Supplementary Materials

Tumor suppressor p53 slides on DNA with low friction and high stability

Anahita Tafvizi* †‡, Fang Huang§, Jason S. Leith†*, Alan R. Fersht§, Leonid A. Mirny*, and Antoine M. van Oijen†||

*Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA; †Harvard Medical School, Dept. of Biological Chemistry and Molecular Pharmacology, Boston, MA, USA; ‡Harvard University, Department of Physics; ¶Harvard University, Program in Biophysics; §University of Cambridge, Dept. of Chemistry, Cambridge, UK

||Corresponding author: antoine_van_oijen@hms.harvard.edu

DNA preparation and flow stretching

Purified DNA from λ phage (New England Biolabs) was linearized and biotinylated at one end by annealing a 3’ biotin-modified oligo (5’AGGTCGCCGCCC3’-biotin; Integrated DNA Technologies) to the complementary λ-phage 5’ overhang. Flow cells (0.1 mm height, 2.0 mm width) with a streptavidin-coated surface were prepared as described previously (1, 2). The streptavidin-coated flow-cell surfaces were blocked by incubation with blocking buffer (Tris 20mM, EDTA 2mM, NaCl 50mM, BSA 0.2mg/ml,
Tween 20 0.005%; pH 7.5) for 20 minutes. Biotin-modified DNA constructs were introduced into the flow cell at a rate of 0.1 mL/min at a concentration of 8 pM for 20 minutes. These conditions resulted in an average density of ~50 surface-tethered DNA molecules per field of view (~50 ×50 µm²).

The single-molecule imaging experiments were performed in an imaging buffer, containing 20 mM HEPES, 0.5 mM EDTA, 2 mM MgCl₂, 0.5 mM DTT, 0.05 mg/mL BSA (pH 7.9), and varying amounts of KCl. Imaging buffer was drawn into the channel by a syringe pump at a flow rate of 0.1 mL/min, creating shear flow near the coverslip surface (3). Single-molecule imaging was done with 1-5 pM p53 in imaging buffer; measurements of protein density on individual DNA molecules were done at higher concentrations (100–300 pM).

**Protein preparation and labeling**

The super stable mutant of human full-length p53 (flp53, residues 1-393) with mutations M133L, V203A, N239Y and N268D (4) was used. Solvent-exposed Cys residues at positions 182, 275 and 227, and the partially buried Cys-124 were all mutated by Ala so that only one exposed Cys (Cys-229) remains. The protein was expressed in *Escherichia coli* and purified as described (5, 6). Cys-229 was labeled with Alexa Fluor 488 maleimide from Invitrogen. The labeling was carried out in phosphate buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.0) with protein concentration of 20-100 mM at 0-4 ºC. 10-fold excess Alexa Fluor 488 maleimide was added after the disulfide bonds were reduced with 1mM of tris(2-carboxyethyl) phosphine (TCEP). The labeling progress was followed by matrix assisted laser desorption/ionization time-of-flight mass
spectrometry (MALDI-TOF MS). The reaction was quenched with 2-10 mM β-mercaptoethenol after ~1 h and the labeled protein was immediately separated from the free dyes on a desalting column. Mass Spectrometry analysis of the purified protein ruled out labeling of the protein at stoichiometric excess.

**Fluorescence imaging**

Fluorescence imaging of the movement of the labeled p53 proteins along DNA was performed by placing the flow cell on top of an inverted microscope (Olympus IX71) and exciting the AlexaFluor 488 label by the 488-nm line from an Ar/Kr laser (Coherent I-70 Spectrum). A high-N.A. microscope objective (Olympus, N.A. = 1.45) was used to illuminate the sample with total-internal reflection. The illuminated area had a diameter of 50 µm at the sample plane. The fluorescence was collected by the same objective and imaged by an EM-CCD camera (Andor iXon), after filtering out scattered laser light. Single-molecule data was analyzed using custom-written particle-tracking MATLAB code, partially using code obtained from http://physics.georgetown.edu/matlab/.

**Particle tracking**

The positions of labeled particles were determined by fitting each single-molecule fluorescence image to a two-dimensional Gaussian distribution. The accuracy of position determination is given by

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

where \( N \) is the number of photons collected (7). Typical signals from individual AlexaFluor 488 labels corresponded to 125±56 photons per 50 ms integration. Using the
standard deviation of the microscope point-spread function $s$ (140 nm for our microscope), the pixel size $a$ (166 nm), and the background level $b$ (5-10 photons), we calculate the standard error of position determination to be $\sigma = 10$–20 nm.

**Determination of drift rates and diffusion coefficients**

We evaluate the presence of any directional bias in protein motion by measuring the net displacement of a protein divided by the duration of its trajectory. In the absence of any drift, the net displacement of a population of molecules undergoing normal diffusion will form a normal distribution around zero. In our experiment, however, we observe a small bias of the proteins’ motion in the direction of the flow. The flow-induced drift distances are about a factor of 5 smaller than the diffusional distances at experimental timescales (see main text) and are likely to have a minimal impact on the analysis of the diffusion properties of the protein. Nonetheless, we evaluate the effect of drift and diffusion as separate contributions by subtracting the mean drift over all trajectories at a particular biochemical condition from that condition’s individual trajectories and calculating the diffusion coefficient from the drift-corrected trajectories (8, 9). This method assumes that drift and diffusion are independent and that their contributions to the displacement of the protein is additive.

The mean drift is the mean of each trajectory’s total displacement over its duration, weighted by the duration of the trajectory. This weighted mean is equivalent to the total drift divided by the total duration of all trajectories, as if they were concatenated into a single long trajectory (Fig. S2, blue line). The standard deviation of the drift for a given biochemical condition is likewise weighted according to the durations of the trajectories.
As shown in Fig. S1, for 162 sliding proteins at 125 mM total salt concentration, the distribution of the drift component of the trajectories is shifted $262 \pm 1144$ bp/sec from zero in the direction of the flow.

Having experimentally obtained trajectories of multiple particles, we determine the diffusion coefficient $D$ of particle by plotting the mean square displacement (MSD) of the particles as a function of time windows $n \Delta t$ (equation 2), and fitting the resulting data to a straight line, whose slope equals $2D$.

$$MSD(N,n) = \frac{\sum_{i=1}^{N-n} (y_{i+n} - y_i)^2}{N - n} = 2Dn\Delta t.$$ \hspace{1cm} (2)

We measure the diffusion coefficient by the slope of the fit to the data corresponding to $n=3-10$ (0.15-0.5 seconds). The upper bound is limited by the typical length of the trajectories, whereas the lower bound is chosen to exclude the effect of short-lived DNA fluctuations. The DNA fluctuations appear on a timescale of less than 0.1 second as determined by tracking particles that appear to not slide on the DNA. The MSD plot of such particles is linear only on timescales less than 0.1 second and shows bounded diffusion for longer timescales. Therefore, by excluding small time windows, we are avoiding the fluctuations of DNA to appear in the calculation of the diffusion coefficient.

Figure S4 shows the MSD data for different proteins. The majority of the proteins display linear MSD versus time plots, indicating normal Brownian diffusion along DNA.

Out of 484 initial trajectories for 125 mM total salt concentration, 327 have trajectories longer than 1.5 seconds, with the reminder being too short to result in reliable diffusion coefficients. Out of these 327, the 235 trajectories with total distance traveled larger than 500 nm are chosen to avoid particles non-specifically bound to the glass surface of the
flow cell, as well as particles on DNA that are not sliding. The MSD curves for the majority of these proteins are linear. To avoid fitting curves corresponding to nonlinear MSD vs. \( t \) curves, we only take into consideration the 162 molecules for which the Pearson correlation coefficient between MSD and \( n\Delta t \) is greater than 0.9. Trajectories with non-linear MSD curves are likely to be non-sliding proteins on DNA with high-amplitude fluctuations. Diffusion and drift coefficients were determined from these 162 final trajectories in 125 mM total salt concentration. Similar proportions of proteins were selected in each of the above steps for other salt concentrations, and the number of final analyzed trajectories was similar across different biochemical conditions. Also the overall shape of the distribution of diffusion coefficient is similar for different salt concentrations (Fig. S3).

**Calculation of activation barrier heights in sliding**

The diffusion coefficient for a spherical object diffusing by purely translational movement in one dimension can be calculated by the Stokes-Einstein relation:

\[
D_{d,\text{lim}} = \frac{k_B T}{6\pi \eta a}. \tag{3}
\]

Where \( \eta \) is the solvent viscosity (8.9×10^{-4} Pa·s for water at 25 °C), \( a \) is the radius of the diffusing p53 protein (3.9 nm (10)), \( k_B \) is the Boltzman constant and \( T \) is the temperature. For p53, this calculation results in a one-dimensional diffusion coefficient of 6.2×10^7 nm²/sec. If the movement along DNA is coupled to a rotational component caused by the protein tracking the helical pitch around the DNA, the Stokes-Einstein relation can be modified as follows (3, 11):
The extra term in the denominator describes the additional energy dissipation caused by rotation of the protein around DNA. For p53, the upper value for $D_{rot+1D}$ is $7.7 \times 10^6$ (bp$^2$/sec). Assuming a step size of 1 base pair, the upper limit of the stepping rate can be calculated as (3, 12)

$$\frac{2D_{\text{lim}}}{<x>^2} = k_{\text{lim}},$$

which results in an upper limit for the stepping rate $k_{\text{lim}}$ of $1.54 \times 10^7$ steps/sec. The measured diffusion constant $D_{\text{exp}}$ has a value of $(2.60\pm2.17) \times 10^6$ bp$^2$/sec, corresponding to a stepping rate of $(5.20\pm4.34) \times 10^6$ steps/sec. From the Arrhenius relation $k_{\text{exp}}/k_{\text{lim}} = \exp(-\Delta G^\ddagger/k_BT)$, we calculate a value for the height of the activation barrier, $\Delta G^\ddagger$, of $1.78 \pm 1.21 \ k_BT$ for the protein-DNA constant energy barrier model.

**Stokes drag force**

In order to calculate the Stokes drag force exerted on a protein bound to the DNA, we need an estimate for the velocity of the buffer flow at the position of the DNA-bound protein. In our flow-stretching, the buffer solution was drawn into the channel by a syringe pump with a flow rate of 0.1 mL/min creating shear flow near the cover slip surface. The flow channel is 100 µm in height and 2mm in width, resulting in an average velocity of the buffer of 0.83 cm/sec. The flow velocity, however, is not a constant throughout the channel, but is zero at the boundaries, yielding a parabolic flow profile (12). The mean distance of the DNA from the coverslip surface is 0.2 µm (3, 12, 13).
With a channel height $h$, the flow velocity $v_y$ at a distance $y$ from the surface can be expressed as:

$$v_{avg} = \frac{2}{3} v_{max}, \text{ and } v_y = v_{max} \left( \frac{hy - y^2}{h^2 / 4} \right) = \frac{3}{2} v_{avg} \left( \frac{hy - y^2}{h^2 / 4} \right). \quad (6)$$

The average velocity of the flow at the center of the DNA, 0.2 $\mu$m above the surface, can be estimated as 100 $\mu$m/sec. The Stokes drag force exerted on an object close to a surface is given by

$$F = 6\pi \eta rv \left( 1 + \frac{9r}{16y} \right), \quad (7)$$

with $\eta$ denoting the viscosity, $r$ the radius and $y$ its distance from the surface (14). The force exerted on a single p53 bound to the DNA at a distance of 0.2 $\mu$m from the surface is calculated to be $\sim 6.6$ fN and is responsible for the bias in protein translocation in the direction of the flow.

**Measuring the protein density on DNA**

To measure the dependence of the binding affinity of p53 for nonspecific DNA, we count the number of DNA-bound proteins per $\lambda$ DNA molecule and divide by the DNA length to obtain a protein density (Fig. 1B and C, main text). The high sensitivity of binding affinity to salt concentration makes it difficult, however, to choose a protein concentration that allows for an unambiguous determination of the number of molecules at the various salt concentrations used. Instead, we measure the number of detected photons per unit length of $\lambda$ DNA as a proxy for the number of proteins bound. The single-molecule sliding experiments provide an average intensity per p53 protein of
125±56 photons/sec, a value that is used to convert intensity per unit length of DNA into number of proteins per unit length of DNA.
Supporting Figures

Figure S1
Figure S4
Supporting Figure legends

**Figure S1**

Weighted histogram of the drift velocity $v$ of 162 individual p53 proteins. Drift for each p53 protein is calculated by dividing the net displacement of each trajectory by its duration. The weight of each trajectory in the histogram is proportional to its duration. The distribution is biased and the mean of the distribution is at 262 bp/s in the direction of the flow. The depicted results are for 125 mM total salt concentration. Similar distributions were observed for other salt concentration indicating that drift is not dependent on the concentration of salt in our assay.

**Figure S2**

Trajectories of all of the particles stitched together. Black line is the displacement in the direction perpendicular to the flow direction. Blue line is the displacement in the direction of the flow and the red line is corrected by reducing the drift effect in the displacement.

**Figure S3**

The distribution of diffusion coefficients for total salt concentration of (A) 25 mM, (B) 75 mM, (C) 125 mM, and (D) 175 mM.

**Figure S4**
Mean square displacement $v_s. n\Delta t$ ($\Delta t=0.05$ s) for different proteins at a total salt concentration of 125 mM. As can be seen from the plot, most of the trajectories show linear dependence of MSD on time, indicating normal Brownian diffusion along DNA. Out of 235 trajectories, 162 were selected based on the selection criteria described in the text.
References: