Published in final edited form as:

J Mol Biol. 2008 January 11; 375(2): 353–366.

Transcription Processing at 1,*N*²-ethenoguanine by Human RNA Polymerase II and Bacteriophage T7 RNA Polymerase†

Alexandra Dimitri¹, Angela K. Goodenough^{2,3}, F. Peter Guengerich^{2,4}, Suse Broyde¹, and David A. Scicchitano^{1,*}

¹Department of Biology, New York University, NY, NY, 10003, USA

²Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37232

³Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37232

⁴Department of Center in Molecular Toxicology, Vanderbilt University, Nashville, Tennessee 37232

Summary

The DNA lesion 1,*N*²-ethenoguanine is formed endogenously as a byproduct of lipid peroxidation or by reaction with epoxides that result from the metabolism of the industrial pollutant vinyl chloride, a known human carcinogen. DNA replication past 1,*N*²-ethenoguanine and site specific mutagenesis studies in mammalian cells have established the highly mutagenic and genotoxic properties of the damaged base. However, there is as yet no information on the processing of this lesion during transcription. Here, we report the results of transcription past a site specifically modified 1,*N*²-ethenoguanine DNA template. This lesion contains an exocyclic ring obstructing the Watson-Crick hydrogen bonding edge of guanine. Our results show that 1,*N*²-ethenoguanine acts as a partial block to the bacteriophage T7 RNA polymerase, which allows nucleotide incorporation in the growing RNA with the selectivity A > G > (C = −1 deletion) >> U. In contrast, 1,*N*²-ethenoguanine poses an absolute block to human RNA polymerase II elongation, and nucleotide incorporation opposite the lesion is not observed. Computer modeling studies show that the more open active site of T7 RNA polymerase allows lesion bypass when the 1,*N*²-ethenoguanine adopts the *syn* conformation. This orientation places the exocyclic ring in a collision free empty pocket of the polymerase, and the observed base incorporation preferences are in agreement with hydrogen bonding possibilities between the incoming nucleotides and the Hoogsteen edge of the lesion. On the other hand, in the more crowded active site of the human RNA polymerase II, the modeling studies show that both *syn* and *anti* conformations of the 1,*N*²-ethenoguanine are sterically impermissible. Polymerase stalling is currently believed to trigger the transcription coupled nucleotide excision repair machinery. Thus, our data suggest that this repair pathway likely is engaged in the clearance of the 1,*N*²-ethenoguanine from actively transcribed DNA.

Keywords

1,*N*²-ethenoguanine lesion; Transcription; DNA damage; Molecular modeling; RNA sequencing

†The work was supported by NIH grants ES010581 to DAS and NIH 2RO1 CA75449 and 5RO1 CA28038 to SB. AKG was supported by an NIEHS training grant PHS NIH T32-ES00728

*corresponding author: David A. Scicchitano: tel. (212) 998–8237, fax (212) 995–4015. E-mail das2@nyu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

The DNA lesion 1,*N*²-ethenoguanine (1,*N*²-εG) (Figure 1) is formed endogenously as a byproduct of lipid peroxidation and oxidative stress¹⁻⁵, and by the reaction of DNA with oxidized metabolites of the exogenous industrial agents vinyl chloride and urethane^{6,7}. Vinyl chloride was identified as a human carcinogen in 1974⁶. A significant association between 1,*N*²-εG and cigarette smoking has been detected in the urine of humans⁸. Etheno-modified (ε) bases have been implicated in carcinogenesis, especially angiosarcomas of the liver⁹. The 1,*N*²-εG lesion contains an exocyclic ring which obstructs the Watson-Crick base-pairing edge of guanine¹⁰. In chromosomal DNA of Chinese hamster ovary cells, 1,*N*²-εG induces deletions, rearrangements, and base pair substitutions¹¹. *In vitro* primer extension studies show highly polymerase-dependent processing of this lesion, with blockage, or mutagenic and even non-mutagenic bypass observed in the different systems^{10,12-16}. Despite the relative wealth of data on replication fidelity past etheno adducts, no transcriptional studies have been reported to date to delineate the behavior of RNA polymerases facing these lesions.

Transcriptional studies using bulky DNA adducts, such as those derived from polycyclic aromatic hydrocarbons (PAH) and 2-aminofluorene and 2-acetylaminofluorene, as well as cyclobutane pyrimidine dimers (CPD), and non-bulky lesions including thymine glycols, 8-oxoguanine, abasic sites and methylated bases, show that transcription complexes can stall or bypass partially or completely at the damaged sites depending on the polymerase and the lesion structure¹⁷⁻²⁸. While the precise fate of the stalled RNA polymerases is still being resolved, it is highly likely that they provide the signal for repair via the transcription-coupled DNA repair pathway (TCR), which clears the lesion from the DNA²⁹⁻³¹. An actively transcribed gene containing a damaged base that is not able to arrest RNA polymerases may actually reduce the ability of the cell to repair the lesion. On the other hand, lesion bypass often causes the DNA-dependent RNA polymerases to insert incorrect bases or delete information from a nascent transcript, a process called 'transcriptional mutagenesis'^{22,32}. Elucidating why some DNA adducts pose complete blocks to transcription while others only partially impede elongation or permit complete bypass is key to understanding the onset of TCR and transcriptional mutagenesis.

Many parameters are known to influence whether or not a particular lesion will impede transcription elongation including lesion size, linkage site, stereochemistry, hydrogen bonding capacity and local DNA sequence surrounding the lesion²². The consequences of the unique structural features of the 1,*N*²-εG lesion, namely its completely obstructed hydrogen bonding edge and the added rigid exocyclic ring, on RNA synthesis during transcription elongation are therefore of interest and remain to be elucidated.

To that end, we have examined transcription past 1,*N*²-εG by both human RNA polymerase II (hRNAPII) and bacteriophage T7 RNA polymerase (T7 RNAP), two structurally different RNA polymerases. Following transcription, the resulting RNA products were sequenced, and computer modeling was employed to provide structural interpretations of the experimental findings. While hRNAPII is completely blocked by the lesion due to its highly confining active site region, the more open T7 RNAP permits partial lesion bypass with incorporation selectivity A > G > (C = -1 deletion) >> U. Hydrogen bonding opportunities between the Hoogsteen edge of the lesion and the incoming NTPs determined in the modeling studies rationalized the observed incorporation preferences.

Results

Unmodified and 1,*N*²-εG modified DNA templates were synthesized and are free of nicks

The 11-base oligomer 5'-CCATCGCTACC-3' containing a site-specific 1,*N*²-εG at the underscored G was built into the transcribed strand of a DNA template, placing the lesion downstream from the CMV enhancer/promoter and the T7 promoter as described in Methods (Figure 2 and Figure S1). Following template synthesis, the integrity of the ligated DNA was verified by digestion with the endonuclease I-PpoI and analyzed on a denaturing PAGE (Figure S2). Complete ligation between the vector and the tail should result in DNA fragments of 187, 191, and 953 bases in length following I-PpoI digestion. In contrast, the presence of incompletely ligated templates would yield additional fragments of 73, 77, 110, or 118 bases in length. Figure S2 shows that the prepared templates were completely ligated as indicated by the presence of the DNA doublet of 191 and 187 bases. The absence of smaller fragments confirmed a dearth of nicked templates.

1,*N*²-εG partially blocks T7 RNA polymerase transcription elongation

DNA templates were transcribed using T7 RNAP, a simple, single-subunit enzyme that requires no ancillary transcription factors for its activity. Transcription of unmodified DNA having an identical sequence to the 1,*N*²-εG modified template produced full-length RNA of 340 bases in length (Figure 3(a), lane 6). In contrast, T7 RNAP transcription of the 1,*N*²-εG modified template resulted in a mixture of two RNA species of approximately 274 and 340 bases in length (Figure 3(a), lane 7). The shorter transcripts coincide with the position of the 1,*N*²-εG lesion on the template strand, indicating that this lesion stalls or blocks T7 RNAP progression.

1,*N*²-εG stalls T7 RNA polymerase transcription but permits eventual bypass

To determine the efficiency of enzyme bypass, time course experiments were performed with T7 RNAP past the unmodified and the 1,*N*²-εG modified templates. Transcription was performed as described in Methods. However, during the reaction, aliquots were taken every ten minutes, quenched with loading dye, and heated to 65 °C for 15 min. A 108-base DNA oligomer radioactively labeled at its 5'-end with [γ -³²P]ATP was added to each sample to act as a standard to minimize loading errors following PAGE. The ratio of band intensity of nascent RNA to the DNA standard was calculated for each time point. As expected, the unmodified template produced increasing amounts of full-length RNA over time. Smaller transcripts, caused by pause sites in the templates, were also observed. In contrast, transcription past the 1,*N*²-εG template showed that after one minute of transcription initiation, the adduct mainly blocked the progression of the polymerase (Figure 4(a) and (b)). Over time, the ratio of full-length to truncated transcripts increased (Figure 4(c)), suggesting that some blocked polymerases could eventually bypass the lesion. If this were not the case, the ratio would have stayed the same over time. Note that smaller molecular weight RNAs accumulate below the truncated transcripts at position 274. These arise when the lesion blocks transcription and the polymerases "back up" behind each other, hence producing shorter transcripts³³.

1,*N*²-εG is an absolute block to Human RNA polymerase II elongation

hRNAPII transcription past 1,*N*²-εG was performed using HeLa nuclear extracts to provide the enzyme and all its accessory factors. The CMV enhancer/promoter present in the pCI-neo-G-less plasmid utilized to construct the DNA template is a strong and very specific promoter for only RNAPII as noted by the manufacturer (Promega Inc.) and by Zhou H et al. (2005)³⁴, Kadener S et al. (2002)³⁵, and others. For single rounds of transcription, GTP was omitted from the nucleotide mix allowing the transcription complex to enter elongation only to the point at which the first GTP was required, which was at the end of the G-less cassette

incorporated into the template³⁶. Heparin was then added to sequester any free hRNAPII. The subsequent addition of GTP permitted the stalled elongating complex to continue translocation along the DNA. For multiple rounds of transcription, all ribonucleotides were supplied in the beginning of the reaction. The results showed that 1,*N*²-εG posed an absolute block to hRNAPII during both single and multiple rounds of transcription, yielding transcripts approximately 316 bases in length (Figure 3(b), lanes 7 and 8). As expected, the unmodified template produced full-length RNA only (Figure 3(b), lanes 5 and 6).

T7 RNAP produces full-length transcripts with extensive misincorporation opposite 1,*N*²-εG and truncated RNAs indicative of stalling at the lesion site

RT-PCR of the RNA population, comprised of full-length and truncated RNA, is shown in Figure S3. Sequencing of the purified PCR product obtained from the T7 RNAP transcripts indicated a heterogeneous RNA population, necessitating subcloning to analyze individual RT-PCR products. In the full-length transcripts (~ 60% of total RNA) (Figure 4(b)) that resulted from lesion bypass, 1,*N*²-εG directed the incorporation of a mixture of bases with a clear preference for purines, mainly adenine (46%) and guanine (23%). One-base deletions were also observed (Table 2). The truncated transcripts (~ 40% of total RNA) (Figure 4(b)) showed that T7 RNAP stalled in the immediate vicinity of the lesion, either before the lesion, at the lesion or right beyond it. The latter involved incorporation of adenine or cytosine followed by two non-template instructed adenines before terminating (Table 3(a)). The unmodified template produced only full-length transcripts having the expected base sequence.

hRNAPII produces only truncated RNAs with no nucleotide incorporation opposite or past 1,*N*²-εG

hRNAPII transcription past 1,*N*²-εG resulted in the formation of truncated RNA alone. The polymerase incorporated the base complementary to the last nucleotide on the 3' side of the lesion with no nucleotide incorporation opposite 1,*N*²-εG (Table 3(b)). About a quarter of the truncated RNA contained a subsequent non-template directed addition of between 15 and 30 nucleotides. These originate from small RNA molecules present in the HeLa extract³⁷ which ligated between the chimera, utilized for ligation mediated (LM)-RT-PCR, and the nascent RNA (see Methods).

Hydrogen bonding and steric effects rationalize the observed lesion processing by T7 RNAP and hRNAPII

Molecular models of 1,*N*²-εG in T7 RNAP—Crystal structures of T7 RNAP³⁸⁻⁴⁰ were utilized as the basis for the modeling studies as described in Methods. The sequences of the template/non-template and the RNA transcript in the models matched the sequence used in the transcription experiments reported here (Figure 5). We focused on transcription elongation with each round of nucleotide addition viewed as a four-phase process (Figure S4): (1) the pre-insertion phase (open complex) in which a ribonucleotide binds to a site located on the O helix paired to the template base, but not yet in a position for insertion; (2) the insertion phase (closed complex) during which formation of the substrate complex occurs by forming a base pair between the ribonucleotide and the template base with rotation of the O helix toward the DNA:RNA hybrid pair; (3) the pre-translocation phase in which the nucleotide is now incorporated into the growing RNA with dissociation of a pyrophosphate molecule; and (4) the post-translocation phase where the RNA molecule has been extended by one nucleotide with the release of the pyrophosphate. Finally, translocation of the 3'-end of the RNA places the next template nucleotide in the flipped out, pre-insertion position to prepare for another round of elongation⁴⁰. This four-phase process was utilized to provide a structural rationale for the experimentally observed findings: (1) transcripts were truncated one residue before the lesion and the majority of these products had the expected sequence at their 3'-end; (2) full-

length transcripts were obtained containing significant base misincorporation events opposite the lesion, with the preference $A > G > (C = -1 \text{ deletion}) \gg U$.

Two glycosidic bond conformations (χ) for the $1,N^2$ - ϵ G nucleoside were investigated, namely *anti* and *syn* (Figure 1). When the $1,N^2$ - ϵ G nucleoside was in the *anti* conformation, our computer models showed that the lesion blocks incoming NTPs, with severe collisions with incoming purines but more modest ones with incoming pyrimidines. The six membered ring of the purines completely overlapped the exocyclic ring of the $1,N^2$ - ϵ G (Figure 6(a)), while pyrimidines were only in close contact with the etheno N^2 - C^{11} - C^{10} edge (Figure 1) via their O^2 - N^3 edge, with atom-atom distances of approximately 0.6 Å at the closest approach (Figure 6(b)). However, when the $1,N^2$ - ϵ G nucleoside was in the *syn* conformation, the lesion could be housed, without collision, in the polymerase active site pocket on the major groove side of the damaged guanine, allowing incorporation of a ribonucleotide opposite the lesion. Furthermore, the *syn* conformation was stabilized by a bifurcated hydrogen bond⁴¹ between the O^6 of $1,N^2$ - ϵ G and the protein side chains of Arg-632 and Gln-649 (Figure 7 and Table S1). Potential hydrogen-bonding interactions between incoming NTPs and *syn* $1,N^2$ - ϵ G were present in the models to varying extents (Figure 8 and Table S2). For ATP, plausible hydrogen bonds between N^6 of adenine and O^6 of $1,N^2$ - ϵ G, and an $N1$ -protonated adenine, observed in a number of crystal structures⁴²⁻⁴⁴, to $N7$ of $1,N^2$ - ϵ G offered an explanation as to why this base could be readily incorporated opposite the lesion. For GTP, a single hydrogen bond between $N1$ of guanine and $N7$ of $1,N^2$ - ϵ G was feasible; hence, incorporation of G opposite the adduct can be rationalized. As for CTP, the possibility of hydrogen bonding between N^4 of cytosine and O^6 of $1,N^2$ - ϵ G was suggested, with a second opportunity for hydrogen bonding between $N3$ -protonated cytosine and $N7$ of $1,N^2$ - ϵ G. $N3$ -protonated cytosines have also been observed crystallographically^{43,45,46}. Finally, in the case of UTP, only one potential hydrogen-bonding possibility was suggested, between $N3$ of uridine and $N7$ of $1,N^2$ - ϵ G. While distances between the potential hydrogen bonding partners involving the cytosine or the uridine and $1,N^2$ - ϵ G were too long by approximately 2 Å, orientations were suitable for incorporation; with modest local structural reorganization, CTP or UTP could hydrogen bond opposite $1,N^2$ - ϵ G. However, UTP has less possibilities of hydrogen-bonding than cytosine, offering a justification as to why UTP incorporation opposite $1,N^2$ - ϵ G was only rarely observed.

We have given detailed results for the substrate insertion complex, but the models were essentially identical for the pre-translocation product complex and quite similar in the post translocation and the substrate pre-insertion complexes. Results for these are given in Table S2.

Molecular models of $1,N^2$ - ϵ G in yRNAP II— $1,N^2$ - ϵ G completely blocked the progression of RNAPII with no nucleotide incorporation opposite the lesion. RNAPII is a larger enzyme than T7 RNAP. It is composed of twelve subunits, Rpb1–12, making it a 550 kD complex that is reminiscent of a claw-like structure with the subunits Rpb1 and Rpb2 forming the active site. Unlike T7 RNAP, hRNAPII requires a set of general transcription factors for it to bind to its promoter and begin transcription⁴⁷⁻⁵⁰. For modeling RNAPII containing the damage, we employed the *S. cerevisiae* RNAPII (yRNAPII) crystal structures because there are none available for the human polymerase. The two polymerases share an 80% overall identity^{48, 49} making the yeast polymerase a reasonable substitute for modeling the human enzyme. Our modeling studies showed that neither the *anti* nor the *syn* conformations of $1,N^2$ - ϵ G could be accommodated by the RNAPII active site. For the *anti* conformation, the original crystal conformation ($\chi = -173^\circ$) was least obstructing since the $1,N^2$ - ϵ G could be accommodated in the minor groove of the DNA/RNA hybrid without collision with protein residues. But, as in T7 RNAP, when incoming rNTPs were inserted, the exocyclic ring of the lesion overlapped parts of the six membered ring of the purines, and was also too close to pyrimidines, with the closest contact between the exocyclic ring of the lesion and the pyrimidine of 0.9 Å. Relative

to the active site of T7 RNAP, the RNAPII active site is densely packed, limiting most opportunities for modest rearrangements that could alleviate even slight close contacts (Figure 9). Furthermore, in the optimal *syn* conformation for 1,*N*²-εG ($\chi = -61^\circ$), there was a close, 1.7 Å contact between the 1,*N*²-εG(H8) and residue Thr-831(Hγ22) of the RNAPII subunit Rpb1. Modeling efforts to relieve this contact produced a cascade of further disturbances which, while accommodating the lesion itself, caused severe derangement of the polymerase: rotation of the key bridge helix by approximately 30° and disruption of two critical hydrogen bonding networks, one between Thr-831 and residues Asp-826, Thr-827, Val-829, and Gly-835 of the same subunit, and another between the bridge helix residues Thr-827 and Gly-823 and trigger loop residue, Thr-1083^{47,51-53}. Moreover, collisions between the bridge helix and adjacent residues that were present below the nucleotide entry site (E site) were engendered (Further details are given Figure S5). Collectively, these data support the simple model that RNAPII simply cannot reorganize in a way that can accommodate the lesion and retain catalytic activity.

Discussion

We have carried out transcriptional assays past the 1,*N*²-εG adduct utilizing T7 RNAP and HeLa nuclear extracts which provide the RNAPII complex with all its necessary factors needed for transcription; these extracts have been widely used for transcriptional studies involving a variety of DNA lesions^{19,23,24}. Our results indicate that 1,*N*²-εG interferes with transcription elongation by both polymerases. T7 RNAP pauses or stalls at the adduct but also bypasses it with significant base misincorporation: A > G > (C = -1 deletion) >> U. In contrast, hRNAPII stalls completely at 1,*N*²-εG resulting in the production of truncated RNA and no full-length transcripts. This adds 1,*N*²-εG to the growing list of lesions that can impede RNA synthesis when present on the transcribed strand of a gene, including adducts derived from ultraviolet radiation, ionizing radiation, polycyclic aromatic hydrocarbons, and oxygen free radicals, among others^{22,54,55}.

1,*N*²-εG blocks T7 RNAP elongation, but the results presented here suggest that eventual bypass can occur. As reaction time increased, the ratio of full-length to truncated transcripts increased. If this ratio had remained the same or decreased over time, it would suggest that blocked transcription is not overcome in such a way as to produce full-length RNA. The fact that the ratio increases implies that some blocked polymerases can eventually bypass the adduct, producing full-length RNA. Such bypass could occur if there were a change in the conformation of the adduct inside the active site of the enzyme, a change that perhaps rearranges the lesion from an *anti* to a *syn* conformation, which the modeling studies suggest would favor lesion bypass. The RNA sequencing data also support this notion. Truncated transcripts typically fell into one of three types: those in which the polymerase stalled one base before the adduct with no nucleotide incorporation opposite the lesion; those where the polymerase stalled opposite the adduct and added an adenine opposite the lesion; and those that stalled opposite the adduct and added cytosine followed by two nontemplate-instructed adenines. The latter two have also been noted for other lesions (e.g.: *O*⁶-meG, 5GNIM) (Dimitri, A and Scicchitano DA, manuscripts in preparation) and may indicate an inherent stalling property of this polymerase.

The observed incorporation preferences for T7 RNAP elongation past 1,*N*²-εG can be qualitatively explained by the extent of hydrogen bonding between the lesion and the four NTPs incorporated in the models: A > G > (C = -1 deletion) >> U. With protonation of the adenine N1, there are two potential hydrogen bonds in the 1,*N*²-εG:A pair and one in the 1,*N*²-εG:G pair, involving the Hoogsteen edge of the purines (Figure 8(a)). For 1,*N*²-εG:C, distances for hydrogen bonding are too large by approximately 2 Å, but orientations are suitable for forming two hydrogen bonds with an N3 protonated cytosine; hence, incorporation of

cytosine would be feasible but an energetic cost would be exacted in the rearrangement to achieve the hydrogen bonding distance. In the case of 1,*N*²-εG:U only one hydrogen bond could be formed, and this one would again necessitate rearrangement to bridge the 2 Å gap permitting hydrogen bond formation (Figure 8(b)). In addition, stacking provides a further structural rationale for the preference of purine over pyrimidine incorporation opposite the lesion in the *syn* conformation, since purine NTPs are better stabilized through stacking than pyrimidines. Moreover, opposite the *syn* 1,*N*²-εG the pocket to house the incoming NTP is spacious, which would also favor the larger purines (Figure 9 and Figure S6). Finally, the observed deletion is consistent with an *anti* 1,*N*²-εG nucleoside skipped by the polymerase (Figure S7); a similar structure/function relationship was seen with this lesion in the Y family DNA polymerase Dpo4 in which deletions were prominently observed and crystal structures showed an unpartnered *anti* 1,*N*²-εG skipped by the incoming nucleotide¹⁴.

Our models showed much less severe steric contact with the *anti* 1,*N*²-εG facing pyrimidines than purines. With a rearrangement of approximately 2 Å these close contacts could be relieved, allowing pyrimidine incorporation opposite the *anti* 1,*N*²-εG. The orientation of hydrogen donor-acceptor groups suggests that one hydrogen bond could be formed between *N*⁴ of cytosine and *O*⁶ of 1,*N*²-εG but none with the 1,*N*²-εG:U pair. While apparently less favorable than the *syn* conformation, the *anti* structure remains a possible scenario for cytosine incorporation opposite the lesion (Table S3).

In contrast to T7 RNAP, hRNAPII elongation is completely blocked by 1,*N*²-εG during both single and multiple rounds of transcription. It is unlikely that the effect observed in our experiments is due to repaired templates or abasic sites caused by the removal of the damage^{12,56}. If complete repair of the template had occurred such that the 1,*N*²-εG was removed and any resulting gap in DNA was repaired via re-synthesis and ligation of DNA, full-length RNA would have been produced, but none was observed. Furthermore, if a glycosylase had removed the lesion, an apurinic site would have been produced, and they do pose blocks to transcription^{27,57}. However, incubation of the templates with HeLa nuclear extract followed by exposure to piperidine, an agent that cleaves abasic sites, showed no observable template degradation, supporting the fact that 1,*N*²-εG was not removed from the template and is the actual block to elongation in these studies (Figure S8).

While T7 RNAP permitted translesion synthesis with varying degrees of fidelity and efficiency, hRNAPII allowed no translesion synthesis or nucleotide incorporation opposite the lesion. Our modeling studies showed that the much more complex and densely packed RNAPII is unable to accommodate the lesion for nucleotide incorporation or extension without collisions, whether the lesion is *syn* or *anti*. Modest enzyme motions to relieve apparently small contacts do not appear to be feasible without producing a cascade of serious disturbances that would abolish enzyme capability for NTP insertion (Figures S5).

While the results reported here are the first to demonstrate the effects of 1,*N*²-εG on transcription, there have been several reports describing DNA polymerase behavior at this adduct^{10,13-16}. The lesion poses a complete block to human DNA polymerase δ, analogous to hRNAPII while the human DNA translesion polymerases η, ι, and κ are able to copy past 1,*N*²-εG. DNA polymerases κ and ι add thymine opposite the adduct and then terminate. In contrast, DNA polymerase η adds bases opposite 1,*N*²-εG in the following order: G > A > C, with no observed addition of T. Thus, the effect of 1,*N*²-εG on DNA synthesis by DNA-dependent DNA polymerases is very much a function of the specific enzyme being studied, as is the case for the RNA polymerases investigated here.

What are the implications for the cell when 1,*N*²-εG blocks transcription elongation? Complete transcriptional blockage could induce ubiquitylation of the stalled RNAP complex in an effort

to remove it from the transcription unit, resulting in enzymatic turnover³⁰. Or, it affords the opportunity for soliciting TCR to clear the damage. If this is the case, 1,*N*²-εG may not only be repaired by the action of a glycosylase during base excision repair¹² but may also be cleared from the transcribed strand of actively expressed genes. In cells, a collection of pathways maintains genomic integrity by removing deleterious lesions and repairing the damaged DNA. Among these pathways are nucleotide excision repair⁵⁸, base excision repair^{59,60}, direct reversal mechanisms⁶¹, and mismatch repair⁶². Each pathway has a requisite mechanism for recognizing DNA damage, which is the first step in clearing a lesion from the genome. However, it is now clear that many damaged bases in DNA are subject to removal by more than one pathway, presumably due to multiple means of damage recognition^{63,64}. These issues need to be addressed further. The fact that 1,*N*²-εG poses a strong threat to transcription provides an additional mechanistic explanation for the toxicity of agents like vinyl chloride and lipid peroxidation products that induce its formation in cellular DNA.

Conclusion

In conclusion, our studies with the 1,*N*²-εG damage show partial lesion bypass using T7 RNAP with base incorporation preference A > G > (C = -1 deletion) >> U. Even though the Watson-Crick edge of the guanine is obstructed by the exocyclic ring, the observed NTP incorporation preferences can be rationalized by hydrogen bonding opportunities between the NTP and the Hoogsteen edge of the damaged guanine in the *syn* conformation. In contrast, the lesion poses an absolute block to hRNAPII elongation with no nucleotide incorporation opposite the 1,*N*²-εG, due to unfavorable steric interactions in this more crowded polymerase complex. The rigid exocyclic ring in the *anti* conformation bars nucleotide entry, while in the *syn* conformation the etheno multi-ringed system collides with nearby polymerase residues in the active site region. This lesion illustrates how hRNAPII can discriminate a damaged base from a normal one due to its size and altered hydrogen bonding capability, further supporting the notion that a transcription complex can serve as a DNA “damage sensor”⁶⁵.

Materials

Unless stated otherwise, all reagents, enzymes, and biologicals were purchased from Promega, Sigma-Aldrich, New England Biolabs, Ambion, or BioRad. DNA oligomers were obtained from Sigma-Aldrich. Radioactive materials were purchased from Perkin Elmers. All DNA sequencing reported in this paper was done using the GeneAmp PCR System 9700 (BioRad) and 3100 Capillary Sequencer (Applied Biosystems).

Methods

Plasmid construction

pCI-neo-G-less was prepared as described earlier⁶⁶, and its sequence was verified. pCI-neo-G-less was further modified with the addition of a promoter for bacteriophage T7 RNAP, which is now closer to the CMV enhancer/promoter element. In brief, a 49 base-pair strand of DNA containing the T7 RNAP promoter was inserted into a SacI site in pCI-neo-G-less (Figure S1 (a)). Ultimately the T7 promoter was positioned 26 bases downstream from the +1 transcription start site for hRNAPII. The sequence of this recombinant plasmid, which was called pCI-neo-G-less-T7, was verified.

Preparation of transcription templates

pCI-neo-G-less-T7 was used to prepare DNA templates for transcription by both hRNAPII and T7 RNAP as described previously²³. Briefly, pCI-neo-G-less-T7 was digested with BbsI, producing a 5'-nonpalindromic, four-base overhang (Figure S1(a)). Synthetic single-strand

DNA oligomers, an 11-mer and a 96-mer, were phosphorylated at the 5'-end using T4 PNK and ATP and annealed to a complementary 90-mer containing a 3'-biotin tag. The assembled oligomers created a tail. Note that the 11-mer was either modified with a site-specific 1,*N*²-εG lesion¹⁰ or unmodified, depending on whether an adduct-modified or control tail was being assembled. DNA oligomer sequences are shown in Table 1. The tail was then ligated to the digested vector in the presence of T4 DNA ligase for 16 h at 16 °C. The resulting DNA templates were purified using streptavidin-linked paramagnetic particles that were added to the biotinylated templates and incubated at room temperature. The particles were washed, and BglII was used to remove all elements of the pCI-neo-G-less-T7 vector that were not necessary for transcription. Finally, the templates were cleaved from the beads with EcoRV, gel purified using a QIAquick Gel Extraction kit (Qiagen) and verified on a non-denaturing 1% agarose matrix (Figure S1(b)). Template integrity was confirmed by digestion with I-PpoI, an enzyme that cuts the DNA adjacent to the site of vector-tail ligation (Figure 2). The digested templates were labeled on the 5'-ends with [³²γ]ATP using T4 PNK in an exchange reaction. The products were analyzed using 7% PAGE containing 8 M urea.

Transcription with T7 RNA polymerase

Transcription using T7 RNAP was performed as described in⁶⁷. Briefly, 100 mM DTT, 10 units RNasin, 2.5 mM each of ATP, GTP, and UTP, 200 μM [α-³²P]CTP (50 Ci/mmol), 20 units T7 RNA polymerase, and 25 ng (0.033 pmoles) of DNA template were mixed in transcription optimized buffer (40 mM Tris-HCl (pH 7.9 at 25 °C), 10 mM NaCl, 6 mM MgCl₂, 2 mM spermidine) to a total volume of 20 μl. The reaction was incubated at 37 °C for 1 h. RNA was extracted using phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v) and then precipitated for 2 h in 100% ethanol containing 15 μg of glycol-blue. The RNA was isolated by centrifugation (15 min at 13,000 g), and the resulting pellet was dried for 60 min to evaporate any residual ethanol. The sample was then resuspended in RNase-free water and loaded on SP6 spin columns (BioRad) to remove unincorporated radiolabeled nucleotides. An equal volume of loading dye (80% formamide, 5mM Tris-boric acid, 1 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol) was added to the sample that was then denatured at 90 °C for 10 min and immediately loaded on a 7% PAGE with 8 M urea. Image Quant software (Molecular Dynamics) was used to visualize and quantify the radiolabeled transcripts.

Transcription with hRNA polymerase II

Transcription reactions using hRNAPII were performed as described in²⁴. Briefly, 8 units of HeLa nuclear extract, 100 ng (0.15 pmols) DNA template, 10 units of RNase inhibitor, 10 mM each of ATP, GTP, and UTP, 0.4 uM [α-³²P]CTP (25 Ci/mmol) was added to transcription buffer (20mM HEPES (pH 7.9 at 25 °C), 100mM KCl, 0.2mM EDTA, 0.5mM DTT, 20% glycerol) to a final volume of 25 μl. The reaction was incubated at 30 °C for 60 min and quenched by the addition of HeLa extract stop solution (0.3M Tris-HCl (pH 7.4 at 25 °C), 0.3M sodium acetate, 0.5% SDS, 2mM EDTA, 3μg/ml tRNA). The same procedures for the extraction, precipitation, and loading of the RNA on PAGE were employed as described for transcription with T7 RNAP.

Single rounds of transcription by RNAPII were carried out with a modification of the above protocol that made use of the G-less cassette of the template. The reaction was initiated by the addition of 10 mM each of ATP and UTP, and 0.4 uM [α-³²P]CTP (25 Ci/mmol) and incubated at 30 °C for 10 min. Heparin was added to a final concentration of 80 μg/ml and incubated at 30 °C for another 10 min. Finally, GTP was added to a final concentration of 400 μM, and the reaction was incubated for 40 min. Isolation and characterization of transcripts were as described earlier for multiple-round transcription reactions.

RNA sequencing

Truncated and full-length RNA were sequenced using reverse transcriptase and the polymerase chain reaction (RT-PCR). Transcription reactions were performed as described earlier, but in the presence of all non-radioactive nucleotides with each at a concentration of 500 μ M for T7 RNAP and 400 μ M for RNAPII. This was followed by digestion with RNase-free DNase at 37 °C for 30 min to destroy any residual DNA template. Primers for the reverse transcription of full-length RNA can be found in Table 1. For the reverse transcriptase reaction, 20 pmols of primers were annealed to the RNA template by heating to 70 °C for 5 min and then chilling the mixture at 0 °C for 5 min to prevent nonspecific binding. ImProm II reverse transcriptase, 10 mM dNTPs, and 20 units RNase inhibitor were added to the primer/template hybrid in a total volume of 15 μ l, and the mixture was incubated at 25 °C for 5 min. The reaction was then incubated at 42 °C for 1 h, and the reverse transcriptase was heat inactivated at 70 °C for 15 min. The residual RNA was digested using 5 units of RNase for 1 h at 37 °C, and the enzyme was subsequently heat inactivated at 65 °C for 15 min. The excess primers and small RNA fragments were eliminated by using Microcon Centrifugal Filter Columns. The columns also served to remove salt and concentrate the cDNA product. Finally, PCR was performed with forward and reverse primers.

To gain a detailed profile of the RNA sequences produced during transcription, the RT-PCR products were cloned and sequenced. Following RT-PCR, the different products were further purified using a 1.2% non-denaturing agarose matrix. The DNA products were subsequently inserted into the pDrive vector, which was developed specifically for cloning PCR products. The cloned PCR products were used to transform Qiagen EZ competent cells grown on agar plates containing kanamycin and IPTG. White colonies were selected, grown overnight in Luria broth supplemented with ampicillin, and after 16 h, plasmids were extracted using Qiagen Miniprep kits. The resulting DNA was sequenced.

To sequence the 3'-ends of truncated transcripts, these ends were ligated to 1 μ M RNA-DNA chimeric oligomer (Table 1) using T4 RNA ligase. The use of chimeric oligomers to sequence RNA as described here was based directly on their use to identify and sequence miRNA from plant cell extracts⁶⁸. The reverse transcriptase primer 22RTrunc (Table 1) was then annealed to the tail, and RT-PCR and cloning were subsequently performed. This approach not only selected truncated transcripts, but also provided additional full-length transcripts to which the chimeric DNA was ligated.

Molecular modeling

All modeling was carried out with the Biopolymer module of INSIGHTII (Accelrys). X-ray crystal structures of T7 RNAP, representing four phases of one cycle of nucleotide incorporation during the elongation phase of transcription, were used as the starting structures for models containing the 1,*N*²- ϵ G adduct. These four structures contained a 30-base DNA template, a 9-base DNA:RNA hybrid, and an incoming ribonucleotide. The structures with Protein Data Bank codes⁶⁹ 1S0V³⁸, 1S76⁴⁰, 1S77⁴⁰ and 1MSW³⁹ represent the substrate complex (pre-insertion phase), the substrate complex (insertion phase), the product complex (pre-translocation phase) and the post translocation complex, respectively (Figure S4). The sequences of the template/non-template and the RNA transcript were remodeled to match the sequence used in the transcription experiments reported here (Figure 5). The structure of the damaged base 1,*N*²- ϵ G was obtained from a crystal structure of this lesion in Dpo4 DNA polymerase with Protein Data Bank code⁶⁹ 2BQ3¹⁴. The unmodified guanine at position i+1 or i+2 (only in the post-translocation complex MSW) of the template strand was replaced by this 1,*N*²- ϵ G structure (Figure 5). The subsequent search for optimal structures employed the following protocol: first, the glycosidic torsion angle (χ) (Figure 1) was rotated continuously utilizing the INSIGHTII utility bumpcheck in a search for structures with minimal collisions;

at this stage, slight close contacts (less than 2 Å) were relieved by modest rotations (less than 20°) of side chains involving Arg-632 and Thr-636; second, when a collision-free range was found, the glycosidic torsion angle was rotated within this range to locate structures with optimal hydrogen bonds between 1,*N*²-εG and the incoming NTP and/or neighboring protein residues. This procedure was carried out for each of the four NTPs investigated. No NTP is present opposite the base at i+2 in the post-translocation complex. Hydrogen bonding criteria, within accepted ranges^{41,45,70-73} were used: (1) donor to acceptor atom distance < 3.9 Å; (2) hydrogen to acceptor atom distance ≤ 2.9 Å; (3) the angle θ defined by the donor atom-hydrogen-acceptor atom > 90°.

For modeling RNAPII containing the damage, we used *S. cerevisiae* RNAPII (yRNAPII) crystal structures because there are none for the human polymerase. The yeast and human RNA polymerase II share an 80% overall identity^{48,49} making the yeast polymerase a good model to study as a substitute for the human enzyme. The X-ray crystal structure of yRNAPII complexed with a 29-base DNA template, 10-base RNA transcript and a 14-base non-template DNA complementary to the template strand downstream of the RNA, Protein Data Bank code⁶⁹ 2NVZ⁴⁷, was used as the starting structure for models containing the damaged base, 1,*N*²-εG. The same procedures and criteria described for T7 RNAP were employed. Rendering of the structures and all images were generated using the PYMOL program⁷⁴.

Acknowledgments

We thank the members of the Scicchitano Lab for critical reading and evaluation of the manuscript. The work was supported by NIH ES010581 to DAS and NIH 2R01 CA75449 and 5R01 CA28038 to SB. AD was supported in part by a graduate fellowship from New York University. AKG was supported by an NIEHS training grant PHS NIH T32-ES00728. Carmelo Rizzo (Department of Biochemistry and the Center in Molecular Toxicology; Vanderbilt University) provided the 1,*N*²-εG-containing 11-mer; support for oligodeoxynucleotide synthesis was provided by PHS NIH ES05355.

The work was supported by NIH grants ES010581 to DAS and NIH 2R01 CA75449 and 5R01 CA28038 to SB. AKG was supported by an NIEHS training grant PHS NIH T32-ES00728

Abbreviations

1,*N*²-εG, 1,*N*²-ethenoguanine
 T7RNAP, T7 RNA polymerase
 hRNAPII, human RNA polymerase II
 TCR, transcription coupled DNA repair
 PAH, polycyclic aromatic hydrocarbons
 CPD, cyclobutane pyrimidine dimers
 CMV, cytomegalovirus
 RT-PCR, reverse transcriptase-polymerase chain reaction
 DTT, dithiothreitol
 BSA, bovine serum albumin
 HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
 EDTA, diaminoethanetetraacetic acid
 PAGE, polyacrylamide gel electrophoresis
 T4 PNK, T4 polynucleotide kinase
 IPTG, Isopropyl β-D-1-thiogalactopyranoside
 BP, benzo[a]pyrene
 BPDE, 7,8-dihydroxy-9,10-epoxybenzo[a]pyrene diol epoxide
 O⁶-meG, O⁶ methyl guanine
 5-GNIM, 5-guanidino-4-nitroimidazole

References

- Guengerich FP, Kim DH, Iwasaki M. Role of human cytochrome P-450 IIE1 in the oxidation of many low molecular weight cancer suspects. *Chem. Res. Toxicol* 1991;4:168–79. [PubMed: 1664256]
- Hecht SS, Young-Sciame R, Chung FL. Reaction of alpha-acetoxy-N-nitrosopiperidine with deoxyguanosine: oxygen-dependent formation of 4-oxo-2-pentenal and a 1,N2-ethenodeoxyguanosine adduct. *Chem. Res. Toxicol* 1992;5:706–12. [PubMed: 1446012]
- Morinello EJ, Ham AJ, Ranasinghe A, Nakamura J, Upton PB, Swenberg JA. Molecular dosimetry and repair of N(2),3-ethenoguanine in rats exposed to vinyl chloride. *Cancer Res* 2002;62:5189–95. [PubMed: 12234983]
- Lee SH, Arora JA, Oe T, Blair IA. 4-Hydroperoxy-2-nonenal-induced formation of 1,N2-etheno-2'-deoxyguanosine adducts. *Chem. Res. Toxicol* 2005;18:780–6. [PubMed: 15833039]
- Douki T, Odin F, Caillat S, Favier A, Cadet J. Predominance of the 1,N2-propano 2'-deoxyguanosine adduct among 4-hydroxy-2-nonenal-induced DNA lesions. *Free Radic. Biol. Med* 2004;37:62–70. [PubMed: 15183195]
- Bolt HM. Vinyl chloride-a classical industrial toxicant of new interest. *Crit. Rev. Toxicol* 2005;35:307–23. [PubMed: 15989139]
- Sodum RS, Chung FL. 1,N2-ethenodeoxyguanosine as a potential marker for DNA adduct formation by trans-4-hydroxy-2-nonenal. *Cancer Res* 1988;48:320–3. [PubMed: 3335007]
- Chen HJ, Chiu WL. Association between cigarette smoking and urinary excretion of 1,N2-ethenoguanine measured by isotope dilution liquid chromatography-electrospray ionization/tandem mass spectrometry. *Chem. Res. Toxicol* 2005;18:1593–9. [PubMed: 16533024]
- Chung FL, Chen HJ, Nath RG. Lipid peroxidation as a potential endogenous source for the formation of exocyclic DNA adducts. *Carcinogenesis* 1996;17:2105–11. [PubMed: 8895475]
- Goodenough AK, Kozekov ID, Zang H, Choi JY, Guengerich FP, Harris TM, Rizzo CJ. Site specific synthesis and polymerase bypass of oligonucleotides containing a 6-hydroxy-3,5,6,7-tetrahydro-9H-imidazo[1,2-a]purin-9-one base, an intermediate in the formation of 1,N2-etheno-2'-deoxyguanosine. *Chem. Res. Toxicol* 2005;18:1701–14. [PubMed: 16300379]
- Akasaka S, Guengerich FP. Mutagenicity of site-specifically located 1,N2-ethenoguanine in Chinese hamster ovary cell chromosomal DNA. *Chem. Res. Toxicol* 1999;12:501–7. [PubMed: 10368312]
- Saparbaev M, Langouet S, Privezentzev CV, Guengerich FP, Cai H, Elder RH, Laval J. 1,N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of Escherichia coli mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase. *J. Biol. Chem* 2002;277:26987–93. [PubMed: 12016206]
- Choi JY, Zang H, Angel KC, Kozekov ID, Goodenough AK, Rizzo CJ, Guengerich FP. Translesion synthesis across 1,N2-ethenoguanine by human DNA polymerases. *Chem. Res. Toxicol* 2006;19:879–86. [PubMed: 16780368]
- Zang H, Goodenough AK, Choi JY, Irimia A, Loukachevitch LV, Kozekov ID, Angel KC, Rizzo CJ, Egli M, Guengerich FP. DNA adduct bypass polymerization by Sulfolobus solfataricus DNA polymerase Dpo4: analysis and crystal structures of multiple base pair substitution and frameshift products with the adduct 1,N2-ethenoguanine. *J. Biol. Chem* 2005;280:29750–64. [PubMed: 15965231]
- Langouet S, Muller M, Guengerich FP. Misincorporation of dNTPs opposite 1,N2-ethenoguanine and 5,6,7,9-tetrahydro-7-hydroxy-9-oxoimidazo[1,2-a]purine in oligonucleotides by Escherichia coli polymerases I exo- and II exo-, T7 polymerase exo-, human immunodeficiency virus-1 reverse transcriptase, and rat polymerase beta. *Biochemistry* 1997;36:6069–79. [PubMed: 9166777]
- Langouet S, Mican AN, Muller M, Fink SP, Marnett LJ, Muhle SA, Guengerich FP. Misincorporation of nucleotides opposite five-membered exocyclic ring guanine derivatives by Escherichia coli polymerases in vitro and in vivo: 1,N2-ethenoguanine, 5,6,7,9-tetrahydro-9-oxoimidazo[1,2-a]purine, and 5,6,7,9-tetrahydro-7-hydroxy-9-oxoimidazo[1,2-a]purine. *Biochemistry* 1998;37:5184–93. [PubMed: 9548749]
- Tornaletti S, Maeda LS, Lloyd DR, Reines D, Hanawalt PC. Effect of thymine glycol on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. *J. Biol. Chem* 2001;276:45367–71. [PubMed: 11571287]

18. Tornaletti S, Maeda LS, Kolodner RD, Hanawalt PC. Effect of 8-oxoguanine on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. *DNA Repair (Amst)* 2004;3:483–94. [PubMed: 15084310]
19. Kuraoka I, Suzuki K, Ito S, Hayashida M, Kwei JS, Ikegami T, Handa H, Nakabeppu Y, Tanaka K. RNA polymerase II bypasses 8-oxoguanine in the presence of transcription elongation factor TFIIS. *DNA Repair (Amst)* 2007;6:841–51. [PubMed: 17374514]
20. Cline SD, Riggins JN, Tornaletti S, Marnett LJ, Hanawalt PC. Malondialdehyde adducts in DNA arrest transcription by T7 RNA polymerase and mammalian RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A* 2004;101:7275–80. [PubMed: 15123825]
21. Bregeon D, Doddridge ZA, You HJ, Weiss B, Doetsch PW. Transcriptional mutagenesis induced by uracil and 8-oxoguanine in *Escherichia coli*. *Mol. Cell* 2003;12:959–70. [PubMed: 14580346]
22. Scicchitano DA, Olesnicki EC, Dimitri A. Transcription and DNA adducts: what happens when the message gets cut off? *DNA Repair (Amst)* 2004;3:1537–48. [PubMed: 15474416]
23. Perlow RA, Kolbanovskii A, Hingerty BE, Geacintov NE, Broyde S, Scicchitano DA. DNA adducts from a tumorigenic metabolite of benzo[a]pyrene block human RNA polymerase II elongation in a sequence- and stereochemistry-dependent manner. *J. Mol. Biol* 2002;321:29–47. [PubMed: 12139931]
24. Schinecker TM, Perlow RA, Broyde S, Geacintov NE, Scicchitano DA. Human RNA polymerase II is partially blocked by DNA adducts derived from tumorigenic benzo[c]phenanthrene diol epoxides: relating biological consequences to conformational preferences. *Nucleic Acids Res* 2003;31:6004–15. [PubMed: 14530448]
25. Roth RB, Amin S, Geacintov NE, Scicchitano DA. Bacteriophage T7 RNA polymerase transcription elongation is inhibited by site-specific, stereospecific benzo[c]phenanthrene diol epoxide DNA lesions. *Biochemistry* 2001;40:5200–7. [PubMed: 11318642]
26. Brueckner F, Hennecke U, Carell T, Cramer P. CPD damage recognition by transcribing RNA polymerase II. *Science* 2007;315:859–62. [PubMed: 17290000]
27. Tornaletti S, Maeda LS, Hanawalt PC. Transcription arrest at an abasic site in the transcribed strand of template DNA. *Chem. Res. Toxicol* 2006;19:1215–20. [PubMed: 16978026]
28. Donahue BA, Fuchs RP, Reines D, Hanawalt PC. Effects of aminofluorene and acetylaminofluorene DNA adducts on transcriptional elongation by RNA polymerase II. *J. Biol. Chem* 1996;271:10588–94. [PubMed: 8631860]
29. Mellon I, Spivak G, Hanawalt PC. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 1987;51:241–9. [PubMed: 3664636]
30. Svejstrup JQ. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem. Sci* 2007;32:165–71. [PubMed: 17349792]
31. Brueckner F, Cramer P. DNA photodamage recognition by RNA polymerase II. *FEBS Lett* 2007;581:2757–60. [PubMed: 17521634]
32. Saxowsky TT, Doetsch PW. RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chem. Rev* 2006;106:474–88. [PubMed: 16464015]
33. Zhou Y, Martin CT. Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced “bumping”. *J. Biol. Chem* 2006;281:24441–8. [PubMed: 16816387]
34. Zhou H, Xia XG, Xu Z. An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi. *Nucleic Acids Res* 2005;33:e62. [PubMed: 15805121]
35. Kadener S, Fededa JP, Rosbash M, Kornblihtt AR. Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc. Natl. Acad. Sci. U S A* 2002;99:8185–90. [PubMed: 12060763]
36. Luo Y, Hagler J, Shuman S. Discrete functional stages of vaccinia virus early transcription during a single round of RNA synthesis in vitro. *J. Biol. Chem* 1991;266:13303–10. [PubMed: 1712778]
37. Kramer A. Fractionation of HeLa cell nuclear extracts reveals minor small nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci. U S A* 1987;84:8408–12. [PubMed: 2960976]
38. Temiakov D, Patlan V, Anikin M, McAllister WT, Yokoyama S, Vassilyev DG. Structural basis for substrate selection by t7 RNA polymerase. *Cell* 2004;116:381–91. [PubMed: 15016373]

39. Yin YW, Steitz TA. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* 2002;298:1387–95. [PubMed: 12242451]
40. Yin YW, Steitz TA. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* 2004;116:393–404. [PubMed: 15016374]
41. Torshin IY, Weber IT, Harrison RW. Geometric criteria of hydrogen bonds in proteins and identification of “bifurcated” hydrogen bonds. *Protein Eng* 2002;15:359–63. [PubMed: 12034855]
42. Brown T, Leonard GA, Booth ED, Chambers J. Crystal structure and stability of a DNA duplex containing A(anti).G(syn) base-pairs. *J. Mol. Biol* 1989;207:455–7. [PubMed: 2754734]
43. Williams LD, Shaw BR. Protonated base pairs explain the ambiguous pairing properties of O6-methylguanine. *Proc. Natl. Acad. Sci. U S A* 1987;84:1779–83. [PubMed: 3470757]
44. Pan B, Mitra SN, Sundaralingam M. Crystal structure of an RNA 16-mer duplex R (GCAGAGUUAUAUCUGC)₂ with nonadjacent G(syn).A+(anti) mispairs. *Biochemistry* 1999;38:2826–31. [PubMed: 10052954]
45. Spratt TE, Levy DE. Structure of the hydrogen bonding complex of O6-methylguanine with cytosine and thymine during DNA replication. *Nucleic Acids Res* 1997;25:3354–61. [PubMed: 9241252]
46. Ginell SL, Kuzmich S, Jones RA, Berman HM. Crystal and molecular structure of a DNA duplex containing the carcinogenic lesion O6-methylguanine. *Biochemistry* 1990;29:10461–5. [PubMed: 2271657]
47. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 2006;127:941–54. [PubMed: 17129781]
48. Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 2001;292:1863–76. [PubMed: 11313498]
49. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 2001;292:1876–82. [PubMed: 11313499]
50. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol* 2004;11:394–403. [PubMed: 15114340]
51. Toulkhonov I, Zhang J, Palangat M, Landick R. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell* 2007;27:406–19. [PubMed: 17679091]
52. Vassilyev DG, Vassilyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 2007;448:163–8. [PubMed: 17581591]
53. Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 2007;448:157–62. [PubMed: 17581590]
54. Scicchitano DA. Transcription past DNA adducts derived from polycyclic aromatic hydrocarbons. *Mutat. Res* 2005;577:146–54. [PubMed: 15922365]
55. Tornaletti S. Transcription arrest at DNA damage sites. *Mutat. Res* 2005;577:131–45. [PubMed: 15904937]
56. Gros L, Ishchenko AA, Saparbaev M. Enzymology of repair of etheno-adducts. *Mutat. Res* 2003;531:219–29. [PubMed: 14637257]
57. Kathe SD, Shen GP, Wallace SS. Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. *J. Biol. Chem* 2004;279:18511–20. [PubMed: 14978042]
58. Costa RM, Chigancas V, Galhardo Rda S, Carvalho H, Menck CF. The eukaryotic nucleotide excision repair pathway. *Biochimie* 2003;85:1083–99. [PubMed: 14726015]
59. Almeida KH, Sobol RW. A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair (Amst)* 2007;6:695–711. [PubMed: 17337257]
60. Wilson DM 3rd, Bohr VA. The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair (Amst)* 2007;6:544–59. [PubMed: 17112792]

61. Pegg AE, Fang Q, Loktionova NA. Human variants of O(6)-alkylguanine-DNA alkyltransferase. *DNA Repair (Amst)*. 2007
62. Houtgraaf JH, Versmissen J, van der Giessen WJ. A concise review of DNA damage checkpoints and repair in mammalian cells. *Cardiovasc. Res* 2006;7:165–72. [PubMed: 16945824]
63. Kovtun IV, McMurray CT. Crosstalk of DNA glycosylases with pathways other than base excision repair. *DNA Repair (Amst)* 2007;6:517–29. [PubMed: 17129768]
64. Yang W. Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst)* 2006;5:654–66. [PubMed: 16574501]
65. Lindsey-Boltz LA, Sancar A. RNA polymerase: The most specific damage recognition protein in cellular responses to DNA damage? *Proc. Natl. Acad. Sci. U S A* 2007;104:13213–4. [PubMed: 17684092]
66. Perlow RA, Schinecker TM, Kim SJ, Geacintov NE, Scicchitano DA. Construction and purification of site-specifically modified DNA templates for transcription assays. *Nucleic Acids Res* 2003;31:e40. [PubMed: 12655028]
67. Choi DJ, Roth RB, Liu T, Geacintov NE, Scicchitano DA. Incorrect base insertion and prematurely terminated transcripts during T7 RNA polymerase transcription elongation past benzo[a]pyrenediol epoxide-modified DNA. *J. Mol. Biol* 1996;264:213–9. [PubMed: 8951371]
68. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 2001;15:188–200. [PubMed: 11157775]
69. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42. [PubMed: 10592235]
70. Fabiola F, Bertram R, Korostelev A, Chapman MS. An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Sci* 2002;11:1415–23. [PubMed: 12021440]
71. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol* 1994;238:777–93. [PubMed: 8182748]
72. Taberner L, Bella J, Aleman C. Hydrogen bond geometry in DNA-minor groove binding drug complexes. *Nucleic Acids Res* 1996;24:3458–66. [PubMed: 8811103]
73. Panigrahi SK, Desiraju GR. Strong and weak hydrogen bonds in the protein-ligand interface. *Proteins*. 2007
74. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA: 2002.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Figure 1.

Structure of (a) guanine and (b) 1,N²-ethenoguanine. The glycosidic torsion angle (γ) is defined by the atoms O4'-C1'-N9-C4. To facilitate comparison between guanine and 1,N²-ethenoguanine, the numbering scheme of 1,N²-ethenoguanine differs from the standard one¹⁰.

Figure 2.

Site specifically modified DNA templates prepared for transcriptional assays. The vector displayed in white is comprised of the cloned T7 promoter and the CMV promoter/enhancer (black boxes). The tail region is formed by annealing a 96-mer, a 90-mer, and an 11-mer (grey boxes) containing the lesion whose position is shown with a black triangle. The arrow heads indicate the +1 transcription start site. I-PpoI cuts at position 948 of the template and is used to check for proper ligation of the vector to the tail. The total size of the template is 1,140 base pairs. The expected sizes of the RNA products obtained during transcription by T7 RNAP and hRNAPII are shown below the template.

Figure 3.

T7 RNAP and hRNAPII transcription of 1, N^2 - ϵ G modified-DNA templates. (a) T7 RNAP transcription past the unmodified DNA template (T_G) resulted in full-length RNA approximately 340 bases in length (lane 6). Transcription past the modified template ($T_{\epsilon G}$) resulted in two RNA species, 274 and 340 bases in length respectively (lane 7). Lane 2 is the RNA produced from a positive control template supplied with the T7 RNAP Kit. Lanes 3–5 are negative experimental controls, in which DNA or T7 RNAP or NTPs were omitted from the reactions. A 50 base pair DNA step ladder is shown in lane 1. (b) Lane 2 is the RNA produced from a positive control template supplied with the HeLa nuclear extract ~ 363 bases. Lanes 3 and 4 are negative experimental controls, in which DNA or HeLa extracts were absent. Single and multiple rounds of transcription using HeLa nuclear extracts with the undamaged template (T_G) generated full-length transcripts of approximately 383 bases in length (lanes 5

and 6) while single and multiple rounds of transcription past the damaged template ($T_{\varepsilon}G$) produced only truncated RNA approximately 316 bases in length (lanes 7 and 8). A 50 base pair DNA step ladder is shown in lane 1.

Figure 4.

Time-course assay of T7 RNAP transcription past the 1,*N*²-εG damaged templates. (a) Transcription past the undamaged template, T_G (lanes 1–9) and the damaged template T_{εG} (lanes 10–18) was performed over 90 minutes. Aliquots were removed every 10 minutes to measure the extent of lesion blockage or bypass over time. Fragments were compared to a 50 bp DNA step ladder (M). The increased intensity across lanes shows the increase in RNA over time as compared to a loading control of 108 bases in length. Transcription past T_{εG} shows that early in transcription (<10 min) the lesion mainly blocks transcription (RNA = 274 bases) and as time progresses, the polymerase bypasses the damage more effectively, forming more full-length RNA (340 bases) (lanes 10–18). (b) The percent of RNA at each time point compared to the loading control (108 mer) is shown. The total RNA is represented with hexagons, the truncated RNA with squares and full-length RNA with triangles. (c) The ratio of full-length to truncated RNA produced over 90 minutes is shown. The ratio increases for the first 60 minutes where it reaches a plateau of 0.6 implying that some truncated RNAs are eventually extended to the full-length product. The number of radioactive CTPs incorporated in each transcript was taken into account in the calculation.

Figure 5.

Sequences employed in the experimental and modeling studies in T7 RNAP and RNAPII complexes. The template nucleotide, G or 1, N^2 - ϵ G, to be transcribed occupies the insertion position *i+1* (cyan). The next nucleotide to be transcribed is in the pre-insertion position *i+2* (grey). The ribonucleotide previously incorporated at the 3'-end of the transcript, R_1 (green). The rNTP to be inserted opposite the lesion is called Incomer (red). The Non-Template DNA strand is colored in purple.

Figure 6.

View of the *anti* 1,*N*²-εG pairing with incoming (a) purines or (b) pyrimidines. (a) Severe collisions with purines are observed as the two nucleosides overlap, while with pyrimidines (b) more modest close contacts are found. Color code is the same as in Figure 5 and Figure S4. In addition, the exocyclic ring of 1,*N*²-εG is displayed in yellow. All nitrogen, oxygen and phosphorous atoms are shown in marine blue, red and violet respectively. *anti*A and *anti*C are shown as representative of purines and pyrimidines.

Figure 7.

Bifurcated H-bond between *syn* 1,*N*²-εG and T7 RNAP residues, Arg-632 and Gln-649, involving O⁶ of 1,*N*²-εG and the Nε2 and Nε of Gln-649 and Arg-632, respectively. The ATP partner is shown for illustration but the same interactions are present for all NTPs opposite *syn* 1,*N*²-ethenoguanine. Hydrogen bonds are shown as dotted black lines and the hydrogen atoms of the backbone were removed for clarity. Color code is the same as in Figure 5.

Figure 8.

Potential hydrogen bonding interactions between *syn* 1,*N*²-εG and (a) purine and (b) pyrimidine rNTPs (see text). (a) *Syn* 1,*N*²-εG:A⁺ pair; εG(N7):A(N1⁺) and εG(O⁶):A(N⁶). *Syn* 1,*N*²-εG:G pair; εG(N7):G(N1). (b) *Syn* 1,*N*²-εG:rC⁺ pair; εG(N7):C(N3⁺) and εG(O⁶):C(N⁴). *Syn* 1,*N*²-εG:U pair; εG(N7):U(N3). Color code is the same as in Figure 5. Details of the potential hydrogen bonding schemes are shown in Table S2.

Figure 9.

Views of (a) T7 RNAP and (b) yRNAPII with a templating guanine opposite a CTP ribonucleotide. Note that the RNAPII active site is much more crowded than the single subunit T7 RNAP. The active site of each enzyme is also shown as an inset. The black boxes outline the area shown in the insets. Color code is the same as in Figure 5.

Table 2

Bases incorporated opposite the 1,*N*²-εG lesion in the full-length RNA generated during transcription by T7RNAP.

rNTP incorporated	A	G	C	Deletion	U
% ^a	46	23	15	15	1

^a: Percentages (%) were calculated after sequencing 100 clones.

Table 3

Bases incorporated opposite the 1,*N*²-εG lesion in the truncated RNA generated during transcription with (a) T7 RNAP and (b) hRNAPII.

(a)			
	Sequence	RNA length	%
Expected RNA Sequence	...UCUUGCGGUAGCGAUGG...		
Truncated 1 bases before εG	...UCUUGCGGUA	271	3
Truncated at εG	...UCUUGCGGUAG	272	66
Misincorporation then truncation	...UCUUGCGGUAGA	273	14
	...UCUUGCGGUAGAG	274	7
	...UCUUGCGGUAGCAA	275	10
(b)			
	Sequence	RNA length	%
Expected RNA Sequence	...UCUUGCGGUAGCGAUGG...		
Truncated 2 bases before εG	...UCUUGCGGU	270	20
Truncated at εG	...UCUUGCGGUAG	272	75
	...UCUUGCGGUAC		5