Published in final edited form as:

Isr J Chem. 2013 June ; 53(6-7): . doi:10.1002/ijch.201300005.

Understanding the Molecular Basis of RNA Polymerase II Transcription

Su Zhang^[a] and Dong Wang^{[a],*}

^[a]University of California, San Diego, Skaggs School of Pharmacy & Pharmaceutical Sciences

Abstract

Synthetic nucleic acid analogues have profoundly advanced our knowledge of DNA and RNA, as well as the complex biological processes that involve nucleic acids. As a pivotal enzyme, eukaryotic RNA polymerase II (Pol II) is responsible for transcribing DNA into messenger RNA, which serves as a template to direct protein synthesis. Chemically modified nucleic acid analogues have greatly facilitated the structural elucidation of RNA Pol II elongation complex and understanding the key chemical interactions governing RNA Pol II transcriptional fidelity. This review addresses major progress in RNA polymerase II mechanistic studies using modified nucleic acid analogues in recent years.

Keywords

transcription; RNA polymerase II; nucleic acids; chemical biology; structural biology

1. Introduction

Nucleic acids are responsible for the genetic information storage, propagation, and regulation in all living organisms ^[1–2]. The intrinsic structural features of nucleic acids provide the molecular basis of specific base pairing recognition (Figure 1A) depicted in Watson and Crick's landmark paper^[3]. Synthetic chemical tools have significantly expanded our understanding of nucleic acids by altering each building block (nucleobase, sugar, and phosphate) of nucleic acids. The increasing number of modifications and versatile functions of synthetic nucleic acid analogues have led to revolutions in many areas of basic research, such as chemistry, biology, material science, and the interfaces between them, as well as innovations in drug discovery, molecular diagnostics and therapeutics^[4–8]. Sixty years after the discovery of double-helix model of DNA structure^[3], advances in nucleic acids chemistry have provided powerful tools for exploring and manipulating the activities of complex living systems.

Despite the versatility of nucleic acid analogues, modern cellular machinery have been evolved to specifically recognize natural DNA and RNA as substrates for genetic information storage and transfer. This is due to the fact that the active site of polymerase has evolved to accommodate the natural nucleotide and it is very selective through forming a network of extensive interactions with nucleic acids to specifically recognize cognate substrates and exclude non-cognate ones. The precision of polymerases thus allows decoding and expressing genetic information in a spatial- and temporal-specific manner. Notably, recent studies reveal that modified substrates with unnatural nucleic acid

*Dong Wang (dongwang@ucsd.edu). University of California, San Diego, Skaggs School of Pharmacy & Pharmaceutical Sciences, 9500 Gilman Drive, La Jolla, CA 92093 (USA) phone: (+1) 858 822 5561 fax: (+1) 858 822 1953.

backbones could be used in enzyme-catalyzed reactions. However, this generally requires extensive engineering of polymerases and reaction conditions with compromised efficiency and fidelity [9].

Eukaryotic RNA Pol II is the key enzyme that is responsible for transcribing DNA into messenger RNA, which serves as a template to direct protein synthesis. RNA Pol II is highly selective in nucleotide triphosphate (NTP) loading and incorporation, which maintains a very low error rate of transcription^[10]. A combination of structural, biochemical, and genetic studies^[11–28] have shed light on our understanding of the RNA polymerase transcription mechanisms at atomic level, as well as the molecular interaction network that is crucial for RNA Pol II transcription. Chemically modified nucleic acids has been widely used as fluorescent probes, crosslinking agents, and structural surrogates to facilitate the elucidation of RNA Pol II structures^[11–12, 16, 29–30], understanding of transcriptional mechanism and its subsequent regulation, along with DNA damage and repair^[31–35]. Since a thorough discussion of RNA Pol II transcription mechanisms is beyond the scope of this review, we direct readers interested in learning more about RNA Pol II catalytic mechanism and transcription regulation to several excellent reviews on this topic^[36–43]. This review will focus on the most recent applications of synthetic nucleic acid analogues to understand the structural basis of transcription and chemical interactions controlling RNA Pol II transcriptional fidelity.

2. Implications of Nucleic Acid Analogues in Structural Elucidation of RNA Pol II Transcription

Landmark work led by Kornberg group provided the first detailed three-dimensional structural information about of RNA Pol II transcribing elongation complex (TEC) was determined by generating RNA Pol II transcribing RNA Pol II in 2001 and revealed key protein features in a Pol II transcribing complex^[44–45]. The first crystal structure elongation complex from a tailed DNA template and specifically withholding the incoming nucleotide triphosphate (UTP)^[44]. The results revealed the existence of an RNA-DNA hybrid at the center of an RNA Pol II transcribing complex. This Pol II elongation complex is at the “pre-translocation” state, in which the newly added nucleotide is still occupying the active site and preventing the next nucleotide addition. In order to proceed to transcription elongation along DNA template for next nucleotide addition, RNA Pol II must translocate from a pre-translocation state one base pair downstream along the DNA template, referred as the post-translocation state (Figure 1B). Westover *et al* reported the first Pol II elongation complex in the post-translocate state using a different approach which involved a direct assembly of RNA/DNA scaffold into Pol II elongation complex^[11]. This structure revealed two novel important features: first, the active site of Pol II complex is empty and poised for next nucleotide addition; second, the RNA and DNA forms a short RNA/DNA hybrid and separated by Pol II lid, rudder and fork loop 1 domains^[11].

A fundamental question in RNA Pol II transcription is how do the correct and mismatched nucleotides bind to the Pol II active site? A major challenge to overcome in structural studies is to separate two crucial events: nucleotide binding and nucleotide addition. These two events are kinetically coupled and complete very rapidly (within 3s time scale), whereas it often takes days to weeks to obtain crystals for structure determination. Two approaches employing chemically modified nucleotides were used to overcome this challenge. These chemically modified nucleotides are able to “trap” Pol II elongation complex at nucleotide binding state and “disallow” nucleotide addition to the 3'-end RNA chain ^[11–12]. In one approach, a chain-terminating 3'-deoxyadenosine triphosphate (3'-dATP, Figure 2, 1) was end-incorporated to the RNA by transcription^[11]. This approach resulted in a transcribing complex at the “post-translocation” state. The crystals were then soaked with either a

matched or mismatched nucleotide^[11]. In both cases, the substrate NTPs were binding RNA Pol II without addition to the RNA chain due to the lacking of 3'-OH at RNA chain. Alternatively, nonhydrolyzable GTP analog guanosine-5'-[(α,β)-methyleno]-triphosphate (GMPCPP, Figure 2, 2) was employed to preventing nucleotide addition^[12, 16]. These structural studies revealed several important key insights in understanding the molecular basis of nucleotide selection and addition by RNA Pol II. First, the structure of TEC with mismatched NTP revealed that the mismatched nucleotide binds to Pol II mainly through its triphosphate moiety coordinating with two Mg^{2+} ions, the base moiety is in an inverted conformation facing away from DNA template, referred as "entry site or E site". Second, the structure TEC with matched NTP revealed that the matched NTP binds to Pol II active site in a different manner, referred as "addition site" (Figure 1C). In this addition site, the base moiety of correct nucleotide forms a Watson-Crick base pair with DNA template base. Third, more importantly, upon the correct NTP binding to the Pol II active site, the trigger loop conformation is switched from an inactive open state to an active closed state^[12]. The extensive network of Pol II residues with all three moieties of correct nucleotide provides the molecular basis of nucleotide selection (Figure 1C). More importantly, the closure of the active site by trigger loop establishes Pol II active site poised for the subsequent Pol II enzyme-catalyzed nucleotide addition reaction^[12-13, 46]. Intriguingly a natural toxin alpha-amanitin slows down Pol II transcription dramatically by trapping the trigger loop in an open inactive conformation^[12-13, 47].

Our understanding of molecular basis of the nucleotide addition cycle (NAC) of Pol II has been greatly advanced by these structural studies employed modified nucleic acid analogues. The NTP substrate first diffuses into Pol II active site through the secondary channel and binds transiently to the entry site^[11, 16, 44]. The bound matched nucleotide can switch to a more stable binding mode by rotating its base to the addition site and form a Watson-Crick base pair with DNA template, whereas the mismatched nucleotide is released from active site. Upon correct NTP binding to the Pol II addition site, the trigger loop conformation is switched from an inactive open state to an active closed state^[12]. The closure of the active site subsequently catalyzes the nucleotide addition reaction^[12-13, 46], followed by release of the pyrophosphate ion (PPi)^[48]. Pol II must translocate one base pair downstream from a pre-translocation state to a post-translocation state in order to proceed to the next nucleotide addition^[34, 44, 48-53] (Figure 3A).

Nucleotide selection and incorporation serves as a major checkpoint for the maintenance of Pol II transcriptional fidelity. In addition, two more checkpoint steps are employed by RNA Pol II to further reduce the frequency of nucleotide misincorporation via slow mismatched extension and proofreading. First, the extension rate from a mismatched RNA/DNA hybrid is extremely slow in comparison to extension from a fully matched RNA/DNA hybrid. Mismatch causes the distortion of RNA/DNA hybrid region and misaligns the 3'-OH for the S_N2 attack of the incoming nucleotide, and thus slows down the next nucleotide addition reaction. Second, misincorporation of RNA often causes Pol II to move in a reverse direction, called backtrack. Pol II can cleave a short oligos containing the mismatched nucleotide from 3'-end via its intrinsic cleavage activity or through TFIIS-stimulated cleavage reaction (Figure 3B). Structural studies of backtracked Pol II and complex with TFIIS revealed that the 3'-end is extruded from active site and shed lights on the proofreading mechanisms^[14-15, 34, 54].

3. Using Nucleotide Analogues to Dissect Chemical Interactions

Controlling Transcriptional Fidelity

While structural studies reveals that correct incorporation of nucleotides into the nascent RNA transcript requires an interaction network involving RNA Pol II residues, NTP

substrate, and template DNA, it is not clear what contributions of individual chemical interactions affect the fidelity of transcription. This is very important not only for us to understand the molecular basis of how Pol II reads DNA template and maintains high transcriptional fidelity on non-damaged DNA template, but also facilitate us to understand how DNA damage affects transcriptional fidelity.

A major challenge is that it is difficult to dissect individual chemical interaction contribution through a direct comparison of a natural nucleotide base and DNA lesions, which often involve the change of several factors (such as functional groups, shape, and size) at the same time. To overcome this challenge, Wang, Kool and coworkers used a series of “hydrogen bond deficient” nucleoside analogues to dissect chemical interactions governing RNA Pol II transcriptional fidelity^[55]. A single nucleotide (T) in the DNA template was substituted with nonpolar thymidine isosteres that closely mimic the shape of thymine but vary by sub-ångstrom increment in size (Figure 4A)^[56–57]. Among all five analogues, F has nearly identical size and shape to T, while H is smaller and L, B, and I are increasingly larger. The overall change in size across this series is only about 1.0 Å. In this way, a direct comparison of absence of strong electrostatic effects at the individual “checkpoint” steps could be examined by omitting any possible alternative hydrogen-bond interactions^[58]. Meanwhile, size effect can be also be evaluated by comparing these five analogues.

The study was systematically performed to evaluate three checkpoint steps that contribute to Pol II transcriptional fidelity (Figure 3C): 1) nucleotide selection and incorporation, 2) nucleotide extension, 3) proofreading cleavage. To test the role of hydrogen bond in the first checkpoint step of Pol II transcription, the active RNA Pol II transcribing complex at post-translocation state was assembled with a nonpolar isostere at the +1 position of template DNA. Substantial incorporation of ATP against the hydrogen bond contribution of nucleotide incorporation for RNA Pol II is in sharp contrast to that of DNA polymerases, which shows that lacking of hydrogen bonds has relatively minor effect on deoxyribose nucleotide incorporation^[59–65].

To test the role of hydrogen bond in the second checkpoint step of Pol II transcription, the extension of RNA transcripts beyond nonpolar template bases were investigated. Surprisingly, replacing A:T base pair with A:F or A:I (match) base pairs does not drastically affect CTP incorporation specificity. These findings are in sharp contrast to the significant role of Watson-Crick hydrogen bonding in the first checkpoint step, revealing that the requirement of hydrogen bonding in Pol II transcription is nonpolar analogues and further RNA transcript elongation have been observed for all five substituents, although the rate of incorporation is much slower comparing with that using a DNA template containing canonical T at the same position. The substrate selectivity of Pol II with templates containing T and other five DNA analogs has been further tested by incubating Pol II transcribing complexes with ATP, UTP, CTP and GTP, respectively. ATP gave the highest incorporation efficiency among the four substrates. A significant amount of UTP incorporation against F, L, B, I, and T templates was also observed, whereas no detectable UTP incorporation with H template was observed. There is hardly any GTP incorporation being observed for H, F, L, B, and I templates. In contrast, a substantial amount of GTP was incorporated against natural T template by forming a wobble pair. Lastly, almost no CTP incorporation could be detected for all DNA templates. The substrate specificities were then systematically measured by pre-stead state kinetic studies. Strikingly, Pol II showed essentially no discrimination between ATP and UTP for F, L, B, and I templates (~1.0) and moderate discrimination for H template (~100 fold), while there is a 4.4×10^5 fold selectivity of ATP over UTP for T template, indicating a significant role of hydrogen bonding in RNA Pol II substrate recognition. Kinetic studies also demonstrated that analogue substitution greatly reduced the matched nucleotide substrate k_{pol} values while

only causing several folds of reduction in apparent binding affinity. Thus, the reducing specificity constant ($k_{\text{pol}}/k_{\text{d,app}}$) is attributed more to decreased incorporation rates than lower substrate binding affinities. This indicates that hydrogen bonding might have a stronger effect on correctly positioning the incoming NTP for phosphodiester bond formation than the binding of the substrate. The step specific. Much more surprising results were seen in the extension from mismatched end. A 250- and 1030-fold specificity increase was observed when CTP was incorporated on the U:F and U:I template (mismatch), respectively. This result reveals, for the first time, the contribution of wobble pair hydrogen bonding in preventing error extension. Taken together, these halogen analogue substitutions caused a significant decrease in discrimination of a matched over a mismatched 3' RNA terminus. A U:T mismatch at 3' end reduces CTP incorporation efficiency by $\sim 10^5$ fold, while U:F and U:I mismatch only result in a decrease of extension efficiency by $\sim 10^2$ fold. These results suggest that while Watson-Crick hydrogen bonding might not be necessary for efficient transcript elongation from a matched 33 RNA end in the second checkpoint step, wobble hydrogen bonding is critical for discrimination against mismatched pairs.

To test the role of hydrogen bond in the third checkpoint step of Pol II transcription, the TFIIIS-stimulated cleavage assays were used to examine the Pol II proofreading through backtracking and cleavage of mismatches at the 3' end of RNA. As expected, Pol II complex with a canonical A:T base pair resulted in the slowest cleavage rate and a U:T mismatch gave a fast cleavage rate. Pol II complex containing analogues H and F produced faster cleavage with both matched (A) and mismatched (U) 3' RNA residues. This indicates that hydrogen bonds are important in proofreading control. On the other hand, Pol II complex containing larger substituent L, B, and I resulted in slow cleavage rates comparable with A:T base pair, regardless of the terminal base pairs. The increased base stacking of these analogues might compensate for the loss of hydrogen bonds at the 3' end of RNA. Thus, the absence of hydrogen bonding and/or weak base stacking both appear to promote RNA backtracking.

Collectively, the study demonstrated that loss of hydrogen bonding with template base results in a lower discrimination between correct and incorrect substrate nucleotide, an increase of mismatch extension, and an increase of cleavage rate from a mismatched 3'-RNA terminus. On the other hand, increased size and base stacking causes a universal decrease in NTP incorporation and cleavage, as well as an increase in extension from a mismatched 3'-RNA terminus.

4. Exploring RNA Transcription Using Noncanonical DNA Nucleobases

As a well-characterized epigenetic modification, DNA methylation serves as an important epigenetic mechanism for gene expression regulation^[66–67]. DNA methylation involves the transfer of a methyl group to the 5 position of cytosine to produce 5-methylcytosine (5mC). More recent research have identified that the Tet family of cytosine oxygenase is responsible for oxidizing 5mC into 5-hydroxymethyl-cytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)^[68–70] (Figure 4B). These DNA methylation variants are proposed to be involved in DNA demethylation pathways^[7, 70]. However, the exact roles of these DNA methylation variants in epigenetic and transcription regulations remain to be established.

Wang and coworkers start to address the question whether Pol II transcription machinery can recognize the subtle change of DNA template base during transcription, motivated by noticing that the new functional groups introduced in 5-hydroxymethyl-cytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) are dramatically different from chemical-inert methyl group (5mC) or hydrogen atom from unmodified cytosine. These new

functional groups could potentially interact differentially with protein residues although they do not interfere with Watson-Crick base pair. To understand the functional effects of five different forms of cytosine in RNA Pol II transcription, DNA templates containing a site-specific substitution of C with one of the variants mentioned above were used to systematically evaluate the rate and substrate specificity during transcription process^[71]. Intriguingly, GTP polymerization rate and substrate specificity against 5fC and 5caC templates were significantly lower than C template, while no noticeable difference was observed among C, 5mC and 5hmC templates. In addition, 5fC and 5caC templates resulted in an increased biphasicity of product formation. This result suggests an equilibrium between two interconvertible states of the RNA Pol II complex: one is poised for rapid GTP incorporation and the other is paused and requires a longer time for GTP incorporation. The latter phase may reflect a rate-limiting isomerization that must occur before substrate incorporation. Furthermore, TFIIIS-mediated cleavage rates for 5fC and 5caC templates are increased comparing with that for C template while there are no remarkable changes in the case of 5mC and 5hmC. Thus, 5fC and 5caC may greatly shift the complex from an active state to a paused state and promote RNA Pol II transient pausing and backtracking. In terms of transcriptional fidelity, ATP misincorporation efficiency is significantly higher than that of UTP and CTP across all five cytosine forms, although it is still relatively low comparing with correct GTP incorporation. 5fC also reduces RNA Pol II substrate discrimination of GTP over ATP by a factor of ~30 comparing with natural C template due to a decreased GTP incorporation specificity.

Taken together, these results indicate that RNA Pol II can indeed differentiate subtle changes in cytosine modifications at 5 position and process them differently, possibly by specific interactions between formyl or carboxyl groups on the base and nearby residues of the RNA Pol II subunits. Decreased rates of RNA Pol II transcription in the templates lead to transient pausing at 5fC and 5caC modification sites. This pausing grants RNA Pol II enough time to interact with other regulatory elements in order to achieve fine-tuned transcriptional dynamics and co-transcriptional process such as splicing. These findings provide new insights into the potential functional interplay between cytosine modification status and transcription activities. A recent genome-wide mapping study also reveal a strong correlation of 5fC sites and high peaks of Pol II binding sites (presumed pausing sites) in the gene bodies, which is consistent with in vitro studies^[72].

5. Summary and Outlook

Synthetic nucleic acid analogues have been broadly used to study the structure and mechanism of RNA Pol II transcription. The utilization of synthetic nucleotide analogues in structural studies of Pol II transcribing complex provides important structural insights of enzyme-substrate recognition and catalytic mechanisms. We and others have revealed NTP loading and selectivity, conformational change at the active site upon substrate binding, and different translocation states during transcription elongation. Moreover, detailed biochemical studies of RNA Pol II activities encountering DNA template containing noncanonical bases also allow us to deepen our understanding of the molecular basis of Pol II transcription fidelity and gain insights into how RNA Pol II reads and discriminates between DNA template and nucleotide substrates with subtle structural differences.

We shall take full advantage of novel synthetic nucleic acid probes to investigate the molecular mechanism of the substrate recognition and catalytic mechanisms of RNA Pol II transcription that cannot be addressed by conventional biochemical or genetic approaches. For example, how does sugar pucker of the substrate influence its binding and incorporation? What kind of structural criteria are needed to induce the correct trigger loop folding and movement? How does the RNA Pol II respond to different epigenetic

modifications and what are the resulting consequences to gene regulation and expression? Understanding these questions serves as the starting point for deciphering the mysteries of the genome and discovering novel knowledge and new drug targets leading to the treatment of transcription related diseases.

Acknowledgments

The authors thank members of the Wang lab for helpful comments and suggestions. D.W. acknowledges the NIH (GM085136) and (GM102362_01A1), Kimmel Scholars award from the Sidney Kimmel Foundation for Cancer Research, start-up funds from Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD and Academic Senate Research Award from UCSD.

References

1. Joyce GF. PLoS Biol. 2012; 10
2. Crick F. Nature. 1970; 227:561–563. [PubMed: 4913914]
3. Watson JD, Crick FHC. Nature. 1953; 171:737–738. [PubMed: 13054692]
4. Benner SA. Acc Chem Res. 2004; 37:784–797. [PubMed: 15491125]
5. Lin C, Liu Y, Yan H. Biochemistry. 2009; 48:1663–1674. [PubMed: 19199428]
6. Carell T, Brandmayr C, Hienzsch A, Muller M, Pearson D, Reiter V, Thoma I, Thumbs P, Wagner M. Angew Chem, Int Ed. 2012; 51:7110–7131.
7. Song CX, Yi CQ, He C. Nat Biotechnol. 2012; 30:1107–1116. [PubMed: 23138310]
8. Opalinska JB, Gewirtz AM. Nat Rev Drug Discov. 2002; 1:503–514. [PubMed: 12120257]
9. Chaput JC, Yu HY, Zhang S. Chem Biol. 2012; 19:1360–1371. [PubMed: 23177191]
10. Ninio J. Biochimie. 1991; 73:1517–1523. [PubMed: 1805967]
11. Westover KD, Bushnell DA, Kornberg RD. Cell. 2004; 119:481–489. [PubMed: 15537538]
12. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Cell. 2006; 127:941–954. [PubMed: 17129781]
13. Kaplan CD, Larsson KM, Kornberg RD. Mol Cell. 2008; 30:547–556. [PubMed: 18538653]
14. Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, Kornberg RD. Science. 2009; 324:1203–1206. [PubMed: 19478184]
15. Kettenberger H, Armache KJ, Cramer P. Cell. 2003; 114:347–357. [PubMed: 12914699]
16. Kettenberger H, Armache KJ, Cramer P. Mol Cell. 2004; 16:955–965. [PubMed: 15610738]
17. Kashkina E, Anikin M, Brueckner F, Pomerantz RT, McAllister WT, Cramer P, Temiakov D. Mol Cell. 2006; 24:257–266. [PubMed: 17052459]
18. Zenkin N, Yuzenkova Y, Severinov K. Science. 2006; 313:518–520. [PubMed: 16873663]
19. Koyama H, Ueda T, Ito T, Sekimizu K. Genes Cells. 2010; 15:151–159. [PubMed: 20088966]
20. Walmacq C, Kireeva ML, Irvin J, Nedialkov Y, Lubkowska L, Malagon F, Strathern JN, Kashlev M. J Biol Chem. 2009; 284:19601–19612. [PubMed: 19439405]
21. Kireeva ML, Nedialkov YA, Cremona GH, Purtov YA, Lubkowska L, Malagon F, Burton ZF, Strathern JN, Kashlev M. Mol Cell. 2008; 30:557–566. [PubMed: 18538654]
22. Koyama H, Ito T, Nakanishi T, Sekimizu K. Genes Cells. 2007; 12:547–559. [PubMed: 17535246]
23. Bregeon D, Doddridge ZA, You HJ, Weiss B, Doetsch PW. Mol Cell. 2003; 12:959–970. [PubMed: 14580346]
24. Thomas MJ, Platas AA, Hawley DK. Cell. 1998; 93:627–637. [PubMed: 9604937]
25. Jeon C, Agarwal K. Proc Natl Acad Sci USA. 1996; 93:13677–13682. [PubMed: 8942993]
26. Nesser NK, Peterson DO, Hawley DK. Proc Natl Acad Sci USA. 2006; 103:3268–3273. [PubMed: 16492753]
27. Reines D, Conaway RC, Conaway JW. Curr Opin Cell Biol. 1999; 11:342–346. [PubMed: 10395562]
28. Erie DA, Yager TD, von Hippel PH. Annu Rev Biophys Biomol Struct. 1992; 21:379–415. [PubMed: 1381976]

29. Touloukianov I, Zhang J, Palangat M, Landick R. *Mol Cell*. 2007; 27:406–419. [PubMed: 17679091]
30. Vassilyev DG, Vassilyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. *Nature*. 2007; 448:163–168. [PubMed: 17581591]
31. Brueckner F, Hennecke U, Carell T, Cramer P. *Science*. 2007; 315:859–862. [PubMed: 17290000]
32. Damsma GE, Alt A, Brueckner F, Carell T, Cramer P. *Nat Struct Mol Biol*. 2007; 14:1127–1133. [PubMed: 17994106]
33. Damsma GE, Cramer P. *J Biol Chem*. 2009; 284:31658–31663. [PubMed: 19758983]
34. Walmacq C, Cheung AC, Kireeva ML, Lubkowska L, Ye C, Gotte D, Strathern JN, Carell T, Cramer P, Kashlev M. *Mol Cell*. 2012; 46:18–29. [PubMed: 22405652]
35. Wang D, Zhu G, Huang X, Lippard SJ. *Proc Natl Acad Sci U S A*. 2010; 107:9584–9589. [PubMed: 20448203]
36. Young RA. *Annu Rev Biochem*. 1991; 60:689–715. [PubMed: 1883205]
37. Sims RJ 3rd, Belotserkovskaya R, Reinberg D. *Genes Dev*. 2004; 18:2437–2468. [PubMed: 15489290]
38. Svejstrup JQ. *Biochim Biophys Acta*. 2013; 1829:1. [PubMed: 23312801]
39. Zhou Q, Li T, Price DH. *Annu Rev Biochem*. 2012; 81:119–143. [PubMed: 22404626]
40. Kaplan CD. *Biochim Biophys Acta*. 2013; 1829:39–54. [PubMed: 23022618]
41. Liu X, Bushnell DA, Kornberg RD. *Biochim Biophys Acta*. 2013; 1829:2–8. [PubMed: 23000482]
42. Svetlov V, Nudler E. *Biochim Biophys Acta*. 2013; 1829:20–28. [PubMed: 22982365]
43. Martinez-Rucobo FW, Cramer P. *Biochim Biophys Acta*. 2013; 1829:9–19. [PubMed: 22982352]
44. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. *Science*. 2001; 292:1876–1882. [PubMed: 11313499]
45. Cramer P, Bushnell DA, Kornberg RD. *Science*. 2001; 292:1863–1876. [PubMed: 11313498]
46. Sosunov V, Sosunova E, Mustaev A, Bass I, Nikiforov V, Goldfarb A. *EMBO J*. 2003; 22:2234–2244. [PubMed: 12727889]
47. Brueckner F, Cramer P. *Nat Struct Mol Biol*. 2008; 15:811–818. [PubMed: 18552824]
48. Da LT, Wang D, Huang X. *J Am Chem Soc*. 2012; 134:2399–2406. [PubMed: 22206270]
49. Svetlov V, Nudler E. *Curr Opin Struct Biol*. 2009; 19:701–707. [PubMed: 19889534]
50. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM. *Nature*. 2005; 438:460–465. [PubMed: 16284617]
51. Guo M, Xu F, Yamada J, Egelhofer T, Gao Y, Hartzog GA, Teng M, Niu L. *Structure*. 2008; 16:1649–1658. [PubMed: 19000817]
52. Kornberg RD. *Proc Natl Acad Sci U S A*. 2007; 104:12955–12961. [PubMed: 17670940]
53. Liu X, Bushnell DA, Silva DA, Huang X, Kornberg RD. *Science*. 2011; 333:633–637. [PubMed: 21798951]
54. Cheung AC, Cramer P. *Nature*. 2011; 471:249–253. [PubMed: 21346759]
55. Kellinger MW, Ulrich S, Chong JN, Kool ET, Wang D. *J Am Chem Soc*. 2012; 134:8231–8240. [PubMed: 22509745]
56. Kim TW, Kool ET. *J Org Chem*. 2005; 70:2048–2053. [PubMed: 15760186]
57. Kim TW, Delaney JC, Essigmann JM, Kool ET. *Proc Natl Acad Sci U S A*. 2005; 102:15803–15808. [PubMed: 16249340]
58. Khakshoor O, Wheeler SE, Houk KN, Kool ET. *J Am Chem Soc*. 2012; 134:3154–3163. [PubMed: 22300089]
59. Moran S, Ren RXF, Rumney S, Kool ET. *J Am Chem Soc*. 1997; 119:2056–2057. [PubMed: 20737028]
60. Kim TW, Briebe LG, Ellenberger T, Kool ET. *J Biol Chem*. 2006; 281:2289–2295. [PubMed: 16311403]
61. Sintim HO, Kool ET. *Angew Chem, Int Ed*. 2006; 45:1974–1979.
62. Silverman AP, Garforth SJ, Prasad VR, Kool ET. *Biochemistry*. 2008; 47:4800–4807. [PubMed: 18366188]

63. Wolfle WT, Washington MT, Kool ET, Spratt TE, Helquist SA, Prakash L, Prakash S. Mol Cell Biol. 2005; 25:7137–7143. [PubMed: 16055723]
64. Washington MT, Helquist SA, Kool ET, Prakash L, Prakash S. Mol Cell Biol. 2003; 23:5107–5112. [PubMed: 12832493]
65. Mizukami S, Kim TW, Helquist SA, Kool ET. Biochemistry. 2006; 45:2772–2778. [PubMed: 16503632]
66. Jaenisch R, Bird A. Nat Genet. 2003; 33(Suppl):245–254. [PubMed: 12610534]
67. Nudler E. Annu Rev Biochem. 2009; 78:335–361. [PubMed: 19489723]
68. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Science. 2011; 333:1300–1303. [PubMed: 21778364]
69. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T. Angew Chem, Int Ed. 2011; 50:7008–7012.
70. Wu H, Zhang Y. Genes Dev. 2011; 25:2436–2452. [PubMed: 22156206]
71. Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. Nat Struct Mol Biol. 2012; 19:831–833. [PubMed: 22820989]
72. Raiber EA, Beraldi D, Ficiz G, Burgess HE, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S. Genome Biol. 2012; 13:R69. [PubMed: 22902005]

Biographies

Dong Wang received his B.S. in 1998 from Peking University (Beijing University), China. He then worked with Prof. Stephen Lippard at Massachusetts Institute of Technology on the mechanism of cisplatin action and cellular responses to DNA damage induced by this anti-cancer drug. After his doctoral study, he joined Prof. Roger Kornberg group at the Stanford University to conduct his postdoctoral research on structural studies of RNA polymerase II transcription. He joined the University of California, San Diego in 2010 as an assistant professor. His research focus on understanding the functional interplays among transcription, epigenetic DNA modifications, and DNA damage processing pathways. He has been selected as one of the recipients for NIH Pathway to Independence Award (K99/R00), the Kimmel Scholar Award in 2012 for cancer research, and Leukemia & Lymphoma Society Special Fellow Award.

Su Zhang received his B.S. in 2005 from Zhejiang University, China. He obtained Ph.D. in Chemistry under the supervision of Prof. John Chaput at Arizona State University, where he developed methodologies to perform *in vitro* evolution using threose nucleic acid (TNA). Currently he is carrying out research in the Wang group as a postdoctoral associate on

dissecting chemical interactions controlling transcriptional fidelity using synthetic nucleic acid analogues.

Figure 1.

A) Hydrogen bonding in canonical Watson-Crick base pairs G:C and A:T. B) Cutaway view of RNA Pol II elongation complex (pdb:1R9T). RNA Pol II is revealed as a tan surface. The nascent RNA, template DNA, and non-template DNA are shown in red, cyan, and green, respectively. C) Detailed structure active site of RNA Pol II transcribing complex bound with a matched GTP (PDB: 2E2H). The bridge helix and trigger loop are shown in magenta and cyan, respectively. RNA and DNA are shown in yellow and GTP is green. Side chains are shown as sticks. Nitrogen, oxygen, and phosphate atoms are highlighted in blue, red, and orange, respectively. Mg²⁺ ions are shown in gold.

Figure 2.

Structures of chemically modified nucleotides, 3'-deoxyadenosine 5'-triphosphate (1, 3'-dATP) and guanosine-5'-[(α,β)-methylene]-triphosphate (2, GMPCPP). These modified nucleotides were used for structural studies and mechanism elucidation of the RNA transcribing complexes.

Figure 3.

A three-dimensional scheme of A) forward translocation and B) proofreading, respectively, across different RNA Pol II registers ($n-1$, n , and $n+1$). Different states of RNA Pol II transcription elongation are shown in colored squares: post-translocation state (orange); pre-translocation state (yellow); nucleotide bound entry site (blue); nucleotide bound addition site (green); backtracked state (grey). Red arrows indicate the direction of transcription elongation. Straight blue arrows indicate the reverse translocation and curved blue arrows indicate the RNA cleavage. C) A two-dimensional scheme of different states of RNA/DNA hybrid within RNA Pol II active site during transcription elongation. Three key checkpoint steps for RNA Pol II transcription fidelity are depicted: (1) nucleotide selection and incorporation, (2) RNA transcript extension, and (3) proofreading. DNA and RNA strands in RNA Pol II transcribing complex are shown in green and red. The matched and mismatched nucleotides, and their template base are shown in purple, grey, and yellow, respectively. Correct incorporation and misincorporation are depicted with n and m , respectively. The positions of 3'-end of matched (n) or mismatched (m) RNA are depicted as registers of $n-1$, n , $n+1$, $n+2$, $m+1$, and $m+2$, respectively. The position of next nucleotide addition is shown in a dotted box. The dotted line indicates a very slow reaction.

Figure 4.

A) The structures of nonpolar nucleoside analogues dH (4), dF (5), dL (6), dB (7), dI (8) compared with thymidine (3, T). B) Five forms of cytosine bases. Cytosine (9) is converted into 5mC (10) by DNA methyltransferase (DNMT). Oxidation of 5mC by ten-eleven-translocation (TET) enzymes produces 5hmC (11), 5fC (12), and 5caC (13).