Mechanism of DNA alkylation-induced transcriptional stalling, lesion bypass, and mutagenesis

Liang Xu[a,b], Wei Wang[a], Jiabin Wu[c], Ji Hyun Shin[a], Pengcheng Wang[d], Ilona Christy Unarta[e], Jenny Chong[a], Yinsheng Wang[c,d,1], and Dong Wang[a,f,1]

[a]Division of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093-0625; [b]Department of Chemistry, Sun Yat-Sen University, Guangzhou 510275, China; [c]Environmental Toxicology Graduate Program, University of California, Riverside, CA 92521-0403; [d]Department of Chemistry, University of California, Riverside, CA 92521-0403; [e]Department of Chemistry, The Hong Kong University of Science and Technology, Hong Kong, China; and [f]Department of Cellular and Molecular Medicine, School of Medicine, University of California, San Diego, La Jolla, CA 92093-0625

Alkylated DNA lesions, induced by both exogenous chemical agents and endogenous metabolites, interfere with the efficiency and accuracy of DNA replication and transcription. However, the molecular mechanisms of DNA alkylation-induced transcriptional stalling and mutagenesis remain unknown. In this study, we systematically investigated how RNA polymerase II (pol II) recognizes and bypasses regioisomeric O^2-, $N3$-, and O^4-ethylthymidine (O^2-, $N3$-, and O^4-EtdT) lesions. We observed distinct pol II stalling profiles for the three regioisomeric EtdT lesions. Intriguingly, pol II stalling at O^2-EtdT and $N3$-EtdT sites is exacerbated by TFIIS-stimulated proofreading activity. Assessment for the impact of the EtdT lesions on individual fidelity checkpoints provided further mechanistic insights, where the transcriptional lesion bypass routes for the three EtdT lesions are controlled by distinct fidelity checkpoints. The error-free transcriptional lesion bypass route is strongly favored for the minor-groove O^2-EtdT lesion. In contrast, a dominant error-prone route stemming from GMP misincorporation was observed for the major-groove O^4-EtdT lesion. For the $N3$-EtdT lesion that disrupts base pairing, multiple transcriptional lesion bypass routes were found. Importantly, the results from the present in vitro transcriptional studies are well correlated with in vivo transcriptional mutagenesis analysis. Finally, we identified a minor-groove–sensing motif from pol II (termed Pro-Gate loop). The Pro-Gate loop faces toward the minor groove of RNA:DNA hybrid and is involved in modulating the translocation of minor-groove alkylated DNA template after nucleotide incorporation opposite the lesion. Taken together, this work provides important mechanistic insights into transcriptional stalling, lesion bypass, and mutagenesis of alkylated DNA lesions.

transcription | DNA alkylation | transcriptional lesion bypass | transcriptional mutagenesis | RNA polymerase II

Transcription is the first key step in gene expression, where RNA polymerase II (pol II) is the central enzyme responsible for accurate pre-mRNA and noncoding RNA synthesis in eukaryotic cells. During pol II-mediated transcription elongation, there are at least three transcription fidelity checkpoints to ensure high accuracy in transcription. These fidelity checkpoints include an insertion step for specific nucleotide selection and incorporation, an extension step for preferential extension from a matched over mismatched 3′-RNA terminus, and a proofreading step for preferential removal of misincorporated nucleotides from the 3′-RNA terminus (1) (Fig. 1A).

The integrity of genomic DNA constantly faces numerous attacks from both endogenous and environmental agents, some of which cause significant structural and chemical alterations to DNA (1–11). These DNA lesions, when situated in actively transcribed genomic regions, can significantly perturb pol II-catalyzed transcriptional elongation (1, 5, 6, 10–16). RNA pol II may bypass these lesions, which are sometimes accompanied with nucleotide misincorporation into the nascent transcripts, termed transcriptional mutagenesis, or stall at these DNA lesions, which may initiate a specialized DNA repair pathway that

preferentially repairs damaged DNA on transcribed template strand, a process known as transcription-coupled repair (TCR) (4–6, 13, 17, 18).

Understanding how pol II recognizes DNA lesions constitutes a crucial step toward elucidating downstream pathways and consequences of transcriptional mutagenesis or TCR. Previous studies have revealed the structural basis for a few types of DNA lesions on transcription efficiency, such as those induced by reactive oxygen species (19, 20), UV light (21, 22), and platinum-based drugs (23, 24). While these studies provided important insights into the recognition of certain classes of DNA damage by pol II, little is known about the mechanisms of transcriptional lesion recognition and processing of alkylated DNA lesions (25).

Alkylation represents a major form of DNA damage that has significant implications in public health and has important clinical relevance (26). For example, tobacco-specific nitrosamines, after metabolic activation, can lead to DNA alkylation (27). It was reported that smokers' tissue contains much higher levels of ethylated DNA lesions than nonsmokers (28–35), and the respective levels of $N3$-, O^2-, and O^4-EtdT in leukocyte DNA of smokers are ~224-, ~10-, ~48-fold higher than those of nonsmokers (32). On the other hand, DNA alkylating agents (e.g., temozolomide and streptozotocin) are among the most widely used chemotherapeutic agents for cancer treatment (26).

Significance

DNA alkylation represents a major form of DNA damage that is of high clinical and human health relevance; however, the molecular mechanisms of transcriptional lesion recognition, stalling, and bypass remain unknown. Herein, we carried out a comprehensive investigation to compare the effects of three regioisomeric EtdT lesions on transcription. Intriguingly, we found that the location of alkyl group dictates transcriptional stalling profile and lesion bypass routes, and we identified a novel minor-groove–sensing motif, termed Pro-Gate, which plays an important role in detecting the minor-groove lesion. This work provides important mechanistic insights into DNA alkylation-induced transcriptional stalling and mutagenesis. Our study also provides knowledge about cancer etiology and for the future design of effective cancer chemotherapeutic agents.
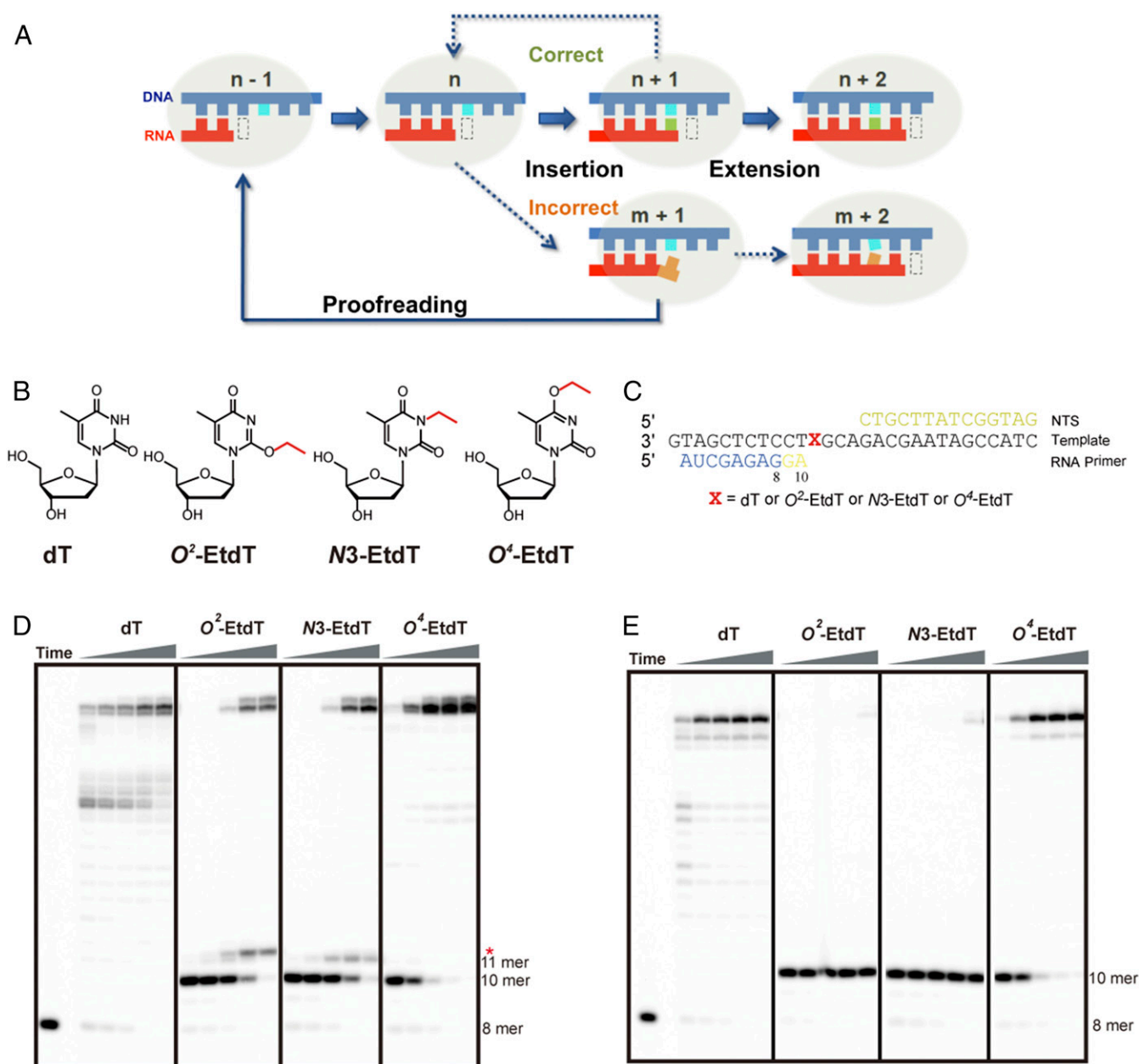
Fig. 1. RNA pol II transcriptional elongation in the damaged template containing an alkylated thymine base. (*A*) A schematic diagram showing the transcription elongation process, and the three steps of transcription checkpoint control, that is, insertion, extension, and proofreading. (*B*) Alkylation of thymine at different positions. (*C*) Scaffolds used in transcription elongation experiments. The position of damaged thymine base is marked as X. (*D* and *E*) Gel analysis of RNA pol II transcriptional elongation in the absence (*D*) and presence (*E*) of TFIIS. The concentration of NTP was 1 mM; the time points were 15 s, 1 min, 5 min, 20 min, and 1 h, respectively. The concentration of TFIIS is 1 μM.

DNA alkylation is unique in that the size of the alkyl groups is diverse, and so are the alkylation sites in DNA. In this vein, the size of alkyl group conjugated to DNA can vary significantly by different alkylating agents. While conjugation of bulky alkyl groups with DNA generally causes severe stalling of pol II-mediated transcription, covalent attachment of small alkyl group with DNA can exert very different effects on transcription (36–44). Additionally, alkyl groups can react with many different sites in DNA, including *N*3 and *N*7 of adenine, O^6 and *N*7 of guanine, *N*3, O^2, and N^4 of cytosine, *N*3, O^2, and O^4 of thymine, as well as the phosphate backbone (45, 46). Intriguingly, the effects of alkylated DNA lesions on transcription appears to be influenced by the locations of the alkyl group on the nucleobase. Indeed, our previous studies using double-stranded plasmids harboring site-specifically inserted and structurally defined ethylated thymidine lesions revealed that the transcriptional bypass efficiency and mutation frequency are modulated by the ethyl group being conjugated with the O^2, *N*3, or O^4 position of thymidine (41, 42). However, the molecular mechanisms of transcriptional lesion bypass (TLB) and mutagenesis of these DNA lesions remain elusive.

Here, we systematically investigated how pol II recognizes and bypasses three regioisomeric ethylated thymidine lesions (O^2-, *N*3-, or O^4-EtdT) (Fig. 1*B*) and how these lesions affect pol II transcription fidelity checkpoints. This systematic investigation on three regioisomeric EtdT lesions provided important mechanistic insights into the recognition of small alkylated DNA lesions by

RNA pol II and revealed different roles of fidelity checkpoints during TLB, which may confer distinct biological consequences. Our study also provides knowledge about cancer etiology and for the future design of effective cancer chemotherapeutic agents.

Results

Distinct Effects of O^2-, $N3$-, and O^4-EtdT on Transcription Elongation Mediated by RNA Pol II. To test the impact of the three regioisomeric O^2-, $N3$-, and O^4-EtdT lesions on RNA pol II-catalyzed transcription elongation, we assembled recombinant pol II elongation complex and performed in vitro transcription assays using the DNA scaffolds containing a downstream site-specific EtdT lesion (47–49) (O^2-, $N3$-, and O^4-EtdT) (Fig. 1C and Figs. S1–S3).

As shown in Fig. 1D, pol II elongation complex was significantly stalled immediately before the O^2-EtdT or $N3$-EtdT site, whereas no stalling was observed at the same position for nondamaged template (dT). Accumulation of 10-mer transcript for the O^2-EtdT– and $N3$-EtdT–harboring templates indicated inefficient nucleotide insertion opposite the lesion sites. Interestingly, after prolonged incubation, pol II was able to move forward by one nucleotide to a second stalling site (11-mer) immediately after the nucleotide incorporation opposite the O^2-EtdT or $N3$-EtdT, suggesting that the extension past the damaged site was also substantially hampered. No further significant stalling was observed once pol II escaped these two pausing/stalling sites. Strikingly, the second pol II stalling band for the O^2-EtdT template (11-mer transcript, marked by a red asterisk in Fig. 1D) exhibited different mobility on the gel from that observed for the $N3$-EtdT template, suggesting that pol II preferentially inserts different nucleotides opposite O^2-EtdT and $N3$-EtdT. In stark contrast to the two strong consecutive stallings of pol II observed for $N3$-EtdT and O^2-EtdT, we only found a single pausing at the 10-mer position for O^4-EtdT. In addition, the bypass across O^4-EtdT site was highly efficient relative to the other two alkylated thymidine derivatives. Taken together, our transcription runoff assay showed that the same size of alkyl group conjugated at the three different positions of dT can exert markedly different effects on transcriptional pausing. While $N3$-EtdT and O^2-EtdT cause two strong consecutive pol II stallings (with a different migration in the second stalling), minimal stalling of RNA pol II elongation was observed for O^4-EtdT.

TFIIS Prevents Transcriptional Bypass of O^2-EtdT and $N3$-EtdT, but Not O^4-EtdT. Transcription factor TFIIS can stimulate transcript cleavage and reactivate backtracked pol II, and therefore it promotes pol II to bypass a variety of DNA lesions and transcription barriers such as pausing sequences and nucleosomes (50–53). Thus, we also examined whether TFIIS could facilitate transcriptional bypass of the three EtdT lesions. Intriguingly, in the presence of TFIIS, we observed prolonged pol II stalling immediately before pol II reaches the O^2-EtdT and $N3$-EtdT sites (10-mer position, Fig. 1E). In sharp contrast to the expected roles of TFIIS in promoting transcriptional bypass, we observed that TFIIS pronouncedly prevents the bypass of O^2-EtdT and $N3$-EtdT, where, in the presence of TFIIS, no full-length products emanating from transcriptional bypass of the two lesions were observed (Fig. 1E, O^2-EtdT and $N3$-EtdT panels, and Fig. S4). Notably, the concentrations of TFIIS we used for in vitro TFIIS-stimulated transcript cleavage assays (200, 400, and 1,000 nM) are comparable to its estimated cellular concentration (~250–400 nM; *Materials and Methods*). This result suggests that the TFIIS-stimulated proofreading activity of pol II (backtrack and transcript cleavage) is much more efficient than the pol II-mediated forward transcriptional bypass of the O^2-EtdT and $N3$-EtdT lesions (i.e., insertion and extension). Consequently, the majority of pol II was stalled at the sites of the two lesions. In contrast to the findings made for the O^2-EtdT and $N3$-EtdT lesions, we observed no apparent effect of TFIIS on the pol II-mediated bypass of O^4-EtdT. As a positive control, we found that TFIIS promotes pol II-mediated bypass of the pausing sequences on the nondamaged template (Fig. 1E, dT panel). Taken together, we found that the stalling of pol II at the O^2-EtdT and $N3$-EtdT sites is exacerbated by TFIIS-stimulated proofreading activity, whereas transcription bypass across O^4-EtdT site is relatively efficient and not affected by the presence of TFIIS.

Impact of EtdT Lesions on the Three Transcriptional Fidelity Checkpoint Steps of RNA Pol II. To further investigate whether the transcriptional fidelity of pol II is altered during lesion bypass, we carried out a systematic investigation on the impacts of the EtdT lesions on the three transcriptional fidelity checkpoint steps: insertion, extension, and proofreading.

For the first fidelity checkpoint step (Fig. 2A, insertion), we observed distinct preferences in nucleotide incorporation opposite the three EtdT lesions (Fig. 2B). As shown in Fig. 2B, with the use of 1 mM nucleoside triphosphates, we observed that AMP could be incorporated opposite all three EtdT lesions. We also detected significant nucleotide misincorporation, to an extent that is comparable to or higher than the canonical AMP incorporation. Intriguingly, GMP was preferentially inserted opposite O^2- and O^4-EtdT, whereas UMP was the most favorable nucleotide inserted opposite $N3$-EtdT (Fig. 2B). To assess quantitatively the nucleotide discrimination in the insertion step, we performed pre–steady-state kinetic experiments under single-turnover conditions to measure the rate constant (k_{pol}) (Fig. S5) and specificity constant (k_{pol}/K_d) for nucleotide insertion opposite all three EtdT lesions or undamaged dT (Fig. 2C, Figs. S5 and S6, and Tables S1 and S2).

These kinetic results, as summarized in Fig. 2C, led to several important observations. First, the rate constants for the incorporations of the correct AMP opposite all three EtdT lesions were greatly diminished (by ~10^5-fold) relative to that for the control undamaged template. Second, we observed unique patterns in nucleotide selection for the three EtdT lesions, indicating that these regioisomeric EtdT lesions elicit distinct impacts on substrate selection in the first checkpoint step (insertion). For the minor-groove O^2-EtdT lesion, AMP and GMP incorporations are strongly favored over CMP and UMP incorporations (Fig. 2C, O^2-EtdT panel). On the other hand, AMP and UMP are preferentially inserted opposite $N3$-EtdT over GMP and CMP (Fig. 2C, $N3$-EtdT panel). For the major-groove O^4-EtdT lesion, incorporation of GMP was highly efficient, that is, at a rate that is ~10-fold higher than that of AMP incorporation (Fig. 2C, O^4-EtdT panel). As a result of this high frequency of GMP misincorporation, O^4-EtdT was the easiest among the three EtdT lesions to be bypassed by pol II. These results indicate the functional importance of regioisomeric alkylation in DNA, where alkylation at different positions of thymine base exerts pronouncedly distinct effects on substrate nucleotide selectivity in the insertion step, thereby differentially modulating transcriptional fidelity in the first step of fidelity checkpoint.

Next, we investigated the impacts of these EtdT lesions on the second step of fidelity checkpoint (Fig. 3A, extension). We found that, for all three EtdT lesions, the efficiency of extension from the matched primer 11A was markedly reduced (by ~10^4- or 10^5-fold) relative to that for the undamaged dT template (Fig. 3B, 11A columns). Furthermore, we observed that the presence of DNA lesions significantly changes the efficiency of extension from mismatched primers. For the undamaged template (dT), apart from the efficient extension from the matched primer 11A, we observed that extension from the mismatched 11G primer was also relatively efficient (~10^{-1} $\mu M^{-1} \cdot min^{-1}$, only ~100-fold less than extension from 11A), whereas extension from the mismatched 11U or 11C primer was extremely inefficient (10^5-fold and 10^4-fold less than extension from 11A, respectively) (Fig. 3B, dT panel). In stark contrast, we found that the presence of O^2-EtdT or $N3$-EtdT greatly disfavored the extension from the mismatched 11G (~10^{-6} $\mu M^{-1} \cdot min^{-1}$, five
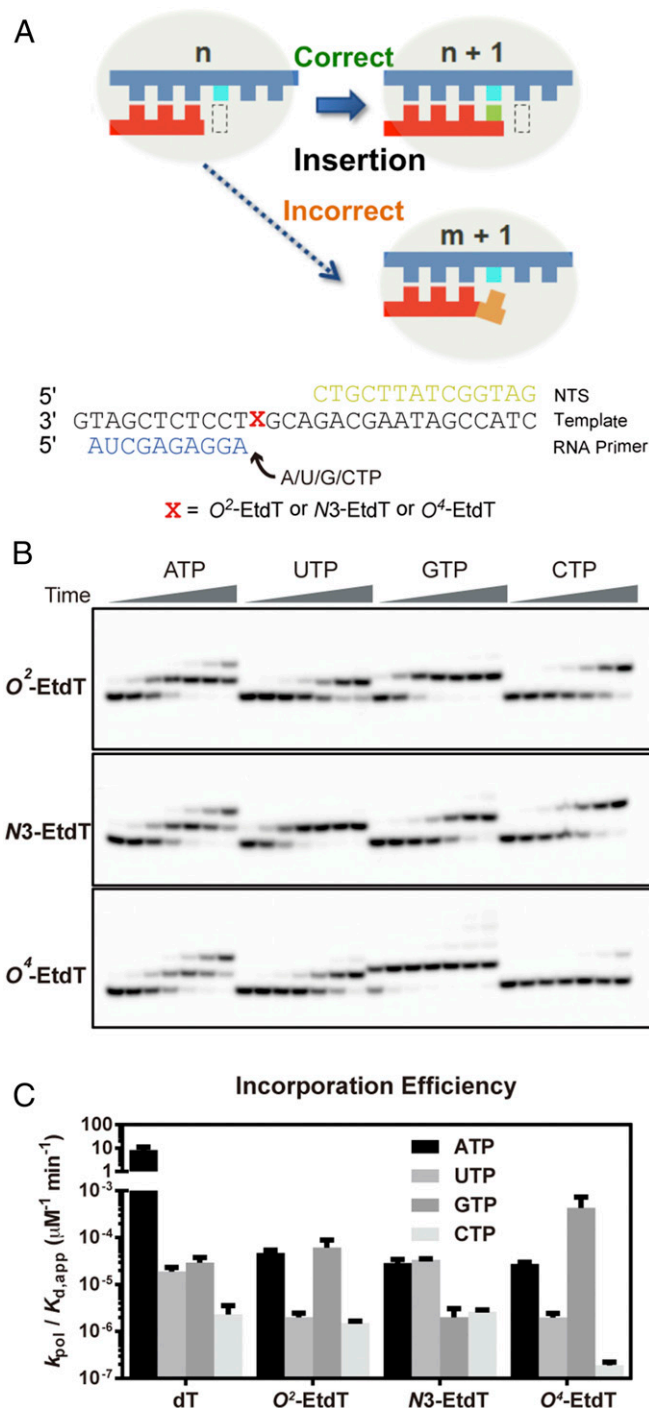
Fig. 2. Nucleotide incorporation opposite the EtdT lesions. (*A*) A scheme illustrating the first fidelity checkpoint step (insertion) and the scaffold used in this assay. (*B*) Representative images of gels for monitoring single-nucleotide addition opposite the three ethylated thymine nucleosides. The concentration of NTP was 1 mM; the time points were 1 min, 5 min, 20 min, 1 h, 3 h, 8 h, and 1 d, respectively. (*C*) Kinetic analysis of single-nucleotide incorporation opposite the ethylated thymidines in comparison with the undamaged thymidine.

orders of magnitude lower than that of 11G extension for the control dT template), whereas these DNA lesions modestly favored 11U and 11C extension (Fig. 3*B*, O^2-EtdT and *N*3-EtdT panels). Strikingly, we found that the extension after GMP incorporation was the most efficient for the O^4-EtdT–containing

template, which is ∼1,000- and 30-fold more efficient than extension from 11A and 11U, respectively (Fig. 3*B*, O^4-EtdT panel). Taken together, the three regioisomeric EtdT lesions significantly alter the extension selectivity in a position-specific manner, thereby influencing differently transcriptional fidelity in the second fidelity checkpoint step.

We then assessed the third transcription fidelity checkpoint step (Fig. 4*A*, proofreading). To this end, we first performed intrinsic transcript cleavage assay, which not only allowed us to measure the transcript cleavage rates but also revealed pol II translocation states based on the transcript cleavage pattern (Fig. 4*B*). It turned out that the presence of EtdT lesions significantly changed the transcript cleavage pattern of pol II complex containing matched 11A. For the matched 11A/dT scaffold, we only observed a single dominant transcript cleavage product (Fig. 4*B*, n-1, black arrows) corresponding to the cleavage from pol II at pretranslocation state (the 3′-terminal rA still occupies the insertion site and forms Watson–Crick base pairing with dT in the template) (Fig. 4*B*, 11A and dT panel). In contrast, we observed a dominant n-2 cleavage product (Fig. 4*B*, red arrows) and a minor n-1 cleavage product (Fig. 4*B*, black arrows) from all three 11A/Et-dT lesion scaffolds, indicating that the presence of the EtdT lesions promotes pol II backtracking, where pol II translocates backward from the pretranslocation to backtracked state and 3′-terminal rA is extruded toward the secondary channel (Fig. 4*B*, 11A panel). In addition, consistent with the previous literature finding that 3′-mismatched
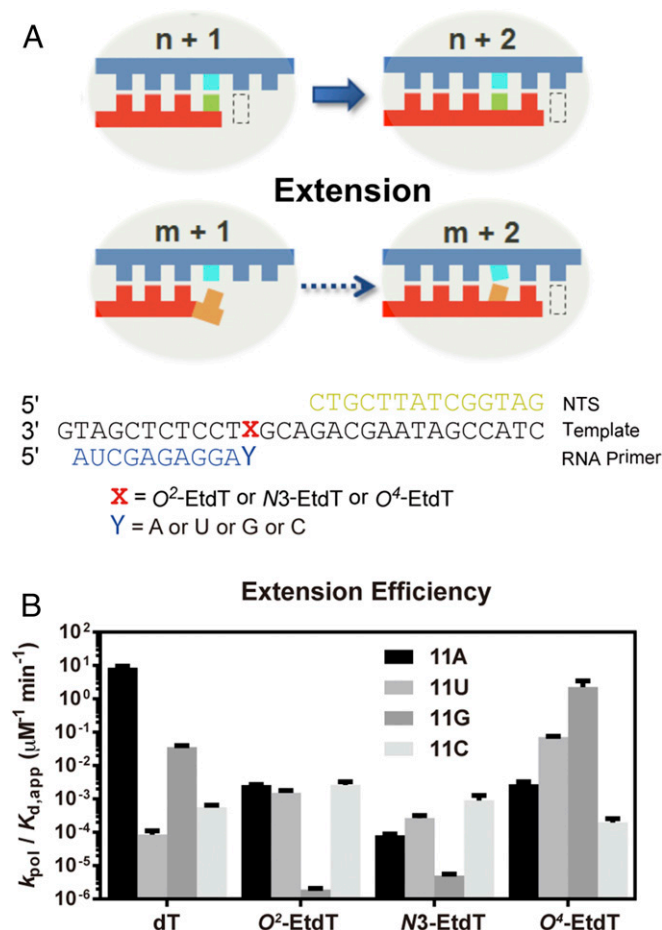

Fig. 3. Analysis of the subsequent extension step after the nucleotide addition opposite the alkylated thymine site. (*A*) Scaffold used in this assay. (*B*) Kinetic analysis of the subsequent extension after the ethylated thymine in comparison with the undamaged thymine base.
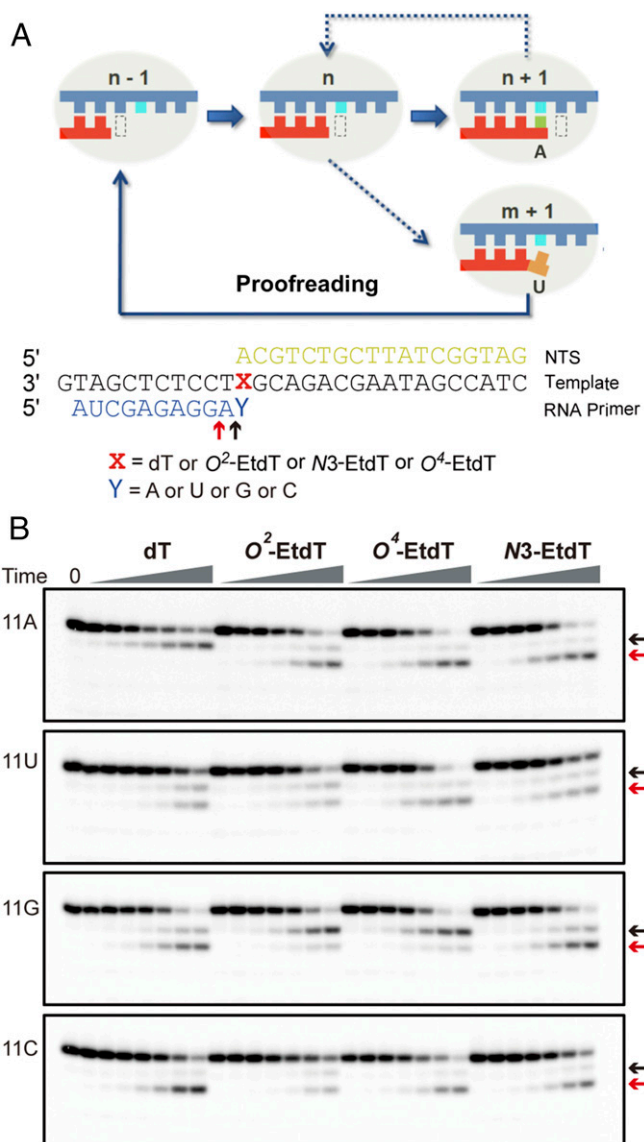
Fig. 4. Backtrack and proofreading of RNA pol II after nucleotide insertion opposite the ethylated thymidine lesions. (A) Scaffold used in the cleavage assay. (B) Results of RNA pol II intrinsic cleavage. The black arrow refers to the pretranslocation cleavage product; the red arrow refers to the backtrack cleavage product. The time points were 5 min, 20 min, 60 min, 3 h, 8 h, 24 h, and 48 h, respectively.

RNA or damaged DNA scaffolds promote pol II backtracking (51), we observed two cleaved transcripts (n-1 and n-2) for all other mismatched or damaged scaffolds (Fig. 4B, 11C, 11G, and 11U panels). Notably, the formation of n-2 cleavage product was the major event observed for all of the mismatched or damaged templates, except for those carrying 11G:O^2-EtdT or 11G:O^4-EtdT (Fig. 4B, 11G). The cleavage products of these two scaffolds mainly arise from pretranslocation cleavage, indicating that their base parings at the 3′-terminus were more stable than other damaged scaffolds. We also performed TFIIS-stimulated cleavage assay and found that, in the presence of TFIIS, pol II can efficiently cleave the transcripts at the damage sites (Fig. S7). Collectively, these results suggest that pol II, in the presence of TFIIS, can efficiently cleave transcripts for the damage-containing scaffolds, but the translocation patterns are altered and depend on the positions on the thymine base where the alkyl groups are situated.

Distinct Transcriptional Bypass Routes of O^2-, $N3$-, and O^4-EtdT Lesions.
By integrating the kinetic data of the three transcriptional fidelity checkpoints for the EtdT lesions, an overall transcriptional bypass scheme can be derived, as summarized in Fig. 5. Alkylation at different positions confers distinct routes for TLB (Fig. 5B). For the DNA template containing a minor-groove O^2-EtdT lesion, both ATP and GTP were favored in the insertion step. However, the subsequent extension after GMP incorporation was extremely disfavored. As a result, AMP incorporation and subsequent extension constituted the major TLB pathway, which essentially maintained the transcriptional fidelity, albeit with compromised bypass efficiency. The lesion bypass routes for the $N3$-EtdT template, however, were very different. In the nucleotide addition step, AMP and UMP could be more readily incorporated than CMP and GMP; nevertheless, the subsequent extension after CMP incorporation was highly efficient. Consequently, all three TLB routes (from AMP, UMP, and CMP) are likely to occur for $N3$-EtdT, where only the GMP pathway is blocked in both the insertion and extension steps. In this scenario, transcriptional fidelity was compromised with potential A→U or A→C mutations (Fig. 5B). Strikingly, a distinct and dominant TLB route was observed for the major-groove O^4-EtdT lesion, where GMP was preferentially inserted in the addition step with over 10-fold higher efficiency than the matched AMP (Fig. 2). Furthermore, the extension after GMP incorporation was remarkably efficient, which is ~1,000- and 100-fold more efficient than the extension after AMP and UMP insertions, respectively (Fig. 3). In addition, the efficiency in extension after GMP insertion is similar to what was observed for the matched nucleotide extension for the undamaged template (Fig. 3). Hence, the GMP pathway predominates the lesion bypass route for O^4-EtdT, and the transcriptional fidelity is almost abolished with a high frequency of A-to-G mutation.

To quantify the alteration in fidelity for TLB, we introduced an index, termed TLB efficiency (see *Materials and Methods* for more details), to gauge the overall lesion bypass efficiency by combining the rates for all three fidelity checkpoints (i.e., insertion, extension, and proofreading). Considering that the proofreading rates for all three lesions are similar (Fig. 4 and Fig. S7), the TLB value is thus mainly controlled by the insertion and extension steps, and it reflects the likelihood of a particular translesion bypass route or stalling. By normalizing the TLB value with "error-free" TLB value for each template (e.g., the AMP route; see *Materials and Methods* for more details), we can directly compare the different TLB routes for the three lesion-bearing DNA templates, as summarized in Fig. 5C. Therefore, for each lesion, we can quantitatively evaluate the likelihood of particular TLB routes (Fig. 5C). A value above 1 indicates that this specific route is more efficient than the error-free TLB route, whereas a value below 1 suggests that this specific route is disfavored. As shown in Fig. 5C, error-free route (AMP route) is strongly favored for transcriptional bypass of the minor-groove O^2-EtdT lesion, whereas a dominant error-prone route (GMP route) is preferred for transcriptional bypass of the major-groove O^4-EtdT lesion. For the $N3$-EtdT lesion that disrupts base pairing, multiple transcriptional bypass routes are possible, which could be error free (AMP route) or error prone (UMP and CMP routes).

Importantly, the aforementioned in vitro transcription results are well correlated with what we observed previously from cell-based transcription studies (41). We found that, in nucleotide-excision repair-deficient human skin fibroblasts, O^4-EtdT induces exclusively A→G mutation and $N3$-EtdT triggers A→C and A→U mutations, whereas the nucleotide incorporation opposite O^2-EtdT is promiscuous (41) (Fig. 5D). Hence, the results obtained from the present in vitro biochemical study are remarkably similar to those from cell-based study, especially in the context that the in vitro assay was conducted with purified RNA pol II in the presence of one nucleotide at a time, whereas
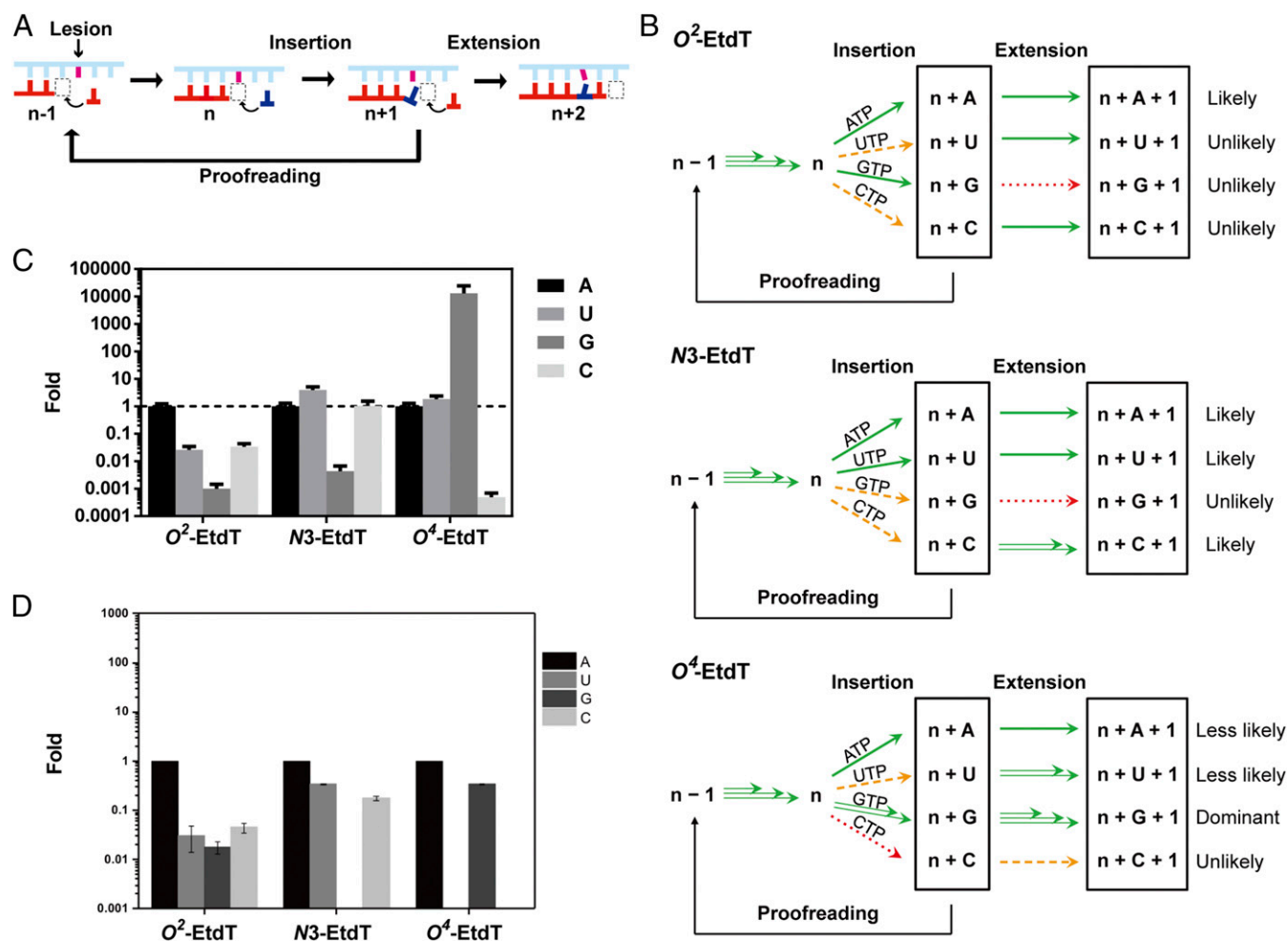
Fig. 5. Alkylation at different positions at the hydrogen bonding face of thymidine leads to distinct bypass preferences by RNA pol II. (*A*) Lesion bypass was controlled by different fidelity checkpoints. (*B*) A summary of transcription lesion bypass for the three regioisomeric EtdT lesions. Green, orange, and red arrows designate ribonucleotide incorporations that are efficient, difficult, and extremely difficult, respectively. This scheme was depicted based on the relative kinetic values of each transcriptional bypass step. (*C*) A comprehensive analysis of relative bypass efficiency of RNA pol II for the three regioisomeric EtdT lesions. (*D*) The distributions of nucleotides inserted opposite the regioisomeric EtdT lesions based on transcription assays conducted in XPA-deficient human skin fibroblasts, where the data represent the mean and SE of results from three independent transfection experiments (the panel was plotted based on data reported in ref. 41).

the cellular transcription reaction is mediated by the entire human transcription machinery in the mutual presence of all four nucleotides.

Structural Insights into Lesion Bypass and Stalling of Major- and Minor-Groove Lesions. The above results showed that the pol II-mediated bypass routes for major-groove and minor-groove alkylation lesions are distinct; while GMP insertion is permitted for both O^2-EtdT and O^4-EtdT in the insertion step (Fig. 2C), the subsequent extension after the GMP insertion is extremely inefficient for O^2-EtdT, but strongly favored for O^4-EtdT (Fig. 3B). To further understand how pol II distinguishes the GMP route for O^2-EtdT and O^4-EtdT lesions during the insertion and extension steps, we modeled, based on previous structural studies (54, 55), the active site of the pol II elongation complex when it encounters the O^2-EtdT and O^4-EtdT lesions, followed by energy minimization (see *Materials and Methods* for details) (Fig. 6). Molecular-modeling results indicate that guanine can form wobble base pairs with both O^2-EtdT and O^4-EtdT in different orientations (Fig. 6B). These hydrogen bonds may stabilize the rG:O^2-EtdT or rG:O^4-EtdT pair, which also finds its support from the intrinsic cleavage patterns as shown in Fig. 4B. Thus, these wobble base pairs at the +1 site of pol II active center could favor

GMP binding and addition in the first insertion step (Fig. 2). Note that because the efficiency of GMP addition is several orders of magnitude lower than that of canonical ATP incorporation for the undamaged dT template (trigger loop-dependent nucleotide addition, in which the trigger loop adopts an active, closed conformation to catalyze nucleotide addition) (56), we speculate that nucleotide addition opposite the lesion likely occurs in a trigger loop-independent manner, in which the trigger loop remains in inactive, open conformations (1, 57).

To proceed with extension after GMP incorporation, the newly formed rG:O^4-EtdT or rG:O^2-EtdT base pair needs to translocate into the −1 site so as to facilitate the subsequent nucleotide addition. Interestingly, during translocation, the ethyl group in the major and minor grooves of DNA could experience drastically different environments (Fig. 6). Upon translocation of the O^4-EtdT from the +1 to −1 site, the ethyl group, which is located in the major groove of DNA, is well tolerated in the active site of pol II without any steric clash. Hence, subsequent extension from the rG:O^4-EtdT pair is highly efficient (Fig. 3B). In contrast, the ethyl group of the O^2-EtdT lesion, which points to the minor groove, encounters a severe steric clash with P448 in Rpb1 when there is no local rearrangement. P448 is located in a
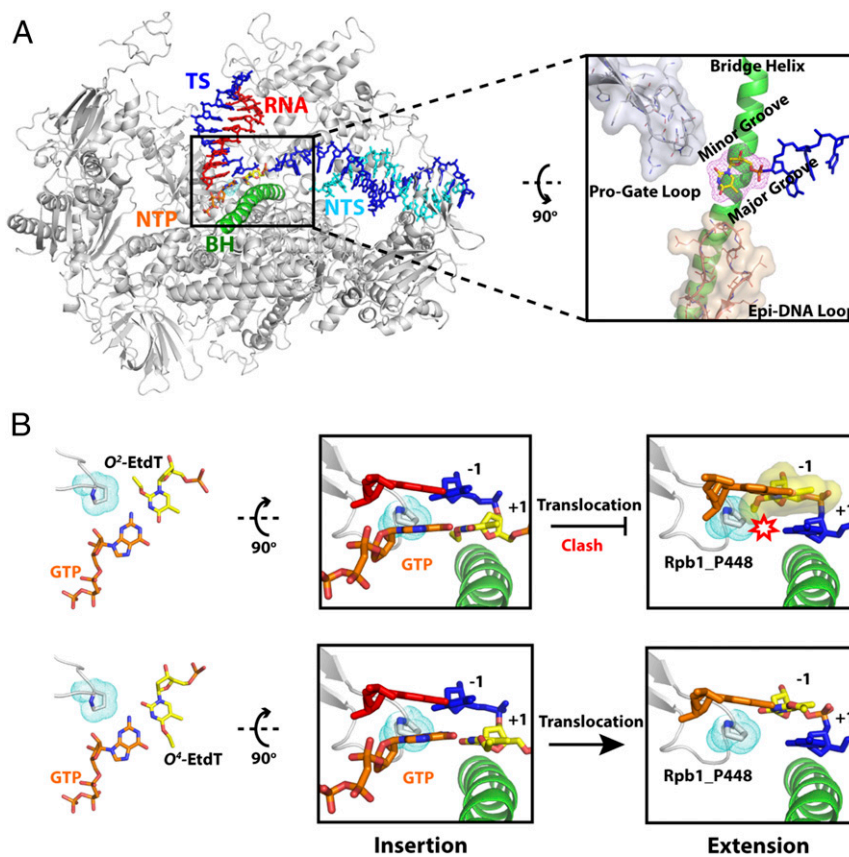
Fig. 6. Transcriptional bypass of O^4- and O^2-EtdT has distinct structural effects. (*A*) During the incorporation of GMP, the guanine base can pair with both O^4- and O^2-EtdT. (*B*) In the extension step, the alkylation in the major groove (O^4-EtdT) has limited disruption during translocation, whereas the minor-groove alkylation (O^2-EtdT) has strong steric clash with the P448 residue, and further altered this minor-groove–interacting loop in Rpb1.

motif consisting of a connection loop between the two β-sheets (β14 and β15) in the active-site domain. Because the P448-containing loop (Rpb1 440–460) faces toward the minor groove of RNA:DNA hybrid and acts as a "steric gate" to prevent or slow down the translocation of nucleotide with alkylation in the minor groove, we termed it the "Pro-Gate" loop. This clash may severely hinder the translocation of the minor-groove O^2-EtdT lesion, which may significantly compromise the extension efficiency (Figs. 3*A* and 6*B*). Notably, our energy minimization results showed that P448 is displaced from the original position to minimize the steric clash, suggesting a potential mechanism for tolerating small minor-groove DNA alkylation and a slow transcriptional bypass (Fig. S8), whereas the presence of bulky minor-groove DNA alkylation could lead to a strong pol II arrest.

Discussion

Our previous plasmid-based cellular transcription assays revealed that the three regioisomeric EtdT lesions exhibited different transcriptional bypass efficiencies and mutation frequencies (41), although the molecular mechanisms underlying the differences were unclear. In this study, we systematically investigated the mechanisms contributing to the regiospecific effects of DNA alkylation on pol II-mediated transcription, elucidating important relationships between DNA lesion structure and transcriptional stalling, lesion bypass, and fidelity. The addition of an ethyl group to different positions (O^2, N3, and O^4) of thymine elicited distinct transcriptional responses of RNA pol II. Here, we revealed how RNA pol II handles regioisomeric DNA alkylation and how such alkylation affects transcriptional fidelity. A comparison of TLB

efficiencies of the three lesions showed that O^2- and N3-alkylation greatly blocked pol II elongation, whereas O^4-alkylation exerted limited blockage effect. The major pausing of O^2- and N3-alkylation occurred in both the insertion and the subsequent extension steps at the damaged site. For O^4-alkylation, only weak pausing at the insertion step was found, and no obvious pausing was observed at the extension step (Fig. 1*D*).

Further mechanistic insights came from our study about the impact of these regioisomeric EtdT lesions on the individual pol II fidelity checkpoints (Fig. 5). Strikingly, although GMP was selected for both O^2-EtdT and O^4-EtdT in the first fidelity checkpoint, the eventual transcriptional readout was different. For O^4-alkylation, a high frequency of A→G mutation was observed, whereas transcriptional fidelity was maintained for the O^2-alkylation template. The opposite consequences were actually elicited by the second fidelity checkpoint (extension step). For the O^4-EtdT template, the extension step was highly efficient owing to pol II's tolerance of alkylation in the DNA major groove. On the contrary, the extension after GMP incorporation opposite O^2-EtdT was extremely difficult, which is potentially attributed to the steric blockage of P488 residue situated in the DNA minor groove during translocation.

Our systematic investigations of alkylation at different positions of thymine base suggested a potentially unique recognition motif of RNA pol II that may serve as a gatekeeper for monitoring the structural integrity of DNA minor groove. Our previous studies unveiled a couple of important recognition motifs in RNA pol II that target the DNA major-groove modifications or bulky, helix-distorting lesions during translocation (10, 11, 58, 59). Furthermore, we recently identified two important residues

(Rpb 1 R1386/H1387) in switch 1 region that are important for the early detection of minor-groove DNA binder Py-Im molecules even before it reaches the downstream bubble edge (59). Here, we found that the Pro-Gate loop containing a highly conserved Pro (i.e., P448 in Rpb 1) residue might play an additional inspective role during translocation to sense the environment of the minor groove after nucleotide incorporation. This postinsertion minor-groove sensor could serve as another checkpoint for minor-groove modifications following the nucleotide insertion step. We proposed that Pro-Gate loop acts as a steric gate to prevent or slow down the translocation of O^2-EtdT lesion after GMP incorporation (GMP route) but permit other extension routes that bypass O^2-EtdT lesions (AMP, UMP, or CMP routes). Based on our intrinsic cleavage assay, we found that GMP is the only nucleotide that can form stable hydrogen bonds with O^2-EtdT in pretranslocation state and thus observed a dominant n-1 cleavage product (Fig. 4B, in 11G:O^2-EtdT panel), whereas other nucleotides (AMP, UMP, or CMP) fail to stably interact with O^2-EtdT. We speculate that the wobble base pairing between guanine and O^2-EtdT may restrain the flexibility of the thymine base, leading to the direct clash between the ethyl group and P448, whereas the pairing of O^2-EtdT with other nucleobases may be flexible enough to allow for local rearrangement of the lesion, thereby avoiding such severe clash. Therefore, we observed extremely slow extension once GMP is incorporated opposite O^2-EtdT, which arises from the blockage of translocation mediated by the Pro-Gate loop and is aggravated by wobble base pairing. Future studies are warranted for understanding the biological roles of P448 in transcribing undamaged template, in modulating the transcriptional bypass and stalling of other minor-groove lesions (e.g., N^2-alkylated dG derivatives) (48), and in stimulating TCR of these lesions. Taken together, our structural analysis provides mechanistic insights into why the extension step blocked the GMP pathway during the bypass of O^2-EtdT, even though this pathway is favorable in the initial insertion step.

The distinct profiles of transcriptional stalling and lesion bypass routes may also confer very different biological consequences. For O^2-EtdT, even though the transcription progression is initially stalled, a subsequent error-free TLB is favored and the transcriptional fidelity is essentially maintained. Therefore, O^2-EtdT may not severely perturb the accuracy in transmission of genetic information during transcription. On the other hand, transcription elongation through N3-EtdT is slow and mutagenic, which can compromise the fidelity of genetic information flow. It is also interesting to note that the pol II transcriptional bypass of N3-EtdT lesion was greatly compromised in the presence of TFIIS in vitro. This persistent TFIIS-induced transcriptional stalling/arrest may provide enough time for the arrested pol II complex to recruit CSB and other repair factors to initiate TC-NER. Indeed, our previous studies showed that N3-alkylated thymidine derivatives could be substrates for TC-NER (41, 42). It would be interesting for future studies to investigate the potential roles of TFIIS in the TC-NER of N3-EtdT lesion. For O^4-EtdT, TLB is relatively efficient and highly mutagenic. Clearly, from our comprehensive analysis of transcriptional effect of alkylation at different positions in thymidine, distinct downstream cellular response could be envisaged. Herein, studies of these regioisomeric alkylated DNA lesions not only elucidated the biochemical basis of transcriptional mutagenesis but also offered fundamental insights into the biological consequences of regioisomeric alkylated thymidine lesions.

Another interesting observation from our study is that the presence of TFIIS causes an elevated transcriptional stalling at N3-EtdT and O^2-EtdT lesions, which is in sharp contrast to the canonical well-established roles of TFIIS in promoting transcriptional bypass of a variety of DNA lesions and transcription barriers such as pausing sequences and nucleosomes (50–53). Mechanis-

tically, TFIIS stimulates the cleavage of backtracked transcript (from the inactive form of pol II) and therefore facilitates the recovery of the backtracked pol II (inactive form) back into the active form of pol II (posttranslocation state) to allow new nucleotide addition. Why does the presence of TFIIS apparently lead to different outcomes (bypass/stalling) for different transcriptional barriers? We think that there is a key difference between the cases of natural pause sites/nucleosomes and O^2-, N3-EtdT covalent DNA lesions that contributes to the different outcomes of pol II bypass in the presence of TFIIS. For the noncovalent transcription barriers, such as natural pause sites or nucleosomes, while these barriers cause a large portion of pol II to backtrack due to steric hindrance, these barriers can "breathe" (such as transient disruption of local histone–DNA contacts for nucleosome barriers) to transiently relieve the steric hindrance and allow pol II to proceed (for new nucleotide addition). In this scenario, the recovery of pol II from the inactive backtracked state is still the rate-limiting step (i.e., the rate of nucleotide incorporation is still much faster than TFIIS-stimulated cleavage rate). As a result, TFIIS promotes the overall pol II's bypass efficiency of these barriers. This is also true in the case of certain types of covalent DNA lesions as long as the recovery of pol II from the inactive backtracked state is still the rate-limiting step. In sharp contrast, for O^2- and N3-EtdT lesions, the forward translocation and nucleotide incorporation step opposite to these lesions is significantly slowed down (by 10^5-fold) and becomes the rate-limiting step (i.e., much slower than the backtracking/cleavage rate). In this scenario, the recovery of pol II from the inactive backtracked state is no longer the rate-limiting step. Because TFIIS stimulates the transcript cleavage and drives the equilibrium further favored for the backtracking/cleavage route over the transcriptional bypass route. As a result, we observed futile cycles of backtracking/TFIIS-stimulated cleavage route that lead to persistent pol II stalling at these DNA lesions.

Alkylating agents are ubiquitously present in the environment and they can also be produced by endogenous metabolism (26). In addition, alkylating agents are among the most frequently prescribed cancer chemotherapeutic agents (26). Hence, alkylation constitutes one of most common forms of DNA damage. The results from the present study suggest that the impact of DNA alkylation on transcriptional efficiency and fidelity could be significantly modulated by the positions in DNA to which the alkyl groups are conjugated. This work reveals that minor-groove DNA alkylation is more effective at blocking pol II transcription than its major-groove DNA alkylation counterpart. Future strategies for the rational design of effective alkylating agents can be envisaged, that is, through modulating the dominant conjugation position and the size of the alkyl group. In addition, altering the level or activity of TFIIS could be another layer to modulate the efficacy and toxicity of DNA alkylating agents. Interestingly, it was reported that modulation of TFIIS level could be a potential strategy for anticancer therapy (60). Thus, our work provides important knowledge for the risk assessment of human exposure to DNA alkylating agents and for guiding the development of cancer therapeutic agents that can maximize their toxicity to tumor cells while minimizing the induction of transcriptional mutagenesis.

Materials and Methods

Preparation of Damaged DNA Templates. Oligodeoxyribonucleotides (ODNs) harboring a site-specifically inserted O^2-, N3-, and O^4-EtdT were synthesized following previously published procedures (49). The identities and purities of the lesion-containing ODNs were confirmed by electrospray ionization–mass spectrometry (ESI-MS) and tandem-MS (MS/MS) analyses (Figs. S1–S3). The synthesized lesion-containing ODNs were subsequently ligated to provide the templates for in vitro transcription studies.

Transcriptional Elongation. The *Saccharomyces cerevisiae* RNA pol II was purified as previously described with IgG affinity column followed by HiTrap Heparin and Mono Q columns for further purification (56, 58). The RNA pol II elongation complexes for in vitro transcription assays were assembled using established methods (57–59, 61–63). Briefly, an aliquot of 5′-^{32}P–labeled RNA (10 μM) was annealed with a 1.5-fold amount of template DNA (15 μM) and twofold amount of nontemplate DNA (20 μM) from 65 °C to room temperature over a 2-h period to form the RNA/DNA scaffold (final stock concentration: 1 μM, defined by RNA concentration) in an elongation buffer, which contained 20 mM Tris·HCl (pH 7.5), 40 mM KCl, and 5 mM MgCl$_2$. An aliquot of the annealed RNA/DNA scaffold (50 nM) was then incubated with a fourfold amount of pol II (200 nM) at room temperature for 10 min to ensure the formation of pol II elongation complex for all RNA/DNA scaffold. The in vitro transcription reaction was initiated by mixing the pol II elongation complex with an equal volume of various twofold concentration of NTP solution. The final reaction mixture contained 25 nM scaffold, 100 nM pol II, 5 mM DTT, 5 mM MgCl$_2$, 40 mM KCl, 20 mM Tris·HCl (pH 7.5), and NTP. The transcription reactions were quenched at various time points by addition of 1 vol of 0.5 M EDTA (pH 8.0). Samples were then mixed with 1 vol of denaturing PAGE running buffer (100% formamide supplemented with 0.25% bromophenol blue and xylene cyanol) and heated to 95 °C for 5 min to denature the scaffolds. The quenched products were then resolved on 16% denaturing urea/TBE polyacrylamide gels and visualized using a storage phosphor screen and quantitated by Molecular Imager PharosFX Plus system (Bio-Rad).

Single-Turnover Nucleotide Incorporation Transcription Assays. The assay was carried out as previously described (61–63). Briefly, nucleotide incorporation assays were conducted by preincubating 50 nM scaffold with 200 nM pol II for 10 min in the elongation buffer at 22 °C. The preincubated enzyme: scaffold complex was then mixed with an equal volume of solution containing 40 mM KCl, 20 mM Tris·HCl (pH 7.5), 10 mM DTT, 10 mM MgCl$_2$, and twofold concentrations of various nucleotides. The final reaction mixture contained 25 nM scaffold, 100 nM pol II, 5 mM MgCl$_2$, and various nucleotide concentrations in the elongation buffer. The transcription reactions were quenched at various time points by addition of 1 vol of 0.5 M EDTA (pH 8.0). Reactions requiring time points shorter than 5 s were quenched using a RQF-3 Rapid Quench Flow (KinTek Corporation) as described (61). Samples were then mixed with 1 vol of denaturing PAGE running buffer (100% formamide supplemented with 0.25% bromophenol blue and xylene cyanol) and heated to 95 °C for 5 min to denature the scaffolds. The quenched products were then resolved on 16% denaturing urea/TBE polyacrylamide gels and visualized using a storage phosphor screen and quantitated by Molecular Imager PharosFX Plus system (Bio-Rad).

Nonlinear-regression data fitting was performed using Prism 6. The time dependence of product formation was fit to a one-phase association equation to determine the observed rate (k_{obs}). The substrate concentration dependence was fit to a hyperbolic equation to obtain values for the maximum rate of NTP incorporation (k_{pol}) and apparent K_d ($K_{d,app}$) governing NTP binding, as previously described (64). The specificity constant was determined by $k_{pol}/K_{d,app}$.

Intrinsic and TFIIS-Stimulated Transcript Cleavage Assays. The elongation complex was assembled as described above in 20 mM Tris·HCl (pH 7.5) without Mg^{2+}, and intrinsic cleavage was initiated by the addition of Mg^{2+}. The mixture for the final intrinsic cleavage reaction contained 20 mM Tris·HCl (pH 9.0), 100 nM pol II, 25 nM scaffold, and 50 mM MgCl$_2$. The reaction was quenched by addition of an equal volume of 0.5 mM EDTA at various time points and analyzed by denaturing PAGE.

Recombinant TFIIS was purified as described (61–63). The elongation complex was assembled as described above in a 20 mM Tris·HCl (pH 7.5) buffer without Mg^{2+}. The solution was then mixed with an equal volume of solution containing TFIIS (from 200 nM to 1 μM as indicated in each figure) and MgCl$_2$ in elongation buffer. The final reaction solution contained 100 nM pol II, 25 nM scaffold, 1 μM TFIIS, and 5 mM MgCl$_2$. The reactions were quenched at various time points by addition of an equal volume of 0.5 M EDTA (pH 8.0). The quenched products were again analyzed by 16% denaturing urea/TBE PAGE and visualized using a storage phosphor screen and Pharos FX imager (Bio-Rad).

Analysis of TLB Efficiency. Transcription bypass efficiency is controlled by the three fidelity checkpoints, that is, nucleotide addition, subsequent extension, and proofreading, where the nucleotide addition and subsequent extension determine the forward transcription elongation behavior, whereas the proofreading induces pol II backtrack (Fig. 5*A*). Thus, a simplified

equation can be generated to describe the efficiency of transcriptional elongation:

$$\text{Transcriptional lesion bypass efficiency(TLB)} = K_1 * K_2 / K_{-1},$$

where K_1, K_2, and K_{-1} designate the efficiencies of nucleotide addition, subsequent extension, and proofreading activity, respectively. Hence, comparing with the bypass after the correct AMP insertion, the relative or normalized TLB efficiency can be described as the following:

$$\text{Normalized transcriptional lesion bypass efficiency(nTLB)}$$
$$= (K_{1x} * K_{2x} / K_{-1x}) / (K_{1A} * K_{2A} / K_{-1A}),$$

where A represents the TLB route after AMP insertion opposite the lesion site and x represents TLB route(s) after the insertion of other nucleotides opposite the lesion site. Considering the proofreading activities of TLB routes were similar for all three EtdT lesions studied here (note: this may not always be the case), the normalized TLB efficiency approximates to the following simplified equation:

$$\text{Normalized transcriptional lesion bypass efficiency(nTLB)}$$
$$\approx (K_{1x} * K_{2x}) / (K_{1A} * K_{2A}).$$

Here, the specificity constants (k_{pol}/K_d) can be used to represent the K values in the aforementioned equations. Thus, the normalized TLB efficiency can be calculated as described below:

$$\text{Normalized transcriptional lesion bypass efficiency}$$
$$\approx ((k_{pol}/K_d)_{1x} * (k_{pol}/K_d)_{2x}) / ((k_{pol}/K_d)_{1A} * (k_{pol}/K_d)_{2A}).$$

Estimation of Cellular Concentration of TFIIS. We estimate cellular concentration TFIIS based on recent global analysis of protein expression and MS-based proteomic data of TFIIS abundance. It was reported that there are 6,260 (65) or 7,254 TFIIS molecules per cell (64), which corresponds to 1.04 × 10^{-20} [6,260/(6.022 × 10^{23})] or 1.20 × 10^{-20} [7,254/(6.022 × 10^{23})] mol of TFIIS per cell, given that the average volume of a *Saccharomyces cerevisiae* cell is around ~30–40 μm^3 (or 3–4 × 10^{-14} L) (66). Therefore, the cellular concentration of TFIIS is estimated to be ≈250 nM [1.04 × 10^{-20} mol/(4 × 10^{-14} L)] to ~400 nM [1.20 × 10^{-20} mol/(3 × 10^{-14} L)], which is comparable to the concentrations of TFIIS we used for in vitro TFIIS-stimulated transcript cleavage assays (i.e., 200, 400, and 1,000 nM).

Molecular Modeling and Energy Minimization Methods. To determine the partial charges of damaged nucleosides (O^2-EtdT and O^4-EtdT) during GMP incorporation and the subsequent translocation, we performed geometry optimization and electrostatic potential calculation for each molecule. Geometry optimizations were performed using Gaussian 03 with the Hartree–Fock method and the 6-31G* basis set. Electrostatic potential calculations were conducted using the same method and basis set. The partial charges were then calculated by RESP method using resp module in AmberTools 13, where partial charges were determined by a two-stage RESP procedure (67). In the first stage, partial charges of base atoms and the phosphate group were fitted to the electrostatic potential while maintaining the partial charges of sugar groups assigned by AMBER 99SB-ILDN (68). The bonded and Lennard–Jones parameters were taken from general Amber force field (69).

Four structural models were prepared: GTP insertion stages in both the O^2-dT and O^4-dT templates and the corresponding extension stages thereafter. Each model was solvated in a dodecahedron box with ~103,000 TIP3P (70) water molecules. For the insertion stages, 408 Na$^+$ and 335 Cl$^-$ were added.

For the extension stages, 404 Na$^+$ and 335 Cl$^-$ were added. Apart from the damaged sites, AMBER 99SB-ILDN (68) force field was used to describe bonded and nonbonded parameters of amino acid residues and nucleotides. Energy minimization for each model was then performed using the steepest descent method. GROMACS 4.5 (71) simulation package was used for energy minimization. Notably, for O^2-dT template with GTP insertion, the first energy minimization was performed by restraining the position of O^2-dT, followed by another minimization without any restraint.

1. Xu L, et al. (2015) RNA polymerase II transcriptional fidelity control and its functional interplay with DNA modifications. *Crit Rev Biochem Mol Biol* 50:503–519.
2. Lindsey-Boltz LA, Sancar A (2007) RNA polymerase: The most specific damage recognition protein in cellular responses to DNA damage? *Proc Natl Acad Sci USA* 104:13213–13214.
3. Ljungman M, Lane DP (2004) Transcription—guarding the genome by sensing DNA damage. *Nat Rev Cancer* 4:727–737.
4. Tornaletti S, Hanawalt PC (1999) Effect of DNA lesions on transcription elongation. *Biochimie* 81:139–146.
5. Saxowsky TT, Doetsch PW (2006) RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis? *Chem Rev* 106:474–488.
6. Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: Two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9:958–970.
7. Belotserkovskii BP, Mirkin SM, Hanawalt PC (2013) DNA sequences that interfere with transcription: Implications for genome function and stability. *Chem Rev* 113:8620–8637.
8. Brégeon D, Doetsch PW (2011) Transcriptional mutagenesis: Causes and involvement in tumour development. *Nat Rev Cancer* 11:218–227.
9. Xu L, et al. (2014) Molecular basis of transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis. *DNA Repair (Amst)* 19:71–83.
10. Shin JH, Xu L, Wang D (2016) RNA polymerase II acts as a selective sensor for DNA lesions and DNA modifications. *Transcription* 7:57–62.
11. Shin JH, Xu L, Wang D (2017) Mechanism of transcription-coupled DNA modification recognition. *Cell Biosci* 7:9.
12. Lainé JP, Egly JM (2006) When transcription and repair meet: A complex system. *Trends Genet* 22:430–436.
13. Svejstrup JQ (2007) Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci* 32:165–171.
14. Hoeijmakers JH (2009) DNA damage, aging, and cancer. *N Engl J Med* 361:1475–1485.
15. Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461:1071–1078.
16. Donahue BA, Yin S, Taylor JS, Reines D, Hanawalt PC (1994) Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proc Natl Acad Sci USA* 91:8502–8506.
17. Lagerwerf S, Vrouwe MG, Overmeer RM, Fousteri MI, Mullenders LH (2011) DNA damage response and transcription. *DNA Repair (Amst)* 10:743–750.
18. Svejstrup JQ (2003) Rescue of arrested RNA polymerase II complexes. *J Cell Sci* 116:447–451.
19. Damsma GE, Cramer P (2009) Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *J Biol Chem* 284:31658–31663.
20. Walmacq C, et al. (2015) Mechanism of RNA polymerase II bypass of oxidative cyclopurine DNA lesions. *Proc Natl Acad Sci USA* 112:E410–E419.
21. Brueckner F, Hennecke U, Carell T, Cramer P (2007) CPD damage recognition by transcribing RNA polymerase II. *Science* 315:859–862.
22. Walmacq C, et al. (2012) Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. *Mol Cell* 46:18–29.
23. Damsma GE, Alt A, Brueckner F, Carell T, Cramer P (2007) Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat Struct Mol Biol* 14:1127–1133.
24. Wang D, Zhu G, Huang X, Lippard SJ (2010) X-ray structure and mechanism of RNA polymerase II stalled at an antineoplastic monofunctional platinum-DNA adduct. *Proc Natl Acad Sci USA* 107:9584–9589.
25. Fu D, Calvo JA, Samson LD (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* 12:104–120.
26. Soll JM, Sobol RW, Mosammaparast N (2017) Regulation of DNA alkylation damage repair: Lessons and therapeutic opportunities. *Trends Biochem Sci* 42:206–218.
27. Hecht SS (1998) Biochemistry, biology, and carcinogenicity of tobacco-specific *N*-nitrosamines. *Chem Res Toxicol* 11:559–603.
28. Swenberg JA, et al. (1984) *O*^{4}-ethyldeoxythymidine, but not *O*^{6}-ethyldeoxyguanosine, accumulates in hepatocyte DNA of rats exposed continuously to diethylnitrosamine. *Proc Natl Acad Sci USA* 81:1692–1695.
29. Godschalk R, et al. (2002) Comparison of multiple DNA adduct types in tumor adjacent human lung tissue: Effect of cigarette smoking. *Carcinogenesis* 23:2081–2086.
30. Den Engelse L, De Graaf A, De Brij RJ, Menkveld GJ (1987) *O*^{2}- and *O*^{4}-ethylthymine and the ethylphosphotriester dTp(Et)dT are highly persistent DNA modifications in slowly dividing tissues of the ethylnitrosourea-treated rat. *Carcinogenesis* 8:751–757.
31. Chen L, Wang M, Villalta PW, Hecht SS (2007) Liquid chromatography-electrospray ionization tandem mass spectrometry analysis of 7-ethylguanine in human liver DNA. *Chem Res Toxicol* 20:1498–1502.
32. Chen HJ, Wang YC, Lin WP (2012) Analysis of ethylated thymidine adducts in human leukocyte DNA by stable isotope dilution nanoflow liquid chromatography-nanospray ionization tandem mass spectrometry. *Anal Chem* 84:2521–2527.
33. Chen HJ, Lee CR (2014) Detection and simultaneous quantification of three smoking-related ethylthymidine adducts in human salivary DNA by liquid chromatography tandem mass spectrometry. *Toxicol Lett* 224:101–107.
34. Chao MR, Wang CJ, Chang LW, Hu CW (2006) Quantitative determination of urinary N7-ethylguanine in smokers and non-smokers using an isotope dilution liquid chromatography/tandem mass spectrometry with on-line analyte enrichment. *Carcinogenesis* 27:146–151.
35. Anna L, Kovács K, Gyorffy E, Schoket B, Nair J (2011) Smoking-related *O*^{4}-ethyl-thymidine formation in human lung tissue and comparisons with bulky DNA adducts. *Mutagenesis* 26:523–527.
36. Cline SD, Riggins JN, Tornaletti S, Marnett LJ, Hanawalt PC (2004) Malondialdehyde adducts in DNA arrest transcription by T7 RNA polymerase and mammalian RNA polymerase II. *Proc Natl Acad Sci USA* 101:7275–7280.
37. Dimitri A, Burns JA, Broyde S, Scicchitano DA (2008) Transcription elongation past *O*^{6}-methylguanine by human RNA polymerase II and bacteriophage T7 RNA polymerase. *Nucleic Acids Res* 36:6459–6471.
38. Dimitri A, et al. (2008) Transcription of DNA containing the 5-guanidino-4-nitroimidazole lesion by human RNA polymerase II and bacteriophage T7 RNA polymerase. *DNA Repair (Amst)* 7:1276–1288.
39. Dimitri A, Goodenough AK, Guengerich FP, Broyde S, Scicchitano DA (2008) Transcription processing at 1,*N*^{2}-ethenoguanine by human RNA polymerase II and bacteriophage T7 RNA polymerase. *J Mol Biol* 375:353–366.
40. Burns JA, Dreij K, Cartularo L, Scicchitano DA (2010) *O*^{6}-methylguanine induces altered proteins at the level of transcription in human cells. *Nucleic Acids Res* 38:8178–8187.
41. You C, Wang P, Dai X, Wang Y (2014) Transcriptional bypass of regioisomeric ethylated thymidine lesions by T7 RNA polymerase and human RNA polymerase II. *Nucleic Acids Res* 42:13706–13713.
42. You C, Wang J, Dai X, Wang Y (2015) Transcriptional inhibition and mutagenesis induced by N-nitroso compound-derived carboxymethylated thymidine adducts in DNA. *Nucleic Acids Res* 43:1012–1018.
43. Tanasova M, et al. (2015) Altered minor-groove hydrogen bonds in DNA block transcription elongation by T7 RNA polymerase. *ChemBioChem* 16:1212–1218.
44. Cheng TF, Hu X, Gnatt A, Brooks PJ (2008) Differential blocking effects of the acetaldehyde-derived DNA lesion *N*^{2}-ethyl-2′-deoxyguanosine on transcription by multisubunit and single subunit RNA polymerases. *J Biol Chem* 283:27820–27828.
45. Liu S, Wang Y (2015) Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem Soc Rev* 44:7829–7854.
46. Shrivastav N, Li D, Essigmann JM (2010) Chemical biology of mutagenesis and DNA repair: Cellular responses to DNA alkylation. *Carcinogenesis* 31:59–70.
47. Yuan B, et al. (2011) The roles of DNA polymerases κ and ι in the error-free bypass of N2-carboxyalkyl-2′-deoxyguanosine lesions in mammalian cells. *J Biol Chem* 286:17503–17511.
48. You C, et al. (2012) A quantitative assay for assessing the effects of DNA lesions on transcription. *Nat Chem Biol* 8:817–822.
49. Andersen N, Wang P, Wang Y (2013) Replication across regioisomeric ethylated thymidine lesions by purified DNA polymerases. *Chem Res Toxicol* 26:1730–1738.
50. Kireeva ML, et al. (2005) Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell* 18:97–108.
51. Charlet-Berguerand N, et al. (2006) RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J* 25:5481–5491.
52. Wang D, et al. (2009) Structural basis of transcription: Backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 324:1203–1206.
53. Cheung AC, Cramer P (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471:249–253.
54. Kalnik MW, Kouchakdjian M, Li BF, Swann PF, Patel DJ (1988) Base pair mismatches and carcinogen-modified bases in DNA: An NMR study of G.T and G.O4meT pairing in dodecanucleotide duplexes. *Biochemistry* 27:108–115.
55. O'Flaherty DK, et al. (2016) Lesion orientation of *O*^{4}-alkylthymidine influences replication by human DNA polymerase η. *Chem Sci* 7:4896–4904.
56. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD (2006) Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell* 127:941–954.
57. Xu L, et al. (2014) Dissecting the chemical interactions and substrate structural signatures governing RNA polymerase II trigger loop closure by synthetic nucleic acid analogues. *Nucleic Acids Res* 42:5863–5870.
58. Wang L, et al. (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* 523:621–625.
59. Xu L, et al. (2016) RNA polymerase II senses obstruction in the DNA minor groove via a conserved sensor motif. *Proc Natl Acad Sci USA* 113:12426–12431.
60. Hubbard K, Catalano J, Puri RK, Gnatt A (2008) Knockdown of TFIIS by RNA silencing inhibits cancer cell proliferation and induces apoptosis. *BMC Cancer* 8:133.
61. Kellinger MW, Ulrich S, Chong J, Kool ET, Wang D (2012) Dissecting chemical interactions governing RNA polymerase II transcriptional fidelity. *J Am Chem Soc* 134:8231–8240.
62. Kellinger MW, et al. (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol* 19:831–833.
63. Kellinger MW, Park GY, Chong J, Lippard SJ, Wang D (2013) Effect of a monofunctional phenanthriplatin-DNA adduct on RNA polymerase II transcriptional fidelity and translesion synthesis. *J Am Chem Soc* 135:13054–13061.
64. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* 11:319–324.
65. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
66. Ferrezuelo F, et al. (2012) The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation. *Nat Commun* 3:1012.
67. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J Am Chem Soc* 115:9620–9631.
68. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–1958.
69. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174.
70. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
71. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854.