Nonspecifically bound proteins spin while diffusing along DNA

Paul C Blainey[1], Guobin Luo[1], S C Kou[5], Walter F Mangel[3], Gregory L Verdine[1,4], Biman Bagchi[2] & X Sunney Xie[1]

It is known that DNA-binding proteins can slide along the DNA helix while searching for specific binding sites, but their path of motion remains obscure. Do these proteins undergo simple one-dimensional (1D) translational diffusion, or do they rotate to maintain a specific orientation with respect to the DNA helix? We measured 1D diffusion constants as a function of protein size while maintaining the DNA-protein interface. Using bootstrap analysis of single-molecule diffusion data, we compared the results to theoretical predictions for pure translational motion and rotation-coupled sliding along the DNA. The data indicate that DNA-binding proteins undergo rotation-coupled sliding along the DNA helix and can be described by a model of diffusion along the DNA helix on a rugged free-energy landscape. A similar analysis including the 1D diffusion constants of eight proteins of varying size shows that rotation-coupled sliding is a general phenomenon. The average free-energy barrier for sliding along the DNA was $1.1 \pm 0.2\ k_B T$. Such small barriers facilitate rapid search for binding sites.

Many nucleic acid enzymes and proteins that act on DNA quickly locate target sites by diffusing along nonspecific DNA. It has been shown that proteins can both hop and slide along double-stranded DNA[1–3], although the microscopic mechanism of protein motion along DNA molecules is still not understood in molecular detail. In particular, the path traced by a sliding protein molecule along the surface of DNA has not been established. Both linear paths, parallel to the DNA axis, and helical paths, following a strand or groove of the DNA around the DNA axis, have been taken as assumptions in biophysical and biochemical models. Although rotation of sliding proteins around the DNA helix was implicitly[4] and explicitly[5,6] anticipated, such rotation was not shown to occur during diffusive sliding. The concept of rotational coupling has also arisen among structural biologists based on concepts of molecular recognition and observations of detailed structural complementarity between proteins and DNA[7–9]. Despite the persistent high profile of this question in the literature, it remains unknown whether sliding proteins track the DNA helix. Such tracking would have major biophysical and biochemical implications: for example, only a limited set of enzyme-helix juxtapositions would need to be considered in questions of protein-DNA interaction. In this work, we examine the dependence on protein size of the diffusion constant for sliding along DNA in order to distinguish pure translational diffusion (**Fig. 1a**) along DNA from rotation-coupled (or -slaved) diffusion (**Fig. 1b**). The result offers insights into the mechanism of target search and recognition of all DNA-binding proteins.

As a protein moves along DNA, it experiences three different frictional forces arising from random collision with the solvent molecules, and all three are proportional to the solvent viscosity, retarding the protein's diffusive motion. One is the friction on colinear motion parallel to the DNA axis. In addition, if the protein spins along the DNA helix, there are two rotational components of the friction: the rotational friction for motion along the offset helical path due to circumnavigation of the DNA axis, and the additional rotational friction that arises from the body-centric protein rotation.

Einstein's treatment of translational diffusion as a Brownian motion, together with Stokes' expression for viscous friction, indicates that the diffusion constant of a protein sliding along DNA should vary with protein size as $1/R$, where R is the radius of the protein. This $1/R$ dependence of the 1D diffusion constant (D_1) is valid if the protein experiences only translational friction as it slides along DNA, regardless of the magnitude of this friction. However, if the protein is constrained to track the DNA helix (for instance, in order to maintain optimum contact between its DNA-binding patch and the surface of the DNA helix), the protein will be forced to rotate while translating, and, as a result, the size dependence of the diffusion constant can be quite different. Inclusion of protein rotational friction leads to a much stronger dependence of the diffusion constant on R, of the order $1/R^3$, typical of rotational diffusion. From the cocrystal structures of DNA-bound protein molecules, we know that many proteins bind DNA with a significant offset from the DNA axis. When a protein so bound undergoes motion along the DNA double helix, the path it traces through space is not a straight line. We recently developed a theory to take the nonlinear path traced by offset protein molecules into

a Translation parallel to the DNA axis
b Rotation-coupled translation along DNA helix

Protein's DNA-binding site always faces the DNA axis

DNA-binding site

R

DNA axis

R

DNA axis

R_{OC}

c Energy landscape for protein undergoing rotation-coupled sliding along DNA

$U(z)$

$-\pi/5$ $+\pi/5$

$\}\varepsilon$

−13.2 −9.9 −6.6 −3.3 0.0 3.3 6.6 9.9 13.2 16.5
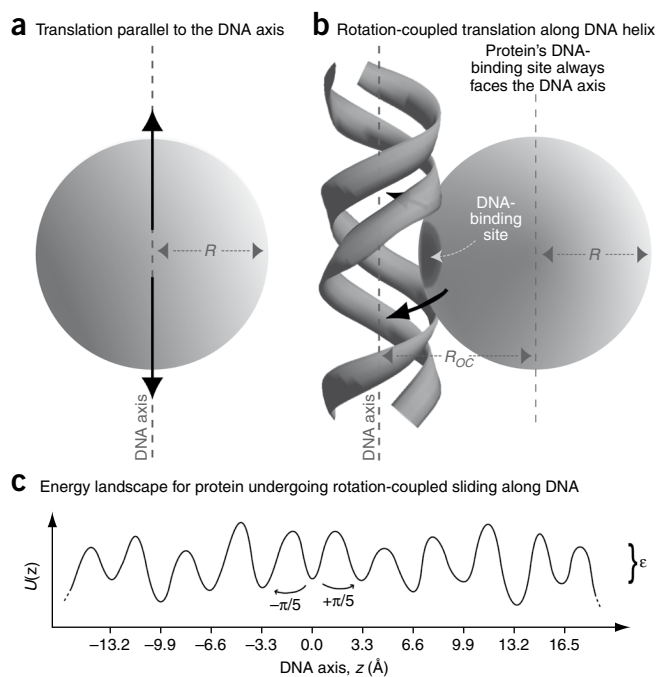
DNA axis, z (Å)

Figure 1 Models and definitions. (**a**) Schematic of linear diffusion model. Protein with radius R translates parallel to DNA axis (D_1 for the linear model is independent of protein offset from the DNA axis; an offset of zero is depicted here). (**b**) Schematic of helical diffusion model. Protein center of mass translates along a helical path. The width of the helical path is parameterized by R_{OC}, the minimum distance between the protein's center of mass and the DNA axis. The protein's DNA-binding site always faces the DNA axis, imposing coupling between protein translation and protein body-centric rotation. (**c**) Cartoon of rugged free-energy landscape $U(z)$ experienced by sliding protein molecule according to the helical diffusion model. System free energy and barrier heights to translocation are heterogeneous, depending on protein position in a heterogeneous base sequence. Each minimum corresponds to binding in register with a particular base pair, while each maximum corresponds to the transition state for sliding to the adjacent position. The rms variation of landscape energy is parameterized as ε. To slide along z to the adjacent base pair register, the protein must rotate by about $2\pi/10$ radians in order to maintain contact between the protein's DNA-binding site and the newly targeted base pair in the DNA helix. The hydrodynamic friction on the protein opposing this rotation-coupled movement along the DNA helix dominates the resistance to translocation of the protein molecule.

account[10]. The new theory differs from Schurr's original treatment[5] by the incorporation of a helical path for sliding, as parameterized by R_{OC}, the minimum distance from the protein center of mass to the DNA axis (**Fig. 1b**). The resulting expression (ref. 10 and below) provides a different numerical estimate of the friction experienced by a sliding protein molecule, although the size dependence remains close to $1/R^3$ when R_{OC} approximates R.

In addition to the frictional forces, interactions between the protein and the DNA retard the diffusion of proteins. The interaction between the protein and the DNA is a sum over a large number of two-body interactions involving the atoms of the protein and the DNA, and it depends both on the distance of separation between the protein and the DNA and on their relative orientation. We can divide this complex protein-DNA interaction into an average part and a fluctuating part. The average part defines a binding potential that governs the pathway along the helix. The fluctuating part gives rise to potential energy barriers along the diffusion pathway. The average of the fluctuating potential is zero; however, because of the heterogeneity of the DNA base sequence and resulting variation in the DNA helical structure, the protein experiences random minima and barriers on its path along the DNA helix. Therefore, it is a rough (or rugged) energy landscape that the protein must navigate by diffusion (**Fig. 1c**).

We now quantify the effects of the rough energy landscape on the diffusion of a protein moving along the DNA helix. It is known that a rough energy landscape with small barriers can retard diffusive motion. We express this retarding factor as $F(\varepsilon)$, as did Zwanzig[11], who showed that the actual diffusion coefficient can be written as the product of the hydrodynamic diffusion constant (the value obtained in the absence of the rough potential) and $F(\varepsilon)$. Zwanzig also showed that if the fluctuating part of the potential obeys a Gaussian distribution, then $F(\varepsilon) = \exp[-(\varepsilon/k_BT)^2]$ where ε denotes the rms variation of the fluctuating portion of the potential, indicating the average energy of the barriers that the protein crosses while sliding. Using the expression for the hydrodynamic friction[10], we arrive at the following expression for the diffusion constant along a helix, which is dependent on the protein radius (R) and

the minimum distance between the protein center of mass and the DNA axis (R_{OC}) (**Fig. 1b** and ref. 10):

$$D_{slide}^{helix} = b^2 \frac{k_BT}{[6\pi\eta Rb^2 + 8\pi\eta R^3 + 6\pi\eta R(R_{OC})^2]}F(\varepsilon) \qquad (1)$$

b^2 describes the effect of the helix pitch on the diffusion constant for sliding. b is given with dimensions distance per full rotation. Specifically, motion following the DNA helix would correspond to a value of 10.5 base pair (bp) for b, or about 3.4 nm. We note that, although ε is most naturally evaluated when the protein translocates a distance of 1 bp along the DNA (**Fig. 1c**), ε is scale invariant (independent of b) because the measured diffusion coefficient (D_{slide}^{helix}) is scale invariant for Brownian motions. If R_{OC} is diminished to a value of zero in equation (1), Schurr's form is recovered (without the $F(\varepsilon)$ term on the right hand side).

For protein-sized objects, the magnitude of the rotational friction dominates the translational friction by 1–3 orders of magnitude, and we may then ignore the translational contribution to obtain the simplified form:

$$D_{slide}^{helix} \approx b^2 \frac{k_BT}{[8\pi\eta R^3 + 6\pi\eta R(R_{OC})^2]}F(\varepsilon) \qquad (2)$$

Many DNA–protein complexes show $R_{OC} \approx R$, in which case $1/R^3$ scaling of D_1 is obtained, albeit with nearly twice the hydrodynamic friction predicted by Schurr. When R_{OC} differs from R, then a different (but still $1/R^3$-like) scaling, $1/[R^3 + ¾R(R_{OC})^2]$, is obtained.

However, if protein rotation is not required for motion along the DNA, then diffusion should exhibit a vastly different dependence, $1/R$, on the size of the diffusing protein molecule. This difference in size dependence, $1/R$ versus $1/[R^3 + ¾R(R_{OC})^2]$, can thus serve as a tool to identify the basic mechanism of protein transport along DNA, and is used below.

RESULTS

We used a single-molecule fluorescence tracking assay to obtain experimental values of D_1 for labeled human oxoguanine DNA glycosylase 1 (hOgg1) ($R = 3.2$ nm, $R_{OC} = 2.5$ nm) sliding along double-stranded DNA[3,12,13] (**Figs. 2** and **3**). The assay takes advantage of a simple flow-stretching method to prepare linear DNA templates and of total internal

Figure 2 Schematic of the flow-stretching and fluorescence imaging apparatus (reproduced from ref. 3) and diffusion of hOgg1. (**a**) Inverted microscope fitted for total internal reflection fluorescence imaging with mounted flow cell. (**b**) Flow cell detail. (**c**) Schematic of flow-stretched λ DNA molecule (not to scale). Buffer solution flows over glass coverslip to which double-stranded λ DNA, 16 μm in length, is attached by one end. (**d**) One-dimensional diffusion trajectories of hOgg1 conjugates of two different sizes, hOgg1–Cy3B (black lines) and hOgg1–PEG–streptavidin–Alexa Fluor 546 (gray lines), diffusing along DNA. Upper traces on left axis correspond to motion along the DNA, '$x(t)$ (bp)'. Lower traces on right axis correspond to motion transverse to DNA, '$y(t)$ (bp)'.
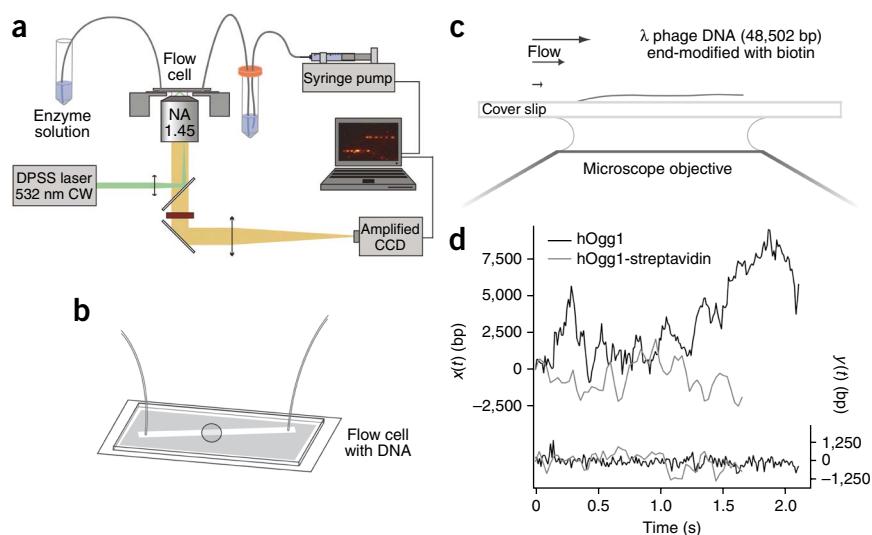
reflection fluorescence excitation to illuminate sliding protein molecules while excluding much of the specific fluorescence background. Because of the limited spatiotemporal resolution of this optical assay, very fine or very fast motions of the protein are not observed and the possible rotation of the protein is not easily discerned from the molecular trajectories, which appear as 1D diffusion. Thus, an indirect method (analysis of R and R_{OC} dependence) that takes advantage of accurately determined apparent 1D diffusion constants is used.

Protein size dependence reveals rotation-coupled sliding

We first chose the human DNA repair protein hOgg1 as a platform to test for protein rotation. hOgg1 has a modest interaction area with DNA[14–16] and is therefore a stringent test case for helical coupling, as rotation is unlikely to be enforced by steric constraints alone. To vary the size of hOgg1, we replaced the small fluorescent label (the Cy3B dye molecule) on hOgg1 with a larger one, streptavidin decorated with dye molecules.

All labeling of hOgg1 reported here was carried out at the protein C terminus. This site is well characterized, having been designed for, and experimentally demonstrated to have, a minimal impact on the hOgg1-DNA interaction and specifically on D_1 (ref. 3). Notably, the strong reactivity of a C-terminal, engineered cysteine residue relative to the internal cysteines was demonstrated by mass spectrometry. This reactivity allows quantitative labeling of hOgg1 at the engineered site with complete specificity. The C terminus of full-length hOgg1 (345 amino acids) is well removed from the DNA-binding interface, although the C terminus can be truncated to bring a label nearer to the DNA-binding interface and active site. In fact, this was done, in order to test the sensitivity of the hOgg1-DNA interaction to C-terminal labeling. hOgg1, C-terminally truncated at position 322, was labeled with small-molecule fluorophores of varying shape, chemistry and charge (both cationic and anionic dyes were tested), and the diffusion constant for sliding along flow-stretched DNA was determined in the single-molecule assay. No difference in D_1 was found with respect to the position or identity of the small-molecule label, demonstrating the notable insensitivity of the hOgg1 DNA-binding interface to C-terminal labeling in general, and specifically to the presence or absence of macromolecular structures in the region of the C terminus, such as the truncated residues 323–345.

We chose streptavidin as a steric and fluorescent label for hOgg1 based on a number of considerations. First, the streptavidin tetramer has a net negative surface charge at the assay pH and very low affinity for DNA. Second, a specific and very robust attachment chemistry, the biotin-streptavidin interaction, is intrinsic to the protein and

simplifies the preparation of monodisperse conjugates. Employing the same excess label–hOgg1–specific affinity chromatography strategy used to generate 1:1 hOgg1–Cy3B conjugates[3], we attached labeled streptavidin to the C terminus of hOgg1 using a flexible, biotinylated poly(ethylene glycol) (PEG) linker, resulting in conjugates with $R = 4.4$ nm and $R_{OC} = 3.3$ nm (**Figs. 2** and **3**).

A number of experimental observations validate our expectations that streptavidin does not interact with DNA in the single-molecule assay and that DNA-binding constant of hOgg1 is not substantially altered by the conjugation of streptavidin. First, in a control experiment with fluorescently labeled streptavidin and hOgg1 (nonbiotinylated) in assay buffer, no binding of strepavidin to DNA was observed. Second, the average numbers of hOgg1–streptavidin and hOgg1–Cy3B molecules bound per DNA at steady state in the single-molecule assay are comparable, given similar experimental conditions (as described for hOgg1 in Online Methods with 50 pM labeled molecules): 0.47 ± 0.79, and 0.78 ± 0.92, respectively. Finally, hOgg1, hOgg1–Cy3B, and hOgg1–streptavidin all elute from GE's HiTrap SP HP resin at about the same salt concentration (130 mM NaCl) at pH 8.0, providing another indication that the electrostatic binding properties of the two conjugates are equivalent and, furthermore, are equivalent to those of the native enzyme.

The PEG linker of roughly 68 ethylene glycol units has several desirable properties that simplify the interpretation of our measurements. The linker is highly flexible and soluble (solvent-like), making it unlikely to impose undesired intramolecular interaction between hOgg1 and streptavidin, or between streptavidin and the DNA, that could affect the diffusion constant in a manner superfluous to the intended hydrodynamic perturbation. Furthermore, the high solubility of PEG gives it well-defined conformational properties that have been extensively studied and described in the polymer physics literature. For instance, the radius of gyration of a 3,000-Dalton PEG chain in aqueous buffer can be accurately calculated at 2.25 nm. The linker's bulk serves to extend the attachment point for streptavidin away from the hOgg1-DNA interface, further reducing the chance of streptavidin-DNA interactions. Finally, the rapid relaxation of PEG chains maintains streptavidin at a constant, well-defined position relative to hOgg1 as hOgg1 undergoes translational and possibly rotational dynamics changes during sliding. Specifically, the maximum relaxation time of the linker is under 50 ns, easily fast enough to keep pace with the hOgg1–streptavidin conjugate, which takes on average more than 200 ns to slide 1 bp based on the data reported below.

Figure 3 Test of the diffusion models by protein size dependence with consistent (hOgg1) DNA-binding interface. (**a**) Test of the linear diffusion model. One-dimensional diffusion constants of hOgg1 and hOgg1–streptavidin sliding along double-stranded DNA versus the inverse protein radius, $1/R$, which is expected to give a linear relationship for molecules sliding along DNA with no rotational coupling. The fit, with just one adjustable parameter, is very poor, giving a reduced χ^2 value of 386. The ordinal error bars represent 95% confidence intervals for D_1 based on the single-molecule data, while the error bars on the abscissa indicate uncertainty in $1/R$. (**b**) Test of the helical diffusion model. The same diffusion constant data from **a** plotted versus $1/[R^3 + \frac{3}{4}R(R_{OC})^2]$, which is expected to give a linear relationship for molecules undergoing rotation-coupled sliding along the DNA helix, where the DNA-binding site on the protein maintains contact with the DNA. Most of the energy dissipated by such a sliding molecule is consumed by friction from protein rotation and circumnavigation of the DNA axis. The fit, with just one adjustable parameter, is satisfactory, giving a reduced χ^2 value of 0.75. The error bars here are similar to those of **a** except that uncertainty in R_{OC} is additionally reflected.
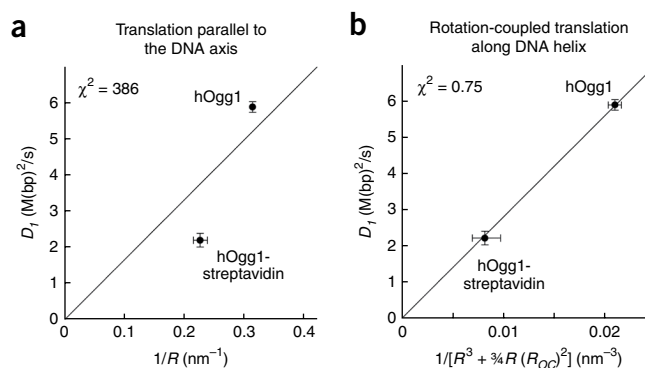
We now introduce a broadly accepted statistical method, the bootstrap[17], and apply it to single-molecule diffusion data to extract diffusion constants. Bootstrap resampling is a simulation-based statistical inference method that can be used to obtain estimates of diffusion constants based on nearly independent, identically distributed protein displacements (see Online Methods and **Supplementary Note** for detail). The bootstrap estimates generally agree with diffusion constants determined using the traditional method, where trajectory average and ensemble average mean-square displacement ($<\Delta x^2(\tau)>$) functions are constructed and subjected to linear regression. However, the bootstrap method is more efficient (with smaller error bars) than regression, and converges to the true value faster than regression with little dependence on model assumptions (for example, Gaussian-distributed errors) while providing a straightforward, empirical means for the analysis of errors (**Supplementary Fig. 1**). To obtain a more accurate estimate of the hOgg1 D_1, we used the bootstrap to reanalyze the hOgg1-Cy3B dataset reported in 2006 (ref. 3), and obtained $D_1 = 5.87 \pm 0.07$ M(bp)^2 s^{-1}, a value that falls within the error bar of the originally reported diffusion constant (4.8 ± 1.1 M(bp)^2 s^{-1}).

Figure 2d depicts the trajectories of hOgg1–Cy3B and the larger conjugate, hOgg1–streptavidin, diffusing along DNA. Clearly, hOgg1–streptavidin diffuses more slowly than hOgg1–Cy3B, with a diffusion constant of only 2.21 ± 0.09 M(bp)^2 s^{-1}. In all cases reported here, the diffusion trajectories lack long-term drift, indicating that there is no significant bias induced by the flow stream used to stretch the DNA.
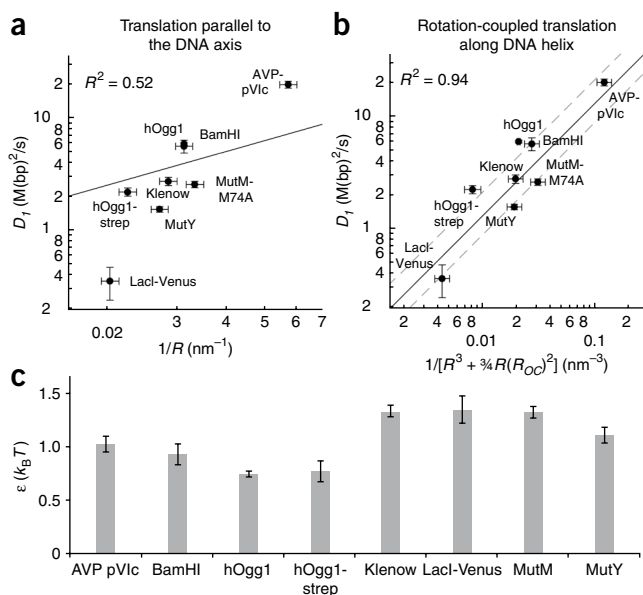
To make an even larger conjugate, we coupled streptavidin-coated quantum dots to hOgg1 (here with limiting hOgg1 and no purification to make 1:1 conjugates), despite concern that the outsized label could cause artifacts such as flow-induced drift, a change in the diffusion mechanism, or interactions with the surface of the flow cell. Furthermore, heterogeneity in the size of the quantum dots could falsely indicate variation in the diffusion constants among individual protein molecules and affect our analysis of size dependence. We obtained trajectories of hOgg1–quantum dot conjugates ($R = 9.0$ nm and $R_{OC} = 10$ nm) sliding on DNA but were unable to determine a consistent estimate of the diffusion constant by using the bootstrap or by regression of mean-square displacements. For this reason, we have excluded the hOgg1–quantum dot data from our quantitative analysis of size dependence. We note that quantum dots have been used recently in studies of protein translocation along DNA, and we caution against the use of semiconductor labels in such studies because of significant reduction in the protein diffusion constant and the possible effect on the translocation mechanism.

To explore the functional dependence of D_1 on R, we plotted D_1 for hOgg1 and hOgg1–streptavidin versus $1/R$ and fit the data to a line that passes through the origin (**Fig. 3a**). D_1 deviates strongly from $1/R$ across the conjugates (reduced $\chi^2 = 386$), indicating that the motion of sliding hOgg1 molecules along DNA is not simple linear diffusion along the chain in a sliding, hopping or hybrid modality. A different theory is clearly required to address the observed nontrivial size dependence.

We applied our theory[10] to understand the single-molecule results, noting that what we measured in the single-molecule experiments is an apparent D_1. The limited temporal and spatial resolution of the experiments preclude direct observation of protein rotation, but the effect of rotation is manifest in the apparent D_1. First, we plotted the measured D_1 versus $1/[R^3 + \frac{3}{4}R(R_{OC})^2]$, testing for the scaling predicted by equation (2) by fitting to a line that passes through the origin (**Fig. 3b**). The size dependence of D_1 can only show $1/[R^3 + \frac{3}{4}R(R_{OC})^2]$ scaling if hOgg1 undergoes persistent sliding with strongly dominant rotation-translation coupling. The fit is excellent (reduced $\chi^2 = 0.75$), indicating that hOgg1 spins while moving along DNA. This result alone indicates that the apparent 1D motion we observe along the DNA is not mediated by hopping or microdiffusion alone, but principally by sliding of hOgg1 in close contact with the DNA, a binding interaction persistent enough to strongly couple high-friction rotation of the protein molecule to its translational motion.

Rotation-coupled sliding is a general translocation mechanism

An expansive understanding of the protein-DNA interaction requires analysis of the free-energy landscape along the protein's path of diffusion and generalization beyond observations of hOgg1 to cases where R and R_{OC} differ significantly. Using the single-molecule method described above, we measured and analyzed the diffusion constants of eight proteins and protein conjugates of various sizes. The panel of diffusion constants consists of two published values, those for hOgg1–Cy3B[3] and *Escherichia coli* LacI–YFP dimers[18], and six newly determined values, those for hOgg1–streptavidin, *Bacillus stearothermophilus* MutY, *E. coli* MutM M74A, the adenoviral AVP–pVIc complex, the BamHI restriction endonuclease dimer, and the Klenow fragment of *E. coli* DNA polymerase I (raw diffusion traces for these proteins are presented in **Supplementary Figure 2**). This group includes proteins with diverse functions, including DNA replication, DNA cleavage, DNA repair, transcriptional regulation and proteolytic activation; variously structured DNA-binding domains, including the classical helix-hairpin-helix, zinc finger and polymerase folds; and diverse organisms of origin, including mammals, bacteria, bacteriophage and human viruses.

Figure 4 Global analysis of a diverse set of DNA-binding proteins according to the linear and helical models. Plotted are values of D_1 for the human hOgg1–Cy3B[3], hOgg1–streptavidin (this work), *B. st.* MutY (this work) and *E. coli* MutM M74A (this work) DNA glycosylases, adenoviral AVP–pVIc complex (this work), *E. coli* LacI–YFP dimers[18], the BamHI restriction endonuclease dimer (this work), and the Klenow fragment of *E. coli* DNA polymerase I (this work). (**a**) Comparison of the data with the linear diffusion model (values of D_1 versus $1/R$). The fit, with just one adjustable parameter, is incongruent with the trend in the data (coefficient of determination, $R^2 = 0.52$), revealing a poor description of the size dependence in this dataset by the linear diffusion model. (**b**) Comparison of the data with the helical diffusion model (values of D_1 versus $1/[R^3 + \frac{3}{4}R(R_{OC})^2]$). This fit, with just one adjustable parameter, quite satisfactorily describes the trend in the data (coefficient of determination, $R^2 = 0.94$), indicating the consistency of these data with the rotation-coupled sliding model. The error bars for each data point describe the same uncertainties represented in **Figure 3**. The gray dashed lines represent bounds on the model prediction reflecting the observational variation in ε. (**c**) ε calculated for sliding on a rough energy landscape[11] for the eight proteins appearing in **b**. The error bars correspond to 95% confidence intervals.

Realizing this set of proteins presents an assortment of sliding interfaces, we make a global analysis of R and R_{OC} dependence, as the effect of differences in ε across the set will be muted across the broad range in size and the greater average difference between R and R_{OC} compared with the hOgg1 conjugates. In **Figure 4** we plot the measured D_1 versus $1/R$ and $1/[R^3 + \frac{3}{4}R(R_{OC})^2]$ for these eight proteins and protein conjugates and fit the dataset with lines that pass through the origin. The result is decidedly characteristic of rotation-coupled diffusion along DNA, as D_1 for this eclectic collection of proteins correlates strongly with $1/[R^3 + \frac{3}{4}R(R_{OC})^2]$ (**Fig. 4a**, coefficient of determination, $R^2 = 0.94$), but not with $1/R$ (**Fig. 4b**, $R^2 = 0.52$). This interesting result allows us to draw two conclusions. First, despite the fact that each protein has its own DNA-binding interface (and hence, its own ε), the $F(\varepsilon)$ values are similar. Second, the broad diversity of the sliding interface among the eight proteins suggests that rotation-translation coupling is a mechanism common to all DNA-binding proteins.

Small energy barriers in helical sliding along DNA helix

We used equation (1) to obtain values of ε for each protein, taking advantage of the experimental values of D_1 and diffusion constants calculated from the helical sliding model using R and R_{OC}, with b fixed at 10.5 bp, indicating helical sliding that matches the helical pitch of the DNA (**Fig. 4c**). The result demonstrates that the values of ε calculated for all these proteins are small and tightly clustered between 0.75 and 1.35 k_BT, averaging 1.1 ± 0.2 k_BT (±s.d.) across the group. The independently derived ε values from the hOgg1 and hOgg1–streptavidin datasets agree closely, consistent with the fact that the two protein–DNA complexes share the same DNA-binding interface. This is a further indication that the labeling of hOgg1 with streptavidin was minimally perturbative of the non-specific binding interaction. Although small barrier heights have been theoretically implicated[19], our measurements represent the first experimental determination, to our knowledge, of the ruggedness parameter, ε.

DISCUSSION

The data presented here reveal small average barriers tightly clustered around 1 k_BT, which explain how these proteins can slide rapidly on the rugged energy landscape in search of their targets. Because of the strong dependence of the diffusion constant on ε, proteins whose cellular function depends on fast sliding have evolved to minimize ε. The remaining small yet statistically significant differences among the calculated ε values within this diverse set of proteins reflect the differing physical and functional constraints under which each protein's sliding activity has evolved. Recent papers reporting sliding by several additional proteins[20–23], including a study implicating helical sliding by proliferating cell nuclear antigen, are consistent with our conclusion that rotation-coupled translation is a general feature of protein sliding.

It is conceivable that differences exist among the helical paths traced by different protein types. Such differences may be important for both sliding activity and protein function. For example, different proteins may track different parts of the DNA helix, such as the major groove, the minor groove, both grooves or neither groove (in the case of backbone tracking). The degree of interaction with the bases is also likely to vary widely with protein type, the DNA state, DNA sequence, the presence or absence of cofactors and the solution conditions. New structural and biophysical approaches using sophisticated biochemical methods and perturbations can reveal the interplay between rotation-coupled sliding activity and protein function (recognition and/or catalytic activity) in thermodynamic and kinetic terms.

We have shown that nonspecifically bound protein molecules diffuse along the helical path defined by DNA and rotate in order to keep the DNA-binding face of the protein in contact with DNA during fast 1D sliding along DNA. Obligate tracking of the DNA helix by sliding protein molecules, and the accompanying 360-degree rotation per helical turn, are prerequisites for efficient recognition of targets in DNA[3,16] and are necessary for the conveyance of some types of information by proteins along DNA, as recently proposed for the type III restriction enzymes[24]. For each protein type, a particular helical track along the DNA represents a locus of low free-energy nonspecific binding states that support rapid sliding along the rugged free-energy landscape with small barriers along the DNA. By maintaining the protein position and orientation with respect to targeted base pairs, the protein–DNA complex is able to retain a kinetically efficient target-recognizing configuration while rapidly scanning a DNA substrate.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Published online at http://www.nature.com/nsmb/.

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. 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).
2. Gowers, D.M. & Halford, S.E. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* **22**, 1410–1418 (2003).
3. 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).
4. Von Hippel, P.H. & Berg, O.G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
5. Schurr, J.M. The one-dimensional diffusion coefficient of proteins absorbed on DNA. Hydrodynamic considerations. *Biophys. Chem.* **9**, 413–414 (1979).
6. Sakata-Sogawa, K. & Shimamoto, N. RNA polymerase can track a DNA groove during promoter search. *Proc. Natl. Acad. Sci. USA* **101**, 14731–14735 (2004).
7. Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
8. Kalodimos, C.G. *et al.* Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386–389 (2004).
9. Von Hippel, P.H. BIOCHEMISTRY: Completing the View of Transcriptional Regulation. *Science* **305**, 350–352 (2004).
10. Bagchi, B., Blainey, P.C. & Xie, X.S. Diffusion constant of a non-specifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B* **112**, 6282–6284 (2008).
11. Zwanzig, R. Diffusion in a rough potential. *Proc. Natl. Acad. Sci. USA* **85**, 2029–2030 (1988).
12. van Oijen, A.M. *et al.* Single-molecule kinetics of λ exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235–1238 (2003).
13. Schroeder, C.M., Blainey, P.C., Kim, S.J. & Xie, X.S. Hydrodynamic flow-stretching assay for single-molecule studies of nucleic acid–protein interactions. In: Ha, Taekjip and Selvin, Paul R., editors *Single-Molecule Techniques: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 461–492 (2008).
14. Bruner, S.D., Norman, D.P.G. & Verdine, G.L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
15. 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).
16. Banerjee, A., Santos, W.L. & Verdine, G.L. Structure of a DNA glycosylase searching for lesions. *Science* **311**, 1153–1157 (2006).
17. Efron, B. & Tibshirani, R.J. An introduction to the bootstrap. *Monographs on Statistics and Applied Probability* **57**. New York: Chapman & Hall, 1993.
18. Elf, J., Li, G.W. & Xie, X.S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
19. Slutsky, M. & Mirny, L.A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
20. Tafvizi, A. *et al.* Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **95**, L01–L03 (2008).
21. Liu, S., Abbondanzieri, E.A., Rausch, J.W., Grice, S.F.J. & Zhuang, X. Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science* **322**, 1092–1097 (2008).
22. Lin, Y. *et al.* Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA. *Biophys. J.* **96**, 1911–1917 (2009).
23. Kochaniak, A.B. *et al.* Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.* **284**, 17700–17710 (2009).
24. Ramanathan, S.P. *et al.* Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. USA* **106**, 1748–1753 (2009).
25. Gupta, S. *et al.* DNA binding provides a molecular strap activating the adenovirus proteinase. *Mol. Cell. Proteomics* **3**, 950–959 (2004).

ONLINE METHODS

Protein labeling. Human hOgg1 and *E. coli* mutM were site-specifically labeled with Cy3B and purified by published methods[3]. Briefly, C-terminal cysteine mutants were engineered, overproduced in *E. coli* and purified by cation exchange chromatography. The purified proteins were incubated at 22–24 °C and pH 8 with excess Cy3B–maleimide before the reaction was quenched by addition of β-mercaptoethanol and chromatographic removal of free dye. Klenow fragment of *E. coli* polymerase I (USB) was site-specifically labeled through reaction between Cy3B–maleimide (12.5 μM) and the unique cysteine in the Klenow fragment (~1 μM) at room temperature for 2 h. Excess Cy3B was removed using a desalting column.

Conjugate formation. hOgg1 residues 12–345 + 346C (where 346C is an engineered cysteine residue at the C terminus) was labeled with biotin–PEG3k–maleimide (Nektar) and purified by published methods[3]. Excess Alexa Fluor 546–labeled streptavidin (Invitrogen) was mixed with limiting hOgg1–PEG–biotin and the 1:1 conjugate (each hOgg1 monomer coupled to a single streptavidin tetramer) purified by cation exchange chromatography. In a separate reaction, purified hOgg1–PEG–biotin was mixed with excess quantum dot–655–streptavidin (Invitrogen) and used without further purification.

DNA flow stretching and fluorescence microscopy. Flow stretching of DNA and total internal reflection fluorescence microscopy were carried out as previously described[3], with the exception of red-shifted fluorescence emission filters for the experiments with quantum dots. λ DNA was tethered to a glass surface at one end and stretched by a laminar flow of buffer (**Figs. 2a–c**). Incubation of the flow cell inner surfaces with 0.1 mg ml^{-1} streptavidin prevents binding of streptavidin-functionalized hOgg1 and streptavidin-coated quantum dots to the biotinylated coverslip surface. Single molecules that bound to the DNA were illuminated by a laser beam (532 nm) and imaged with the aid of an inverted fluorescence microscope (Olympus IX70) and sensitive EMCCD camera (Photometrics Cascade:512B).

In general, the labeled proteins and conjugates were infused at concentrations of 0.01–5 nM at 10–50 ml h^{-1}. The assay buffers were formulated as follows: hOgg1, 10 mM Tris or phosphate buffered at pH 7.8–8.0 (both buffers give equivalent results), 10 mM NaCl, 0.1 mM EDTA, 5% (v/v) glycerol, 0.5 mg ml^{-1} BSA and 5 mM β-mercaptoethanol; LacI, 10 mM phosphate buffered at pH 7.3, 2 mM NaCl, 0.05 mM EDTA, 20% (v/v) glycerol, 0.5 mg ml^{-1} BSA, 0.01% (v/v) Tween and 1 mM β-mercaptoethanol; AVP-pVIc, 10 mM MES buffered at pH 6.5, 2–25 mM NaCl, 0.05 mM EDTA, 20% ethanol and 5% (v/v) glycerol; Klenow, BamHI and MutM, 5 mM phosphate buffered at pH 7.5, 5 mM NaCl and 0.5 mg ml^{-1} BSA. In some experiments with MutM, 0.5 mg ml^{-1} casein was used in place of BSA for better suppression of surface adsorption.

Data analysis. The fluorescence images were processed using DiaTrack particle tracking software (Semasopht) for centroid determination as previously described[3]. Diffusion constants were estimated by applying the bootstrap to molecular displacement data[17]. The bootstrap is a simulation-based modern statistical inference method. Since its introduction in the early 1980s, the bootstrap method has swept through statistics and has been applied to a wide range of datasets, including particle physics, and in the simulation studies of statistical physics, phase transition, and more. Briefly, the distribution of $<\Delta x^2(\tau)>$ is built up empirically by resampling the data (with replacement) several thousand times. Model parameters and error estimates are obtained from the distributions of boostrap-resampled $<\Delta x^2(\tau)>$. In general, the diffusion constants obtained using the bootstrap method agree with those determined from the slope of ensemble-average mean-square displacement plots according to traditional regression methods but yield more accurate estimates of the true values and errors. An extended introduction to the bootstrap is given in the **Supplementary Note**.

R and R_{OC} were determined from available structural data, with the aid of light-scattering measurements for hOgg1 (ref. 3), and synchrotron footprinting data to locate the likely DNA-binding site for AVP-pVIc[25]. Proteins were modeled as ellipsoids of revolution, with R_{OC} taken as the minimum distance from the ellipsoid center to the center of the DNA axis. For hOgg1–PEG–streptavidin, we assume that the label extends from the DNA–protein complex in the direction of the hOgg1 C-terminal helix. The relative errors in R and R_{OC} are estimated at 10% (95% CI), except for hOgg1–streptavidin, for which the errors are estimated at 15%. Values and errors of ε were also computed using the bootstrap, with R and R_{OC} sampled as normal random variables in the bootstrap resampling steps.