Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models

Erik G. Marklund, Anel Mahmutovic, Otto G. Berg, Petter Hammar, David van der Spoel, David Fange, and Johan Elf

Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, 75124 Uppsala, Sweden

Edited by Peter H. von Hippel, University of Oregon, Eugene, OR, and approved October 18, 2013 (received for review April 26, 2013)

Transcription factors search for specific operator sequences by alternating rounds of 3D diffusion with rounds of 1D diffusion (sliding) along the DNA. The details of such sliding have largely been beyond direct experimental observation. For this purpose we devised an analytical formulation of umbrella sampling along a helical coordinate, and from extensive and fully atomistic simulations we quantified the free-energy landscapes that underlie the sliding dynamics and dissociation kinetics for the LacI dimer. The resulting potential of mean force distributions show a fine structure with an amplitude of 1 k_B T for sliding and 12 k_B T for dissociation. Based on the free-energy calculations the repressor slides in close contact with DNA for 8 bp on average before making a microscopic dissociation. By combining the microscopic molecular-dynamics calculations with Brownian simulation including rotational diffusion from the microscopically dissociated state we estimate a macroscopic residence time of 48 ms at the same DNA segment and an in vitro sliding distance of 240 bp. The sliding distance is in agreement with previous in vitro sliding-length estimates. The in vitro prediction for the macroscopic residence time also compares favorably to what we measure by single-molecule imaging of nonspecifically bound fluorescently labeled LacI in living cells. The investigation adds to our understanding of transcription-factor search kinetics and connects the macro-/mesoscopic rate constants to the microscopic dynamics.

facilitated diffusion | lac operon | lac repressors | gene regulation

Gene expression is regulated by transcription factors (TFs) that recognize specific sequences, operators, in DNA. The TFs are passively transported, yet they find their target operators faster than 3D diffusion allows in vitro (1). Theoretical efforts to reconcile the short search times with passive transport have resulted in the facilitated diffusion model, which describes the search process as alternating rounds of 3D diffusion in the cytoplasm and 1D diffusion along DNA (3). The latter means that the TF is capable of interacting sequence independently and “sliding” on or very close to the DNA helix. Single-molecule microscopy has provided evidence for facilitated diffusion in vitro (4) and recently also in vivo (5). Single-molecule studies in vitro indirectly suggest that the 1D diffusion along the DNA follows a helical trajectory. This is inferred from how the 1D-diffusion rate constant scales with the size of the sliding protein. The measurements also predict that a free-energy profile for sliding has a roughness of ~1 k_B T for several different TFs (6). Helical sliding on DNA has been explored at various levels of molecular simulation over the last decade (7–13). Although there are several lines of evidence for helical diffusion of TFs along the DNA, the atomic details of the translocation remain unknown, e.g., how hydrogen bonds change while sliding and how water and ions are displaced. These microscopic interactions will govern how the TF probes for the specific operator, how long a time the TF is nonspecifically bound to DNA, and thus how far the molecule slides before dissociating.

The lac operon has been a model system for gene regulation in bacteria for 50 y and is well characterized with respect to in vitro biochemistry, in vivo regulation, and structure (14–16). The wild-type lac operon repressor is a LacI homotetramer, a dimer of dimers, with two DNA-binding interfaces composed of two head domains each. As such, the tetramer can bind to two DNA strands simultaneously. A C-terminal truncation of one α-helix prevents tetramerization and results in LacI dimers with retained capacity for binding to one operator (5). Structure models from X-ray crystallography of LacI (16, 17) show how inducers or anti-inducers promote structural changes that modulate the affinity for the operator. There is, however, no complete structure model of LacI bound to nonoperator DNA, but an NMR ensemble of the head domain bound to an “anti-operator” indicates the structure of the head domain, how it differs from the operator-bound case, and the interactions it makes with DNA (18).

Here we use molecular dynamics (MD) simulations to reveal how dimeric LacI move along the DNA in search for the operator. We construct an atomistic structure model for dimeric LacI bound to nonoperator DNA from an existing experimental X-ray structure of the operator-bound LacI dimer and the nonspecific conformation seen in the NMR structure of the head domains. We then pull LacI in the direction of the DNA axis (steered molecular dynamics) without applying any angular forces and find that a helical path is preferred over nonhelical sliding on the DNA surface. On the basis of our findings from the axial pulling simulations we devise an umbrella-sampling method for calculating a potential of mean force (PMF) along a helical reaction coordinate and apply it to dimeric LacI bound to nonoperator DNA. We compare this to the free energy of dissociating from DNA to determine whether LacI remains in contact with DNA during the search process or whether translocation along DNA mainly occurs via macroscopic hops. We show that our simulation results agree well with experiments by connecting the free-energy differences at microscopic length scales and timescales with experimentally

Significance

Transcription factors (TFs) are the major regulators of gene expression. We have used a combination of computer simulations and theoretical methods to explore the atomic details for how a model TF, the lac repressor, moves on DNA in search for its specific binding site. We find that it slides along the DNA along a helical path, which allows it to probe for specific binding in a major groove. The uniqueness in the study lies in that the fully atomistic molecular dynamics model allows us to estimate the microscopic interaction energies from the TF-DNA structure and that we have developed a theoretical tool to derive macroscopic, experimentally testable, predictions for DNA residence time and sliding lengths from the microscopic interaction energies.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Frequently available online through the PNAS open access option.

1To whom correspondence should be addressed. E-mail: johan.elf@icm.uu.se.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1307905110/-/DCSupplemental.
measurable macroscopic quantities, using Brownian dynamics simulation including rotational diffusion to bridge the mesoscopic length scales.

Results

The Nonspecifically Bound Structure. The DNA-binding interface of the nonspecifically bound LacI dimer rapidly attained a structure close to that of the NMR-derived nonspecific structure during the targeted MD. During further equilibration without any applied forces its structure remained (rmsd of its Cα was 1.58 ± 0.20 Å) and the mutual center of mass (COM) of the head domains was located 1.33 nm from the central axis of the DNA helix during preparation. The structure of the core domains changed little between the specifically bound starting conformation and the equilibrated nonspecifically bound conformation. An asymmetry that can also be seen in the crystal structure, with the core domains tilting slightly toward one head domain, persisted during equilibration. In subsequent simulations, however, the core domains “swayed” freely (Fig. S1); hence we assume that the interface between the core and the head domains escaped any nonequilibrium configuration.

Axial COM Pulling. When LacI was pulled one base pair axially for 100 ns on AT-DNA the COM of the head domains underwent a rotational movement corresponding to slightly less than 1/10th of a full helical turn. Keeping the reference planes, toward which the head domains were pulled, fixed at their final positions for another 100 ns allowed for LacI to complete the rotation. The mutual COM of the head domains explored a region in the plane spanned by the angle and axial position (the φr plane) that corresponds to what is expected for a helical path; in both directions, the COM movement spanned 1 bp and almost 1/10th revolution (Fig. 1B). The COM remained at 1.38 ± 0.11 nm (SD) from the DNA axis throughout the simulation, implying that LacI remained close to the DNA while traveling 1 bp.

Although the pulling simulations are inherently nonequilibrium, the fact that LacI followed the DNA helix indicated this to be the dominant path for sliding, motivating further investigations of helical sliding and dissociation.

Helical Umbrella Sampling. The free-energy profile G(x) relative to an arbitrary G0 along a reaction coordinate X is directly related to how X is populated. Application of an additional known potential U in the X direction in MD simulations can enable sampling of regions where G is high. Running simulations where slightly different U focuses the sampling to different parts of X is known as umbrella sampling and allows the calculation of the entire free-energy profile along X as a PMF (19). Because umbrella sampling can be used to directly sample energy barriers, it allows for the study of reactions that occur on timescales substantially longer than those of the simulations. Still, if the simulations are too short, they will be biased by the “memory” of the starting configuration and the first part of the data collected in this way should be omitted from analysis. Also, a proper choice of reaction coordinate is crucial for obtaining meaningful results. The result from the axial pulling experiment guided our choice of reaction coordinate and as a consequence our design of umbrella force is as follows:

i) The umbrella force should be applied only in the helical direction. This allows for orthogonal movements and makes comparison with the PMF in the radial direction possible.

ii) The tangent vector of a helical path depends on the radial distance from the helix axis; the projection of the unit tangent vector onto the helix axis decreases monotonically with distance and becomes zero in the limit of infinite distance. This is analogous to a spiral staircase being steeper near its center than on its outer rim; an infinitely wide staircase would have zero inclination on its outmost parts.

iii) The (umbrella) force must be a conservative field, i.e., the negative of the gradient of a scalar potential. If this is not the case, then the work required to reach an arbitrary point in conformational space will be path dependent.

iv) From i and iii follows that the umbrella potential is invariant to the radial distance only exactly on the helical path; in positions away from the helical path the potential will vary with the distance to the helical axis. As a consequence, for the force to be a conservative field a radial component to the umbrella force necessarily arises (Fig. S2).

In Support Information, Helical Umbrella Sampling, we derive all equations necessary for implementing a harmonic helical umbrella potential and explore some of their features. The umbrella potential is U(φr, r) = kθr/L(ξ), and its force components are Fφ = −2πkur(0L−2πn)/L and Fr = −8πkur(0L−2πn)/L. Here k is a force constant; u and r denote the deviation from a reference point ξ in the helical and the radial direction, respectively; L is the r-dependent length of a full helical turn; and θ is the angular distance to a reference point.

Helical PMF. The PMF along the helical path had a fine structure with an amplitude of ∼1 kBT. Two peaks stand out with a period of 1 bp and a height of ∼3.5 kBT relative to their immediate surroundings (Fig. 2C). The longitudinal SD—the roughness—of the PMF was 1.2 kBT for AT repeats and 0.8 kBT for GC repeats. The SD at each point given by bootstrapping was ∼1 kBT in most of the PMF. The overall features were the same for the AT and GC systems although AT repeats have been reported to bind significantly stronger to LacI (20). Specific interactions with AT result from structural rearrangements of the DNA (21) that are prevented by the periodic boundaries for the DNA in the simulations.

Fig. 1. (A) The regions explored by the head domains’ centers of mass in the pulling simulations have been repeated to illustrate the helical sliding path over many base pairs. The head domains are separated by approximately one helical turn, and hence their paths overlap. (B) Region sampled by the head domains’ mutual center of mass when a force is applied in the axial direction. In blue and red are points from dragging the head domains away from and toward a reference point on the DNA, respectively. An ideal B-DNA helix is shown as a black line for reference.
of the latter. The DNA-binding interface was along the helix (Fig. 2) and the total number of protein N termini. Within the span of single trajectories they hydrogen bonded to DNA for a long time, but it did so only in a minority of simulations (Tyr47, Asn50, and Gln54). In addition to previously identified residues in the head domains, there were many hydrogen bonds (a minimum of 1.37 SD) for LacI along a helical path on AT repeats (blue solid line) and GC repeats (red dashed line). The average number of hydrogen bonds between LacI and AT-DNA per umbrella window along the helical path, and as there was less sampling near the peaks, the underlying statistics for the hydrogen bonds were less reliable there. Nevertheless, there were maxima ~ 1 bp apart.

Interactions with Water and Ions. Ions and water are displaced as LacI translocates on DNA, based on the helical umbrella-sampling simulations. Between the head domains the amount of water nearly reaches the background level around DNA (Fig. 3A) and there is an enrichment of ~1 Na+ (Fig. 3B).

During the dissociation process the water recondenses on DNA until the COM separation between the DNA and the DNA-binding interface reaches ~2.7 nm (Fig. S4A).

Radial PMF and the Microscopic Residence Time. The PMF as a function of the COM distance between the head domains of LacI and the DNA directly below can be seen in Fig. 2E. The PMF was virtually barrierless, inclined almost monotonically with increasing distance, and the dissociated state was ~12 k_BT above the bound state.

At a distance of r = 2.8 nm we consider the DNA-binding domain to be microscopically dissociated from DNA because water is recondensed at DNA and the head domains start moving independently of the DNA (Figs. S4A and B and S5). From Eqs. S7 and S8 the mean time to reach the endpoint r = ρ from any starting distribution within the interval r_0 ≤ r ≤ ρ, where r_0 corresponds to LacI closely bound to DNA, can be calculated as

$$\tau_d = \int_{r_0}^{\rho} \frac{e^{-G_b(r)}}{D_3} \frac{e^{-G_b(r')}}{\rho} \int_{r_0}^{r'} c(r') dr' dr'' dr''.$$  \[1\]

The integrals were evaluated numerically, using the radial PMF for G_b(r), the equilibrium distribution c(r) = e^{-G_b(r)} / \int_{r_0}^{\rho} e^{-G_b(r')} dr', the distance at the maximum in the radial PMF (2.8 nm) for \(r_0 \lesssim \rho \lesssim r_0\), and D_3 = 50 μm^2/s based on the Stokes–Einstein equation, yielding \(\tau_d = 69 \mu s\) or, in terms of the microscopic dissociation rate constant, \(k_d^{micro} = \tau_d^{-1} = 1.45 \times 10^4 \text{s}^{-1}\).

The Macroscopic Residence Time. The next challenge is to derive properties from the microscopically simulated system that can be directly compared with experimental measurements. We start by calculating the residence time on one DNA segment including the inevitable rebinding events that is a consequence of diffusion-controlled dissociation (22), the effect being much enhanced in the case of leaving a 1D object in 3D. This involves calculating the number of rebinding events a TF makes to the same DNA segment before it reaches a distance R_c, where rebinding is uncorrelated to the microscopic dissociation event (3).

For this reason we consider interactions with both GC and AT repeats to be representative of nonspecific interactions with DNA.

In all umbrella-sampling simulations, with a total time of ~9 μs, the head domains remained in close contact with the DNA, the COM of the former being 1.37 ± 0.14 nm (SD) from the axis of the latter. The DNA-binding interface was flexible, whereas the individual head domains were relatively rigid (Fig. S3).

Hydrogen Bonding While Sliding. The hydrogen bonding of individual amino acids is seen in Fig. 2F. The two head domains hydrogen bond to the DNA to a similar degree, although subtle differences can be seen. The occupancy of the hydrogen bonds that are identified as part of the nonspecific binding (18) differed between neighboring umbrella windows along the helical path, which could explain some of the variation in the helical PMF (Fig. 2F). The ones characteristic of specific binding were either completely absent (Ala53) or occurring only with low occupancy in a minority of simulations (Tyr47, Asn50, and Glu54). In addition to previously identified hydrogen-bonding residues we found that Gly58 can bind for a long time, but it did so only in a minority of simulations. Met1 and Lys2 are part of the flexible N termini. Within the span of single trajectories they hydrogen bonded to several different sites on DNA, phosphate groups on both strands, and the DNA bases.

There were striking similarities between the negative of the total number of protein–DNA hydrogen bonds and the PMF along the helix (Fig. 2B and C), but also notable differences. The number of hydrogen bonds had maxima (N_max = 25.1) near the integer number of base pairs traveled and minima in between (N_min = 9.9), resembling the overall shape of the PMF. Near 0.75 bp traveled, however, there were many hydrogen bonds (a minimum in Fig. 2B) but a peak in the PMF. The hydrogen bonds were calculated per window, and as there was less sampling near the peaks, the underlying statistics for the hydrogen bonds were less reliable there. Nevertheless, there were maxima ~1 bp apart.

The integrals were evaluated numerically, using the radial PMF for G_b(r), the equilibrium distribution c(r) = e^{-G_b(r)} / \int_{r_0}^{\rho} e^{-G_b(r')} dr', the distance at the maximum in the radial PMF (2.8 nm) for \(r_0 \lesssim \rho \lesssim r_0\), and D_3 = 50 μm^2/s based on the Stokes–Einstein equation, yielding \(\tau_d = 69 \mu s\) or, in terms of the microscopic dissociation rate constant, \(k_d^{micro} = \tau_d^{-1} = 1.45 \times 10^4 \text{s}^{-1}\).
To consider both translational and rotational aspects of the diffusion process we have developed a Monte Carlo scheme for the diffusion of the repressor from the microscopically dissociated state to $R_c$ or back to the bound state. The details of the simulation are described in Supporting Information, Stochastic Reaction-Diffusion Simulations (see also Fig. S6), but in principle the repressor displays rotational diffusion with the diffusion rate constant $D_0 = 3 S / 4 a^2$ and translational diffusion away from the DNA or around the DNA with the diffusion rate constant $D_S$. The repressor binds back to the DNA with the microscopic rate $k_{\text{micro}} = k_{\text{macro}} / 700 = 21 \text{s}^{-1}$, which corresponds to a mean in vitro residence time of 48 ± 12 ms.

**Sliding Length.** The 1D-diffusion coefficient depends on the variance $\sigma^2$ of the underlying potential (25). In the case of a sizeable particle on a helical path (6) and a potential with fluctuations that obey a Gaussian distribution we have

$$D_1 \approx \frac{b^2 k_B T}{8 \eta (R^3 + 2 R (R_{OC}^3))} e^{-\sigma^2 / (k_B T)}, \tag{2}$$

where $b$ is the pitch per radian, $\eta$ is the viscosity of the medium, and $R$ and $R_{OC}$ are the particle’s radius and distance from the DNA axis. From our structure we get $R = 4.2 \text{ nm}$, $R_{OC} = 4.55 \text{ nm}$, and $b = 3.4 / (2 \pi) \text{ nm}$. $\eta = 8.94 \times 10^{-4} \text{ Pa s}$ for water at 300 K, and from our PMF we have $e \approx 1 \text{ k_BT}$, which yields $D_1 = 1.23 (0.40-2.50) \times 10^5 \text{ bp}^2/\text{s}$. Finally, the sliding length is given by

$$s_{MD} = \sqrt{D_1 / k_{\text{macro}}} = 240 \pm 105 \text{ bp}. \tag{3}$$

comparable to the value of ~150 bp inferred from in vitro experiments on the tetramer conducted under conditions corresponding to those from which we have taken the in vitro parameters (14).

**The in Vivo Residence Time.** We may now ask whether the macroscopic residence time on DNA based on in vitro parameters has any relevance in the intracellular situation where differences in pH, ionic strength, and cytoplasmic crowding may change the dissociation process and make the nonspecific DNA interactions significantly different. For this reason we used single-molecule in vivo imaging of fluorescently labeled LacI dimers (LacI-Venus) to estimate nonspecific residence time on chromosomal DNA in a bacterial strain where the specific operator sites were removed. The assay is based on how long a fraction of the fluorescent molecules can be seen at different exposure times. If a molecule is dissociated, it will move so fast that it cannot be observed as a diffraction-limited spot. We find that nonspecifically bound molecules are seen only for exposure times <20 ms when the inducer isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) is present. Without IPTG LacI interacts longer with DNA, presumably because of partially specific interactions that are lost upon addition of IPTG (Fig. 4A). When we consider that the repressor will interact nonspecifically with several different DNA segments before leaving a diffraction-limited region, we infer that the fully nonspecific interactions with one DNA segment typically lasted shorter than 20 ms. A lower bound for the residence time can be estimated by the recent in vivo sliding-length measurements (5), where $\sqrt{D_1 t} \approx 32 \text{ bp}$, which gives $t > 3 \text{ ms}$ if we consider that the in vivo diffusion constant for sliding is lower than $D_1 = 4 \times 10^5 \text{ bp}^2/\text{s}$ measured in vitro (4). As a consequence the residence time determined from MD simulations is comparable to what can be estimated for in vivo conditions.

**Conclusions**

When the lac repressor is pulled slowly along the DNA, it spontaneously follows a helical path along the grooves although no forces are applied to turn the molecule or to move it radially. This rules out a more unstructured sliding on the DNA surface and makes biological sense because it aligns the DNA-binding interface of the repressor to the DNA, allowing for the interrogation of hydrogen-bonding partners presented by the bases (26). The result from the pulling simulation also motivated us to characterize the free-energy landscape along the helical coordinate, using umbrella sampling. However, to know whether the repressor actually slides along the DNA before dissociation, we also need to compare the free-energy barriers for sliding 1 bp to the energy for spontaneously dissociating. When we compare the average time for microscopic dissociation, $\tau_{\text{d}}$, with the calculated diffusion constant for sliding at 27 °C, we find that LacI is about 100 times more likely to slide 1 bp than to dissociate.

The atomistic modeling also allows us to keep track of structural features (Fig. S7) as well as individual hydrogen bonds, water molecules, and ions at the LacI–DNA interface. Ions and water are displaced as LacI slides along DNA (Fig. 3). However, they recondense at the other side of each DNA-binding domain and the sliding process is thus neutral with respect to the number of coordinated ions and water molecules. With increasing distance between LacI and DNA on the other hand, up to the reaction radius, the solvent recondenses on the DNA (Fig. S4A). The individual head domains are, apart from loops and the N-terminal tail, not particularly flexible during the sliding (Fig. S4B). Their position relative to each other is flexible, however, which indicates that the DNA-binding interface is insensitive to fluctuations in DNA structure and that LacI can slide also on (slightly) bent DNA. The link between the core and the head domains is flexible and allows large motions of the core domains, which could facilitate formation of DNA loops.

The simulations show that there are multiple ways for the side chains at the DNA-binding interface of LacI to form hydrogen bonds with DNA. In particular, we find that the hydrogen-bonding network is much more dynamic than that seen in Kalodimos et al.’s NMR ensemble (18). The dynamics of the hydrogen bonds while moving along the DNA are illustrated in Movie S1, where a number of structures from different umbrella ensembles were shown. The flexibility in the hydrogen-bonding network yields a nearly flat energy landscape for sliding, which facilitates fast diffusion. The complex dynamics of the hydrogen bonds limit the quality of calculated PMFs from simulations because the conformational exploration will be slow within the individual umbrella windows. Our simulated dataset is already massive, and increasing the length of the simulations by another order of magnitude to increase sampling is unfortunately beyond what is practically feasible with the computer hardware at our disposal. The overall correspondence between our results and macro-/mesoscopic measurements in the literature is therefore an indication that we capture the important characteristics of the free-energy landscape.

The results provide constraints in terms of microscopic thermodynamics and geometry that need to be considered when modeling the search process. First of all, the relatively large free-
energy penalty of dissociating from the DNA explains how and why other DNA-binding proteins hinder the passage of TFs in the search for the operator (5). Second, the helical sliding path means that the TF is in a position to recognize the specific binding site when sliding by. This increases the reaction cross-section significantly compared with diffusing in any orientation on the DNA surface.

It is remarkable that the free-energy profiles of the bound and free states meet at nearly the same level at the reaction radius in Fig. 2E; they were calculated in very different ways. Whereas the calculated potential for the bound state, i.e., inside the reaction radius, depends on the MD force field, the potential for the free state depends largely on the equilibrium binding constant obtained from experiments in vitro. The uncertainty in the equilibrium binding constant alone makes it likely that the striking agreement with the PMF at the reaction radius is partly a coincidence. In fact, the Brownian simulations of the dissociation processes suggest that the angular dependence on the LacI association to DNA is important for binding and that not all relevant rotation angles are tested before leaving DNA for an uncorrelated position. This implies that we do not expect a flat PMF at the reaction radius where the rotational degrees of freedom come into play. Instead, we expect a slight barrier in the radial reaction coordinate corresponding to a loss in entropy when the freely rotating LacI is locked into the binding configuration.

We have presented an important step toward a comprehensive understanding of TF search kinetics. The consequences will become important in the future elucidation of transcription-factor binding to the specific operator site and in the potential atomistic explanation of the low probability of binding to the operator site as recently reported (5). More generally, the results and the presented theoretical tools strongly connect macro-/mesoscopic events at the nanometer level, which enables a deeper interpretation of experimental observations.

Materials and Methods

Structure Preparation. No experimental structure exists in the literature for the complete LacI dimer bound to nonoperator DNA. To obtain a starting structure for the simulations herein the structural information for the non-specifically bound head domain (NS) and that for the full operator-bound LacI dimer (S) were combined as follows in a targeted-MD approach. The DNA sequence of S was replaced by the anti-operator from NS with MacroMolecule Builder (27), followed by targeted-MD simulation with harmonic position restraints applied to the DNA phosphate groups and the backbone of the helices in the head domains. Reference positions for the restraints were taken from NS and the force constant was set to 200 kJ·mol⁻¹·nm⁻². As such, the head domains were driven toward the non-specific conformation whereas the core domains were free to relax in response to this change. The system was simulated for 50 ns under these conditions. Protonation states of amino acids were corroborated with the H++ server (28). Two new systems were created by exchanging the anti-operator with 30 bp ideal B-DNA composed of alternating A and T in the first case and G and C in the second. In both systems the DNA helix was made periodic so that its one end was bonded to the other end of the neighboring periodic copy. The molecular systems simulated from hereon were contained in a rectangular simulation box measuring 10,143 x 13,019 x 13,019 nm³, comprising 172,752 (AT) and 172,662 (GC) atoms, including 53,116 water molecules, 181 Na⁺, and 115 Cl⁻. The systems were then simulated for 20 ns without position restraints to allow for additional equilibration.

Axial Pulling Simulations. LacI on AT repeats was used as a starting point for simulations, where LacI was pulled in the direction of the DNA axis, relative to 10 bp DNA that was immediately under LacI. Pulling forces were applied to the three α-helices of the head domains in the axial direction to avoid biasing the translocation toward a helical path. A force constant of 10,000 kJ·mol⁻¹·nm⁻² was used with one plane of reference per head domain. The planes, on which the pulling potential was centered, were slowly moved in the axial direction 0.34 nm (corresponding to 1 bp) over a period of 100 ns, after which the reference planes were held still for another 100 ns. Separate pulling simulations were done in both directions of the DNA axis.

Helical Umbrella Sampling. With LacI kept fixed, the DNA was shifted and rotated around its DNA axis in steps of 0.01 helical revolutions to cover a total of 0.2 revolutions along the helix, corresponding to a translocation of 1 bp in each direction. Every such starting structure was resolvated in 0.154 M NaCl. Energy minimization and 20 ns of MD simulation with a Berendsen thermostat (29) and a v-rescale thermostat (30) were followed to equilibrate all systems in a helical umbrella potential. In total there were 21 umbrella potentials for each of the two DNA sequences. All umbrella windows were simulated for 200 ns under conditions identical to the equilibration simulations except that a Parrinello–Rahman barostat (31) with time constant 0.5 ps was used and the helical umbrella potential was applied to the mutual COM of the head domains' helices. Force constants for the umbrella sampling were 2,000 kJ·mol⁻¹·turn⁻². Additional simulations were run with higher force constants (i.e., 3,000 kJ·mol⁻¹·turn⁻²) around points on the reaction coordinate that were poorly sampled in the first round of simulations. The first 50 ns of every simulation were omitted to allow for further equilibration before analysis. All umbrella-sampling data were combined through weighted-histogram analysis with the g_wham tool (32) into PMFs for AT and GC repeats. SDs were estimated through bootstrapping and the histograms were weighted according to the inverse of the autocorrelation times of the underlying data. The PMFs were forced to be cyclic because the DNA sequences were periodic.

Radial Umbrella Sampling. The energy-minimized starting structure for LacI on AT repeats was equilibrated for 200 ps using a v-rescale thermostat with a time constant of 2 ps. Then, to equilibrate the pressure, we simulated another 200 ps, where a Parrinello–Rahman barostat was used with a time constant of 0.5 ps. Umbrella potentials that moved away from DNA at a rate of 10⁻⁶ nm/ps and with force constants of 5,000 kJ·mol⁻¹·nm⁻² were then applied to the three helices of each head domain in the direction orthogonal to the DNA axis. The reference group for the pulling was 20 bp of DNA directly below LacI. The pulling proceeded for 45 ns, after which 12 configurations were selected for the subsequent umbrella sampling simulation, where each umbrella window was simulated for 200 ns. An additional 10
Supporting Information

Marklund et al. 10.1073/pnas.1307905110

Overall Approach and Workflow

Molecular-dynamics simulations where LacI was pulled along the DNA axis were used to determine the sliding reaction coordinate. From these simulations a helical coordinate was deemed the most physical (Fig. 1B). Subsequent umbrella sampling along this helical reaction coordinate allowed for the determination of the potential of mean force (PMF) (i.e., the free-energy profile), from which the one-dimensional diffusion constant $D_1$ was derived via Eq. 2. Another set of umbrella-sampling simulations, this time along the radial coordinate, allowed for the determination of the PMF in the radial direction and the microscopic residence time, i.e., the time to reach the microscopic reaction radius with the LacI in a specific orientation to the DNA. Brownian-dynamics simulations yielded an estimate of how many times LacI binds back to the DNA without losing rotational and translational correlations. The macroscopic residence time is the microscopic residence time times the number of rebinding events, and when combined with $D_1$ it gives an estimate for the sliding length.

Helical Umbrella Sampling

To express the umbrella potential and forces we define a coordinate system on the surface of a cylinder as described below and depicted in Fig. S2 A and B. We find it illustrative to imagine this cylinder to be cut open and unrolled into a flat surface. Let $\mathcal{O}$ be a line from the DNA axis, stretching to infinity and orthogonal to the DNA axis, intersecting the helical path. On $\mathcal{O}$ we center an umbrella potential fulfilling criteria $i$–$iv$ that were formulated in the main text. For a given radius $r$ from the DNA axis we define a cylindrical surface $C_r$ and denote the point where $\mathcal{O}$ intersects $C_r$ by $\mathcal{O}_r$. We define the two unit vectors $\mathbf{u}$ and $\mathbf{v}$ on the cylinder $C_r$; $\mathbf{u}$ is parallel to the helical path and $\mathbf{v}$ is orthogonal to $\mathbf{u}$ ($\mathbf{v} = \mathbf{u} \times \mathbf{r}$). $(\mathbf{u}, \mathbf{v})$ form an orthonormal basis on $C_r$ that describes any point on $C_r$ as $i\mathbf{u} + j\mathbf{v}$. The u component describes how far along the helical path a point is relative to $\mathcal{O}_r$ and allows us to formulate the umbrella potential $U$,

$$U(u, r) = k \frac{u^2}{L(r)},$$

where $L \equiv \sqrt{A^2 + r^2 B^2}$ is the length of a helical turn. $A$ and $B$ are the axial and angular components of a full turn; hence $B \equiv 2\pi$ whereas $A$ depends on the helical pitch. For canonical B-DNA we get $L \equiv \sqrt{3.4^2 + 0^2} \text{ nm}$. This choice of $L$ means that the unit for the reaction coordinate is the number of full helical turns, which in the case of canonical B-DNA corresponds to 10 bp. Other definitions of $L$, and consequently of $A$ and $B$, work equally well but yield other units for the reaction coordinate. We chose to have an umbrella potential with no explicit $v$ dependence. In principle we could apply forces also in the $v$ direction (with further consequences for the radial force) but we prefer to allow the transcription factor (TF) to be free to diverge from the helical path in $v$ because this “disentangles” the helical reaction coordinate from the radial reaction coordinate that describes the dissociation from DNA.

The umbrella force is by definition $\mathbf{F} = -\nabla U$. Its helical component, i.e., in the $u$ direction, is given by differentiating $U$ by $u$:

$$F_u = -2k u \frac{u}{L}.$$  

Similarly, the radial component is obtained by differentiating $U$ by $r$,

$$F_r = -k B u \left(\frac{2L}{L^2} + \frac{L}{2L} \right),$$

where $\gamma \equiv 2\pi r$ and $\theta$ is in turn the angle of the TF relative to the DNA. For the sake of completeness we define the unit vectors $\gamma$ pointing in the angular direction, and $\mathbf{Z} \equiv (\hat{z}, \mathbf{\gamma})$ is an additional orthonormal basis in $C_r$ that we do not depend on directly, but find useful for inspecting the umbrella forces and potential. For $B = 2\pi$ Eq. S1c becomes

$$F_r = -k B u \left(\frac{2L}{L^2} + \frac{L}{2L} \right).$$

Eqs. S1a–S1d allow us to apply the umbrella potential and forces in simulations and to produce PMFs through, e.g., weighted histogram analysis. The umbrella potential, the force components, and the relative magnitude of the latter are further illustrated in Fig. S2 C–F.

As seen in Fig. S2D, $F_r = 0$ on the helical path (i.e., on $v = 0$) and on $u = 0$. $F_r$ is also centrosymmetric around the umbrella reference point $\mathcal{O}_r$ and antisymmetric when reflected in, or orthogonal to, the helical path. The centrosymmetry arises from the fact that both $\gamma$ and $u$ change sign for any point that is inverted around $\mathcal{O}_r$. For reactions that do not follow the ideal helical path there may, within a single umbrella window, be a slight net force in the radial direction. In the system under study here the radial force is substantially smaller than the helical force; for typical values of $r$ [F$_u$] remains $>50$ times larger than $|F_r|$ within $0.4$ nm from the helical path and $>80$ times larger within $0.25$ nm. As such, we expect the radial force to have a negligible impact on the calculated PMFs, especially because the axial pulling resulted in deviations of just a few Ångströms from the ideal helical path (Fig. S2B). We note that $\lim_{u \to 0} u = \gamma$ (a consequence of item ii in the main text) and that $\lim_{v \to 0} F_r = 0$ as the difference in the numerator of Eq. S1d approaches zero. The denominator furthermore becomes $r^4$ and the terms in the numerator scale as $r^2$, which also results in a vanishing $F_r$. Similarly, setting $A = 0$ yields the special case of a pure angular potential with $F_r = 0$ for all $r$.

Diffusion Time in a Potential

The relatively slow variation of the free energy of the bound state along the reaction coordinate (the radial PMF in Fig. 2E in the main text) suggests that dissociation is not limited by an activation process, e.g., like a conformational change. Thus, the microscopic dissociation rate can be calculated from the diffusion time out of the potential well. Consider a single molecule freely diffusing along a reaction coordinate $r$ with rate $D_3$ in a potential $G_r(r)$. Calculate the mean time to reach the endpoint at $r = \rho$ from any starting distribution within the interval $r_0 < r < \rho$. If $G_r(r)$ is given in units of $k_BT$, the 1D-diffusion equation is

$$\frac{\partial c}{\partial t} = D_3 \frac{\partial^2 c}{\partial r^2} + D_3 \frac{\partial c}{\partial r} \frac{dG_r}{dr} + D_3 \frac{\partial^2 c}{dr^2}.$$  

At time $t = 0$, the initial probability distribution for the molecule in the interval is $c_0(r)$. The boundary conditions are reflecting at $r = r_0$ and absorbing at $r = \rho$:  

Marklund et al. www.pnas.org/cgi/content/short/1307905110

1 of 8
\[
\left[ \frac{dc}{dr} + \frac{dG_b}{dr} \right] = 0, \tag{S3a}
\]

\[
c(r = \rho, t) = 0. \tag{S3b}
\]

Integrating Eq. S2 over time from 0 to infinity gives

\[
-c_0(r) = D_3 \frac{d^2c}{dr^2} + D_1 \frac{dc}{dr} - D_3 \frac{d^2G_b}{dr^2}. \tag{S4}
\]

Here \( \hat{c} = \int_0^\infty c(r, t) dt \). This can be integrated over \( r \) to give

\[
\int_0^r c_0(r) dr = -D_3 \hat{c} \left[ \frac{d^2}{dr} - D_3 \frac{d^2G_b}{dr} \right]. \tag{S5}
\]

This has the solution for \( \hat{c} \),

\[
\hat{c}(r) = \frac{1}{D_3} \int_0^r e^{-G_b(r)} \left[ \int_0^r e^{-G_b(r)} dr \right] \hat{c}(r) dr. \tag{S6}
\]

As \( \hat{c}(r) dr \) is the mean residence time in the interval \( (r, r + dr) \), the total residence time in the whole interval \( (r_0, \rho) \) is

\[
\tau_d = \int_0^\rho \hat{c}(r) dr. \tag{S7}
\]

The effective dissociation rate is \( k_{1\rightarrow 0}^{\text{macro}} = 1/\tau_d \). If the initial distribution is an equilibrium distribution, we have

\[
c_0(r) = e^{-G_b(r)} \int_0^\rho e^{-G_b(r^*)} dr. \tag{S8}
\]

This satisfies the equilibrium requirement and is normalized to 1 (a single molecule) in the interval \( (r_0, \rho) \).

**Transforming a 2D Free-Energy Landscape to an Equivalent 1D Description Along the Reaction Coordinate**

The interaction between a TF and nonspecific DNA is essentially a two-dimensional problem as translation in the third dimension (along the DNA axis) is invariant. The reaction coordinate, \( r \), can be defined by the distance between the center of mass of the TF and the helix axis of the DNA. Here we consider the free-energy profile experienced by the TF as it moves along the reaction coordinate. In the region \( r_0 < r < \rho \), the molecules are considered bound and in the region \( \rho < r < R_b \), they are unbound. The bound state is characterized by very stringent orientational constraints. The free-energy potential in the bound state, \( G_b(r) \), has been determined (the radial PMF) at distances \( r, r_0 < r < \rho \). Thus, this potential includes the entropic contributions from whatever “wiggle room” is allowed by the steric constraints. At equilibrium, the 2D concentration in the free state is constant = \( c_2 \) (molecules per unit area). Along the reaction coordinate, the corresponding probability distribution will be \( c_1(r) = 2\pi r c_2 \). Similarly, at equilibrium, the 1D concentration along the bound-state coordinate will be \( c_b(r) = C e^{-G_b(r)} \). \( C \) and \( c_2 \) are unknown constants required for normalization of the probability distribution. In the transition region where the TF goes from bound to free, the orientational constraints are relaxed. This could lead to a sharp increase in configurational entropy and therefore to a decrease in the free energy, \( G_f(r) \), of the free state. This is compensated by the concomitant release of binding interactions. However, the details of the free-energy profile in the transition region are of little importance for the equilibrium distribution as long as \( R_b \gg \rho \). At equilibrium the chemical potential along the reaction coordinate must be equal everywhere:

\[
\mu(r) = \ln(c_b(r)) + G_b(r) = \ln(C); \text{ for } b < r < \rho
\]

\[
\mu(r) = \ln(2\pi r c_2) + G_b(r); \text{ for } \rho < r < R_b.
\]

Thus, by demanding that \( \mu(r) \) be constant, one finds \( C = 2\pi rc_2 e^{G_b(r)} \) or, equivalently, \( G_f(r) = \text{constant} - \ln r \). This gives rise to an entropic radial force \( \tau = (1/r) \). For the bound state the probability is \( P_b = \int_0^\rho e^{-G_b(r)} dr \), and for the free state it is \( 1 - P_b = \int_0^\rho c_1(r) dr = c_2 \pi (R_b^2 - \rho^2) \). These probabilities are related to the nonspecific binding constant \( K \) as

\[
K = \frac{P_b}{\tau R_b^2} = \frac{2\pi c_2 \int_0^\rho e^{-G_b(r)} dr}{R_b^2 - \rho^2}. \tag{S10}
\]

Here \( \tau = 3.4 \times 10^{-4} \mu \text{m} \) is the width of a base pair and \( K \) must be expressed as solution volume per molecule rather than per mole. The concentration of base pairs is \( 1/(\tau R_b^2) \); for a given DNA concentration, this defines \( R_b \). Now \( G_f \) can be solved for and, if \( R_b^2 \gg \rho^2 \), this gives

\[
G_f(r) = \ln \left[ \frac{K}{2\pi \tau R_b^2} \int_0^\rho e^{-G_b(r)} dr \right]. \tag{S11}
\]

In this expression, \( \rho \) must be chosen sufficiently large so that for \( r \geq \rho \), all orientations are freely available and all interactions with DNA released. This value for \( \rho \) may be larger than the endpoint for the calculated PMF. However, \( e^{-G_b(r)} \) is totally dominated by the contributions for small \( r \) where \( G_b(r) \) is the smallest, and the exact choice for the upper limit in the integral in Eqs. S10 and S11 matters little. Exactly what happens with the free-energy landscape in the transition region around \( r = \rho \) is uncertain, as discussed above, but at distances beyond \( r = \rho \), the free energy along the reaction coordinate will decrease as given by Eq. S11.

It should be noted that the reaction coordinate, \( r \), in the free state is defined as the distance between the center of mass of the TF and the helix axis of the DNA. In the bound state, however, the PMF was calculated using a reaction coordinate defined as the distance between the center of mass of the head groups and the DNA axis. In the bound state, the center of mass of the TF is located \( R_{OC} \approx 3.1 \text{ nm} \) farther away from the DNA than are the head groups. Thus, if the dissociation distance is chosen as \( \rho = 2.8 \text{ nm} \) in the bound state (Estimating the Reaction Radius and Electrostatic Considerations section), this would correspond to an association distance (reaction radius) in the free state of \( ca. \rho_f = 5.9 \text{ nm} \). This is the value of \( \rho_f \) that enters the calculations of \( k_{assoc} \) and \( k_{dissoc} \) below and in the main text. However, because nonspecific association is a 2D process, these rates depend very weakly, only logarithmically, on \( \rho_f \).

It is remarkable that the free-energy profiles of the bound and free states meet at nearly the same level at the reaction radius; they were calculated in very different ways. Although the details of the free-energy landscape in the transition region are uncertain, this agreement is consistent with the proposition that nonspecific association for LacI is determined by diffusion. If the levels had been very different, e.g., if \( G_f(r) \ll e^{-G_b(r)} \), then association would not be diffusion limited. Alternatively, if
\( e^{G_f(\eta)} \gg e^{G_s(\rho)} \), dissociation into solution would require a distinct second step and the value of the dissociation distance \( \rho \) would not have been chosen adequately.

**Microscopic vs. Macroscopic Rates**

The present molecular dynamics (MD) calculations describe the microscopic parameters, specifically \( D_1 \) and \( k_{d\text{micro}} \), which refer to the bound complex. These results can be set in context via the theory of facilitated diffusion (2, 3), which couples the motion of the TF along the DNA while bound with its motion free in solution. In the following we give a short summary of the parts of the theory that pertain to the present work. The free diffusion can be described by a 2D diffusion equation for the TF,

\[
\frac{\partial c_2}{\partial t} = D_2 \frac{\partial^2 c_2}{\partial r^2}.
\]

This is coupled to the bound state via a boundary condition at the DNA cylindrical surface, \( \rho_f \):

\[
2\pi D_3 \rho_f \frac{\partial c_2}{\partial r} \bigg|_{r=\rho_f} = \frac{k_{\text{micro}}}{\ell} c_2(\rho_f). \tag{S13}
\]

\( k_{d\text{micro}}/\ell \) serves as a local association rate constant for the TF once it is at the reaction radius. The factor \( 1/\ell \) simply serves to determine the units of \( k_{d\text{micro}} \) as a rate constant per base pair. Beyond the correlation distance, \( R_c \), from one DNA segment, the TF is more likely to approach another segment. This can be accounted for by an absorbing boundary condition

\[
c_2(r = R_c) = 0. \tag{S14}
\]

\( R_c \) can be defined from the density, \( D_T \), of nonspecific DNA in solution such that \( D_T = 1/(\pi R_c^2) \). Consider a TF that has just dissociated; What is the probability, \( P_{\text{diss}} \), that it does not return to the same segment it just left? Using an initial distribution, \( c_2^0(r) = \delta(r-r_f)/(2\pi\rho_f) \), and integrating the diffusion equation over all \( t \), \( \tilde{c}_2(t) = \int_0^\infty c_2(r, t)dr \), one finds

\[
\frac{\delta(r-r_f)}{2\pi\rho_f} = D_2 \frac{\partial}{\partial r} \left( r \frac{\partial \tilde{c}_2}{\partial r} \right).
\]

The integrated net flux across the outer boundary at \( r = R_c \) is found to be

\[
P_{\text{diss}} = -2\pi D_3 R_c \frac{\partial \tilde{c}_2}{\partial r} \bigg|_{r=R_c} = \frac{1}{1 + \alpha \ln \left( R_c/\rho_f \right)}, \tag{S16}
\]

where \( \alpha = k_{d\text{micro}}/(2\pi D_3) \). Thus, a TF that is present just outside the reaction radius has a probability \( P_{\text{diss}} \) of completing the dissociation, and the macroscopic dissociation constant will be

\[
k_{d\text{macro}} = \frac{k_{d\text{micro}}}{1 + \alpha \ln \left( R_c/\rho_f \right)}. \tag{S17}
\]

With probability \( 1 - P_{\text{diss}} \), the TF simply returns to the same DNA segment. The corresponding macroscopic association rate constant can be found from the nonspecific binding constant

\[
K = \frac{k_{d\text{macro}}}{k_{d\text{micro}}} = \frac{k_{d\text{macro}}}{k_{d\text{micro}}}. \tag{S18}
\]

which gives

\[
k_{d\text{macro}} = \frac{k_{d\text{micro}} k_{u\text{macro}}}{k_{d\text{micro}} + k_{u\text{micro}}} = \frac{k_{d\text{micro}} k_{u\text{macro}}}{1 + \alpha \ln \left( R_c/\rho_f \right)} = \frac{2\pi D_3}{\ell} \left( 1/\alpha + \ln \left( R_c/\rho_f \right) \right). \tag{S19}
\]

In the limit \( \alpha \gg 1 \), the macroscopic association rate constant reaches its maximum value \( 2\pi D_3/[\ln(\rho_f/\rho_s)] \), which corresponds to the Smoluchowski limit, \( 4\pi D_3 \rho_s \), for the association of spheres. Effectively \( \alpha \) serves as a parameter that reduces the macroscopic rates from their maximal diffusion-limited values. Therefore, \( \alpha \) is referred to as the extent of diffusion control. However, in the system studied here, we have calculated \( k_{d\text{micro}} \) from a diffusion process along the reaction coordinate. Thus, the calculated \( \alpha \) is independent of the diffusion rate (i.e., independent of viscosity changes) and is not strictly a measure of diffusion control in this case. Nevertheless, \( \alpha \) is the main determinant for the dissociation probability, Eq. S16. For very large values of \( \alpha \), almost all microdissociations simply lead to an almost immediate rebinding to the same base pairs that the TF had just left. These short excursions contribute very little to the movement of the TF along the DNA axis and the sliding distance during the macroscopic lifetime of the complex can be calculated as \( s = \sqrt{D_1 k_{d\text{micro}}} \).

**Orientation Constraints in 2D Diffusion**

Not only is the association of a TF with nonspecific DNA dependent on the molecules “touching” each other at the reaction radius, but also their relative orientations are important. Although it is difficult to find a useful analytical expression that describes the influence of all possible orientation angles on the DNA and the protein, it is straightforward to calculate the effect from one of them. Consider the diffusion flux of fully reactive circular particles onto a stationary circular target (the DNA) with reaction radius \( \rho_f \) and reactive orientation angle \( |\phi| < \phi_0 \). In polar coordinates \((r, \phi)\), the steady-state concentration, \( c \), of particles outside the reactive target will satisfy

\[
\frac{1}{r} \frac{\partial}{\partial r} \left( r^2 \frac{\partial c}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 c}{\partial \phi^2} = 0. \tag{S20}
\]

At the target boundary, \( r = \rho_f \), the particle will bind if it approaches from the correct orientation. This can be described by the radiation boundary condition, Eq. S13,

\[
2\pi D_3 \rho_f \frac{\partial c}{\partial r} \bigg|_{r=\rho_f} = \kappa c \left( \rho_f, \phi \right): \quad |\phi| < \phi_0. \tag{S21a}
\]

Here \( \kappa \) is the local (microscopic) reaction rate at the target surface. In the wrong orientation, the boundary condition is reflective:

\[
\frac{\partial c}{\partial r} \bigg|_{r=\rho_f} = 0; \quad \phi_0 < |\phi| < \pi. \tag{S21b}
\]

Furthermore, the concentration is kept constant, \( c_0 \), at the outer boundary

\[
c(r = R_c, \phi) = c_0. \tag{S22}
\]

Under these conditions the association flux is given by

\[
J = D_3 \rho_f \int_{-\pi}^{\pi} \frac{\partial c}{\partial r} \bigg|_{r=\rho_f} d\phi = \frac{\kappa}{2\pi} \int_{-\phi_0}^{\phi_0} c(\rho_f, \phi) d\phi. \tag{S23}
\]

Once \( J \) is determined, the association rate constant is defined by
\[ k_a = J/c_0. \]  

The general solution to Eq. S20 that satisfies the boundary condition [S22] at the outer boundary is

\[ c(r, \phi) = c_0 + A_0 \ln(r/R_c) + \sum_{n=1}^{\infty} A_n R_c^n \left[ (r/R_c)^n - (r/R_c)^{-n} \right] \cos(n\phi). \]  

Following the approach of Shoup et al. (4), which works very well for the corresponding problem in 3D, we simplify the boundary conditions [S21a] and [S21b] by assuming that the influx is homogenous over the reactive region, \(-\theta_0 < \phi < \theta_0\); i.e.,

\[ D_3 \frac{\partial c}{\partial r} \bigg|_{r=r_0} = \begin{cases} J/2\theta_0 & |\phi| < \theta_0 \\ 0 & \theta_0 < |\phi| < \pi. \end{cases} \]  

The unknown coefficients \(A_n\) can now be determined by applying Eq. S26 to the general solution [S25]. This gives

\[ A_0 = J/2\pi D_3 \]

\[ A_n = \frac{J}{\pi D_3} \frac{\sin(n\theta_0)}{n\theta_0} \frac{1}{n^2 \left( 1 + \left( R_c/\rho \right)^n \right)}; \quad n > 0. \]  

Now the solution \(c(\rho, \phi)\) is expressed in the still unknown association flux \(J\). This can be solved for by setting \(J = \int_0^{\pi} c(\rho, \phi) d\phi\) from Eq. S23. This gives for the association rate constant

\[ \frac{1}{k_a} = \frac{\pi}{\theta_0 \kappa} + \frac{\ln(R_c/\rho)}{2\pi D_3} + \frac{1}{\pi D_3} \sum_{n=1}^{\infty} \left( \frac{\sin(n\theta_0)}{n\theta_0} \right)^2 \frac{\tan[n \ln(R_c/\rho)]}{n}. \]  

When the third dimension along the DNA helix is introduced, the rate constants should be multiplied by the base pair length, \(t\), to give the appropriate units per concentration of base pairs. Thus,

\[ \frac{1}{k_a^{\text{macro}}} = \frac{1}{k_a^{\text{micro}}} + \frac{1}{2\pi D_3 t} \left( \ln(R_c/\rho) \right) + \frac{1}{2} \left( \frac{\sin(n\theta_0)}{n\theta_0} \right)^2 \frac{\tan[n \ln(R_c/\rho)]}{n}. \]  

\[ k_a^{\text{micro}} = \pi \theta_0 / \kappa \]  is defined from the boundary condition \(\kappa\) such that \(k_a^{\text{macro}} \rightarrow k_a^{\text{micro}}\) in the limit when diffusion through solution is infinitely fast. Eq. S29 can be rewritten as

\[ k_a^{\text{macro}} = \frac{2\pi D_3 t}{\alpha \ln(R_c/\rho) + f(\theta_0)}. \]  

Here \(f(\theta_0)\) is the effect of the angular constraint as given by the sum in Eq. S29 and \(\alpha = k_a^{\text{macro}} / 2\pi D_3 t\) as above. The probability that a microscopic dissociation will lead to a macroscopic one is

\[ P_{\text{diss}}^{\text{macro}} = \frac{k_a^{\text{macro}}}{k_a^{\text{micro}}} + \frac{1}{1 + \alpha \left( \ln(R_c/\rho) + f(\theta_0) \right)}. \]  

When the whole target is reactive, \(\theta_0 = \pi, f(\theta_0) = 0\), and the previous result, Eq. S19, is recovered. It can be noted that \(f(\theta_0)\) is largely independent of the radii as long as \(\ln(R_c/\rho) > 2\) or so. In this case, the angular effect could be considered as a reduction in the effective reaction radius such that \(\rho_1\) in Eqs. S16–S19 would be replaced by \(\rho_{\text{eff}} = \rho_1 \exp(-f(\theta_0))\). The result in Eq. S31 agrees very well with the simulations described below (Fig. S6).

**Stochastic Reaction–Diffusion Simulations**

To deduce how the macroscopic dissociation is affected by a reactive patch also on the TF we resorted to stochastic reaction–diffusion simulations. The 2D system described above is spatially discretized into concentric shells, with DNA in the center and the TF diffusing by radial jumps between shells, and implemented as a reaction-diffusion master equation (RDME) (5). In addition to radial diffusion we explicitly model circular diffusional motion of the TF around the DNA with rotational angle \(\phi\) and rotational diffusion of the TF around its own axis with angle \(\theta\). For each sampled time step in the RDME simulation new \(\phi\) and \(\theta\) angles are sampled according to the diffusion equation. Here the distributions of angular rotations are given by

\[ c(\delta \phi, t) = \frac{1}{\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} \cos(n \delta \phi) \]  

for the circular motion of the TF around the DNA and

\[ c(\delta \theta, t) = \frac{1}{\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} \cos(n \delta \theta) \]  

for the rotation of the TF around its own axis. \(\delta \phi\) and \(\delta \theta\) are the increments in \(\phi\) and \(\theta\) angles after a time \(t\). \(r\) is the radial distance of the TF to the center of DNA. \(D_3\) is the rotational diffusion constant related to \(D_3\) as \(D_3 = 3D_3/4R^2\), where \(R\) is the radius of a spherical particle. Care must be taken such that the time steps are short to limit the error due to discretization of the rotational motions in relation to TF binding to DNA. The radial discretization is made so fine that after the TF has on average jumped from the innermost shell to its adjacent neighbor the SDs of the increments in rotational angles are at most 0.03.

The binding reaction between TF and DNA is restricted to a half-angulard width of \(\theta_0\) of the DNA and furthermore to a half-angulard width of \(\theta_0\) on the TF. In simulations this restriction is accounted for by making the binding propensity nonzero only when both the conditions \(|\phi| < \theta_0\) and \(|\phi - \theta| < \theta_0\) are fulfilled. For estimation of the probability of reaching the outer boundary of radius \(R_c = 100\) nm, one TF is started at time 0 in the innermost shell with both the rotational angles \(\phi = \theta = 0\) and the fraction of times the TF jumps out of the outer boundary instead of binding back to the DNA. However, for the case with a reactive patch on DNA alone (\(\theta_0 = \pi/10\) and \(\theta_0 = \pi\)) the initial angle was instead sampled from a uniform distribution between \(-\pi/10\) and \(+\pi/10\) to allow for comparison with the theory above. \(R = 4.5\) nm is the radius of the TF and the radius of the DNA is 1 nm, which together form a reactive radius \(\rho_0 = 5.5\) nm.

The results of simulations for different values of the microscopic association rate constant are shown in Fig. S6. For the cases of completely reactive boundaries (\(\theta_0 = \pi\) and \(\theta_0 = \pi\)) and a reactive patch on the DNA alone (\(\theta_0 = \pi/10\) and \(\theta_0 = \pi\)) the simulated result is compared with the analytical counterparts as described above. We find that also in the presence of reactive patches on both the DNA and the TF \(1/P_{\text{diss}}^{\text{macro}}\) follows a linear relationship in the microscopic association constant. This relation is then used to find \(1/P_{\text{diss}}^{\text{macro}}\) values in the computationally intractable regimes set by the microscopic association constant.
The microscopic association rate constant is calculated from the microscopic dissociation rate constant and the equilibrium constant for TF binding to DNA in vitro: \( k_{\text{ass}}^{\text{micro}} = k_{\text{diss}}^{\text{micro}} \approx 4.8 \text{ M}^{-1} \cdot \text{s}^{-1} \) when \( N = 2 \cdot 10^4 \text{ M}^{-1} \approx 3.3 \cdot 10^4 \text{ M}^{-1} \), and \( k_{\text{diss}}^{\text{micro}} = 1.45 \cdot 10^3 \text{ s}^{-1} \). For the reactive angles of \( \theta_0 = \pi/2 \) and \( \theta_0 = \pi/10 \) this gives the number of rebinding events before macroscopic dissociation as \( n = 1/k_{\text{diss}}^{\text{micro}} \approx 520 \) and for reactive angles \( \theta_0 = \pi/3 \) and \( \theta_0 = \pi/20 \) the number of rebinding events is \( N \approx 880 \). The estimate is admittedly crude because we do not allow for diffusion and rotations out of the plane and should be considered a first-order approximation of the geometric effects that LacI and DNA are not reactive over all their surfaces.

**Estimating the Reaction Radius and Electrostatic Considerations**

When LacI dissociates, we would expect water to recondense on DNA where LacI was bound. Indeed, the number of water molecules around the DNA segment increases with the center of mass (COM) distance between the head domains of LacI and the DNA (Fig. S4A) and starts to level off around umbrella window 11, which corresponds to a COM separation of 2.7 nm. The reaction radius is not necessarily synonymous with broken direct contacts, however. Rather, the concept of a bound state is one founded in thermodynamics and correlation between molecules (6). In the bound state the motions of LacI and the DNA are highly correlated, a correlation that is lost in the unbound state. The distance at which this correlation is lost indicates the reaction radius \( \rho \). When we restrain the radial distance \( r \) between the head domains and the reference DNA segment to 2.7 nm or more (umbrella window 11 and beyond), we observe an increased mobility for LacI relative to the DNA segment, reflected by the increased root mean square deviation (rmsd) for the axial component of their COM distance (Fig. S4B). Beyond this distance the interactions of both head domains with the DNA are disrupted such that LacI is no longer associated primarily with the reference DNA segment to which it was originally bound (Fig. S5). Smaller deviations can be seen already at 2.25 nm, but these can be attributed to excursions made by a single head domain while the other one remains associated with the DNA segment. Based on these observations we conclude that the reaction radius is near 2.7 nm and set the reaction radius \( \rho \approx 2.8 \text{ nm} \) in our calculations to be on the safe side. Setting \( \rho \) too large will in the absence of a free-energy barrier make no or little difference in other results depending on \( \rho \). Note that this is a “soft” boundary between the bound and the unbound states because of the absence of an energy barrier in the PMF. The PMF (Fig. 2D) is relatively flat around 2.8 nm; hence other results depending on \( \rho \) are relatively insensitive to moderate errors in the value of \( \rho \).

For distances where LacI is no longer associated to the DNA segment where it started the measured force in our umbrella-sampling simulations is misrepresentative for our reaction coordinate and the PMF can therefore not be calculated in a reliable way beyond that point. We know that the short-range interactions between DNAs are weak and transient beyond \( \rho \), so they will not contribute to the PMF farther out. Electrostatic interactions can be important on longer distances and therefore in principle determine the shape of the PMF beyond \( \rho \). The net charge of LacI in our simulation is \(-6\) elementary charge units, which means that there are no attractive long-range charge–charge interactions between LacI and the negatively charged DNA. The potential of other, higher-order, electrostatic interactions declines rapidly with distance, \( 1/r^n \) where \( n \geq 2 \), which means that their contributions to the PMF on longer distances are likely negligible. As such, the important features of the PMF are revealed by our umbrella-sampling simulations, with the possible exception of entropic effects arising from rotational degrees of freedom becoming more accessible beyond \( \rho \) as short-range interactions are severed between LacI and DNA. Finally, in classical simulations such as ours with nonpolarizable force fields, the electrostatic interactions involving ionic species are slightly overestimated (7). This could in principle have bearing for the electrostatic contribution to the PMFs that we have calculated. We expect the effect to be relatively small, however, because the interactions are not solely composed of electrostatics and because only part of the electrostatics is between ionic residues. It follows that the consequences for the sliding length, residence time, \( \alpha \), and the conclusions that we draw therefore are minor.

**Estimating the Impact of Uncertainty in the PMF’s Roughness on Sliding Length**

The nonlinear dependence of the sliding length on the roughness in the PMF makes rigorous estimates of the error difficult. The error estimates for different parts of the PMF, obtained by bootstrapping, are furthermore not independent, which makes them even more involved. To estimate an upper error bound we therefore scale the amplitude of the PMF for AT repeats so that the largest peak touches the upper error bar (i.e., by a factor 1.46) and calculate how this changes the 1D diffusion coefficient and the sliding length. Similarly, we estimate the lower error bound by scaling down the amplitude so that the largest peak touches the lower error bar (i.e., by a factor 0.54). This makes the error bounds 0.57 and 1.43 times our estimated sliding length in addition to the uncertainty from the diffusion–reaction simulations.

**Flexibility of the DNA-Binding Interface**

To assess the structural flexibility of the DNA-binding interface during the sliding we analyzed the helical umbrella-sampling trajectories by calculating the root mean square deviations (rmsd) for the \( C_\alpha \) of the head domains’ helices. rmsd were calculated for the two head domains individually and together after superimposing them on the first structure of Kalodimos et al.’s NMR ensemble (8). As seen in Fig. S3 each individual head domain remained close to the NMR structure whereas the relative orientations of the two displayed much larger variations. This means that, although the structures of the individual head domains are relatively static, the DNA-binding interface as a whole is flexible and suggests that LacI can remain bound also in the presence of large-scale fluctuations in DNA structure.

During our simulations, and the end of the trajectory. The MC optimization minimized the number of rebinding events before macroscopic dissociation as \( n = 1/k_{\text{diss}}^{\text{micro}} \) and, to create a smooth trajectory they were sorted using DNA coordinates, allowing creation of a movie of a head domain moving over the DNA. All these conformations were concatenated, and, to create a smooth trajectory they were sorted using a Monte Carlo (MC) algorithm. This algorithm was implemented in the GROMACS program \( g \_\text{cluster} \). First, the two structures with the largest rmsd between proteins were put at the beginning and the end of the trajectory. The MC optimization minimized the rmsd between adjacent frames in an attempt to find the shortest path between the two extremes (Fig. S7). Finally, to smooth the movie further a linear interpolation of the coordinates of three frames between each trajectory frame was done using the \( g \_\text{morph} \) program in GROMACS. The hydrogen bonds between protein and DNA in Movie S1 were computed on the fly during movie generation in PyMOL (Schrödinger Software) and may be different from those calculated for noninterpolated structures.

Marklund et al. www.pnas.org/cgi/content/short/1307905110

5 of 8

Fig. S1. The core domain, which is leaning toward the head domain of the A chain in the crystal structure, moves with respect to the head domains in the simulations. The two images are the final frames from two neighboring umbrella-sampling simulations.

Fig. S2. (A) The helical path on a cylinder and the coordinate system (\(u,v\)) used for expressing the reaction coordinate and the helical umbrella potential. (B) The unrolled cylinder and how (\(u,v\)) changes with \(r\). (C–F) Helical umbrella potential and forces around \(\Omega\), for a force constant of 1 a.u. and \(r = 1.4\) nm (typical for the LacI head domains on dsDNA). (C) Helical umbrella potential \(U\). (D) Helical force component \(F_u\). (E) Radial force component \(F_r\). (F) Comparison between force components, i.e., \(|F_u/F_r|\). The radial force arises from the helix’s \(r\) dependence, but its magnitude is dwarfed by that of the helical force close to the helical path for \(r = 1.4\) nm.
Fig. S3. rmsd for all C\textalpha of the head domains’ helices with respect to the Kalodimos et al. NMR structure (8). Each individual head domain was found to be relatively static, but their orientation relative to each other varied substantially. As such, the DNA-binding interface is flexible.

Fig. S4. The radial umbrella-sampling simulations. (A) Amount of water around the DNA segment to which LacI was originally bound, for umbrella windows at increasing distance. Water recondenses on the DNA as LacI is kept farther from the DNA. (B) rms displacement of the head domains in the direction of the DNA axis. The interactions between LacI and DNA are weakened to a point where LacI is free to move with respect to the DNA. Around umbrella window no. 6 the mobility starts to increase, which is due to the transient dissociation of single head domains, whereas in umbrella window no. 11 both head domains dissociate from DNA, which is reflected by a jump in the rmsx.

Fig. S5. (A and B) Final frames from the radial umbrella-sampling simulation nos. 10 (A) and 11 (B), with LacI shown in gray, the DNA segment in yellow, and other parts of DNA in orange. In simulation no. 11, where the COM distance between the head domains and the DNA segment is restrained to 2.7 nm, LacI detaches from the DNA segment. The same phenomenon occurs in simulations with larger restraint distances, implying that the reaction radius is close to 2.7 nm.

Fig. S6. Linear dependence between number of rebinding events and microscopic association rate constant. Squares show simulated data. Solid lines show analytical solutions as described in Eqs. S16 (black) and S29 (green). Dashed lines show linear regressions of the corresponding simulated data. The first three points in each curve were simulated with 1,000 spatially discrete shells whereas the remaining points were simulated using 10,000 shells.
Fig. S7. rmsd matrix between the sorted frames from all helical umbrella-sampling trajectories combined into one long trajectory.

Movie S1. Shown is one of the head domains sliding (backward) over the DNA piece. The lysine side chains in the loops of the head domain interact intermittently with the DNA, highlighting why extremely long simulations were necessary to obtain a converged potential of mean force.

Movie S1