Single-molecule analysis reveals TDG exhibits multiple modes of linear diffusion to process 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, 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.

Abstract

DNA methylation plays a key role in epigenetics, with 60-80% of CpG sites containing 5-methylcytosine. Base excision repair (BER) is suggested to be the main pathway involved in active DNA demethylation. 5-formylctyosine (5fC), an oxidized moiety of methylated cytosine, is recognized and removed by thymine DNA glycosylase (TDG) to generate an abasic site. TDG binds avidly to abasic sites and is product inhibited. Using single molecule fluorescence experiments, we saw TDG interact with DNA containing 5fC specifically and non-specifically with lifetimes of 72.9 and 7.5 seconds, respectively. These results indicate that TDG cleaves the 5fC and stays bound for an extended time at the generated abasic site. Mean squared displacement analysis and a two color TDG experiment indicate that TDG exhibits multiple modes of linear diffusion, including hopping and sliding, in search of a lesion. The catalytically crippled variants, N140A and R275A/L, have a reduced binding lifetime compared to wild type and Mean Squared Displacement (MSD) analysis indicates that R275L/A moves on the DNA with a faster diffusivity. These results indicate that mutating R275, but not N140 interferes with damage recognition by TDG. Our findings give insight into how TDG searches for its lesions in long stretches of undamaged DNA.

Introduction

DNA can be damaged by a variety of endogenous and exogenous sources and these potentially mutagenic bases are removed via base excision repair (BER). In humans, BER is initiated by one of eleven damage specific DNA glycosylases that recognize the lesion 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-carbxylcytosine (5caC) [1-4]. Interestingly, is has been shown that depletion of TDG results in embryonic lethality in mice, showing the significant role that TDG plays in active demethylation in addition to its role in removing lesions caused by environmental DNA damage [5,6].

While TDG has been shown to preferentially remove T from G:T mismatches in the context of a CpG, other substrates do not show this sequence context dependence [7,8]. One such lesion is 5fC, which has an abundance of $6x10^4$ per human genome, and the removal of 5fC was similar in the context of a CpG and outside a CpG [9-11]. The lack of CpG sequence preference for 5fC is due to strong interactions of TDG amino acid side chains with the formyl group [3,4,12]. In addition to TDG specificity for G:T mismatches and fC, TDG is able to bind to DNA non-specifically with an affinity of 293 ± 64 nM for one molecule and a second molecule can bind with an affinity of 1172 ± 254 nM [13].

Previous studies have shown that N140 is essential for the catalytic activity on 5fC because of interactions with the base and the nucleophile. The N140A mutation greatly reduces the catalytic activity 16,000-fold [4]. It has also been shown that R275 also 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. Mutating to R275A and R275L decreases the excision rate 8-fold and 30-fold, respectively, and also reduces the binding affinity to G:T containing substrates 3-fold [14].

While the roles for these key active site amino acids in catalysis have been well studied, it remains unknown if they play any role in TDG searching for substrates embedded in long DNA molecules in which there is an excess of undamaged DNA.

Here, we utilized a new single-molecule method (SMADNE) to watch the search process of fluorescently labeled TDG in real-time on defined DNA tethers using a LUMICKS C-trap [15]. By observing TDG scanning the DNA as it searched for 5fC, we determined that TDG searches for damaged sites using a combination of sliding and hoping. We have also found that specific amino acid side chains that are essential for catalytic activity also impact the dwell time on DNA and alter the non-specific diffusion.

Materials and Methods

DNA substrate preparation

Nick translation for confocal imaging. Lambda DNA was biotinylated as described previously [15]. 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 hour at 50°C then heat inactivated for 20 minutes 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 minutes at 37°C. Reaction was inactivated by heating at 75°C for 20 minutes 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 minutes at 65°C and allowed to cool to room temperature slowly.

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. Briefly, the following oligonucleotide sequences were ordered from Trilink and IDT, respectively:

Top strand (5fC): 5'-phosphate-*CAAC*ACCAGTCCATCGCTCA**5fC**GTACAGAGCTG–3'

Bottom strand: 5'- *ACCA*CAGCTCTGTACGTGAGCGATGGACTGGT-3'

Annealing reactions were heated at 95°C for 5 minutes 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 minutes and allowed to slowly cool to room temperature. Substrate was stored at 4°C and protected from light.

Plasmid Constructs

Wild type TDG-HaloTag and TDG-HaloTag variants were expressed from mutant plasmids constructed utilizing Gene Universal Inc with the pHTN HaloTag CMV-neo vector (Promega).

Nuclear Extract

Nuclear extracts were prepared as described previously [15]. 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 hours 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.

Nuclear Extract Quantification

Western Blotting. Different amounts of purified proteins and nuclear extracts were denatured at 95°C for 10 minutes. 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 minutes before developing using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific #34095).

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.

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-JF-552 were excited with a 561 nm laser and the emission was collected with a 575–625 nm band pass filter, and HaloTag-JF-635 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.

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 [16]. Mean squared displacement (MSD) analysis was performed using a custom script from LUMICKS based on the following equation:

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

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 [17]. The diffusion coefficient (D) was determined by fitting a linear model of one-dimensional diffusion to the MSD plots:

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

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.

Results

TDG binds DNA specifically and non-specifically.

In order to visualize TDG binding to DNA, we utilized the new approach termed single-molecule analysis of DNA repair proteins from nuclear extracts (SMADE) [15]. A plasmid containing full length TDG with a HaloTag on the C-terminus was expressed in mammalian cells, labeled with JF-635, and a nuclear extraction was performed and used on the C-trap (supplemental figures). In order to generate a DNA substrate containing 5fC, biotinylated λ DNA was treated with Nt. BspQI to generate 10 nick sites where d5fCTP and fluorescein-labeled

dUTP are incorporated via nick translation (Fig. 1A). Of the 10 sites, only eight are observed because two of the sites are within 400 base pairs and cannot be resolved on the C-trap and one of the sites is near the end of λ DNA and thus too close to the bead to be observed (Supplemental Figure). Utilizing the LUMICKS C-trap, 4.4 - 4.8 μm streptavidin coated beads are(?) trapped by optical tweezers. Then the DNA is flowed into the flow cell and one molecule of DNA is tethered between the beads. Next, the nuclear extract containing TDG-HaloTag-JF635 is diluted 1:10 and added into the flow cell. Once flow is stopped, the tethered DNA is brought over to the inlet port of channel 4 to begin collecting kymographs (Fig. 1B). Over the course of each kymograph, stationary events were observed (12.6%) as well as events that were motile (i.e., exhibited 1D diffusion) (87.4%), with all particles associating and dissociating during the course of a 5-10-minute kymograph.

Looking further into the motile and stationary events, we were able to directly measure specificity. By analyzing TDG-HaloTag events co-localized with the fiducial marker in the DNA (fluorescein dUTP), indicative of the location of the 5fC, we determined the stationary events were TDG bound to sites that contained 5fC, and the motile events represent TDG binding nonspecifically to the DNA. Fitting the motile events to cumulative residence time distribution plot (CRTD) with a one-phase exponential decay determined a lifetime of 7.5 ± 0.07 seconds. The stationary events were fit to a CRTD with two-phase exponential decay with lifetimes of 14.7 ± 4 and 101.1 ± 17 seconds (Fig. 1 C-D). We also measured the binding of TDG to undamaged DNA and only motile events were observed with lifetimes of 15.6 sec, supporting that when TDG binds to a lesion it is stationary (Fig. 1E -F). We hypothesize that the longer lifetime is TDG binding to and cleaving the 5fC, as it was previously determined biochemically it has a k_{max} of 0.61 min^{-1}, meaning that it takes ~ 98 seconds to cleave 5fC at 22°C [1]. The shorter binding events exclusively occur after a long-lived event and could represent TDG binding to an abasic site generated by its catalytic function or a nicked abasic site processed by APE1 in the extracts [15]. Since the nick translated DNA contains multiple sites of 5fC, we generated a DNA

substrate that contains a single 5fC to more easily observe if lifetime differs over time, indicating turnover. By having only one lesion in the center of the substrate, we can determine if the behavior we saw for DNA containing multiple 5fC is observed. Using the DNA tethering kit from LUMICKS, we ligated a 36 bp duplex containing one central 5fC into the handles and were able to tether them to 1.76 μm streptavidin coated beads. Upon analyzing the binding events, we again saw a long-lived event, 41.8 ± 28 sec, that is then followed by shorter binding events, 0.6 ± 0.1 sec, supporting the idea that TDG is rebinding to the generated abasic site (Fig. 1G-H).

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

After observing TDG bind and move on DNA non-specifically, we sought to determine the mode of linear diffusion TDG exhibits on DNA. Linear diffusion can be broken down into two main forms: sliding and hopping that have very distinct behaviors. When a protein is sliding on DNA, it is spiraling around the helix without dissociating, and the translational and rotational diffusions are coupled [18,19]. Alternatively, when a protein is hopping on DNA, it is micro-dissociating as it translocates along the DNA and the translational and rotational diffusions are not coupled. In order to differentiate between sliding and hopping, we studied TDG at varying ionic concentrations. If a protein is sliding on DNA, it is expected that the rate of diffusion will be independent of salt concentration. If a protein is hopping on DNA, the observed linear diffusion is expected to increase as hopping is dependent on a protein's affinity to nonspecific DNA, which is impacted by increasing salt conditions [20-22]. We observe 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 (Figure 2A). Of note, the theoretical limit of diffusion of TDG-HaloTag sliding along the DNA was calculated to be 2.56×10^{-3} $\mu m^2/s$ (see supplemental description), suggesting that faster diffusion reflects hopping events. While there was an increase in the diffusion, none of the values were above the theoretical limit to diffusion to unambiguously signify hopping. In order to directly visualize hopping by TDG, we labeled and mixed TDG-HaloTag with two different colored labels, JF-635 (red) and JF-552 (green). Two types of behavior were observed. Red

TDG and a green TDG were observed to collide, in which the two orthogonally labeled TDG proteins did not bypass one another, and another mode in which red TDG and green TDG molecules bypass one another (Figure 3B). 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 indicate R275 is essential for base detection.

It was previously determined that N140 and R275 are both essential for the catalytic activity of TDG-HaloTag, but it remains unclear how they affect TDG's ability to search for DNA damage [2,14,23]. Therefore, we investigated the roles these key amino acids in binding to and diffusing on DNA containing 5fC and undamaged DNA. On DNA containing 5fC, all three variants bound to the DNA non-specifically and were motile, and also bound specifically and were stationary, as was observed for the TDG-HaloTag. The N140A motile, non-specific events bound with a lifetime of 17.4 ± 0.5 seconds (84.8%) and the stationary, specific events bound with a lifetime of 31.6 ± 6.4 seconds (15.2%) (Fig. 3A). The R275A motile, non-specific events bound with a lifetime of 1.07 ± 0.03 seconds (92.5%) and the stationary, specific events bound with a lifetime of 10.0 ± 2.7 seconds (7.5%) (Fig. 3B). The R275L motile, non-specific events bound with a lifetime of 0.7 ± 0.02 seconds (88.7%) and the stationary, specific events bound with a lifetime of 1.6 ± 0.13 seconds (11.3%) (Fig. 3C). N140A, R275A, and R275L bound to undamaged DNA with lifetimes of 1.9 ± 0.13, 2.8 ± 0.22, and 1.8 ± 0.03 seconds, respectively (Fig. 3A-C), which is shorter than the lifetime of TDG-HaloTag binding to undamaged DNA. Together this shows that disrupting key active site residues disrupts the ability of TDG to bind and search along undamaged DNA, as well as demonstrating a reduced binding lifetime at sites of 5fC, in agreement with previous bulk biochemical studies.

Since we observed the binding lifetimes of the variants to differ from wild type TDG, we examined the diffusion behavior by MSD analysis. N140A had a diffusion coefficient of 2.2×10^{-2} $\mu m^2/s$, which was not significantly different from wild type (Fig. 3D). This indicates that N140 does not play a significant role in the movement of TDG on DNA, but due to binding lifetime

being shorter, a smaller amount of DNA will be sampled. The diffusion coefficient for R275A ($1.3x10^{-1}$ $\mu m^2/s$) and R275L ($6.5x10^{-2}$ $\mu m^2/s$) was significantly faster than both wild type and N140A variant (Fig. 3A). While this faster diffusion increases the amount of DNA that is sampled with each binding event, it may be that the rapid search is less effective and that TDG is more likely to bypass lesions, similar to behavior observed with other glycosylases where wedge residues are mutated [17].

Discussion

In this study, we utilized single-molecule methods to determine how TDG searches for DNA damage as well as the role for key active site residues in this search process. These single-molecule experiments have shown that TDG is able to form stable interactions with the undamaged DNA to facilitate its search. Despite the ongoing 1D search, TDG primarily binds its damage sites directly out of solution in a 3D diffusion mechanism, with a limited number of events that bound and briefly slide into the lesion. 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. Furthermore, we found that TDG bound specific lesions sites much longer than undamaged DNA. Examining TDG under different 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 glycosylases [24-26].

Having the ability to switch between sliding and hopping leads to the question of what conditions favor TDG hopping versus sliding on DNA. If TDG has a sequence preference, when it encounters the sequence, it may adopt a confirmation in which tight binding resulting in direct sliding as TDG further investigates that region of DNA. While preferred sequences by TDG Is not currently known, it is possible that sequences enriched in C and G nucleotides could be likely candidates for initiating sliding behaviors, because these regions would have a heavier burden of DNA damage. This is because, firstly, an enrichment of CG dinucleotides would generate more opportunities for damage to form, secondly, TET enzymes localize to C-G rich

CpG islands, oxidizing 5-methylC groups for their removal by TDG and subsequent transcriptional activation [27]. The ability of TDG to switch between sliding and hopping may also be advantageous in situations where TDG encounters other factors on the DNA, including other transcription factors or nucleosomes. With a sliding-exclusive method, full dissociation would need to occur in order to sample the DNA after the blockage, but hopping would allow TDG to circumvent the blockage and continue scanning on the other side of it. It has been previously determined the glycosylase OGG1 slides along the DNA in search of a lesion [24], whereas UNG, a uracil DNA glycosylase family member, performs 3D diffusion, short range sliding, and hopping to find a lesion [25]. Thus, TDG which exhibits multiple modes of diffusion 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.

We have also shown that N140A, R275A, and R275L variants of TDG exhibited a different behavior on DNA than the wild type. 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 [15]. On the other hand, R275A/L variants both exhibited a shorter dwell time as well as a faster diffusion. It is possible that the lack of the positive charge causes TDG to have lower binding energy 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 mutated to alanine, then faster diffusivity was observed indicating that efficient recognition of a base is dependent on a single active site residue [17]. 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 contributes to efficient damage detection and repair.

Acknowledgements

This work was supported by NIH R35ES031638 (BVH), T32GM088119 (BLS), the Hillman

Postdoctoral Fellowship for Innovative Cancer Research and F32ES034982 (MAS), 2P30CA047904 to the UPMC Hillman Cancer Center, and the major equipment grant S10OD032158-01A1 (BVH).

Competing interests

None declared.

Figure Legends

Figure 1: TDG binds DNA specifically and non-specifically.

(A) Cartoon schematic showing how 5fC was incorporated into nick translated λ DNA. (B) A diagram depicting the order reagents, which are under laminar flow, are captured in the flowcell. (C) 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. (D) CRTD analysis fit to a two-phase decay of TDG binding DNA containing 5fC specifically (n = 70) and nonspecifically (n = 487). (E) A cartoon depiction of the undamaged DNA substrate with an example kymograph of TDG binding and moving. (F) CRTD analysis fit to a one-phase decay of TDG binding undamaged λ DNA (n = 155). (G) An example kymograph of TDG binding to a single 5fC. (H) CRTD analysis fit to a two-phase decay of TDG binding to 5fC (n = 28).

Figure 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 undamaged λ. Dashed line, D_{lim}, theoretical limit to free diffusion for TDG-HaloTag. ****$p < 0.0001$ by two-way ANOVA. (B) An example kymograph of the two-color TDG experiment. White arrows indicate bypass events and yellow arrows indicate collision events.

Figure 3: TDG catalytic variants indicate R275 is essential for base detection.

(A) CRTD analysis fit to a one-phase decay of N140A TDG binding to undamaged (n = 134), specifically to 5fC (n = 30) and nonspecifically (n = 168). (B) CRTD analysis fit to a one-phase decay of R275A TDG binding to undamaged (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

undamaged (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 undamaged λ. Dashed line, D_{lim}, theoretical limit to free diffusion for TDG-HaloTag. ****$p < 0.0001$ by two-way ANOVA.

References

1. 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).
2. 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**, 15813-22 (2013).
3. 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).
4. Pidugu, L.S. et al. Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase. *Biochemistry* **55**, 6205-6208 (2016).
5. Cortázar, D. et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419-23 (2011).
6. Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67-79 (2011).
7. Bennett, M.T. et al. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J Am Chem Soc* **128**, 12510-9 (2006).
8. 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**, 27578-86 (2007).
9. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3 (2011).
10. 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**, 831-3 (2012).
11. DeNizio, J.E. et al. TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites. *J Mol Biol* **433**, 166877 (2021).
12. 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).
13. 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**, 2319-29 (2011).
14. 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).
15. Schaich, M.A. et al. Single-molecule analysis of DNA-binding proteins from nuclear extracts (SMADNE). *Nucleic Acids Res* **51**, e39 (2023).
16. 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).
17. 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).
18. 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).
19. 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).
20. 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).

21. 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).
22. 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).
23. 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).
24. 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).
25. 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).
26. Friedman, J.I. & Stivers, J.T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957-67 (2010).
27. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep* **13**, 28-35 (2011).

Figure 1

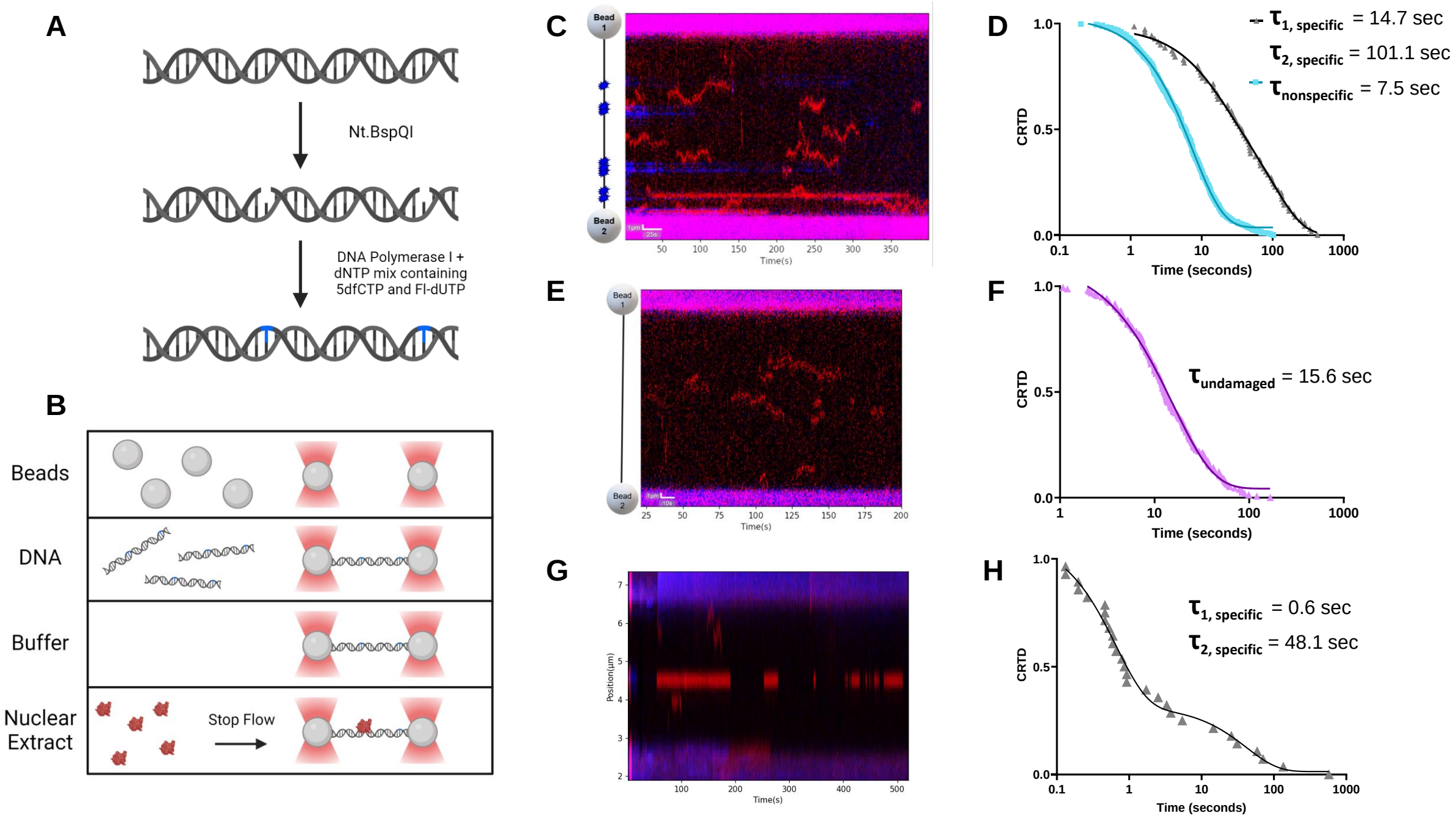

A
Nt.BspQI
DNA Polymerase I +
dNTP mix containing
5dfCTP and Fl-dUTP
B
Beads
DNA
Buffer
Nuclear
Extract
Stop Flow
C
Bead 1
Bead 2
Time(s)
D
$\tau_{1, specific}$ = 14.7 sec
$\tau_{2, specific}$ = 101.1 sec
$\tau_{nonspecific}$ = 7.5 sec
CRTD
Time (seconds)
E
Bead 1
Bead 2
Time(s)
F
$\tau_{undamaged}$ = 15.6 sec
CRTD
Time (seconds)
G
Position(µm)
Time(s)
H
$\tau_{1, specific}$ = 0.6 sec
$\tau_{2, specific}$ = 48.1 sec
CRTD
Time (seconds)

Figure 2

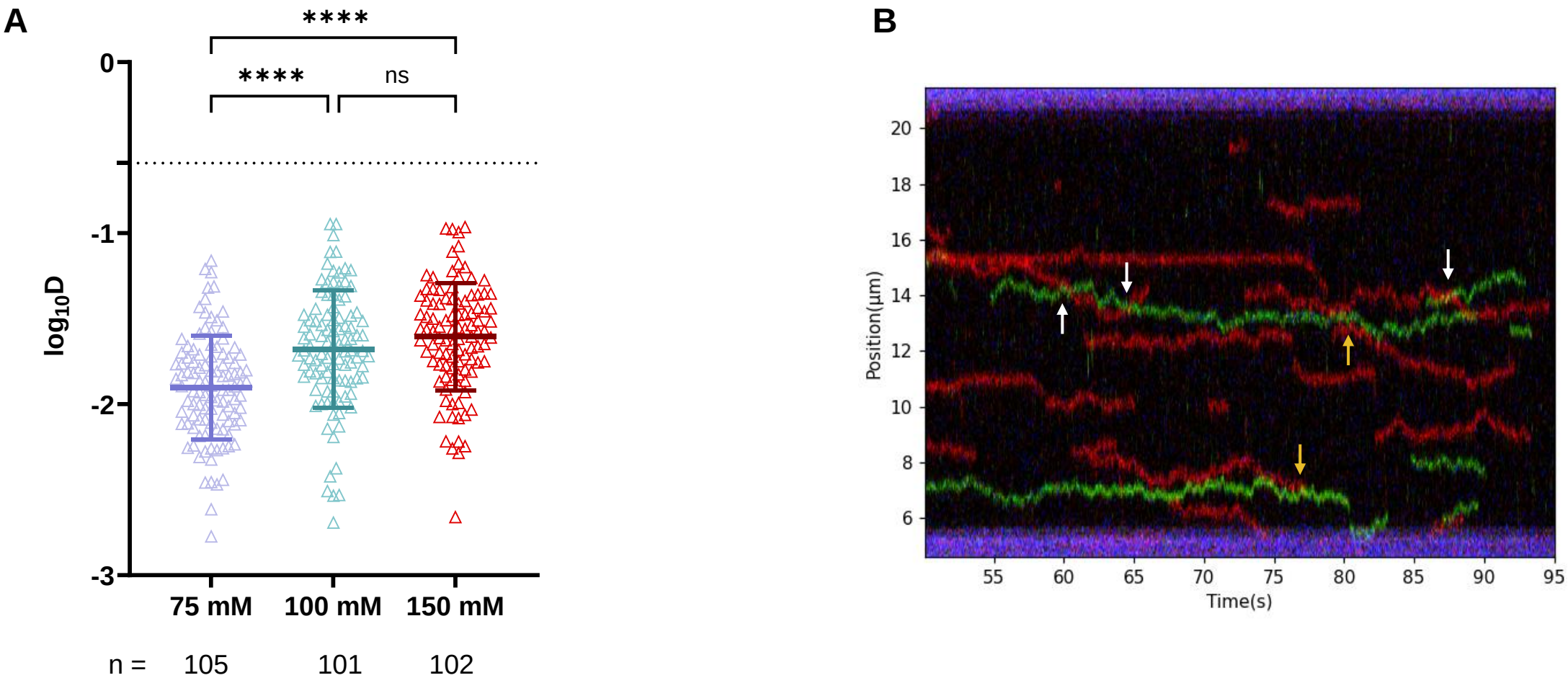
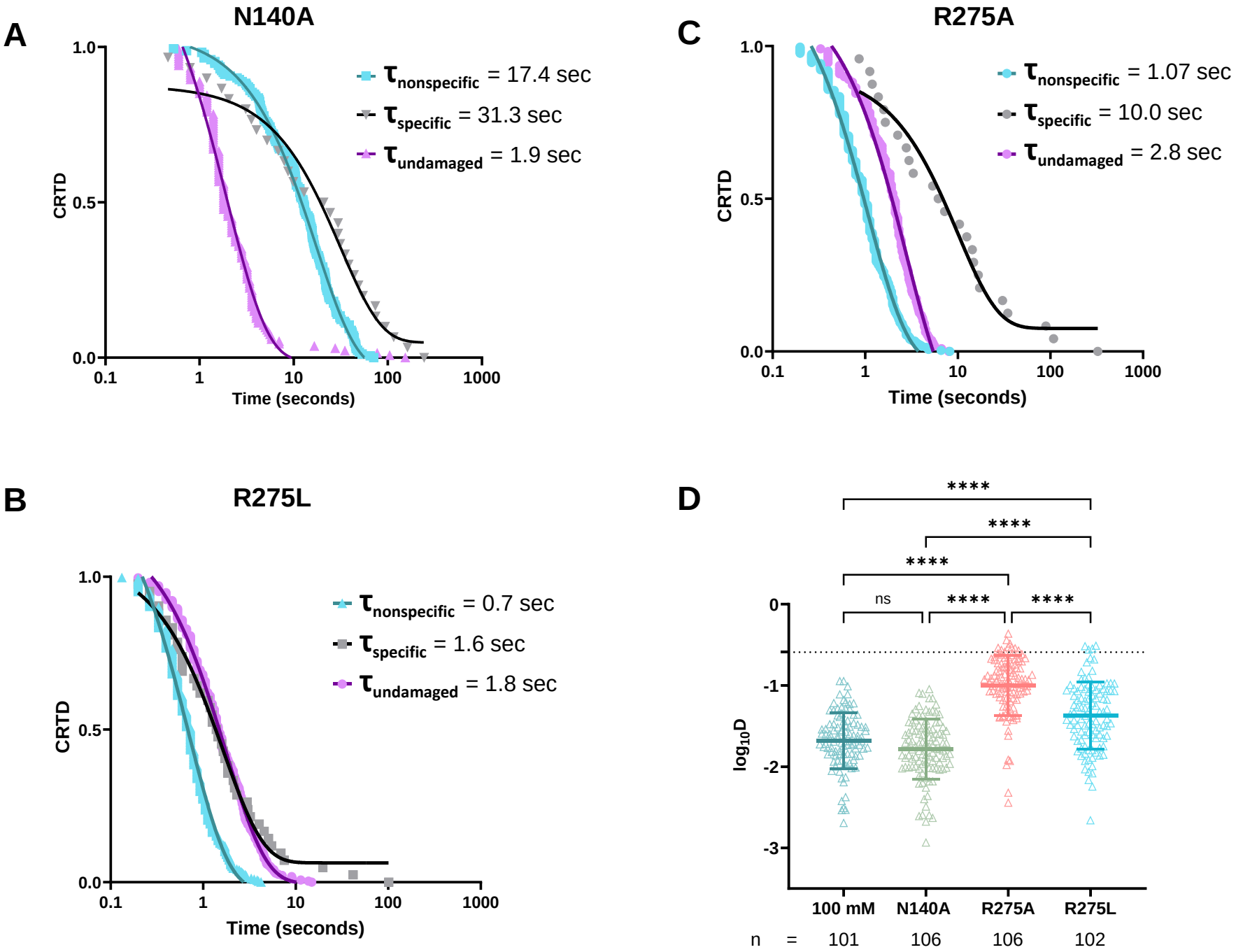

A
****
****
ns
0
-1
-2
-3
log10D
75 mM
100 mM
150 mM
n =
105
101
102
B
Position(µm)
20
18
16
14
12
10
8
6
55
60
65
70
75
80
85
90
95
Time(s)

Figure 3


A
N140A
CRTD
Time (seconds)
τnonspecific = 17.4 sec
τspecific = 31.3 sec
τundamaged = 1.9 sec
C
R275A
τnonspecific = 1.07 sec
τspecific = 10.0 sec
τundamaged = 2.8 sec
B
R275L
τnonspecific = 0.7 sec
τspecific = 1.6 sec
τundamaged = 1.8 sec
D
log10D
ns
****
100 mM
N140A
R275A
R275L
n = 101 106 106 102