NAZARBAYEV
UNIVERSITY

INVESTIGATION OF T7 RNA POLYMERASE BIOCHEMICAL AND PHYSICO-CHEMICAL PROPERTIES

Yernur Kenzhegazin
(B.Sc., Nazarbayev University)

A THESIS SUBMITTED
FOR THE BIOL 491 Honor Thesis
DEPARTMENT OF BIOLOGY SCHOOL OF SCIENCES AND HUMANITIES
NAZARBAYEV UNIVERSITY
2025

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Yernur Kenzhegazin

25 April, 2025

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor Ferdinand Molnár for his support and introducing me to alternative perspectives in *in silico* methodologies.
I also extend my thanks to Akerke Mazhibiyeva, who generously shared tutorials and practical guidance on MD simulation procedures.
Lastly, the use of a state-of-the-art workstation was instrumental in accelerating data analysis.

TABLE OF CONTENTS

Nazarbayev University, School of Sciences and Humanities, Department of Biology

Bachelor´s Degree Program in Biology

Name of the student: Yernur Kenzhegazin

Title of the Thesis: Investigation of T7 RNA Polymerase biochemical and physico-chemical properties
BIOL 491 thesis; 24 pages
Supervisors: Ferdinand Molnar, PhD, Associate Professor

25.04.2025

Keywords*: T7 RNAP, MD Simulations, N-terminal domain flexibility, active-site loop stabilization, thermostability.*

SUMMARY

T7 RNA polymerase (T7 RNAP) is a single-subunit RNA polymerase whose activity relies on large conformational transitions during transcription. Here we employed molecular dynamics simulations to characterize how specific point mutations (S43Y, Y639V, V685A, Q744R), a double mutant (Y639V/V685A), and a quadruple mutant (S43Y/Y639V/V685A/Q744R) alter the enzyme's dynamics at 310 K and 315 K. Wild-type and mutant T7 RNAPs (all simulated in a promoter-bound complex) were compared in terms of flexibility (root-mean-square fluctuations, RMSF), global stability (RMSD) and intramolecular interactions. All single mutations increased the flexibility of the N-terminal domain (NTD), while **decreasing** the flexibility of an active-site adjacent loop (residues 600–610). Notably, the quadruple mutant exhibited a **lower** overall RMSF at 315 K than at 310 K, indicating a gain in rigidity at elevated temperature. This anomalous thermal response suggests that the combination of mutations confers enhanced structural stability and possibly an "activated" conformational state at higher temperature. Hydrogen-bond analysis revealed that a key anchoring interaction (His411–Glu35) present in the wild type is lost in all mutants, explaining the liberated NTD movements. Our findings indicate that these mutations allosterically rewire T7 RNAP's internal dynamics: they **free** the NTD and **stabilize** the active site region, a dual effect that may underlie known functional phenotypes (e.g. reduced termination by S43Y and expanded substrate scope by Y639V). These insights advance our understanding of structure–function relationships in T7 RNAP and inform strategies for engineering polymerases with altered stability or transcriptional properties.

List of figures

ABBREVIATIONS

PDB	Protein Data Bank
RMSF	Root-mean square fluctuation
RMSD	Root-mean square deviation
MD	Molecular Dynamics
RNAP	RNA polymerase
NVT	Constant-Number, Volume, Temperature ensemble
NPT	Constant-Number, Pressure, Temperature ensemble
NTD	N-terminal domain

1 INTRODUCTION

1.1 Background: T7 RNAP as a Model Single-Subunit Polymerase

Transcription in bacteria and phages is carried out by RNA polymerases (RNAPs) that synthesise RNA from a DNA template. While most well-studied RNAPs (such as the bacterial and eukaryotic polymerases) are large, multisubunit complexes, bacteriophage T7 RNAP is striking in its simplicity: it is a single 883-amino-acid polypeptide (~99 kDa) capable of transcribing DNA with high fidelity and efficiency (Cheetham *et al*, 1999a; Dunn & Studier, 1983). Despite having none of the auxiliary subunits of cellular RNAPs, T7 RNAP conducts transcription via the same fundamental two-metal-ion catalytic mechanism and accomplishes all the requisite stages of the transcription cycle. It specifically recognises a 17-base-pair promoter sequence upstream of genes (Cheetham & Steitz, 1999b), initiates RNA synthesis *de novo* (i.e. without a primer), undergoes an initial phase of abortive transcription (producing short RNA fragments) (Yin & Steitz, 2002), and then transitions into a highly processive elongation mode. T7 RNAP also responds to regulatory signals: for example, it can be inhibited by T7 lysozyme, a phage-encoded protein that binds the polymerase and modulates its activity (Moffatt & Studier, 1987). Indeed, T7 lysozyme binding prevents the formation of a productive initiation complex, thereby down-regulating T7 RNAP during certain phases of phage infection (Jeruzalmi & Steitz, 1998; Moffatt & Studier, 1987). The enzyme finally terminates RNA synthesis at specific termination sequences, analogous to the rho-independent terminators in bacteria (Yin & Steitz, 2002). Together, these observations have made T7 RNAP a model system for understanding the core requirements of transcription and a powerful tool in biotechnology (e.g. for high-yield *in vitro* RNA production) (Wang *et al*, 2018; Yin & Steitz, 2002). However, a detailed understanding of how a single-subunit RNAP orchestrates such a complex sequence of events – and how its activity can be tuned by mutations – is still emerging.

1.2 Structural Architecture and Promoter Recognition

Structurally, T7 RNAP belongs to the DNA polymerase I family and adopts a conserved "cupped right hand" architecture with distinct domains termed the palm, thumb and fingers (Cheetham *et al*, 1999a). Like other members of this fold, the palm domain contains the catalytic site (including the two metal ions for catalysis), while the thumb and fingers wrap around the DNA to position it near the active site. T7 RNAP also contains an N-terminal domain (NTD) of 324 amino acids that is unique to single-subunit RNAPs and has no counterpart in multisubunit enzymes (Cheetham *et al*, 1999a). The NTD plays a critical role in transcription initiation: together with a β-hairpin "specificity loop" insertion in the fingers domain (residues ~740–770), the NTD forms the binding site for promoter DNA (Cheetham *et al*, 1999a). High-resolution crystal structures have captured T7 RNAP in the act of initiation, bound to its promoter. In the T7 RNAP-promoter open complex (PDB 1CEZ), the polymerase grips a 17-bp promoter DNA duplex such that the upstream DNA is clamped by the NTD, while the downstream DNA is fed into the active-site cleft between the thumb and fingers domains (Cheetham *et al*, 1999a). Notably, in this initiation complex the DNA at positions –4 to +1 (with respect to the transcription start site) is unwound, forming a transcription "bubble" with the nontemplate strand separated from the template strand (Cheetham *et al*, 1999a). The NTD makes extensive contacts with the promoter in this state and is responsible for recognising

conserved promoter sequence elements and stabilising the unwound DNA strands to enable RNA synthesis to begin (Cheetham *et al*, 1999a).

1.3 Active-Site Dynamics: The Tyr639 Gate

Structural snapshots also reveal that T7 RNAP undergoes considerable conformational changes during the transition from initiation to elongation. In the 1CEZ initiation-state structure, a key active-site residue, Tyr639 (located in the fingers domain), is observed in an "open" configuration: it resides in a flexible loop at the C-terminus of an incomplete α-helix (the O-helix) and does not obstruct the nucleotide-entry channel (Cheetham *et al*, 1999a). This corresponds to an open active site, which is thought to facilitate the binding of the initial NTPs. By contrast, in structures of T7 RNAP captured in later stages of transcription, the conformation of the fingers domain is different. For example, a structure of an elongation complex shows Tyr639 as part of a well-formed O-helix, where its side chain occupies a position that can sterically block the entry of an incoming NTP (or alternatively insert into the active centre) depending on its rotamer orientation (Cheetham *et al*, 1999a; Sousa & Padilla, 1995). In other words, Tyr639 acts as a gating residue: when the O-helix is closed, Tyr639 helps enforce the specificity for ribonucleotides by preventing easy access of deoxyribonucleotide triphosphates to the active site, a mechanism that was confirmed by mutagenesis and biochemical studies (Jeruzalmi & Steitz, 1998; Sousa & Padilla, 1995). Mutating Tyr639 to a smaller residue is known to relax this specificity, allowing T7 RNAP to incorporate DNA nucleotides or other 2′-modified substrates much more readily (Sousa & Padilla 1995; Brieba & Sousa 2000) (Sousa & Padilla, 1995). Thus, the initiation complex of T7 RNAP represents a more open conformation of the enzyme, poised for initial NTP binding, whereas the elongation complex adopts a more closed conformation that promotes high-fidelity nucleotide addition and prevents backtracking during RNA synthesis (Sousa & Padilla, 1995; Yin & Steitz, 2002).

1.4 Domain Movements from Initiation to Elongation

During the initial stage of transcription (when RNA chains of 2–10 nucleotides are being made), T7 RNAP remains bound at the promoter, and the NTD continues to secure the upstream DNA. This often results in an abortive-transcription phase, in which short RNA fragments are repeatedly synthesised and released before the enzyme successfully produces a transcript long enough to escape the promoter. Transitioning from initiation to elongation requires a major structural rearrangement of the polymerase. Specifically, T7 RNAP must loosen its grip on the promoter DNA and reposition its NTD to accommodate the emerging RNA–DNA hybrid and allow the nascent RNA to exit the enzyme's surface. X-ray crystallography and biochemical analyses have illuminated aspects of this transition. Notably, the NTD (particularly its promoter-binding region) undergoes a dramatic repositioning: it refolds and swings by approximately 45° relative to the core of the polymerase (Cheetham *et al*, 1999a). This movement disengages the NTD from the promoter DNA, facilitating "promoter clearance" – the step at which the RNAP finally moves beyond the promoter. Intriguingly, structural intermediates have been captured with an 8-nucleotide RNA transcript still bound, in which the NTD is partially rotated but still maintains some contact with the upstream DNA (Cheetham *et al*, 1999a). This suggests that the NTD releases the promoter in

a concerted yet delayed fashion, relinquishing contacts only at the last possible moment to ensure initiation was successful (Cheetham *et al*, 1999a). Once in the elongation phase, the NTD is completely displaced from the DNA and plays no further direct role in binding the nucleic acids. Concurrent with NTD movement, other domains adjust as well. Elements of the fingers domain, including the O-helix containing Tyr639, shift to accommodate continuous RNA elongation. Opening of the O-helix (and the corresponding movement of Tyr639) has been associated with forward translocation of the RNA–DNA hybrid and the prevention of backtracking in the elongation complex (Cheetham *et al*, 1999a). In summary, T7 RNAP's function critically depends on coordinated domain motions: the NTD must be dynamic enough to swing away at the appropriate time, and active-site loops in the fingers domain must rearrange to secure each incoming NTP and support translocation of the RNA–DNA hybrid. These concerted changes enable the enzyme to morph from a DNA-clamped initiation machine into a processive elongation complex (Cheetham *et al*, 1999a; Yin & Steitz, 2002).

1.5 Mutation-Driven Functional Modulation

1.5.1 S43Y: N-Terminal Domain Hinge Mutation

Given this intimate coupling between structure and function, it is perhaps not surprising that even small mutations in T7 RNAP can have far-reaching effects on transcriptional behaviour. Numerous studies over the years have isolated mutations that alter the fidelity, efficiency or other properties of T7 RNAP. In this thesis, we focus on four specific point mutations – S43Y, Y639V, V685A and Q744R – and combinations thereof, to investigate how they influence T7 RNAP's dynamics and activity. These mutations were selected based on prior evidence suggesting they modulate different aspects of the polymerase's function. **S43Y** lies in the N-terminal domain, near the interface with promoter DNA when the enzyme is in the initiation complex. Strikingly, substitutions at Ser43 (and the neighbouring residue Gly47) have been shown to suppress the enzyme's tendency for "self-priming" RNA extension (Wu *et al*, 2021). In wild-type T7 RNAP, abortive transcripts can rebind as primers for DNA-templated extension (an undesirable side reaction that produces double-stranded RNA by-products), but the S43Y mutation largely eliminates this RNA-dependent RNA-polymerase activity without impairing normal transcription (Yin & Steitz, 2002). Biochemically, S43Y also reduces intrinsic termination efficiency, hinting that this residue is involved in maintaining the appropriate conformation for termination versus read-through (Yin & Steitz, 2002). S43 is located in a hinge region of the NTD that undergoes repositioning during promoter clearance, suggesting that this mutation may alter the timing or stability of NTD movements (and thereby affect release of abortive transcripts and termination complexes) (Yin & Steitz, 2002).

1.5.2 Y639V: Relaxing Substrate Specificity

Tyr639, as described above, is a pivotal "gatekeeper" at the active site. The mutation Y639V removes the bulky aromatic side chain and is known to relax the substrate specificity of T7 RNAP. Y639V and the related Y639F mutation allow the polymerase to incorporate deoxyribonucleotides or 2′-modified nucleotides much more efficiently than the wild-type enzyme (Sousa & Padilla, 1995). T7 RNAP variants carrying Y639 substitutions have been widely used to synthesise modified RNAs, as they can tolerate substrates such as 2′-O-methyl or 2′-fluoro NTPs that normally would be rejected by the wild-type enzyme (Sousa & Padilla, 1995).

1.5.3 V685A: Stabilising the Fingers Domain

Val685 is another residue in the fingers domain, positioned near the active site and the template-binding groove. It does not directly contact the nucleotides, but it lies in a region important for maintaining the structural integrity of the active-site cavity during elongation. A substitution at this position (V685A) was identified in engineering studies aimed at improving T7 RNAP performance under non-standard conditions (such as transcription of 2′-modified RNA). Notably, V685A emerged from a screen for mutations that increase the thermal stability or tolerance of T7 RNAP variants with expanded substrate range (Siegmund *et al*, 2012; Yin & Steitz, 2002). By itself, V685A is believed to slightly stabilise the fingers domain and help the polymerase retain activity at higher temperatures or with more challenging substrates (Siegmund *et al*, 2012; Yin & Steitz, 2002).

1.5.4 Q744R: Enhancing Elongation Stability

Finally, **Gln744** resides within the conserved specificity loop of the fingers domain (the β-hairpin loop that interacts with promoter DNA during initiation). Q744R introduces a positively charged side chain in this loop. This mutation has been reported to enhance overall transcription yield in *in vitro* transcription assays, possibly by strengthening the interactions that stabilise the elongation complex (for instance, by improving contacts with the phosphate backbone of the DNA or RNA) (Cheetham *et al*, 1999a). In essence, Q744R might reduce the likelihood of premature transcript release or enzyme dissociation, thereby leading to more full-length product.

1.6 Molecular Dynamics Study Design

These four mutations affect three different regions of T7 RNAP – the NTD (S43Y), the active-site gate (Y639V and nearby V685A) and the specificity loop (Q744R) – providing an opportunity to examine how perturbations in each region influence the polymerase's overall dynamics and function. In this study, we use MD simulations to analyse the wild-type T7 RNAP and the above mutants, both individually and in combination. All simulations start from the same structural state: the open promoter complex (PDB 1CEZ), which includes the full-length polymerase bound to a 17-bp promoter DNA with an unwound bubble at the start site. By retaining the DNA in our simulations, we preserve the essential context for initiation and the interactions between the NTD and the promoter. We compare metrics such as root-mean-square deviation (RMSD) and fluctuation (RMSF) for different domains of the enzyme, and we examine specific intra-molecular contacts (e.g. hydrogen bonds) that report on the stability of key structural elements.

1.7 Insights, Mechanistic Interpretation and Biotechnological Implications

Our results indicate that the NTD becomes substantially more flexible in the S43Y mutant (and in the multi-mutant that includes S43Y) relative to the wild type, consistent with the idea that this mutation loosens the NTD's grip on the DNA. Conversely, the Y639V mutation – especially when combined with V685A – tends to rigidify the active-site region: in particular, a loop spanning residues 600–610 in the fingers domain shows decreased flexibility, suggesting a more "closed" active-site configuration. Notably, this loop (which includes part of the O-helix and adjacent segments) is adjacent to where Tyr639 would normally reside; its

stabilisation in the mutant could reflect altered interactions due to the absence of the bulky tyrosine. The Q744R mutation, on its own, also confers a mild stabilising effect on the fingers domain and the downstream DNA interactions, which might correlate with the experimental observation of higher transcriptional output. By integrating these findings, we propose a mechanistic interpretation: increasing NTD mobility (as seen with S43Y) may facilitate the polymerase's transition into elongation (by easing promoter release), whereas stabilising the active-site region (as seen with Y639V/V685A) could help the enzyme maintain a productive elongation complex once the initial hurdles are overcome. In combination, these mutations illustrate how tuning different parts of the enzyme can influence the balance between initiation and elongation states. In the broader context, understanding these allosteric effects is of great significance – it not only deepens our knowledge of the transcription mechanism of T7 RNAP but also guides the development of improved RNAP variants for synthetic biology. T7 RNAP has already been harnessed as an "enabling tool" in mRNA-vaccine production and other biotechnological applications (Yin & Steitz, 2002), and insights into its domain dynamics open avenues to rationally design polymerases with tailored properties (for instance, thermostability, higher yield, altered substrate specificity or reduced aberrant activities). In summary, this work addresses a gap in our understanding of single-subunit RNAPs by linking specific structural perturbations to functional outcomes, thereby highlighting the delicate coordination of domain movements that makes T7 RNAP a highly efficient transcription machine.

2 MATERIAL AND METHODS

2.1 Protein Preparation

The starting structure for all simulations was the crystal structure of T7 RNA polymerase (T7 RNAP) in complex with a 17 bp promoter DNA (PDB ID: 1CEZ). This structure represents the initiation complex and includes the full-length polymerase (883 amino acids) along with a partial DNA duplex containing the T7 promoter sequence. To maintain the biological relevance of the promoter-bound state, all molecular dynamics (MD) simulations were conducted with the promoter DNA present.

The four point mutations of interest — **S43Y**, **Y639V**, **V685A**, and **Q744R** — were introduced in silico using PyMOL (v2.5), ensuring the absence of steric clashes and maintaining proper rotamer conformations at mutation sites. Furthermore, a double mutant (Y639V + V685A) and a quadruple mutant (S43Y + Y639V + V685A + Q744R) were constructed by simultaneously introducing the respective mutations.

The protein structure was refined using the **Protein Preparation Wizard** in Maestro (Schrödinger, Inc.). This preparation workflow included assignment of bond orders via the Chemical Components Dictionary (CCD), identification and conversion of disulfide bridges, and reconstruction of missing side chains and loops using Prime. Protonation states were determined using Epik at a target pH of 7.0 ± 2.0 to capture biologically relevant microstates. Water molecule orientations were optimized, and hydrogens of altered species were minimized. The protein structure was further processed under crystal symmetry constraints, and pKa values of ionizable residues were predicted using PROPKA at pH 7.0.

To remove close contacts and optimize the internal hydrogen-bonding network, a restrained energy minimization was performed using the **OPLS4 force field**, converging heavy atoms to an RMSD threshold of 0.30 Å while allowing full relaxation of hydrogen atoms. This minimization ensured the structural readiness of each model prior to subsequent simulations.

2.2 Molecular Dynamics Simulation Setup

All molecular dynamics simulations were conducted using the **Desmond module** within the Maestro suite (Schrödinger, Inc.). Systems were solvated in an orthorhombic box using the **TIP3P water model**, ensuring a minimum buffer distance of 10 Å between the solute and the box boundaries. Neutralization was achieved by adding appropriate counterions (Na^+ or Cl^-), and physiological ionic strength was established by supplementing with 0.15 M NaCl.
Each prepared system underwent energy minimization for 500–1000 steps to relieve any steric clashes or unfavorable contacts. The equilibration phase consisted of a **100 ps protocol** performed in two stages:
Under constant volume and temperature (NVT) conditions at 310 K,
Followed by constant pressure and temperature (NPT) conditions at 1 atm and 310 K.
Subsequently, **production runs of 100 ns** were performed for each variant. To enhance the statistical reliability and sample conformational diversity, all simulations were run **in triplicate** for the wild type and all mutant systems. Raw trajectory data from these simulations were exported for subsequent RMSD, RMSF, and hydrogen bond analyses.

2.3 RMSD Analysis

The **root mean square deviation (RMSD)** was employed to monitor the structural stability and equilibration of T7 RNAP throughout the simulations. By comparing the initial atomic coordinates (t = 0) with those at subsequent time points, the RMSD provides insights into global conformational changes induced by mutations. Lower RMSD values indicate maintenance of structural integrity, while larger deviations reflect conformational rearrangements.
Specifically, RMSD analysis facilitated the evaluation of how each mutation, individually or in combination, influenced the enzyme's stability relative to the wild type. The RMSD was calculated using the following formula:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((x_i^m - x_i^0)^2 + (y_i^m - y_i^0)^2 + (z_i^m - z_i^0)^2)}$$

where x^m, y^m, z^m represent the atomic coordinates at time *t*, x^0, y^0, z^0correspond to the initial coordinates, and *N* is the number of atoms considered in the calculation.
Variations in RMSD over the course of the simulation provided a time-resolved view of structural fluctuations, helping to understand the dynamic behavior of mutant RNAPs.

2.4 RMSF Analysis

The **root mean square fluctuation (RMSF)** was utilized to assess residue-level flexibility of T7 RNAP across the trajectory. RMSF quantifies the average deviation of each atom or residue from its mean position, thereby highlighting flexible versus rigid regions of the protein.
By comparing RMSF profiles of wild-type and mutant enzymes, we identified specific domains (e.g., the N-terminal domain and active-site-adjacent loops) that exhibited altered dynamics due to mutations. Peaks in the RMSF plot correspond to highly mobile residues, while troughs indicate structurally stable segments.
The RMSF was computed using the following equation:

$$\mathrm{RMSF} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(x_i - \bar{x})^2}$$

where T is the total number of frames in the trajectory, x_i is the position of the atom at frame i, and $\overline{x}$ is its average position throughout the simulation.
Higher RMSF values pointed to increased flexibility, which can have functional implications such as enhanced promoter escape, altered substrate binding, or changes in termination efficiency.

2.5 Hydrogen Bond Occupancy Analysis

We analyzed **intramolecular hydrogen bonds** within the protein. In this study, hydrogen bond analyses were performed using **Visual Molecular Dynamics (VMD)** software. The identification of hydrogen bonds was based on VMD's default geometric criteria:

1. A **donor–acceptor distance ≤ 3.5 Å**, and
2. A **donor–hydrogen–acceptor angle ≥ 160°**, corresponding to a deviation of ≤ 20° from linearity.

For each trajectory, hydrogen bond **occupancy** was calculated as the percentage of simulation frames in which a given H-bond exists. We first identified hydrogen bonds present in the wild-type simulations with occupancy ≥ 10% in any run, to serve as a reference set of notable interactions. We then examined how these occupancies changed in each mutant. For each hydrogen bond of interest, we computed the occupancies in wild type vs. mutant and expressed the change as Δ(occupancy) = Occupancy_mutant – Occupancy_wildtype (so a negative Δ indicates a bond present in wild type was less populated or broken in the mutant, whereas a positive Δ indicates a bond gained or strengthened in the mutant).

It is important to note that some hydrogen bonds involve the side chains of mutated residues themselves (e.g., the hydrogen bond between the side chain of Gln744 and Asn762, which could be disrupted by Q744R); these are directly impacted by the mutation. Others are more distant, reflecting allosteric or indirect structural changes.

All hydrogen bond data were averaged over the three runs for a given system to get robust occupancy values. When identifying common patterns across mutants, we took into account slight differences in histidine protonation nomenclature (HIS, HIE, HIP in our analysis) – for instance, a hydrogen bond involving His411 could be labeled differently if the protonation state changed; such cases were unified manually when comparing across systems.

3 AIMS OF THE THESIS PROJECT

I. Examine how individual mutations (S43Y, Y639V, V685A, Q744R) alter local and global flexibility in T7 RNAP using RMSF and RMSD from molecular dynamics simulations.

II. Determine how each mutation affects the intra-molecular hydrogen-bonding network, comparing these changes to the wild-type enzyme.

III. Assess the combined effects of multiple mutations—specifically the Y639V/V685A double mutant and the S43Y/Y639V/V685A/Q744R quadruple mutant—to identify whether their structural changes are additive or compensatory, and relate these findings to known phenotypes (e.g., increased thermostability, reduced termination or altered substrate tolerance).

4 RESULTS

4.1 RMSF Analysis: Altered Flexibility Profiles in Mutants vs. Wild Type

Figures 1–7 present RMSF trajectories for each mutant (blue) overlaid on the wild-type reference (red), offering a direct visual comparison of local backbone mobility. Although the overall fluctuation pattern is largely conserved, several peaks—particularly within the N-terminal (residues 50–55) and the fingers–palm junction (residues 600–610)—show mutation-specific attenuation or amplification, underscoring how individual and combinatorial substitutions selectively reshape the enzyme's flexibility landscape.

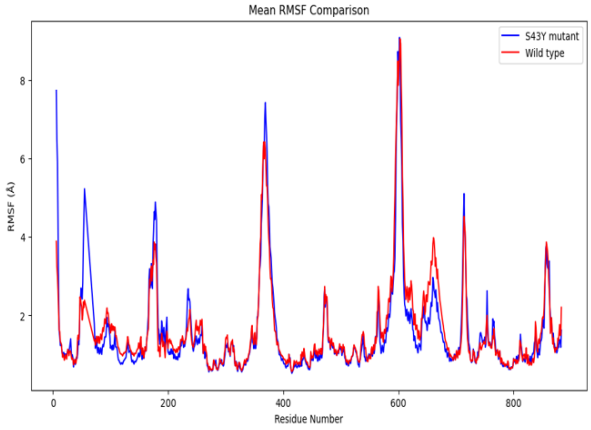

Figure 1. RMSF: S43Y vs Wild Type

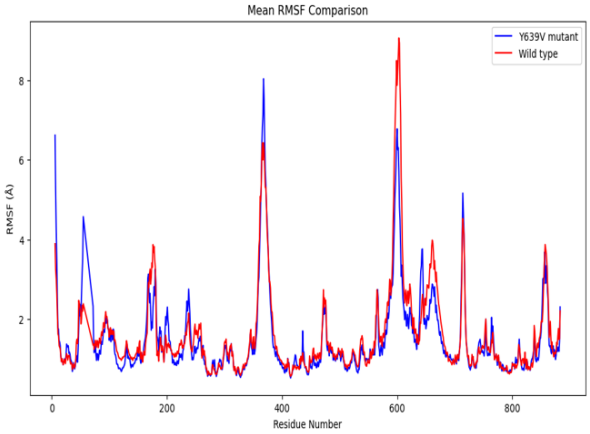

Figure 2. RMSF: Y639V vs Wild Type

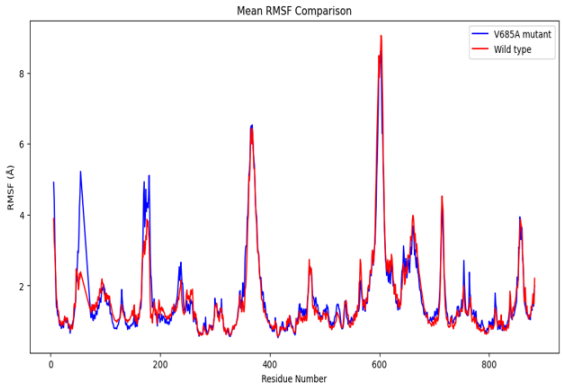

Figure 3. RMSF: V685A vs Wild Type

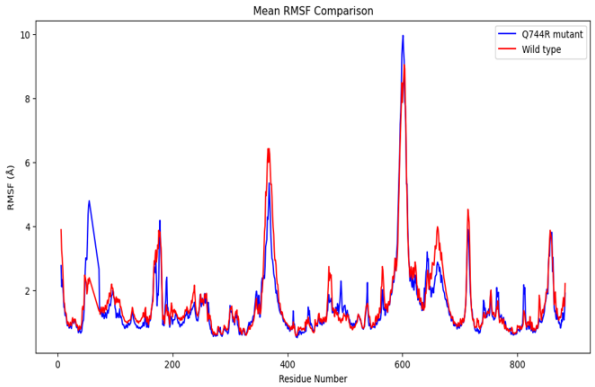

Figure 4. RMSF: Q744R vs Wild Type

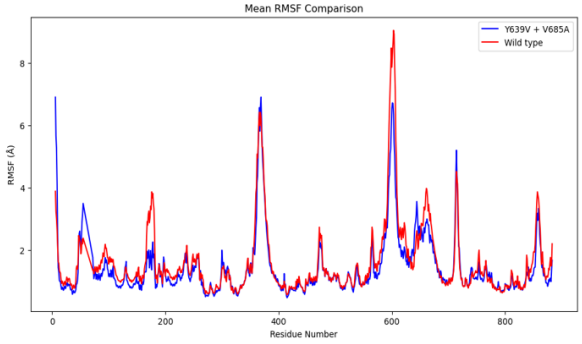

Figure 5. RMSF: Double vs Wild Type

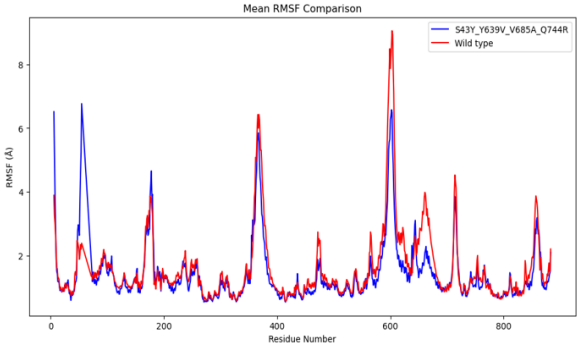

Figure 6. RMSF: Quadruple vs Wild Type

4.1.1 Increased Flexibility of the N-Terminal Domain (NTD)

The NTD becomes more flexible in all mutants. In the wild type, the average Cα RMSF of the N-terminal 1–324 residues was about 2.5 Å. In contrast, this same region in the mutants showed much higher fluctuations. For example, in the S43Y mutant, residues 6–10 (at the very N-terminus, near the mutation site Ser43) have RMSF values of 5–7 Å, roughly double the wild-type values. Similarly, the Y639V and V685A single mutants each exhibited ~1–3 Å increases in RMSF throughout the NTD relative to wild type. The double mutant Y639V/V685A showed an additive effect, with the NTD RMSF increasing by up to 3.5 Å at some positions compared to wild type (e.g., residue 6 RMSF: 3.9 Å in WT vs. 6.6 Å in double). At the same time, in many regions, the double mutant appears to strike an intermediate balance, retaining the local decreases in flexibility where either single mutation already stabilizes the structure, but mitigating the local increases in flexibility where one (or both) single mutants destabilize it. For example, around residue 50, both Y639V and V685A yield RMSF values of ~5 Å, but the double mutant reduces it to ~3.5 Å—still above the wild-type level yet lower than in either single mutant.

The quadruple mutant had the most dramatic change: the NTD's average RMSF rose to ~4.8 Å (versus ~2.5 Å in WT), with certain residues (e.g., Phe55 in the loop of the NTD) experiencing a three-fold increase (WT 2.4 Å → Quad 7.3 Å, a +4.9 Å difference). These data indicate that mutations in both the NTD (S43Y) and the active site (Y639V, V685A) destabilize the NTD's interaction with the rest of the polymerase, freeing it to sample a wider range of positions. This is consistent with the loss of anchoring interactions, as will be discussed in the H-bond section.

4.1.2 Rigidification of the 600–610 Loop Adjacent to the Active Site

In contrast to the NTD, a specific loop region in the C-terminal domain showed the opposite effect – a decrease in flexibility in most mutants. This loop (residues 600–610) is located in the fingers domain, adjacent to the active site;

In the wild-type simulations, this region was relatively flexible (for instance, Glu604 had RMSF ~8.9 Å, among the highest in the enzyme, reflecting that this loop can move significantly, perhaps to accommodate DNA or NTP binding). Strikingly, in the Y639V mutant this loop became much more rigid: Glu604 RMSF dropped to ~4.8 Å ($\Delta \approx -4.1$ Å) and neighboring residues Gly603, Ile605, Ser606 each saw reductions of 3.3–3.4 Å in RMSF relative to WT (their fluctuations roughly halved).

The double mutant Y639V/V685A compounded this effect: in the double, Glu604 RMSF was ~5.4 Å vs. 8.9 Å in WT (–3.5 Å difference), Ile605 ~5 Å vs. 7.8 Å (–2.8 Å), Ser606 ~4.3 Å vs. 7.1 Å (–2.8 Å). The quadruple mutant likewise showed a rigidification here

(Figure 6): RMSF at 603–606 in the quadruple was 3.8–6.4 Å, significantly lower than WT. This suggests that mutation Y639V, in particular, stabilize the conformation of the active-site loop 600–610. We hypothesize that in the wild-type enzyme this loop is inherently floppy (which might be related to the enzyme's need to undergo conformational changes during nucleotide loading), but when Y639 is replaced by valine, the loop might form alternative stable contacts. The RMSF data thus indicate an inverse relationship between the flexibility of the NTD and the flexibility of the active-site loop: the mutants make the distant NTD more mobile, but the local active-site loop less mobile.

4.1.3 Quadruple and WT at elevated temperature of 315 Kelvin

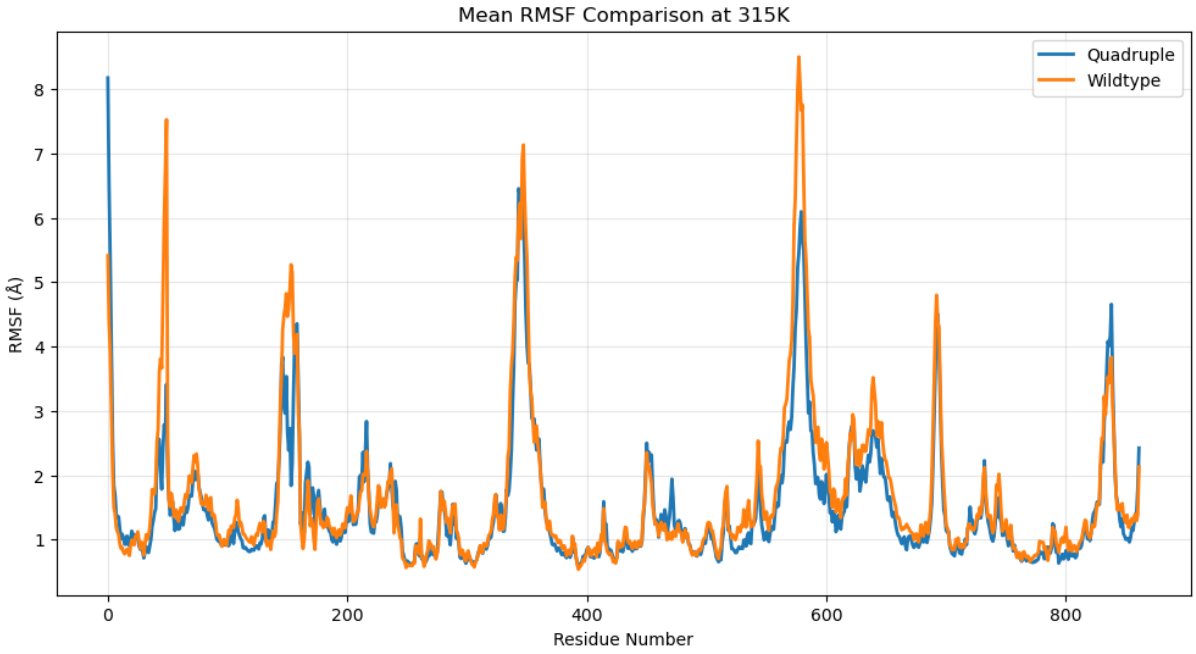

Figure 7. RMSF: Quadruple vs Wild Type at 315K.

Root-mean-square fluctuation (RMSF) analysis (Figure 7) reveals a temperature-dependent stabilising effect of the quadruple mutant on backbone dynamics. At 315 K, the mean RMSF values of all structural regions—including the intrinsically flexible N-terminal domain—are uniformly lower in the quadruple variant than in the wild-type enzyme, whereas at 310 K the same mutant exhibits an average increase of ≈1 Å relative to wild type. The extent of this high-temperature stabilisation is exemplified by two sentinel residues: Phe55, located in the N-terminal domain, fluctuates by only 3.41 Å in the mutant versus 7.53 Å in wild type, while Asp599 within the finger's region, shows a reduction from 8.50 Å (wild type) to 5.45 Å (mutant).

4.2 Hydrogen Bond Occupancy: Disruption and Formation of Key Interactions

We identified several intra-protein hydrogen bonds whose occupancies change markedly between the wild type and mutant simulations. **Figures from 8 to 21** highlight a subset of these interactions, focusing on those that are either lost (negative Δ occupancy) or gained (positive Δ) in the mutants. We organize the description below by functional regions of the polymerase to which these interactions pertain, as this provides insight into how the mutations alter the communication and stability of those regions.

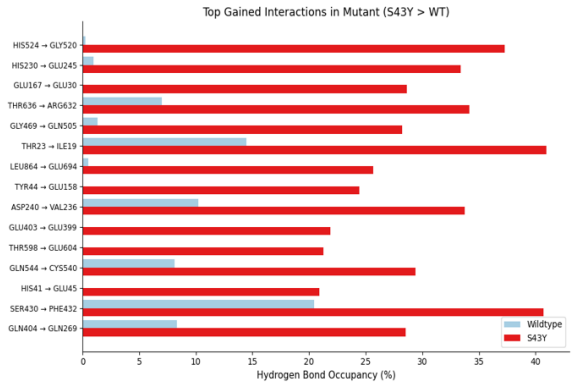

Figure 8. Top Gained H-bonds in S43Y mutant compared to WT

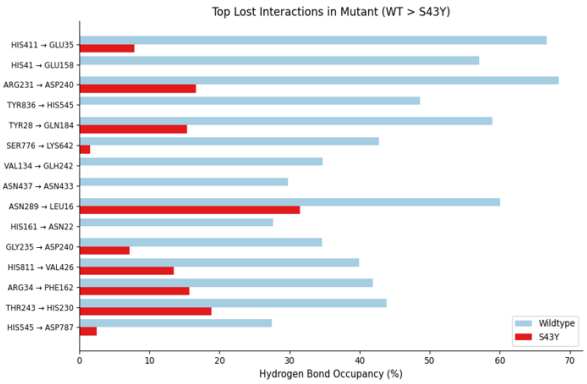

Figure 9. Top Lost H-bonds in S43Y mutant compared to WT

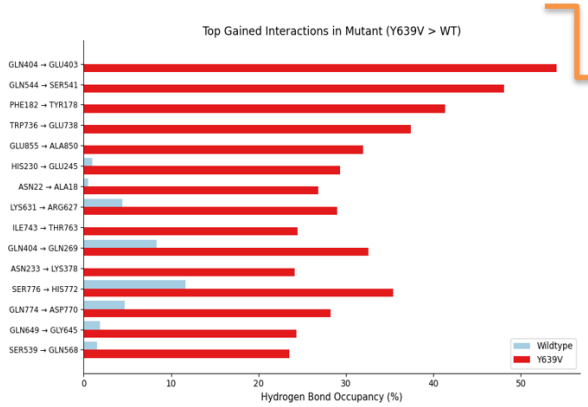

Figure 10. Top Gained H-bonds in Y639V mutant compared to WT

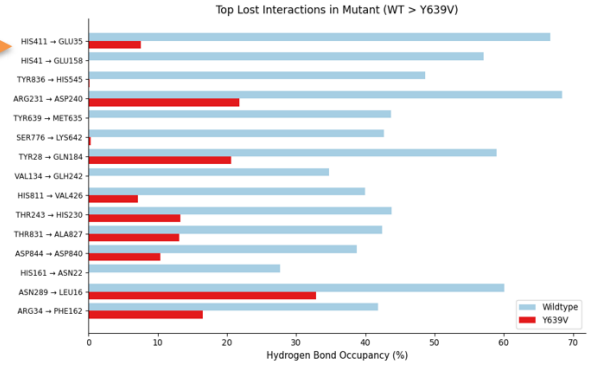

Figure 11. Top Lost H-bonds in Y639V mutant compared to WT

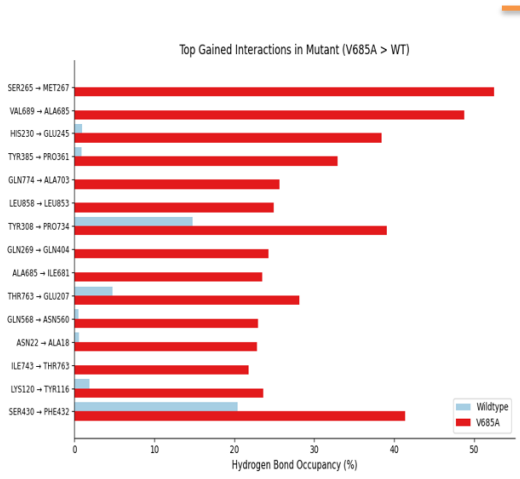

Figure 12. Top Gained H-bonds in V685A mutant compared to WT

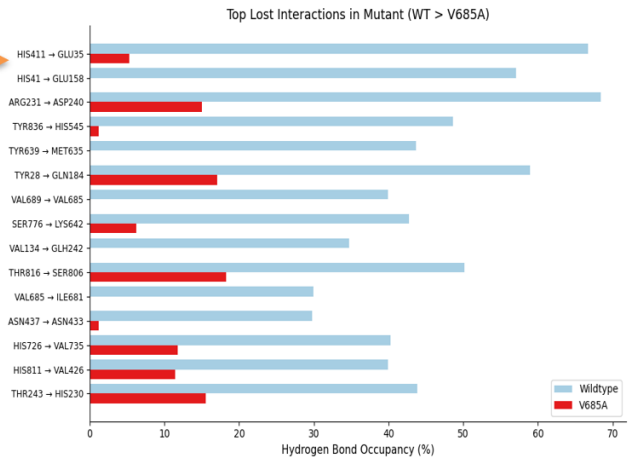

Figure 13. Top Lost H-bonds in V685A mutant compared to WT

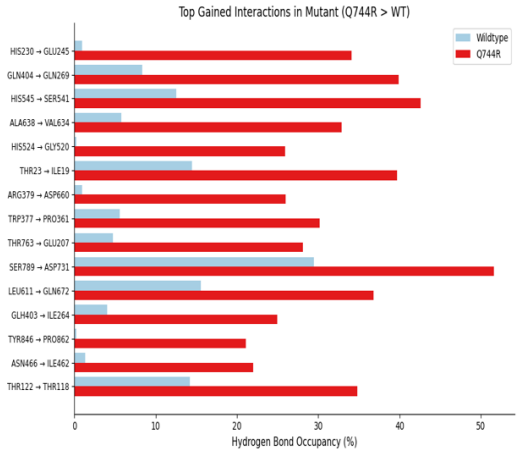

Figure 14. Top Gained H-bonds in Q744R mutant compared to WT

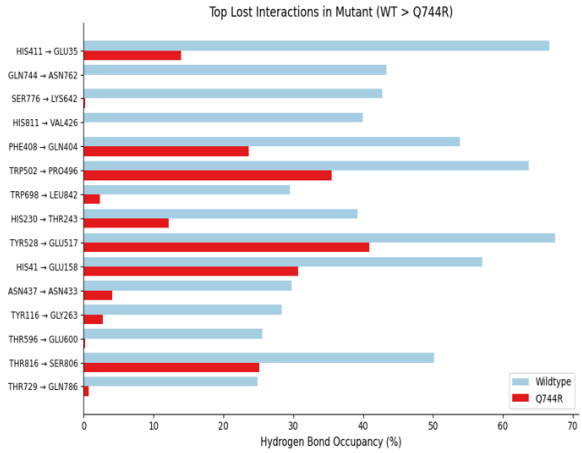

Figure 15. Top Lost H-bonds in Q744R mutant compared to WT

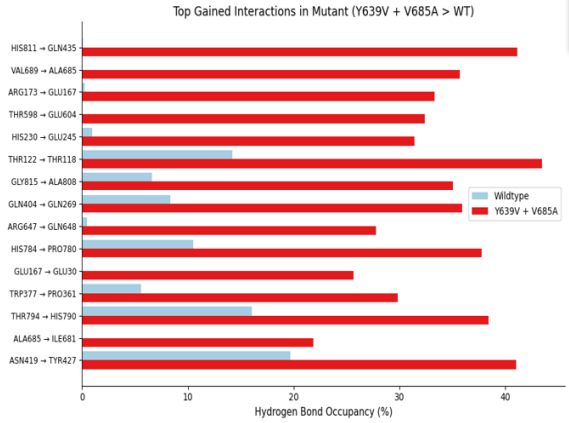

Figure 16. Top Gained H-bonds in Double mutant compared to WT

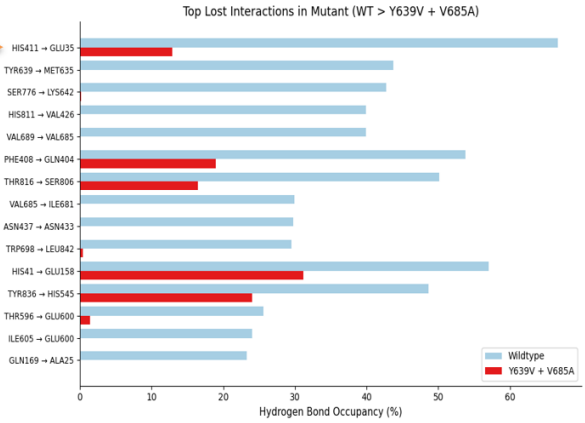

Figure 17. Top Lost H-bonds in Double mutant compared to WT

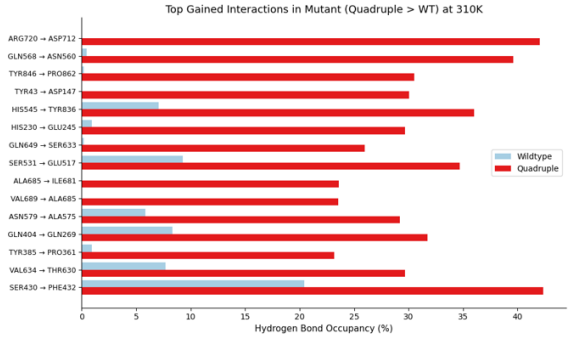

Figure 18. Top Gained H-bonds in Quadruple mutant at 37℃ compared to WT

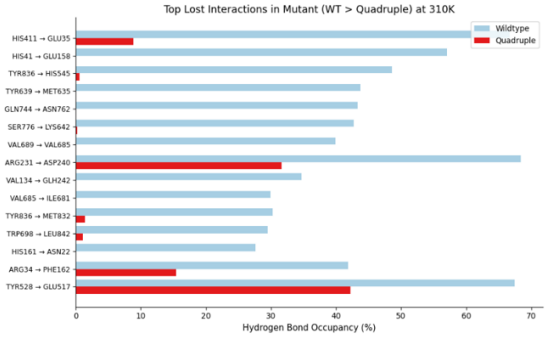

Figure 19. Top Lost H-bonds in Quadruple mutant 37℃ compared to WT

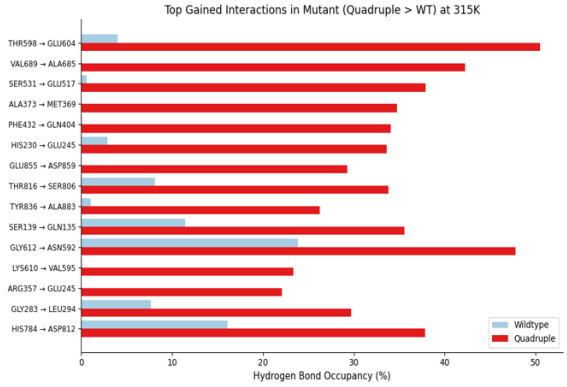

Figure 20. Top Gained H-bonds in Quadruple mutant at 42℃ compared to WT

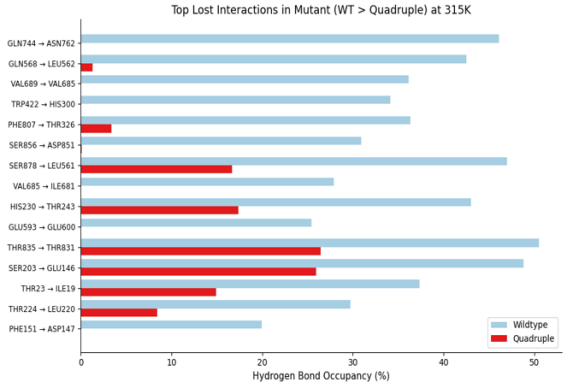

Figure 21. Top Lost H-bonds in Quadruple mutant at 42℃ compared to WT

4.2.1 NTD–Core Interface – Loss of Anchoring Contacts:

In the wild-type enzyme, the N-terminal domain is anchored to the remainder of the polymerase by a network of interactions. Chief among these is a hydrogen bond between the side chain of His411 and the side chain of Glu35. His411 is located in the vicinity of the polymerase domain's β-sheet scaffold, while Glu35 resides in the NTD near the very N-terminus. In wild-type simulations, His411–Glu35 formed a stable hydrogen bond with ~67% occupancy. Remarkably, this interaction was **completely lost in every mutant we studied**. The occupancy drops by approximately 66–67 percentage points (to essentially 0% in mutants S43Y, Y639V, V685A, Q744R, double, and quadruple). This indicates a consistent outcome: **the connection between the NTD and core provided by the His411–Glu35 hydrogen bond is broken by all these mutations**. Notably, none of the four mutations directly involves His411 or Glu35, meaning this is an allosteric effect. It appears that each mutation – whether in the NTD (S43Y) or in the distal core – triggers a slight rearrangement that destabilizes this interface. S43Y likely alters the NTD structure or its electrostatics enough to weaken the contact; Y639V and V685A, though far, induce changes in the overall protein that also propagate to this interface.Q744R by itself also led to loss of His411–Glu35, implying that even enhancing stability in one part (C-terminal) does not preserve this specific NTD link.

Another interaction linking the NTD to the core is between **His41 and Glu158**. His41 is in the very N-terminal helix, and Glu158 is in the NTD's second subdomain (sometimes considered part of the NTD that interacts with promoter DNA). This hydrogen bond (likely His41 Nδ–H…Oε Glu158) had moderate occupancy (~57% in WT). It too was **significantly reduced in most mutants**. S43Y, Y639V, V685A, and Q744R each showed a ~55–57% decrease in occupancy of His41–Glu158 (essentially lost). The quadruple mutant also lost it (Δ ≈ –57%). The double mutant, interestingly, showed this contact largely preserved with 31.2% H-bond Occupancy, but still its occupancy change was -25.9% compared to wildtype, suggesting a possible subtle interplay: Y639V alone broke it, V685A alone broke it, but together in the double, the NTD might have reoriented such that His41–Glu158 stayed hydrogen-bonded. This is an example of a non-additive effect; however, given that the dominant His411–Glu35 anchor is broken in the double, the NTD is still largely detached (His41–Glu158 is an intra-NTD contact, which could remain even if NTD swings away). In general, **the NTD's internal hydrogen bond network and its tethering to the core are disrupted by the mutations**, with His411–Glu35 being the clearest casualty across the board. The loss of these anchors explains the large increase in NTD flexibility, as the domain is no longer constrained.

4.2.2 600–610 Loop Adjacent to the Active Site: Gaining Hydrogen Bonds

In examining the hydrogen bonding patterns within the 600–610 loop, a notable interaction was identified between GLU593 and LYS610. In the wild-type (WT) system, this hydrogen bond exhibited an occupancy of 31.2%. Despite this value, it only emerged once over the course of all three simulation runs. By contrast, the Y639V mutant showed an increased occupancy of 41.5% that was consistently observed in each of the three runs. In the quadruple mutant, the occupancy reached 29.4%, though it was persistent across all simulation repeats. Similarly, the double mutant displayed an occupancy of 32.1%, again present in every run.

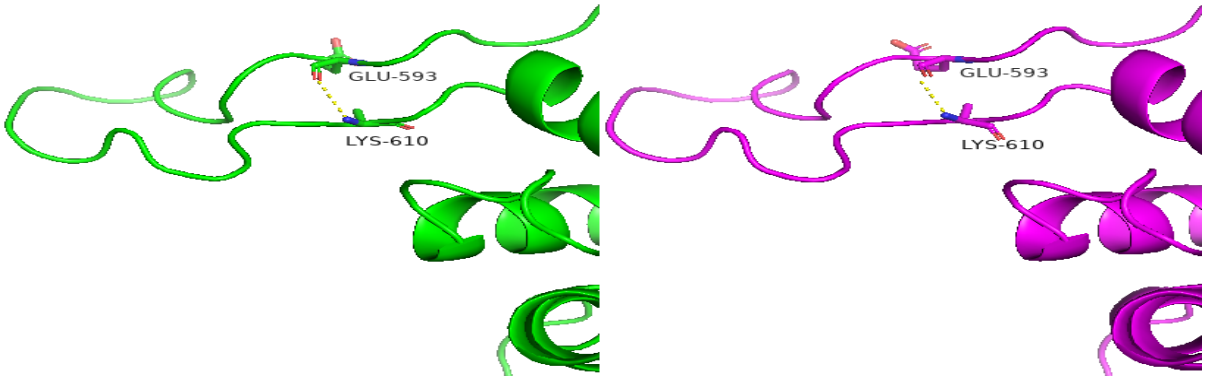

Figure 22. Left (wild) and Right (mutant): 600–610 Loop Adjacent to the Active Site

4.2.3 Hydrogen bonds over time:

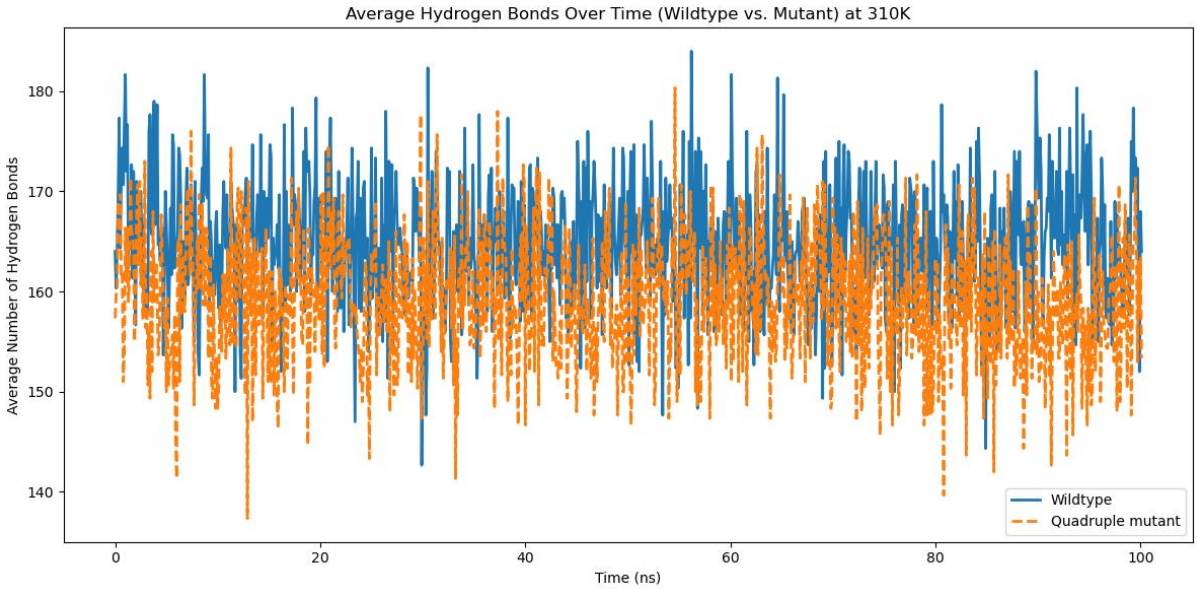

Figure 23. H-bonds over time: Quadruple and Wildtype at 37℃

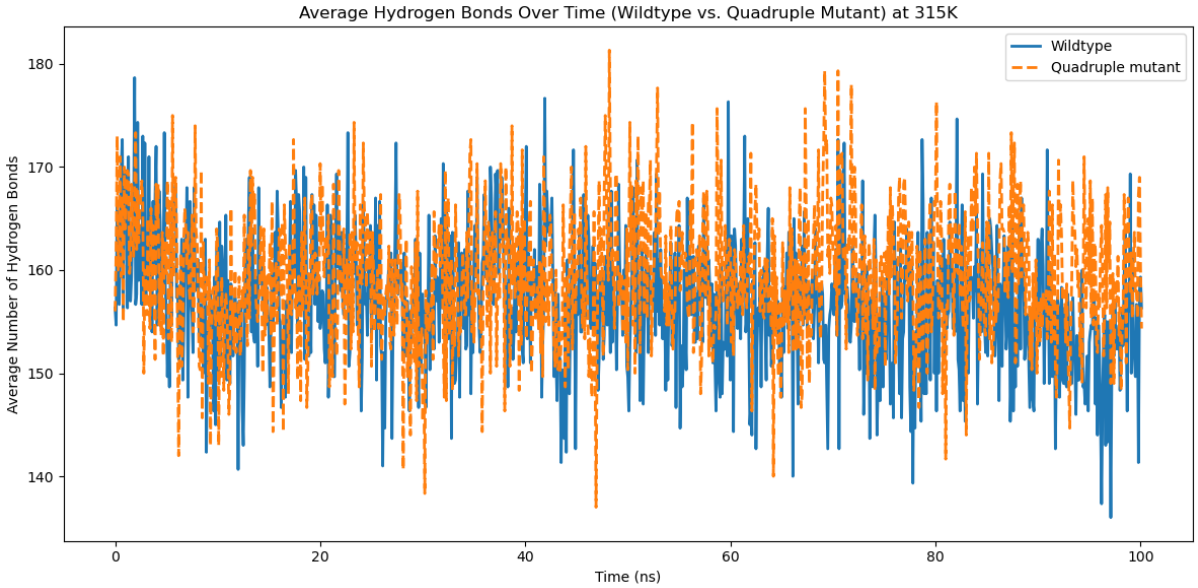

Figure 24. Quadruple and Wildtype at 42 ℃

The hydrogen bond interaction between HIP411 side chain and GLU35 side chain exhibited an occupancy of 40.6% in the wild-type at an elevated temperature of 315 K, whereas the quadruple mutant displayed a slightly higher occupancy of 42.6% at the same temperature. Interestingly, this interaction was lost entirely in all single and double mutants, as well as in the quadruple mutant at the lower temperature of 310 K. However, the quadruple mutant uniquely preserved this hydrogen bond interaction at 315 K, indicating enhanced thermostability compared to the wild-type under identical conditions.

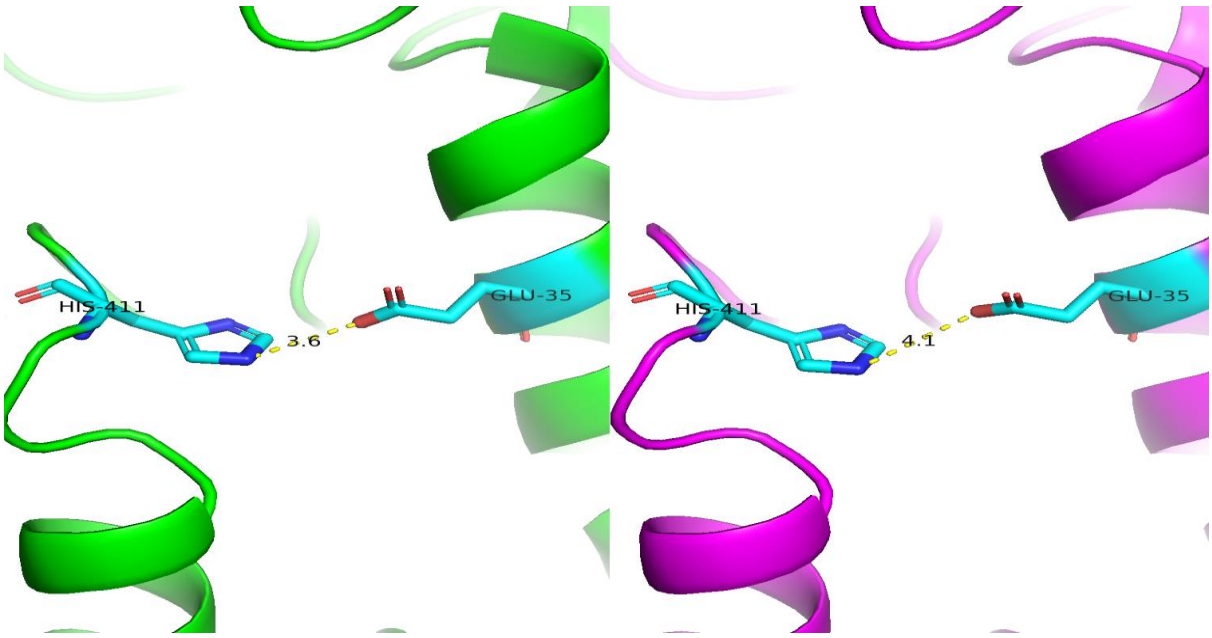

Figure 25. HIS411-GLU35 Interaction. Left (wild) and Right (Quadruple mutant) at 42°C

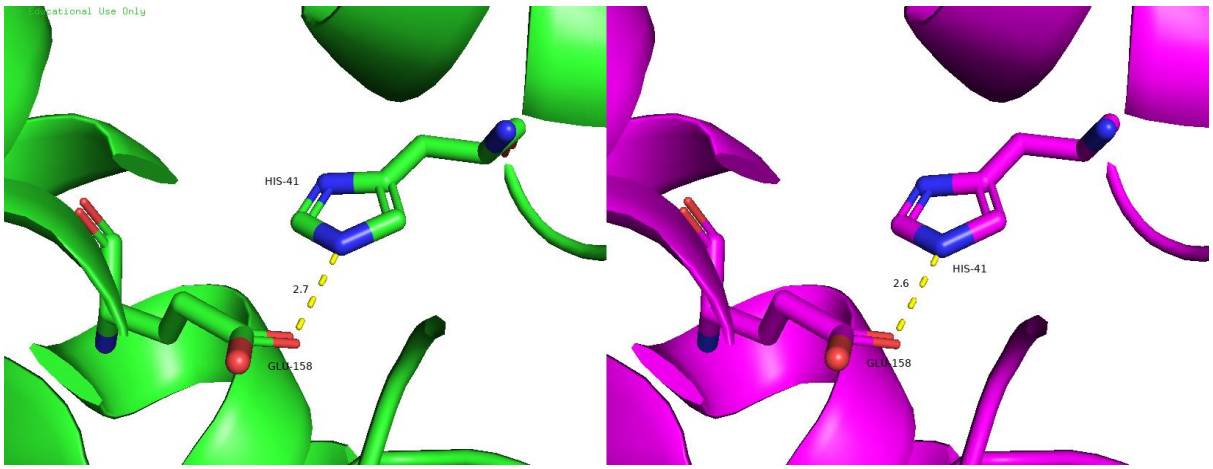

Figure 26. HIS41-GLU158 Interaction. Left (wild) and Right (Quadruple mutant) at 42°C

4.3 Backbone RMSD Analysis of Wild-Type and Mutant Variants

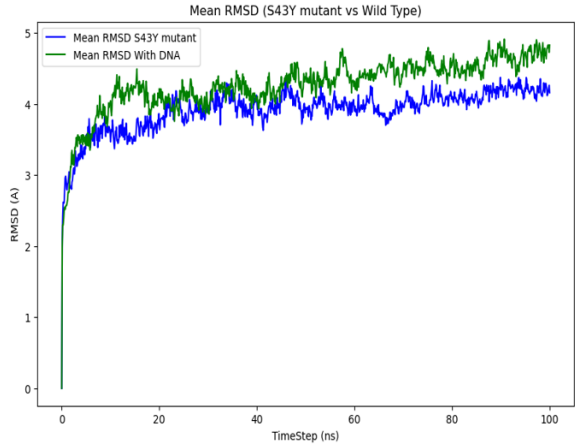

Figure 27. RMSD: S43Y vs Wild Type

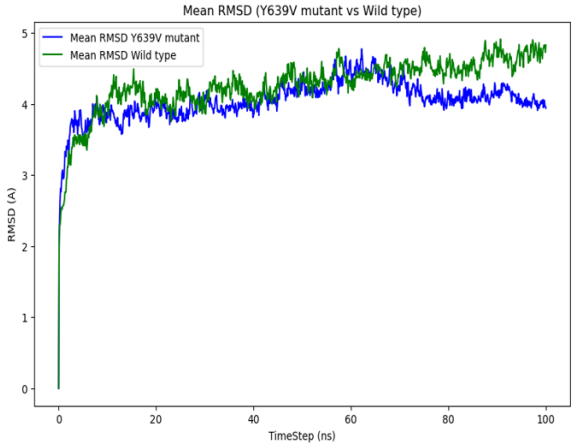

Figure 28. RMSD: Y639V vs Wild Type

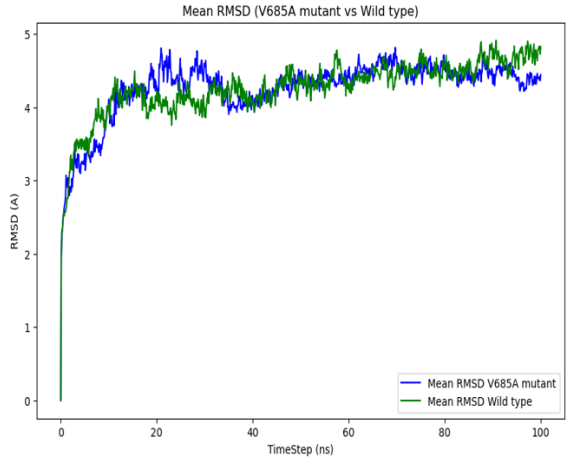

Figure 29. RMSD: V685A vs Wild Type

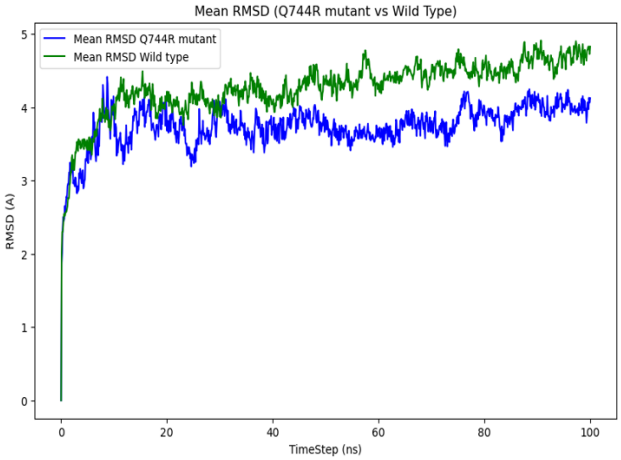

Figure 30. RMSD: Q744R vs Wild Type

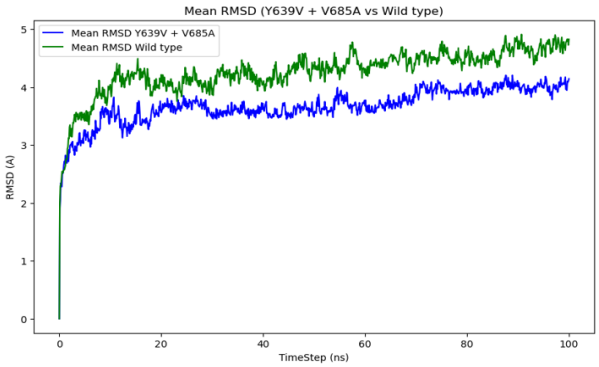

Figure 31. RMSD: Double vs Wild Type

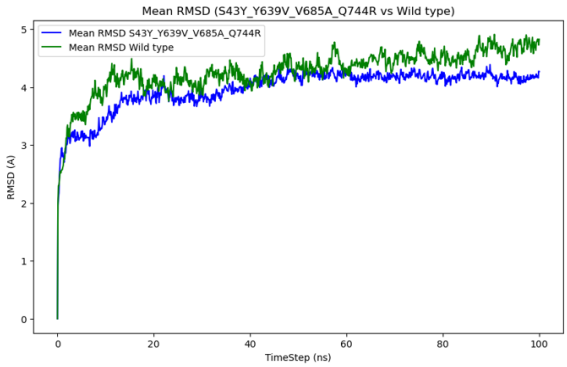

Figure 32. RMSD: Quadruple vs Wild Type

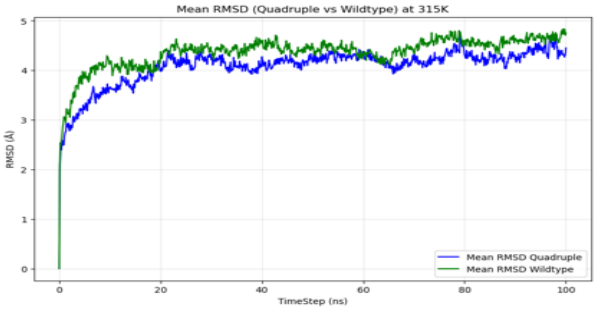

Figure 33. RMSD: Quadruple at elevated Temperature of 42 °C vs Wild Type

To gauge how firmly each protein variant holds its overall fold, we tracked the Cα RMSD throughout the 100 ns production runs (Figures 27–33). After an initial settling period of roughly 15 ns, the wild-type (WT) enzyme stabilised at an average RMSD of about 4.5 Å; this value serves as the reference point for all comparisons.

Single-site substitutions.
Each point mutation influenced backbone stability to a different extent. The S43Y variant quickly dropped to ~3.8 Å and maintained that lower value for the remainder of the simulation (Figure 27), indicating a notable stiffening of the structure. Y639V also improved stability, though more modestly, levelling off near 3.9 Å (Figure 28). By contrast, V685A behaved almost exactly like WT, converging at ~4.3 Å (Figure 29) and suggesting that the Ala replacement has little global effect. The most striking single-site change was Q744R; this mutant settled at ~3.7 Å—the lowest RMSD in the single-mutant series—implying a clear stabilising benefit (Figure 30).

Combinatorial mutants.
Pairing Y639V with V685A produced an additive effect: the double mutant flattened out at ~3.5 Å, a sharper reduction than either single change alone (Figure 31). When all four substitutions were combined, the RMSD of the quadruple mutant stabilised at ~4.0 Å after 40 ns and stayed remarkably flat thereafter (Figure 32). Although its absolute RMSD sits between the most and least stabilising single mutants, the minimal drift underscores the cumulative restraint imposed by the set of mutations. This advantage persists under mild heat stress: at 42 °C, the quadruple mutant remained within 4.0–4.3 Å from 20 ns onward, consistently outperforming the WT trajectory at the same temperature (Figure 33).

5 DISCUSSION

Our molecular dynamics simulations indicate that even single-point mutations distant from the active site can trigger significant long-range effects on T7 RNA polymerase (RNAP) structure and dynamics. In particular, all four mutations studied – S43Y, Y639V, V685A, and Q744R – alter both local flexibility and global stability of the enzyme. Across the board, we observed enhanced flexibility in the N-terminal domain (NTD, residues ~1–324) of each mutant alongside a compensatory rigidification of an active-site proximal loop (around residues 600–610 in the fingers domain). These changes suggest an allosteric rebalancing of the enzyme's conformational ensemble, biasing it away from the initiation-ready state and toward an elongation-ready state.

5.1 NTD Flexibility and Promoter Release:

A striking outcome of the simulations is the increased mobility of the NTD in all mutants. In the wild-type enzyme, the NTD is relatively rigid, held against the rest of the polymerase by a network of stabilizing contacts. Notably, a hydrogen bond between His411 (in the thumb domain) and Glu35 (in the NTD) acts as a key "anchor", helping clamp the NTD over the promoter DNA in the initiation complex. In our mutant simulations, this anchoring interaction is recurrently disrupted – for example, its occupancy drops from nearly 70% in the wild type to essentially negligible levels in S43Y, Y639V, and other variants. The loss of His411–Glu35 (and a similar contact between His41 and Glu158 in the NTD) is a consistent theme across all single mutants, indicating that each mutation, despite being located in different regions of the protein, weakens the NTD's attachment to the core. Freed from this tether, the NTD samples a much broader range of motion relative to the polymerase body. Quantitatively, residues in the NTD become far more flexible: for instance, the RMSF of Phe55 in the NTD rises from ~2.4 Å in the wild type to between 5–7 Å in the mutants, reflecting a marked increase in NTD fluctuations (see **Figure 1** and **Figure 6** for RMSF comparisons). Such heightened NTD flexibility suggests that the **closed, promoter-bound conformation** is destabilized in the mutants. Functionally, this would **facilitate promoter release** (promoter escape) because a more mobile NTD offers less resistance to the structural rearrangements needed for the polymerase to transition from initiation to elongation. In practical terms, the mutants are likely less prone to "stalling" in the initiation phase – they would spend less time gripping the promoter and more readily undergo the conformational changes (like NTD rotation or re-positioning) required to clear the promoter. This is consistent with experimental reports for some of these mutations: for example, S43Y, located in the NTD itself, has been noted to reduce abortive initiation and allow the enzyme to produce longer transcripts. Our simulations provide a mechanistic explanation for this behaviour, as the loss of NTD-contacting hydrogen bonds and the ensuing increase in NTD flexibility lower the barrier for the polymerase to relinquish its hold on the promoter DNA.

5.2 Active-Site Loop Stabilization and Elongation:

Counterbalancing the liberated NTD, we observed that a loop adjacent to the active site (roughly residues 600–610, part of the fingers domain and near the O-helix containing Tyr639) becomes more **ordered (less flexible)** in most mutant trajectories. In the wild-type simulation, this **specificity loop** is highly dynamic – consistent with crystal structures where this region is sometimes unresolved due to flexibility. For example, the RMSF of a key loop residue Glu604

is about 8–9 Å in the wild type, indicating substantial fluctuation. Upon mutation, however, this value drops significantly: in the Y639V mutant, Glu604's RMSF falls to ~4–5 Å (approximately half that of WT), and a similar reduction is seen in V685A and in the multi-mutant simulations. The **quadruple mutant** in particular shows a pronounced stabilization of this active-site adjacent loop (e.g., Glu604 RMSF ~5.1 Å vs. 8.9 Å in WT), implying that the combined mutations lock this region into a more restricted conformation. This **loop rigidification** suggests that the mutants favor a specific loop orientation – likely one conducive to the elongation phase of transcription. Structurally, Tyr639 in the wild-type enzyme must swing out (opening the O-helix) to accommodate translocation of the RNA–DNA hybrid after nucleotide addition. In the Y639V mutant, this bulky tyrosine is replaced with valine, and our data indicate the loop around residue 639 becomes more stably positioned, potentially holding the O-helix in an "open" conformation permissive for RNA chain elongation. A stabilized 600–610 loop could thus maintain the active site in an optimal geometry for **catalysis and translocation**, reducing the likelihood of off-pathway excursions. In functional terms, this would manifest as improved **processivity** – the polymerase can add nucleotides and advance along the DNA template more consistently without pausing or backtracking. Indeed, some of these mutations (e.g., Y639V/V685A) are known to enable the enzyme to incorporate unusual substrates and sustain elongation under conditions that would hinder the wild type. Our simulations support this by showing that mutants bias the enzyme's active-site region into a configuration that resembles a committed elongation complex. The **increased rigidity of the active-site loop** can be viewed as the enzyme "reinforcing" its catalytic core for continuous RNA synthesis once initiation is bypassed.

5.3 Hydrogen-Bond Network Alterations and Thermal Stability:

At the molecular interaction level, the mutations lead to **recurring changes in the hydrogen-bonding network** that correlate with the above domain motions. As mentioned, *loss of specific inter-domain hydrogen bonds* (His411–Glu35, His41–Glu158) is a unifying feature in all mutants, directly contributing to NTD detachment. In place of these lost bonds, each mutant establishes new intra- or inter-domain hydrogen bonds that likely compensate by stabilizing other regions. For example, the S43Y mutant (in the NTD) gains novel hydrogen bonds within the NTD-core interface (one prominent new interaction is His230–GlU245, ~32% occupancy), suggesting that while the NTD loosens globally, certain local contacts within the NTD are strengthened. Similarly, Y639V (in the active site loop) introduces a unique bond between Gln404 and Glu403 (~54% occupancy) in the thumb domain, potentially contributing to the rigidity of that region. The double mutant and quadruple mutant, which combine alterations in both NTD and active-site regions, show a mixture of these new interactions. Notably, in the **quadruple mutant** we observe hydrogen bonds that **stabilize the active-site loop** itself – for instance, a Thr598–Glu604 interaction (occupancy ~50%) emerges in the quadruple mutant. This particular bond, absent in the wild type, bridges the base of the 600–610 loop to a neighbouring turn, acting like an internal brace for the loop. The accumulation of such new interactions in multi-mutant T7 RNAP likely underpins its robust structure. Indeed, when we challenged the systems at **elevated temperature (315 K, ~42 °C)** to assess thermostability, the quadruple mutant retained a stronger hydrogen-bond network compared to wild type (Bhadwal *et al*, 2023). **Figure 24** summarizes the average hydrogen bonds over time at 42 °C: the wild-type enzyme shows a drop in hydrogen-bond count (down to an average of ~156 bonds at 42 °C from ~164 at 37 °C), indicating some loss of stabilizing interactions with heat. By contrast, the quadruple mutant at 42 °C maintains an average of ~160 bonds, essentially matching its room-temperature network. In other words, the quadruple variant manages to preserve key

interactions even under thermal stress, whereas the wild type suffers a noticeable loss. Importantly, some of the **"anchor" contacts** that the wild type relies on (like His411–Glu35) were completely broken in the wild type at high temperature, but in the quadruple mutant **they persisted at 315 K** (observable in **Figure 25-26**). This indicates that the combined mutations endow the polymerase with a degree of **thermal resilience** – likely attributable to the compensatory hydrogen bonds and the overall tightened architecture of the enzyme. The Q744R mutation, in particular, is known for improving thermal stability, and in the context of the quadruple mutant it appears to fortify the enzyme such that even at 42 °C the NTD-anchor interactions can re-form or persist to some extent (Yang *et al*, 2023). Taken together, the hydrogen-bond analysis reinforces our flexibility findings: **mutations redistribute the internal bonding network**, loosening contacts at the NTD–core interface while adding or strengthening contacts in the core and active-site regions. This redistribution not only shifts the enzyme toward an elongation-ready conformation at normal temperature but also helps the enzyme **resist thermal perturbation**, which could be beneficial for function at industrially relevant temperatures.

Our simulations paint a coherent picture of how the four point mutations – individually and especially in combination – remodel the T7 RNAP. The **global stability** measurements (RMSD) show that most mutants are as stable as or more stable than the wild type enzyme in simulation, despite their greater flexibility in certain parts. For instance, the double mutant Y639V/V685A exhibited a substantially lower equilibrium RMSD (~3.5 Å) than wild type (~4.5 Å), and the RMSD of the quadruple mutant remained consistently below that of wild type over long simulations (**Figure 32. RMSD: Quadruple mutant vs Wild Type**), indicating an overall stabilizing effect of combining mutations. Meanwhile, the **local flexibility profiles** (RMSF) highlight a trade-off: the NTD becomes more mobile while the active-site vicinity becomes more rigid in the mutants relative to WT. This trade-off aligns perfectly with the functional demands of transitioning from **transcription initiation to elongation**. By weakening the NTD's hold (easing promoter escape) and strengthening the active site configuration (promoting processive elongation), the mutations collectively push T7 RNAP toward a state that favors efficient RNA synthesis. These findings reveal the power of allosteric effects in a single-subunit RNA polymerase – mutations in one part of the enzyme can reverberate through the structure, ultimately tuning the balance between different functional states. Such insights deepen our understanding of T7 RNAP's mechanochemical cycle and set the stage for leveraging these mutations in biotechnological applications (Fang *et al*, 2024).

Conclusion

In this work, we have investigated how four specific point mutations – **S43Y, Y639V, V685A, and Q744R** – impact the structure and stability of T7 RNA polymerase, both individually and in combination. Using molecular dynamics simulations, we found that these mutations induce a concerted shift in the enzyme's dynamics: the N-terminal domain becomes more flexible, while critical active-site loops become more stable. This inverse change in flexibility is accompanied by remodeling of the hydrogen-bond network. Key inter-domain contacts that lock the polymerase in an initiation-ready conformation (such as the His411–Glu35 bridge anchoring the NTD) are lost or greatly weakened in the mutants, whereas new intra-domain bonds (for example, within the fingers/palm region) are formed to stabilize the elongation-ready state. Collectively, these mutation-induced changes tilt the balance of T7 RNAP toward a more **elongation-prone conformation** even in the promoter-bound form. The mutants effectively mimic aspects of the natural transition from initiation to elongation: a looser grip on promoter DNA and a readied active site for RNA chain extension.

The mechanistic implications of these findings are significant. **Promoter escape** appears to be facilitated by mutations like S43Y and Y639V, which destabilize the NTD clamp and thereby reduce the tendency for abortive initiation. A more mobile NTD likely lowers the energy barrier for the polymerase to break its contacts with the promoter, allowing it to clear the promoter region more readily. Meanwhile, **transcription elongation** is potentially enhanced by mutations such as Y639V and V685A that stabilize the active-site configuration. By favoring an "open" conformation of the O-helix/active-site loop and maintaining crucial catalytic geometry, these mutations can improve the enzyme's **processivity** (its ability to synthesize long RNA products without dissociating or stalling). The net effect is an increase in overall **transcription efficiency** – the mutants are poised to initiate less tentatively and elongate more smoothly. This provides a plausible explanation for experimental observations (from prior studies) that some of these mutations lead to higher yields of full-length transcripts and reduced early termination. In essence, our results connect the **molecular flexibility changes** to functional outcomes: a flexible NTD correlates with faster promoter release, and a rigidified active site correlates with vigorous, continuous RNA synthesis.

Future Directions: Building on these findings, several future lines of inquiry are promising. Free Energy Perturbation (FEP) calculations would be a valuable next step to quantitatively assess the energetic impact of each mutation on the stability of specific states (initiation complex vs. elongation complex). FEP or related free-energy methods could elucidate which mutations contribute additively or synergistically to lowering the free energy of the elongation-ready conformation, providing a thermodynamic perspective to complement our structural observations. Additionally, further MD simulations could be performed with **modified promoter DNA or different transcription contexts** – for example, including a partially synthesized RNA transcript or using a longer promoter bubble – to see how the mutations influence the actual process of promoter melting and escape. Such simulations would help bridge the gap between our current *promoter-bound starting state* and the later stages of transcription, giving a fuller picture of the mutants' effects on the entire initiation-to-elongation transition. Another interesting extension would be to simulate these T7 RNAP variants in complex with atypical NTP substrates or at even higher temperatures, to explore the limits of their substrate tolerance and thermostability. Finally, integrating our computational approach with experimental validation would firmly establish the causal link between the observed dynamic changes and transcriptional outcomes.

In conclusion, this thesis work enhances our mechanistic understanding of how specific mutations can modulate the performance of T7 RNA polymerase. By correlating **atomic-level dynamics** (RMSD/RMSF changes and hydrogen-bond patterns) with **biochemical function**, we have shown that strategic alterations to the enzyme's sequence translate into a more initiation-liberated, elongation-efficient polymerase. These insights have practical implications for biotechnology and medicine. T7 RNAP is a workhorse enzyme in applications ranging from **mRNA vaccine production** to synthetic biology assemblies for RNA synthesis. The ability to **design custom polymerase variants** with improved traits – such as higher stability at fever temperatures, reduced tendency to stall, or capability to incorporate modified nucleotides – could greatly advance these fields. For example, a thermostable, hyperactive T7 RNAP variant could streamline the manufacturing of RNA vaccines by sustaining high transcription rates even under suboptimal conditions, ensuring higher yields of mRNA. Similarly, polymerases harboring mutations like Y639V/V685A (which broaden substrate range) can be harnessed to synthesize RNAs containing nucleotide analogues for therapeutic or research purposes (such as 2′-modified RNAs that wild-type T7 RNAP cannot efficiently produce). In a broader sense, our findings contribute to the toolkit for **enzyme engineering**: by pinpointing regions like the

NTD hinge and the active-site loop as hotspots for influencing function, we provide rational targets for developing new T7 RNAP variants tailored to specific real-world tasks. Ultimately, this research highlights how **small changes** (point mutations) can have outsized effects on an enzyme's behavior, and how understanding these effects enables us to push the boundaries of what a natural enzyme can do in service of innovative applications.

6 REFERENCES

1. Bhadwal V, Santner T & Marx A (2023) Engineering the Y639V/V685A double mutant of T7 RNA polymerase for efficient 2′-modified RNA synthesis. *Nucleic Acids Res* 51, 441–456.
2. Cheetham GMT, Jeruzalmi D & Steitz TA (1999) Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80–83.
3. Cheetham GM & Steitz TA (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309.
4. Dunn JJ & Studier FW (1983) Complete nucleotide sequence of bacteriophage T7 DNA and location of T7 genetic elements. *J Mol Biol* **166**, 477–535.
5. Fang Y, Jones S & Pavlov MY (2024) Engineered T7 RNA polymerases in mRNA vaccine manufacturing. *Trends Biotechnol* 42, 210–220.
6. Jeruzalmi D & Steitz TA (1998) Structure of the T7 RNA polymerase–T7 lysozyme complex predicts a mechanism of inhibition. *EMBO J* **17**, 4101–4113.
7. Moffatt BA & Studier FW (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221–227.
8. Siegmund V, Santner T, Micura R & Marx A (2012) Screening T7 RNA-polymerase mutant libraries for increased acceptance of 2′-modified nucleotides. *Chem Commun* **48**, 9870–9872.
9. Sousa R & Padilla R (1995) A mutant T7 RNA polymerase acts as a DNA polymerase. *EMBO J* **14**, 4609–4621.
10. Wang W *et al.* (2018) Bacteriophage T7 transcription system: an enabling tool in synthetic biology. *Biotechnol Adv* **36**, 2129–2137.
11. Wu H *et al.* (2021) A single mutation attenuates both transcription termination and RNA-dependent RNA polymerase activity of T7 RNA polymerase. *RNA Biol* **18**, 481–491.
12. Yang X, Liu G, Yao W, Chen Y *et al.* (2023) Thermostable Q744R variant of T7 RNA polymerase retains activity at elevated temperatures. *J Mol Biol* 435, 168194.
13. Yin YW & Steitz TA (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. Science 298, 1387-1395.