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

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], and Bennett Van Houten[1,2,3]

[1]Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburg, PA 15260, USA
[2]UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA 15213, USA
[3]Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA
[4]Department of Biochemistry and Molecular Biology, Department of Cancer Biology, University of Kansas Medical Center, Kansas City, KS 66160, USA
[5]Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.

Corresponding author
Bennett Van Houten, PhD
5117 Centre Ave
Hillman Cancer Center
Research Pavilion, Suite 2.6
Pittsburgh, PA 15213
Phone: 412-623-7762
Fax: 412-623-7761
Email: vanhoutenb@upmc.edu

Abstract

Base excision repair is the main pathway involved in active DNA demethylation. 5-formylctyosine 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 studied TDG in the presence and absence of 5-formylctyosine. TDG exhibits multiple modes of linear diffusion, including hopping and sliding, in search of a lesion. We probed TDG active site variants and truncated N-terminus revealing how these variants alter the lesion search and recognition mechanism of TDG. On DNA containing an undamaged nucleosome, TDG was found to either bypass, colocalize with, or encounter but not bypass the nucleosome. However, truncating the N-terminus reduced the number of interactions with the nucleosome. Our findings provide unprecedented mechanistic insights into how TDG searches for DNA lesions in chromatin.

45 **Introduction**

46 DNA can be damaged by a variety of endogenous and exogenous sources and these
47 potentially mutagenic bases are removed via base excision repair (BER)[1]. In humans, BER is
48 initiated by one of eleven damage specific DNA glycosylases that recognize a lesion and cleave
49 the glycosidic bond to leave an abasic site. The repair is completed by the actions of several
50 proteins that restore the correct base pair. Thymine DNA glycosylase (TDG) is a monofunctional
51 glycosylase that initiates BER by recognizing a specific repertoire of lesions, including 5-
52 methylcytosine (5mC) that has been deaminated to form a G:T mismatch or 5mC that has been
53 oxidized by the ten eleven translocation (TET) family of enzymes into 5-formylcytosine (5fC) and
54 5-carboxylcytosine (5caC)[2-5]. Interestingly, TDG is the only known mammalian DNA glycosylase
55 that creates an embryonic lethal phenotype in mice upon depletion: presumably due to the
56 significant role that TDG plays in active demethylation, in addition to the action on
57 environmental DNA damage[6,7].

58 While TDG is known to preferentially remove T from G:T mismatches in the context of a
59 CpG, other substrates do not show this sequence context dependence [8,9]. One such lesion is
60 5fC, which has an abundance of $6x10^4$ per human genome, and the removal of 5fC was similar
61 in the context of a CpG and outside a CpG[10-12]. Previous studies have shown that N140 of TDG
62 is essential for the catalytic activity on 5fC because of interactions with the water nucleophile.
63 The N140A variant greatly reduces the catalytic activity 16,000-fold[5]. It has been shown that
64 R275 plays an essential role in catalysis by filling in the void in the DNA that is generated once
65 the substrate nucleobase is flipped into the active site[13]. In addition to TDG specificity for G:T
66 mismatches and 5fC, TDG is able to bind to DNA nonspecifically with an affinity of 293 ± 64 nM
67 and has a footprint of 10 base pairs[14,15]. While the roles for these key active site amino acids in
68 catalysis is well-established, their role in mediating TDG searching for DNA lesions in the
69 presence of excess unmodified DNA is unknown.

70 A major unexplored consideration for any glycosylase search process in a cell is the
71 presence of chromatin. DNA packaged into nucleosomes would interrupt a DNA sliding
72 mechanism and inhibit the accessibility of specific sites in the absence of nucleosome sliding
73 and/or remodeling. Among DNA glycosylases this phenomenon is particularly relevant to TDG,
74 as the methylated DNA where its substrates are formed are associated with higher degree of
75 chromatinization[16]. TDG has also been shown to directly interact with nucleosome core
76 particles, however kinetic studies also demonstrated TDG activity was only reduced ~two-fold
77 for uracil substrates present on a nucleosome core particle[17,18].
78 Here, we utilized single-molecule analysis of DNA-binding proteins from nuclear extracts
79 (SMADNE)[19] to directly observe the search process of fluorescently labeled TDG in real-time on
80 DNA containing 5fC substrates or nucleosome core particles. By observing TDG scanning the
81 DNA as it searches for 5fC, we determined that TDG searches for damaged sites using a
82 combination of sliding and hopping and that TDG diffusivity was highly dependent on DNA
83 tension. This combined search mechanism was also observed on DNA containing NCPs. In
84 some instances, TDG hopped over nucleosomes to continue searching the DNA, but in other
85 cases it approached the NCP but did not bypass. We have also found that specific amino acid
86 side chains that are essential for catalytic activity also impact the dwell time and the ability of
87 TDG to diffuse on DNA. Additionally, although ΔN-term TDG (residues 82-410) is catalytically
88 active, we demonstrate that the N-terminus plays a significant role in binding and searching
89 unmodified DNA as well as playing a key role in engaging undamaged NCPs.

90 **Results**

91 *TDG binds DNA specifically and nonspecifically.*

92 TDG-HaloTag was overexpressed in U2OS cells, and the nuclear extracts were
93 quantified using western blots and SDS gels to determine a 60-fold difference in overexpressed
94 fluorescently tagged WT TDG to endogenous TDG (Fig. S2). To determine the binding behavior
95 of TDG on DNA containing a lesion of interest, we initially performed the single-molecule

96 imaging with TDG-HaloTag-JF635 and DNA containing 5fC generated by nick translation (Fig.
97 1a, c). Analysis of the resulting kymographs revealed complex binding dynamics, including non-
98 motile (12.6%) and motile TDG binding events (i.e., exhibited 1D diffusion, 87.4%). Further co-
99 localization analysis of red TDG-HaloTag-JF635 with the blue 5fC fiducial marker in the DNA
100 (fluorescein dUTP) revealed the stationary events were primarily TDG bound to 5fC sites (Fig.
101 1d). The motile events primarily represent TDG binding to the DNA nonspecifically. Of note,
102 diffusion events that are less than 0.3 seconds are not detected. Fitting the motile events to a
103 cumulative residence time distribution plot (CRTD) with a one-phase exponential decay, a
104 binding lifetime of 7.5 ± 0.3 seconds was obtained (Fig. 1d-e, Table 1). Only motile events with a
105 binding lifetime of 7.9 ± 0.1 sec were observed in additional single-molecule imaging
106 experiments with TDG-HaloTag-JF635 and unmodified DNA, supporting that the stationary
107 events reflect TDG bound to a lesion (Fig. 1f-g, Table 1). In contrast to the motile events, the
108 CRTD plot of stationary events that co-localized with the 5fC fiducial marker were fit with a two-
109 phase exponential decay with binding lifetimes of 14.7 ± 1.7 and 101.1 ± 7.1 seconds (Fig. 1d-e,
110 Table 1). We hypothesize that these longer TDG lifetimes represent binding to and cleaving the
111 5fC. This is consistent with previous kinetic experiments that indicate it takes ~ 68 secs (k_{max} of
112 0.61 min^{-1}) to cleave 5fC at 22°C, plus the time for TDG to dissociate from the generated abasic
113 site at[2]. Notably, the shorter 14.7 ± 1.7 second binding events, that co-localized with the 5fC
114 fiducial marker, exclusively occur after a long-lived event. These shorter events likely represent
115 TDG binding to an abasic site generated by its catalytic function or a nicked abasic site
116 processed by apurinic/apyrimidinic endonuclease 1 (APE1) in the extracts[19].
117 Since the nick translated DNA contains multiple sites with a region of 5fC, we generated
118 a 12 kb DNA substrate that contains a single center-positioned 5fC to more easily observe if
119 lifetime differs over time, indicating turnover. We then performed the correlative optical
120 tweezers-fluorescence microscopy (CTFM) experiment using TDG-HaloTag-JF635 with the
121 single 5fC DNA. A CRTD plot of stationary events that co-localized with the 5fC were fit with a

122 two-phase exponential decay with binding lifetimes of 0.67± 0.06 sec and 41.8 ± 12.2 sec (Fig.
123 1d-e,h-i, Table 1), consistent with the short- and long-lived binding lifetimes observed on the
124 DNA substrate containing 10 5fC sites (Fig. 1 e). These data further indicate that long-lived
125 events represent TDG binding to the substrate and the short-lived events are mostly likely TDG
126 rebinding to the catalytic product abasic site (Fig. 1h-i).

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

128 After observing TDG bind and move on DNA nonspecifically, we sought to determine the
129 mode of linear diffusion TDG uses on DNA. Linear diffusion occurs through sliding and hopping
130 mechanisms that have distinct behaviors. Sliding is when a protein tracks the DNA helix without
131 dissociating, leading to coupled translational and rotational diffusion[20,21]. Alternatively, hopping is
132 when a protein micro-dissociates as it translocates along the DNA, leading to uncoupled
133 translational and rotational diffusions. Notably, the sliding and hopping mechanisms can be
134 distinguished by determining whether the diffusion coefficient of TDG changes as a function of
135 salt concentration: diffusion via sliding is unaffected by salt concentration, whereas diffusion for
136 hopping is sensitive to salt concentration[22-24]. To further differentiate whether TDG uses a
137 sliding, hopping, or multimodal 1D diffusion, we performed additional single-molecule imaging
138 experiments with TDG-HaloTag-JF635 and unmodified DNA at varying ionic concentrations (75
139 to 150 mM NaCl). We observed an increase in the diffusion from 1.6×10^{-2} $\mu m^2/s$ at 75 mM NaCl,
140 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,
141 Supplemental Fig. 5a-d). The theoretical limit of diffusion of TDG-HaloTag sliding along the DNA
142 was calculated to be 2.56×10^{-1} $\mu m^2/s$ (see Supplemental Note). Any rates that exceed this
143 theoretical limit of diffusion would reflect hopping events. While there was an increase in the
144 diffusion with increasing ionic strength and a widening of the distribution of diffusivities, none of
145 the values were above the theoretical limit of diffusion to unambiguously signify hopping. Since
146 the crystal structures (PDBs 2RBA and 4Z7B) show TDG interacts with both strands of the helix
147 when bound to a lesion, we labeled and mixed TDG-HaloTag with two different colored labels,

148 JF635 (red) and JF552 (green) to directly visualize hopping. Two types of behavior were
149 observed: (1) JF635-TDG and JF552-TDG collide, in which the two orthogonally labeled TDG
150 proteins did not bypass one another (62.5%), and (2) JF635-TDG and JF552-TDG molecules
151 bypass one another (37.5%) (Fig. 2b). The only way it is possible to observe both of these
152 events is for TDG to have the ability to not only slide on the DNA but also to hop.
153 *TDG catalytic variants reveal R275 is essential for base detection.*
154 N140 is essential for activating the water molecule for catalysis to occur and R275 fills in
155 the hole in the duplex that is generated when the base is flipped into the active site. The
156 difference in expression of each variant compared to endogenous TDG is 72-fold for N140A, 45-
157 fold for R275A, and 21-fold for R275L (Fig. S2). Mutating to R275A and R275L decreases the
158 excision of T from G:T pairs by 8-fold and 30-fold, respectively, and also reduces the binding
159 affinity of either variant to G:T containing substrates 3-fold[13]. It is unknown if either variant
160 affects the ability of TDG to search for DNA lesions[3,13,25]. To investigate the role of these key
161 amino acids on the lesion search and recognition mechanism of TDG, we performed additional
162 single-molecule imaging with TDG-HaloTag-N140A, R275L, and R275A variants and DNA
163 containing regions of 5fC. Consistent with the WT TDG, we observed motile and non-motile
164 events for all three variants tested, which represent nonspecific and 5fC binding, respectively.
165 The N140A nonspecific events (motile) bound with a lifetime of 15.9 ± 0.23 seconds (84.8%)
166 and the specific events (non-motile) bound with a lifetime of 52.9 ± 1.1 seconds (15.2%) (Fig.
167 3a, Supplemental Figs. 1d and 6a-b). This shorter lifetime for specific events (compared to WT)
168 and the absence of short-lived rebinding events, further suggesting that wild type TDG is able to
169 cleave the 5fC moiety and rebind to the abasic site product. The R275A nonspecific events
170 (motile) bound with a lifetime of 1.07 ± 0.02 seconds (92.5%) and the specific events (non-
171 motile) bound with a lifetime of 10.0 ± 1.1 seconds (7.5%) (Fig. 3b, Supplemental Figs. 1e and
172 6c-d). The R275L nonspecific events (motile) bound with a lifetime of 0.7 ± 0.02 seconds
173 (88.7%) and the specific events (non-motile) bound with a lifetime of 1.7 ± 0.06 seconds (11.3%)

174 (Fig. 3c, Supplemental Figs. 1e and 6e-f). N140A, R275A, and R275L bound to unmodified DNA
175 with lifetimes of 7.7 ± 0.24, 2.8 ± 0.13, and 1.8 ± 0.02 seconds, respectively (Fig. 3a-c, Table 1),
176 with the R275A and R275L lifetimes shorter than the lifetime of TDG-HaloTag-JF635 binding to
177 unmodified DNA. Together these data reveal that disrupting key active site residues disrupts the
178 ability of TDG to bind DNA as well as demonstrating a reduced binding lifetime at sites of 5fC, in
179 agreement with previous bulk biochemical studies[13,26].

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

190 We can estimate the residence time per base pair by utilizing the stepping rate (see
191 Supplemental Note). The base pair residence time for wild type was estimated to be 2.1 ± 1.6
192 μs. For N140A, the base pair residence time was estimated to be 2.6 ± 2.2 μs, which is similar
193 to the time for wild type TDG (2.1 ± 1.6 μs). In contrast, the R275A and R275L TDG variants had
194 shorter residence times of 0.5 ± 0.3 μs and 0.7 ± 0.5 μs, respectively (Supplemental Table 1).
195 This may indicate that the R275A/L variants are no longer able to flip out and sample base
196 pairs, which agrees with previous biochemical studies that indicated R275 plays a role in
197 promoting or stabilizing the flipped nucleotide[13]. While the faster diffusion observed for the
198 R275A and R275L TDG variants increases the amount of DNA sampled during each binding

199 event, the rapid search may be less effective if TDG is more likely to miss the lesion or modified
200 base, similar to behavior observed with other DNA glycosylases wedge residue variants[27].
201 *The N-terminus of TDG facilitates its movement on DNA.*
202 Previous work identified the first 81 amino acids are not essential for the catalytic activity
203 of TDG[28]. However, we hypothesized that the intrinsically disordered N-terminus may play a key
204 role for TDG search of unmodified DNA (Supplemental Fig. 1a, c). To investigate the role of the
205 TDG N-terminal region on lesion search and lesion recognition mechanism, we performed
206 additional single-molecule imaging experiments with TDG-HaloTag containing a truncation of
207 the N-terminal 81 residues (ΔN-term-TDG-HaloTag) and DNA containing regions of 5fC
208 modification. The ΔN-term-TDG-HaloTag was expressed 50-fold higher compared to
209 endogenous TDG (Fig S2). We observed similar behavior with ΔN-term-TDG-HaloTag as full-
210 length TDG-HaloTag, where long-lived, stationary events were followed by shorter lived
211 stationary events (Fig. 4a). Interestingly, the nonspecific lifetime of ΔN-term-TDG-HaloTag was
212 1.2 ± 0.02 seconds which is three-fold shorter than full-length TDG-HaloTag (Fig. 4b). This
213 reduced lifetime on nonspecific DNA supports the hypothesis that the N-terminus of TDG is
214 important for searching the DNA.
215 A previous AFM study determined that TDG bends the DNA ~30° when searching for a
216 lesion, whereas the DNA is bent ~70° when TDG is bound to the lesion[29]. We hypothesized
217 based on the published AFM study[29] that the necessity of DNA bending by TDG may result in
218 shorter lifetimes at higher DNA tensions because TDG cannot fully engage with the extended
219 DNA, but our data showed this is not the case. In our prior experiments we used a tension of 10
220 pN. By increasing the tension from 5 pN to 40 pN with full-length TDG, we observed similar
221 lifetimes of 6.49 ± 0.02, 6.45 ± 0.06, and 6.25 ± 0.20 seconds at 5, 10 and 20 pN, respectively
222 (Fig. 4c, Table 2). At 30 pN and 40 pN, we observed lifetimes that were significantly ~two-fold
223 higher than at lower tensions (Fig. 4c, Table 2). For ΔN-term TDG, we also observed a two-fold
224 increase in the binding lifetimes with increasing tension. Additionally, ΔN-term TDG stays bound

225 to the unmodified DNA with a 3-5-fold shorter lifetime than full-length at the respective tensions
226 (Table 2). Shorter lifetimes for ΔN-term compared to full-length further supports that idea that
227 the N-terminus is essential for prolonged interaction with nonspecific DNA. MSD analysis
228 indicated a decrease in diffusivity with increasing DNA tension for both full-length and ΔN-term
229 TDG, showing an ~ 5-fold decrease from 5 to 40 pN (Fig. 4e-f, Table 2). Longer lifetimes and a
230 decrease in diffusivity at higher tensions could indicate that TDG is better at interrogating the
231 DNA for the base modification. Furthermore, these data indicate that bending of the DNA is not
232 essential for TDG to bind productively to DNA. This idea is further supported by an increase in
233 the residence time per base pair (Supplemental Note, Supplemental Table 2). The anomalous
234 diffusion exponent, α, provides further insight into the movement of a particle, with α = 1 is
235 random diffusion, α < 1 sub-diffusive (constrained motion), and α > 1 is super diffusive. For full-
236 length WT TDG a decrease in the ~~anomalous diffusion exponent, α,~~ was observed but we did
237 not see a significant decrease in α for ΔN-term TDG (Fig. 4e-f, Table 2). The TDG variants,
238 N140A, R275L, and R275A, were also investigated at 40 pN to further investigate base
239 interrogation investigate base interrogation at the idea of base interrogation. TDG N140A had a
240 similar behavior to WT TDG: a decrease in both alpha and diffusivity were observed with an
241 increase in tension (Fig. 4g). At high tension, the diffusivity of TDG R275L significantly
242 decreased compared to low tension, but no significant difference was observed for TDG R275A.
243 (Fig. 4h-i), supporting the idea that the size of the side chain aids in flipping in the base. These
244 data indicate that DNA tension, perhaps through increasing the distance between stacked bases
245 (~5% increase in distance at 40 pN compared to 10 pN), provides better interaction of the N-
246 terminal domain to increase the affinity to DNA increasing the dwell time and slowing the rate of
247 diffusion to a mode of constrained motion.

248 *N-terminus of TDG mediates nucleosome interactions.*

249 We sought to understand how chromatin structure impacts the TDG search mechanism
250 by reconstituting an undamaged nucleosome using a 601 sequence[30]and ligating the

251 nucleosome into 6.2 kb biotinylated handles using the DNA tethering kit from LUMICKS
252 (Supplemental Fig. 7). We then performed single-imaging experiments using TDG-HaloTag and
253 the Cy3 labeled mononucleosome-containing DNA substrate. We observed three different
254 behaviors when TDG encountered the nucleosome: (1) TDG approaches from one side and
255 hops over the nucleosome (bypass) (Fig. 5a), (2) TDG collides with but does not move past the
256 nucleosome (no bypass) (Fig. 5b), and (3) TDG colocalizes with the nucleosome after the initial
257 encounter (Fig. 5c). Further analysis revealed that TDG bypasses the nucleosome during 33.3%
258 of encounters, did not bypass the nucleosome during 39.2% of encounters, and colocalized with
259 the nucleosome during 27.5% of the encounters. CRTD analysis fit to a one-phase exponential
260 indicated that TDG bound to nucleosomes with a lifetime of 16.5 ± 1.9 seconds (Fig. 5e), which
261 is 2-fold longer than non-damaged DNA, and 4.5-fold shorter than binding to 5fC.
262 Since we observed the N-terminus is important for movement on DNA, we also sought to
263 see if the N-terminus plays a role in binding or bypassing nucleosomes. To address this, we
264 performed additional single-imaging experiments using TDG-HaloTag-ΔN-term and the
265 mononucleosome-containing DNA substrate. These experiments revealed a significant
266 decrease in the frequency of ΔN-term colocalized with the nucleosome (8.3%), an increase in
267 the number of events that approached but did not bypass the nucleosome (61.1%), and saw a
268 similar frequency for bypass (30.6%) (Fig. 5e). This decrease in TDG colocalization with the
269 nucleosome indicates that the N-terminus plays an important role in nucleosome binding.

270 **Discussion**

271 In this study, we utilized single-molecule methods to determine how TDG searches for
272 DNA damage as well as uncovering how key active site residues, domains, and the presence of
273 nucleosomes alter this search process. Similar to other DNA glycosylases, TDG forms
274 moderately stable interactions with unmodified DNA to optimize its search[31,32]. Despite the
275 ongoing 1D search, in our single molecule regime TDG primarily binds its modified sites directly
276 out of solution in a 3D diffusion mechanism, with a limited number of events that bind

277 unmodified DNA first and then slide into the lesion. Of note, sliding interactions of less than 100
278 base pairs are not discernable on our system so these events that appear as 3D may have
279 some component of 1D diffusion not observed, as was observed for human uracil DNA
280 glycosylase hUNG2[33,34]. Furthermore, we found that TDG bound specific lesions sites 10-fold
281 longer than unmodified DNA. Since TDG is able to bind a wide array of modified bases,
282 including 5caC and G:T mismatch, it is plausible that a similar behavior would be observed:
283 TDG will bind to the modified base and become nonmotile, with rebinding to the generate abasic
284 site. While it is currently not known how long TDG will bind to different lesions, it is likely that
285 TDG will bind to a specific base modification for a time that is directly correlated to the
286 biochemically determined cleavage rate followed by subsequent rebinding to an abasic site.
287 Examining TDG under different ionic conditions, as well as two color experiment, demonstrated
288 that TDG is able to both slide and hop along the DNA, which appears to be common amongst
289 bacterial and mammalian glycosylases[33,35,36]. We estimated TDG's search length on λ DNA
290 utilizing the observed diffusion constant in $\frac{bp^2}{sec}$ and the nonspecific lifetime of 7.9 sec (See
291 Supplemental Note) to be 5225 bp or ~10% of λ DNA per encounter.
292 The complex lesion search mechanism used by TDG has several potential benefits
293 during interrogation of the genome for DNA lesions. Due to TDG's high affinity for nondamaged
294 DNA, the ability to switch between sliding and hopping may also be advantageous in situations
295 where TDG encounters other factors on the DNA, including other transcription factors or
296 nucleosomes (Fig. 6). With a sliding-exclusive method, full dissociation would need to occur to
297 sample the DNA after the block. However, hopping would allow TDG to circumvent the block
298 and continue scanning on the other side of an obstacle. This use of dual linear diffusion modes
299 leads to the question of whether certain cellular conditions favor TDG hopping versus sliding on
300 DNA. For example, if TDG has a sequence preference, it may adopt a conformation at these
301 preferred sequences where tight binding leads to direct sliding as TDG further investigates that

302 region of DNA. While it is currently unclear whether TDG has sequence preference, it is
303 possible that sequences enriched in CG dinucleotides could result in initiating sliding behaviors.
304 This type of increased binding to particular sequences would be highly advantages in regions
305 with a heavier burden of DNA modification either due to (1) an enrichment of methylated CG
306 dinucleotides or (2) TET enzymes localizing to methylated CpG rich islands to oxidize 5-methlyC
307 groups for their removal by TDG and subsequent transcriptional activation[37]. It has been
308 previously determined the glycosylase OGG1 slides and hops along the DNA in search of a
309 lesion[31,35], whereas hUNG2, a uracil DNA glycosylase family member, performs 3D diffusion,
310 short range sliding, and hopping to find a lesion[33]. Thus, TDG exhibiting multiple modes of
311 diffusion is in agreement with the behavior observed for other mammalian glycosylases and
312 appears to be essential to efficiently scan the DNA and the ability to switch between modes may
313 be dependent upon sequence or other cellular factors (Fig. 6a).
314 Our single molecule work also revealed the active site variants N140A, R275A, and
315 R275L change the behavior of TDG on DNA. N140A had a shorter dwell time on 5fC but did not
316 exhibit a faster diffusion than wild type. This contrasts other previous studies with the
317 glycosylase OGG1, where the catalytically dead glycosylase bound much longer than that of the
318 WT[19,38]. On the other hand, R275A/L variants both exhibited a shorter dwell time as well as a
319 faster diffusion. One explanation for this behavior is that R275 probes the DNA sequence during
320 the sliding, enabling efficient detection of DNA damage but also slowing the speed of sliding
321 interactions. Further, it is possible that the lack of the positive charge causes TDG to have lower
322 binding affinity and thus shorter dwell times and faster rates of diffusion. A previous study on
323 bacterial glycosylases Fpg, Nei, and Nth has also shown that if a residue that interrogates the
324 DNA is changed to alanine, faster diffusivity was observed, indicating that efficient recognition of
325 a base is dependent on a single active site residue[27]. Together, these findings provide an in-
326 depth description of how key catalytic residues of TDG contribute to its search for DNA damage

327 and provides insight into how its mixed mode of hopping and sliding searches contribute to
328 efficient damage detection and repair.

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

347 We have shown that not only is TDG able to hop over a nucleosome, but it also interacts
348 with a nucleosome with an apparent dwell time of 16.5 ± 1.9 seconds. Deleting the N-terminus
349 significantly reduces the binding frequency to nucleosomes (Fig. 6b). This interaction of TDG
350 with nucleosomes may act as a scaffold to help facilitate removal of oxidized cytosine in
351 heterochromatic regions. Alternatively, perhaps TDG associates with nucleosomes prior to the
352 arrival of TET enzymes, and thus would increase the efficiency of oxidative demethylation by

353 already being in the correct location prior to the initial 5mC oxidation step in heterochromatic
354 regions. Efficient active oxidative demethylation is vital for life and of interest for potential
355 therapeutic applications[47], and this work gives key mechanistic insight on this process. TDG
356 interacting with a nucleosome could also alter nucleosome dynamics and may interfere with fork
357 progression and DNA replication, and thus might explain its degradation during S-phase in
358 mammalian cells.

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

367 **Online Methods**

368 ***DNA substrate preparation***

369 *Nick translation for confocal imaging.* Lambda DNA was biotinylated as described previously[19].
370 Briefly, DNA was purchased from New England Biotechnologies and treated with Klenow
371 fragment polymerase (NEB) and a dNTP mix containing biotinylated dCTP. This results in one
372 side of the DNA containing four biotins and the other containing six biotins.
373 Next, 1 µg of the biotinylated lambda DNA is treated with Nt.BspQI (NEB) for 1 hour at 50°C
374 then heat inactivated for 20 minutes at 80°C, which will generate single strand breaks at 10
375 different sites along the DNA by cutting 3' of the recognition site 5'-GCTCTTCN-3'. To
376 incorporate 5fC at the 10 sites, 800 ng nicked DNA was incubated with a dNTP mix containing
377 d5fCTP, fluorescein labeled dUTP, dATP, and dGTP were incubated and 10 units of DNA
378 polymerase 1 (NEB) for 6 minutes at 37°C. Reaction was inactivated by heating at 75°C for 20

379 minutes and allowed to slowly cool to room temperature. To seal the nicks, a reaction containing
380 800 ng of the nick translated DNA, 8 mM ATP and 10 units of T4 ligase (NEB) were incubated
381 overnight at 16°C. Reaction was heat inactivated for 10 minutes at 65°C and allowed to cool to
382 room temperature slowly. Of the 10 d5fCTP sites, only eight d5fCTP sites can be resolved as
383 two of the sites are within 400 bp of each other and one site is near the end of λ DNA and thus
384 too close to the bead (Supplemental Fig. 4a-c).
385 *Arms ligation for confocal imaging.* Using the LUMICKS DNA tethering kit and protocol, a
386 sequence of interest is ligated into two handles that are each 6.3 kb in length to generate a
387 substrate that is 12.6 kb in total. For the defined 5fC, the following oligonucleotide sequences
388 were ordered from Trilink and IDT, respectively:
389 Top strand (5fC): 5'-phosphate-*CAAC* ACC AGT CCA TCG CTC A**5fC**G TAC AGA GCT G–3'
390 Bottom strand: 5'- phosphate – *ACCA* CAG CTC TGT ACG TGA GCG ATG GAC TGG T-3'
391 Annealing reactions were heated at 95°C for 5 minutes in buffer containing 10 mM Tris-HCl (pH
392 8.0) and 100 mM KCl, and the heat block was turned off and reactions then cooled slowly to
393 room temperature. In a 10 µL reaction, 25 nM of the duplex, 2.5 µL of each handle, 1 µL of 10X
394 ligase buffer, and 0.5 µL of ligase were mixed and incubated at 16°C overnight. Reaction was
395 heat inactivated at 65°C for 10 minutes and allowed to slowly cool to room temperature.
396 Substrate was stored at 4°C and protected from light.
397 Undamaged nucleosome were reconstituted on a 601 sequence as described
398 previously[30]. Briefly, Human histone H3 C96S C110A, H2A K119C, H2B, and H4 were ordered
399 from the Histone Source at Colorado State University. Incubate H2A K119C and H2B in 2
400 mg/mL guanidinium buffer at room temperature for 2 hours. Mix equimolar amounts of H2A and
401 H2B and dialyze against high salt refolding buffer a total of 3 times, and at least 8 hours for
402 each exchange at 4°C. Repeat the same process for H3 C96S C110A and H4 to refold the
403 tetramer. Purify the H2A K119C/H2B dimer and the H3 C96S C110A/H4 tetramer over Superdex
404 200 column. Incubate the H2A K119C/H2B dimer with 0.7mM TCEP for 20 minutes at 4°C. Cy3-

405 maleimide dye was added to the H2A K119C in a 2:1 molar ratio and incubated at room
406 temperature for 1 hour while rocking. Reaction was quenched with 10 mM DTT, the dimer was
407 purified over a Superdex S200 column, and frozen in 50% glycerol. After confirming the
408 stoichiometry with SDS gel, add equal volume of 100% glycerol to store the H2A K119C/H2B
409 dimer and the H3 C96S C110A/H4 tetramer at -20°C.
410 To reconstitute the nucleosome on DNA, the following ultramer sequences were ordered from
411 IDT:
412 Top Strand: 5'-phosphate-*CAAC* TGA GAC CAT GTA CCC AGT TCG AAT CGG ATG TAT ATA
413 TCT GAC ACG TGC CTG GAG ACT AGG GAG TAA TCC CCT TGG CGG TTA AAA CGC GGG
414 GGA CAG CGC GTA CGT GCG TTT AAG CGG TGC TAG AGC TGT CTA CGA CCA ATT GAG
415 CGG CCT CGG CAC CGG GAT TCT CGA TAA CTC AGC AAT AGT GGG TCT CA – 3'
416 Bottom strand: 5'- phosphate -*ACCA* TGA GAC CCA CTA TTG CTG AGT TAT CGA GAA TCC
417 CGG TGC CGA GGC CGC TCA ATT GGT CGT AGA CAG CTC TAG CAC CGC TTA AAC GCA
418 CGT ACG CGC TGT CCC CCG CGT TTT AAC CGC CAA GGG GAT TAC TCC CTA GTC TCC
419 AGG CAC GTG TCA GAT ATA TAC ATC CGA TTC GAA CTG GGT ACA TGG TCT CA – 3'
420 The annealed DNA, H2A K119C/H2B dimer, and the H3 C96S C110A/H4 tetramer are mixed in
421 a 1:2:1 molar ratio and equilibrate in dialysis tubing against high salt buffer for 30 minutes. To
422 remove the high salt buffer, a series of dialysis steps occur to transition from 1.5 M NaCl to
423 0.125 M NaCl. The reconstituted nucleosome is concentrated, and heat shocked at 55°C for 30
424 minutes, prior to spinning over a 10-40% sucrose gradient for 40 hours at 125,000 xG at 4°C.
425 Fractions containing reconstituted nucleosomes are combined, buffered exchanged into TE
426 buffer, and concentrated to ~1 μM and stored at 4°C. To ligate into the LUMICKS handles, 25
427 nM of the reconstitute nucleosome, 2.5 μL of each handle, 1 μL of 10X ligase buffer, and 0.5 μL
428 of ligase were mixed and incubated at 16°C overnight. Substrate was stored at 4°C and
429 protected from light.
430 ***Plasmid Constructs***

431 Wild type TDG-HaloTag and TDG-HaloTag variants were expressed from plasmids constructed
432 utilizing Gene Universal Inc with the pHTN HaloTag CMV-neo vector (Promega). (Supplemental
433 Figs. 1a-e)

434 ***Nuclear Extract***

435 Nuclear extracts were prepared as described previously[19]. Briefly, U2OS cells were cultured in
436 5% CO2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/L glucose,
437 10% fetal bovine serum (Gibco), and 5% penicillin/streptavidin (Life Technologies). To obtain
438 transient overexpression of TDG-HaloTag, cells were transfected with 4 µg of plasmid per 4
439 million cells using the lipofectamine 2000 reagent and protocol (Thermo Fisher Cat# 11668019).
440 Nuclear extraction was performed 24 hours after transfection utilizing nuclear extraction kit from
441 Abcam (ab113474). Following the protocol from the kit, 10 µL single-use aliquots were made,
442 flash-frozen and, stored at -80°C. Extracts were diluted in buffer containing 20 mM HEPES pH
443 7.5, 100 mM NaCl, 0.2 mM EDTA, 0.1 mg/ml BSA, 1 mM DTT, 1 mM Trolox, and 5% glycerol.

444 ***Nuclear Extract Quantification***

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

452 *Western Blotting.* Different amounts of purified proteins and nuclear extracts were denatured at
453 95°C for 10 minutes. Equal volumes were loaded on a 4-20% tris-glycine polyacrylamide gel
454 (Invitrogen XP04202BOX). Proteins were transferred onto a polyvinylidene difluoride membrane
455 and blocked in 20% nonfat dry milk in PBST (1x phosphate-buffered saline containing 0.1%

456 Tween 20) for 1 hr at room temperature. Membranes were probed with TDG primary antibody
457 (1:500 Sigma HPA052263) overnight at 4˚C. Membranes were washed 3 times for 14 min in
458 PSBT and incubated with peroxidase conjugated rabbit secondary antibody for 1 h at room
459 temperature. Membranes were washed again 3 times for 15 minutes before developing using
460 SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific #34095).
461 (Supplemental Figure 2d),

462 ***Confocal Imaging in the LUMICKS C-trap***

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

473 Single-molecule experiments were then performed with a LUMICKS C-trap using a 5-
474 channel flow chamber. For experiments using λ or nick translated DNA, ~4 μm streptavidin
475 coated beads and experiments using the LUMICKS handles, ~1.7 μm streptavidin coated beads
476 were used. The beads were captured using the dual optical tweezers and then transferred to a
477 second chamber containing the biotinylated DNA and the biotinylated DNA molecule tethered
478 between the streptavidin coated beads. Once the DNA molecule is tethered, it is transferred to a
479 final chamber containing a 1:10 dilution of the nuclear extract with TDG-HaloTag-JF635. To
480 determine the concentration of the protein in the flow cell, background photon intensities were
481 determined to generate the standard curve of purified HaloTag labeled with JF635 or JF552.

482 The background intensities of the protein from the nuclear extract were measured and the curve
483 was utilized to determine protein concertation (Supplemental Figure 3a-b).
484 *Analysis.* Kymographs were analyzed with custom software from LUMICKS, Pylake. The
485 kymographs were exported as .h5 files from the C-trap and viewed using the utility "C-Trap .h5
486 Visualization GUI (2020)" by John Watters (harbor.lumicks.com). Line tracking was performed
487 using a custom script from LUMICKS based on a Gaussian fit over the line to determine its
488 positions[48]. Only events that associated and dissociated during the course of a 5-to-10-minute
489 kymograph were tracked. Events are plotted on cumulative residence time distribution (CRTD)
490 plot in order to determine the binding lifetime. Specific binding event curves are fit with a two-
491 phase exponential decay, for catalytically active WT protein. One-phase exponential decays
492 were sued for the variants, which have little or no catalytic activity, and nonspecific binding
493 events for the variants and WT. The two-phase fit for WT represents a second rebinding event to
494 the catalytic product abasic site with a shorter dwell time. The goodness of the fit for the CRTD
495 curves range from R^2 = 0.96 – 0.99. Mean squared displacement (MSD) analysis was
496 performed using a custom script from LUMICKS based on the following equation:

497 $$MSD\ (n\Delta t) = \frac{1}{N-n}\sum_{i=1}^{N-n}(x_{i+n} - x_i)^2 \quad (1)$$

498 Where N is the total number of frames in the phase, n is the number of frames at a given time
499 step, Δt is the time increment of one frame, and x_i is the particle position in the ith frame[27]. The
500 diffusion coefficient (D) was determined by fitting a linear model of one-dimensional diffusion to
501 the MSD plots:

502 $$MSD\ (n\Delta t) = 2D(n\Delta t)^{\alpha} + y \quad (2)$$

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

506 All data are available from the corresponding author upon reasonable request.

507 **Code Availability**

508 Python scripts used to calculate lifetime and MSD analysis have been deposited at

509 https://harbor.lumicks.com/scripts.

510 **Contributions**

511 B.L.S. and B.V.H. conceived the research. B.L.S. performed all Ctrap and protein concentration

512 determination experiments. V.R. performed all western blot experiments. B.L.S analyzed all

513 single-molecule data. T.M.W. reconstituted nucleosomes. B.L.S., M.A.S. and B.V.H. drafted the

514 paper, which was reviewed, discussed, and edited by all authors.

515 **Acknowledgements**

516 We greatly appreciate the discussions with the lab members during the course of these

517 experiments. This work was supported by NIH R35ES031638 (BVH), T32GM088119 (BLS), the

518 Hillman Postdoctoral Fellowship for Innovative Cancer Research and F32ES034982 (MAS),

519 2P30CA047904 to the UPMC Hillman Cancer Center, the major equipment grant

520 S10OD032158-01A1 (BVH), R35-GM136225 (ACD), R35GM128652 (BDF), and F32GM140718

521 (TMW).

522 **Competing interests**

523 None declared.

524 **Figure Legends**

525 **Figure 1: TDG binds DNA specifically and nonspecifically.**

526 (A) Cartoon schematic showing how 5fC was incorporated into nick translated λ DNA. (B)

527 Cartoon schematic showing 28 base pair duplex DNA containing a single 5fC (orange) ligated

528 into 6 kb LUMICKS handle kit, with handle 2 containing ATTO 488 (blue). (C) A diagram

529 depicting the order of reagents, which are under laminar flow, are captured in the flowcell. (D) A

530 cartoon depiction of the DNA substrate used for TDG binding, with 5fC sites shown in blue, and

531 an example kymograph with TDG binding shown in red. Specific event indicated with grey

532 asterisk and nonspecific event indicated with teal asterisk. (E) Cumulative Resident Time
533 Distribution (CRTD) analysis fit to a two-phase decay of TDG binding DNA containing 5fC
534 specifically (n = 70) and nonspecifically (n = 487). (F) A cartoon depiction of the unmodified DNA
535 substrate with an example kymograph of TDG binding and moving. (G) CRTD analysis fit to a
536 one-phase decay of TDG binding unmodified λ DNA (n = 155). (H) An example kymograph of
537 TDG binding to a single 5fC. (I) CRTD analysis fit to a two-phase decay of TDG binding to 5fC
538 (n = 28). Created with BioRender.com.

539 **Figure 2: TDG exhibits linear diffusion on DNA.**
540 (A) Scatter plot of the diffusion coefficient ($\log_{10}D$) calculated for TDG with increasing ionic
541 strength on unmodified λ. Dashed line, D_{lim}, theoretical limit to free diffusion for TDG-HaloTag.
542 ****$p < 0.0001$ by two-way ANOVA. (B) An example kymograph of the two-color TDG
543 experiment. Separate TDG-HaloTag extracts were labeled with either JF635 (red) or JF552
544 (green) and mixed at a 1:1 ratio. White arrows indicate bypass events and yellow arrows
545 indicate collision events.

546 **Figure 3: TDG catalytic variants indicate R275 is essential for base detection.**
547 Experiments conducted with unmodified or nick translated DNA containing 5fC. (A) CRTD
548 analysis fit to a one-phase decay of N140A TDG binding to unmodified (n = 163), specifically to
549 5fC (n = 30) and nonspecifically (n = 168). (B) CRTD analysis fit to a one-phase decay of R275A
550 TDG binding to unmodified (n = 110), specifically to 5fC (n = 24) and nonspecifically (n = 296).
551 (C) CRTD analysis fit to a one-phase decay of R275L TDG binding to unmodified (n = 269),
552 specifically to 5fC (n = 42) and nonspecifically (n = 330). (D) Scatter plot of the diffusion
553 coefficient ($\log_{10}D$) calculated for each variant TDG on unmodified λ. Dashed line, D_{lim},
554 theoretical limit to free diffusion for TDG-HaloTag. ****$p < 0.0001$ by two-way ANOVA.

555 **Figure 4: The N-terminus of TDG is significant for its movement on DNA. Tension alters**
556 **diffusivity and α for ΔN-term and full-length TDG.**

557 (a) A cartoon depiction of the DNA substrate with 5fC sites shown in blue, and a representative
558 kymograph with TDG binding shown in red. Break in kymograph to show location of fiducial
559 markers. (b) CRTD analysis fit to a two-phase decay of TDG binding DNA containing 5fC
560 specifically (n = 32). CRTD analysis fit to a one-phase decay of TDG binding nonspecifically (n =
561 243). (c) CRTD analysis fit to a one-phase decay of full-length TDG binding to unmodified DNA
562 at 5 pN (n = 172), 10 pN (n = 27), 20 pN (n = 53), 30 pN (n = 169), and 40 pN (n = 102). (d)
563 CRTD analysis fit to a one-phase decay of ΔN-term TDG binding to unmodified DNA at 5 pN (n
564 = 113), 10 pN (n = 201), 20 pN (n = 197), 30 pN (n = 293), and 40 pN (n = 203) (e) Plot of
565 diffusion coefficients (*D*) versus alpha (α) for full-length for 5 pN (n = 50), 10 pN (n = 41), 20 pN
566 (n = 41), 30 pN (n = 57), and 40 pN (n = 33). (f) Plot of diffusion coefficients (*D*) versus alpha (α)
567 for ΔN-term for 5 pN (n = 30), 10 pN (n = 36), 20 pN (n = 24), 30 pN (n = 22), and 40 pN (n =
568 35). (g) Plot of diffusion coefficients (*D*) versus alpha (α) for N140A for 10 pN (n = 33) and 40 pN
569 (n = 34). (h) Plot of diffusion coefficients (*D*) versus alpha for 10 pN (n = 19) and 40 pN (n = 27).
570 (α) for R275L. (i) Plot of diffusion coefficients (*D*) versus alpha (α) for R275A for 10 pN (n = 14)
571 and 40 pN (n = 19).

572 **Figure 5: N-terminus of TDG is important for interacting with nucleosomes.**

573 A cartoon depiction of the DNA substrate with Cy3 labeled nucleosome. Representative
574 kymograph with full-length TDG (shown in red) (a) bypassing, (b) does not bypass, or (c)
575 colocalizes with the nucleosome shown in green. (d) Representative kymograph of ΔN-term
576 TDG (red) colocalizing with nucleosome (green). (e) CRTD analysis fit to a one-phase decay of
577 TDG interacting with the NCP (n = 28). (f) Stacked bar graph showing the fraction of bypass
578 (green), no bypass (grey), and colocalized (teal) events for full-length and ΔN-term. shows there
579 is $p = 0.0011$ by X^2.

580 **Figure 6: Working model.**

581 TDG uses multiple modes of linear diffusion to efficiently search for 5fC. (a) WT TDG (gray)
582 fused to HaloTag (red) scans dsDNA with sliding and hopping interactions, with a lifetime of 7.9

583 ± 0.06 sec and diffusivity of 0.028 ± 0.02 μm2/sec. Upon 5fC engagement (orange), TDG binds
584 with a lifetime of 72.9 ± 5.4 sec. These interactions are dependent on active site residues N140,
585 R275, and the N-terminal domain. (b) When TDG encounters a nucleosome core particle, it can
586 bypass, not bypass, or colocalize with the NCP (16.5 ± 1.9 sec). Importantly the N-terminal
587 domain of TDG stabilizes the interaction with the NCP, resulting in a four-fold increase in
588 colocalization frequency. Created with BioRender.com.

589

590

591 **Table 1. Binding lifetimes and diffusivity**

Protein	DNA	Lifetime(s) and Percentages[b]	Average binding lifetime (τ_{avg})[b]	Diffusivity on undamaged DNA ($\mu m^2/s$)[c]
Wild Type	Unmodified	7.9 ± 0.06	72.9 ± 5.4	0.028 ± 0.02
	5fC, specific[a]	101.1 ± 7.1 (67.3 ± 3%)		
		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		

592 [a] Nick translated DNA
593 [b] The tau is the best fit value ± the standard error of the fit to the observed data
594 [c] The diffusivity is the mean ± the standard error of the mean
595
596
597
598
599
600

601 **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

602 [a] The tau is the best fit value ± the standard error of the fit to the observed data
603 [b] The diffusivity is the mean ± the standard error of the mean
604 [c] The alpha is the mean ± the standard error of the mean
605

606 **References**

607 1. Whitaker, A.M., Schaich, M.A., Smith, M.R., Flynn, T.S. & Freudenthal, B.D. Base
608 excision repair of oxidative DNA damage: from mechanism to disease. *Front Biosci*
609 *(Landmark Ed)* **22**, 1493-1522 (2017).
610 2. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine
611 and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J*
612 *Biol Chem* **286**, 35334-35338 (2011).
613 3. Maiti, A., Michelson, A.Z., Armwood, C.J., Lee, J.K. & Drohat, A.C. Divergent
614 mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from
615 DNA. *J Am Chem Soc* **135**, 15813-22 (2013).
616 4. Pidugu, L.S., Dai, Q., Malik, S.S., Pozharski, E. & Drohat, A.C. Excision of 5-
617 Carboxylcytosine by Thymine DNA Glycosylase. *J Am Chem Soc* **141**, 18851-18861
618 (2019).
619 5. Pidugu, L.S. et al. Structural Basis for Excision of 5-Formylcytosine by Thymine DNA
620 Glycosylase. *Biochemistry* **55**, 6205-6208 (2016).
621 6. Cortázar, D. et al. Embryonic lethal phenotype reveals a function of TDG in maintaining
622 epigenetic stability. *Nature* **470**, 419-23 (2011).
623 7. Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation
624 by linked deamination-base excision repair. *Cell* **146**, 67-79 (2011).
625 8. Bennett, M.T. et al. Specificity of human thymine DNA glycosylase depends on N-
626 glycosidic bond stability. *J Am Chem Soc* **128**, 12510-9 (2006).
627 9. Morgan, M.T., Bennett, M.T. & Drohat, A.C. Excision of 5-halogenated uracils by human
628 thymine DNA glycosylase. Robust activity for DNA contexts other than CpG. *J Biol Chem*
629 **282**, 27578-86 (2007).
630 10. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-
631 carboxylcytosine. *Science* **333**, 1300-3 (2011).
632 11. Kellinger, M.W. et al. 5-formylcytosine and 5-carboxylcytosine reduce the rate and
633 substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol* **19**, 831-3
634 (2012).
635 12. DeNizio, J.E. et al. TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites. *J*
636 *Mol Biol* **433**, 166877 (2021).
637 13. Maiti, A., Morgan, M.T. & Drohat, A.C. Role of two strictly conserved residues in
638 nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase.
639 *J Biol Chem* **284**, 36680-36688 (2009).
640 14. Morgan, M.T., Maiti, A., Fitzgerald, M.E. & Drohat, A.C. Stoichiometry and affinity for
641 thymine DNA glycosylase binding to specific and nonspecific DNA. *Nucleic Acids Res*
642 **39**, 2319-29 (2011).
643 15. Maiti, A., Morgan, M.T., Pozharski, E. & Drohat, A.C. Crystal structure of human thymine
644 DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition.
645 *Proc Natl Acad Sci U S A* **105**, 8890-5 (2008).
646 16. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. & Cedar, H. The role of DNA
647 methylation in setting up chromatin structure during development. *Nat Genet* **34**, 187-92
648 (2003).
649 17. Tarantino, M.E., Dow, B.J., Drohat, A.C. & Delaney, S. Nucleosomes and the three
650 glycosylases: High, medium, and low levels of excision by the uracil DNA glycosylase
651 superfamily. *DNA Repair (Amst)* **72**, 56-63 (2018).
652 18. Deckard, C.E. & Sczepanski, J.T. Reversible chromatin condensation by the DNA repair
653 and demethylation factor thymine DNA glycosylase. *Nucleic Acids Res* **49**, 2450-2459
654 (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**, 6282-4 (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**, 6929-48 (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 U S A* **108**, 563-8 (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 U S A* **111**, E2091-9 (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**, 2716-29 (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**, 9295-303 (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**, 7487-98 (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 U S A* **105**, 10791-6 (2008).
34. Schonhoft, J.D. & Stivers, J.T. Timing facilitated site transfer of an enzyme on DNA. *Nat Chem Biol* **8**, 205-10 (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 U S A* **103**, 5752-7 (2006).
36. Friedman, J.I. & Stivers, J.T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957-67 (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 (Amst)* **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 U S A* **119**, e2206513119 (2022).

40. Bercy, M. & Bockelmann, U. Hairpins under tension: RNA versus DNA. *Nucleic Acids Res* **43**, 9928-36 (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 U S A* **116**, 8350-8359 (2019).
43. Wang, H.Y., Elston, T., Mogilner, A. & Oster, G. Force generation in RNA polymerase. *Biophys J* **74**, 1186-202 (1998).
44. Rouzina, I. & Bloomfield, V.A. Force-induced melting of the DNA double helix 1. Thermodynamic analysis. *Biophys J* **80**, 882-93 (2001).
45. Williams, M.C., Rouzina, I. & Bloomfield, V.A. Thermodynamics of DNA interactions from single molecule stretching experiments. *Acc Chem Res* **35**, 159-66 (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**, 1948-57 (2016).

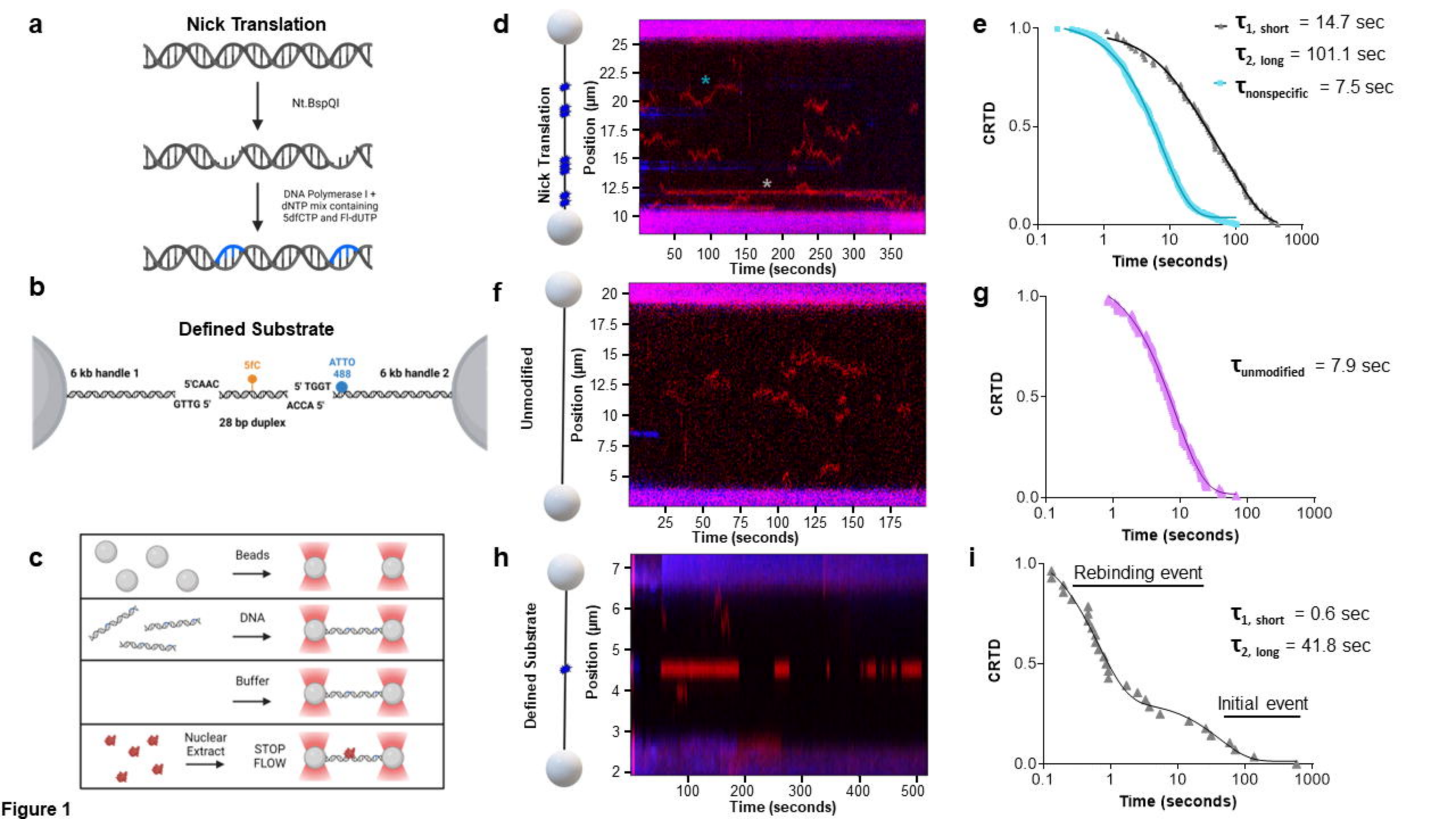

a
Nick Translation
Nt.BspQI
DNA Polymerase I +
dNTP mix containing
5dfCTP and Fl-dUTP
b
Defined Substrate
6 kb handle 1
5'CAAC
GTTG 5'
5fC
28 bp duplex
ATTO
488
5' TGGT
ACCA 5'
6 kb handle 2
c
Beads
DNA
Buffer
Nuclear
Extract
STOP
FLOW
d
Nick Translation
Position (µm)
Time (seconds)
e
CRTD
Time (seconds)
τ1, short = 14.7 sec
τ2, long = 101.1 sec
τnonspecific = 7.5 sec
f
Unmodified
Position (µm)
Time (seconds)
g
CRTD
Time (seconds)
τunmodified = 7.9 sec
h
Defined Substrate
Position (µm)
Time (seconds)
i
CRTD
Time (seconds)
Rebinding event
Initial event
τ1, short = 0.6 sec
τ2, long = 41.8 sec

Figure 1

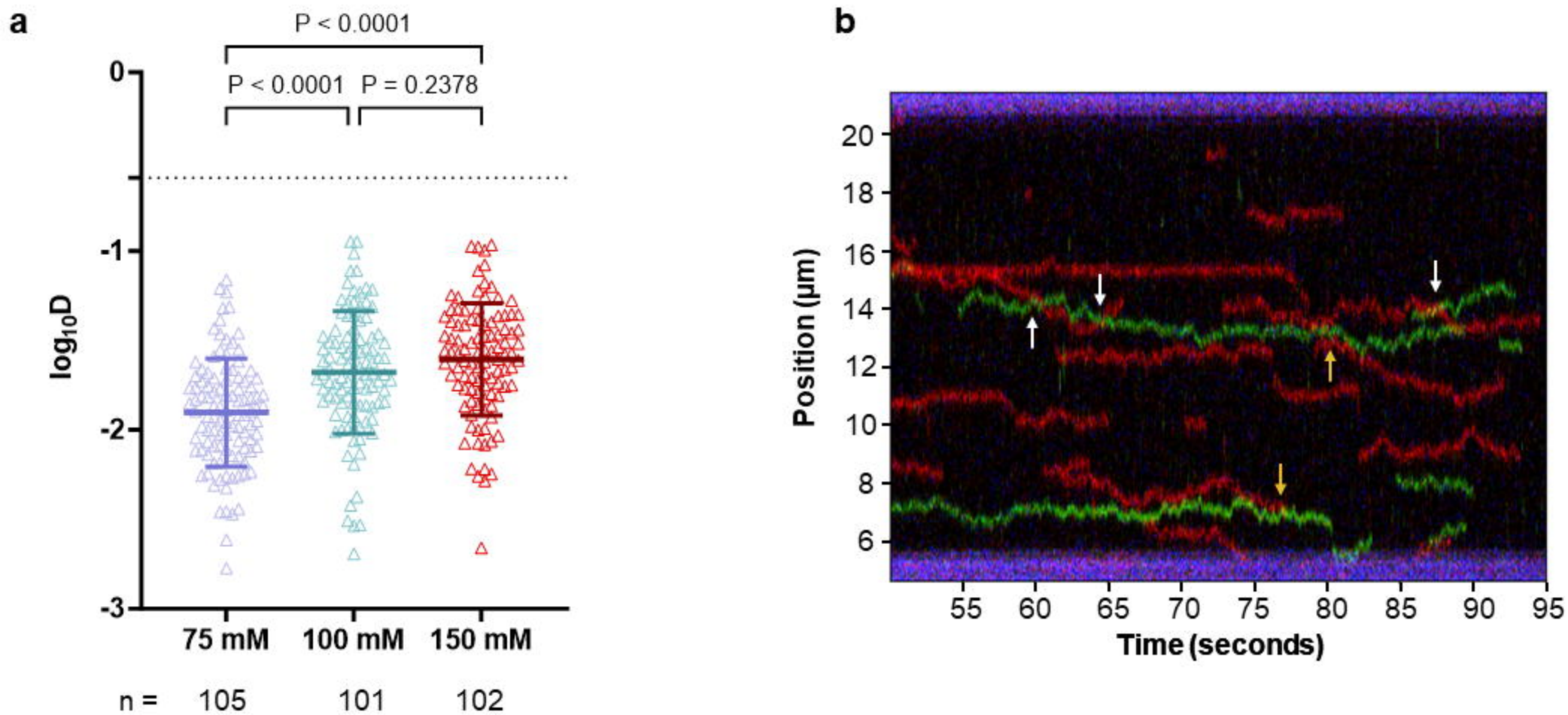

a
P < 0.0001
P < 0.0001
P = 0.2378
0
-1
-2
-3
log10D
75 mM
100 mM
150 mM
n =
105
101
102
b
Position (µm)
20
18
16
14
12
10
8
6
55
60
65
70
75
80
85
90
95
Time (seconds)

Figure 2

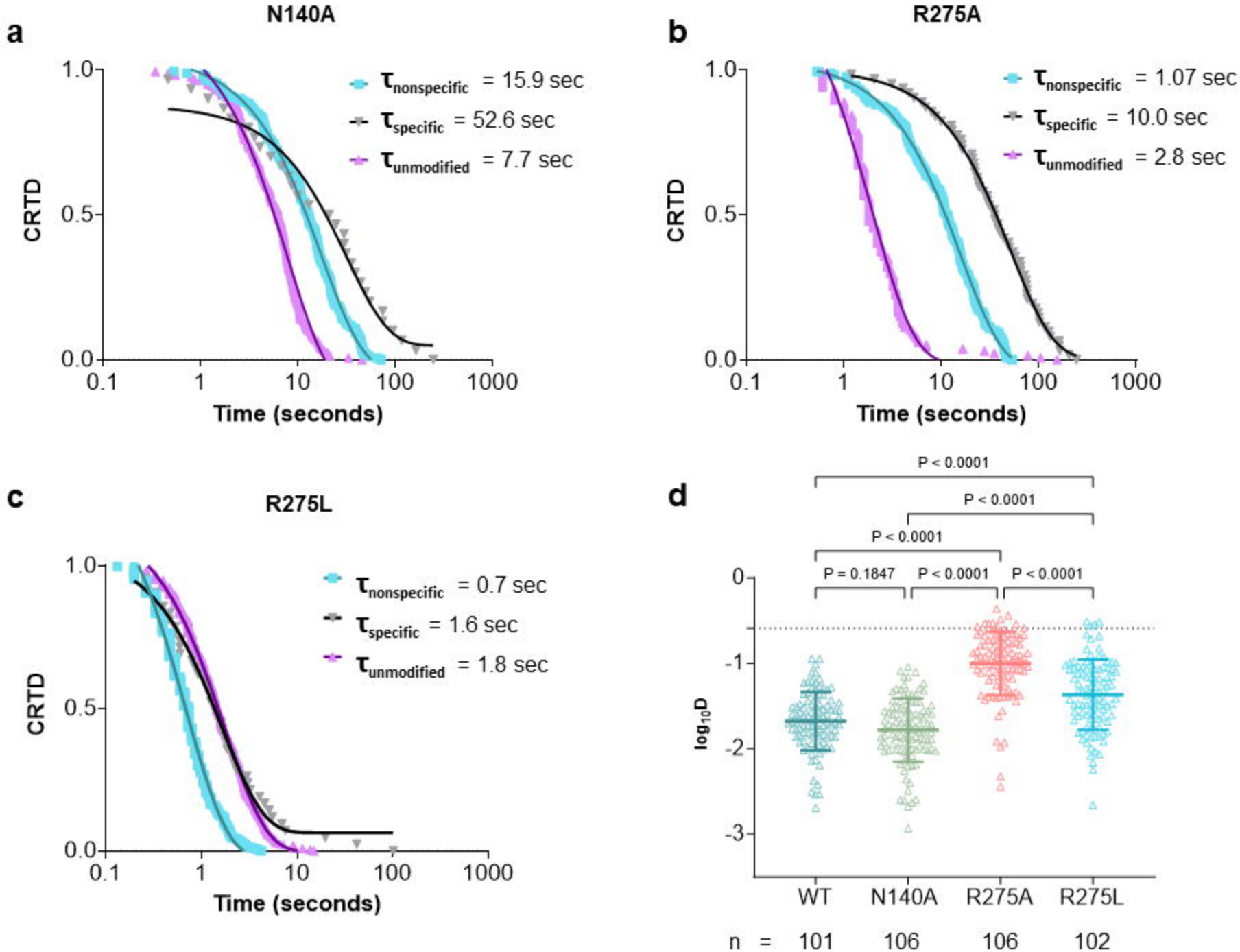
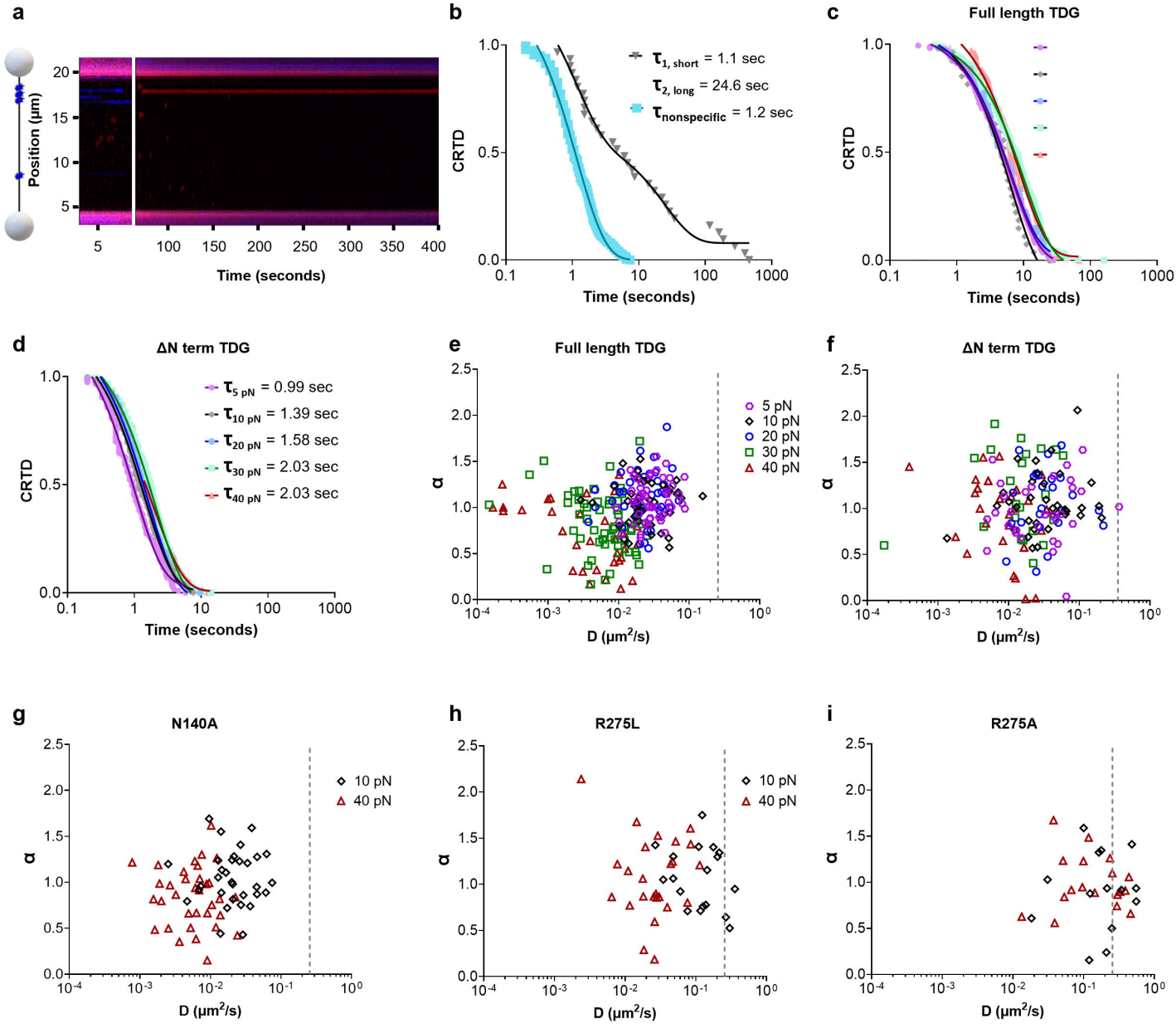

a
N140A
CRTD
Time (seconds)
τnonspecific = 15.9 sec
τspecific = 52.6 sec
τunmodified = 7.7 sec
b
R275A
τnonspecific = 1.07 sec
τspecific = 10.0 sec
τunmodified = 2.8 sec
c
R275L
τnonspecific = 0.7 sec
τspecific = 1.6 sec
τunmodified = 1.8 sec
d
P < 0.0001
P < 0.0001
P < 0.0001
P = 0.1847
P < 0.0001
P < 0.0001
log10D
WT
N140A
R275A
R275L
n = 101 106 106 102

Figure 3


a
Position (μm)
Time (seconds)
b
CRTD
Time (seconds)
τ1, short = 1.1 sec
τ2, long = 24.6 sec
τnonspecific = 1.2 sec
c
Full length TDG
CRTD
Time (seconds)
d
ΔN term TDG
CRTD
Time (seconds)
τ5 pN = 0.99 sec
τ10 pN = 1.39 sec
τ20 pN = 1.58 sec
τ30 pN = 2.03 sec
τ40 pN = 2.03 sec
e
Full length TDG
α
D (μm2/s)
5 pN
10 pN
20 pN
30 pN
40 pN
f
ΔN term TDG
α
D (μm2/s)
g
N140A
α
D (μm2/s)
10 pN
40 pN
h
R275L
α
D (μm2/s)
10 pN
40 pN
i
R275A
α
D (μm2/s)

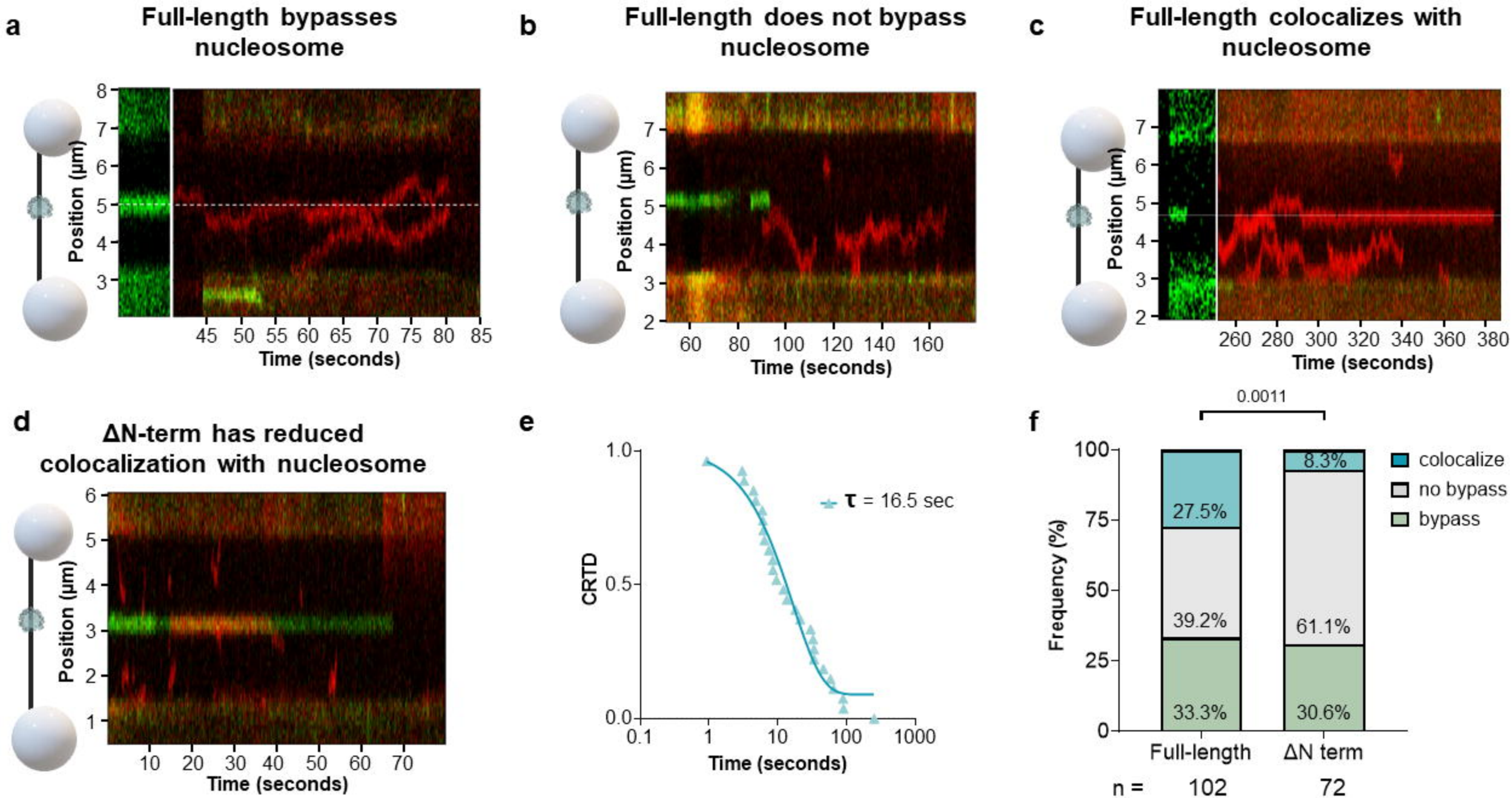

a
Full-length bypasses nucleosome
Position (µm)
Time (seconds)
b
Full-length does not bypass nucleosome
Position (µm)
Time (seconds)
c
Full-length colocalizes with nucleosome
Position (µm)
Time (seconds)
d
ΔN-term has reduced colocalization with nucleosome
Position (µm)
Time (seconds)
e
τ = 16.5 sec
CRTD
Time (seconds)
f
0.0011
Frequency (%)
colocalize
no bypass
bypass
27.5%
39.2%
33.3%
8.3%
61.1%
30.6%
Full-length
ΔN term
n = 102 72

Figure 5

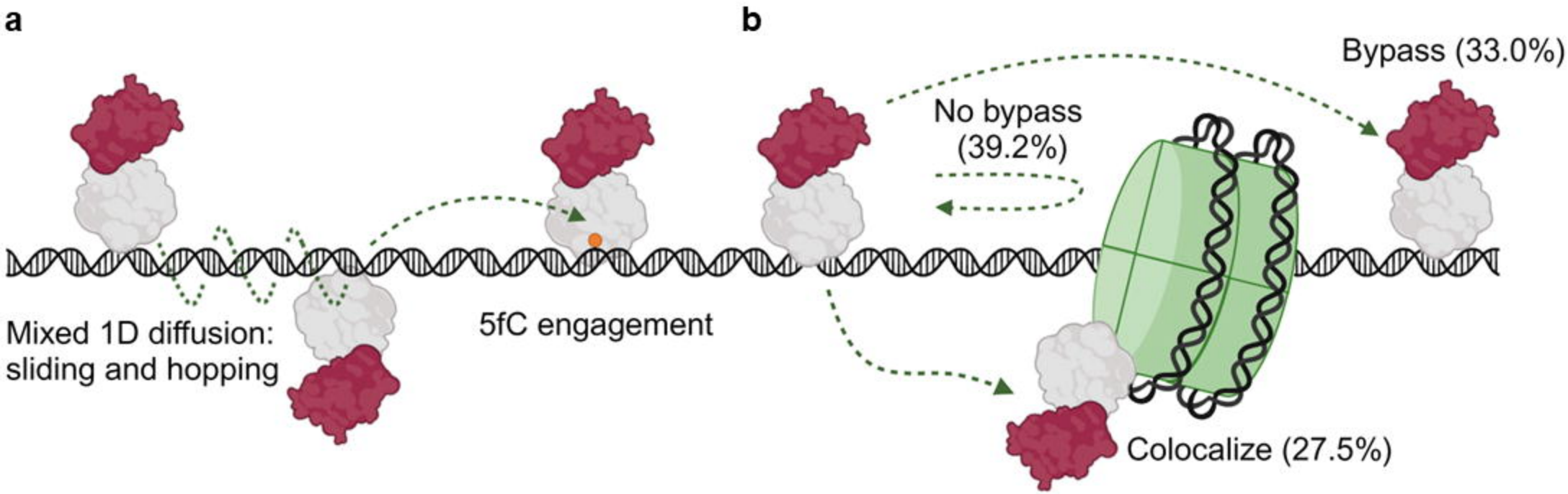

a
b
Bypass (33.0%)
No bypass
(39.2%)
Mixed 1D diffusion:
sliding and hopping
5fC engagement
Colocalize (27.5%)

Figure 6