Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG

Thomas Dodd[a], Chunli Yan[a], Bradley R. Kossmann[a,1], Kurt Martin[a], and Ivaylo Ivanov[a,2]

[a]Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302

Edited by Michael L. Klein, Temple University, Philadelphia, PA, and approved April 27, 2018 (received for review February 23, 2018)

Thymine DNA glycosylase (TDG) is a pivotal enzyme with dual roles in both genome maintenance and epigenetic regulation. TDG is involved in cytosine demethylation at CpG sites in DNA. Here we have used molecular modeling to delineate the lesion search and DNA base interrogation mechanisms of TDG. First, we examined the capacity of TDG to interrogate not only DNA substrates with 5-carboxyl cytosine modifications but also G:T mismatches and nonmismatched (A:T) base pairs using classical and accelerated molecular dynamics. To determine the kinetics, we constructed Markov state models. Base interrogation was found to be highly stochastic and proceeded through insertion of an arginine-containing loop into the DNA minor groove to transiently disrupt Watson–Crick pairing. Next, we employed chain-of-replicas path-sampling methodologies to compute minimum free energy paths for TDG base extrusion. We identified the key intermediates imparting selectivity and determined effective free energy profiles for the lesion search and base extrusion into the TDG active site. Our results show that DNA sculpting, dynamic glycosylase interactions, and stabilizing contacts collectively provide a powerful mechanism for the detection and discrimination of modified bases and epigenetic marks in DNA.

molecular dynamics | Markov state models | genome maintenance | epigenetics | DNA glycosylase

Genome maintenance occurs in the context of chromatin and it is becoming increasingly apparent that epigenetic regulation is intricately intertwined with the DNA damage response in ensuring genome stability. Understanding how epigenetic marks are recognized, distinguished from exogenous or endogenous DNA lesions, and processed by the canonical DNA repair machinery is a topic of great current interest. Here our focus is on the base excision repair (BER) pathway, which in addition to an established role in genome maintenance, is associated with many other cellular processes (1), including a recently discovered critical role in epigenetic regulation (2–5). The most prominent epigenetic modification in mammalian genomes is cytosine methylation, which typically occurs at CpG islands and enhances chromatin packing to promote gene silencing (6). Consequently, 5-methylcytosine (5mC) demethylation is crucial for resuming the transcription of silenced genes. Notably, unbalanced cytosine methylation is a hallmark of cancer (7–9). In cancer, predominantly demethylated regions of the genome could become hypermethylated leading to the silencing of tumor suppressor genes. Furthermore, 5-methylcytosine deamination results in G:T mismatches that could cause C-to-T transition mutations during DNA replication. It is estimated that nearly a third of cancer mutations found in coding regions of the genome arise from C and 5mC deamination at CpG sites (2). There is also a clear link between aging and methylation levels in GpG islands (10). The importance of maintaining the methylation state of the genome requires tight regulation of pathways controlling the levels of 5mC. Removal of 5mC bases (Fig. 1) is known to proceed through successive steps of oxidation by enzymes from the ten-eleven-translocase (TET) family, producing 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) intermediates (5, 11, 12). Unlike methylcytosine, these intermediates are substrates for thymine DNA glycosylase (5, 13, 14), a classic DNA repair enzyme. While thymine DNA glycosylase (TDG) is important for the repair of mutagenic DNA lesions, it has an even more prominent role in ensuring epigenetic stability. In this capacity, TDG activity is vital during embryonic development (15). TDG also interacts with numerous protein partners engaged in epigenetic regulation [e.g., DNMT3a (16) and CBP/p300 acetylase (17)] and transcription [transcription factors, nuclear receptors (18)] and is intricately involved in the regulation of gene expression.

A second pathway to process 5mC is through deamination followed by the action of MBD1–4 glycosylases (19, 20). The resulting abasic DNA is then channeled through the BER pathway. BER efficiency relies on a remarkably discriminating search for modified bases among an enormous background of normal DNA. The search is followed by damage-specific base extrusion into the enzyme's active site, removal of the damaged bases, and handoff of the product DNA to downstream pathway participants.

Here we establish a basis to understand the key principles underpinning the extraordinary power of the TDG glycosylase to discriminate in favor of modified bases against a backdrop of normal genomic DNA. We further elucidate the protein–nucleic acid interactions ensuring specificity for lesions or epigenetic marks. Key to selectivity is nucleotide extrusion, a process involving a nucleotide swinging out of the DNA helix and being accommodated in the catalytic pocket of TDG. Nucleotide extrusion (21–23) is a major determinant of glycosylase selectivity,

Significance

The most prominent epigenetic modification in mammalian genomes is cytosine methylation at position 5 on the pyrimidine ring. Thymine DNA glycosylase (TDG) plays a central role in the pathways for 5-methyl cytosine removal and thus influences gene silencing, stem cell differentiation, and alterations in normal development. Additionally, methylation abnormalities in DNA are often observed in diseases, specifically cancer. Here we examine the mechanisms by which TDG detects, extrudes, and excises modified bases in DNA. Using path-sampling methodologies, we compute minimum free energy paths for TDG base extrusion. The computed paths reveal a unique mechanism underpinning TDG selectivity for DNA lesions or modified bases, which involves DNA sculpting, global protein dynamics, conformational gating, and specific protein–nucleic acid interactions.

Author contributions: I.I. designed research; T.D., C.Y., B.R.K., K.M., and I.I. performed research; T.D., C.Y., B.R.K., K.M., and I.I. analyzed data; and T.D. and I.I. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Present address: 360i, Atlanta, GA 30309.

[2]To whom correspondence should be addressed. Email: iivanov@gsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1803323115/-/DCSupplemental.

Published online May 21, 2018.

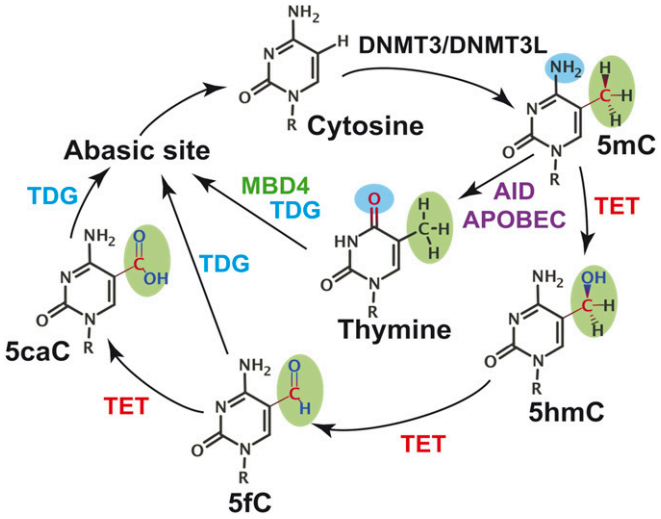

Fig. 1. Schematic representation of the two active demethylation pathways known in mammals. MBD4-mediated pathway is shown with blue arrows; the TET-mediated pathway is shown with red arrows.

with potential for selection or rejection of substrates at each intermediate along the base eversion path. Glycosylases (22, 24) also employ DNA sculpting strategies (e.g., DNA bending and loop insertion) to lower the energetic barrier of base extrusion and thus increase the efficiency of the dynamic lesion search. Whether glycosylases employ active or passive strategies in this search process has been a topic of considerable debate. NMR evidence has suggested glycosylases could act as passive kinetic traps for spontaneously exposed extrahelical bases (25, 26). Conversely, evidence from molecular crystallography (MX) has pointed to active base extrusion mechanisms. Numerous glycosylase structures (20, 22, 27–29) have shown that DNA binding is accompanied by a multitude of conformational changes preceding active site chemistry: (*i*) DNA sculpting through interactions with the enzyme DNA-binding groove; (*ii*) DNA bending, minor groove compression, and backbone distortion at the lesion site; (*iii*) residue insertion into the DNA stack to expel the lesion base and stabilize the orphaned base; and (*iv*) base flipping into lesion-specific recognition pockets that sterically exclude nonlesion bases. Cross-linking strategies have, in rare instances, captured crystallographic snapshots of base extrusion intermediates (30) and could, in principle, provide insight into short-lived species along base extrusion paths. Nonetheless, base flipping is inherently dynamic and therefore not easily construed from static crystallographic snapshots. Therefore, molecular modeling studies have been extensively used to complement structural biology approaches and have proven enormously valuable in unraveling detailed dynamics of glycosylase enzymes and the origins of selectivity in BER (23, 29, 31).

Results and Discussion

To explore whether TDG employs an active or passive mechanism, we carried out simulations on TDG/5caC-DNA complexes. As starting points for computational modeling, we utilized existing structures of TDG/5caC-DNA in a preextrusion and postexcision state [Protein Data Bank (PDB) ID codes: 2RBA and 5HF7] (32, 33). The following initial models were created: (*i*) preextrusion state (5caC accommodated in the DNA base stack); (*ii*) fully extruded state (5caC inserted in the TDG active site); and (*iii*) an initial interrogation complex. To address base interrogation, we started from systems with initially separated TDG and 5caC-DNA and simulated complex formation. We also simulated TDG in the presence of a G:T mismatch and with normal DNA.

Our first goal was to delineate the accessible conformational space for TDG/5caC-DNA and to assess the capacity of TDG to interrogate not only 5caC but also G:T mismatches and nonmismatched (A:T) base pairs. To this end, we carried out accelerated molecular dynamics (aMD) (34) runs on the TDG interrogation complexes. The aMD method enhances sampling of the torsional degrees of freedom to accelerate phase space exploration and facilitate transitions over high energy barriers. Surprisingly, we observed that the presence of TDG induces multiple transient base-opening events that occur within 200 ns of aMD sampling (*SI Appendix*, Fig. S1*A*). Base interrogation was found to be highly stochastic and appeared to proceed through insertion of an arginine-containing loop (Arg275) into the DNA minor groove to transiently disrupt Watson–Crick pairing. To ascertain that these events are also detectable in

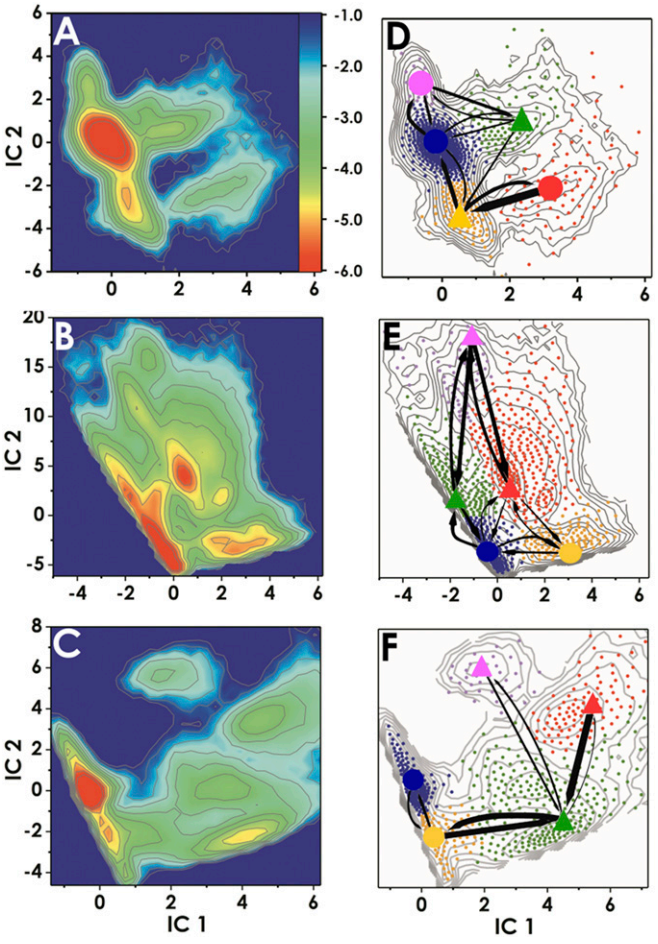

Fig. 2. Conformational dynamics of TDG/DNA complexes during base interrogation. Computed free energy profiles projected onto the first two ICs for (*A*) 5caC-DNA (*B*) G:T mismatch, and (*C*) A:T base pairs. The color bar *Inset* denotes the ΔG scale in kcal/mol. Results from MSM analysis for (*D*) 5caC-DNA, (*E*) G:T mismatch, and (*F*) A:T base pair systems. *D–F* show the positions of all microstate clusters in the space of the two ICs (small dots); how these small clusters were agglomerated into macrostates (denoted by large dots for the intrahelical or triangles for the extrahelical states) and the probabilities of transitions between macrostates. Microstates (dots) are colored by the macrostate they belong to. To allow easy comparison of the three simulation systems, corresponding states are denoted with the same color. The relative thickness of the arrows connecting the macrostates denotes the macrostate transition probabilities computed using transition path theory.

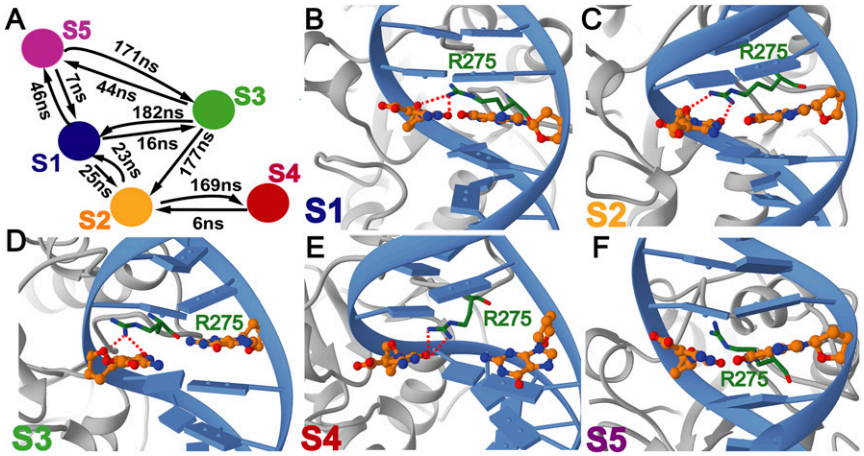

Fig. 3. Representative structures selected from each macrostate of the Markov state model corresponding to 5caC base interrogation by TDG. (*A*) Transition timescales determined from transition path theory; representative structures of (*B*) macrostate S1; (*C*) macrostate S2; (*D*) macrostate S3; (*E*) macrostate S4; and (*F*) macrostate S5. Each state is colored according to the color scheme in Fig. 2*D*; panels are labeled by macrostate designation. TDG is shown in gray; DNA is shown in blue. The intercalating Arg275 residue at the tip of the insertion loop is shown in ball and stick representation and colored in green. The extruded 5caC base and the orphaned base are shown in ball and stick representation and colored in orange. The calculated transition timescales determined from transition path theory are included in *A*.

unbiased MD (*SI Appendix*, Fig. S1*B*), we carried out multiple trajectory regular MD runs for an aggregate simulation time of 8 μs. Analogous runs were performed on the G:T mismatch and unmodified DNA systems. DNA backbone torsion angles for the interrogated base pair and distances between the two bases and the guanidinium group of Arg275 were selected as coordinates for time-lagged independent component analysis (TICA) (35, 36). The combined trajectories were projected onto the first two independent components (ICs). Different energy minima (metastable states) are present and clearly separated in the TICA projections (Fig. 2 *A*–*C*). All trajectory frames were then clustered in the projected space of the two ICs using the *k*-means algorithm, producing 800 clusters (i.e., microstates). From this data we constructed Markov state models (MSMs) (37, 38) and evaluated the kinetics of TDG base interrogation using transition path theory (39). Results are presented in Fig. 2 *D*–*F*. Several conclusions are immediately apparent from our analysis. First, TDG probes DNA bases nonspecifically, interrogating not only 5caC-modified bases but also G:T mismatches and normal base pairs. Among the identified kinetically distinct macrostates we distinguish two low-populated extrahelical states with Arg275 inserted into the DNA stack in two different orientations (Fig. 3 and *SI Appendix*, Fig. S3). These states are accessed through two kinetic intermediates: (*i*) an intermediate with TDG-induced local torsional shift and intact Watson–Crick pairing; and (*ii*) an intermediate with partially broken Watson–Crick pairing and Arg275 inserted between the extruded and the orphaned base (Fig. 3 *C* and *D*). The second key observation is that the extrahelical states are extremely short lived, and thus not readily detectable by NMR (Fig. 3*A* and *SI Appendix*, Figs. S2*A* and S3*A*). The free energy landscapes in Fig. 2 reflect this, with barriers to the extrahelical states not exceeding 4 kcal/mol. Third, the H bonding between Arg275 and 5caC is variable and differs from the pattern for normal DNA (Fig. 3 and *SI Appendix*, Fig. S3). This observation rationalizes the differences between the free energy landscapes in Fig. 2 *A* and *C* with a lower barrier for a 5caC modified base to access the extrahelical state. We also analyzed DNA structural parameters using the Curves+ code (40) to determine whether TDG exploits local DNA deformation to facilitate lesion interrogation and selection. We found very little difference between the interbase parameters as well as the total bend of the 5caC and the A:T intrahelical states (*SI Appendix*, Fig. S4*A* and Table S5). However, analysis of the transient extrahelical states revealed changes to both the shift and tilt values (~2 Å and ~6.5°) at the interrogation site and flanking base of 5caC (*SI Appendix*, Fig.

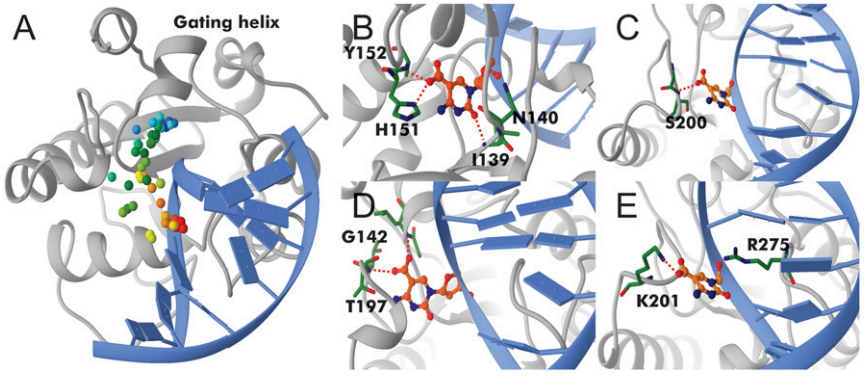

Fig. 4. Optimized base extrusion path for 5caC in TDG. (*A*) The path is indicated by colored dots tracking the position of the carbon atom of the 5caC carboxylate group along the path. Color denotes the replica index from initial (red) to final (blue). Four snapshots along the path are shown. (*B*) Early intermediate with Arg275 intercalating into the DNA; (*C*) 5caC at the Pro198 loop; (*D*) 5caC positioned next to the gating helix; and (*E*) 5caC extruded into the TDG active site.

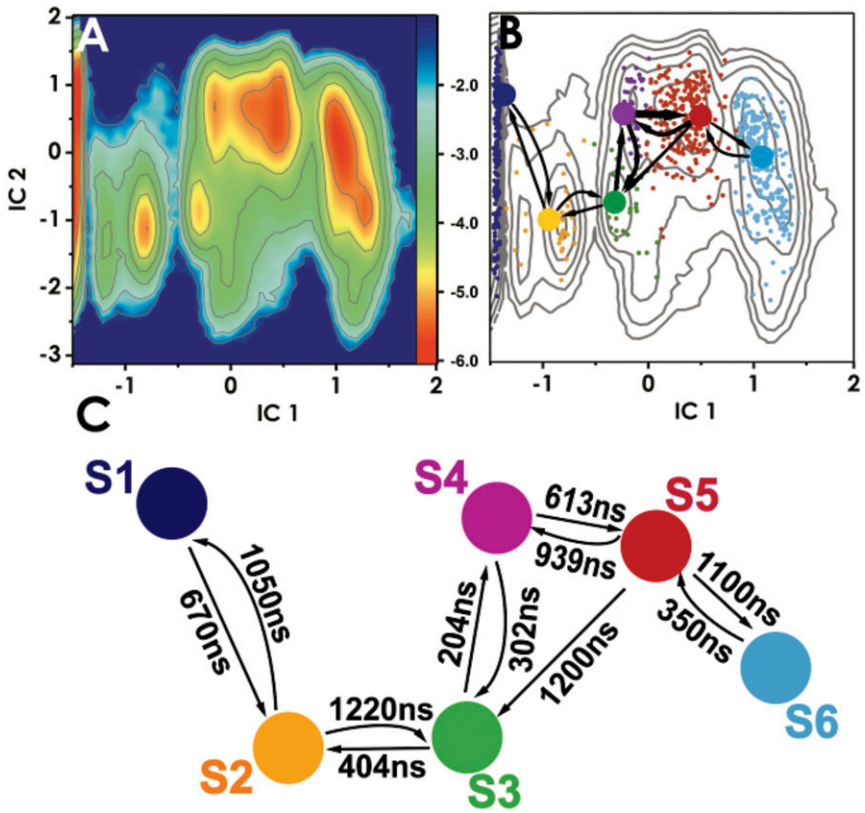

Fig. 5. Conformational dynamics of TDG/5caC-DNA complex during base eversion. (*A*) Free energy profile projected onto the first two ICs. Color bar *Inset* denotes ΔG scale in kcal/mol. (*B*) Results from MSM analysis. Microstates (dots) are colored by macrostate they belong to. Probability fluxes between macrostates from transition path theory are shown by arrows. (*C*) Calculated macrostate transition timescales.

S4*B*). Importantly, the negative charge on 5caC provides a convenient handle for TDG to stabilize an extrahelical intermediate using an opportunely positioned Lys201 (Fig. 4*E*).

G:T mismatches could also rapidly transition between nonextruded and extruded states (Fig. 2*E*), exhibiting two well-defined metastable states with disrupted base pairing. The H-bonding pattern of Arg275 in the observed extrahelical states also differs from the pattern for normal DNA (*SI Appendix*, Figs. S2 and S3). This can be rationalized by the fact that G:T mismatches form "wobble" hydrogen-bonding pairs, which require a sideways shift of one base relative to Watson–Crick positioning. Our structural analysis of the interrogation site confirms this, with intrahelical shift and twist values differing significantly (by ∼2 Å and ∼5°, respectively) from the ones measured for the intrahelical states of the 5caC and A:T (*SI Appendix*, Fig. S4*A*). This leads to increased propensity for bending and base pair disruption at the G:T site (41–44), which TDG takes advantage of through backbone distortion alone. Thus, TDG has shown stronger G:T mismatch repair activity, in vitro, compared with modified substrates (45). Interestingly, Curves+ analysis of the interrogation states for the G:T mismatch system indicates that base stacking is disrupted not only for the transient extrahelical macrostates but also for the intrahelical basin that contains the majority of conformers from the unbiased MD trajectories. Thus, unlike the regular (A:T) DNA or 5caC, the G:T mismatch has a local structural distortion at the very outset of interrogation. This leads to an energetically destabilized initial state and lower energetic cost for base extrusion in the G:T mismatch system. These results substantiate the differences in the transition timescales, with G:T base extrusion occurring on a faster timescale than both the 5caC and A:T substrates (Fig. 3 and *SI Appendix*, Figs. S2*A* and S3*A*).

The results also highlight the role of the TDG insertion loop (Ala270–Pro280) and particularly Arg275, which stabilizes the extrahelical intermediates by stepwise replacement of the Watson–Crick hydrogen bonds. Several glycosylases have been proposed to utilize similar mechanisms in which both DNA sculpting and loop insertion is exploited. Structures of hOGG1, its bacterial homolog MutM, and MBD4 all exhibit Arg-loop insertion (20, 29, 46, 47). Similarly, single molecule experiments have shown the *Escherichia coli* repair enzymes Fpg, Nei, and Nth to utilize DNA sculpting and intercalating loop strategies to interrogate and extrude damage bases (48, 49).

As the next step in our analysis, we determined the complete base eversion path for TDG/5caC-DNA, starting with the interrogation complexes and ending with the fully extruded state. We identified the key intermediates imparting selectivity and also computed an effective free energy landscape for this transition. Our results show that base eversion in TDG is a gated process that involves motions of several flexible loops and a gating helix (Fig. 4). Therefore, intuitive reaction coordinates (e.g., pseudo torsions) are, in this case, not practical. Recently, there has been considerable progress in methods (50–52) to optimize minimum energy paths (MEPs) when the initial and final states are known. We leveraged two of these methods, the partial nudged elastic band (PNEB) (53, 54) and finite temperature string method (55, 56), to investigate recognition of modified bases by TDG. Both methods define the MEP as a chain of replicas of the system connecting the initial and final configurations. First, we optimized a MEP between the preextrusion and the fully extruded states using PNEB. Gradually spreading the replicas from these two states allowed the optimization process to discover the path in an unbiased way. The PNEB optimized path (Movie S1) served as a starting point for further optimization with the finite temperature string method, which could provide more extensive path sampling. Since the string method works in projected collective variable (CV) space, a preliminary PNEB step was necessary to provide an unbiased initial path and to select CVs for the string method. Using this protocol for TDG/5caC-DNA, we completed 25 ns of PNEB optimization

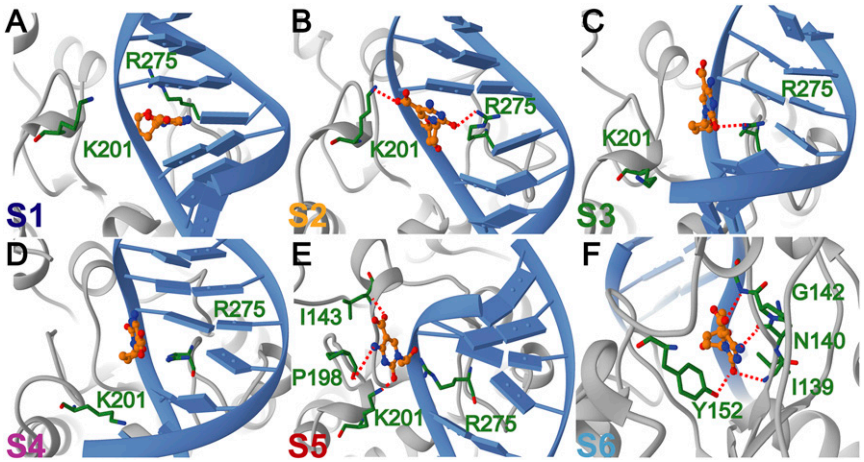

Fig. 6. Representative structures selected from each macrostate along the optimized TDG base eversion path for 5caC DNA. Representative structures of (*A*) macrostate S1; (*B*) macrostate S2; (*C*) macrostate S3; (*D*) macrostate S4; (*E*) macrostate S5; and (*F*) macrostate S6. Each state is colored according to the color scheme in Fig. 5*B*; panels are labeled by macrostate designation. TDG is shown in gray; DNA is shown in blue. Residues contacting the extruded base are explicitly shown and labeled in green. The 5caC base is shown in ball and stick representation and colored in orange. Hydrogen bonds to the extruded base are denoted as red dash lines.

(with 28 replicas) and 224 ns (200 iterations) of the string method (Fig. 4). After MEP convergence, we released each replica and sampled an aggregate of 11.2 μs of unrestrained MD trajectories along the base eversion path, which we further analyzed to construct an MSM.

Our results reveal an intricate network of protein–DNA contacts necessary to accommodate the 5caC base during its passage from the DNA base stack into the TDG active site. Importantly, these contacts significantly lower the free energy barriers for base extrusion to ∼4 kcal/mol (Fig. 5*A*). By comparison, umbrella sampling simulations of base eversion in the absence of the glycosylase result in barriers of at least 12 kcal/mol (*SI Appendix*, Fig. S5), which is consistent with previously published values for base extrusion barriers in DNA (21–23). From the MSM analysis we identify six kinetically distinct states along the 5caC eversion path (Figs. 5 *B* and *C* and 6). State S1 has the modified base accommodated in the stack; state S2 is an early intermediate wherein 5caC is inserted between the intercalating Arg275 and Lys201. The positive charge on the lysine stabilizes the negative charge on the 5caC carboxyl group. States S3–S5 correspond to configurations wherein 5caC interacts with residues of the Pro198 loop of TDG. Indeed, access to these three relatively rapidly interconverting states is gated by the motion of the Pro198 loop and the adjacent helix (Ser205–Lys221). The TDG gating helix (Leu143–Lys148) serves as a secondary gate by closing over the active site after the 5caC base is inserted. Thus, base eversion by TDG is a global conformational transition and conformational gating is necessary for the 5caC base to access the active site.

Conclusions

Our computational modeling reveals a mechanism underpinning TDG selectivity for DNA lesions (G:T mismatches) or modified bases (e.g., 5caC), which involves DNA sculpting, global protein dynamics, conformational gating, and specific protein–nucleic acid interactions that stabilize the extruded base along the path from the DNA stack to the TDG active site. Our model for base extrusion by TDG bears certain similarities to an earlier proposed mechanism ("pinch-push-pull") for human UNG (57–60). In this model, the pinch involved compression of the DNA backbone such that the distances between the phosphates flanking the uracil base were reduced by ∼4 Å. Three static enzyme loops were proposed to mediate DNA recognition: the minor groove reading loop (His268–Ser273), the Prorich loop (Pro165–Pro168) and the Gly–Ser loop (Gly246–Ser247). Nucleotide flipping was proposed to be facilitated by the intercalation (push) of Leu272 into the DNA base stack. The final step was the pulling of the uracil base and ribose ring deep into the uracil recognition pocket, resulting in hydrogen bonding to every polar atom of the uracil and in face-to-face π-stacking with Phe158 and Tyr147. In a similar scenario for TDG, the pinch step is achieved by DNA sculpting via protein–DNA interactions and dynamic Arg275-loop insertion. This leads to kinking of the DNA substrate and compression of the distance between the flanking phosphates above and below the extrusion site by up to 3 Å in the extrusion intermediates (*SI Appendix*, Table S1). The push involves insertion of an interrogation loop into the DNA minor groove, intercalation of an arginine (Arg275) from the tip of the interrogation loop into the DNA stack, and stepwise replacement of Watson–Crick H bonds to lower the energetic barrier for base flipping. Finally, the pull step is achieved by accommodation of the extrahelical base via specific residue interactions in four stable intermediates along the extrusion path (Fig. 6). However, there are also important differences with the previous model. First, unlike Leu272 in UNG, which plays the role of a steric plug, Arg275 has the capacity to actively disrupt Watson–Crick hydrogen bonding. Stepwise replacement of Watson–Crick H bonds between the extruded and orphaned base lowers the barrier to reach the most populated extrahelical state during interrogation. Second, the pinch-push-pull model originated from molecular crystallography and emphasized the role of static enzyme loops. By contrast, we show that protein dynamics and global gating motions of TDG are essential. Specifically, transitioning the 5caC base into the active site requires gating motions of the Pro198 loop and the adjacent helix (Ser205–Lys221) as well as motions of the TDG gating helix (Leu143–Lys148). These motions cannot be easily construed from static crystal structures. Collectively, our results shed light on the key determinants of glycosylase selectivity and uncover universal rules governing this class of enzymes.

Materials and Methods

Models for the pre- and postextrusion states were constructed from two TDG/DNA crystal structures (PDB ID codes: 5HF7 and 2RBA) (32, 33). To examine base interrogation, we built systems of initially separated TDG and DNA for 5caC:G, G:T, and A:T base pairs. All systems were then minimized, heated in the NVT ensemble, and then equilibrated in the NPT ensemble (1 atm and 300 K). One hundred nanoseconds of MD equilibration was followed by 200 ns of accelerated MD. Snapshots of the aMD trajectories were then replicated and released for an aggregate simulation time of 8 μs for each

system. Detailed simulation protocols can be found in *SI Appendix, Supplementary Methods*. The base eversion path for TDG/5caC was represented by 28 replicas. For path optimization, all heavy atoms of the TDG/5caC complex were included. We first performed PNEB optimization of the replicas for 25 ns with a simulated annealing protocol. We then further optimized the path using the finite-temperature string method with swarms of trajectories. PNEB/string method protocols and the definition of collective variables are provided as *SI Appendix*. After string convergence, each image underwent 200 ns of unrestrained MD for an aggregate simulation time of 11.2 μs. The combined trajectories were used for MSM analysis and computing the effective free energy profiles. TICA calculations were then performed on the systems and the trajectory frames were then clustered in projected space using the *k*-means algorithm. From the resultant clusters, MSMs were constructed and used to estimate probability fluxes, as well as transition timescales in and out of each PCCA+ macrostate. For the DNA structure analysis 1,000 frames were selected from each macrostate and analyzed using the Curves+. All TICA and MSM analysis details are provided in *SI Appendix, Supplementary Methods*.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant GM110387 and National Science Foundation Grant MCB-1149521. Computational resources were provided in part by allocations from the National Science Foundation's Extreme Science and Engineering Discovery Environment program CHE110042 and the National Energy Research Scientific Computing Center supported by the Department of Energy Office of Science Contract DE-AC02-05CH11231.

1. Zharkov DO (2008) Base excision DNA repair. *Cell Mol Life Sci* 65:1544–1565.
2. Bellacosa A, Drohat AC (2015) Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA Repair (Amst)* 32:33–42.
3. Liu MY, DeNizio JE, Schutsky EK, Kohli RM (2016) The expanding scope and impact of epigenetic cytosine modifications. *Curr Opin Chem Biol* 33:67–73.
4. Breiling A, Lyko F (2015) Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* 8:24.
5. Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502:472–479.
6. Eden S, Hashimshony T, Keshet I, Cedar H, Thorne AW (1998) DNA methylation models histone acetylation. *Nature* 394:842.
7. Jones PA, Laird PW (1999) Cancer epigenetics comes of age. *Nat Genet* 21:163–167.
8. Baylin SB, Jones PA (2011) A decade of exploring the cancer epigenome–Biological and translational implications. *Nat Rev Cancer* 11:726–734.
9. Issa JP (2004) CpG island methylator phenotype in cancer. *Nat Rev Cancer* 4:988–993.
10. Issa JP (2014) Aging and epigenetic drift: A vicious cycle. *J Clin Invest* 124:24–29.
11. Ito S, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–1303.
12. Pastor WA, Aravind L, Rao A (2013) TETonic shift: Biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* 14:341–356.
13. Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J Biol Chem* 286:35334–35338.
14. He YF, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303–1307.
15. Cortázar D, et al. (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* 470:419–423.
16. Li YQ, Zhou PZ, Zheng XD, Walsh CP, Xu GL (2007) Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res* 35:390–400.
17. Tini M, et al. (2002) Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol Cell* 9:265–277.
18. Um S, et al. (1998) Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J Biol Chem* 273:20728–20736.
19. Hashimoto H (2014) Structural and mutation studies of two DNA demethylation related glycosylases: MBD4 and TDG. *Biophysics (Nagoya-shi)* 10:63–68.
20. Hashimoto H, Zhang X, Cheng X (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: Structural basis and implications for active DNA demethylation. *Nucleic Acids Res* 40:8276–8284.
21. Banavali NK, MacKerell AD, Jr (2002) Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J Mol Biol* 319:141–160.
22. Hitomi K, Iwai S, Tainer JA (2007) The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst)* 6:410–428.
23. Priyakumar UD, MacKerell AD, Jr (2006) Computational approaches for investigating base flipping in oligonucleotides. *Chem Rev* 106:489–505.
24. Jacobs AL, Schär P (2012) DNA glycosylases: In DNA repair and beyond. *Chromosoma* 121:1–20.
25. Cao C, Jiang YL, Krosky DJ, Stivers JT (2006) The catalytic power of uracil DNA glycosylase in the opening of thymine base pairs. *J Am Chem Soc* 128:13034–13035.
26. Cao C, Jiang YL, Stivers JT, Song F (2004) Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat Struct Mol Biol* 11:1230–1236.
27. Bruner SD, Norman DP, Verdine GL (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* 403:859–866.
28. Fromme JC, Verdine GL (2003) DNA lesion recognition by the bacterial repair enzyme MutM. *J Biol Chem* 278:51543–51548.
29. Qi Y, et al. (2009) Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* 462:762–766.
30. Parker JB, et al. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 449:433–437.
31. Knips A, Zacharias M (2017) Both DNA global deformation and repair enzyme contacts mediate flipping of thymine dimer damage. *Sci Rep* 7:41324.
32. Maiti A, Morgan MT, Pozharski E, Drohat AC (2008) Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci USA* 105:8890–8895.
33. Coey CT, et al. (2016) Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res* 44:10248–10258.
34. Hamelberg D, de Oliveira CA, McCammon JA (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J Chem Phys* 127:155102.
35. Schwantes CR, Pande VS (2013) Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J Chem Theory Comput* 9:2000–2009.
36. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F (2013) Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139:015102.
37. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106:19011–19016.
38. Scherer MK, et al. (2015) PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J Chem Theory Comput* 11:5525–5542.
39. Metzner P, Schütte C, Vanden-Eijnden E (2006) Illustration of transition path theory on a collection of simple examples. *J Chem Phys* 125:084110.
40. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res* 37:5917–5929.
41. Hunter WN, et al. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. *J Biol Chem* 262:9962–9970.
42. Tsutakawa SE, Jingami H, Morikawa K (1999) Recognition of a TG mismatch: The crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* 99:615–623.
43. Lamers MH, et al. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* 407:711–717.
44. Obmolova G, Ban C, Hsieh P, Yang W (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 407:703–710.
45. Zhu B, et al. (2000) 5-methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* 28:4157–4165.
46. Qi Y, Spong MC, Nam K, Karplus M, Verdine GL (2010) Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM. *J Biol Chem* 285:1468–1478.
47. Chung SJ, Verdine GL (2004) Structures of end products resulting from lesion processing by a DNA glycosylase/lyase. *Chem Biol* 11:1643–1649.
48. Nelson SR, Dunn AR, Kathe SD, Warshaw DM, Wallace SS (2014) Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc Natl Acad Sci USA* 111:E2091–E2099.
49. Dunn AR, Kad NM, Nelson SR, Warshaw DM, Wallace SS (2011) Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res* 39:7487–7498.
50. Vanden-Eijnden E, Venturoli M (2009) Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J Chem Phys* 130:194103.
51. Májek P, Elber R (2010) Milestoning without a reaction coordinate. *J Chem Theory Comput* 6:1805–1817.
52. Mathews DH, Case DA (2006) Nudged elastic band calculation of minimal energy paths for the conformational change of a GG non-canonical pair. *J Mol Biol* 357:1683–1693.
53. Kossmann B, Ivanov I (2014) Alkylpurine glycosylase D employs DNA sculpting as a strategy to extrude and excise damaged bases. *PLoS Comput Biol* 10:e1003704.
54. Bergonzo C, Campbell AJ, Walker RC, Simmerling C (2009) A partial nudged elastic band implementation for use with large or explicitly solvated systems. *Int J Quantum Chem* 109:3781–3790.
55. Pan AC, Sezer D, Roux B (2008) Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B* 112:3432–3440.
56. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J Chem Phys* 125:24106.
57. Slupphaug G, et al. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* 384:87–92.
58. Parikh SS, et al. (2000) Uracil-dna glycosylase-DNA substrate and product structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc Natl Acad Sci USA* 97:5083–5088.
59. Parikh SS, et al. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J* 17:5214–5226.
60. Mol CD, et al. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: Protein mimicry of DNA. *Cell* 82:701–708.