Biophysical *Journal*
Article

Biophysical Society

Switching promotor recognition of phage RNA polymerase in silico along lab-directed evolution path

Chao E,[1,4] Liqiang Dai,[1,2,4] and Jin Yu[3,*]
[1]Beijing Computational Science Research Center, Beijing, China; [2]Shenzhen JL Computational Science and Applied Research Institute, Shenzhen, Guangdong, China; and [3]Department of Physics and Astronomy, Department of Chemistry, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, Irvine, California

ABSTRACT In this work, we computationally investigated how a viral RNA polymerase (RNAP) from bacteriophage T7 evolves into RNAP variants under lab-directed evolution to switch recognition from T7 promoter to T3 promoter in transcription initiation. We first constructed a closed initiation complex for the wild-type T7 RNAP and then for six mutant RNAPs discovered from phage-assisted continuous evolution experiments. All-atom molecular dynamics simulations up to 1 μs each were conducted on these RNAPs in a complex with the T7 and T3 promoters. Our simulations show notably that protein-DNA electrostatic interactions or stabilities at the RNAP-DNA promoter interface well dictate the promoter recognition preference of the RNAP and variants. Key residues and structural elements that contribute significantly to switching the promoter recognition were identified. Followed by a first point mutation N748D on the specificity loop to slightly disengage the RNAP from the promoter to hinder the original recognition, we found an auxiliary helix (206–225) that takes over switching the promoter recognition upon further mutations (E222K and E207K) by forming additional charge interactions with the promoter DNA and reorientating differently on the T7 and T3 promoters. Further mutations on the AT-rich loop and the specificity loop can fully switch the RNAP-promoter recognition to the T3 promoter. Overall, our studies reveal energetics and structural dynamics details along an exemplary directed evolutionary path of the phage RNAP variants for a rewired promoter recognition function. The findings demonstrate underlying physical mechanisms and are expected to assist knowledge and data learning or rational redesign of the protein enzyme structure function.

SIGNIFICANCE Lab-directed evolution switches protein enzyme functions without revealing physical mechanisms. Our computational work here shows physically how a viral RNA polymerase and its mutants switch promoter recognition function from one promoter to another along a representative lab-directed evolution path. We found that protein-DNA electrostatic interactions well characterize the RNAP-promoter bias to initiate transcription. We also identified key residues and structural elements (e.g., an auxiliary helix) at the RNAP-promoter interface that dictate the promoter bias and switch. Our work thus reveals detailed mechanisms on the RNAP-promoter recognition as well as provides physical basis for further experimental data learning and rational design.

INTRODUCTION

Lab-directed evolution technologies in recent years have made substantial advancements in functional design or redesign of biomolecular systems, in particular on protein enzyme activities and specificities (1–5). In the lab-directed evolution, sequence mutations and recombination are intensively promoted and followed by high-throughput screening or selection to target on specific protein functions (6–9). For example, in phage-assisted continuous evolution (PACE), bacteriophage with modified life cycle is employed to transfer evolving genes between bacterial host cells to promote fast-replicating phage populations that contain gene mutations toward certain favored enzyme activities (10,11). With technology advancements, not only individual protein enzymes with certain functions can be designed, but also pathways or protein interaction networks can be modulated or rewired (12–17). Meanwhile, rational design of protein functions based on molecular structures and biochemical

Submitted August 6, 2021, and accepted for publication January 10, 2022.
[4]These authors contributed equally
*Correspondence: jin.yu@uci.edu
Editor: Mark Williams.
https://doi.org/10.1016/j.bpj.2022.01.007

properties of the protein have always been pursued (18–22), which usually demand physical understanding and computational exploration on optimal solutions in the high-dimensional space of protein sequence or conformation evolution. Since biomolecules or enzymes are intrinsically complex systems evolved with complicated structure-function relations, straightforward physical or rational approaches can be highly challenging. The lab-directed evolution studies, however, provide abundant data on designed and redesigned on-path leading to end products with desired functions, which can be particularly interesting to learn and to infer the underlying structure-function relation to further support physically based rational approaches. In this work, we use in silico methods, i.e., molecular modeling and all-atom molecular dynamics (MD) simulation to study promoter recognition of viral RNA polymerase (RNAP) variants discovered from lab-directed evolution. In particular, we take advantage of the PACE achievements on switching a bacteriophage RNAP from recognizing its original promoter to another one in a similar phage system (9,23).

The recognition and binding of RNAP to the promoter occur at the initial stage of gene transcription, which essentially determines the promoter activity or productivity of followed gene expression. In eukaryote cells, the transcription initiation is highly regulated, conducted by a multi-subunit RNAP in coordination with a large number of transcription factors (24,25). In contrast, single-subunit viral RNAP from bacteriophage T7, which is constantly utilized in the lab gene expression system, is able to complete transcription from initiation to termination in the absence of additional factors (26,27). Such a viral RNAP system is accordingly ideal for studying elementary key transcription functions. In particular, T7 RNAP and its closely related single-subunit viral RNAPs from other bacteriophages (e.g., T3, SP6, and K11) demonstrate high specificities in their individual promoter activities (28–33). Meanwhile, mutant phage RNAPs have also been identified to show modulated promoter specificities, e.g., certain mutations of T7 RNAP lead to switching of its specificity from T7 promotor to T3 or SP6 promoter (29,32,34,35). Hence, study of specific promotor recognition in phage RNAP transcription initiation, in particular on how the wild-type (or wt) RNAP and its mutant (or mt) RNAPs or variants change their promoter specificities from the original DNA promotor to the promoter of alternated DNA sequences, would be of high interest to reveal physical mechanisms underlying specific protein-DNA sequence recognition.

To study the phage RNAP promoter recognition using in silico approach, high-resolution atomic structures of the corresponding systems are needed. The high-resolution crystal structure of the T7 RNAP-DNA-promoter-binding complex has been resolved (36,37), with the RNAP in association with an incomplete transcription bubble that demonstrates an open form (see Fig. 1 *A*). However, at the stage of

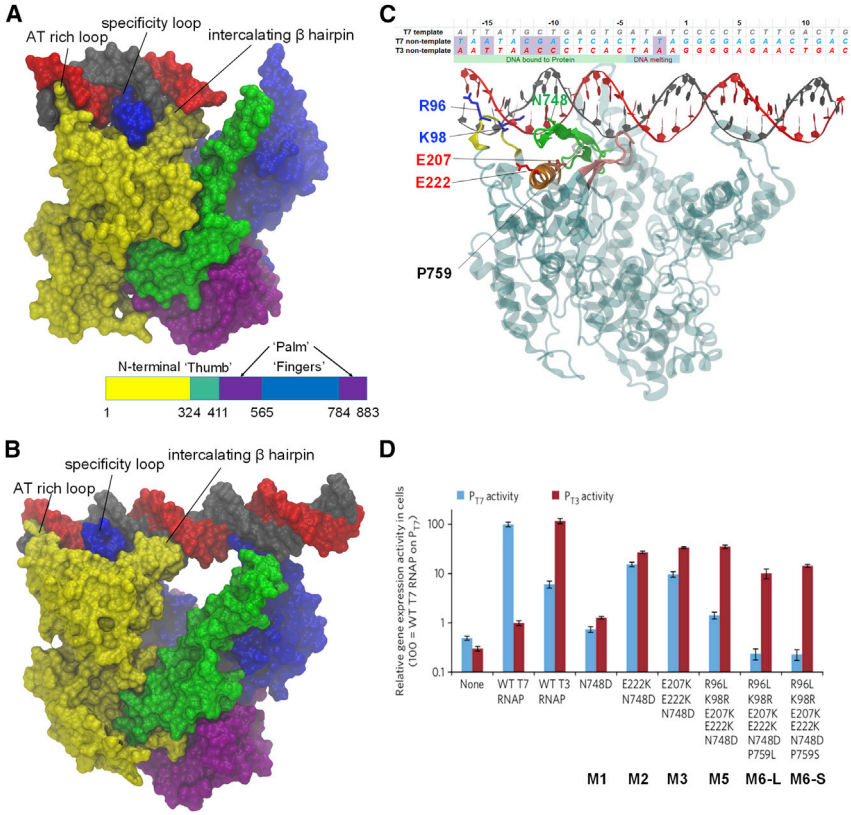

FIGURE 1 Structures of the closed and open complexes of T7 RNAP during transcription initiation and mutant RNAPs from recognizing T7 DNA promotor to T3 promotor identified from directed evolution experiment (23). (*A*) The crystal structure of the T7 RNAP taken from the open complex (PDB: 1QLN), in the absence of transcript bubble. (*B*) The constructed model of the closed T7 RNAP initiation complex in this work. (*C*) The T7 RNAP promotor recognition or the closed initiation complex, with amino acids mutated in the directed evolution shown. The AT-rich loop (ATL) (residues 93–101, yellow), specificity loop (SPL) (residues 739–770, green), intercalating β hairpin (INB) (residues 230–245, pink), and an auxiliary helix (AXH) (residues 206–225, orange) are shown. The protein is shown in transparent cyan. The template and non-template T7 DNA promoter strands are shown in gray and red, respectively, with corresponding sequence listed for both T7 and T3 promoters (differences shaded). (*D*) The gene expression activities of T7 RNAP variants containing subsets of mutations (labeled M1–M6) in the evolved clones from the PACE, working on the T7 (blue bars) and T3 promoters (red). The data show mean values ± SE. The data images are adapted from the experimental work (23).

RNAP initial binding and recognition on the promotor, the promoter DNA still remains closed; hence, a closed promotor complex of T7 RNAP is needed for this study. Employing molecular docking and modeling techniques, we constructed a closed transcription initiation complex of T7 RNAP (see Fig. 1 *B*) so that the specific binding characteristics of T7 RNAP to its promoter can be directly investigated by using all-atom MD simulations. Subsequently, in silico mutations of a small number of protein residues (up to 5 to 6 mutations) were conducted to the wt-RNAP to obtain multiple RNAP variants, which have then been studied; individually also via the all-atom MD simulations.

Earlier studies show that the T7 DNA promoter is mainly composed of two functional domains, a protein-binding region upstream (from −17 to −5) and a transcription initiation region downstream (from −4 to +6; see Fig. 1 *C*; (29,38)). In general, experiments found that the substitution of bases in the upstream binding region has significant impact on the RNAP binding (39,40) but little impact on the initiation; the replacement of bases in the downstream region, however, mainly affects the initiation, but not the RNAP binding (31). In addition, studies have shown that the DNA region responsible for specific binding and recognition in phage T7 promoter is largely via −12 to −8, whereas the region that distinguishes the T7 and T3 promoters locates mainly around −12 to −10 (40–42).

In the PACE experiments system (9,23), the wt-T7 RNAP that recognizes T7 promoter is evolved into variant RNAPs that can finally recognize the promoter from phage T3, with their corresponding promoter activities documented ((23); see Fig. 1 *D*). At an initial evolution stage, the wt-T7 RNAP can transcribe from the T7 promoter, but not from the T3 promoter. Next, a single-point mutant (or M1: N748D) appears, which recognizes neither the T7 nor the T3 promoter, as both promoter activities are low. Further, a double mutant (or M2: E222K and N748D) and a triple mutant (or M3: E207K, E222K, and N748D) show increased promoter activities on both T7 and T3 promoters, with activities on the T3 promoter slightly higher than those on the T7 promoter. Finally, a five-point mutant (or M5: R96L, K98R, E207K, E222K, and N748D) and two six-point mutants (M6-L or M6-S: R96L, K98R, E207K, E222K, N748D, and P759L or P759S) are evolved, which are significantly active on the T3 promoter but function marginally on the T7 promoter, demonstrating prominent bias toward recognizing the T3 promoter. Note that all involved amino acid mutations are located at an AT-rich loop (ATL) (residues 93–101), an auxiliary helix (AXH) (residues 206–225), an intercalating beta hairpin (INB) (residues 230–245), and a specificity loop (SPL) (residues 739–770), which are all located at the RNAP-promoter-DNA-binding interface or nearby (see Fig. 1 *C*; (37)). The ATL and SPL had been previously recognized as the key structural elements for promoter recognition in T7 RNAP (37,43), whereas the mutations on the AXH were discovered from the directed evolution experiment system (9,23) and are particularly analyzed in this work. All these mutations are accordingly studied in current in silico investigations, with seven RNAPs (wt T7 RNAP and six variants) modeled and simulated at the atomic resolution, with each RNAP in complex with T7 and T3 promoters, respectively.

By following the phage T7 RNAP protein and its variants on the lab-directed evolution path obtained from the PACE, rewired toward recognizing the phage T3 promoter, we intend to understand how the protein-DNA recognition is achieved, comparatively, and how the specific recognition function is modulated via point mutations, under pressure of the directed evolution. To do that, we conducted all-atom MD simulations up to 1 μs for each of the above RNAP-promoter DNA complexes. We comparatively studied protein-DNA interactions at the promotor recognition sites for all these systems. We found that electrostatic interactions between the RNAP protein and DNA provide an effective measure on the protein-DNA recognition preference in current systems. Certain residues seem to play key roles in switching the promoter recognition function of the protein. In addition to the ATL and SPL structural elements previously identified, the AXH element turns out to contribute significantly to switch the promoter binding and recognition. In particular, the coordination and competition among these essential structural elements to the promotor DNA are also examined via hydrogen bonding and salt-bridge interactions.

MATERIALS AND METHODS

Obtaining an apo T7 RNAP protein structure

From a crystal structure of T7 RNAP (PDB: 1ARO) (44) that contains additionally a T7 lysozyme (see Fig. S1), we removed the lysozyme and used MODELLER (45) to fill in the missing gaps (residues 60–72, 165–182, 234–240, 345–384, and 590–611) in the protein. The obtained structure was then compared with an apo T7 RNAP structure containing Cα atoms only (PDB: 4RNP) ((46); see Fig. S1), and consistency between the two was found.

Docking the apo T7 RNAP onto double-stranded DNA promoter to construct a closed initiation complex

Using 2.0 version of web 3DNA (w3DNA 2.0) Interface (47), we generated standard B-form double-stranded DNA (dsDNA) containing the T7 promoter, with the template strand consisting of 30 nucleotides (5′-GTCAGTTCTCCCCTATAGTGAGTCGTATTA-3′).

Then, using Hdock Server (48), an online software for protein-protein and protein-DNA and RNA docking based on a hybrid algorithm of template-based modeling and ab initio free docking (http://hdock.phys.hust.edu.cn/), we docked the apo T7 RNAP structure onto the 30-bp dsDNA containing the T7 promotor. First, 100 complex structures were generated from Hdock. Next, a fast-Fourier-transform-based global docking program (HDOCK lite) was used to globally sample putative binding modes in the HDOCK server, in which an improved shape-based pairwise scoring function has been used, and the best-scored top 10 structures were provided (as

shown in Fig. S2). From the top 10 scored models, we selected the three structures that show similar DNA promoter positioning to that from the crystal structure of the T7 RNAP open initiation complex (37). We then performed a 1-μs all-atom MD simulation and calculated the hydrogen bonds between the SPL and promoter in the structure. Finally, we selected the highest scored structure, which has protein-DNA hydrogen bonding interactions well represented, according to the existing open initiation complex structure of T7 RNAP (see Fig. S3). The modeled (and further equilibrated) RNAP structures (wt-RNAP and mutants constructed below) can be made available upon request.

Construction of structural models of mutant RNAPs from directed evolution

According to the lab-directed evolution (23), the wt-T7 RNAP gradually evolved to a series of mt-RNAPs that recognize less the T7 promoter but more the T3 promoter. Based on those mutants, we constructed 14 simulation systems with 7 types of T7 RNAPs, including the wt-RNAP and variants (labeled as M1–M6), and 2 types of dsDNA (containing T7 or T3 promoter). For the mutation, the Tleap method in AmberTools was used to change the amino acid in the protein (49). AmberTools was also used to mutate the DNA basepairs from the T7 promoter to those in the T3 promoter with the nucleic acid backbone unchanged. All these constructed structures were subjected to substantial energy minimization (20,000 steps energy minimization was conducted) and then followed by equilibration MD simulations.

Setup of atomistic MD simulations

All MD simulations were performed using GROMACS-5.1 software package (50–52). The AMBER99sb-2012 force field with PARMBSC0 nucleic acid parameters (53–55) was used to describe the system. The minimum distance from the protein to the border of the simulation box was set to 13 Å. To neutralize the system and keep the ion concentration at an ionic strength of 0.15M salt, 163 Na^+ ions and 119 Cl^- ions were added. The simulation system contained a total of ∼156,000 atoms. The cutoff distance of van der Waals force (vdW) and short-range electrostatic interaction was set to 10 Å. Long-distance electrostatic interactions were handled using the particle net Ewald method (56). The neighbor list of interactions was updated every five steps with a time step of 2 fs. The following procedures were then performed for running each simulation: 1) 20,000 steps energy minimization was conducted using the steepest descent algorithm; 2) 200 ps equilibrium simulation under canonical ensemble (constant number of particles, volume, and temperature) was conducted; followed by 3) 500 ps equilibrium simulation under isothermal-isobaric ensemble (constant number of particles, pressure, and temperature) by position restraining the heavy atoms with a force constant of 1,000 kJ mol^{-1} nm^{-2}. The temperature was maintained at 310K using a velocity-rescaling thermostat with a coupling constant of 0.1 ps^{-1} (57). 4) Finally, a 1-μs MD simulation under isothermal-isobaric ensemble was conducted at 310K and 1 atm using the velocity-rescaling thermostat and the Parrinello-Rahman barostat, respectively (58,59).

Calculation of protein-DNA hydrogen bonds and salt bridges

The hydrogen bonding interactions and salt-bridge interactions formed (>10%) in the last 800 ns of the simulation between the RNAP and the bound region of the promotor DNA. To determine the hydrogen bonds, the distance between the donor atom and the acceptor atom is less than 3.5 Å, and the angle of the donor atom-hydrogen atom-acceptor atom is greater than 140°. The salt bridge is defined when the distance between the most positively charged N atom of the protein Arg or Lys residue and the most negatively charged two oxygen atoms on DNA phosphate group is less than 5 Å.

Calculation of protein-DNA interaction energetics

The protein-DNA interactions were calculated between two residue groups: one group is the promotor DNA (dsDNA −17 to −1) and the other group is the core part of protein within 25 Å of the promoter dsDNA (−17 to −1). The electrostatic (ele) and vdW interactions were re-calculated from simulated trajectories with the water-bearing model, using the g_energy module in Gromacs. The convergences of the calculated energetics (electrostatics) with different sizes of the protein included and different simulation time used are shown in Fig. S4. The energetics between RNAP and T7/T3 promoters and their differences for the key residues are recorded in Table S1, *A* and *B*.

Construction of the CG model and setup of the CG simulations

The coarse-grained (CG) simulations were performed by the CafeMol 3.0 software (60). The initial structure of T7 RNAP was obtained from the crystal structure (PDB: 1ARO) (45). The CG protein structure was constructed by using the off-lattice Go model (61). Each CG particle was located on the Cα atom to represent one amino acid and with the conformations biased toward the native structure (crystal structure here) under the Go-model potential.

In the CG model of dsDNA (200 bp in length), each nucleotide is represented by three CG particles, corresponding to base, sugar, and phosphate groups via the 3SPN.1 model (62,63), in which the bond stretching, angle bending, dihedral angle twisting, base-base interaction, excluded volume effect, solvation energy, and electrostatic energy are considered. The electrostatic interactions and excluded volume effects are considered. All CG simulations are performed by Langevin dynamics under constant temperature with Berendsen thermostat.

RESULTS

Protein-DNA electrostatic interaction energetics provides quantitative measures for the RNAP-promoter recognition

To probe whether protein-DNA interactions that stabilize the RNAP at the promoter also contribute to the promoter recognition and differentiation, we calculated the ele (E^{ele}) and vdW (E^{vdW}) interaction energies between the RNAP protein and the promoter-binding region of the DNA (−17 to −5) for wt T7 RNAP and all mutants (14 simulation systems). The interactions were calculated between atoms from protein and DNA at a cutoff distance ∼25 Å (the results converge for cutoff >20 Å) (see materials and methods and Fig. S4*A*). The convergences of the energetic calculations over simulation time are shown for the wt-RNAP (Fig. 2 *A*) and for the RNAP variants (Fig. S4 *B*–*F*). The energetics obtained by averaging 1 μs of simulation trajectories for individual systems are presented (Fig. 2 *B*). Additionally, we conducted repeating simulations on wt-RNAP and one directed mutant (M5) in complex with T7 promoter and found that the overall protein-DNA

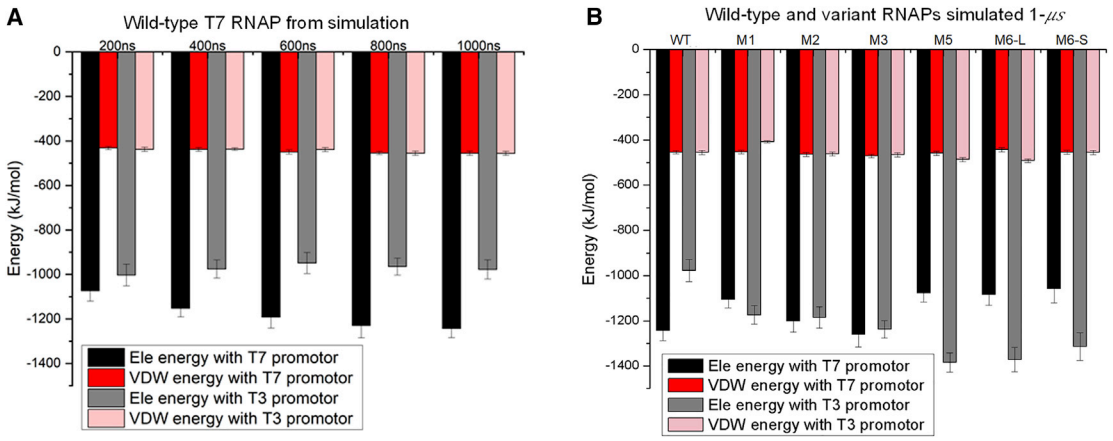

FIGURE 2 Protein-DNA promoter interaction energetics calculated from all-atom MD simulations for the wt T7 RNAP and six RNAP variants in the lab-directed evolution. The interaction energetics include electrostatic (ele) (E^{ele}) and van der Waals (vdW) (E^{vdW}) contributions averaged from simulations. (*A*) The interaction energetics between the wt-RNAP and T7 and T3 promoters from two 1-μs equilibrium MD simulations with data analyses conducted at 0–200 ns, to 400 ns, … and to 1,000 ns or 1 μs. Convergence of energetics shows after ∼400 ns. (*B*) The interaction energetics averaged over the 1-μs individual trajectories for 14 simulation systems, i.e., wt-RNAP and six mt-RNAPs (M1–M6) in complex with T7 and T3 promotors (see Fig. S4 for convergence of RNAP-DNA promoter interaction energetics for all simulation systems).

interaction energetics converges similarly (>400–600 ns) in the repetitive simulation systems (see Fig. S5).

The results indicate that the bias of E^{ele} between the RNAP and the promoter-binding region well characterizes the recognition preferences of the RNAP, as compared with the experimental results (Fig. 1 *D*). For example, the wt T7 RNAP has lower electrostatic interaction energies E^{ele} with the T7 promoter than with the T3 promoter, i.e., it binds more stably in electrostatic energetics to the T7 promoter, consistent with it having higher activities or recognizing better the T7 than T3 promoter. For M1, M2, and M3 RNAP variants, they bind T7 and T3 promoters with similar and slightly weak E^{ele}, consistent with their low promoter activities and differentiation between the T7 and T3. Notably, for M5 and M6-L/S RNAPs that demonstrate high activities on the T3, but not T7, promoter, the protein-DNA electrostatic energetics E^{ele} is significantly lower on the T3 promoter than on the T7 promoter. Meanwhile, the vdW energetics also shows a similar tendency in stabilizing the protein with the promoter of higher activities, though the trend is not as significant as for the electrostatic energetics (see Fig. 2).

Individual residue contributions to the RNAP-promoter electrostatic bias and corresponding dynamics at the protein-DNA binding interface

Since for certain RNAP, the protein-DNA energetic difference between the T7 and T3 promoters well characterizes the promoter recognition preference. We calculated $\Delta E^{ele} = E^{ele}_{T7} - E^{ele}_{T3}$ for each system (E^{ele}_{T7} and E^{ele}_{T3} are the RNAP interaction energetics with the T7 and T3 promoters, respectively) and projected the contributions to ΔE^{ele} onto individual amino acids (aas), as ΔE^{ele}_i for the i-th aa from the RNAP (Fig. 3; or see notable individual aa contributions to respective E^{ele}_{T7} and E^{ele}_{T3} in Fig. S5 and numerical values in Table S1). In particular, we found that the key aas contributing significantly to the promoter recognition, i.e., with large amplitudes of ΔE^{ele}_i, locate mainly on the ATL (aas 93–101), AXH (206–225), INB (230–245), and SPL (739–770). For the wt-RNAP, ATL-R96, AXH-R215, INB-R231, and SPL-R746/756 stabilize both T7 and T3 promoters (see Fig. S5), more to T7 and less to T3; Q135 (located between INB and AXH) only interacts noticeably with T7. Correspondingly, ATL-R96, Q135, AXH-R215, INB-R231, and SPL-R746 have $\Delta E^{ele}_i < 0$, i.e., biasing the RNAP to be more stabilized on the T7 promotor (Fig. 3 *B*). In addition, SPL-N748 shows bias toward the T7 promoter, though its respective interactions with T7 and T3 promoters are not particularly strong (Fig. S6). AXH-E218, however, biases toward the T3 promoter (with $\Delta E^{ele}_i > 0$), without noticeable interactions with the respective promoters either.

In the M1 (N748D) (Fig. 3 *C*), the mutation itself places an immediate energetic bias toward stabilizing the T3 promoter (it is indeed D748 in T3 RNAP) (64). However, energetic contributions of N748D toward either T7 or T3 promoter are still insignificant (see Fig. S6). Though ATL-G97 (and T101), Q135, and INB-R231 still have $\Delta E^{ele}_i < 0$, ATL-K95 and K98 (along with R96 and R99), AXH-H211, and SPL-N748D (along with T745 and R746) start having notable $\Delta E^{ele}_i > 0$, i.e., biasing toward the T3 promoter. The M1 variant thus has the SPL destabilization introduced directly via the N748D mutation, which then perturbs the promoter bias from both ATL and AXH toward T3 (see Fig. 4, *A* and *B*).

The M2 (N748D + E222K) (Fig. 3 *D*) then has an additional AXH-E222K mutation. Now, ATL-R96 (instead of

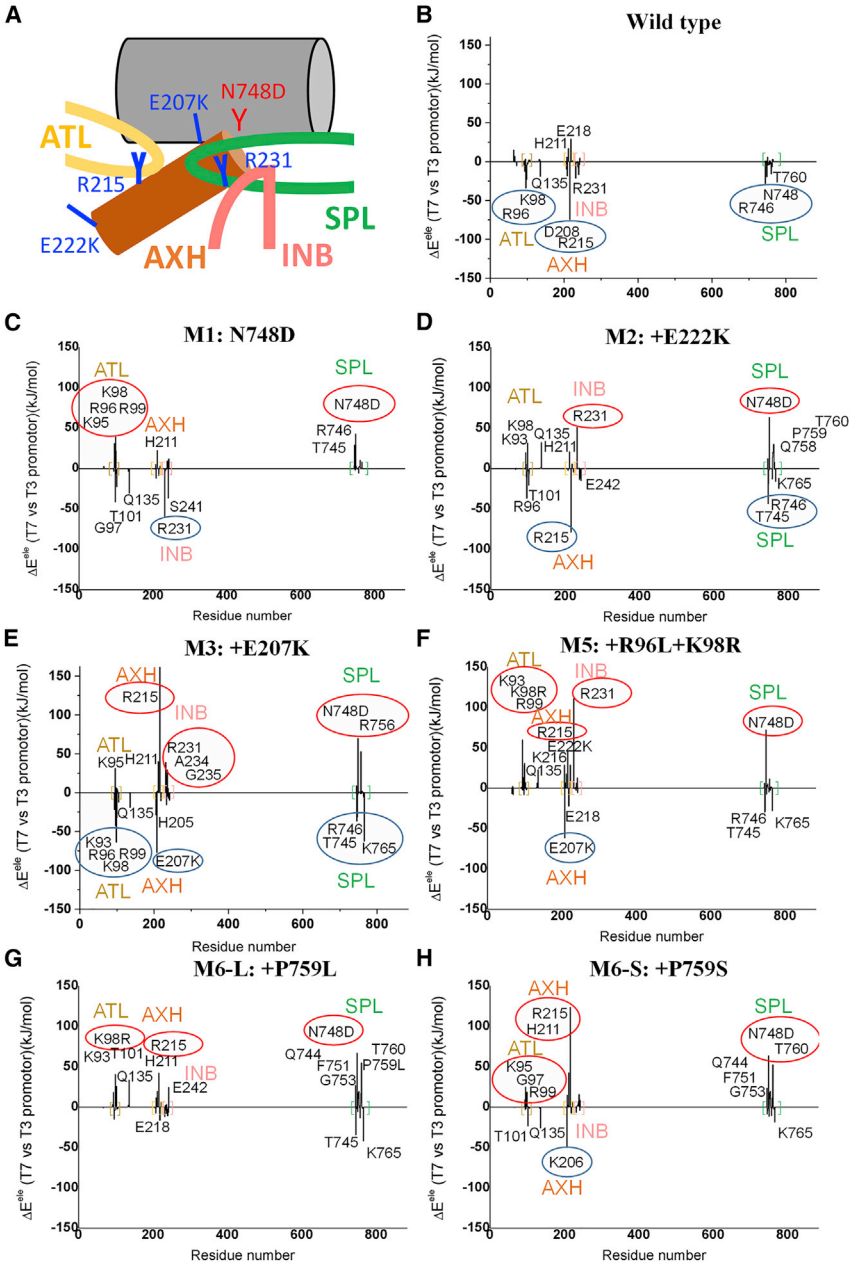

FIGURE 3 The energetic contributions ΔE_i^{ele} from individual amino acids that bias or stabilize the RNAP association with the T7 promotor ($\Delta E_i^{ele} < 0$) or with the T3 promotor ($\Delta E_i^{ele} > 0$), as $\Delta E^{ele} \equiv E_{T7}^{ele} - E_{T3}^{ele}$ is the relative protein-DNA electrostatic interaction energy calculated between the RNAPs and the two promotors.; (*A*) A carton representation shows the key residues and key structural elements. The ATL (AT-rich loop), AXH (auxiliary helix), INB (intercalating beta hairpin), and SPL (specificity loop) are colored in yellow, orange, red, and green, respectively. (*B–H*) The energetic contributions are demonstrated for the wt- to mt-RNAPs (M1 to two M6). Key residues with notable contributions to the energy bias are labeled (with $|\Delta E^{ele}| > \sim 15$ KJ/mol).

G97) and T101, AXH-R215, and SPL-T745 and R746 (along with K765) contribute $\Delta E_i^{ele} < 0$, while ATL-K98, Q135, AXH-H211, INB-R231, and SPL-N748D (along with Q758, P759, and T760) have $\Delta E_i^{ele} > 0$. It seems that AXH-E222K mutation leads to the switch of the promoter preference for both Q135 and INB-R231 (see Fig. 4 *C*). Close examinations show that INB-R231 can switch its side chain upside down from M1 to M2 to be away from the T7 promoter, as E222K brings its side chain toward the T7 promoter, but not toward the T3 promoter (see Fig. S7); accordingly, R231 associates much better with the T3 promoter than the T7 promoter in M2.

Next, for M3 (N748D + E222K + E207K) (Fig. 3 *E*) with a third mutation AXH-E207K, one obtains ATL-K98 (along with R96 and R99), H205, AXH-E207K, SPL-T745, R746, and K765 having $\Delta E_i^{ele} < 0$ to stabilize the T7 promoter; and ATL-K95, AXH-H211 and R215, INB-R231 (along with A234 and G235), and SPL-N748D (along with R756) having $\Delta E_i^{ele} > 0$ bias toward the T3 promoter. N748D has a significant stabilizing interaction with the T3 promoter then (larger than that in M1 or M2), and INB-R231 interacts closely with both T7 and T3 promoters (Fig. S6) yet maintains its bias toward T3. This time, although AXH-E207K itself largely stabilizes the T7 promoter, it facilitates

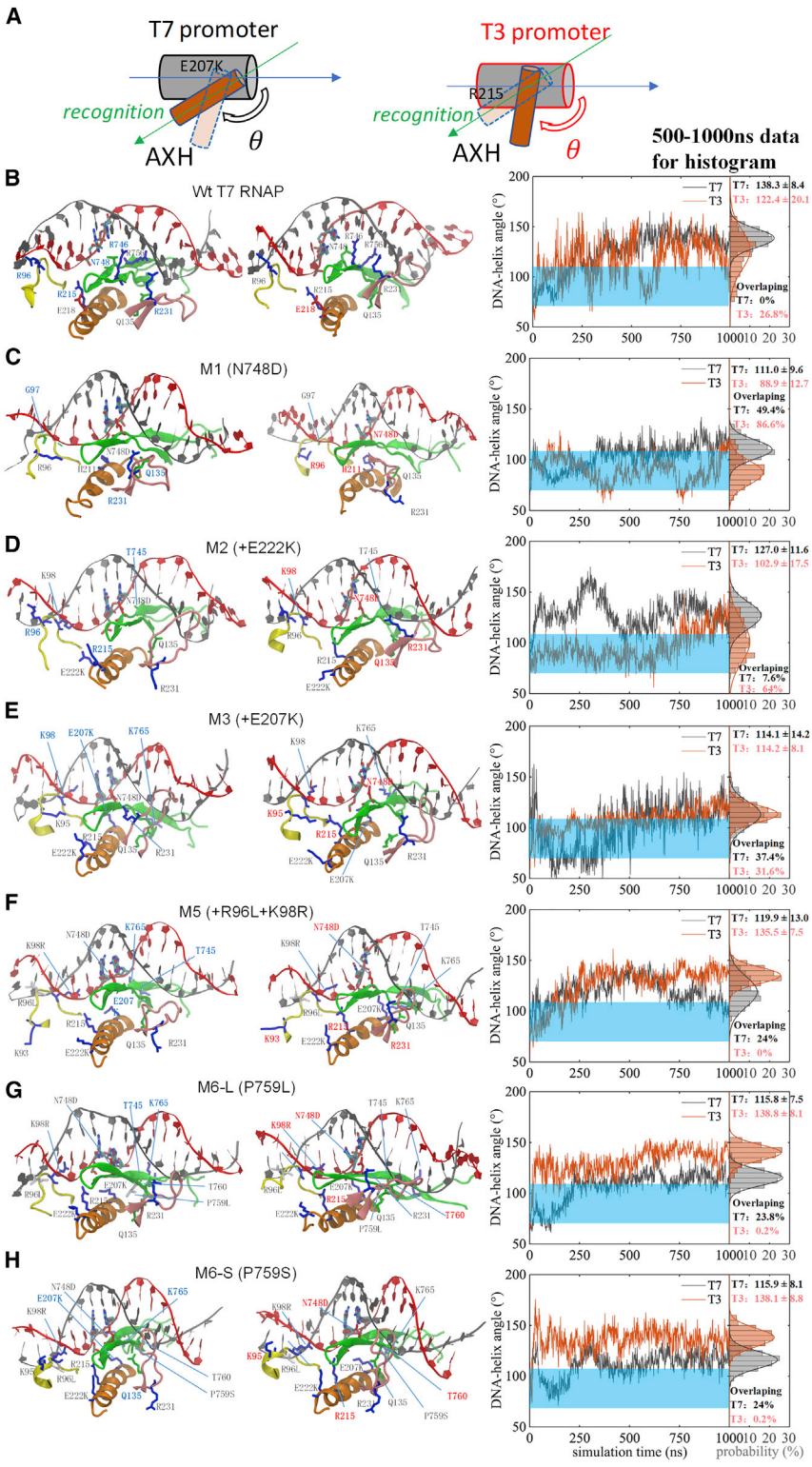

FIGURE 4 Structural views of the protein-DNA binding interface for wt-RNAP and six RNA variants (M1–M6) in complexes with T7 and T3 promoters. The ATL (yellow), SPL (green), INB (pink), and AXH (orange) in close association with the DNA promoter are shown. (*A*) A cartoon representation shows how the AXH orientation angle (θ) is measured, along with prominent angular changes of the AXH as well as E207K and R215 coordination (see text) on the T7 and T3 promoter DNA shown, respectively. (*B–E*) wt-RNAP and M1–M3 early mutants in the evolution. (*F–H*) RNAP directed mutants M5 to two M6 late in the evolution, which recognize preferentially the T3 promoter. In (*B–H*), the left panel and middle panel show the structure views of protein-DNA interface for T7 and T3 promoter , respectively, and the right panel shows the orientation angles (θ) measured correspondingly.The orientation angle (θ) between the AXH and the DNA long axis is measured from the simulations (the time series to the right of the structural views for each system, with black and red data for T7 and T3 promoter systems and blue bars indicating perpendicular positioning or configurations of the AXH (θ∼70°–110°) at disadvantage to the RNAP promoter recognition). The energy values, standard deviations, and overlapping percentages are calculated for 500–1,000 ns and shown in the histograms provided on the right.

AXH-R215 to bind preferably toward the T3 promoter. A competition between E207K and R215 toward the DNA promoter actually shows, as E207K succeeds binding more closely than R215 to the T7 promoter, R215 actually binds more closely than E207K toward the T3 promoter (see Figs. 4 *E* and S8). Due to promoter interactions with E207K and R215 from the N- and C-terminal of the AXH, respectively, the AXH orientation angle (θ) with the DNA

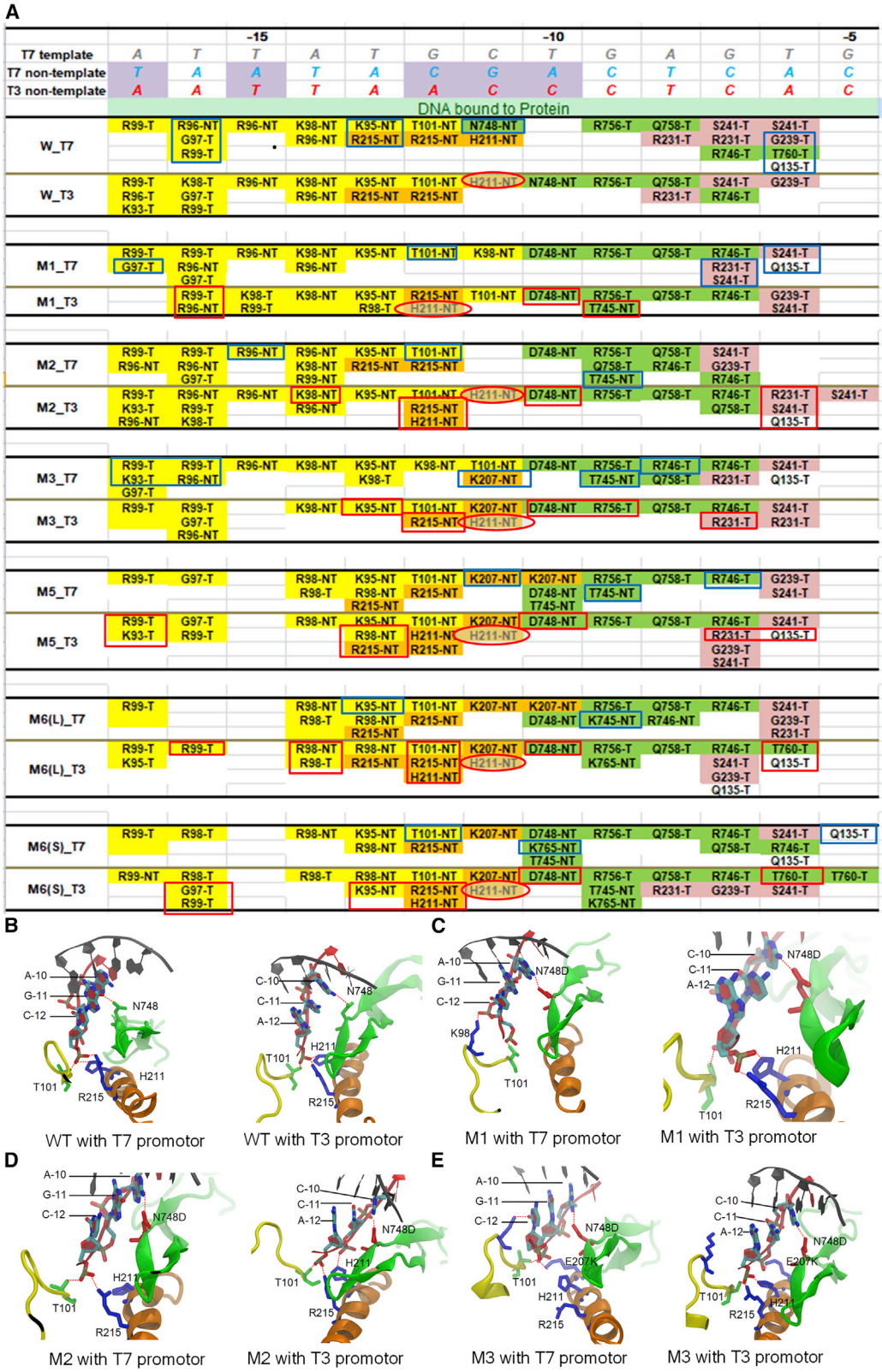

FIGURE 5 The hydrogen bond patterns between RNAP and the promotor DNA from the simulation systems. (*A*) The hydrogen bonding interactions formed (>10%) in the last 0.8 μs of the simulation between the RNAP and the bound region of the promotor DNA. The T7 DNA template (gray) and non-template (blue) chains are displayed schematically with sequences on top, with T3 non-template chain (red) shown as well (the different sequences between T7 and T3 promoters are highlighted in purple). The amino acids at the ATL, SPL, INB, and AXH are highlighted by yellow, green, pink, and orange,

(legend continued on next page)

long axis substantially changes (see Fig. 4), with θ_{T7} decreasing from 127° ± 12° (in M2) to 114° ± 14° in M3 (toward a perpendicular range 70°–110° around 90°) and θ_{T3} increasing from 103° ± 18° (in M2) to 114° ± 8° in M3 (away from the perpendicular range). The θ_{T7} change thus brings the AXH more perpendicular to the T7 promoter DNA, and the trend persists into M5 and M6. Meanwhile, an opposite trend shows for θ_{T3} (i.e., the AXH aligned less perpendicular or better with the T3 promoter DNA). Overall, the M2 and M3 do not differentiate much between the T7 and T3 promoters, yet they prepare for the necessary key residue configurations for the promoter recognition in M5/M6. In particular, R215 biasing toward the T3 promoter and the accompanied AXH orientational changes with respect to the promoter DNA appear to be essential. Such orientational changes of AXH toward more perpendicular to the T7 promoter but aligning better with the T3 promoter demonstrate well into M5 and M6 (see Fig. 4 and overlapping statistics, which measure how much percentile the sampled AXH reorients about perpendicularly on the promoter). Since AXH angles show significant fluctuations, we extended individual simulations up to 1,500 ns and found similar trends of AXH re-orientating differently on the T7 and T3 promoters from wt-RNAP to the directed mutants (M5 and M6; see Fig. S9).

In comparison, the promoter recognition and differentiation become prominent in M5 and M6, in which R96L + K98R on the ATL additionally occur (M5) (Fig. 3 *F*), and then SPL-P759L/S (M6) (Fig. 3, *G* and *H*). In both cases, there are more residues contributing to $\Delta E^{ele}>0$, i.e., to stabilize the T3 promoter. In M5, AXH-E207K (along with E218) and SPL-T745 (along with R746 and K765) remain for $\Delta E_i^{ele} < 0$, whereas ATL-K93 (along with K98R and R99), AXH-R215 and E222K, INB-R231, and SPL-N748D all have $\Delta E_i^{ele} > 0$. Since it is exactly R96 and K98 bias toward the T7 promoter in M2 and M3, respectively, mutations of both largely abolish the ATL bias on the T7 promoter (though K98R still closely interacts with both T7 and T3 promoters) (see Fig. S6). The SPL-K765 stabilization toward the T7 promoter also disappears comparing with M3.

In the M6-L (P759L), SPL-T745 and SPL-K765 can still contribute to $\Delta E_i^{ele} < 0$, although ATL-K98R, AXH-R215 and H211, and SPL-N748D (along with several residues from Q744 to T760) have $\Delta E_i^{ele} > 0$. In the M6-S (P759S), ATL-T101, AXH-K206, and SPL-K765 contribute to $\Delta E_i^{ele} < 0$, and ATL-K95 to R99, AXH-R215 and H211, and SFL-N748D (and residues from Q744 to T760) have $\Delta E_i^{ele} > 0$. It seems that SPL-P759L/S mutation induces the Q744 to T760 region to stabilize the T3 promoter further, although the flexible INB-R231 does not necessarily bias toward the T3 promoter (it interacts with the promoter closely with both T7 and T3 in M6-L) (see Fig. S6). The AXH-E207K stabilization to the T7 promoter does not persist in M6 anymore. Marginal ATL/AXH/SPL stabilization to the T7 promoter still exists in the M5 or M6. Meanwhile, SPL-N748D (started from M1) and AXH-R215 (triggered in M3) contribute robustly toward the T3 promoter. Note that AXH-R215 binds much less stabilized on the T7 promoter in the M6-S(P759S) than in the M6-L(P759L); hence, its overall bias toward the T3 promoter is much more significant in the M6-S than M6-L.

Analyzing hydrogen bonding at the RNAP-DNA promoter interface to probe further aa contribution to recognition

To investigate more specific or detailed recognition of the RNAP on the promoter, we checked the corresponding hydrogen bonds and salt-bridge interactions at the RNAP-promoter interface for each of the simulation systems (see materials and methods). Since most hydrogen bonds are fluctuating and highly dynamical, we recorded hydrogen bonds with at least ∼10% of the occupancy during the microsecond simulation (∼0.8 μs). The corresponding results are summarized in Fig. 5 *A*. Note that the DNA sequences of the T7 and T3 promoters differ only at −17, −15, −12, −11, and −10 positions (around the DNA major groove −15 to −10; see Fig. 1 *D*), and position −12 to −10 are crucial for the promoter specificity (40).

The hydrogen bonds formed by the ATL and DNA span extensively from the upstream minor groove region (−17 to −15) to the major groove in the middle (−15 to −12). In the wt-RNAP, ATL forms approximately four hydrogen bonds (with T7) and seven hydrogen bonds (with T3) at the upstream region (−17 to −15), mainly with the template strand (denoted T); one hydrogen bond forms far downstream as T101: NT-12 (NT denotes the non-template strand) on both the T7 and T3 promoters. In the single-point mutant (M1), ATL forms approximately six hydrogen bonds (with T7) and approximately four hydrogen bonds (with T3) upstream; two hydrogen bonds T101:NT-12 and K98:NT-11 are formed on the T7 promoter, and the two hydrogen bonds switch to T101:NT-11 and K98:NT-12 on the T3 promoter. In the double and triple mutants (M2 and M3), ATL maintains approximately six hydrogen bonds (with T7) but four to seven hydrogen bonds (with T3) upstream; the most downstream hydrogen bond forms as T101:NT-12/-11 on the T7 promoter or as T101:NT-12 on the T3 promoter. In the T3 promoter preferred mutants (M5 and M6), the

respectively. Blue and red rectangles are placed to show residues simultaneously contributing significantly to electrostatic stabilization (as in Fig. 3) to the T7 and T3 promotors, respectively. The hydrogen bonds contributed from the AXH-H211-NT is circled red, as H211-NTs always bias toward the T3 promoter in the mutant RNAPs. (*B–E*) The hydrogen bonds formed between the RNAP and the central bound region (−12 to −10) of promotor DNA for the wt-RNAP (*B*) and for the early or transition mt-RNAPs M1–M3 (*C–E*), respectively, on the T7 promoter (left) and T3 promoter (right).

upstream ATL hydrogen bonds reduced significantly: one to two hydrogen bonds for T7 and three to four hydrogen bonds for T3; the most downstream hydrogen bond is always maintained as T101:NT-12, no matter on which promoter. Hence, it seems that the ATL hydrogen bond association with the upstream template DNA strand is less with T7 in the wt-RNAP, although the trend switches somehow in the transitional mutants (M1 to M3) and recovers a bit but with the overall ATL hydrogen bond association with the promoter weakened in the directed mutants (M5 and M6, due to R96L and K98R mutations on the ATL). In particular, the ATL association most downstream T101:NT-12 can move to NT-11 in the transitional mutants (M1 and M3) but recovers to stable T101:NT-12 in the directed mutants.

As for the aas on the AXH, all hydrogen bonds are concentrated in the middle of the major groove on the non-template DNA strand (NT-13 to -11). In the wt-RNAP, AXH forms three same hydrogen bonds (R215:NT-13, R215:NT-12, and H211:NT-11) (see Fig. 5 *B*) with both T7 and T3 promoters. Hence, the hydrogen bonds do not seem to contribute to promoter differentiation. In M1, AXH loses the hydrogen-bonding association with the T7 promoter entirely, whereas two AXH hydrogen bonds (R215:NT-12 and H211:NT-12) maintained with the T3 promoter (Fig. 5 *C*). R215 and H211 hydrogen bonds recover somehow in M2 (upon E222K), with H211 persistently forming hydrogen bond biasedly on the T3 promoter, as well as in M3 (upon E207K and K207:NT-11 formed for both promoters) (Fig. 5, *D* and *E*). K207 continues to associate with NT-11 or even NT-10 into the directed mutants (M5 and M6), on both T7 and T3 promoters, with R215 forming hydrogen bonds with NT-13/-12 on both promoters; H211 remains associating preferentially with the T3 promoter (M5 and M6) (see Fig. S10). Hence, it seems that M1 (or SPL-N748D) critically breaks a balance of the AXH hydrogen-bonding association with the promoter DNA between the two species (T7 and T3) and enables AXH-H211 to maintain hydrogen bonding with the T3 promoter, but not with the T7 promoter anymore, and then, the mutations E222K and E207K, i.e., directly emerge on the AXH, enhance the AXH association with the DNA, as well as support the hydrogen bonding preference of H211 to the T3 promoter, persistently into the directed mutants.

The hydrogen bonds formed by the SPL and the promoter DNA are mainly located far downstream (mainly on template T-10 to -7), up to NT-11 (only for N748 in the wt-RNA with T7) or down to NT-6 and -5 (T760 in the wt-RNAP with T7 or in the M6 with T3). Notably, N748D emerges as a first and most critical mutation and moves the base-specific hydrogen-bonding N748:NT-11 with T7 and N748:NT-10 with T3 in the wt-RNAP to D748:NT-10 for all the RNAP variants (from M1 to M6) (see Figs. 5 and S8). D748 is the only residue that forms a hydrogen bond with the DNA base (NT-10) in the transitional mutants (M1-M1) and into the directed mutants (M5-M6) on both T7 and T3 promoters; NT-10 associates with E207K and/or T745 additionally, only on T7, but not T3 promoter. Hence, N748D seems to be essential to re-position hydrogen bonding interaction for the promoter recognition and support the AXH for the re-wired promoter differentiation. Other hydrogen-bond-forming residues in the SPL seem to maintain stable contacts in all systems, including the wt-RNAP and variants (R756:G-9, Q758:A-8, and R746:G-7). Note that T760:T-6 exists in the wt-RNAP with the T7 promoter preference; it then switches to the T3 promoter preference (M6), due to the mutation P759L/S.

Finally, the INB region forms hydrogen bonds mainly via R231, Q239, and S241 with template -7 and -6 position (occasionally with R231 to -8 or S241 to -5). In the wt-RNAP, INB forms a couple more hydrogen bonds with the T7 promoter than with the T3 promoter. Such a bias reduces or even reverses slightly in the RNAP variants. It appears that INB can play some role still. In particular, INB-R231 shows a transient role in promoting bias toward the T3 promoter (in M2 and M5), energetically or via forming hydrogen bonds with the DNA, yet in general, R231 side chain is highly flexible and frequently swings, without sustained bias.

In addition, we also examined the salt-bridge interactions at the RNAP-promoter DNA interface (see Fig. S11). ATL at upstream contributes dominantly to the salt-bridge interactions, which shows no obvious differentiation between T7 and T3 promoters. Nevertheless, in the wt-RNAP, the ATL salt-bridge R96-NT-16 with the T7 promoter does contribute to energetically stabilize the T7 system; in the M2/M3, the salt-bridge K98-T-15/T-14 on the T7 promoter does as well (see Fig. 2 *B*). Interestingly, such stabilization and bias abolish in the M5/M6 with R96L and K98R, which indicate that mutations on the ATL exactly promote the specificity to the T3 promoter. Meanwhile, at the -12 to -10 region key for the promoter differentiation, there are salt bridges from the AXH (e.g., R215-NT for all RNAPs except for M3 with T7; K207-NT starting from M3), from ATL (K98 or R98 in all mt-RNAPs except for M3 with T7; occasionally K95-NT in M1 with T3 and M3 with T7), and from SPL (occasionally K765-NT for M1/M5/M6 with T7). In particular, one can see that the AXH involved with more salt-bridge interactions with DNA starting from M1-M2 upon the mutations and the AXH salt bridges extend further in M3–M6.

DISCUSSION

The promoter binding of an RNAP plays a primary regulatory role in gene transcription. Since T7 RNAP can conduct initiation without transcription factor, it is expected that the RNAP can also search and locate the promoter, possibly via 1D diffusion along DNA, as detected experimentally (65). Indeed, we also tested in this research the apo T7 RNAP search on the DNA non-specifically, using CG

modeling and MD simulation (60). We found that T7 RNAP diffuses processively along DNA with the SPL and the ATL structural elements making particularly frequent contacts with DNA (see Fig. S12), due to protein-DNA electrostatic interactions (with an implicit solvent modeled at an ionic strength 0.15 M). Hence, it seems that the SPL and ATL can be the most important structural elements for the RNAP to locate the promoter sequences for initiation as the RNAP conducts diffusional search nonspecifically along DNA.

The directed evolution experiment had been designed to train the RNAP from recognizing the T7 promoter to recognize the T3 promoter instead and demonstrated a representative evolutionary path following the wt-RNAP → M1 (N748D) → M2 (+E222K) → M3 (+E207K) → M5 (+R96L and K98R) → M6-L (+P759L) or M6-S (+P759S). Based on the promoter activities and differentiation, one can divide these RNAPs into four groups: the wt-RNAP, which recognizes the T7, but not T3, promoter; M1, low promoter activities on both promoters; M2 and M3, notable promoter activities on both promoters yet no differentiation; and M5 and two M6s, which recognize the T3, but not T7, promoter. We analyzed the mechanism of switching the specific protein-DNA recognition along the above directed evolution path by conducting MD simulation of individual RNAPs along the path. According to previous experimental work (40) and the alignment of the T7 and T3 promoter sequence, it is noted that the −12 to 10 region of the promoter DNA mainly determines the specific sequence recognition. In particular, N748 from the SPL forms the only specific hydrogen bonding contact with the DNA base NT-11G (or NT-10C) on this region to the T7 (or T3) promoter. Accordingly, one expects that the hydrogen bonding between residue 748 and NT-11 or NT-10 is key to the specific recognition. Meanwhile, in the wt-RNAP, ATL extends from upstream to form hydrogen bonding contact T101:NT-12 downstream while AXH competitively binds −13 to −11 region (R215:NT-13 and -12 and H211:NT-11) similarly on both T7 and T3 promoters. Following, one sees that four critical protein-DNA binding and recognition transitions along the directed evolution path that play important roles.

The first mutation N748D from SPL breaks up the original binding and specificity to T7 promoter from the wt-RNAP as transiting to the M1 RNAP. The mutation N748D indeed shifts the specific hydrogen bonding contact from NT-11G to NT-10A on the T7 promoter, although there is no shift on the T3 promoter yet (N748D:NT-11C maintains). Along with this, K98 and T101 from the upstream ATL extend hydrogen bonding from (NT-14 and NT-12 originally) to NT-11 downstream on the T7 and T3 promoters, respectively. It appears as if ATL "pulls" on the −12 to −10 promoter DNA toward upstream (see Fig. 6 schematics). AXH then behaves in response by abolishing R215 and H211 hydrogen bonds with NT-13 to −10 on the T7 promoter, although it adjusts the R215 and H211 hydrogen bonds from NT-13 to −11 to NT-12 altogether on the T3 promoter. One thus sees that the "deactivation" of the M1 RNAP on the T7 promoter simply results from

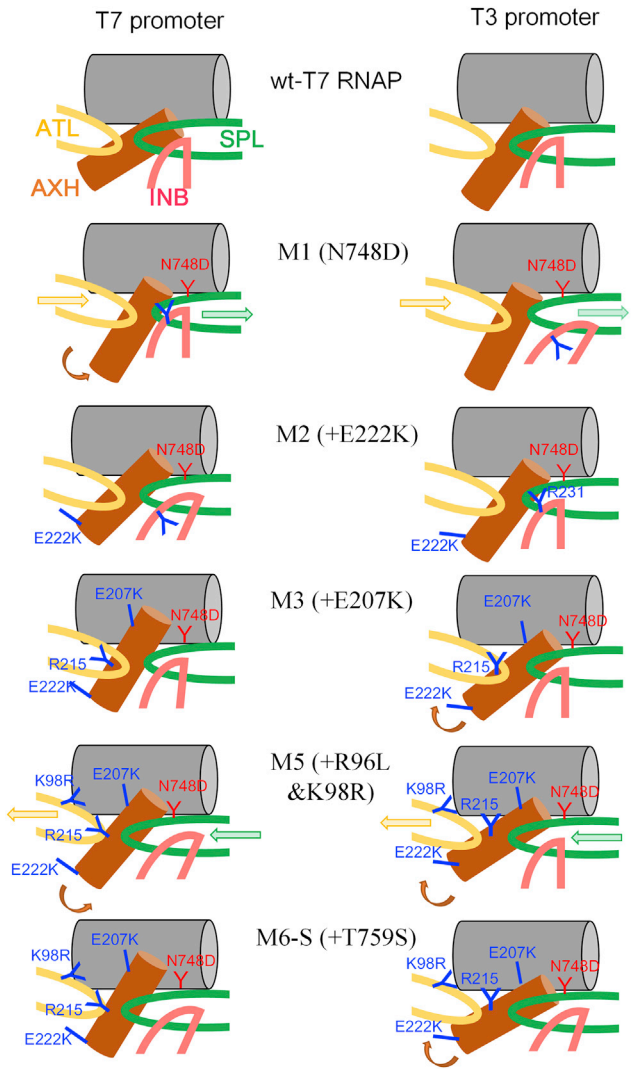

FIGURE 6 Schematics on the key structural elements and residues at RNAP-promoter DNA interface along the directed evolution path of switching the RNAP from recognizing the T7 promoter to recognize the T3 promoter. The key structural elements ATL (yellow), AXH (orange), INB (red), and SPL (green) are presented along with most key residues in mutation and response. In the wt-RNAP to M1 transition (N748D bias toward T3), ATL and SPL shift downstream and AXH rotates (toward perpendicular) on the T7 promoter. Upon M1 → M2 (+E222K), AXH moves toward the DNA promoter slightly and starts to play a bigger role, although INB fluctuates to allow R231 stabilization toward the T3 promoter, but not the T7. Upon M2 → M3 (E207K), AXH-R215 moves toward the T3 promoter to align AXH better along the DNA axis, although R215 cannot compete well with E207K on the T7 promoter. Upon M3 → M5 (+R96L and K98R), ATL loses the biased R96 and K98 interaction with T7; the ATL or SPL withdraws or extends upstream, so that AXH becomes even better aligned with the T3 promoter DNA but less with the T7 promoter DNA, which supports the switched RNAP recognition to T3 promoter. M5 → M6 (+T759S) further locks the T3 promoter specificity.

perturbing some critical hydrogen bonds (SPL-N748D and AXH-R215 and H211). The consequent energetic impacts are noticeable, as the RNAP-T7 promoter interaction is destabilized although RNAP-T3 promoter interaction stabilized, electrostatically (the vdW still stabilizes or biases toward T7) (see Fig. 2 *B*), with ATL-K95, AXH-H211, and SPL-N748D stabilizing toward T3.

Next, AXH starts to play a bigger role via mutation E222K from M1 to M2 RNAP. The mutation of the negatively charged E222 on the AXH to the positively charged K222 allows the AXH to move closer to the promoter DNA (for both T7 and T3) than in the M1. Consequently, AXH recovers its hydrogen bonds with the T7 promoter via R215:NT-13 and 12, which also energetically stabilizes the T7 promoter binding. H211 additionally forms a hydrogen bond with NT-11 (aside with NT-12 in M1) on the T3 promoter. Energetically, E222K also triggers an immediate orientational switch of the side chain R231 from the INB, allosterically (as the distance between the two residues is about 15–22 Å), taking advantage of the flexible side-chain motions of R231 on the INB-loop region so to stabilize R231 association with the T3 promoter. As R215 and R231 energetically bias toward the T7 and T3 promoters, respectively, there is no obvious energetic bias (either electrostatic or vdW) to T7 and T3. M2 RNAP (N748D + E222K) accordingly shows notable promoter activities but does not display bias or differentiation between two promoters.

Then, AXH takes over the promoter binding via E207K from M2 to M3. The additional charge conversion from negative to positive on the AXH allows it to move even closer to the promoter than M2. E207K indeed well stabilizes energetic association of the M3 RNAP with the T7 promoter, and it also forms hydrogen bonding with NT-11 on both T7 and T3 promoters. Interestingly, however, the responses of R215 and consequent competition between K207 and R215 in association with the promoter reveal differently in the T7 and T3 systems. On the T7 promoter, R215 stays far from the DNA promoter and K207 dominates the energetic association and forms hydrogen bonds with NT-11; on the T3 promoter, however, R215 stays similarly close to the DNA promoter as K207, and they form hydrogen bonds with NT-12 and NT-11, respectively. Thus, residue R215 originally favors the T7 promoter association (in the wt-RNAP) and then switches to stabilize more or bias toward the T3 promoter upon the E207K mutation, and notably, such R215 bias toward T3 then maintains robustly along the followed evolution path. One sees that, upon the second and the third mutations on the AXH, it is mainly the local electrostatic charge interactions that compete for the RNAP-promoter association, even though the overall energetic contributions between the two promoter systems are still in balance and show no bias. A dominant role in switching electrostatic energetic bias from T7 to T3 promoter at the protein-DNA interface can indeed be found for AXH (see Fig. S13), which significantly stabilizes T7 promoter in wt-RNAP yet switches to stabilize T3 promoter notably from transient to the directed mutants (M3 to M6), assisted by three other elements (ATL, INB, and SPL).

Further, the ATL modulation via R96L + K98R fully switches the promoter bias to T3, transiting from M3 to M5 (and M6). Although the most key residues N748K and R215 (enabled by E207K) energetically biasing toward the T3 promoter have established robust association in M3, the further energetic stabilization to enable the promoter specificity is achieved by the ATL. Before, ATL remains energetically stabilizing to the T7 promoter and R96/K98 contributes to that, i.e., via the electrostatic interactions. Although R96L simply reduces the ATL-promoter association due to loss of electrostatic attraction, K98R abolishes hydrogen bonding with NT-12 so that ATL withdraws upstream on the T7 promoter, having T101 hydrogen bonding shifted from NT-11 (in M3) to NT-12 (in M5). Meanwhile, SPL-T745 extends upstream to form hydrogen bonding with NT-10 (along with D748) on the T7 promoter. Such changes may accordingly destabilize the AXH association with the T7 promoter. On the other hand, AXH seems to associate with the T3 promoter more extensively as H211 forms an additional hydrogen bonding with NT-12 (aside from NT-11), and R215 occasionally forms an additional hydrogen bonding with NT-13 (aside from NT-12), aside from its robust energetic bias toward T3. With another mutation P759L/S (M6-L or M6-S) from the SPL, small energy bias from T760 to the T3 promoter is brought about, and the energetic competition of K207 (weakening the bias toward T7) and R215 (strengthening the bias toward T3) on the AXH can be further tuned to bias toward the T3 promoter. Hence, the last stage enabling the promoter specificity seems to be achieved by further balancing a variety of hydrogen bonds and local charge interactions among ATL, AXH, and SPL.

CONCLUSIONS

We utilized all-atom MD simulations to reveal physical mechanisms of viral T7 RNAP variants rewiring promoter recognition along the lab-directed evolution path, as the promoter recognition of the RNAP switches from the original T7 promoter to the slightly different T3 promoter. As the first point mutation N748D emerges on the SPL (specificity loop) of T7 RNAP to bias toward the T3 promoter, it critically shifts its hydrogen bond to 1 nt downstream and perturbs the hydrogen bonding patterns at the protein-DNA interface, breaking the balance between the SPL downstream and the ATL upstream and slightly dissociating the RNAP from the promoter to hinder the original promoter recognition. Notably, the current study identifies an auxiliary helix (AXH 206–225) that takes over switching the RNAP-promoter recognition via the second and third

mutations (E222K and E207K) of the RNAP along the directed evolution path, as AXH interacts more closely with the promoter mainly via the charge interactions upon the two mutations and then reorientates differently on the T7 and T3 promoters, mainly due to competition between E207K and R215 in binding to the promoter. The promoter specificity is thus facilitated and finally switched upon mutations on the ATL (R96L + K98R), which adjust the protein-DNA hydrogen bonding and salt-bridge patterns further and resets the balance between the ATL and SPL. Final mutation on the SPL (R759L or R759S) modulates the RNAP-promoter interactions additionally and strengthens the promoter specificity. Such structural dynamics details revealed from the simulations, in particular, in the transitional mutants (M1, M2, and M3) can be tested experimentally via high-resolution structural characterization or fluorescence measurements. The revealed mechanisms may assist structure-function information learning of the system to promote further rational design on specific RNAP-promoter recognition.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.01.007.

AUTHOR CONTRIBUTIONS

J.Y. conceived and designed the research. C.E. and L.D. performed the simulation. C.E., L.D., and J.Y. analyzed the data. J.Y. and C.E. wrote the paper.

ACKNOWLEDGMENTS

This work has been supported by National Natural Science Foundation of China grant nos. 11775016 and 11635002. J.Y. has been supported by the National Science Foundation-Simons Center for Multiscale Cell Fate Research of University of California, Irvine via NationalScience Foundation Division of Mathematical Sciences 1763272, the Simons Foundation grant no. 594598, and a start-up fund from Universityof California, Irvine. We also acknowledge the computational support from the Beijing Computational Science Research Center (CSRC).

REFERENCES

1. Arnold, F. H. 2018. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed. Engl.* 57:4143–4148.
2. Bornscheuer, U. T., B. Hauer, …, U. Schwaneberg. 2019. Directed evolution empowered redesign of natural proteins for the sustainable production of chemicals and pharmaceuticals. *Angew. Chem. Int. Ed. Engl.* 58:36–40.
3. Kwon, K. K., D.-H. Lee, …, S.-G. Lee. 2018. Evolution of enzymes with new specificity by high-throughput screening using DmpR-based genetic circuits and multiple flow cytometry rounds. *Sci. Rep.* 8:1–9.
4. Bornscheuer, U. T. 1998. Directed evolution of enzymes. *Angew. Chem. Int. Ed. Engl.* 37:3105–3108.
5. Cherry, J. R., and A. L. Fidantsef. 2003. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* 14:438–443.
6. Arnold, F. H. 1998. Design by directed evolution. *Acc. Chem. Res.* 31:125–131.
7. Yuan, L., I. Kurek, …, R. Keenan. 2005. Laboratory-directed protein evolution. *Mol. Biol. Rev.* 69:373–392.
8. Turner, N. J. 2009. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* 5:567–573.
9. Esvelt, K. M., J. C. Carlson, and D. R. Liu. 2011. A system for the continuous directed evolution of biomolecules. *Nature.* 472:499–503.
10. Packer, M. S., and D. R. Liu. 2015. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16:379–394.
11. Badran, A. H., and D. R. Liu. 2015. In vivo continuous directed evolution. *Curr. Opin. Chem. Biol.* 24:1–10.
12. Cobb, R. E., R. Chao, and H. Zhao. 2013. Directed evolution: past, present, and future. *AIChE J.* 59:1432–1440.
13. Chatterjee, R., and L. Yuan. 2006. Directed evolution of metabolic pathways. *Trends Biotechnol.* 24:28–38.
14. Crook, N., J. Abatemarco, …, H. S. Alper. 2016. In vivo continuous evolution of genes and pathways in yeast. *Nat. Commun.* 7:1–14.
15. Shen, Y.-P., L. San Fong, …, J.-Z. Liu. 2019. Combining directed evolution of pathway enzymes and dynamic pathway regulation using a quorum-sensing circuit to improve the production of 4-hydroxyphenylacetic acid in Escherichia coli. *Biotechnol. Biofuels.* 12:1–11.
16. Yokobayashi, Y., R. Weiss, and F. H. Arnold. 2002. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U S A.* 99:16587–16591.
17. Bonsor, D. A., and E. J. Sundberg. 2011. Dissecting protein− protein interactions using directed evolution. *Biochemistry.* 50:2394–2402.
18. Bryson, J. W., S. F. Betz, …, W. F. DeGrado. 1995. Protein design: a hierarchic approach. *Science.* 270:935–941.
19. Hellinga, H. 1997. Rational protein design: combining theory and experiment. *Proc. Natl. Acad. Sci. U S A.* 94:10015–10017.
20. Tiwari, M. K., R. Singh, …, J. K. Lee. 2012. Computational approaches for rational design of proteins with novel functionalities. *Comput. Struct. Biotechnol. J.* 2:e201204002.
21. Davey, J. A., A. M. Damry, …, R. A. Chica. 2017. Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* 13:1280–1285.
22. Korendovych, I. V. 2018. Rational and semirational protein design. *Methods Mol. Biol.* 1685:15–23.
23. Carlson, J. C., A. H. Badran, …, D. R. Liu. 2014. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* 10:216–222.
24. Cramer, P. 2002. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* 12:89–97.
25. Kornberg, R. D. 1999. Eukaryotic transcriptional control. *Trends Cell Biol.* 9:M46–M49.
26. Sousa, R., S. Mukherjee, and M. Kivie. 2003. T7 RNA polymerase. *Prog. Nucleic Acid Res. Mol. Biol.* 73:1–41.
27. McAllister, W. 1997. Transcription by T7 RNA polymerase. *In* Mechanisms of Transcription. Springer, pp. 15–25.
28. Skinner, G. M., C. G. Baumann, …, J. G. Hoggett. 2004. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. *J. Biol. Chem.* 279:3239–3244.
29. Rong, M., B. He, …, R. K. Durbin. 1998. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U S A.* 95:515–519.
30. Jorgensen, E. D., R. K. Durbin, …, W. McAllister. 1991. Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *J. Biol. Chem.* 266:645–651.
31. Újvári, A., and C. T. Martin. 1997. Identification of a minimal binding element within the T7 RNA polymerase promoter. *J. Mol. Biol.* 273:775–781.
32. Raskin, C. A., G. Diaz, …, W. T. McAllister. 1992. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase

at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* 228:506–515.

33. Ikeda, R. A., L. L. Chang, and G. S. Warshamana. 1993. Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences. *Biochemistry.* 32:9115–9124.

34. Sousa, R., Y. Chung, …, E. Lafer. 1990. Single crystals of a chimeric T7/T3 RNA polymerase with T3 promoter specificity and a nonprocessive T7 RNAP mutant. *J. Biol. Chem.* 265:21430–21432.

35. Raskin, C. A., G. A. Diaz, and W. T. McAllister. 1993. T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl. Acad. Sci. U S A.* 90:3147–3151.

36. Cheetham, G. M., and T. A. Steitz. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science.* 286:2305–2309.

37. Cheetham, G. M., D. Jeruzalmi, and T. A. Steitz. 1999. Structural basis for initiation of transcription from an RNA polymerase–promoter complex. *Nature.* 399:80–83.

38. Chapman, K. A., S. I. Gunderson, …, R. R. Burgess. 1988. Bacteriophage T7 late promoters with point mutations: quantitative footprinting and in vivo expression. *Nucleic Acids Res.* 16:4511.

39. Lee, S., and C. Kang. 1992. A two-base-pair substitution in T7 promoter by SP6 promoter-specific base pairs alone abolishes T7 promoter activity but reveals SP6 promoter activity. *Biochem. Int.* 26:1–5.

40. Klement, J. F., M. B. Moorefiedl, …, W. T. McAllister. 1990. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. *J. Mol. Biol.* 215:21–29.

41. Bailey, J. N., J. F. Klement, and W. T. McAllister. 1983. Relationship between promoter structure and template specificities exhibited by the bacteriophage T3 and T7 RNA polymerases. *Proc. Natl. Acad. Sci. U S A.* 80:2814–2818.

42. Li, T., H. H. Ho, …, C. T. Martin. 1996. Major groove recognition elements in the middle of the T7 RNA polymerase promoter. *Biochemistry.* 35:3722–3727.

43. Steitz, T. A. 2009. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr. Opin. Struct. Biol.* 19:683–690.

44. Jeruzalmi, D., and T. A. Steitz. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* 17:4101–4113.

45. Šali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.

46. Sousa, R., Y. J. Chung, …, B.-C. Wang. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature.* 364:593–599.

47. Li, S., W. K. Olson, and X.-J. Lu. 2019. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* 47:W26–W34.

48. Yan, Y., D. Zhang, …, S.-Y. Huang. 2017. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 45:W365–W373.

49. Case, D. A., T. A. Darden, …, W. Zhang. 2008. Amber 10. University of California.

50. Hess, B., C. Kutzner, …, E. Lindahl. 2008. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

51. Van Der Spoel, D., E. Lindahl, …, H. J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

52. Berendsen, H. J., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

53. Guy, A. T., T. J. Piggot, and S. Khalid. 2012. Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophys. J.* 103:1028–1036.

54. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.

55. Joung, I. S., and T. E. Cheatham, III. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* 112:9020–9041.

56. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

57. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

58. Nosé, S., and M. Klein. 1983. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50:1055–1076.

59. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.

60. Kenzaki, H., N. Koga, …, S. Takada. 2011. CafeMol: a coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* 7:1979–1989.

61. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

62. de Pablo, J. J. 2011. Coarse-grained simulations of macromolecules: from DNA to nanocomposites. *Annu. Rev. Phys. Chem.* 62:555–574.

63. Hinckley, D. M., J. P. Lequieu, and J. J. de Pablo. 2014. Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* 141:035102.

64. McGraw, N., J. Bailey, …, W. McAllister. 1985. Sequence and analysis of the gene for bacteriophage T3 RNA polymerase. *Nucleic Acids Res.* 13:6753–6766.

65. Kim, J. H., and R. G. Larson. 2007. Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res.* 35:3848–3858.