JBC RESEARCH ARTICLE

Thymine DNA glycosylase is an RNA-binding protein with high selectivity for G-rich sequences

Received for publication, December 28, 2022, and in revised form, February 17, 2023 Published, Papers in Press, March 6, 2023,
https://doi.org/10.1016/j.jbc.2023.104590

Lauren A. McGregor[1], **Baiyu Zhu**[1], **Allison M. Goetz**[1], **and Jonathan T. Sczepanski**[1,*]

From the [1]*Department of Chemistry, Texas A&M University, College Station, Texas, USA*

Reviewed by members of the JBC Editorial Board. Edited by Patrick Sung

Thymine DNA glycosylase (TDG) is a multifaceted enzyme involved in several critical biological pathways, including transcriptional activation, DNA demethylation, and DNA repair. Recent studies have established regulatory relationships between TDG and RNA, but the molecular interactions underlying these relationships are poorly understood. Herein, we now demonstrate that TDG binds directly to RNA with nanomolar affinity. Using synthetic oligonucleotides of defined length and sequence, we show that TDG has a strong preference for binding G-rich sequences in single-stranded RNA but binds weakly to single-stranded DNA and duplex RNA. TDG also binds tightly to endogenous RNA sequences. Studies with truncated proteins indicate that TDG binds RNA primarily through its structured catalytic domain and that its disordered C-terminal domain plays a key role in regulating TDG's affinity and selectivity for RNA. Finally, we show that RNA competes with DNA for binding to TDG, resulting in the inhibition of TDG-mediated excision in the presence of RNA. Together, this work provides support for and insights into a mechanism wherein TDG-mediated processes (*e.g.*, DNA demethylation) are regulated through the direct interactions of TDG with RNA.

DNA glycosylases initiate the base excision repair pathway by recognizing specific nucleobase lesions in DNA and catalyzing cleavage of the corresponding glycosidic bond (1–3). Thymine DNA glycosylase (TDG), a member of the uracil DNA glycosylase superfamily, has been shown to recognize mismatched pyrimidine bases from G•T and G•U pairs and initiate BER at these sites (4, 5). Like all enzymes in this family, TDG uses a base-flipping mechanism to place the damaged base into its active site to catalyze the cleavage of the glycosidic bond. Regardless of the identity of the substrate, the resulting abasic site (AP) is processed by the canonical BER pathway.

TDG is the only glycosylase known to be capable of removing 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) from DNA in mammals (6, 7) and thus plays a central role in the active DNA demethylation pathway responsible for reversing 5-methylcytosine (5mC) (8, 9). Deletion of TDG from embryonic stem cells results in a five- to tenfold increase in the levels of 5fC and 5caC(10), which preferentially accumulate at genomic features associated with important gene-regulatory functions, such as promoters and enhancers (10–15). During mammalian development, TDG's catalytic activity is required for maintaining epigenetic stability of many developmental and tissue-specific genes, and deletion of TDG causes embryonic lethality due, in part, to hypermethylation of promoters within these genes (16, 17). Aside from its catalytic role, TDG has also been shown to potentiate transcription through physical interactions with various transcription factors and activating histone modifiers (18, 19). Notably, TDG is essential for recruiting the histone acetyltransferases (HATs) CBP/p300 to a subset of hormone-responsive genes to facilitate histone acetylation and subsequent transcriptional activation (16, 17, 20). Together, these findings support the notion that TDG plays an important role in regulating and maintaining appropriate genetic and epigenetic states. However, the mechanisms surrounding these important regulatory functions of TDG remain poorly understood.

RNAs, such as long noncoding RNAs (lncRNA) and enhancer RNAs (eRNAs), are known to play key roles in transcriptional regulation and DNA methylation control. Therefore, it is not surprising that emerging evidence indicates potential regulatory relationships between TDG and RNA. TDG was shown to be essential in mediating the transcription of eRNAs at the enhancers of a subset of E2-responsive genes (21), which have also been shown to assemble into RNA-dependent ribonucleoprotein complexes (22). Furthermore, the lncRNA *TARID* was shown to target a TDG/GADD45A (growth arrest and DNA-damage-inducible, alpha) complex to the *TCF21* gene promoter for subsequent DNA demethylation, demonstrating the ability of lncRNAs to guide TDG-mediated activities (23). A very recent study provided evidence that the lncRNA TETILA physically interacts with TDG to target a DNA demethylation complex to specific genes (24). Although TDG lacks a canonical RNA-binding domain, it does have several positively-charged surfaces that could facilitate electrostatic interactions with RNA (Fig. S1). Moreover, TDG's disordered N- and C-terminal domains are enriched in cationic and/or polar residues, sequence features that are known to promote RNA binding and the self-assembly of ribonucleoprotein complexes (Fig. S1) (25–28). Despite these observations and the link between RNA and TDG-mediated processes, direct evidence that TDG binds to RNA is lacking.

* For correspondence: Jonathan T. Sczepanski, jon.sczepanski@chem.tamu.edu.

Herein, we carry out quantitative *in vitro* binding studies with synthetic RNAs of defined length and sequence and demonstrate, for the first time, that TDG is a *bona fide* RNA binding protein. We show that TDG binds preferably and tightly to G-rich sequences in single-stranded RNAs but has a weak affinity for ssDNA and duplex RNA. Studies with truncated proteins indicate that TDG binds RNA primarily through its structured catalytic domain and that its disordered C-terminal domain plays a key role in regulating TDG's affinity and selectivity for RNA. Finally, we demonstrate that RNA competes with DNA for binding to TDG, providing a novel mechanism through which RNA can regulate TDG-mediated excision.

Results

TDG binds RNA with a preference for G-rich sequences in vitro

To begin our investigation, we synthesized a series homopolymer 30-mer RNA oligonucleotides of different base compositions and measured TDG binding by an agarose gel-based electrophoretic mobility shift assay (EMSA) (Fig. 1). Because the affinity of RBPs for RNA can be dependent on RNA length, especially for promiscuous RBPs, RNA of the same length was compared. Furthermore, all synthetic RNAs used in this study were cyanine 3 (Cy3)-labeled at their 5′ ends in an identical manner (Table S1). Interestingly, within this series, only poly[G] (G_{30}) had a measurable affinity for TDG, with an apparent K_d of 434 nM (95% confidence interval (95% CI), 396–472 nM) (Fig. 1 and Table 1). While we were able to detect a small TDG-RNA shift in the EMSAs for A_{30}, C_{30}, and U_{30}, the K_d for these interactions greatly exceeded 5 μM, which was the highest protein concentration tested (Fig. 1*A*). While these data suggested that guanines promoted TDG binding, it was also possible that other bases (*e.g.*, adenines) prevented binding. To test this, we asked whether the insertion of Gs into A_{30} could rescue binding by TDG. We prepared a series of RNAs containing different arrangements of Gs and As, $(GA)_{20}$, $(GGAA)_{10}$, and $(G_3A_4)_4$ and found that they all bound TDG at least as well as G_{30} (Fig. 1*C* and Table 1). Their binding affinities were measured as $GA_{20} > (G_3A_4)_4 > (GGAA)_{10}$. The

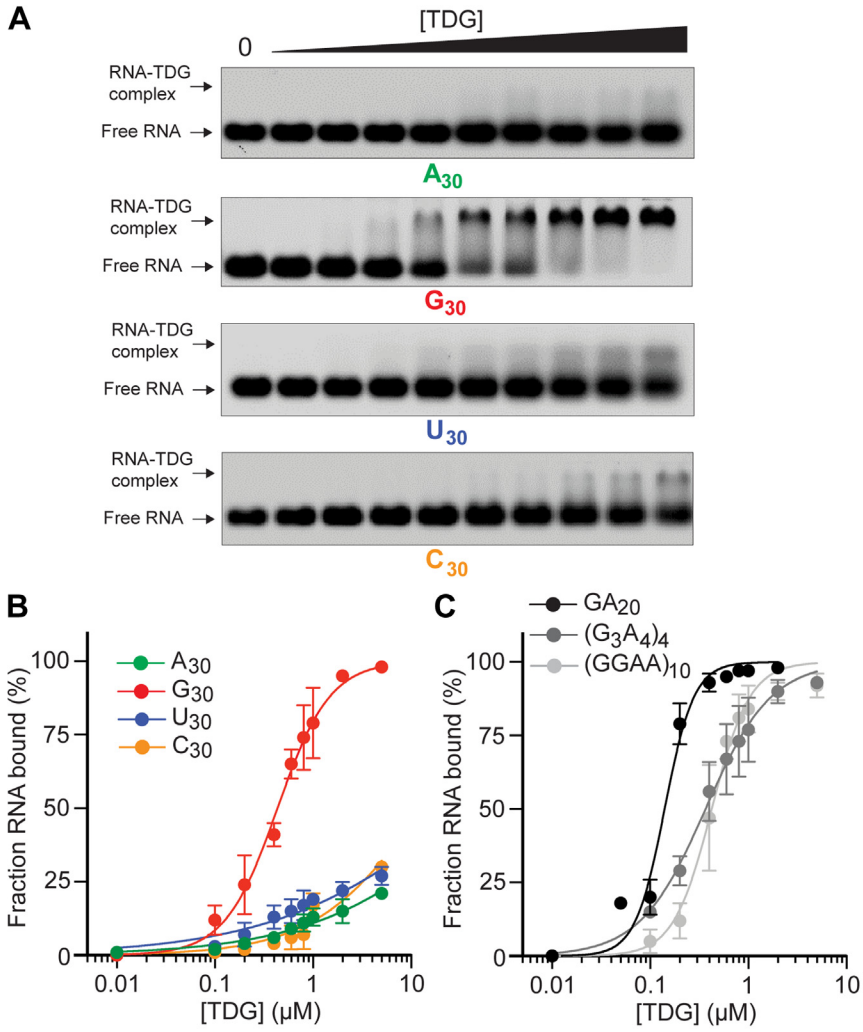


Figure 1. TDG binds preferentially to G-rich RNA *in vitro*. *A*, Representative EMSA data for homopolymeric RNA sequences binding to TDG (0–5 μM). Uncropped gel images are presented in Fig. S11. *B*, Saturation plots for binding of TDG to homopolymeric RNA. Data are mean ± SD (n = 3). *C*, saturation plots for binding of TDG to RNAs containing different arrangements of Gs and As. Data are mean ± SD (n = 3). Representative EMSA data for panel (*C*) are shown in Fig. S2*A*. EMSA, electrophoretic mobility shift assay; TDG, Thymine DNA glycosylase

Table 1
Equilibrium dissociation constants and Hill coefficients (*h*) for TDG binding to various RNA and DNA strands.

Sequence	K_d (nM)	95% CI	Hill (*h*)	95% CI
G_{30}	434	396–473	1.6	1.4–1.9
C_{30}	>5000	—	—	—
A_{30}	>5000	—	—	—
U_{30}	>5000	—	—	—
$(GA)_{20}$	140	118–165	2.9	2.2–4.0
$(GA)_{15}$	410	375–447	1.7	1.5–2.0
$(GA)_{10}$	2420	1890–3278	1.1	0.9–1.4
$(GA)_5$	>8097	6704–10,506	1.2	1.0–1.5
$d(GA)_{20}$	1448	1289–1656	0.9	0.9–1.0
$(GU)_{20}$	319	301–338	2.2	1.9–2.5
$(G_3A_4)_4$	366	338–394	1.3	1.2–1.4
$(GGAA)_{10}$	423	376–472	2.2	1.7–2.9
$d(GGAA)_{10}$	2287	2106–2494	2.3	2.0–2.7
$L\text{-}(GGAA)_{10}$	532	505–559	3.4	2.9–4.0
GC_{HP}	2896	3242–4842	2.2	1.7–2.9
MUT_{HP}	139	127–153	2.9	2.0–3.8
HOTAIR	39	36–41	2.7	2.3–3.2
TFF1e	104	100–108	3.7	3.2–4.4

95% CI, 95% confidence interval; TDG, Thymine DNA glycosylase; TFF1e, *TFF1* enhancer RNA transcript; HOTAIR, HOX antisense intergenic RNA.

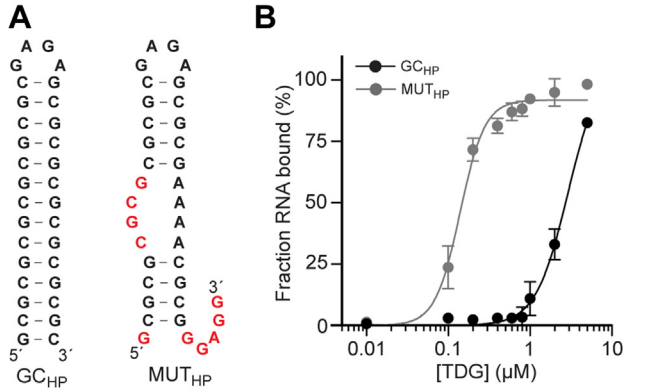


Figure 2. TDG avoids binding to double-stranded RNA. *A*, sequence and secondary structures of CG_{HP} and MUT_{HP}. Differences in MUT_{HP} relative to CG_{HP} are emphasized in *red* text. *B*, saturation plots for binding of TDG to CG_{HP} and MUT_{HP}. Data are mean ± SD (n = 3). Representative EMSA data are shown in Fig. S4. EMSA, electrophoretic mobility shift assay; TDG, Thymine DNA glycosylase.

narrow range of dissociation constants among these sequences (~100–400 nM) relative to TDG's affinity for A_{30} (>5 μM) indicated that the specific arrangement of Gs is not a critical factor for binding. Another sequence containing only G and U, $(GU)_{15}$, also bound tightly to TDG (Fig. S2 and Table 1). Thus, we concluded that TDG binds preferentially to G-rich RNAs.

TDG has a low affinity for double-stranded RNA

In addition to its sequence, the secondary structure of RNA is an important factor to consider when studying protein–RNA interactions. RBPs often exhibit structure-specific RNA binding that is important for function (29). In particular, a number of RBPs have been shown to bind preferentially to G-quadruplex (G4)-forming RNA sequences (30–32). G-rich RNAs having at least four closely spaced G-tracks can form intramolecular G4 structures comprised of planar stacks of Hoogsteen-bonded G-quartets. Previous studies (30), including our own work (33), have shown that $(GGAA)_{10}$ and $(G_3A_4)_4$, but not $(GA)_{20}$, form G4 structures under buffer conditions similar to those used in our binding assays. Indeed, circular dichroism (CD) spectroscopy confirmed that $(GGAA)_{10}$ and $(G_3A_4)_4$ form strong G4 structures under our EMSA buffer conditions, whereas $(GA)_{20}$ does not (Fig. S3*A*). The fact that TDG bound tightly to these sequences with a narrow range of affinities (Fig. 1*C* and Table 1) suggested that the presence of a G4 had little influence on TDG's ability to bind G-rich RNAs.

In addition to G4s, we also examined whether TDG could bind perfect double-stranded RNA (dsRNA) (34). We prepared a G-rich RNA hairpin, GC_{HP}, designed to form a 14 bp stem connected by a tetraloop (Fig. 2*A*). GC_{HP} had exclusively G/C base pairs within its stem and an overall base composition of 50% G. Hairpin formation was confirmed by CD spectroscopy (Fig. S3*B*). The affinity of GC_{HP} for TDG (K_d ~ 3 μM) was drastically reduced relative to the other G-rich sequences tested, which have similar G content (Fig. 2*B* and Table 1). This result indicated that TDG has low affinity for dsRNA, even if it is G-rich. Consistently, insertion of a few helix defects into GC_{HP}, which resulted in an internal loop and 3′ tail giving ~40% single-stranded nucleotides (MUT_{HP}), was sufficient to greatly increase its affinity for TDG (K_d = 139 nM, 95% CI 127–153 nM) (Fig. 2 and Table 1).

RNA length dependence on TDG binding

We next examined the relationship between RNA length and TDG binding. We prepared a series of incrementally shorter versions of $(GA)_n$, where n = 20, 15, 10, or 5. (Table 1). As shown in Figure 3, as the length of the RNA decreased so too did its affinity for TDG. The K_d increased >10-fold when the length of the RNA was halved from 40-nt to 20-nt and the 10-mer $(GA)_5$ had a K_d > 5 μM (Table 1). Thus, in addition to its structure (above), the length of the RNA is an important factor for binding. This direct relationship between dissociation constant and RNA length suggests that in the context of a larger RNA, such as a

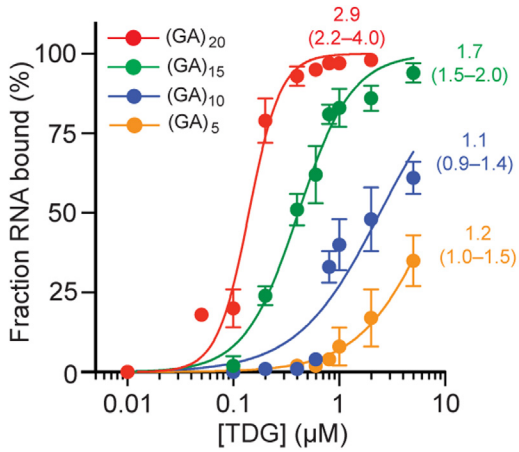


Figure 3. TDG binds RNA in a length-dependent manner. Saturation plots for binding of TDG to $(GA)_{20}$ and its truncations. Binding curves were fit to three independent replicates. The Hill coefficient (*h*) is given, with 95% CIs in parentheses. Representative EMSA data are shown in Fig. S5. EMSA, electrophoretic mobility shift assay; TDG, Thymine DNA glycosylase.

lncRNA or eRNA, the length of a G-rich motif (and potentially the density of G within it) will play an important role in determining where and how tightly TDG binds. Interestingly, the Hill coefficient, *h*, used to describe the cooperativity of binding in systems with multiple simultaneous binding events, also decreased as the RNA was shortened. This overall behavior is consistent with a protein binding non-sequence-specifically and cooperatively to multiple sites on the RNA (35, 36). Indeed, similar behavior has been observed for other promiscuous RNA-binding proteins, including those that bind preferentially to G-rich sequences (37, 38). Crystal structures show that TDG binds DNA with a footprint of ~10 bps, with the majority of interactions occurring at the periphery (39, 40). One intriguing possibility is that TDG binds RNA using the same surface as for DNA, resulting in an RNA footprint of similar length. This could explain the very weak affinity of TDG for short RNAs (≤10 nt), as well as its higher affinity and cooperativity for longer RNAs (≥30 nt), which can more easily accommodate multiple TDG molecules. Future studies are needed to test this hypothesis.

TDG binds endogenous RNA sequences

Up to this point, our experiments had primarily employed short, homopolymeric sequences. Therefore, we sought to demonstrate TDG's ability to bind long, native RNA sequences that have mixed base composition and are capable of folding into intricate secondary structures *in vivo*. For these experiments, we chose the lncRNA HOX antisense intergenic RNA (HOTAIR) and the *TFF1* enhancer RNA transcript (TFF1e) (Table S3). HOTAIR has been shown to function as a molecular scaffold for several RNA-binding proteins (41), including the G-rich RNA-binding polycomb repressive complex 2 (PRC2) (30, 37), making it a good model for studying such interactions. TFF1e RNA is transcribed from the *TFF1* gene enhancer that is also bound by TDG, suggesting the possibility of an interaction *in vivo*. We transcribed a 419-nt fragment of HOTAIR and a 316-nt fragment of TFF1e that each contained ~60% G content (Table S3). As shown in Figure 4 and Table 1, TDG bound tightly to both RNAs, but its K_d for HOTAIR (39 nM, 95% CI 36–41 nM) was ~threefold lower than for TFF1e (104 nM, 95% CI 100–108 nM). Based on our observations above, this likely can be attributed to differences in sequence and folded secondary structure between these RNAs. These results confirm that TDG has the capacity to bind endogenous RNA sequences and that our prior results are not artifacts resulting from the use of short, homopolymeric sequences.

TDG discriminates between single-stranded RNA and single-stranded DNA

Having shown TDG's preference for single-stranded RNA (ssRNA), we considered whether TDG could bind single-stranded DNA (ssDNA). Although ssDNA occurs rarely in cells, we expected that any potential differences in binding between these polymers could provide further insights into the nature of TDG-RNA interactions. Therefore, we compared the RNAs $(GGAA)_{10}$ and $(GA)_{20}$ to their DNA counterparts, $d(GGAA)_{10}$ and $d(GA)_{20}$, respectively, for binding to TDG. We found that TDG had much less affinity for the ssDNAs than their corresponding RNAs (Fig. 5 and Table 1), indicating that TDG has a preference for ssRNA over ssDNA.

Although poorly understood, the ability of RBPs to discriminate ssRNA from ssDNA of the same sequence are most often attributed to specific interactions with the ribose sugar (*e.g.*; hydrogen bonding with the 2′-hydroxyl group) and/or differences in conformational flexibility between the two polymers, with ssRNA being less flexible than ssDNA (42–45). In order to probe the involvement of specific interactions with the ribose sugar, we examined whether TDG could bind the enantiomer of $(GGAA)_{10}$, L-$(GGAA)_{10}$, composed entirely of L-ribose sugar units. If contacts with the sugar play an important role in binding, then we expected that its complete stereochemical inversion would alter the formation of these interactions and potentially TDG's ability to bind RNA. Instead, we found that TDG bound L-$(GGAA)_{10}$ about as well as native $(GGAA)_{10}$ (Fig. 5*A*). Assuming TDG binds both enantiomers through a similar set of residues, these data suggest that TDG's ability to discriminate between ssDNA and ssRNA is not due to structure-specific interactions with the ribose sugar.

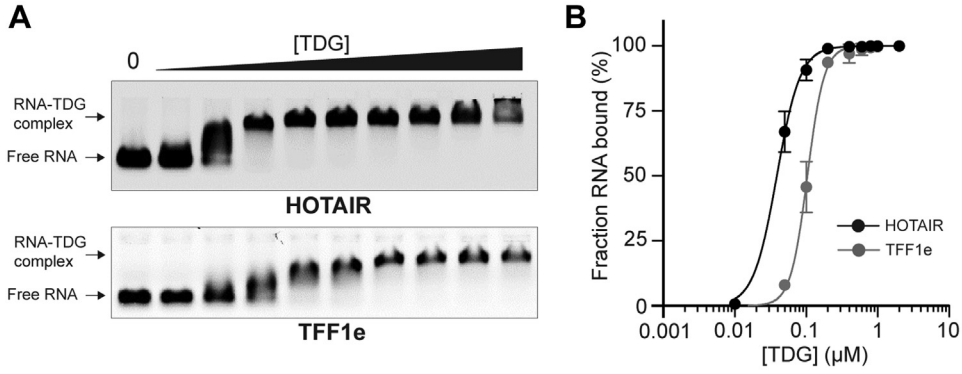


Figure 4. TDG binds endogenous RNA sequences. *A*, representative EMSA data for HOTAIR and TFF1e RNA sequences binding to TDG (0–2 μM). Uncropped gel images are presented in Fig. S12. *B*, saturation plots for binding of TDG to HOTAIR and TFF1e RNAs. Data are mean ± S.D. (n = 3). EMSA, electrophoretic mobility shift assay; HOTAIR, HOX antisense intergenic RNA; TDG, Thymine DNA glycosylase; TFF1e, *TFF1* enhancer RNA transcript.

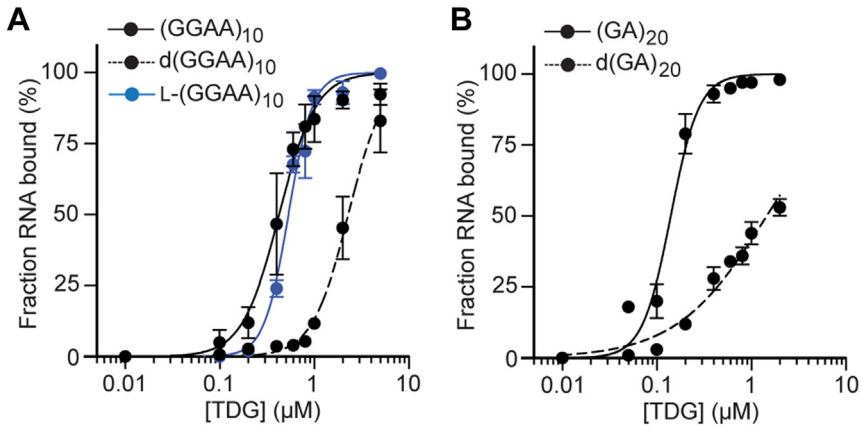
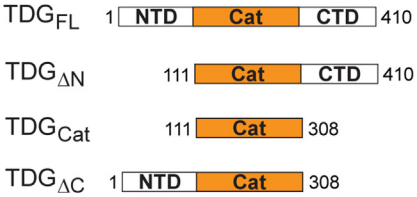


Figure 5. TDG prefers binding RNA over ssDNA. Saturation plots for binding of TDG to $(GGAA)_{10}$ (*A*) and $(GA)_{20}$ (*B*) RNAs and their variants. Data are mean ± SD (n = 3). Representative EMSA data are shown in Fig. S6. TDG, Thymine DNA glycosylase; ssDNA, single-stranded DNA; EMSA, electrophoretic mobility shift assay. EMSA, electrophoretic mobility shift assay; HOTAIR, HOX antisense intergenic RNA; TDG, Thymine DNA glycosylase; TFF1e, *TFF1* enhancer RNA transcript.

TDG's N- and C-terminal domains modulate RNA affinity and selectivity

Work from our laboratory (46) and others (39, 47, 48) have demonstrated the importance of TDG's disordered N-terminal domain (NTD; residues 1–111) and C-terminal domain (CTD: residues 309–410) for DNA binding and catalysis (Fig. 6*A*). To determine whether the NTD and/or CTD contribute to RNA binding, and to gain insight into the overall RNA-binding mechanism, we examined TDG variants lacking one or both of these disordered domains. As shown in Figure 6*B*, deletion of TDG's NTD alone ($TDG_{\Delta N}$) had little effect on its affinity and selectivity for RNA. Thus, despite the cation nature of the NTD, especially residues 82 to 110, which are known to promote nonspecific DNA binding (39, 47, 48), the NTD does not appear to contribute to RNA binding in the context of the CTD. Similarly, deletion of both the NTD and CTD from TDG (*i.e.*, the catalytic domain; TDG_{Cat}) had little effect on poly[G] binding, although this truncation had increased affinity for A_{30} and C_{30} relative to TDG_{FL} and $TDG_{\Delta N}$ (Fig. 6*B*). This result also indicates that the primary interactions between TDG and G-rich RNA occur through the structured catalytic domain. The most dramatic effects on RNA binding were observed upon deletion of the CTD alone ($TDG_{\Delta C}$). The affinity of $TDG_{\Delta C}$ for G_{30} was increased nearly tenfold (K_d = 45 nM, 95% CI 44–47 nM) compared to the other variants tested (Fig. 6*B*). Moreover, $TDG_{\Delta C}$ bound tightly (K_d < 500 nM) and indiscriminately to the other, non-G homopolymers A_{30}, C_{30}, and U_{30}. Together, these data suggest that, while TDG-RNA interactions occur primarily through the catalytic domain in the context of the full-length protein, deletion of the CTD results in an additional, nonspecific RNA binding capacity through the cationic NTD. This model is consistent with the known ability of the CTD to destabilize interactions between cationic residues within the NTD and negatively charged DNA (46, 47), resulting in impaired DNA binding. Our results now suggest that the CTD functions similarly in the context of RNA

B



	TDG_{FL}	$TDG_{\Delta N}$	TDG_{Cat}	$TDG_{\Delta C}$
A_{30}	>5	>5	1220 (1150–1280)	422 (385–458)
G_{30}	434 (396–473)	323 (288–360)	497 (470–525)	45 (44–47)
U_{30}	>5	>5	>5	313 (270–360)
C_{30}	>5	>5	1480 (1320–1670)	345 (320–370)

Figure 6. The influence of TDG's NTD and CTD on RNA binding. *A*, TDG domains discussed in this work. The basic patch within the NTD (residues 82–110) is highlighted. *B*, TDG truncations tested (*left*) and their affinity for homopolymeric RNAs (*right*). Shown are K_d values (nM) with 95% CIs in parentheses (n = 3). Representative EMSA data and saturation binding plots are shown in Figs. S7–S9. EMSA, electrophoretic mobility shift assay; TDG, Thymine DNA glycosylase; NTD, N-terminal domain.

binding by preventing the cationic NTD from engaging RNA. The increased affinity of TDG_{Cat} for A_{30} and C_{30} relative to TDG_{FL} and $TDG_{\Delta N}$ also suggests that the CTD influences RNA binding by the structured catalytic domain. Thus, the CTD plays an important role in modulating TDG's affinity and selectivity for RNA through several mechanisms.

RNA competes with DNA for binding to TDG

The observation that TDG's catalytic domain alone (TDG_{Cat}) binds tightly to RNA raised the possibility that RNA and DNA compete for the same binding site on TDG. Therefore, we carried out a competition experiment by titrating unlabeled RNA with pre-formed TDG-DNA complexes (Fig. 7*A*). For these experiments, we employed a DNA duplex (DNA_{FU}) containing the non-cleavable substrate analogue 2′-deoxy-2′-fluoroarabinouridine (^{F}U) to monitor TDG binding in the absence of base excision (Table S1) (49). Combining DNA_{FU} (100 nM) with twofold excess TDG (200 nM) resulted in the formation of both 1:1 and 2:1 TDG/DNA_{FU} complexes as observed by EMSA (Fig. 7*A*). Previous studies have shown that the 1:1 complex corresponds to a single TDG subunit bound tightly to the ^{F}U site (substrate complex), whereas the 2:1 complex corresponds to a second TDG subunit bound weakly to an adjacent undamaged site (nonspecific complex) (40, 50). Importantly, both the substrate and nonspecific complexes are formed through the same binding site on their respective TDGs. As shown in Figure 7*A*, the addition of G_{30} RNA led to a concentration-dependent decrease in both the 1:1 and 2:1 TDG-DNA_{FU} complexes. In agreement with affinity measurements (50), disruption of the tighter 1:1 complex between TDG and ^{F}U required much higher RNA concentrations when compared to the weaker, nonspecific 2:1 complex. In contrast to G_{30}, A_{30}, which binds very weakly to TDG, had a very modest effect on TDG-DNA_{FU} complexes, even at the highest RNA concentration tested. Together, these data support the conclusion that DNA and RNA share the same or mutually exclusive binding sites on TDG.

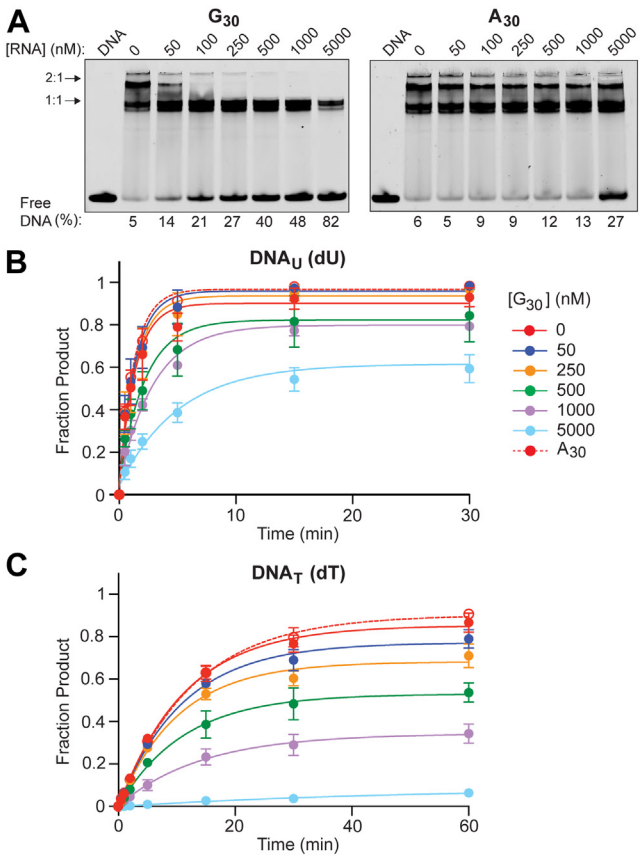


Figure 7. RNA and DNA compete for binding to TDG. *A*, preformed complexes between DNA_{FU} (100 nM) and TDG (200 nM) are disrupted by G_{30} RNA (*left*) but not A_{30} RNA (*right*). The "DNA" marker indicates DNA_{FU} in the absence of TDG and RNA. The percent free (unbound) DNA_{FU} is indicated underneath each lane. *B* and *C*, TDG-mediated excision of DNA_U (*B*) and DNA_T (*C*) is inhibited by RNA. For each reaction, the DNA substrate (100 nM) was mixed with the indicated concentration of G_{30} RNA or 5 μM A_{30} RNA (A_{30} in the legend) followed by the addition of TDG (200 nM). Data are mean ± S.D. (n = 3). TDG, Thymine DNA glycosylase.

The competition between DNA and RNA for binding to TDG suggested that RNA can inhibit base excision by competing against DNA for the enzyme. Therefore, we monitored TDG-mediated excision of DNA containing G•T and G•U mismatches (DNA_T and DNA_U, respectively) in the presence of RNA (Fig. 7*B* and *C* and Table S1). For both lesions, we found that the extent of excised DNA was reduced by the presence of G_{30} RNA in a concentration-dependent manner. The effect of RNA on TDG-mediated excision was greater for G•T mismatches (DNA_T) compared to G•U mismatches (DNA_U), with the presence of 5 μM G_{30} leading to nearly complete inhibition of G•T processing. Again, this observation is consistent with the much higher affinity of TDG for G•U relative to G•T mismatches (K_d = 0.6 nM and 18 nM, respectively) and further supports a competitive binding model (50). Excision of both mismatches by TDG was also inhibited by native TFF1e RNA (Fig. S10), demonstrating that these effects are not an artifact of G_{30} RNA. As expected, A_{30}, which binds very weakly to TDG and competes weakly for DNA binding, had little effect on TDG-mediated excision of both mismatches at the highest RNA concentration tested (5 μM) (Fig. 7, *B* and *C*).

Discussion

In this study, we provided the first quantitative measurements of the affinity of TDG to RNA, clearly demonstrating that RNA binding is a fundamental property of TDG. We showed that TDG's interactions with RNA are dependent on its sequence and secondary structure, with TDG binding preferentially to G-rich RNA sequences. Moreover, we demonstrated that RNA competes with DNA for binding to TDG, providing a novel mechanism through which RNA can regulate TDG-mediated excision. While a link between RNA and TDG-mediated processes (*e.g.*, DNA demethylation) has been established in cells (21, 23, 24, 51), the molecular interactions underlying these relationships have remained poorly understood. The results presented herein provide support for and new insights into a mechanism wherein TDG-mediated processes are regulated through the direct interactions of TDG with RNA.

The average human cell contains around 10 pg total RNA (52). Assuming a cellular volume of ~3000 fl (for HeLa cells) (53–55), we estimate the cellular concentration of a 40-mer unit of RNA to be ~250 μM. Thus, the affinity of TDG to G-rich RNA (K_d ~ 200–500 nM) is likely to be biologically meaningful, even if only a small fraction of this RNA is available for binding. Moreover, our data show that TDG binds G-rich RNA as well as non-specific double-stranded DNA *in vitro* (K_d ~ 300 nM). Considering that the vast majority of cellular DNA is wrapped into nucleosomes, which inhibit DNA binding by TDG, it is reasonable to predict that the competition between DNA and RNA binding could be widespread in cells and may play an important role in localizing TDG to specific sites throughout the genome (discussed below). A competitive binding model also has important implications for the regulation of TDG activity. As we showed previously, RNA had a much greater influence on G•T processing compared to G•U processing, suggesting a role for RNA in directing TDG's substrate selectivity. Regardless of the nucleobase substrate, however, TDG binds very tightly (K_d ~ 1 nM) to its AP site product (50, 56, 57). Thus, even in the presence of very high RNA concentrations, TDG is likely to remain bound to the AP site following excision to facilitate the handover of this reactive intermediate to the downstream-acting enzymes of BER. Of course, it is entirely likely that we have not yet identified the ideal RNA substrate for TDG and that sequences with much higher affinity exist in cells. Our results also do not consider the influence of other proteins or ribonucleoproteins. Therefore, particular biological situations may exist wherein TDG's affinity for RNA is substantially greater than that reported herein.

While further biochemical studies are underway to uncover the molecular mechanisms of RNA recognition by TDG, our results provide initial insight into how TDG binds RNA. In particular, TDG's strong preference for binding to G-rich RNA sequences and its inability to differentiate between a D- and L-ribose sugar backbone strongly indicate that nucleobases play a more dominant role than the sugar/phosphate backbone in TDG-RNA interactions. Our observation that TDG binds weakly to double stranded RNA is also consistent with a binding mode that is reliant on interactions with the nucleobases, which are disfavored with duplex RNA (58, 59). This also implies that the interactions between TDG and RNA are not primarily electrostatic. Instead, TDG may rely more heavily on hydrophobic and/or π-interactions to bind RNA. RNA–protein π interactions, which are predominantly formed between RNA nucleobases and π-containing amino acids, have been shown to play particularly central roles in RNA–protein complexes and contribute considerable stability and selectivity to protein–RNA binding (42, 44, 60). Analyses of π interactions occurring in protein-RNA crystal structures consistently find that most contacts occur with phenylalanine (F) and tyrosine (Y), which prefer a stacked orientation (*i.e.*, π-stacking) relative to the RNA nucleobases (44, 61, 62). Interestingly, the catalytic core of TDG, which appears to be the primary binding site for RNA, has an abundance of these residues (F = 11, Y = 8; 10% of the total amino acids). The catalytic domain also contains eight arginine (R) residues, including R275 within the "insertion loop", which contacts with both strands of the DNA substrate (40). Coincidentally, the insertion loop appears to confer CpG sequence specificity through its interactions with guanine. Like F and Y, R has the propensity to form π–π interactions with nucleobases (44), as well as cation–π interactions, which have been shown to favor G (63). Arginine is also among the most frequent amino acid to form base-specific hydrogen bonds with RNA, providing a potential strategy for G selectivity (58, 59, 64, 65). These observations suggest a potential starting point for identifying key TDG-RNA contacts.

Our results also point to an important role for the CTD in modulating TDG's affinity and selectivity for RNA. We found that deletion of the CTD resulted in an overall increase in RNA affinity, especially when the cationic NTD was present (Fig. 6). The CTD has been shown to destabilize interactions between cationic residues within the NTD and negatively charged DNA, possibly through direct interactions, resulting in impaired DNA binding (47). Our results now suggest that the CTD functions similarly in the context of RNA binding by preventing the cationic NTD from engaging RNA. The proposed model for RNA binding depicted in Figure 8 summarizes our findings. This model raises the intriguing possibility that posttranslational modification of the CTD (*e.g.*, SUMOylation) and/or its interactions with other proteins could enable TDG-RNA interactions to be tuned to fit specific biological contexts (48). Additional studies are needed to explore these ideas further.

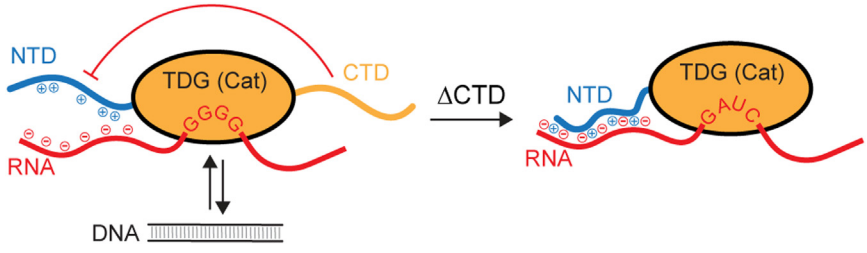


Figure 8. Proposed model for TDG binding to RNA. TDG binds preferentially to G-rich RNA sequences through its structure catalytic domain, resulting in a competition between RNA and DNA for binding to TDG. The CTD prevents nonspecific electrostatic interactions between RNA and the cationic NTD, possibly through direct interactions between the two domains, promoting G-selectivity at the cost of RNA affinity. Upon deletion of the CTD, electrostatic interactions between the NTD and RNA are thus increased, leading to tighter overall binding but reduced selectivity for G. CTD, C-terminal domain; NTD, N-terminal domain; TDG, Thymine DNA glycosylase.

Finally, our findings have important implications for the role of RNA in the active DNA demethylation pathway, in which TDG plays an essential role (8, 9). Active DNA demethylation occurs at specific promoter and enhancer sequences in response to developmental or environmental signals and is often restricted to a few CpG dinucleotides (16, 17, 21). Yet, it remains unclear how this precision is achieved. While interactions between sequence-specific transcription factors and proteins involved in DNA demethylation, including TDG, are known to play a role in targeting DNA demethylation to specific genes, emerging evidence also supports the involvement of RNA (18, 19, 66). Several studies have shown that lncRNAs, which can bind to complementary DNA sequences, can serve as guides to recruit components of the DNA demethylation machinery to specific locations on the genome (23, 24, 67). The lncRNA TETILA, for example, was shown to function as a scaffold to recruit TET2 and TDG to the matrix metalloproteinase-9 promoter (MMP-9), leading to promoter-specific demethylation and MMP-9 expression (24). Our results provide further support for this model and, specifically, a role for RNA in directing TDG activity. TDG's selectivity toward single-stranded G-rich sequences also supports this view, as any model involving RNA-mediated targeting implies some level of specificity. In addition to recruitment, our results suggest a role for RNA in modulating TDG-mediated removal of 5fC/5caC from DNA through a competitive binding mechanism. Indeed, TDG binds 5fC/5caC-containing CpG dinucleotides with a K_d similar to G•T mismatches (50), indicating that RNA will also compete with these substrates for binding to TDG. Thus, for genes that require TDG-mediated demethylation for activation, the local concentration of RNA may prove to be important not only for TDG occupancy but also for coordinating the timing of 5fC/5caC removal.

Experimental Procedures

Reagents

Oligonucleotides were either purchased from Integrated DNA Technologies (IDT) or prepared in-house by solid-phase synthesis on an Expedite 8909 DNA/RNA synthesizer using protocols recommended by the manufacturer. Nucleoside phosphoramidites and other solid-phase synthesis reagents were purchased from Glen Research. Sulfo-Cyanine3 (Cy3) and Cyanine5 (Cy5) hydroxysuccinimide (NHS) ester dyes (cat. nos. 21,320, 23,320) were purchased from Lumiprobe Life Science Solutions, *In vitro* transcribed RNA was purified using the Monarch RNA Cleanup Kit (cat. no. T2030L) purchased from New England BioLabs.

Expression and purification of TDG

Full-length human TDG and truncated TDG variants used herein were expressed and purified as described previously (68, 69). Purified proteins were stored at −80 °C in HP50 buffer (50 mM HEPES, pH 8, 50 mM NaCl, 10% glycerol, 10 mM BME, 1 mM PMSF) until use.

RNA synthesis and purification

All synthetic RNA oligonucleotides (Table S1) were prepared in house using protocols recommended by the manufacturer. Prior to use, oligonucleotides were purified by 20% denaturing PAGE (19:1 acrylamide:bisacrylamide). Excised bands containing the purified oligonucleotides were eluted overnight at 23 °C in a buffer containing 200 mM NaCl, 10 mM EDTA, and 10 mM Tris (pH 7.6). The gel fragments were removed from the solution by centrifugation, and the eluted oligonucleotides were concentrated using a 3 kDa pore-size Amicon Ultra Centrifugal Filter (MilliporeSigma). The concentrated samples were then desalted by ethanol precipitation, and the final concentration was determined using the absorbance at 260 nm on a NanoDrop 2000c (ThermoFisher). The identity of all novel oligonucleotides was confirmed by MS (Figs. S20–S27).

Oligonucleotides were fluorescently labeled using Sulfo-Cy3 and -Cy5 NHS ester dyes *via* conjugation to the 3' end of oligonucleotides using a 3′-amino modification (3′-amino-modifier C6 CPG) incorporated during synthesis. The conjugation was completed by combining the amino-modified oligonucleotide at 100 μM with a 5 mM final concentration of dye NHS ester in 0.1 M sodium bicarbonate buffer (pH 8.5). The conjugation reaction was shaken overnight at 23 °C. To remove excess dye, samples were ethanol precipitated and purified once more *via* 20% denaturing PAGE, as described previously. Stock solutions were prepared by dilution of the oligonucleotide to 100 nM in folding buffer (25 mM NaCl, 10 mM HEPES (pH 7.8), 0.1 mM PMSF, and 50 mM KCl).

HOTAIR and TFF1e RNAs were prepared by *in vitro* transcription from the corresponding DNA templates (Tables S2 and S3). The DNA template used to transcribe TFF1e RNA, referred to as TFF1e-DNA, was generated by PCR amplification of 150 ng of human genomic DNA using primers TFF1eFWD and TFF1eREV (Table S2) employing Phusion High-Fidelity DNA Polymerase according to the manufacturer's instructions. Similarly, the DNA template used to transcribe HOTAIR RNA, referred to as HOTAIR-DNA, was generated by PCR amplification of 150 ng of human genomic DNA using primers HOTAIRFWD and HOTAIRREV (Table S2). Following PCR amplification, TFF1e-DNA and HOTAIR-DNA were purified and desalted using the GenCatch Advanced PCR Extraction kit and used in an *in vitro* transcription reaction without further purification. Transcription reactions were carried out using 200 pmol of DNA template per 100 μl total reaction volume. The reaction mixture consisted of 10 U/μl T7 RNA polymerase, 0.001 U/μl Inorganic pyrophosphatase (IPP), 25 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 40 mM Tris (pH 7.9), 5 mM of each of the four NTPs, and 0.5 mM 5-aminoallyl-UTP. The reaction mixture was incubated at 37 °C for 2 h followed by the addition of 2 U of Turbo DNAses (Life Tech, Carlsbad, CA). After 30 min at 37 °C, the transcribed RNA was purified using the Monarch RNA Cleanup Kit. Labelling was carried out as described earlier using Sulfo-Cy5 NHS ester dye. After the labeling was complete, RNAs were ethanol precipitated and purified *via* 20% denaturing PAGE as described above.

Electrophoretic mobility shift assays

The dissociation constant (K_d) of TDG and its truncated variants for the various RNA ligands were determined by EMSA as described previously (30). Briefly, 5 to 50 nM Cy5-labeled RNA was mixed with various concentration of TDG in a reaction mixture containing 37.5 mM NaCl, 12.5 mM KCl, 10 mM HEPES (pH 7.8), 2.5 mM BME, and 5% glycerol. After incubating for 30 min at 23 °C, an aliquot was removed and loaded on a nondenaturing 1% agarose gel buffered with 1×TBE at 4 °C. Electrophoresis was carried out for 45 min at 6 to 8 V/cm (0.75 mm thick gel) and the gel visualized using a Typhoon FLA-9500 Molecular Imager (General Electric Co). Images were quantified using ImageQuant TL software imager (GE Healthcare Lifesciences). The area around the unbound species was tightly boxed, whereas the area for the bound species included both the discretely shifted band(s) and the area between the bands. Thus, any intermediate species were included in the bound fraction. Equations for specific binding with Hill slope were fit using GraphPad Prism 9 Version 9.4.1. Uncropped gel images are presented in the Supporting Information (Figs. S11–S19).

Circular dichroism spectroscopy

Oligonucleotides (5 μM) were folded in a buffer containing 37.5 mM NaCl, 12.5 mM KCl, 10 mM HEPES (pH 7.8), 2.5 mM BME, and 5% glycerol. Circular dichroism spectra were obtained from a 350 μl sample in a quartz cuvette using an Applied Photophysics Chirascan spectrophotometer (Leatherhead) at 1 nm intervals from 220 to 370 nm at room temperature.

DNA competition experiments

The double-stranded DNA substrate DNA_{FU} was prepared by annealing 6 μM each FU-FWD and FU-REV (Table S1) in a buffer consisting of 50 mM NaCl and 20 mM HEPES (pH 7.8). The reaction mixture was heated at 95 °C for 1 min and slowly cooled down to room temperature. DNA competition experiments were analyzed using electrophoretic mobility shift assays (EMSAs). To generate preformed DNA_{FU}/TDG complexes, DNA_{FU} (100 nM) was mixed with TDG_{FL} (200 nM) in a buffer consisting of 25 mM NaCl, 10 mM Tris (pH 7.6), and 5% glycerol. The reaction mixture was incubated at 30°C for 30 min before the indicated concentrations (50–5000 nM) of RNA (G_{30} or A_{30}) were added. After incubating at 30 °C for another 30 min, an aliquot was loaded onto a 10% native PAGE (19:1 acrylamide:bisacrylamide) containing 0.5× TBE and 5% glycerol. Electrophoresis was carried out for 45 min at 6 to 8 V/cm (0.75 mm thick gel) at 4 °C, and the gel was visualized using a Typhoon FLA-9500 Molecular Imager as described above.

Glycosylase assays

The double-stranded DNA substrates DNA_U and DNA_T were prepared as described above by annealing 6 μM each dU-FWD with dU-REV and dT-FWD with dT-REV, respectively (Table S1). Glycosylase reactions were prepared by mixing 100 nM DNA substrate with the indicated concentration of RNA in a buffer consisting of 25 mM NaCl, 2.5 mM $MgCl_2$, and 10 mM HEPES (pH 7.8). Experiments were initiated by adding TDG (200 nM) to the buffered substrates and were allowed to incubate at 30 °C. Aliquots (2 μl) were removed at the indicated times and quenched by the addition of a solution (2 μl) of 1% SDS in water. The AP site product was then cleaved by the addition of an equivalent volume of 0.2 M NaOH, followed by the addition of 8 μl of denaturing loading buffer (90% formamide, 10 mM EDTA, pH 8). Products were resolved by 20% denaturing PAGE (19:1 acrylamide:bisacrylamide), and the gel was visualized using a Typhoon FLA-9500 Molecular Imager as described previously.

Data Availability

The data generated during all experiments is available from the author upon reasonable request.

Supporting information—This article contains supporting information. Figures S1–S27 and Tables S1–S3 (33, 70).

Author contributions—L. A. M. and B. Z.: Methodology, Investigation, Formal analysis, Writing – Review & Editing; A. G.: Investigation, Resources; J. T. S.: Conceptualization, Supervision, Funding acquisition, Writing – Original draft, Methodology.

Funding and additional infromation—This material is based upon work supported by the National Science Foundation [NSF2126416]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflict of interests—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: AP, abasic site; CTD, C-terminal domain; eRNA, enhancer RNAs; TDG, Thymine DNA glycosylase; EMSA, electrophoretic mobility shift assay; HOTAIR, HOX antisense intergenic RNA; lncRNA, long noncoding RNAs; MMP-9, matrix metalloproteinase-9 promoter; NTD, N-terminal domaint; SSRNA, single-stranded RNA; SSDNA, single-stranded DNA; TFF1e, TFF1 enhancer RNA transcrip.

References

1. Schermerhorn, K. M., and Delaney, S. (2014) A chemical and kinetic perspective on base excision repair of DNA. *Acc. Chem. Res.* **47**, 1238–1246
2. Brooks, S. C., Adhikary, S., Rubinson, E. H., and Eichman, B. F. (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* **1834**, 247–271
3. Drohat, A. C., and Coey, C. T. (2016) Role of base excision "repair" enzymes in erasing epigenetic marks from DNA. *Chem. Rev.* **116**, 12711–12729
4. Wiebauer, K., and Jiricny, J. (1990) Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 5842–5845

5. Neddermann, P., and Jiricny, J. (1994) Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 1642–1646
6. Maiti, A., and Drohat, A. C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338
7. He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307
8. Schuermann, D., Weber, A. R., and Schär, P. (2016) Active DNA demethylation by DNA repair: facts and uncertainties. *DNA Repair* **44**, 92–102
9. Wu, X., and Zhang, Y. (2017) TET-Mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534
10. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, Ana C., Fung, H.-L., *et al.* (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706
11. Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., *et al.* (2015) Single-base resolution analysis of 5-Formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.* **10**, 674–683
12. Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014) Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **32**, 1231–1240
13. Song, C.-X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., *et al.* (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691
14. Xia, B., Han, D., Lu, X., Sun, Z., Zhou, A., Yin, Q., *et al.* (2015) Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Met.* **12**, 1047–1050
15. Sun, Z., Dai, N., Borgaro, J. G., Quimby, A., Sun, D., Corrêa, I. R., *et al.* (2015) A sensitive approach to map genome-wide 5-Hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Mol. Cell* **57**, 750–761
16. Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., *et al.* (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423
17. Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., *et al.* (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79
18. Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., and Schär, P. (2007) The enigmatic thymine DNA glycosylase. *DNA Repair* **6**, 489–504
19. Sjolund, A. B., Senejani, A. G., and Sweasy, J. B. (2013) MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles. *Mutat. Res.* **743-744**, 12–25
20. Tini, M., Benecke, A., Um, S.-J., Torchia, J., Evans, R. M., and Chambon, P. (2002) Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell* **9**, 265–277
21. Kolendowski, B., Hassan, H., Krstic, M., Isovic, M., Thillainadesan, G., Chambers, A. F., *et al.* (2018) Genome-wide analysis reveals a role for TDG in estrogen receptor-mediated enhancer RNA transcription and 3-dimensional reorganization. *Epigenetics Chromatin* **11**, 5
22. Nair, S. J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M. J., *et al.* (2019) Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat. Struct. Mol. Biol.* **26**, 193–203
23. Arab, K., Park, Y. J., Lindroth, A. M., Schäfer, A., Oakes, C., Weichenhan, D., *et al.* (2014) Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 *via* GADD45A. *Mol. Cell* **55**, 604–614
24. Zhou, L., Ren, M., Zeng, T., Wang, W., Wang, X., Hu, M., *et al.* (2019) TET2-interacting long noncoding RNA promotes active DNA demethylation of the MMP-9 promoter in diabetic wound healing. *Cell Death Dis.* **10**, 813
25. Wiedner, H. J., and Giudice, J. (2021) It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. *Nat. Struct. Mol. Biol.* **28**, 465–473
26. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406
27. Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., *et al.* (2018) A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699
28. Alshareedah, I., Kaur, T., Ngo, J., Seppala, H., Kounatse, L.-A. D., Wang, W., *et al.* (2019) Interplay between short-range attraction and long-range repulsion controls reentrant liquid condensation of ribonucleoprotein–RNA complexes. *J. Am. Chem. Soc.* **141**, 14593–14602
29. Seemann, S. E., Mirza, A. H., Hansen, C., Bang-Berthelsen, C. H., Garde, C., Christensen-Dalsgaard, M., *et al.* (2017) The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Res.* **27**, 1371–1383
30. Wang, X., Goodrich, K. J., Gooding, A. R., Naeem, H., Archer, S., Paucek, R. D., *et al.* (2017) Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell* **65**, 1056–1067
31. Huang, Z.-L., Dai, J., Luo, W.-H., Wang, X.-G., Tan, J.-H., Chen, S.-B., *et al.* (2018) Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J. Am. Chem. Soc.* **140**, 17945–17955
32. Dumas, L., Herviou, P., Dassi, E., Cammas, A., and Millevoi, S. (2021) G-quadruplexes in RNA biology: recent advances and Future directions. *Trends Biochem. Sci.* **46**, 270–283
33. Deckard, C. E., and Sczepanski, J. T. (2018) Polycomb repressive complex 2 binds RNA irrespective of stereochemistry. *Chem. Commun.* **54**, 12061–12064
34. Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R. M., *et al.* (2019) RNA structure drives interaction with proteins. *Nat. Commun.* **10**, 3246
35. Kowalczykowski, S. C., Paul, L. S., Lonberg, N., Newport, J. W., McSwiggen, J. A., and Von Hippel, P. H. (1986) Cooperative and noncooperative binding of protein ligands to nucleic acid lattices: experimental approaches to the determination of thermodynamic parameters. *Biochemistry* **25**, 1226–1240
36. Epstein, I. R. (1979) Kinetics of nucleic acid–large ligand interactions: exact Monte Carlo treatment and limitingl cases of reversible binding. *Biopolymers* **18**, 2037–2050
37. Davidovich, C., Zheng, L., Goodrich, K. J., and Cech, T. R. (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257
38. Wang, X., Schwartz, J. C., and Cech, T. R. (2015) Nucleic acid-binding specificity of human FUS protein. *Nucl. Acids Res.* **43**, 7535–7543
39. Coey, C. T., Malik, S. S., Pidugu, L. S., Varney, K. M., Pozharski, E., and Drohat, A. C. (2016) Structural basis of damage recognition by thymine DNA glycosylase: key roles for N-terminal residues. *Nucl. Acids Res.* **44**, 10248–10258
40. Maiti, A., Morgan, M. T., Pozharski, E., and Drohat, A. C. (2008) Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8890–8895
41. Bhan, A., and Mandal, S. S. (2015) LncRNA HOTAIR: a master regulator of chromatin dynamics and cancer. *Biochim. Biophys. Acta - Rev. Cancer* **1856**, 151–164
42. Pal, A., and Levy, Y. (2019) Structure, stability and specificity of the binding of ssDNA and ssRNA with proteins. *PLoS Comput. Biol.* **15**, e1006768
43. Dickey, T. H., Altschuler, S. E., and Wuttke, D. S. (2013) Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* **21**, 1074–1084
44. Wilson, K. A., Holland, D. J., and Wetmore, S. D. (2016) Topology of RNA-protein nucleobase-amino acid π-π interactions and comparison to analogous DNA-protein π-π contacts. *RNA* **22**, 696–708
45. Chen, H., Meisburger Steve, P., Pabit Suzette, A., Sutton Julie, L., Webb Watt, W., and Pollack, L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 799–804
46. Deckard, C. E., III, and Sczepanski, J. T. (2021) Reversible chromatin condensation by the DNA repair and demethylation factor thymine DNA glycosylase. *Nucl. Acids Res.* **49**, 2450–2459

47. Smet-Nocca, C., Wieruszeski, J.-M., Chaar, V., Leroy, A., and Benecke, A. (2008) The Thymine–DNA glycosylase regulatory domain: residual structure and DNA binding. *Biochemistry* **47**, 6519–6530
48. Steinacher, R., and Schär, P. (2005) Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr. Biol.* **15**, 616–623
49. Maiti, A., Morgan, M. T., and Drohat, A. C. (2009) Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase. *J. Biol. Chem.* **284**, 36680–36688
50. Morgan, M. T., Maiti, A., Fitzgerald, M. E., and Drohat, A. C. (2011) Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. *Nucl. Acids Res.* **39**, 2319–2329
51. Boland, M. J., and Christman, J. K. (2008) Characterization of dnmt3b: thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J. Mol. Biol.* **379**, 492–504
52. Finka, A., Sood, V., Quadroni, M., Rios, P. D. L., and Goloubinoff, P. (2015) Quantitative proteomics of heat-treated human cells show an across-the-board mild depletion of housekeeping proteins to massively accumulate few HSPs. *Cell Stress Chaperones* **20**, 605–620
53. Zhao, L., Kroenke, C. D., Song, J., Piwnica-Worms, D., Ackerman, J. J. H., and Neil, J. J. (2008) Intracellular water-specific MR of microbead-adherent cells: the HeLa cell intracellular water exchange lifetime. *NMR Biomed.* **21**, 159–164
54. Cohen, L. S., and Studzinski, G. P. (1967) Correlation between cell enlargement and nucleic acid and protein content of hela cells in unbalanced growth produced by inhibitors of DNA synthesis. *J. Cell Physiol.* **69**, 331–339
55. Park, K., Jang, J., Irimia, D., Sturgis, J., Lee, J., Robinson, J. P., *et al.* (2008) 'Living cantilever arrays' for characterization of mass of single live cells in fluids. *Lab Chip* **8**, 1034–1041
56. Waters, T. R., Gallinari, P., Jiricny, J., and Swann, P. F. (1999) Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* **274**, 67–74
57. Fitzgerald, M. E., and Drohat, A. C. (2008) Coordinating the initial steps of base excision repair: Apurinic/apyrimidinic endonuclease 1 actively stimulates thymine dna glycosylase by disrupting the product complex. *J. Biol. Chem.* **283**, 32680–32690
58. Gupta, A., and Gribskov, M. (2011) The role of RNA sequence and structure in RNA–protein interactions. *J. Mol. Biol.* **409**, 574–587
59. Allers, J., and Shamoo, Y. (2001) Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **311**, 75–86
60. Corley, M., Burns, M. C., and Yeo, G. W. (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* **78**, 9–29
61. Morozova, N., Allers, J., Myers, J., and Shamoo, Y. (2006) Protein–RNA interactions: Exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinform* **22**, 2746–2752
62. Jones, S., Daley, D. T. A., Luscombe, N. M., Berman, H. M., and Thornton, J. M. (2001) Protein–RNA interactions: a structural analysis. *Nucl. Acids Res.* **29**, 943–954
63. Zhang, H., Li, C., Yang, F., Su, J., Tan, J., Zhang, X., *et al.* (2014) Cation-pi interactions at non-redundant protein-RNA interfaces. *Biochemistry (Moscow)* **79**, 643–652
64. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucl. Acids Res.* **29**, 2860–2874
65. Han, K., and Nepal, C. (2007) PRI-Modeler: extracting RNA structural elements from PDB files of protein–RNA complexes. *FEBS Lett.* **581**, 1881–1890
66. Hassan, H. M., Kolendowski, B., Isovic, M., Bose, K., Dranse, H. J., Sampaio, A. V., *et al.* (2017) Regulation of active DNA demethylation through RAR-mediated recruitment of a TET/TDG complex. *Cell Rep.* **19**, 1685–1697
67. Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., *et al.* (2019) GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat. Genet.* **51**, 217–223
68. Deckard, C. E., Banerjee, D. R., and Sczepanski, J. T. (2019) Chromatin structure and the pioneering transcription factor FOXA1 regulate TDG-mediated removal of 5-formylcytosine from DNA. *J. Am. Chem. Soc.* **141**, 14110–14114
69. Neddermann, P., and Jiricny, J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.* **268**, 21218–21224
70. Maiti, A., Noon, M. S., MacKerell, A. D., Jr, Pozharski, E., and Drohat, A. C. (2012) Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc. Natl. Acad. Sci. USA* **109**, 8091–8096