Exploring the conformational dynamics of RNA dependent RNA polymerase of SARS-CoV-2 in the presence of various nucleotide analogues

Shruti Koulgi[#], Vinod Jani[#], Mallikarjunachari Uppuladinne V N[#], Uddhavesh Sonavane and Rajendra Joshi*

High Performance Computing-Medical and Bioinformatics Applications Group, Centre for Development of Advanced Computing (C-DAC), Panchavati, Pashan, Pune, India 411-008

*corresponding author: rajendra@cdac.in

[#]joint first authorship

Abstract

RNA dependent RNA polymerase (RdRP) from positive stranded RNA viruses has always been a hot target for designing of new drugs as it is responsible for viral replication. The major class of drugs that are targeted against RdRP are nucleotide analogues. An extensive docking and molecular dynamics study describing the role of natural nucleotides (NTPs) and its analogues in imparting an inhibitory effect on the RdRP has been presented here. RdRP simulations in its apo, NTP-bound and analogue-bound form have been performed for a cumulative time of 1.9 μs. The conformational flexibility of the RdRP molecule was explored using Principal Component Analysis (PCA) and Markov State Modeling (MSM) Analysis. PCA inferred the presence of correlated motions along the conserved motifs of the RdRP. The ligand binding motif F and template binding motif G showed motions that are negatively correlated with one another. LYS 551, ARG 553 and ARG 555 which are a part of the motif F appear to form strong interactions with the ligand molecules. ARG 836, a primer binding residue was observed to strongly bind to the nucleotide analogues. The MSM analysis helped to observe different conformational states explored by the RdRP. The ensemble docking of the ligands on the Markov states suggested the involvement of the above residues in ligand interactions. The Markov states obtained clearly demarcated the open and closed conformations. The closed states were observed to have more favorable docking of the ligands. MSM analysis predicted a probable inhibitory mechanism involving the closing of the template entry site by reduction in the distance between the flanking finger and thumb subdomain.

Keywords: SARS-CoV-2, remdesivir, favipiravir, galidesivir, sofosbuvir

Introduction

The current global pandemic, COVID-19 which is caused due to the novel coronavirus, Severe Acute Respiratory Syndrome - Coronavirus - 2 (SARS-CoV-2) belongs to the group of Coronaviruses (CoVs) known to cause majorly respiratory tract infections across multiple species.[1] The classification of SARS-CoV-2 states that it belongs to the subfamily *Coronavirinae*, family *Coronaviridae*, order *Nidovirales*. The CoVs that are known to infect the human species are known as human coronaviruses (HCoVs). HCoVs has been observed to cause infections leading to both mild effects such as common cold, bronchitis and life threatening pneumonia.[2,3] Prior, to the identification of SARS-CoV-2 , two HCoVs namely, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) were responsible for causing the SARS and MERS disease. The mortality rate in case both these diseases was significant enough to raise a need for therapeutics.[4-6] Similarly, the COVID-19 caused by SARS-CoV-2 has gained importance for development of drugs/vaccine owing to its epidemiology. It is an enveloped virus with positive stranded RNA and its genome ranges up to 30,000 nucleotides and codes for around 29 proteins.[7-9] These proteins consist of structural and non-structural proteins (nsp) and are responsible for the overall functioning of the virus. The protein, RNA dependent RNA polymerase (RdRP) which is responsible for viral replication is observed to be highly conserved amongst the group of coronaviruses. The RdRP is one of the nsp and is known as the nsp 12. The RdRP of SARS-CoV-2 is known to share a similarity of 96.3 % with the RdRP of SARS-CoV.[10] The first experimental structure of SARS-CoV-2 RdRP was elucidated using cryo-Electron Microscopy (cryo-EM) in May 2020.[11] The structure of the viral RdRP resembles to that of a right cupped hand formation. The cryoEM structure of SARS-CoV-2 RdRP is in complex with the nsp7 and nsp8. The polymerase domain of RdRP is sub-divided into functionally important subdomains and motifs. Figure 1 A represents different subdomains and motifs present in the RdRP molecule. The right cupped hand formation of the RdRP comprises of three major subdomain, the finger subdomain 1 and 2 (FD1, FD2), palm subdomain (PD) and thumb subdomain (TD). The FD1 and FD2 has the residues within the range of 398-581 and 621-679 respectively. The PD has the residues within the range of 582-627 and 688-815. PD is the largest subdomain in comparison to the others. TD consists of residues within the range 816-919 which has helical structures and a functionally critical β-loop. A nidovirus RdRP-associated nucleotidyltransferase domain (NiRAN), is present in the residue range from 115 to 250. This is followed by a linker region between it and the FD1 from the residue range 260-370. RdRP being a highly conserved protein across the viral family has 7 conserved motifs, A-G within the polymerase domain.[12,13,14] Figure 1B represents these motifs present in the polymerase domain. The motif A ranges from residue 613 to 626, this motif has catalytic residues. The motif B is a part of the PD and ranges from residue 675 to 710. This motif consists of residues that facilitate binding to the RNA template. The motif C consists of residues within the range 753-767 and like motif B is a part of the PD. This motif consists of the SER 759, ASP 760 and ASP 761 (SDD) triplet site, which is crucial for primer binding. The motif D ranges from 772 to 796 and is a part of PD. The motif E ranges from 811 to 821 and is part of the TD. This motif consists of the CYS 813, SER 814 and GLN 815 (CSQ) site which is crucial for primer binding. The motif F consists of residues within the range 544-555 and is known to bind to the inhibitor molecules. This motif consists of highly polar LYS and ARG residues that

initiate strong interaction with the inhibitor molecules. The motif G comprises of residues within the range 500-514, which have residues that play a critical role in RNA template binding. The RdRP is structurally well characterized hence, there is a scope for development of drug molecules that may target the functionally important subdomains and motifs and lead to an inhibitory effect. Various drug development strategies have been targeted towards inhibition of the RdRP of the SARS-CoV-2 to minimize the viral replication.[15-18] One of the major efforts has been put in identifying nucleotide analogues which would help in pre-mature termination of the RNA by interfering the RNA replication process.[19,20] Nucleotide analogues are resultant of the modifications in the natural nucleotides namely, Adenosine-5'-triphosphate (ATP), Guanosine-5'-triphosphate (GTP), Cytidine-5'-triphosphate (CTP), Thymidine-5'-triphospate (TTP) and Uridine-5'-triphosphate (UTP). Few of the known FDA approved nucleotide analogues are, favipiravir (FPV), galidesivir (GDV), lamivudine (LMD), ribavirin (RBV), remdesivir (RDV) and sofosbuvir (SBV). The prodrug form of these nucleotide analogues has been approved by the FDA. These prodrugs get metabolized into the tri-phosphate form which is the active metabolite that interferes with the activity of the RdRP enzyme.[21,22] GDV and RDV which are known adenosine analogues have been proved successful in treating viral infections. The former one is considered as a broad spectrum antiviral and has been used for treating hemorrhagic fevers.[23] RDV, is currently being used as an emergency drug for treating critically ill COVID-19 patients.[24-28] The earlier use of RDV was observed in treating the EBOLA disease.[29] Experimental efforts to study its effect on inhibiting the current SARS-CoV-2 suggest high potency.[26] An announcement by the US National Institute of Allergy and Infectious Diseases (NIAID) had recommended the used of RDV for faster recovery of SARS-CoV-2 infected patients.[27,28] FPV and RBV, are known guanosine analogues are currently being studied to understand its inhibitory effect on the RdRP protein.[30-32] Earlier these analogues have been used to treat influenza in Japan, Hepatitis C and viral hemorrhagic fevers.[23] LMD, is a cytidine analogue and has earlier been used to treat HIV and Hepatitis B infections.[33] This analogue is known for its anti-retroviral activity. SBV, is a uridine analogue and has been extensively used to treat Hepatitis C viral infections.[33] Drug repurposing strategies to identify potential drugs that can be targeted against the RdRP are being developed world-wide.[35-38] Molecular docking studies of RBV, RDV, GDV and SBV suggest the role of crucial residues from the different subdomains of the RdRP.[38] The molecular dynamics (MD) simulation studies have enabled the understanding the conformational dynamics of the RdRP molecule.[39,40] MD simulation studies on earlier known RdRP molecules from viruses related to SARS-CoV-2 reveal the role of inter-domain molecular motions responsible for the protein functioning.[39,40] Correlated motions among the different motifs and domains of the RdRP result in the proper viral replication.[40,41] Hence, the work presented here explores the conformational variation of RdRP molecule of SARS-CoV-2 in the presence of natural nucleotides (NTPs) and their analogues. The MD trajectories were analyzed to understand individual ligand effect on RdRP conformational dynamics and impact of interesting interacting residues. The 932 residue long RdRP molecule has been studied using molecular docking and MD simulation studies. The five NTPs namely, ATP, GTP, TTP,CTP and UTP and six nucleotide analogues namely, FPV, GDV, LMD, RBV, RDV and SBV have been docked on to the active site of the RdRP and further simulated for a cumulative of 150 ns each. A ligand-free RdRP, has also been simulated for a cumulative of 250 ns. This apo-form of RdRP

has been used as a control to perform comparative analysis of the ligand-bound simulations. Principal component analysis (PCA) of the simulation data enabled to identify the functionally crucial regions of the RdRP that undergo changes in the ligand-bound form. PCA also enabled to point the correlated motions within the different motifs of RdRP in the presence of either NTPs or analogues. Free energy analysis helped in identification of residues that strongly interact with the ligand molecule and form stable interactions. Interactions specific to NTPs and nucleotide analogues were identified through hydrophilic and hydrophobic interactions between the ligand and the RdRP residues. In order to obtain a holistic view on the flexibility of the RdRP molecule in accommodating the binding of the different ligands, Markov state modeling (MSM) analysis of the simulation was performed which gave an insight on the transition of the RdRP molecule from a functionally active conformation to a probable inhibitory conformation. There have been studies reported where in MSM analysis has been found to be beneficial in identifying intermediate states visited by biomolecules through MD simulations.[42-46] A probable mechanism of inhibition was identified based on the variation of the conformational parameters across different Markov states obtained. An ensemble docking of the above mentioned NTPs and nucleotide analogues was performed on the states obtained through MSM which revealed RdRP residues of importance in initiating the inhibitory effect.

Methodology

Systems Preparation

The co-ordinates for the RdRP protein were obtained from the SWISSMODEL. The template used for building the RdRP molecule was PDB ID: 7BTF.[11] The 7BTF structure consists of RdRP (nsp12) in complex with nsp7 and nsp8, only the nsp12 co-ordinates were used (Figure 1 A). Three missing residues namely, SER, ALA and ASP from the N-terminal were added using the *xleap* module of AmberTools17.[47] The RdRP molecule consisted of two Zinc atoms in co-ordination with CYS and HIS residues. One of the Zinc atom was in co-ordination with CYS 301, CYS 306, CYS 310 and HIS 295 and the other in co-ordination with CYS 487, CYS 645, CYS 646 and HIS 642. Both the Zinc atoms formed a tetrahedral co-ordination complex. A total of 12 RdRP systems were prepared for the study. Table 1 shows the details and notations of these 12 simulation systems. The active metabolite forms of the natural nucleotides and their analogues were obtained from the PubChem database (Figure 2). The five natural nucleotides studied were, namely, adenosine-5'-triphosphate (PubChem CID: 5957), guanosine-5'-triphosphate (PubChem CID: 135398633), cytidine-5'-triphosphate (PubChem CID: 6176), thymidine-5'-triphospate (PubChem CID: 64968) and uridine-5'-triphosphate (PubChem CID: 6133). The six nucleotide analogues studied were, namely, favipiravir-triphosphate (PubChem CID: 5271809), galidesivir-triphosphate (PubChem CID: 146047139), lamivudine-triphosphate (PubChem CID: 454110), ribavirin-triphosphate (PubChem CID: 122108), remdesivir-triphosphate (PubChem CID: 121304016) and sofosbuvir-triphosphate (PubChem CID: 23725128).

Molecular Docking

The molecular docking of all the natural nucleotides and its analogues was performed on the ligand-free structure of RdRP as explained above. The flexible docking protocol was followed using the DOCK 6 software.[48] The parameters for the RdRP molecule were generated using UCSF Chimera. The addition of hydrogen atoms and assignment of charges for the RdRP molecule was performed. The AMBERFF14SB force field was used to assign atom types and internal co-ordinates for the protein molecule.[49] The active site of RdRP was determined using the *sphgen* module of DOCK 6.[48] The cavity selection is done based on clustering of spheres generated along the surface of the protein. The largest cluster of the spheres was considered as the active site of RdRP. This step is followed by generation of a grid along the active site using the *grid* module of DOCK 6.[48] This was followed by individual docking of each of 11 nucleotides to the receptor molecule. The standard parameters were used for the docking of ligand molecules on the protein.

Molecular Dynamics

The molecular dynamics (MD) simulations were performed for all the systems specified in Table 1. The AMBER16 simulation package was used for the same. The force field used was AMBERFF14SB and the Zinc Amber Force Field (ZAFF).[49,50] ZAFF was used for handling the two Zinc co-ordination complexes present in the RdRP molecule. The parameters for all the ligand molecules were generated using the *antechamber* module of AmberTools17.[51] The general AMBER force field was used to assign atom types and bonded parameters to the ligand molecules.[52] Each of the 12 systems were solvated using the TIP3P water model and the entire system was neutralized using Na+ atoms. The *xleap* module of AmberTools17 was used for the same.[49] Minimization was performed using the steepest descent and the conjugate gradient method in two steps. Initially, the solvent and gradually the entire simulation system was minimized in 10000 steps. Similarly, the entire system was gradually heated to 300 K using the Langevin thermostat. The hydrogen atoms were constrained using the SHAKE algorithm. The system was equilibrated for 1 ns at constant temperature and pressure of 300 K and 1 atm. The production run was performed for 50 ns. Five replicates of 50 ns each were simulated for the RdRP-APO system, resulting in a cumulative of 250 ns. Three replicates of 50 ns each were simulated for all the eleven ligand bound systems. Hence, each ligand bound system was simulated for a cumulative of 150 ns each. Hence, a cumulative of 1.9 μs data was simulated for the RdRP systems explained in Table 1.

Analysis

The analysis included root mean square fluctuation (RMSF), principal component analysis (PCA), molecular mechanics/generalized Born surface area (MM-GBSA) free energy, hydrogen bonding and salt bridge interactions, multivariate PCA, and Markov state modeling (MSM) analysis. The RMSF and PCA were performed using the *cpptraj* module of AmberTools17.[53] The MM-GBSA free energy was performed using the *MMPBSA.py* module of AmberTools17.[54] The hydrogen bonding and salt bridge interactions were calculated using the PLIP software.[55]

Multivariate PCA was performed using the R package.[56] MSM analysis was performed using PyEMMA software.[57]

Results & Discussion

Conformational variation in the RdRP molecule

The conformational changes occurring in the RdRP molecule in each of the simulations were compared by performing the PCA. All the atoms except for hydrogen were considered as reaction coordinates. Figure S1 shows the distribution of the conformers along the principal component (PC) 1 and PC 2 for the RdRP-APO, RdRP-ATP, RdRP-TTP, RdRP-GTP, RdRP-CTP and RdRP-UTP systems. The conformers in the RdRP-APO system were observed to vary between the eigen values -100 to 200 along the PC 1. However, this distribution was observed to be constricted between the eigen values -100 to 100 along the PC 2. Maximum variation in the conformations was captured by PC 1, although no distinct cluster of conformers was observed to be formed. The RdRP-ATP, RdRP-TTP and RdRP-CTP were observed to form three distinct population of conformers along the PC 1. In case of RdRP-ATP and RdRP-CTP the three populations was observed to overlap the distribution of the RdRP-APO system. However, in case of RdRP-TTP, one population was observed to be distinct with no overlap in the RdRP-APO distribution. The other two populations were observed to be partially overlapped with the RdRP-APO distribution. The RdRP-GTP system was observed to form two distinct populations along the PC 1. One of these populations overlapped with the distribution of RdRP-APO system. The other one was observed to be widely distributed along PC 2, with few conformers spanning region distinct to that of the RdRP-APO. The RdRP-UTP system was observed to vary the least in comparison to the other RdRP and NTP complexes along the PC 1. Most of the conformers were distributed in the same region as that of the RdRP-APO. These observations suggest that the conformational variation in the NTP-bound RdRP systems significantly sampled distinct populations as compared to the widespread population distribution observed for the RdRP-APO. The presence of the NTPs may have induced dominant changes that result in distinct conformations of the RdRP molecule which are minimally explored in the ligand-free state. Figure S2 shows the population distribution for the RdRP and nucleotide analogue systems. The RdRP-LMD and RdRP-RBV systems which have the cytidine and guanosine analogues respectively, were observed to sample three distinct population of conformers along the PC 1. Although, the distribution of RdRP-LMD conformers was constricted to shorter range of eigen values as compared to that of RdRP-RBV. The region of distribution along PC 1 and 2 for RdRP-LMD matched to that of RdRP-CTP. In contrast to RdRP-RBV system, where in the population was more distributed and sampled new conformations as compared to RdRP-GTP. The RdRP systems with adenosine analogues, namely, RdRP-GDV and RdRP-RDV showed three dominant populations. The distribution of RdRP-RDV system matched to that of RdRP-ATP. However, in case of RdRP-GDV few new conformers were also sampled. The RdRP-FPV which is also a known purine analogue was observed to have distribution of the population similar to that of the RdRP-ATP. The RdRP-SBV appeared to sample populations in the similar to that of RdRP-APO. However, the distribution was less dispersed in case of RdRP-SBV as compared to RdRP-APO. The population distribution along the first two PCs for the nucleotide

analogue-bound RdRP systems showed more densely populated clusters as compared to the RdRP-APO. Three of the nucleotide analogues namely, RdRP-LMD, RdRP-RDV and RdRP-FPV showed similar distribution along the first two PCs as compared to their parent natural nucleotides. The sampling of significant populations along the first two PCs for all the ligand-bound RdRP systems suggests that the presence of ligand enabled exploring statistically significant conformations of the RdRP molecule. In order to map these conformational changes on to the actual protein, the RMSF of the residues along the PC 1, 2 and 3 was calculated. Figure 3 and 4 shows the RMSF averaged over the entire residue for the RdRP-NTPs and RdRP-nucleotide analogue systems along the PC 1 (black), 2 (red) and 3 (green) respectively. The first row of the Figure 3 shows the residue-wise RMSF of the RdRP-APO system along PC 1, 2 and 3. The remaining five rows from top to bottom, represent the same for RdRP-ATP, RdRP-TTP, RdRP-GTP, RdRP-CTP and RdRP-UTP respectively. The regions shaded are the ones that were observed to be have higher RMSF values along the entire RdRP-protein. The N-terminal region ranging from residue 20-100 was observed to be fluctuating the most in the RdRP-ATP system. RMSF values ranging 5-9 Å were seen in this region along all the first three PCs. The RdRP-APO, RdRP-TTP, RdRP-GTP and RdRP-CTP showed fluctuations in this region which were captured by PC 2 and 3. This suggests that the conformational changes in this region were more dominant in RdRP-ATP as compared to the other RdRP-NTP systems. The FD1 region ranging from 398-581 residues also appeared to fluctuate the most in most of the systems. The magnitude of fluctuation captured by PC 1 ranged within 2-4 Å for the RdRP-APO, RdRP-TTP and RdRP-CTP systems. However, slightly higher values of RMSF were observed in case of RdRP-ATP, RdRP-GTP and RdRP-TTP. This fluctuation was observed to be reduced in RdRP-GTP and RdRP-CTP systems as captured by PC 2. The other systems showed significant fluctuation in this region along PC 2. The third most fluctuating region was the TD region ranging from 816 to 919 residues. The RdRP-APO system fluctuated the most in this region. The magnitude of fluctuation for all the RdRP-NTP systems was within the range from 2-4 Å. The fluctuation was observed to be captured by the first three PCs. Figure 4 shows the RMSF values averaged over the residues captured by the PCs 1, 2 and 3 for the RdRP-FPV, RdRP-GDV, RdRP-LMD, RdRP-RBV, RdRP-RDV and RdRP-SBV systems. All the three regions of the RdRP namely, N-terminal region, FD1 region and TD region which were observed to fluctuate the most in RdRP-NTP systems, showed similar results for the RdRP nucleotide analogue bound systems. However, the magnitude of fluctuations in the N-terminal region was observed to be more in PC 2 as compared PC 1. This suggests that this variation was not one of the dominant motions in the RdRP. The FD1 region showed RMSF in the range of 3-5 Å in all the RdRP-nucleotide analogue systems for the PC 1. The fluctuations in this region extended till the PC 3 too. The TD region also fluctuated in the range of 3-6 Å as captured by PC1. The fluctuation appeared to reduce in PC 2 but was again captured by PC 3. Apart from these regions which were in common with the RdRP-NTP systems. A linker region between the NiRAN subdomain and finger subdomain ranging within the residues 260-370 was observed to fluctuate within 3-6 Å in most of the RdRP systems bound to nucleotide analogues. RdRP-FPV and RdRP-LMD showed fluctuation in this linker region within the range of 5-6 Å captured through PC 1. The remaining nucleotide analogue RdRP systems showed fluctuations that were captured by the PC 2 and 3. In case of the RdRP-NTP complexes, this region was observed to fluctuate significantly in case of RdRP-UTP

as the RMSF values were high in the first PC itself. The RdRP-TTP too showed fluctuation in this region although it was captured only by PC 3. The observations made through PCA may infer that the N-terminal, finger subdomain 1 (FD1) and thumb domain (TD) may be the responsible to sample distinct conformations of RdRP in the ligand bound systems. The opening and closing of the template entry site involves two of these domains namely, FD1 and TD. The motions in these two domains play a crucial role.[41]

Changes in the template entry site of RdRP

The changes in the template entry site of RdRP were investigated by calculating the distance between the finger subdomain 1 (FD1) and thumb subdomain (TD). The centre of mass of the residues ranging from 398-581 (FD1) and 816-919 (TD) were considered as the two points for calculating the distance. Figure 5 A and B shows the histogram plot for the distance between the FD1 and TD subdomains of the RdRP for the RdRP-NTPs and the RdRP-nucleotide analogue systems respectively. The distance between the two subdomains for the apo crystal structure with PDB ID 7BV1 is 28.12 Å. The black dotted line shows this distance for reference. It was clearly observed that the largest population in the RdRP-APO system showed the distance to be 28 Å which is considerably near to the experimental value of the RdRP in the apo form. A distance more or less than 28.12 Å indicates opening or closing of the template entry site respectively. In case of the NTPs, the RdRP-ATP and RdRP-CTP, considerable number of conformers showed opening of the template entry site. RdRP-ATP sampled more number of conformers that had the FD1-TD distance more than 28.12 Å. In contrast to RdRP-CTP which sampled more of conformations at a distance of 27 Å indicating closing of the template entry site. The remaining three RdRP-NTP systems, RdRP-TTP, RdRP-GTP and RdRP-UTP showed more number of conformers with distance around 27.5 Å inferring minimal closing of the template entry site. In case of the nucleotide analogues (Figure 5 B), the RdRP-FPV showed maximum number of conformers at a distance of 27 ± 0.2 Å. The RdRP-RBV system showed the population of conformers distributed between 27.7 - 28 Å. These two RdRP systems with nucleotide analogues showed significant population leading to closing of the template entry site. The RdRP-GDV and RdRP-LMD appeared to sample similar number of conformers greater and lower than the experimental 28.12 Å. This suggests that in the presence of these nucleotide analogues, both the opening and closing motions of the template entry site occurred. The RdRP-SBV system showed an exclusive distribution of conformers spanning a range of 25-31 Å followed by RdRP-RDV system with a range of 25-30 Å for the calculated distance. However, in case of the former one more number of conformations were sampled with distance values ranging more than 28.12 Å. The vice-versa was observed for the RdRP-RDV system. These observations suggest that the maximum opening and closing of the template entry site was observed in the RdRP-SBV and RdRP-RDV systems respectively as compared to all the other ligands. The comparison between the RdRP-NTPs and RdRP-nucleotide analogues infers that the analogues exhibited closing of the template entry site better than the NTPs.

Changes in the conserved motifs

The inter-motif distances between the seven conserved motifs, A-G were calculated. Among these distances, the correlation of distances between the motifs A-E and G with F and A-F with

G were observed to show highest correlation (data not shown for other inter-motif distances). Multivariate PCA was performed for the six distances between F and rest of the motifs, and between G and rest of the motifs. The conserved motif F plays a role in ligand binding and motif G plays a role in binding to the RNA template. Figure S3 gives the co-efficient of correlation between the inter-domain distances. Figure 6 gives the projections of each of inter-domain distances along PC 1 and 2. More than 80 % of the variance was captured by the first two PCs. In in case of all the inter-motif distances, where, motif F was involved, namely, AF, BF, CF, DF, and appeared to have similar PC projections, as they were clustered together, depicted in circled region 1 of Figure 6. However, the distance FG, was observed to have a projection completely in a different region of PC 1 and 2 distribution. Figure S3 shows that the correlation of coefficient values were more than 0.85 within the distances AF, BF, CF, and DF. The co-efficient of correlation for distance EF with all the other distances with motif F were significantly lower but had a positive value. However, the negative correlation value for distance FG with all other distances of motif F clearly supports its distinct projection on the first two PCs. The positive and negative correlation values infer similar and different behavior of the parameters under consideration. This may infer that the molecular motions along the motif A-E were significantly in sync with the motif F. Whereas, contrasting molecular movements may have occurred in case of motifs F and G. Similarly, in case of the motif G, the distances AG, BG, CG, DG and EG have positive correlation values greater than 0.75 among one another (Figure S3). These inter-motif distances clustered together when distributed along PC 1 and 2, depicted in circled region 2 of Figure 6. The distance FG too had positive values but they were lower than 0.55. Hence, it can be inferred that the regions that were crucial for template (motif G) and ligand (motif F) binding showed non-synonymous molecular motions with respect to one another. This may also be a resultant effect induced due to the presence of ligand.

Binding interactions of NTPs and analogues

Understanding the binding of the NTPs and nucleotide analogues would help to deduce the factors responsible to induce changes in the conformation of the RdRP molecule. The MM-GBSA free energy of binding between the ligands and the RdRP molecule was calculated. The important residues that contribute to the binding either through hydrogen bond or salt bridge formation were calculated using the PLIP software. The residues from RdRP that were involved in either of the two interactions having a percentage occupancy above 20 were identified. A total of twenty eight such residues were observed to occur within all the RdRP ligand-bound systems. These residues belonged to the finger subdomain 1 (FD1), palm subdomain 1 and 2 and the thumb subdomain. Seven residues namely, LYS 545, SER 549, ALA 550, LYS 551, ARG 553, ALA 554 and ARG 555 belonged to the conserved motif F. Motif F is also known as the ligand binding motif consisting of ten residues, out of which seven were obtained through the simulations too. In addition to these seven residues, seven more residues were observed to interact with both the NTPs as well as nucleotide analogues. Eight residues of RdRP were observed to interact exclusively with NTPs. Similarly, six residues of RdRP, selectively interacted only with the nucleotide analogues. The free energy contribution in binding of these twenty eight residues has been shown in Figure 7 and 8 for the NTPs and the nucleotide analogues respectively. LYS 551 and ARG 555 from the motif F were observed to strongly

contribute in binding to the NTPs as well as the nucleotide analogues. Figure 9 shows the start positions of the ligands and the most interacting residues from the RdRP. In case of the NTPs, the LYS 551 contributed the most in binding to CTP followed by TTP, ATP, GTP and UTP. In case of the nucleotide analogues, LYS 551 contributed the most in binding to SBV followed by RBV, GDV, FPV, RDV and LMD. ARG 555 contributed the most in binding to TTP, followed by GTP, CTP, ATP and UTP within the NTPs. It contributed the most in binding to RDV, followed by LMD, GDV, RBV, FPV and SBV within the nucleotide analogues. ARG 553, was observed to contribute significantly more in binding to the nucleotide analogues as compared to the NTPs. The best free energy contribution in binding by ARG 553 was observed for FPV, followed by SBV, GDV, LMD, RBV and RDV. The SER 759 and ASP 760 of the catalytic site consisting of SER 759, ASP 760 and ASP 761 (SDD) were also observed to contribute in binding to the ligands. SER 814 from the CSQ site from the motif E consisting of CYS 813, SER 814 and GLN 815 is known to bind to the template. However, SER 814 was also one of the residues observed in binding to the ligands. ARG 836 was observed to contribute in binding to all the nucleotide analogues and only TTP among the NTPs. This residue is one of the primer binding residues from the thumb subdomain. This residue was observed to contribute the most in binding of RDV to the RdRP. The role of ARG 836 in binding to the nucleotide analogues and was one of the novel findings obtained from the simulations. The role of this residue in binding to the nucleotide analogues may contribute to the closing of the template entry site.

Capturing overall RdRP flexibility

The flexibility attained by the RdRP molecule was further investigated using the Markov State Modeling (MSM) analysis. MSM analysis helps in statistically sampling the data and capturing changes that occur along a slow or rate limiting step. The parameter along which such a change occurs is known as the collective variable (CV). The co-ordinates of the RdRP were extracted from the ligand-bound and the ligand-free simulations. The CV was selected based on variational approach for Markov processes (VAMP) score. The details of the CV selection have been given in the Supplementary Information SI. The backbone torsion was considered as the CV along which the different states visited by the RdRP molecule were predicted (Supplementary Figure SI.1). In addition to the collective variable (CV), the choice of lag time also holds importance in MSM analysis. Time independent component analysis (tICA) was used for performing the dimensionality reduction. tIC 1 and 2 both showed that two transitions occurred in the entire simulation data (Supplementary Figure SI.2). Figure 10 shows the four distinct clusters obtained along tIC 1 and tIC 2. The color distribution in the Figure 10 indicates the free energy, which is inversely proportional to the number of conformations obtained. Figure 10 also shows the superimposed RdRP structures obtained from these MSM clusters/states on to the apo form of RdRP, PDB ID 7BV1. The backbone RMSD of the MSM states 1, 2, 3 and 4 with reference to 7BV1 was 3.748 Å, 3.899 Å, 3.03 Å, and 4.531 Å respectively. These representatives of each of the states were compared to the known PDB structures of SARS-CoV-2 RdRP. However, in order to understand the characteristics of these four states, the distance between the FD1 and TD subdomain of the RdRP was calculated (Figure 11). This parameter was chosen as in case of the MD simulations, and this distance differentiated well between the RdRP-APO and ligand-bound systems. Figure 10 shows the population obtained for the four

MSM states owing to the distance calculated. It was clearly observed that, the MSM states 1 and 2 showed majority of conformations in the open state of the template entry site. Whereas, the MSM states 3 and 4 showed majority of conformations in the closed state of the template entry site. The black dotted line in the Figure 11 signifies the distance between FD1 and TD for the PDB ID 7BV1. The snapshots of each of the MSM state representatives shown in the Figure 11 have the regions circled which showed significant deviation from the apo form of RdRP, PDB ID 7BV1. The circled regions denoted by A, B, and C were the ones that showed clear deviation from the 7BV1 experimental structure. The region A belongs to the linker region between the NiRAN and finger subdomain 1 of the RdRP. This region was observed to have high RMSF values in the RdRP systems bound to the nucleotide analogues (Figure 3). This region appeared to significantly deviate from the 7BV1 in case of MSM states 1 and 2. The regions denoted by B and C belonged to the TD subdomain of the RdRP. The work done by Appleby and co-workers suggests that there is an opening and closing movement of the β-loop of TD subdomain during template and nucleotide loading step of the replication.[41] Region C circled in Figure 11, denotes this particular β-loop, and a good overlap of this region between the MSM states 1, 2 and 7BV1 was observed. The MSM states 3 and 4 clearly showed the inward movement of this β-loop in comparison to the 7BV1. Similarly, the region denoted as B also showed clear inward movement in case of the MSM states 3 and 4 as compared to 1 and 2. The MSM transition predicted also suggested that the conformations followed the path of state 1 → 2 → 3&4. The findings obtained for RdRP bound to nucleotide analogue systems matched to the states 3 and 4 and the ones for the RdRP-APO and few of the NTP-bound systems matched the states 1 and 2. The inhibitory mechanism proposed in previously reported studies of RdRP very successfully sampled in the simulation data presented here.[40,41] The closing of the template entry site involving the movement of the thumb subdomain's β-loop was observed through the simulations. The linker region between the NiRAN and the finger subdomain 1 may also be playing a crucial role in the opening and closing event of the template entry site.

Ensemble docking of NTPs and nucleotide analogues

The 5 NTPs and 6 nucleotide analogues were flexibly docked using DOCK 6 on each of the four MSM state representatives. Figure 12 shows the grid scores for each of these ligands when docked against the four MSM states. The NTPs namely, ATP, TTP and UTP were observed to dock favorably to all the four states. The nucleotide analogues, namely, LMD, RBV and RDV were observed to dock favorably to all the four states. MSM state 1 was found to favor stable binding of 7 out of the 11 ligands. Whereas, the MSM state 4 was found to favor stable binding of all the ligands except for CTP. The ligand molecules, LMD, SBV, GTP and GDV were observed to bind with the best grid score to the MSM states 1, 2, 3 and 4 respectively. LMD is a cytidine analogue, strongly bound to the RdRP as compared to the CTP. SBV, a uridine analogue bound better than UTP in three of the states namely, 2, 3 and 4. Although, GTP was observed to bind strongly to the MSM state 3 in comparison to others, unfavorable binding was observed for the states 1 and 2. Its analogue RBV, was observed to bind strongly with states 1 and 2 and with 0.6 kcal/mole higher value of grid score from GTP for state 4. The states 3 and 4 exhibited favorable binding of most of the ligands and with significantly better grid scores compared to the other two states. The hydrogen bonding and salt bridge interactions between the docked

ligand molecules and the RdRP residues were observed. LYS 551, ARG 553 and ARG 555 from the conserved motif F were observed to bind and form salt bridge with the ligand molecules in all the four states. ALA 550 was observed to form interactions with the ligand molecules only in the states 3 and 4. ARG 836 from the TD subdomain was observed to form hydrogen bonding or salt bridge interactions with the only the nucleotide analogues GDV, LMD, RBV, and RDV and single NTP, TTP.

Conclusion

The MD simulation studies on the ligand-free, NTP-bound and nucleotide analog-bound RdRP reveal the role of functionally important residues and significantly correlated molecular motions responsible for inducing a probable inhibitory effect. The principal component analysis suggests significant conformational changes in the finger subdomain 1 and thumb subdomain irrespective of presence of the NTPs or analogues. The linker region (residue range 260-370), structurally forming a β-hairpin loop between the NiRAN and finger subdomain 1 proved to fluctuate considerably in the presence of nucleotide analogues. The principal component analysis on inter-motif distances also revealed the correlated motions of the motifs A-E with F and G. The molecular motions of motif F and G were observed to be negatively correlated. The former one being crucial for inhibitor binding and the latter one for binding of the RNA template. The non-synchronous motions of motifs F and G with respect to one another may have affected the conformation of the template entry site. As both these motifs, are present in the regions flanking the template entry site. The opening and closing of the template entry site was clearly visible in all the simulations. The nucleotide analogues tend to sample more conformations depicting closure of the template entry site. Hydrogen bond and salt bridge calculations disclosed twenty-eight residues to be involved in strongly interacting with either of the eleven ligand molecules. The residues LYS 551, ARG 553, ARG 555 were observed to actively involved in forming significant stable interactions with all the ligand molecules. ARG 836 was identified to be specifically involved in binding to the nucleotide analogues. The Markov state modeling analysis exhibited the probable inhibitory mechanism involving the closing of the template entry site in the presence of nucleotide analogues. The four states obtained clearly demarcated the inhibitory state of RdRP based on the distance between the finger subdomain 1 and thumb subdomain, the flanking regions near the template entry site. The ensemble docking of these states by the NTPs and nucleotide analogues supported the findings obtained through simulations about the critical residues involved in the ligand interactions. These findings on domain movements, correlated motif motions and ligand interacting residues would prove to be helpful in designing molecules that exhibit inhibitory effect on the RdRP of the viruses.

Conflict of Interest Statement

The authors declare no conflict of interest.

Acknowledgement

The authors would like to acknowledge the National Supercomputing Mission, Ministry of Electronics and Information Technology (MeitY), Government of India for funding for this work. The authors would also like to thank the National PARAM Supercomputing Facility (NPSF) and Bioinformatics Resource and Applications Facility (BRAF) at Centre for Development of Advanced Computing (C-DAC), Pune for providing the computational infrastructure.

References

1. Pene F, Merlat A, Vabret A, Rozenberg F, Buzyn A, Dreyfus F, Cariou A, Freymuth F, Lebon P. Coronavirus 229E-related pneumonia in immunocompromised patients. *Clin Infect Dis.* 2003 Oct 1;37(7):929-32. doi: 10.1086/377612. Epub 2003 Sep 8. PMID: 13130404; PMCID: PMC7107892.

2. van der Hoek L. Human coronaviruses: what do they cause? *Antivir Ther.* 2007;12(4 Pt B):651-8. PMID: 17944272.

3. Lim YX, Ng YL, Tam JP, Liu DX. Human Coronaviruses: A Review of Virus-Host Interactions. *Diseases.* 2016 Jul 25;4(3):26. doi: 10.3390/diseases4030026. PMID: 28933406; PMCID: PMC5456285.

4. Graham RL, Donaldson EF, Baric RS. A decade after SARS: strategies for controlling emerging coronaviruses. *Nat Rev Microbiol.* 2013 Dec;11(12):836-48. doi: 10.1038/nrmicro3143. Epub 2013 Nov 11. PMID: 24217413; PMCID: PMC5147543.

5. Kim Y, Cheon S, Min CK, Sohn KM, Kang YJ, Cha YJ, Kang JI, Han SK, Ha NY, Kim G, Aigerim A, Shin HM, Choi MS, Kim S, Cho HS, Kim YS, Cho NH. Spread of Mutant Middle East Respiratory Syndrome Coronavirus with Reduced Affinity to Human CD26 during the South Korean Outbreak. *mBio.* 2016 Mar 1;7(2):e00019. doi: 10.1128/mBio.00019-16. PMID: 26933050; PMCID: PMC4810480.

6. Hemida MG, Chu DKW, Perera RAPM, Ko RLW, So RTY, Ng BCY, Chan SMS, Chu S, Alnaeem AA, Alhammadi MA, Webby RJ, Poon LLM, Balasuriya UBR, Peiris M. Coronavirus infections in horses in Saudi Arabia and Oman. *Transbound Emerg Dis.* 2017 Dec;64(6):2093-2103. doi: 10.1111/tbed.12630. Epub 2017 Mar 13. PMID: 28296228; PMCID: PMC7169745.

7. Wang H, Li X, Li T, Zhang S, Wang L, Wu X, Liu J. The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis.* 2020 Apr 24:1–7. doi: 10.1007/s10096-020-03899-4. Epub ahead of print. PMID: 32333222; PMCID: PMC7180649.

8. Gaurav A, Al-Nema M. Polymerases of Coronaviruses: Structure, Function, and Inhibitors. *Viral Polymerases.* 2019:271–300. doi: 10.1016/B978-0-12-815422-9.00010-3. Epub 2018 Nov 2. PMCID: PMC7149517.

9. Jia H, Gong P. A Structure-Function Diversity Survey of the RNA-Dependent RNA Polymerases From the Positive-Strand RNA Viruses. *Front Microbiol.* 2019 Aug 22;10:1945. doi: 10.3389/fmicb.2019.01945. PMID: 31507560; PMCID: PMC6713929.

10. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, Wang Y, Guo X. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses.* 2020 Feb 22;12(2):244. doi: 10.3390/v12020244. PMID: 32098422; PMCID: PMC7077191.

11. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science.* 2020 May 15;368(6492):779-782. doi: 10.1126/science.abb7498. Epub 2020 Apr 10. PMID: 32277040; PMCID: PMC7164392.

12. Srinivasan S, Cui H, Gao Z, Liu M, Lu S, Mkandawire W, Narykov O, Sun M, Korkin D. Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. *Viruses.* 2020 Mar 25;12(4):360. doi: 10.3390/v12040360. PMID: 32218151; PMCID: PMC7232164.

13. Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res.* 2003 Dec 15;31(24):7117-30. doi: 10.1093/nar/gkg916. PMID: 14654687; PMCID: PMC291860.

14. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, Wang Y, Guo X. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses.* 2020 Feb 22;12(2):244. doi: 10.3390/v12020244. PMID: 32098422; PMCID: PMC7077191.

15. Zhu W, Chen CZ, Gorshkov K, Xu M, Lo DC, Zheng W. RNA-Dependent RNA Polymerase as a Target for COVID-19 Drug Discovery. *SLAS Discov.* 2020 Jul 13:2472555220942123. doi: 10.1177/2472555220942123. Epub ahead of print. PMID: 32660307.

16. Aftab SO, Ghouri MZ, Masood MU, Haider Z, Khan Z, Ahmad A, Munawar N. Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. *J Transl Med.* 2020 Jul 7;18(1):275. doi: 10.1186/s12967-020-02439-0. PMID: 32635935; PMCID: PMC7339606.

17. Wang Y, Anirudhan V, Du R, Cui Q, Rong L. RNA-dependent RNA polymerase of SARS-CoV-2 as a therapeutic target. *J Med Virol.* 2020 Jul 7. doi: 10.1002/jmv.26264. Epub ahead of print. PMID: 32633831.

18. Prajapat M, Sarma P, Shekhar N, Prakash A, Avti P, Bhattacharyya A, Kaur H, Kumar S, Bansal S, Sharma AR, Medhi B. Update on the target structures of SARS-CoV-2: A systematic review. *Indian J Pharmacol.* 2020 Mar-Apr;52(2):142-149. doi: 10.4103/ijp.IJP_338_20. Epub 2020 Jun 3. PMID: 32565603; PMCID: PMC7282679.

19. Chien M, Anderson TK, Jockusch S, Tao C, Li X, Kumar S, Russo JJ, Kirchdoerfer RN, Ju J. Nucleotide Analogues as Inhibitors of SARS-CoV-2 Polymerase, a Key Drug Target for COVID-19. *J Proteome Res.* 2020 Jul 21. doi: 10.1021/acs.jproteome.0c00392. Epub ahead of print. PMID: 32692185.

20. van Hemert FJ, Zaaijer HL, Berkhout B. In silico prediction of ebolavirus RNA polymerase inhibition by specific combinations of approved nucleotide analogues. *J Clin Virol.* 2015 Dec;73:89-94. doi: 10.1016/j.jcv.2015.10.020. Epub 2015 Nov 5. PMID: 26587786.

21. Parang K, El-Sayed NS, Kazeminy AJ, Tiwari RK. Comparative Antiviral Activity of Remdesivir and Anti-HIV Nucleoside Analogs Against Human Coronavirus 229E (HCoV-229E). *Molecules.* 2020 May 17;25(10):2343. doi: 10.3390/molecules25102343. PMID: 32429580; PMCID: PMC7287735.

22. Konkolova E, Dejmek M, Hřebabecký H, Šála M, Böserle J, Nencka R, Boura E. Remdesivir triphosphate can efficiently inhibit the RNA-dependent RNA polymerase from various flaviviruses. *Antiviral Res.* 2020 Aug 4:104899. doi: 10.1016/j.antiviral.2020.104899. Epub ahead of print. PMID: 32763313.

23. De Clercq E. New Nucleoside Analogues for the Treatment of Hemorrhagic Fever Virus Infections. *Chem Asian J.* 2019 Nov 18;14(22):3962-3968. doi: 10.1002/asia.201900841. Epub 2019 Aug 7. PMID: 31389664; PMCID: PMC7159701.

24. Li Z, Wang X, Cao D, Sun R, Li C, Li G. Rapid review for the anti-coronavirus effect of remdesivir. *Drug Discov Ther.* 2020;14(2):73-76. doi: 10.5582/ddt.2020.01015. PMID: 32378648.

25. Liang C, Tian L, Liu Y, Hui N, Qiao G, Li H, Shi Z, Tang Y, Zhang D, Xie X, Zhao X. A promising antiviral candidate drug for the COVID-19 pandemic: A mini-review of remdesivir. *Eur J Med Chem.* 2020 Sep 1;201:112527. doi: 10.1016/j.ejmech.2020.112527. Epub 2020 Jun 6. PMID: 32563812.

26. Gordon CJ, Tchesnokov EP, Woolner E, Perry JK, Feng JY, Porter DP, Götte M. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem.* 2020 May 15;295(20):6785-6797. doi: 10.1074/jbc.RA120.013679. Epub 2020 Apr 13. PMID: 32284326; PMCID: PMC7242698.

27. Ledford H. Hopes rise for coronavirus drug remdesivir. *Nature.* 2020 Apr 29. doi: 10.1038/d41586-020-01295-8. Epub ahead of print. PMID: 32350436.

28. Ko WC, Rolain JM, Lee NY, Chen PL, Huang CT, Lee PI, Hsueh PR. Arguments in favour of remdesivir for treating SARS-CoV-2 infections. *Int J Antimicrob Agents.* 2020 Apr;55(4):105933. doi: 10.1016/j.ijantimicag.2020.105933. Epub 2020 Mar 6. PMID: 32147516; PMCID: PMC7135364.

29. Tchesnokov EP, Feng JY, Porter DP, Götte M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses.* 2019 Apr 4;11(4):326. doi: 10.3390/v11040326. PMID: 30987343; PMCID: PMC6520719.

30. Delang L, Abdelnabi R, Neyts J. Favipiravir as a potential countermeasure against neglected and emerging RNA viruses. *Antiviral Res.* 2018 May;153:85-94. doi: 10.1016/j.antiviral.2018.03.003. Epub 2018 Mar 7. PMID: 29524445.

31. Furuta Y, Komeno T, Nakamura T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc Jpn Acad Ser B Phys Biol Sci.* 2017;93(7):449-463. doi: 10.2183/pjab.93.027. PMID: 28769016; PMCID: PMC5713175.

32. Khalili JS, Zhu H, Mak NSA, Yan Y, Zhu Y. Novel coronavirus treatment with ribavirin: Groundwork for an evaluation concerning COVID-19. *J Med Virol.* 2020 Jul;92(7):740-746. doi: 10.1002/jmv.25798. Epub 2020 Apr 10. PMID: 32227493; PMCID: PMC7228408.

33. Cattaneo D, Capetti A, Rizzardini G. Drug-drug interactions of a two-drug regimen of dolutegravir and lamivudine for HIV treatment. *Expert Opin Drug Metab Toxicol.* 2019 Mar;15(3):245-252. doi: 10.1080/17425255.2019.1577821. Epub 2019 Feb 11. PMID: 30704313.

34. Jácome R, Campillo-Balderas JA, Ponce de León S, Becerra A, Lazcano A. Sofosbuvir as a potential alternative to treat the SARS-CoV-2 epidemic. *Sci Rep.* 2020 Jun 9;10(1):9294. doi: 10.1038/s41598-020-66440-9. PMID: 32518317; PMCID: PMC7283245.

35. Elfiky AA. SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: an in silico perspective. *J Biomol Struct Dyn.* 2020 May 6:1-9. doi: 10.1080/07391102.2020.1761882. Epub ahead of print. PMID: 32338164; PMCID: PMC7222627.

36. Ruan Z, Liu C, Guo Y, He Z, Huang X, Jia X, Yang T. SARS-CoV-2 and SARS-CoV: Virtual screening of potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12). *J Med Virol.* 2020 Jun 24:10.1002/jmv.26222. doi: 10.1002/jmv.26222. Epub ahead of print. PMID: 32579254; PMCID: PMC7361265.

37. Balasubramaniam M, Reis RJS. Computational target-based drug repurposing of elbasvir, an antiviral drug predicted to bind multiple SARS-CoV-2 proteins. *ChemRxiv [Preprint].* 2020 Apr 8. doi: 10.26434/chemrxiv.12084822. PMID: 32511290; PMCID: PMC7263767.

38. Elfiky AA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sci.* 2020 Jul 15;253:117592. doi: 10.1016/j.lfs.2020.117592. Epub 2020 Mar 25. PMID: 32222463; PMCID: PMC7102646.

39. Koulgi S, Jani V, Uppuladinne V N M, Sonavane U and Joshi R. Remdesivir-bound and ligand-free simulations reveal the probable mechanism of inhibiting the RNA dependent RNA polymerase of severe acute respiratory syndrome coronavirus 2. *RSC Adv.*, 2020, 10, 26792-26803. DOI: 10.1039/D0RA04743K

40. Moustafa IM, Shen H, Morton B, Colina CM, Cameron CE. Molecular dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity. *J Mol Biol.* 2011 Jul 1;410(1):159-81. doi: 10.1016/j.jmb.2011.04.078. Epub 2011 May 7. PMID: 21575642; PMCID: PMC3114172.

41. Appleby TC, Perry JK, Murakami E, Barauskas O, Feng J, Cho A, Fox D 3rd, Wetmore DR, McGrath ME, Ray AS, Sofia MJ, Swaminathan S, Edwards TE. Viral replication. Structural

basis for RNA replication by the hepatitis C virus polymerase. *Science.* 2015 Feb 13;347(6223):771-5. doi: 10.1126/science.1259210. PMID: 25678663.

42. Jani, Vinod & Sonavane, Uddhavesh & Joshi, Rajendra. (2019). Detecting early stage structural changes in wild type, pathogenic and non-pathogenic prion variants using Markov state model. *RSC Adv.* 9. 14567-14579. 10.1039/C9RA01507H.

43. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. Markov models of molecular kinetics: generation and validation. *J Chem Phys.* 2011 May 7;134(17):174105. doi: 10.1063/1.3565032. PubMed PMID: 21548671.

44. Chodera JD, Noé F. Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol.* 2014 Apr;25:135-44. doi: 10.1016/j.sbi.2014.04.002. Epub 2014 May 16. Review. PubMed PMID: 24836551; PubMed Central PMCID: PMC4124001.

45. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys.* 2007 Apr 21;126(15):155101. PubMed PMID: 17461665.

46. Sirur A, De Sancho D, Best RB. Markov state models of protein misfolding. *J Chem Phys.* 2016 Feb 21;144(7):075101. doi: 10.1063/1.4941579. PubMed PMID: 26897000; PubMed Central PMCID: PMC4760979.

47. D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, AMBER 2016, University of California, San Francisco. 2016.

48. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, Case DA, Kuntz ID, Rizzo RC. DOCK 6: Impact of new features and current docking performance. *J Comput Chem.* 2015 Jun 5;36(15):1132-56. doi: 10.1002/jcc.23905. PMID: 25914306; PMCID: PMC4469538.

49. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput.* 2015 Aug 11;11(8):3696-713. doi: 10.1021/acs.jctc.5b00255. Epub 2015 Jul 23. PMID: 26574453; PMCID: PMC4821407.

50. Peters MB, Yang Y, Wang B, Füsti-Molnár L, Weaver MN, Merz KM Jr. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput.* 2010 Sep 14;6(9):2935-2947. doi: 10.1021/ct1002626. PMID: 20856692; PMCID: PMC2941202.

51. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model.* 2006 Oct;25(2):247-60. doi: 10.1016/j.jmgm.2005.12.005. Epub 2006 Feb 3. PMID: 16458552.

52. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem.* 2004 Jul 15;25(9):1157-74. doi: 10.1002/jcc.20035. Erratum in: J Comput Chem. 2005 Jan 15;26(1):114. PMID: 15116359.

53. Roe DR, Cheatham TE 3rd. Parallelization of CPPTRAJ enables large scale analysis of molecular dynamics trajectory data. *J Comput Chem.* 2018 Sep 30;39(25):2110-2117. doi: 10.1002/jcc.25382. Epub 2018 Oct 3. PMID: 30368859; PMCID: PMC7313716.

54. Miller BR 3rd, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput.* 2012 Sep 11;8(9):3314-21. doi: 10.1021/ct300418h. Epub 2012 Aug 16. PMID: 26605738.

55. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W443-7. doi: 10.1093/nar/gkv315. Epub 2015 Apr 14. PMID: 25873628; PMCID: PMC4489249.

56. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

57. M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé: PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models, *J. Chem. Theory Comput.* 11, 5525-5542 (2015)

Figure Legends

Figure 1: RdRP-APO model considered for the study. The subdomains finger NiRAN, linker region, 1 and 2, palm, thumb.

Figure 2: The 2D representation of the five NTPs namely, ATP, GTP, TTP, CTP and UTP and six nucleotide analogues namely, FPV-triphosphate, GDV-triphosphate, LMD-triphosphate, RBV-triphosphate, RDV-triphosphate and SBV-triphosphate

Figure 3: Residues-wise RMSF for the RdRP-APO, RdRP-ATP, RdRP-TTP, RdRP-GTP, RdRP-CTP and RdRP-UTP along PC 1 (black), PC 2 (red) and PC 3 (green)

Figure 4: Residues-wise RMSF for the RdRP-FPV, RdRP-GDV, RdRP-LMD, RdRP-RBV, RdRP-RDV and RdRP-SBV along PC 1 (black), PC 2 (red) and PC 3 (green)

Figure 5: Population of conformers along the distance between the center of mass of finger subdomain (FD1) and thumb subdomain (TD) for (A) RdRP-NTPs and (B) RdRP-nucleotide analogs. Dotted line is the value of this distance for 7BV1

Figure 6: Projections of the inter-motif distances, AF, BF, CF, DF, EF, FG, AG, BG, CG, DG and EG along the first two principal components PC 1 and 2

Figure 7: Residue-wise free energy contribution in binding to the respective ligand molecules for the (A) RdRP-ATP, (B) RdRP-TTP, (C) RdRP-GTP, (D) RdRP-CTP and (E) RdRP-UTP systems.

Figure 8: Residue-wise free energy contribution in binding to the respective ligand molecules for the (A) RdRP-FPV, (B) RdRP-GDV, (C) RdRP-LMD, (D) RdRP-RBV, (E) RdRP-RDV and (F) RdRP-SBV systems.

Figure 9: Start positions of the ligands with the most interacting residues namely, LYS 551, ARG 553, ARG 555, LYS 798 and ARG 836

Figure 10: (A) Four macrostates (MSM states) obtained along the time independent component 1 and 2 using MSM analysis. (B) Representative of each state (color coded) superimposed on 7BV1 denoting the template entry site, finger subdomain 1 (FD1) and thumb subdomain (TD) regions.

Figure 11: Population of conformers along the distance between finger subdomain 1 and thumb subdomain for the four MSM states. Structures represent each state superimposed on 7BV1. Circled regions A, B and C are the most deviating regions with respect to 7BV1

Figure 12: Grid scores obtained for all the ligands after docking to the four representative RdRP structures of MSM states 1 (blue), 2 (red), 3 (green) and 4 (orange)

Table 1: RdRP systems studied and the original indication of the prodrug form of the nucleotide analogues

System	Notation used	Analogue of this NTP	Original Indication
Ligand-free	RdRP-APO	-	-
Natural nucleotides (NTPs)			
Adenosine-5'-triphosphate	RdRP-ATP	-	-
Thymidine-5'-triphosphate	RdRP-TTP	-	-
Guanosine-5'-triphosphate	RdRP-GTP	-	-
Cytidine-5'-triphosphate	RdRP-CTP	-	-
Uridine 5'-triphosphate	RdRP-UTP	-	-
Nucleotide analogues (antivirals)			
Favipiravir triphosphate	RdRP-FPV	GTP	Treat influenza in Japan (Avigan)
Galidesivir triphosphate	RdRP-GDV	ATP	Broad spectrum antiviral
Lamivudine triphosphate	RdRP-LMD	CTP	Anti-retroviral used for HIV/Hepatitis B
Ribavirin triphosphate	RdRP-RBV	GTP	Antiviral used against RSV, Hepatitis C and viral hemorragic fevers
Remdesivir triphosphate	RdRP-RDV	ATP	Treat EBOLA infection
Sofosbuvir triphosphate	RdRP-SBV	UTP	Antiviral used against Hepatitis C

Figure 1

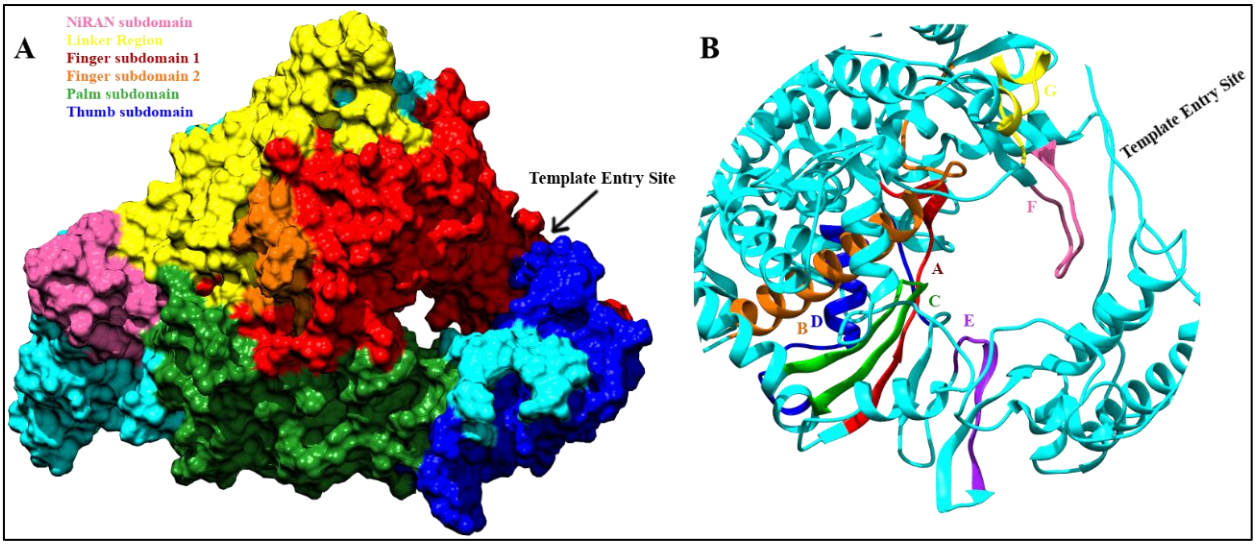

A
NiRAN subdomain
Linker Region
Finger subdomain 1
Finger subdomain 2
Palm subdomain
Thumb subdomain
Template Entry Site
B
G
Template Entry Site
F
A
C
B
D
E

Figure 2

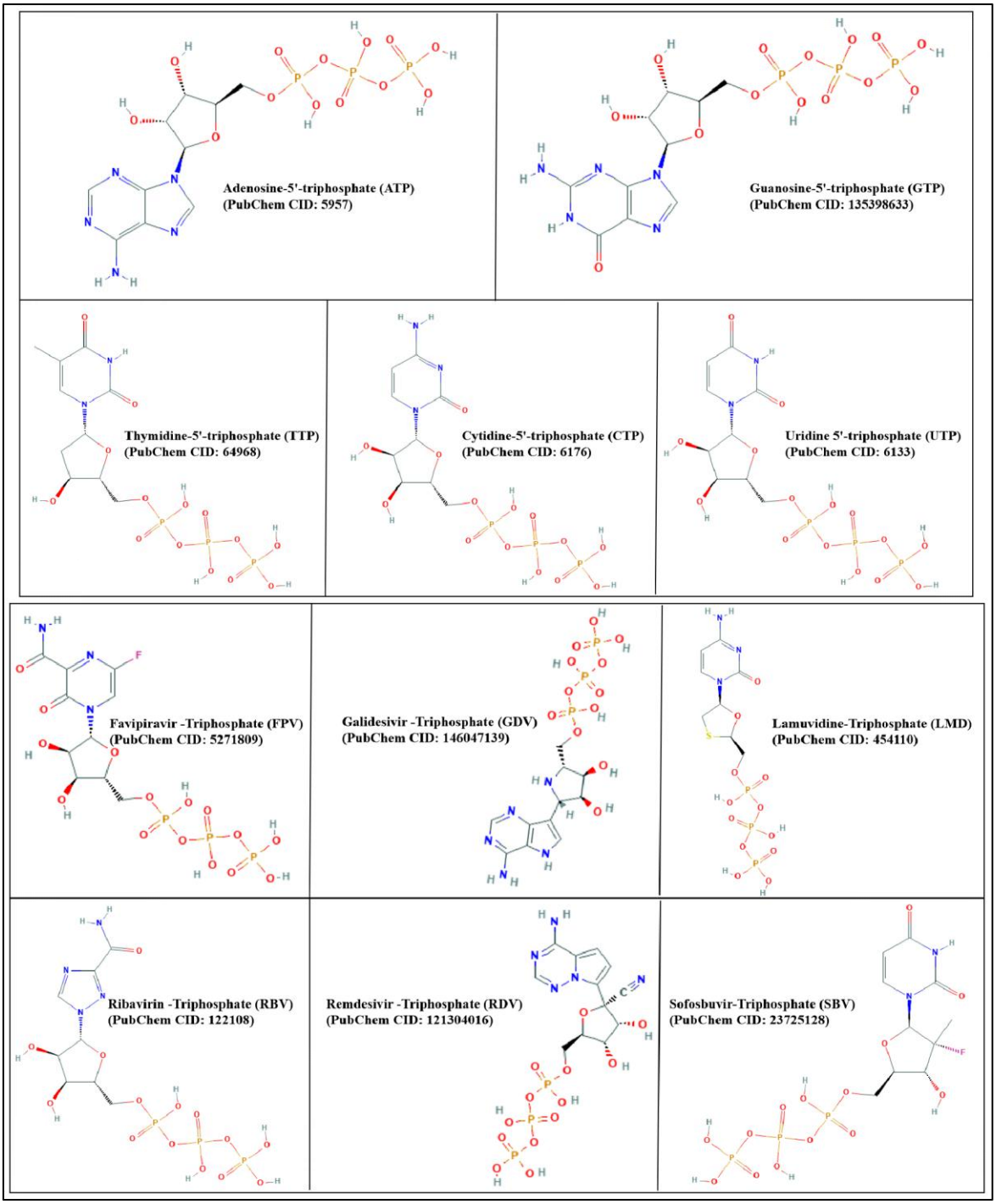

Adenosine-5'-triphosphate (ATP)
(PubChem CID: 5957)
Guanosine-5'-triphosphate (GTP)
(PubChem CID: 135398633)
Thymidine-5'-triphosphate (TTP)
(PubChem CID: 64968)
Cytidine-5'-triphosphate (CTP)
(PubChem CID: 6176)
Uridine 5'-triphosphate (UTP)
(PubChem CID: 6133)
Favipiravir -Triphosphate (FPV)
(PubChem CID: 5271809)
Galidesivir -Triphosphate (GDV)
(PubChem CID: 146047139)
Lamuvidine-Triphosphate (LMD)
(PubChem CID: 454110)
Ribavirin -Triphosphate (RBV)
(PubChem CID: 122108)
Remdesivir -Triphosphate (RDV)
(PubChem CID: 121304016)
Sofosbuvir-Triphosphate (SBV)
(PubChem CID: 23725128)

Figure 3

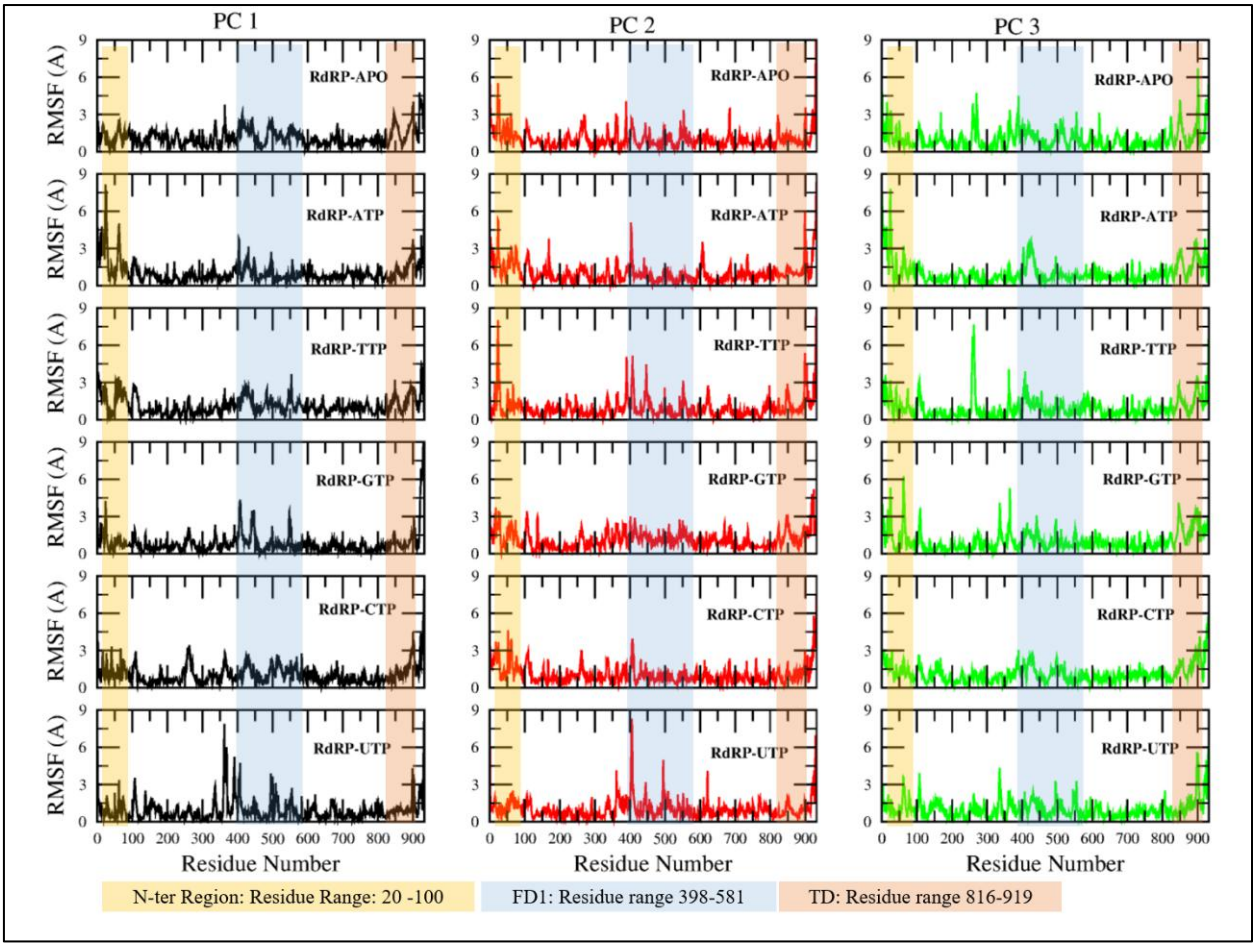

PC 1
PC 2
PC 3
RMSF (A)
RdRP-APO
RdRP-ATP
RdRP-TTP
RdRP-GTP
RdRP-CTP
RdRP-UTP
0 3 6 9
0 100 200 300 400 500 600 700 800 900
Residue Number
N-ter Region: Residue Range: 20 -100
FD1: Residue range 398-581
TD: Residue range 816-919

Figure 4

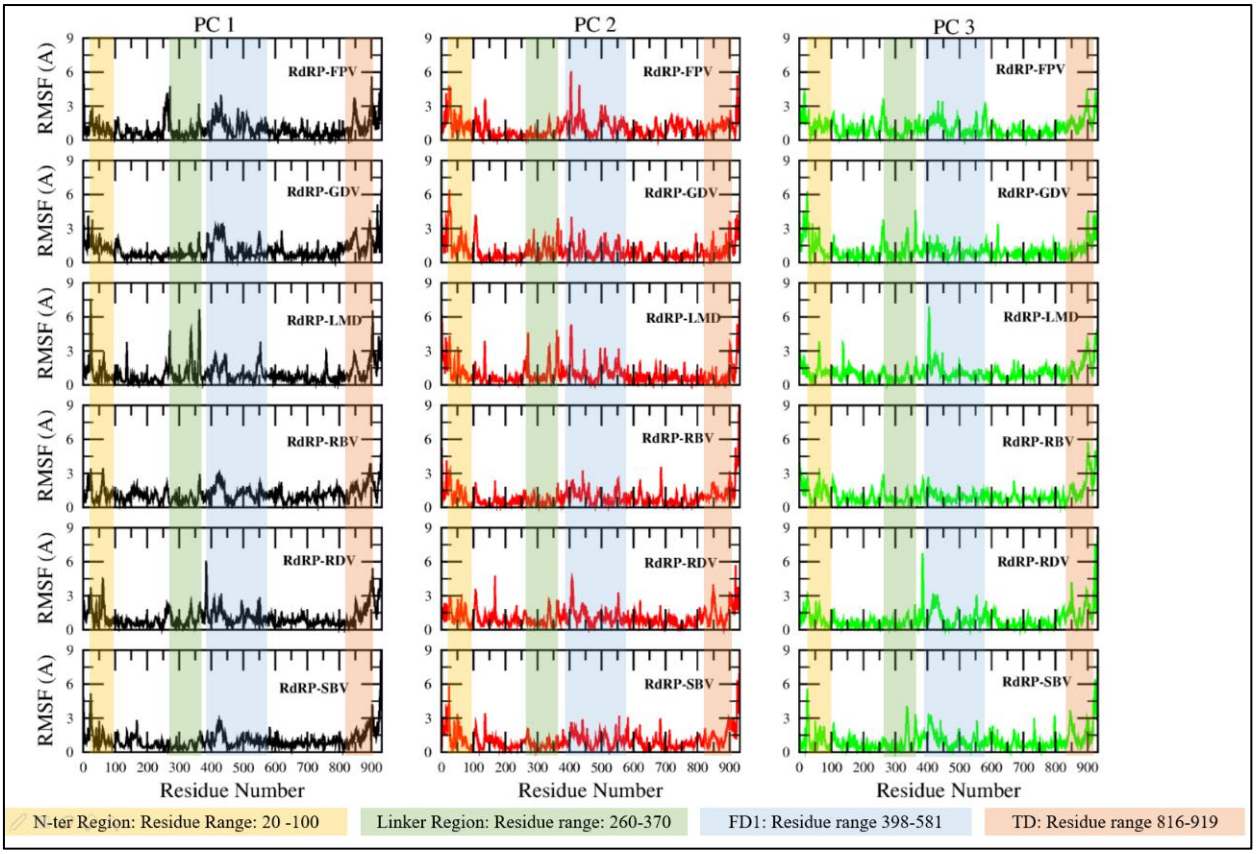

PC 1
PC 2
PC 3
RMSF (A)
RdRP-FPV
RdRP-GDV
RdRP-LMD
RdRP-RBV
RdRP-RDV
RdRP-SBV
0 3 6 9
0 100 200 300 400 500 600 700 800 900
Residue Number
N-ter Region: Residue Range: 20 -100
Linker Region: Residue range: 260-370
FD1: Residue range 398-581
TD: Residue range 816-919

Figure 5

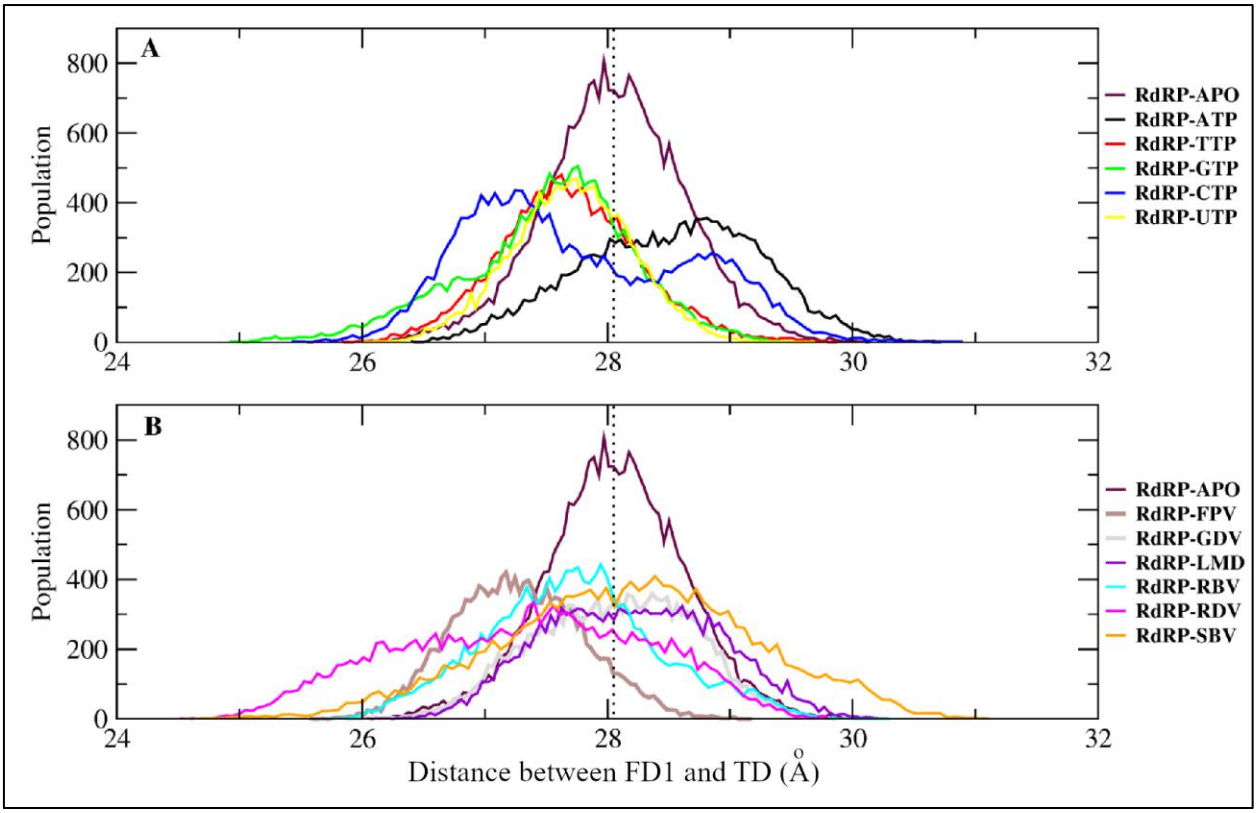

A
B
Population
800
600
400
200
0
24
26
28
30
32
RdRP-APO
RdRP-ATP
RdRP-TTP
RdRP-GTP
RdRP-CTP
RdRP-UTP
RdRP-FPV
RdRP-GDV
RdRP-LMD
RdRP-RBV
RdRP-RDV
RdRP-SBV
Distance between FD1 and TD (Å)

Figure 6

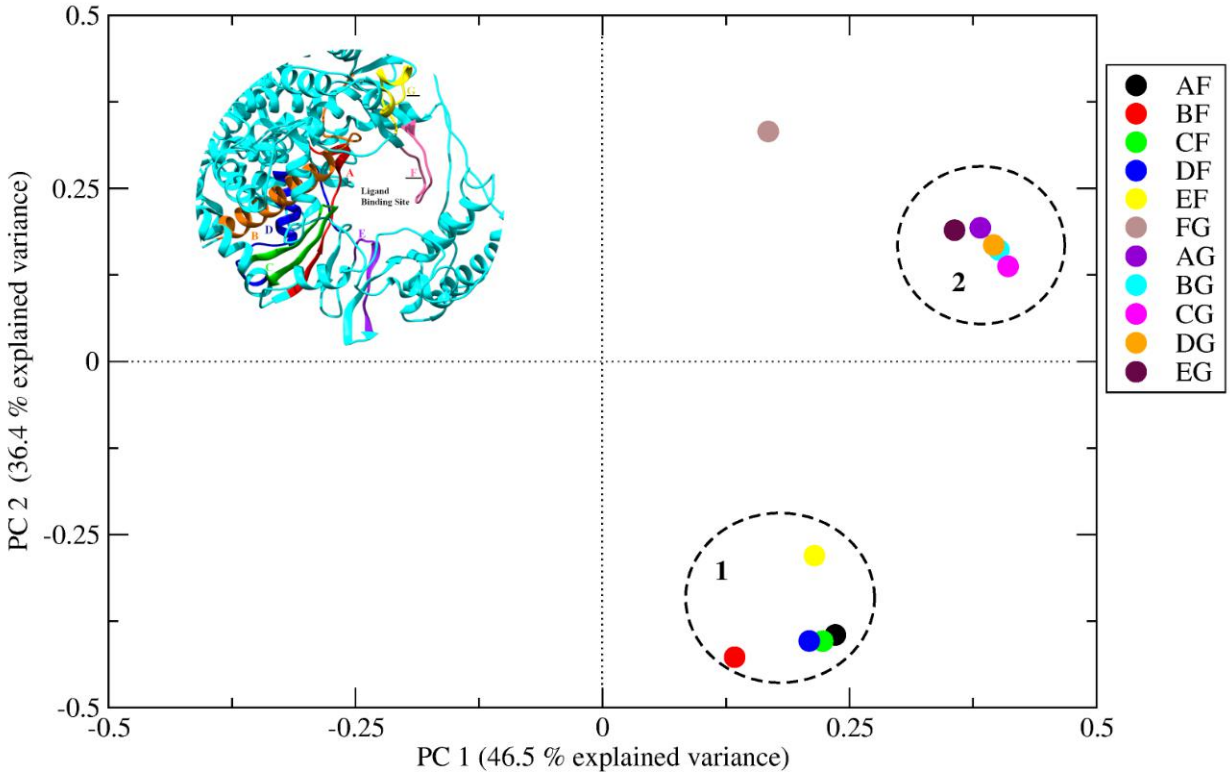

0.5
0.25
0
-0.25
-0.5
-0.5
-0.25
0
0.25
0.5
PC 2 (36.4 % explained variance)
PC 1 (46.5 % explained variance)
Ligand Binding Site
1
2
AF
BF
CF
DF
EF
FG
AG
BG
CG
DG
EG

Figure 7

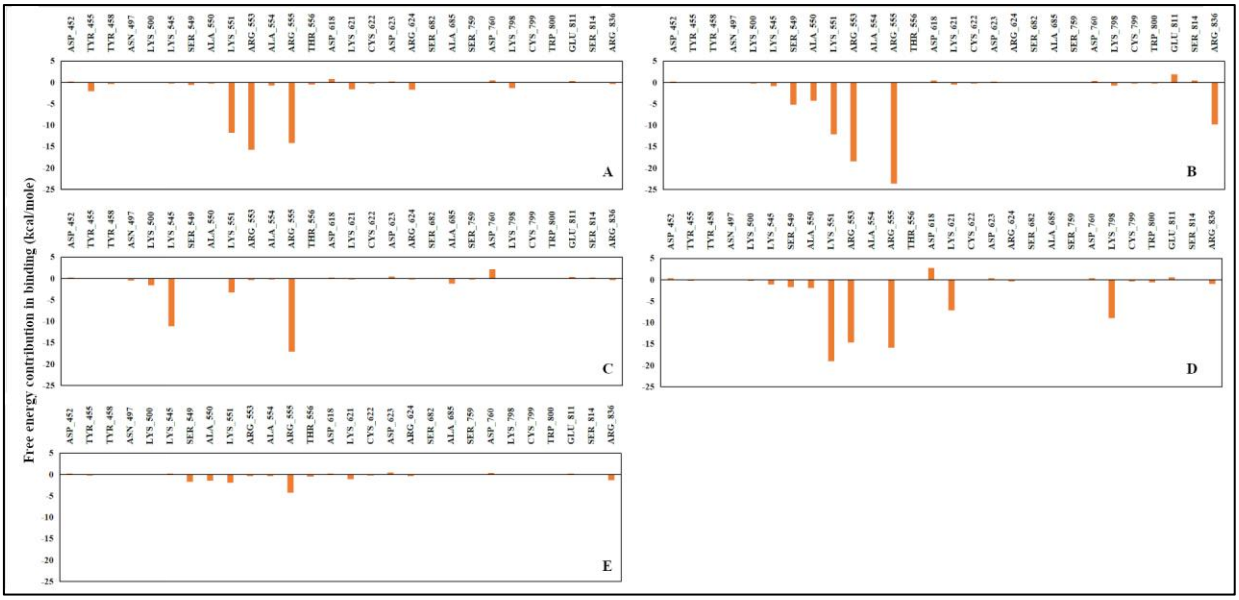

Free energy contribution in binding (kcal/mole)
ASP_452 TYR_455 TYR_458 ASN_497 LYS_500 LYS_545 SER_549 ALA_550 LYS_551 ARG_553 ALA_554 ARG_555 THR_556 ASP_618 LYS_621 CYS_622 ASP_623 ARG_624 SER_682 ALA_685 SER_759 ASP_760 LYS_798 CYS_799 TRP_800 GLU_811 SER_814 ARG_836
5 0 -5 -10 -15 -20 -25
A
B
C
D
E

Figure 8

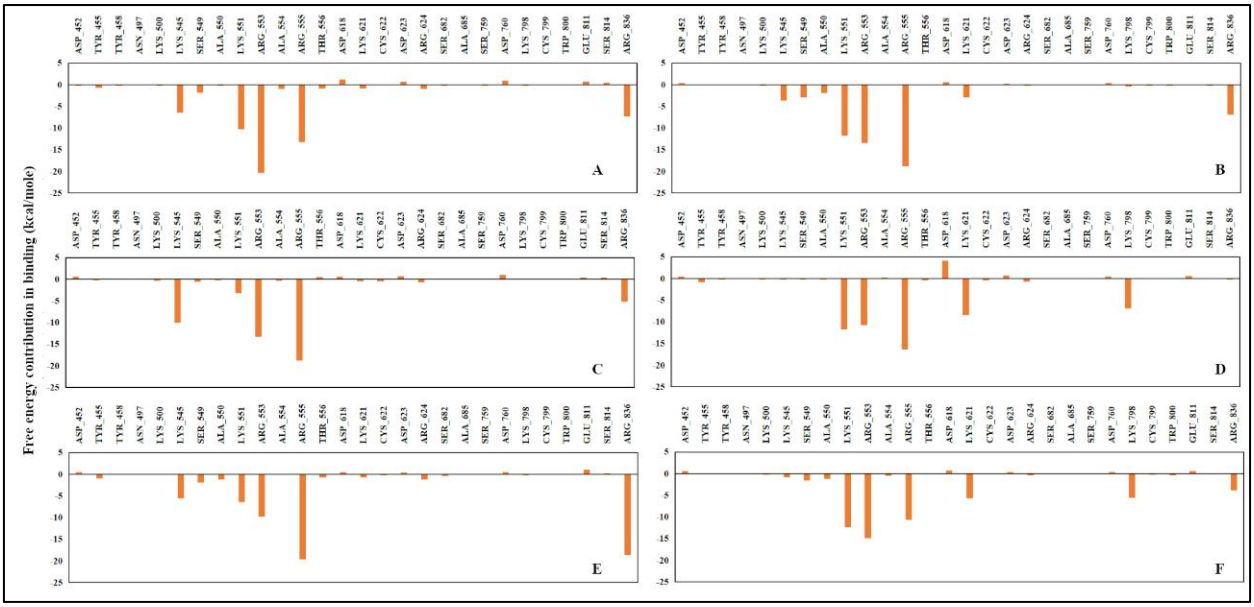


Figure 9

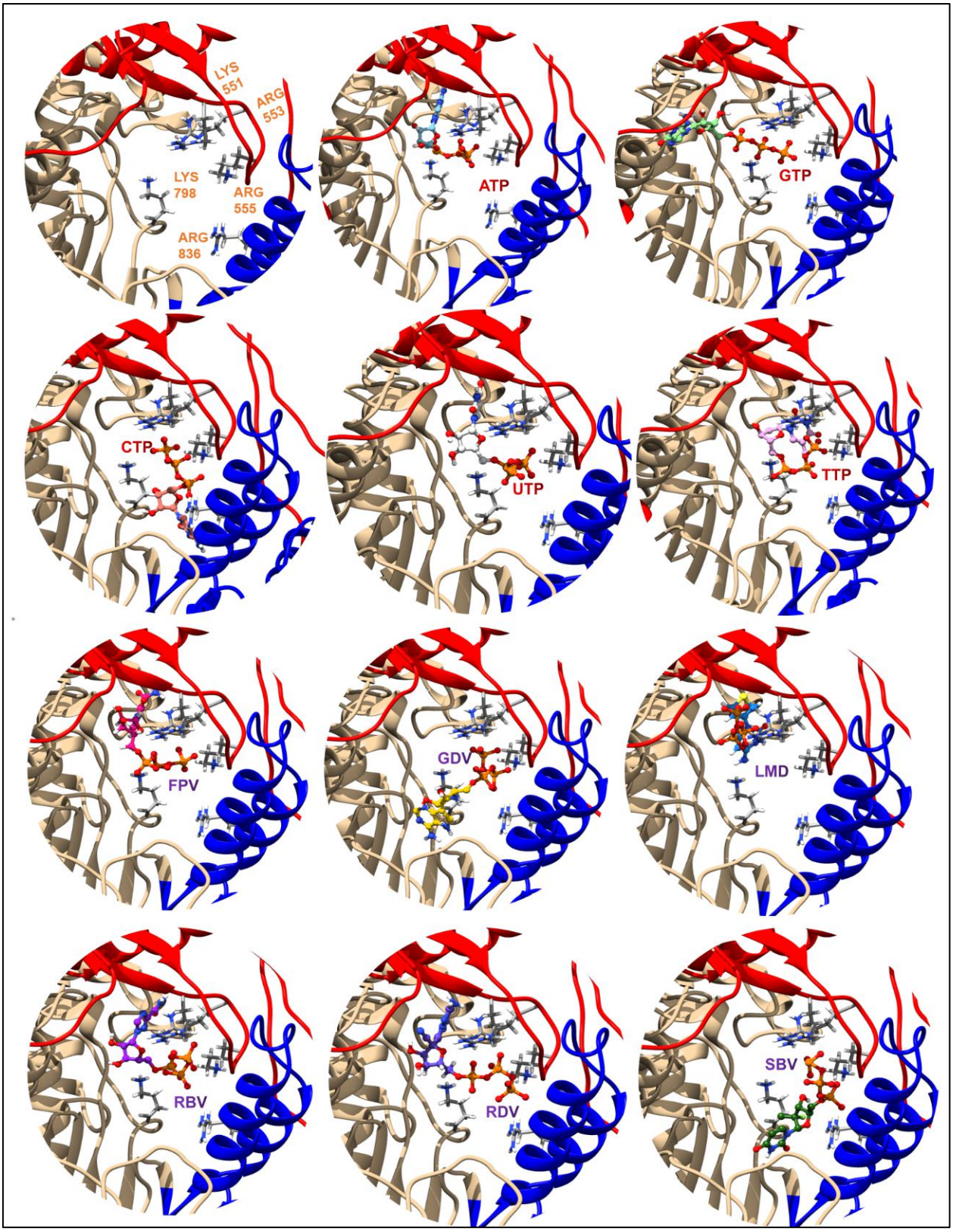

LYS 551
ARG 553
LYS 798
ARG 555
ARG 836
ATP
GTP
CTP
UTP
TTP
FPV
GDV
LMD
RBV
RDV
SBV

Figure 10

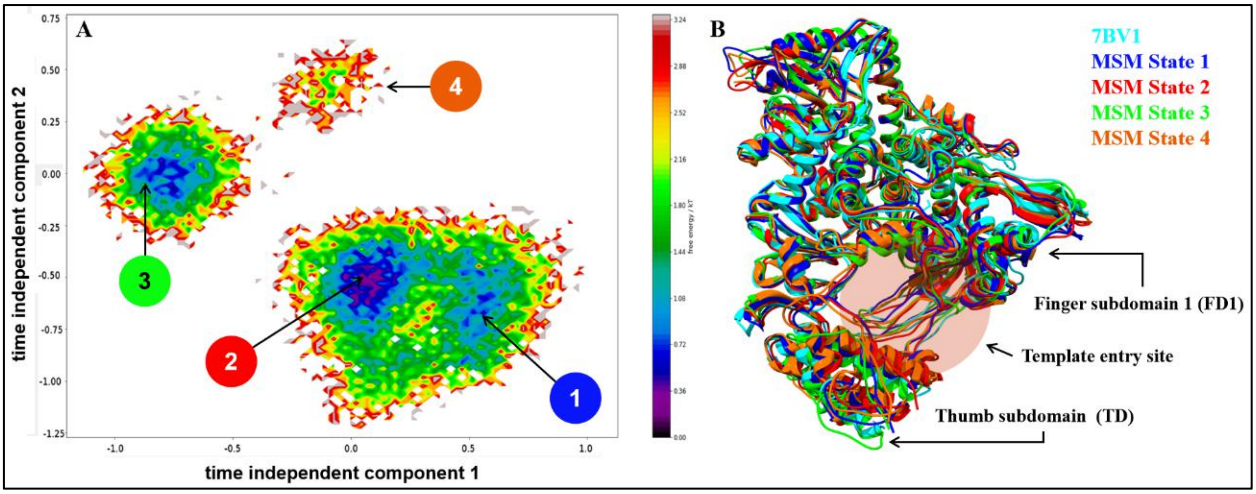

A
B
time independent component 2
time independent component 1
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-1.0
-0.5
0.0
0.5
1.0
1
2
3
4
7BV1
MSM State 1
MSM State 2
MSM State 3
MSM State 4
Finger subdomain 1 (FD1)
Template entry site
Thumb subdomain (TD)

Figure 11

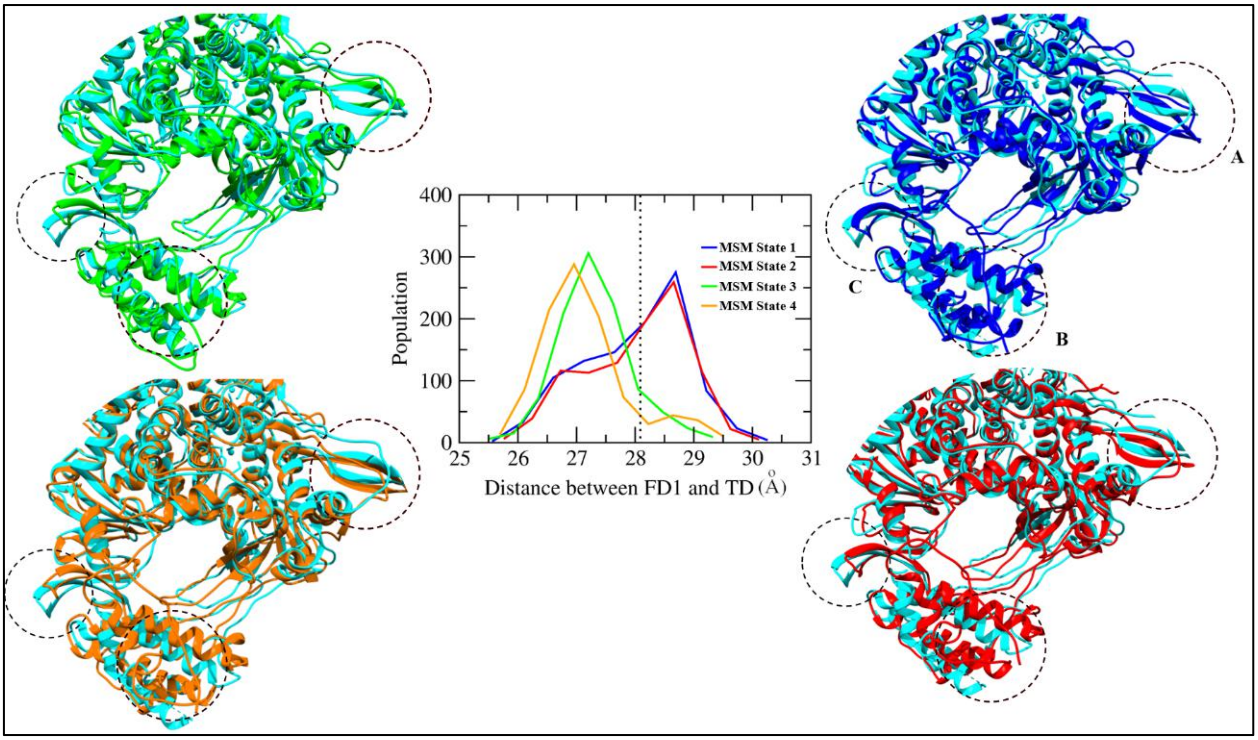
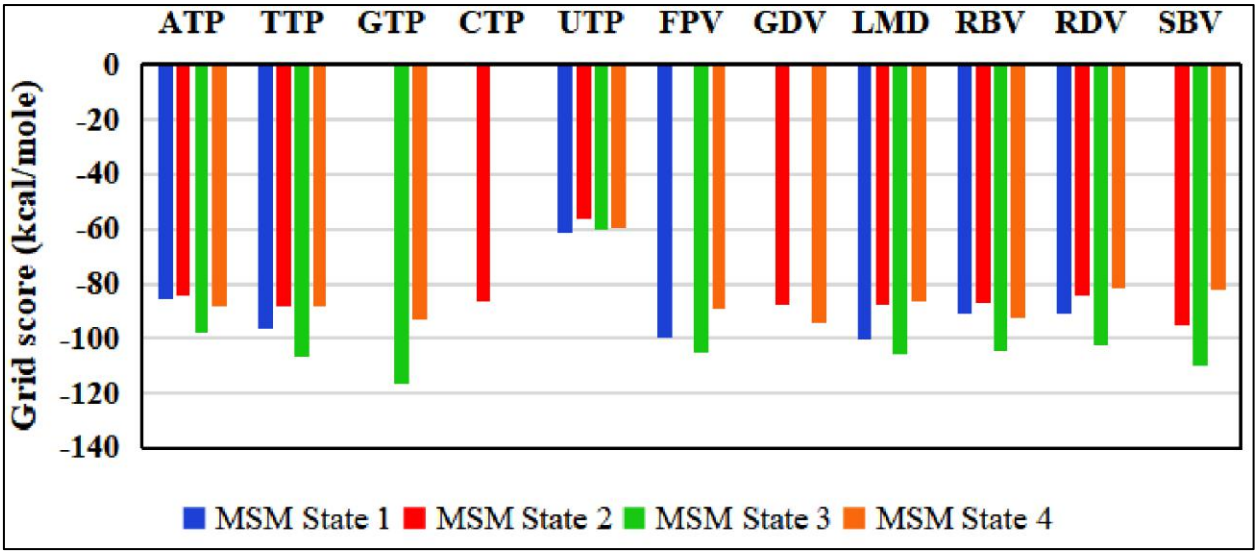

A
B
C
400
300
200
100
0
Population
25
26
27
28
29
30
31
Distance between FD1 and TD (Å)
MSM State 1
MSM State 2
MSM State 3
MSM State 4

Figure 12


ATP
TTP
GTP
CTP
UTP
FPV
GDV
LMD
RBV
RDV
SBV
0
-20
-40
-60
-80
-100
-120
-140
Grid score (kcal/mole)
MSM State 1
MSM State 2
MSM State 3
MSM State 4