV2-E system was simulated for a total of 130 ns. The five 100 ns trajectories (two of SARS-Cov2-E+CDC, two of SARS-Cov2-E+UDC and one of SARS-Cov2-E) were comparatively analysed to study the effect of CDC/UDC binding to SARS-CoV2-E. Figure 1 shows the overall workflow of the study. All trajectories were analysed with simulation even analysis, simulation interaction diagrams of Desmond.

3 RESULTS AND DISCUSSION

3.1 Comparative modelling of SARS-CoV2-E pentameric structure

The 3D structure of SARS-CoV2-E was modelled using the SWISS-MODEL server. Upon submitting the FASTA sequence of SARS-CoV2-Ein SWISS-MODEL, initially eight templates were identified for the SARS-CoV2-E and out of these 2mm4.1A which is a monomer of the conserved Golgi complex targeting signal in coronavirus Envelope protein was used as template for further modelling. The modelled monomer was assessed on many parameters for its quality in SWISS-MODEL itself and with other popularly used protein structure quality prediction programs. Figure 2 shows the quality assessment results for the model pre and post refinement. Global Quality Estimate as predicted by QMEAN comprising of four individual terms or Z-scores which represent the interaction potential between Cβ atoms only, all atoms, the solvation potential and the torsion angle potential(Figure S1a). This value represents the degree of native-ness of a protein and is recommended to be ideal,
when it is close to zero and higher than -4. The initial model generated for SARS-CoV2-E was having a QMEAN of -3.27, indicating its acceptability.

Sequence identity: 93%

Figure 2. Homology modelling and structure quality assessment results
The Comparison plot shown in Figure S1b assesses the model quality by relating the quality scores of individual models with scores obtained for experimental structures of similar size. The position of the modelled protein (red star) in this plot was found within the acceptable range. The structure was submitted to the PROCHECK server to obtain the Ramachandran plot of the phi and psi dihedrals and G-factors for the main chain bond lengths, angles and dihedrals. The initial model was found to have 99% of residues in the allowed regions. To determine if there is any disordered region of the protein, the model was submitted to the ERRAT server. ERRAT verifies the structures by statistically comparing the intra-molecular non-covalent interactions of the model with that of the reported high-resolution structures. The overall quality factor should be more than 91% for a reasonably good structure. The overall quality factor of the initial model was found to be low initially with several disordered residues in both the terminal regions. From these structure quality scores it was clear that further structure refinement of the initial model was highly required as it might not be suitable to be used for further studies. The Prime Loop refinement program of Schrodinger was used to perform a thorough refinement of the disordered regions of the initial model (Figure 2). Structure refinement and ERRAT analysis were iteratively done till we obtained a good quality final structure with an overall quality factor of 94 (Figure 2). This monomeric structure was further submitted to the SymmDock server to perform a geometry-based docking for generation of a pentameric assembly of SARS-Cov2-E. This server returned 20 best pentameric structures, each of them was superposed with the pentameric structure of the homologous envelope protein of SARS-CoV. The root mean squared deviations (RMSD) ranged from 4.47 to 3.44 Å. The pentameric form with the lowest RMSD i.e., 3.44 Å was considered for docking calculations of CDC and UDC and MD simulations.

The reliability of the modelled structure used for our study was also validated by aligning it with the very recently reported NMR structure 7K3G (only the trans membrane region is
resolved) and the alignment score was found to be as low as 0.124 and 0.245 for the monomer and the pentameric structures respectively.

3.2 Design of Model Systems of SARS-Cov2-E bound to CDC and UDC and Molecular Dynamics Simulations

Molecular docking with the Glide program returned a total of 26 poses for CDC and 6 poses for UDC, the XP docking scores ranging from -5.37 to -7.32 for CDC and from -6.87 to -7.36 for UDC. The average RMSD among the poses were below 1.5 Å and almost all of them showed H-bond with T30 of Chain A. Figure S2 shows the top scoring binding pose of both CDC and UDC with SARS-Cov2-E. Both of them make H-bond with T30 of Chain A, however the difference was, in case of CDC the atom O27 (-OH group attached to the 10th carbon position) acts as a H-bond donor while the atom O26 (-OH group of the fragment substituent at the 6th carbon position) acts as a H-bond donor in case of UDC. The binding energies of the complexes were calculated to be -40.83 and -39.94 with CDC and UDC respectively. MD simulations were performed on the complexes of CDC and UDC with SARS-Cov2-E as well as SARS-Cov2-E without any ligand in order to explore the effect of CDC and UDC binding on the structural dynamics of SARS-Cov2-E. Atomistic MD approach has been employed here also to monitor and enhance the stabilities of the protein-ligand (CDC/UDC) interactions under dynamical conditions. SARS-Cov2-E being a membrane protein, the simulations were performed in a DPPC membrane environment surrounded by TIP3P water box. Figure S3 shows the initial set up of the model systems.

Most MD methods are known to be based on heuristic approximations and they do not sample the entire conformational space of a protein-ligand system effectively. In order to address these limitations in our study, the MD simulations were carried out in two different phases. The initial phase of 30 ns simulations were run on the three systems, where the
receptors (SARS-CoV2-E) had the same initial coordinates. In the final phase, 100 ns simulations were carried out on one lowest energy structure obtained from the initial 30 ns MD trajectories of system SARS-CoV2-E and two lowest energy structures obtained from each of the initial 30 ns MD trajectories of systems SARS-CoV2-E+CDC and SARS-CoV2-E+UDC. These five 100ns trajectories were named as SARS-CoV2-E (2), SARS-CoV2-E+CDC (2), SARS-CoV2-E+CDC (3), SARS-CoV2-E+UDC (2) and SARS-CoV2-E+UDC (3). The initial coordinates of the receptors and ligands in all the above five simulations were different, which enabled us to sample the conformational spaces more robustly. Variations of different structural properties of the protein and the bound ligands were analysed independently for the initial and final phase MD simulations.

3.3 Structural Dynamics of SARS-CoV2-E, with and without bound CDC/UDC: Observations from the Initial 30ns Simulations

Figure S4 (a-h) show variations in different structural properties of the three systems in the initial 30 ns simulations. RMSD graphs of the systems in the initial 30 ns run indicated that the structures were stabilized within the first 5 ns only and there was no further significant change in the structure for next 25 ns. However, the systems SARS-Cov2-E + CDC and SARS-Cov2-E + UDC were stabilized at about 4-5 Å, while SARS-Cov2-E without any ligand was stabilized at a higher RMSD. This observation indicated that somehow ligand binding conferred structural stability to SARS-Cov2-E. Figure S4b shows the root mean squared fluctuations (RMSF) of residue (all the heavy atoms of each residue were considered) of the model systems during the simulations. Vertical lines against the residues indicate presence of ligand contact with the corresponding residue. RMSF quantifies the displacement of the centre of mass of the residues from a mean position during the simulation. The RMSF plot shows that, the terminal residues of all the chains except chain C show high fluctuations in the ligand bound systems as compared to the system without
ligand. The secondary structure analysis of the three complexes from the MD simulation results shows that the total secondary structure elements (SSE) were maintained throughout the MD Simulation in all the three complexes (Figure S5). However, minor changes in the arrangement of alpha-helices in terminal residues were observed in UDC and CDC bound complexes. Comparing the ligand contacts, it was observed that, CDC mostly makes residue contacts with chains C, D and E while UDC mostly makes contacts with chains A and B. Radius of gyration (RГyr) is used as a measure of the compactness of a system. In Figure S4d, RГyr profile of SARS-Cov2-E was clearly observed to be lower than that of the ligand bound systems, indicating that ligand binding makes the membrane protein less compact. RMSD of the ligands with respect to their own initial structures (Figure S4f) were found to be quite lower suggesting that there were very less fluctuations or conformational changes in the ligands’ internal configuration, but the RMSD with respect to the protein structures (Figure S4e) were as high as the RMSD of the overall protein structures (Figures S4a, S4b). However, the small difference in the RMSDs of the ligands with respect to themselves and with respect to the protein indicates that though the ligands translate from their initial positions along with the conformational changes of the overall protein, they do not show tendencies to diffuse out of the protein channel. The RГyr and solvent accessible surface area (SASA) of CDC (Figures S4g and S4h) showed stable profiles. In case of UDC, both these values were found to drop after 15 ns and then, were maintained stably till 30 ns.

3.4 Effect of CDC and UDC binding on the structure and dynamics of SARS-Cov2-E:

Observations from the final MD simulations

As discussed earlier, the final MD simulations were carried out on the lowest energy snapshots from the initial MD trajectories. This enabled us to evaluate the structural evolutions of the system from different starting points with a robust sampling the conformational spaces of the complexes. As observed in the RMSD plot, the SARS-CoV2-E
and the UDC bound structures equilibrated within 2.5-3 Å, while the CDC bound structures showed deviations up to 3.5-4 Å. The plot indicated that, the CDC bound systems underwent higher conformational changes as compared to the UDC bound structures. The RMSF plots show, slightly high fluctuations up to 3Å in the residues 40-49 (extracellular side) of the C-Chain of SARS-CoV2-E and A, B and E chains of the CDC and UDC bound systems. Figures 3c and 3d show the solvent accessible surface area (SASA) of the ligand binding transmembrane regions and the RGyr of the systems. These plots clearly show a significant increase in the SASA of the transmembrane regions and high RGyr (less compact or loose structures) upon CDC/UDC binding. The total energies of the CDC/UDC bound structures were observed to be lower than the structure without any ligand (Figure 3e).

Figure 3. Variations of different structural properties of the protein-ligand complexes in different model systems during final simulations.
The ligands CDC/UDC did not show high RMSD with respect to themselves throughout the simulation. RMSD of UDC with respect to the receptor was found to be as high as 4 Å as compared to CDC. However, CDC or UDC were not diffused out of the protein and the higher values for UDC was due to the translational movements of the ligand in the transmembrane region. SASA of UDC in the system SARS-CoV2-E+UDC (3) was shown to be slightly higher than the other systems. This also shows lower RGyr as compared to the other ligands. The interaction energy of CDC with SARS-COV2-E was found to be lower as compared to that of UDC indicating a more stable binding of CDC (Figure 4).

**Figure 4.** Variations in structural and enthalpic properties of ligands in all the model systems during the final MD run.
In order to understand the degrees of correlated movements between the monomers of SARS-CoV2-E in presence and absence of bound CDC/UDC were analysed by calculating the covariance matrices shown in Figure 5 (Roy and Post 2012). The covariance matrix of the system SARS-CoV2-E + CDC (2) shows highly anti-correlated movement of residues of chains B and C with respect to chain A and Chain E. Similar high anti-correlated movements were observed among chains B, C and A, D, E in the system SARS-CoV2-E + UDC (3). Such anti-correlated movements might be due to passage of high number of water molecules inside the transmembrane region in the CDC and UDC bound systems.

Figure 5. Covariance matrices of the residues of systems a) SARS-CoV2-E+CDC (2), b) SARS-CoV2-E+CDC (3), c) SARS-CoV2-E+UDC (2), d) SARS-CoV2-E+UDC (3), e) SARS-CoV2-E (2). The residue numbers are plotted in both the axes and the Chains are denoted as different colours. A positive correlation coefficient indicates a correlated motion, while a negative correlation coefficient indicates an anti-correlated movement between the residues. A threshold of 0.25 for the correlation coefficient has been recommended (Roy and Post 2012) and is used in the present study to understand correlated movements in the protein domains in various systems. Dark purple patches correspond to the anticorrelated movements of the corresponding residues and the orange/pink ones the correlated movements.
3. 5 Differential binding of CDC and UDC with SARS-Cov2-E

CDC and UDC have same molecular formula, same molecular scaffold and functional groups as the latter is a 7[beta] hydroxyl epimer of the former. Though their conformations differ only at one chiral centre, this small conformational difference attributes to their differential binding with SARS-Cov2-E. Figure S6 (a-b) show the Interactions of CDC and UDC with SARS-Cov2-E during the initial 30ns MD run. H-bond interactions with >50% occupancy with a particular residue is shown in these interaction diagrams. Both CDC and UDC formed stable H-bonds with T30 of E and B chains respectively. CDC also showed stable hydrophobic contacts and water bridge interactions of the C, D and E chain residues whereas, UDC mostly interacted with the A and B chain residues through H-bond and water bridge interactions in the initial 30 ns MD run. Figure S6c shows MM/GBSA energy of CDC and UDC binding throughout the MD simulation, calculated from snapshots saved at each 3 ns. The plots clearly show that ligand binding has been enhanced as compared to the initial poses, when subjected to MD simulations. The binding energies were significantly improved from -40 to as low as -72 kcal/mol and from -39 to as low as -66 kcal/mol for CDC and UDC respectively during the initial simulations. Two snapshots with the lowest MM/GBSA binding energies were picked from each trajectory for the final 100 ns MD simulations (marked by circles in the Figure S6c). Figure6 gives a clear picture of the ligand binding pattern and type of interactions of both the ligands during the final MD simulations. In the docked pose, both the ligands were making H-bonds with T30 of chain A, but the binding pattern significantly changed throughout the initial and final MD simulations. H-bond of both CDC and UDC with T30 was maintained in the MD simulations, but CDC showed stable H-bond with T30 of chain E, while UDC maintained stable H-bond with T30 of chain A.
Figure 6. Interactions of CDC and UDC with SARS-Cov2-E in different model systems. H-bond interactions with >50% occupancy with a particular residue is shown in the interaction diagrams.
CDC binding was mostly stabilized by hydrophobic contacts and water bridge interactions of the C, D and E chain residues (C:F26, E:L18, E:A22, E:F23) whereas, UDC binding was mostly stabilized by H-bond and water bridge interactions with the A, B, C and E chain residues (B:T30, B:R38, B:S55, B:Y59, C:F23, E:T30). As observed from Figure 6, all the atoms of both the ligands were exposed to the solvent throughout the simulations. The interaction energy plot from the final simulations also showed that CDC binds more favourably with interaction energies varying between 50-65 kcal/mol as compared to UDC, whose interaction energies vary between 40-60 kcal/mol.

3.6 Alteration of inter-chain interactions and water permeability of SARS-Cov2-E upon CDC and UDC binding

SARS-Cov2-E has a pentameric quaternary assembly to form a transmembrane channel. In order to understand whether CDC or UDC binding has any effect on the quaternary structural arrangements of SARS-Cov2-E, we calculated the number of H-bonds formed between the individual chains (FigureS7). Figure 7a shows the average number of H-bond formed between the constituent chains of SARS-Cov2-E in presence and absence of the ligands CDC and UDC. It can be observed from Figure 7a, that in the native SARS-Cov2-E (2) pentameric structure an average of 3-6 H-bonds are formed between the adjacent chains i.e., between A and B, B and C, C and D, D and E and E and A (or A and E). Upon CDC binding, it was observed that the average number H-bonds between adjacent chains A-B, B-C, C-D and D-E have decreased significantly, while the average number of H-bonds between A-E has increased as compared to the SARS-CoV2-E (2) system. Upon UDC binding, the average number of H-bonds between all the adjacent chains significantly increases as compared to the SARS-CoV2-E (2) system except for those between D-E. Such imbalance in the number of inter-chain interactions due to ligand binding might be the reason for the reduced compactness of the ligand bound systems, which correlates with their higher RGyr
profiles (Figure 3d) of SARS-CoV2-E+CDC (2), SARS-CoV2-E+CDC (3), SARS-CoV2-E+UDC (2), SARS-CoV2-E+UDC (3). Such altered structural assemblies can lead to altered permeabilities through the transmembrane channel. To study this effect in a detailed manner, we calculated the number of water molecules present in the transmembrane channel of SARS-Cov2-E in ligand bound and unbound forms.

Figure 7. a) Average number of inter-chain hydrogen bonds calculated from MD results in all the three complexes b) Total number of water molecules inside the transmembrane channel c) graphical representation of the water molecules inside the transmembrane channel.

Figure 7b shows the total number of water molecules inside the transmembrane channel throughout the final simulations, calculated from the MD trajectories SARS-CoV2-E (2), SARS-CoV2-E+CDC (2), SARS-CoV2-E+CDC (3), SARS-CoV2-E+UDC (2), SARS-CoV2-E+UDC (3). Figure 7c gives a graphical representation of presence of water molecules
in the channel. It is very interesting to observe that, the number of water molecules that could enter the channel were significantly higher (ranging between 75 to 100) in case of the ligand bound systems, while it is significantly low (between 50 to 75), and shows a decreasing trend for the system without any bound ligand. The explanation to this can be as follows. The transmembrane region of SARS-Cov2-E comprises mostly of hydrophobic or non-polar amino acid residues creating a compact hydrophobic environment making it difficult for the water molecules to enter. When CDC/UDC binds, firstly, it loosens the structural assembly by altering the inter-chain H-bonds and makes the transmembrane channel wider. Secondly, the hydrophilic polar functional groups of CDC/UDC bind to and attract more water molecules into the channel. These effects of CDC/UDC can be exploited to explore them further as inhibitors of the SARS-Cov2-E channel or as effective carriers to facilitate entrance of polar inhibitors inside the viral cells.

The pharmacokinetics profile and the druglikeness of CDC/UDC were predicted using SwissADME server (Daina, Michielin et al. 2017). Both the drugs pass all the druglikeness filters such as Lipinski, Ghose, Veber etc. and also showed good bioavailability score along with high GI absorption (Table S1).

Conclusions

In the current study, we have attempted to explore CDC and UDC as potential drug repurposing candidates, which might act through binding to the SARS-CoV2-E protein. We have generated a good quality homo-pentamer of SARS-CoV2 using homology modelling and symmetry docking. Complexes of CDC/UDC with SARS-CoV2-E were generated using molecular docking. The structures of SARS-CoV2-E in free and CDC/UDC bound forms were subjected to a total of 130 and 230 ns molecular dynamics simulations in DPPC membrane environments in two phases. By comparative analyses of the MD trajectories, we
observed that, both CDC and UDC bind to the transmembrane regions of SARS-CoV2-E forming thermodynamically stable complexes. In spite of having the same molecular formula, molecular scaffold and functional groups, CDC binding is stabilized by mostly hydrophobic interaction and water bridges while UDC binds through hydrogen bonds and water bridges. T30 residue (belonging to any of the chains) was found to be a key residue for both CDC and UDC binding. Both CDC and UDC are found to disrupt inter-chain H-bonds between adjacent chains and loosen the overall structure of SARS-Cov2-E pentameric assembly and allow passage of more number of water molecules into the transmembrane region. In addition to this, the polar functional groups of CDC and UDC aid to attract more number of water molecules inside the transmembrane region. In view of all the above observations we suggest these two molecules can be of particular importance to disrupt the SARS-Cov2-E and also facilitate the entry/flow of water and other polar drugs/inhibitors into the viral cell. CDC/UDC are already known for their drug absorption enhancer characteristics as they increase the drug bioavailability either by increasing solubility and dissolution rate of drugs or by increasing the membrane fluidity and permeability (Pavlovic, Golocorbin-Kon et al., 2018). Our observations are in close agreement with the membrane permeability enhancing properties of CDC and UDC as mentioned in literature. These initial computational findings form the background to consider these endogenous surfactants/detergents as potential candidates for future laboratory-based studies for targeting SARS-Cov2-E.

**Acknowledgements:** CC thanks Department of Science and Technology for financial assistance in the form of DST-INSPIRE Faculty award (IFA16-LSBM-170) and Schrodinger for providing short term licenses for some of the modules. RY acknowledges Council of Scientific and Industrial Research, India for providing financial support in the form of Junior Research Fellowship.

**Declaration of interest:** The authors declare that there is no conflict of interest
References

Abdulrab, S., S. Al-Maweri, et al. Ursodeoxycholic acid as a candidate therapeutic to alleviate and/or prevent COVID-19-associated cytokine storm, Med Hypotheses. 2020 May 30;143:109897. doi: 10.1016/j.mehy.2020.109897.

Cai, Q., M. Yang, et al. (2020). "Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study." Engineering 18(10): 007.

Caly, L., J. D. Druce, et al. (2020). "The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro." Antiviral Res 178(104787): 3.

Carino, A., F. Moraca, et al. (2020). Hijacking SARS-Cov-2/ACE2 receptor interaction by natural and semi-synthetic steroidal agents acting on functional pockets on receptor binding region.

Cascella, M., M. Rajnik, et al. (2020). Features, Evaluation, and Treatment of Coronavirus.

Chen, Y., Q. Liu, et al. (2020). "Emerging coronaviruses: Genome structure, replication, and pathogenesis." J Med Virol 92(4): 418-423.

Choudhury, C. (2020). "Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease." J Biomol Struct Dyn 1: 1-14.

Choudhury, C., U. Deva Priyakumar, et al. (2016). "Structural and Functional Diversities of the Hexadecahydro-1H-cyclopenta[a]phenanthrene Framework, a Ubiquitous Scaffold in Steroidal Hormones." Mol Inform 35(3-4): 145-157.

Choudhury, C., U. D. Priyakumar, et al. (2015). "Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase." J Chem Inf Model 55(4): 848-860.

Choudhury, C., U. D. Priyakumar, et al. (2016). "Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors." Journal of Chemical Sciences 128(5): 719-732.

Colovos, C. and T. O. Yeates (1993). "Verification of protein structures: patterns of nonbonded atomic interactions." Protein Sci 2(9): 1511-1519.

Corse, E. and C. E. Machamer (2003). "The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction." Virology 312(1): 25-34.

Daina, A., O. Michielin, et al. (2017). "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules." Scientific Reports 7(1): 42717.

Hirota, I., K. Chijiwa, et al. (1992). "Effect of chenodeoxycholate and ursodeoxycholate on nucleation time in human gallbladder bile." Gastroenterology 102(5): 1668-1674.

Khailany, R. A., M. Safdar, et al. (2020). "Genomic characterization of a novel SARS-CoV-2." Gene Rep 19(100682): 100682.

Laskowski, R. A., V. V. Chistyakov, et al. (2005). "PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids." Nucleic Acids Res 33(Database issue).

Ninomiya, R., K. Matsuoka, et al. (2003). "Micelle formation of sodium chenodeoxycholate and solubilization into the micelles: comparison with other unconjugated bile salts." Biochim Biophys Acta 15(3): 116-125.

Pushpakom, S., F. Iorio, et al. (2019). "Drug repurposing: progress, challenges and recommendations." Nat Rev Drug Discov 18(1): 41-58.

Roda, A., C. Cerré, et al. (1995). "Experimental evaluation of a model for predicting micellar composition and concentration of monomeric species in bile salt binary mixtures." J Pharm Sci 84(5): 593-598.

Roy, A. and C. B. Post (2012). "Detection of Long-Range Concerted Motions in Protein by a Distance Covariance." Journal of Chemical Theory and Computation 8(9): 3009-3014.
Ruch, T. R. and C. E. Machamer (2011). "The hydrophobic domain of infectious bronchitis virus E protein alters the host secretory pathway and is important for release of infectious virus." J Virol 85(2): 675-685.

Schneidman-Duhovny, D., Y. Inbar, et al. (2005). "PatchDock and SymmDock: servers for rigid and symmetric docking." Nucleic Acids Res 33(Web Server issue).

Shivakumar, D., J. Williams, et al. (2010). "Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field." J Chem Theory Comput 6(5): 1509-1519.

Subramanian, S., T. Iles, et al. (2020). "Merit of an Ursodeoxycholic Acid Clinical Trial in COVID-19 Patients." Vaccines 8(2).

Teoh, K. T., Y. L. Siu, et al. (2010). "The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis." Mol Biol Cell 21(22): 3838-3852.

Wang, D.-Y., Z. Jiang, et al. (2019). "Molecular stratification within triple-negative breast cancer subtypes." Scientific Reports 9(1): 19107.

Waterhouse, A., M. Bertoni, et al. (2018). "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic Acids Res 46(W1): W296-W303.

Wilson, L., C. McKinlay, et al. (2004). "SARS coronavirus E protein forms cation-selective ion channels." Virology 330(1): 322-331.

Yao, X., F. Ye, et al. (2020). "In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)." Clin Infect Dis 71(15): 732-739.
Graphical Abstract

Virtual repurposing of Ursodeoxycholate and Chenodeoxycholate as lead candidates against SARS-CoV2-Envelope protein: A molecular dynamics investigation

Reena Yadav¹#, Chinmayee Choudhury¹#, Yashwant Kumar², Alka Bhatia¹*

¹Department of Experimental Medicine and Biotechnology, and ²Department of Immunopathology, Postgraduate Institute of Medical Education and Research, Chandigarh, India