nature communications

Article

https://doi.org/10.1038/s41467-024-53497-7

Thymine DNA glycosylase combines sliding, hopping, and nucleosome interactions to efficiently search for 5-formylcytosine

Received: 19 April 2024

Accepted: 10 October 2024

Published online: 25 October 2024

Check for updates

Brittani L. Schnable[1,2], Matthew A. Schaich[2,3], Vera Roginskaya[2,3], Liam P. Leary[2,3], Tyler M. Weaver[4], Bret D. Freudenthal[4], Alexander C. Drohat[5] & Bennett Van Houten[1,2,3]

Base excision repair is the main pathway involved in active DNA demethylation. 5-formylcytosine and 5-carboxylcytosine, two oxidized moieties of methylated cytosine, are recognized and removed by thymine DNA glycosylase (TDG) to generate an abasic site. Using single molecule fluorescence experiments, we study TDG in the presence and absence of 5-formylcytosine. TDG exhibits multiple modes of linear diffusion, including hopping and sliding, in search of base modifications. TDG active site variants and truncated N-terminus, reveals these variants alter base modification search and recognition mechanism of TDG. On DNA containing an undamaged nucleosome, TDG is found to either bypass, colocalize with, or encounter but not bypass the nucleosome. Truncating the N-terminus reduces the number of interactions with the nucleosome. Our findings provide mechanistic insights into how TDG searches for modified DNA bases in chromatin.

DNA can be damaged by a variety of endogenous and exogenous sources and these potentially mutagenic bases are removed via base excision repair (BER)[1]. In humans, BER is initiated by one of eleven damage-specific DNA glycosylases that recognize a base modification and cleave the glycosidic bond to leave an abasic site. The repair is completed by the actions of several proteins that restore the correct base pair. Thymine DNA glycosylase (TDG) is a monofunctional glycosylase that initiates BER by recognizing a specific repertoire of lesions, including 5-methylcytosine (5mC) that has been deaminated to form a G:T mismatch or 5mC that has been oxidized by the ten eleven translocation (TET) family of enzymes into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[2–5]. Interestingly, TDG is the only known mammalian DNA glycosylase that creates an embryonic lethal phenotype in mice upon depletion: presumably due to the significant role that TDG plays in active demethylation, in addition to the action on environmental DNA damage[6,7].

While TDG is known to preferentially remove T from G:T mismatches in the context of a CpG, other substrates do not show this sequence context dependence[8,9]. One such base modification is 5fC, which has an abundance of 6×10^4 per human genome, and the removal of 5fC was similar in the context of a CpG and outside a CpG[10–12]. Previous studies have shown that N140 of TDG is essential for the catalytic activity on 5fC because of interactions with the water nucleophile. The N140A variant greatly reduces the catalytic activity 16,000-fold[5]. It has been shown that R275 plays an essential role in catalysis by filling in the void in the DNA that is generated once the substrate nucleobase is flipped into the active site[13]. In addition to TDG specificity for G:T mismatches and 5fC, TDG is able to bind to DNA nonspecifically with an affinity of 293 ± 64 nM and has a footprint of 10 base pairs[14,15]. While the roles for these key active site amino acids in catalysis is well-established, their role in mediating TDG searching for DNA base modifications in the presence of excess unmodified DNA is unknown.

[1]Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburg, Pittsburgh, PA, USA. [2]UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. [3]Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [4]Department of Biochemistry and Molecular Biology, Department of Cancer Biology, University of Kansas Medical Center, Kansas City, KS, USA. [5]Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD, USA. e-mail: vanhoutenb@upmc.edu

A major unexplored consideration for any glycosylase search process in a cell is the presence of chromatin. DNA packaged into nucleosomes would interrupt a DNA sliding mechanism and inhibit the accessibility of specific sites in the absence of nucleosome sliding and/or remodeling. Among DNA glycosylases this phenomenon is particularly relevant to TDG, as the methylated DNA where its substrates are formed are associated with higher degree of chromatinization[16]. TDG has also been shown to directly interact with nucleosome core particles, however kinetic studies also demonstrated TDG activity was only reduced ~two-fold for uracil substrates present on a nucleosome core particle[17,18].

Here, we utilized single-molecule analysis of DNA-binding proteins from nuclear extracts (SMADNE)[19] to directly observe the search process of fluorescently labeled TDG in real-time on DNA containing 5fC substrates or nucleosome core particles. By observing TDG scanning the DNA as it searches for 5fC, we determined that TDG searches for damaged sites using a combination of sliding and hopping and that TDG diffusivity was highly dependent on DNA tension. This combined search mechanism was also observed on DNA containing NCPs. In some instances, TDG hopped over nucleosomes to continue searching the DNA, but in other cases it approached the NCP but did not bypass. We have also found that specific amino acid side chains that are essential for catalytic activity also impact the dwell time and the ability of TDG to diffuse on DNA. Additionally, although ΔN-term TDG (residues 82–410) is catalytically active, we demonstrate that the N-terminus plays a significant role in binding and searching unmodified DNA as well as playing a key role in engaging undamaged NCPs.

Results

TDG binds DNA specifically and nonspecifically

TDG-HaloTag was overexpressed in U2OS cells, and the nuclear extracts were quantified using western blots and SDS gels to determine a 60-fold difference in overexpressed fluorescently tagged WT TDG to endogenous TDG (Supplementary Fig. 2). To determine the binding behavior of TDG on DNA containing a base modification of interest, we initially performed the single-molecule imaging with TDG-HaloTag-JF635 and DNA containing 5fC generated by nick translation (Fig. 1a, c). Analysis of the resulting kymographs revealed complex binding dynamics, including non-motile (12.6%) and motile TDG binding events (i.e., exhibited 1D diffusion, 87.4%). Further co-localization analysis of red TDG-HaloTag-JF635 with the blue 5fC fiducial marker in the DNA (fluorescein dUTP) revealed the stationary events were primarily TDG bound to 5fC sites (Fig. 1d). The motile events primarily represent TDG binding to the DNA nonspecifically. Of note, diffusion events that are <0.3 s are not detected. Fitting the motile events to a cumulative residence time distribution plot (CRTD) with a one-phase exponential decay, a binding lifetime of 7.5 ± 0.3 s was obtained (Fig. 1d, e and Table 1). Only motile events with a binding lifetime of 7.9 ± 0.1 sec were observed in additional single-molecule imaging experiments with TDG-HaloTag-JF635 and unmodified DNA, supporting that the stationary events reflect TDG bound to a base modification (Fig. 1f, g and Table 1). In contrast to the motile events, the CRTD plot of stationary events that co-localized with the 5fC fiducial marker were fit with a two-phase exponential decay with binding lifetimes of 14.7 ± 1.7 and

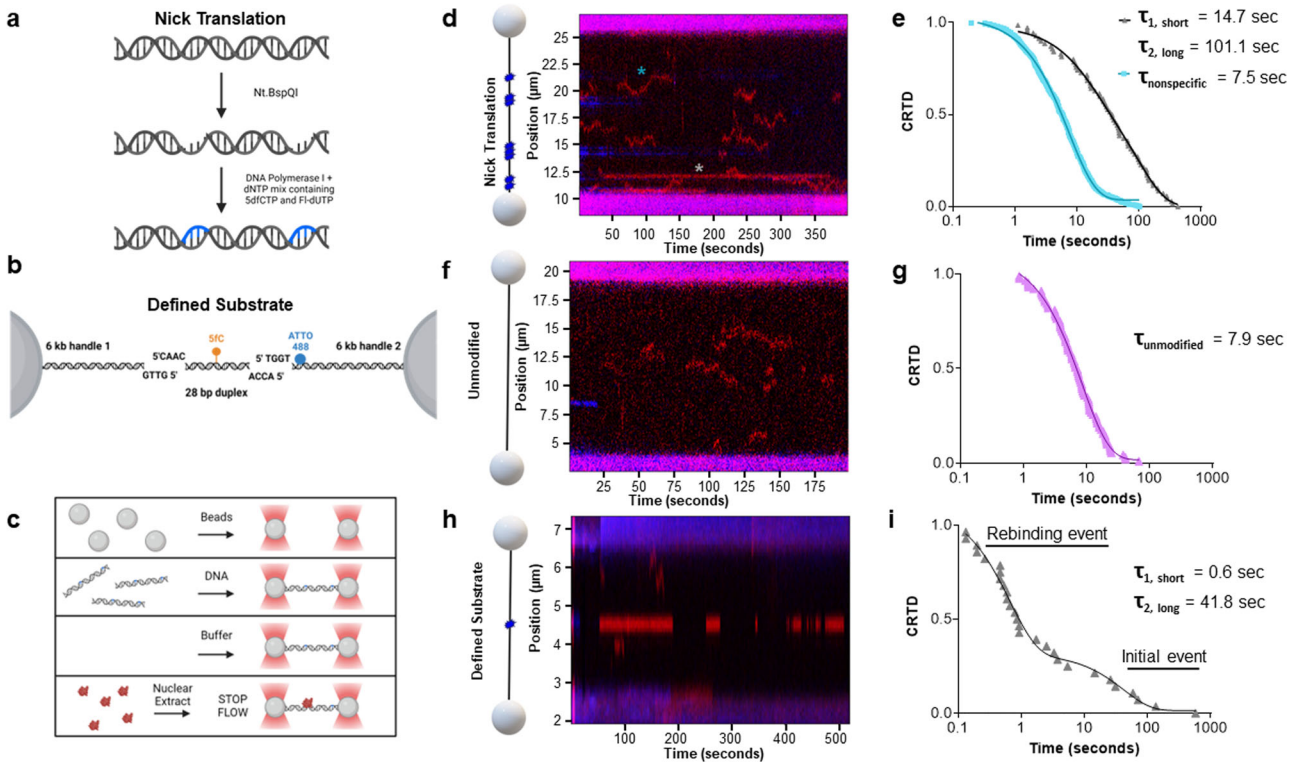

Fig. 1 | TDG binds DNA specifically and nonspecifically. a Cartoon schematic showing how 5fC was incorporated into nick translated λ DNA. Created in BioRender. Schnable, B. (2023) BioRender.com/q21t266. **b** Cartoon schematic showing 28 base pair duplex DNA containing a single 5fC (orange) ligated into 6 kb LUMICKS handle kit, with handle 2 containing ATTO 488 (blue). Created in BioRender. Schnable, B. (2024) BioRender.com/z84c645. **c** A diagram depicting the order of reagents, which are under laminar flow, are captured in the flowcell. Created in BioRender. Schnable, B. (2023) BioRender.com/k78o196. **d** A cartoon depiction of the DNA substrate used for TDG binding, with 5fC sites shown in blue, and an example kymograph with TDG binding shown in red. Specific event indicated with gray asterisk and nonspecific event indicated with teal asterisk. **e** Cumulative Resident Time Distribution (CRTD) analysis fit to a two-phase decay of TDG binding DNA containing 5fC specifically ($n = 70$) and nonspecifically ($n = 487$). Data represents the mean ± SEM from four independent experiments. **f** A cartoon depiction of the unmodified DNA substrate with an example kymograph of TDG binding and moving. **g** CRTD analysis fit to a one-phase decay of TDG binding unmodified λ DNA ($n = 155$). Data represents the mean ± SEM of the fit from three independent experiments. **h** An example kymograph of TDG binding to a single 5fC. **i** CRTD analysis fit to a two-phase decay of TDG binding to 5fC ($n = 28$). Data represents the mean ± SEM of the fit from three independent experiments.

101.1 ± 7.1 s (Fig. 1d, e and Table 1). We hypothesize that these longer TDG lifetimes represent binding to and cleaving the 5fC. This is consistent with previous kinetic experiments that indicate it takes ~ 68 secs (k_{max} of 0.61 min^{-1}) to cleave 5fC at 22 °C, plus the time for TDG to dissociate from the generated abasic site at[2]. Notably, the shorter 14.7 ± 1.7 s binding events, that co-localized with the 5fC fiducial marker, exclusively occur after a long-lived event. These shorter events likely represent TDG binding to an abasic site generated by its catalytic function or a nicked abasic site processed by apurinic/apyrimidinic endonuclease 1 (APE1) in the extracts[19].

Since the nick translated DNA contains multiple sites with a region of 5fC, we generated a 12 kb DNA substrate that contains a single center-positioned 5fC to more easily observe if lifetime differs over time, indicating turnover. We then performed the correlative optical tweezers-fluorescence microscopy (CTFM) experiment using TDG-HaloTag-JF635 with the single 5fC DNA. A CRTD plot of stationary events that co-localized with the 5fC were fit with a two-phase exponential decay with binding lifetimes of 0.67 ± 0.06 sec and 41.8 ± 12.2 sec (Fig. 1d, e, h, i and Table 1), consistent with the short- and long-lived binding lifetimes observed on the DNA substrate containing 10 5fC sites (Fig. 1e). These data further indicate that long-lived events represent TDG binding to the substrate and the short-lived events are mostly likely TDG rebinding to the catalytic product abasic site (Fig. 1h, i).

Table 1 | Binding lifetimes and diffusivity

Protein	DNA	Lifetime(s) and Percentages[a]	Average binding lifetime (τ_{avg})[a]	Diffusivity on undamaged DNA ($\mu m^2/s$)[b]
Wild Type	Unmodified	7.9 ± 0.06		0.028 ± 0.02
	5fC, specific[c]	101.1 ± 7.1 (67.3 ± 3%)	72.9 ± 5.4	
		14.7 ± 1.7 (32.7 ± 3%)		
	5fC, nonspecific	7.5 ± 0.3		
	Undamaged NCP	16.5 ± 1.9		
N140A	Unmodified	7.7 ± 0.24		0.022 ± 0.02
	5fC, specific	52.6 ± 1.1		
	5fC, nonspecific	15.9 ± 0.23		
R275A	Unmodified	2.8 ± 0.13		0.130 ± 0.09
	5fC, specific	10.0 ± 1.1		
	5fC, nonspecific	1.07 ± 0.02		
R275L	Unmodified	1.8 ± 0.02		0.087 ± 0.07
	5fC, specific	1.7 ± 0.06		
	5fC, nonspecific	0.7 ± 0.02		
ΔN term	5fC, specific	24.55 ± 3.57	10.2 ± 1.6	
		1.11 ± 18		
	5fC, nonspecific	1.2 ± 0.02		

[a]The tau is the best fit value ± the standard error of the fit to the observed data.
[b]The diffusivity is the mean ± the standard error of the mean.
[c]Nick translated DNA.

TDG exhibits mixed modes of linear diffusion during its search for DNA damage

After observing TDG bind and move on DNA nonspecifically, we sought to determine the mode of linear diffusion TDG uses on DNA. Linear diffusion occurs through sliding and hopping mechanisms that have distinct behaviors. Sliding is when a protein tracks the DNA helix without dissociating, leading to coupled translational and rotational diffusion[20,21]. Alternatively, hopping is when a protein micro-dissociates as it translocates along the DNA, leading to uncoupled translational and rotational diffusions. Notably, the sliding and hopping mechanisms can be distinguished by determining whether the diffusion coefficient of TDG changes as a function of salt concentration: diffusion via sliding is unaffected by salt concentration, whereas diffusion for hopping is sensitive to salt concentration[22–24]. To further differentiate whether TDG uses a sliding, hopping, or multimodal 1D diffusion, we performed additional single-molecule imaging experiments with TDG-HaloTag-JF635 and unmodified DNA at varying ionic concentrations (75–150 mM NaCl). We observed an increase in the diffusion from 1.6×10^{-2} $\mu m^2/s$ at 75 mM NaCl, to 2.8×10^{-2} $\mu m^2/s$ at 100 mM NaCl, and 3.1×10^{-2} $\mu m^2/s$ at 150 mM NaCl (Fig. 2a, Table 1, and Supplementary Fig. 5a–d). The theoretical limit of diffusion of TDG-HaloTag sliding along the DNA was calculated to be 2.56×10^{-1} $\mu m^2/s$

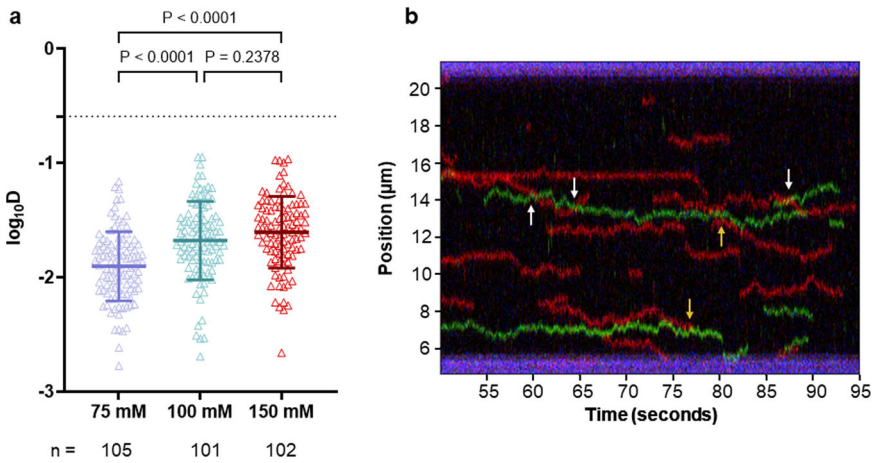

Fig. 2 | TDG exhibits linear diffusion on DNA. a Scatter plot of the diffusion coefficient ($\log_{10}D$) calculated for TDG with increasing ionic strength on unmodified λ. Dashed line, D_{lim}, theoretical limit to free diffusion for TDG-HaloTag (75 mM NaCl $n = 105$; 100 mM NaCl $n = 101$; 150 mM NaCl $n = 102$). *P*-values determined by two-way ANOVA. Data represents the mean ± SD from three independent experiments. **b** An example kymograph of the two-color TDG experiment. Separate TDG-HaloTag extracts were labeled with either JF635 (red) or JF552 (green) and mixed at a 1:1 ratio. White arrows indicate bypass events and yellow arrows indicate collision events.

https://doi.org/10.1038/s41467-024-53497-7

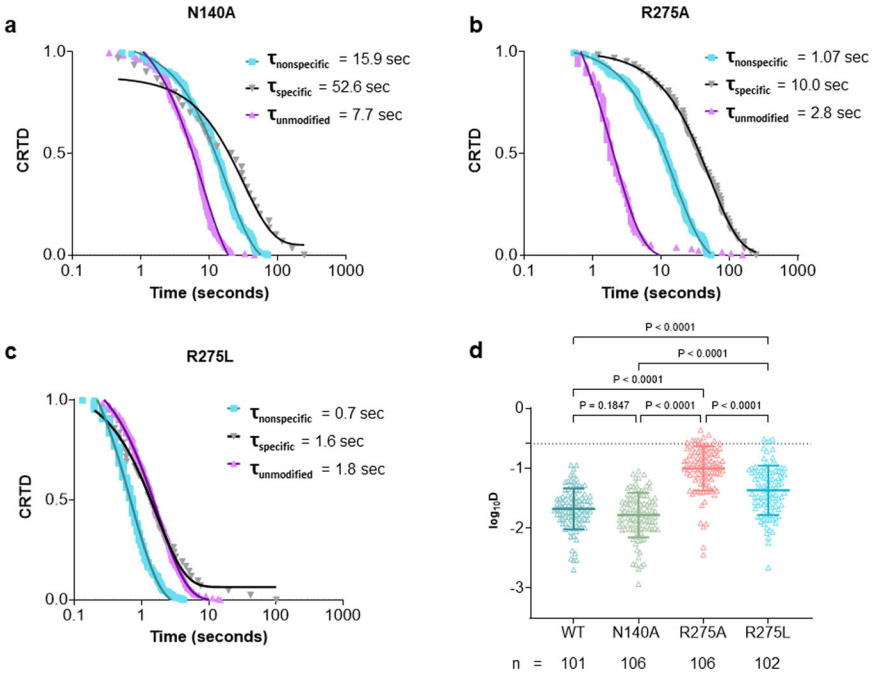

Fig. 3 | TDG catalytic variants indicate R275 is essential for base detection. Experiments conducted with unmodified or nick translated DNA containing 5fC. **a** CRTD analysis fit to a one-phase decay of N140A TDG binding to unmodified ($n = 163$), specifically to 5fC ($n = 30$) and nonspecifically ($n = 168$). **b** CRTD analysis fit to a one-phase decay of R275A TDG binding to unmodified ($n = 110$), specifically to 5fC ($n = 24$) and nonspecifically ($n = 296$). **c** CRTD analysis fit to a one-phase decay of R275L TDG binding to unmodified ($n = 269$), specifically to 5fC ($n = 42$) and nonspecifically ($n = 330$). **d** Scatter plot of the diffusion coefficient ($\log_{10}D$) calculated for each variant TDG on unmodified λ (WT $n = 101$; N140A $n = 106$; R275A $n = 106$; R275L $n = 102$). Dashed line, D_{lim}, theoretical limit to free diffusion for TDG-HaloTag. *P*-values determined by two-way ANOVA. Data represents the mean ± SD from three independent experiments.

(see Supplementary Note). Any rates that exceed this theoretical limit of diffusion would reflect hopping events. While there was an increase in the diffusion with increasing ionic strength and a widening of the distribution of diffusivities, none of the values were above the theoretical limit of diffusion to unambiguously signify hopping. Since the crystal structures (PDBs 2RBA and 4Z7B) show TDG interacts with both strands of the helix when bound to a base modification, we labeled and mixed TDG-HaloTag with two different colored labels, JF635 (red) and JF552 (green) to directly visualize hopping. Two types of behavior were observed: (1) JF635-TDG and JF552-TDG collide, in which the two orthogonally labeled TDG proteins did not bypass one another (62.5%), and (2) JF635-TDG and JF552-TDG molecules bypass one another (37.5%) (Fig. 2b). The only way it is possible to observe both of these events is for TDG to have the ability to not only slide on the DNA but also to hop.

TDG catalytic variants reveal R275 is essential for base detection

N140 is essential for activating the water molecule for catalysis to occur and R275 fills in the hole in the duplex that is generated when the base is flipped into the active site. The difference in expression of each variant compared to endogenous TDG is 72-fold for N140A, 45-fold for R275A, and 21-fold for R275L (Supplementary Fig. 2). Mutating to R275A and R275L decreases the excision of T from G:T pairs by 8-fold and 30-fold, respectively, and also reduces the binding affinity of either variant to G:T containing substrates 3-fold[13]. It is unknown if either variant affects the ability of TDG to search for DNA base modifications[3,13,25]. To investigate the role of these key amino acids on the base modification search and recognition mechanism of TDG, we performed additional single-molecule imaging with TDG-HaloTag-N140A, R275L, and R275A variants and DNA containing regions of 5fC. Consistent with the WT TDG, we observed motile and non-motile events for all three variants tested, which represent nonspecific and 5fC binding, respectively. The N140A nonspecific events (motile) bound with a lifetime of 15.9 ± 0.23 s (84.8%) and the specific events (non-motile) bound with a lifetime of 52.9 ± 1.1 s (15.2%) (Fig. 3a and Supplementary Figs. 1d and 6a, b). This shorter lifetime for specific events (compared to WT) and the absence of short-lived rebinding events, further suggesting that wild type TDG is able to cleave the 5fC moiety and rebind to the abasic site product. The R275A nonspecific events (motile) bound with a lifetime of 1.07 ± 0.02 s (92.5%) and the specific events (non-motile) bound with a lifetime of 10.0 ± 1.1 s (7.5%) (Fig. 3b and Supplementary Figs. 1e and 6c, d). The R275L nonspecific events (motile) bound with a lifetime of 0.7 ± 0.02 s (88.7%) and the specific events (non-motile) bound with a lifetime of 1.7 ± 0.06 s (11.3%) (Fig. 3c, Supplementary Figs. 1e and 6e, f). N140A, R275A, and R275L bound to unmodified DNA with lifetimes of 7.7 ± 0.24, 2.8 ± 0.13, and 1.8 ± 0.02 s, respectively (Fig. 3a–c and Table 1), with the R275A and R275L lifetimes shorter than the lifetime of TDG-HaloTag-JF635 binding to unmodified DNA. Together these data reveal that disrupting key active site residues disrupts the ability of TDG to bind DNA as well as demonstrating a reduced binding lifetime at sites of 5fC, in agreement with previous bulk biochemical studies[13,26].

Since we observed the binding lifetimes of the TDG variants to differ from wild-type TDG, we examined the diffusion of these variants by mean squared displacement (MSD) analysis. MSD analysis measures the displacement of the particle over a given time window, over the course of the event length to determine the diffusivity and α, the anomalous diffusion coefficient. N140A had a diffusion coefficient of 2.2×10^{-2} μm²/s, which was not significantly different from wild type (Fig. 3d). This indicates that N140 does not significantly affect the movement of TDG on unmodified DNA, but the decreased binding lifetime leads to less DNA being sampled before dissociation. The diffusion coefficient for R275A (1.3×10^{-1} μm²/s) and R275L (8.7×10^{-2} μm²/s) were significantly faster than both the wild type and N140A variant (Fig. 3a and Table 1), suggesting that R275 is more important for searching unmodified DNA than N140.

We can estimate the residence time per base pair by utilizing the stepping rate (see Supplementary Note). The base pair residence

time for wild type was estimated to be 2.1 ± 1.6 μs. For N140A, the base pair residence time was estimated to be 2.6 ± 2.2 μs, which is similar to the time for wild-type TDG (2.1 ± 1.6 μs). In contrast, the R275A and R275L TDG variants had shorter residence times of 0.5 ± 0.3 μs and 0.7 ± 0.5 μs, respectively (Supplementary Table 1). This may indicate that the R275A/L variants are no longer able to flip out and sample base pairs, which agrees with previous biochemical studies that indicated R275 plays a role in promoting or stabilizing the flipped nucleotide[13]. While the faster diffusion observed for the R275A and R275L TDG variants increases the amount of DNA sampled during each binding event, the rapid search may be less effective if TDG is more likely to miss the lesion or modified base, similar to behavior observed with other DNA glycosylases wedge residue variants[27].

The N-terminus of TDG facilitates its movement on DNA

Previous work identified the first 81 amino acids are not essential for the catalytic activity of TDG[28]. However, we hypothesized that the intrinsically disordered N-terminus may play a key role for TDG search of unmodified DNA (Supplementary Fig. 1a, c). To investigate the role of the TDG N-terminal region on the base modification search and recognition mechanism, we performed additional single-molecule imaging experiments with TDG-HaloTag containing a truncation of the N-terminal 81 residues (ΔN-term-TDG-HaloTag) and DNA containing regions of 5fC modification. The ΔN-term-TDG-HaloTag was expressed 50-fold higher compared to endogenous TDG (Supplementary Fig. 2). We observed similar behavior with ΔN-term-TDG-HaloTag as full-length TDG-HaloTag, where long-lived, stationary events were followed by shorter lived stationary events (Fig. 4a). Interestingly, the

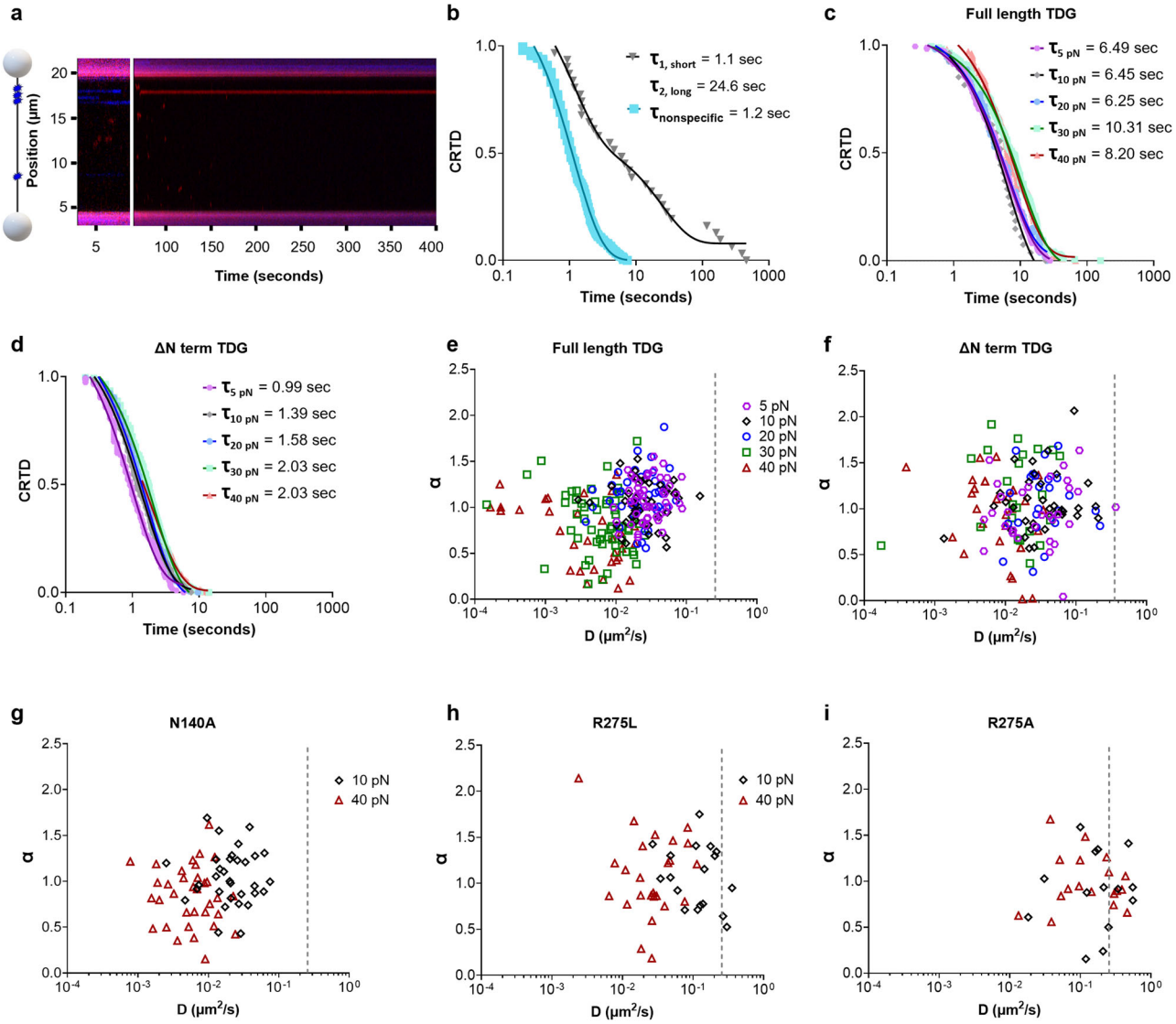

Fig. 4 | The N-terminus of TDG is significant for its movement on DNA. Tension alters diffusivity and α for ΔN-term and full-length TDG. **a** A cartoon depiction of the DNA substrate with 5fC sites shown in blue, and a representative kymograph with TDG binding shown in red. Break in kymograph to show location of fiducial markers. **b** CRTD analysis fit to a two-phase decay of TDG binding DNA containing 5fC specifically ($n = 32$). CRTD analysis fit to a one-phase decay of TDG binding nonspecifically ($n = 243$). **c** CRTD analysis fit to a one-phase decay of full-length TDG binding to unmodified DNA at 5 pN ($n = 172$), 10 pN ($n = 27$), 20 pN ($n = 53$), 30 pN ($n = 169$), and 40 pN ($n = 102$). **d** CRTD analysis fit to a one-phase decay of ΔN-term TDG binding to unmodified DNA at 5 pN ($n = 113$), 10 pN ($n = 201$), 20 pN ($n = 197$), 30 pN ($n = 293$), and 40 pN ($n = 203$). Data in (**b**–**d**) represents the mean ± SEM of the fit from 3 independent experiments. **e** Plot of diffusion coefficients (D) versus alpha (α) for full-length for 5 pN ($n = 50$), 10 pN ($n = 41$), 20 pN ($n = 41$), 30 pN ($n = 57$), and 40 pN ($n = 33$). **f** Plot of diffusion coefficients (D) versus alpha (α) for ΔN-term for 5 pN ($n = 30$), 10 pN ($n = 36$), 20 pN ($n = 24$), 30 pN ($n = 22$), and 40 pN ($n = 35$). **g** Plot of diffusion coefficients (D) versus alpha (α) for N140A for 10 pN ($n = 33$), and 40 pN ($n = 34$). **h** Plot of diffusion coefficients (D) versus alpha for 10 pN ($n = 19$) and 40 pN ($n = 27$). (α) for R275L. **i** Plot of diffusion coefficients (D) versus alpha (α) for R275A for 10 pN ($n = 14$) and 40 pN ($n = 19$). Data in (**e**–**i**) represents the mean ± SD from three independent experiments.

Table 2 | Binding lifetimes and diffusivities dependent upon tension

Protein	Tension (pN)	Lifetime(s)[a]	Diffusivity ($\mu m^2/s$)[b]	α[c]
Wild Type	5	6.49 ± 0.04	0.034 ± 0.019	1.07 ± 0.22
	10	6.45 ± 0.06	0.032 ± 0.027	1.05 ± 0.24
	20	6.25 ± 0.20	0.026 ± 0.016	1.11 ± 0.27
	30	10.31 ± 0.11	0.008 ± 0.006	0.83 ± 0.32
	40	8.20 ± 0.27	0.007 ± 0.006	0.72 ± 0.35
ΔN term	5	0.99 ± 0.02	0.049 ± 0.067	0.99 ± 0.25
	10	1.39 ± 0.02	0.050 ± 0.049	1.04 ± 0.35
	20	1.58 ± 0.02	0.047 ± 0.052	0.98 ± 0.40
	30	2.03 ± 0.02	0.020 ± 0.015	1.19 ± 0.46
	40	2.03 ± 0.02	0.012 ± 0.009	0.99 ± 0.40
N140A	10	5.9 ± 0.32	0.025 ± 0.017	1.05 ± 0.29
	40	5.52 ± 0.12	0.007 ± 0.005	0.85 ± 0.32
R275A	10	0.7 ± 0.03	0.239 ± 0.179	0.90 ± 0.43
	40	0.73 ± 0.04	0.277 ± 0.403	1.01 ± 0.29
R275L	10	1.4 ± 0.37	0.143 ± 0.095	1.07 ± 0.34
	40	0.66 ± 0.02	0.037 ± 0.031	1.05 ± 0.45

[a]The tau is the best fit value ± the standard error of the fit to the observed data.
[b]The diffusivity is the mean ± the standard error of the mean.
[c]The alpha is the mean ± the standard error of the mean.

nonspecific lifetime of ΔN-term-TDG-HaloTag was 1.2 ± 0.02 s which is three-fold shorter than full-length TDG-HaloTag (Fig. 4b). This reduced lifetime on nonspecific DNA supports the hypothesis that the N-terminus of TDG is important for searching the DNA.

A previous AFM study determined that TDG bends the DNA ~ 30° when searching for a lesion, whereas the DNA is bent ~70° when TDG is bound to the lesion[29]. We hypothesized based on the published AFM study[29] that the necessity of DNA bending by TDG may result in shorter lifetimes at higher DNA tensions because TDG cannot fully engage with the extended DNA, but our data showed this is not the case. In our prior experiments we used a tension of 10 pN. By increasing the tension from 5 pN to 40 pN with full-length TDG, we observed similar lifetimes of 6.49 ± 0.02, 6.45 ± 0.06, and 6.25 ± 0.20 s at 5, 10 and 20 pN, respectively (Fig. 4c and Table 2). At 30 pN and 40 pN, we observed lifetimes that were significantly ~two-fold higher than at lower tensions (Fig. 4c and Table 2). For ΔN-term TDG, we also observed a two-fold increase in the binding lifetimes with increasing tension. Additionally, ΔN-term TDG stays bound to the unmodified DNA with a 3–5-fold shorter lifetime than full-length at the respective tensions (Table 2). Shorter lifetimes for ΔN-term compared to full-length further supports that idea that the N-terminus is essential for prolonged interaction with nonspecific DNA. MSD analysis indicated a decrease in diffusivity with increasing DNA tension for both full-length and ΔN-term TDG, showing an ~ 5-fold decrease from 5 to 40 pN (Fig. 4e, f and Table 2). Longer lifetimes and a decrease in diffusivity at higher tensions could indicate that TDG is better at interrogating the DNA for the base modification. Furthermore, these data indicate that bending of the DNA is not essential for TDG to bind productively to DNA. This idea is further supported by an increase in the residence time per base pair (Supplementary Note and Supplementary Table 2). The anomalous diffusion exponent, α, provides further insight into the movement of a particle, with $\alpha = 1$ is random diffusion, $\alpha < 1$ sub-diffusive (constrained motion), and $\alpha > 1$ is super diffusive. For full-length WT TDG a decrease in the α was observed but we did not see a significant decrease in α for ΔN-term TDG (Fig. 4e, f and Table 2). The TDG variants, N140A, R275L, and R275A, were also investigated at 40 pN to further investigate base interrogation investigate base interrogation at the idea of base interrogation. TDG N140A had a similar behavior to WT TDG: a decrease in both alpha and diffusivity were observed with an increase in tension (Fig. 4g). At high tension, the diffusivity of TDG R275L significantly decreased compared to low tension, but no significant difference was observed for TDG R275A. (Fig. 4h, i), supporting the idea that the size of the side chain aids in flipping in the base. These data indicate that DNA tension, perhaps through increasing the distance between stacked bases (~5% increase in distance at 40 pN compared to 10 pN), provides better interaction of the N-terminal domain to increase the affinity to DNA increasing the dwell time and slowing the rate of diffusion to a mode of constrained motion.

N-terminus of TDG mediates nucleosome interactions

We sought to understand how chromatin structure impacts the TDG search mechanism by reconstituting an undamaged nucleosome using a 601 sequence[30]and ligating the nucleosome into 6.2 kb biotinylated handles using the DNA tethering kit from LUMICKS (Supplementary Fig. 7). We then performed single-imaging experiments using TDG-HaloTag and the Cy3 labeled mononucleosome-containing DNA substrate. We observed three different behaviors when TDG encountered the nucleosome: (1) TDG approaches from one side and hops over the nucleosome (bypass) (Fig. 5a), (2) TDG collides with but does not move past the nucleosome (no bypass) (Fig. 5b), and (3) TDG colocalizes with the nucleosome after the initial encounter (Fig. 5c). Further analysis revealed that TDG bypasses the nucleosome during 33.3% of encounters, did not bypass the nucleosome during 39.2% of encounters, and colocalized with the nucleosome during 27.5% of the encounters. CRTD analysis fit to a one-phase exponential indicated that TDG bound to nucleosomes with a lifetime of 16.5 ± 1.9 s (Fig. 5e), which is 2-fold longer than non-damaged DNA, and 4.5-fold shorter than binding to 5fC.

Since we observed the N-terminus is important for movement on DNA, we also sought to see if the N-terminus plays a role in binding or bypassing nucleosomes. To address this, we performed additional single-imaging experiments using TDG-HaloTag-ΔN-term and the mononucleosome-containing DNA substrate. These experiments revealed a significant decrease in the frequency of ΔN-term colocalized with the nucleosome (8.3%), an increase in the number of events that approached but did not bypass the nucleosome (61.1%), and saw a similar frequency for bypass (30.6%) (Fig. 5e). This decrease in TDG colocalization with the nucleosome indicates that the N-terminus plays an important role in nucleosome binding.

Discussion

In this study, we utilized single-molecule methods to determine how TDG searches for DNA damage as well as uncovering how key active site residues, domains, and the presence of nucleosomes alter this search process. Similar to other DNA glycosylases, TDG forms moderately stable interactions with unmodified DNA to optimize its search[31,32]. Despite the ongoing 1D search, in our single-molecule regime TDG primarily binds its modified sites directly out of solution in a 3D diffusion mechanism, with a limited number of events that bind unmodified DNA first and then slide into the base modification. Of note, sliding interactions of less than 100 base pairs are not discernable on our system so these events that appear as 3D may have some component of 1D diffusion not observed, as was observed for human uracil DNA glycosylase hUNG2[33,34]. Furthermore, we found that TDG bound specific lesions sites 10-fold longer than unmodified DNA. Since TDG is able to bind a wide array of modified bases, including 5caC and G:T mismatch, it is plausible that a similar behavior would be observed: TDG will bind to the modified base and become nonmotile, with rebinding to the generate abasic site. While it is currently not known how long TDG will bind to different base modifications, it is likely that TDG will bind to a specific base modification for a time that is directly correlated to the biochemically determined cleavage rate followed by subsequent rebinding to an abasic site. Examining TDG under different

https://doi.org/10.1038/s41467-024-53497-7

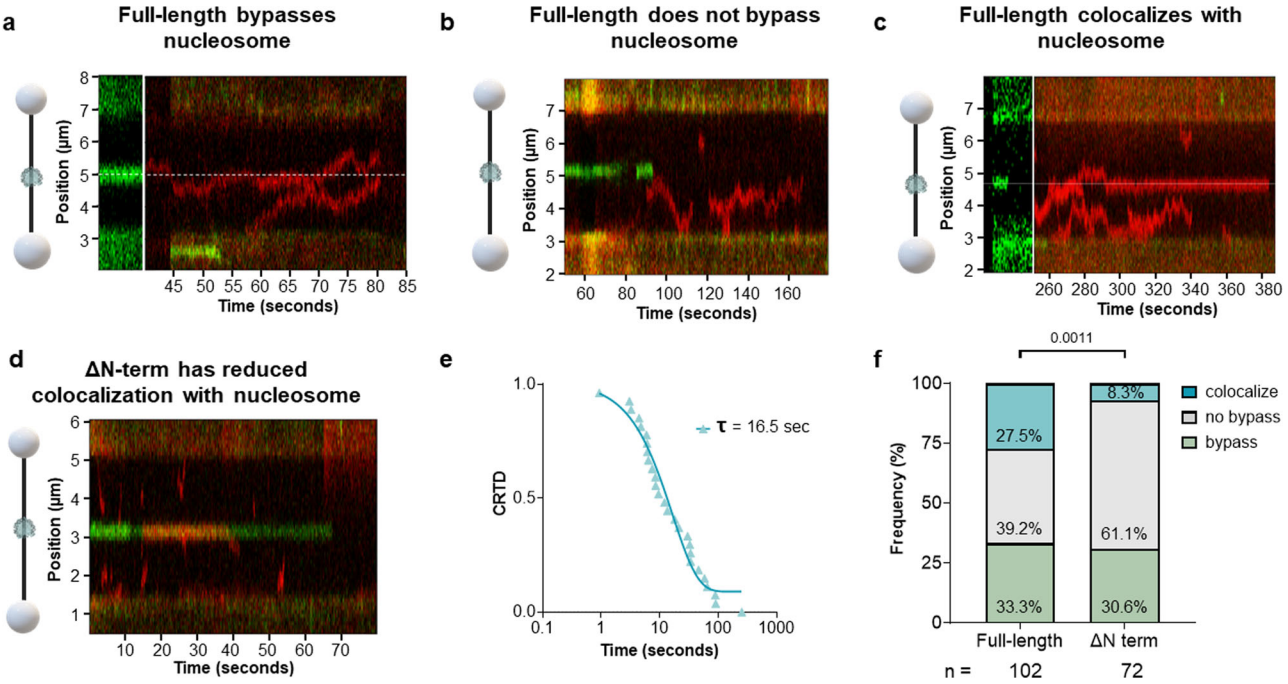

Fig. 5 | N-terminus of TDG is important for interacting with nucleosomes. A cartoon depiction of the DNA substrate with Cy3 labeled nucleosome. Representative kymograph with full-length TDG (shown in red) (**a**) bypassing, (**b**) does not bypass, or (**c**) colocalizes with the nucleosome shown in green. (**d**) Representative kymograph of ΔN-term TDG (red) colocalizing with nucleosome (green). **e** CRTD analysis fit to a one-phase decay of TDG interacting with the NCP ($n = 28$). **f** Stacked bar graph showing the fraction of bypass (green), no bypass (gray), and colocalized (teal) events for full-length and ΔN-term ($p = 0.0011$ by X^2).

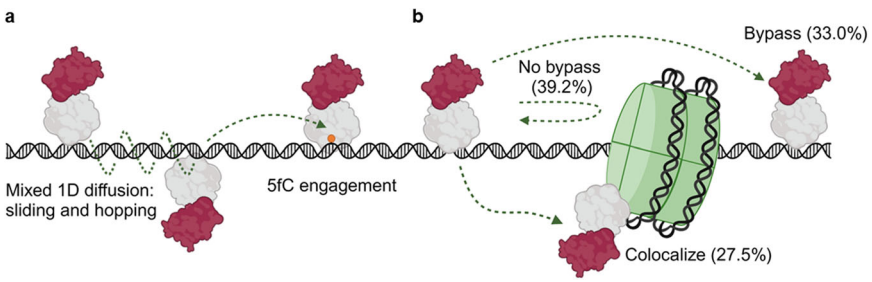

Fig. 6 | Working model. TDG uses multiple modes of linear diffusion to efficiently search for 5fC. **a** WT TDG (gray) fused to HaloTag (red) scans dsDNA with sliding and hopping interactions, with a lifetime of 7.9 ± 0.06 s and diffusivity of 0.028 ± 0.02 μm²/s. Upon 5fC engagement (orange), TDG binds with a lifetime of 72.9 ± 5.4 s. These interactions are dependent on active site residues N140, R275, and the N-terminal domain. **b** When TDG encounters a nucleosome core particle, it can bypass, not bypass, or colocalize with the NCP (16.5 ± 1.9 s). Importantly the N-terminal domain of TDG stabilizes the interaction with the NCP, resulting in a four-fold increase in colocalization frequency. Created in BioRender. Schaich, M. (2024) BioRender.com/y41i875.

ionic conditions, as well as two color experiment, demonstrated that TDG is able to both slide and hop along the DNA, which appears to be common amongst bacterial and mammalian glycosylases[33,35,36]. We estimated TDG's search length on λ DNA utilizing the observed diffusion constant in $\frac{bp^2}{sec}$ and the nonspecific lifetime of 7.9 sec (See Supplementary Note) to be 5225 bp or ~10% of λ DNA per encounter.

The complex search mechanism used by TDG has several potential benefits during interrogation of the genome for DNA base modifications. Due to TDG's high affinity for nondamaged DNA, the ability to switch between sliding and hopping may also be advantageous in situations where TDG encounters other factors on the DNA, including transcription factors or nucleosomes (Fig. 6). With a sliding-exclusive method, full dissociation would need to occur to sample the DNA after the block. However, hopping would allow TDG to circumvent the block and continue scanning on the other side of an obstacle. This use of dual linear diffusion modes leads to the question of whether certain cellular conditions favor TDG hopping versus sliding on DNA. For example, if TDG has a sequence preference, it may adopt a conformation at these preferred sequences where tight binding leads to direct sliding as TDG further investigates that region of DNA. While it is currently unclear whether TDG has sequence preference, it is possible that sequences enriched in CG dinucleotides could result in initiating sliding behaviors. This type of increased binding to particular sequences would be highly advantages in regions with a heavier burden of DNA modification either due to (1) an enrichment of methylated CG dinucleotides or (2) TET enzymes localizing to methylated CpG rich islands to oxidize 5-methlyC groups for their removal by TDG and subsequent transcriptional activation[37]. It has been previously determined the glycosylase OGG1 slides and hops along the DNA in search of a lesion[31,35], whereas hUNG2, a uracil DNA glycosylase family member, performs 3D diffusion, short-range sliding, and hopping to find a lesion[33]. Thus, TDG exhibiting multiple modes of diffusion is in agreement with the behavior observed for other mammalian glycosylases and appears to be essential to efficiently scan the DNA and the ability to switch between modes may be dependent upon sequence or other cellular factors (Fig. 6a).

Our single molecule work also revealed the active site variants N140A, R275A, and R275L change the behavior of TDG on DNA. N140A had a shorter dwell time on 5fC but did not exhibit a faster diffusion than wild type. This contrasts other previous studies with the glycosylase OGG1, where the catalytically dead glycosylase bound much longer than that of the WT[19,38]. On the other hand, R275A/L variants both exhibited a shorter dwell time as well as a faster diffusion. One explanation for this behavior is that R275 probes the DNA sequence during the sliding, enabling efficient detection of DNA damage but also slowing the speed of sliding interactions. Further, it is possible that the lack of the positive charge causes TDG to have lower binding affinity and thus shorter dwell times and faster rates of diffusion. A previous study on bacterial glycosylases Fpg, Nei, and Nth has also shown that if a residue that interrogates the DNA is changed to alanine, faster diffusivity was observed, indicating that efficient recognition of a base is dependent on a single active site residue[27]. Together, these findings provide an in-depth description of how key catalytic residues of TDG contribute to its search for DNA damage and provides insight into how its mixed mode of hopping and sliding searches contribute to efficient damage detection and repair.

Deleting the disordered N-terminus of TDG, consisting of the first 81 amino acids, did not reduce the ability to bind to 5fC. However, the ability of ΔN-term TDG to search nonspecific DNA was significantly reduced, due to the three-fold shorter lifetime of ΔN-term TDG bound to DNA nonspecifically. While the tension on DNA at any given point in the cell is not well known, it is known that nucleosomes begin to unwrap at 5 pN[39], hairpins begin to pull apart 10–15 pN[40,41], G-quadruplexes begin to unfold 15-20 pN[42], RNA polymerase has been shown to exert a force of 25-30 pN on DNA[43] and double-strand DNA begins to unwind at ~50 pN[44,45]. With increasing tension, the λ DNA may begin to have regions that may potentially allow for TDG to interrogate the DNA more easily perhaps due to widening of the base steps and less stacking interactions, allowing better interrogation of the DNA at higher tensions. The observed decrease in α, indicating more constrained motion on DNA, for full-length but not ΔN-term TDG suggests that the disordered N-terminus may play a role in allowing TDG to thoroughly interrogate the DNA for a base modification. The N-terminus may also play a role in TDG switching between hopping and sliding. If TDG needs to form for tighter interactions with the DNA in order to slide, the N-terminus, which contains several positively charged amino acid residues, may adopt a confirmation that limits hoping to better probe an area. Similarly, a study has shown that the N-terminus of hUNG2 has an increase in contact with DNA during molecular crowding which results in an efficient sliding and prevents hUNG2 from dissociating and diffusing into bulk solution[46].

We have shown that not only is TDG able to hop over a nucleosome, but it also interacts with a nucleosome with an apparent dwell time of 16.5 ± 1.9 s. Deleting the N-terminus significantly reduces the binding frequency to nucleosomes (Fig. 6b). This interaction of TDG with nucleosomes may act as a scaffold to help facilitate removal of oxidized 5-methylcytosine in heterochromatic regions. Alternatively, perhaps TDG associates with nucleosomes prior to the arrival of TET enzymes, and thus would increase the efficiency of oxidative demethylation by already being in the correct location prior to the initial 5mC oxidation step in heterochromatic regions. Efficient active oxidative demethylation is vital for life and of interest for potential therapeutic applications[47], and this work gives key mechanistic insight on this process. TDG interacting with a nucleosome could also alter nucleosome dynamics and may interfere with fork progression and DNA replication, and thus might explain its degradation during S-phase in mammalian cells.

Taken together, our studies offer new insight into how key active site residues and domains of TDG contribute to its efficient search for DNA damage, and how chromatin can act either as a roadblock or a scaffold during the search process. Of note, these studies with nucleosome core particles represent a unique look at how base excision repair enzyme interact with nucleosomes that do not have free DNA ends present. Future studies will need to be completed to determine the impact nucleosome unwrapping, chromatin remodelers, and how post translational modifications of TDG will impact its ability to efficiently search for its substrates.

Methods

DNA substrate preparation

Nick translation for confocal imaging. Lambda DNA was biotinylated as described previously[19]. Briefly, DNA was purchased from New England Biotechnologies and treated with Klenow fragment polymerase (NEB) and a dNTP mix containing biotinylated dCTP. This results in one side of the DNA containing four biotins and the other containing six biotins.

Next, 1 μg of the biotinylated lambda DNA is treated with Nt.BspQI (NEB) for 1 h at 50 °C then heat inactivated for 20 min at 80 °C, which will generate single strand breaks at 10 different sites along the DNA by cutting 3′ of the recognition site 5′-GCTCTTCN-3′. To incorporate 5fC at the 10 sites, 800 ng nicked DNA was incubated with a dNTP mix containing d5fCTP, fluorescein-labeled dUTP, dATP, and dGTP were incubated and 10 units of DNA polymerase 1 (NEB) for 6 min at 37 °C. Reaction was inactivated by heating at 75 °C for 20 min and allowed to slowly cool to room temperature. To seal the nicks, a reaction containing 800 ng of the nick-translated DNA, 8 mM ATP and 10 units of T4 ligase (NEB) were incubated overnight at 16 °C. Reaction was heat inactivated for 10 min at 65 °C and allowed to cool to room temperature slowly. Of the 10 d5fCTP sites, only eight d5fCTP sites can be resolved as two of the sites are within 400 bp of each other and one site is near the end of λ DNA and thus too close to the bead (Supplementary Fig. 4a–c).

Arms ligation for confocal imaging. Using the LUMICKS DNA tethering kit and protocol, a sequence of interest is ligated into two handles that are each 6.3 kb in length to generate a substrate that is 12.6 kb in total. For the defined 5fC, the following oligonucleotide sequences were ordered from Trilink and IDT, respectively:

Top strand (5fC): 5′-phosphate-*CAAC* ACC AGT CCA TCG CTC A**5fC**G TAC AGA GCT G–3′

Bottom strand: 5′- phosphate – *ACCA* CAG CTC TGT ACG TGA GCG ATG GAC TGG T-3′

Annealing reactions were heated at 95 °C for 5 min in buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM KCl, and the heat block was turned off and reactions then cooled slowly to room temperature. In a 10 μL reaction, 25 nM of the duplex, 2.5 μL of each handle, 1 μL of 10X ligase buffer, and 0.5 μL of ligase were mixed and incubated at 16 °C overnight. Reaction was heat inactivated at 65 °C for 10 min and allowed to slowly cool to room temperature. Substrate was stored at 4 °C and protected from light.

Undamaged nucleosomes were reconstituted on a 601 sequence as described previously[30]. Briefly, human histone H3 C96S C110A, H2A K119C, H2B, and H4 were ordered from the Histone Source at Colorado State University. H2A K119C and H2B were incubated in 2 mg/mL guanidinium buffer at room temperature for 2 h. Equimolar H2A and H2B were mixed and dialyzed against high salt refolding buffer a total of 3 times, and at least 8 h for each exchange at 4 °C. The same process was repeated for H3 C96S C110A and H4 to refold the tetramer. H2A K119C/H2B dimer and the H3 C96S C110A/H4 tetramer were purified over Superdex 200 column. The H2A K119C/H2B dimer was incubated with 0.7 mM TCEP for 20 min at 4 °C, then Cy3-maleimide dye was added to the H2A K119C in a 2:1 molar ratio and incubated at room temperature for 1 h while rocking. Reaction was quenched with 10 mM DTT, the dimer was purified over a Superdex S200 column, and frozen in 50% glycerol. After confirming the stoichiometry with SDS gel, equal volume of 100% glycerol was added to

store the H2A K119C/H2B dimer and the H3 C96S C110A/H4 tetramer at −20 °C.

To reconstitute the nucleosome on DNA, the following ultramer sequences were ordered from IDT:

Top Strand: 5′-phosphate-*CAAC* TGA GAC CAT GTA CCC AGT TCG AAT CGG ATG TAT ATA TCT GAC ACG TGC CTG GAG ACT AGG GAG TAA TCC CCT TGG CGG TTA AAA CGC GGG GGA CAG CGC GTA CGT GCG TTT AAG CGG TGC TAG AGC TGT CTA CGA CCA ATT GAG CGG CCT CGG CAC CGG GAT TCT CGA TAA CTC AGC AAT AGT GGG TCT CA – 3′

Bottom strand: 5′- phosphate -*ACCA* TGA GAC CCA CTA TTG CTG AGT TAT CGA GAA TCC CGG TGC CGA GGC CGC TCA ATT GGT CGT AGA CAG CTC TAG CAC CGC TTA AAC GCA CGT ACG CGC TGT CCC CCG CGT TTT AAC CGC CAA GGG GAT TAC TCC CTA GTC TCC AGG CAC GTG TCA GAT ATA TAC ATC CGA TTC GAA CTG GGT ACA TGG TCT CA – 3′

The annealed DNA, H2A K119C/H2B dimer, and the H3 C96S C110A/H4 tetramer are mixed in a 1:2:1 molar ratio and equilibrate in dialysis tubing against high salt buffer for 30 min. To remove the high salt buffer, a series of dialysis steps occur to transition from 1.5 M NaCl to 0.125 M NaCl. The reconstituted nucleosome is concentrated, and heat shocked at 55 °C for 30 min, prior to spinning over a 10-40% sucrose gradient for 40 h at 125,000 xG at 4 °C. Fractions containing reconstituted nucleosomes are combined, buffered exchanged into TE buffer, and concentrated to ~1 μM and stored at 4 °C. To ligate into the LUMICKS handles, 25 nM of the reconstitute nucleosome, 2.5 μL of each handle, 1 μL of 10X ligase buffer, and 0.5 μL of ligase were mixed and incubated at 16 °C overnight. Substrate was stored at 4 °C and protected from light.

Plasmid constructs

Wild type TDG-HaloTag and TDG-HaloTag variants were expressed from plasmids constructed utilizing Gene Universal Inc with the pHTN HaloTag CMV-neo vector (Promega). (Supplementary Fig. 1a–e)

Nuclear extract

Nuclear extracts were prepared as described previously[19]. Briefly, U2OS cells were cultured in 5% CO2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/L glucose, 10% fetal bovine serum (Gibco), and 5% penicillin/streptavidin (Life Technologies). To obtain transient overexpression of TDG-HaloTag, cells were transfected with 4 μg of plasmid per 4 million cells using the lipofectamine 2000 reagent and protocol (Thermo Fisher Cat# 11668019).

Nuclear extraction was performed 24 h after transfection utilizing nuclear extraction kit from Abcam (ab113474). Following the protocol from the kit, 10 μL single-use aliquots were made, flash-frozen and, stored at −80 °C. Extracts were diluted in buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 0.1 mg/ml BSA, 1 mM DTT, 1 mM Trolox, and 5% glycerol.

Nuclear extract quantification

SDS gel. A total of 2.5 μg of each nuclear extract was loaded in duplicate on a 4-12% Bis Tris gels (Invitrogen NP0323BOX) with a dye free sample buffer (4X buffer: 40% glycerol, 200 mM Tris-Cl (pH 6.8), and 8% SDS). The gel was imaged using a laser scanner for Cy5 (Typhoon, Amersham). Gels were then stained with Coomassie blue and imaged. The total intensity for each lane was measured and the lanes were averaged together in order to normalize for loading. Finally, each of the variant extracts were normalized to the wild type TDG in order to look at expression levels and the level of free dye. (Supplementary Fig. 2a–c)

Western blotting. Different amounts of purified proteins and nuclear extracts were denatured at 95 °C for 10 min. Equal volumes were loaded on a 4-20% tris-glycine polyacrylamide gel (Invitrogen XP04202BOX). Proteins were transferred onto a polyvinylidene difluoride membrane and blocked in 20% nonfat dry milk in PBST (1x phosphate-buffered saline containing 0.1% Tween 20) for 1 hr at room temperature. Membranes were probed with TDG primary antibody (1:500 Sigma HPA052263) overnight at 4 °C. Membranes were washed 3 times for 14 min in PSBT and incubated with peroxidase-conjugated rabbit secondary antibody for 1 h at room temperature. Membranes were washed again 3 times for 15 min before developing using Super-Signal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific #34095). (Supplementary Fig. 2d),

Confocal Imaging in the LUMICKS C-trap

Imaging. Fluorophores and dyes utilized were excited with the laser closest to their maximum intensity. Fluorescein was excited with a 488 nm laser and emission collected with a 500–550 nm band pass filter, Cy3 and HaloTag-JF552 were excited with a 561 nm laser and the emission was collected with a 575–625 nm band pass filter, and HaloTag-JF635 was excited with a 638 nm laser and emission collected with a 650– 750 nm band pass filter. All data were collected with a 1.2 NA 60x water emersion objective and photons measured with single-photon avalanche photodiode detectors. The laser power was set to 5% and kymographs were collected continuously with 0.1 msec exposure for each pixel of size 100 nm, which resulted in approximately 30 frames/sec for λ DNA. All data was collected with a minimum of two different preparations of nuclear extracts across multiple days.

Single-molecule experiments were then performed with a LUMICKS C-trap using a 5-channel flow chamber. For experiments using λ or nick-translated DNA, ~4 μm streptavidin-coated beads and experiments using the LUMICKS handles, ~1.7 μm streptavidin-coated beads were used. The beads were captured using the dual optical tweezers and then transferred to a second chamber containing the biotinylated DNA and the biotinylated DNA molecule tethered between the streptavidin coated beads. Once the DNA molecule is tethered, it is transferred to a final chamber containing a 1:10 dilution of the nuclear extract with TDG-HaloTag-JF635. To determine the concentration of the protein in the flow cell, background photon intensities were determined to generate the standard curve of purified HaloTag labeled with JF635 or JF552. The background intensities of the protein from the nuclear extract were measured and the curve was utilized to determine protein concertation (Supplementary Fig. 3a, b).

Analysis. Kymographs were analyzed with custom software from LUMICKS, Pylake. The kymographs were exported as.h5 files from the C-trap and viewed using the utility "C-Trap.h5 Visualization GUI (2020)" by John Watters (harbor.lumicks.com). Line tracking was performed using a custom script from LUMICKS based on a Gaussian fit over the line to determine its positions[48]. Only events that associated and dissociated during the course of a 5-to-10-minute kymograph were tracked. Events are plotted on cumulative residence time distribution (CRTD) plot in order to determine the binding lifetime. Specific binding event curves are fit with a two-phase exponential decay, for catalytically active WT protein. One-phase exponential decays were sued for the variants, which have little or no catalytic activity, and nonspecific binding events for the variants and WT. The two-phase fit for WT represents a second rebinding event to the catalytic product abasic site with a shorter dwell time. The goodness of the fit for the CRTD curves range from $R^2 = 0.96 - 0.99$. Mean squared displacement (MSD) analysis was performed using a custom script from LUMICKS based on the following equation:

$$MSD\,(n\Delta t) = \frac{1}{N-n}\sum_{i=1}^{N-n}(x_{i+n} - x_i)^2 \quad (1)$$

Where N is the total number of frames in the phase, n is the number of frames at a given time step, Δt is the time increment of one frame, and x_i is the particle position in the ith frame[27]. The diffusion coefficient (D)

was determined by fitting a linear model of one-dimensional diffusion to the MSD plots:

$$MSD\,(n\Delta t) = 2D(n\Delta t)^{\alpha} + y \quad (2)$$

where α is the anomalous diffusion coefficient and y is a constant (y-intercept). Fittings resulting in $R^2 < 0.8$ or using <10% of the MSD plot were not considered.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data are available from the corresponding author upon reasonable request. Source data are provided with this paper. Large kymograph data are available from the corresponding author upon request. Source data are provided with this paper.

Code availability

Python scripts used to calculate lifetime and MSD analysis have been deposited at https://github.com/schaichm.

References

1. Whitaker, A. M., Schaich, M. A., Smith, M. R., Flynn, T. S. & Freudenthal, B. D. Base excision repair of oxidative DNA damage: from mechanism to disease. *Front. Biosci. (Landmark Ed.)* **22**, 1493–1522 (2017).
2. Maiti, A. & Drohat, A. C. 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 (2011).
3. Maiti, A., Michelson, A. Z., Armwood, C. J., Lee, J. K. & Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **135**, 15813–15822 (2013).
4. Pidugu, L. S., Dai, Q., Malik, S. S., Pozharski, E. & Drohat, A. C. Excision of 5-carboxylcytosine by thymine DNA glycosylase. *J. Am. Chem. Soc.* **141**, 18851–18861 (2019).
5. Pidugu, L. S. et al. Structural basis for excision of 5-formylcytosine by thymine DNA glycosylase. *Biochemistry* **55**, 6205–6208 (2016).
6. Cortázar, D. et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
7. Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
8. Bennett, M. T. et al. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128**, 12510–12519 (2006).
9. Morgan, M. T., Bennett, M. T. & Drohat, A. C. Excision of 5-halogenated uracils by human thymine DNA glycosylase. Robust activity for DNA contexts other than CpG. *J. Biol. Chem.* **282**, 27578–27586 (2007).
10. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
11. Kellinger, M. W. et al. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
12. DeNizio, J. E. et al. TET-TDG active DNA demethylation at CpG and non-CpG sites. *J. Mol. Biol.* **433**, 166877 (2021).
13. Maiti, A., Morgan, M. T. & Drohat, A. C. 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 (2009).
14. Morgan, M. T., Maiti, A., Fitzgerald, M. E. & Drohat, A. C. Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. *Nucleic Acids Res.* **39**, 2319–2329 (2011).
15. Maiti, A., Morgan, M. T., Pozharski, E. & Drohat, A. C. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc. Natl Acad. Sci. USA* **105**, 8890–8895 (2008).
16. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. & Cedar, H. The role of DNA methylation in setting up chromatin structure during development. *Nat. Genet.* **34**, 187–192 (2003).
17. Tarantino, M. E., Dow, B. J., Drohat, A. C. & Delaney, S. Nucleosomes and the three glycosylases: high, medium, and low levels of excision by the uracil DNA glycosylase superfamily. *DNA Repair* **72**, 56–63 (2018).
18. Deckard, C. E. & Sczepanski, J. T. Reversible chromatin condensation by the DNA repair and demethylation factor thymine DNA glycosylase. *Nucleic Acids Res.* **49**, 2450–2459 (2021).
19. Schaich, M. A. et al. Single-molecule analysis of DNA-binding proteins from nuclear extracts (SMADNE). *Nucleic Acids Res.* **51**, e39 (2023).
20. Bigman, L. S., Greenblatt, H. M. & Levy, Y. What are the molecular requirements for protein sliding along DNA? *J. Phys. Chem. B* **125**, 3119–3131 (2021).
21. Bigman, L. S. & Levy, Y. Protein diffusion along protein and DNA lattices: role of electrostatics and disordered regions. *Annu. Rev. Biophys.* **52**, 463–486 (2023).
22. Bagchi, B., Blainey, P. C. & Xie, X. S. Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B* **112**, 6282–6284 (2008).
23. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
24. Tafvizi, A., Huang, F., Fersht, A. R., Mirny, L. A. & van Oijen, A. M. A single-molecule characterization of p53 search on DNA. *Proc. Natl Acad. Sci. USA* **108**, 563–568 (2011).
25. Dow, B. J., Malik, S. S. & Drohat, A. C. Defining the role of nucleotide flipping in enzyme specificity using (19)F NMR. *J. Am. Chem. Soc.* **141**, 4952–4962 (2019).
26. Coey, C. T. & Drohat, A. C. Defining the impact of sumoylation on substrate binding and catalysis by thymine DNA glycosylase. *Nucleic Acids Res.* **46**, 5159–5170 (2018).
27. Nelson, S. R., Dunn, A. R., Kathe, S. D., Warshaw, D. M. & Wallace, S. S. Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl Acad. Sci. USA* **111**, E2091–E2099 (2014).
28. Coey, C. T. et al. Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* **44**, 10248–10258 (2016).
29. Buechner, C. N., Maiti, A., Drohat, A. C. & Tessmer, I. Lesion search and recognition by thymine DNA glycosylase revealed by single molecule imaging. *Nucleic Acids Res.* **43**, 2716–2729 (2015).
30. Ryan, B. J., Weaver, T. M., Spencer, J. J. & Freudenthal, B. D. Generation of recombinant nucleosomes containing site-specific DNA damage. *Methods Mol. Biol.* **2701**, 55–76 (2023).
31. Rowland, M. M., Schonhoft, J. D., McKibbin, P. L., David, S. S. & Stivers, J. T. Microscopic mechanism of DNA damage searching by hOGG1. *Nucleic Acids Res.* **42**, 9295–9303 (2014).
32. Dunn, A. R., Kad, N. M., Nelson, S. R., Warshaw, D. M. & Wallace, S. S. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.* **39**, 7487–7498 (2011).
33. Porecha, R. H. & Stivers, J. T. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl Acad. Sci. USA* **105**, 10791–10796 (2008).
34. Schonhoft, J. D. & Stivers, J. T. Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.* **8**, 205–210 (2012).

35. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA* **103**, 5752–5757 (2006).
36. Friedman, J. I. & Stivers, J. T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957–4967 (2010).
37. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep.* **13**, 28–35 (2011).
38. Schaich, M. A., Weaver, T. M., Roginskaya, V., Freudenthal, B. D. & Van Houten, B. Single-molecule analysis of purified proteins and nuclear extracts: Insights from 8-oxoguanine glycosylase 1. *DNA Repair* **134**, 103625 (2024).
39. Díaz-Celis, C. et al. Assignment of structural transitions during mechanical unwrapping of nucleosomes and their disassembly products. *Proc. Natl Acad. Sci. USA* **119**, e2206513119 (2022).
40. Bercy, M. & Bockelmann, U. Hairpins under tension: RNA versus DNA. *Nucleic Acids Res.* **43**, 9928–9936 (2015).
41. Liu, Y., Galior, K., Ma, V. P. & Salaita, K. Molecular tension probes for imaging forces at the cell surface. *Acc. Chem. Res* **50**, 2915–2924 (2017).
42. Mitra, J. et al. Extreme mechanical diversity of human telomeric DNA revealed by fluorescence-force spectroscopy. *Proc. Natl Acad. Sci. USA* **116**, 8350–8359 (2019).
43. Wang, H. Y., Elston, T., Mogilner, A. & Oster, G. Force generation in RNA polymerase. *Biophys. J.* **74**, 1186–1202 (1998).
44. Rouzina, I. & Bloomfield, V. A. Force-induced melting of the DNA double helix 1. Thermodynamic analysis. *Biophys. J.* **80**, 882–893 (2001).
45. Williams, M. C., Rouzina, I. & Bloomfield, V. A. Thermodynamics of DNA interactions from single molecule stretching experiments. *Acc. Chem. Res.* **35**, 159–166 (2002).
46. Rodriguez, G. et al. Disordered N-terminal domain of human uracil DNA glycosylase (hUNG2) enhances DNA translocation. *ACS Chem. Biol.* **12**, 2260–2263 (2017).
47. Nuñez, J. K. et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503–2519.e17 (2021).
48. Mangeol, P., Prevo, B. & Peterman, E. J. KymographClear and KymographDirect: two tools for the automated quantitative analysis of molecular and cellular dynamics using kymographs. *Mol. Biol. Cell* **27**, 1948–1957 (2016).

Acknowledgements

We greatly appreciate the discussions with the lab members during the course of these experiments. This work was supported by NIH R35ES031638 (BVH), T32GM088119 (B.L.S.), the Hillman Postdoctoral Fellowship for Innovative Cancer Research, and F32ES034982 (M.A.S.), 2P30CA047904 to the UPMC Hillman Cancer Center, the major equipment grant S10OD032158-01A1 (B.V.H.), R35-GM136225 (A.C.D.), R35GM128562 (B.D.F.), and F32GM140718 (T.M.W.).

Author contributions

B.L.S., B.V.H., B.D.F., and A.C.D. conceived the research. B.L.S. performed all C-trap and protein concentration determination experiments. L.P.L. performed maxi preps and sequencing of plasmids. V.R. performed all western blot experiments. B.L.S. analyzed all single-molecule data. T.M.W. reconstituted nucleosomes. B.L.S., M.A.S., and B.V.H. drafted the paper, which was reviewed, discussed, and edited by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53497-7.

Correspondence and requests for materials should be addressed to Bennett Van Houten.

Peer review information *Nature Communications* thanks Mark Glover and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.