



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# In-Silico analysis reveals lower transcription efficiency of C241T variant of SARS-CoV-2 with host replication factors MADP1 and hnRNP-1

Armi Chaudhari<sup>a</sup>, Minal Chaudhari<sup>a</sup>, Sapna Mahera<sup>a</sup>, Zuber Saiyed<sup>a</sup>, Neelam M. Nathani<sup>a</sup>, Shantanu Shukla<sup>b</sup>, Dhaval Patel<sup>c</sup>, Chirag Patel<sup>d</sup>, Madhvi Joshi<sup>a,\*</sup>, Chaitanya G. Joshi<sup>a,\*\*</sup>

<sup>a</sup> Gujarat Biotechnology Research Centre (GBRC), Sector- 11, Gandhinagar - 382011, Gujarat, India

<sup>b</sup> Genome Science and Technology, The University of Tennessee, Knoxville, TN - 37996, United States

<sup>c</sup> Department of Bioinformatics and Structural Biology, Indian Institute of Advanced Research, Gandhinagar, 382426, Gujarat, India

<sup>d</sup> Department of Botany, Bioinformatics and Climate Change Impacts Management, University School of Sciences, Gujarat University, Ahmedabad, 380009, India

## ARTICLE INFO

### Keywords:

MADP1  
hnRNP1  
5'UTR  
C241T  
SARS-CoV-2  
Protein-RNA complexes  
MD simulations  
MMPBSA

## ABSTRACT

Novel severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) has claimed more than 3.3 million lives worldwide and still counting. As per the GISAID database, the genomics of SARS-CoV-2 has been extensively studied, with more than 500 genome submissions per day. Out of several hotspot mutations within the SARS-CoV-2 genome, recent research has focused mainly on the missense variants. Moreover, significantly less attention has been accorded to delineate the role of the untranslated regions (UTRs) of the SARS-CoV-2 genome in the disease progression and etiology. One of the most frequent 5' UTR variants in the SARS-CoV-2 genome is the C241T, with a global frequency of more than 95 %. In the present study, the effect of the C241T mutation has been studied with respect to the changes in RNA structure and its interaction with the host replication factors MADP1 Zinc finger CCHC-type and RNA-binding motif 1 (hnRNP1). The results obtained from molecular docking and molecular dynamics simulation indicated weaker interaction of C241T mutant stem-loops with the host transcription factor MADP1, indicating a reduced replication efficiency. The results are also correlated with increased recovery rates and decreased death rates of global SARS-CoV-2 cases.

## 1. Introduction

COVID-19 pandemic has affected over 40 million individuals globally and has been fatal for nearly 3 million people [1]. SARS-CoV-2 is a ssRNA virus belonging to the *Coronaviridae* family. This family of viruses comprises the lethal viruses like SARS, MERS, and SARS-CoV-2 that causes infections in the respiratory tract [2].

The host-pathogen interaction that follows an initial viral infection results in a struggle to survive between both the host and the pathogen and can result in counter-mechanisms that enable one to prevail and the other to perish. A majority of ssRNA viruses contain their genome replication through 5' cap or have some secondary structures of RNA forming internal ribosomal entry sites (IRES) to recruit host replication factors [3,4]. For viral replication, the 5' untranslated region (UTR) makes a substantial contribution by acting as the main lever that regulates various host replication initiation factors in the RNA viruses, especially in picornaviruses [5]. Another mechanism involves non-AUG

(methionine) initiation wherein a virus targets eukaryotic initiation factor-2 (eIF2), inhibiting the host cell protein synthesis. Additionally, some viruses use the CUG (leucine) initiation mode to overcome the stress of specific antigenic peptides [6]. Furthermore, some viruses employ an alternative for translation initiation, like eIF2-independent shutoff mechanism whereby these viruses use the formation of downstream helix after initiation codon for efficient translation of viral mRNA [7]. These are some of the most important mechanisms by which viruses utilize the host factors for their replication and translation. Thus, these targets can be efficient candidates for preventing viral infections by shutting down the viral translation machinery [8]. SARS-CoV-2 likely performs its replication through a 5' cap-dependent mechanism [9]. Nonstructural protein 16 (nsp16) associates with Nsp 10 to form heterodimer which performs two important functions for viral proliferation in the host, One methylates 5' end of mRNA encoded by virus and second 2' O- methylation of very first nucleotide of viral mRNA [10].

SARS-CoV-2, after entry into the human host through association

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [madhvimicrobio@gmail.com](mailto:madhvimicrobio@gmail.com), [jd1-gbrc@gujarat.gov.in](mailto:jd1-gbrc@gujarat.gov.in) (M. Joshi), [dir-gbrc@gujarat.gov.in](mailto:dir-gbrc@gujarat.gov.in) (C.G. Joshi).

<https://doi.org/10.1016/j.imu.2021.100670>

Received 11 June 2021; Received in revised form 7 July 2021; Accepted 13 July 2021

Available online 21 July 2021

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

with ACE2 (angiotensin converting enzyme) receptors, enters the lung epithelial cells and injects its RNA into the host cytoplasm [11,12]. Full-length negative-sense RNA copies are generated at the initial stage of the infection, which further acts as a template for synthesizing positive-sense genomic RNA [2]. During the discontinuous replication process of the viral genome, RNA stem-loop (SL) structures SL1, SL2, TLR-S (Translation leading region) play an important role by interacting with host RNA binding proteins like the zinc finger CCHC-type and the RNA-binding motif 1 MADP1. These interactions play an important role in viral replication and translation [13–16]. 5' UTR of SARS-CoV-2 have various stem loops reported as SL1, SL2, SL3, SL4A, SL4B and SL5 [17]. With respect to SARS-CoV-1, SL1 contains a bipartite sequence possibly involved in the fine-tuning of viral RNA interaction with the host protein. SL2 contains a highly conserved loop sequence of 5'-CUUCU (N)-3' among most of the coronavirus species, which is likely to form a tetra loop-like structure [18,19]. SL3 is conserved among the group of gamma and beta coronavirus. SL3 comprises TRS-L (transcription regulating sequence) 5'-UCUCAAA-3', which is exposed in the core region of SL3 and is meant to be highly involved in virus replication and transcription [20]. SL4 is one of the main hairpin loops containing domain among the coronaviruses, which is located downstream to the TRS-L sequence and is prone to mutations. Even a single or a double SNP in SL4 can change the frame of the sequences and can curtail viral replication [18,21,22].

The genome sequence of SARS-CoV-2 is around 29–31 kb, and more than 151,000 whole genome sequences have been deposited in GISAID (Global initiative on sharing all influenza data) by researchers across the globe. Top variants in the SARS-CoV-2 genome include high-frequency variants including C241T, C1059T, C3037T, C14408T, A23403G, G25563T, and G28883C, out of which the C241T variant had a 99 % frequency with 0.505 entropy by October 2020 [23]. Though a lot of research has been performed about the missense variants of the genome, to the best of our knowledge, no detailed reports on the effect of the 5' UTR variant C241T on host factor binding are available [24]. Study also shows that mutations in spike helps the virus in escape of immune evasion [25,26]. But overall decrease in death rate and increase in recovery rate can be due to some mutations favoring host immune system [27].

Host factors involved in coronavirus proliferation include Annexin A2, which regulates IBV (immune bronchitis virus) frameshifting [28], hnRNP1, which binds with high affinity at MHV (–) strand of the TLR-s leader sequence of RNA, PAPBs binds at poly-A tract of 3'UTR of BCov, TBV and TEGV [29–31], MADP1 Zinc finger CCHC-type and RNA-binding motif 1 binds at stem-loops structures at 5' UTR of SARS-CoV-1 and IBV [16]. The involvement of these factors is proven by wet-lab experiments involving *in vitro* pull-down assay and gene expression studies [13,32]. In recent viral nsp1 is reported to recognize self RNA through binding with 5' UTR's SL1 and mediates recruitment of 40 S ternary ribosome complex for viral proteins translation [33].

In this article, molecular docking and MD (Molecular Dynamics) simulation studies were performed using the wild type (241 T) and mutant (241C) sequences from 5'UTR of SARS-CoV-2 with host factor MADP1 and hnRNP1 to decipher the effect of viral virulence. Further, the wild type and mutant RNA-Protein complexes were compared for their stability by interpreting H-bonds formations, binding energy using MMGBSA, MMBPSA, dynamics cross-correlation matrix, and Principal Component Analysis. Overall, this work helps establish the interactions of host factors with viral 5'UTR stem-loops and understanding the effect of C241T mutation on the dynamics of SL1, SL2, SL3 SL4A, SL4B and SL5 with respect to host-factors.

## 2. Material and methods

### 2.1. RNA secondary structure prediction

RNA secondary and tertiary structure predictions are very important to perform during *in silico* experiments to reach accurate conclusions.

The FASTA sequence of the 5' UTR of the wildtype SARS-CoV-2 was retrieved from the Rfam (RNA sequence family) database with a reference id RF03117 and was used to predict its secondary structure. The mutant RNA sequence was manually generated using the wild-type sequence by replacing cytosine (C) at position 241 with Uracil (U), which is mentioned as C241T in this article. Strains of various SARS-CoV-2 have already reported the presence of various stem-loops structure and bulges which is documented in the RFAM database; we downloaded FASTA sequence (around 300 base pairs) from the RFAM database (id RF03117) for studying 5' UTR sequences of beta coronaviruses [34].

Generation of RNA 3D structure was done using stand-alone and web-based programs such as 3DNA, X3-DNA-DSSR, and RNA composer [35–37]. Mutation in RNA sequence was generated by C241T in 3DNA based mut\_RNA platform [35]. For RNA secondary and tertiary structures were generated in RNA composer that uses the RNA FRABASE dictionary with the CHARMM force field for generating effective minimized energy RNA structures [37]. An advantage of RNA composers is the ease of predicting the ternary structures of long RNA sequences, which is a limitation of many RNA structure prediction software. The effect of single nucleotide variation in base-pairing probabilities was checked in mutaRNA [38]. Final confirmation of folded RNA structure was confirmed in the X3DNA-DSSR Linux-based operating system which gives accurate root mean square deviation (RMSD) between predicted and experimentally verified sequences.

### 2.2. Prediction of RNA-protein binding sites

As MADP1 binds to SL1 of SARS-CoV-2, this region was selected as a potential protein binding site for the 5'UTR region. Additionally, since it is also known that TRS-L binds to hnRNP1 sequence 5' CGGCGUC 3', this sequence was chosen as a protein binding sequence [3,13]. Prediction of RNA residues binding to protein was also done by RNAPin web server [39], where above mentioned sites were also falling in the protein binding region.

### 2.3. Docking of protein-RNA complexes

Molecular docking of host factors and RNA was performed in HADDOCK 2.2 [40,41]. HADDOCK performs docking using a data-driven approach. Docking encompasses three steps: rigid-body docking, semi-flexible refinement, and water refinement. In stage one of the docking of protein and RNA in their bound conformation, a total of 1000 structures were generated using default settings of the program. Systematic sampling of 180 rotated solutions were used for the rigid-body docking to minimize the occurrence of false positives. The best 20 % structures generated from the rigid-body docking were subsequently used in the second stage of semi-flexible refinement. The semi-flexible refinement is carried out in three-stages, followed by a rigid body torsion angle dynamic, and a simulated annealing stage at different MD steps. In the final simulated annealing stage where 1000 MD steps were performed from 300 K to 50 K with 2 fs time steps. In the final stage, both the side chain and backbone of protein residues at interface and RNA can move except the terminal base of RNA. This final stage consists of a gentle refinement (100 MD heating steps at 100, 200, and 300 K followed by 750 sampling steps at 300 K and 500 MD cooling steps at 300, 200, and 100 K all with 2 fs time steps) in an 8 Å shell of TIP3P water molecules [41]. Docking of protein-RNA complexes were also performed in various servers like PATCHDOCK [42], NPDOCK [43], and HDOCK [44] to compare the docking score between two structures.

### 2.4. Molecular dynamics simulation

MADP1 is supposed to bind at SL1, SL2 region, moreover binding with residues covering SL3 region. Dynamics of SL4 and SL5 seems to be very crucial for influencing the binding of host factors with SL1, SL2 and

SL3. C241T mutation is affecting SL4. We carried out unbiased simulations in recently updated forcefield OPLS4 and Amber to conform the overall dynamics of each stem loop in terms of binding of MADP1 and affinity among both variants of RNA.

Molecular dynamics simulations of Protein-RNA complexes were performed in Desmond implemented in Schrodinger [45,46]. Selected protein-RNA complexes were first added to a TIP3P water box, extending 10 Å beyond any of the complex's atoms. Counter ions ( $\text{Na}^+$ , and  $\text{Cl}^-$  ions) were added to neutralize charges. Salt concentration was set to 0.15 M sodium, and chloride ions to approximate physiological condition. The MD was performed in the NPT ensemble at a temperature of 310.5 K and 1.00 bar pressure over 100 ns with recording intervals of 1.2 ps for energy and 24 ps for trajectory. Simulation box contains ~331, 517 atoms in Desmond trajectories, with box size in volume ( $\text{Å}^3$ ) 3413425. Simulations were run with the latest OPLS4 and OPLS3e force field compatible for Protein and nucleic acids. Plots and figures were generated in PyMOL DeLano, W. L. 2009.

Additional MD simulations were performed in Gromacs (version 2018.1) for MADP1-WT complex [47]. Protein and RNA topologies were built with AMBER force filed [48]. MADP1-RNA complexes were simulated using three site TIP3P water model in rectangular box consisting of atoms ~1979400. Box configuration was set with 1 nm distance from edge of RNA in all the directions. Subsequently, equal number of counter ions  $\text{Na}^+/\text{Cl}^-$  were added to neutralize the overall charge of the system. Bond lengths were constrained using LINCS algorithm [49]. Steepest descent algorithm was used to perform energy minimization to remove any steric clashes and bad contacts (50,000 steps max). Particle Mesh Ewald (PME) summation was applied to account for the long range interactions, with a cut off of 1 nm [50]. Minimization was followed by equilibration with position restraint using NVT (constant number [N], constant volume [V] and constant temperature [T]) and NPT (constant number [N], constant volume [V] and constant temperature [T]) before starting the production run. Berendsen thermostat algorithm was used for maintaining a constant temperature of 300 K in NVT based equilibrium [51]. Post NVT equilibrium followed by, NPT equilibration was performed using Parrinello Rahman barostat at a constant pressure (1 bar) for 100 ps [52]. VMD and UCSF-Chimera were used for visualization of MD trajectory [53,54]. Gromacs analysis scripts `g_rmsd`, and `g_rmsf` were used for plotting root mean square deviation (RMSD) and root mean square fluctuations (RMSF) within the systems through their entire trajectories. VMD was used to analyze hydrogen bonds with 3 Å cutoff and 30° cut off for donor-acceptor values. Xmgrace was used for plotting graphs (<https://plasma-gate.weizmann.ac.il/Grace/>).

Wild-type RNA and mutated variant with C241T, were docked with two different host factors MADP1 and hnRNP1, total four complexes were generated. In our studies MADP1 was our major concern as it is known to bind SARS-CoV-2 5'UTR region. MADP1-RNA complexes were simulated in OPLS4 of Schrodinger and AMBER force filed in Gromacs twice. While hnRNP1-RNA complexes were simulated in Schrodinger only. Over all 10 complexes were simulated for 100 ns?

### 2.5. Binding energy for protein RNA complex is calculated using the following equations

Binding energies of the four protein RNA complexes were calculated using PRIME module in Schrodinger using `thermal_mmgbsa.py` and `residue-wise decomposition using breakdown_mmgbsa_by_residue.py` script. The binding energy was calculated for equally spaced 1000 conformations from the MD trajectories.

$$\Delta G_{\text{Bind}} = \Delta G_{\text{SA}} + \Delta G_{\text{Solv}} + \Delta E_{\text{MM}}$$

$\Delta G_{\text{SA}}$  is the difference in the surface area energies for the protein – RNA complex and the sum of the surface area energies in the protein and RNA.  $\Delta E_{\text{MM}}$  is the difference in the minimized energies between the

protein–RNA complexes.  $\Delta G_{\text{Solv}}$  is the difference in the GBSA solvation energy of the protein–RNA complex and the sum of the solvation energies for the protein and RNA.  $\Delta G_{\text{SA}}$  is the difference in the surface area energies for the complex and sum of the surface area energies in the protein and RNA. Binding energy was also calculated for trajectories from Gromacs using MM-PBSA approach Residual decomposition and energy distribution within residues were calculated using `MmPbSaDecomp.py` and `energy2bfac` [55].

### 2.6. Principal component analysis

A vivid graphical presentation of dominant correlated motions of atoms present in the protein-RNA complex is obtained by the covariance matrix of the Cartesian coordinate data set. To generate covariance matrix of elements  $C_{ij}$  for coordinates  $i$  and  $j$  is given by bellow mentioned formula.

$$C_{ij} = (r_i - \langle r_i \rangle) * (r_j - \langle r_j \rangle)$$

Here,  $r_i$  and  $r_j$  were Cartesian coordinates of  $i$ th and  $j$ th  $\text{C}\alpha$  atoms,  $\langle r_i \rangle$  and  $\langle r_j \rangle$  stood for the time average over all the configurations derived from the molecular dynamic simulation. This analysis was performed using the Bio3D library as implemented in R [56–59].

### 2.7. Dynamics cross-correlation matrix

A cross-correlation matrix was used to study the effect of C241T mutation on Protein-RNA complex dynamics by analyzing how atomic displacements were coupled. A cross-correlation coefficient  $C_{ij}$  was calculated from  $\text{C}\alpha$  atoms by the following equation.

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{(\langle \Delta r_i \rangle^2 \langle \Delta r_j \rangle^2)^{1/2}}$$

In this equation,  $\Delta r_i$  and  $\Delta r_j$  are the displacements from the mean position of the  $i$ th and  $j$ th residues (or atoms), respectively. The angular brackets “ $\langle \rangle$ ” represents the time average of the entire trajectory. The value of  $C_{ij}$  is from  $-1$  to  $1$ . A positive value is assigned for positively correlated motion atom (moving in the same direction) indicated in cyan and the negative value represents negatively correlated motion (moving in the opposite direction) indicated in magenta. The DCCM analysis is carried out using the Bio3D packages of R [56–59].

## 3. Results & discussion

Viral replication is highly dependent on the host factors, especially the transcription factors in the case of RNA viruses. Novel SARS-CoV-2, a positive-strand RNA virus from the family *Coronaviridae* deploys a 5' CAP-dependent discontinuous replication within the human host (24, 46). In the present study, molecular dynamics and interaction of two human transcription factors MADP1 and hnRNP1 were studied with a most common 5' UTR mutation C241T in the viral genome. Previous reports on SARS-CoV-1 infection, MADP1 from the Zinc finger protein motif 1 plays a very crucial role in deciding the replication efficiency within host cells [16]. While another transcription factor, hnRNP-1 is extensively studied with respect to the SARS-CoV-2, having multiple roles in viral proliferation like RNA replication, transcription, export, import, and translation. hnRNP-1 uses RGG (Arginine-glycine-glycine) motif for RNA binding; it also binds with viral nucleocapsid protein (N) [60]. While hnRNP1 is binding to the TLRS sequence (Pyrimidine rich) of most the coronaviruses like TGEV and MHV [3], its involvement in SARS-CoV-2 is yet to be studied.

MADP1 and hnRNP1 shares highly structural similarities with superimposition RMSD 1.54 Å, (Supplementary Figure S1). Due to similarities in tertiary structure and role of hnRNP1 in binding at TLR-s of other coronaviruses leads us to incorporate this protein in our study with respect to binding at TLR-s present in SL3. Since the experimental data

available for the MADP1 interaction within SL1 and SL2 region of the 5'UTR of SARS-CoV-1, docking is performed based of selecting residues within these regions. In this study, major focus was kept on MADP1 interaction within 5'UTR due to published data with respect to SARS-CoV-1, while no experimental data was available for hnRNP-1 interaction in SARS-CoV-1 and SARS CoV-2. HnRNP1 is known to bind at TLR-s other coronaviruses and they share similar sequence at TLR-s it was hypothesized that hnRNP1 might also plays important role in terms of viral transcription. Simulation with respect to MADP1 was performed into two different force fields, OPLS4 released in Schrodinger 2021.1 and AMBER99SB in Gromacs18.1 to explore the effect of mutation in terms of affinity and dynamics of stem loops SL1, SL2, SL3 were same in terms of binding with host factor or not. Simulations were performed in duplicates, because of two reasons. One, to check the reproducibility of simulations data in terms of stability of the system. Second, to understand if the dissociation of the host factor MADP1 from Mutant RNA is actually an observed phenomenon or was merely an artefact from the Desmond run. In the case of hnRNP1, simulations were performed only in Schrodinger.

### 3.1. Structural difference in 5'UTR of SARS-CoV-2

Nucleotide sequences of the wild type and mutant (C241T) viral 5' UTR sequences were taken from GenBank accession id MN908947.3 and secondary structures were generated (Supplementary Table S1). RNA secondary structure was predicted using X3-DNA-DSSR in dot-bracket format [36] and the data were visualized in Varna-GUI and structural differences due to C241T were interpreted. Further, RNA composer was used for generating tertiary structure of 5'UTR to correlate the results obtained though X3-DSSR [37]. Representation of the variation in RNA structure due to SNP is elucidated in Table 1. Base pair number was decreased by the change of two in wild-type RNA compare to mutant one, which in turn also affects the number of helices, bulges and internal loops present in overall 5'UTR sequence. The topological difference in both RNA structures is shown in Supplementary Figure S2. While structural features of RNA stem-loops are described in Fig. 1. In mutant RNA, one major change was observed in SL-4 (101G-111U to 101G-112U), which further leads to a change in the loop structure of SL4 (Fig. 1B). Change in SL4 also changes the tertiary structure of RNA; further inducing difference in geometry of SL1, SL2, and SL3 (Fig. 1D and E). Both RNA sequences showed structural and folding differences within SL1, SL2, and TLR sequences (Fig. 1D-3D & 1E-3E). Wild-type RNA structure shared nearby geometry of SL1, SL2, and SL3, Mutated RNA shared faraway geometry for the same. SL1, SL2, SL3 and SL4

**Table 1**  
Tabular Representation showing the output of X3-DNA-DSSR in context of structural difference in both RNA sequences. Wild-type RNA sequence have changed in number of base pairs, bulges, internal loops, and atom based capping function compare to mutant. Change in above mentioned parameters induce a change in folding of SL1, SL2, and SL3 (shown in Fig. 1).

Parameters	Wild-type	Mutant (C241T)
Total number of nucleotides	300	300
<b>Total number of base pairs</b>	<b>107</b>	<b>109</b>
Total number of multiplets	8	8
Total number of helices	8	7
Total number of stems	17	17
<b>Total number of G-U Wobble pairs</b>	<b>7</b>	<b>8</b>
<b>Total number of atom-base capping interaction</b>	<b>9</b>	<b>10</b>
Total number of splayed-apart dinucleotide	33	33
Consolidated into units	15	15
Total number of hairpin loops	7	7
<b>Total number of bulges</b>	<b>3</b>	<b>5</b>
<b>Total number of internal loops</b>	<b>6</b>	<b>5</b>
Total number of junctions	1	1
Total number of non-loop single-stranded segments	6	6
Total number of extended A-minor (type X) motifs	3	3

seems to form a cavity for binding of transcription factor. Based on output secondary structure generated in dot-bracket form using X3-RNA-DSSR are matching exactly with the same reported experimental data representing stem loops and helices [20,22,32]. To correlate these variations in RNA structure with the favourable or non-favourable binding of host transcription factors, protein-RNA docking and the resulting binding energy was calculated to identify the effects of binding between both sequences wild-type and mutant (C241T) RNA.

### 3.2. Docking of host transcription factors with 5'UTR variants

MADP1 is known to bind at SL1 and SL2 of RNA with help of Nsp1 protein for viral replication and transcription [16]. Another protein hnRNP1 is known to bind at the TLR-S sequence of IBV, TGEV, and MHV [3]. Invitro studies of MADP1 binding to 5' UTR of SARS-CoV-1 were available but not for hnRNP1 [16]. 5'UTR of SARS-CoV-2 is folded such that SL1, SL2, SL3, and SL4 shared adjacent folding, which may enhance the surface area for binding of these proteins. Our major focus was on MADP1 mediated binding at the SL1 and SL2 regions of both variants of the 5' UTR region due to confirmatory Invitro studies were available in SARS-CoV-2. In a majority of *Coronaviridae* family viruses, hnRNP1 is known to bind at TLR-S sequence of 5' UTR [13,30]. It was hypothesized that hnRNP1 might also bind to TLR-S of SARS-CoV-2 for viral proliferation. After docking studies, a change in binding was also observed for host factors with both RNA sequences. Additionally, the stability of host factors-RNA complexes was also analyzed to contrast the affinity among both 5' UTR variants.

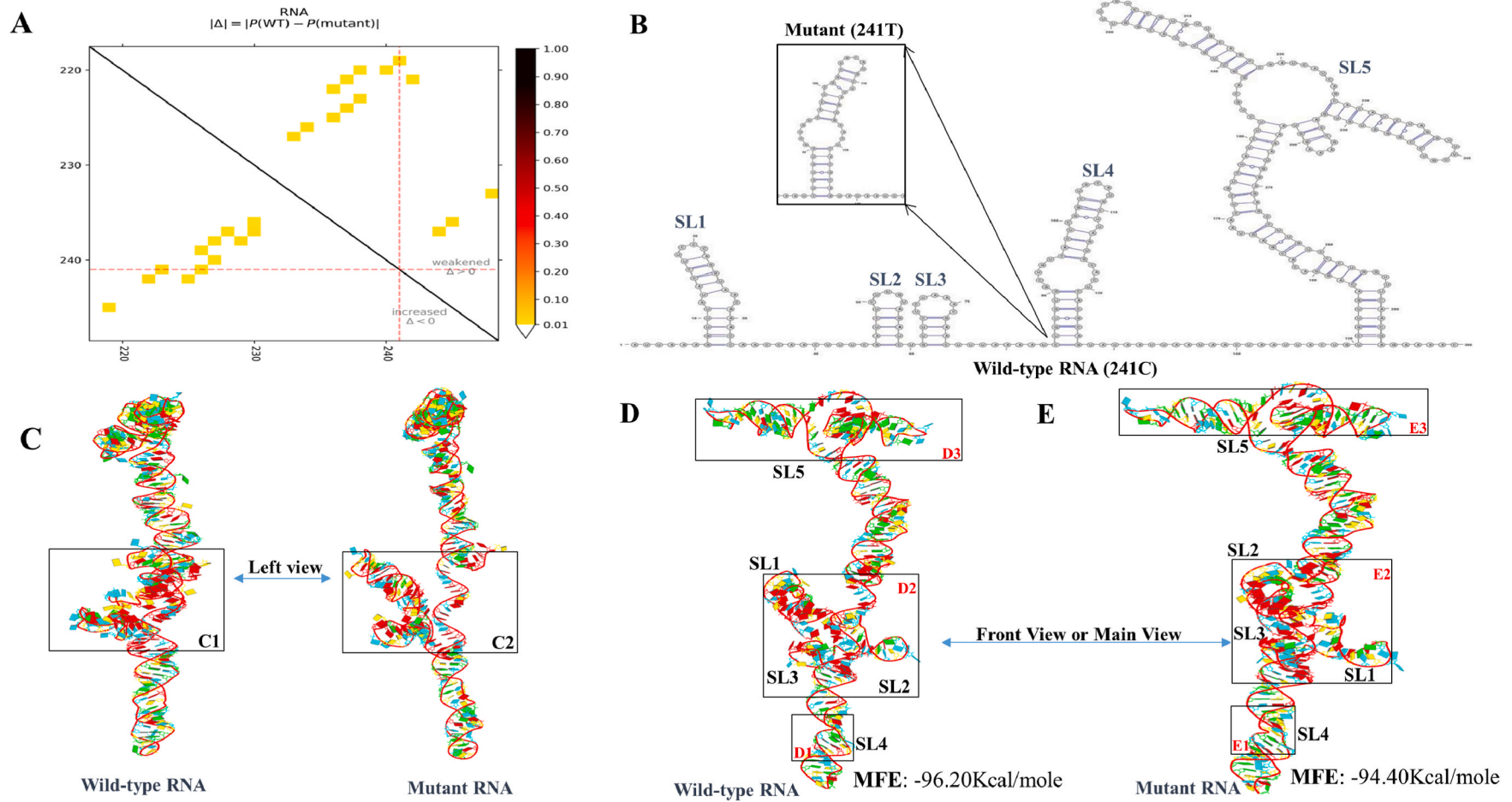
Docking results for wild-type and mutant RNA with host factor protein from HADDOCK depicted the changes in the folding of stem-loops SL1, SL2 and SL3 in the tertiary structures of RNA, which in turn mediates difference in binding within all four complexes. For MADP1, a total of 117 structures in 8 cluster(s) were clustered in HADDOCK for wild type complex representing 58.5% of the water-refined models, while 119 structures in 8 cluster(s), were generated for the mutant sequence (C241T) which represented 59.5% of the water-refined models. For HNRNP1, HADDOCK clustered 107 structures in 13 cluster(s), which represented 53.5% of the water-refined models generated for wild-type complex, while 92 structures in 14 cluster(s), which represents 46.0% of the water-refined models were generated for the mutant complex.

Based on the docking score, wild-type RNA showed to have a better binding profile with a high binding constant value (more negative) of  $-143.0 \pm 3.5$  compared to mutant complex  $-138.4 \pm 1.6$ , as well as the RMSD value for MADP1 was  $0.9 \pm 0.6$  compared to the mutated (C241T) sequence was  $1.1 \pm 0.7$  (see Table 2 for details). The binding pose of MADP1 with RNA structure is shown in Fig. 2. As shown in Fig. 2E and F, MADP1 is covering higher portion of SL1, SL2 and SL3 for binding compare to mutated complex. Hydrogen bond formation within these regions also showing better binding of MADP1 in wild-type complex (H-bonds: 17) compare to the mutated one (H-bonds: 14) depicted in Fig. 2A and B.

For hnRNP1, no significant change was observed for docking score and RMSD values. However, the average standard deviation (negative Z score) among clusters of wild-type RNA  $-2.5$  with host factors indicates better docking among complexes compared to the mutant with  $-1.5$ . (Table 2). Docking results clearly narrate that electrostatic energy and van der Waals energy favored the protein-RNA complex formation while desolvation energy hindered the same. Docked pose with respect to both host factors is shown in Fig. 2.

Docking was also performed in various platforms like NP Dock, H-Dock, and PATCH Dock (Supplementary table: 2). Based on the docking score obtained from HADDOCK and other docking platforms, Wild-type RNA displayed better binding compare to the Mutant RNA.

To further validate these findings, the molecular dynamics of WT and MUT protein-RNA complexes were studied to find binding energy and stability.



**Fig. 1.** Structural changes in the wild-type and mutant variant C241U (C241T) of 5' UTR of SARS-CoV-2. **A:** The differential heat map showing the base pair probability difference between the wild-type and mutant (C241T). The dark red (orange) color indicates a decrease (increase) in the pairing potential induced by the mutation. **B.** RNA 2D structure models generated in X3DNA-SSR to evaluate hairpin loops, bulges, stem loops, multiplets, number of wobble base pairs, extended A-minor (type X) motifs, and many more. Difference due to SNP C241U is elucidated in Table 1. **C.** Left view showing change in overall geometry of stem loops in 5'UTR of both sequences. RNA Backbone is shown in red color, while nucleotides Adenine (A), Guanine (G), Uracil (U), and Cytosine (C) were shown red, green, blue, and yellow color respectively. **C1:** Left view of wild-type RNA. **C2:** Left view of mutant RNA. **D.** Main-view of wild-type RNA. **D2:** Main view, generated from X3-DSSR for wild type RNA where SL1 and SL3 share *nearby* geometry. **E.** Main-view of wild-type RNA. **E2:** Main view generated from X3-DSSR for Mutant RNA where SL1 and SL3 share *far* from each other geometry.

Table 2

**Molecular docking profiles of wild-type vs. mutant RNA interaction with host proteins.** HADDOCK score for wild-type complex with respect to both host transcription factors is more negative compare to mutant. Parameters obtained from docking clearly instigates that wild-type complex seems to me more stable compare to mutant in case of both host factors, MADP1 and hnRNP1.

Protein-RNA complexes	Cluster used in study	HADDOCK score	Cluster size	RMSD	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	Z-score
Wild-type MADP1	Cluster-2	-143.0 ±3.5	37	0.9 ±0.6	-88.8±8.6	-459.4± 22.5	26.6± 1.1	110.5±41.67	2277.5± 86.7	-1.8
Mutant (C241T) MADP1	Cluster 3	-138.4 ±1.6	23	1.1 ±0.7	-75.3±12.7	-519.7±74.5	33.2±3.6	76.7±31.34	2389.6±183.2	-1.4
Wild-type hnRNP1	Cluster-4	-79.7±3.6	9	9.3 ±0.2	-40.0±5.6	-335.3±26.9	18.5±5.5	88.8±19.91	1413.4±165.7	-2.4
Mutant (C241T) hnRNP-1	Cluster-1	-78.3±5.0	12	9.4 ±0.2	-47.4 ±3.1	-248.0±6.0	9.7±5.1	89.5±29.66	1348.1±86.8	-1.5

### 3.3. Change in host factors affinity with respect to dynamics of stem loops in RNA variants through MD simulation

Molecular dynamics (MD) simulations were performed on two variants (C241T) of 5' UTR with two different host transcription factors MADP1 and hnRNP1 after performing molecular docking in HADDOCK. MD simulations of four complexes were performed for 100ns because early dissociation of MADP1 with Mutant (C241T) was observed at 98–99 ns (ns) (Supplementary Videos S1 and S2). Early dissociation in mutant complex is one of the important factors that suggests the stability of RNA-protein complex within both variants. The narrowing of the cavity within SL1 and SL3 affected the binding of MADP1 drastically, which further lead to early dissociation of the RNA-protein complex. Various graphs captured during simulation events of both the wild type and variant in the simulation event analysis package of Schrodinger are described in Figs. 4 and 5. Data from Gromacs trajectories were plotted in Xmgrace package.

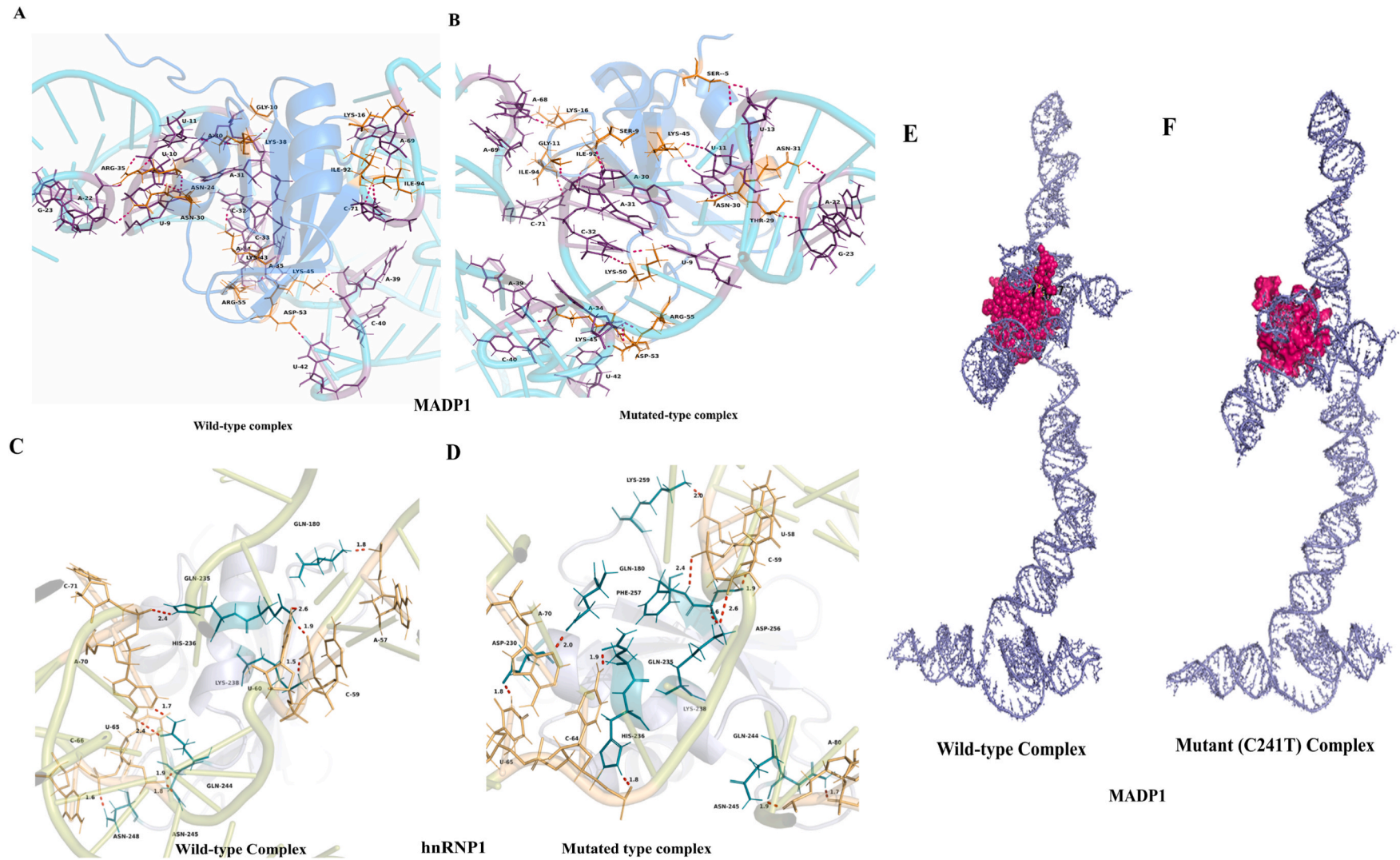
With respect to MADP1, RMSD (Root mean square deviation) of wild-type complex was  $3.89 \pm 1.18$  nm while for mutant it was  $5.18 \pm 1.57$  nm in opsl4 force field. In the case of the Wild-type RNA-protein complex, for the first 20 ns trajectory was fluctuating but after 30ns system reached a plateau showing that the MADP1-RNA complex had formed a stable structure. While the trajectory of the mutant complex was observed to be fluctuating and stabilizing around 60 ns with an average RMSD fluctuation of (Fig. 3A). As shown in Fig. 3, orientation of SL1-SL3 is shown in grey for the wild-type complex and in pink for mutant complex. From OPLS4 trajectory it is clearly visible that wild-type complex is having MADP1 interaction with SL1-SL3 throughout the simulation and mutant showing clearly less interactions or showing pivotal binding within SL1 only. In Fig. 2E and F, initial structure at 0ns and protein was shown in licorice to show the difference area covered by MADP1 in initial structure. One can visualize the orientation of SL1, with respect to MADP1 binding is more favourable in wild-type, compare to mutant complex. From 10ns to 100 ns, in wild-type complex SL1 and SL3 started moving as such to create a cavity around MADP1, while in mutant complex SL1 and SL3 were moving in opposite direction. Black arrows indicate the direction of dynamics of stem loops during the simulation (Figure: 3 A). Dynamics of SL4, where the mutation C241T was observed is important to create this enhance cavity around wild-type complex which is further elaborated from RMSF studies. At the end of simulation both duplicates of mutant complex had shown dissociation (Supplementary Videos 2 and 3).

RMSD values obtained through AMBER99SB were shown in Fig. 3B, wild-type and mutant complexes were showing  $3.86 \pm 0.78$  and  $5.73 \pm 0.89$  nm respectively. Initial structure at 0ns was same for both forcefield which is shown in Fig. 2E and F. In Fig. 3B, SL3 and SL1, were also following same dynamics which was observed in Opls4. As simulation goes on, SL1 and SL3 create a cavity around host factor (madp1) which further enhance affinity for MADP1 with respect to wild-type RNA. At

the same time period 75 ns, mutant complex has formed a binding cavity which is not as compact as wild-type complex. In Fig. 3B, MADP1 is covering more surface area in wild-type RNA as compare to mutant RNA. These findings can be related with over all dynamics of another stem loops which were SL4 and SL5. Dynamics of SL4 and SL5 is influencing the binding of MADP1 with SL1-3, which is shown by RMSF, PCA and porcupine plots for motion mode analysis (Fig. 4).

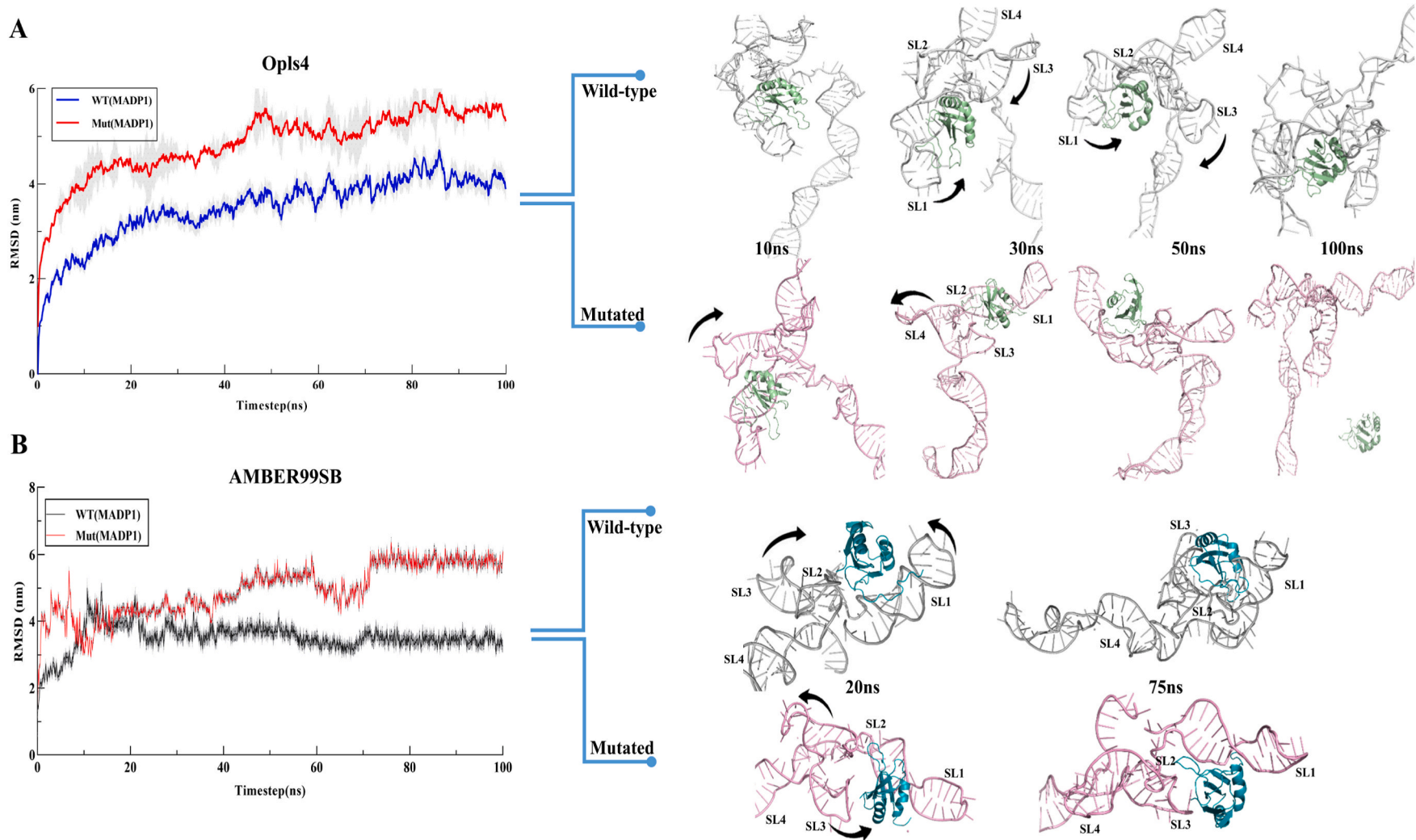
Fig. 3A, is showing RMSD plotted for wild-type and mutant complex in terms of OPLS4 forcefield, which is specific for proteins. RNA structures were fluctuating at higher rate in Schrodinger trajectory, which might lead to increasing RMSD, but system get stable (equilibrate) after ~80 ns? Different behavior protein and nucleic acid in two unlike forcefield may leads to this behavior of trajectories. Major aim was to justify the effect of C241T mutation in two different forcefields. From both forcefields C241T mutation affects the change in 5'UTR which leads decreased stability of 5'UTR with host transcription factors.

RMSF (root mean square fluctuation) values of MADP1 and RNA from both complexes were calculated. Fig. 4A, explains the MADP1 residual fluctuation in both complexes, where amino acid residues from 1 to 60 show less fluctuation in wild type complex compared to the mutant (C241T) complex. Less fluctuation in RMSF value indicates more interaction between RNA-protein complexes (49, 50, 51) which is also evident here in the case of a wild-type complex. Residues which were less fluctuating are actually interacting with RNA, which is further discussed in hydrogen bond analysis section. In Fig. 4A, residual dynamics cross-correlation network for MADP1 were shown, light purple color shows correlated and red color shows anti-correlated motions with respect to RNA variants. In mutant, MADP1 is showing less correlated and more anti correlated motions were observed compare to wild-type. Also, porcupine plots are matching with the RMSF graph. In porcupine plots blue arrows shows low while red shows high and white arrows show moderate atomic displacements. In Figs. 4A, 1a and 1b (also 2a and 2 b) showed red arrows and showing high RMSF values. While 1 b and 1 d in mutant and 2 b and 2 d in wild-type are showing (blue) low atomic displacements covering low RMSF values. In AMBER99SB, Fig. 4C, MADP1 in wild-type was showing high atomic displacement in region covering residues from 1 to 10 (1a) and 50–55 (1 b) shows increase length of arrows in porcupine plots. Mutant complex was showing high atomic displacements in 2a region and 2 b region which is narrated in RMSF plot also. Same results were observed in case of RNA from both variants. In Fig. 5B and D, RMSF of RNA variants in Opls4 and AMBER99SSB is shown respectively. Now one can clearly visualize how the dynamics of SL4 and SL5 is affecting MADP1 binding with SL1-SL3. In Figs. 5D and 1b (wild-type) and 2 b (mutant) of porcupine plots SL4 is showing uniform diffraction in wild-type, while in mutant SL4 is showing highly disturbed motions, which can directly influence the binding of MADP1 with SL1-SL3 as all stem loops are interconnected and part of one RNA structure. Same observations can be made from 1c and 2c regions of Fig. 5D.

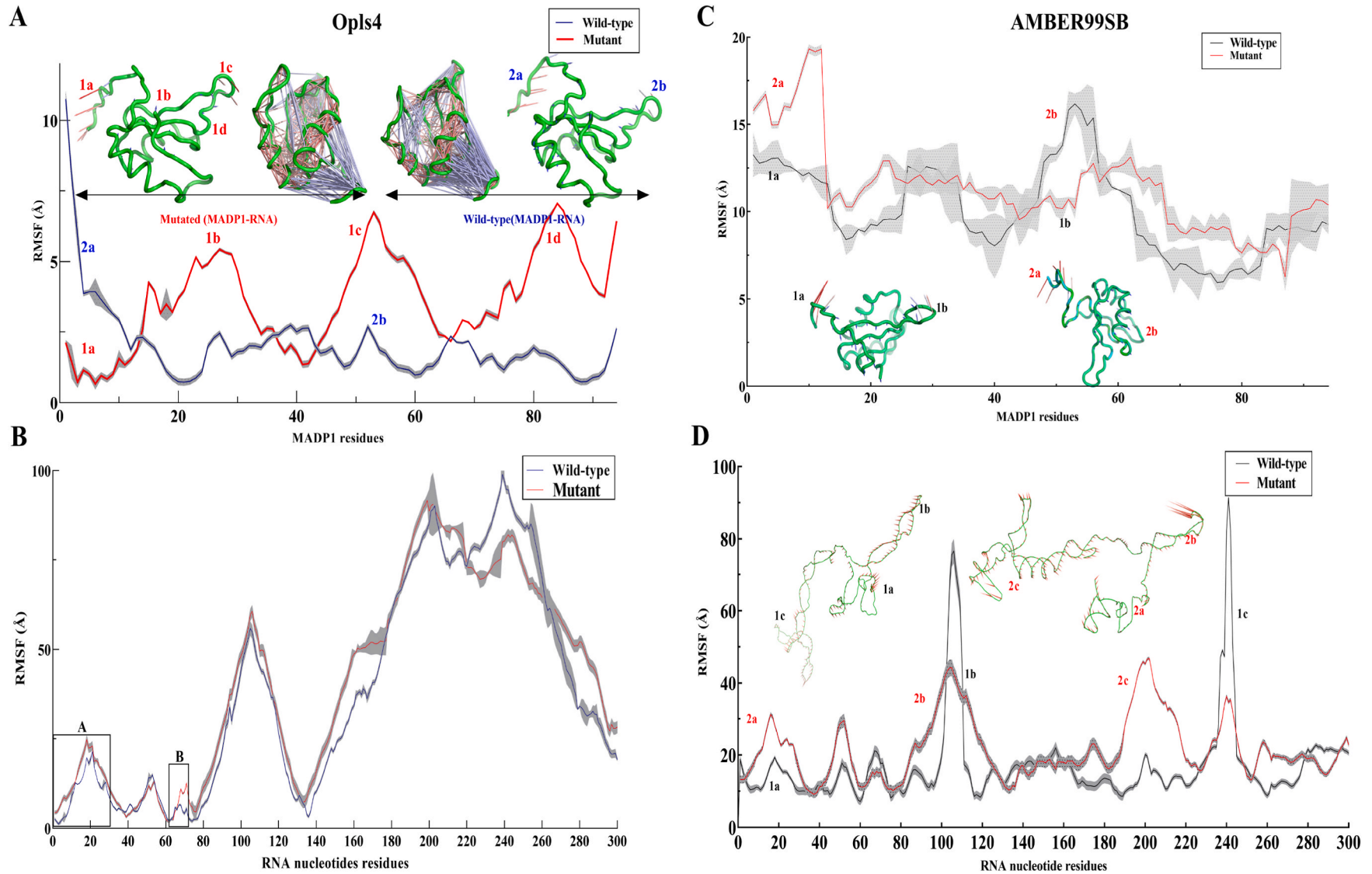


**Fig. 2. Docking pose generated from HADDOCK for MADP1-RNA and hnRNP1-RNA complex.** A & B: MAPD1-wild-type RNA complex having pivotal interactions at Stem-loops region. MADP1 seems to have interfacial hydrogen bonds and stacking interactions with RNA residues within SL1, SL2 and SL3 region of WT and mutant RNA. RNA and protein residues wild-type RNA are shown in blue and marine color. RNA residues interacting with MADP1 (shown in orange sticks) are shown in purple color with sticks. C & D: hnRNP1-WT RNA complex having unique interactions in SL3. RNA is shown in lime color and protein is shown in wheat color. RNA residues interacting with hnRNP1 (shown in green sticks) are shown in orange color with sticks. E&F: Wild-type complex and mutant complex with respect to MADP1. MADP1 is shown in magenta color in surface representation. RNA is shown in cartoon form in slate-blue color.

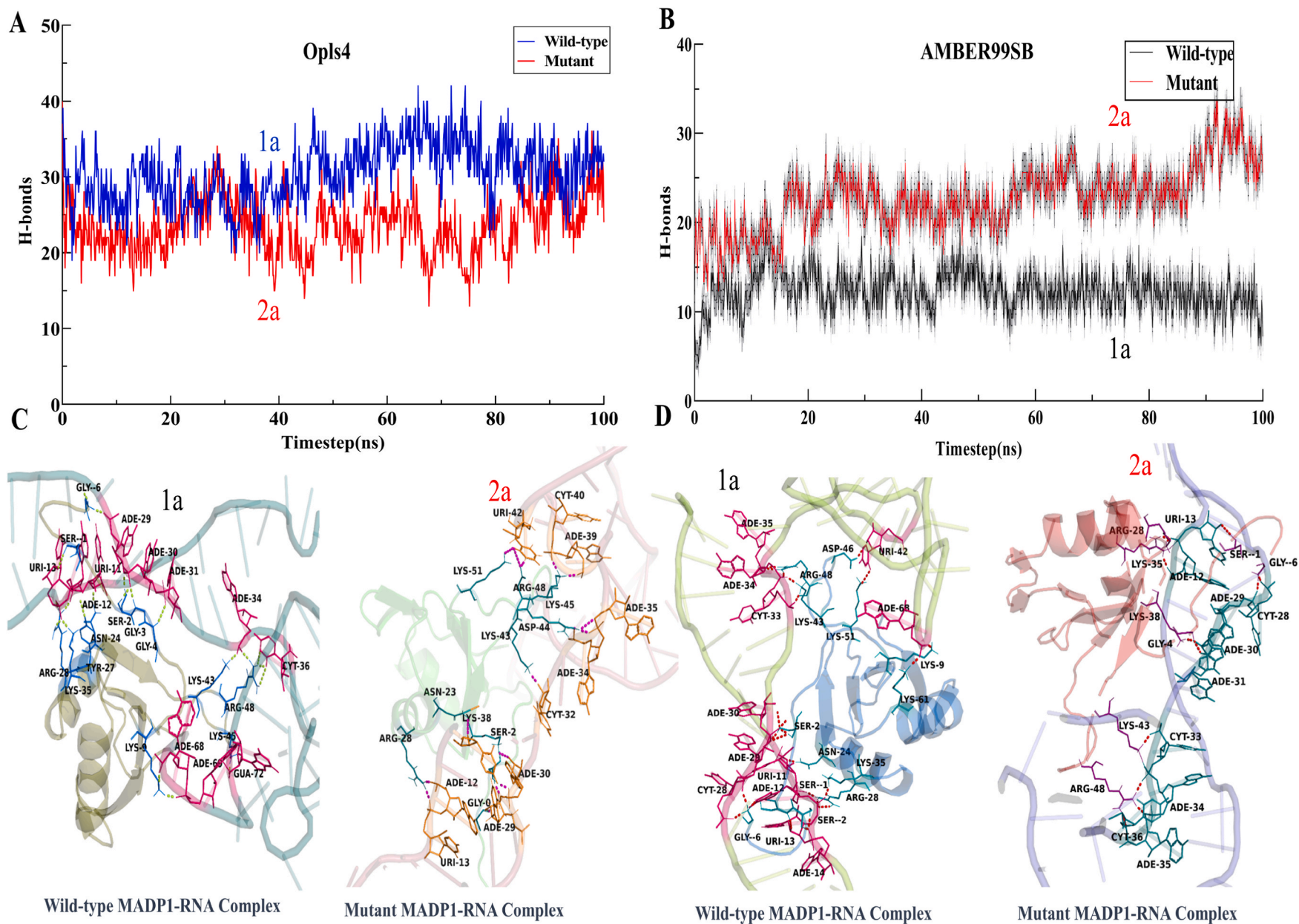




**Fig. 3. RMSD obtained from simulation of MADP1-RNA complexes in OPLS4 and AMBER99SB.** Wild-type RNA is shown in grey color and mutant RNA is shown in pink color. MADP1 is shown in green color. Black colored arrows were depicting direction of stem loops while interacting with MADP1. A. Wild type complexes with respect to MADP1 shows lower RMSD values compare to mutant complex in both force field. Transparent grey region shows standard error obtained through duplicates of simulations. Change in Dynamics of SL1, SL2, and SL3 is different in both RNA which in term affects the binding on MADP1. In Wild-type RNA it shows interaction with SL1, SL2, and SL3; while in Mutant it is showing interaction with SL1 and SL2 only. B. As dynamics of various stem loops are affecting the binding of MADP1 with RNA, same pattern followed in Schrodinger was also observed in Gromacs. MADP1 is showing strong interaction with stem loops 1, 2 and 3 in wild-type complex, while in mutant it is showing with only 1 and 2.



**Fig. 4. RMSF of MADP1 and RNA residues within the complexes. 4A:** RMSF of MADP1 amino acid residues in wild type complex (blue) shows lower residual fluctuation compare to mutant complex (red) in Schrodinger. **4B:** RMSF of in SL1 and SL2 region in MADP-RNA complex shows lower fluctuation in wild-type complex, leading to efficient or compact binding of MADP1 in wild-type RNA compare to mutant RNA in Schrodinger. **4C:** RMSF of MADP1 residues in Gromacs with Amber forcefield. **4D:** RMSF of RNA residues in Gromacs Amber force filed, where again residues which are binding MADP1 with more compactness are showing lower residual fluctuations in wild-type complex. Same things are also shown in form of porcupine plots. Where length of arrows shows degree of fluctuation and direction of arrows shows direction of the same.



**Fig. 5.** Hydrogen bond formation within both trajectories. **5A:** Hydrogen bond formation ( $>3 \text{ \AA}$ ) within wild-type and mutant complexes in Schrodinger. **5B:** Hydrogen bond formation ( $>3 \text{ \AA}$ ) within wild-type and mutant complexes in Schrodinger. **5C:** Pivotal interactions of MADP1 with wild-type and mutant complexes in OPLS4 force field shown using PYMOL. **5D:** Pivotal interactions of MADP1 with wild-type and mutant in AMBER99SSB force field shown using PYMOL.

Another host factor, hnRNP1 and RNA complexes were also analyzed with respect of stability. RMSD of the wild-type complex is 1.45 nm which is much lower compared to mutant complex 2.35 nm (Supplementary Figure 3A). Overall residual fluctuation is shown in Fig. 3B and D. Protein residues that are binding to RNA's TLR-S sequence have overall less fluctuation compared to non-bound residues, among them wild-type hnRNP1 have lower fluctuation compared to mutant (Supplementary Figure 3B-box A). TLR-S (U59–C66) of the wild-type complex has significantly lower fluctuation compare to the mutant complex (Supplementary Figure 3D).

For both host replication factors, wild-type complex seems to be more favourable for the binding compare to the mutant complex based on RMSD and RMSF of both complexes in different force field.

### 3.4. Hydrogen bonds occupancy within Protein-RNA complexes

To study the interaction of amino acids in Protein-RNA complexes, hydrogen bond analysis of those amino acid residues interacting with RNA was analyzed. Protein and RNA interaction with respect to hydrogen bond occupancy is shown in Supplementary Table S3. C241T mutation leads to structural changes in the folding of RNA (Supplementary Figure S2) which in turn affects the binding of MADP1 with SL1 sequence of 5'UTR and hnRNP1 with TLR sequence within SL3. H-bonds occupancy within 3 Å distance cutoff and 20° distance was calculated using VMD [53].

Change in the tertiary structure of RNA leads to a drastic change in the binding of amino-acid residues for both of the transcription factors. Widening of a gap within the Wild-type complex leads to the binding of diverse MADP1 amino-acid residues with RNA, enhancing the stability compared to only one amino acid binding to two nucleotides. As shown in Fig. 5A, hydrogen bond formation in wild-type complex (blue) is higher as compare to mutant complex (red). Positive amino acids residues Lys-45, Arg 48, Lys-9, Arg-28, Lys-51, Arg48, Gly-3, and Ser-2 are showing higher interaction with negatively charged backbone of phosphate and other nucleotide residues, favoring more spontaneous binding due to opposite charges present on binding components. Occupancy of hydrogen bonds also increases in such type of interaction which is correlated with percentage occupancy results of Hydrogen bonds with MADP1 using VMD and decreased in bond length (Supplementary table: S3).

Some hydrogen bonds existed for the only period of time, named dynamic hydrogen bonds, while those that existed throughout the simulation are named as static hydrogen bonds. In SL1 there is a major static hydrogen bonds formation within amino acids and nucleotides. Although, overall amino acid residues showed interactions were the same but change in position of interaction was observed, which is favoring higher percentage occupancy of hydrogen bonds within wild-type complexes. Other interactions were also observed in wild-type complex, forming higher occupancy hydrogen bonds with amino acid residue in Lys-35, Lys-43, Arg-48, and Lys-9 with U13, A34, C36, and A68 respectively forming static hydrogen bonds (Figure: 5C). Wild-type complex seems to cover wide range of nucleotides with different amino acids, which might make it more prone to efficient protein RNA complex formation. Vast varieties of dynamic hydrogen bonds formed during the simulation are formed by amino acids.

Interesting results were observed when MADP1 is showing binding within TLR sequence in both complexes. In the 3D folded structure of 5'UTR, TLR-S residues share a nearby position with respect to SL1 and SL2. This unique interaction leads to one of the parameters to yet study, whether MADP1 is interacting in SARS-CoV-2 TLR-S sequence for viral proliferation or another factor is required for the same. At 100ns major change involved is in the mutant complex where protein is getting dissociated from RNA, compared to wild type complex (Supplementary Videos S1 and S2). This can lead to the conclusion that the mutant complex is having relatively less stable interaction with host factor MADP1 compared to the wild type complex in terms of number

hydrogen bonds 25 and 34 respectively.

Other hypothesized protein-RNA complex (hnRNP1) is analyzed for commenting on the structural stability of RNA with respect to protein binding and interaction. It was surprising for us that Hydrogen bond occupancy in wild-type complex is much higher compared to the Mutant complex (Supplementary Table S3) shown in supplementary figures 8 and 9. Here also the same thing is repeated with respect to MADP1, where numbers of nucleotide residues binding to HNRNP1 were more in wild-type complex compared to the mutant. Nucleotide residues 57 A, 58U, 59C, 60U, 64C, 65U, 69 A, 70 A were binding in wild type complex while 57 A, 58U, 64C, 65U only binding in the mutant complex (Supplementary figure S7 & S8). However, no dissociation among both HNRNP1 and RNA was observed at 100 ns?

RMSD, RMSF, H-Bonds and binding energy in form of MMGBSA indicated weak interaction of mutant protein-RNA complex than wild-type. Results were further correlated with DCCM and binding energy.

### 3.5. Difference in interactions network of protein-RNA complexes through dynamics cross-correlation matrix

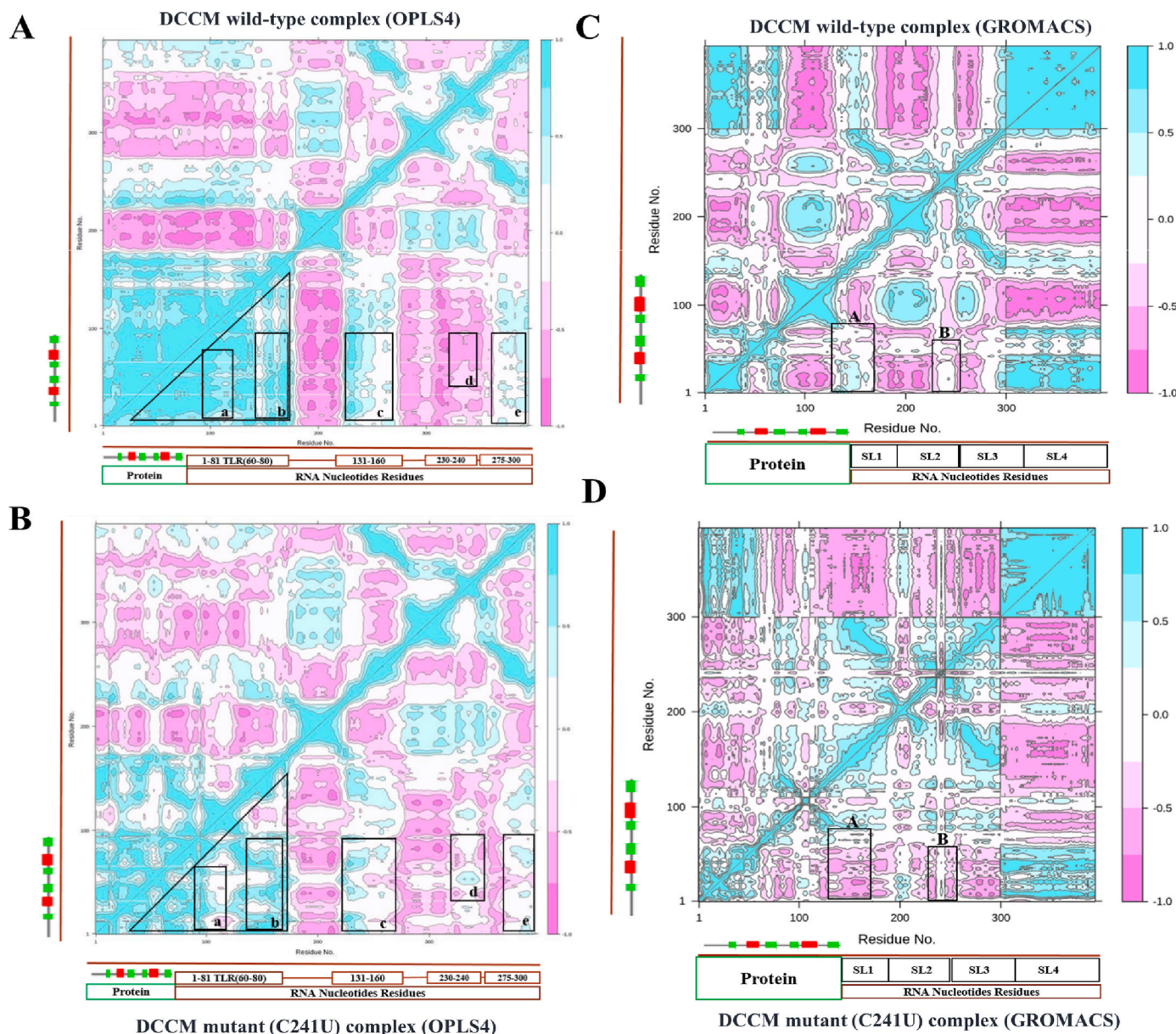
Hydrogen bond interactions were correlated with dynamics cross-correlation matrix, where the interaction between protein and RNA with respect to four complexes is calculated throughout the trajectory of 100 ns?

As shown in Fig. 6, Dynamics cross-correlation matrix was calculated between inter and intra protein motions in both wild-type and mutant (C241T) complexes. As previously discussed in paper, SL1 and SL2 sequences are main interaction point in host-protein MADP1 for SARS-CoV-2 RNA during viral replication. DCCM of 1000 snapshots was performed. In DCCM matrix cyan color indicates positive cross-correlation, magenta color indicates negative cross-correlation, and white color indicates no cross-correlation. DCCM was created for protein-protein, Protein-RNA and RNA-RNA interaction of C $\alpha$  atoms in both complexes. DCCM for MADP1-RNA complexes were shown in Fig. 6.

SL1 and SL2 residues in wild-type complex are showing much higher cyan color intensity with respect to mutant complex (shown in box: an of figure: 6 A). While in Box b nucleotide residues from SL3 were shown. These residues are TLR-S residues among 5'UTR sequence. MADP1 showing unique interaction within TLR-S sequence which is also depicting in DCCM (Box: b of figure: 6 A). Residues in Box a were showing positive correlation with C alpha atoms of protein at C $_{ij}$  value of 0.25–0.85 in wild-type complex, while in mutant complex C $_{ij}$  value was decreased to 0.00 to 0.125. RNA nucleotide residues no 59 to 75 shows positive cross-correlation with protein residues Lys-16, Phe57, Ile92, Ile94, Ala79, Lys-45, Thr-54, Arg-55, Arg35, Met-8, Ser-6, Gly-7, Pro14 and Lys68 in wild-type protein-RNA complexes with higher frequency compare to mutant complex; which is depicted in box b in both Fig. 6A and B. Whereas for Box b, C $_{ij}$  values for mutant complex (0.39) were lower than wild type complex (0.87). Wild-type complex's RNA showed positive correlation in beta sheet residues (green colored) Lys52, Thr47, and Arg35 of MADP1, where this correlation was absent in mutant complex. Lys42, showed more positive correlation with wild-type RNA compared to Mutant RNA (C241T). From DCCM, it is evident that intensity of positive cross correlation in SL1-SL2 of wild type RNA was higher compared to the mutant RNA with MADP1 (cyan).

DCCM of hnRNP1 is shown in Fig. 6C and D, where RNA nucleotide residues on C59, U60, G61, and U62 show interaction with protein amino acid residues Gln180, Lys283, Lys206, Gln253, Ile246 and Asn-48, Gln244 by the formation of hydrogen bonds within 3 Å respectively. The intensity of TLR-S sequence is clearly visible that TLR sequence shows a strong positive correlation with wild-type RNA compared to the mutant which is shown in box A, B, C of both complexes (Fig. 6).

From DCCM it is clearly visible that residue interaction with host initiation factors shows strong positive correlation with nucleotide residue of wild-type complex compares to mutant complexes.



**Fig. 6. Dynamics cross-correlation for MADP1-RNA complexes in Schrodinger and Gromacs.** DCCM was calculated according to time average of  $C\alpha$  atoms within the complex. The whole range of correlation from  $-1$  to  $+1$  is represented in three ranges: cyan color corresponding to positive correlation values ranging from  $0.25$  to  $1$ ; magenta color corresponding to negative correlation values ranging from  $-0.25$  to  $-1$ ; and white color corresponding to weak or no-correlation values ranging from  $-0.25$  to  $+0.25$ . The extent of correlation or anti-correlation is indicated by variation in the intensity of respective cyan or magenta color. **6A:** DCCM of wild-type complex in Schrodinger, box a, b, and c show relatively higher frequency of cyan color compare to mutant complex in SL1-3 region. **6B:** DCCM in Schrodinger with respect to MADP1-mutant RNA complex. **6C:** DCCM of wild-type complex in Gromacs. Box -A, shows positive correlation of MADP1 in form of cyan within TLR-S sequences. **6D:** DCCM of mutant complex in Gromacs, box A shows negative correlation of MADP1 in magenta, within TLR-S sequence.

### 3.6. Principle component analysis (PCA)

PCA was used to study the distinct protein conformational states in a principal component (PC) phase space during the MD simulations. The conformational transitions of the complexes were studied by projecting their trajectories onto a two-dimensional subspace spanned by the first three eigenvectors (PC1, PC2, and PC3) with respect to  $C\alpha$  residues. In [Supplementary Figure S5 & S6](#), both complexes attained two conformational states on the subspace (shown in red and blue). The intermediate state located between these two conformations is shown with white dots.

For hnRNP1, PC1/PC2 and PC3/PC1 of wild-type shows thermodynamically distinct periodic jumps with a substantial energy barrier. While PC1/PC2 and PC2/PC1 of Mutant complex shows overlapping PC

subspaces that lack energy barriers. Value PC1 in the wild-type complex is  $71.95\%$ , compare to  $31.86\%$  of PC1 in the mutant complex. The conjoined distributions of PC1/PC2, PC1/PC3, and PC2/PC3 of the Mutant complex shows energetically unstable complex (more scattered blue dots) compared to PC1/PC2, PC1/PC3, and PC2/PC3 of wild-type complex. More scattered system in the Mutant complex leads conclusion that, it is energetically less stable compared to wild-type. PC1/PC2 and PC3/PC1 of wild-type shows thermodynamically distinct periodic jumps with a substantial energy barrier. While PC1/PC2 and PC2/PC1 of Mutant complex shows overlapping PC subspaces that lack energy barriers ([Supplementary Figure S6](#)).

Same thing is observed in case of MADP1-RNA complexes, PC1 in wild-type complex is  $64.35\%$  and  $73.42\%$  of PC1 in mutant complex. More scattered system is observed PC2/PC3 suggests those principle

modes are unstable in both complexes. PC3/PC1 of wild-type shows periodic jumps with sustainable energy barrier, in mutant scatterings suggests unstable system (Supplementary Fig. S5).

Overall wild-type complex seems to be more stable compared to mutant (C241T) complex with respect to both proteins.

### 3.7. Binding energy components for Protein-RNA complexes

MMGBSA (generalized born surface area) and MMPBSA (Poisson-Boltzmann surface area) are calculated for both protein-RNA complexes. First output structure containing minimized energy was taken in to study for both complexes. Residual decomposition was analyzed to get the contribution of each residue within the complexes for favourable or unfavourable binding. Where  $\Delta G$  bind is higher for wild-type complex compared to mutant complexes for both the proteins MADP1 and hnRNP1 (Table 3). To get insights regarding to energy components contributing more favourable binding within protein- RNA complexes like, electrostatic, H-bonds, van der Waals and ligand strain energy were analyzed.

For each protein-RNA complex, electrostatic energy is major energy component contributing to favourable binding within both complexes, while Solv GB (Generalized Born electrostatic solvation energy) significantly opposes the protein-RNA interaction and complex formation. Both are rather substantial in magnitude but compensate each other, overall resulting in favourable total electrostatic contributions.

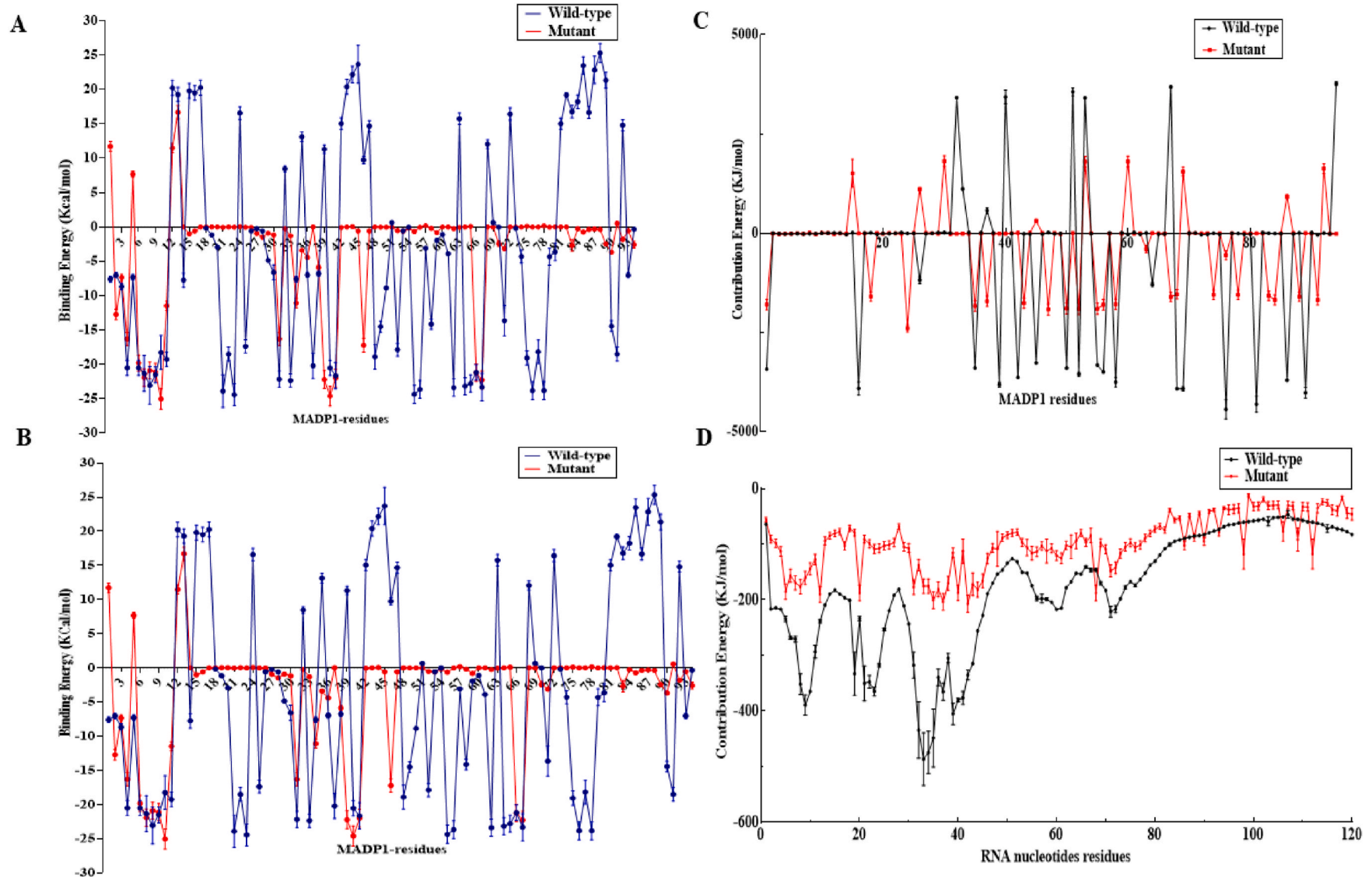
Our major aim was to study the effect the mutation C241T with respect to host replication factor MADP1, docking and molecular simulation studies shows the effect of mutation in favour of host. Binding energy calculations are very crucial for commenting on the complex formation concerning both wild-type and mutant RNA. In MMGBSA, for MADP1 binding energy for WT and MUT complexes were  $-73.58$  kcal/mol and  $-59.66$  kcal/mol respectively. From binding energy, wild-type complex seems to have more efficient complex formation compare to mutant. In Fig. 8A, 8A1 and 8B1, it is clearly visible that majority part of amino acid residues is falling in lime green-blue zone which represents near to zero value for energy contribution in mutant complex. While in the wild-type complex, residues favouring negative spontaneous energy are significantly high. In wild-type complex, amino acid residues interacting more spontaneous are Lys-32, Arg-69, Gly-7, Asn-66, Asn-73, Gln-65, Asn-62, Lys-45, Arg-55 and Lys-51, showing interfacial hydrogen bonds and intermolecular stacking interactions with nucleotide residues covering SL1 and SL2. Residues decomposition graph for MADP1 residues and WT-RNA shows higher negative energy compare to other basic residues within the complex. In wild type complex MADP-1 residues Lys-68 ( $25.32 \pm 2.3$  kcal/mol), Lys-38 ( $-16.72 \pm 3.1$  kcal/mol), Gln-72 ( $-27.30 \pm 3.1$  kcal/mol), and Asn-79 ( $-27.27 \pm 1.3$  kcal/mol) shows 4-5-fold higher energy decomposition compared to other amino acids residues also showing direct contact with RNA backbone (Figs. 7 and 8). In mutant complex amino acid residues interacting with more spontaneous energy are Ser-9, Gly-1, Lys-42, Tyr-34, Arg-55, and Lys53 (showing direct contact with nucleotide residues) (Figs. 8B and 7A). In MUT, minimized structure residues number 1–11

are sharing direct contact with RNA so it is showing less flexibility which is shown in RMSF graph also while in WT there is total opposite case. From Table 3, it is clear that the majority of amino acids residues where electrostatic potential, van Der Waals and H-bonds interaction are high also shows high MMGBSA energy. MADP1-WT complex have 20 hydrogen bonds are forming while in mutant complex only 10 hydrogen bonds are forming, which clearly narrate more compact binding in WT complex compare to MUT (Fig. 8B and G). MADP1 shows a higher binding affinity to SL1 region in WT RNA compare to mutant RNA. In MMPBSA, there is  $-385.722 \pm 22.25$  kJ/mol and  $-85.9493 \pm 30.25$  kJ/mol binding energy for WT and MUT complex respectively (Table 4). Over all residual decomposition for WT complex is much higher compare to MUT for both protein and RNA residues (Fig. 7C and D). Scale bar showing residual contribution with respect to MADP1 and RNA residues is also showing higher energy release in WT residues compare to mutant (Fig. 8). Through MMGBSA and MMPBSA wild-type complex is showing much higher energy compare mutant complex, hence WT seems to have more stable protein-RNA interactions.

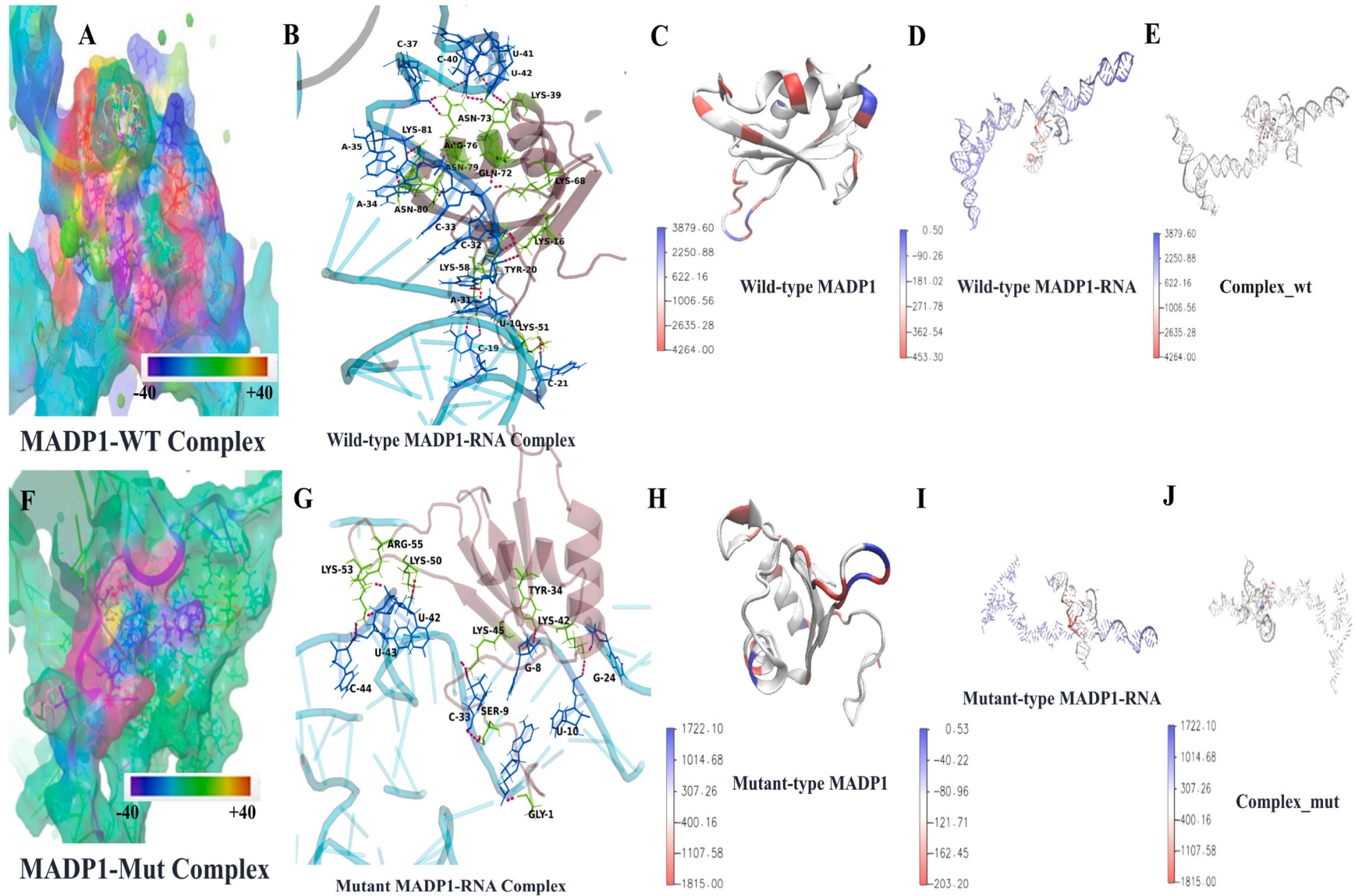
For hnRNP1, binding energy for wild-type and mutant complex is  $-69.85$  kcal/mol and  $-45.85$  kcal/mol. Again wild-type complex seems to bind target with higher affinity compare to mutant in case of hnRNP1 also. Other major contributing energy components for complexes are van Der Waals energy and hydrogen-bond energy (Table 3). These three energies are favoring complex formation more efficiently so energy potencies of each residue in complex were analyzed. As electrostatic energy is favoring the complex formation, residues taking part in interaction are shown in term of electrostatic energy (Supplementary Figs. S7D and S8D). With respect to protein residues, heat maps were shown to correlate amino acid residues with respect to their energy for complex formation. Protein interface residues binding to RNA are showing positive electrostatic potential complimentary to negative electrostatic potential of the phosphate backbone of RNA (Supplementary Figs. S7E and S8E). Extra positive electrostatic potential of residues, Lys 218, L, Lys-259, Arg-277, Arg-254 and Arg185 favors enhanced binding to TLR-S sequence of RNA, henceforth favouring efficient binding of wild-type complex (shown in blue) (Supplementary Figs. S7 and S8). To identify key residue important for binding, residual decomposition was analyzed for both complexes. For wild type complex RNA residues, U47, C59, U60, U62, U63, C64, A79, A81, C123 and G124 are interacting hnRNP1, covering TLR-S sequence -UCUU-, SL2 and SL4. In mutant complex RNA residues, U11, U10, C32, A31, C66, A69, A68, A73 and A74 seem to be interact with hnRNP1. It was quite surprising that the minimized structure of mutant was interacting within SL1 region, more efficiently compare to TLR-S region. HnRNP1 is supposed to bind at TLRs region for viral proliferation, mutated complex (energy minimized structure) is binding less efficiently to TLR-S region. Superimposition of both minimized structures gives the clear idea for the above statements, wt-hnRNP1 is near to SL3 having TLR-S and SL4, while mut-hnRNP1 is near to SL2 and SL3 (Supplementary Fig. S10). From free energy decomposition, it was unsurprised that residues which were having more hydrogen bonds energy and electrostatic energy will possess higher decomposition energy which correlating in

**Table 3**  
Binding energy components from protein-RNA complex.

Energy components	WT MADP1	C241U MADP1	WT hnRNP1	C241U hnRNP1
$\Delta G$ Binding	-65.5821	-54.6634	-69.2351	-45.8545
$\Delta G$ Electrostatic energy	-830.265	-188.509	-1253	-1049.28
$\Delta G$ Covalent energy	5.812749	2.775439	3.256262	11.40654
$\Delta G$ H-bond energy	-10.2371	-6.53	-13.99	-10.1914
$\Delta G$ Lipophilic energy	-4.05164	-4.36723	-2.8212	-4.96648
$\Delta G$ pi piinteraction energy	-9.59606	-3.17418	1.616874	-0.23087
$\Delta G$ selfcontactcorrelation	-0.06446	-0.00658	-0.02245	-0.17093
$\Delta G$ Solv_GB	861.3016	193.0531	1288.897	1101.954
$\Delta G$ vdW energy	-94.3725	-93.1687	-47.9052	-78.4821



**Fig. 7. Residue wise decomposition obtained through MMGBSA for each four complexes. 7A:** Residue wise decomposition for MADP1 residues in MADP1-RNA wild-type and mutant complexes simulated in Opls4. **7B:** Residues wise decomposition of RNA in both complexes in Opls4. **7C:** Residue wise decomposition of MADP1 residues in MADP1 RNA complexes simulated in AMBER99SSB. **7D:** Residues wise decomposition of RNA in both complexes AMBER99SSB.



**Fig. 8.** Minimized structures generated for both MADP1-RNA complexes in Opls4 and AMBER99SSB. **8A:** MMGBSA minimized structure of wild-type MADP1 complex generated in Schrodinger. **8B:** Minimized structure of wild-type MADP1 generated in using MMPBSA approach, dissipating presence of 17 hydrogen bonds. **8C & 8D:** Residual contribution shown in term of B-factor filed using *energy2bfac* in MMPBSA for MADP1 and RNA in wild-type complex. **8E:** Average residual decomposition within the wild-type complex **8F:** MMGBSA minimized structure of mutant MADP1 complex generated in Schrodinger. **8G:** Minimized structure of mutant MADP1 generated in using MMPBSA approach, dissipating presence of only 9 hydrogen bonds. **8H & 8I:** Residual contribution shown in term of B-factor filed using *energy2bfac* in MMPBSA for MADP1 and RNA in mutant complex. **8J:** Average residual decomposition within the wild-type complex.



**Table 4**  
**Binding energy components from protein-RNA complex in MMPBSA.**

Energy components (KJ/mol)	WT-MADP1	C241T -MADP1
ΔG Binding	-385.722 ± 22.25	-85.9493 ± 30.25
Electrostatic energy	-82241.588 ± 142.63	-34726.569 ± 197.567
vdw energy	-420.350 ± 20.36	-356.214 ± 40.25
Polar solvation energy	82250.203 ± 168.25	33987.84 ± 9.45
Non polar solvation energy	-1125.013 ± 10.75	-923.891 ± 10.23

**Supplementary Fig. S9.** In wild-type complex RNA residues within the TLR-S sequence possess more spontaneous, decomposition energy U60 (21.36 kcal/mol), U62 (16.89 kcal/mol) and C64 (-14.56 kcal/mol) compare to the average of -1.5-2 kcal/mol with same residues of the mutant complex. While in mutant residues, C66 (-11.53 kcal/mol), A69 (-5.24kcal/mole), A73 (-10.2 kcal/mol), and A74 (-15.2 kcal/mol) (**Supplementary Fig. S9**). Wild-type sequence seems to have more favourable binding in TLR-s sequence compare to mutant, so 241C seems to be favourable for viral proliferation compare to 241 T.

### 3.8. Epidemiological data

Structural analysis of both RNA, reveal changes in TLR-S region while folding in 3D form, which is affecting the binding of host replication factor MADP1. According to results obtained through MD analysis Mutant, the complex seems to be less stable compare to wild type based on RMSD and Hydrogen bonding between protein RNA complexes. *In-Silico* findings suggest weaker interaction of mutant RNA with host factor MADP1 compared to wild-type. This suggests lesser replication efficiency in mutant SARS-CoV-2 strain. The 5' UTR variant C241T has emerged in March 2020 and its one of the most observed variants in genomes sequenced in 2020, with a frequency of 0.505 [20]. To further correlate this finding, epidemiological data of global cases, death rate, and the recovery rate was obtained from World Health Organization covid-19 dashboard. The recovery rate in global cases has increased from 2.2 % in March 2020 to 70 % to date whereas the overall decline in death rate is observed (**Supplementary Fig. S11**). Although there are many other reasons for increased recovery rate, but the effect of most frequent mutation in the genome can also be not neglected. According to WHO, before the ending of March death rate is 7.1 % while recently in October the death rate is 1.46%.

The fact that if viral RNA is binding less efficiently with host replication factors due to structural changes, the efficiency of virus replication within the host decreases. However, these findings can be further validated experimentally using cell lines and other *In-vitro* methods.

## 4. Conclusion

5' UTR interactions with host factors were studied and it was concluded that C241T changes the SL4, which overall changes the folding of RNA. MADP1 and hnRNP-1 reduce binding shows a decrease in efficiency of the virus to replicate in a host, which in terms decrease the death rate (increasing the recovery rate). Results were correlated with the epidemiological data which is also showing an increased recovery rate all over the world. hnRNP-1 interaction which SARS-CoV-2 is not yet been studied. From our studies, it can be hypothesized that hnRNP1 is binding to the -UCUU sequence present in SARS-CoV-2.

### 4.1. Availability

1. Epidemiological statistics for SARS-CoV-2 was obtained from WHO (world health organization) link: <https://covid19.who.int>

## 5. <https://trackthevirus.info/>

### 5.1. Accession numbers

Public repository accession numbers of Protein and RNA sequences used for this study are listed below.

1. Crystal structure of a Raver1 PRI3 peptide in complex with polypyrimidine tract binding protein RRM2 (**PDB: ID 3ZZY**).
2. Solution structure of RNA binding domain in Zinc finger CCHC-type and RNA binding motif 1, MADP1 (**PDB: ID 2E5H**).
3. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome **GenBank: MN908947.3**.

## 6. Abbreviations

1. SARS-CoV-2: Severe Acute Respiratory Syndrome-CoronaVirus-2
2. UTR: Untranslated Regions
3. SARS-CoV-1: Severe Acute Respiratory Syndrome-CoronaVirus-1
4. MERS: Middle East Respiratory Syndrome
5. GISAID: Global Initiative on Sharing All Influenza Data
6. ssRNA: Single stranded Ribonucleic Acid
7. eIF2: Eukaryotic initiation factor-2
8. IRES: Internal Ribosomal Entry Site
9. ACE2: Angiotensin Converting Enzyme-2
10. TLR-S: Translation leading regions

## 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

We are extremely grateful to the Schrodinger team for providing us an evaluation licence. We are also thankful to Dr Prajwal Nandedkar for providing constant help regarding to the Schrodinger software access and use. We also acknowledge Dr Saumya Patel for manuscript proofreading.

## Appendix A. Supplementary data

Supplementary Data are available at Computers in biology and medicine online.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2021.100670>.

## Funding

Funding is provided by Department of Science and Technology (DST), Government of Gujarat, Gandhinagar, India.

## References

- [1] WHO WHO. Situation report - 65 - coronavirus disease 2019. World Health Organization; 2019. p. 2633. 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [2] V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 2020;155-170 (2021). <https://doi.org/10.1038/s41579-020-00468-6>.
- [3] Sola I, Mateos-Gomez PA, Almazan F, Zuñiga S, Enjuanes L. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol* 2011;8: 237-48. <https://doi.org/10.4161/rna.8.2.14991>.
- [4] Müller C, Schulte FW, Lange-Grünweller K, Obermann W, Madhugiri R, Pleschka S, Ziebuhr J, Hartmann RK, Grünweller A. Broad-spectrum antiviral activity of the eIF4A inhibitor silvestrol against corona- and picornaviruses. *Antivir Res* 2018; 150:123-9. <https://doi.org/10.1016/j.antiviral.2017.12.010>.

- [5] Martínez-Salas E, Francisco-Velilla R, Fernández-Chamorro J, Embarek AM. Insights into structural and mechanistic features of viral IRES elements. *Front Microbiol* 2018;8:1–15. <https://doi.org/10.3389/fmicb.2017.02629>.
- [6] Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N. Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* 2004;2. <https://doi.org/10.1371/journal.pbio.0020366>.
- [7] Toribio R, Ventoso I. Inhibition of host translation by virus infection in vivo. *Proc Natl Acad Sci U S A* 2010;107:9837–42. <https://doi.org/10.1073/pnas.1004110107>.
- [8] Deng JX, Nie XJ, Lei YF, Ma CF, Xu DL, Li B, Xu ZK, Zhang GC. The highly conserved 5' untranslated region as an effective target towards the inhibition of Enterovirus 71 replication by unmodified and appropriate 2'-modified siRNAs. *J Biomed Sci* 2012;19:1. <https://doi.org/10.1186/1423-0127-19-73>.
- [9] makin s, nakagawa k, lokugamage kg. Free information in English and Mandarin on the novel coronavirus COVID- viral and cellular mRNA translation in coronavirus- infected cells. *Advances in viral research*. ISSN 0065; 2020. p. 77.
- [10] Viswanathan T, Arya S, Chan SH, Qi S, Dai N, Misra A, Park JG, Oladunni F, Kovalsky D, Hromas RA, Martínez-Sobrido L, Gupta YK. Structural basis of RNA cap modification by SARS-CoV-2. *Nat Commun* 2020;11:4–10. <https://doi.org/10.1038/s41467-020-17496-8>.
- [11] Zhou P, Lou Yang X, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Di Jiang R, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–3. <https://doi.org/10.1038/s41586-020-2012-7>.
- [12] García LF. Immune response, inflammation, and the clinical spectrum of COVID-19. *Front Immunol* 2020;11:4–8. <https://doi.org/10.3389/fimmu.2020.01441>.
- [13] Fung TS, Liu DX. *Human Coronavirus : host-pathogen interaction*. 2019.
- [14] Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Kringsveld J, Hentze MW. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012;149:1393–406. <https://doi.org/10.1016/j.cell.2012.04.031>.
- [15] Abramo JM, Reynolds A, Crisp GT, Weurländer M, Söderberg M, Scheja M, Hult H, Wernerson A, Emacs A, Distribution UE, Makes W, Like A, Text O, Interface TA, Sets DC, Look TR, Veterans E, Bindings K, Aquamacs C, Support TEX, Previewer TEX, Beaumont C, O'Doherty M, Shannon L, Baume D, \* MY, Coffey M, Bhatara A, Laukka P, Levitin DJ, Darrow A-A, Hornby W, Gingras B, Symposium I, Isbn PS, Author T, All a EC, Majid A, Boroditsky L, Gaby A, Petre M, Rugg G. Host factors in coronavirus replication. Springer International Publishing; 2012. <https://doi.org/10.1007/82>.
- [16] Tan YW, Hong W, Liu DX. Binding of the 50-untranslated region of coronavirus RNA to zinc finger CCHC-type and RNA-binding motif 1 enhances viral replication and transcription. *Nucleic Acids Res* 2012;40:5065–77. <https://doi.org/10.1093/nar/gks165>.
- [17] Miao Z, Tidu A, Eriani G, Martin F. Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol* 2021;18:447–56. <https://doi.org/10.1080/15476286.2020.1814556>.
- [18] Madhugiri R, Fricke M, Marz M, Ziebuhr J. 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. Elsevier Connect , the company ' s public news and information; 2020.
- [19] Yang D, Liu P, Giedroc DP, Leibowitz J. Mouse hepatitis virus stem-loop 4 functions as a spacer element required to drive. *Subgenomic RNA Synthesis* 2011;85: 9199–209. <https://doi.org/10.1128/JVI.05092-11>.
- [20] Rangan R, Zheludev IN, Das R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *Rna*; 2020. <https://doi.org/10.1261/rna.076141.120>.
- [21] MiguelRamos-Pascual, Coronavirus SARS. CoV-2: Analysis of subgenomic mRNA transcription, 3CLpro and PL2pro protease cleavage sites and protein synthesis Corresponding autor. MiguelRamos-Pascual; 2019. p. 1–32.
- [22] Guan B. 5' -proximal cis-acting RNA signals for coronavirus genome replication. 2010.
- [23] Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. NextStrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34:4121–3. <https://doi.org/10.1093/bioinformatics/bty407>.
- [24] Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol* 2020;11:1–13. <https://doi.org/10.3389/fmicb.2020.01800>.
- [25] Chaudhari AM, Kumar D, Joshi M, Patel A, Joshi C. E156G and Arg158, Phe-157/del mutation in NTD of spike protein in B.1.617.2 lineage of SARS-CoV-2 leads to immune evasion through antibody escape. *BioRxiv* 2021. <https://doi.org/10.1101/2021.06.07.447321>. In press.
- [26] Zhang L, Jackson C, Mou H, Ojha A, Rangarajan E, Izard T, Farzan M, Choe H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *BioRxiv*. The Preprint Server for Biology; 2020. <https://doi.org/10.1101/2020.06.12.148726>.
- [27] WHO. Covid-19. India situation report status across states. *Who* 2021;11–3.
- [28] Kwak H, Park MW, Jeong S. Annexin A2 binds RNA and reduces the frameshifting efficiency of Infectious Bronchitis Virus. *PLoS One* 2011;6. <https://doi.org/10.1371/journal.pone.0024067>.
- [29] Spagnolo JF, Hogue BG. Requirement of the poly(A) tail in coronavirus genome replication. *Adv Exp Med Biol* 2001;494:467–74. [https://doi.org/10.1007/978-1-4615-1325-4\\_68](https://doi.org/10.1007/978-1-4615-1325-4_68).
- [30] Galán C, Sola I, Nogales A, Thomas B, Akoulitchev A, Enjuanes L, Almazán F. Host cell proteins interacting with the 3' end of TGEV coronavirus genome influence virus replication. *Virology* 2009;391:304–14. <https://doi.org/10.1016/j.virol.2009.06.006>.
- [31] Wu HY, Ke TY, Liao WY, Chang NY. Regulation of coronavirus poly(A) tail length during infection. *PLoS One* 2013;8:1–12. <https://doi.org/10.1371/journal.pone.0070548>.
- [32] Guan B. TRACE : Tennessee research and creative exchange 5' -proximal cis-acting RNA signals for coronavirus genome replication. 2010.
- [33] Vankadari N, Jeyasankar NN, Lopes WJ. Structure of the SARS-CoV-2 nsp1/5'-untranslated region complex and implications for potential therapeutic targets, a vaccine, and virulence. *J Phys Chem Lett* 2020;11:9659–68. <https://doi.org/10.1021/acs.jpclett.0c02818>.
- [34] Madhugiri R, Karl N, Petersen D, Lamkiewicz K, Fricke M. 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. Elsevier Connect , the company ' s public news and information; 2020.
- [35] Lu XJ, Olson WK. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* 2008;3:1213–27. <https://doi.org/10.1038/nprot.2008.104>.
- [36] Lu XJ, Olson WK. 3DNA. A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 2003; 31:5108–21. <https://doi.org/10.1093/nar/gkg680>.
- [37] Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 2012;40:1–12. <https://doi.org/10.1093/nar/gks339>.
- [38] Miladi M, Raden M, Diederichs S. MutARNA : analysis and visualization of mutation-induced changes in. *RNA structure* 2020;48:287–91. <https://doi.org/10.1093/nar/gkaa331>.
- [39] Panwar B, Raghava GPS. Identification of protein-interacting nucleotides in a RNA sequence using composition profile of tri-nucleotides. *Genomics* 2015;105: 197–203. <https://doi.org/10.1016/j.ygeno.2015.01.005>.
- [40] Van Dijk M, Visscher KM. Solvated protein – DNA docking using HADDOCK. 2013. p. 51–63. <https://doi.org/10.1007/s10858-013-9734-x>.
- [41] Van Dijk M, Van Dijk ADJ, Hsu V, Rolf B, Bonvin AMJJ. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res* 2006;34:3317–25. <https://doi.org/10.1093/nar/gkl412>.
- [42] Schneidman-duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock : servers for rigid and symmetric docking 2005;33:363–7. <https://doi.org/10.1093/nar/gki481>.
- [43] Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM. NPdock: a web server for zymonucleic acid docking. *Nucleic Acids Res* 2015;43:W425–30. <https://doi.org/10.1093/nar/gkv493>.
- [44] Yan Y, Zhang D, Zhou P, Li B, Huang S. HDock : a web server for protein – protein and protein – DNA/RNA docking based on a hybrid strategy. 45; 2017. p. 365–73. <https://doi.org/10.1093/nar/gkx407>.
- [45] Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE. Scalable algorithms for molecular dynamics simulations on commodity clusters. 2006.
- [46] Beard H, Cholleti A, Pearlman D, Sherman W, Loving KA. Applying Physics-based Sampling to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-protein Complexes, 8; 2013. p. 1–11. <https://doi.org/10.1371/journal.pone.0082849>.
- [47] GROMACS development team. GROMACS Documentation Release 2019 GROMACS development team. GROMACS Documentation Release; 2019. 2019.
- [48] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics* 2010;78:1950–8. <https://doi.org/10.1002/prot.22711>.
- [49] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINC: a linear constraint solver for molecular simulations. *J Comput Chem* 1997;18:1463–72. [https://doi.org/10.1002/\(sici\)1096-987x\(199709\)18:12<1463::aid-jcc4>3.0.co;2-h](https://doi.org/10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h).
- [50] Darden T, York D, Pedersen L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–92. <https://doi.org/10.1063/1.464397>.
- [51] Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–90. <https://doi.org/10.1063/1.448118>.
- [52] Parrinello M, Rahman A. Crystal structure and pair potentials: a Molecular-Dynamics Study. *Phys Rev Lett* 1980;45:1196–9. <https://doi.org/10.1103/PhysRevLett.45.1196>.
- [53] Humphrey W, Dalke A, Schulten K. VMD : Visual Molecular Dynamics 1996;7855: 33–8.
- [54] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12. <https://doi.org/10.1002/jcc.20084>.
- [55] Kumari R, Kumar R, Lynn A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 2014;54:1951–62. <https://doi.org/10.1021/ci50020m>.
- [56] Grant BJ, Skjaerven L, Yao XQ. The Bio3D packages for structural bioinformatics. *Protein Sci* 2020:1–11. <https://doi.org/10.1002/pro.3923>.
- [57] Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Cavas LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 2006;22: 2695–6. <https://doi.org/10.1093/bioinformatics/bt461>.

- [58] Skjaerven L, Yao XQ, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinf* 2014;15:1–11. <https://doi.org/10.1186/s12859-014-0399-6>.
- [59] Skjaerven L, Jariwala S, Yao XQ, Grant BJ. Online interactive analysis of protein structure ensembles with Bio3D-web. *Bioinformatics* 2016;32:3510–2. <https://doi.org/10.1093/bioinformatics/btw482>.
- [60] Luo H, Chen Q, Chen J, Chen K, Shen X, Jiang H. The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS (Fed Eur Biochem Soc) Lett* 2005;579:2623–8. <https://doi.org/10.1016/j.febslet.2005.03.080>.