SARS-CoV-2 RdRp Follows Asynchronous Translocation Pathway for Viral Transcription and Replication

Xiaowei Wang[1], Yuan Yao[1], Xin Gao[2,3], Lu Zhang[4,5,6]*

[1] Department of Chemical and Biological Engineering and Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong
[2] Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[4] State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian, China
[5] Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Fujian, China
[6] University of Chinese Academy of Sciences, Beijing, China

* Corresponding author Email: luzhang@fjirsm.ac.cn

ABSTRACT

RNA-dependent RNA polymerase (RdRp) is the replicase machinery for SARS-CoV-2 and thereby it has become one of the most promising drug targets to combat the pandemic as well as the healthy threat posed by the novel coronavirus. Translocation is one essential step for RdRp to exert the viral replication and transcription, and it describes the dynamic process in which the double-stranded RNA moves upstream by one base pair position to empty the active site for the continuous substrate incorporation. However, the molecular mechanisms underlying the dynamic translocation of SARS-CoV-2 RdRp remain elusive. In the current study, we have elucidated the molecular insights into the translocation dynamics of SARS-CoV-2 RdRp by constructing a Markov State Model based on extensive molecular dynamics simulations. We have identified two previously uncharacterized intermediates which pinpoint an asynchronous and rate-limiting translocation of the nascent-template duplex. The movement of the 3'-terminal nucleotide in the nascent strand lags behind its upstream nucleotides due to the uneven protein environment while the translocation of template strand is delayed by the hurdle residue K500. Although the motions of the two strands are not synchronous, they share the same "ratchet" to stabilize the system in the post-translocation state, suggesting a coupled Brownian-ratchet model. Overall, our study has provided the intriguing insights into the translocation dynamics with unprecedented molecular details, which would significantly deepen our understanding about the transcriptional mechanisms of SARS-CoV-2.

KEYWORDS: *Markov State Model, Molecular dynamics simulation, SARS-CoV-2, RNA polymerase, RNA translocation*

INTRODUCTION

The spread of 2019 novel coronavirus (COVID-19) has caused a global pandemic and more than 6.2 million deaths by the end of May, 2022[1]. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the positive-sense single-stranded RNA virus that has caused the pandemic and its genome contains around 20 kilobases[2]. Transcription and replication of such a large RNA genome are the fundamental steps of the viral life cycle. RNA-dependent RNA polymerase (RdRp) containing non-structural protein 12 (nsp12) together with nsp7 and nsp8 is the minimal replicase machinery for SARS-CoV-2 and thereby it has become one of the most promising targets for drug exploration and design[3]. Elucidation of the transcriptional mechanisms of the core molecular machine RdRp would not only facilitate our understanding about viral replication and transcription, but also guide the rational design of potential antiviral drugs or inhibitors[4–9].

The structures of SARS-CoV-2 RdRp at different transcriptional stages have been captured by cryo-electron microscopy (cryo-EM), and they have provided valuable structural basis for understanding how this machinery performs its functions[6,10–15] (Figure 1a). The core component of SARS-CoV-2 RdRp is the catalytic subunit nsp12, which is responsible for the extension of RNA strand. Nsp7 and nsp8 are two accessory cofactors that can facilitate the template binding and the processivity of RdRp[11]. During the RNA elongation, RdRp is processing consecutive nucleotide addition cycles (NACs)[16], where translocation is one essential step that the template-nascent duplex moves to upstream by one base pair position and the active site is vacant for the arrival of the next nucleotide triphosphate (NTP) substrate and initiating the next round of NAC. Although the cryo-EM structures of both the pre- and post-translocation states are available[6,10,11], the static structures alone are insufficient to provide the information of the dynamic translocation.

Two models have been proposed for depicting the translocation process. The “Brownian-ratchet” model suggests that the complex can spontaneously oscillate between the pre- and post-translocation states under thermal fluctuations without the input of any additional chemical energy[17–19]. Under such circumstance, the NTP loaded into the active site would act as the “ratchet” to stabilize the complex in the post-translocation state. In contrast, “power-stroke” is an alternative model suggesting that the thermal dynamics of the complex alone is insufficient to accomplish the translocation while the pyrophosphate ion (PPi) release would provide the chemical energy to power up this process. So far, the translocation mechanisms have been investigated for several RNA polymerases. For example, a previous computational

study has proposed that the translocation of yeast RNA polymerase II (Pol II) follows the "Brownian-ratchet" model, as the thermal oscillation of the bridge helix between straight and bent conformations would guide the translocation of DNA:RNA hybrid[17]. In a distinct scenario, the X-ray crystallographic structure has captured a conformational change of O-helix upon PPi release and thus suggested that the translocation of T7 RNA polymerase (RNAP) abides by the "power stroke" mechanism[20]. However, such a coupling between the PPi release and O-helix has not been observed in other studies and "Brownian-ratchet" model has thus been suggested for describing the translocation in T7 RNAP[21,22]. In a more intriguing scenario, a combination of "Brownian-ratchet" and "power-stroke" models has been proposed for enterovirus 71 (EV71) RdRp[18,19], of which the translocation intermediates have been resolved by X-ray crystallography. Although these studies have provided valuable insight into the translocation dynamics in diverse RNA polymerases, it remains obscure which model SARS-CoV-2 RdRp adopts during its translocation. Since the architecture of the active site of SARS-CoV-2 RdRp is different from that in either Pol II or T7 RNAP, it becomes challenging to directly predict the translocation dynamics in SARS-CoV-2 RdRp based on merely the previous models. Furthermore, the protein environment encircling the double-stranded RNA (dsRNA) in SARS-CoV-2 RdRp is different from that in EV71 viral RdRp, and thus distinct translocation dynamics could be implicated. Therefore, although the dynamic translocation is one essential step for viral replication and transcription in SARS-CoV-2, the underlying molecular mechanism remains elusive.

Molecular dynamics (MD) simulation has become a useful tool to resolve the dynamics of biological macromolecules at the atomic resolution. Recently, it has been widely utilized for drug screening and examining the mechanisms of nucleotide analogs on the nucleotide addition in SARS-CoV-2 RdRp[4,5,8,23–28]. However, the timescale for one turnover of nucleotide addition cycle has been estimated at several milliseconds for viral RdRps[29,30], which is difficult to be directly accessed by conventional MD simulations usually covering the timescales up to hundreds of nanoseconds or a few microseconds for the biological macromolecules. To bridge the timescale gap, the Markov state model (MSM) has emerged as a tailor-made kinetic network model and demonstrated its capability to predict the long-timescale dynamics based on numerous while relatively short MD simulations[31-48]. Specifically, the conformations sampled from the extensive simulations would first be split into hundreds or a few thousands of microstates according to a pre-defined distance metric. The transition probability matrix $T(\tau)$ would then be constructed with each individual entry T_{ij} representing the transition

probability from state i to state j after a certain lag time τ , which should be sufficiently long to ensure that the next state the system would arrive after lag time τ only depends on the current state while is irrelevant with any history. In this scenario, the system becomes memoryless and the long-timescale kinetics can be obtained by propagating the transition probability matrix. So far, the MSM has been widely applied to study the transcriptional dynamics in RNA polymerases, and exhibited its power in predicting the conformational changes and elucidating the underlying molecular mechanisms at the timescales of tens to hundreds of microseconds based on MD simulations in the scale of tens to hundreds of nanoseconds per trajectory[17,22,46,49–51].

In this study, we have constructed a MSM based on extensive all-atomic MD simulations of SARS-CoV-2 RdRp in explicit solvent (~212,000 atoms) with an accumulated simulation time of ~24 μs to elucidate the translocation dynamics in SARS-CoV-2 RdRp. Two previously uncharacterized intermediates have been identified as the key joint states along the translocation pathways, and they have demonstrated an asynchronous translocation. Moreover, the rate-limiting step has been characterized by our MSM and corresponds to the translocation of the template strand as well as the 3'-terminal of nascent strand. Furthermore, K500 has been pinpointed as the hurdle residue responsible for the asynchronous movement between the template strand and the nascent strand, while the intricate interaction network around the 3'-terminal of the nascent strand also contributes to the asynchronous movement within the nascent strand. Overall, our MSM has demonstrated that although each strand of the template-nascent duplex follows its own Brownian motion separately, both strands share the same "ratchet" to stabilize the system in the post-translocation state. These observations have altogether suggested that the motions of the template and nascent strands are still coupled and the translocation dynamics of SARS-CoV-2 RdRp follows a coupled Brownian-ratchet model.

RESULTS

The Translocation of the Nascent and Template Strand Is Asynchronous

We have constructed a robust MSM based on extensive all-atom MD simulations to elucidate the underlying molecular mechanisms of SARS-CoV-2 translocation dynamics. Specifically, the cryo-EM structures[10] (PDB ID: 7c2k and 7bzf) lay the structural basis for the pre- and post-translocation states, respectively. The preliminary translocation pathways have been generated using the modified Climber algorithm[52], which gradually drives the system for the forward translocation from the pre- to post-translocation state as well as the backward translocation.

Subsequently, representative conformations have been selected from the preliminary pathways to initiate the MD simulations and the total ~1M MD snapshots (with an integrated simulation time of ~24 μs) have been obtained for MSM construction (see Methods and SI Section 1 and 2 for details of structural model construction, MD simulations and MSM construction).

Our MSM has revealed four metastable states along the translocation pathway of SARS-CoV-2 RdRp (Figure 1). Besides the pre- (S1) and post-translocation (S4) states that have been captured by cryo-EM structures[10], we have identified two intermediate states which have not been characterized before (S2 and S3 states in Figure 1b). Intriguingly, these states have exhibited an asynchronous translocation pattern, where the nascent strand translocates prior to the template strand (Figure 2). In particular, from S1 to the first intermediate state S2, template strand does not translocate and adopts a similar configuration as in the S1 state (Figure 2b). On a sharp contrary, the nascent strand has demonstrated an obvious movement to the upstream direction (Figure 2d) although its 3'-terminal nucleotide has lagged behind and still occupied the active site as in the pre-translocation (S1) state (Figure 2e). From S2 to the second intermediate state S3, the template strand and the 3'-terminal nucleotide of the nascent strand both start to translocate over a significant distance of ~3 Å (Figure 2b and 2e), while the remaining upstream nucleotides in the nascent strand almost reach their configuration as in the post-translocation state (Figure 2d). Finally, the template strand accomplishes its one-base-pair translocation with a further movement from S3 to S4 state (Figure 2b).

Interestingly, the motion of the transition nucleotide (shown in orange in Figure 2a) is distinct from that of the template-nascent duplex. Specifically, both the backbone and base of transition nucleotide start to move from S1 to S2 state, while only its base shows a significant conformational change from S2 to S3 state (Figure 2c). To examine the movement of the base of transition nucleotide, we have computed the dihedral angle formed by the base of transition nucleotide, the backbone of template strand and the base of upstream nucleotide (Figure S1). If the base of transition nucleotide and the base of its upstream nucleotide are on the same side of their backbones, the dihedral angle would be in the range between -90° and +90°. We have found that the dihedral angles have sampled a broad distribution between -180° and 180° in both S1 and S2 states (Figure S1b), suggesting that the base of transition nucleotide fluctuates significantly and can point to both sides relative to its backbone. From S2 to S3 state, the base of transition nucleotide obviously reduces its flexibility and maintains at the same side as its upstream nucleotide of the template strand, with the dihedral angle centered at 0° in S3 state (Figure S1b). Afterwards, both the base and the backbone of the transition nucleotide move in

sync with the whole template strand until it reaches the canonical *+1* position ready for base pairing with the next NTP substrate (Figure 2a, 2c and S1).

Overall, the four metastable states elucidated by our MSM have demonstrated an asynchronous translocation of the template-nascent duplex in SARS-CoV-2 RdRp. In particular, the nascent strand moves in advance than the template strand, while the upstream nucleotides in the nascent strand translocate prior to the 3'-terminal nucleotide. In contrast to the movement of the duplex, the transition nucleotide shows its distinct translocating motion. Further investigations are required to elucidate how these conformational changes couple with the translocation dynamics.

The Translocation from S2 to S3 State Is the Rate-limiting Step

To further understand the translocation dynamics in SARS-CoV-2 RdRp, we have synthesized five 250-ms trajectories based on the transition probability matrix of our microstate-MSM and observed that RdRp could oscillate between the pre- and post-translocation states multiple times (see SI Section 2.4 for details and Figure S2). Accordingly, the free energy landscape and the populations of the metastable states have been obtained (Figure 1c and Table S1). We have found that the pre-translocation state (S1) can form a relatively fast equilibrium with the first intermediate state S2, with a mean first passage time (MFPT) of ~10 μs (Figure 1c and Table S2). Intriguingly, the transition from S2 to the second intermediate state S3 has defined the rate-limiting step and would happen with a MFPT of ~50 μs (Figure 1c). These two intermediate states have been characterized as essential joint states for the translocation of SARS-CoV-2 RdRp, as the transition flux from the pre- to post-translocation state is entirely through S2 and S3 states according to the transition pathway analysis (see SI Section 2.5 for details). Afterwards, the RdRp would further translocate to S4 state with a MFPT of ~10 μs. These observations have suggested that the transition from S2 to S3 state needs to overcome an energy barrier, and thus establishes the rate-limiting step for the translocation in SARS-CoV-2 RdRp. As shown in Figure 2b and 2e, in this rate-limiting transition from S2 to S3, the 3'-terminal nucleotide of the nascent strand starts to catch up with its upstream nucleotides (Figure 2e), and the template strand initiates its translocation (Figure 2b). However, it remains obscure why the movement of 3'-terminal nucleotide of the nascent strand as well as the template strand is hindered in S1 and S2 states.

Our further analysis has suggested that the protein environment alongside the nascent strand could mediate the asynchronous movement of the 3'-terminal nucleotide relative to its upstream nucleotides. Specifically, in the pre-translocation state, the 3'-terminal nucleotide of

the nascent strand has formed an intricate interaction network with four residues in motif A, B and motif F, where all the averaged interacting distances lie around 3.5 Å (Figure 3a and 3b). In particular, the ribose O2' and O3' atoms of the nucleotide forms hydrogen bonding interactions with D623, while its base is hydrogen bonded with S682, K545, and its backbone forms hydrogen bonding and electrostatic interactions with R555 (Table S3). On a contrary, each nucleotide in the upstream region can only form relatively loose interactions with one to two residues (Figure 3a and 3b). Such an uneven distribution of interactions has thus stalled the 3'-terminal nucleotide to adopt the conformation as in the pre-translocation state and rationalized the asynchronous translocation within the nascent strand from S1 to S2 state (Figure 2a, 2d and 2e, S3a and S3b). Moreover, the translocation of the 3'-terminal nucleotide requires disrupting its complex interactions with the protein environment, which thereby contributes to the energy barrier in the rate-limiting transition from S2 to S3 state.

Besides the movement of 3'-nascent strand, another critical conformational change during the rate-limiting step corresponds to the translocation of the template strand. Different from the scenario for the nascent strand, the upstream and downstream region of the template strand translocate synchronously (Figure S3a and S3c). Such an observation is further rationalized by an even distribution of protein environment encircling the template strand (Figure 3a, 3c and Table S4). Specifically, each nucleotide interacts with one to two protein residues, the interaction strength of which is similar with that in the upstream nascent strand. These observations have suggested that the protein interactions alongside the template strand seldom contribute to the asynchronous movement of template strand relative to that of the nascent strand. Therefore, there must exist other components to determine the translocation dynamics of template strand in the rate-limiting step.

The Hurdle Residue K500 Hinders the Translocation of Template Strand in the Rate-limiting Step from S2 to S3

Our further investigations have revealed that K500 could play a critical role in determining the translocation dynamics of template strand. In the pre-translocation (S1) state, the residue K500 locates between *+1* and *-1* site of the template strand and in a close contact (~3.0Å) with the monophosphate backbone (Figure 4c), which could thus create a steric obstacle to inhibit the forward translocation. To examine this conjecture, we have computed the dihedral angle formed by K500 and the backbone of template strand (Figure 4a and S4). If the sidechain of K500 lays perpendicular to the backbone of the template strand, the dihedral would be around

90°. On the contrary, if the side chain of K500 is parallel to the template strand backbones, the dihedral would be around 0°. Our calculations have demonstrated that the dihedral angles center around 60° and 54° in S1 and S2 states, respectively (Figure 4a), suggesting that the sidechain of K500 lies across the transition pathway of the template and thus hinders the translocation of the template strand in S1 and S2 states (Figure 4c). From S2 to S3 state, the side chain of K500 rotates away and points to the direction parallel to the backbone of the template strand to relieve the obstacle, with the dihedral angle sharply decreasing to ~0° (Figure 4a and 4c). With a further translocation to the S4 state, the side chain of K500 rotates back again to align across the backbone of template strand, with the dihedral increased to ~40° (Figure 4a). These observations have suggested that the side chain of K500 creates a steric barrier to hinder the translocation of template strand in S1 and S2 state, and thus determines the rate-limiting translocation of template strand from S2 to S3 state.

To further validate the role of K500 as a hurdle residue in the rate-limiting step of translocation, we have performed mutant simulations by mutating K500 to Alanine. Such a mutation has significantly reduced the size of the chain side and thus the steric effect originally created by K500 would be accordingly mitigated. To evaluate the influence of K500A mutation on the translocation dynamics of SARS-CoV-2 RdRp, we have chosen a transition state between S2 and S3 (Figure S5b) and examined its translocation propensity by MD simulations upon K500A mutation (see SI Section 3 for details). Simulations of wildtype RdRp with K500 have also been performed as a control, and we have found out that most of the wildtype RdRp conformations stay around the initial configuration even though the distribution is closer to the S2 state (left panel of Figure 4b). On a sharp contrary, the Lys-to-Ala mutation has stimulated more conformations towards S3 state (right panel of Figure 4b), suggesting a reduced barrier between S2 and S3 states and the translocation is favored upon K500A mutation. These results have further consolidated our observations that the bulky side chain of the hurdle residue K500 hinders the translocation of template strand, as reducing the size of the side chain would relieve the hindrance and facilitate the translocation. Therefore, the mutant simulations have again underlined that K500 is responsible for the asynchronous movement between the nascent and template strand. Interestingly, such an effect of the hurdle residue K500 on the asynchronous translocation has also been observed for the counterpart residue T114 in the translocation dynamics of EV71 RdRp, where T114 has been proposed to hamper the translocation of template strand and contribute to the asynchronous movement between the template strand and nascent strand[18,19].

The Translocation of SARS-CoV-2 RdRp Follows the Coupled Brownian-ratchet Model

Although the translocation dynamics of the nascent strand is distinct from that of the template strand during the transition from S1 to S2 state, the motions of the two strands are still coupled after overcoming the rate-limiting step from S2 to S3 state. Specifically, the translocation within the nascent strand is not synchronous as the 3'-terminal nucleotide would be stalled by its intricate interactions with the protein environment when the upstream nucleotides have translocated from S1 to S2 states. The translocation of the template strand is not synchronous with that of nascent strand, as it would be anchored to the same position as in the pre-translocation state due to the steric obstacle created by the hurdle residue K500. Nonetheless, the thermal fluctuations of the duplex are still sufficient to drive both the template strand and the 3'-terminal nucleotide of the nascent strand to overcome their respective energy barrier and reach the intermediate state S3. Afterwards, they both adopt the Brownian motions to form an equilibrium between the intermediate state S3 and the post-translocation state S4. Therefore, although both the template strand and the nascent strand demonstrate Brownian motions, their translocation dynamics are distinct and asynchronous. Even so, they still share the same "ratchet" i.e. the loading of substrate NTP, to stabilize the configuration in the post-translocation state[53]. Therefore, the SARS-CoV-2 RdRp follows a coupled Brownian-ratchet model in its translocation (Figure 5).

Motif F in SARS-CoV-2 RdRp has been proposed to act as a "pawl" to facilitate the translocation due to its similar position as the bridge helix in the eukaryotic RNA polymerase[12,54] (Figure S6). However, such a role of motif F has not been observed herein. In SARS-CoV-2 RdRp, motif F is constituted by two hinges (residues 541-550 and 556-562) as well as a loop (residues 551-555) (Figure S6b). Our simulations have demonstrated that the loop region (K551-R555) of motif F is relatively flexible (for example, the RMSF of R553 is as high as 4.5Å) (Figure S6c). However, they are distant from the template strand (averaged distance > 7.0 Å) and thus difficult to be directly coupled with the translocation dynamics. Although Q541 and N543 in the hinge regions are in a proximity to the template strand (averaged distance ~ 3.9 Å), our calculations have shown that they are relatively rigid (RMSF<2Å in Figure S6c). Further calculations have indicated that their motions seldom couple with that of transition nucleotide, as the Pearson correlation coefficients are always lower than 0.2 (see SI Section 4 for details). We have also noticed that V557 on motif F corresponds to I176 in EV71 viral RdRp, which has been proposed to facilitate the translocation by moving simultaneously with the template strand[18] (Figure S7a). However, our

simulations have demonstrated that the flexibility of V557 in SARS-CoV-2 is relatively low with a RMSF of ~1.4 Å (Figure S6c). Moreover, the side chain of V557 stacks with the base of template nucleotide in S1 and S2 state while their intermolecular distance sharply increases from 4.2 Å to 5.5 Å when the template strand translocates from S2 to S3 state (Figure S7c). Such an observation has indicated that V557 does not move concurrently with the template strand during S2 to S3 transition (Figure S7b and S7c), suggesting the motion of V557 has no obvious interplay with the movement of template strand. Overall, although motif F in SARS-CoV-2 RdRp locates in a similar position as the bridge helix in Pol II, our analysis indicates it seldom directly mediates the translocation dynamics of template-nascent duplex.

We have also noticed that previous works have proposed that the highly conserved "SG" sequence on motif B (Figure S8a) could be involved in the viral RdRp translocation[16,55–57]. To examine the influence of "SG" residues on the translocation dynamics, we have assessed if the movement of the template strand could couple with that of S682 and G683 in SARS-CoV-2 RdRp. Specifically, we have first calculated the Pearson correlation coefficient between the motion of S682 and the nucleotide located at one site upstream than the transition nucleotide (Figure S8a). Our results have shown that the coefficient lies below 0.25 in all the four metastable states (Figure S8b), suggesting S682 seldom participates in the translocation of template strand. In a sharp contrast, an obvious correlation has been observed between G683 and template nucleotide in the intermediate S3 state, with the Pearson correlation coefficient of ~0.45 and obviously higher than that in the other metastable states (Figure S8c). Consistently, the hydrogen bonding probability between G683 and template nucleotide in S3 state almost doubles that in S1 and S2 states (Figure S8d), indicating G683 could facilitate the translocation of template strand by stabilizing the RdRp in the intermediate S3 state. With a further translocation to the post-translocation state, the template nucleotide moves away from G683 and thus the hydrogen bond vanishes. Together, these observations have suggested that G683 could provide stabilization energy to the template nucleotide in the intermediate S3 state, which would thus facilitate the Brownian motion of RdRp to overcome the barrier and translocate from S2 to S3 state.

DISCUSSION

In this study, we have constructed a MSM based on extensive MD simulations to elucidate the translocation dynamics of SARS-CoV-2 RdRp. Our MSM has indicated that the translocation of the template strand and the nascent strand is asynchronous in the SARS-CoV-2 RdRp.

Interestingly, such an asynchronous translocation has also been observed in the human EV71 viral RdRp[18,19], where the translocation intermediates have been captured by X-ray crystallography. Specifically, the crystal structure of EV71 RdRp is a mixture of two states "S6A" and "S6B" as named in the original publication and has featured an asynchronous movement of the template and nascent strands[18]. After comparing these two states of EV71 RdRp with the metastable states of SARS-CoV-2 RdRp identified by our MSM (Figure S9), we have found out that the EV71 "S6A" state corresponds to our pre-translocation (S1) state due to the similar backbone and base configurations in both the template and nascent strands. The EV71 "S6B" state matches with the intermediate state S2 elucidated by our model, demonstrating that the template strand stays immobile relative to the pre-translocation state while the nascent strand has translocated. Overall, our model is consistent with the crystal structures of EV71 RdRp, both demonstrating that the nascent strand translocates in advance of the template strand. Moreover, we have also noticed that in both RdRps, the stalled movement of template strand is caused by hurdle residue (K500 in SARS-CoV-2 and T114 in EV71[18]). Even so, discrepancy still exists between the two RdRps for the asynchronous translocation. In EV71, the more concentrated interactions in the downstream region in EV71 RdRp further impedes the translocation of the template strand and thus contributes to the asynchronous translocation[19]. By contrast, in SARS-CoV-2 RdRp, the protein interactions alongside the template strand are evenly distributed (Figure 3a and 3c, S3a and S3c), and thus would not govern the asynchronous translocation.

Besides viral RdRps, the asymmetric movement of template and nascent strands has also been proposed in other classes of polymerases[22,53]. For example, a computational study of bacteriophage T7 RNAP has demonstrated that the nascent strand almost accomplished its translocation while the template strand just began to translocate due to its stacking with the Y helix residue F644[22]. Although both T7 RNAP and SARS-CoV-2 RdRp have demonstrated the asynchronous translocation, the dynamics of transition nucleotide is different. In T7 RNAP the transition nucleotide moves together with the template strand, whereas in SARS-CoV-2 RdRp the transition nucleotide starts to translocate in S1 state when the template strand seldom moves. Besides T7 RNAP, the eukaryotic Pol II also adopts the asymmetric translocation mechanism[53]. However, herein the definition of asymmetry is different from that in SARS-CoV-2 RdRp or T7 RNAP, and mainly describes the different motions between the transition nucleotide and the DNA:RNA duplex. In Pol II, the template DNA strand and the nascent RNA strand translocate simultaneously and almost accomplish their translocation motions during the rate-

limiting step from the pre-translocation state to the first intermediate state, whereas the transition nucleotide lags behind and only moves by around 1/3 of the whole translocation pathway due to its stacking interaction with the bridge helix residue Y836. Overall, the asymmetric translocation is found not only in SARS-CoV-2 RdRp but also in other RNA polymerases, suggesting it could be a common feature. Even so, there exist obvious discrepancies in diverse RNA polymerases such as the characterization of "asymmetry" and the underlying molecular mechanisms governing the asymmetry.

Our study has suggested that the translocation of the SARS-CoV-2 RdRp adopts a coupled Brownian-ratchet model, where template and nascent strand follows its Brownian motion individually but shares the same ratchet to stabilize the post-translocation state. This is consistent with a recent single molecular experiment, in which the researchers have found out the elongation kinetic data of SARS-CoV-2 RdRp could be well fitted by the force-dependent rate equation derived from the Brownian-ratchet model[58]. Actually, the Brownian-ratchet model has also been adopted by the translocation of the DNA-dependent RNA polymerases. For example, it has been suggested that yeast Pol II can oscillate between the pre- and post-translocation states driven by the thermal fluctuation of the bridge helix, the motion of which is highly correlated with that of transition nucleotide and thus facilitates the translocation[17]. Besides the RNA polymerase, the DNA polymerase has also been suggested to follow the Brownian-ratchet model as the single-molecule experiments have demonstrated that the dsDNA could thermally diffuse forward and backward until the dNTP as the "ratchet" to dock and stabilize the post-translocation state[59,60]. Therefore, the Brownian-ratchet model is commonly valid for characterizing the translocation not only in the RNA polymerases but also in the DNA polymerases. Distinct from the "Brownian-ratchet" model, the "power-stroke" model highlights that the translocation could not be accomplished spontaneously but require the input of extra chemical energy. A previous work based on the crystal structures has suggested that the dissociation of PPi in T7 RNAP would pump the energy to power the O-helix closer to the transition nucleotide and thereby drive the DNA-RNA heteroduplex to the post-translocation state[20]. However, this picture has been challenged by computational works which have proposed that the PPi release did not couple tightly with the O-helix opening and the thermal oscillations of the O-helix are sufficient to drive the translocation[21]. In a more complex scenario, EV71 RdRp has been proposed to adopt a combination of the "Brownian-ratchet" model and the "power stroke" model. On one hand, the nascent strand was subject to the Brownian motion with a fast equilibrium between the pre- and the post-translocation states.

On the other hand, the presence of hurdle residues as well as uneven distributed nucleotide-protein interactions alongside the template strand together hinder the translocation of the downstream region of the template strand, and the chemical energy released by PPi dissociation would power the template strand to accomplish the last step of the translocation[18,19]. Although both our model of SARS-CoV-2 RdRp and previous works of EV71 RdRp have shown that the hurdle residue could create an energy barrier to impede the translocation of the template strand, in the current study, the thermal fluctuations of duplex are sufficient for the SARS-CoV-2 RdRp to cross over the barriers and accomplish the translocation. In this regard, our study has suggested that the translocation of SARS-CoV-2 RdRp still follows the Brownian-ratchet model with the motions of the two strands distinct while simultaneously coupled.

Our MSM has demonstrated that the hurdle residue K500 in the motif G of SARS-CoV-2 RdRp serves as the key structural component to determine the rate-limiting step of the translocation. Interestingly, this residue is conserved in some viral RdRps[61], including Murine hepatitis virus (MHV) and human coronavirus 229E (HCoV-229E), suggesting that these RdRp may adopt a similar translocating mode as SARS-CoV-2 RdRp. Intriguingly, this Lysine has been replaced by an alternative positively-charged residue Arginine in some viral RdRps, for example, R459 in the reovirus $\lambda 3$ (LAM3) RdRp and R217 in the bovine viral diarrhea virus (BVDV). We speculate that the bulkier side chain of Arginine may cause stronger steric effect than Lysine and thus raise the energy barrier for the respective translocation. It has also been noticed that the position of K500 in SARS-CoV-2 RdRp can also been accommodated by the polar but neutral protein residues in other RdRps, for example, T114 in poliovirus RdRp and EV71 RdRp, as well as S96 in hepatitis C virus (HCV) RdRp. The role of hurdle residue T114 in EV71 RdRp has been proposed in a previous work[19], suggesting its side chain could also act as a hindrance for the translocation. However, we speculate that the resulting energy barrier may be lower than that caused by the positively charged residue which could form extra electrostatic attraction with the template strand. We have also noticed that some viral RdRps hold small amino acids in the same position as K500 in SARS-CoV-2 RdRp. For example, the counterpart residue is A400 in the rotavirus, G220 in the classical swine fever virus and G851 in the bacteriophage Qβ. In such cases, the steric effect may be obviously reduced and the mechanism of the translocation in these RdRps could be different from the mechanism proposed in the current study. Therefore, the translocation mechanism of SARS-CoV-2 RdRp may be applicable for viral RdRps with polar or positively charged residues in the same position as

K500 in SARS-CoV-2 RdRp, while viral RdRps with small amino acids may utilize alternative strategies to control the translocation.

We have also noticed that the timescale of the translocation estimated by our MSM occurs at tens of microseconds, which is relatively fast in comparison with the timescales of ~6ms/bp for one turnover of NAC measured by the magnetic tweezer experiment[30,58]. This variation is mainly due to that translocation is only one step constituting the NAC while the period of ~6ms measured by the experiment contains several steps of conformational changes in the NAC, such as the catalysis and active site isomerization which have been proposed as the rate-limiting steps in RNA polymerases such as Pol II[62]. Even so, we speculate that the timescale of translocation obtained from our MSM could still be underestimated as our structural model in the pre-translocation state contains only 9 base pairs from *+1* site to *-8* site. However, a recent cryo-EM structure has resolved the SARS-CoV-2 RdRp with the RNA duplex up to *-28* site which are flanked by two nsp8 extensions[11]. The extra interactions between the nsp8 extensions and dsRNA in the upstream region beyond *-8* site may lead to a slower translocation dynamics than that observed in the current study. Therefore, our results can set a lower limit for the translocation timescale in SARS-CoV-2. Even so, the relative thermodynamics and kinetics between the translocation metastable states identified in our current study can still be valid and have provided intriguing molecular insights into the translocation of SARS-CoV-2 RdRp.

CONCLUSIONS

Overall, our study has investigated the translocation dynamics of the template-nascent duplex in SARS-CoV-2 RdRp by constructing MSM based on numerous MD simulations. We have found out that the translocation is asynchronous not only within the nascent strand, but also between the template and nascent strand. Such an asynchronous translocation pattern is contributed by the uneven protein environment surrounding the nascent strand as well as the hurdle residue K500 nearby the template strand, which together underline the rate-limiting step along the translocation pathway. In the synthesized trajectories of hundreds of milliseconds based on MSM, we have observed the thermal oscillations of RdRp between the pre- and post-translocation states multiple times through the two intermediate states, which characterizes the Brownian-ratchet model. Our study has not only offered valuable insights into the translocation dynamics of RdRp, but also shed light on the transcriptional mechanisms of SARS-CoV-2, which may further orientate the rational design of inhibitors targeting at the dynamic processes in the NAC of viral RdRp[4–9,27,63,64].

COMPUTATIONAL METHODS

Modelling the pre-translocation state and the post-translocation state. We have modelled the pre- and post-translocation states based on the cryo-EM structures[10] of SARS-CoV-2 RdRp (PDB ID: 7c2k and 7bzf[10], respectively). The remdesivir monophosphate at the *-1* site of the nascent strand in the cryo-EM structure of pre-translocation state was modified to adenosine monophosphate. To ensure that the pre- and post-translocation states demonstrate one-base-pair translocation, the nucleotides in the post-translocation state were mutated by Coot 0.8.9[65] software to be consistent with the sequence of the pre-translocation state, and the length of template-nascent duplex in two states were modified to ensure the two states have the same length of dsRNA. See SI Section 1.1 for details.

Generating the preliminary pathway. We have obtained two preliminary pathways connecting the pre- and post-translocation states using the modified Climber algorithm[52,53]. Specifically, the template-nascent duplex was gradually driven from the pre-translocation state to the post-translocation state for the forward translocation by an external energy, which is constituted by a series of harmonic potentials on the atom pairs between the nucleotides and surrounding protein residues. Following the same scheme, we have also generated the preliminary backward translocation pathway. In each pathway, twelve representative conformations were selected to initiate the unbiased MD simulations. See SI Section 1.2 for details.

MD simulations. For each of the 24 chosen conformations from the preliminary pathways, we have solvated it by TIP3P water[66] in a dodecahedron box of 146.1Å×146.1Å×146.1Å. 30 Na^+ ions were added to neutralize the system and the complete system contains 212,325 atoms. After energy minimization and equilibration, each conformation was used to seed five 200ns trajectories under the NVT ensemble (T=298K) with different initial velocities. The Amber99sb-ildn force field[67] was used to simulate the whole system, and all the MD simulations were performed using GROMACS 5.0.4 package[68]. See SI Section 1.3 for details.

Constructing and validating the Markov State Model. After removing the first 20ns in each trajectory, we have obtained 120×180ns MD conformational ensemble for constructing MSM. In particular, the time-structure independent components analysis (tICA)[69–71] was utilized to define the distance metric for splitting the MD conformations into microstates. The Spectral oASIS method[72,73] was used to choose the representative structural features for tICA, which are composed of the atom pairs between the nucleotides and protein motifs. Other hyper

parameters for tICA including the number of tICs, the tICA lag time as well as the number of microstates were determined by the generalized matrix Rayleigh quotient (GMRQ) method[74,75] to ensure the essential and slowest relaxing degrees of freedom were captured. Accordingly, we have computed the correlation between atom pairs at a tICA lag time of 6 ns and used the first 2 tICs to reduce the dimension of freedom. Subsequently, the MD conformations were split into 800 microstates using the K-Center clustering algorithm[76] and a lag time of 50 ns was chosen for MSM construction after validation by the implied timescales and self-transition probability derived from the Chapman-Kolmogorov test[77]. To interpret the molecular mechanisms underlying the translocation dynamics, we have performed the kinetic lumping using the PCCA+ algorithm implemented in MSMBuilder 3.8.0[78–80] to further group the 800 microstates into 4 metastable states. The populations of the metastable states and the mean first message times (MFPTs) were computed using the Markov-Chain Monte Carlo (MCMC) simulations based on the transition probability matrix of the microstate-MSM. The flux from pre- to post-translocation states was analyzed using the transition path theory (TPT) algorithm[81] implemented in the PyEMMA package 2.5.2[82]. See SI Section 2 for details.

Mutant simulations. One conformation in the transition area between S2 state and S3 state was chosen as the structural basis for performing the K500A mutant simulations. After mutating the Lysine to Alanine, we have set up the system following a similar procedure as used for the above wildtype RdRp simulations and performed three 30 ns MD simulations in parallel under NVT ensemble (T=298K). As a control, the same conformation of wildtype RdRp representing the transition state between S2 and S3 states was also used for 3×30ns NVT simulations. See SI Section 3 for details.

AUTHOR INFORMATION

Corresponding Author

Lu Zhang – *State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian 35000, China; Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Fuzhou, Fujian 35000, China; University of Chinese Academy of Sciences, Beijing 100190, China;*

Authors

Xiaowei Wang – *Department of Chemical and Biological Engineering and Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong 999077;*

Yuan Yao – *Department of Chemical and Biological Engineering and Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong 999077;*

Xin Gao – *Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia; Computer, Electrical and Mathematical Sciences and Engineering Division, KAUST, Thuwal 23955, Saudi Arabia;*

Author contributions

X. W. performed MD simulations, constructed the MSM and analyzed the data. L. Z. conceived the research and designed the data analysis. X. G. and Y. Y. arranged the computational environment for the use of the KAUST supercomputer and X-GPU cluster, respectively. X. W., Y. Y., X. G., and L. Z. discussed the results and wrote the manuscript.

Notes

The authors declare no competing interests.

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the National Key R&D program of China (2021YFA1502300), the National Natural Science Foundation of China (21733007) and the Hong Kong Research Grant Council NSFC/RGC Joint Research Scheme (N_HKUST635/20). This research made use of the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST), as well as the computing resources of the X-GPU cluster supported by the Hong Kong Research Grant Council Collaborative Research Fund: C6021-19EF. The authors would like to thank Prof. Xuhui Huang (University of Wisconsin-Madison) for helpful discussions.

REFERENCES

(1) WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data https://covid19.who.int/ (accessed Jul 11, 2021).

(2) Acter, T.; Uddin, N.; Das, J.; Akhter, A.; Choudhury, T. R.; Kim, S. Evolution of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) as Coronavirus

Disease 2019 (COVID-19) Pandemic: A Global Health Emergency. *Sci. Total Environ.* **2020**, *730*, 138996.

(3) Malone, B.; Urakova, N.; Snijder, E. J.; Campbell, E. A. Structures and Functions of Coronavirus Replication–Transcription Complexes and Their Relevance for SARS-CoV-2 Drug Design. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (1), 21–39.

(4) Li, Y.; Zhang, D.; Gao, X.; Wang, X.; Zhang, L. 2′- and 3′-Ribose Modifications of Nucleotide Analogues Establish the Structural Basis to Inhibit the Viral Replication of SARS-CoV-2. *J. Phys. Chem. Lett.* **2022**, *13* (13), 4111–4118.

(5) Byléhn, F.; Menéndez, C. A.; Perez-Lemus, G. R.; Alvarado, W.; Pablo, J. J. de. Modeling the Binding Mechanism of Remdesivir, Favilavir, and Ribavirin to SARS-CoV-2 RNA-Dependent RNA Polymerase. *ACS Cent. Sci.* **2021**, *7* (1), 164–174.

(6) Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A. Höbartner, C.; Cramer, P. Mechanism of SARS-CoV-2 Polymerase Stalling by Remdesivir. *Nat. Commun.* **2021**, *12* (1), 279.

(7) Kwong, C. H. T.; Mu, J.; Li, S.; Fang, Y.; Liu, Q.; Zhang, X.; Kam, H.; Lee, S. M. Y.; Chen, Y.; Deng, F. Reviving Chloroquine for Anti-SARS-CoV-2 Treatment with Cucurbit[7]Uril-Based Supramolecular Formulation. *Chinese Chem. Lett.* **2021**, *32* (10), 3019–3022..

(8) Zhang, L.; Zhang, D.; Wang, X.; Yuan, C.; Li, Y.; Jia, X.; Gao, X.; Yen, H.-L.; Pak-Hang Cheung, P.; Huang, X. 1′-Ribose Cyano Substitution Allows Remdesivir to Effectively Inhibit Nucleotide Addition and Proofreading during SARS-CoV-2 Viral RNA Replication. *Phys. Chem. Chem. Phys.* **2021**, *23* (10), 5852–5863.

(9) Luo, X.; Xu, T.; Gao, X.; Zhang, L. Alternative Role of Motif B in Template Dependent Polymerase Inhibition. *Chinese J. Chem. Phys.* **2022**, *35* (3).

(10) Wang, Q.; Wu, J.; Wang, H.; Gao, Y.; Liu, Q.; Mu, A.; Ji, W.; Yan, L.; Zhu, Y.; Zhu, C. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **2020**, *182* (2), 417–428.

(11) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of Replicating SARS-CoV-2 Polymerase. *Nature* **2020**, *584* (7819), 154–156.

(12) Chen, J.; Malone, B.; Llewellyn, E.; Grasso, M.; Shelton, P. M. M.; Olinares, P. D. B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **2020**, *182* (6), 1560–1573..

(13) Yan, L.; Ge, J.; Zheng, L.; Zhang, Y.; Gao, Y.; Wang, T.; Huang, Y.; Yang, Y.; Gao, S.; Li, M. Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell* **2021**, *184* (1), 184–193.

(14) Yan, L.; Yang, Y.; Li, M.; Zhang, Y.; Zheng, L.; Ge, J.; Huang, Y. C.; Liu, Z.; Wang, T.; Gao, S. Coupling of N7-Methyltransferase and 3′-5′ Exoribonuclease with SARS-CoV-2 Polymerase Reveals Mechanisms for Capping and Proofreading. *Cell* **2021**, *184* (13), 3474–3485.

(15) Yan, L.; Zhang, Y.; Ge, J.; Zheng, L.; Gao, Y.; Wang, T.; Jia, Z.; Wang, H.; Huang,

Y.; Li, M. Architecture of a SARS-CoV-2 Mini Replication and Transcription Complex. *Nat. Commun.* **2020**, *11* (1), 5874.

(16) Gong, P.; Peersen, O. B. Structural Basis for Active Site Closure by the Poliovirus RNA-Dependent RNA Polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (52), 22505–22510.

(17) Silva, D. A.; Weiss, D. R.; Avila, F. P.; Da, L. T.; Levitt, M.; Wang, D.; Huang, X. Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (21), 7665–7670.

(18) Wang, M.; Li, R.; Shu, B.; Jing, X.; Ye, H. Q.; Gong, P. Stringent Control of the RNA-Dependent RNA Polymerase Translocation Revealed by Multiple Intermediate Structures. *Nat. Commun.* **2020**, *11*, 2605.

(19) Shu, B.; Gong, P. Structural Basis of Viral RNA-Dependent RNA Polymerase Catalysis and Translocation. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (28), E4005–E4014.

(20) Yin, Y. W.; Steitz, T. A. The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. *Cell* **2004**, *116* (3), 393–404.

(21) Da, L. T.; E, C.; Duan, B.; Zhang, C.; Zhou, X.; Yu, J. A Jump-from-Cavity Pyrophosphate Ion Release Assisted by a Key Lysine Residue in T7 RNA Polymerase Transcription Elongation. *PLoS Comput. Biol.* **2015**, *11* (11), e1004624.

(22) Da, L.-T.; E, C.; Shuai, Y.; Wu, S.; Su, X.-D.; Yu, J. T7 RNA Polymerase Translocation Is Facilitated by a Helix Opening on the Fingers Domain That May Also Prevent Backtracking. *Nucleic Acids Res.* **2017**, *45* (13), 7909–7921.

(23) Wakchaure, P. D.; Ghosh, S.; Ganguly, B. Revealing the Inhibition Mechanism of RNA-Dependent RNA Polymerase (RdRp) of SARS-CoV-2 by Remdesivir and Nucleotide Analogues: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2020**, *124* (47), 10641–10652.

(24) Zhang, L.; Zhou, R. Structural Basis of the Potential Binding Mechanism of Remdesivir to SARS-CoV-2 RNA-Dependent RNA Polymerase. *J. Phys. Chem. B* **2020**, *124* (32), 6955–6962.

(25) Ribaudo, G.; Ongaro, A.; Oselladore, E.; Zagotto, G.; Memo, M.; Gianoncelli, A. A Computational Approach to Drug Repurposing against SARS-CoV-2 RNA Dependent RNA Polymerase (RdRp). *J. Biomol. Struct. Dyn.* **2020**, *40* (3), 1101–1108.

(26) Mohammad, A.; Al-Mulla, F.; Wei, D. Q.; Abubaker, J. Remdesivir Md Simulations Suggest a More Favourable Binding to SARS-CoV-2 RNA Dependent RNA Polymerase Mutant P323l than Wild-Type. *Biomolecules* **2021**, *11* (7), 919.

(27) Elfiky, A. A. SARS-CoV-2 RNA Dependent RNA Polymerase (RdRp) Targeting: An in Silico Perspective. *J. Biomol. Struct. Dyn.* **2021**, *39* (9), 3204–3212.

(28) Yuan, C.; Goonetilleke, E. C.; Unarta, I. C.; Huang, X. Incorporation Efficiency and Inhibition Mechanism of 2′-Substituted Nucleotide Analogs against SARS-CoV-2 RNA-Dependent RNA Polymerase. *Phys. Chem. Chem. Phys.* **2021**, *23* (36), 20117–20128.

(29) Arnold, J. J.; Cameron, C. E. Poliovirus RNA-Dependent RNA Polymerase (3Dpol):

Pre-Steady-State Kinetic Analysis of Ribonucleotide Incorporation in the Presence of Mg2+. *Biochemistry* **2004**, *43* (18), 5126–5137.

(30) Seifert, M.; Bera, S. C.; Nies, P. van; Kirchdoerfer, R. N.; Shannon, A.; Le, T.-T.-N.; Grove, T. L.; Papini, F. S.; Arnold, J. J.; Almo, S. C. Signatures and Mechanisms of Efficacious Therapeutic Ribonucleotides against SARS-CoV-2 Revealed by Analysis of Its Replicase Using Magnetic Tweezers. *bioRxiv* **2020**, 2020.08.06.240325.

(31) Wang, W.; Liang, T.; Sheong, F. K.; Fan, X.; Huang, X. An Efficient Bayesian Kinetic Lumping Algorithm to Identify Metastable Conformational States via Gibbs Sampling. *J. Chem. Phys.* **2018**, *149* (7), 072337.

(32) Zhang, L.; Pardo-Avila, F.; Unarta, I. C.; Cheung, P. P. H.; Wang, G.; Wang, D.; Huang, X. Elucidation of the Dynamics of Transcription Elongation by RNA Polymerase II Using Kinetic Network Models. *Acc. Chem. Res.* **2016**, *49* (4), 687–694.

(33) Lane, T. J.; Bowman, G. R.; Beauchamp, K.; Voelz, V. A.; Pande, V. S. Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories. *J. Am. Chem. Soc.* **2011**, *133* (45), 18413–18419.

(34) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018,** *140*(7)*,* 2386–2396.

(35) Malmstrom, R. D.; Lee, C. T.; Van Wart, A. T.; Amaro, R. E. Application of Molecular-Dynamics Based Markov State Models to Functional Proteins. *J. Chem. Theory Comput.* **2014**, *10* (7), 2648–2657.

(36) Wang, X.; Unarta, I. C.; Cheung, P. P. H.; Huang, X. Elucidating Molecular Mechanisms of Functional Conformational Changes of Proteins via Markov State Models. *Curr. Opin. Struct. Biol.* **2021**, *67*, 69–77.

(37) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105.

(38) Zhang, B. W.; Dai, W.; Gallicchio, E.; He, P.; Xia, J.; Tan, Z.; Levy, R. M. Simulating Replica Exchange: Markov State Models, Proposal Schemes, and the Infinite Swapping Limit. *J. Phys. Chem. B* **2016**, *120* (33), 8289–8301.

(39) Morcos, F.; Chatterjee, S.; McClendon, C. L.; Brenner, P. R.; López-Rendón, R.; Zintsmaster, J.; Ercsey-Ravasz, M.; Sweet, C. R.; Jacobson, M. P.; Peng, J. W. Modeling Conformational Ensembles of Slow Functional Motions in Pin1-WW. *PLoS Comput. Biol.* **2010**, *6* (12), e1001015.

(40) Buchete, N. V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112* (19), 6057–6069.

(41) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25,* 135–144.

(42) Feng, Y.; Zhang, L.; Wu, S.; Liu, Z.; Gao, X.; Zhang, X.; Liu, M.; Liu, J.; Huang, X.; Wang, W. Conformational Dynamics of Apo-GlnBP Revealed by Experimental and Computational Analysis. *Angew. Chemie - Int. Ed.* **2016**, *55* (45), 13990–13994.

(43) Peng, S.; Wang, X.; Zhang, L.; He, S.; Zhao, X. S.; Huang, X.; Chen, C. Target Search and Recognition Mechanisms of Glycosylase AlkD Revealed by Scanning FRET-FCS

and Markov State Models. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (36), 21889–21895.

(44) Zhang, L.; Jiang, H.; Sheong, F. K.; Pardo-Avila, F.; Cheung, P. P. H.; Huang, X. Constructing Kinetic Network Models to Elucidate Mechanisms of Functional Conformational Changes of Enzymes and Their Recognition with Ligands. In *Methods in Enzymology*; Academic Press Inc., **2016,** *578,* 343–371.

(45) Zhang, L.; Wu, S.; Feng, Y.; Wang, D.; Jia, X.; Liu, Z.; Liu, J.; Wang, W. Ligand-Bound Glutamine Binding Protein Assumes Multiple Metastable Binding Sites with Different Binding Affinities. *Commun. Biol.* **2020**, *3* (1), 419.

(46) Unarta, I. C.; Cao, S.; Kubo, S.; Wang, W.; Cheung, P. P. H.; Gao, X.; Takada, S.; Huang, X. Role of Bacterial RNA Polymerase Gate Opening Dynamics in DNA Loading and Antibiotics Inhibition Elucidated by Quasi-Markov State Model. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (17), e2024324118.

(47) Nuria Plattner Gianni De Fabritiis G, Frank Noe, S. D. Complete Protein–Protein Association Kinetics in Atomic Detail Revealed by Molecular Dynamics Simulations and Markov Modelling. *Nat. Chem.* **2017**, *9* (10), 1005–1011.

(48) Heo, L.; Feig, M. Experimental Accuracy in Protein Structure Refinement via Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (52), 13276–13281.

(49) Da, L. T.; Pardo Avila, F.; Wang, D.; Huang, X. A Two-State Model for the Dynamics of the Pyrophosphate Ion Release in Bacterial RNA Polymerase. *PLoS Comput. Biol.* **2013**, *9* (4), e1003020.

(50) Wang, B.; Feig, M.; Cukier, R. I.; Burton, Z. F. Computational Simulation Strategies for Analysis of Multisubunit RNA Polymerases. *Chem. Rev.* **2013,** *113*(11), 8546–8566.

(51) Da, L. T.; Pardo-Avila, F.; Xu, L.; Silva, D. A.; Zhang, L.; Gao, X.; Wang, D.; Huang, X. Bridge Helix Bending Promotes RNA Polymerase II Backtracking through a Critical and Conserved Threonine Residue. *Nat. Commun.* **2016**, *7* (1), 11244.

(52) Weiss, D. R.; Levitt, M. Can Morphing Methods Predict Intermediate Structures? *J. Mol. Biol.* **2009**, *385* (2), 665–674.

(53) Silva, D.-A.; Weiss, D. R.; Pardo Avila, F.; Da, L.-T.; Levitt, M.; Wang, D.; Huang, X. Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (21), 7665–7670.

(54) Yang, H.; Rao, Z. Structural Biology of SARS-CoV-2 and Implications for Therapeutic Development. *Nat. Rev. Microbiol.* **2021**, *19* (11), 685–700.

(55) Jia, H.; Gong, P. A Structure-Function Diversity Survey of the RNA-Dependent RNA Polymerases from the Positive-Strand RNA Viruses. *Front. Microbiol.* **2019**, *10* (8), 1945.

(56) Wu, J.; Liu, W.; Gong, P. A Structural Overview of RNA-Dependent RNA Polymerases from the Flaviviridae Family. *Int. J. Mol. Sci.* **2015**, *16* (6), 12943–12957.

(57) Sholders, A. J.; Peersen, O. B. Distinct Conformations of a Putative Translocation

Element in Poliovirus Polymerase. *J. Mol. Biol.* **2014**, *426* (7), 1407–1419.

(58) Bera, S. C.; Seifert, M.; Kirchdoerfer, R. N.; van Nies, P.; Wubulikasimu, Y.; Quack, S.; Papini, F. S.; Arnold, J. J.; Canard, B.; Cameron, C. E. The Nucleotide Addition Cycle of the SARS-CoV-2 Polymerase. *Cell Rep.* **2021**, *36* (9), 109650.

(59) Morin, J. A.; Cao, F. J.; Lázaro, J. M.; Arias-Gonzalez, J. R.; Valpuesta, J. M.; Carrascosa, J. L.; Salas, M.; Ibarra, B. Mechano-Chemical Kinetics of DNA Replication: Identification of the Translocation Step of a Replicative DNA Polymerase. *Nucleic Acids Res.* **2015**, *43* (7), 3643–3652.

(60) Arias-Gonzalez, J. R. A DNA-Centered Explanation of the DNA Polymerase Translocation Mechanism. *Sci. Rep.* **2017**, *7* (1), 7566.

(61) Gorbalenya, A. E.; Pringle, F. M.; Zeddam, J. L.; Luke, B. T.; Cameron, C. E.; Kalmakoff, J.; Hanzlik, T. N.; Gordon, K. H. J.; Ward, V. K. The Palm Subdomain-Based Active Site Is Internally Permuted in Viral RNA-Dependent RNA Polymerases of an Ancient Lineage. *J. Mol. Biol.* **2002**, *324* (1), 47–62.

(62) Yu, J.; Da, L. T.; Huang, X. Constructing Kinetic Models to Elucidate Structural Dynamics of a Complete RNA Polymerase II Elongation Cycle. *Phys. Biol.* **2015**, *12* (1), 016004.

(63) Li, L.; Chang, S. H.; Xiang, J. F.; Li, Q.; Liang, H. H.; Tang, Y. L.; Liu, Y. F. NMR Identification of Anti-Influenza Lead Compound Targeting at PA C Subunit of H5N1 Polymerase. *Chinese Chem. Lett.* **2012**, *23* (1), 89–92.

(64) Fu, Y. H.; Liu, Y. R.; Zheng, Y. P.; Jiang, N.; Yue-Ying-Jiao; Li, W.; Peng, X. L.; He, J. S. An RNA Polymerase I-Driven Human Respiratory Syncytial Virus Minigenome as a Tool for Quantifying Virus Titers and Screening Antiviral Drug. *Chinese Chem. Lett.* **2017**, *28* (1), 131–135.

(65) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D* **2004**, *60* (12), 2126–2132.

(66) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960.

(67) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958.

(68) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(69) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9* (4), 2000–2009.

(70) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1), 15102.

(71) Naritomi, Y.; Fuchigami, S. Slow Dynamics of a Protein Backbone in Molecular Dynamics Simulation Revealed by Time-Structure Based Independent Component

Analysis. *J. Chem. Phys.* **2013**, *139* (21), 215102.

(72) Litzinger, F.; Boninsegna, L.; Wu, H.; Nüske, F.; Patel, R.; Baraniuk, R.; Noé, F.; Clementi, C. Rapid Calculation of Molecular Kinetics Using Compressed Sensing. *J. Chem. Theory Comput.* **2018**, *14* (5), 2771–2783.

(73) Yao, Y.; Cui, R. Z.; Bowman, G. R.; Silva, D. A.; Sun, J.; Huang, X. Hierarchical Nyström Methods for Constructing Markov State Models for Conformational Dynamics. *J. Chem. Phys.* **2013**, *138* (17), 174106.

(74) McGibbon, R. T.; Pande, V. S. Variational Cross-Validation of Slow Dynamical Modes in Molecular Kinetics. *J. Chem. Phys.* **2015**, *142* (12), 124105.

(75) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014**, *10* (4), 1739–1752.

(76) Hochbaum, D. S.; Shmoys, D. B. Best Possible Heuristic for the K-Center Problem. *Math. Oper. Res.* **1985**, *10* (2), 180–184.

(77) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schtte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105.

(78) Deuflhard, P.; Weber, M. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra Appl.* **2005**, *398* (1–3), 161–184.

(79) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7* (2), 147–179.

(80) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **2011**, *7* (10), 3412–3419.

(81) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (45), 19011–19016.

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

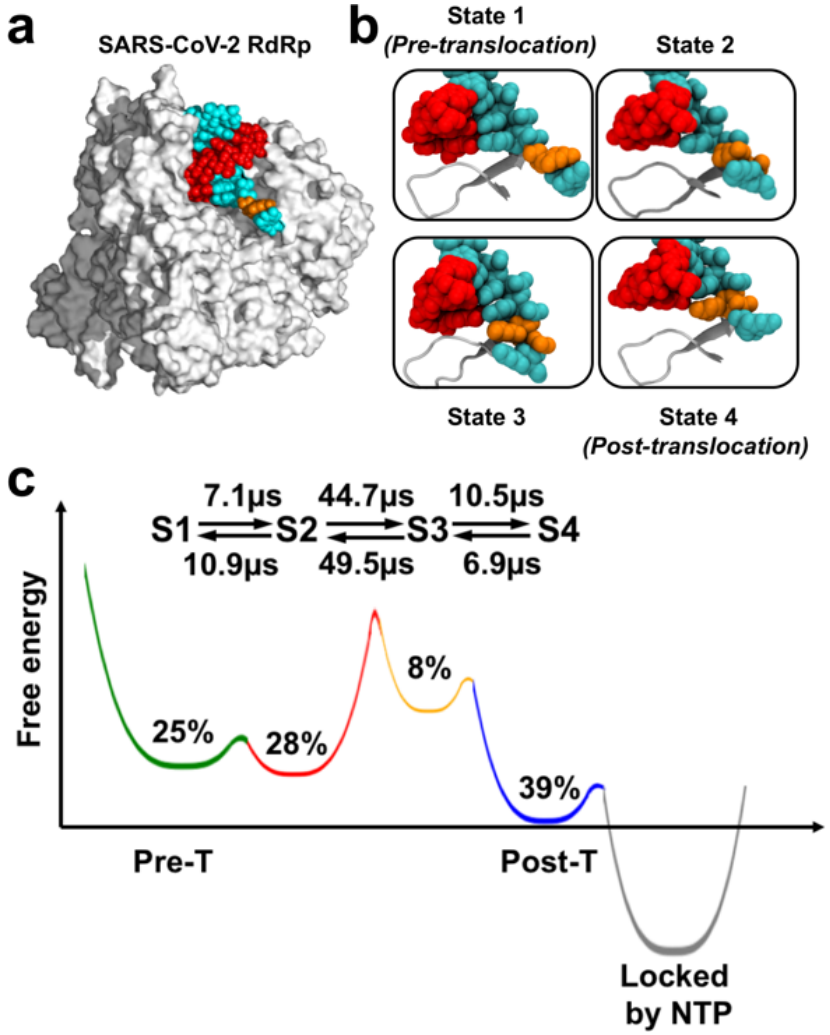

Figure 1. The four metastable states and the free energy landscape that describe the translocation dynamics of SARS-CoV-2 RdRp elucidated by MSM. a The structure of SARS-CoV-2 RdRp in the pre-translocation state. The template and the nascent strand is shown in cyan and red with vdW representation, respectively. The transition nucleotide (TN) is highlight in orange and the protein is shown in white as the background. **b** Representative conformations of four metastable states identified by our MSM. The motif F is shown in grey cartoon as a reference to demonstrate the relative motion of the template-nascent duplex. **c** The schematic free-energy landscape of SARS-CoV-2 RdRp translocation. The equilibrium populations and the mean first passage times are labelled.

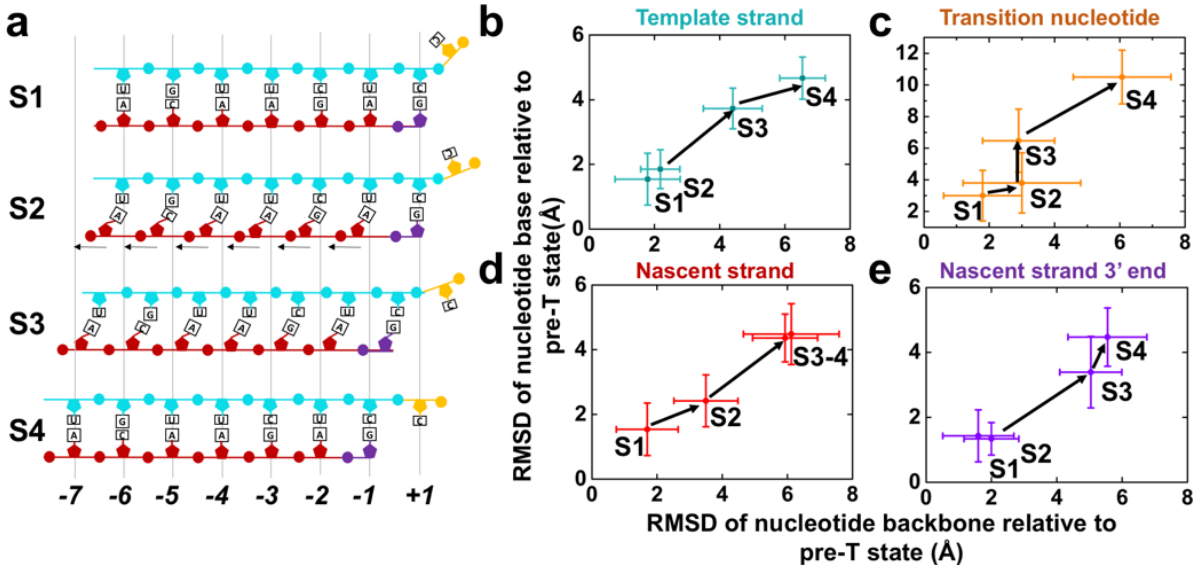

Figure 2. The template-nascent duplex demonstrates an asynchronous translocation. a A schematic cartoon showing the configuration of the template strand (in cyan) and the nascent strand (in red) in the four metastable states elucidated by our model. The TN is shown in orange while the 3'-terminal nucleotide of the nascent strand is shown in purple. The 5'-terminal nucleotide of the template strand is not shown for clarity. **b** The root mean square deviation (RMSD) of the template strand (shown in cyan in **a**) in the four metastable states relative to that in the pre-translocation state. **c** Similar to **b** but for the RMSD of TN. **d** RMSD of the nascent strand (excepting the 3'-terminal nucleotide) in the four states against its conformation in the pre-translocation state. **e** Similar to **d** but for the 3'-terminal nucleotide of the nascent strand. In **b-e**, the x-axis shows the RMSD of the nucleotide backbone phosphate atom relative to those in the pre-translocation state while the y-axis displays for the RMSD of oxygen and nitrogen atoms in nucleotides' bases. The standard deviation and the mean value were estimated using all the MD conformations in each state.

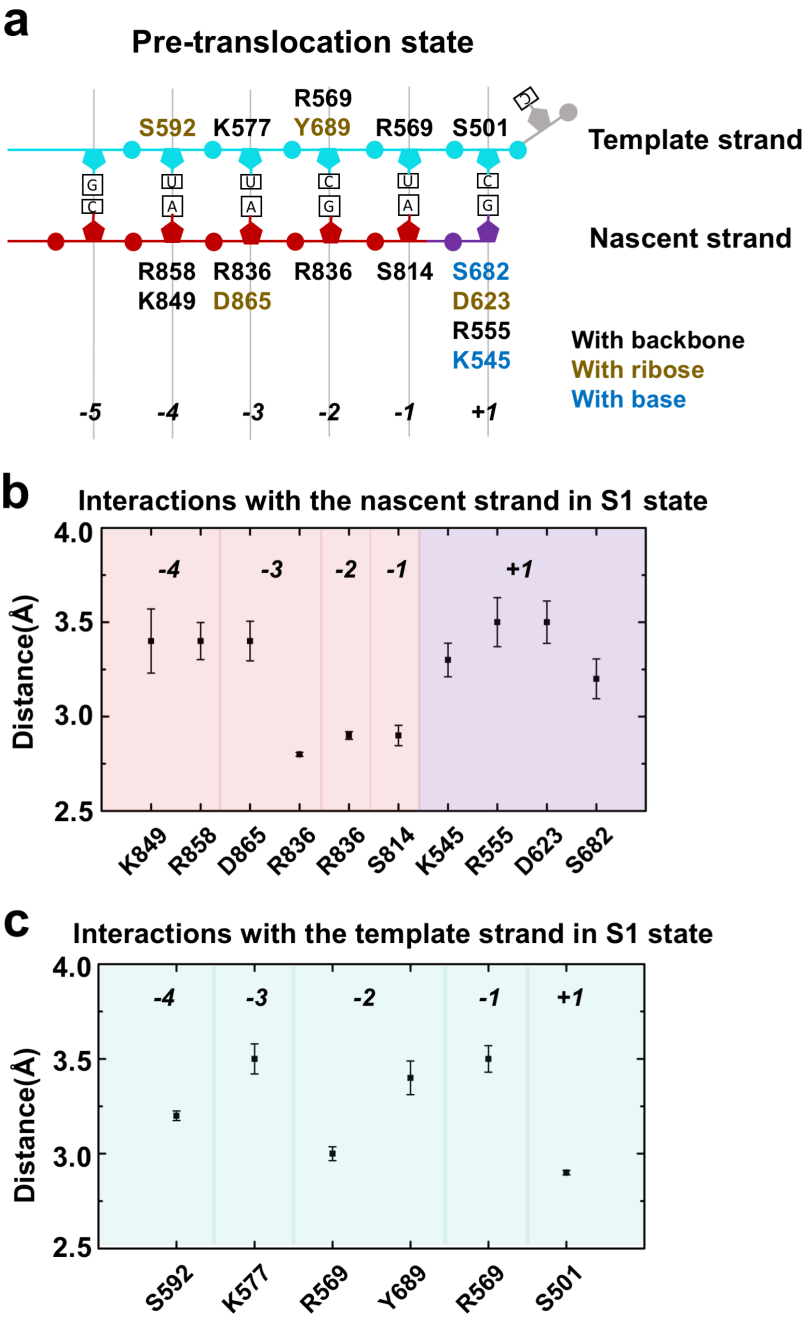

Figure 3. Protein environment surrounding the template-nascent duplex in the pre-translocation (S1) state. a Interactions between nucleotides in the template-nascent duplex from *+1* to *-5* site and their surrounding protein residues in the pre-translocation state. The residues were labelled and colored according to the specific nucleotide moiety (the phosphate backbone (black), ribose (brown) and base (blue)) involved in the interactions with the protein residues. The atoms included for the distance calculation are listed in the Table S3 and S4. **b-c** The mean values and the standard deviations of the distances between the nucleotides and protein residues in the pre-translocation state. The background is colored according to the location of the nucleotides**.** The mean values and standard deviations were obtained by 20 bootstrapping samples. In each sample, 120 trajectories were randomly selected with replacement from the total 120 trajectories.

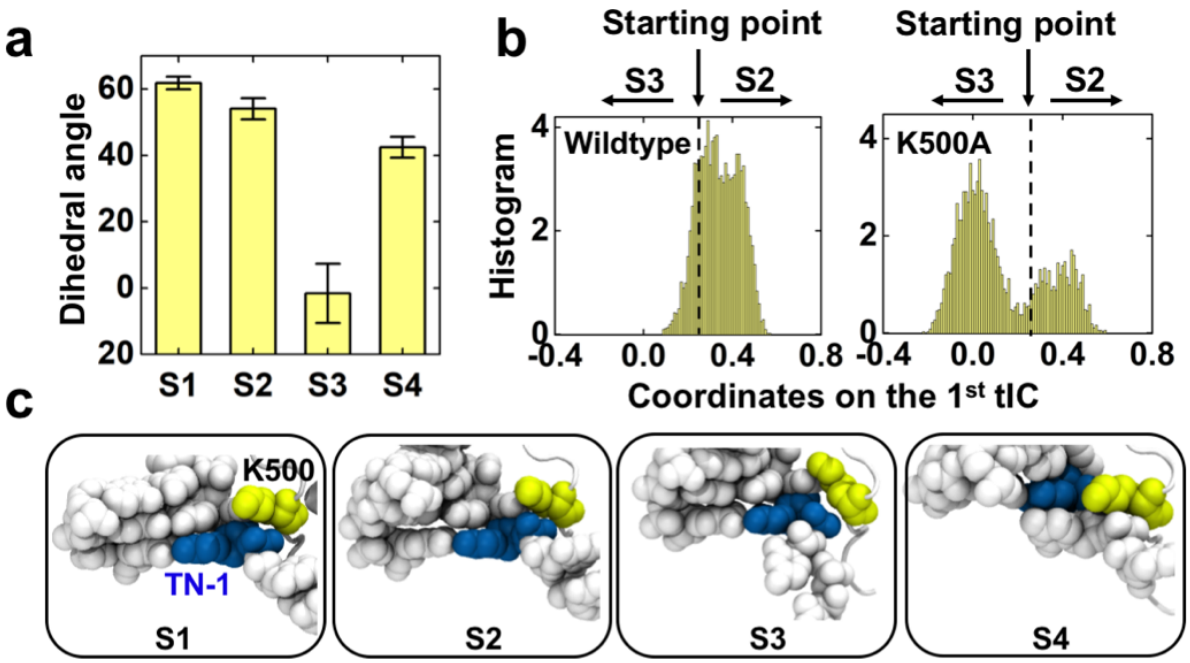

Figure 4. The hurdle residues K500 hinders the translocation of the template strand from S2 and S3. a The dihedral angle representing the side chain orientation of K500 relative to the nucleotide backbones of the template strand in the four metastable states. The dihedral angle was defined in Figure S5. The mean values and standard deviations were estimated by 20 bootstrapping samples. In each sample, 120 trajectories were randomly selected with replacement from the total 120 trajectories. **b** Comparison of the MD conformations of K550A mutant simulations (right panel) with that of the wildtype RdRp simulations (left panel). The MD conformations were projected onto the 1st tIC (x-axis) and the respective histogram (y-axis) was plotted. The dashed line denotes the projection of the initial conformation representing the transition state. The two horizontal arrows pointing to the right and left indicate the direction to the intermediate state S2 and S3, respectively. **c** Representative conformations are shown to demonstrate the orientation of hurdle role of K500 relative to TN-1 (the nucleotide one site upstream than the transition nucleotide (TN)) in four metastable states. K500 and TN-1 are shown in yellow and blue spheres, respectively.

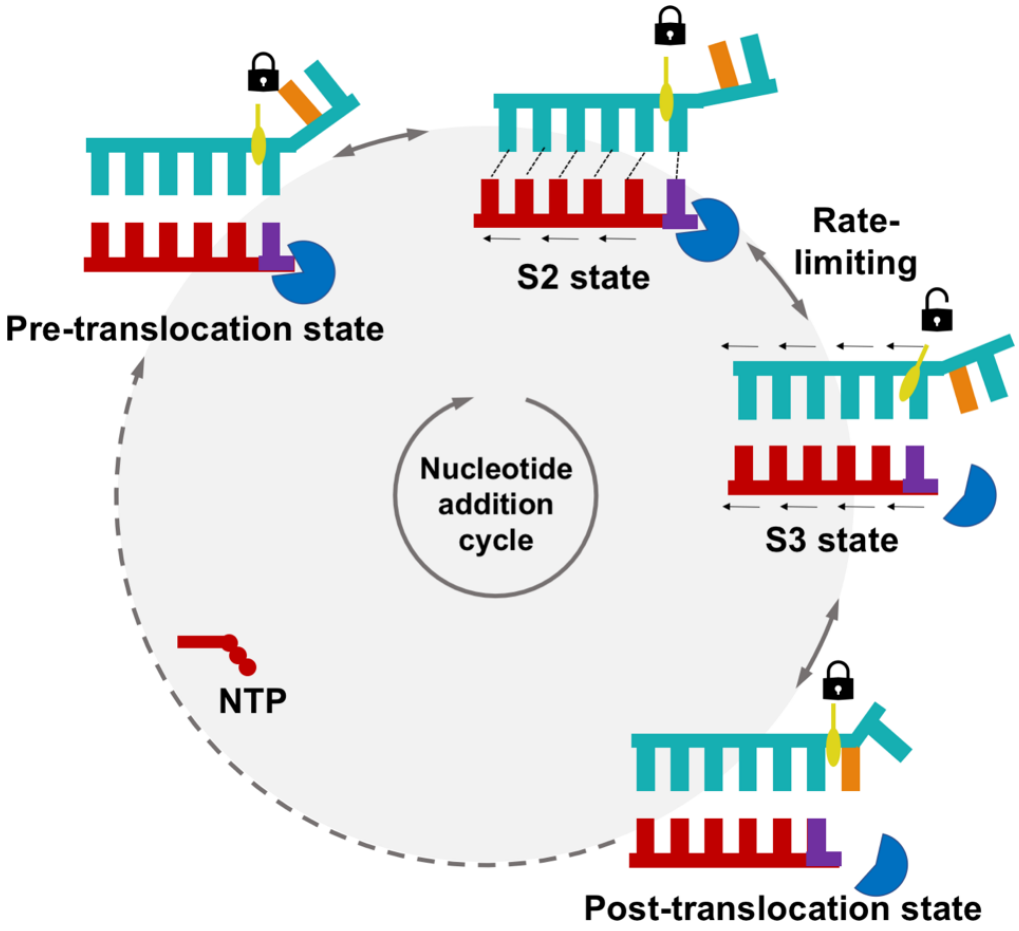

Figure 5. A diagram describing the coupled Brownian-ratchet model of SARS-CoV-2 RdRp translocation. The template strand, the transition nucleotide, the nascent strand and its 3' -terminal nucleotide are displayed in cyan, orange, red and purple, respectively. The protein interaction network surrounding the 3' -terminal nucleotide of the nascent strand is highlighted in blue pie shape, with the ~3/4 pie describing a tight interaction while the ~ ½ pie depicting a loose interaction. The hurdle residue K500 (in yellow) acts as a lock and hinders the movement of template strand from the pre-translocation state to S2 state. The hindrance is released from S2 to S3 and template strand starts to move from S2 to S3 state, corresponding to the rate-limiting step. K500 returns to "lock" the template strand again in the post-translocation state, where the active site is empty and ready for the NTP loading as the common "ratchet" for stabilizing the system in the post-translocation state and initiating the next round of NAC. The dashed and curved arrow from post- to pre-translocation states indicates multiple-step conformational changes (such as NTP loading, active site isomerization and catalysis, etc.) are required.