Impact of Remdesivir Incorporation Along the Primer Strand on SARS-CoV-2 RNA-dependent RNA polymerase

Sehr Naseem-Khan,† Madison B. Berger,†‡ Emmett M. Leddin,† Yazdan Maghsoud,†‡ and G. Andrés Cisneros*,†¶,‡†

†Department of Chemistry, University of North Texas, Denton, Texas, 76201, USA
‡Present address: Department of Chemistry and Biochemistry, University of Texas at Dallas, TX 75080, USA
¶Present address: Department of Physics, University of Texas at Dallas, TX 75080, USA

E-mail: andres@utdallas.edu

Abstract

Remdesivir was the first antiviral drug that received emergency use authorization from the United States Food and Drug Administration to treat COVID-19. Remdesivir is a nucleotide analogue that targets the RNA-dependent RNA polymerase (RdRp) of coronaviruses, including SARS-CoV-2. The solution of multiple RdRp structures has been one of the main axes of research in the race against the SARS-CoV-2 virus. Several hypotheses of the mechanism of inhibition of RdRp by remdesivir have been proposed, although open questions remain. This work explores the impact of up to four incorporations of remdesivir along the primer strand on RdRp by means of molecular dynamics (MD) simulations. The simulation results suggest that the overall structure and dynamical behavior of RdRp is destabilized by remdesivir. The strongest impact
on the structure and dynamics are observed after three incorporations, when remdesivir is located at position -A3. Our results are consistent with previously reported experimental and computational results, and help provide further atomic-level detail on the role played by remdesivir on the disruption of RNA synthesis by RdRp.

1 Introduction

The COVID-19 pandemic has garnered significant attention from the global scientific community. This has led to a world-wide effort to understand and develop effective therapeutics to treat severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus infection. Multiple approaches are being explored and have led to the development of various vaccines and small-molecule drugs. At the time of writing this paper, several vaccines have been developed and are currently being administrated to the global population. This is a major milestone and the first line of defense against SARS-CoV-2. The development of antivirals is also important due to multiple factors including (but not limited to) lack of vaccines for segments of the population, allergic reactions to vaccines, etc.

The SARS-CoV-2 virus is part of the betacoronavirus family. Its genome encodes 29 proteins, three of which have received the most attention as therapeutic targets: the spike glycoprotein, main protease (M\textsuperscript{pro}), and RNA-dependent RNA polymerase (RdRp). The glycoprotein, which drives the entry into human cells, has been a vaccine target, whereas M\textsuperscript{pro} and RdRp have been prominent drug targets due to their roles in replication and transcription of the virus (although they are not the only protein targets).

RdRp is composed of three main domains (nsp12, nsp7, nsp8) and the RNA template-product duplex. The polymerase region, which can be subdivided into the finger, palm and thumb subdomains, is localized in the nsp12 domain, as well as the NiRAN (nidovirus RdRp-associated nucleotidyltransferase) and Interface regions. Several cryo-electron microscopy (cryo-EM) structures of RdRp have been resolved including PDB 6M71 and PDB
7BTF from Gao et al.\textsuperscript{19} and PDB 6YYT from Hillen et al.\textsuperscript{20} The latter is represented in Figure 1A showing the replicating form of the RdRp structure with elongated nsp8 and RNA template-product duplex. Yin et al. have also determined cryo-EM structures with remdesivir incorporated on the primer RNA strand (PDB 7BV2).\textsuperscript{21} However, none of these structures explicitly include a complex with a nucleotide triphosphate in the active site.

![Replicating SARS-CoV-2 polymerase structure](image)

Figure 1: (A) Representation of the replicating SARS-CoV-2 polymerase structure (PDB: 6YYT). (B) Model of the RdRp active site with ATP as the incoming nucleotide triphosphate.

Remdesivir was the first drug that received emergency use authorization by the United States FDA (Food and Drug Administration)\textsuperscript{22} to treat COVID-19 patients in the USA.\textsuperscript{23-26} Remdesivir was originally developed against the Ebola virus, and it started the clinical trial process before the 2019 COVID-19 outbreak.\textsuperscript{27-28} Because of this, the drug remdesivir has been the focus of many experimental and computational research studies in order to establish its possible mechanism of action and efficacy against the SARS-CoV-2 virus.\textsuperscript{21,29-34}
When entering human cells, the remdesivir prodrug is reduced to an active triphosphate form through metabolic reactions. This active form of remdesivir (REM) is very similar to its natural analogue, ATP. One of its main differences from ATP is the presence of a cyano group at the C1′ position. REM’s overall similarity to ATP allows remdesivir to be incorporated into the primer strand during RNA synthesis. Remdesivir has been proposed to act as a chain terminator by blocking synthesis as it binds in the incoming nucleotide position, or as an RNA translocation blocker at the -A4 position, after the incorporation of four nucleotides with the first being remdesivir. This stalling (delayed) mechanism is proposed to occur due to a potential steric clash between remdesivir at -A4 and SER861. On the other hand, QM/MM studies have suggested that remdesivir at -A4 and SER861 could form a covalent bond.

Here, we have used computational simulations based on all-atom classical molecular dynamics (MD) to investigate the impact of remdesivir incorporation at multiple positions along the primer strand. REM and two of its analogues (termed CID and CYD) were compared with ATP in the incoming nucleotide position to investigate the possibility of any of these drugs acting as chain terminators (Figure 2). Here, REM is in its trisphosphate form of the prodrug remdesivir, and CID is formed from the prodrug CID (PubChemID: 44468491). The remainder of this article is as follows: in the next section we describe the development of force field parameters for all non-standard nucleotides and simulation details. Next, we discuss the results for the investigation of effects of different inhibitors in the incoming nucleotide position compared with ATP, followed by the results of the impact of remdesivir on the primer strand and concluding remarks.
2 Computational Details

2.1 Structure and MD simulations set-up

We constructed our model using the replicating SARS-CoV-2 RdRp structure obtained from cryo-EM by Hillen et al. (PDB 6YYT). The missing residues A4–V30, T51–T76, and K98–Q117 were incorporated after overlaying the cryo-EM RdRp structure by Gao et al. (PDB 7BTF). We have modelled the active site using the positions of the incoming nucleotide and the two Mg$^{2+}$ cations from the Norovirus Polymerase structure by Zamyatkin et al. (PDB 3H5Y) (Figure 1B, and see also Figure S1). The RdRp structure was then solvated in TIP3P water using a 10 Å cutoff from the center mass of RdRp, and 61 K$^+$ counterions were added to neutralize the water box with the LEaP program. The force fields used to describe the protein and RNA are Amber-ff14SB and RNA-$\chi$OL3, re-
We have also used the ZAFF force field to describe the two Zn$^{2+}$ cations and first coordination sphere ligands.\textsuperscript{48} Finally, the parameters of the incoming nucleotide studied here (CID, CYD, and REM) were obtained through pyRED calculations on the R.E.D server.\textsuperscript{47-50} The ATP parameters were obtained from Meagher et al.\textsuperscript{51} AMBER20 was used for MD simulations of all systems.

Each system was run in triplicate. Initially, each system was heated from 0 K to 100 K for 100 ps, followed by a 500 ps simulation to reach the production temperature of 300 K with a 2 fs time step, and a collision frequency of 5 ps$^{-1}$. During these stages the protein/substrate/cofactor atoms were restrained with a force constant of 10 kcal mol$^{-1}$ Å$^{-2}$. After equilibration, restraints were removed except for the zinc cations and their coordinating ligands, which were restrained at 10 kcal mol$^{-1}$ Å$^{-2}$ during the whole MD simulation. Production simulations for all systems were performed using the NPT ensemble with a 2 fs time step at 300 K for 350 ns, using the Langevin thermostat\textsuperscript{52} with a collision frequency of 1 ps$^{-1}$, and the Berendsen barostat\textsuperscript{53} using the default weak-coupling scheme. A 9 Å cutoff was employed for non-bonded interactions, and the smooth particle mesh Ewald (PME) approach was used to describe long-range electrostatic effects.\textsuperscript{54}

### 2.2 Trajectory analysis

Cptraj\textsuperscript{55} was used to compute root mean squared deviation and fluctuation (RMSD/RMSF), matrix correlations, and distance analyses. Normal mode analysis was performed using the ProDy code.\textsuperscript{56} NCI (Non-Covalent Interactions) analysis was computed with the NCIPLoT program\textsuperscript{57,58} using its default color code (RGB) for plotting isosurfaces, where red, green and blue represent repulsive, attractive and weak attractive forces respectively. Network analysis was carried out using the dynamic network analysis (DyNetAn) python package via a Jupyter Notebook from Melo et al.\textsuperscript{59} Gephi\textsuperscript{60} was used to compose the node plots. The UCSF Chimera\textsuperscript{61} and VMD\textsuperscript{62} software programs were used to render the images.
2.3 Energy Decomposition Analysis

Energy Decomposition Analysis (EDA) was performed to understand the nature of the intermolecular interactions between two fragments using an open-source Fortran90 program. This approach is based on the calculation of the average non-bonded inter-molecular interaction energy, \( \langle E_{ij} \rangle \), between every pair of fragments \( i \) and \( j \). Subsequently, the alterations (if any) in inter-molecular interactions as a function of changes between different systems can be obtained:

\[
\Delta E_{ij}^{AB} = \langle E_{ij}^B \rangle - \langle E_{ij}^A \rangle,
\]  

where \( A \) denotes the reference system, \( B \) denotes the target system, and \( \langle E_{ij}^A \rangle \) denotes the average non-bonded inter-molecular interaction energy of fragment \( i \) in system \( A \) with all other fragments \( j \) in \( A \) (and likewise for \( B \)). Based on this data, it is possible to investigate the non-bonded inter-molecular interactions as a function of a particular reference residue \( i \). Here, all EDA results correspond to the difference of non-bonded intermolecular interactions, \( \Delta E_{ij}^{AB} \) (henceforth \( \Delta E_{nb} \)), with the ATP system obtained after averaging over the three replicates of each system; that is, the ATP system has been used as reference for all \( \Delta E_{nb} \) calculations.

2.4 Dynamic Network Analysis

Dynamic network analysis was carried out exclusively on the protein and RNA atoms. Due to the large number of atoms in all of these systems, cpptraj was used to strip the water and counterions. In addition, the combined trajectory for each system was trimmed by saving every 25th frame. Amino acids were represented by one node placed on the alpha carbon. Ribonucleotides were represented with two nodes; the first node was placed on the phosphorus and the second on the nitrogen of the nitrogenous base. 4.5 Å was used for the cutoff distance and 0.75 for contact persistence. Once the network was calculated,
the Floyd–Warshall algorithm, as implemented in DyNetAn\textsuperscript{59} was used to calculate the optimal path between two selected nodes: SER861 and the incoming nucleotide in the active site (either ATP, CID, CYD or REM). The optimal path can provide information as to how the allosteric communication between nodes throughout the protein may change. The node plots that will be further discussed in the Results and Discussion section were generated from the network data. The size of the circles reflect the number of node–node interactions within the designated subunit. The size of the edge that connects the circles refers to the number of interactions between nodes in those respective subunits.

3 Results and Discussion

3.1 Effects of the incoming nucleotide

As described above, remdesivir has shown a strong potential to stop the RNA synthesis at the incoming nucleotide position (chain terminator). Byléhn et al. have recently computed the binding energies of remdesivir, favilavir, and ribavirin in their monophosphate form.\textsuperscript{33} They have found that the binding energy of remdesivir is more favorable, thus showing its ability to be incorporated into the RNA primer strand. Here, we compare remdesivir and its two analogues (CID & CYD), with respect to ATP (i.e the natural nucleotid).

All systems were observed to be stable throughout the simulation time, without major deviations in individual subunits or significant active site changes (Figure S2). Figures 3A,C and E show the difference residue-wise correlation with respect to the ATP system computed as a function of the incoming nucleotide, namely ATP, CID, CYD, and REM, as a heatmap projected on the protein structure (see also Figure S4). We observed that the finger and palm domains are anti-correlated with respect to ATP for the three systems. EDA results also show that two residues close to the incoming nucleotide, K551 and R555, have positive $\Delta E_{nb}$ values across the three systems. This suggests that all three inhibitors show a destabilizing interaction within the active site compared with ATP.
estingly, NCIPLLOT analysis indicates multiple van der Waals and attractive forces around the active site that aid in the stability of the incoming nucleotide in the active site.

Overall, the CYD system differs the most compared to CID and REM. For CYD, we observe mostly non-correlated movements, with respect to ATP, for the whole structure. REM and CID systems behave similarly having correlated movements with respect to ATP such as the nsp7, nsp8, and thumb regions. Additionally, the directionality of the first normal mode of CYD is in the opposite direction compared to both REM and CID. Thus, the CYD system seems to have a stronger destabilizing effect on the RdRp structure.
Figure 3: Comparison of the heatmap correlation matrix difference with the ATP system computed as a function of the incoming nucleotide for REM, CID and CYD systems. The black arrow indicates the motion of the first normal mode. The isosurfaces represent NCI-PLOT results near the active site (isovalue=0.4, cutoff=5 Å). The values of the difference of EDA results with respect to ATP, computed as a function of the incoming nucleotide, are indicated in purple.
3.2 Role of remdesivir within the RNA primer strand

We next assessed the effect of REM incorporated at different positions from -A1 to -A4 in the RNA primer strand. We denote by REM-A[n] all the RdRp systems studied in this work: REM-A[4], REM-A[1,4], REM-A[1,2,4], REM-A[1,3,4] and REM-A[1,2,3,4] (Figure 2). As previously stated, the steric effect between SER861 and REM at position -A4, due to the cyano group of REM on the sugar ring, has been proposed to be the main cause for the disruption of RNA translocation from -A3 to -A4. For that reason, we aim here to shed a light on the delayed mechanism hypothesis.

To do so, we first measured the distance ($d_{c_{\beta}-N}$) between the beta carbon ($c_{\beta}$) of SER861 and the nitrogen atom of the cyano-group of REM located at the -A4 position (Figure 4F). For all of the REM-A[n] systems, $d_{c_{\beta}-N}$ stays near 3.5 Å during the entire MD simulation. Conversely, for the ATP, CID, CYD and REM systems, a guanine base is originally located at the -A4 position, and the distance between $c_{\beta}$ of SER861 and the oxygen atom (O2’) of the guanine fluctuates between 3 and 5 Å during the MD simulation (Figure S3). Additionally, most of the REM-A[n] systems have smaller RMSF values for the REM located at the -A4 position and SER861 residues.

Mainly vdw interactions, as displayed by the green isosurfaces from the NCIPLOT analysis, are observed for all systems. They are more spread out when the number of REM positioned in the RNA primer strand increases (Figures 4A-E, and Figure S5). This suggests an increase in the number of interactions between REM and SER861 at the -A4 position; however, it is not possible to determine whether these are attractive or repulsive from NCIPLOT. EDA analysis computed with respect to SER861 reveals that this intermolecular interaction is mainly repulsive with respect to ATP. $\Delta E_{nb}$ varies between 1.4 kcal mol$^{-1}$ to 3.3 kcal mol$^{-1}$ for all the REM-A[n] systems except REM-A[1,2,4]. In this case, $\Delta E_{nb}$ is $-1.1$ kcal mol$^{-1}$, which slightly favors an attractive interaction between REM and SER861. Overall, these results are consistent with the steric repulsion hypothesis caused by the cyano group on REM when it is located at the -A4 position, albeit only
Figure 4: Intermolecular interaction and structural analysis at the -A3 and -A4 positions of the RNA strand. (A-E): NCIPLOT analysis indicates mainly van der Waals intermolecular interaction between SER861 and its nearby residues (green isosurfaces, isovalue=0.4, cutoff=5 Å). EDA results are displayed for relevant residues using SER861 (values in purple) and REM at the -A3 position (values in red) respectively as the reference residue. EDA shows that the nature of the intermolecular interaction between SER861 and REM at -A4 position is mainly repulsive for most of the REM-A[n] systems, and that the salt bridge formed between K593 and D865 is stronger in the case of REM-A[1,3,4] and REM-A[1,2,3,4]. RMSF values are displayed in blue for SER861 and the residue at the -A4 position. The heatmap correlation matrix difference with the ATP system as a function of SER861 is also plotted. (F): Analysis of the distance between the beta carbon of SER861 and the nitrogen atom of the cyano-group of REM at the -A4 position.

On the other hand, it has been shown that the formation of a salt bridge between K593 and D865 could stabilize REM positioned at the -A3 position. Here, we also observe a strong attractive interaction between K593 and D865 in the case of the REM-A[1,3,4] and REM-A[1,2,3,4] systems (Figures D-E), as opposed to REM-A[n] systems having a uracil nucleobase at the -A3 position (Figures A-C). Indeed, EDA results computed as a function
of the nucleotide at the -A3 position show a significant increase of $\Delta E_{nb}, \sim 20 \text{ kcal mol}^{-1}$ for K593, and $\sim 10 \text{ kcal mol}^{-1}$ for D865 (in magnitude). The favorable intermolecular force induced by the salt bridge helps stabilize REM at the -A3 position. Finally, it is interesting to note that the movements of residues from the REM-A[n] system are anti-correlated with respect to ATP, except for the REM-A[1,2,3,4] system. In the latter, REM residues at -A3 and -A4 positions are correlated with ATP.

### 3.3 Dynamic effects of remdesivir throughout the RdRp structure

Lastly, dynamic network analysis was carried out to gain insights into how incorporation of REM at different positions along the RNA primer strand may have effects throughout the entire structure. Specifically, we investigated how the number of node-to-node interactions within subunits and between subunits changed for the different systems. While much of this data is qualitative due to the need to sample small windows throughout the trajectory, some key trends were seen. A greater number of interactions within subunits was found for most of the REM-A[n] and CID/CYD systems. More specifically, we see an increase in intra-subunit interactions within the polymerase, interface, nsp8a subunits, as well as the RNA, with a few exceptions. This can be seen in Figure 5 by the purple circles representing subunits. The opposite is true for inter-subunit interactions. We see a large decrease in inter-subunit interactions, specifically interface-nsp8a, polymerase-nsp8a, polymerase-nsp8b, nsp8a-RNA and nsp8a-nsp7 for most of the REM-A[n] systems. This loss of inter-subunit interaction is represented by the green lines connecting subunits in Figure 5 (see also Figure S6 and Tables S1 and S2). From this, it can be inferred that there is an overall loss of allosteric communication between the different subunits and RNA of RdRp but a gain of communication within subunits when compared with the ATP system. The only outlier system found from these trends was the REM-A[4] which exhibited similar interactions to the ATP system.
Figure 5: Representation of the difference edge–node plots, with respect to the ATP system from the dynamic network analysis. The size of the circles is proportional to the number of communications between the residues within the considered subunit. The thickness of the edges is proportional to the number of communication between the subunits. Communications favored in the ATP system are colored green, and those favored in the studied system are colored purple.

To gather further insight into how the allosteric communication changed between the different REM-A[n] and CID/CYD systems, we looked at the optimal pathways between SER861 and the incoming nucleotide. Figure 6 represents the two main optimal paths found across all the studied systems. We observed that one optimal path follows the RNA primer strand (A), while the other path (B) goes through the protein. The optimal path was calculated for each window of the trajectory for all replicates. The ATP, REM-A[4] and REM-A[1,2,4] almost equally favored both paths. All other systems exclusively favored a pathway through the RNA primer strand. This result corroborates the node–node interactions previously discussed in this section. Most of the REM-A[n] systems, except for
REM-A[4], have lost communication between subunits but have maintained or increased the number of interactions within each isolated subunit. That is, when remdesivir or an analogue of it is introduced into the RNA primer strand, the protein experiences a major loss in allosteric communication, which could explain the decrease in RNA synthesis.

Figure 6: The two dominating optimal pathways found between SER861 and the incoming nucleotide throughout all RdRp systems are shown. Path (A) follows the RNA, whereas path (B) favors a path through the protein. As an example, we show here (C) the optimal paths connecting SER861 and REM.

4 Conclusions

We have investigated the structural, intermolecular, and dynamical effects of several RdRp systems. In line with previous studies, we have also shown that remdesivir can be incorporated within the RNA primer strand, thanks to its similarities with ATP. We have also demonstrated that the steric effect at play between SER861 and REM at the -A4 position of the RNA primer strand is a weak repulsive intermolecular interaction. This could imply that the further translocation of RNA from -A3 to -A4 could be a possibility in the case of a very weak or non-existent steric effect induced by SER861. However, our dynamical network analysis suggests strong destabilization of the whole RdRp structure. Particularly,
we have found a loss of communication between the nsp domains with both the RNA and polymerase for most systems, with respect to the ATP system. As previously mentioned, Aranda and Orozco have discussed the probability of the formation of a covalent bond between the SER861 and remdesivir at the -A4 position. In this way, the full stop of the translocation of RNA would be ensured. Certainly, the cyano-group of remdesivir could be revisited or modified to increase its antiviral reactivity.

Acknowledgement

This work was funded by NIH Grant No. R01GM108583. Computational time was provided by the University of North Texas CASCaM CRUNTCh3 high-performance cluster partially supported by NSF Grant No. CHE-1531468 and by XSEDE Project No. TG-CHE160044. Additional computational time from the COVID-19 HPC Consortium is gratefully acknowledged. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC05-00OR22725. MBB thanks MolSSI for support provided via a software development fellowship.

Supporting Information Available

Additional RMSD, distance analysis, heatmap correlation matrix plots, active site, and edge–node data are provided in the SI.

References


(39) Zamyatkin, D. F.; Parra, F.; Machín, Á.; Grochulski, P.; Ng, K. K.-S. Binding of 2′-amino-2′-deoxycytidine-5′-triphosphate to norovirus polymerase induces rearrangement of the active site. Journal of Molecular Biology 2009, 390, 10–16.


Graphical TOC Entry

Communications in the REM system:
- More in REM
- More in ATP