Role of Bacterial RNA Polymerase Gate Opening Dynamics in DNA Loading and Antibiotics Inhibition Elucidated by *quasi*-Markov State Model

Ilona Christy Unarta[1,2], Siqin Cao[2,3], Shintaroh Kubo[4], Wei Wang[2,3], Peter Pak-Hang Cheung[2,3], Xin Gao[5], Shoji Takada[4], Xuhui Huang[*1,2,3,6]

[1]*Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong.*
[2]*Center of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Kowloon, Hong Kong*
[3]*Department of Chemistry, The Hong Kong University of Science and Technology, Kowloon, Hong Kong.*
[4]*Department of Biophysics, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake, Sakyo Kyoto, 606-8502, Japan*
[5]*Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.*
[6]*Hong Kong Center for Neurodegenerative Diseases, Hong Kong Science Park, Hong Kong*

* To whom correspondence should be addressed. Email: xuhuihuang@ust.hk

Author Contributions: I.C.U. and X.H. designed research; I.C.U., S.K., performed research; I.C.U., S.C., and W.W. analyzed data; I.C.U., P.P.C. X.G., S.T. and X.H. wrote the paper.

Competing Interest Statement: The authors declare no conflict of interest.

Classification: Physical Sciences/Biophysics and Computational Biology

Keywords: Bacterial RNA Polymerase, Markov State Model, Transcription Initiation

Abstract

To initiate transcription, the holoenzyme (RNA Polymerase in complex with σ factor) loads the promoter DNA via the flexible loading gate created by the Clamp and β-lobe, yet their roles in DNA loading have not been characterized. We employed *quasi*-Markov State Model (qMSM) built from extensive molecular dynamics simulations to elucidate the dynamics *Thermus aquaticus* holoenzyme's gate opening. We showed that during gate opening, β-lobe oscillates 4 orders of magnitude faster than the Clamp, whose opening depends on the Switch 2's structure. Myxopyronin, an antibiotic that binds to Switch 2 was shown to undergo a conformational selection mechanism to inhibit Clamp opening. Importantly, we reveal a critical but undiscovered role of β-lobe, whose opening is sufficient for DNA loading even when the Clamp is partially closed. These findings open the opportunity for development of antibiotics targeting β-lobe of RNAP. Finally. we have shown that our qMSMs, which encode non-Markovian dynamics based on the generalized master equation formalism, hold great potential to be widely applied to study biomolecular dynamics.

Significance

The initiation of bacterial transcription requires DNA loading through the flexible loading gate of the RNA Polymerase complex, consisting of the Clamp and β-lobe domain. While the Clamp has been shown to be important in transcription initiation, the role of β-lobe remains unclear. Using *q*MSMs constructed from molecular dynamics simulations, we showed that the opening of β-lobe is orders of magnitude faster than that of the Clamp, which depends on the structural transition of Switch 2. Strikingly, the opening of the β-lobe is sufficient geometrically to accommodate DNA loading even when the Clamp is partially closed. These two observations highlight β-lobe's critical role to allow DNA loading during initiation. Thus, β-lobe is a promising target for bacterial antibiotics. The qMSMs also provides a promising tool to investigate biomolecular dynamics.

Introduction

RNA Polymerase (RNAP) is the enzyme responsible for transcription. Upon promoter binding, the multi-step process of promoter melting occurs, whereby the promoter double-stranded DNA (dsDNA) is being separated into two single-stranded DNAs. The recognition between holoenzyme (RNAP and σ factor) and promoter DNA consists of a series of intermediate states that varies with the species, the promoter sequence, transcription factors, and environmental stress (1–4). For instance, a cryo-EM study of the promoter melting of *M. tuberculosis* (*Mtb*) holoenzyme reveals two stable intermediates (5). However, when *E. coli* holoenzyme is in complex with the transcription activator protein TraR, the populations of promoter melting intermediates would increase, enabling the capture of as many as seven intermediate structures by cryoEM (3). Importantly, it is widely acknowledged that partial or complete closure of the Clamp is required for the stabilization of the initial holoenzyme-dsDNA complex (closed complex/RPc) (3, 5–7). Lastly, the Clamp must be closed in the open complex (Rpo) to stabilize the separated DNA (8–10).

The crab claw conformation of RNAP consists of two highly flexible pincers, the Clamp and β-pincers (β-protrusion & β-lobe) (8, 9, 11), which are important for transcription initiation (12). Indeed, the overall crab claw conformation of RNAP is highly conserved in all prokaryotes (11). During promoter melting, the dsDNA must be separated before being loaded into the active site to form the transcription bubble. In the first step, the holoenzyme binds to the promoter, followed by the separation of the -10 element by β-protrusion and σ-factor (3, 6, 13). Then, the downstream element of the promoter DNA, either in the dsDNA or ssDNA form, must be loaded into the inner cleft of RNAP via the loading gate, which constitutes the space between the Clamp and the β-lobe (12) (Figure 1A). To accommodate the large conformational changes of DNA during its loading through the loading gate into the RNAP inner cleft (12), the gate needs to be first opened and then closed. The current literature suggests that gate opening is essentially synonymous with Clamp opening (12), while the dynamics of promoter melting involving the conformational changes of the loading gate, particularly the role of β-lobe remain elusive.

Previous transcription initiation studies have been focused on the Clamp opening and closure (12), even though β-lobe is an indispensable component of the loading gate. This is likely due to the deceivingly stable conformation of β-lobe, while the Clamp clearly adopts different conformations in the crystal and cryoEM structures during promoter melting (3, 5, 6, 14). However, the larger B-factors (protein flexibility) of the β-lobe in these structures suggest that it is more flexible than the Clamp (15–17). Thus, the β-lobe conformation reported in these structures could be an apparent average of a highly flexible domain (15, 17–19). Hence, both Clamp and β-lobe need to be considered to elucidate the mechanisms of the DNA loading during initiation. Furthermore, dynamic transitions between resolved intermediate states at atomic resolution during promoter melting remain unresolved. The importance of the β-lobe was hypothesized in the recent cryoEM study of *Mtb* RNAP during promoter melting (5). In this study, a cryoEM structure of an intermediate of promoter melting with DNA loaded was captured even though the Clamp is locked in a partially closed conformation by an inhibitor, suggesting β-lobe's important role (instead of

the Clamp) in the first steps of dsDNA loading. Revealing these processes can provide mechanistic insights on how the Clamp and β-lobe serve as the pincers of the loading gate coordinate to allow efficient DNA loading.

The Clamp domain has been the target for antibiotics such as Fidaxomicin, Lipiarmicin, Myxopyronin (Myx), Corallopyronin (Cor), and Squaramides with strong efficacies on inhibiting bacterial gene transcription initiation (2, 8, 20, 21). These antibiotics all bind to the Switch regions, which lie at the base of the Clamp and function to modulate Clamp opening (8, 9). Myx, Cor and Squaramides bind to the Switch 2 region and confine the Clamp in a partially closed conformation (5, 20–22). Hence, inhibition of the Switch 2 region by these antibiotics have been shown to disrupt Clamp motion and transcription initiation (20, 21). Furthermore, this class of inhibitors has been shown to have negligible effects on the activity of eukaryotic RNA Polymerases (20) so that it can be used as an antibiotics for human use. Even though structural studies have described the binding interactions between RNAP and Myx, the dynamic processes for antibiotics to recognize RNAP remain unclear (20, 21). Interestingly, these antibiotic-bound structures are all locked at the unfolded configuration of the Switch 2 region (8, 20, 21), which is different from the folded configuration of the Switch 2 region in the apo holoenzyme structures with the Clamp being open or closed. Based on these observations from static structures, two distinct recognition mechanisms between antibiotics and RNAP can be proposed. In the conformational selection mechanism (23–31), the Switch 2 region can oscillate between folded and unfolded conformations, and the antibiotics then selectively bind to an unfolded conformation and subsequently trap the RNAP in a conformation with the Clamp partially closed. In the induced fit model (23–31), antibiotics bind to a folded Switch 2 region, and this binding further induces the unfolding of the Switch 2 region, hence stabilizing the RNAP-antibiotics bound complex. To resolve these competing recognition mechanisms, it is important to understand the dynamics of molecular recognition of Switch 2 region by antibiotics. Understanding these recognition mechanisms can provide insights to improve the antibiotics design targeting the bacterial RNAP clamp motion.

To complement experiments, all-atom Molecular Dynamics (MD) simulations in the explicit solvent provide a powerful approach to model protein dynamics at atomic resolution. However, the timescale accessible to MD simulations of biomolecular complexes, such as RNAP (at microseconds or shorter), remains orders of magnitude shorter than that of functional conformational changes of interest (often at milliseconds or longer). Markov State Model (MSM) is a popular method that can bridge this timescale gap (32–41), in which continuous dynamics are modeled as Markovian transitions among metastable conformational states at discrete time intervals (lag times). MSMs have been widely applied to study protein conformational changes (40, 42–58), including those involved in the elongation of multi-subunit RNA polymerases (44, 55, 57–63). MSMs must be constructed with long enough lag times to allow Markovian interstate transitions. This imposes a main challenge for MSM studies of RNAP since the lag time is bound by the length of relatively short MD simulations to estimate transition probabilities. To address this challenge, we recently developed the quasi-MSM (qMSM) method based on the generalized master equation formalism, which encodes non-Markovian dynamics into memory kernel

functions (64). qMSM provides a promising approach to study the global conformational changes involved in the RNAP loading gate motion.

In this work, we elucidated the dynamics of the loading gate opening modulated by both the Clamp and the β-lobe. To simulate the loading gate dynamics that occur at millisecond timescale (9), we constructed qMSM from an aggregation of 61.2μs all-atom MD simulations of the holoenzyme (~540,000 atoms). We determined that the clamp domain is the rate-limiting step of the loading-gate opening and closing, as the Clamp opening and closing is four orders of magnitude slower than the β-lobe. Our qMSM reveals two novel intermediate states of Clamp closing, one of which is bound by antibiotics. Myx (an inhibitor of Clamp domain) recognizes RNAP via the conformational selection model for a metastable partially closed state containing an unfolded Switch 2 region. We further elucidated the complete molecular mechanisms by which Clamp domain opens and closes. Excitingly, we discovered that β-lobe is critical to promoter melting. Our results clearly demonstrated the coordination of both Clamp and β-lobe to control the loading gate and generate a sufficiently opened loading gate for DNA loading.

Results and Discussions

Construction and Validation of qMSM for the Loading Gate Dynamics of RNAP

To understand the functional dynamics of the loading gate of *Thermus aquaticus* holoenzyme, we followed our previously published protocol (65) to construct MSMs from extensive all-atom MD simulations. The loading gate opening is a complex global conformational change involving opening as well as swivelling of both Clamp and β-lobe. These complex motions render the initial pathways highly difficult to obtain in order to fully sample the multidimensional flexibility of both pincers. To address this challenge, we adopted the two-basin Go-model Coarse-grained MD simulations (CG-MD) (66) to efficiently drive the loading gate opening/closing via Go-potential, while fully exploring the dynamics of Clamp and β-lobe (see Methods and SI Sec. 2.1). Indeed, our residue-based CG-MD simulations successfully sampled transitions between the open and closed RNAP clamp conformations (Figure S1A). We then back-mapped representative CG-MD structures to all-atom structures, and initiated MD simulations (accumulated 61.2μs all-atom MD simulations of the system with ~540,000 atoms, see Methods for details). To construct MSMs, we applied the time-lagged Independent Component Analysis (tICA) (67–69) to reduce the high dimensional MD conformations to three collective variables and subsequently grouped them into 100 microstates using the k-center clustering (70). Please refer to Methods for more details of microstate MSM construction and validation.

To interpret biological mechanisms underlying loading gate opening, we utilized our recently developed qMSM (64) method to build a model containing only 4 metastable macrostates (State S1 to S4 in Figure 1) via kinetic lumping from 100 microstates (see Methods). To describe the slow timescale of a complex biomolecular conformational changes (occurring at millisecond timescale), markovian MSMs were typically built with hundreds of microstates, making the interpretation of biological significance highly difficult. Here, qMSM, which can recover the slow timescale only with a few numbers of states, is utilized to describe the slow opening and closing

of the loading gate. Different from MSM, qMSM considers non-Markovian dynamics via time-dependent memory kernels ($\boldsymbol{K}(\tau)$) using the generalized master equation (Eq. (1)). For our system, the memory kernels are shown to relax at τ_K = 30ns (see Methods and Figure S6A), and qMSM built at τ_K can accurately predict the dynamics as compared with MD simulations (Figure S6C). Strikingly, we show that qMSM greatly outperforms 4-state MSMs, which failed the Chapman-Kolmogorov test even when built with a lag time as long as 100 ns (Figure S6B) and predicted faster dynamics than all-atom MD simulations (Figure S6C-D).

Dynamics of Opening and Closing of Clamp and β-lobe Revealed by qMSM

To examine the inherent oscillations of the opening and closing of the Clamp and β-lobe, we employed the apo structure of *Thermus aquaticus* holoenzyme in the absence of DNA to cover all possible conformations. We conducted a qMSM study on the holoenzyme to reveal the timescale of the Clamp and the β-lobe opening. Using qMSM, we identified four metastable states that correspond to different Clamp conformations, indicating that Clamp domain motion constitutes the slowest motions of the holoenzyme (Figure 1B). We observed that the estimated equilibrium populations consist mostly of the state with open Clamp (S1: 71.4%, similar with PDB ID: 5TJG (6)). This is consistent with the crystal structures of holoenzyme that are captured in the open Clamp conformation (2, 71). The remaining conformations consist of the partially closed Clamp α-helical Switch 2 (S2), the partially closed Clamp π-helical Switch 2 (S3) and closed Clamp state (S4, similar with PDB ID: 4XLN (7)) constituting 16.7%, 7.7%, and 4.2% of the total equilibrium population, respectively (Figure 1C). We note that the closed Clamp state is the minority in the RNAP holoenzyme system, which is consistent with sm-FRET experiments (9). In addition, the closed Clamp conformation is observed by crystallography and cryo-EM only when RNAP is in complex with DNA (7). This implies that the binding of DNA helps to close the Clamp. Interestingly, we identified two partially closed intermediate states (S2 and S3) which have not been captured previously in the crystal structures of the holoenzyme but play an important role for antibiotics binding as we will discuss in the next section. We found that the Clamp conformations of these two intermediate states are similar but differ by the conformations of the Switch 2 region (a helical motif positioned directly under the Clamp domain, see Figure 1A-B). In S2, the Switch 2, which resembles the open state S1, is in α-helix conformation (Figure 2B). In sharp contrast, the conformations of Switch 2 in S3, which resemble the closed state S4, is in the bulkier π-helix conformations (Figure 2B).

Interestingly, among the three switch regions (Switch 1 to 3, Figure 2A), we found that only the local refolding of Switch 2 region is strongly correlated with the global conformational changes of the Clamp closing (Figure 2C). This strong correlation is consistent with our observation from qMSM, where the refolding of the Switch 2 region from α-helix to π-helix is the rate limiting step for the Clamp to close (see Fig. 1C & 2B). Our results demonstrate the critical role of Switch 2 region in Clamp opening and closing. We also examine the interactions of Switch 2 region that regulate the Clamp domain closing. The helical structure of Switch 2 (β'610-620) in each state is stabilized by the interactions between Switch 2 and its surrounding, particularly Switch 1 (β'1431-1443) and Switch 3 (β1010-1032) (Figure S8 & Table S2). Switch 1 and Switch 3 regions are

adjacent to Switch 2 and are also located directly under the Clamp domain (Figure 2A). Especially, in S1 & S2, Switch 2 mostly interacts with Switch 1, while Switch 2 mostly interacts with Switch 3 in S3 and S4 (Figure S8 & Table S2). These observations suggest that the helical structure and interactions experienced by the Switch 2 region affect its relative position which then determines the conformation of the RNAP clamp.

Further analysis of structures in metastable states identified by qMSM revealed residues that are important for the stabilization of switch 2 helical conformation. In particular, residues βR1031, β'L619, and βQ1019 form specific interactions that stabilize the π-helix conformation in S3 and S4 (Figure S8C&D, Table S2). Therefore, we predict that replacement of these two residues with those ones with shorter sidechains (e.g. β'Q1019A) will lead to the disruption of the π-helix and negatively impact transcription initiation. Indeed, previous experimental study shows that mutants of residues βR1031 and β'L619 (βR1269H and β'L343I in *E. coli* RNAP) exhibited impairment of promoter melting at early stage (72).

We further elucidate the timescale of the Clamp opening and closing, which is shown to be in millisecond from our qMSM. By calculating the mean first passage time (MFPT) between pairs of metastable states (Table S1), we determined that the transition between the states with similar conformation of Switch 2 are fast: i.e. transitions between open (S1) and S2 or between S3 and closed (S4) (Figure 1C). On the other hand, the transitions between states with different conformations of Switch 2 are slow: i.e. transitions from S1/S2 to S3/S4 (Figure 1C & Table S1). These observations indicate that the rate limiting step of Clamp closing corresponds to the conformational change of Switch 2 region that occurs at ~2 milliseconds (i.e. local refolding from α to π-helix, Figure 2B). These observations agree with the previous hypothesis that Switch 2 region acts as the hinge under the Clamp that regulates the Clamp domain (8, 20).

In our simulations, the β-lobe is more dynamics and flexible than previously suggested by structural studies (15–17), underlining its potential role in the promoter melting not reported before. To understand the role of β-lobe during promoter melting, we further analyse the β-lobe conformations in our qMSM metastable states and observed that the opening of β-lobe is much faster than the Clamp. First, we noticed that the distributions of β-lobe opening angles are similar in all four metastable states (Figure 3B), suggesting that β-lobe's opening and closing may occur in all states. Then, we calculated the timescale of β-lobe opening and closing, which are consistently at the timescale of a hundred nanoseconds in all metastable states (Figure 3C & Figure S9B). These results indicate that β-lobe is much more flexible than the Clamp domain and can dynamically oscillate between the open and closed states at significantly faster (4 orders of magnitudes) timescale than that of the Clamp. The Clamp opening is a much slower process, as it requires substantial local refolding of the secondary structure of Switch 2 region. However, the β-lobe is connected to the rest of RNAP only by two coil chains, thus exhibiting substantially higher flexibility than the Clamp. This observation is also consistent with the crystal structures of holoenzyme, where the B-factor of β-lobe is always larger than that of the Clamp (15–17). Overall, our qMSM study reveals novel atomistic details on the dynamics of loading gate that could not be

revealed by structural studies alone. This includes the vast difference in the timescale of the opening between the Clamp and β-lobe, implicating their functional significance and the structural changes of the Switch 2 helix accounting for the opening and closing of the Clamp.

Elucidation of the Clamp Domain Dynamics Reveals the Recognition Mechanisms of Myx

Our discovery of the atomistic mechanisms of Clamp closing involving Switch 2 allows us to further explore the mechanisms by which a class of antibiotics such as Myx inhibit the initiation of transcription. We first showed that partially unfolded Switch 2 is transient only in a partially closed state (S2), and this unfolded conformation can sufficiently accommodate binding to Myx. We further analyse the Switch 2 structures in each qMSM metastable state. First, we calculate the distances between β'617-620 of Switch 2 and β'1466-1467 or β1084-1085, corresponding to D3 or D4, respectively (Figure 4A). D3 is the distance between centers of mass of C-α atoms of β'617-620 (part of Switch 2) and β1084-1085. D4 is calculated as the distance between centers of mass of C-α atoms of β'617-620 (part of Switch 2) and β'1466-1467. As shown in crystal structures, the binding pocket of Myx is formed by the unfolding of Switch 2 region and these two helices (Figure 4A) (20–22). Thus, for Myx to bind, enough space needs to be created between Switch 2 and these helices, which is shown by relatively large values D3 and D4 in the co-crystal structure of Myx bound to RNAP (+ symbol in Figure 4A&B). Our qMSM reveals that, only in S2, a fraction of the Switch 2 structures displays sufficiently large D3 and D4, corresponding to the detached and partially unfolded Switch 2 region (Figure 4B).

To further determine if the observed partially unfolded structures can indeed accommodate binding of Myx, we performed molecular docking to the transient partially unfolded (selected from qMSM) and folded structures (observed in crystal structures). Indeed, only docking to the structures from S2 results in a binding pose similar to the co-crystal structure, as indicated by large number of native contacts (orange dots in Figure 4E). Specifically, only binding to the detached (or unfolded) Switch 2 structure from S2, where both D3 and D4 are as large as the co-crystal structure (Δ symbol in Figure 4B; orange dots in Figure S11D). Evidently, the best docking pose with the largest number of native contacts is strikingly similar to the co-crystal structure (Figure 4D) and has several similar interactions with the co-crystal structure: i.e. Myx with residue β'1463, β'1467, β1084 and β'621 (Figure S11A&B). On the other hand, folded Switch 2 structure does not provide sufficient space for the binding of Myx, as Myx binds outside of the supposed binding pocket (∇ symbol in Figure 4B&C) yielding docking poses dissimilar to the co-crystal structure (brown dots in Figure S11D).

The above observations suggest a conformational selection recognition mechanism of the antibiotics. In this mechanism, Myx initially binds to the partially unfolded Switch 2 region, especially only in one of the partially closed clamp states: S2, where binding pocket can form spontaneously in RNAP. The binding of Myx will confine the clamp in a partially closed clamp conformation, and hinders it to either close or open (8, 9). Notably, there are consistent interactions between Myx and residues in RNAP in the initial bound conformation (Figure S11A&B). We

expect that upon initial binding, additional structural adjustments will further optimize these interactions to fully resemble the co-crystal structures.

Previous experimental results support the conformational selection mechanism of Myx that we proposed here. In particular, our model suggests that mutation of a phenylalanine, F614, derived from published experimental study (21), a Switch 2 region residue that is in contact with the βL1088 and β'A1438, to a less bulky alanine may reduce the hydrophobic interactions and thus increase the chance of Switch 2 unfolding. This may subsequently result in more available space in the binding pocket and thus higher affinity for the binding of Myx. Indeed, previous in-vitro IC50 measurement has validated this prediction and shown that mutation at β'F614A (F614A) causes hypersensitivity to Myx (21). To further investigate the binding of Myx to the F614A mutant, we also performed docking of Myx to the Switch 2 region with F614A mutation of structures from S2 (both folded and detached/unfolded). Docking to the unfolded and detached F614A conformations result in larger number of docking poses that are more similar to the co-crystal structure (large number of native contacts of light blue dots, Figure S11D). This implies that F614A does allow for more favourable binding with Myx by providing more space in the binding pocket. Interestingly, docking to the folded F614A conformations does not result in shift of fraction of native contact compared with the wild type (dark blue dots Figure S11D). Thus, we anticipate that F614A only affects the partially unfolded structures and unfolding is required to cause hypersensitivity of F614A. This result further supports the conformational selection mechanism. Furthermore, it has been shown that Myx cannot inhibit transcription initiation if the transcription bubble has already been formed in RPo complex (20, 21). In this complex, the Clamp and Switch 2 are locked in closed conformation. Thus, the conformational selection mechanism of Myx recognition can explain this experimental result, since Myx binding requires the transition of the Clamp domain to the partially closed conformation.

The Role of β-lobe on Promoter DNA Loading

We next inspected the loading gate (i.e. the space between β-lobe and Clamp) conformations that can accommodate dsDNA loading. To achieve this, we modelled dsDNA in the loading gate conformations with different degrees of opening (Figure 5A&B). We found that even when the Clamp is partially closed, opening of the β-lobe is sufficient to accommodate dsDNA. To select representative conformations, we grouped the Clamp and β-lobe conformations from our MD simulations into 3 and 4 states, respectively according to structural similarities based on the Principal Component Analysis (see Methods and SI Sec. 7 for details). As a result, the Clamp conformations are clustered to open, partially closed, and closed conformations (see Figure S12 and SI Sec. ?? for details), while the β-lobe conformations are grouped into the open, partially closed, partially closed shifted, and closed conformations (see Figure S13 and SI Sec. ?? for details). Combining these two sets of conformations results in 12 combinations of Clamp and β-lobe conformations (Figure 5B). We adopted the distance between the Clamp and the β-lobe to describe the openness of the loading gate (D1 in Figure 5C). We then examined if the loading gate in these 12 conformations has sufficient space to allow the loading of dsDNA. Interestingly, we found that only three combinations of Clamp and β-lobe conformations can allow the loading of

Commented [XH1]: Ilona: please add the choice of the # of clusters in Clamp and Beta-lobe to SI. They are too technique.

Commented [XH2]: Please check and update these two sentences if necessary

the dsDNA: open Clamp and open β-lobe, open Clamp and partially closed β-lobe, partially closed Clamp and open β-lobe (Figure 5C-E). In contrast, dsDNA will experience steric clashes when it passes through the loading gate in all other combinations of Clamp and β-lobe conformations (Figure S14 & S15). At the beginning of the promoter melting, the Clamp has to be closed or at least partially closed (3, 5–7). Then, dsDNA of the downstream promoter must pass through the loading gate representing the space between Clamp and β-lobe to enter the inner cleft of RNAP. Based on our results, either Clamp or β-lobe needs to open to provide enough space for the loading of the promoter dsDNA. If the Clamp is already in partially closed state at the beginning of the promoter melting, the opening of β-lobe will be sufficient to facilitate dsDNA loading geometrically.

Our proposed mechanism is supported by the cryoEM of *Mtb* holoenzyme with promoter dsDNA already loaded to the inner cleft and corallopyronin (Cor) bound (5). Since Cor confines the Clamp domain in partially closed conformation, this cryoEM structure suggests that Clamp opening may not be required during the initial loading of dsDNA. We note that our proposed mechanism only explains the initial loading of dsDNA during promoter melting, where the transcription bubble has not been fully formed. In the later stage of the promoter melting, the loaded dsDNA will be separated into single-stranded, and this step may require clamp opening. Furthermore, we also note that the structure of the promoter dsDNA may be more complicated than the DNA model used here, where the upstream promoter DNA (position -100) was found to wrap around the *E. coli* RNAP during promoter melting (73, 74). In this situation, opening of the Clamp was required even at the beginning of promoter melting. Thus, as mentioned, different species of bacteria may follow different promoter melting pathways and have different free energy landscapes of the promoter melting. Nonetheless, the β-lobe is always more flexible than the Clamp across different bacterial species as observed by the B-factors in crystal and cryoEM structures (15–17), suggesting that opening of β-lobe is always easier than the Clamp. Therefore, it is important to consider β-lobe during the promoter melting. Importantly, these findings highlight the importance of β-lobe during transcription initiation, which open the opportunities to the development of antibiotics that target β-lobe. Interestingly, a recent Cryo-EM study suggests that an antibiotic, Sorangicin, may inhibit the bacterial transcription via affecting the movement of β-lobe through allosteric interactions (75).

Conclusions

In this study, we have applied qMSM built from extensive all-atom MD simulations of bacterial RNAP to elucidate the dynamics of DNA loading gate opening and closing. Using qMSM, we have determined that the timescale of the β-lobe opening is orders of magnitude faster than the Clamp, suggesting the significant role of β-lobe during the DNA loading owing to its much higher mobility for gate opening or closing. Moreover, we discovered that the opening of the β-lobe is geometrically sufficient to allow initial loading of promoter dsDNA during the promoter melting, even when the Clamp is in the partially closed conformation. Furthermore, we have shown that the Clamp domain opening is mostly regulated by local refolding of Switch 2 region from an α-helix to a π-helix. We further show that the spontaneous refolding and detachment of Switch 2 occurring in a partially closed Clamp state provide enough space to allow binding of antibiotics

Myx via a conformational selection mechanism. Considering that the Clamp, β-lobe, and Switch regions are conserved features in all prokaryotic species, we expect that the findings in this work could be generalized to other species of bacteria. Our qMSM, which adopts the memory kernels based on the generalized master equation, provides a promising approach in generating models containing a handful of states that can greatly facilitate the interpretation of the biological mechanisms. Overall, our work yields new insights into the loading gate dynamics and highlight the undiscovered role of β-lobe, which provides a potential novel target for future development of antibiotics.

Methods

All-atom MD simulation setup for RNAP structures with Open and Closed Clamp

The open Clamp conformation of the holoenzyme (RNAP and σ^A factor) was obtained from the crystal structure of *Thermus aquaticus* RNAP holoenzyme in complex with downstream-fork promoter (PDB ID: 5TJG (6)). The closed Clamp structure was modelled from crystal structures of *Taq* RNAP holoenzyme bound to the complete transcription bubble (PDB ID: 4XLN (7)). The DNA chains are removed from the crystal structure in all the following MD simulations. For both crystal structures, residues 162-452 of β' domain were replaced by the GAG sequence. The part replaced is not conserved and located far from the inner surface of the Clamp domain. The modelled structures were solvated in a dodecahedron simulation box, and additional Na^+ and Cl^- ions were added to neutralize the system and reach the 0.15 M salt concentration. The total number of atoms in the simulation box is consistently set to be 543,237 for all all-atom MD simulations initiated from different structures. GROMACS 4.5 simulation package (76) and AMBER 99SB-ILDN force field (77) are used for all-atom MD simulations. Please refer to SI Sec. 1: "All-atom MD simulation set-up" for more details of the system set-up and all-atom MD simulations.

Microstate-MSM construction and validation

We followed our previously published protocol (65) to construct microstate-MSM to investigate the opening/closing dynamics of the RNAP loading gate as described below:

(a) Initial pathways connecting the open and closed RNAP clamp conformations: We obtained initial pathways using a coarse-grained (AICG2+) model (66, 78), in which each bead represents an amino acid residue and the two-basin Go-potential was applied to drive the conformational change between two desired structures of interest. As shown in Figure S1A, we performed 5 CG-MD simulations using the CafeMol software (79) for each direction of transition, i.e. closed to open and open to closed, of RNAP clamp conformation (see SI Sec. 2.1 for details of CG-MD simulations). We chose 86 CG-MD conformations (Figure S1B) to further reconstruct them to all-atom conformations using the back-mapping algorithm based on restrained MD simulations (80). Please refer to Figure S2 and SI Sec. 2.2 for more details of the back-mapping procedure.

(b). All-atom MD sampling along the initial pathways: We performed extensive sampling that results in a MD dataset consisting of 306 200-ns MD simulations initiated from conformations along the initial pathways as well as from open and closed RNAP clamp structure (~61.2μs MD simulations in total, see SI Sec. 2.2. for more details).

(c). tICA (67–69) for dimensionality reduction: We applied the Force Distribution Analysis (FDA) (81) to identify a set of pairwise atomic distances that correlate with the RNAP clamp motion: i.e. those experience the most different force distributions between the open and closed clamp (see SI Sec. 2.3 for details of FDA analysis). We further merged these atom-pair distances with those between RNAP motifs that have been previously suggested to be correlated with the clamp motion (20, 82, 83). As a result, we obtained a list of 1,770 pairwise distances (involving C-α atoms from residues in the Clamp, β-lobe, β-protrusion, Switch regions, and active site, see Figure S4) to input to the tICA analysis. We have further validated our selection of atom-pair distances by showing that it outperforms a series of alternative atom-pair distances by systematically removing certain atom pairs from the list (Figure S3B). The hyper-parameters for the tICA analysis (i.e. 3 tICs and tICA relaxation time of 10 ns) and clustering (i.e. 100 microstates via k-center clustering (70)) were obtained by the Generalized Matrix Rayleigh Quotient (GMRQ (84)) method (see Figure S5A-C and SI Sec. 2.4 for details).

(d) We validated the 100-microstate MSM by the Chapman-Kolmogorov test, where we showed that the residence probabilities from individual microstate predicated by the microstate-MSM agrees with those directly computed from all-atom MD simulations (see Figure S5E and SI Sec. 2.5).

Macrostate *quasi*-MSM construction and validation

To facilitate the interpretation of biological mechanisms, we applied our recently developed qMSM (64) algorithm to construct a model containing only 4 states (S1 to S4). To construct qMSM, we first performed the kinetic lumping to group 100 microstates into 4 macrostates using the PCCA+ (85, 86) algorithm. We chose 4 macrostates because a stable gap is observed between 3rd and 4th slowest implied timescale (Figure S5D). Our qMSM applies the generalized master equation formalism to encode the non-Markovian dynamics in time-dependent memory kernels ($\boldsymbol{K}(\tau)$):

$$\dot{\boldsymbol{T}}(t) = \boldsymbol{T}(t)\dot{\boldsymbol{T}}(0) + \int_0^{\min\{t,\tau_K\}} \boldsymbol{T}(t-\tau)\boldsymbol{K}(\tau)d\tau, \qquad (1)$$

where τ_K refers to the memory kernel relaxation time (or characteristic decay time): $\boldsymbol{K}(t > \tau_K) \approx \boldsymbol{0}$, and $\boldsymbol{T}(t)$ is the transition probability matrix (TPM). To numerically compute $\boldsymbol{K}(t)$ at $t = n\Delta t$ (Δt=1ns), we utilized a discretized form of the generalized master equation (see Eq. (S1)-(S2)) (64). We chose $\tau_K = 30$ns because the mean integration of memory kernel (MIK) converges at this lag time (see Eq. (S5) and Figure S6A). Please refer to SI Sec. 3 for additional details of qMSM construction.

We validated our qMSM using the Chapman-Kolmogorov test by showing that the qMSM with $\tau_K = 30$ns can accurately predict residence probabilities for all four states (Figure S6C). In sharp contrast, we showed that the 4-macrostate MSM without considering the memory kernels is not Markovian even with a lag time up to ~100ns (Figure S6B). We further showed that these first-order MSMs predict significantly faster dynamics than those of the MD simulations in the Chapman-Kolmogorov test (Figure S6C-S6D), and also predicted substantially shorter MFPTs

than qMSM (Figure S7B and S7D). To compute MFPTs and to identify major pathways for RNAP clamp closing by the transition path theory (87, 88) , we utilized the TPM: $T(\tau = 5\mu s)$ obtained from the discretized generalized master equation in our qMSM (see SI Sec. 4 for additional details). We note that the inclusion of memory functions does not increase the number of states in qMSMs. Instead, rather than using the TPM as in MSMs, qMSMs model the dynamics using the transition tensors (64). As a result, qMSMs has substantial advantages over MSMs in interpreting biological mechanisms by yielding models with a handful of states.

Molecular Docking of Myxopyronin to RNAP

To determine which conformations of Switch 2 can accommodate the binding of Myx, we performed molecular docking to several representative structures from each metastable state. For the structures from S2, we separated the unfolded and detached switch 2 conformations from the folded ones. To check the similarity of the docking poses to the co-crystal structures of *Taq* holoenzyme and Myx, we defined and calculated the number of native contacts for each docking pose. The number of native contacts is the number of interactions between *Taq* holoenzyme and Myx in a docking pose which is also found in the co-crystal structure (PDB ID: 3DXJ (20)) (see SI Sec. 6 for further details of molecular docking).

Selection of representative loading gate conformations and modeling of dsDNA

To determine which conformation of Clamp and β-lobe geometrically allow dsDNA loading, we selected representative structures from our MD simulations based on the degree of opening of the Clamp and β-lobe conformations. First, we performed Principal Component Analysis (PCA) (89) of Clamp and β-lobe separately, to characterize the Clamp and β-lobe opening motions (see SI Sec. 7 for details of PCA). Based on the PCA results, we performed independent k-means clustering to group Clamp and β-lobe conformations to 3 and 4 clusters, respectively (Figure 5 A&B). By merging the results from the two sets of clustering, we obtained 12 clusters of different combinations of Clamp and β-lobe conformations (Figure 5C). For each cluster, we selected the structure closest to the cluster center and modelled dsDNA in the loading gate using The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC (Figure S11&S12). The steric clash for each models were checked by MDWeb webserver (90).

Acknowledgement

X.H. was supported by the Hong Kong Research Grant Council (16303919, 16307718, AoE/P-705/16, AoE/M-09/12, and T13-605/18-W), and Hong Kong Innovation and Technology Commission (ITCPD/17-9 and ITC-CNERC14SC01). X.G. was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. FCC/1/1976-23, FCC/1/1976-26, URF/1/4098-01-01, URF/1/4352-01-01, REI/1/4473-01-01 and REI/1/0018-01-01. This research made use of the computing resources of the Supercomputing Laboratory at King Abdullah University of Science and Technology and the X-GPU cluster supported by the Hong Kong Research Grant Council Collaborative Research Fund: C6021-19EF.

References

1. Ruff EF, Thomas Record M, Artsimovitch I (2015) Initial events in bacterial transcription initiation. *Biomolecules* 5(2):1035–1062.
2. Boyaci H, et al. (2018) Fidaxomicin jams mycobacterium tuberculosis RNA polymerase motions needed for initiation via RBPA contacts. *Elife* 7(e34823):1–19.
3. Chen J, et al. (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol Cell* 78(2):275-288.e6.
4. Danson AE, Jovanovic M, Buck M, Zhang X (2019) Mechanisms of σ54-Dependent Transcription Initiation and Regulation. *J Mol Biol* 431(20):3960–3974.
5. Boyaci H, Chen J, Jansen R, Darst SA, Campbell EA (2019) Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* 2:1.
6. Feklistov A, et al. (2017) RNA polymerase motions during promoter melting. *Science (80-)* 356(6340):863–866.
7. Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* 4(e08504):1–23.
8. Chakraborty A, et al. (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science (80-)* 337(6094):591–595.
9. Duchi D, Mazumder A, Malinen AM, Ebright RH, Kapanidis AN (2018) The RNA polymerase clamp interconverts dynamically among three states and is stabilized in a partly closed state by ppGpp. *Nucleic Acids Res* 46(14):7284–7295.
10. Duchi D, et al. (2018) Conformational heterogeneity and bubble dynamics in single bacterial transcription initiation complexes. *Nucleic Acids Res* 46(2):677–688.
11. Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9(2):85–98.
12. Mazumder A, Kapanidis AN (2019) Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. *J Mol Biol* 431(20):3947–3959.
13. Feklistov A, Darst SA (2011) Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* 147(6):1257–1269.
14. Hubin EA, et al. (2017) Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife* 6. doi:10.7554/eLife.22520.
15. Murakami KS (2013) X-ray crystal structure of Escherichia coli RNA polymerase σ70 holoenzyme. *J Biol Chem* 288(13):9126–34.
16. Chen J, et al. (2019) E. Coli trar allosterically regulates transcription initiation by altering RNA polymerase conformation. *Elife* 8:1–29.
17. Vassylyev DG, et al. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* 417(6890):712–719.
18. Murakami KS, Masuda S, Darst SA (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science (80-)* 296(5571):1280–4.
19. Kouba T, et al. (2019) The Core and Holoenzyme Forms of RNA Polymerase from Mycobacterium smegmatis. *J Bacteriol* 201(4):1–12.
20. Mukhopadhyay J, et al. (2008) The RNA Polymerase "Switch Region" Is a Target for Inhibitors. *Cell* 135(2):295–307.
21. Belogurov G a, et al. (2009) Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* 457(7227):332–335.
22. Molodtsov V, et al. (2015) X-ray Crystal Structures of Escherichia coli RNA Polymerase with Switch Region Binding Inhibitors Enable Rational Design of Squaramides with an

Improved Fraction Unbound to Human Plasma Protein. *J Med Chem* 58(7):3156–3171.
23. Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. *Protein Eng* 12(9):713–720.
24. Kumar S, Ma B, Tsai C-J, Sinha N, Nussinov R (2008) Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci* 9(1):10–19.
25. Tsai C-J, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. *Protein Sci* 8(6):1181–1190.
26. Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci* 44(2):98–104.
27. Ma B, Shatsky M, Wolfson HJ, Nussinov R (2009) Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein Sci* 11(2):184–197.
28. Arora K, Brooks CL (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci U S A* 104(47):18496–18501.
29. Bui JM, McCammon JA (2006) Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism. *Proc Natl Acad Sci U S A* 103(42):15451–15456.
30. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* 17(6):633–640.
31. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci U S A* 96(18):9970–2.
32. Prinz JH, et al. (2011) Markov models of molecular kinetics: Generation and validation. *J Chem Phys* 134(17):174105.
33. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* 126(155101):1–17.
34. Pan AC, Roux B (2008) Building Markov state models along pathways to determine free energies and rates of transitions. *J Chem Phys* 129(064107):1–8.
35. Zhang BW, et al. (2016) Simulating Replica Exchange: Markov State Models, Proposal Schemes, and the Infinite Swapping Limit. *J Phys Chem B* 120(33):8289–8301.
36. Morcos F, et al. (2010) Modeling conformational ensembles of slow functional motions in pin1-WW. *PLoS Comput Biol* 6(12):1–13.
37. Huang X, Bowman GR, Bacallado S, Pande VS (2009) Rapid equilibrium sampling initiated from nonequilibrium data. *Proc Natl Acad Sci U S A* 106(47):19765–19769.
38. Bowman GR, Huang X, Pande VS (2009) Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* 49(2):197–201.
39. Malmstrom RD, Lee CT, Van Wart AT, Amaro RE (2014) Application of molecular-dynamics based markov state models to functional proteins. *J Chem Theory Comput* 10(7):2648–2657.
40. Buchete NV, Hummer G (2008) Coarse master equations for peptide folding dynamics. *J Phys Chem B* 112(19):6057–6069.
41. Chodera JD, Noé F (2014) Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol* 25:135–144.
42. Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci* 106(45):19011–19016.

43. Da LT, Wang D, Huang X (2012) Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J Am Chem Soc* 134(4):2399–2406.
44. Silva D-A, et al. (2014) Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc Natl Acad Sci U S A* 111(21):7665–70.
45. Plattner N, Doerr S, De Fabritiis G, Noé F (2017) Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat Chem* 9(10):1005–1011.
46. Vanatta DK, Shukla D, Lawrenz M, Pande VS (2015) A network of molecular switches controls the activation of the two-component response regulator NtrC. *Nat Commun* 6:1–7.
47. Silva DA, Bowman GR, Sosa-Peinado A, Huang X (2011) A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comput Biol* 7(5). doi:10.1371/journal.pcbi.1002054.
48. Malmstrom RD, Kornev AP, Taylor SS, Amaro RE (2015) Allostery through the computational microscope: CAMP activation of a canonical signalling domain. *Nat Commun* 6(May). doi:10.1038/ncomms8588.
49. Lawrenz M, Shukla D, Pande VS (2015) Cloud computing approaches for prediction of ligand binding poses and pathways. *Sci Rep* 5(7918). doi:10.1038/srep07918.
50. Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci U S A* 108(25):10184–10189.
51. Beauchamp KA, McGibbon R, Lin YS, Pande VS (2012) Simple few-state models reveal hidden complexity in protein folding. *Proc Natl Acad Sci U S A* 109(44):17807–17813.
52. Lane TJ, Shukla D, Beauchamp KA, Pande VS (2013) To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr Opin Struct Biol* 23(1):58–65.
53. Bowman GR, Voelz VA, Pande VS (2011) Taming the complexity of protein folding. *Curr Opin Struct Biol* 21(1):4–11.
54. Kohlhoff KJ, et al. (2014) Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat Chem* 6(1):15–21.
55. Da L-T, et al. (2016) Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat Commun* 7:11244.
56. Jiang H, et al. (2015) Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding followed by Structural Re-arrangement. *PLoS Comput Biol* 11(7).
57. Da LT, Pardo Avila F, Wang D, Huang X (2013) A Two-State Model for the Dynamics of the Pyrophosphate Ion Release in Bacterial RNA Polymerase. *PLoS Comput Biol* 9(4). doi:10.1371/journal.pcbi.1003020.
58. Da LT, et al. (2015) A Jump-from-Cavity Pyrophosphate Ion Release Assisted by a Key Lysine Residue in T7 RNA Polymerase Transcription Elongation. *PLoS Comput Biol* 11(11):1–22.
59. Zhang L, et al. (2016) Elucidation of the Dynamics of Transcription Elongation by RNA Polymerase II using Kinetic Network Models. *Acc Chem Res* 49:687–694.
60. Unarta IC, et al. (2018) Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Curr Opin Struct Biol* 49:54–62.
61. Wang W, Cao S, Zhu L, Huang X (2017) Constructing Markov State Models to elucidate

the functional conformational changes of complex biomolecules. *Wiley Interdiscip Rev Comput Mol Sci* e1343. doi:10.1002/wcms.1343.
62. Da LT, Wang D, Huang X (2012) Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J Am Chem Soc* 134(4):2399–2406.
63. Wang B, Sexton RE, Feig M (2017) Kinetics of nucleotide entry into RNA polymerase active site provides mechanism for efficiency and fidelity. *Biochim Biophys Acta - Gene Regul Mech* 1860(4):482–490.
64. Cao S, Montoya-Castillo A, Wang W, Markland TE, Huang X (2020) On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *J Chem Phys* 153(014105):1–13.
65. Wang W, Cao S, Zhu L, Huang X (2018) Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdiscip Rev Comput Mol Sci* 8(1). doi:10.1002/wcms.1343.
66. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* 103(32):11844–9.
67. Schwantes CR, Pande VS (2013) Improvements in Markov State Model construction reveal many non-native interactions in the folding of NTL9. *J Chem Theory Comput* 9(4):2000–2009.
68. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F (2013) Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139(1):015102.
69. Naritomi Y, Fuchigami S (2013) Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J Chem Phys* 139(21):215102.
70. Hochbaum DS, Shmoys DB (1985) A Best Possible Heuristic for the k-Center Problem. *Math Oper Res* 10(2):180–184.
71. Artsimovitch I, et al. (2005) Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell* 122(3):351–363.
72. Rutherford ST, Villers CL, Lee JH, Ross W, Gourse RL (2009) Allosteric control of escherichia coli rRNA promoter complexes by DksA. *Genes Dev* 23(2):236–248.
73. Sreenivasan R, et al. (2020) Fluorescence-Detected Conformational Changes in Duplex DNA in Open Complex Formation by E. coli RNA Polymerase: Upstream Wrapping and Downstream Bending Precede Clamp Opening and Insertion of the Downstream Duplex . *Biochemistry*. doi:10.1021/acs.biochem.0c00098.
74. Gries TJ, Kontur WS, Capp MW, Saecker RM, Record MT (2010) One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proc Natl Acad Sci U S A* 107(23):10418–10423.
75. Lilic M, et al. (2020) The antibiotic sorangicin A inhibits promoter DNA unwinding in a Mycobacterium tuberculosis rifampicin-resistant RNA polymerase. *Proc Natl Acad Sci* 117(48):30423–30432.
76. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.
77. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber

ff99SB protein force field. *Proteins Struct Funct Bioinforma* 78(8):1950–1958.
78. Takada S, et al. (2015) Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc Chem Res* 48(12):3026–3035.
79. Kenzaki H, et al. (2011) CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J Chem Theory Comput* 7(6):1979–1989.
80. Peng J, Yuan C, Ma R, Zhang Z (2019) Backmapping from Multiresolution Coarse-Grained Models to Atomic Structures of Large Biomolecules by Restrained Molecular Dynamics Simulations Using Bayesian Inference. *J Chem Theory Comput* 15(5):3344–3353.
81. Stacklies W, Seifert C, Graeter F (2011) Implementation of force distribution analysis for molecular dynamics simulations. doi:10.1186/1471-2105-12-101.
82. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292(5523):1876–82.
83. Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. *Science (80-)* 292(5523):1863–1876.
84. McGibbon RT, Pande VS (2015) Variational cross-validation of slow dynamical modes in molecular kinetics. *J Chem Phys* 142(12). doi:10.1063/1.4916292.
85. Deuflhard P, Weber M (2005) Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl* 398(1–3):161–184.
86. Röblitz S, Weber M (2013) Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification. *Adv Data Anal Classif* 7(2):147–179.
87. Vanden-Eijnden E (2006) Towards a Theory of Transition Paths. *J Stat Phys* 123(3). doi:10.1007/s10955-005-9003-9.
88. Metzner P, Schütte C, Vanden-Eijnden E (2009) Transition Path Theory for Markov Jump Processes. *Multiscale Model Simul* 7(3):1192–1219.
89. Abdi H, Williams LJ (2010) Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2(4):433–459.
90. Hospital A, et al. (2012) MDWeb and MDMoby: An integrated web-based platform for molecular dynamics simulations. *Bioinformatics* 28(9):1278–1279.

Figures:

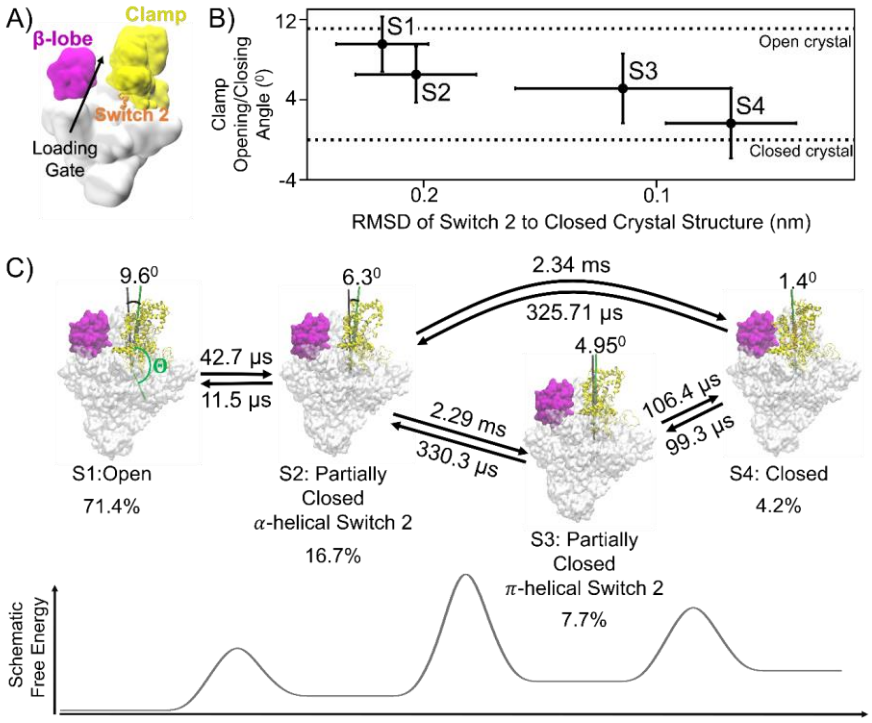

Figure 1. qMSM identifies four metastable states corresponding to the opening and closing of RNAP clamp domain. A) The overall cartoon representation of bacterial holoenzyme and its domains, Clamp, β-lobe, and Switch 2 region are shown as yellow, magenta, and orange, respectively. B) Four macro states of the Clamp domain of RNAP are identified by MSM. The metastable states are represented as Clamp opening/closing angles and RMSD of Switch 2 region (β'611-620) aligned to the Switch 2 region of the closed Clamp in the crystal structure. The Clamp opening/closing angle (Θ) is defined as the torsion angle between four centers of mass of C-α of residues i) ω 60-68, ii) β'1461-1468, iii) β1031-1033 & β'620-622, iv) β'568-573 and is offset by the angle of closed Clamp crystal structure ($\Theta - \Theta_{\text{closed}}$). C) The transition time between metastable states and the stationary population of each state is shown. The transitions between S1 and S2, as well as S3 and S4, are fast. The transitions between S1 or S2 to either S3 or S4 are slow, as indicated by the higher free energy barrier. The full list of transition time (MFPT) is shown in Table S1.

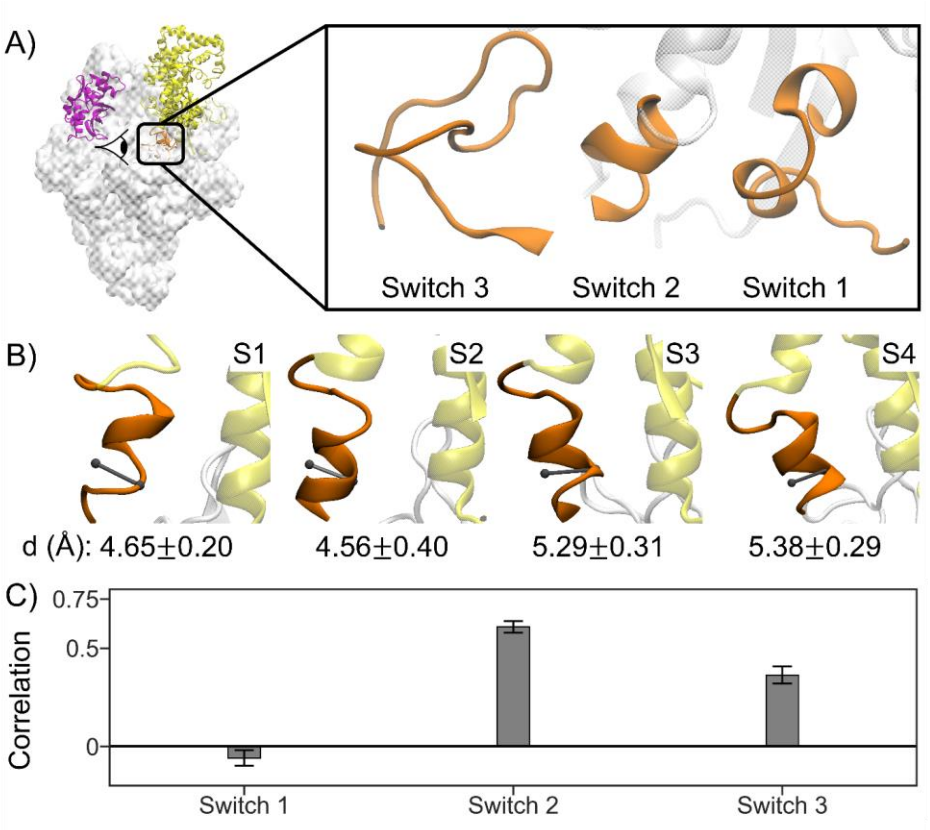

Figure 2. Switch 2 region under the clamp acts as a hinge of the Clamp domain opening and closing. A) Switch regions are located under the Clamp domain. There are three switch regions, Switch 1, Switch 2, and Switch 3, which are adjacent to each other. The viewing angle of the switch regions is shown by eye symbol in the left figure. B) Diameter of the α-helix of the Switch 2 region in each metastable state is shown. The viewing direction of these figures is perpendicular to the viewing direction of A) and into the protein core. The diameter is the distance between center of mass of C-α atoms of residue β'615 & 619 and C-α atom of residue β'617. Small diameter of helix at ~4.5Å indicates α-helical conformation, while large diameter of helix at ~5.2Å indicates π-helical conformation. C) Pearson's correlation coefficient of the Clamp opening/closing angle with the RMSD of Switch 1, 2, 3 regions aligned to the C-α atoms of the individual Switch regions in the closed Clamp crystal structure. For the calculation of RMSD of switch regions, the C-α atoms of the following residues are included: i) β'1431-1443 for Switch 1, ii) β'610-620 for

Switch 2, iii) β1010-1032 for Switch 3. Based on correlation analysis, only conformational change of Switch 2 is strongly correlated with the Clamp opening/closing motion.

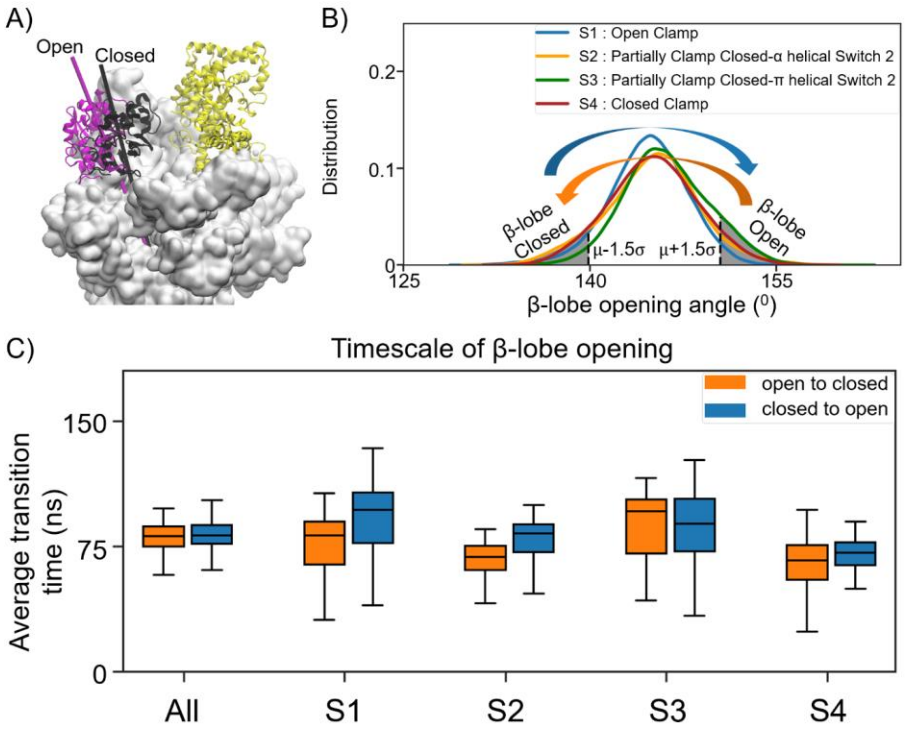

Figure 3. β-lobe is a highly dynamic domain compared to Clamp. A) The open and closed conformations of β-lobe are shown as magenta and black, respectively. The β-lobe opening angle is defined as the dihedral angle formed by the center of mass of four groups of residues represented by the C-α atoms: i) β'877-889, ii) β'938-943, iii) β'786-793, iii) β142-325. B) Distribution of β-lobe opening angle in each metastable state is shown. The open β-lobe state consists of structures having β-lobe opening angle larger than μ+1.5σ, where μ and σ is the average and standard deviation of all MD trajectories. The closed β-lobe state consists of the structures having β-lobe opening angle smaller than μ-1.5σ. C) Average transition time values of β-lobe opening in all the dataset and each metastable state. The opening of the β-lobe is two to three orders of magnitude faster than the Clamp opening.

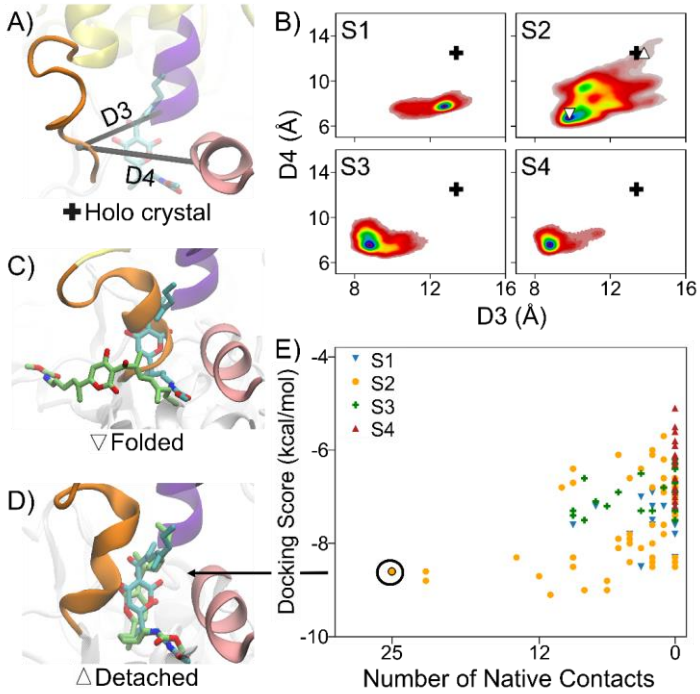

Figure 4. Spontaneous detachment or unfolding of Switch 2 region only occurs in a partially closed state, S2, which allows binding of Myx. A) The structure of Switch 2 region when bound to Myx is shown. Switch 2, helix 1084-1090, helix β'1463-1467 are shown in orange, purple, and pink cartoon. Myx is shown as cyan sticks representation. D3 or D4 are the distance between center of mass represented by C-α atoms of β'617-620 (part of Switch 2) and β1084-1085 (D3) or β'1466-1467 (D4), respectively. B) The free energy landscape represented as D3 and D4 are shown for each metastable state. The plus sign (+) represents the co-crystal structure of RNAP bound with Myx (PDB ID:3DXJ). The detached and folded Switch 2 structure in S2 is shown as (Δ) and (∇). C) The docking pose of Myx to a folded (Δ) Switch 2 is shown as lime sticks representations. D) The docking pose of Myx to a detached (Δ) Switch 2 is shown as lime sticks representation, which also corresponds to the best docking pose (largest number of native contacts). In Figure A), C), and D), In both C) and D), Myx in the co-crystal structure (PDB ID: 3DXJ) is overlaid and shown as cyan sticks representation. E) Docking results of Myx to RNAP conformations in each metastable state are plotted as docking score and number of native contacts.

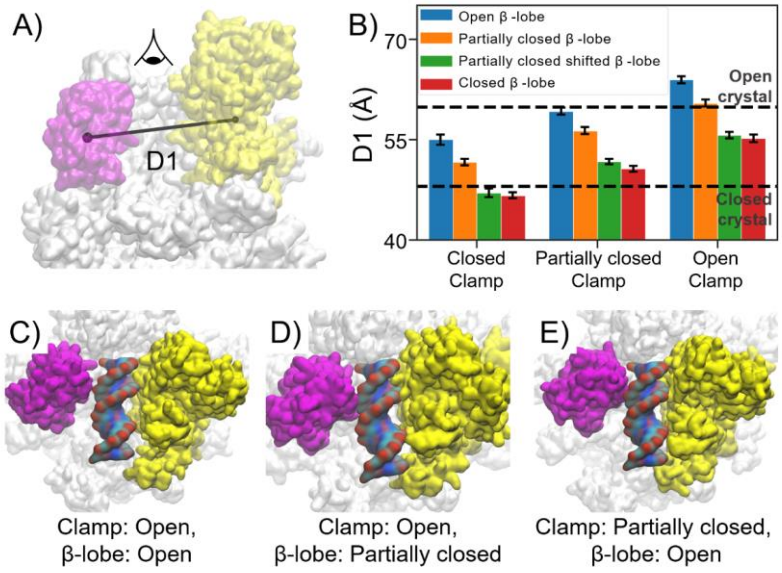

Figure 5. Opening of β-lobe is sufficient for DNA loading even when the Clamp is partially closed. A) D1 is the distance between center of mass of Clamp and β-lobe. The atoms included for the calculation of center of mass is the same as the ones included for PCA. B) Merging the clustering in A and B results in 12 combinations of Clamp and β-lobe conformations. The D1 values for the structure closest to the cluster centers from each combination are shown as bar graph. There are three combinations of Clamp and β-lobe conformation that can accommodate dsDNA, which are C) open and open, D) open and partially closed, E) partially closed and open, respectively.