







ARTICLE

<https://doi.org/10.1038/s41467-018-07797-4>

OPEN

Breaking the speed limit with multimode fast scanning of DNA by Endonuclease V

Arash Ahmadi ¹, Ida Rosnes^{1,2}, Pernille Blicher¹, Robin Diekmann ^{3,4}, Mark Schüttpelz ³, Kyrre Glette ⁵, Jim Tørresen ⁵, Magnar Bjørås^{2,6}, Bjørn Dalhus ^{1,2} & Alexander D. Rowe^{1,7}

In order to preserve genomic stability, cells rely on various repair pathways for removing DNA damage. The mechanisms how enzymes scan DNA and recognize their target sites are incompletely understood. Here, by using high-localization precision microscopy along with 133 Hz high sampling rate, we have recorded EndoV and OGG1 interacting with 12-kbp elongated λ -DNA in an optical trap. EndoV switches between three distinct scanning modes, each with a clear range of activation energy barriers. These results concur with average diffusion rate and occupancy of states determined by a hidden Markov model, allowing us to infer that EndoV confinement occurs when the intercalating wedge motif is involved in rigorous probing of the DNA, while highly mobile EndoV may disengage from a strictly 1D helical diffusion mode and hop along the DNA. This makes EndoV the first example of a monomeric, single-conformation and single-binding-site protein demonstrating the ability to switch between three scanning modes.

¹Department of Medical Biochemistry, Institute for Clinical Medicine, University of Oslo, NO-0372 Oslo, Norway. ²Department of Microbiology, Oslo University Hospital HF, Rikshospitalet and University of Oslo, PO Box 4950 Nydalen, NO-0424 Oslo, Norway. ³Biomolecular Photonics, Department of Physics, University of Bielefeld, Universitätsstraße 25, DE-33615 Bielefeld, Germany. ⁴European Molecular Biology Laboratory (EMBL), Cell Biology and Biophysics Unit, Meyerhofstraße 1, DE-69117 Heidelberg, Germany. ⁵Department of Informatics, University of Oslo, PO Box 1080 Blindern, NO-0316 Oslo, Norway. ⁶Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), PO Box 8905, NO-7491 Trondheim, Norway. ⁷Department of Newborn Screening, Division of Child and Adolescent Medicine, Oslo University Hospital, PO Box 4950 Nydalen, NO-0424 Oslo, Norway. Correspondence and requests for materials should be addressed to B.D. (email: bjornda@medisin.uio.no) or to A.D.R. (email: alerow@ous-hf.no)

Chemical alterations of DNA can introduce permanent damage and mutations, the accumulation of which may cause genomic instability—a primary cause of both cancer and aging^{1–3}. In order to maintain genomic stability, DNA damage is corrected by several classes of DNA repair proteins³. Despite decades of intensive research, a detailed understanding of the molecular mechanisms by which these proteins identify and repair errors in DNA with incredible fidelity, while relying only on thermal energy to reach their targets, remains elusive. Facilitated diffusion is the widely accepted term used to describe the mechanism by which DNA binding proteins stochastically and passively locate target sites^{4–6}. In this model the linear structure of DNA provides a track to which the proteins associate, enabling them to scan DNA with a dramatically increased rate of damage recognition compared to a random three-dimensional (3D) search^{5,6}. According to an early theoretical outline of this model⁴, proteins initially undergoing free 3D diffusion in solution, and bind to DNA at nonspecific sites. They then follow one of several strategies to scan DNA. (i) Proteins remain bound and scan the DNA driven by Brownian diffusion, following the DNA helix—known as one-dimensional (1D) helical sliding. Previous single-molecule studies have shown that helical sliding is the preferred strategy for several DNA repair proteins^{7–9}. (ii) Proteins undergo repeated microscopic dissociation-reassociation steps during which they disengage from, but remain proximate to, the DNA with a high probability of reassociation at another site within the same region—this is referred to as hopping¹⁰ or jumping¹¹. (iii) Proteins transfer between two different segments of DNA due to close contact of those segments—known as intersegmental transfer¹². (iv) Proteins macroscopically dissociate from DNA and continue with free 3D diffusion in solution. Despite some single-molecule studies showing that selected proteins choose either sliding or hopping for scanning^{7–10}, several bulk experiments^{13,14} and theoretical^{15–18} studies strongly support the idea that the most efficient target site recognition mechanism incorporates sliding interspersed with varying lengths or forms of hopping. Moreover, recent single-molecule studies have described proteins that can switch between hopping and sliding. However, this has so far only been attributed either to multimeric, ring-shaped proteins which encircle the DNA, keeping the unbound protein in close contact with DNA^{11,19–21}; to proteins with two distinct binding sites²²; or to proteins which undergo cofactor-induced conformational changes^{12,23,24}. It remains to be determined whether interspersed 1D helical sliding and hopping is a strategy limited only to the classes of protein and complexes for which it has already been described, or whether it is also a strategy which may be adopted by single-conformation, monomeric proteins with a single DNA binding site.

In addition to nonspecifically associating with the DNA and performing rapid scanning, proteins must ideally be able to accurately interrogate the substrate to find specific target sites which they can stably bind. A central theoretical study addresses this issue by introducing two different modes of scanning, which they term “search” and “recognition” modes¹⁷. In the search mode, proteins interact with a smooth binding energy landscape of around $1 k_B T$, over which they can easily perform efficient helical sliding. When the binding energy landscape is rougher ($>2 k_B T$), scanning becomes extremely slow, and with increased roughness, the protein–DNA target complex tends towards a stable association–recognition mode. These two modes are originally described for two distinct protein–DNA binding conformations which may be alternated between on account of small or large conformational changes to either protein, DNA or both components. These conformational changes alter the roughness of the binding energy landscape and are coupled to the switch between specific and nonspecific protein–DNA interactions. Several examples exhibiting these modes have been examined in single-molecule scanning experiments^{7–9,11,22,25} and the roughness of the energy landscape for helical sliding is consistently

determined to be between 0.6 and $1.78 k_B T$. The ubiquity of these parameters across different classes of enzymes which scan DNA, and the role of specific structural elements in the choice of scanning mode, are therefore central to our understanding of these interactions.

One such structural element is seen in the crystal structures of endonuclease V (EndoV) in complex with damaged DNA, which have revealed a strand-separating wedge motif at the protein surface, consisting of the four residues: Pro79, Tyr80, Ile81 and Pro82 (PYIP)²⁶. This highly conserved PYIP wedge is crucial for EndoV’s ability to recognize helical distortions in DNA, since the DNA strands are split exactly at the weak point in DNA²⁷, and can therefore be reasonably expected to play a role in switching from search to recognition mode. EndoV recognizes and cleaves the DNA strand next to a wide variety of DNA damage, ranging from deaminated bases^{28–34}, apurinic/apyrimidinic (AP) sites^{29,32}, base mismatches^{32,35}, insertion–deletion (ID) loops, hairpins, flaps and pseudo-Y structures³⁶. In addition, the enzyme has also been shown to bind to, but not cleave, a variety of branched DNA structures such as forks, three-way junctions and Holliday junctions³⁷.

In the present study the scanning mechanism of EndoV has been characterized at the single-molecule level, and compared with the well-studied baseline human 8-oxoguanine DNA glycosylase 1 (hOGG1), which represents a typical protein with pure helical sliding behavior⁷. Two aspects of the scanning mechanism are investigated here. (i) The role of EndoV’s wedge motif in switching between search and recognition modes is examined by comparison of wild-type EndoV (wt-EndoV) and a wedge-deficient mutant EndoV (wm-EndoV). (ii) The ability of these proteins to switch between hopping and sliding modes, despite being monomeric proteins with a single conformation and single DNA binding site and none of the other properties common to enzymes which have previously displayed this behavior. We show that the intercalating wedge motif is involved in rigorous probing of the DNA, and that the proteins scan DNA at speeds clearly exceeding the limit for helical sliding, making EndoV the first example of a monomeric protein able to switch between a strictly 1D helical diffusion and hopping along the DNA independent of a conformational change in the protein or by binding of a cofactor.

Results and discussion

Protein–DNA interaction visualization. To examine protein–DNA interactions at the single-molecule level, DNA tracks were produced by attaching a 12 kbp fragment of λ -DNA to a microscope coverglass at one end and a polystyrene bead held in an optical trap at the other end (Fig. 1), allowing the DNA to be elongated to ~95% of its theoretical length. Protein molecules were fluorescently labeled with ATTO 647N, and the emission signal was detected using a high light collection efficiency optical setup³⁸ combined with effective noise reduction strategies, including surface passivation and illumination of a thin layer of the sample³⁹. Once DNA was localized and linearized (Fig. 1c, top) the labeled proteins were injected into the sample cell of a custom-made flow system and interactions between proteins and DNA were recorded with high temporal resolution (7.5–23.5 ms) (Fig. 1c, middle and Supplementary Movie 1). All signals in a frame were localized with a spatial precision of between 20 and 42 nm, and the trajectories—defined as the path followed by a protein during a single uninterrupted DNA binding event—of proteins moving along DNA were analyzed using custom analysis routines⁴⁰ (Fig. 1c, bottom). We took several important steps to remove potential sources of error. In the current study, DNA binding events take place in the absence of flow, DNA-intercalating dyes and large antibodies and quantum dots, which heavily affect the protein’s diffusive properties. By making use of photo-stable dyes with high quantum yield along with the above-mentioned strategies to increase signal-to-noise ratio, we have been able to exploit the advantages of long-lived bright

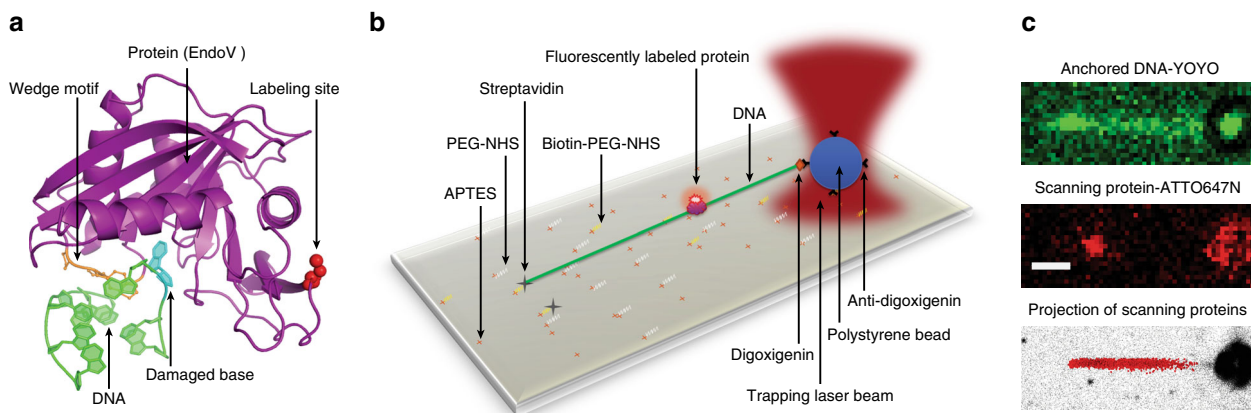


Fig. 1 Protein-DNA interaction experiment. **a** Crystal structure of *Thermotoga maritima* (Tma) EndoV (purple) in complex with deaminated DNA (green). The protein labeling site (red spheres) does not interfere with DNA binding and damage recognition (cyan base in DNA). The PYIP-wedge motif (orange) consists of the four residues: Pro79, Tyr80, Ile81 and Pro82. **b** Single-molecule experimental setup. APTES (3-Aminopropyl)triethoxysilane, PEG-NHS *N*-hydroxylsuccinimide (NHS) functionalized polyethylene glycol, Biotin-PEG-NHS biotinylated PEG-NHS. The position of the polystyrene bead is controlled by a steerable optical trap, and used to linearize the DNA. **c** Upper panel: epifluorescence microscopy image of linearized DNA labeled with intercalating dye YOYO-1 (not present during the protein scanning experiments). Middle panel: a single fluorescently labeled EndoV molecule interacting with linearized DNA. Lower panel: projection of all detected trajectories of one data set of EndoV along the DNA. White scale bar equals 1 μm

signals without compromising the biological relevance of our results by drastically increasing the size of the diffusing units.

Instantaneous diffusion rate distributions. The proteins displayed heterogeneous scanning behavior, pausing intermittently between qualitatively fast and slow runs on the DNA (Fig. 2a, insets). Instantaneous diffusion rates were calculated with moving windows 5 steps in length for individual trajectories (Fig. 2a and Supplementary Fig. 1), and were compared against a single-mode simulated random walk trajectory of a particle with an average diffusion coefficient equal to that measured for the corresponding protein ($D_{\text{ave-hOGG1}} = 0.127 \pm 0.269 \mu\text{m}^2 \text{s}^{-1}$, $D_{\text{ave-wt-EndoV}} = 0.645 \pm 0.977 \mu\text{m}^2 \text{s}^{-1}$, $D_{\text{ave-wm-EndoV}} = 0.701 \pm 0.936 \mu\text{m}^2 \text{s}^{-1}$). The histogram in Fig. 2a shows the distribution of calculated instantaneous diffusion rates for the three proteins, while the solid line of the overlaid density plot shows the distribution of a single-mode simulated random walk. The apparent shape of the distribution of proteins shows the deviation from that which would be expected in the case of a monomodal random walk. The instantaneous diffusion coefficients for proteins span a range a whole order of magnitude larger than the corresponding simulated random walks, and there are clear indications of multimodality.

Classification of scanning by activation energy. To investigate the existence of these modes we began by considering the observed 1D diffusion in terms of the binding energy landscape. According to a pure hydrodynamic model^{41,42}, when proteins follow a helical path along DNA, there is an upper limit for the observed 1D diffusion rate, calculated as $0.89 \mu\text{m}^2 \text{s}^{-1}$ and $1.3 \mu\text{m}^2 \text{s}^{-1}$ for hOGG1 and EndoV, respectively (Supplementary Note 1). In this model, friction between the DNA and the protein is assumed to be zero—ideal sliding with no activation energy barrier. In reality, proteins experience an energy barrier landscape of the order of $k_B T$ ¹⁷, yielding measured diffusion constants which are markedly lower than the theoretical upper limit. Relating this ideal upper limit to the measured diffusion rates gives the corresponding value of the instantaneous activation energy barrier. Considering the sliding process in terms of the kinetics of a sequence of interactions between the protein and adjacent DNA bases gives a rate constant (k) for the interaction of $k = \frac{1}{t} = 2D / \langle x^2 \rangle$, where t is time (s), the mean square displacement $\langle x^2 \rangle$ is equal to 1 bp^2 and the diffusion constant D is

expressed in terms of $\text{bp}^2 \text{s}^{-1}$. From the Arrhenius equation we know that $k = A e^{-E_a/k_B T}$, and for ideal sliding at the upper diffusion rate limit where $E_a = 0$, $A = k_{\text{ideal}}$. The activation energy barrier is obtained as $E_a = \ln\left(\frac{k_{\text{ideal}}}{k}\right) k_B T$. Using the instantaneous diffusion rate results we are able to calculate E_a for every step in the trajectories, as shown in Fig. 2a. Based on the value of E_a these steps are classified to belong to one of these three ranges representing different modes of diffusion (Supplementary Movie 2). (i) $0.5 k_B T < E_a < 2 k_B T$ —based on several theoretical and experimental studies^{7-9,11,17,22,25,42}, where the diffusion process is expected to be dominated by helical sliding. (ii) $E_a > 2 k_B T$ —describes segments where proteins apparently paused locally on the DNA or where step sizes are below the spatial resolution limit of our instrumentation, and this mode is referred to as recognition¹⁷ or interrogation mode⁹. (iii) $E_a < 0.5 k_B T$ —in this mode the protein slides very close to the upper limit of the diffusion rate, an observation which has not been reported previously for helical sliding. In the case of EndoV, in around 50% of the time in this mode, the protein exceeds the upper limit for diffusion by helical sliding ($E_a < 0$). This is a clear indication that in this mode, the protein diffuses in a manner which is incompatible with the limitations of helical sliding, while remaining clearly associated with the 1D axis of the DNA.

Classification of scanning by hidden Markov model. To further validate the three-mode interpretation of our scanning results, avoiding kinetic assumptions related to the activation energy barrier classification, we used a recently published variational Bayes single-particle tracking (vbSPT) method⁴³ to identify and classify diffusion states using a hidden Markov model (HMM) approach. The model assumes a given number of diffusion states with memoryless transitions occurring between them. Each state represents an independent Gaussian distributed component of the overall diffusion spectrum. By finding the best fit to the experimental data, the ideal number of states, their associated diffusion coefficients, occupancies and transition probabilities are determined using a variational Bayesian method. The vbSPT software determines the ideal number of states by over-specifying the initial number of hidden states, and converging on the most suitable value. While the vbSPT reliably selected a single mode as the best fit for our simulated random walks, the software failed to converge on a reasonable number of states for our experimental

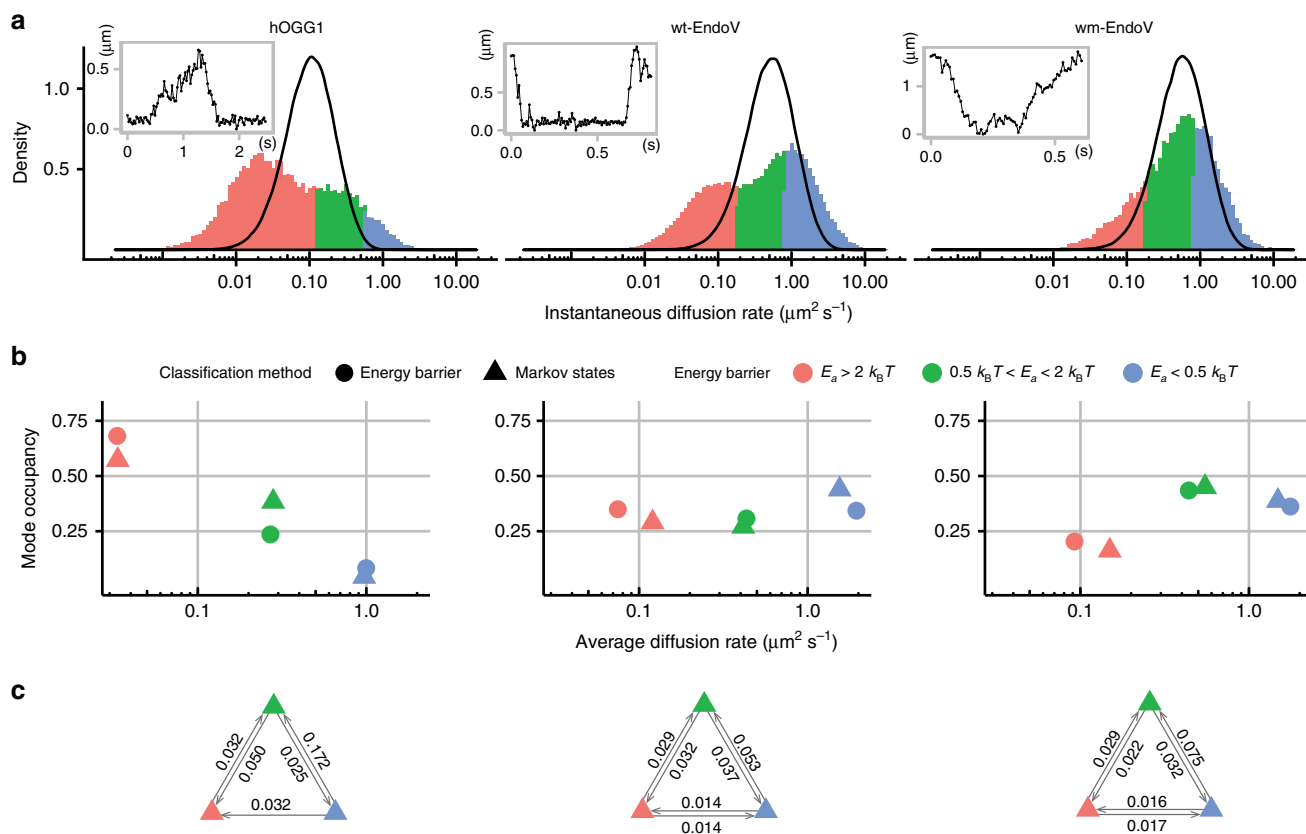


Fig. 2 Diffusion analysis. **a** Distribution of instantaneous diffusion coefficients computed for hOGG1 (697 trajectories) with sliding window of 117.5 ms, and for wt- and wm-EndoV (wt-EndoV: 3443 trajectories, and wm-EndoV: 1099 trajectories) using sliding window of 37.5 ms. Insets: examples of single trajectories for each protein with time on the x-axis and displacement on the y-axis. **b** Comparison of classifications based on activation energy barrier (circles) and on Markov hidden state models (triangles), showing strong agreement between methods. The state occupancy is the proportion of time that proteins spend in each of the classified modes. The standard error of the mean of the average diffusion rate for all points is below $0.02 \mu\text{m}^2 \text{s}^{-1}$. **c** Inter-state transition probabilities of the hidden Markov models. Arrows between triangles show transition probabilities for the transitions between the respective scanning modes, colored according to the energy barrier

data, and always selected more than three states to be the best fit for all proteins (8 states for wm- and wt-EndoV, and 6 states for hOGG1). We investigated the dynamics of mode switching for the three-state HMM classification. Looking at all possible transitions for wt- and wm-EndoV between three consecutive states, we find that while the transition matrix indicates roughly equal probabilities for molecules in state 2 (helical sliding) to switch into state 1 or state 3, the relative probability of switching back to the same state from which they entered state 2 is double the relative probability of switching into the third state (Supplementary Fig. 2). Thus, transitions from state 2 are dependent upon the transition into state 2 which may resemble some level of molecular memory. This manifestation of non-Markovian transitions could explain why vbSPT fails to converge on a reasonable number of states, instead adding states to approximate the non-Markovian behavior.

To further investigate the origin of excess states in vbSPT, we looked into the transition probability matrices for all HMM with 4 or more states to investigate the possibility of population heterogeneity. For a total of 13 models, all but 2 represented reducible transition matrices, while the last two models were nearly reducible (4-state models for wt- and wm-EndoV). In contrast to this, the 3-state models for all proteins investigated had irreducible transition matrices. Interestingly, we also find that vbSPT clearly prefers a 3-state hidden Markov model to a 2-state model, which resonates well with the energy barrier-based classification we employ here.

The results of both the energy barrier and the HMM classifications are shown in Fig. 2b. The average diffusion rate of each mode in the energy barrier-based classification is calculated by averaging the instantaneous diffusion rates over all steps within that particular mode. Since we use a moving window of five frames to calculate the instantaneous diffusion rate, the last four frames of each trajectory are not assigned any value for the instantaneous diffusion rate, and are thus excluded from the energy barrier classification. Because of this, we excluded the same four frames of each trajectory after the HMM classification, and recalculated the diffusion rate and occupancy of each Markov state. The recalculated values are very close to the original values reported by the HMM without frame correction (Supplementary Fig. 3), and the average diffusion rate and occupancy of each of the three Markov states are in good agreement with the initial energy barrier-based classification (Fig. 2b).

This consistency extends to the classification of individual points in trajectories, with 73% classification accuracy and a strong correlation between the results of the two independently performed classifications. The transition probabilities between the hidden Markov states for the three-state classifications are shown in Fig. 2c. For all three proteins the probability of switching from state 2 (helical sliding) to state 1 (interrogation mode) is consistently larger than for switching from state 3 (high mobility hopping) to interrogation mode. We also find that the transition

from high mobility hopping to helical sliding is more likely than going directly from hopping to interrogation mode. We interpret this as a clear indication that a helical sliding mode is central to efficient and reliable base interrogation and damage recognition.

Role of EndoV wedge motif in scanning. Despite an ideal expectation that the amplified fragments of DNA used in our experiments were undamaged, proteins clearly entered intermittent periods of confinement at randomly distributed locations along the DNA. These periods are represented by the $E_a > 2 k_B T$, or interrogation (recognition) mode of scanning. As is shown in Fig. 2a, a large proportion of the instantaneous diffusion rate distribution for both wt-EndoV and hOGG1 lies within the interrogation mode, and a separate peak is clearly visible, around one order of magnitude higher. There is a clear distinction between the instantaneous diffusion rate distributions for these two proteins and the distributions of their corresponding single-mode simulated random walks. In contrast, the distribution for the wedge-deficient mutant EndoV (wm-EndoV) more closely resembles the corresponding simulated random walk than the others. The distribution for wm-EndoV lacks the clear interrogation peak, and this is reflected in the much lower interrogation-state occupancy calculated by both classification methods in Fig. 2b. In addition, using the classification of points along each trajectory from the hidden Markov-based classification, we can estimate the frequency of switching to the interrogation mode per 1000 bp DNA traversed. This is measured to be 1.51, 0.68 and 0.27 for hOGG1, wt-EndoV and wm-EndoV, respectively. These clear disparities between the scanning modes of wt- and wm-EndoV strongly suggest that the wedge motif plays a fundamental role in switching between search and interrogation mode. The crystal structure of wt-EndoV in complex with both deaminated DNA²⁶ and an insertion–deletion mismatch loop²⁷ show that the wedge penetrates the DNA helix to partly separate the two DNA strands at the site of lesion. The wedge locks the protein at the site of damage as it binds to an inherent weak point in the DNA. This is in line with the observation that EndoV might also recognize and process a large variety of DNA structures with anomalous DNA stacking such as AP-sites^{29,32}, base mismatches^{32,35}, ID loops, hairpins, flaps and

pseudo-Y structures.³⁶ The high frequency of switching to the interrogation mode and the high occupancy of this mode for hOGG1, as calculated by both classification methods (Fig. 2b), corresponds well with the hypothesis that, in the interrogation mode, hOGG1 examines flipped-out bases one by one in an exosite pocket on the protein surface^{44,45}.

Salt dependence of diffusion rate and hopping. It has been suggested that friction between proteins and the DNA tracks on which they move typically reduces the diffusion rate by a factor of 2 to 5 compared to the theoretically calculated upper limit of diffusion for helical sliding⁴²; this corresponds to an activation energy barrier of around $0.7\text{--}1.6 k_B T$. The lowest experimental activation energy barrier reported for helical sliding has been $0.6 k_B T$ ²², therefore we assume that an activation energy barrier of $0.5 k_B T$ is a reasonable threshold, below which proteins cannot purely follow a helical path. In our experiments, EndoV and hOGG1 spend around 35% and 8% of the time in $E_a < 0.5 k_B T$ respectively, of which ~50% belongs to a range where $E_a < 0$. This is clear evidence that for a considerable proportion of its interaction lifetime, EndoV is able to travel faster along DNA than would be possible if strictly limited to helical sliding. We hypothesize that in this mode, the protein is disengaged but remains electrostatically confined to the immediate vicinity of the DNA, enabling a striking increase in mobility through 1D diffusion utilizing the mechanism of hopping⁴ (Fig. 3 and Fig. 4). Since higher ionic strengths increase the chance of dissociation, and reduce the chance of reassociation between DNA and protein, the occupancy of any diffusion state which depends on dissociation of the protein from the DNA should be elevated by higher salt concentrations. To verify whether significant hopping exists for the proteins under study, we investigated the salt dependence of the average diffusion rates for each of the three scanning modes classified according to the activation energy barrier. The average diffusion constant of EndoVs for scanning mode with $E_a < 0.5 k_B T$ shows a 1.5-fold increase at higher salt concentrations (Fig. 3), while all other modes of diffusion for EndoV were unaffected. None of the diffusion modes for hOGG1 were salt dependent. This is strong evidence that EndoV utilizes hopping during the $E_a < 0.5 k_B T$ mode, since a higher salt concentration

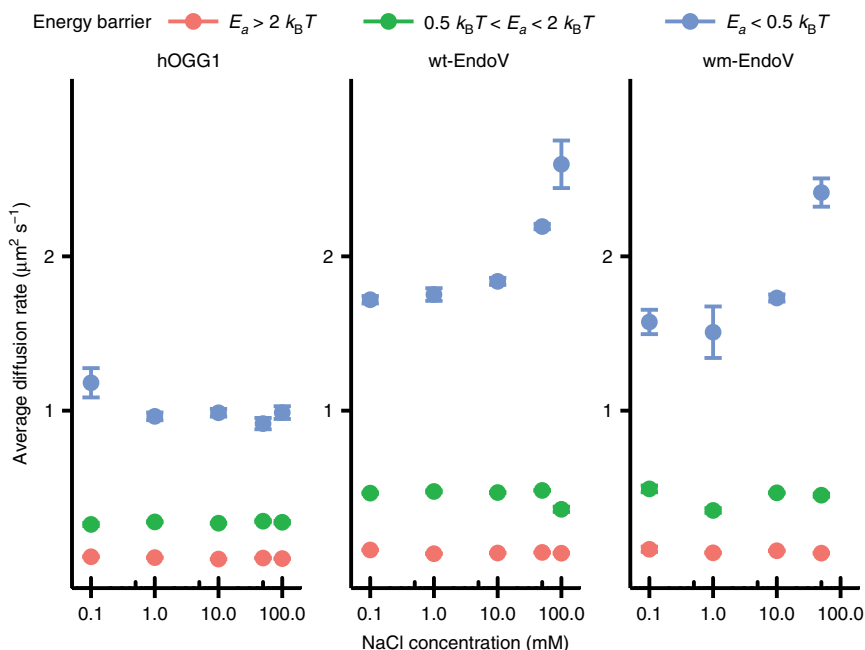


Fig. 3 Diffusion rate analysis by buffer salt concentration. Average diffusion rates for each mode of diffusion as a function of buffer salt concentration. Error bars show the standard error of the mean (SEM) and are not visible when smaller than the size of the points in the plot

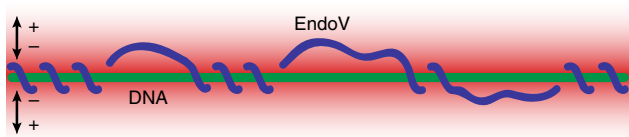


Fig. 4 Model of interspersed helical sliding and hopping on DNA. The protein trajectory (blue trace) is confined within the radial electrostatic field (red to white) surrounding the DNA (green) to achieve rapid reassociation to DNA and efficient target localization during DNA scanning. Number of turns and length of hopping steps are unknown parameters and not to scale in this scheme

increases the amount of time spent in a higher mobility state, thereby increasing the average diffusion rate. On the other hand, low occupancy of the $E_a < 0.5 k_B T$ state for hOGG1 indicates that helical sliding is the dominant scanning strategy for hOGG1, which is consistent with previous reports⁷. If hopping does occur for hOGG1, it is very short-lived and salt dependence cannot be detected within our experimental and analytical resolution.

Multimode fast scanning and mode switching. We have shown that the three modes of scanning based on activation energy barrier $E_a > 2 k_B T$, $0.5 k_B T < E_a < 2 k_B T$, $E_a < 0.5 k_B T$ are representative of DNA interrogation, helical sliding, and hopping, respectively. The wedge motif of EndoV plays an important role in switching from scanning to interrogation mode, and the protein uses a mix of helical sliding and short helically decoupled hopping in the fast DNA search mode; this suggests that the interactions between protein and DNA alternates between states with markedly different activation energy barriers, allowing distinct diffusion modes. Crystallographic data show that the free and DNA-bound EndoV are structurally similar²⁶, which rules out structural changes of the protein as the mechanism for switching between 1D hopping and helical sliding during scanning. Moreover, EndoV is a monomeric, single-binding-site protein too small to clamp around DNA, or delegate hopping and sliding to alternate binding sites or different conformational forms of the protein regulated by binding of a cofactor such as adenosine triphosphate (ATP), as seen for other proteins that are able to switch between helical sliding and hopping^{11,12,19,20,22–24}. Despite this, the experimental evidence consistently shows interspersed phases of helical sliding and rapid 1D hopping for EndoV. This is compatible with a model for scanning which allows microscopic dissociation–reassociation events during which the protein remains constrained within a radial electrostatic field surrounding the DNA. Diffusing in 1D, but decoupled from the constraints of the helical path, the protein is able to adopt an even faster means of scanning the DNA (Fig. 4 and Supplementary Movie 2). These results have shown that there is significant detail in these protein–DNA interactions which has not previously been revealed due to technical limitations, and it will be extremely instructive to determine how ubiquitous these modes of interaction are among other classes of DNA-associated proteins.

Methods

Protein expression and purification. *Escherichia coli* BL21 (DE3) RIPL Codon Plus cells (Stratagene) containing pET28b plasmids with the full-length sequences of *Thermotoga maritima* EndoV mutant D43A (wt-EndoV) or D43A/wedge mutant (⁷⁹PYIP⁸²→GGGG; wm-EndoV) were incubated in LB medium supplemented with 50 $\mu\text{g ml}^{-1}$ kanamycin at 37 °C. Protein expression was induced by adding 0.5 mM IPTG when the OD_{600nm} reached ~0.75, followed by overnight incubation at 18 °C. Cells were harvested by centrifugation at 27,000 $\times g/4$ °C for 20 min and resuspended in buffer A (50 mM NaCl, 20 mM MES, pH 6.5). Cells were disrupted by sonication and centrifuged at 27,000 $\times g/4$ °C for 30 min. The protein extracts were incubated at 75 °C for 15 min in order to precipitate a large fraction of *E. coli* proteins, followed by a second centrifugation step. The supernatants were loaded onto a 5 ml HiTrap SP XL column (GE Healthcare) equilibrated with buffer A. The protein was eluted using a linear salt gradient to 1 M NaCl in buffer A. Fractions rich in *Tma* EndoV were pooled, concentrated and applied onto a Superdex75 size exclusion column (GE Healthcare) equilibrated

with buffer B (100 mM NaCl, 20 mM MES, pH 6.5). Protein fractions containing pure *Tma* EndoV were identified by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel.

The wedge mutant (⁷⁹PYIP⁸²→GGGG) was generated using the Quik-Change Site Directed Mutagenesis kit (Stratagene) and performed according to the manufacturer's protocol. The four glycine residues were simultaneously introduced into the full-length D43A EndoV sequence by using a complementary primer pair containing the glycine-specific codons (underlined) in the wedge position: 5'-AGGGGAGAGATAACTTTTGGCGGCGGTGGGGGGCTCCTTGCTTTCAGAGAAGG-3' and 5'-CCTTCTCTGAAAGCAAGGAGCCCCACCGCCGCAAAAGTTATCTCTCCCCT-3' (Eurofins MWG/Operon). The mutant construct was verified by sequencing, and the expression plasmid was subsequently transformed into *E. coli* BL21 (DE3) RIPL Codon Plus cells for protein expression. The inactive D43A mutant was used to avoid any disruptive cleavage of DNA substrates during scanning experiments.

An S118C mutant of a truncated 12–327 hOGG1 protein was generated using the Quik-Change Site Directed Mutagenesis kit (Stratagene) according to the manufacturer's protocol. The S118 was mutated to a Cys residue to allow attachment of the fluorescent ATTO N647 reporter dye. The primers used to produce the mutation were as follows: 5'-CACTGGGTTGCCGTGGACTGCCACTTCCAAGAGTG-3' and 5'-CACCTCTTGGAAGTGGCAGTCCACGGAACCCAGTG-3'. The mutant construct was verified by sequencing and transformed into *E. coli* BL21 (DE3) RIL Codon Plus cells. The protein was purified as previously described⁴⁶.

Fluorescent labeling of the proteins. To directly observe the position of single proteins in real time, the fluorescent dye ATTO 647N (ATTO-TEC) was bound by maleimide-NH coupling to a native cysteine in the EndoV (C154), and a genetically engineered cysteine in Ogg1 (C118). The cysteines in both EndoV and hOGG1 are located on the protein surfaces far from the damage recognition pockets (active sites). Prior to labeling, all proteins were extensively dialyzed against 1 \times phosphate-buffered saline (PBS) buffer for 12–18 h at 4 °C. For the labeling procedure, 1 ml of 40–100 μM protein in 1 \times PBS buffer was mixed with 1.3-fold molar excess of ATTO 647N and incubated at room temperature in the dark for between 0.5 and 3 h. Free dye molecules were separated from the labeled proteins using a NAP-5 column (GE Healthcare) with 1 \times PBS as elution buffer. Fractions rich in labeled proteins were identified by measuring the absorbance at 280 and 647 nm (A_{280} and A_{647}) using a NanoDrop One instrument (Thermo Scientific). The labeling efficiency for different batches of proteins varied between 30 and 80% as determined using A_{280} , A_{647} and the molar extinction coefficients of the protein and the dye, respectively.

λ -DNA substrate preparation. In order to anchor one end of the DNA to the surface of a coverslip and attach the other end to a polystyrene bead, as shown in Fig. 1b, a 12 kbp linear fragment of λ -DNA with biotin and digoxigenin tags at each end was designed. This DNA substrate was prepared by PCR amplification of λ -DNA using primers modified with 5' biotin (5'-bio-ACTTCGCTTCTTCC CATT-3') and 5' digoxigenin (5'-dig-ATCTCGCTTCCACTCCAGA-3') (Eurofins MWG/Operon). The PCR reaction was performed in a 1 \times LongAmp Taq Reaction Buffer, with a final volume of 50 μl , 300 μM of each dNTP, 0.4 μM of each primer, 2 units LongAmp Taq DNA polymerase (New England Biolabs) and 0.1 ng μl^{-1} of λ -DNA template. The PCR included an initial denaturation step at 94 °C for 3 min, 35 cycles of denaturation (94 °C for 15 s), annealing (60 °C for 60 s) and primer extension (65 °C for 16 min), followed by a final extension step at 65 °C for 10 min. The quantity and quality of the PCR product was analyzed by spectrophotometry (260/280 absorbance measurements) and gel electrophoresis. The PCR product was stored in aliquots at 4 °C and diluted immediately before use.

Polystyrene bead functionalization. In order to attach a polystyrene bead to the digoxigenin-modified end of the λ -DNA, the beads were coated with anti-digoxigenin antibody. The method is mainly adapted from the manufacturers' manuals for the various materials used, in addition to Bangs laboratories' protocol for covalent coupling. Next, 50 μl of 100 mg ml^{-1} carboxylate-modified polystyrene beads (diameter = 0.9 μm , Sigma-Aldrich) was suspended in 500 μl of MES buffer (50 mM MES, pH 6), mixed well and centrifuged at 10 $\times g$ for 5 min before the supernatant was discarded from the tube. This washing step was repeated three times. To make the surface of the carboxylate-modified beads amino-reactive, the beads were resuspended and incubated with 100 μl of 50 mg ml^{-1} EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, ThermoFisher) in MES buffer for 30 min at room temperature. To quench the reaction and remove the excess EDC, the beads were washed twice with MES buffer as explained above. Next, to facilitate the reaction of amino-reactive beads with the amine-containing antibody, the pH was adjusted to 8.5 by washing twice with borate buffer (0.1 mM sodium tetraborate, pH 8.5 adjusted with HCl). Then, the beads were resuspended and incubated with 100 μl of 0.5 mg ml^{-1} anti-digoxigenin (Fab fragments from sheep, Roche) in borate buffer for 4 h with mild shaking at room temperature. After incubation, the beads were washed once with borate buffer, then washed and incubated with 50 mM TRIS at pH 8 for 2 h at room temperature. Finally, the beads were washed once and stored in 1 \times PBS (pH 7.5), 2 mg ml^{-1} bovine serum

albumin (BSA) and 0.1% Tween-20. Batches with functionalized beads were stored at 4 °C and could be used for up to 1 month without any significant degradation of the functionality.

Surface preparation for DNA anchoring. In order to prepare coverslip surfaces for the anchoring of biotin-tagged DNA (Fig. 1b), and to passivate the surface preventing nonspecific binding, we performed a series of surface treatment operations including cleaning, functionalization and passivation. The method is adapted from previously published protocols^{47–51} and further modified after extensive experimentation, and with input from the manufacturers' manuals.

(I) Initially, 6–8 coverslips (24 × 60 mm, Menzel Gläser) were placed in a glass-staining dish and rigorously cleaned in the following sequence: (i) sonication in pure ethanol for 30 min; (ii) rinsing with MiliQ filtered (MQF) water 3 times; (iii) sonication in 1 M potassium hydroxide for 30 min; and (iv) rinsing with MQF water 3 times. This series of cleaning steps (i)–(iv) were repeated 3 times over. Afterwards, the coverslips were carefully rinsed with acetone to remove any traces of previous solutions.

(II) To functionalize the surface with amine groups, the cleaned coverslips were incubated with 2% (v/v) solution of 3-aminopropyltriethoxysilane (Sigma-Aldrich) in dry acetone at room temperature for 2–4 min, immediately followed by thorough rinsing with MQF water to remove any residual acetone, then dried in a 100 °C oven for 30 min.

(III) Simultaneous functionalization for DNA binding and passivation towards nonspecific binding was achieved by saturating the coverslip surface with a mixture of amino-reactive polyethylene glycol (PEG-NHS, molecular weight (MW) = 5000 Da, Nanocs) and biotinylated PEG-NHS (Biotin-PEG-NHS, MW = 5000 Da, Nanocs). Then, 120 μ l of a mixture of 150 mg ml⁻¹ PEG-NHS and 0.1 mg ml⁻¹ Biotin-PEG-NHS in 0.1 M sodium bicarbonate solution was sandwiched between two amino-functionalized coverslips and kept in a humid chamber overnight. The following day, the two coverslips were separated and washed carefully with MQF water and blow-dried with nitrogen gas. These PEGylated surfaces were stored in a vacuum and could be used for up to 3 weeks after preparation.

(IV) Prior to the main single-molecule experiment, 120 μ l of streptavidin (from *Streptomyces avidinii*, Sigma-Aldrich) with a concentration of 0.01 mg ml⁻¹ in 1 × PBS was added to the PEGylated surface of one coverslip, and sandwiched with another clean but not functionalized coverslip for 3–15 min, depending on the required density of binding sites. The PEGylated coverslip was washed, blow-dried with nitrogen and immediately used in the construction of the flow chamber.

Flow chamber and surface-DNA-bead construction. To construct the experimental flow chamber with a sample volume of around 20–30 μ l, a streptavidin-coated coverslip was fused to a pre-cleaned microscope slide containing two holes for inlet and outlet flow on its surface, using double-sided tape. The chamber was connected to a 100 μ l syringe (Model 1710 LT SYR, Hamilton) using silicon and PEEK tubes (1/16-inch outer diameter, Sigma-Aldrich). The syringe was connected to a pump (Harvard Apparatus PHD 2000) with sensitive control over the flow rate. A three-way valve (V100T, Upchurch Scientific) was positioned in the flow circuit allowing the user to switch the flow between a supply tank of solution and the reaction chamber. After assembly, the flow chamber was sequentially washed by flowing 200 μ l of washing buffer (25 mM TRIS, pH 7.5) followed by 200 μ l of blocking buffer (25 mM Tris-HCl, pH 7.5, 2 mM EDTA, 1–3 mg ml⁻¹ BSA, 0.01% (v/v) Tween-20) with a flow rate of 50 μ l min⁻¹. The flow chamber was incubated with blocking buffer for 1–2 h. Following this, 200 μ l of 5–50 ng ml⁻¹ of λ -DNA substrate was injected with a flow rate of 10 μ l min⁻¹, and incubated for 5–30 min depending on the required density of DNA on the surface. Unbound DNA was washed away by the injection of 400 μ l of 10% blocking buffer in washing buffer with a flow rate of 10 μ l min⁻¹. Anti-digoxigenin-coated polystyrene beads with a concentration of 10–100 μ g ml⁻¹ in 10% blocking buffer in washing buffer were mixed and sonicated for 30 s and injected with a flow rate of 5 μ l min⁻¹. Depending on the overall efficiency of the sample preparations, beads bound to the anchored DNA molecules within 30–90 min. It is worth mentioning that high flow rates (more than 50 μ l min⁻¹) were avoided to prevent ruptures of the surface-DNA-bead construction. In addition, sudden changes in the flow rate and air bubbles in the system were avoided to prevent anchored DNAs from being torn off the surface.

Excess unattached beads were carefully washed off with 200 μ l washing buffer (containing the appropriate concentration of NaCl, depending on the experiment) with a flow rate of 5–20 μ l min⁻¹. Finally, the fluorescently labeled proteins were added to the flow chamber with a concentration of 0.1–2 nM, in the appropriate assay buffer at a rate of 5–20 μ l min⁻¹, and the interaction with the trapped DNA was recorded as described below.

High-concentration batches of labeled proteins (1–100 μ M) were kept on ice, and immediately injected after dilution in assay buffer and subsequent thermal equilibration to room temperature. Note that care should be taken to avoid too much protein entering the system, since excess amount of protein in the flow chamber can lead to accumulation of protein molecules on the DNA and/or the coverslip surface, with an increase in the noise level and a drastically lower ability to detect and record single-molecule trajectories. It was also critical to constantly supply the reaction chamber with fresh protein.

Holographic optical tweezers and imaging setup. Individual flow chambers were fixed onto the piezo-steerable stage of a custom-built combined holographic optical tweezer and subdiffraction resolution microscope which has previously been described³⁸. In brief, a spatial light modulator was used to steer the infrared trapping laser within the sample. The trapping beam and excitation laser were coupled into the microscope light path and the objective lens, allowing simultaneous trapping and fluorescence microscopy. The entry position of the beam into the objective lens was controlled using an adjustable mirror on the excitation light path, making it possible to generate a highly inclined and laminated optical sheet³⁹ for illumination of the sample. This resulted in concentration of laser power in the vicinity of the surface and a drastic drop in background noise from the sample volume, and was critical to achieving a high signal-to-noise ratio (SNR). The light emitted from the fluorophores was collected with the same objective lens and imaged on a high sensitivity EMCCD camera controlled by Micro-Manager software⁵². Taking advantage of high SNR along with high light collection efficiency detection, the exposure time of imaging was pushed down to 7.5 ms for EndoVs and 23.5 ms for hOGG1, allowing the observation and extraction of transient behaviors on very short times scales in the scanning process. With these exposure times, an average of 47–143, 42–132 and 45–173 photons were collected per frame for single wt-EndoV, wm-EndoV and hOGG1 molecules, respectively.

Image processing and trajectory tracking. Each data set (video stream recorded as consecutive TIFF files containing between 50,000 and 200,000 frames) was processed using the ThunderSTORM⁵³ plugin in FIJI⁵⁴. All signals within each frame were localized by Gaussian fitting and registered according to the frame number (as a time series). The lower panel of Fig. 1c shows the projection of all signals in one data set into a single image. Localization precision (the average standard deviation of the peak of the Gaussian fits) of several immobile hOGG1, wt-EndoV and wm-EndoV were measured between 20–33 nm, 23–39 nm and 22–42 nm, respectively, to confirm localization precision.

Later, using code custom-developed⁴⁰ in R, the signal from diffusing proteins were tracked, separated from the background noise and registered as uninterrupted trajectories. For each data set (containing up to hundreds of trajectories occurring separately along a single DNA strand) in which the position of the bead and DNA remain unchanged, the beads were localized and several trajectories were visually inspected; these two details were then used to localize the DNA. Applying a rotation matrix, the coordinates were rotated such that the localized DNA was aligned along the *x*-axis, meaning that all trajectories along DNA were similarly oriented, and a boundary filter (200 nm either side of the DNA) was set on the *y*-axis to exclude molecules localized far from the target DNA. Within these limits, trajectories with signals that are present in more than 5 consecutive frames, and which had moved at least 300 nm along the DNA (at least once) during the trajectory, were retained and the remainder of analysis was conducted on this subset of the data. The effect of blinking of the fluorescent dyes was mitigated slightly by allowing two temporally adjacent trajectories to be connected if they were separated by only one frame, and the detected protein in these adjacent frames were localized within 600 nm of one another. Supplementary Movie 1 shows a sequence of trajectories which are cut from different data sets, concatenated and visualized using the TrackMate⁵⁵ plugin in FIJI.

Instantaneous and average diffusion analysis. In order to compute the instantaneous diffusion coefficient, a moving window of width 5 frames (37.5 ms for EndoVs and 117.5 ms for hOGG1) was used. As the window moves along the trajectories at each step the mean squared displacement ($\langle x^2 \rangle$) was calculated for the next 5 steps. From the diffusion equation⁵⁶ for 1D diffusion $\langle x^2 \rangle = 2Dt$, the diffusion coefficient was calculated as the instantaneous diffusion coefficient at that particular step. To compute the average diffusion coefficient of proteins in different diffusion modes the instantaneous diffusion coefficient was averaged over particular segments of trajectories belonging to either of modes.

Code availability. All source codes used in the analysis pipeline are available on GitHub: <https://doi.org/10.5281/zenodo.1487773>

Data availability

The time series position data for all trajectories are available on Figshare: <https://figshare.com/s/59402119aa83c4abb4c9>. All data including the raw image data captured for this study are available from the corresponding authors upon reasonable request.

Received: 4 April 2018 Accepted: 27 November 2018

Published online: 19 December 2018

References

1. Friedberg, E. C. et al. *DNA Repair and Mutagenesis* (ASM Press, Washington, 2006).

2. Hoeijmakers, J. H. J. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).
3. Wood, R. D., Mitchell, M., Sgouros, J. & Lindahl, T. Human DNA repair genes. *Science* **291**, 1284–1289 (2001).
4. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
5. Riggs, A. D., Bourgeois, S. & Cohn, M. The lac repressor-operator interaction. III. Kinetic studies. *J. Mol. Biol.* **53**, 401–417 (1970).
6. Tafvizi, A., Mirny, L. A. & Van Oijen, A. M. Dancing on DNA: kinetic aspects of search processes on DNA. *Chemphyschem* **12**, 1481–1489 (2011).
7. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA* **103**, 5752–5757 (2006).
8. Blainey, P. C. et al. Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16**, 1224–1229 (2009).
9. Dunn, A. R., Kad, N. M., Nelson, S. R., Warshaw, D. M. & Wallace, S. S. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.* **39**, 7487–7498 (2011).
10. Komazin-Meredith, G., Mirchev, R., Golan, D. E., van Oijen, A. M. & Coen, D. M. Hopping of a processivity factor on DNA revealed by single-molecule assays of diffusion. *Proc. Natl Acad. Sci. USA* **105**, 10721–10726 (2008).
11. Bonnet, I. et al. Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* **36**, 4118–4127 (2008).
12. Gorman, J. et al. Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl Acad. Sci. USA* **109**, E3074–E3083 (2012).
13. Gowers, D. M., Wilson, G. G. & Halford, S. E. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl Acad. Sci. USA* **102**, 15883–15888 (2005).
14. Porecha, R. H. & Stivers, J. T. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl Acad. Sci. USA* **105**, 10791–10796 (2008).
15. Bauer, M. & Metzler, R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **102**, 2321–2330 (2012).
16. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
17. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
18. Bénichou, O., Kafri, Y., Sheinman, M. & Voituriez, R. Searching fast for a target on DNA without falling to traps. *Phys. Rev. Lett.* **103**, 1–4 (2009).
19. Kochaniak, A. B. et al. Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.* **284**, 17700–17710 (2009).
20. Cuculis, L., Abil, Z., Zhao, H. & Schroeder, C. M. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* **6**, 7277 (2015).
21. Cuculis, L., Abil, Z., Zhao, H. & Schroeder, C. M. TALE proteins search DNA using a rotationally decoupled mechanism. *Nat. Chem. Biol.* **12**, 831–837 (2016).
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. USA* **108**, 563–568 (2011).
23. Schwarz, F. W. et al. The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science* **340**, 353–356 (2013).
24. Etson, C. M., Hamdan, S. M., Richardson, C. C. & van Oijen, A. M. Thioredoxin suppresses microscopic hopping of T7 DNA polymerase on duplex DNA. *Proc. Natl Acad. Sci. USA* **107**, 1900–1905 (2010).
25. Tafvizi, A. et al. Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **95**, L01–L03 (2008).
26. Dalhus, B. et al. Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **16**, 138–143 (2009).
27. Rosnes, I. et al. Structural basis of DNA loop recognition by endonuclease v. *Structure* **21**, 257–265 (2013).
28. Demple, B. & Linn, S. On the recognition and cleavage mechanism of Escherichia coli endodeoxyribonuclease V, a possible DNA repair enzyme. *J. Biol. Chem.* **257**, 2848–2855 (1982).
29. Yao, M., Hatahet, Z., Melamed, R. J. & Kow, Y. W. Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli. *J. Biol. Chem.* **269**, 16260–16268 (1994).
30. Yao, M. & Kow, Y. W. Interaction of deoxyinosine 3'-endonuclease from Escherichia coli with DNA containing deoxyinosine. *J. Biol. Chem.* **270**, 28609–28616 (1995).
31. Weiss, B. Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation. *DNA Repair (Amst.)*, **7**, 205–212 (2008).
32. Yao, M. & Kow, Y. W. Further characterization of Escherichia coli endonuclease V. *J. Biol. Chem.* **272**, 30774–30779 (1997).
33. He, B., Qing, H. & Kow, Y. W. Deoxyxanthosine in DNA is repaired by Escherichia coli endonuclease V. *Mutat. Res. DNA Repair* **459**, 109–114 (2000).
34. Hitchcock, T. M., Gao, H. & Cao, W. Cleavage of deoxyoxanosine-containing oligodeoxyribonucleotides by bacterial endonuclease V. *Nucleic Acids Res.* **32**, 4071–4080 (2004).
35. Yao, M. & Kow, Y. W. Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3'-endonuclease from Escherichia coli. *J. Biol. Chem.* **269**, 31390–31396 (1994).
36. Yao, M. & Kow, Y. W. Cleavage of insertion / deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from Escherichia coli. *J. Biol. Chem.* **271**, 30672–30676 (1997).
37. Fladeby, C. et al. The human homolog of Escherichia coli endonuclease V is a nuclear protein with affinity for branched DNA structures. *PLoS ONE* **7**, 1–11 (2012).
38. Diekmann, R. et al. Nanoscopy of bacterial cells immobilized by holographic optical tweezers. *Nat. Commun.* **7**, 13711 (2016).
39. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 1–7 (2010).
40. Ahmadi, A., Dalhus, B. & Rowe, A. D. Analysis of DNA scanning by EndoV and hOGG1 proteins: the source code. <https://doi.org/10.5281/zenodo.1487773> (2018).
41. Schurr, J. M. The one-dimensional diffusion coefficient of proteins absorbed on DNA. Hydrodynamic considerations. *Biophys. Chem.* **9**, 413–414 (1975).
42. Bagchi, B., Blainey, P. C. & Sunney Xie, X. Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B* **112**, 6282–6284 (2008).
43. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods* **10**, 265–269 (2013).
44. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
45. Lee, S., Radom, C. T. & Verdine, G. L. Trapping and structural elucidation of a very advanced intermediate in the lesion-extrusion pathway of hOGG1. *J. Am. Chem. Soc.* **130**, 7784–7785 (2008).
46. Dalhus, B. et al. Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase. *Structure* **19**, 117–127 (2011).
47. Sofia, S. J., Premnath, V. & Merrill, E. W. Poly(ethylene oxide) grafted to silicon surfaces: grafting density and protein adsorption. *Macromolecules* **31**, 5059–5070 (1998).
48. Ha, T. et al. Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature* **419**, 638–641 (2002).
49. van Oijen, A. M. et al. Single-molecule kinetics of lambda DNA exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235–1238 (2003).
50. Groll, J. et al. Biofunctionalized, ultrathin coatings of cross-linked star-shaped poly(ethylene oxide) allow reversible folding of immobilized proteins. *J. Am. Chem. Soc.* **126**, 4234–4239 (2004).
51. Tanner, N. A. & van Oijen, A. M. Visualizing DNA replication at the single-molecule level. *Methods Enzymol.* **475**, 259–278 (2010).
52. Edelstein, A. D. et al. Advanced methods of microscope control using µManager software. *J. Biol. Methods* **1**, 10 (2014).
53. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).
54. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
55. Tinevez, J. Y. et al. TrackMate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
56. Berg, H. C. *Random Walks in Biology* (Princeton University Press, Princeton, 1993).

Acknowledgements

We acknowledge South-East Norway Regional Health Authorities (grants 2014034 and 2015095 to B.D.), the Research Council of Norway (program FRIMEDBIO to M.B., program DAAD project 281255 to B.D.), the German Academic Exchange Service (grant no. 57402206 to M.S.) and the University of Oslo, MLS^{UIO} program (PhD grant to A.A.) for funding. We thank Professor Oddmund Bakke at Oslo NorMIC imaging platform, University of Oslo, for access for initial testing of flow cells, DNA and proteins.

Author contributions

A.A., M.B., B.D. and A.D.R. conceived the study; A.A., I.R., P.B. and B.D. prepared DNA and proteins; A.A., I.R., R.D., M.S. and A.D.R. contributed to single-molecule DNA scanning data collection; A.A., K.G., J.T., M.B., B.D. and A.D.R. participated in flow-cell design; A.A. and A.D.R. processed and analyzed the data; A.A., M.B., B.D. and A.D.R. interpreted the data, and discussed results and wrote the paper with contributions from the other authors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-07797-4>.

Competing interests: The authors declare no competing interests. I.R. is currently employed by F. Hoffmann-La Roche (Roche Norge AS). The data included in this paper are based on research which has had no influence or involvement by F. Hoffmann-La Roche by any means. Any personal views of I.R. should not be understood or quoted as being made on behalf of or reflecting the position of F. Hoffmann-La Roche.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018