Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Molecular Graphics and Modelling 109 (2021) 108021

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

Investigation of structural analogs of hydroxychloroquine for SARS-CoV-2 main protease (Mpro): A computational drug discovery study

Saima Reyaz [a,1], Alvea Tasneem [a,1], Gyan Prakash Rai [a,1], Hridoy R. Bairagya [b,*]

[a] Department of Computer Science, Jamia Millia Islamia, New Delhi, 110025, India
[b] Department of Biophysics, All India Institute of Medical Sciences, New Delhi, 110029, India

ARTICLE INFO

Keywords:
Hydroxychloroquine (HCQ)
Structural analogs of HCQ
Computational binding study
MD Simulation
Water displacement method
Main protease
nCOVID-19

ABSTRACT

The main protease (Mpro) is the key enzyme of nCOVID-19 and plays a decisive role that makes it an attractive drug target. Multiple analysis of crystal structures reveals the presence of W1, W2, and W3 water locations in the active site pocket of Mpro; W1 and W2 are unstable and are weakly bonded with protein in comparison to W3 of Mpro-native. So, we adopt the water displacement method to occupy W1 or W2 sites by triggering HCQ or its analogs to inactivate the enzyme. Virtual screening is employed to find out best analogs of HCQ, molecular docking is used for water displacement from catalytic region of Mpro, and finally, MD simulations are conducted for validation of these findings. The docking study reveals that W1 and W2 are occupied by respective atoms of ZINC28706440 whereas W2 by HCQ and indacaterol. Finally, MD results demonstrate (i) HCQ occupies W1 and W2 positions, but its analogs (indacaterol and ZINC28706440) are inadequate to retain either W1 or W2 (ii) His41 and Asp187 are stabilized by W3 in Mpro-native and His41, Cys145 and HCQ by W7 in ZINC28706440, and W4, W5, and W6 make water mediated bridge between indacaterol with His41. The structural, dynamical, and thermodynamic (WFP and J value) profiling parameters suggest that W3, W4, and W7 are prominent in their corresponding positions in comparison with W5 and W6. The final results conclude that ZINC28706440 may act as a best analog of HCQ with acceptable physico-chemical and toxicological scores and may further be synthesized for experimental validation.

1. Introduction

The World Health Organization (WHO) declared on March 11th, 2020, that the outbreak of "Novel CoronaVirus Disease 2019" (nCOVID-19) had become a pandemic and was one of the highly pathogenic that infected humans [1]. Since the onset of the pandemic, there were more than 200 million people confirmed nCOVID-19 cases globally, leading to over more than 4 million deaths were reported to WHO [2]. A single or combined test that provides virological and serological output would be ideal to identify individuals who had been infected. The novel coronavirus had a rapid spread to most of the countries till date and caused deadly deaths with a 2% mortality rate. Risk factors for mortality include advanced age, obesity, diabetes, and hypertension. However, there are no clinically approved or specific therapeutic drugs that are available for nCOVID-19 [3]. Intensive research on SARS-CoV-2 is urgently required for the treatment of nCOVID-19. The test, trace, and treat (T3) programs will become mandatory and may act as effective drugs for safe therapies [4].

Vaccines might not be enough to quell the pandemic, although many repurposed drug candidates are being evaluated, many are redundant, and lack a strong rationale for clinical development [5]. Earlier in the course of the disease, due to lack of specific treatment for nCOVID-19, pre-clinical investigation explored the usage of hydroxychloroquine (HCQ), an anti-malarial drug, against nCOVID-19. Many clinical reports and research articles have explored the effectiveness of HCQ for the treatment of nCOVID-19 [6–9]. However, HCQ is a well-known to have some complications and side-effects in some cases. The cause for a failure of HCQ treatment should be investigated by testing the isolated

Abbreviations: SARS, severe acute respiratory syndrome; CoV, coronavirus; nCOVID-19, novel coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; HCQ, hydroxychloroquine; Mpro, main protease; 3CLpro, 3C-like protease; HTS, high-throughput screening; WFP, water finding probability.

* Corresponding author. Department of Biophysics, All India Institute of Medical Sciences, New Delhi, India.
E-mail address: hbairagya@gmail.com (H.R. Bairagya).

[1] Authors (S.R., A.T., and G.P.R.) contributed equally to this manuscript and shared first authorship.

https://doi.org/10.1016/j.jmgm.2021.108021
Received 12 April 2021; Received in revised form 18 August 2021; Accepted 2 September 2021
Available online 9 September 2021
1093-3263/

SARS-CoV-2 strains of the non-respondents and analyzing their genome that may be associated with the metabolism of HCQ. Medicine like remdesivir, HCQ, lopinavir, and interferon regimens had a small or no effect on hospitalized patients with nCOVID-19, as recorded by overall mortality, initiation of ventilation, and duration of hospital stay [10]. The nCOVID-19 Treatment Guidelines Panel recommended against the use of chloroquine or HCQ with or without azithromycin for the treatment of nCOVID-19 in hospitalized and nonhospitalized patients. Furthermore, the Panel also advised against the use of high-dose chloroquine (600 mg twice daily for 10 days) for the treatment of nCOVID-19. However, despite demonstrating antiviral activity in some in vitro systems, HCQ with or without azithromycin did not reduce upper or lower respiratory tract viral loads or demonstrate clinical efficacy in a rhesus macaque model [11]. The trial enrollment ended early on 5th June 2020, after an independent data-monitoring committee recommended reviewing the unblinded data, and the investigators and trial-steering committee also concluded that the data showed no beneficial effect of hydroxychloroquine [12,13].

According to the phylogenetic and genomic information, the coronaviruses are genotypically categorized into four genera: α-CoV, β- CoV, γ- CoV, and δ-CoV. Furthermore, its β-type is subdivided into four viral lineages of A, B, C, and D. However, CoV has a substantial single-stranded RNA virus genome (27–32 kb) [14] and its virion contains two major components: genomic RNA and a protein capsid. The RNA genome of n-Cov-19 has 29891 nucleotides in length that encode 9860 amino acids. Moreover, its genome contains the following items: two flanking untranslated regions (UTRs), a single open reading frame (7096-aa), a non-structural polyprotein (7096-aa), four structural proteins – Spike (S) (1273-aa), Envelope (E) (75-aa), Membrane (M) (222-aa), Nucleocapsid (N) (419-aa), and five accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10) [15].

The main protease (Mpro) or 3C-like protease (3CLpro) is another key enzyme in the life cycle of nCOVID-19, playing a pivotal role in mediating viral replication, and transcription that makes it an attractive drug target for this virus [16]. Mpro resembles the structure of cysteine proteases with two antiparallel β-barrels domains [17]. The active site of this enzyme comprises a catalytic dyad His41 and Cys145, and lacks the third catalytic residue, which is replaced by a stable water molecule [18] (Fig. S1). This Cys-His catalytic dyad of this enzyme is located in a cleft between the domains I and II. The N-terminal residues (res. ID. from 1 to 7) of Mpro are considered to play an eminent role in the proteolytic activity, whereas the –C terminal (domain III) is reported to be required for dimerization [19]. Crystal or NMR structures for Mpro protein with HCQ or its analogs are not yet available because the determination of such complex structures may be difficult. In the face of global economic recession, the developments of infrastructure for such experimental trails are really challenging for most of the researchers due to high cost. Thus, the approach adopted in our investigation is to search structural analogs of HCQ for possible inhibition of SARS-CoV-2. The functional differences between HCQ and its analogs may translate into performance variances that could potentially influence the decision process in either HCQ or its analogs and can ultimately become a viable therapeutic agent for nCOVID-19.

As we know, water molecules play a vital role in protein structure, enzyme catalysis, protein architecture, conformational stability, and protein plasticity, stabilization of salt bridges, ligand binding, and selectivity for specific interactions [20–27]. However, drug repurposing has been a strategy adopted by several research groups to provide effective treatment in a short period. Therefore, few computational studies have been reported including a massive virtual screening, molecular docking, followed by molecular dynamics (MD) simulation for obtaining new anti-COVID molecules. The static structural data do not explain the functional and dynamical nature of water molecules in the Mpro protein, changes of geometrical and electronic variations at the ligand binding site after binding HCQ. In this context, it is important to trace the stepwise movement of the biochemical mechanism of Mpro from native to ligand bound state, its dynamics, and conformational changes using MD simulation. During the MD simulations, the movements of ligands in respect of active site residue could explain interesting dynamical phenomena that are generally inaccessible in crystal structure of Mpro.

The estimation of binding free energy for the ligand may play a decisive role to achieve experimentally determined value for target proteins, but accurate binding energy calculation of ligand remains a formidable challenge. In structure-based drug design, the free energy perturbation and thermodynamic integration methods are not commonly used to compute binding energy calculation of ligand because these methods are implemented numerically, require sufficient statistical sampling, are time-consuming, and are computationally expensive [28]. Hence, linear interaction energy (LIE) and molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA) are the alternative and popular computational methods that are broadly used in structure-based drug discovery to calculate the binding energy of ligand in a fast and rational way [29]. In this study, we can therefore evaluate and compare the performance of LIE-D and MM/PBSA methods with respect to their ability and efficacy. In the current work, we investigated whether the LIE-D can achieve similar accuracy compared to MM/PBSA in calculating binding affinities of HCQ and its analogs? These approaches are efficient, less time consuming, and are also computationally inexpensive than other methods mentioned above.

Numerous computational methods like molecular dynamics simulations with inhomogeneous fluid solvation theory [30] allow us to identify the specific geometrical and electronic locations of invariant water molecules in the active site of the protein and also compute their thermodynamic properties [31]. Multiple analyses of crystal structures of the native form of Mpro protein provide suggestions into which water molecules can be targeted for displacement. The medicinal chemistry methods explain that tightly bound and well-ordered water molecules may gain entropy and are difficult to displace them by ligands, whereas weakly bound water molecules are mostly favorable for replacement by ligands that may expect to increase binding affinities [32]. To the best of our knowledge this is the first computational study that explores the results of molecular docking on the basis of crystal water displacement in the active site pocket of Mpro and these data are validated by MD analysis by focusing on conformational dynamics of HCQ and its analogs. Moreover, the integrated findings of single trajectory simulations are the starting structures which were obtained from molecular docking study from specific binding poses of HCQ or its analogs. No MD simulation study has yet been reported on structural transition of Mpro from its native to HCQ or its analogs bound state to capture the conformational transitions. Our computational techniques act as a complementary approach to synthesis of new compounds or repurposing of drugs. The vision of present study is to evaluate the inhibitors that are structurally similar to HCQ, and it will be compared to Mpro protein in the complex. In this aspect, our computational study provides a testable hypothesis for the binding mechanism of HCQ or its structurally similar analogs that may be synthesized experimentally for further validation and also guide us to solve the crystal structure of the Mpro with HCQ bound complex.

2. Materials and methods

2.1. Structures collection and identification of conserved water molecules at catalytic region of Mpro

In the present study, the eleven (PDB ID: 6YB7, 6Y84, 7ALH, 7ALI, 7BRO, 7JR3, 7KPH, 7C2Y, 7CWB, 7JUN, and 7JVZ) ligand-free crystal structures of the main protease were obtained from the Protein Data Bank [33]. The PDB ID: 6YB7 (resolution 1.25 Å) was taken as a template for further molecular docking and MD simulation study because it adopted an open conformation to accommodate ligands in the active site (Table S1). Three conserved water molecules (W1, W2, and W3) were

identified at the catalytic region of these ligand-free crystal structures using the 3dSS program [34]; besides, the conserved positions of these water molecules were also verified by the Swiss PDB Viewer program [35]. Two water molecules (oxygen atoms) or two atoms (oxygen atoms between two water molecules or oxygen atom of water molecule and specific atom of ligand) whose center-to-center distance was within 1.80 Å, [36], in between reference (PDB ID: 6YB7), and movable structures were assigned as conserved [37,38]. The W1 is H-bonded with ND1 of His41 in 6YB7, 6Y84, 7ALH, 7ALI, 7BRO, 7JR3, and 7KPH crystal structures, W2 is also H-bonded only with W1, and W3 makes H-bonded with His41 and Asp187 in 7C2Y, 7CWB, 7JUN, and 7JVZ; therefore, they were assigned as conserved water positions at the catalytic region of native Mpro protein. The solvent-accessible surface area (SASA) of W1, W2, and W3 water molecules were measured using the in-house TCL script with a probe radius of 1.4 Å. The normalized B factor (BNorm) for all conserved water molecules were calculated using the formula, BNorm = (Bi –)/σ(B), where Bi is the B factor of each atom of protein, is the mean B factor of the protein molecule, and σ(B) is the standard deviation of the B factors [39].

2.2. Virtual screening and QSAR property analysis

The SMILES [CCN(CCO)CCCC(C)NC1=C2C=CC(Cl)=CC2=NC=C1] of HCQ was obtained from the Drug-Bank database (v5.1.5) [40] and it was used for high throughput screening to investigate structural analogs of HCQ by Swiss Similarity program [41]. Four chemical libraries (Ligands from PDB, ChEMBL, Zinc-drug like, and Zinc-lead like) were chosen for the high throughput screening (HTS) to achieve the best structural analogs of HCQ. After HTS, the fifty structural analogs of HCQ are observed with reasonably good similarity scores like HCQ and finally, the best 14 HCQ analogs (mentioned in the supplementary section of Fig. S2) were obtained on the basis of similarity scores for further molecular docking study with 6YB7 crystal.

The Osiris property explorer [42] and Swiss-ADME [43] programs have been used to compare the pharmacokinetics and drug-likeness scores between the HCQ molecule and its analogs (14 compounds). Each molecule was filtered out on the basis of six molecular properties (cLogP, solubility, molecular weight, TPSA, drug-likeness, and drug-score) from the Osiris program and four characters (molar refractivity, Lipinski, bioavailability score, and synthetic accessibility) from Swiss-ADME program.

2.3. Molecular docking study

Displacing the W1 (PDB ID: W639) and W2 (PDB ID: W864) positions by HCQ or its analogs from the 6YB7 crystal structure is the main aim of the present docking study. The W1 and W2 locations are weakly bound and have less interaction energy with protein, hence they may displace well, in contrast the W3 which is tightly bound to protein with strong interaction energy and it seems to be unfavorable for water displacement. The “A” chain was obtained (excluding water molecules) from 6YB7 crystal structure and it was considered as a receptor. The HCQ molecule and its 14 structural analogs were also taken as a ligand for further molecular docking study using the AutoDock Vina program (v 1.1.2) [44]. The required PDBQT file of the receptor protein and 15 ligand molecules were assigned Kollman united atom charges [45] and Gasteiger partial atomic charges, respectively [46]. To obtain all the possible binding orientations, rigid and flexible dockings were performed not only for HCQ but also for its 14 analogs, since they should have a similar binding mode. Therefore, all the rotatable bonds of each ligand were kept to be flexible during docking, and residues inside the binding pocket were made rigid initially, and then flexible docking was performed by keeping the side-chains of His41 of the receptor protein made flexible. Grid point spacing was set at 1 Å and grid points 30 were taken in each direction of grid box and the concern box was centered at ND1 of His41 (in 6YB7). Best five binding modes of each ligand were identified after cluster analysis according to their binding affinity. The molecular docking results produced that the drug HCQ, indacaterol, and ZINC28706440 show a desirable result and these three docked structures were further consider for molecular dynamics simulation study.

2.4. Molecular dynamics simulations and trajectory analysis

The initial docked structure of HCQ, indacaterol, ZINC28706440, and native conformation of 6YB7 were obtained for further MD simulation studies to identify the dynamic properties of water molecules in the ligand-binding pocket of Mpro protein and also to decipher its structural changes from native to different ligand-bound conformations. Using the AutoPSF module of Visual Molecular Dynamics (VMD v.1.9.3.) program [47], missing hydrogen atoms were added to each complex structure. Furthermore, the ligand HCQ, indacaterol, and ZINC28706440 were parameterized using the SwissParam program [48] to obtain topology and parameter files for the CHARMM force field [49]. All these structures were then solvated in a cubic box of around 10,000 TIP3P water molecules extending at least 10 Å from the protein surface. Sodium and chloride ions were employed to neutralize the overall charge of the system; the resulting system consisted of around 33,000 atoms. MD simulation was performed for four solvated structures (HCQ, indacaterol, and ZINC28706440, and Mpro-native) using the Nanoscale Molecular Dynamics (NAMD v.2.11) program [50] by assigning the CHARMM-36 all-atom force field (with map correction) [51] for protein and the CHARMM General Force Field (CGenFF) for all ligands [52]. Three step minimizations were adopted to stabilize the system using the steepest-descent method. Initially, energy minimizations were performed on the solvent molecules for 5000 steps with the fixed solute. Two additional minimization processes were applied; restrained (CA atoms of protein) and unrestrained minimizations were carried out for 2000 and 1000 cycles, respectively. Subsequently, water molecules and ions were equilibrated for 2 ns by fixing solute to ensure the removal of any remaining steric clashes. Stepwise heating was carried out from 0 to 310 K for 50 ps in three steps at constant volume and with a gradually reduced restraint. Then each system was equilibrated for 500 ps under *NPT* restraints followed by *NVT* conditions reverting to the *NPT* ensemble for a period of 1 ns [53]. We ensured that each system was brought to equilibrium before continuing our simulations by verifying that each system reached a point where the energy fluctuations and RMSD were stable and PCA data are not scattered over a larger area [54].

To mimic physiological conditions, the temperature was kept at 310 K using Langevin dynamics with a damping coefficient of 5 ps^{-1}. The pressure was maintained at 1 atm using the Langevin piston Nose–Hoover method, with a piston period of 100 fs and a decay time of 50 fs. During the simulations, a Nosé-Hoover Langevin piston barostat [55] and Langevin thermostat [56] were used to enforce constant pressure and temperature. The SHAKE algorithm [57] was used to keep bonds involving H atoms at their equilibrium length, allowing a 2 fs time step. The van der Waals interactions were truncated at 12.0 Å with switching from 10.0 Å. Electrostatic interactions were modeled accordingly with a dielectric constant of 1.0 throughout the equilibration and production runs. The standard Particle–Mesh Ewald method was used with periodic boundary conditions to compute the long-range electrostatic interaction of the system by specifying an appropriate Particle–Mesh Ewald grid size. Finally, production runs of all-atom MD for each structure were performed in the *NPT* ensemble for 50 ns. The atomic coordinates were recorded at every 2 ps for further data analysis (25,000 frames). Each MD trajectory, especially 12,500 recorded snapshots (frame Id. 12,500 to 25,000) were considered for further data analysis using the VMD program. Average root mean square deviations (RMSD) and root mean square fluctuations (RMSF) of CA atoms of MD structures (12,500 frames) were calculated (for the reference structure) using an in house TCL script.

2.5. Identification of water sites in MD structures and calculation of their thermodynamic and structural dynamic properties

To investigate the identification of water sites (W) between each ligand and His41 residue in the catalytic region of native MD structure, several amino acids were identified, especially relevant, in determining the binding with those water molecules (Table 1). Water molecules (W3 to W7) were investigated from corresponding HCQ, indacaterol, and ZINC28706440 bound MD structures by the clustering algorithm of WATCLUST [58] program in VMD. The clustering parameters of wat-numbermin, dist, dr, WFRr, and pop were assigned as 15% of the total frame, 50.0, 0.16, 0.20, 0.60, and 0.90 Å, respectively. The visual analysis of the WATCLUST results provide us identification of each water site (W) that has maximum WFP values and is occupied with the H-bonding distance, either, between each ligand and their adjacent atoms of the protein or only respective atoms of protein. A water molecule is defined as being inside that corresponding water site if its oxygen atom distance to the water center is less than 0.6 Å, a value approximately corresponding to a volume of 1 $Å^3$ for the water site [59]. However, among the five water sites (W3 to W7) in MD structures, W3, W4, and W7 are found in both the X-ray and MD structures, whereas W1 and W2 were observed only in X-ray. Therefore, the remaining W5 and W6 were assigned a new numbering scheme to distinguish them from water molecules in X-ray structures.

To calculate the average potential energy, combine, with the interaction of water molecules inside the water site (W) with protein and the remaining solvent, electrostatic and van der Waals interactions were computed. The interaction energy of W3 to W7 water sites in the HCQ, indacaterol, and ZINC28706440 bound structures were calculated by sampling of 25,000 snapshots. For each snapshot, we computed the average interaction energy with both the protein and all other (bulk) water molecules with the VMD program using the namd-energy plugin [60]. We calculated the energy ΔE shown in the following equation

$$\Delta E = E_{WP} - ½ (E_{W\text{-}bulk})$$

where ΔE is the energy difference, E_{WP} is the mean interaction energy between a water molecule in the hydration site and protein, and E_{W-bulk} is the mean total interaction energy of a water molecule in bulk.

The water site is a region where water molecules are more likely to be found than in bulk. Therefore, the probability of finding water molecules or WFP is assigned as a function in volume V normalized with respect to the bulk solvent density for each MD structures. Using the inverse Boltzmann relationship, we obtain the following equation and characterize each water site from each conformation and also quantify this magnitude in energetic terms, by defining J as

$$J = -RT\ ln\ (WFP)$$

To compare this parameter for different water sites, the values were reported in Table 1 and were calculated as an arbitrarily chosen volume of 1 $Å^3$ [61].

2.6. Analysis of ligand binding pocket

The Pocket Volume Measurer program (POVME: version 3.0) [62] was employed to identify and characterize the volume, shape, and size of ligand binding pocket in four MD structures. Using POVME, we quantitatively compared the volume of ligand-binding pockets as a function of time, which is inaccessible experimentally. The program uses a single inclusion sphere by assigning the ND1 atom at the active site residue (His41) as a center of the inclusion sphere, which encompasses the entire binding cleft of ligand. The POVME algorithm calculates the volume of pocket by subtracting the space occupied by protein atoms in each frame from the inclusion sphere volume. This procedure was used for all MD structures after superimposing each one and aligning their trajectories to their first frame. Around 500 frames from each trajectory were used to measure the average volume of each pocket. The size of the inclusion sphere was set to 11 Å, and the remaining values of van der Waals radius, grid spacing, and contiguous cutoff parameters were set to 4.50, 0.50, and 1.09 Å, respectively [62].

2.7. Binding energy calculation using MM/PBSA and LIE-D method

The binding energy of ligand HCQ, indacaterol, and ZINC28706440 were evaluated by the results from protein−ligand MD trajectories using the molecular mechanics combined with the Poisson−Boltzmann and surface area (MM/PBSA) method in the program **C**alculation of **F**ree **E**nergy (CaFE) [63]. The single-trajectory approach is more widespread owing to its simplicity, precision, efficiency, and accuracy compared to the multi-trajectory setup. The MD trajectories of the protein-ligand complex as obtained in the one-average strategy were used to evaluate each free energy term on ΔG_{Bind} within the following equation:

$$\Delta G_{Bind} = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand}).$$

$\Delta G_{complex}$ represents the free energy of the complex, and $\Delta G_{protein}$ and ΔG_{ligand} indicate the free energies of the unbound protein and ligand, respectively. Three different energy terms are calculated by the MM/PBSA method which includes energy difference between the complex and separate receptor and ligand molecule in the gas phase from NAMD. In MM/PBSA, polar solvation free energy was calculated using the APBS program [64] by solving the Poisson−Boltzmann equation by assigning the PB radii as bondi, PB boundary condition as mdh (multiple Debye-Huckel), and PB chgm (charges are mapped to the grids) as spl2 (cubic B-spline). The nonpolar term ($\Delta G_{nonpolar}$) is often approximated by a solvent accessible surface area (SASA) and were estimated by assigning SA radii as bondi using a simple empirical relation: $(\Delta G_{nonpolar}) = \gamma A + b$, where A is the solvent-accessible surface area that was estimated using the VMD program, gamma (γ) and b are empirical constants, and in this work, we used 0.0054 and 0.92 kcal/mol for γ and b, respectively. The change in conformational entropy (−TΔS) is generally computed by normal-mode analysis on a set of conformational snapshots taken from MD simulations. Due to the high computational demand and large standard error compared to other contributing terms, the entropic term was neglected in the MM/PBSA models when only the

Table 1
Thermodynamic and structural parameters of water molecules at the catalytic site (ligand binding) in different conformations of MD simulated structures of Mpro protein.

MD-Structures	Conserved Water Sites	Reference Residue ID	E_{WP}	½ E_{W-bulk}	Δ*E*	WFP	J
Native	W3	H41, D187, H164	−18.75	−3.50	−15.25	24.00	−1.90
HCQ	W3	H41, D187, H164	−20.73	0.15	−20.88	26.00	−1.95
Indacaterol	W4	H41	−8.27	−4.29	−2.82	6.16	−1.09
	W5	W4, W6	1.60	−6.14	7.74	1.89	−0.38
	W6	S46, LIG	−1.92	−5.45	2.37	3.58	−0.76
ZINC28706440	W7	H41, C145	−10.49	−3.61	−6.88	11.00	−1.43

E_{WP}, Conserved Water with protein; E_{W-bulk}; Conserved Water with bulk water; LIG; indacaterol.
*The calculation is defined previously in Methods in Section 2.5. The energies (E_{WP}, $E_{W-bulk,}$ Δ*E*, WFP, and J interaction value) are in kcal/mol.

relative binding free energies of similar ligands are needed and presented in this work [65].

By default, the dielectric constant of solute (ε_{solute}) was set to 1 to examine the sensitivity of calculated binding free energies to the solute dielectric constant. The 2,500 frames were used to estimate the contribution of the binding free energy. In the current work, we evaluated the LIE-D method [66] to achieve the binding energy of each ligand to compare the result from the MM/PBSA model. The binding free energies of ligand HCQ, indacaterol, and ZINC28706440 were estimated from protein–ligand MD structures using the modified linear interaction energy method (LIE-D) in the CaFE plugin of VMD. LIE-D is widely used to calculate non-covalent binding free energies of ligands to proteins. To estimate the binding free energy of ΔG (kcal/mol), we used the following scoring function:

$$\Delta G_{bind}^{LIE\text{-}D} = \beta\,(V_{ele\text{-}complex} - V_{ele\text{-}ligand}) + \alpha\,(V_{vdw\text{-}complex} - V_{vdw\text{-}ligand}) + (-0.95\,[\beta\,(V_{ele\text{-}complex} - V_{ele\text{-}ligand}) - \alpha\,(V_{vdw\text{-}complex} - V_{vdw\text{-}ligand})] - 2.06)$$

The coefficients α and β are adjustable empirical parameters for the van der Waals and electrostatic contribution of the non-bonded interaction energy to the free energy of binding, respectively. The balance (difference) between electrostatic (polar) and van der Waals (non-polar) contributions to the binding free energy was defined as $D = \beta\,(V_{ele\text{-}complex} - V_{ele\text{-}ligand}) - \alpha\,(V_{vdw\text{-}complex} - V_{vdw\text{-}ligand})$ [66]. The final 5 ns trajectories from each MD simulation were considered for calculating the binding energy of each ligand using LIE-D. The empirical α and β coefficients were initially set to 0.18 and 0.44, respectively.

3. Results and discussion

3.1. Analysis of ligand-binding pocket in Mpro-crystal structures

The eleven crystal structures of the Mpro-native are obtained from the Protein Data Bank either in ligand-bound or ligand-free forms. These structures have been independently solved by several research groups using PHENIX or REFMAC or BUSTER or nCNS refinement programs at 6.1–7.5 pH. Only those structures determined in C 1 2 1 space groups were considered in the present computational study to eliminate crystal lattice bias (except PDB ID: 7C2Y in P 2_1 2_1 2_1) because insufficiencies of structural data in different space groups may not conclude a comparative analogous study. In this context, the present study revealed the existence of W1, W2, and W3 water molecules at the catalytic region of the Mpro which may have plausible roles in the structure, function, and stabilization of Mpro-native. Furthermore, the superimposed complex structures between 6YB7, 6Y84, 7ALH, 7ALI, 7BRO, 7JR3, and 7KPH have clearly displayed the presence of W1 and W2 water molecules those are occupied a bit far (<3.50 Å) from Asp187, in contrast, to the superpositions of 7C2Y, 7CWB, 7JUN, and 7JVZ structures showing the position of W3 site that makes H-bond with Asp187. Interestingly, W1 forms H-bonds with ND1 of His41 and oxygen of W2, W2 associates only W1 site, and W3 makes H-bonds with ND1 of His41 and OD2 of Asp187 (presented in Fig. 1). The geometrical positions of W639, W615, W573, W563, W499, W484, and W629 are found to be conserved in the W1 site with B_{Norm} value from 0.42 to 5.07 Å^2 and average SASA value ~8.97 Å^2, moreover, W864, W800, and W691 are located at W2 position with B_{Norm} ranging from 2.24 to 3.89 Å^2 and average of SASA value ~23.87 Å^2, and similarly, W346, W441, W409, and W429 are occupied in W3 position with B_{Norm} values from −0.78 to −1.54 Å^2 and average of SASA value 0 0.0 Å^2. Thus, the B_{Norm} of W1 and W2 positions are >0.0 Å^2 and SASA value > 2.5 Å^2 (except for few structures, PDB ID: 6Y84, 7JR3, and 7KPH), on the contrary, the corresponding value for B_{Norm} in W3 site is < 0.0 Å^2 and SASA value ~ 0.0 Å^2 (detail mentioned in Table S2). Especially, the average Pearson's correlation coefficient r_p (R^2) between B_{Norm} and SASA values in W1 and W2 are −0.61 (0.37) and −0.78 (0.61), respectively, that represents an opposite correlations. Moreover, water molecules with an accessible surface area <2.5 Å^2 were considered to be buried water molecules [67]. Hence, the observed atomic displacement parameters (a.d.p) usually referred as B_{Norm} and SASA values of water molecules revealed that the W3 site is considerably stable, invariant, and deeply buried in crystal structures whereas the locations of W1 and W2 are dynamic and highly exposed. According to suggestion from Carugo et al. [68], if B_{Norm} is greater than or ~ 1.20 Å^2 for any hydration site, then that will be fully occupied by water molecule. This information indicates the W3 location in PDB ID: 7CWB and 7JUN can also be occupied fully by a water molecule.

The crystallographic structural parameters (c.s.p.) such as Matthews coefficient, solvent content, and calculated mean B-factors of protein are found maximum (2.83, 56.48, and 33.56 Å^2) in 7C2Y and minimum (1.91, 35.50, and 16.87 Å^2) in 6YB7. Moreover, the observed ratio of the number of protein atoms with respect to water molecules (N_{PROT}/N_{HOH}) in 11 Mpro-native (at room temperature with 1.25–2.50 Å resolution) are ranges from 5.34 to 51.45 that are suggesting 7JVZ ensemble is actually random compared to remaining structures [69] (mentioned in Table S2). Remarkably, the average Pearson's correlation coefficient r_p (R^2) between B_{Norm} and Matthews coefficient in W1, W2, and W3 are −0.18 (0.03), −0.83 (0.69), and 0.62 (0.39) whereas the r_p (R^2) between B_{Norm} and N_{PROT}/N_{HOH} in W1, W2, and W3 are −0.79 (0.63), −0.99 (0.98), and 0.13 (0.02). In the present study, an attempt has been made to establish a relationship between a.d.p. of three water molecules and c.s.p of 11 crystal structures that indicates 6YB7 is the stable structure but its W1 and W2 site are considerably unstable compared to W3.

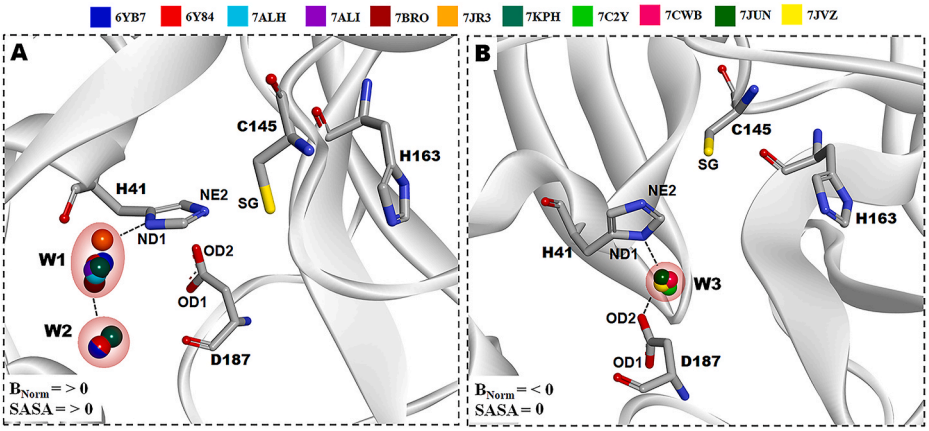

Fig. 1. Superposition of 11 native conformations of X-ray structures reveal the positions of conserved water molecules inside the catalytic pocket in Mpro protein. **(A)** Water molecule W1 and W2 is conserved with respect of D187 in PDB ID: 6YB7, 6Y84, 7ALH, 7ALI, 7BRO, 7JR3, and 7KPH. **(B)** W3 is conserved and also makes H-bond with respect of D187 in PDB ID: 7C2Y, 7CWB, 7JUN, and 7JVZ.

3.2. Binding mode prediction of HCQ, its analogous and analysis of their pharmacological properties

Our computational drug repurposing workflow against Mpro-native was started with a docking of 4 FDA-approved drugs and 11 new ligands (analogous HCQ) using AutoDock Vina program (v1.1.2). In this approach, we elucidate how the ligands are occupying the positions of W1 or W2 in the binding site? In water-based drug discovery technique, the docking results can be considered a "good fit" if ligands grip the location of water sites and occupy those positions accordingly [31]. To do so, we exhaustively searched the proper and accurate orientation of each docked ligand, so that corresponding atoms of the respective ligand can occupy the positions of water sites in docked structure. To probe the validity of this procedure, we obtained the docked structures accordingly. The HCQ, indacaterol, and ZINC28706440 were only considered for further study because they were displaced of those water molecules in the binding pocket of Mpro whereas remaining 12 HCQ analogs were not feasible. Analysis of molecular docking results showed HCQ, indacaterol, and ZINC28706440 formed short contacts with ND1 of His41 and their binding energies were −5.80, −8.60, and −5.60 kcal/mol respectively (Table S3). Furthermore, HCQ, indacaterol, and ZINC28706440 have three; five, and four H-bonding interactions with the residues of Mpro respectively, and they also get short contacts with Cys145 (except ZINC28706440). The indacaterol has high affinity for Mpro protein rather than remaining 14 ligands, but HCQ and ZINC28706440 have similar binding affinities in docking study.

To assess the statistical significance of this docking result, the physico-chemical and toxicity comparison was performed for 15 ligands. In the present study, OSIRIS Property Explorer was used to predict the drug-likeness score of all compounds and comparing them with the drug HCQ. The program calculated the risks of side effects, such as mutagenic, irritant, tumorigenic, and reproductive effects, as well as drug-relevant properties. Moreover, the overall drug-score was estimated from summation of cLogP, solubility, TPSA, drug-likeness, drug-score, and molar refractivity parameters. Interestingly, the potential drug-likeness score of HCQ, chloroquine, methylergometrin, indacaterol, and ZINC77520192 were significantly higher (ranging from 6.54 to 8.03) than remaining ligands. However, except HCQ, other three compounds, produced, insignificant docking results. However, the values of physico-chemical parameters referred to as drug-score, cLogP, TPSA, drug-likeness, and synthetic-accessibility are 0.48, 3.08, 48.39, 6.54, and 2.82 in HCQ, 0.66, 2.81, 81.59, 2.12, and 3.68 in indacaterol, and 0.50, 1.60, 57.18, 1.74, and 2.01 in ZINC28706440, respectively (Table S4). Mostly, the drug-score of HCQ, indacaterol, and ZINC28706440 were significantly higher (ranging from 0.48 to 0.66) and their predicted *in silico* toxicity-risk properties were also higher than other ligands (Table S5).

3.3. Analysis of MD trajectories

3.3.1. Ligand- Mpro interaction profiling

The findings of MD simulations have been explored in terms of dynamics and provide new insights to discriminate the structural and functional properties between native and ligand bound Mpro conformations. We observed that HCQ interacts with His41, Cys145, His164, Val186, Arg188, and Gln189, indacaterol with His41, Ser46, Glu47, Asn142, and Gln189, and ZINC28706440 with His41, His163, His164, Glu166, Arg188, and Gln189. Ligand-residue interactions profiling were measured on final 2,500 MD-structures and were presented in Table S6. During simulation, conformational changes upon binding HCQ, indacaterol, and ZINC28706440 to Mpro protein suggest that HCQ and ZINC28706440 have similar binding pose and identical interactions with His41, His164, Arg188, and Gln189 but indacaterol has different interactional patterns in respect of His41 (Fig. S3).

The present investigation focuses on the residues of ligand-binding pocket (His41, Cys145, and His163) along with non-catalytic sites and seems significant. The flexibility (RMSF) of each residue in the Mpro-native, HCQ, indacaterol, and ZINC28706440-bound MD structures are described accordingly, where the N- and –C terminal regions of this protein exhibit more flexibility (upper panel of Fig. 2). In MD structure, atomic fluctuation (CA atom only) of Mpro-native is 0.39–3.00 Å, HCQ is 0.37–2.50 Å, indacaterol is 0.38–3.00 Å, and ZINC28706440 is 0.37–3.00 Å. The RMSF values for active site residues of His41, Cys145, and His163 are 0.54, 0.50, and 0.48 Å in Mpro-native, 0.53, 0.49, and 0.43 Å in HCQ, 0.54, 0.50, and 0.46 Å in indacaterol, and 0.56, 0.49, and 0.45 Å in ZINC28706440, Whereas, residues like Glu47, Leu50, Asn51, Ser139, Tyr154, Pro168, Ala191, Ala193, Asn277, and Gly278 have reasonably high values ranging from 1.73 to 1.89 Å. Therefore, the atomic deviations of these residues in Mpro-native and indacaterol structures seem more flexible than remaining structures. Moreover, the comparative trajectory analysis of RMSD (CA atom) with respect to time (ns) among the MD structures of Mpro-native, HCQ, indacaterol, and ZINC28706440-bound forms indicate that the RMSD value is within 0.54–2.92 Å in Mpro-native, 0.55 to 2.17 Å in HCQ, 0.51 to 2.73 Å in indacaterol, and 0.52 to 2.59 Å in ZINC28706440-bound structures (lower panel of Fig. 2). A comparative analysis of these four MD structures show the average RMSD of Mpro-native (1.84 Å) and indacaterol (1.93 Å) have large fluctuations and display more dynamic nature than HCQ (1.37 Å) and ZINC28706440 (1.68 Å) -bound structure.

The average volumes of ligand-binding pockets, especially the free space are ~820 $Å^3$ in Mpro-native crystal structures. As a rough approximation, we calculated the size of the pocket required to bind a ligand by measuring the volume of a "minimal sphere." During the conformational transition from Mpro-native to HCQ or indacaterol or ZINC28706440, the volume of ligand-binding pocket increases from 862 to 948 $Å^3$ in HCQ, 1991 $Å^3$ in indacaterol, and 1088 $Å^3$ in ZINC28706440 (Fig. 3). Therefore, MD data suggest that the volume of the ligand-binding pocket expands significantly in ligand-bound conformation relative to Mpro-native, even though, the volumes are enlarged similarly in HCQ or ZINC28706440 than indacaterol.

3.3.2. Structural, dynamical, and thermodynamic characterization of water molecules

The present MD simulation studies of ligand-binding pocket demonstrate and confirm the positions of five (W3 to W7) water molecules in the protein. The thermodynamic and structural properties of these water molecules were characterized and mentioned in Methods in Section 2.5. These water sites establish that the H-bonds with ligands or residues or conserved water molecules (described in Table 1) and by assigning them as references (explained in the method section), we were capable to find these water sites with significant water occupation probability. Due to the inherent mobility of water molecules in MD structures, it is difficult to determine and visualize their positions. Therefore, to identify the presence of water molecules in the ligand-binding pocket in X-ray crystallographic structure this is the actual source for obtaining structural information. We therefore compare the water positions in MD conformations with their respective crystal structures in Mpro-native. We noticed that the positions of water site W3 (W604 in 6Y84), W4 (W861 in 6Y84), and W7 (W816 in 6Y84) were available in both MD and X-ray structures whereas W5 and W6 water locations were only produced during MD simulation. If the crystal structures of Mpro do not show any explicit water molecules in the ligand-binding site, this does not necessarily mean that they are really absent in structures. Perhaps, these water sites are either in rapid exchange with bulk solvent or relatively mobile within the binding site. Therefore, they were inaccessible in crystallographic studies, but may still play a major role in the protein function.

As described in Table 1, water site W3, W4, and W7 have high WFP values and ranging from 26.00 to 6.16 because they have, (i) strong E_{WP} (−20.73 to −8.27 kcal/mol), and (ii) weak E_{W-bulk} interaction energy (0.15 to −4.29 kcal/mol), therefore, their ΔE (−20.88 to −2.82 kcal/mol) are also high. In contrast, the W5 and W6 sites contain less WFP

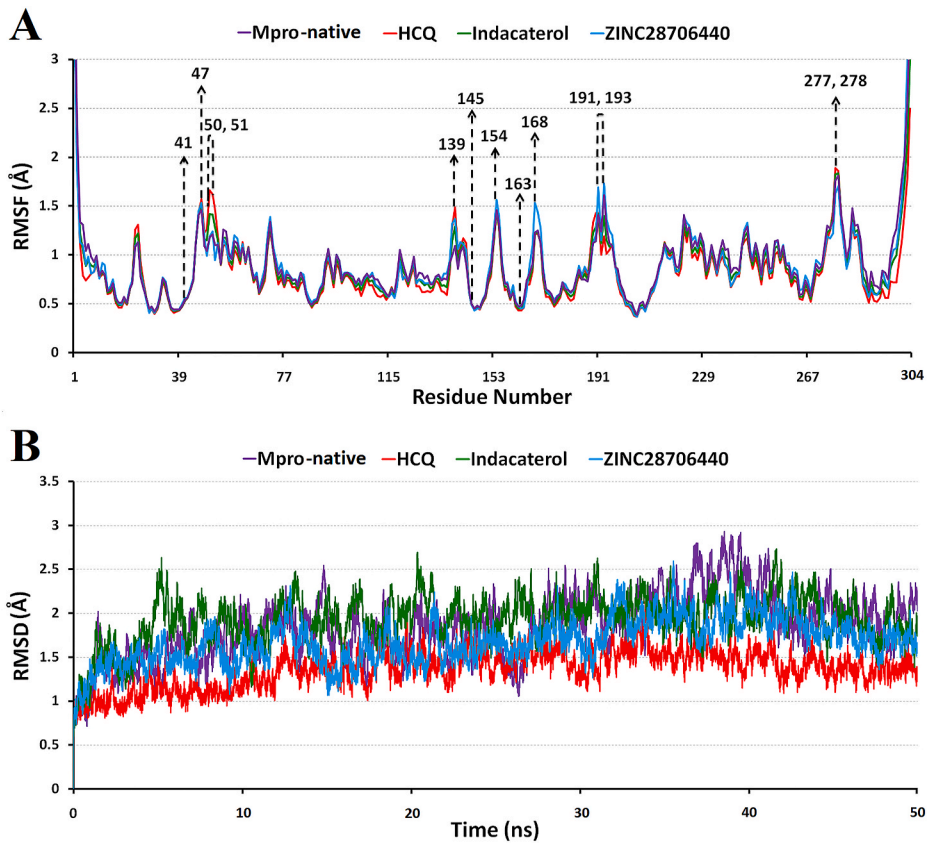

Fig. 2. Top: Comparison of RMSF (root-mean-square fluctuation) (Å) of CA atoms in MD structure of Mpro-native (violet), HCQ (red), indacaterol (green), and ZINC28706440 (blue) bound form. The final 25,000 frames for each conformation were used. **Bottom**: Comparison of RMSD (root-mean-square deviation) (Å) of CA atoms in MD structure of Mpro-native (violet), HCQ (red), indacaterol (green), and ZINC28706440 (blue) bound form.

values (lower than 4.00) with high E_{W-bulk} (−6.14 and −5.45 kcal/mol), and weak E_{WP} interaction energy (1.60 and −1.92 kcal/mol), as a result, ΔE are also weak (7.74 and 2.37 kcal/mol). The corresponding WFP values expressed in the terms of interaction energy (J) are distinct and are also important. W3, W4, and W7 are also highly prominent and stable in their corresponding positions because they have the lowest J energy value ranging from −1.95 to −1.09 kcal/mol in comparison with W5 and W6 whose corresponding values are −0.38 and −0.76 kcal/mol. Remarkably, the average Pearson's correlation coefficient r_p (R^2) of W3 with W4, W3 with W7, and W4 with W7 are significantly high [0.97 (0.94) to 0.99 (0.99)] in respect of E_{WP}, E_{W-bulk}, and WFP and explains these corresponding water sites could be present predominately in MD structures.

The ligand-binding site is stabilized by either W3 or W7 in Mpro-native, HCQ, and ZINC28706440-bound conformation, whereas W4, W5, and W6 water sites are observed in indacaterol. To identify the interactions of water molecules with ligand or catalytic residues in each MD-structure, we observed the region of His163—Ser144-Cys145—His41—W3—Asp187 in Mpro-native, His163—Ser144-Cys145/HCQ—His41—W3—Asp187 in HCQ, His163—Ser144-Cys145—W4—His41—His164/Asp187 and His41—W4—W5—W6—indacaterol in indacaterol, and His163—Ser144-Cys145—W7—His41—ZINC28706440 in ZINC28706440-bound conformation (Fig. 4). To explain the possible role of each water molecule from each MD structure, we assumed that W3 water can act as a catalytic partner of Mpro by stabilizing the thiolate–imidazolium (Cys145-SH—NE-His41) and imidazole-chlorine (HCQ-Cl—NE-His41) interactions in the Mpro-native and HCQ structure respectively. Meanwhile, W4 makes H-bond with the imidazole group of His41, W5 also makes a water mediated bridge between W4 and W6, and W6 interacts with indacaterol (N1 atom). Moreover, W7 is observed to occupy Cys-His dyad in ZINC28706440 to inactivate the thiolate–imidazolium ion-pair.

3.3.3. Occupying the positions of W1 or W2 by ligand in X-ray crystal structure

In the present study, we investigated experimentally known data in order to produce the crystallographically predicted conserved water molecules in Mpro-native on a basis of statistical and thermodynamic analysis. Thus, the interaction energy of W1 and W2 sites in 6YB7 and W3 in 7CWB crystal structure are computed accordingly. The corresponding values expressed in the terms of interaction energy ΔE for W1, W2, and W3 are −4.79, −3.90, and-17.19 kcal/mol while their interaction energy with protein (ΔE_{WP}) are −2.93, −1.98, and −17.43 kcal/mol respectively. The displacement of W1 or W2 water site from the catalytic region could inactivate the Mpro enzyme and these water sites are less stable and dynamic due to $B_{Norm} > 0$ Å^2, are highly exposed with solvent-accessible surface areas because of SASA value > 0 Å^2, and their electrostatic (ΔE) and bulk water (E_{W-bulk}) energy are higher than W3 position. Therefore, taking up the positions of W1 or W2 by the proposed ligand in Mpro using molecular docking study will be a challenging task. This structural result can help us for more detailed physico-chemical characterizations of the conserved water molecules that play an important role in molecular docking study. For investigating water displacement approach regarding the ligand-water interactions; we superimposed the HCQ, indacaterol, and ZINC28706440 on the 6YB7 crystal structure. Findings from this docking study reveals the C11 of HCQ and C22 of indacaterol occupy the position of W2, while C2 and O1 of ZINC28706440 grip the locations of W1 and W2, respectively. Molecular docking study may produce insignificant information regarding the rearrangements of W1, W2, and W3 water molecules after binding to the ligand, thus, MD-simulation is the best alternative option to justify the adjustment of this water in terms of dynamics at the active site of

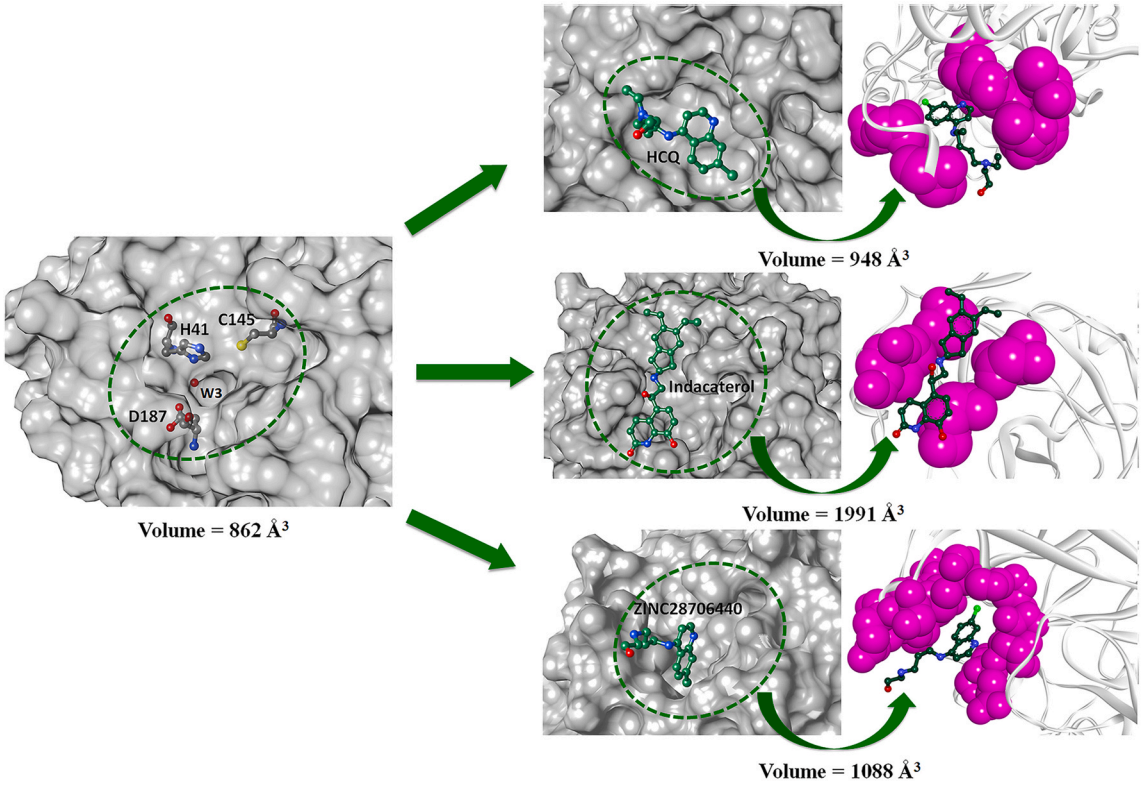

Fig. 3. Approximate volume (Å^3) of the binding pocket for ligand HCQ, indacaterol, and ZINC28706440 during structural transitions from native to ligand-bound form in the MD simulation. Each molecular surface of the panel marks by a green circle represents ligand-binding groove of the Mpro protein. The changes in volume of each MD structure in **(A)** native Mpro, **(B)** HCQ, **(C)** indacaterol, and **(D)** ZINC28706440 are marked below the structures.

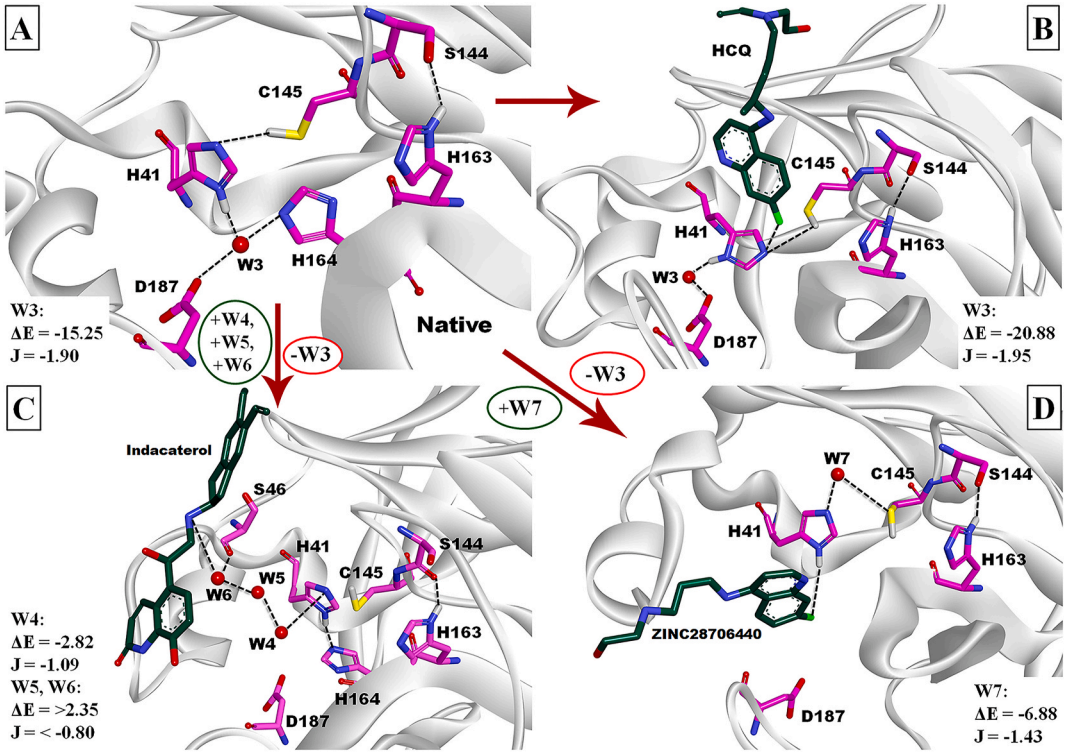

Fig. 4. MD simulations of native, HCQ, indacaterol, and ZINC28706440 inside the active-site pocket of Mpro protein. Catalytic residues His41, Cys145, His163, and ligands are highlighted by the stick models, while water molecules represent in the pink ball. **(A)** The interactions of His163—Ser144-Cys145—His41—W3—Asp187 are shown inside the active-site pocket in Mpro-native, **(B)** His163—Ser144-Cys145/HCQ—His41—W3—Asp187 are shown in HCQ-bound MD structure, **(C)** His163—Ser144-Cys145—W4—His41—His164/Asp187 and His41—W4—W5—W6—indacaterol in indacaterol-bound MD structure, and **(D)** His163—Ser144-Cys145—W7—His41—ZINC28706440 in ZINC28706440-bound conformation.

pocket. Finally, the MD-simulation suggests that the C11 and N2 atoms in HCQ obtain the positions of W1 and W2, respectively, and binding energy of HCQ comparatively augments the docked conformation (−5.80 to −9.76 kcal/mol). Consistently, indacaterol grip the W2 location during docking but it cannot occupy either W1 or W2 position during MD-simulation and gets more stable by losing energy (−8.60 to −43.77 kcal/mol). Similarly, ZINC28706440 captures the locations of W1 and W2 in the docking study but it moves away from these water locations in MD-simulation and obtains stability by releasing the energy from −5.60 to −22.28 kcal/mol (Fig. 5 and Table S7).

3.3.4. Comparative analysis of LIE-D and MM/PBSA energy of ligand for Mpro

The LIE-D and MM/PBSA binding free energies and their decompositions for the three ligands were compared and mentioned in Table 2. As a model in MM/PBSA, the atomic radii plays a significant role in the performance of solvation energy calculations. In the present work, we used bondi radii because the chlorine atom (in HCQ and ZINC28706440) is parameterized accordingly and assigned the value 1.76 Å [70].

We observed that the average lig-RMSD is 1.74 in HCQ, 1.96 in indacaterol, and 1.71 Å in ZINC28706440. The lig-RMSD is < 2 Å suggesting a stable conformation of ligand thus it reflects HCQ, indacaterol and ZINC28706440 that are relatively stable. For MM/PBSA, the mean value of electrostatic energy ($\Delta E^{ele} + \Delta G^{PB}$) is 12.26 in HCQ, 3.95 in indacaterol, and 1.46 kcal/mol in ZINC28706440 but for LIE-D, the value of ΔE^{ele} is −10.36 in HCQ, −94.92 in indacaterol, and −136.11 (kcal/mol) in ZINC28706440. However, the electrostatic energies are significantly different for the three binding modes. The overall correlation value r_p (R^2) of $\Delta E^{ele} + \Delta G^{PB}$ with $\Delta G^{MM/PBSA}$ and ΔE^{ele} with ΔG^{LIE-D} are −0.43 (0.19) and 0.98 (0.95) in HCQ, 0.62 (0.39) and 1.00 in indacaterol, and 0.95 (0.90) and 1.00 in ZINC28706440. The contribution of electrostatic energy to ΔG in MM/PBSA method is poor in HCQ and strong in ZINC28706440 while the LIE-D predicts these energies contributions are strong and significant in the respective three ligands. Similarly, for the van der Waals and hydrophobic interaction energies ($\Delta E^{vdw} + \Delta E^{SA}$) in MM/PBSA and ΔE^{vdw} values in LIE-D are −33.30 and −28.90 in HCQ, −14.14 and −11.12 in indacaterol, and −26.37 and −22.28 kcal/mol in ZINC28706440. For the three binding modes, the van der Waals energies are very different, which means HCQ and ZINC28706440 have a good hydrophobic contact than indacaterol. The correlation value r_p (R^2) of $\Delta E^{vdw} + \Delta E^{SA}$ with $\Delta G^{MM/PBSA}$ and ΔE^{vdw} with ΔG^{LIE-D} are 0.80 (0.63) and 0.89 (0.80) in HCQ, 0.73 (0.53) and −0.31 (0.10) in indacaterol, −0.72 (0.52) and −0.92 (0.84) in ZINC28706440 that reveals this relationship of two energy terms with ΔG that are very high in HCQ and poor in ZINC28706440.

In this regard, it is concluded that both the electrostatic and the van der Waals interactions have significant role for binding of ligand, but the relative electrostatic interaction contributes more. The electrostatic energy and lig-RMSD of HCQ and ZINC28706440 are apparently similar and has been predicted by LIE-D. Our findings also support the results of different research groups [71] to indicate that the electrostatic interaction energy may be a good predictor for identifying best structural analogous of ligands and it is the predominant factor for determining the different binding orientations. Only when comparing the ΔG correlation results of HCQ versus indacaterol and HCQ versus ZINC28706440, we observed that the $r_p(R^2)$ values are −0.74 (0.54) and 0.47 (0.22) in MM/PBSA and 0.41 (0.61) and 0.72 (0.52) in LIE-D. According to the correlation findings (r_p and R^2), the MM/PBSA and LIE-D methods suggest that the ΔG value of HCQ is correlated strongly with ZINC28706440 and insignificantly with indacaterol. Hence, HCQ has a strong correlation with ZINC28706440 than indacaterol and thus these statistical findings in respect of r_p values also demonstrate that ZINC28706440 is the best structural analogous of HCQ.

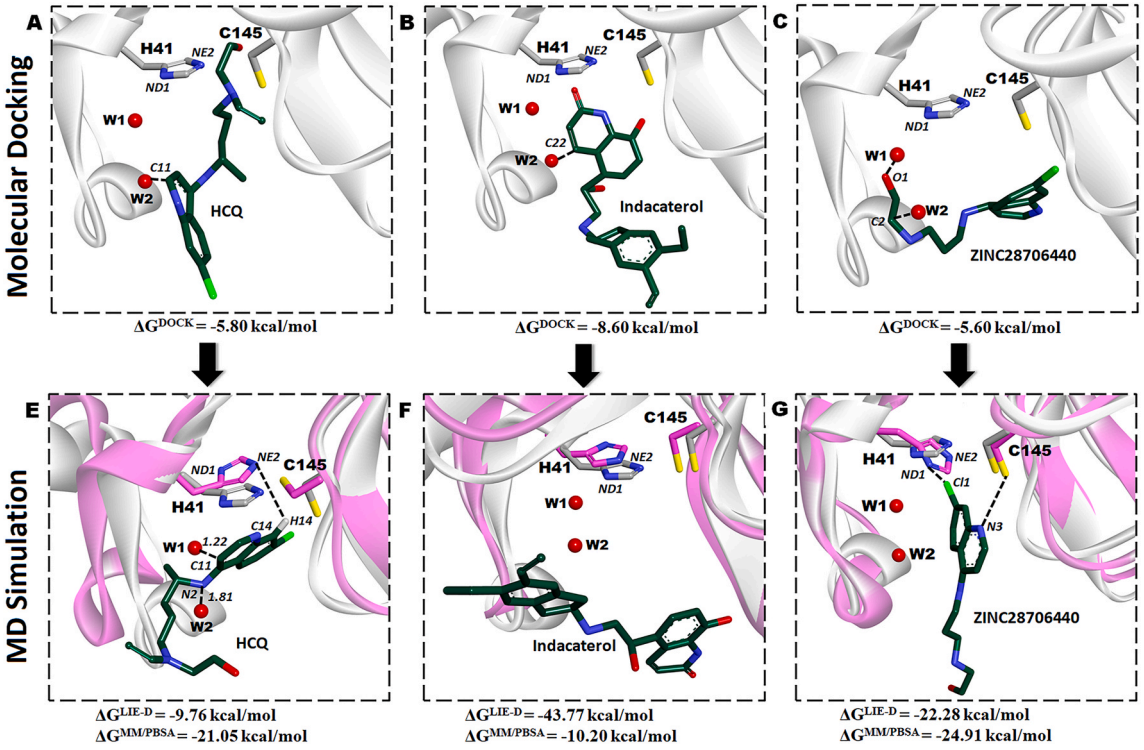

Fig. 5. Top: Superimposed complex structures of HCQ, indacaterol, and ZINC28706440 (docked conformation) with Mpro-native (PDB ID: 6YB7). The C11 of HCQ and C22 of indacaterol occupy the W2 location, but C2 and O1 of ZINC28706440 take the W1 and W2 sites, respectively. **Bottom:** Superposition of the HCQ, indacaterol, and ZINC28706440 bound-MD conformations on 6YB7 crystal structure show the C11 and N2 atom of HCQ grip the W1 and W2 locations, but indacaterol, and ZINC28706440 are remotely occupied without displacing W1 and W2 site.

Table 2
The binding free energy (in kcal/mol) of the ligand HCQ, Indacaterol, and ZINC28706440 are calculated by the LIE-D and MM/PBSA method. The parameters α and β are used in LIE-D, whereas SD is calculated from MM/PBSA.

	LIE-D					MM/PBSA							
		α	β	ΔE^{Vdw}	ΔE^{ele}	ΔG^{LIE-D}	ΔE^{Vdw}	ΔE^{ele}	ΔE^{PB}	ΔE^{SA}	ΔG^{Pol}	ΔG^{np}	$\Delta G^{MM/PBSA}$	SD	*Lig-RMSD*
HCQ	50ns	0.18	0.44	−30.54	−10.05	−9.92	**−30.54**	−9.82	24.05	−4.52	14.22	−35.06	−20.84	3.31	1.84
	49ns	0.18	0.44	**−30.82**	−12.05	−10.85	−30.82	−11.87	**25.53**	−4.57	13.65	−35.40	−21.75	**2.44**	1.73
	48ns	0.18	0.44	−30.16	**−12.60**	**−10.97**	−30.16	**−12.05**	23.99	−4.49	11.94	−34.66	**−22.72**	3.54	1.65
	47ns	0.18	0.44	−25.83	−9.04	**−8.62**	**−25.83**	−8.72	19.28	−4.21	10.55	−30.05	**−19.49**	3.84	**1.85**
	46ns	0.18	0.44	**−27.17**	**−8.04**	**−8.43**	−27.17	**−8.24**	**19.13**	−4.18	10.88	−31.35	−20.46	**4.47**	**1.65**
Indacaterol	50ns	0.18	0.44	−11.55	**−156.31**	**−70.86**	−11.55	**−182.79**	**183.73**	**−3.56**	**0.93**	−15.12	**−14.18**	3.44	**2.18**
	49ns	0.18	0.44	**−8.55**	−122.39	−55.39	**−8.55**	−145.90	148.30	−2.81	2.39	**−11.36**	−8.96	**5.77**	2.10
	48ns	0.18	0.44	−9.09	−68.87	−31.94	−9.08	−94.45	100.45	**−2.51**	**5.99**	−11.60	**−5.61**	3.68	1.97
	47ns	0.18	0.44	−13.00	−79.66	−37.39	−13.00	−107.87	112.73	−3.18	4.86	−16.18	−11.32	3.86	1.85
	46ns	0.18	0.44	**−13.40**	**−47.39**	**−23.26**	**−13.40**	**−88.49**	**94.03**	−3.08	5.53	**−16.48**	−10.95	**2.24**	**1.72**
ZINC28706440	50ns	0.18	0.44	−20.81	**−158.61**	**−73.53**	−20.81	**−177.34**	174.32	−4.06	**−3.02**	−24.87	**−27.89**	4.26	1.79
	49ns	0.18	0.44	**−19.26**	−154.63	−71.50	**−19.26**	−175.83	**174.37**	−4.09	−1.46	**−23.35**	−24.82	**5.05**	1.82
	48ns	0.18	0.44	−22.31	−149.70	−69.88	−22.31	−172.53	171.10	**−4.12**	−1.42	−26.43	−27.86	4.08	**1.39**
	47ns	0.18	0.44	−22.71	−138.38	−64.97	−22.71	−170.96	173.19	−4.10	2.22	−26.81	−24.58	5.46	1.69
	46ns	0.18	0.44	**−26.33**	**−79.24**	**−39.60**	**−26.33**	**−104.53**	**115.51**	**−4.04**	**10.97**	**−30.37**	**−19.40**	**2.82**	**1.87**

*The SD column shows the standard deviation of the various terms in the MM/PBSA method, Lig-RMSD indicates root-mean-square deviations of ligand−atom positions (Å) between initial pose and each MD structures.

4. Conclusion

The present study focuses on the ligand-binding pocket of the Mpro protein to characterize its stabilization and the role of water molecules during the enzymatic mechanism. The comprehensive analysis of multiple crystal structures of Mpro-native reveals the presence of three conserved water locations (W1, W2, and W3) that maintain the architecture of the active site pocket when ligands are absent in the enzyme. The crystallographic structural parameters of those concern sites and their interaction energy profiling conclude that the W1 and W2 are dynamic in nature, considerably unstable, weakly bonded with protein in low interaction energy, and are highly exposed. Whereas W3 position is apparently stable, invariant, tightly bonded with the protein by strong interaction energy, and deeply buried in crystal structures of the Mpro-native. Therefore, we have planned to displace W1 or W2 water sites by triggering HCQ or its analogs to inactivate the enzyme and thus virtual screening with molecular docking studies have been adopted to determine the analogs for HCQ. The findings demonstrated that indacaterol and ZINC28706440 are the best analogs of HCQ with reasonably acceptable physico-chemical and toxicological values. The docking study reveals that HCQ and indacaterol occupy the W2 position and ZINC28706440 grip the locations of W1 and W2 with binding energy -5.80, -8.60 and -5.60 kcal/mol. Finally, the MD-simulations results show that the HCQ obtains the positions of W1 and W2 whereas its analogs (indacaterol and ZINC28706440) could not retain them and leave the position of either W1 or W2. These ligands gain stability by releasing energy compared to HCQ from docking to MD simulation study. The MD simulation results again consistent with the existence of five water molecules in the ligand binding pocket of the enzyme, especially water site W3, which is observed in Mpro-native and HCQ, W4, W5, and W6 in indacaterol, and W7 in ZINC28706440-bound structure. The W3 and W7 stabilize the thiolate–imidazolium and imidazole-chlorine interaction in Mpro-native and HCQ respectively and W7 also stabilizes His41—Cys145 in ZINC28706440-bound conformation. Interestingly, W4, W5, and W6 in indacaterol make a water mediated H-bond between ligand and His41. During conformational transition from Mpro-native to indacaterol or ZINC28706440, W3 is departed from Mpro-native and W4, W5, and W6 enter the binding pocket in indacaterol and similarly W7 arrives in ZINC28706440-bound structure. The computational evidence encourages us to identify structural, dynamical, and thermodynamic discriminative features of ligand binding site for comparative analysis of Mpro-native with HCQ or its analogs. In this context, WFP, J value, and other thermodynamic profiling parameters suggest that W3, W4, and W7 are highly prominent in their corresponding positions in comparison with W5 and W6. Our results appeared to be significant and evident that we decided to explore our important findings with the medical community in the current pandemic situation. We, therefore, recommend that ZINC28706440 is the best structural analog for HCQ that should be further synthesized, and require proper experimental and pre-clinical investigation for the possible treatment of nCOVID-19.

Authors contributions

Saima Reyaz: Methodology, Software handling (Molecular Docking, Screening, and ADMET), Formal Analysis, Data Curation, Writing-Review & Editing. **Alvea Tasneem:** Methodology, Software handling (Molecular Docking, and MD Simulation), Formal Analysis, Data Curation and Validation, Writing-Review & Editing. **Gyan Prakash Rai:** Methodology, Software handling (Molecular Docking, Screening, and ADMET), Formal Analysis, Investigation, Writing Review and Editing. **Hridoy R. Bairagya:** Conceptualization, Supervision and full correspondence, writing original draft preparation.

Data availability statement

Supplementary data of this article is provided in the supplementary material.

Ethical approval

The manuscript does not contain experiments on animals and humans; hence ethical permission not required.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

HRB is thankful to C.S.I.R. (Council of Scientific and Industrial Research) for providing Senior Research Associateship (Pool Scientist scheme). HRB gratefully acknowledges the Department of Biophysics, A. I.I.M.S., New Delhi and S.R., A.T., and G.P.R. also thankfully acknowledges the Department of Computer Science, Jamia Millia Islamia, Central University-New Delhi, for providing computational resources. We also acknowledge the support of High-performance computing facility at I.I.T. Delhi for MD simulations study. We are grateful to Prof. David J. Wales for his comments on the original manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2021.108021.

References

[1] World Health Organization, WHO Director-General's Opening Remarks at the Media Briefing on COVID-19. 2020, 2020. March 11. World Health Organization, https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020.
[2] WHO COVID-19 dashboard, Available online: https://covid19.who.int/. (Accessed 17 August 2021). accessed.
[3] M.A. Martinez, Lack of effectiveness of repurposed drugs for COVID-19 treatment, Front. Immunol. 12 (2021) 653, https://doi.org/10.3389/fimmu.2021.653371.
[4] T. Asselah, D. Durantel, E. Pasmant, G. Lau, R.F. Schinazi, COVID-19: discovery, diagnostics and drug development, J. Hepatol. 74 (2020) 168–184, https://doi.org/10.1016/j.jhep.2020.09.031.
[5] S.M. Bartsch, K.J. O'Shea, M.C. Ferguson, M.E. Bottazzi, S.N. Cox, U. Strych, J. A. McKinnell, P.T. Wedlock, S.S. Siegmund, P.J. Hotez, B.Y. Lee, How Efficacious Must a COVID-19 Coronavirus Vaccine Be to Prevent or Stop an Epidemic by Itself, medRxiv, 2020, https://doi.org/10.1101/2020.05.29.20117184.
[6] D. Zhou, S.M. Dai, Q. Tong, COVID-19: a recommendation to examine the effect of hydroxychloroquine in preventing infection and progression, J. Antimicrob. Chemother. 75 (2020) 1667–1670, https://doi.org/10.1093/jac/dkaa114.
[7] J. Liu, R. Cao, M. Xu, X. Wang, H. Zhang, H. Hu, Y. Li, Z. Hu, W. Zhong, M. Wang, Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro, Cell Discov 6 (2020) 1–4, https://doi.org/10.1038/s41421-020-0156-0.
[8] P. Gautret, J.C. Lagier, P. Parola, L. Meddeb, M. Mailhe, B. Doudier, J. Courjon, V. Giordanengo, V.E. Vieira, H.T. Dupont, S. Honoré, Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial, Int. J. Antimicrob. Agents 56 (2020) 105949, https://doi.org/10.1016/j.ijantimicag.2020.105949.
[9] R. Derwand, M. Scholz, Does zinc supplementation enhance the clinical efficacy of chloroquine/hydroxychloroquine to win today's battle against COVID-19? Med. Hypotheses 142 (2020) 109815, https://doi.org/10.1016/j.mehy.2020.109815.
[10] WHO Solidarity Trial Consortium, Repurposed antiviral drugs for COVID-19—interim WHO SOLIDARITY trial results, N. Engl. J. Med. 384 (2021) 497–511, https://doi.org/10.1056/NEJMoa2023184.
[11] E.S. Rosenberg, E.M. Dufort, T. Udo, L.A. Wilberschied, J. Kumar, J. Tesoriero, P. Weinberg, J. Kirkwood, A. Muse, J. DeHovitz, D.S. Blog, B. Hutton, D. R. Holtgrave, H.A. Zucker, Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with COVID-19 in New York state, J. Am. Med. Assoc. 323 (2020) 2493, https://doi.org/10.1001/jama.2020.8630.
[12] RECOVERY Collaborative Group, Effect of hydroxychloroquine in hospitalized patients with Covid-19, N. Engl. J. Med. 383 (2020) 2030–2040, https://doi.org/10.1056/nejmoa2022926.
[13] P.W. Horby, M. Mafham, L. Linsell, J.L. Bell, N. Staplin, J.R. Emberson, M. Wiselka, A. Ustianowski, E. Elmahi, B. Prudon, A. Whitehouse, T. Felton, J. Williams, J. Faccenda, J. Underwood, J.K. Baillie, L. Chappell, S.N. Faust, T. Jaki, K. Jeffery, W.S. Lim, A. Montgomery, K. Rowan, J. Tarning, J.A. Watson, N.J. White, E. Juszczak, R. Haynes, M.J. Landray, Effect of Hydroxychloroquine in Hospitalized Patients with COVID-19: Preliminary Results from a Multi-centre, Randomized, Controlled Trial, MedRxiv, 2020, https://doi.org/10.1101/2020.07.15.20151852.
[14] Z. Song, Y. Xu, L. Bao, L. Zhang, P. Yu, Y. Qu, H. Zhu, W. Zhao, Y. Han, C. Qin, From SARS to MERS, thrusting coronaviruses into the spotlight, Viruses 11 (2019) 59, https://doi.org/10.3390/v11010059.
[15] T. Phan, Novel coronavirus: from discovery to clinical diagnostics, Infect. Genet. Evol. 79 (2020) 104211, https://doi.org/10.1016/j.meegid.2020.104211.
[16] W. Cui, K. Yang, H. Yang, Recent progress in the drug development targeting SARS-CoV-2 main protease as treatment for COVID-19, Front. Mol. Biosci. 7 (2020) 616341, https://doi.org/10.3389/fmolb.2020.616341.
[17] T.K. Nandi, H.R. Bairagya, B.P. Mukhopadhyay, K. Sekar, D. Sukul, A.K. Bera, Conserved water-mediated H-bonding dynamics of catalytic Asn 175 in plant thiol protease, J. Biosci. 34 (2009) 27–34, https://doi.org/10.1007/s12038-009-0006-6.
[18] A. Tasneem, G.P. Rai, S. Reyaz, H.R. Bairagya, The possible molecular mechanism of SARS-CoV-2 main protease: new structural insights from computational methods, Sci. Med. J. 2 (2021) 108–126, https://doi.org/10.28991/scimedj-2020-02-si-11.
[19] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs, Science 300 (2003) 1763–1767, https://doi.org/10.2210/pdb1p9u/pdb.
[20] H.R. Bairagya, A. Tasneem, G.P. Rai, S. Reyaz, Structural and dynamical impact of water molecules at substrate-or product-binding sites in human GMPR enzyme: a study by molecular dynamics simulations, J. Phys. Chem. B 125 (2020) 1351–1362, https://doi.org/10.1021/acs.jpcb.0c08818.
[21] H.R. Bairagya, B.P. Mukhopadhyay, A.K. Bera, Conserved water mediated recognition and the dynamics of active site Cys 331 and Tyr 411 in hydrated structure of human IMPDH-II, J. Mol. Recogn. 24 (2011) 35–44, https://doi.org/10.1002/jmr.1021.
[22] B.P. Mukhopadhyay, B. Ghosh, H.R. Bairagya, A.K. Bera, T.K. Nandi, S.B. Das, Modeling study of rusticyanin-cytochrome C4 complex: an insight to possible H-bond mediated recognition and electron—transfer process, J. Biomol. Struct. Dyn. 25 (2007) 157–164, https://doi.org/10.1080/07391102.2007.10507164.
[23] H.R. Bairagya, B.P. Mukhopadhyay, An insight to the dynamics of conserved water-mediated salt bridge interaction and interdomain recognition in hIMPDH isoforms, J. Biomol. Struct. Dyn. 31 (2013) 788–808, https://doi.org/10.1080/07391102.2012.712458.
[24] A. Banerjee, H.R. Bairagya, B.P. Mukhopadhyay, T.K. Nandi, D.K. Mishra, Conserved water mediated H-bonding dynamics of Ser117 and Thr119 residues in human transthyretin–thyroxin complexation: inhibitor modeling study through docking and molecular dynamics simulation, J. Mol. Graph. Model. 44 (2013) 70–80, https://doi.org/10.1016/j.jmgm.2013.04.010.
[25] B. Chakrabarti, H.R. Bairagya, P. Mallik, B.P. Mukhopadhyay, A.K. Bera, An insight to conserved water molecular dynamics of catalytic and structural Zn+ 2 ions in matrix metalloproteinase 13 of human, J. Biomol. Struct. Dyn. 28 (2011) 503–516, https://doi.org/10.1080/07391102.2011.10508591.
[26] A. Banerjee, H.R. Bairagya, B.P. Mukhopadhyay, T.K. Nandi, A.K. Bera, Structural Insight to Mutated Y116S Transthyretin by Molecular Dynamics Simulation, 2010. https://nopr.niscair.res.in/handle/123456789/10118.
[27] B.P. Mukhopadhyay, H.R. Bairagya, Protein folding: grand challenge of nature, J. Biomol. Struct. Dyn. 28 (2011) 637–638, https://doi.org/10.1080/073911011010524971.
[28] J. Wang, P. Morin, W. Wang, P.A. Kollman, Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA, J. Am. Chem. Soc. 123 (2001) 5221–5230, https://doi.org/10.1021/ja003834q.
[29] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J.Z. Zhang, T. Hou, End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design, Chem. Rev. 119 (2019) 9478–9508, https://doi.org/10.1021/acs.chemrev.9b00055.
[30] T. Lazaridis, Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory, J. Phys. Chem. B102 (1998) 3531–3541, https://doi.org/10.1021/jp9723574.
[31] J. Wahl, M. Smieško, Thermodynamic insight into the effects of water displacement and rearrangement upon ligand modifications using molecular dynamics simulations, ChemMedChem 13 (2018) 1325–1335, https://doi.org/10.1002/cmdc.201800093.
[32] C. Bissantz, B. Kuhn, M. Stahl, A medicinal chemist's guide to molecular interactions, J. Med. Chem. 53 (2010) 5061–5084, https://doi.org/10.1021/jm100112j.
[33] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.
[34] K. Sumathi, P. Ananthalakshmi, M.M. Roshan, K. Sekar, 3dSS, 3D structural superposition, Nucleic Acids Res. 34 (2006) W128–W132, https://doi.org/10.1093/nar/gkl036.

[35] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723, https://doi.org/10.1002/elps.1150181505.
[36] K. Ogata, S.J. Wodak, Conserved water molecules in MHC class-I molecules and their putative structural and functional roles, Protein Eng. Des. Sel. 15 (2002) 697–705, https://doi.org/10.1093/protein/15.8.697.
[37] H.R. Bairagya, D.K. Mishra, B.P. Mukhopadhyay, K. Sekar, Conserved water-mediated recognition and dynamics of NAD+ (carboxamide group) to hIMPDH enzyme: water mimic approach toward the design of isoform-selective inhibitor, J. Biomol. Struct. Dyn. 32 (2014) 1248–1262, https://doi.org/10.1080/07391102.2013.812982.
[38] H.R. Bairagya, B.P. Mukhopadhyay, S. Bhattacharya, Role of the conserved water molecules in the binding of inhibitor to IMPDH-II (human): a study on the water mimic inhibitor design, J. Mol. Struc.-Theochem. 908 (2009) 31–39, https://doi.org/10.1016/j.theochem.2009.04.037.
[39] D.K. Smith, P. Radivojac, Z. Obradovic, A.K. Dunker, G. Zhu, Improved amino acid flexibility parameters, Protein Sci. 12 (2003) 1060–1072, https://doi.org/10.1110/ps.0236203.
[40] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082, https://doi.org/10.1093/nar/gkx1037.
[41] V. Zoete, A. Daina, C. Bovigny, O. Michielin, SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening, J. Chem. Inf. Model. 56 (2016) 1399–1404, https://doi.org/10.1021/acs.jcim.6b00174.
[42] T. Sander, OSIRIS Property Explorer, Organic Chemistry Portal, 2001. https://www.organic-chemistry.org/prog/peo/.
[43] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 1–13, https://doi.org/10.1038/srep42717.
[44] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461, https://doi.org/10.1002/jcc.21334.
[45] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, https://doi.org/10.1002/jcc.21256.
[46] G.M. Morris, R. Huey, A.J. Olson, Using autodock for ligand-receptor docking, Curr. Protoc. Bioinf. 24 (2008) 8–14, https://doi.org/10.1002/0471250953.bi0814s24.
[47] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, https://doi.org/10.1016/0263-7855(96)00018-5.
[48] V. Zoete, M.A. Cuendet, A. Grosdidier, O. Michielin, SwissParam: a fast force field generation tool for small organic molecules, J. Comput. Chem. 32 (2011) 2359–2368, https://doi.org/10.1002/jcc.21816.
[49] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S.A. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217, https://doi.org/10.1002/jcc.540040211.
[50] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, NAMD2: greater scalability for parallel molecular dynamics, J. Comput. Phys. 151 (1999) 283–312, https://doi.org/10.1006/jcph.1999.6201.
[51] J. Huang, A.D. MacKerell Jr., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, J. Comput. Chem. 34 (2013) 2135–2145, https://doi.org/10.1002/jcc.23354.
[52] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell Jr., CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (2010) 671–690, https://doi.org/10.1002/jcc.21367.
[53] P. Manoharan, N. Ghoshal, Fragment-based virtual screening approach and molecular dynamics simulation studies for identification of BACE1 inhibitor leads, J. Biomol. Struct. Dyn. 36 (2018) 1878–1892, https://doi.org/10.1080/07391102.2017.1337590.
[54] A. Grossfield, S.E. Feller, M.C. Pitman, Convergence of molecular dynamics simulations of membrane proteins, Proteins: Struct. Funct. Bioinf. 67 (2007) 31–40, https://doi.org/10.1002/prot.21308.
[55] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms, J. Chem. Phys. 101 (1994) 4177–4189, https://doi.org/10.1063/1.467468.
[56] J. Pimthon, R. Willumeit, A. Lendlein, D. Hofmann, Membrane association and selectivity of the antimicrobial peptide NK-2: a molecular dynamics simulation study, J. Pept. Sci. 15 (2009) 654–667, https://doi.org/10.1002/psc.1165.
[57] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, 2017, https://doi.org/10.1093/oso/9780198803195.001.0001. Oxford Scholarship Online.
[58] E.D. Lopez, J.P. Arcon, D.F. Gauto, A.A. Petruk, C.P. Modenutti, V.G. Dumas, M. A. Marti, A.G. Turjanski, WATCLUST: a tool for improving the design of drugs based on protein-water interactions, Bioinformatics 31 (2015) 3697–3699, https://doi.org/10.1093/bioinformatics/btv411.
[59] D.F. Gauto, S. Di Lella, D.A. Estrin, H.L. Monaco, M.A. Martí, Structural basis for ligand recognition in a mushroom lectin: solvent structure as specificity predictor, Carbohydr. Res. 346 (2011) 939–948, https://doi.org/10.1016/j.carres.2011.02.016.
[60] D.J. Huggins, M. Marsh, M.C. Payne, Thermodynamic properties of water molecules at a protein–protein interaction surface, J. Chem. Theor. Comput. 7 (2011) 3514–3522, https://doi.org/10.1021/ct200465z.
[61] S. Di Lella, M.A. Martí, R.M.S. Álvarez, D.A. Estrin, J.C.D. Ricci, Characterization of the galectin-1 carbohydrate recognition domain in terms of solvent occupancy, J. Phys. Chem. B 111 (2007) 7360–7366, https://doi.org/10.1021/jp068989k.
[62] J.R. Wagner, J. Sørensen, N. Hensley, C. Wong, C. Zhu, T. Perison, R.E. Amaro, POVME 3.0: software for mapping binding pocket flexibility, J. Chem. Theor. Comput. 13 (2017) 4584–4592, https://doi.org/10.1021/acs.jctc.7b00500.
[63] H. Liu, T. Hou, CaFE: a tool for binding affinity prediction using end-point free energy methods, Bioinformatics 32 (2016) 2216–2218, https://doi.org/10.1093/bioinformatics/btw215.
[64] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041, https://doi.org/10.1073/pnas.181342398.
[65] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J.Z. Zhang, T. Hou, End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design, Chem. Rev. 119 (2019) 9478–9508, https://doi.org/10.1021/acs.chemrev.9b00055.
[66] W.E. Miranda, S.Y. Noskov, P.A. Valiente, Improving the LIE method for binding free energy calculations of protein–ligand complexes, J. Chem. Inf. Model. 55 (2015) 1867–1877, https://doi.org/10.1021/acs.jcim.5b00012.
[67] S.P. Kanaujia, K. Sekar, Structural and functional role of water molecules in bovine pancreatic phospholipase A2: a data-mining approach, Acta Crystallogr. D Biol. Crystallogr. 65 (2009) 74–84, https://doi.org/10.1107/S0907444908039292.
[68] O. Carugo, Correlation between occupancy and B factor of water molecules in protein crystal structures, Protein Eng. 12 (1999) 1021–1024, https://doi.org/10.1093/protein/12.12.1021.
[69] O. Carugo, D. Bordo, How many water molecules can be detected by protein crystallography? Acta Crystallogr. D Biol. Crystallogr. 55 (1999) 479–483, https://doi.org/10.1107/S0907444998012086.
[70] A.V. Bondi, van der Waals volumes and radii, J. Phys. Chem. 68 (1964) 441–451, https://doi.org/10.1021/j100785a001.
[71] S. Keretsu, S.P. Bhujbal, S.J. Cho, Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation, Sci. Rep. 10 (2020) 1–14, https://doi.org/10.1038/s41598-020-74468-0.