UNIVERSITY OF CALIFORNIA SAN DIEGO

Transcription on a Hydrophobic Unnatural Base Pair by T7 RNA Polymerase

A Thesis submitted in partial satisfaction of the requirements

for the degree Master of Science

in

Chemistry

by

Haoqing Xu

Committee in charge:

Professor Dong Wang, Chair
Professor Tatiana Mishanina
Professor Navtej Toor

2023

The Thesis of Haoqing Xu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

RNAP	RNA polymerase
UBP	Unnatural base pair
Pol II	RNA polymerase II
EC	Elongation complex
PIC	Preinitiation complex
GTF	General transcription factor
mRNA.	Messenger RNA

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Dong Wang for his support as the chair of my committee. I would also like to express my gratitude to Dr. Juntaek Oh, Dr. Peini Hou and Dr. Qingrong Li for their research mentorship and guidance in writing this thesis. This work was supported by National Institutes of Health (R01 GM102362 and GM148476 to D.W.).

Chapter 2 is co-authored with Dr. Oh, Juntaek; Dr. Kimoto, Michiko; Chong, Jenny; Dr. Hirao, Ichiro and Dr. Wang, Dong. The thesis author was the primary author of this chapter.

Chapter 3 is co-authored with Dr. Oh, Juntaek; Dr. Kimoto, Michiko; Chong, Jenny; Dr. Hirao, Ichiro and Dr. Wang, Dong. The thesis author was the primary author of this chapter.

ABSTRACT OF THE THESIS

Transcription on a Hydrophobic Unnatural Base Pair by T7 RNA Polymerase

by

Haoqing Xu

Master of Science in Chemistry

University of California San Diego, 2023

Professor Dong Wang, Chair

Bacteriophage T7 RNA polymerase (T7 RNAP) is frequently used for RNA synthesis with unnatural base pairs (UBPs) and specific synthetic alterations in a wide range of biotechnological and medicinal applications. As a knowledge gap in the area, however, the molecular basis for recognition and processing of UBPs by T7 RNAP during transcription remains poorly understood. We explored how the hydrophobic Ds–Pa pair is recognized and processed as a third base pair by T7 RNAP during transcription elongation. T7 RNAP integrates DsTP opposite Pa with great efficiency, equivalent to that of natural nucleotides, although the kinetics of PaTP incorporation opposite Ds is significantly slower. Using structural biology approach, we discovered that T7 RNAP recognizes unnatural substrates differently than its natural substrates. We found distinct unnatural nucleoside triphosphate binding sites for PaTP and DsTP at the pre-insertion state. We identified several separate-of-function mutants of T7 RNAP that affect UBP transcription selectively but not normal nucleic acid transcription. These results provide molecular insights into the recognition of UBP transcription by T7 RNAP. Our investigations also provide important information for the creation of the next generation of UBPs for efficient transcription and other applications.

CHAPTER 1

1.1 Introduction to transcription

Transcription is an essential step in gene expression and it is carefully regulated. Changes of transcription of specific genes can have significant effects on an organism's development and function[1].

The transcription is catalyzed by RNA polymerase (RNAP). During transcription, RNA polymerase moves along the template strand from 3' to 5' direction and synthesizes the complementary RNA chain from 5' to 3' direction by adding new nucleotides to the 3' end of RNA[2].

Transcription can be briefly divided into three stages: initiation, elongation and termination[3]. Take RNA polymerase II (Pol II) as an example of eukaryotic transcription. At initiation stage, a preinitiation complex (PIC) containing Pol II and general transcription factors (GTFs) is assembled around promoter sequences (Fig. 1[4]). The promoter DNA is unwound by PIC to allow *de novo* RNA synthesis. Pol II then escapes from the promoter and keeps on extending its pre-messenger RNA (pre-mRNA) product and enters productive elongation stage. At elongation stage, Pol II recruits elongation factors and splicing factors that are responsible for the co-transcriptional maturation of the RNA products. At termination stage, the pre-mRNA is polyadenylated and cleaved and Pol II is released from DNA template. A mature mRNA transcript is then transported out of the nucleus and translated into proteins by ribosome in the cytoplasm.

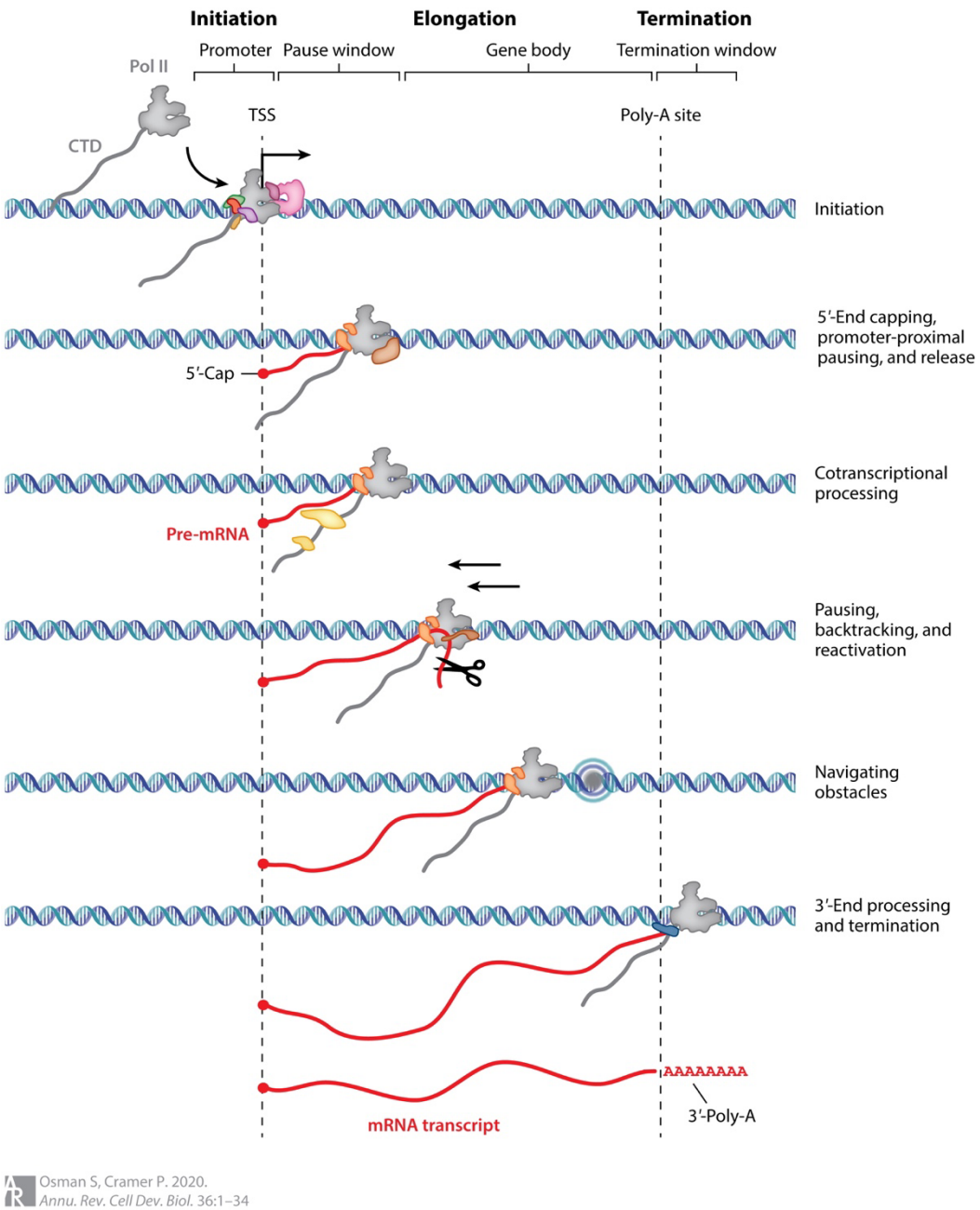

Figure 1: Pol II transcription cycle. Pol II transcription follows a defined cycle that starts with initiation at promoter sequences; in this cycle, Pol II and the GTFs TBP, TFIIA, TFIIB, TFIIE, and TFIIH assemble to coordinate the unwinding of DNA and the initiation of RNA chain synthesis. Following initiation, Pol II escapes from the promoter to extend its RNA product. During elongation, cotranscriptional processes such as RNA 5′-end capping, promoter-proximal pausing and pause release, backtracking and reactivation, and navigating obstacles take place. Finally, Pol II terminates the transcription cycle, releasing a mature mRNA transcript. This figure is cited from Osman, S.; Cramer, P. Structural biology of RNA polymerase II transcription: 20 years on. *Annual Review of Cell and Developmental Biology* 2020, *36*, 1-34.[1]

1.2 T7 RNA polymerase

There are two distinct RNA polymerase superfamilies, Single-subunit RNAPs and multi-subunit RNAPs. Single-subunit RNAPs include mitochondrial RNAPs and bacteriophage single-subunit, are part of the single subunit 'right-handed' RNAP superfamily. All cellular muti-subunit RNAPs belong to second superfamilty with an active site created at the junction of two double-barrel motifs, are members of two-ψ β-barrel polymerases[5].

T7 RNA polymerase (T7 RNAP) belongs to single-subunit 'right-handed' RNA polymerase superfamily. T7 RNAP is encoded in the bacteriophage T7 genome. It is a highly specific enzyme that only recognizes T7 promoter and transcribes the T7 phage DNA, and it does not bind to or transcribe host cell DNA. T7 RNA polymerase consists of 883 amino acids and has a molecular weight of approximately 98 kDa.

T7 RNAP is composed of three main subdomains (Fig. 2): thumb subdomain, palm subdomain, and finger subdomain[3]. The thumb subdomain consists of residues 330–410 and forms a long α-helix on one side of the template-binding cleft. Structural and function analyses suggest that the thumb subdomain stabilizes the transcription complex and maintains its processivity during transcript elongation. Part of thumb domain (residues 345–385) is conformationally disordered in apo RNAP and this part of thumb domain becomes organized when T7 RNAP forms a complex with the promoter DNA[6].

The palm subdomain comprises the most catalytically important residues (386–448, 532–540, and 788–838). Although T7 RNAP is highly α-helical in structure, the surface of the palm subdomain facing the DNA and RNA chain is largely β-sheet. A β-hairpin is formed by residues 805–817, and D812, which coordinate the catalytic magnesium ions, lies at the tip of it. Along with D537 in another β-strand-turn helix, both of conserved aspartic acids are the most

catalytically critical residues in active site. These two aspartic acids (D537 and D812) coordinate the two magnesium ions, which catalyze the formation of phosphodiester bonds in the now widely accepted "two-metal" mechanism[7]. The metal ions stabilize the negatively charged penta-covalent transition state centered on the α-phosphate, and the negative charge on the 3'-OH increases its nucleophilicity.

The finger subdomain makes up the majority of contacts between the template base and the template strand immediately downstream of the template base and incoming NTP. These contacts are produced by residue 636–640, the N-terminal residues 642–654, and the residues that connects these helices. This part of finger subdomain (627–640) corresponds to the precision of interaction between the polymerase and the template/substrate base pair.

Unlike muti-subunit cellular RNAPs, which requires other transcription factors to recognize the promoter region, T7 RNAP is able to identify and bind to the T7 promoter itself during transcription. After binding to the promoter, the enzyme unwinds the DNA double helix and initiates the synthesis of a complementary RNA molecule in a DNA template-dependent manner[2]. RNA synthesis occurs at the active site of the N-terminal domain. The RNA molecule is released by the RNA exit channel in the C-terminal domain as it is produced.

Due to its high selectivity and effectiveness at synthesizing specific RNA molecules, T7 RNA polymerase is a valuable tool in molecular biology and biotechnology[8]. It has been used extensively in the synthesis of large amounts of RNA *in vitro* for a variety of applications, including the production of messenger RNA (mRNA) for protein expression, the synthesis of

ribozymes and other RNA enzymes, all of which are crucial components of the gene expression process.

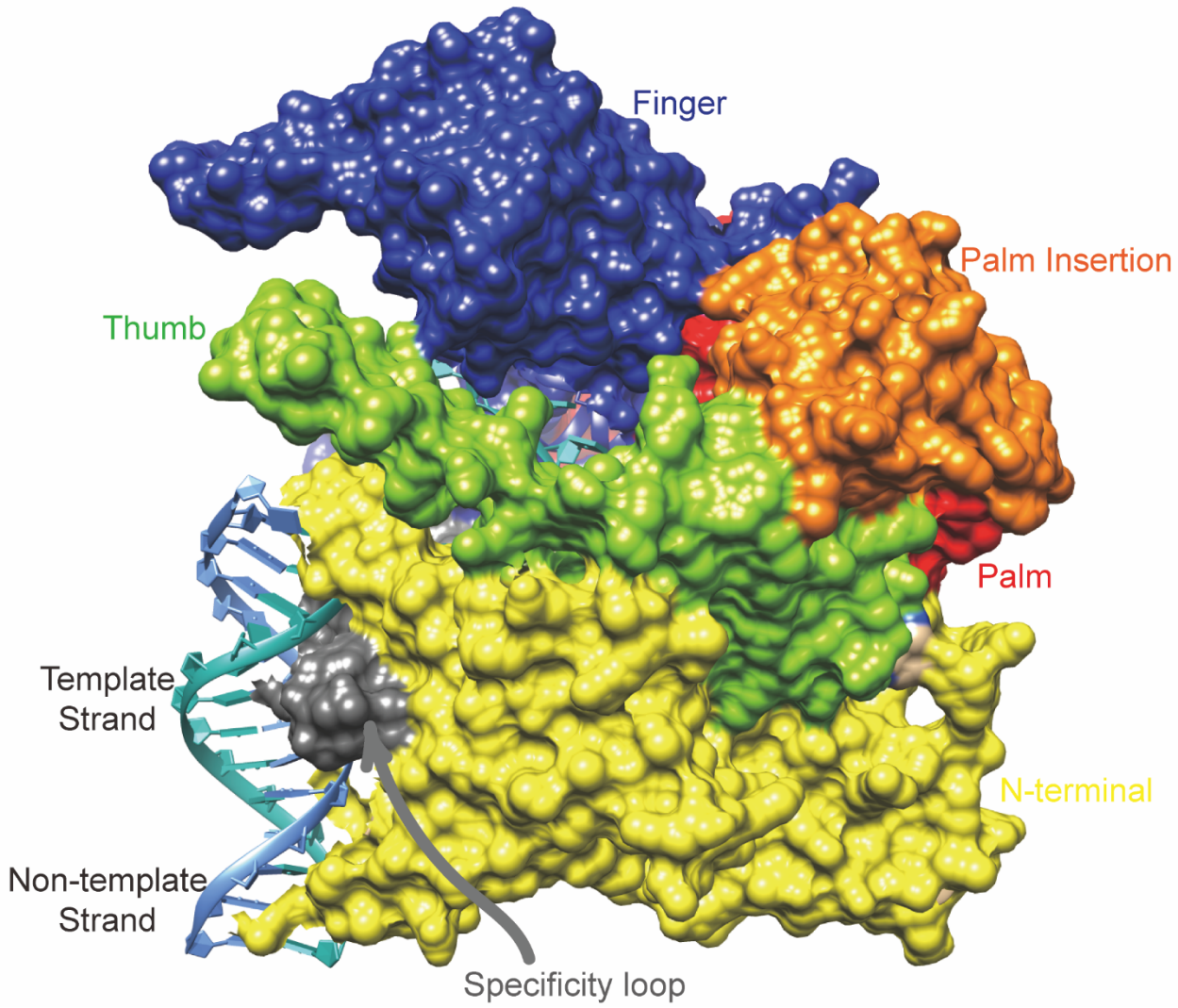

Figure 2: Overall surface model of a T7 RNA polymerase complex. The thumb, palm, finger, palm insertion, N-terminal, and the specific loop are labelled as green, red, blue, orange, yellow and grey, respectively. The non-template strand is labeled blue, and template strand is labelled green. This figure is generated from PDB file 1QLN.

1.3 Transcription of T7 RNA polymerase

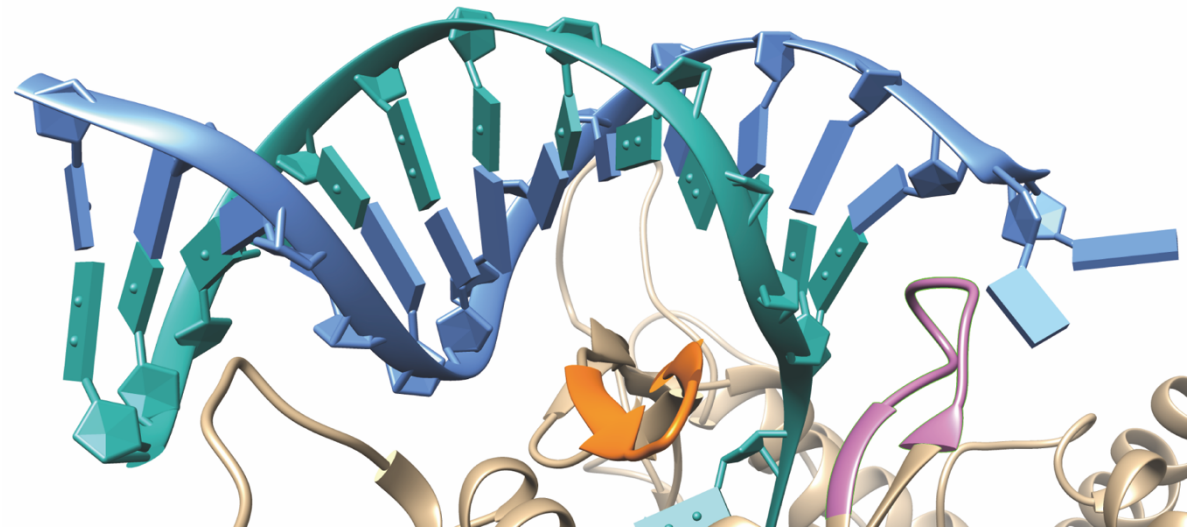

Figure 3: A promoter recognition region is shown here. The beta hairpin is labelled orange, and the intercalating loop is labelled pink.

Similar to other RNA polymerases, the transcription process of T7 RNA polymerase can be divided into three stages, initiation, elongation and termination[3].

Promoter recognition is the first step of initiation[9]. The consensus sequence for the most active Class III T7 promoters consists of 17 base pairs upstream and 6 base pairs downstream of the transcription start point[10]. This promoter (Class III) is sequence-specifically recognized by the sequence between -5 and -17. As shown in Fig. 3, a protein beta hairpin from the C-terminal domain rests on the N-terminal platform and makes contacts with the central major groove from about -7 to -11[3]. Residues 232 to 242, which is also known as the intercalating hairpin, inserts between the template strand and non-template strand. They help unwind the DNA duplex and stabilize the transcription bubble by Val 237 on the tip of the hairpin stacking on the -5 base pair[3]. When the initial transcription bubble forms, T7 RNAP will begin to search for cytosine and start forming the first phosphodiester bond. The preference of GTP at this stage may relate to the hydrogen bond between His 784 and 2'-amino group on GTP[3].

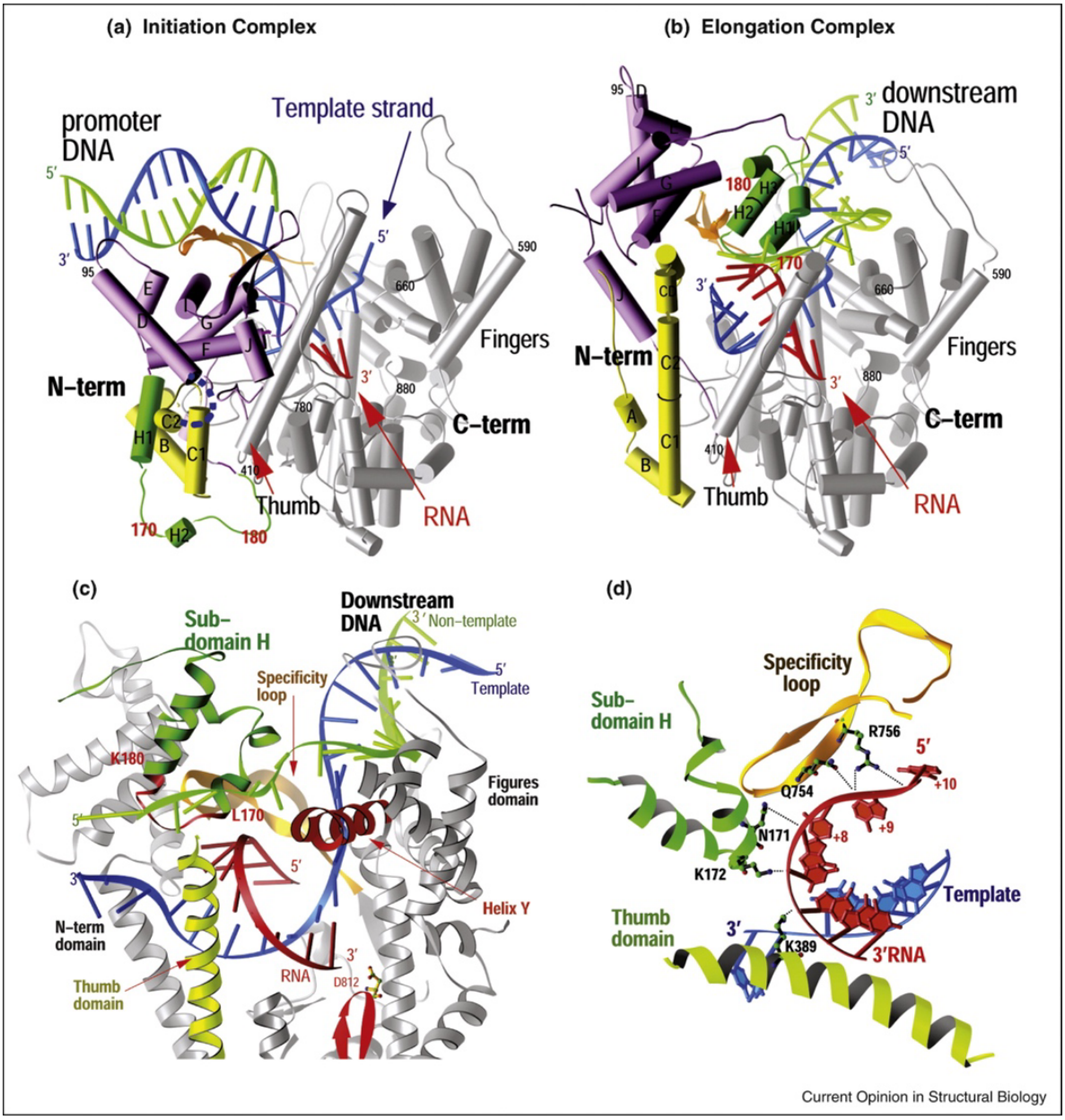

Figure 4: Comparison of the structures of the T7 RNAP initiation and elongation complexes (A) and (B) and views of the transcription bubble (C) and (D). This figure is cited from Yin, Y. W.; Steitz, T. A. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **2002**, *298* (5597), 1387-1395.[11]

The initial transcription bubble only has room for 3 base pairs, whereas the normal transcription bubble has a length of 8 to 9 base pairs[11]. Thus, a large conformational change should occur in this process. Comparing the structure of initiation and elongation (Fig. 4) revealed that the N-terminal domain may be regarded of as consisting of three subdomains. In

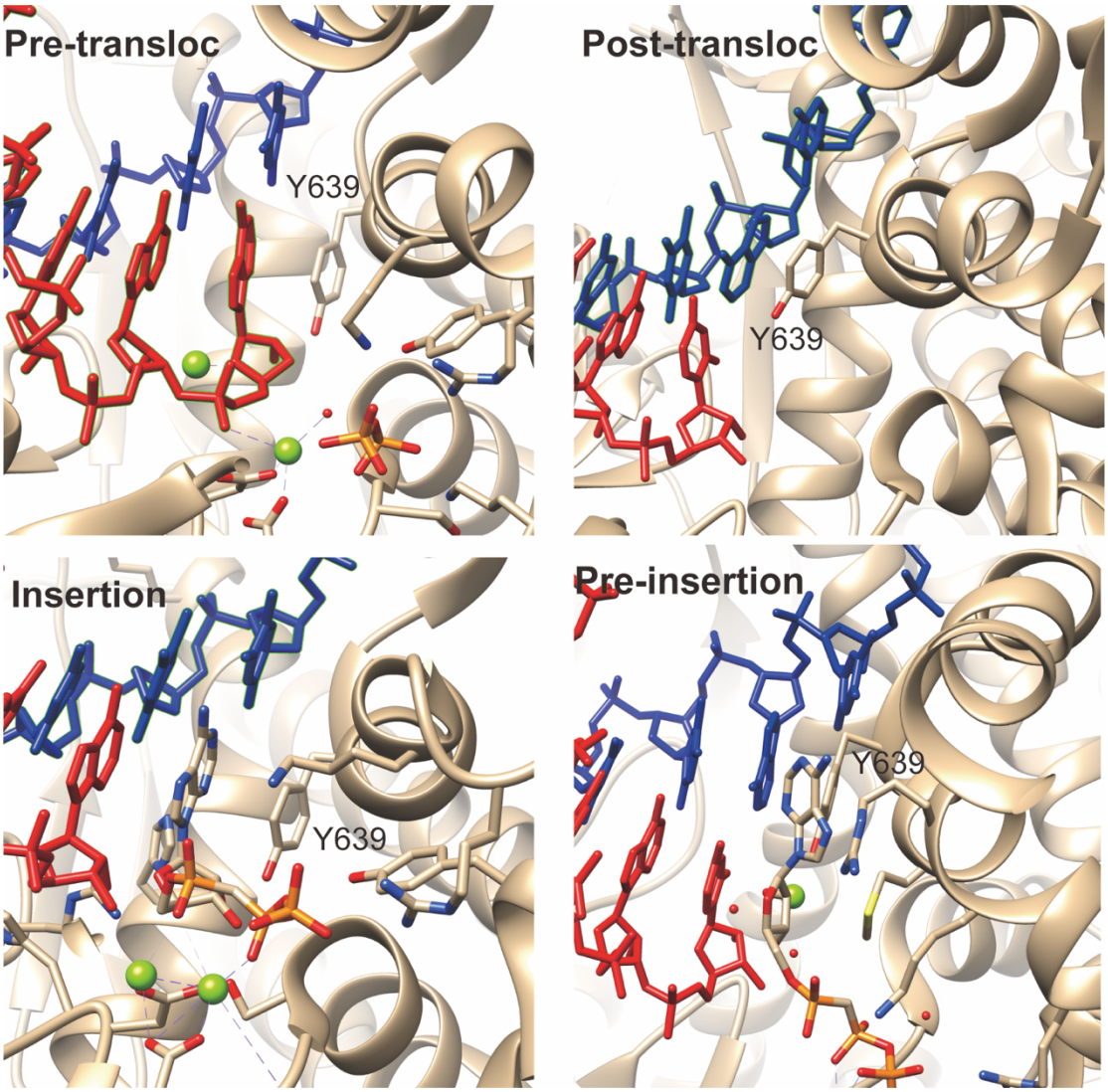

Figure 5: 4 stages of transcription elongation complex are shown here. Pre-translocation stage is a closed state from PDB: 1S77. Post-translocation state is an open state, made from PDB: 1H38. Pre-insertion is an open state, made from PDB: 1S0V. Insertion is a close state, made from PDB: 1S76.

particular, an assembly of six helices rotates 220° and shifts 30 Å as a rigid body, creating space for the elongating product and positioning the subdomain in the area formerly occupied by the promoter DNA. A loop of protein that moves from one side of the N-terminal domain to the opposite side of the 6-helix subdomain and refolds into a two-helix H domain, allowing the RNA product to exit, is perhaps the most remarkable structural arrangement observed during this

transition. The resulting N-terminal domain lacks a specificity loop, rendering the elongation complex incapable of recognizing promoter DNA and sequence-insensitive[11].

During transcription elongation, the enzyme adds a base per nucleotide addition cycle, then translocates (moves forward) along the DNA in preparation for next round of nucleotide addition[12]. In the pre-translocated state shown in Fig. 5, the nucleotide addition site is occupied by newly added RNA 3'-end. To allow next nucleotide addition, the enzyme proceeds one base forward to reach a post-translocated state. The substrate NTP initially binds to an "open" conformation of the enzyme at a location on the O helix[11], termed as pre-insertion state. At pre-insertion state, the Y639 residue of O helix remains stacking with upstream DNA template base, the matched substrate appears to be base-paired with the following template base even though both substrate and template base is not fully loaded at active site. Then, T7 RNAP undergoes significant conformational changes and the Y639 residue of O helix moves away, the matched nucleotide and the templating base are loaded at active site to establish base stacking with upstream RNA:DNA hybrid. This conformational transition is stabilized by the formation of ionic links between the triphosphates moiety of the NTP, the two magnesium ions, and an R627 on the O helix in the insertion stage (closed conformation)[3]. T7 RNAP adopts a fully closed state, termed insertion state. After nucleotide incorporation, the pyrophosphate is dissociated, and the enzyme remains in the closed conformation due to the interaction network between a magnesium ion and R627 on the O helix. This cycle concludes with the dissociation of pyrophosphate, which leads to the formation of an open complex and the translocation of T7 RNAP[11].

The rotation of a 5-helix bundle, which includes the O helix, along a pivot axis constitutes the majority of the conformational shift in the enzyme associated with the return to the open

state[11]. Tyrosine residue (Y639) moves 3.5 Å and stacks on the primer-template bases when the bundle is rotated to the "open" conformation. In fact, this modification stabilizes the product duplex's translocated position (Figure 5). In addition to blocking the insertion site for the subsequent incoming NTP, the position of this tyrosine in this open state of the enzyme sterically prevents the T7RNAP from returning to its pre-translocation position[13]. The tyrosine side chain does not move out of the way until the NTP binds to the pre-insertion site on the O helix in the open state. The pre-insertion state then promotes the formation of the closed state, allowing the incoming NTP base and the template base to move in the insertion site and replace it[14]. The translocated state may be related to the open conformation being more stable than the closed conformation, following dissociation of the pyrophosphate and removal of the corresponding interaction networks. The rotating 5 helix bundle encloses hydrophobic surface area in the open conformation, which likely supplies energy to support the change to open conformation and the accompanying translocation upon PPi dissociation[11].

When the polymerase hits a terminator, the T7 RNAP transcription reaction stops and releases the mRNA. There are two distinct types of terminators, termed class I and class II terminators[15]. Class I terminator encodes a G-C rich sequence, which forms a hairpin secondary structure, on the RNA sequence, followed by a long chain of U residues. This is similar to rho-independent intrinsic terminators of *E. coli* RNAP[16]. It appears that the function of hairpin formation at these terminators is to impair T7 RNAP interactions with the single-stranded region of RNA located 8–14 nt away from the RNA 3'-end[17]. Additionally, hairpin formation may disturb certain base pairing in the RNA:DNA hybrid. Once the hairpin breaks these contacts, the remaining U-rich hybrid may be insufficient to maintain the RNA in the transcription complex, allowing for transcription to terminate. The U-rich section may also play a crucial role in slowing

transcription through the terminator to allow for isomerization in RNA secondary structure or RNAP conformation[17]. In contrast to class I terminators, class II terminators appear to lack an unique secondary structure. Initial identification of the Class II termination signal recognized by T7 RNAP was made in the cloned human prepro-parathyroid hormone (PTH) gene. These signals (class II signals) do not encode RNAs with a consistent secondary structure, but they do share a common sequence (ATCTGTT, non-template)[16]. Subsequently, more members of this type were discovered in the concatemer junction of replicating T7 DNA, in the *E. coli* rrnB T1 terminator, in a cDNA copy of the intergenic region of vesicular stomatitis RNA, in adenovirus DNA, and perhaps in bacteriophage lambda DNA[17].

1.4 Unnatural base pair

All life on earth contains four genetic alphabets in their DNA genome - dG, dC, dA, and dT. Scientists have been striving to construct novel synthetic nucleotides that form an unnatural base pair (UBP). The long-term goal in this field is to expand the genetic alphabet and enable the ability to create semi-synthetic organisms with additional information in their DNA and the potential for new traits or functions[18].

Over last few decades, several groups developed different strategies for UBPs. Dr. Steven A. Benner group started developing UBPs in the mid-1980s[19]. He found that Watson-Crick complementary pairing between natural base pairs follows two major principles:

1) Shape complementarity-always between a larger purine base and a smaller pyrimidine base.
2) Hydrogen bond pairing - hydrogen bonds are always formed through the interaction of a hydrogen bond donor on one base with a hydrogen bond acceptor on the other base.

Based on the these two principles, Benner designed a series of UBP with hydrogen bond donor and hydrogen bond acceptor groups of nucleotide bases are reshuffled. The purine and

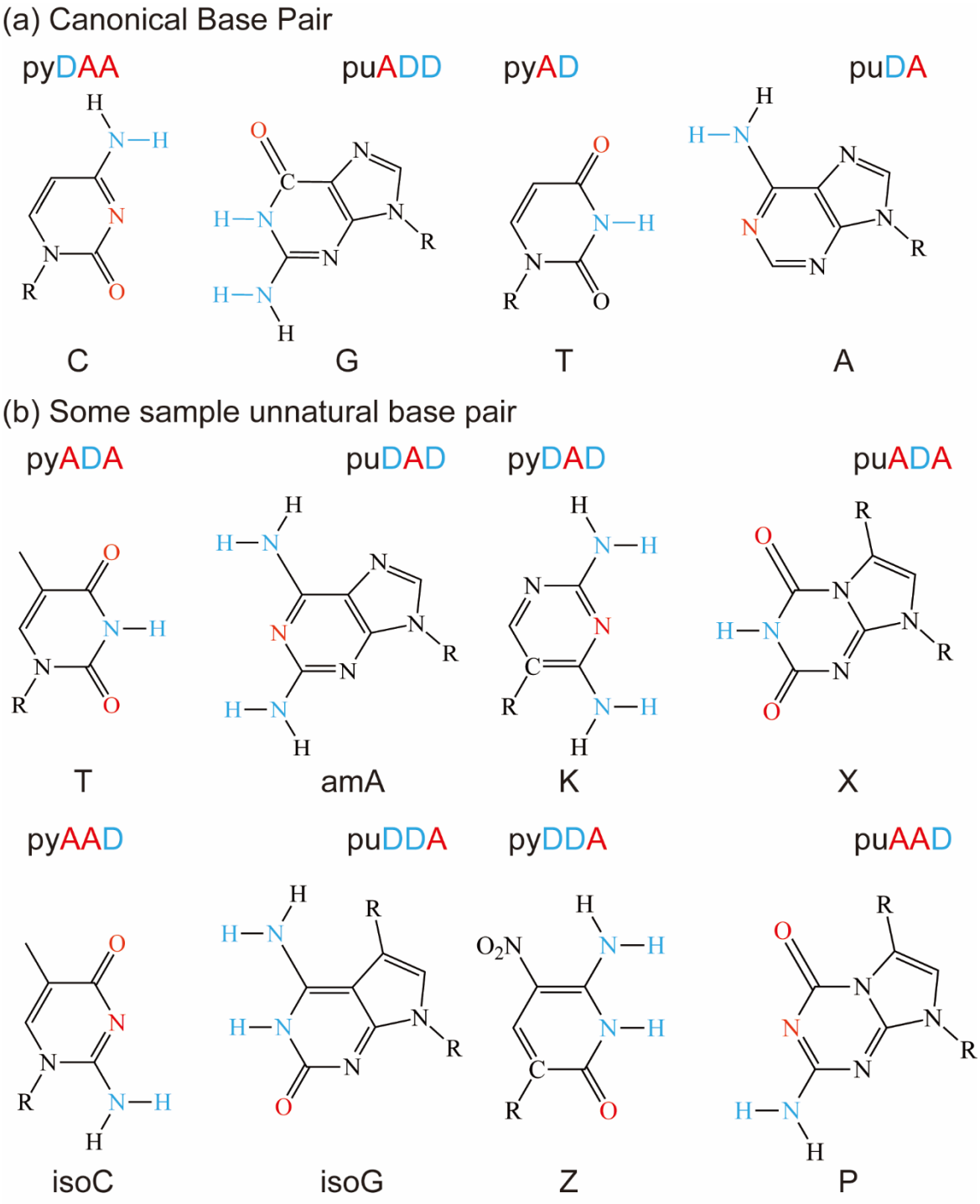

Figure 6: Unnatural base pairs invented by Benner's group are listed. The abbreviation py means pyrimidine, and pu means purine. The red A represent a hydrogen bond acceptor, and the blue N represents a hydrogen bond donor.

pyrimidine rings can also be suitably modified (e.g., K-X, Z-P base pairs), which in turn can greatly increase the variety of UBPs as shown in Figure 6[19]. Benner group first successfully achieved *in vitro* replication and transcription of DNA containing the UBP isoG-isoC in 1989, and then *in vitro* replication and transcription of DNA containing another UBP, K-X, in 1990[20]. Subsequently, in 1992, using isoG-isoC complementary pairs and with the help of synthetic tRNA containing the anticodon CU (isoG), the Benner group further translated the artificial extended codon (isoC) AG containing isoC into a non-canonical amino acid 3-iodotyrosine by an *in vitro* translation system and successfully incorporated it into a small peptide. When the fidelity of *in vitro* replication of these two UBPs was examined by primer extension experiments, the results were not very satisfactory. Since the late 1980s, the Benner group has been searching for better artificial base extensions, and in 2006, the Benner group finally discovered a better UBP, Z-P[21].

The pyDAA-puADD hydrogen bond pairing of Z and P (Fig. 6) enables the Z-P base pair to be more tightly bound and to form a more stable hydrogen bond than C-G. More importantly, the Z-P base pairs designed based on the Watson-Crick complementary pairing principle are more easily recognized by DNA polymerase (Fig. 7). This was also demonstrated by primer extension and PCR experiments, where most commercial DNA polymerases recognized Z-P base pairs with 99.8% fidelity[18].

In 1997-1998, Dr. Kool at Stanford University created a hydrophobic UBP (Z-F)[22]. He argued that hydrogen bond pairing was not necessary for the design of UBPs, but rather the complementarity of chemical structures, base stacking and electrostatic repulsion were more important. Using this idea, his group succeeded in creating the first hydrophobic UBP, Z-F (an

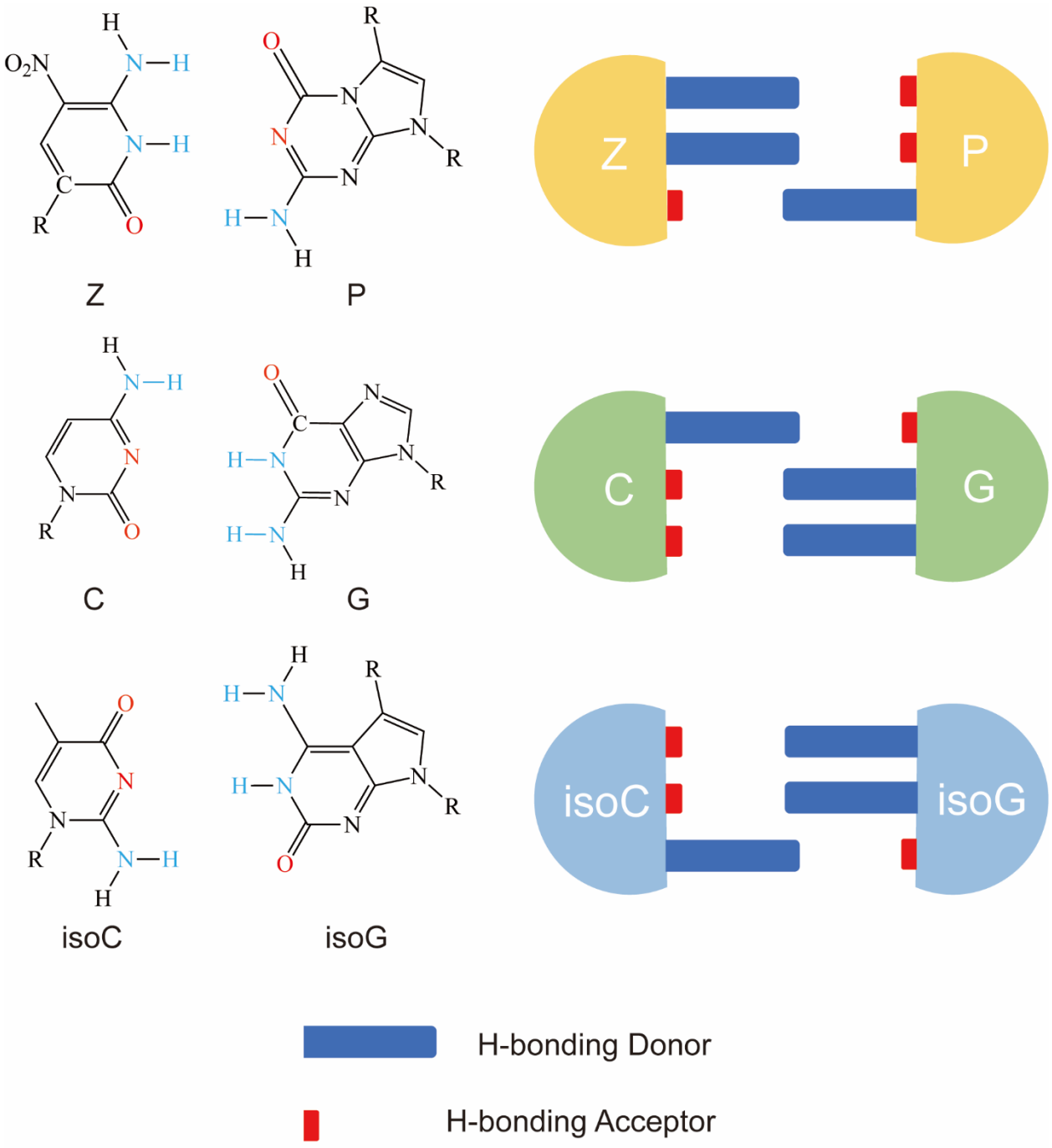

Figure 7: The hydrogen bonding of Z-P, C-G and isoC-sioG base pairs.

analogue of the T-A base pair) (Fig. 8). However, the fidelity of the Z-F pair was not good, while the pairing between A-F and Z-T was more efficient[22].

In 2006, Hirao et al. developed a new hydrophobic UBP, Ds and Pa (Fig. 8)[23, 24]. However, they soon found that hydrophobic UBPs generally have self-pairing problems. By chance, an

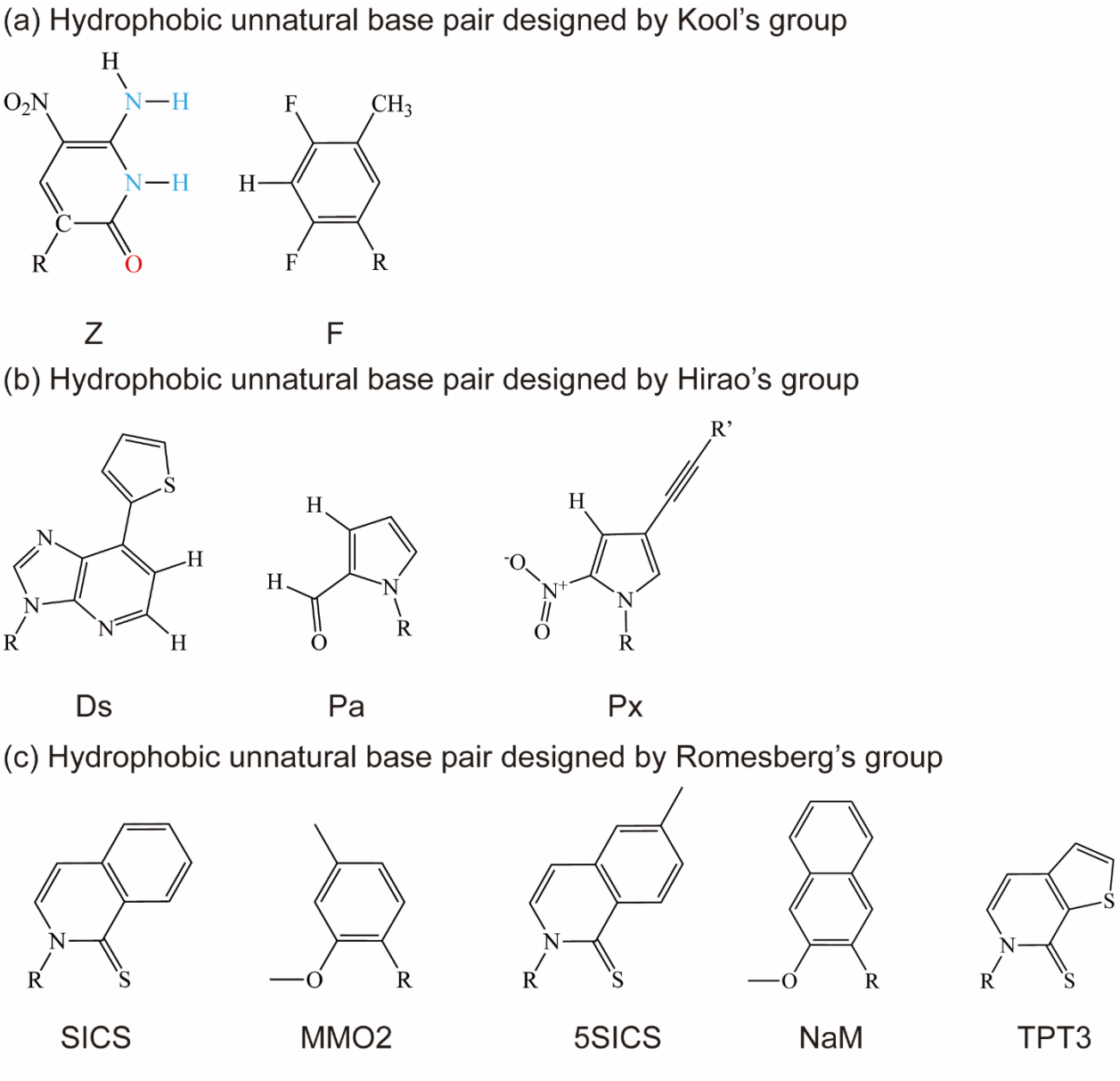

Figure 8: Some representative hydrophobic unnatural base pairs invented by Kool's group, Hirao's group, and Romesberg's group.

experimental oversight led Hirao et al. to accidentally synthesize a derivative of γ-amino dDsTP (dDsTPNH2) during the synthesis of dDsTP. Out of curiosity, they performed primer extension experiments using this derivative instead of dDsTP and found that in this case, the Ds only doped into the position corresponding to Pa on the template chain but not into the position corresponding to Ds. Based on these results, in 2006, the group reported efficient PCR

amplification of artificially expanded DNA containing Ds-Pa and *in vitro* transcription of artificially expanded DNA containing this base pair and its derivatives (mainly Pa containing various modifications). This work enabled the first truly efficient *in vitro* replication and transcription of artificially expanded DNA containing a non-natural base pair[24].

In 2009, Romesberg's team selected a hydrophobic base pair, dMMO2-dSICS, from the 60 hydrophobic artificial bases they had previously developed for efficient replication *in vitro* (Fig. 8)[25]. However, they encountered a similar problem to that encountered by Hirao's group in developing Ds-Pa, the self-pairing problem of dSICS[24]. However, they successfully overcame this problem by introducing a methyl group at the 5-position of dSICS to create a new UBP, dMMO2-d5SICS (Fig. 8)[26]. Based on this, the team further developed the aUBP dNaM-d5SICS (Fig. 8). In the same year, Romesberg's group reported PCR amplification[18] containing these two hydrophobic artificial base pairs as well as *in vitro* transcription[27]. They achieved high-fidelity PCR amplification (99.5%- 99.6% per cycle) of templates containing two consecutive or discontinuous dNaM- d5SICS base pairs in addition to PCR amplification of templates containing a single dNaM-d5SICS base pair[28]. By further modification of the d5SICS bases, the team developed a more superior non-natural base pair, dNaM-dTPT3 (Fig. 8). When using OneTaq for PCR amplification, the fidelity of dNaM-dTPT3 replication exceeded 99.98% per cycle, which is somewhat close to the replication fidelity of natural bases (10^{-4}-10^{-7} error rate)[18].

CHAPTER 2

2.1 Structural basis of UBP transcription by T7 RNA polymerase

Previous structural investigations have uncovered the structural basis for substrate selection and nucleotide addition cycle by T7 RNAP for the natural base pair system[13, 29]. T7 RNAP initiates substrate selection in the pre-insertion state, where the entering ATP substrate forms hydrogen bonds with the template dT base before both template base and substrate are completely loaded into the catalytic center of T7 RNAP. At this stage, ATP substrate is bound along with open O-helix approximately 10 Å from the active center, template +1 base is sequestered in a protein pocket formed by O/O'-helices (away from the active center), and the gate residue Tyr639 is stacked with upstream template -1 base obstructing the loading of template +1 base. In addition to nucleobase selection via hydrogen-bonding base-pairing, T7 RNAP chooses rNTP over dNTP via Mg^{2+}-mediated interaction of tyrosine 639 (Y639) with 2'-OH of substrate ribose in the pre-insertion state. Therefore, the pre-insertion state is a crucial checkpoint for the fidelity of T7 RNAP transcription. Upon correct substrate binding in the pre-insertion state, T7 RNAP undergoes coordinated conformational changes in order to transition to the insertion state for chemical reaction. These coordinated actions include of the loading of template +1 base and incoming NTP to the active center, the relocation of blocking Y639, and the rotation of the O-helix subdomain to seal the active site.

The substrate selection specificity of T7 RNAP in its pre-insertion state depends on complementary hydrogen bonding between Watson-Crick base pairs. Given that hydrophobic UBPs, such as Ds–Pa pairs, lack hydrogen bonds, an exciting unresolved issue arises: what is the chemical mechanism by which T7 RNAP recognizes and processes Ds–Pa pairs? This is a crucial knowledge gap in the field. This question was addressed by utilizing a

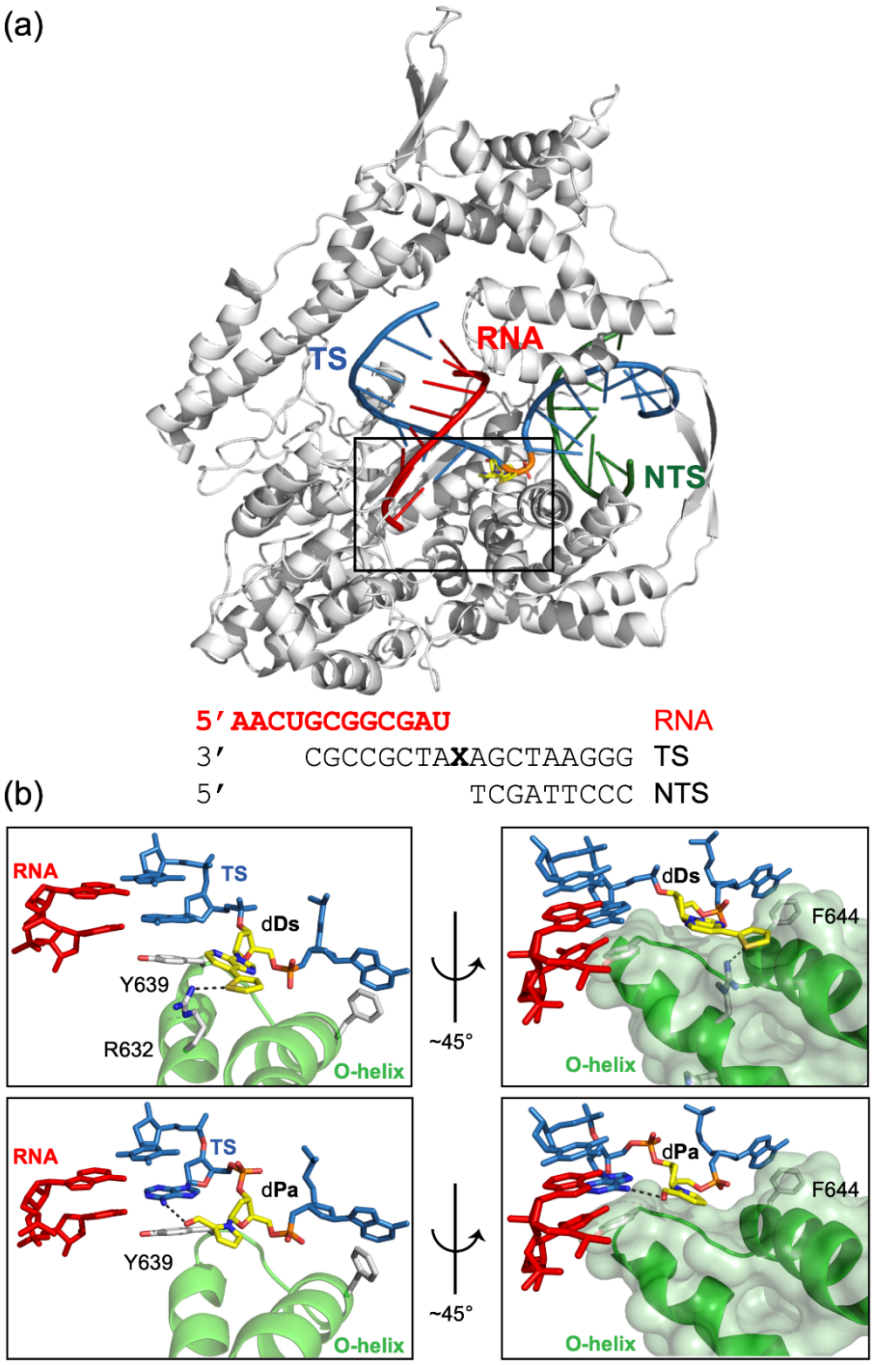

Figure 9: Unnatural template loading by T7 RNA polymerase. (a) Scaffold used for structural study and overall structure of T7 RNA polymerase. Template strand DNA, RNA and non-template strand DNA is colored blue, red and green, respectively. (b) Active site of apo dDs or dPa harboring T7 RNAP (left panel). Right panel shows surface of O-helix and UBP together to indicate fitting of UBP above the O-helix. O-helix and its surface is colored in green. +1 template base is colored in yellow. Other interacting residues are colored in white. Potential hydrogen bonding distance between UBP and T7 elements are labeled with black dash. This figure is cited from Oh, J.; Kimoto, M; Xu, H.; Chong, J.; Hirao, I.;Wang, D. Structural basis transcription recognition of a unnatural base pair by T7 RNA polymerase. *Nature Communications* **2023**, *14*(1), 195.[30]

combination of enzyme kinetics, structural biology, modeling, and mutagenesis. Using kinetic analysis, we determined if and how the T7 RNAP elongation complex selectively identifies dDs or dPa templates and incorporates corresponding unnatural nucleotide substrates (PaTP or DsTP) during transcription. To comprehend the structural basis of UBP transcription, several structures of T7 RNAP–UBP complexes with or without their respective substrates (dDs, dPa apo structures, dDs–PaTP, dPa–DsTP, dDs–ATP, dPa–ATP complex structures) were solved. Our findings revealed that both dDs and dPa prefer unconventional pairing partners (with variations in incorporation efficiency and fidelity). We detected a separate way of unnatural nucleoside triphosphate binding for PaTP and DsTP, indicating a distinct substrate recognition mechanism of UBP during T7 RNAP elongation. Using structures as a guide, we have identified a T7 RNAP mutant that affects UBP transcription preferentially but not normal nucleic acid transcription. Collectively, these findings offer mechanistic insights on UBP transcription and prospective design strategies for the next generation of UBPs.

2.2 Structural analysis of T7 RNA polymerase elongation complex with UBP

Six different T7 RNAP elongation complex structures containing a site-specific dDs or dPa at the +1 site of the template strand were solved[30]. The overall structure of T7 RNAP elongation complexes including UB are remarkably similar to those of natural T7 RNAP (Fig. 9)[12]. T7 RNAP is caught in the post-translocation state, in which the elongation complex is awaiting incoming substrate and the gating Y639 was stacking with the -1 template strand base. Both dDs and dPa are accommodated in the same binding pocket over the O-helix as the natural bases (Figs. 9)[12]. Intriguingly, we discovered UB-specific interactions between the unnatural base and T7 RNAP. The dDs base appears to interact with R632 through the formation of hydrogen bonds

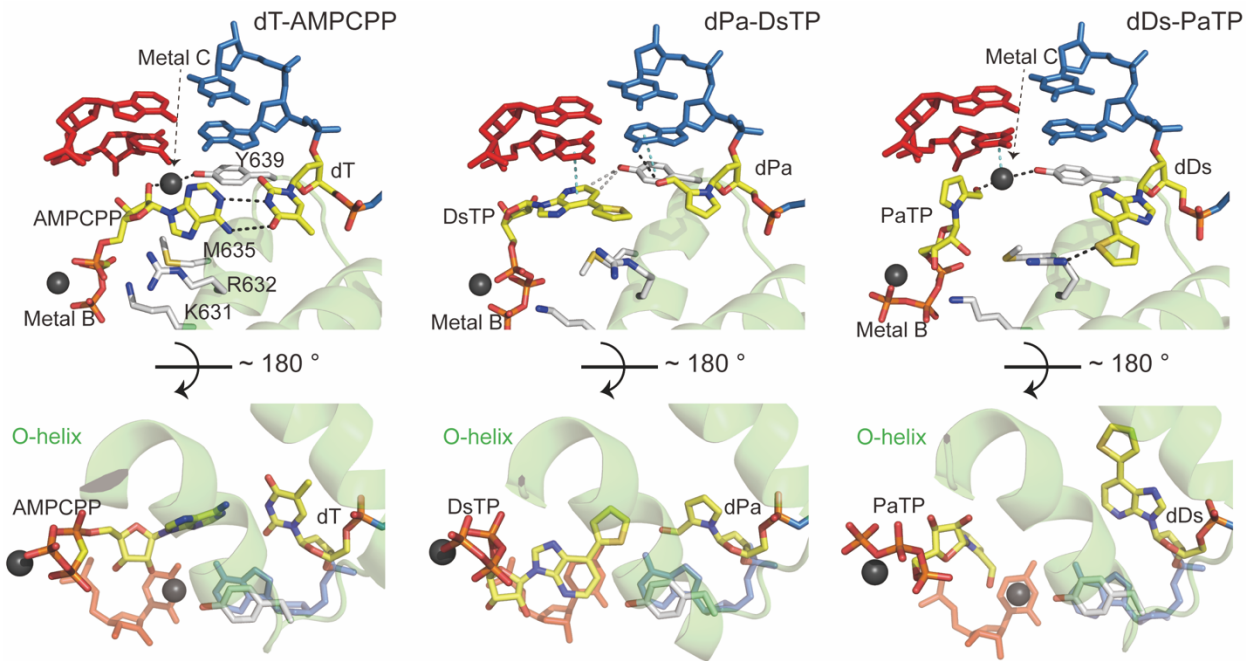

Figure 10: Comparison of pre-insertion state between natural and unnatural base pairs (UBPs). Pre-insertion state with dT–AMPCPP, dPa–DsTP and dDs–PaTP pair (PDB code: 1S76). Key residues for recognizing incoming nucleoside triphosphate as well as template strand, were shown in white. +1 template base and substrate are shown in yellow. Magnesium ions (Metal B: interact with triphosphate moiety. Metal C: interact with Y639) are colored in dark gray. Hydrophilic and stacking interactions are shown in black and cyan dash. This figure is cited from Oh, J.; Kimoto, M; Xu, H.; Chong, J.; Hirao, I.;Wang, D. Structural basis transcription recognition of a unnatural base pair by T7 RNA polymerase. *Nature Communications* **2023**, *14*(1), 195. [30]

between the sulfur atom in Ds and the carboxy group in R632. We also discovered the aldehyde group of dPa base at +1 position creates hydrogen bonding with 6-amine group of adenine base at the -1 position, suggesting particular nucleobase of flanking sequence may contribute to the constraint of the Pa conformation during the template loading.

T7 RNAP recognizes the UBP substrate in the pre-insertion stage. Prior research shown that the cognate natural nucleotide makes hydrogen bonds with its corresponding template base in the pre-insertion state, a vital fidelity checkpoint state before T7 RNAP enters the insertion state and commits to nucleotide incorporation[13, 29]. At the pre-insertion state, the incoming nucleotide substrate forms first hydrogen bonds and base pairs with the +1 template base, whilst the gate residue Y639 continues to stack with the -1 template base. The +1 base pair (+1 template base

and incoming substrate) is not yet fully loaded to form base stacking interactions with upstream RNA:DNA hybrid, and the O-helix is in an open conformation.

Because the Ds–Pa pair lacks hydrogen bonds but nonetheless allows reliable transcription *in vitro*, we are interested in determining how T7 RNAP recognizes an incoming unnatural nucleoside triphosphate. Four T7 RNAP substrate-bound structures (dDs–PaTP, dDs–ATP, dPa–DsTP and dPa–ATP structures) were generated. (Fig. 10). Intriguingly, we found that both DsTP and PaTP substrates adopt novel conformations not seen in previous T7 RNAP structures with natural substrates[29]. Therefore, the substrate recognition patterns of DsTP and PaTP are quite different from those of their natural nucleotide counterparts.

Previous research has shown that the identification of natural ATP in the pre-insertion state is achieved by three phases of crucial interactions[29]. The base moiety of incoming NTP first forms hydrogen bonds with the +1 template base. Second, the ribose portion of the substrate is identified by R632 and magnesium ion (metal C) – Y639 bridges. The triphosphate moiety and magnesium ion (metal B) are ultimately recognized by D471, K474, R627, and K631. In contrast, both DsTP and PaTP maintain only the contacts necessary to hold magnesium ion (metal B) and its triphosphate moiety (Fig. 10).

For the T7 RNAP dPa–DsTP complex, we identified a unique intermediate state, termed the "primed" state, between the pre-insertion and insertion states. At insertion state, DsTP is moved closer to the binding site. The enlarged nucleobase of DsTP is compatible with the 3'-end nucleobase of RNA. In its primed state, Y639 continues to stack with a template base of -1. The geometry of Ds aromatic ring is complementary to that of Y639's benzene ring, resulting in a perfect edge-to-edge base pair contact with Y639 and stacking with -1 RNA:DNA base pair (Fig. 10).

For the T7 RNAP structure including dDs–PaTP, we observed that the ribose and base moieties of PaTP rotated 180 degrees while magnesium ion (metal B) and triphosphate moiety interactions were maintained. As a result of ribose moiety rotation, we discovered that R632 interacts with the sulfur atom of template dDs rather than the O4 atom of nucleotide substrate's ribose ring. K631, rather than R632, interacts with PaTP's 3' hydroxyl and phosphate groups (Fig. 10). Intriguingly, a unique interaction network was found between the aldehyde group of the Pa nucleobase moiety, metal C, Y639, and the -1 nucleobase of 3' RNA (Figs. 10). By direct coordination with the aldehyde group of Pa nucleobase and the hydroxyl group of Y639 as well as the base of 3' RNA, the positioned metal C was stabilized (Figs. 10). Intriguingly, a metal C was detected at a similar position in the pre-insertion state of the atom in its normal state. The magnesium ion is misplaced with respect to the center of the -1 base ring and is coordinated by Y639 and 2'-OH of the entering AMPCPP (Fig. 10).

Chapter 2 is co-authored with Dr. Oh, Juntaek; Dr. Kimoto, Michiko; Chong, Jenny; Dr. Hirao, Ichiro and Dr. Wang, Dong. The thesis author was the primary author of this chapter.

CHAPTER 3

3.1 Function of key residues in recognizing UBPs

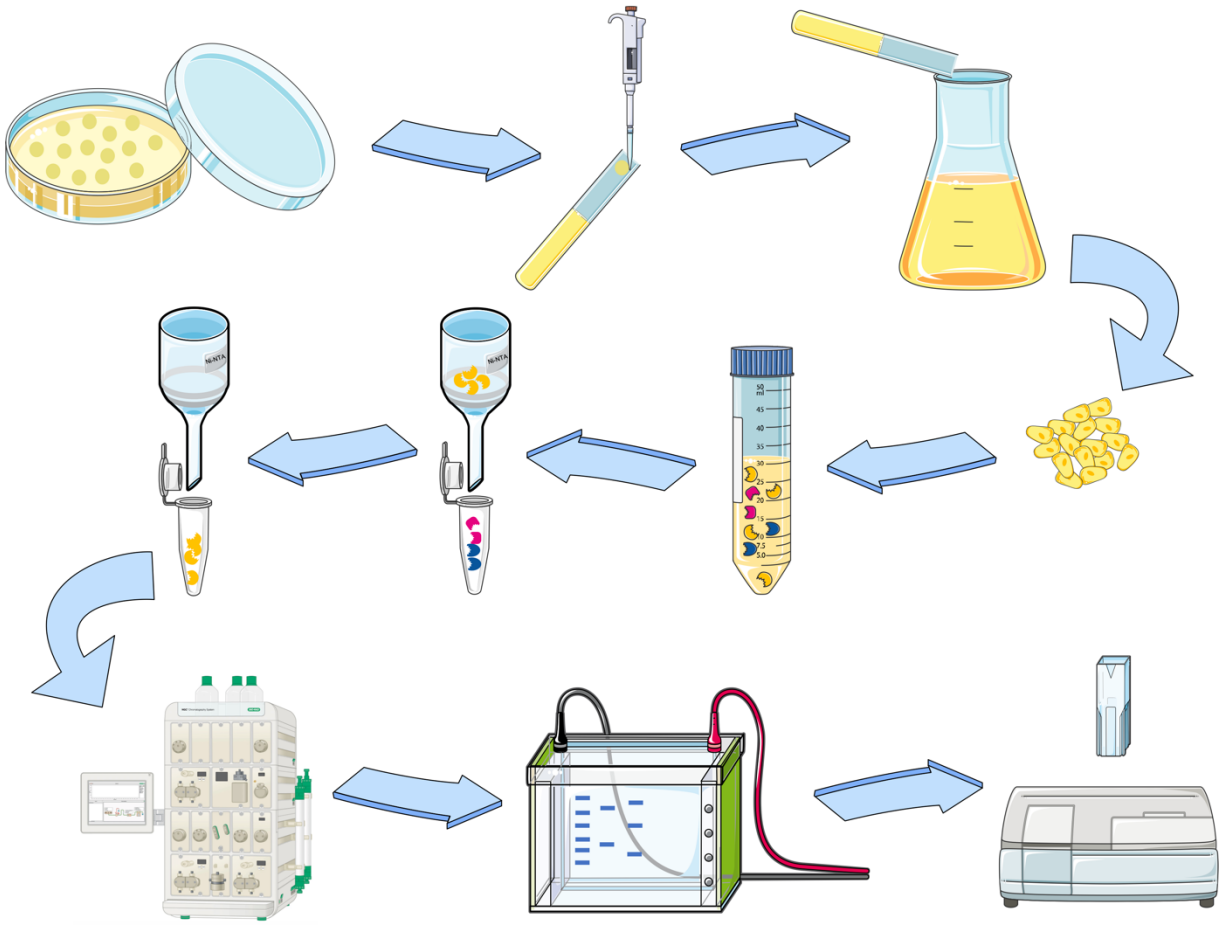

Figure 11: Scheme of purification process.

As we analyze the structures of T7 RNAP complexes including UBP, we've observed that several critical residues behave differently compared to their native pre-insertion state. For example, Y639, which identifies the 2' hydroxyl group in natural transcription, was interacting via a magnesium ion with the aldehyde group of PaTP (and close distance to DsTP). R632, which previously interacted with ribose's oxygen, now interacts with dDs's sulfide group. M635 appears to stabilize the entering substrate through a hydrophobic action. To determine their functions in UBP transcription, we produced T7 RNAP mutants with a single-residue

substitution, Y639F, R632A, or M635A/Y/H/F/L/I/W, and analyzed their single-nucleotide incorporation efficiency.

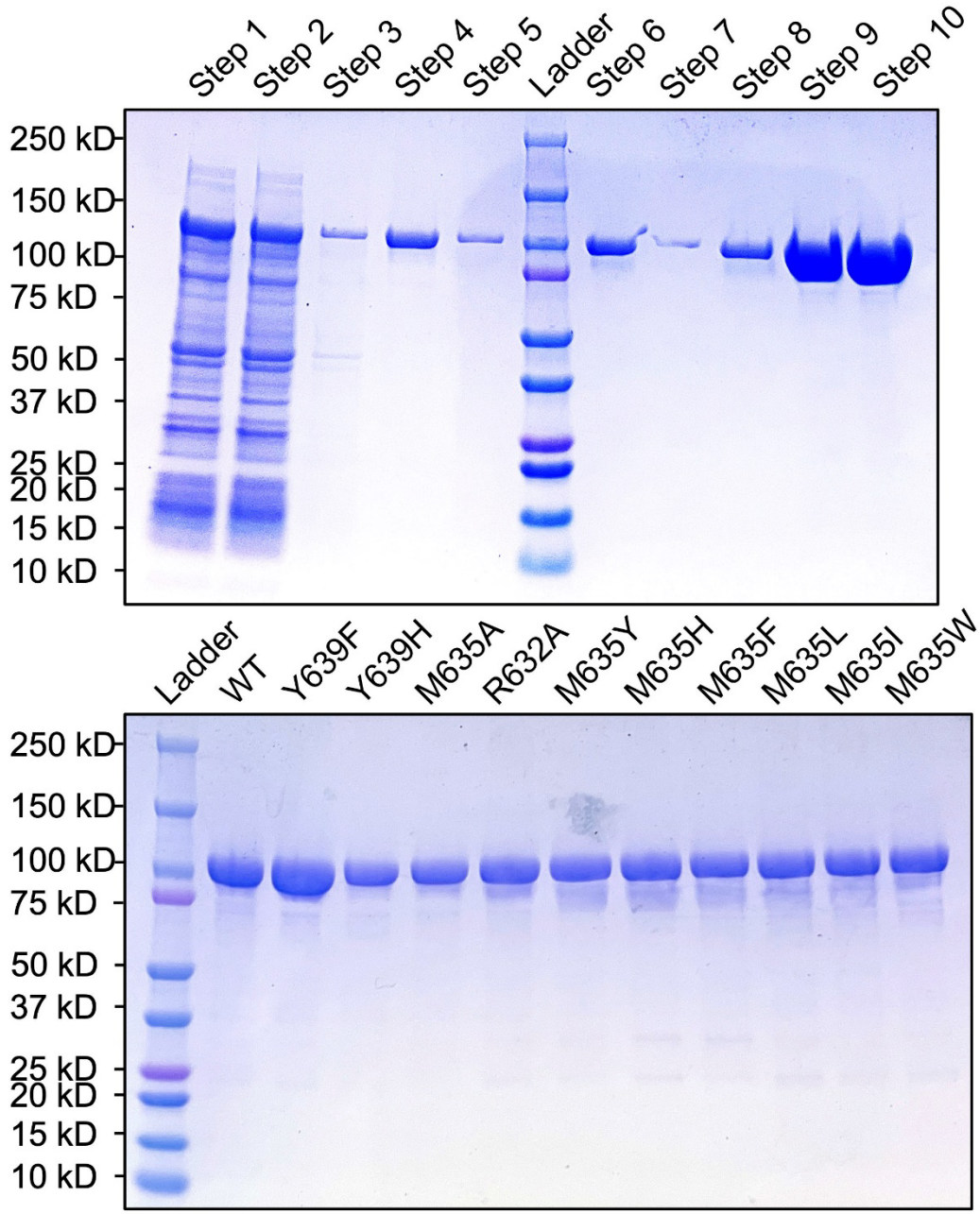

Figure 12: (a) SDS-page gel of each purification process. Step 1: Mix lysate supernatant with Ni-NTA resin at 4 °C for 1 hour. Step 2: Use gravity column to separate resin from mixture. Step 3: Wash with lysis buffer with additional 20 mM imidazole. Step 4: Wash with lysis buffer with additional 30 mM imidazole. Step 5: Add 2 mg TEV protease and incubate at 4 °C overnight. Collect flow through the next day. The band is the resin after collecting flow through. Step 6: The band is the flow through of Step5. Step 7: Load Step 5 flow through to resin and collect flow through. Step 8: The band is the flow through of Step 7. Step 9: Dilute and load to Heparin HP column and harvest peak. Step 10: Concentrate with 50 kDa filter system. (b)SDS-page gel of all final products.

3.2 Generation of T7 RNAP mutants

For T7 RNAP mutants, we back mutated pQE-T7-FAL plasmid provided by Andrew Ellington Lab to wild-type pQE-T7[31] through PCR site mutagenesis and confirmed by sequencing. WT pQE-T7 was then mutated to Y639F, R632A, or M635A/Y/H/F/L/I/W, respectively in the same way.

Each T7 mutant plasmid was transformed into BL21 (non-DE3) to ensure WT-T7 RNAP free expression and purification. Fig. 11 shows the process of T7 RNAP and its mutants purification process. In LB broth, cells were grown at 37 °C until $O.D._{600}$ reach 0.6 to 0.8. 0.5 mM IPTG was added at this time to ensure the expression of protein at 20 °C.

After shaking in the incubator for 20 hours, cells were lysed by a microfluidizer in lysis buffer (20 mM Tris pH 7.5, 200 mM NaCl, 2 mM BME). Lysate supernatant was applied to a gravity Ni-NTA column and washed with 20 mM imidazole buffer (lysis buffer+20 mM imidazole pH 8.0) and lysis buffer with additional 30 mM imidazole buffer (lysis buffer+30 mM imidazole pH 8.0). After checking the flowthrough solution with Bradford protein assay, purified protein was eluted by lysis buffer with additional 250 mM imidazole buffer. Elution was further purified by Heparin HP column (Cytiva). For protein used for crystallization, TEV protease was added after washing with 30 mM imidazole buffer and incubated at 4 °C overnight to remove His tag. The column flowthrough was collected the next day and washed with 20 mM imidazole buffer until the Bradford protein assay showed grey. All flowthrough was loaded to a new Ni-NTA column pre-equilibrium with 20 mM imidazole buffer. The final flowthrough was further purified by Heparin HP column (Cytiva). Purified T7 RNAP was concentrated to different concentrations (Table 3.1), respectively, in storage buffer (20 mM Tris pH 7.5, 450 mM NaCl, 5 mM BME, 10% Glycerol). Each aliquot was flash-frozen and stored in -80 °C. Protein was

confirmed by SDS-page gel (Fig. 12), and the concentrations were determined by Bradford protein assay.

Table 3.1: Protein concentration.

Mutant Name	Final concentration(mg/ml)
WT	20.0
Y639F	18.2
M635A	31.4
R632A	32.1
M635W	65.9
M635F	31.6
M635Y	20.8
M635H	32.2
M635L	25.2
M635I	35.7

3.3 In vitro transcription assay

In vitro transcription assay is a laboratory technique used to measure the activity of a specific gene by measuring the amount of RNA produced from it in a test tube or culture dish. The assay typically involves isolating the DNA of interest, introducing it into a cell-free system containing the necessary enzymes and cofactors for transcription, and measuring the resulting RNA using techniques such as gel electrophoresis or real-time PCR[32]. This method allows researchers to study gene expression and regulation in a controlled environment, and can be used to investigate the effects of mutations, drugs, and other factors on gene activity.

Here, we performed the transcription assay with a mini scaffold. The 18-mer DNA templates (5'-GGGAATCGA**X**ATCGCCGC, **X** = Ds or Pa) with unnatural bases were chemically synthesized via an H8 DNA/RNA Synthesizer (K&A Laborgerate), using phosphoramidites

reagents for the natural and Ds or Pa bases, followed by a purification with denaturing gel electrophoresis. The UBPs, Ds and Pa phosphoramidites, and Ds and Pa triphosphates (DsTP and PaTP), were prepared as described previously [33]. Mini-scaffold was prepared by annealing 18-mer template-strand DNA (5'-GGGAATCGAXATCGCCGC, X = Ds or Pa), non-template DNA (5'-TCGATTCCC) and P^{32} -labeled RNA (5'-AACUGCGGCGAU) at 2:3:1 molar ratio in a buffer containing 10 mM Tris pH 8.1, 200 mM NaCl, 20 mM $MgCl_2$ and 5 mM BME. The scaffolds were heated to 80 °C for 5 mins and cooled down gradually to room temperature.

For the transcription assay with mini-scaffold, 4 μM T7 RNAP was incubated with pre-assembled mini-scaffold in elongation buffer (EB, 20 mM Tris pH 7.5, 40 mM KCl, 5 mM $MgCl_2$, and 5 mM DTT) at room temperature for 20 min to ensure the formation of a stable T7 RNAP elongation complex. To start the transcription reaction, an equal volume of Pol II elongation complex was mixed with various concentration of rNTPs or unnatural base triphosphate. Final reaction condition was 120 nM of T7 RNAP, 20 nM of scaffold, and varying concentrations of rNTPs in elongation buffer.

For each settled time point, four volumes of the stop buffer (90 % formamide, 10% 0.5 M EDTA pH 8.0 with bromophenol blue and xylene cyanol dye) was added to the reaction mixture to stop the reaction. The zero time point was generated by adding stop buffer to the elongation complex without NTPs. After quenching, the samples were incubated at 95 °C for 15 min to denature the T7 RNAP elongation complex and scaffold. After denaturation, the RNA transcripts were ready for denaturing urea/PAGE gel analysis. The quenched products (RNA transcripts) were analyzed by denaturing urea/PAGE gel with 12% of acrylamide in 1 × TBE at 200 V for 2.5–3.5 h. RNA transcripts are visualized using a storage phosphor screen and Typhoon imager (GE)[32].

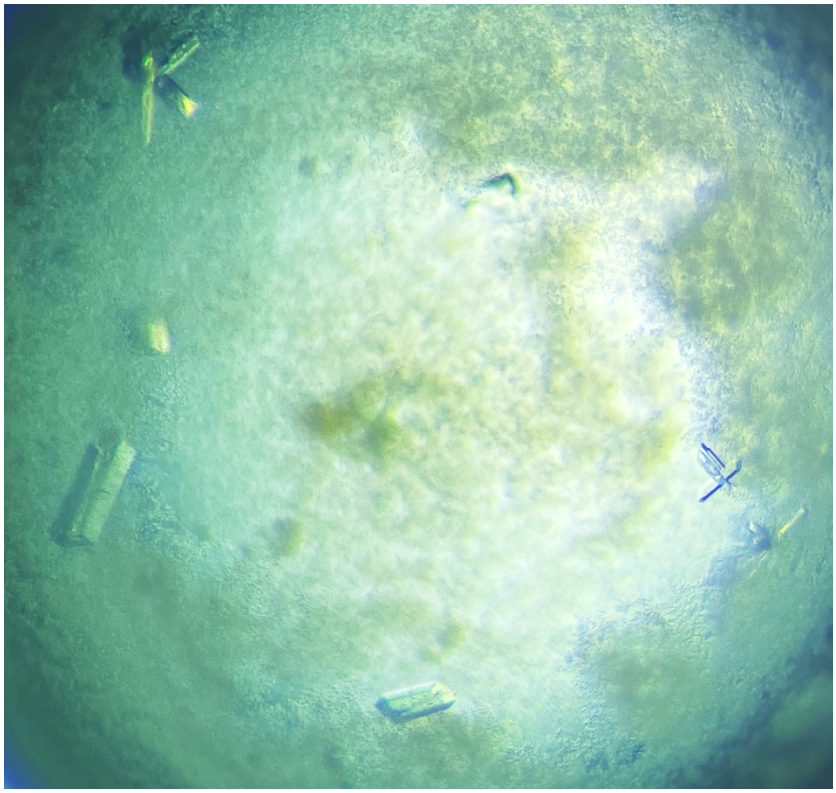

Figure 13: Shape of T7 RNAP elongation complex crystal. This photo was taken 1 week after setting up crystal.

3.4 T7 RNAP crystal set up

Firstly, mini-scaffold was assembled by mixing 18-mer template strand DNA (5'-GGGAATCGAXATCGCCGC, X=Ds or Pa), non-template strand DNA (5'-TCGATTCCC) and 3' deoxy RNA (5'-AACUGCGGCGAU) at 1:1:1.2 molar ratio in a buffer containing 10 mM Tris pH 8.1 (filtered by 0.22 µm PES), 200 mM NaCl (filtered by 0.22 µm PES), 20 mM $MgCl_2$ (filtered by 0.22 µm PES) and 5 mM BME (freshly diluted). 3' deoxy RNA was used to avoid nucleotide addition, and the T7 RNAP elongation could pause at pre-insertion stage. The mixture of nucleic acid strands was heated at 80 °C for 10 mins and cooled down gradually to room temperature. T7 RNAP elongation complex was obtained by mixing enzyme and scaffold at 1:1.2 molar ratio and allowed them to assemble on ice for 30 mins. Microcentrifuge was cooled down to 4 °C at this time. T7 RNAP elongation complex was centrifuged at 13200 rpm for 10

mins to remove precipitated complex. Then, T7 RNAP elongation complex (10 mg/ml) was crystalized by mixing with equal amount of crystallization buffer in hanging drop vapor diffusion plate at 22 °C. The crystallization buffer consists of 100 mM Tris pH 8.5 (Hampton Research), 10–15% PEG 8000 (Hampton Research), 8% glycerol (Hampton Research) and 5 mM BME (freshly diluted). Fig. 13 shows the shape of crystal under microscope.

3.5 Results and discussion

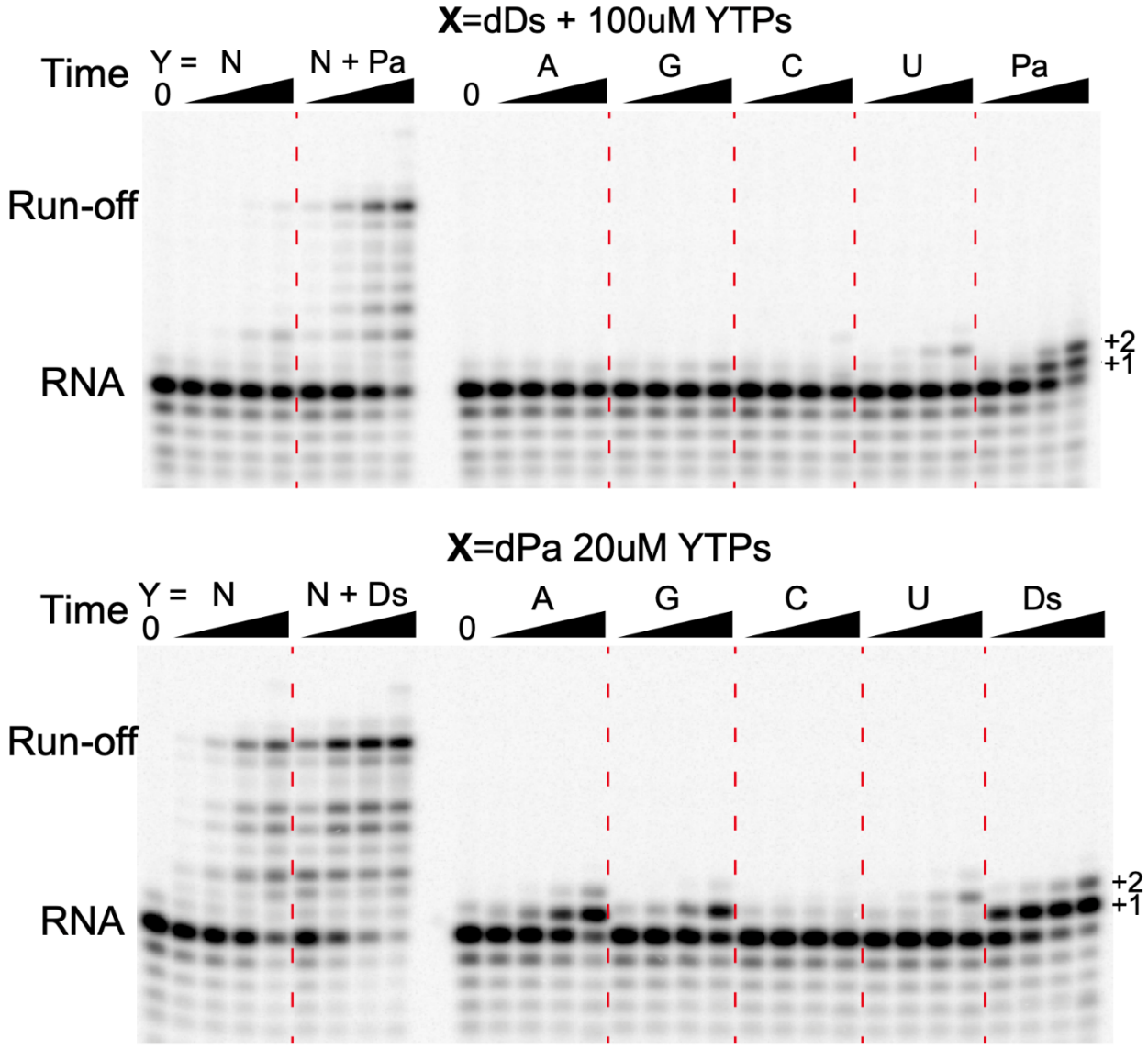

Figure 14: Single-nucleotide incorporation analysis. X indicates template-strand DNA base, while Y indicates added nucleoside triphosphate. Time points were 15 sec, 1 min, 5 min and 30 min. This figure is cited from Oh, J.; Kimoto, M; Xu, H.; Chong, J.; Hirao, I.;Wang, D. Structural basis transcription recognition of a unnatural base pair by T7 RNA polymerase. *Nature Communications* **2023**, *14*(1), 195. [30]

Fig. 14 shows that WT T7 RNAP an identify both dDs and dPa templates, and preferentially inserting their designed cognate partner, PaTP and DsTP, respectively. To be specific, PaTP is the only substrate that is incorporated opposite the dDs template (Figs. 14). Consistently, PaTP is needed for T7 RNAP to bypass the Ds site and generate the run-off transcript during elongation

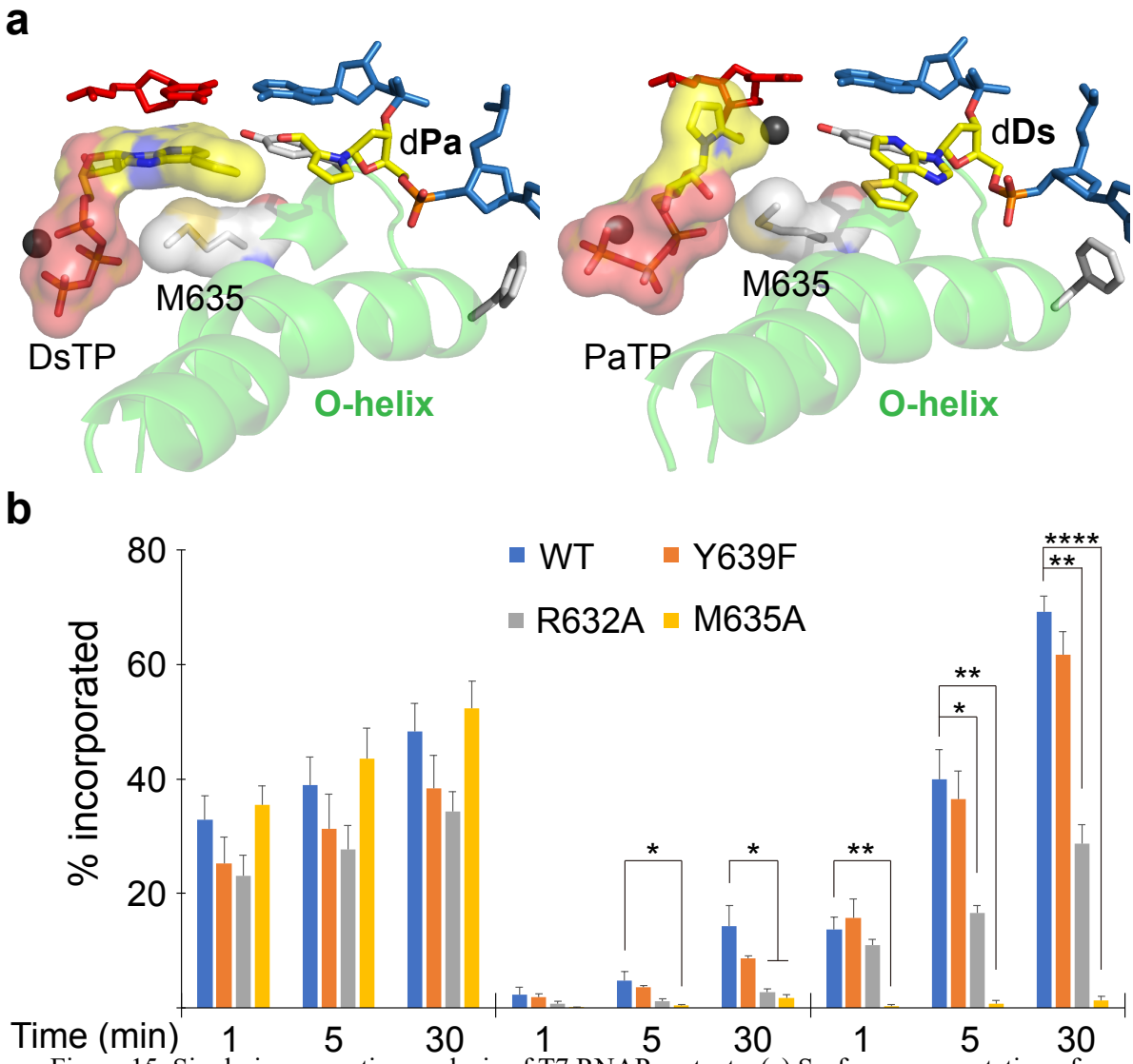

Figure 15: Single-incorporation analysis of T7 RNAP mutants. (a) Surface representation of incoming DsTP or PaTP and M635. Surface was colored by atom type. Carbon of M635 and DsTP were colored as white and yellow, respectively. (b) Relative incorporation efficiency of T7 RNAP WT and mutants. All data in this figure were obtained and quantified from at least three independent experiments (*P < 0.05, **P < 0.01, ****P < 0.0001, two-tailed Student's t-test). This figure is cited from Oh, J.; Kimoto, M; Xu, H.; Chong, J.; Hirao, I.;Wang, D. Structural basis transcription recognition of a unnatural base pair by T7 RNA polymerase. *Nature Communications* **2023**, *14*(1), 195. [30]

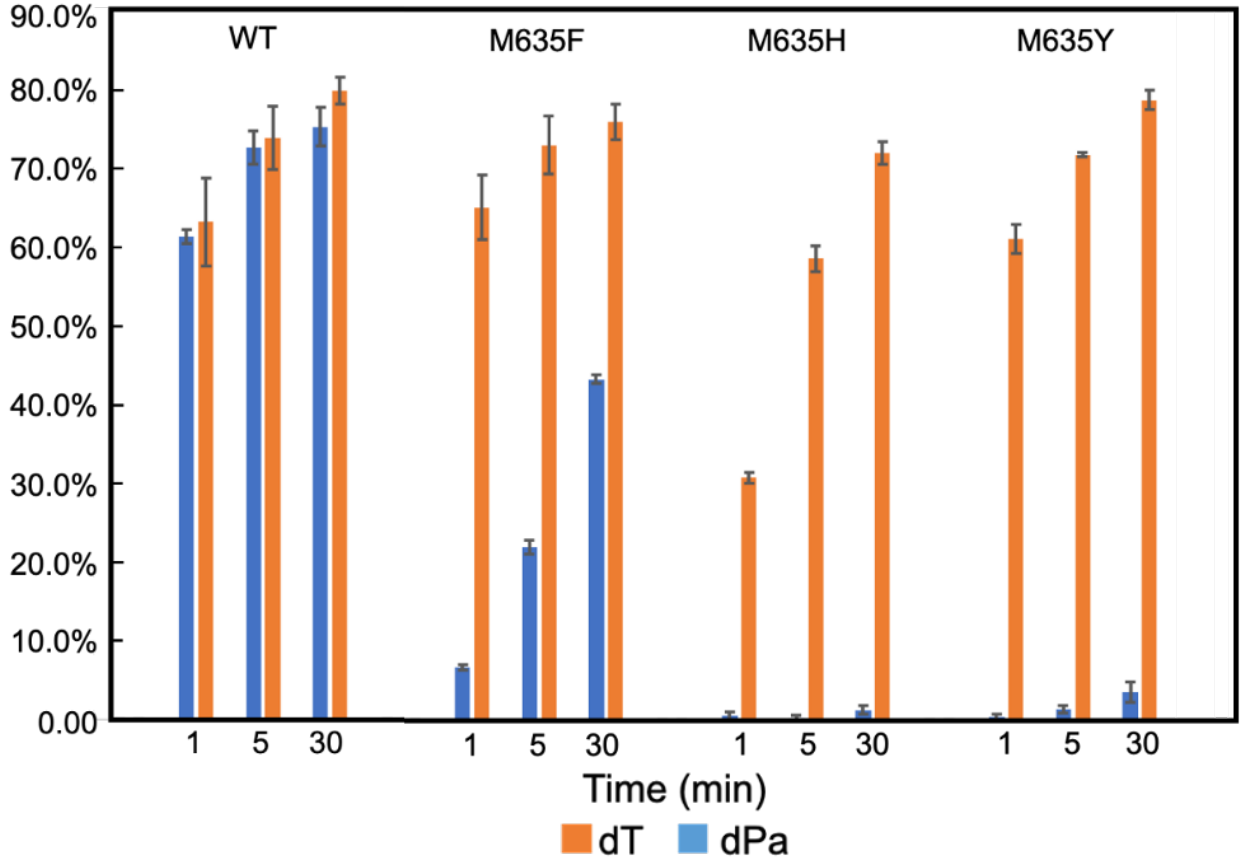

Figure 16: Quantitation of M635 mutants single incorporation test.

process (Fig. 14). In contrast, rapid DsTP addition and slow addition for ATP and GTP were observed for the dPa template (with longer incubation time) (Fig 14). In addition, WT T7 RNAP prefers dPa template than dDs template, less DsTP concentration is required for efficient transcription as well as producing run-off products.

Fig. 15 shows that Y639F substitution results in a comparable small drop in dT–ATP scaffold and maintain about 50% activity in UBP transcription. Intriguingly, the R632A mutation acts differently in each scaffold. In nature scaffold, the R632A mutant is slightly weaker than Y639F mutant. However, in dDs–PaTP incorporation, the R632A mutation inhibits by more than 5-fold, and in dPa–DsTP incorporation, it's even more active than WT T7 RNAP. M635A totally eliminates UBP incorporation (dDs–PaTP and dPa–DsTP) but has nearly no effect on

spontaneous ATP incorporation against a dT template. This finding suggests that T7 RNAP residues may have a different role in mediating transcription from UBP scaffolds as opposed to natural scaffolds. Specifically, we discovered that M635 becomes overwhelmingly crucial for Ds–Pa pair transcription (Fig. 15).

Because M635 seems vital to UBP transcription, more mutants, such as M635F, M635H, M635Y were generated to help further testing how this residue would affect T7 RNAP transcribing UBP. Natural scaffold dT with ATP incorporation, and UBP dPa with DsTP incorporation were tested by these mutants. Fig. 16 shows the quantitation of in vitro transcription assay of single incorporation efficiency of these mutants. Among all of the 4 mutants, M635F has conserved most, about 55%, of enzyme activity when incorporating DsTP to dPa template. Intriguingly, M635H and M635Y almost kills the enzyme when DsTP incorporated.

3.6 Conclusion and future direction

Here, we discovered that DsTP or PaTP assumes a distinct pre-insertion state that is not found in any other natural substrate structure. Although the pre-insertion states of UBP differ from those of natural pairs, our biochemical assays indicate that these pre-insertion states are capable of allowing the specific nucleotide addition reaction. Among all of the mutations we tested, M635 is crucial for hydrophobic UBPs transcription recognition. Even if M635 seems to provide hydrophobic force in structure, mutation to more hydrophobic side chain doesn't improve the efficiency UBP transcription. Unfortunately, crystallization trial of these mutants was not successful. If guided by crystal structure, more clear clues may possibly be discovered. Future research may focus on the crystal structure of important mutants, such as M635A, Y639F, R632A and M635F with nature or UBP incorporation. Also, biochemical assay of double

mutants and other mutants may be considered as well. In addition, other RNA polymerase, like Pol II and *E. Coli.* RNA polymerase, may be investigated for comparison.

Chapter 3 is co-authored with Dr. Oh, Juntaek; Dr. Kimoto, Michiko; Chong, Jenny; Dr. Hirao, Ichiro and Dr. Wang, Dong. The thesis author was the primary author of this chapter.

REFERENCES

1. Geiduschek, E. P.; Tocchini-Valentini, G. P., Transcription by RNA polymerase III. *Annual review of biochemistry* **1988,** *57* (1), 873-914.

2. Cheetham, G. M.; Steitz; A, T., Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **1999,** *286* (5448), 2305-2309.

3. Sousa, R.; Mukherjee, S., T7 RNA polymerase. *Progress in nucleic acid research and molecular biology* **2003**, 1-41.

4. Osman, S.; Cramer, P., Structural biology of RNA polymerase II transcription: 20 years on. *Annual Review of Cell and Developmental Biology* **2020,** *36*, 1-34.

5. Sauguet, L., The extended "two-barrel" polymerases superfamily: structure, function and evolution. *Journal of molecular biology* **2019,** *431* (20), 4167-4183.

6. Bonner, G.; Lafer, E. M.; Sousa, R., The thumb subdomain of T7 RNA polymerase functions to stabilize the ternary complex during processive transcription. *Journal of Biological Chemistry* **1994,** *269* (40), 25129-25136.

7. Steitz, T. A., A mechanism for all polymerases. *Nature* **1998,** *391* (6664), 231-232.

8. Dousis, A.; Ravichandran, K.; Hobert, E. M.; Moore, M. J.; Rabideau, A. E., An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nature Biotechnology* **2022**, 1-9.

9. Imburgio, D.; Rong, M.; Ma, K.; McAllister, W. T., Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **2000,** *39* (34), 10419-10430.

10. Rong, M.; He, B.; McAllister, W. T.; Durbin, R. K., Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences* **1998,** *95* (2), 515-519.

11. Yin, Y. W.; Steitz, T. A., Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **2002,** *298* (5597), 1387-1395.

12. Tahirov, T. H.; Temiakov, D.; Anikin, M.; Patlan, V.; McAllister, W. T.; Vassylyev, D. G.; Yokoyama, S., Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* **2002,** *420* (6911), 43-50.

13. Yin, Y. W.; Steitz, T. A., The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **2004,** *116* (3), 393-404.

14. Long, C.; E, C.; Da, L. T.; Yu, J., Determining selection free energetics from nucleotide pre-insertion to insertion in viral T7 RNA polymerase transcription fidelity control. *Nucleic acids research* **2019,** *47* (9), 4721-4735.

15. Jeng, S. T.; Gardner, J. F.; Gumport, R. I., Transcription termination in vitro by bacteriophage T7 RNA polymerase. The role of sequence elements within and surrounding a rho-independent transcription terminator. *Journal of Biological Chemistry* **1992,** *267* (27), 19306-19312.

16. Lyakhov, D. L.; He, B.; Zhang, X.; Studier, F. W.; Dunn, J. J.; McAllister, W. T., Pausing and termination by bacteriophage T7 RNA polymerase. *Journal of molecular biology* **1998,** *280* (2), 201-213.

17. Macdonald, L. E.; Durbin, R. K.; Dunn, J. J.; McAllister, W. T., Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. *Journal of molecular biology* **1994,** *238* (2), 145-158.

18. Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E., PCR with an expanded genetic alphabet. *Journal of the American Chemical Society* **2009,** *131* (41), 14620-14621.

19. Sismour, A. M.; Benner, S. A., The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Research* **2005,** *33* (17), 5640-5646.

20. Leal, N. A.; Kim, H. J.; Hoshika, S.; Kim, M.-J.; Carrigan, M. A.; Benner, S. A., Transcription, reverse transcription, and analysis of RNA containing artificial genetic components. *ACS synthetic biology* **2015,** *4* (4), 407-413.

21. Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A., Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Research* **2006,** *34* (21), 6095-6101.

22. Moran, S.; Ren, R. X. F.; Kool, E. T., A thymidine triphosphate shape analog lacking Watson–Crick pairing ability is replicated with high sequence selectivity. *Proceedings of the National Academy of Sciences* **1997,** *94* (20), 10506-10511.

23. Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S., An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nature Methods* **2006,** *3* (9), 729-735.

24. Leconte, A. M.; Romesberg, F. E., Amplify this! DNA and RNA get a third base pair. *Nature Methods* **2006,** *3* (9), 667-668.

25. Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E., Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *Journal of the American Chemical Society* **2008,** *130* (7), 2336-2343.

26. Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E., Optimization of an unnatural base pair toward natural-like replication. *Journal of the American Chemical Society* **2009,** *131* (9), 3246-3252.

27. Seo, Y. J.; Matsuda, S.; Romesberg, F. E., Transcription of an expanded genetic alphabet. *Journal of the American Chemical Society* **2009,** *131* (14), 5046-5047.

28. Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E., Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proceedings of the National Academy of Sciences* **2012,** *109* (30), 12005-12010.

29. Temiakov, D.; Patlan, V.; Anikin, M.; McAllister, W. T.; Yokoyama, S.; Vassylyev, D. G., Structural basis for substrate selection by T7 RNA polymerase. *Cell* **2004,** *116* (3), 381-391.

30. Oh, J.; Kimoto, M.; Xu, H.; Chong, J.; Hirao, I.; Wang, D., Structural basis of transcription recognition of a hydrophobic unnatural base pair by T7 RNA polymerase. *Nature Communications* **2023,** *14* (1), 195.

31. Meyer, A. J.; Garry, D. J.; Hall, B.; Byrom, M. M.; McDonald, H. G.; Yang, X.; Yin, Y. W.; Ellington, A. D., Transcription yield of fully 2 '-modified RNA can be increased by the addition of thermostabilizing mutations to T7 RNA polymerase mutants. *Nucleic Acids Res.* **2015,** *43* (15), 7480-7488.

32. Oh, J.; Xu, J.; Chong, J.; Wang, D., Structural and biochemical analysis of DNA lesion-induced RNA polymerase II arrest. *Methods* **2019,** *159*, 29-34.

33. Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S., An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat. Methods* **2006,** *3* (9), 729-35.