

# Binding mechanism of SARS-CoV-2 spike protein with human ACE2 receptor

Rajendra P. Koirala<sup>1</sup>, Bidhya Thapa<sup>2</sup>, Shyam P. Khanal<sup>1</sup>, Jhulan Powrel<sup>1</sup>,  
Rajendra P. Adhikari<sup>3</sup> & Narayan P. Adhikari<sup>1,\*</sup>

\*narayan.adhikari@cdp.tu.edu.np.

<sup>1</sup>Central Department of Physics, Tribhuvan University, Kathmandu, Nepal

<sup>2</sup>Padma Kanya Multiple Campus, Tribhuvan University, Kathmandu, Nepal

<sup>3</sup>Department of Natural Sciences, Kathmandu University, Nepal

## Abstract

SARS-CoV-2 virus interacts via C-terminal domain of spike protein to human cell receptor protein hACE2. Amino acid residues residing at the interface play vital role in binding of SARS-CoV-2 CTD to hACE2. The detailed atomic level investigation of interactions at binding interface of SARS-CoV-2 CTD/hACE2 provides indispensable information on better understanding of location for drug target. In the present work, we have studied the dynamical behaviour of the complex by analyzing the molecular dynamics (MD) trajectories. The major interacting residues of SARS-CoV-2 CTD and hACE2 have been identified by analyzing the non-bonded interactions such as hydrogen bondings, salt bridges, hydrophobic interactions, van der Waals interactions etc. Umbrella sampling method has been used to estimate the binding free energy for in-depth understanding of binding mechanism between virus protein and host receptor. The binding free energy difference, key residues at the interface, important atomic interactions and contact surface areas have been compared with the molecular complex of SARS-CoV and hACE2. Relatively larger contact surface area, more non-bonded interactions as well as greater binding free energy provide the evidence for favorable binding of SARS-CoV-2 with hACE2 receptor than SARS-CoV.

## Introduction

Corona virus disease (COVID-19) pandemic, caused by severe acute respiratory syndrome (SARS)-like coronavirus (SARS-CoV-2), is a serious health concern for the global community [1, 2, 3]. Although the origin of the virus is still unclear, it has been spread all over the world threatening the human civilizations after its initial outbreak from China in December 2019. Till date, more than 23 millions infected population has been reported globally and more than eight hundred thousand people have lost their lives [4]. There is no approved drug or vaccine against the COVID-19, even though several antiviral drugs have been proposed and are also in clinical trials [5].

SARS-CoV-2 has more than 70 percent of structural similarity with SARS-CoV; most of the residues at binding interface are similar [7, 6]. SARS-CoV-2 similar with SARS-CoV and also other coronavirus utilizes human angiotensin converting enzyme 2 (hACE2) receptor to enter into human cell. This entry process is mediated by the spike(s) glycoprotein which are embedded in the capsid of SARS-CoV-2 [8]. The spike protein (S) of SARS-CoV-2 in most cases is cleaved by host into the subunits S1 and S2 which are responsible for receptor recognition and membrane fusion respectively [9]. Similar to SARS-CoV, C-Terminal Domain(CTD) of S1 subunit of spike protein in SARS-CoV-2

acts as receptor binding domain (RBD) [10, 11]. Even though both SARS-CoV and SARS-CoV-2 have same binding domain, the binding affinity of SARS-CoV-2 is greater than SARS-CoV [7, 12].

Immediately after the COVID-19 outbreak, several researches have been carried out to identify the nature and location of binding of SARS-CoV-2 CTD with hACE2 using static crystal structure [12, 13]. To our best knowledge molecular dynamics (MD) study of SARS-CoV-2 CTD/hACE2 has not been carried out to understand the binding mechanism during dynamics and also to estimate the free energy differences. The free energy calculation provides in-depth insight on the binding mechanism between the protein molecules [14]. There are several experimental techniques of measuring binding free energy such as isothermal titration calorimetry (ITC) [15], fluorescence resonance energy transfer (FRET) [16], nuclear magnetic resonance (NMR) [17], surface plasmon resonance (SPR) [18] and many others; however, these experimental methods are not easily accessible and time consuming. The computational approach can be the best complement for large scale investigations [19, 20, 21].

Out of many computational approach, umbrella sampling is one of the widely used method for the calculation of free energy in large molecular system [22, 23]. It improves the sampling system by designing and implementing the biasing potentials as a function of reaction coordinates [24, 25]. If an energy barrier exists in between two regions of configuration states, there may be poor sampling, despite the long simulation run being carried out. The applied biasing potential bridges such configuration states and makes it easier in searching local or global minima, which can be considered as the structurally favorable state in the molecular complex [26].

In this study, we have carried out molecular dynamics (MD) simulations for the comprehensive study of binding mechanism of SARS-CoV-2 CTD with hACE2. In addition, umbrella sampling method has been executed to estimate the binding free energy of SARS-CoV-2 CTD/hACE2. Required windows for the umbrella sampling have been taken from steered molecular dynamics (SMD) [27] simulations. In SMD, SARS-CoV-2 CTD has been translated, taking the hACE2 as the reference molecule. The quantitative estimation of binding affinity between the targeted molecules facilitates in silico-drug designing. Then, we have performed comparative study of various non-covalent atomic interactions at the binding interface of SARS-CoV-2 and SARS-CoV with hACE2. Furthermore, the binding free energy and contact surface area are compared of these complexes.

## Results

MD simulations were performed to investigate the binding mechanisms of SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 complexes by analyzing non-bonded interactions at the interface, contact surface area measurement and binding free energy calculation. This will be valuable to understand the entry mechanism of the virus to host receptor hACE2 and to provide molecular level insights on dynamic behavior of the complex.

**Root Mean Square Deviation (RMSD).** We checked the thermodynamic states of both systems (SARS-CoV-2 CTD/hACE2 complex and SARS-CoV RBD/hACE2 complex) to ensure whether the systems are conformationally stable or not. During 10 ns equilibration run, both systems were found conformationally stable after 2 ns simulation run. Then, RMSD of the backbones of both complexes were plotted with respect to time

dependent function of MD simulation as shown in Figure 1. From Figure 1, it is seen that the average RMSD of SARS-CoV-2 CTD/hACE2 complex was observed slightly smaller than that of SARS-CoV RBD/hACE2 complex, i.e., the complex of SARS-CoV-2 CTD and hACE2 is more stable than the complex of SARS-CoV RBD and hACE2.

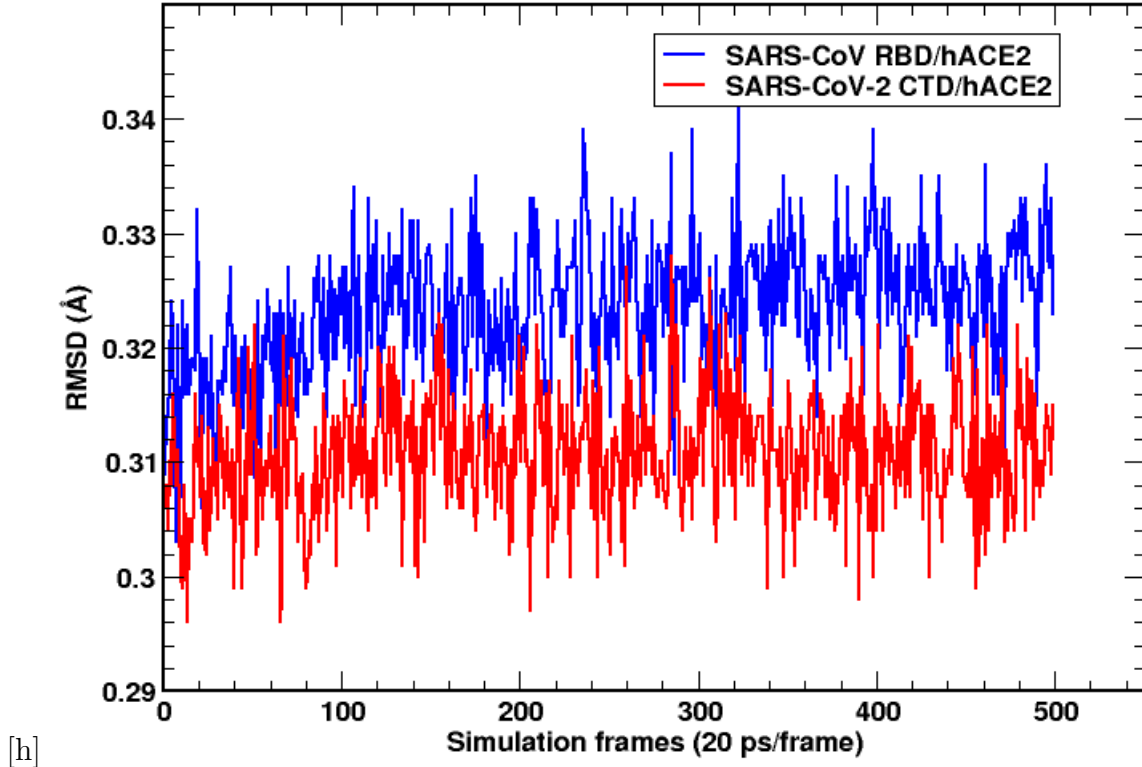


Figure 1: RMSD of SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2.

**Contact Surface Area.** We also analyzed the contact surface area between the spike protein CTD/RBD and hACE2 receptor using MD trajectories. Contact area is the surface buried at the interface between two proteins which contributes to bind and stabilize the protein-protein complexes. Larger contact surface indicates greater stability of the structure [28]. Figure 2 shows the contact surface area for SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 complexes. The contact surface area for SARS-CoV-2 CTD is more in comparison to SARS-CoV RBD indicating the greater binding affinity of SARS-CoV-2 with receptor. The estimated values of contact surface area are presented in Table 1. From the Table 1, it has been observed that SARS-CoV-2 CTD/hACE2 has larger contact area than SARS-CoV RBD/hACE2 by  $77.0194 \text{ \AA}^2$ .

Table 1: Calculation of contact surface area;  $A_1$ ,  $A_2$  &  $A_3$  are the average solvent accessible surface area (SASA) of hACE2, SARS-CoV-2 CTD and their complexes; and  $A'_1$ ,  $A'_2$  &  $A'_3$  are the average solvent accessible surface area (SASA) of hACE2, SARS-CoV RBD and their complexes.

Complex	SASA ( $\text{\AA}^2$ ) for				
	hACE2 ( $A_1$ )	SARS-CoV-2 CTD ( $A_2$ )	complex( $A_3$ )	$A_1 + A_2 - A_3$	Net contact area ( $A$ ) [28]
SARS-CoV-2 CTD/hACE2	29100.3613	11227.9717	38549.0195	1779.3135	889.6568
SARS-CoV RBD/hACE2	29092.6152	10859.5420	38326.8825	$A'_1 + A'_2 - A'_3$	Net contact area ( $A'$ )
				1625.2747	812.6374

**Non-bonded Interactions.** Furthermore, we studied in details the hydrogen bonds,

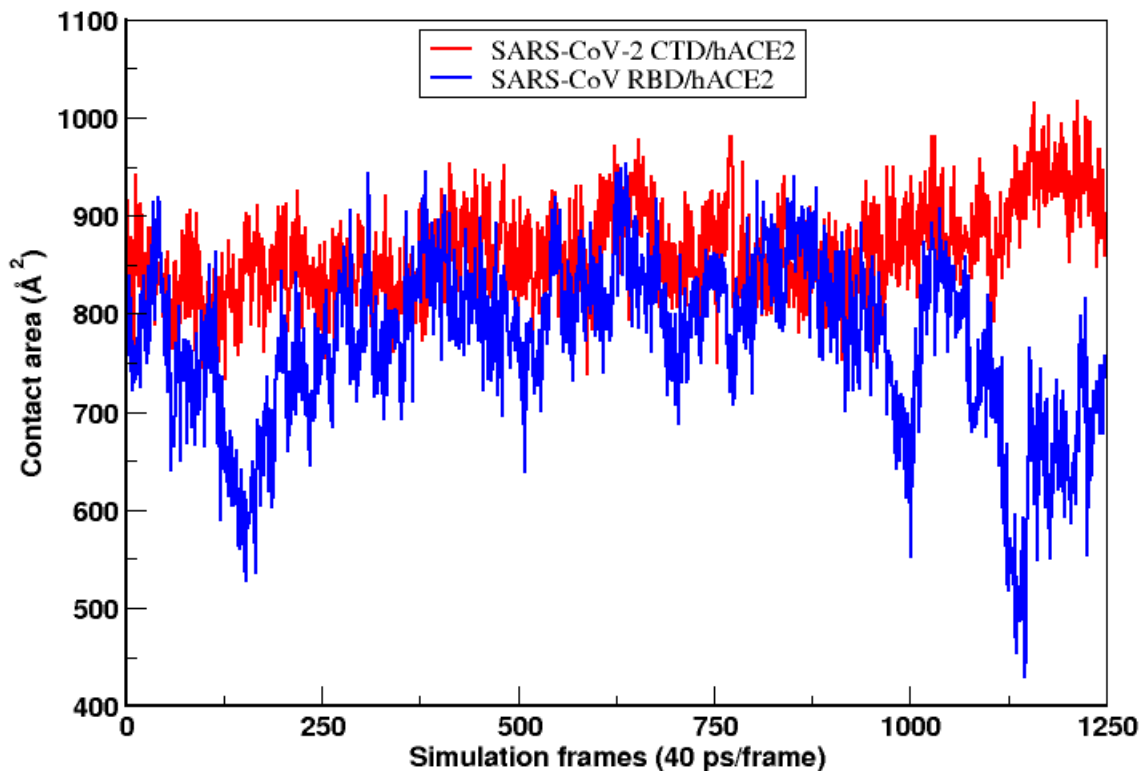


Figure 2: Comparison of contact surface area at the interface of SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2.

salt-bridges, hydrophobic, electrostatic and vdw interactions between the residues residing at the interface between SARS-CoV-2 CTD/hACE2; and also compared with SARS-CoV RBD/hACE2.

**Hydrogen Bonds:** Wang et al. (2020) and Lan et al. (2020) have studied the atomic interactions at the interface of static crystal structure of SARS-CoV-2 CTD/hACE2 complex [12, 13], whereas we have investigated the hydrogen bonds at the interface of two complexes by analyzing the MD simulations trajectories. The cut-off distance for hydrogen bond was taken 3.5 Å [12]. We monitored the time evolution of number of hydrogen bonds formed at the interface between SARS-CoV-2 CTD/hACE2 and also compared with that of SARS-CoV RBD/hACE2 as shown in Figure 3 (Also see supplementary Table S1). Hydrogen bonds were found to be consistently working throughout the simulation. All potential interfacial hydrogen bonds formed in the SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 during the 100 ns of simulations are as shown in Table 2. Total hydrogen bonds formed during the simulations were seen to be more in case of SARS-CoV; however the strength and life time of potential hydrogen bonds were found to be greater in case of SARS-CoV-2. Unlike SARS-CoV-2 CTD, some residues of SARS-CoV RBD form hydrogen bonds with the carbohydrates attached to the N-terminal domain of hACE2. The bond formed between VAL404 of SARS-CoV RBD with CARA-BMAN3 is significant.

There is no hydrogen bond formed between SER19 of hACE2 with any residues of SARS-CoV RBD in static structures [12, 13], however our results show two potential hydrogen bonds formed between main chain and side chain of SER19 of hACE2 with side chain of ASP463 of SARS-CoV RBD as in Figure 4(a). The hydrogen bonds formed at three key

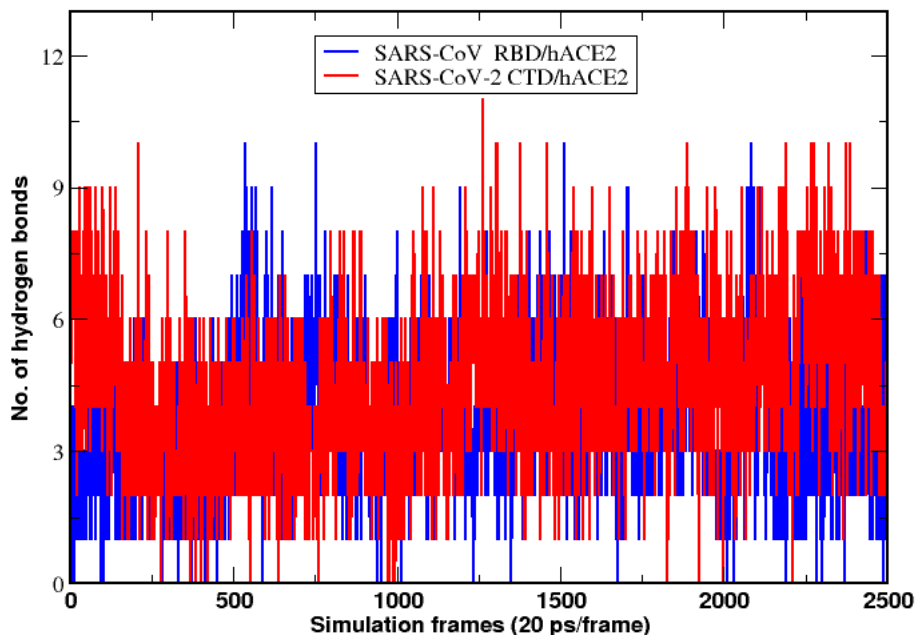


Figure 3: Comparison of time evolution of number of hydrogen bonds at the interfaces of SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2.

Table 2: Comparison of hydrogen bond occupancy percentage in SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2.

H-bond occupancy percentage	SARS-CoV-2 CTD/hACE2	SARS-CoV RBD/hACE2
< 1	47	80
1-10	19	22
10-20	1	2
20-30	2	4
30-40	4	1
40-50	1	1
50-60	2	1

region of interface between SARS-CoV RBD/hACE2 at 0<sup>th</sup> frame are shown in Figures 4(a), 4(b) and 4(c). We have observed some difference in the atomic interactions at the interface of both virus proteins and hACE2 than that of static crystal structure. The two different approaches might be the reason of variation in the number of interactions.

At the binding interface of SARS-CoV-2 CTD and hACE2 receptor, there are three key regions where most of the polar contacts between the interfacial residues take place. In region one, SER19, GLN24 and TYR83 of hACE2 form hydrogen bonds with ALA475 and ASN487 of SARS-CoV-2 CTD as in Figure 5(a) (also see supplementary Figures: S1(a), S2(a), S3(a) & S4(a)). In region two, there are interactions between the residues LYS417, TYR453 and GLN493 of SARS-CoV-2 CTD forming hydrogen bonds with ASP30, LYS31, HIS34 and GLU35 of hACE2 Figure 5(b) (also see supplementary Figures: S1(b), S2(b), S3(b) & S4(b)). Similarly, in region three, there is extensive network of hydrogen bonds between SARS-CoV-2 CTD residues GLY446, TYR449, GLY496, GLN498, THR500, ASN501, GLY502 and TYR505 with the hACE2 residues GLU37, ASP38, TYR41, GLN42, LYS353, ASP355 and ARG393 Figure 5(c) (also see supple-

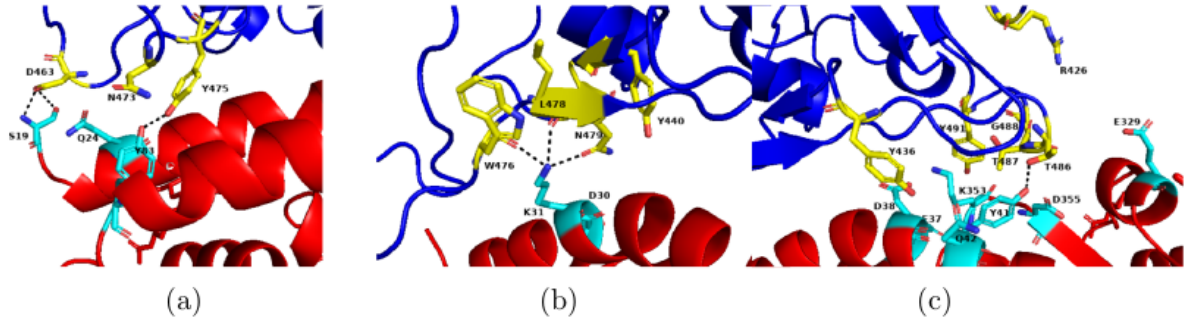


Figure 4: Details of hydrogen bondings at the three key contact sites on binding interface between SARS-CoV RBD and hACE2 receptor. hACE2 receptor is shown in red and SARS-CoV RBD is shown in blue color. The contact residues which can form potential hydrogen bonds are shown as sticks & labelled by 1 residue letter (for 0<sup>th</sup> frame).

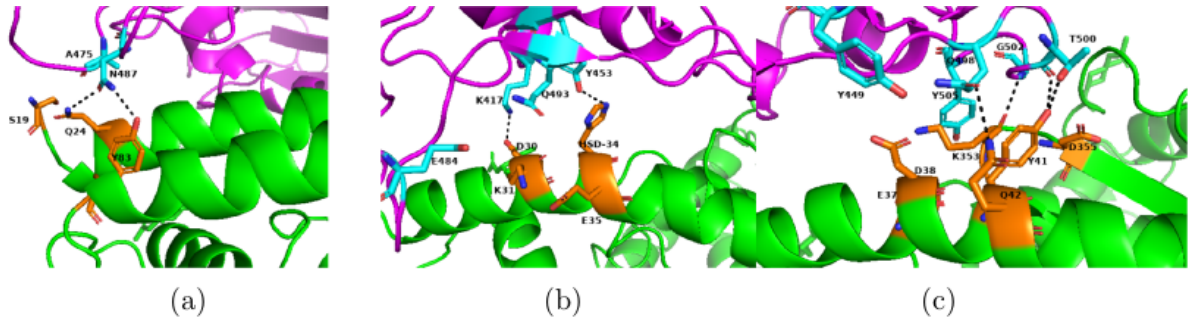


Figure 5: Details of hydrogen bondings at the three key contact sites on binding interface between SARS-CoV-2 CTD and hACE2 receptor. hACE2 receptor is shown in green and SARS-CoV-2 CTD is shown in magenta color. The contact residues which can form potential hydrogen bonds are shown as sticks & labelled by 1 residue letter (for 0<sup>th</sup> frame).

mentary Figures: S1(c), S2(c), S3(c) & S4(c)). Because of dynamical nature of our system, there is continuous breaking and formation of hydrogen bonds during the simulations. The variation of hydrogen bonds formed during the five representative frames of 100 ns simulation of SARS-CoV-2 has been shown in Figure 5 and supplementary Figures S1-S4.

**Salt-bridges:** In addition to extensive network of interfacial hydrogen bonds, another important contribution to protein-protein binding comes from salt-bridge interactions. MD trajectory analysis has shown four salt-bridges, having different bond length and strength, formed at the interface of SARS-CoV-2 CTD/hACE2. The salt-bridge formed between the residue LYS417 of SARS-CoV-2 CTD with ASP30 of hACE2 is found to be the strongest one among them owing to its short bond length. The remaining residues GLU484, LYS458 and ARG403 of SARS-CoV-2 CTD have formed salt-bridges with LYS31, GLU23 and GLU37 of hACE2 respectively. In contrast, we found only one salt-bridge formed between ARG426 of SARS-CoV RBD and GLU329 of hACE2 which is much weaker than that of SARS-CoV-2 because of large bond length as in Figure 6. The mean contact score of each salt-bridge formed at both interfaces calculated using Pycontact (see supplementary Figure S5).

**Hydrophobic Interactions:** Hydrophobic patch created in the interface between SARS-CoV-2 CTD and hACE2 contributes to enhance the binding of protein complexes. In

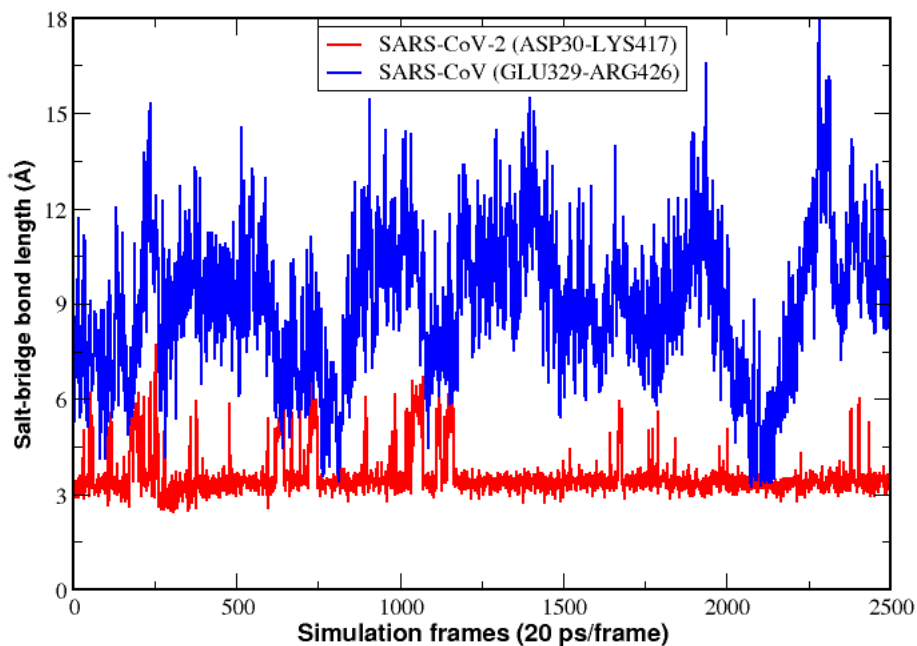


Figure 6: Comparison of time evolution of salt-bridge bond length at the interface of SARS-CoV RBD/hACE2 and SARS-CoV-2 CTD/hACE2.

SARS-CoV-2 CTD, PHE486 and TYR489 have hydrophobic contacts with the residues MET82, LEU79 and TYR89 of hACE2 (see supplementary Figure S6). It has been observed that PHE486 inserts into the hydrophobic groove created by the MET82, LEU79 and TYR89. This residue in SARS-CoV RBD is replaced by LEU472 being oriented in another direction has very weak interactions as compared to PHE486. Moreover, TYR489 has formed close and stable contact with PHE28 of hACE2. On the contrary, in SARS-CoV RBD only one residue TYR484 has formed strong and persistent hydrophobic contact with the TYR41 of hACE2 (see supplementary Figure S6).

**Electrostatic and van der Waals (vdw) Interactions:** The electrostatic and van der Waals (vdw) interactions in two complexes have been studied using NAMD Energy plugin package in VMD. Figure 7 depicts the comparative analysis of energy due to electrostatic and vdw interactions as a function of time for the two complexes. In the beginning of simulations, the electrostatic contributions of SARS-CoV-2 CTD/hACE2 was distinctly higher than SARS-CoV RBD/hACE2, however after 16 ns simulation time lapsed, the contributions were almost equal. In addition, the potential energy contributed by vdw interactions were consistently almost equal for both the systems throughout the simulations. It reveals that electrostatic and vdw interactions are almost equally contributed in binding both the complexes.

**Free Energy.** To investigate the energetic difference in binding of hACE2 with SARS-CoV-2 CTD and SARS-CoV RBD, the free energy differences have been estimated using Umbrella sampling technique. Umbrella windows were taken from the trajectories of SMD simulations. The interactions between the molecules in SARS-CoV-2 CTD/hACE2 were found terminated after traversing 9 Å distance away from the original position. To incorporate all interacting pathways, ten windows with 1 Å distance separation were taken for every successive window. On the other hand, the interactions between the

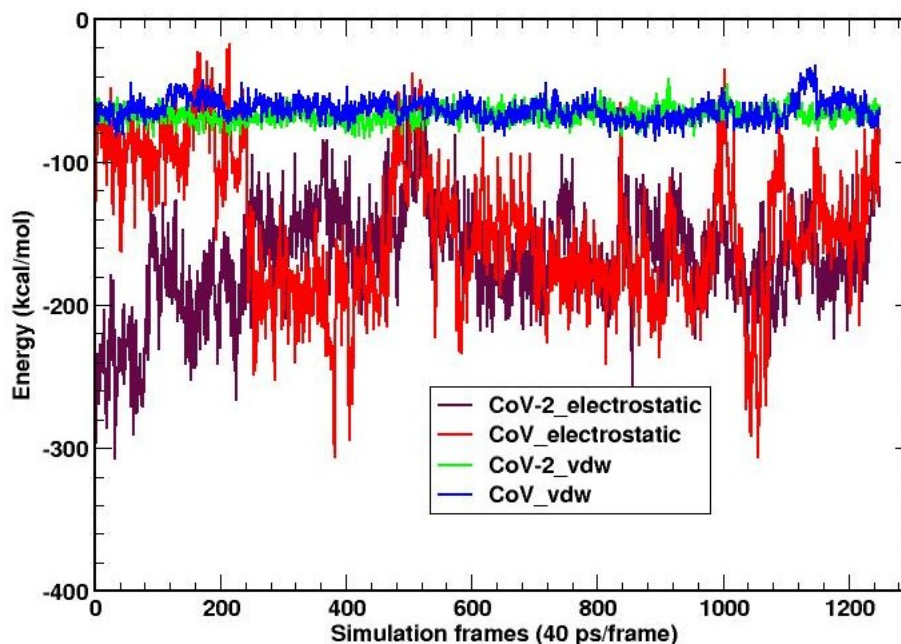


Figure 7: Comparison of electrostatic and vdw interactions of SARS-CoV-2 CTD and SARS-CoV RBD with hACE2.

molecules in SAR-CoV RBD/hACE2 were found ceased after traversing  $5 \text{ \AA}$  distance from the original position. Therefore, six windows were prepared separating  $1 \text{ \AA}$  distance away for every successive window. Figure 8 shows the change in free energy during the translation of virus CTD from active pocket of hACE2 for both complexes. The center of mass (COM) distance as a reaction coordinate allows us to track the free-energy changes for SARS-CoV-2 CTD in complex with hACE2 and SAR-CoV RBD in complex with hACE2 and compare the differences in the binding affinity for the two complexes. The SARS-CoV-2 CTD in complex with hACE2 is found to have the greater binding free energy of  $\sim 1.91 \text{ kcal/mol}$  compared to the SAR-CoV RBD in complex with hACE2. This, as well as the nature of the free-energy curve, provides an insight on binding mechanisms of the complexes.

## Discussion

COVID-19 pandemic has seriously threatened public health throughout the globe. Since there is no approved drug till date to combat the people against the SARS-CoV-2 virus, more comprehensive study is essential through the various aspects at molecular level. The fundamental necessity is to understand the entry mechanism of the virus into the human cell, which is really helpful to discover the drug against the virus. To deal the entry mechanisms and dynamical characteristics of the virus cell in complex with hACE2 receptor, we used various computational techniques. C-Terminal Domain (CTD) of S1 subunit of spike protein, being the active interacting region, has been taken into consideration in SARS-CoV-2. We performed the comparative analysis of the key residues and atomic interactions responsible for the binding of the SARS-CoV-2 CTD and SARS-CoV RBD with human ACE2 receptor.

Estimation of structural variation during the simulation is the foremost judgement of molecular stability in molecular dynamics study. RMSD is the measure of stability of molecular structure in the cellular environment. Well equilibrated system with consistent



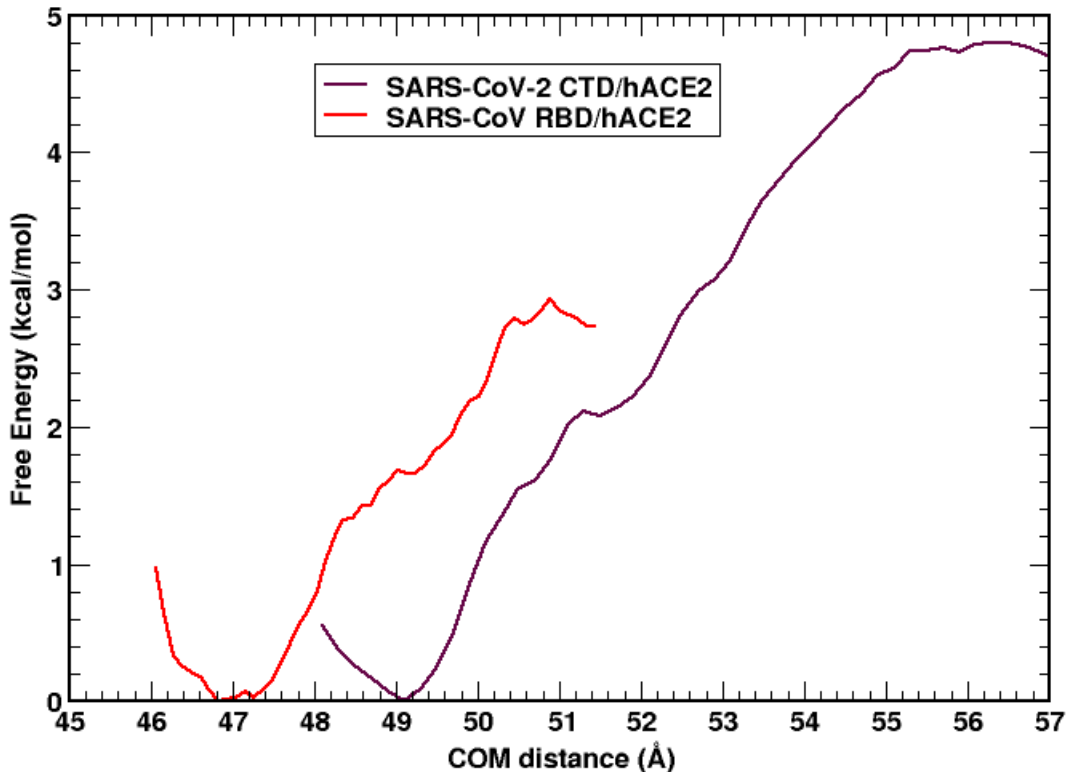


Figure 8: Free energy curve for SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 during the translation of spikes CTD/RBD from hACE2.

RMSD ensures us to proceed for the further study of binding affinity and energy variations of the molecular complexes [29]. Moreover, contact surface area between the molecules identifies the binding strength of the complex. Therefore, we have obtained the contact surface area of both complexes calculating the SASA. SASA has been determined from time evolution data generated from the 100 ns NVT run. Then, average value of contact area for both the systems have been presented in Table 1 and are interpreted graphically in Figure 2. Larger contact surface area in SARS-CoV-2 CTD/hACE2 complex depicts the stronger binding of this complex than that of SARS-CoV RBD/hACE2 [30].

Our results show considerable similarity in the binding sites, interfacial residues and important atomic interactions in both viral protein receptor binding domain (i.e., SARS-CoV-2 CTD and SARS-CoV RBD). However, there is some structural variation in a loop between two structures in the binding region and some residues at the binding sites are different. This facilitates more and stronger atomic contacts between SARS-CoV-2 CTD and hACE2 interface and thereby enhancing its binding affinity. Polar residues residing at the interface form an extensive network of hydrogen bonds and salt-bridge interactions [31, 32, 33, 34]. Our study reveals that interfacial hydrogen bonds, salt-bridges and hydrophobic interactions play an important role in the binding of SARS-CoV-2 CTD to host cell receptor. Furthermore, comparative analyses of the binding mechanism of two

viral proteins with hACE2 show that binding affinity of SARS-CoV-2 is greater than that of SARS-CoV. Notably, more residues are engaged in the binding of SARS-CoV-2 CTD with hACE2. We find the greater number of potential hydrogen bonds formed in the case of SARS-CoV-2 CTD which contributes to higher binding affinity. More and stronger salt-bridges formed in case of SARS-CoV-2/hACE2 establish stronger binding to the receptor. Additionally, we observe hydrophobic interactions are stronger in case of SARS-CoV-2 which also contribute to enhanced binding.

The contributions of electrostatic and vdw contacts are significant to form a stable protein-protein complex [35, 36]. The potential energy in binding the virus CTD/RBD and host receptor are compared in both the systems. Though, initially the electrostatic energy is observed relatively larger in SARS-CoV-2 CTD/hACE2 than that of SARS-CoV RBD/hACE2, the dynamical results show almost equal contributions in both the complexes. This shows that the contributions of hydrogen bonds, salt bridges and hydrophobic interactions are responsible to provide the greater binding strength in SARS-CoV-2 CTD/hACE2.

The binding mechanisms of the complexes are further analyzed estimating free energy differences from umbrella sampling method. SMD trajectories are taken for the appropriate samples that ensure the sufficient overlapping on windows [37]. In SMD, the virus CTD/RBD are pulled upto that distance, beyond which no interactions persists. We find the interactions of molecules in complex SARS-CoV RBD/hACE2 have been terminated after the displacement of RBD by 5 Å from host receptor, whereas the interactions sustain upto 9 Å displacement from the initial position in SARS-CoV-2 CTD/hACE2. Comparisons of free energy of two complexes has provided the insight of bonding affinity between the virus CTD/RBD and hACE2 molecules. The greater free energy difference between SARS-CoV-2 CTD in complex with hACE2 depicts the stronger binding strength than the complex of SARS-CoV RBD and hACE2. As the further investigation, we plan to calculate the solvation free energy of SARS-CoV-2 and SARS-CoV molecule in the aqueous environment.

## Methods

**System Set Up.** Two molecular structures, PDB IDs 6LZG and 2AJF, were taken for the molecular dynamics simulations. The PDB ID 6LZG contains the complex of SARS-CoV-2 CTD and hACE2 receptor protein (i.e., SARS-CoV-2 CTD/hACE2 complex) and that of PDB ID 2AJF contains the complex of SARS-CoV RBD and hACE2 receptor protein (i.e., SARS-CoV RBD/hACE2 complex) [38]. CHARMM-GUI [39] was used to create the pdb and psf files of these complexes. Then, both the complex structures were solvated using TIP3P [40] water and electrically neutralized by adding NaCl. A cubical box of dimensions  $144 \times 144 \times 144 \text{ \AA}^3$  was prepared for NVT simulation of the complex SARS-CoV-2 CTD/hACE2 and another cubical box of dimensions  $131 \times 131 \times 131 \text{ \AA}^3$  was prepared for NVT simulation of the complex SARS-CoV RBD/hACE2. Further, two equal sized orthorhombic simulation boxes were prepared in order to estimate the free energy differences of above complexes by changing the dimensions to  $250 \times 90 \times 90 \text{ \AA}^3$  and electrically neutralized by adding NaCl.

**Molecular Dynamics Simulation.** All atom molecular dynamics (MD) simulations were performed using NAMD [27] simulation package. The CHARMM36m [41] force

field was used for each simulations. The Particle Mesh Ewald (PME) [42] was used for the long-range interactions with a 12.0 Å non-bonded cut-off. The energy minimization was performed for 10,000 steps, using the conjugate gradient and line search algorithm [43, 44]. The system was then equilibrated at 310 K for 10 ns with harmonically restrained heavy atoms taking 1 fs time step. Finally, the production run was propagated for 100 ns simulation under NVT conditions by using Langevin dynamics with a damping constant of 1 ps<sup>-1</sup> taking time step of 2 fs.

**Steered Molecular Dynamics and Umbrella Sampling.** To perform the umbrella sampling, sample windows were chosen from steered molecular dynamics (SMD) trajectories. During SMD, CTD/RBD of SARS-CoV-2 CTD/SARS-CoV RBD were pulled correspondingly towards the negative x-direction with constant velocity pulling method of velocity 0.00005 Å/fs. In this process, the alpha carbons of hACE2 protein were taken as the fixed atoms and alpha carbons in CTD/RBD part of spike protein of the systems were taken as the dummy atoms; and were pulled from their center of mass (COM) along the negative x-direction with constant velocity ( $\vec{v} = d\vec{x}/dt$ ) in water and ions environment so that the SMD atom experiences the force  $F(\vec{t}) = k(\vec{v} t - \Delta\vec{x})$ , providing the external potential energy,

$$U(x, t) = \frac{1}{2}k(\vec{v} t - \Delta\vec{x} \cdot \hat{n})^2 \quad (1)$$

where,  $k$  ( $= 5 \text{ kcal mol}^{-1} \text{Å}^{-2}$ ) is the spring constant and gives the stiffness of the applied harmonic restraining force, and  $\Delta\vec{x}(t) = \vec{x}(t) - \vec{x}_0$ , is the displacement of SMD molecules from initial position  $\vec{x}_0$  to instantaneous position  $\vec{x}(t)$  and  $\hat{n}$  is the unit vector along the direction of pulling.

Umbrella sampling was performed to investigate the free energy difference during the translation of SARS-CoV-2 CTD from hACE2 protein for system SARS-CoV-2 CTD/hACE2; and identical condition is applied for system SARS-CoV RBD/hACE2. SMD trajectories were used to select the appropriate windows. Identifying the information on the termination of molecular interactions from SMD, we estimated the number of umbrella windows in both the systems. Ten windows were prepared in SARS-CoV-2 CTD/hACE2 and six windows were prepared for SARS-CoV RBD/hACE2 complexes. Every successive window was taken from the SMD trajectories during the translation of 1 Å along the negative x-direction. The window size ensures the sufficient overlapping of successive windows to cover the entire reaction coordinate space. The reaction coordinate was chosen as the distance between the center of mass (COM) of hACE2 and CTD/RBD spike along the negative x-axis. To make the necessary overlapping reaction coordinates, a bias potential  $V(x)$  was used to force the system to fluctuate in coordinate space, which is given by,

$$V(x) = \frac{1}{2}k_i(x - x_0)^2 \quad (2)$$

where  $x_0$  is the harmonic constraint defining a center of window  $i$  ( $i = 1$  to 10 for SARS-CoV-2 and 1 to 6 for SARS-CoV), and force constant  $k_i$  is the window width. We used the force constant of  $1.5 \text{ kcal mol}^{-1} \text{Å}^{-2}$ .

**Data Analysis.** Visual Molecular Dynamics (VMD) [45] and Pymol [46] were used to visualize as well as generate images of the complex structures. VMD analysis tools were

used to identify and analyze non-bonded interactions by using the simulation trajectories. The NAMD energy plugin, available in VMD, was used to calculate the non-bonded interaction energy contributions. Pycontact [47] software package was used to analyze the hydrophobic interactions and salt bridges between the targeted protein residues in CTD/RBD of spike protein and ACE2. Weighted Histogram Analysis Method (WHAM) program [48] was used to estimate the free energy from umbrella sampling simulation. The free energy calculation of large molecular system is generally computationally demanding. This method minimizes the statistical errors as well as increases the efficiency of computational simulation. Moreover, it has the advantage of multiple overlapping of probability distributions for obtaining better estimation of free energy calculations.

## References

- [1] Yan, R. et al. Structural basis for the recognition of the 2019-nCoV by human ACE2. *Science* **367**, 1444–1448 (2020).
- [2] Guan, W. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
- [3] Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* (2020).
- [4] <https://covid19.who.int/table>.
- [5] Folegatti, P. M. et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet* (2020).
- [6] Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* (2020).
- [7] Shang, J. et al. Clinical characteristics of coronavirus disease 2019 in China. *Nature* **581**, 221–224 (2020).
- [8] Lu, G., Wang, Q. & Gao, G. F. Bat-to-human: spike features determining host jump of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* **23**, 8, 468–478 (2015).
- [9] Lai M. Coronaviridae. *Fields virology* 1305–1318 (2007).
- [10] Lu, G. et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 7461, 227–231 (2013).
- [11] Li, F. et al. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 5742, 1864–1868 (2005).
- [12] Wang, Q. et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* (2020).
- [13] Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 7807, 215–220 (2020).

- [14] Peng, Y. et al. Predicting protein–DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver. *Bioinformatics* **34**, 5, 779–786 (2018).
- [15] Velázquez-Campoy, A et al. Isothermal titration calorimetry. *Curr. Protoc. Cell Biol.* **23**, 1, 17–8 (2004).
- [16] Hillisch, A., Lorenz, M. & Diekmann, S. Recent advances in FRET: distance determination in protein–DNA complexes. *Curr. Opin. Struct. Biol.* **11**, 2, 201–207 (2001).
- [17] Campagne, S., and Gervais, V. & Milon, A. Nuclear magnetic resonance analysis of protein–DNA interactions. *J. R. Soc. Interface.* **8**, 61, 1065–1078 (2011).
- [18] Teh, H., et al. Characterization of protein- DNA interactions using surface plasmon resonance spectroscopy with various assay schemes. *Biochemistry* **46**, 8, 2127–2135 (2007).
- [19] Donald, J. E., Chen, W. W. & Shakhnovich, E. I. Energetics of protein–DNA interactions. *Nucleic Acids Res.* **35**, 4, 1039–1074 (2007).
- [20] Jones, S. et, al. Protein-DNA interactions: a structural analysis. *J. Mol. Biol.* **287**, 5, 877–896 (1999).
- [21] Zhang, C. et, al. A knowledge-based energy function for protein- ligand, protein-protein, and protein- DNA complexes. *J. Med. Chem.* **48**, 7, 2325–2335 (2005).
- [22] Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 2, 187–199 (1977).
- [23] Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 6, 932–942 (2011).
- [24] Zheng, L., Chen, M. & Yan W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci.* **105**, 51, 20227–20232 (2008).
- [25] Yang, M., et al. Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface. *J. Chem. Phys.* **141**, 4, 07B618.1 (2014).
- [26] Yang, Y., et al. Efficient sampling over rough energy landscapes with high barriers: A combination of metadynamics with integrated tempering sampling. *J. Chem. Phys.* **144**, 9, 094105 (2016).
- [27] Phillips, J. et, al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 16, 1781–1802 (2005).
- [28] Zou, X. et al. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res.* **40**, 6, 2747–2758 (2012).

- [29] Kufareva, I. & Ruben, A. Methods of protein structure comparison. *Homology Modeling*, 231–257 (2011).
- [30] Ma, B. et al. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci.* **100**, 10 5772–5777 (2003).
- [31] Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* **2**, 7 527–541 (2003).
- [32] Chen, J., Sawyer, N. & Rayan, L. Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Science* **22**, 4 510–515 (2013).
- [33] Jones, S. & Thornton, J. M. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**, 1 121–132 (1997).
- [34] Xu, D., Tsai, C. & Nussinov, R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* **10**, 9 999–1012 (1997).
- [35] Jones, S. et al. Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Res.* **31**, 24, 7189–7198 (2003).
- [36] DiStasio Jr, R. A., Gobre, V. V. & Tkatchenko, A. Many-body van der Waals interactions in molecules and condensed matter. *J. Phys. Condens. Matter* **26**, 21 213202 (2014).
- [37] Isralewitz, B. et al. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* **19**, 1, 13–25 (2001).
- [38] Berman, H. et, al. The protein data bank. *Nucleic Acids Res.* **28**, 1, 235–242 (2000).
- [39] Lee, J. et, al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**, 1, 405–413 (2016).
- [40] Jorgensen, W. et, al. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 2, 926–935 (1983).
- [41] Huang, J. et, al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **14**, 1, 71–71 (2017).
- [42] Harvey, M. J., Giupponi, G. & Fabritiis, G. De. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *Nature methods* **5**, 6, 1632–1639 (2009).
- [43] Khanal, S. P., Kandel, Y. P. & Adhikari, N. P. Transport properties of zwitterion glycine, diglycine, and triglycine in water. *AIP Advances* **9**, 6, 065303–10 (2019).
- [44] Koirala, R. P. , Bhusal, H. P. , Khanal, S. P. & Adhikari, N. P. Effect of temperature on transport properties of cysteine in water. *AIP Advances* **10**, 2, 025122 (2020).

- [45] Humphrey, W. et al. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 1, 33–38 (1996).
- [46] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
- [47] Scheurer, M. et al. PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. *Biophysical journal* **114**, 3, 577–583 (2018).
- [48] Kumar, S. et al. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 8, 1011–1021 (1992).

**Acknowledgements:** RPK & SPK acknowledge the partial financial support from Nepal Academy of Science and Technology (NAST). NPA acknowledges the UGC Award no. CRG-73/74-S&T-01 and TWAS research grants RG 20-316. We acknowledge the computing facilities of Supercomputer Centre Kathmandu University, which was established with equipment donated by CERN and the Arkansas High Performance Computing Center which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission.

**Competing Interests:** The authors declare that they have no competing financial interests.