Target search and recognition mechanisms of glycosylase AlkD revealed by scanning FRET-FCS and Markov state models

Sijia Peng\textsuperscript{a,1}, Xiaowei Wang\textsuperscript{b,1}, Lu Zhang\textsuperscript{c,e,1,2}, Shanshan He\textsuperscript{d}, Xin Sheng Zhao\textsuperscript{d,2}, Xuheui Huang\textsuperscript{b,2,3}, and Chunlai Chen\textsuperscript{a,2}\textsuperscript{e,2}

DNA glycosylase is responsible for repairing DNA damage to maintain the genome stability and integrity. However, how glycosylase can efficiently and accurately recognize DNA lesions across the enormous DNA genome remains elusive. It has been hypothesized that glycosylase translocates along the DNA by alternating between a fast but low-accuracy diffusion mode and a slow but high-accuracy mode when searching for DNA lesions. However, the slow mode has not been successfully characterized due to the limitation in the spatial and temporal resolutions of current experimental techniques. Using a newly developed scanning fluorescence resonance energy transfer (FRET)–fluorescence correlation spectroscopy (FCS) platform, we were able to observe both slow and fast modes of glycosylase AlkD translocating on double-stranded DNA (dsDNA), reaching the temporal resolution of microsecond and spatial resolution of subnanometer. The underlying molecular mechanism of the slow mode was further elucidated by Markov state model built from extensive all-atom molecular dynamics simulations. We found that in the slow mode, AlkD follows an asymmetric diffusion pathway, i.e., rotation followed by translation. Furthermore, the essential role of Y27 in AlkD diffusion dynamics was identified both experimentally and computationally. Our results provided mechanistic insights on how conformational dynamics of AlkD–dsDNA complex coordinate different diffusion modes to accomplish the search for DNA lesions with high efficiency and accuracy. We anticipate that the mechanism adopted by AlkD to search for DNA lesions could be a general one utilized by other glycosylases and DNA binding proteins.

Significance

DNA glycosylase repairs DNA damage to maintain the genome integrity, and thus it is essential for the survival of all organisms. However, it remains a long-standing puzzle how glycosylase diffuses along the genomic DNA to locate the sparse and aberrant lesion sites efficiently and accurately in the genome containing numerous base pairs. Previously, only the high-speed–low-accuracy search mode has been characterized experimentally, while the low-speed–high-accuracy mode is undetectable. Here, we observed the low-speed mode of glycosylase AlkD translocating, and further dissected its molecular mechanisms. To achieve this, we developed an integrated platform by combining scanning FRET-FCS with Markov state model. We expect that this platform can be widely applied to investigate other glycosylases and DNA-binding proteins.

Edited by Donald G. Truhlar, University of Minnesota, Minneapolis, MN, and approved July 28, 2020 (received for review February 17, 2020)

\textsuperscript{a}School of Life Sciences, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structure, Tsinghua University, Beijing 100084, China; \textsuperscript{b}Department of Chemistry, Center of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Kowloon 999077, Hong Kong; \textsuperscript{c}State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian 350002, China; \textsuperscript{d}Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Department of Chemical Biology, College of Chemistry and Molecular Engineering, and Biomedical Pioneering Innovation Center, Peking University, Beijing 100871, China; \textsuperscript{e}University of Chinese Academy of Sciences, Beijing 100049, China

Author contributions: X.S.Z., X.H., and C.C. designed research; S.P., X.W., L.Z., and S.H. performed research; S.P., X.W., and L.Z. contributed new reagents/analytic tools; S.P., X.W., L.Z., X.S.Z., X.H., and C.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Supporting Information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2002971117/-/DCSupplemental.

First published August 20, 2020.

www.pnas.org/cgi/doi/10.1073/pnas.2002971117

PNAS | September 8, 2020 | vol. 117 | no. 36 | 21889–21895
Single-molecule fluorescence resonance energy transfer (FRET) is more sensitive to the variations within small distance and has been successfully applied to monitor the 1-bp relative motions between DNA and nucleosome (22). However, according to our estimation, to investigate the fast and slow diffusion modes of glycosylase for lesion sites search simultaneously, both nanometer spatial resolution and microsecond time resolution are required. Therefore, commonly used camera-based microscopy is not suitable to implement single-molecule FRET measurements due to its limitation in time resolution (~10 ms).

In this regard, we chose to combine FRET with our previously developed scanning fluorescence correlation spectroscopy (FCS) based on a confocal microscope (23) to design a scanning FRET-FCS method. This method can provide protein dynamics at subnanometer spatial resolution and microsecond temporal resolution, three orders of magnitude improvement from millisecond resolution provided by previous methods (16–19), serving as a useful platform to monitor the one-dimensional (1D) diffusion modes of glycosylase on double-stranded DNA (dsDNA).

In this study, we have utilized the high-resolution scanning FRET-FCS to successfully capture both the fast and slow modes within short-distance diffusion (21–99 bp) of Bacillus cereus alkylpurine glycosylase D (AlkD), an alkylpurine DNA glycosylase targeting a range of diverse alkylpurine substrate (24, 25). We found that the fast mode has a similar diffusion timescale (~0.12 μs per bp) as that found for other protein–DNA systems (19, 26–29). More interestingly, we directly captured a slow mode (~15 μs per bp) of AlkD diffusion along the dsDNA, which has not been characterized before experimentally. To further dissect its underlying molecular mechanism, we constructed Markov state model (MSM) (30–49) based on extensive MD simulations with an accumulated time of ~15 μs initiated from the crystal structure of AlkD–DNA complex (50). One 1-ms synthetic trajectory was generated according to the transition probability matrix of the MSM in order to visualize continuous cycles of AlkD diffusion over dsDNA, the timescale of which cannot be reached by conventional MD simulations. We further discovered that the slow mode is constituted by a rate-limiting rotational motion, followed by a sequential translational motion of AlkD on dsDNA. Moreover, the diffusion rate per base pair estimated by MSM is consistent with that of the slow mode measured by scanning FRET-FCS, suggesting the slow-diffusing AlkD adopts a conformation that closely resembles the crystal structure of AlkD in complex with the DNA lesion (50). Furthermore, we have pinpointed the essential role of residue Y27 in determining the AlkD’s dynamic diffusion on dsDNA. In summary, we have combined scanning FRET-FCS measurements and MSM to elucidate how diffusion modes are regulated by conformational dynamics of AlkD–DNA complexes to achieve efficient target searching, paving the way to investigate the lesion recognition and excision in AlkD and other DNA-binding proteins.

**Results**

**AlkD Displays 1D Diffusion on dsDNA via Two Different Modes.** As shown in Fig. 1A, Cy3-labeled AlkD was present in the solution to transiently bind and diffuse laterally (1D) on immobilized Cy3-labeled dsDNA. A home-built inverted confocal fluorescence microscope equipped with a piezo stage was used to capture scanning FRET-FCS curves, which were calculated via cross-correlation between Cy3 and Cy5 signals.

From 0.03 to 100 ms, our scanning FRET-FCS curves displayed the rise of cross-correlation, which was caused by processes leading to changes of Cy3–Cy5 FRET signals at this timescale. Three anticorrelation components were extracted from the scanning FRET-FCS curves in this range (Fig. 1B). We assigned the slowest component as the process of binding and dissociation between AlkD and immobilized dsDNA, because its relaxation time (τ_dis = 44 ± 4 ms) agreed well with the relaxation time (66 ± 18 ms) estimated independently from the protein–dsDNA binding and dissociation rates (SI Appendix, Table S1) using our single-molecule total internal reflection fluorescence microscope (see SI Appendix for details). As expected, when the length of DNA duplex decreased to 8 bp to completely abolish 1D diffusion of AlkD, scanning FRET-FCS curve displayed only one anticorrelation component caused by binding and dissociation between AlkD and dsDNA (SI Appendix, Fig. S1A). Additional control experiments also showed that the two fast anticorrelation components in this range were not produced by breathing of dsDNA or motion of dyes and linkers (SI Appendix, Fig. S1B).

One anticorrelation component caused by binding and dissociation between AlkD and dsDNA (SI Appendix, Fig. S1C and D), indicating that they were not caused by heterogeneity of immobilized molecules. Therefore, we anticipated that the remaining two components (τ_fast and τ_slow) were caused by 1D diffusion of AlkD on dsDNA.

Next, we aimed to dissect the origin of the two diffusion components. To achieve this, we performed Monte Carlo simulations (see SI Appendix for details) to model random diffusion trajectories of AlkD and to generate FRET-FCS curves by assuming that the FRET efficiency depends on both the relative distance and dipole orientation between Cy3 and Cy5 (SI Appendix, Fig. S1E). This enables us to examine whether diffusion-coupled rotational motions of AlkD around dsDNA are the
cause for the two diffusion components in our measured FRET-FCS curves. The simulated FRET-FCS curves displayed two anticorrelation components, whose relaxation times were about 200-fold different from each other (SI Appendix, Fig. S1F). The simulated results did not agree with our measured relaxation times ($\tau_{fast} = 0.22 \pm 0.01$ ms and $\tau_{slow} = 4.1 \pm 0.3$ ms, which differ by about 20 times), disproving the assumed model and suggesting that two diffusion components were not caused by diffusion-coupled rotational motions of AlkD. This is understandable, as both dyes were linked to biomolecules by a single covalent bond, and thus the orientations of the dipoles of Cy3 and Cy5 are speculated to move most freely. As a result, their FRET efficiency only depended on their relative distance. Last, we discovered that the salt concentration affected two diffusion components differently (Fig. 1C and SI Appendix, Fig. S1G), further consolidating that the two diffusion components are caused by different modes. Specifically, high ionic strength, which weakens interactions between AlkD and dsDNA, reduced the relaxation times of both diffusion components and enhanced the contributions of the fast component. Therefore, the relaxation time of the fast component is more sensitive toward salt concentration than the slow component. Together, these results suggested that the two diffusion components, whose relaxation times are about 10- to 20-fold different from each other, were likely caused by two 1D diffusion modes of AlkD–dsDNA complexes possessing different interacting conformations and strengths.

Residence Times of AlkD per Base Pair Derived from Relaxation Times. We first examined the influence of dsDNA length on the relaxation times $\tau_{fast}$ and $\tau_{slow}$ (Fig. 2A and SI Appendix, Fig. S2). When the length of dsDNA decreased from 99 to 21 bp, $\tau_{fast}$ decreased from $\sim 0.3$ to $\sim 0.1$ ms and the proportion of the fast component also decreased from $\sim 80\%$ to $\sim 20\%$. On the contrary, $\tau_{slow}$ was less sensitive to the length of dsDNA and always remained at around 3 ms (Fig. 2A). In this regard, the residence time of AlkD on each base pair in the fast mode can be estimated by exploring the FRET-FCS curves of dsDNA with different lengths. In particular, we generated scanning FRET-FCS curves of AlkD on dsDNA of different lengths in the Monte Carlo simulations (see SI Appendix for details). Two parameters, namely the residence times of AlkD on an internal base pair ($\tau_{int}$) and on a terminal base pair ($\tau_{ter}$), were introduced to perform the Monte Carlo simulations. When $\tau_{int} = 0.12 \pm 0.08$ ms and $\tau_{ter} = 11 \pm 2$ ms, the timescales estimated from Monte Carlo simulation yielded the best fit to the experimentally measured overall timescales for $\tau_{fast}$ at different lengths of dsDNA (Fig. 2B and SI Appendix, Fig. S3A). Based on this result, the residence time of AlkD is $0.12 \pm 0.08$ ms per bp in its fast 1D diffusion mode, and the corresponding diffusion coefficient $D = (8 \pm 5) \times 10^{-10}$ bp/s is within the range of reported values for other glycosylase proteins (19).

We found that $\tau_{slow}$ is highly sensitive to the DNA sequences by using two dsDNAs of the same length but with different sequences (21 bp vs. 21 bp-2, Fig. 2A). When the DNA sequence changed from 21 bp to 21 bp-2, $\tau_{slow}$ was extended from 2 to 4 ms and the proportion of the slow component also increased from $\sim 60\%$ to $\sim 80\%$. By sharp contrast, the DNA sequence had marginal influence on $\tau_{fast}$. Because $\tau_{slow}$ is not sensitive to the length of dsDNA (Fig. 2A), we cannot uniquely estimate $\tau_{int}$ and $\tau_{ter}$ of the slow component using the method described above (SI Appendix, Fig. S3A). To provide a rough estimation of the residence time of AlkD on each base pair in the slow mode, we fit the experimental FRET-FCS data of 45-bp dsDNA by setting $\tau_{int}$:$\tau_{ter} = 1:1$, 1:10, and 1:100, respectively, in the Monte Carlo simulations. Accordingly, simulation produced three respective sets of parameters $\tau_{int} = 17 \pm 1$ µs and $\tau_{ter} = 17 \pm 1$ µs for $\tau_{int}$:$\tau_{ter} = 1:1$, $\tau_{int} = 11 \pm 1$ µs and $\tau_{ter} = 110 \pm 8$ µs for $\tau_{int}$:$\tau_{ter} = 1:10$, and $\tau_{int} = 2.2 \pm 0.2$ µs and $\tau_{ter} = 220 \pm 20$ µs for $\tau_{int}$:$\tau_{ter} = 1:100$, all of which could match the experimental values of the slow mode ($4.1 \pm 0.3$ ms) (SI Appendix, Fig. S3B). Although $\tau_{int}$:$\tau_{ter}$ differs significantly in three sets of parameters, the residence times of AlkD on an internal base pair ($\tau_{int}$) are similar and serve as a suitable metric for estimating the diffusion timescales. Therefore, we estimated that AlkD spends 2 to 17 µs per bp in the slow mode.

Diffusion Dynamics Characterized by Markov State Model. A recent MD simulation study proposed that the AlkD–dsDNA complex can alternate between a loose-interacting conformation for general search along dsDNA and a tightly interacting conformation in the crystal-like structure when initiating excision (13). To establish the connection between the conformation of the AlkD–dsDNA complex and dynamics of 1D diffusion that we captured, we constructed MSM based on extensive all-atom MD simulations published on a public structure of AlkD (Protein Data Bank [PDB] ID: 5CL3) (50). MD simulations with aggregated simulation time of 15 µs were performed (see SI Appendix for details), and more than 700,000 MD conformations were utilized for the MSM construction. AlkD–dsDNA complex in our simulations adopted configurations closely resembling the crystal structure (PDB ID: 5CL3). First, dsDNA was bent in an extent quite similar to that of the crystal structure (Fig. 3A). The bend angle is a common metric to measure the bending of dsDNA (51) (see SI Appendix for details). We found that the bend angle of dsDNA in our simulations covered a wide range between 20° and 100°, and the distribution centered at $\sim 59°$. This matched well with the bend angle ($\sim 55°$) of the crystal structure. Second, most of the protein–DNA interactions observed in the crystal structure were well maintained in our MD simulations (Fig. 3B and see SI Appendix for details). For example, the hydroxyl group of Y27 was hydrogen bonded with the nitrogen atoms in the nucleobase of DNA. Also, nitrogen atoms in the side chain of R148 formed salt bridges with the phosphate group of the adjacent adenine. Furthermore, the hydrophobic ring of W109 and W187 formed CH-π interaction with the neighboring nucleosugar. Overall, dsDNA was bound to AlkD in a similar fashion as in the crystal structure.

To fully elucidate the diffusion dynamics of AlkD along dsDNA, we used time-structure independent components analysis (tICA) (34, 52, 53) to determine the slowest relaxing degrees of freedom and constructed MSM using MSBuilder (54) (see SI Appendix for details). The parameters for constructing MSM were validated by generalized matrix Rayleigh quotient (40, 41) (SI Appendix, Figs. S4 and S5). We identified three metastable states when AlkD diffuses over one base pair. Besides the pre- (S1) and post-translocation (S3) states, we identified a translocation intermediate state (S2, Fig. 3C). To further visualize the
Fig. 3. Markov state modeling of AlkD–dsDNA complex. (A) Distribution of dsDNA bend angle sampled by MD simulations (red curve). For comparison, the bend angle calculated using the crystal structure (PDB ID: 5CL3) is shown in black line. (B) Comparison of protein–DNA distances between MD simulations (red circle with error bar) and the crystal structure (black square) (see SI Appendix for details). “G6” and “A7” represent G with residue ID 6 and A with residue ID 7 in our simulation model, respectively. (C) Representative conformations of three major metastable states from two different views. dsDNA and AlkD are shown in indigo and gray, respectively. The helices close to the dsDNA are shown in individual colors, as well as one phosphate in blue sphere to demonstrate the conformational change of AlkD. The mean first passage times estimated from MSM are presented beside the arrows. (D) Projections of AlkD–dsDNA conformations onto the first tIC as a function of time for the 1-ms trajectory generated based on MSM. The state index is labeled at the right-hand side. The projection on 1st tIC is expanded beneath. (E) Projection of MD conformations onto the angle describing the rotation of AlkD around the dsDNA (y axis) and the translating distance of AlkD along the dsDNA (x axis). Contour plots of each metastable state are also shown with the population in the parenthesis. Asymmetric Translocation Pathway Over One Base Pair. Our MSM demonstrated that the translocation over one base pair followed an asymmetric pathway with a rotational motion followed by a sequential translational motion. As shown in Fig. 3 C and E, from S1 to S2, AlkD rotated ∼34° around the dsDNA; from S2 to S3, AlkD slide ∼3.5 Å along the dsDNA, which corresponds to the length of one base pair (55). Furthermore, we found that the timescales of these two conformational transitions were drastically different (Fig. 3C). The transition from S1 to S2 caused by the rotation of AlkD around dsDNA was the rate-limiting step, which occurred at 13 to 17 μs, while transition from S2 to S3 caused by the sliding motion of AlkD only took 1.3 to 1.5 μs, 10 times faster than that from S1 to S2. This could be attributed to the extent of hydrogen bond network reconfiguration during the conformational change (SI Appendix, Fig. S6). From S1 to S2, ∼4.5 hydrogen bonds were broken while ∼7.0 hydrogen bonds were formed simultaneously; from S2 to S3, elimination of ∼3.1 hydrogen bonds was accompanied by the formation of ∼3.0 hydrogen bonds. In this regard, rotation of AlkD around dsDNA, identified as S1-to-S2 transition, involved more significantly reconfigurations of hydrogen bonds to complete its conformational change and therefore became the rate-limiting step in the asymmetric translocation pathway.

Effect of Residue Y27 on Diffusion Dynamics. Previous studies have pinpointed several amino acid residues that are essential for recognizing and excising the damaged DNA (56–58). However, roles of any AlkD residues on protein diffusion dynamics have not been examined. Herein, we focused on residue Y27, which maintained its hydrogen bond with nucleotides during the whole cycle of diffusion (SI Appendix, Fig. S7). The hydroxyl group of hydrophilic residue Y27 inserts into the minor groove of DNA and contacts the nucleobase instead of the backbone (57). To separately examine the effects of hydrogen bonds and steric interactions, we designed Y27F and Y27A mutants, respectively. Both mutants eliminated hydrogen bond while Y27A mutation could further reduce the size of the side chain. The experimental scanning FRET-FCS curve indicated that Y27A accelerated the diffusion in both fast and slow modes, while Y27F mutation barely influenced the translocation rate (Fig. 4A and SI Appendix, Fig. S8), suggesting that it is the steric effect rather than the hydrogen bond that plays a key role in determining the diffusion dynamics. To further dissect the molecular mechanism of Y27 on...
the diffusion dynamics, mutant simulations were initiated from the conformational region corresponding to the rate-limiting rotation step in the translocation. It is worthy to note that our simulations were initiated from the crystal structure with DNA damage, and thereby can only examine how the mutants affect the rate of the slow mode. Specifically, we selected 10 conformations from the transition area between S1 state and S2 state. For each conformation, we generated the corresponding Y27 mutation to initiate two 50-ns MD simulations (see SI Appendix for details). As a comparison, these 10 conformations were also used to seed wild-type simulations. Consistent with the experimental observation, MD simulations also showed that the Y27A mutation promoted the rate (smaller resistance value; see SI Appendix for details) while Y27F mutation demonstrated a similar rate as wild type (Fig. 4A). As the steric effect of Y27 played a dominant role in determining the diffusion dynamics of AlkD along dsDNA, we examined the van der Waals interactions between Y27 and its neighboring protein residues, as well as nucleic acids (Fig. 4B and SI Appendix, Figs. S9 and S10). We found the wild-type Y27 experienced the strongest van der Waals interactions with its surroundings, which were slightly weakened in Y27F but greatly reduced in Y27A. This result supported the above assertion that the bulky side chain of Y27 or Y27F led to strong steric effect and thus hindered the 1D diffusion, while Y27A has smaller volume and accelerated diffusion. Further analysis pinpointed protein residues M24, M28, and W109 contributed significantly to the steric effect caused by Y27 or Y27F (Fig. 4C and SI Appendix, Fig. S9).

Discussion

Glycosylase AlkD is an essential enzyme responsible for repairing DNA damage by utilizing a unique non–base-flipping recognition mechanism (50, 56, 58). However, it has been a long-standing question how DNA glycosylases can locate the DNA lesions efficiently and accurately. It has been proposed that DNA glycosylases such as AlkD can alternative between a fast mode and a slow mode to search for DNA lesions. Unfortunately, existing single-molecule fluorescence imaging techniques do not have sufficient resolutions to capture the slow mode. In this study, our newly developed scanning FRET-FCS platform with improved spatial and temporal resolutions enabled us to examine the diffusion of AlkD along dsDNA at much higher resolutions, which captured the fast and slow 1D diffusions of AlkD simultaneously. In the fast mode, AlkD spends ~0.12 μs on each base pair, which corresponds to 1D diffusion constant of ~8 × 10^6 bp·s⁻¹. This is similar to the reported diffusion rates of other proteins, including HOG1, MutY, and MutM glycosylases, along dsDNA (19). In the slow mode, AlkD spends 2 to 17 μs on each base pair with 100 mM NaCl, which corresponds to a 1D diffusion constant of 6 × 10^4 to 5 × 10^5 bp·s⁻¹. On average, AlkD spends ~70 ms on dsDNA during each binding event and diffuses across 70 to 190 bp in its slow mode before dissociation, which is beyond the detection limit of existing single-molecule imaging methods. This emphasizes the advantages of our FRET-FCS platform.

A recent MD simulation study identified two distinct AlkD–dsDNA conformations, a loose-binding state and a tight-binding state, which were hypothesized to cause different diffusion rates (13). Similarly, a DNA binding domain EngHD has been shown to switch between two different conformations to modulate its diffusion rates (11). Our MD simulations demonstrated that AlkD–dsDNA adopts a crystal-structure–like, tight-binding conformation and MSM estimated a diffusion timescale of 15 to 20 μs per bp, which agreed with the measured value of the slow mode. This confirmed our assertion that the slow mode corresponds to the tight-binding conformation as observed in the crystal structure and in our MD simulations, while the fast mode is likely caused by the loose-binding conformation identified by previous MD simulation (13). Collectively, evidence from experiments and computations support the model that AlkD–dsDNA switches between different conformations during target searching (Fig. 5).

Combining our findings with previous studies, we proposed the following kinetic scheme to elucidate how conformational dynamics of AlkD and dsDNA interplay with each other to enable efficient target searching and recognition (Fig. 5). AlkD–dsDNA adopts the loose-binding conformation in the fast diffusion mode, which is a general search mode to screen through unmodified dsDNA with high efficiency. Although AlkD can occasionally transit into its slow mode on an unmodified dsDNA, distortions of dsDNA structure caused by modified or mismatched bases are likely to promote the slow diffusion mode by stabilizing the protein into its crystal-like conformation to form a tight-binding complex, which gives AlkD sufficient time to thoroughly search near the lesion (13). After locating the modified base, AlkD–dsDNA excision complex is formed to complete the target recognition and to initiate subsequent lesion excision. We hypothesized that similar search mechanisms are used by human glycosylase AAG, which also presents a loose-binding and a strong-binding structure with DNA (12).

Our MSM constructed from all-atom MD simulations provided additional insights into the diffusion mechanisms of AlkD.

![Mechanistic scheme of diffusion and target search of AlkD. Modified bases are shown in red.](image-url)
First, AlkD displays rotational motions around the helical structure of dsDNA during its 1D diffusion, a mechanism used by many DNA binding proteins (19). Our MSM further demonstrated that diffusion of AlkD over one base pair contained two sequential motions in the slow mode. The ~34° rotation motion is the rate-limiting step, which is 10 times slower than the sliding motion over one base pair. Therefore, mutations affecting the rate-limiting rotational motion shall also impact the overall diffusion rate of the slow mode. Indeed, our mutant simulations indicate that Y27A can accelerate the rotational motion, and thus should also increase the overall rate of the slow diffusion mode. Our scanning FRET-FCS measurements of the Y27A mutant display a faster diffusion rate compared to wild type, and thus validated our theoretical prediction based on the asymmetric translocation model. This agreement between simulations and experiments on the mutant’s effect on AlkD’s diffusion rate indirectly supports our proposed mechanism of the asymmetric movement. In addition, the relaxation times of the fast and slow modes, which were 10 to 20 times different from each other, were affected by Y27A and Y27F mutations in a similar trend (SI Appendix, Fig. S8), suggesting that AlkD in the fast and slow modes may share similar interaction network surrounding residue Y27. We also note several previous studies have achieved direct comparisons between MSM predictions and experimental observations including FRET (59), and particularly the augmented Markov model framework has been developed to combine experimental and simulation data to improve the accuracy of the predicted biomolecular dynamics by MSMs (60). In this study, our MSMs only investigate metastable states in our MSMs and experimental observables.

Here, combining our newly developed scanning FRET-FCS method with MSM constructed from extensive MD simulations, we developed an integrative platform to elucidate atomistic diffusion dynamics of proteins on dsDNA. Our results suggested that AlkD utilizes a general fast diffusion mode to pass through unmodified bases with high efficiency and a slow diffusion mode to thoroughly search near the lesion site. Conformational dynamics of AlkD–dsDNA complex regulate transformation between two diffusion modes to ensure efficiency and accuracy of the target search processes. Similar mechanisms are likely being used by other glycosylases and DNA binding proteins.

Materials and Methods

Scanning FRET-FCS Measurements. PEG-passivated slides were incubated with 0.05 mg/mL streptavidin for 3 min. Then biotinylated dsDNA was specifically attached to PEG-passivated slides via interactions between biotin and streptavidin. Free-diffusing Cy5-AlkD was added to microscope flow cells to examine interaction between Cy5-dsDNA and Cy5-AlkD. Unless stated otherwise, scanning FRET-FCS experiments were performed at 25 °C in buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 1% BSA) with an oxygen scavenging system, containing 3 mg/mL glucose, 100 μg/mL glucose oxidase, 40 μg/mL catalase, 1 mM cyclodextratetraene, 1 mM 4-nitrobenzylalcohol, and 1.5 mM 6-hydroxy-2,5,7,8-tetramethyl-chromane-2-carboxylic acid (61, 62). Scanning FRET-FCS measurements were performed on a home-built confocal microscope, based on a Zeiss AxiO Observer D1 fluorescence microscope with an oil-immersion objective (Zeiss; 100×; numerical aperture, 1.4), and solid-state 488-, 532-, and 640-nm excitation lasers (Coherent; OBIS Smart Lasers). The piezo nanopositioning stage (PI) was assembled on the stage of the microscope and controlled by a multifunction I/O card (National Instruments; PCI-6289) with a custom Labview script (National Instruments) to achieve accurate nanoscale scanning and positioning. For the scanning FRET-FCS experiments, the scan speed was 0.1 μm/s. Laser powers at the samples were ~5 μW. Fluorescence signals from the sample passed through a pinhole (diameter, 50 μm) and were separated by a dichroic mirror (T635pxr, Chroma), which were further filtered by bandpass filters ET580/65m (for Cy5, Chroma) and ET700/75m (for Cy3, Chroma) before detected by two avalanche photodiode (APD) detectors (Excilitas; SPCM-AQRH-14). The cross-correlation of fluorescence signals was calculated by a correlator (Correlator.com; Flex2-01D). For each experimental condition, three or more identical replicates were performed. Each replicates were collected for 5 min.

Scanning FRET-FCS curves between 30 ms and 250 ms were fitted by the following:

$$G(r) = A_{fast} \cdot e^{-\tau_{fast}} + A_{slow} \cdot e^{-\tau_{slow}} + A_{dis} \cdot e^{-\tau_{dis}},$$

to extract three relaxation times. Percentage of fast diffusion mode was calculated via $A_{fast} / (A_{fast} + A_{slow} + A_{dis}).$

MD Simulations. The crystal structure of AlkD [PDB ID: 5CL3 (50)] was used as the structural basis for constructing the pre-translocation model and the post-translocation model for investigation the diffusion of AlkD over one base pair. Manual modifications were made to ensure the length and sequence of base pairs match between the pre- and post-translocation states. Preliminary pathways were generated using Climber algorithm (63, 64) and 30 representative conformations were selected from the preliminary pathways (SI Appendix, Fig. S11) to seed the first round of MD simulations. For each conformation, energy minimization and equilibration were performed, followed by one 50-ns NVT simulations at 298 K. The conformations in the first round of simulations were collected and used as the conformational basis to seed the extensive MD simulations for the MSM construction. Amber99sb-ildn force field (65) was utilized to simulate both the protein and nucleotides. All MD simulations were performed by Gromacs 5.0.4 package (11). See SI Appendix for the details about the model construction and MD simulations.

Markov State Model. The conformational ensemble from the 300 × 50-ns MD simulations served as the structural basis to construct MSM using MSmBUILDER (54). In particular, tICA (34, 52, 53) was used to determine the slowest relaxing degrees of freedom (see SI Appendix for details). Generalized matrix Rayleigh quotient (40, 41) was utilized to validate the parameters for constructing MSM, including the atom pairs, tIC number, tIC lag time, and the number of microstates (SI Appendix, Figs. S4 and S5). With the optimal parameters, the MD conformational ensemble was finally divided into 1,000 microstates by K-center algorithm (66, 67) (SI Appendix, Figs. S5A and S52) and a proper lag time of 15 ns was chosen to render the model Markovian (SI Appendix, Fig. S13). To gain the molecular insight, the 1,000 microstates were further grouped into macrostates based on their kinetic similarity. See SI Appendix for the details about MSM construction and validation.

For further experimental and computational details, see SI Appendix.

Data Availability. Raw data used to plot all figures can be found in Dataset S1. All study data are included in the article and SI Appendix.

ACKNOWLEDGMENTS. This project was supported by funds from the National Natural Science Foundation of China (Grants 21922704, 21877069, and 31570574 to C.C.; 21233002 and 21521003 to X.S.Z.; and 21723307 and 21903071 to L.Z.), the National 1000 Youth Talents Program of China to L.Z., and Hong Kong Research Grant Council (Grants 16307718, 16318161, AoE/P-705/16, and T13-60518/18-W to X.H.).

L. S. E. Jones Jr., Differential effects of reactive nitrogen species on DNA base excision repair initiated by the alkyladenine DNA glycosylase. Carcinogenesis 30, 2123–2129 (2009).

J. L. Friedman, J. T. Stivers, Detection of damaged DNA bases by DNA glycosylase enzymes. Biochemistry 48, 4957–4967 (2010).

Y. Y. Zhang, L. Chang, P. J. O’Brien, Repair of alkylation damage in eukaryotic chromatin depends on searching ability of alkyladenine DNA glycosylase. ACS Chem. Biol. 10, 2606–2615 (2015).

1. E. C. Friedberg et al., DNA repair: From molecular mechanism to human disease. DNA Repair (Amst.) 5, 986–996 (2006).

2. W. P. Roos, B. Kaina, DNA damage-induced cell death by apoptosis. Trends Mol. Med. 12, 440–450 (2006).

3. P. A. Jeggo, L. H. Pearl, A. M. Carr, DNA repair, genome stability and cancer: A historical perspective. Nat. Rev. Cancer 16, 35–42 (2016).

4. M. A. Petri, T. Tulika, L. M. Carmona-Marin, M. Scheibye-Knudsen, Protecting the aging genome. Trends Cell Biol. 30, 117–132 (2020).

5. S. L. E. Jones Jr., Differential effects of reactive nitrogen species on DNA base excision repair initiated by the alkyladenine DNA glycosylase. Carcinogenesis 30, 2123–2129 (2009).

6. J. L. Friedman, J. T. Stivers, Detection of damaged DNA bases by DNA glycosylase enzymes. Biochemistry 48, 4957–4967 (2010).

7. Y. Y. Zhang, L. Chang, P. J. O’Brien, Repair of alkylation damage in eukaryotic chromatin depends on searching ability of alkyladenine DNA glycosylase. ACS Chem. Biol. 10, 2606–2615 (2015).
21. S. Peng, R. Sun, W. Wang, C. Chen, Single-molecule photoactivation FRET: A general
27. A. Granéli, C. C. Yeykal, R. B. Robertson, E. C. Greene, Long-distance lateral diffusion
Peng et al. PNAS

36. B. E. Husic, V. S. Pande, Markov state models: From an art to a science. J. Chem. Theory Comput. 10, 2648–2657 (2014).

37. F. Sittel, G. Stock, Perspective: Identification of collective variables and metastable
process by the inducible transcription factor Egr-1. Proc. Natl. Acad. Sci. U.S.A. 109, E1724–E1732 (2012).

38. W. Wang, S. Q. Cao, L. Z. Zhu, X. H. Huang, Constructing Markov state models to
electivate the functional conformational changes of complex biomolecules. Wires Comput Mol Sci e1343 (2018).

39. B. W. Zhang et al., Simulating replica exchange: Markov state models, proposal

40. R. T. McGibbon, V. S. Pande, Variational cross-validation of slow dynamical modes in
molecular kinetics. J. Chem. Phys. 142, 124105 (2015).

41. F. Nüske, B. G. Keller, G. Pérez-Hernández, A. S. Mey, F. Noé, Variational approach to
molecular kinetics. J. Chem. Theory Comput. 10, 1739–1752 (2014).

42. R. D. Malmström, C. T. Lee, A. Van Wart, R. E. Amaro, On the application of molecular
dynamics based Markov state models to functional proteins. J. Chem. Theory Comput. 10, 2648–2657 (2014).

43. F. Morcos et al., Modeling conformational ensembles of slow functional motions in

44. N. V. Buchete, G. Hummer, Coarse master equations for peptide folding dynamics.
J. Phys. Chem. B 112, 6057–6068 (2008).

45. A. C. Pan, B. Roux, Building Markov state models along pathways to determine free
energies and rates of transitions. J. Chem. Phys. 129, 064107 (2008).

46. J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, W. C. Swope, Automatic discovery of
metastable states for the construction of Markov models of macromolecular conformational
dynamics. J. Chem. Phys. 126, 155101 (2007).

47. G. R. Bowman, V. S. Pande, F. Noé, An Introduction to Markov State Models and Their
Application to Long Timescale Molecular Simulation, (Springer Science and Business Media, 2013).

48. E. A. Mullins et al., The DNA glycosylase AlkD uses a non-base-flipping mechanism to
excise bulky lesions. Nature 527, 254–258 (2015).

49. A. Vologodskii, M. D. Frank-Kamenetski, Strong bending of the DNA double helix.
Nucleic Acids Res. 41, 6785–6792 (2013).

50. Y. Y. Chen, S. Fuchigami, Slug dynamics of a protein backbone in molecular dy-
namics simulation revealed by time-based structure independent component analysis.
J. Chem. Phys. 139, 215102 (2013).

51. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow
molecular order parameters for Markov model construction. J. Chem. Phys. 139, 015102 (2013).

52. M. P. Harrigan et al., MSMBuilder: Statistical models for biomolecular dynamics.
Biophys. J. 110, 10–15 (2015).

53. G. G. Privé et al., Helix geometry, hydration, and G.A mismatch in a B-DNA dimer.
Science 238, 498–504 (1987).

54. E. H. Robinson, A. S. Gowda, T. E. Spratt, B. Gold, B. F. Eichman, An unprecedented
nucleic acid capture mechanism for excision of DNA damage. Nature 468, 406–411
(2010).

55. E. A. Mullins, H. E. Robinson, B. F. Eichman, The substrate binding interface of al-
kylpurine DNA glycosylase AlkD. DNA Repair (Amst.) 13, 50–54 (2014).

56. E. A. Mullins et al., AlkC and AlkD. Structure and flexibility adaptation in nonspecific and specific
DNA repair of alkylated DNA by a new superfamily of AlkD-DNA glycosylase.
J. Mol. Biol. 414, 372–393 (2012).

57. E. A. Mullins, H. E. Robinson, B. F. Eichman, Toxicity and repair of DNA adducts produced by
the natural product yataytubine. J. Med. Chem. 50, 3707–3715 (2007).

58. F. Noé et al., Dynamical fingerprints for probing individual relaxation processes in
biomolecular dynamics with simulations and kinetic experiments. Proc. Natl. Acad. Sci.
U.S.A. 108, 4822–4827 (2011).

59. S. Olsson, H. Wu, F. Paul, C. Clementi, F. Noé, Combining experimental and simulation data of molecular processes via augmented Markov models. Proc. Natl. Acad. Sci.
U.S.A. 114, 8265–8270 (2017).

60. B. Wu et al., Translocation kinetics and structural dynamics of ribosomes are modu-
lated by the conformational plasticity of downstream pseudoknots. Nucleic Acids Res. 46, 9736–9748 (2018).

61. L. Zhang et al., Conformational dynamics and cleavage sites of Cas12a are modulated
by complementarity between crRNA and DNA. Science 19, 492–503 (2019).

62. D. A. Silva et al., Millisecond dynamics of RNA polymerase II translocation at atomic
resolution. Proc. Natl. Acad. Sci. U.S.A. 111, 7665–7670 (2014).

63. D. R. Weiss, M. Levitt, Can morphing methods predict intermediate structures? J. Mol.
Biol. 385, 665–674 (2009).

64. K. Lindorff-Larsen et al., Improved side chain torsion potentials for the Amber ff99SB
protein force field. Proteins: Struct., Funct., Biol. 78, 1950–1956 (2010).

65. T. F. Gonzalez, Clustering to minimize the maximum intercluster distance. Theor.
Comput. Sci. 38, 293–306 (1985).

66. Y. Zhao, F. K. Sheong, J. Sun, P. Sander, X. Huang, A fast parallel clustering algorithm for
molecular simulation trajectories. J. Comput. Chem. 34, 95–104 (2013).

31. B. Roux, K. A. Dill, Understanding structure and dynamics of biomolecules with
molecular dynamics simulations. Annu. Rev. Phys. Chem. 55, 657–702 (2004).

32. J. D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. 
Curr. Opin. Struct. Biol. 25, 135–144 (2014).

33. J. H. Printz et al., Markov models of molecular kinetics: Generation and validation. 
J. Chem. Phys. 134, 174105 (2011).

34. C. R. Schwantes, V. S. Pande, Improvements in Markov state model construction reveal
many non-native interactions in the folding of NTL9. J. Chem. Theory Comput. 9, 2000–2009 (2013).

35. F. Noé, E. Rosta, Markov models of molecular kinetics. J. Chem. Phys. 151, 194001 (2019).

36. B. E. Husic, V. S. Pande, Markov state models: From an art to a science. J. Am. Chem.
Soc. 140, 2386–2396 (2018).

37. F. Sittel, G. Stock, Perspective: Identification of collective variables and metastable
states of protein dynamics. J. Chem. Phys. 149, 150901 (2018).