Taylor & Francis
Taylor & Francis Group

Check for updates

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

Reena Yadav[a†], Chinmayee Choudhury[a†], Yashwant Kumar[b] and Alka Bhatia[a]

[a]Department of Experimental Medicine and Biotechnology, Postgraduate Institute of Medical Education and Research, Chandigarh, India; [b]Department of Immunopathology, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Communicated by Ramaswamy H. Sarma

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 employs state of art computational methods like homology modelling, molecular docking and molecular dynamics simulations to evaluate the potential of two widely used surfactant drugs namely chenodeoxycholate(CDC) and ursodeoxycholate (UDC), to bind to the envelope protein of SARS-Cov2(SARS-Cov2-E).The monomeric unit of SARS-Cov2-E was modelled from a close homologue ($>$90% sequence identity) and a pentameric assembly was modelled using symmetric docking, followed by energy minimization in a DPPC membrane environment. The minimized structure was used to generate best scoring SARS-Cov2-E–CDC/UDC complexes through blind docking. These complexes were subjected to 230 ns molecular dynamics simulations in triplicates in a DPPC membrane environment. Comparative analyses of structural 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 with the transmembrane-channel residues.T30 was observed to be a key residue for CDC/UDC binding. CDC/UDC binding affected the H-bonding pattern between adjacent monomeric chains, slackening the compact transmembrane region of SARS-Cov2-E. Additionally, the polar functional groups of CDC/UDC facilitated entry of a large number of water molecules into the channel. These observations suggest CDC/UDC as potential candidates to hinder the survival of SARS-Cov2 by disrupting the structure of SARS-Cov2-E and facilitating the entry of solvents/polar inhibitors inside the viral cell.
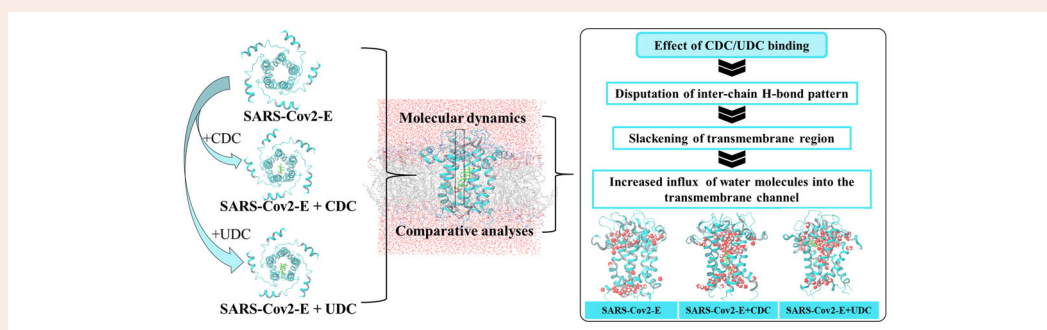
Abbreviations: ACE2: Angiotensin-converting enzyme 2; CDC: Chenodeoxycholate; SARS-Cov2: Severe Acute Respiratory Syndrome Corona Virus 2; UDC: Ursodeoxycholate

1. Introduction

Coronavirus disease 2019 (COVID-19) is the currently prevailing pandemic which is caused by the 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. The virus isolated and

sequenced in January 2020 was found to show 79.5% and 96% genetic similarity to SARS-Cov and bat SARS-Cov respectively (Chen et al., 2020). Molecular level studies have shown 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 a single stranded positive sense RNA. The RNA genome is 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 et al., 2020). Although many candidate vaccines and potential drugs with probable efficacy against the virus are in the 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. Drug repurposing is an alternative strategy for investigating already approved drugs, for new indications beyond the scope of their original usage. This strategy has grabbed greater attention due to its various advantages over conventional drug development, such as, low failure risk (as the repurposed drug is already safe in preclinical models and human trials) and lesser time and cost involvement (Pushpakom et al., 2019). In the case of SARS-Cov2 also, the first line of treatment is the repurposed drugs. Based on the past experiences with other SARS, antiviral drugs like ribavirin and lopinavir-ritonavir are being used (Caly et al., 2020) currently. An FDA approved anti-parasitic, Ivermectin has also been shown to inhibit SARS-Cov2 *in vitro*. Another antiviral drug remdesivir which has been shown to be effective in both patients and *in vitro* conditions is also emerging as a potent treatment for COVID-19(Cascella & Rajnik, 2020). Other repurposed drugs currently used in SARS-Cov2 treatment include hydroxylchloroquine, chloroquine, favipiravir and nitazoxanide (Cai et al., 2020; Wang et al., 2019; Yao 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 (Kurumurthy & Rao, 2011; Choudhury et al., 2015; Choudhury et al., 2016; Bhardwaj & Purohit, 2020; Bhardwaj et al., 2020; Choudhury, 2020; Singh, Bhardwaj et al., 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 the intestines (Roda 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 a steroid nucleus and hydroxyl and carboxyl groups which makes them amphiphilic (Scheme 1). Due to their structure, they act as natural biosurfactants/detergents having micellar properties (Ninomiya 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 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 (Bhardwaj et al., 2020; Choudhury 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 complex or chemically conjugated drugs due to their drug absorption enhancer characteristics. They increase the drug bioavailability either by increasing the solubility and dissolution rate of drugs or by increasing the membrane fluidity and permeability (Pavlović & Goločorbin-Kon, 2018). Few recent communications have also pointed towards the anti-inflammatory, immuno-modulatory and anti-apoptotic effects of UDC to tackle the cytokine storm syndrome which is strongly associated with severe tissue damage and fatal outcomes of COVID-19 (Abdulrab et al., 2020; Subramanian 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, 2020; Kumar & Yadav, 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-Eis investigated here using computational methods.

SARS-Cov2-Eshares 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 & Machamer, 2003; Ruch & Machamer, 2011; Teoh et al., 2010). The hydrophobic region of transmembrane domain contains amphipathic α-helix that oligomerizes to form an ion conductive pore in the 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 et al., 2004). In the current study, the effects of UDC and CDC on the overall structural dynamics and solvent permeability of SARS-Cov2-E have been computationally investigated. The observations of this study indicate the capability of UDC and CDC as promising 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 et al., 2018) was used to model the 3 D 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-
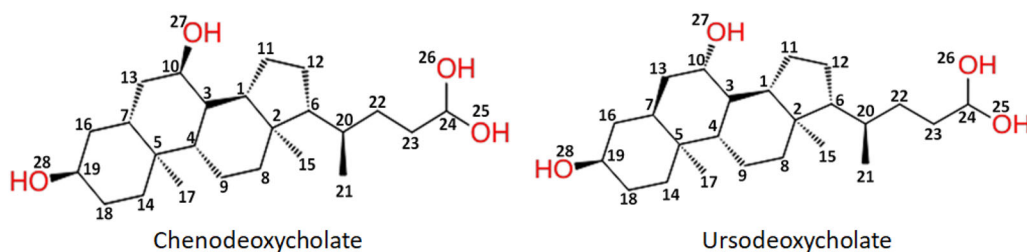
Scheme 1. 2D structure of CDC and UDC

Cov2-E. The generated model was further submitted for structure quality assessment in SWISS-MODEL, ERRAT (Colovos & Yeates, 1993) and PROCHECK (PDBsum) servers (Laskowski & Chistyakov, 2005). The modelled structure was also submitted to the ProSA server (Wiederstein & Sippl, 2007) for additional quality assessment. Based on these assessments, the disordered regions of the proteins were refined by iteratively performing the 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, 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 (5 × 29.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 (Li et al., 2005) and all the newly added hydrogen atoms were minimized *impref* to avoid steric clashes. The resultant pre-processed structure was further used for the preparation of grids followed by docking and molecular dynamics (MD) simulations. Before proceeding for docking, the pentameric model was put inside a predefined dipalmitoyl phosphatidyl choline (DPPC) membrane (at 325 K) taking residues 17–37 as the transmembrane atoms using the system builder tool of Desmond. The membrane embedded protein was then solvated with TIP3P water and 10000 steps of steepest descent energy minimization were performed on the system to remove steric clashes and generate a low energy conformer of the protein in the 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 Epik ionizer, 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 20 Å for 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 (optimized potentials for liquid simulations)(Shivakumar et al., 2010) was used for docking with all default parameters. Two protein-ligand complexes taking the 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 Born Surface Area (MM/GBSA) based binding free energy (ΔG_{bind}) was computed for the complexes and MD simulations.

2.3. Molecular dynamics simulations

MD simulations were implemented 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) (Bowers, Chow 2006) of Schrodinger. The OPLS_2005 force field was employed for the protein-ligand complexes. All the three model systems were first embedded inside a predefined DPPC membrane (at 325 K) 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 the Nose-Hoover Chain thermostat method for 5 ns and 5 ns of pressure relaxation with the Martyna-Tobias-Klein barostat method. The production run was performed in two phases. Initially, production run was carried out for 30 ns under NPT ensemble on each system using a cut-off distance of 12 Å for non-bonded interactions. Coordinates were saved at every 30 ps to generate trajectories of 1000 frames each. MM/GBSA binding energies (Kumari et al., 2014) 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).

Each of these four complexes were subjected to final phase MD simulations (in DPPC membrane, with similar parameters as described above) for 100 ns, to confirm the stable binding of CDC/UDC with SARS-Cov2-E. 100 ns 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 theSARS-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 and simulation interaction diagrams of Desmond.

3. Results and discussion

3.1. Comparative modelling of SARS-Cov2-E pentameric structure

The 3 D structure of SARS-Cov2-E was modelled using the SWISS-MODEL server. The SWISS-MODEL server initially identified eight templates forSARS-Cov2-E upon submitting its FASTA sequence. The SARS coronavirus Envelope protein (2MM4.pdb) was selected as the template for homology modelling of SARS-Cov2-E.Quality of the modelled monomeric structure was assessed by SWISS-MODEL itself and with other popularly used protein structure quality prediction programs. The QMEAN value estimates the global structural quality of the model. QMEAN has four individual terms or Z-scores representing the interaction potential between i) $C\beta$ atoms only, ii)all atoms, iii) the solvation potential and iv) 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 -3.2, indicating its acceptability. 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 model was also submitted to the ProSA for overall quality assessment. The model returned a Z-score of -0.01, which falls well within the Z-score ranges of known protein structures of the same size. The structure was further submitted to the PROCHECK server to obtain the Ramachandran plot for 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. Structure refinement and ERRAT analysis were iteratively done till we obtained a good quality final structure with an overall quality factor of 94. Figure S2 shows the quality assessment results for the model before and after refinement. This monomeric structure was subjected to symmetric docking using the SymmDock server to perform a geometry-based docking to generate a pentameric assembly of SARS-Cov2-E. This server returned 20 best pentameric structures, each of them was superposed on the pentameric structure of the homologous envelope protein of SARS-Cov (5 × 29.pdb) and the root mean squared deviations (RMSD)ranged from 3.44 to 4.77 Å. The pentameric form with the lowest RMSD (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 transmembrane 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 was below 1.5 Å and almost all of them showed H-bond with T30 of Chain A. Figure S3 shows the top scoring binding pose of both CDC and UDC with SARS-Cov2-E. Both of them make H-bonds 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 an H-bond donor while in case of UDC, the atom O26 (-OH group of the fragment substituent at the 6th carbon position) acts as the H-bond donor. 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 aTIP3P water box. Figure S4 shows the initial set up of the model systems. Most MD methods are known to be based on heuristic
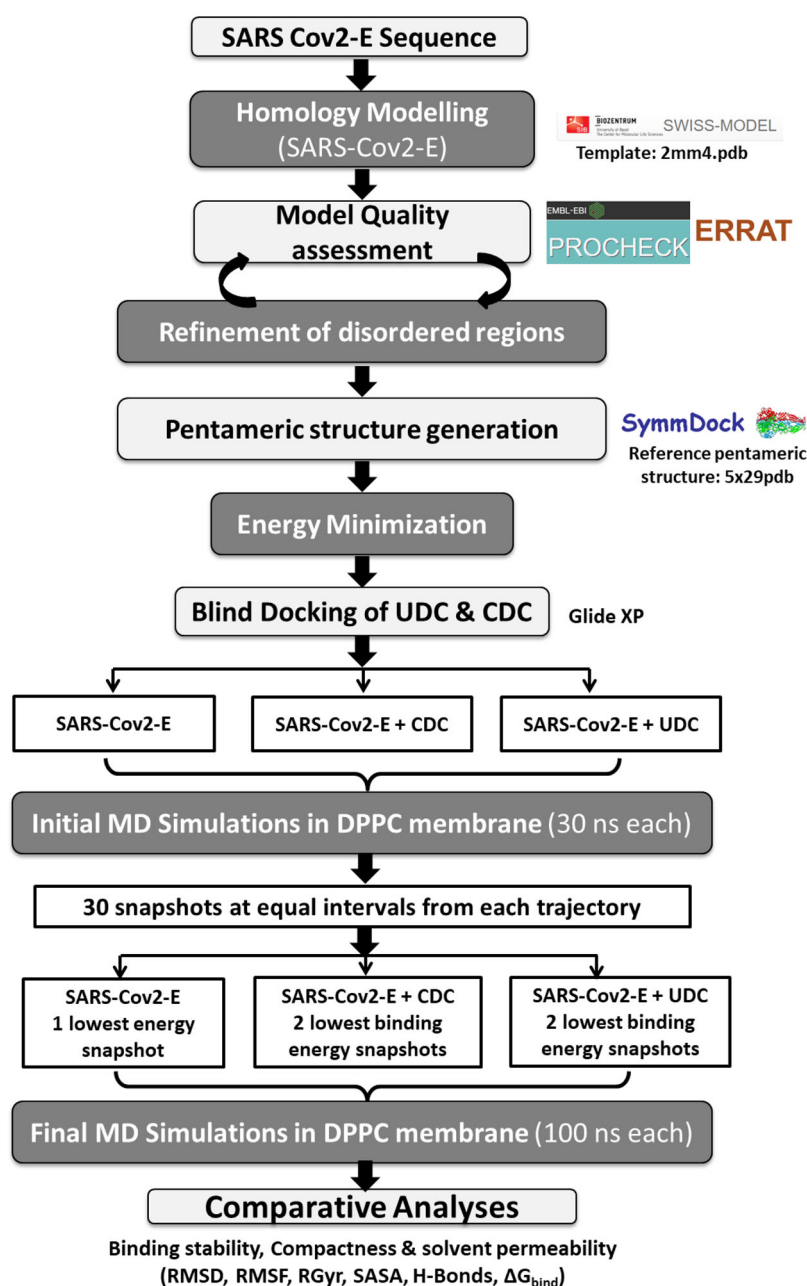
Figure 1. Overall Workflow of the study.

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 100 ns 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 30 ns simulations

Figure S5 **(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 the next 25 ns. However, the systems SARS-Cov2-E + CDC and SARS-Cov2-E + UDC were stabilized at about 4–5 Å, while
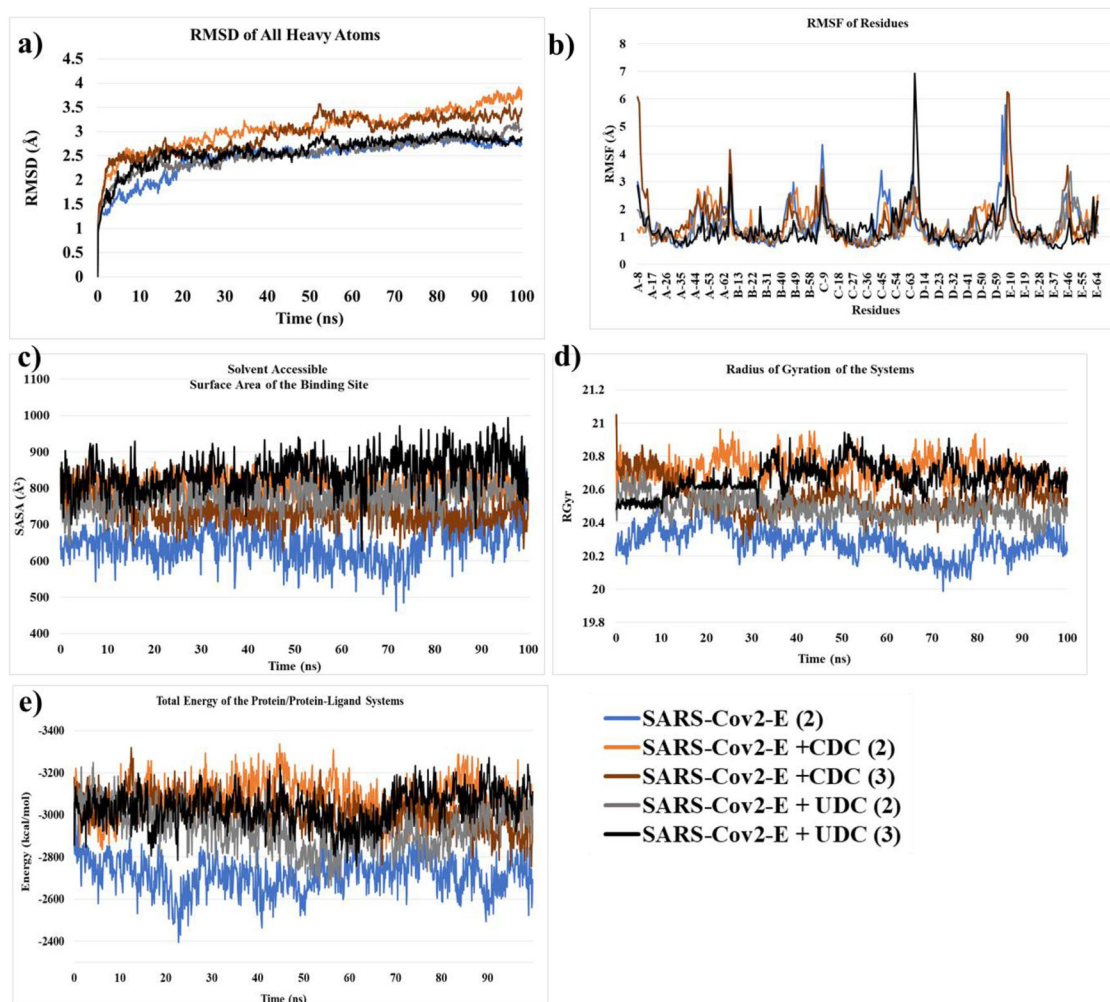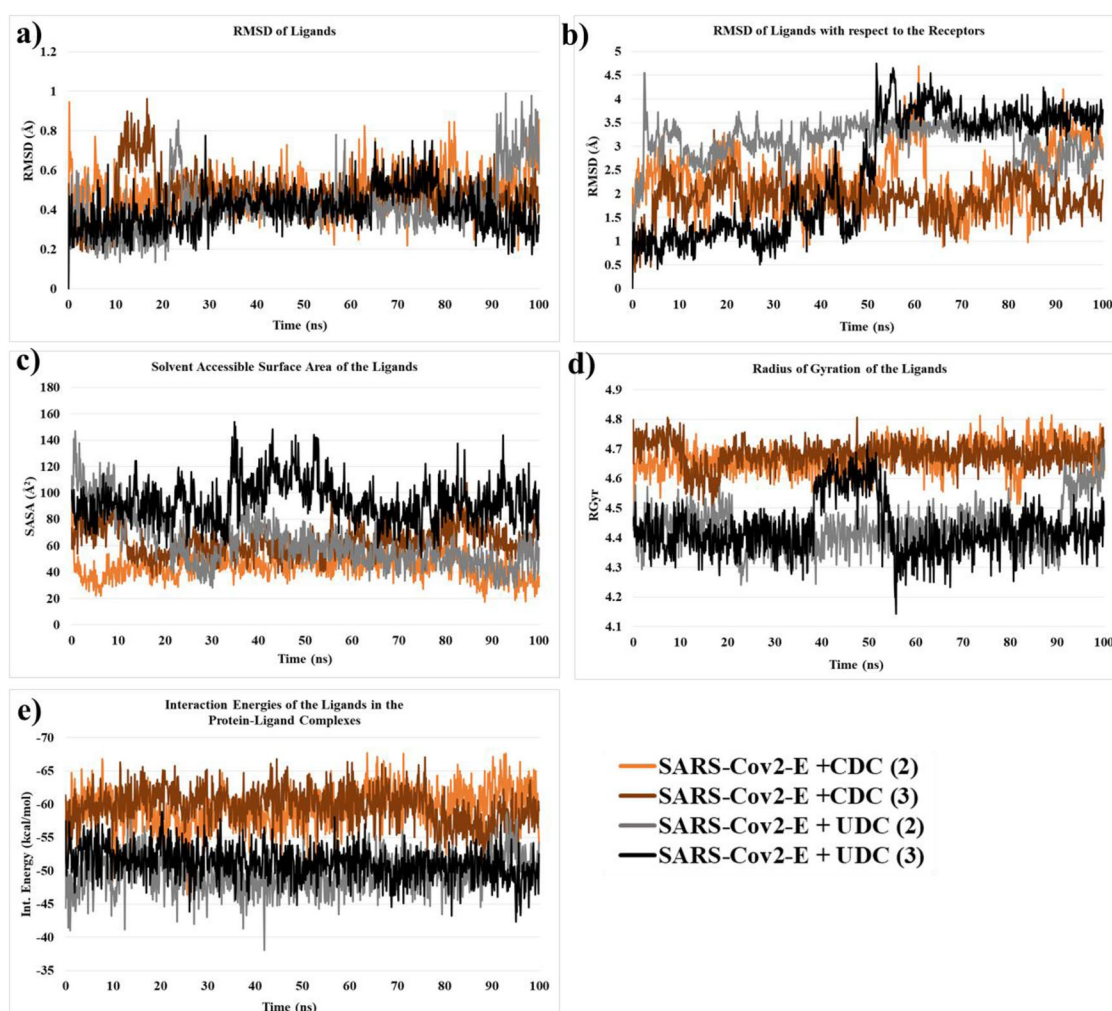
Figure 2. Variations of different structural properties of the protein-ligand complexes in different model systems during final simulations.

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 S5c 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 the presence of ligand contacts 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 S6). 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 (RGyr) is used as a measure of the compactness of a system. In Figure S5d, the RGyr profile of SARS-Cov2-E was 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 initial structures (Figure S5f) 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 S5e) were as high as the RMSD of the overall protein structures (Figures S5a, S5b). However, the small differences in the RMSDs of the ligands with respect to themselves and with respect to the protein indicate 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 RGyr and solvent accessible surface area (SASA) of CDC (Figures S5g and S5h) 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

Figure 3. Variations in structural and enthalpic properties of ligands in all the model systems during the final MD run.

trajectories. This enabled us to evaluate the structural evolutions of the system from different starting points with a robust sampling of 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 3 Å 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. Figures2c and 2d 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 2e).

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 were 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 3).

In order to understand the degrees of correlated movements between the monomers of SARS-Cov2-E in the presence and absence of bound CDC/UDC, the complexes were analysed by calculating the covariance matrices shown in Figure 4 (Roy & Post, 2012).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 & 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 anti-correlated movements of the corresponding residues and the orange/pink ones the correlated movements.

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 present due to
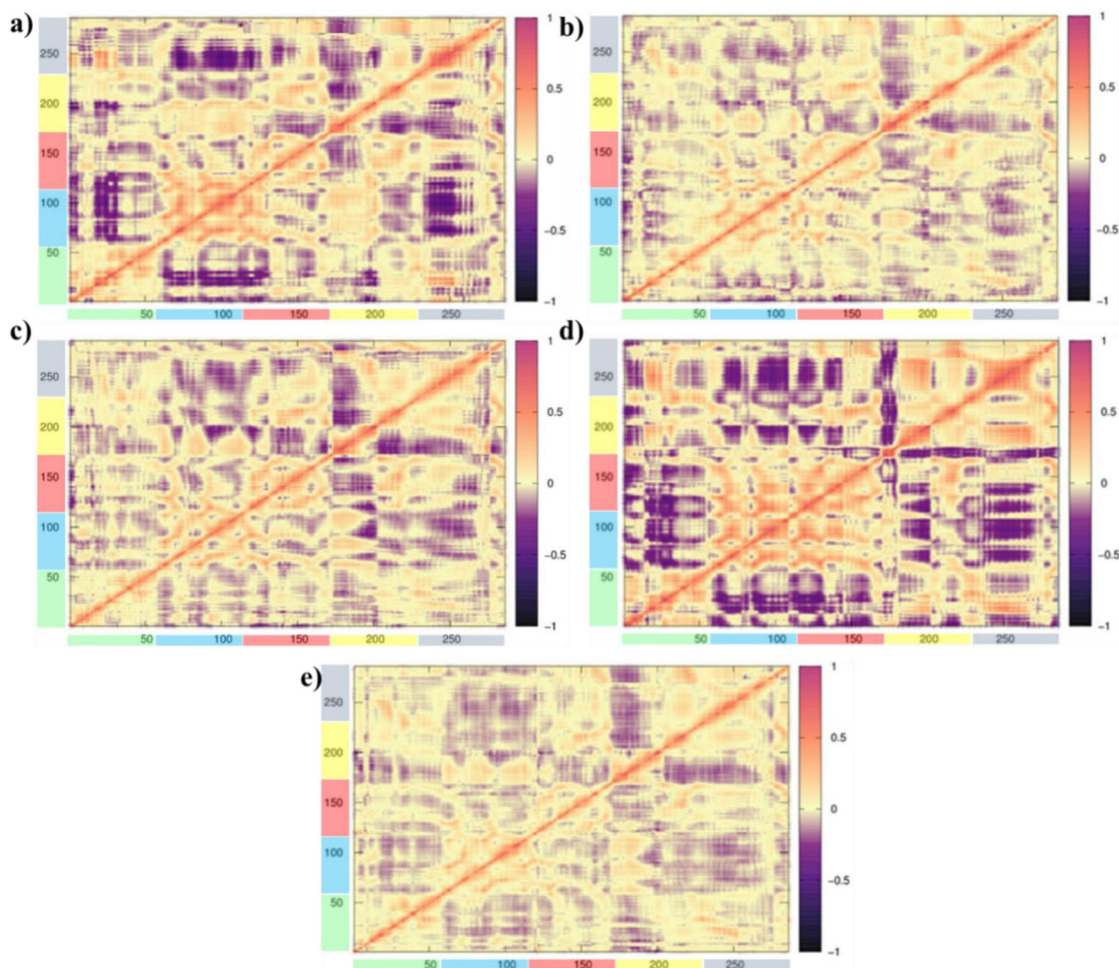
Figure 4. Covariance matrices of the residues of systems a) SARS-Cov2-E + CDC (2), b) SARS-Cov2-E + CDC (3), c) SARS-Cov2-E + UDC (2), d) SARS-Cov2-E + UDC (3), e) SARS-Cov2-E (2). The residue numbers are plotted in both the axes and the Chains are denoted as different colours.

passage of high number of water molecules inside the trans-membrane region in the CDC and UDC bound systems.
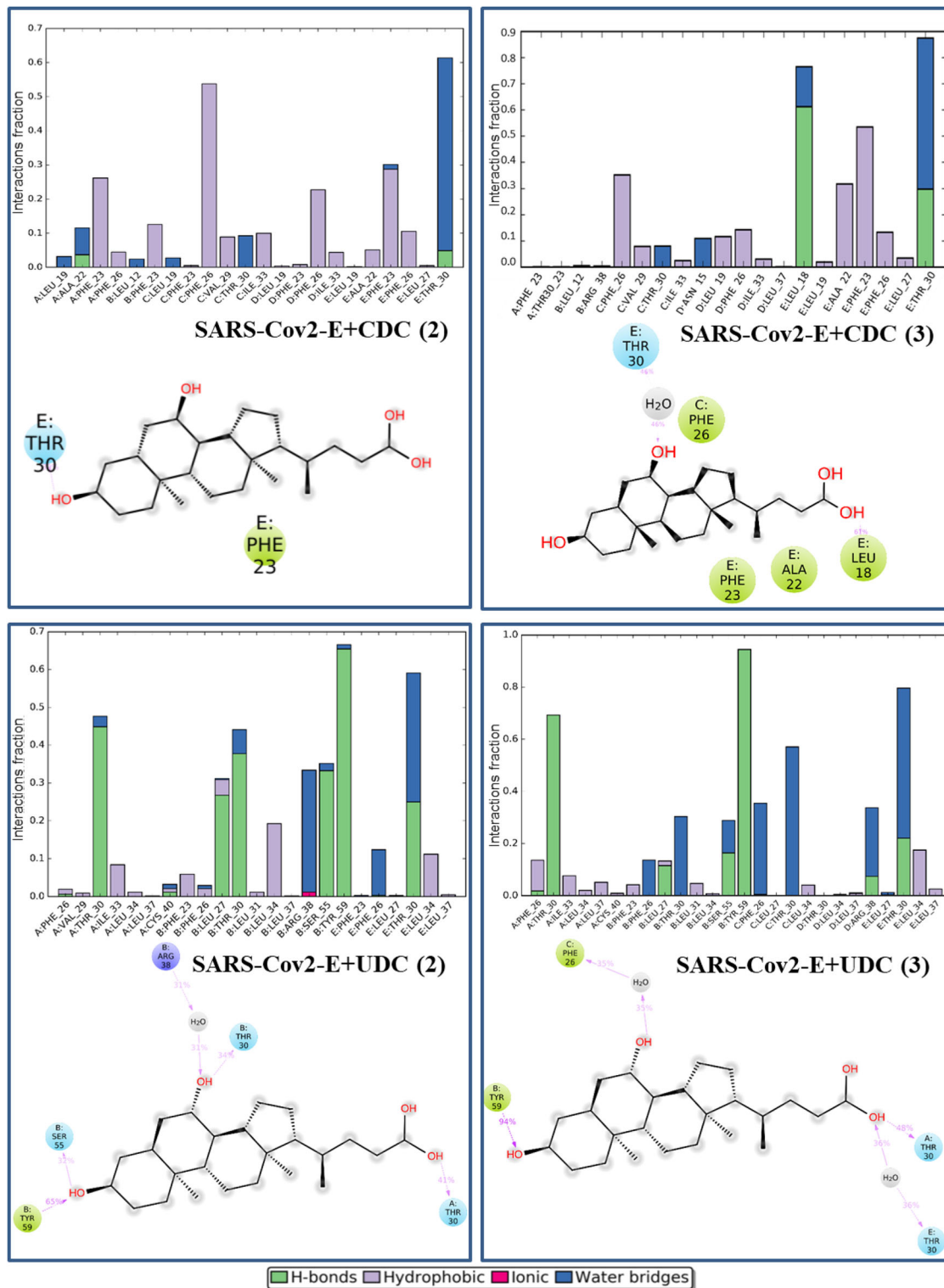
3.5. Differential binding of CDC and UDC with SARS-Cov2-E

CDC and UDC have the 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 S7 (a-b) show the interactions of CDC and UDC with SARS-Cov2-E during the initial 30 ns 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 S7c shows MM/GBSA energy of CDC and UDC binding throughout the MD simulation, calculated from snapshots saved at every 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 Figure S7c). Figure5 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. 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 5, all the atoms of both 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 40–60kcal/mol.

Figure 5. 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.

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 (Figure S8). Figure 6a shows the average number of H-bond formed between the constituent chains of SARS-Cov2-E in the presence and absence of the ligands CDC and UDC. It can be observed from Figure 6a, that in the native SARS-Cov2-E (2) pentameric structure an average of 3–6 H-bonds are formed between the adjacent chains, such as, between A
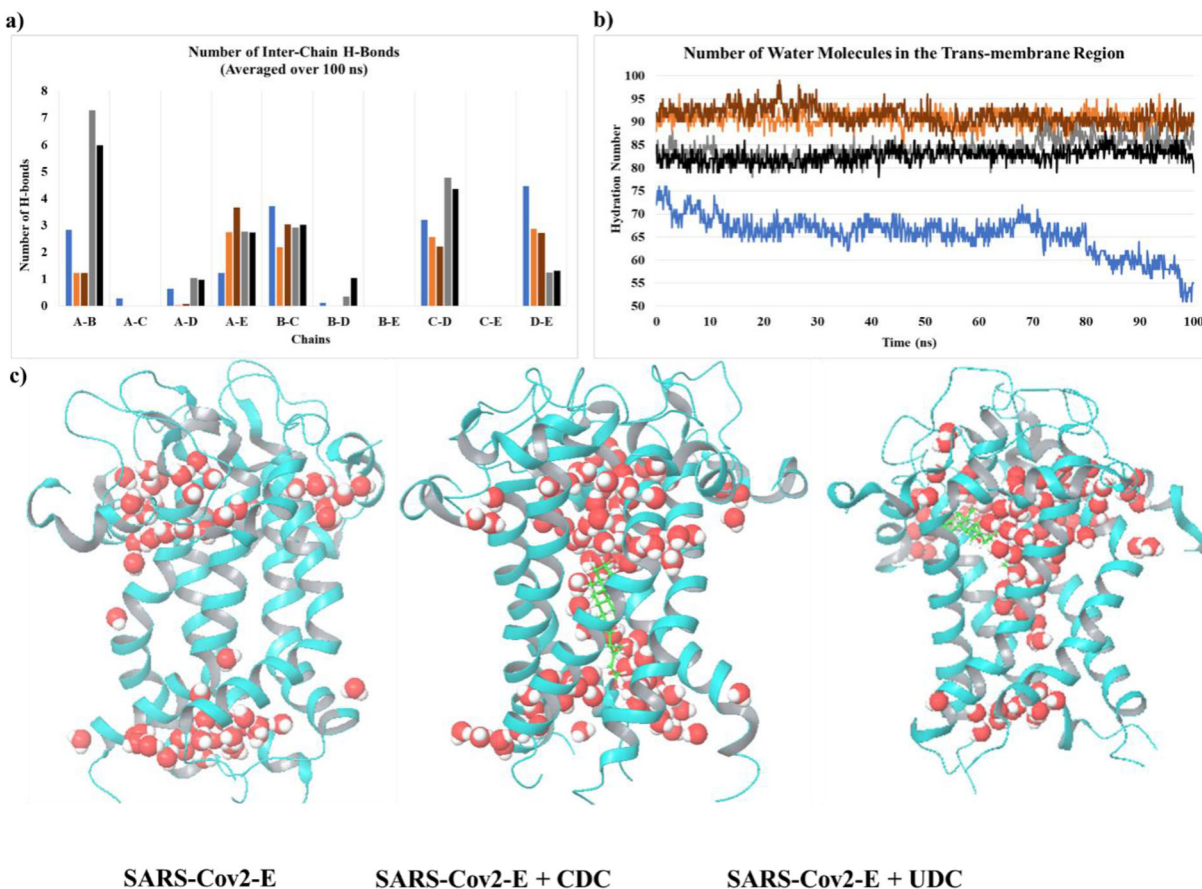
Figure 6. a) Average number of inter-chain hydrogen bonds calculated from MD results in all the three complexes b) Total number of water molecules inside the transmembrane channel c) graphical representation of the water molecules inside the transmembrane channel.

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 of 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 higher RGyr profiles (Figure 2d) 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 permeability 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 6b shows the total number of water molecules inside the transmembrane channel throughout the final simulations, calculated from the MD trajectories of 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 6c gives a graphical representation of the 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 the 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 for 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 interchain 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 the entrance of polar inhibitors inside the viral cells.

The pharmacokinetics profile and the druglikeness of CDC/UDC were predicted using the SwissADME server(Daina et al., 2017). Both the drugs pass all the duglikeness filters such as Lipinski, Ghose, Veber, Egan and Muegge filters. These drugs also showed good bioavailability scores along with high GI absorption (Table S1). The amphiphilic nature of CDC and UDC helps them in easy membrane permeation. They are suitable for adsorption in the hydrophobic membrane environments by virtue of their high lipophilicity, while the hydrophilic functional groups facilitate the H-bond and salt bridge interactions with some of the polar transmembrane channel residues. These polar functional groups also attract water molecules into the channel.

4. 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-E 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 MD 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. Despite 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 interchain H-bonds between adjacent chains and loosen the overall structure of SARS-Cov2-E pentameric assembly and allow passage of large number of water molecules into the transmembrane region. In addition to this, the polar functional groups of CDC and UDC aid to attract a 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 virus. 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 (Pavlović & Goločorbin-Kon, 2018). Our observations are in close agreement with the membrane permeability enhancing properties of CDC and UDC as mentioned in the 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

Disclosure statement

The authors declare that there is no conflict of interest

References

Abdulrab, S., Al-Maweri, S., & Halboub, E. (2020). Ursodeoxycholic acid as a candidate therapeutic to alleviate and/or prevent COVID-19-associated cytokine storm. *Medical Hypotheses*, *143*, 109897. https://doi.org/10.1016/j.mehy.2020.109897

Bhardwaj, V., & Purohit, R. (2020). Computational investigation on effect of mutations in PCNA resulting in structural perturbations and inhibition of mismatch repair pathway. *J Biomol Struct Dyn*, *38*(7), 1963–1974. https://doi.org/10.1080/07391102.2019.1621210

Bhardwaj, V. K., Singh, R., Sharma, J., Das, P., & Purohit, R. (2020). Structural based study to identify new potential inhibitors for dual specificity tyrosine-phosphorylation-regulated kinase. *Comput Methods Programs Biomed*, *194*, 105494. https://doi.org/10.1016/j.cmpb.2020.105494

Bhardwaj, V., Singh, R., Singh, P., Purohit, R., & Kumar, S. (2020). Elimination of bitter-off taste of stevioside through structure modification and computational interventions. *Journal of Theoretical Biology*, *486*, 110094. https://doi.org/10.1016/j.jtbi.2019.110094

Bowers, K. J., Chow, D. E. (2006). Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*.

Cai, Q., Yang, M., Liu, D., Chen, J., Shu, D., Xia, J., Liao, X., Gu, Y., Cai, Q., Yang, Y., Shen, C., Li, X., Peng, L., Huang, D., Zhang, J., Zhang, S., Wang, F., Liu, J., Chen, L., … Liu, L. (2020). Experimental treatment with favipiravir for COVID-19: An open-label control study. *Engineering*, *6*(10), 1192–1198. https://doi.org/10.1016/j.eng.2020.03.007

Caly, L., Druce, J. D., Catton, M. G., Jans, D. A., & Wagstaff, K. M. (2020). The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research*, *178*(104787), 104787. https://doi.org/10.1016/j.antiviral.2020.104787

Carino, A., & Moraca, F. (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., & Rajnik, M. (2020). Features, Evaluation, and Treatment of Coronavirus.

Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, *92*(4), 418–423. https://doi.org/10.1002/jmv.25681

Choudhury, C. (2020). Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease. *Journal of Biomolecular Structure and Dynamics*, *1*, 1–14.

Choudhury, C., Deva Priyakumar, U., & Narahari Sastry, G. (2016). Structural and functional diversities of the hexadecahydro-1H-cyclopenta[a]phenanthrene framework, a ubiquitous scaffold in steroidal hormones. *Molecular Informatics*, *35*(3–4), 145–157. https://doi.org/10.1002/minf.201600005

Choudhury, C., Priyakumar, U. D., & Sastry, G. N. (2015). Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. *Journal of Chemical Information and Modeling*, *55*(4), 848–860. https://doi.org/10.1021/ci500737b

Choudhury, C., Priyakumar, U. D., & Sastry, G. N. (2016). Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors. *Journal of Chemical Sciences*, *128*(5), 719–732. https://doi.org/10.1007/s12039-016-1069-1

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science : a Publication of the Protein Society*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Corse, E., & Machamer, C. E. (2003). The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology*, *312*(1), 25–34. https://doi.org/10.1016/S0042-6822(03)00175-2

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1), 42717. https://doi.org/10.1038/srep42717

Hirota, I., Chijiiwa, K., Noshiro, H., & Nakayama, F. (1992). Effect of chenodeoxycholate and ursodeoxycholate on nucleation time in human gallbladder bile. *Gastroenterology*, *102*(5), 1668–1674. https://doi.org/10.1016/0016-5085(92)91728-M

Khailany, R. A., Safdar, M., & Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, *19*(100682), 100682. https://doi.org/10.1016/j.genrep.2020.100682

Kumar, Y., & Yadav, R. (2020). Can natural detergent properties of bile acids be used beneficially in tackling coronavirus disease-19? *Future Virology*. https://doi.org/10.2217/fvl-2020-0210

Kumari, R., Kumar, R., & Lynn, A., Open Source Drug Discovery Consortium (2014). g_mmpbsa-a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Kurumurthy, C., & Rao, P. (2011). A facile and single pot strategy for the synthesis of novel naphthyridine derivatives under microwave irradiation conditions using ZnCl 2 as catalyst, evaluation of AChE inhibitory activity, and molecular modeling studies. *Medicinal Chemistry Research - MED CHEM RES*, *21*, 1785-1795.

Laskowski, R. A., & Chistyakov, V. V. (2005). PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Research*, *33*(Database issue), D266-8.

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, *61*(4), 704–721. https://doi.org/10.1002/prot.20660

Ninomiya, R., Matsuoka, K., & Moroi, Y. (2003). Micelle formation of sodium chenodeoxycholate and solubilization into the micelles: Comparison with other unconjugated bile salts. *Biochimica et Biophysica Acta.*, *1634*(3), 116–125. https://doi.org/10.1016/j.bbalip.2003.09.003

Pavlović, N., & Goločorbin-Kon, S. (2018). Bile acids and their derivatives as potential modifiers of drug release and pharmacokinetic profiles. *Frontiers in Pharmacology*, *9*, 1283-1283.

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2019). Drug repurposing: Progress, challenges and recommendations. *Nature Reviews. Drug Discovery*, *18*(1), 41–58. https://doi.org/10.1038/nrd.2018.168

Roda, A., Cerré, C., Fini, A., Sipahi, A. M., & Baraldini, M. (1995). Experimental evaluation of a model for predicting micellar composition and concentration of monomeric species in bile salt binary mixtures. *Journal of Pharmaceutical Sciences*, *84*(5), 593–598. https://doi.org/10.1002/jps.2600840514

Roy, A., & Post, C. B. (2012). Detection of long-range concerted motions in protein by a distance covariance. *Journal of Chemical Theory and Computation*, *8*(9), 3009–3014. https://doi.org/10.1021/ct300565f

Ruch, T. R., & Machamer, C. E. (2011). The hydrophobic domain of infectious bronchitis virus E protein alters the host secretory pathway and is important for release of infectious virus. *Journal of Virology*, *85*(2), 675–685. https://doi.org/10.1128/JVI.01570-10

Schneidman-Duhovny, D., & Inbar, Y. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research.*, *33*(Web Server issue), W363-7.

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., & Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *Journal of Chemical Theory and Computation*, *6*(5), 1509–1519. https://doi.org/10.1021/ct900587b

Singh, R., Bhardwaj, V., Das, P., & Purohit, R. (2020). Natural analogues inhibiting selective cyclin-dependent kinase protein isoforms: A computational perspective. *Journal of Biomolecular Structure & Dynamics*, *38*(17), 5126–5135. https://doi.org/10.1080/07391102.2019.1696709

Subramanian, S., Iles, T., Ikramuddin, S., & Steer, C. J. (2020). Merit of an ursodeoxycholic acid clinical trial in COVID-19 patients. *Vaccines*, *8*(2), 320. https://doi.org/10.3390/vaccines8020320

Teoh, K.-T., Siu, Y.-L., Chan, W.-L., Schlüter, M. A., Liu, C.-J., Peiris, J. S. M., Bruzzone, R., Margolis, B., & Nal, B. (2010). The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Molecular Biology of the Cell*, *21*(22), 3838–3852. https://doi.org/10.1091/mbc.E10-04-0338

Wang, D.-Y., Jiang, Z., Ben-David, Y., Woodgett, J. R., & Zacksenhaus, E. (2019). Molecular stratification within triple-negative breast cancer subtypes. *Scientific Reports*, *9*(1), 19107. https://doi.org/10.1038/s41598-019-55710-w

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290

Wilson, L., McKinlay, C., Gage, P., & Ewart, G. (2004). SARS coronavirus E protein forms cation-selective ion channels. *Virology*, *330*(1), 322–331. https://doi.org/10.1016/j.virol.2004.09.033

Yao, X., Ye, F., Zhang, M., Cui, C., Huang, B., Niu, P., Liu, X., Zhao, L., Dong, E., Song, C., Zhan, S., Lu, R., Li, H., Tan, W., & Liu, D. (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). *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *71*(15), 732–739. https://doi.org/10.1093/cid/ciaa237