University of Groningen

A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA

Blainey, Paul C.; van Oijen, Antoine M. ; Banerjee, Anirban; Verdine, Gregory L.; Sunney Xie, X.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Supporting Information

Signal Tracking and Data Analysis. DIATRACK software (Semasopht, North Epping, Australia) is used to fit images of the moving protein molecules with two-dimensional Gaussian functions and create displacement trajectories. Trajectories are analyzed using MATLAB software (MathWorks, Natick, MA). Trajectories from signals appearing for five or more frames in the region of the stretched DNA molecule that exhibit transverse Brownian motion consistent with the DNA fluctuations in that direction are retained. Fig. 4*A* shows ensemble-averaged mean-square displacement based on raw trajectories. Upon conversion of the mean-square displacement fit slopes to one-dimensional diffusion constants (appearing in Fig. 4 *B* and *C*), a small correction (+10%) is applied to compensate for incomplete stretching of the DNA in our assay. Diffusion constant error estimates are determined from the quality of the linear fits and the distribution of fit parameter values measured from different data sets collected under the same experimental condition.

Accuracy of Centroid Determination. The accuracy with which we can determine the centroid position of a photon source in our samples is principally limited by the number of photons collected. We collect between 1,000 and 4,000 photons per second from Cy3B. Given the integration time of 10–50 ms, an average of 10–200 photons appear per molecule per frame (N). Using the width of the microscope point-spread fuction (s), pixel size in the image (a), and the standard deviation of the background level (b), we calculate the standard error of positioning human oxoguanine DNA glycosylase 1 (hOgg1) (σ) to be 10–50 nm according to (1):

$$\sigma = \left[\frac{s^2}{N} + \frac{a^2/12}{N} + \frac{8\pi s^4 b^2}{a^2 N^2} \right]^{\frac{1}{2}}$$

(observations of labeled protein molecules adsorbed on the coverslip surface confirm the theoretical spatial resolution).

Determination of Solution Flow Velocity at the DNA. The strength of polymer stretching under shear flow can be characterized by the dimensionless Weissenberg number (*Wi*), defined as the product of the polymer relaxation time ($\tau = 0.2$ s for λ) and the magnitude of the solution velocity gradient, or "shear rate" ($\dot{\gamma}$) (2). Observations of DNA-bound labeled protein molecules reveal that this flow rate stretches the DNA to $\approx 80\%$ of its contour length, which corresponds to *Wi* = 100 (2). With this information, we can calculate the $\dot{\gamma}$ to be 500 per s in our assay according to $\dot{\gamma} = Wi/\tau$. Knowing the shear rate of 500 per s, the mean distance of the DNA from the coverslip surface, 0.2 μm (from the Brownian dynamics simulation), and that the flow speed is zero at the surface, we can calculate the flow speed to be 0.2 μm × 500 per s = 100 μm/s at the DNA.

Brownian Dynamics Simulations. In principle, the observed movements of DNA-bound proteins can arise either from facilitated diffusion or fluctuation of the DNA extension. It is known that both the amplitude and time scale of DNA extension fluctuation depend on the flow strength (2). We perform single-molecule experiments at *Wi* = 100, where the DNA conformational fluctuation is fast and does not complicate our measurements of protein diffusion. To determine the time scale of these fluctuations under our experimental condition, Brownian dynamics simulations of the tethered DNA including intramolecular hydrodynamic interactions and a steeply repulsive interaction with the surface are conducted (3).

The relaxation time of the simulated DNA is determined by release from 35% extension. Averaging over the results from 60 repetitions, single-exponential relaxation of the DNA extension with a time constant of 0.192 s is found. For simulation under flow, the 500 per s shear rate matching the assay condition is chosen. The simulation with shear flow reproduces the observed mean fractional extension of 80%. The maximum tension on the DNA, which occurs where a DNA molecule is tethered to the surface, is <2 pN, and the mean DNA segment positions are all within 350 nm of the coverslip surface. The standard deviation of the DNA position in the transverse dimension scales from zero at the anchor point up to 300 nm near the free end. The time scale of transverse fluctuations

is <0.1 s as determined by autocorrelation of DNA segment positions. These transverse fluctuations of the DNA position do not interfere with our measurement of enzyme positions along a DNA molecule. Deviation of the longitudinal position of DNA segments scale to 400 nm at the extreme end of the DNA. The time scale of the longitudinal fluctuations is measured at 0.0158 s near the tether point (red points and fitting function, Fig. 6*A*), and at 0.0172 s at the free end of the DNA (blue points and fitting function, Fig. 6*A*). The noise in the two curves is correlated because the same simulation run was used to generate both. The rapid decay of these autocorrelation functions indicates that fluctuations of the DNA extension do not contribute to our measurements of protein displacement in the longitudinal direction on the time scales 0.1 – several seconds.

Experimental Measurement of DNA Extension Fluctuation Time Scale. As a further control, we performed experiments to monitor the extension of DNA stretched by flow at Wi = 100. We measured the extension of tethered λ DNA molecules stained with intercalating dye for comparison with the simulation results. The fluctuation time scale of stained DNA is expected to be longer than that of unstained DNA by a factor of 1.93 based on the difference in Kuhn step lengths and contour lengths. Fig. 6*B* shows the autocorrelation of two DNA extension versus time trajectories with decay constants 0.0399 and 0.0497 s. When divided by the factor relating fluctuation time scales of stained and unstained DNA, time constants of 0.0207 and 0.0258 s result, revealing approximate agreement with the simulation results.

Susceptibility of Hopping Proteins to Flow. We observe negligible drift of the population of diffusing enzymes in the flow direction (Fig. 3*B*). This observation constitutes additional evidence against hopping as the mechanism underlying hOgg1's one-dimensional diffusion. We can use information about the protein and the flow to calculate the expected population drift for hypothetical hopping protein molecules.

We compute the fraction of time hOgg1 would spend diffusing in solution from the ratio of the observed one-dimensional diffusion constant to the three-dimensional diffusion

coefficient (D_3). Dynamic light scattering measurements show that the protein diffuses with $D_3 = 76$ μm^2/s when free in solution (data not shown) and reveal no evidence of protein aggregation, even at concentrations up to seven orders of magnitude higher than used in single-molecule assay. This diffusion constant is consistent with a calculation based on the Stokes–Einstein relationship using the radius 3.2 nm estimated from the protein crystal structure. We use the measured one-dimensional diffusion constant at pH 7.5, 0.31 μm^2/s, to calculate the fraction of time the proteins spend diffusing in three dimensions (f_{free}) to be 0.0041 according to: $D_1 = D_3 \times f_{free}$. Multiplying by the measured mean binding lifetime for the data shown in Fig. 3B, 0.56 s, we calculate that hypothetical hopping enzymes spend an average of 0.0023 s free of the DNA and subject to the flow of buffer in this experiment. It is a good assumption that the protein will be affected by the full force of flow as soon as it is unbound from the DNA since DNA's behavior as a "slender" hydrodynamic body at high extensions is well known. The expected drift of the population is the product of the time exposed to flow and the rate of flow, or 0.0023 s $\times$ 100 μm/s $= 0.230$ μm. This contradicts the observed mean displacement of the population in the experiment represented in Fig. 3B, which is 0.013 μm. In all experiments conducted, the hOgg1 populations' mean net drift is 0.0 ± 0.030 μm.

Calculation of Mean Sliding Length. Fig. 8 shows a histogram of 62 observed hOgg1 binding times on undamaged DNA at the highest salt concentration tested (0.1 M NaCl, pH 7.0; first bin with 41 samples thrown out to exclude the effect of missed fast events). These data are well fit using a single-exponential function with parameter $1/\lambda = 0.0246 \pm 0.0013$ s. Thus, we consider the protein binding time to be a random variable T distributed according to $f(t) = \lambda e^{-\lambda t}$. To compute the expected sliding length of a sample of proteins sliding with lifetime distributed according to $f(t)$, we consider the density of the random variable $X = (2DT)^{1/2}$, where D is the one-dimensional diffusion constant found for hOgg1 at physiological pH, 5×10^6 bp^2/s. To do so, we take advantage of the cumulative distribution function of X: $F_X(x) = P(X \leq x)$. Because we know the relationship between X and T, we can obtain $F_X(x)$, the integrated density of X: $F_X(x) =$

$P(X \le x) = P(T \le x^2/2D) = 1 - e^{-\frac{\lambda}{2D}x^2}$. Differentiating $F_X(x)$ to obtain the density of X, we

find $f(x) = \dfrac{\lambda}{D}xe^{-\frac{\lambda}{2D}x^2}$. $\langle X \rangle$, the expected sliding length, is easily shown to equal $\sqrt{\dfrac{\pi D}{2\lambda}}$, or 440 bp given $1/\lambda = 0.0246$ seconds and $D = 5 \times 10^6$ bp^2/s.

Experiments with Damaged DNA. WT and K249Q hOgg1 were also tested in the assay using lesion-containing DNA. We created lesions at random locations in λ DNA by photolysis in the presence of riboflavin, a sensitizer known to yield 8-oxoguanine with good specificity (4). We exposed 20 pM λ DNA to 532 nm laser radiation at 50 mW/cm^2 for 60 s (path length = 1.0 cm) in the presence of 10 μM riboflavin at pH 7.5 (0.05 M Tris/0.05 M NaCl). The experiment was then conducted as usual. WT and K249Q hOgg1 presented subdiffusion in the mean-square displacement analysis, with some individual displacement versus time traces showing the sliding activity to cease and hOgg1 to remain fixed at one site on the DNA (within the limits of experimental resolution) until photobleaching of the dye molecule label (data not shown). This stopping activity, observed consistently in experiments with lesion-containing DNA but not in experiments with undamaged DNA, indicates that hOgg1 is capable of recognizing oxoG under the conditions of the single-molecule assay.

1. Thompson, R. E., Larson, D. R. & Webb, W. W. (2002) *Biophys. J.* **82,** 2775–2783.

2. Doyle, P. S., Ladoux, B. & Viovy, J. L. (2000) *Phys. Rev. Lett.* **84,** 4769–4772.

3. Schroeder, C. M., Shaqfeh, E. S. G. & Chu, S. (2004) *Macromolecules* **37,** 9242–9256.

4. Adam, W., Andler, S., Saha-Möller, C. R. & Schönberger, A. (1996) *J. Photochem. Photobiol. B* **34,** 51–58.