Rapid Search for Specific Sites on DNA Through Conformational Switch of Nonspecifically Bound Proteins

Huan-Xiang Zhou*

Department of Physics and Institute of Molecular Biophysics, Tallahassee, Florida 32306, USA

Supplementary Information Appendix

Search Time t_s and Bimolecular Rate Constant k_a for the Two-State Model. In this model, illustrated by the unimolecular scheme of the main text, the DNA-binding protein is assumed to transfer stochastically between the bulk solution and the DNA surface; in the latter region the protein has a finite chance of being captured by the specific site. These steps are modeled as ordinary chemical kinetics, with rate constants κ_3, κ_{3-}, and κ_1. Correspondingly the average lifetime of P_b, the protein in the bulk solution, is

$$\bar{t}_{3d} = \frac{1}{\kappa_3}, \qquad [S1]$$

and the average lifetime of P_{ns}, the nonspecifically bound protein, is

$$\bar{t}_{1d} = \frac{1}{\kappa_{3-} + \kappa_1}. \qquad [S2]$$

Note that P_{ns} can decay via two pathways: to P_b with rate constant κ_{3-}, and to P_s (the target-bound protein) with rate constant κ_1. That is why the denominator on the right-hand side of Eq. S**2** is the sum of the two rate constants. Between the two pathways, the probability for the latter is

$$\eta = \frac{\kappa_1}{\kappa_{3-} + \kappa_1}. \qquad [S3]$$

Starting from the bulk solution, the probability of the protein reaching P_s after n rounds of cycling between P_b and P_{ns} is (1)

$$p_n = (1-\eta)^{n-1}\eta. \qquad [S4]$$

The average rounds of cycling between P_b and P_{ns} that the protein goes through before being captured by the specific site is therefore

$$\bar{n} = \sum_{n=1}^{\infty} nP_n = \frac{1}{\eta} \qquad [S5a]$$

$$= \frac{\kappa_{3-} + \kappa_1}{\kappa_1}. \qquad [S5b]$$

The average total search time is

$$t_s = \bar{n}(\bar{t}_{3d} + \bar{t}_{1d}) \qquad [S\mathbf{6a}]$$

$$= \frac{\kappa_{3-} + \kappa_1}{\kappa_1}\left(\frac{1}{\kappa_3} + \frac{1}{\kappa_{3-} + \kappa_1}\right) \qquad [S\mathbf{6b}]$$

$$= \frac{1}{\kappa_3} + \left(1 + \frac{\kappa_{3-}}{\kappa_3}\right)\frac{1}{\kappa_1}. \qquad [S\mathbf{6}c]$$

Eq. S**6c** can also be derived from solving the rate equations corresponding to unimolecular reaction scheme. Denoting the probability of the protein being in the P_b state at time t as $[P_b](t)$ and the analogous quantity for the P_{ns} state as $[P_{ns}](t)$, we have

$$\frac{d[P_b](t)}{dt} = -\kappa_3[P_b](t) + \kappa_{3-}[P_{ns}](t), \qquad [S\mathbf{7a}]$$

$$\frac{d[P_{ns}](t)}{dt} = \kappa_3[P_b](t) - (\kappa_{3-} + \kappa_1)[P_{ns}](t). \qquad [S\mathbf{7b}]$$

The sum of the probabilities that the protein is still in P_b and P_{ns} at time t is the survival probability, i.e., the probability that the protein has not been captured:

$$S(t) = [P_b](t) + [P_{ns}](t). \qquad [S\mathbf{8}]$$

The search time t_s is the average time that the protein takes to reach P_s for the first time, and is given by

$$t_s = \int_0^\infty dt S(t). \qquad [S\mathbf{9}]$$

Solving Eqs. S**7**, e.g., by Laplace transform, subject to the initial condition $[P_b](0) = 1$ and $[P_b](0) = 0$, and carrying out the integral of Eq. S**9**, we obtain Eq. S**6c**.

We now derive the bimolecular rate constant k_a from the two-state model. For the moment let the volume of the bulk solution around the DNA to be V; shortly we will take the limit $V \to \infty$. The ratio κ_3/κ_{3-} is the equilibrium constant between P_b and P_{ns}. Assuming that the potential of mean force $U(\mathbf{r})$ is nonzero only when the protein is in the P_{ns} state, we have

$$\frac{\kappa_3}{\kappa_{3-}} = \frac{\int_{\text{surface region}} dv e^{-\beta U(\mathbf{r})}}{\int_{\text{bulk solution}} dv} = \frac{\mathscr{K}_{ns}}{V}, \qquad [S\mathbf{10}]$$

where *dv* is a volume element, and

$$\mathscr{K}_{\mathrm{ns}} = \int_{\text{surface region}} dv e^{-\beta U(\mathbf{r})} \qquad [\mathbf{S11}]$$

is the equilibrium constant for forming the protein-DNA nonspecific complex:

$$\mathrm{P_b} + \mathrm{D} \underset{\kappa_{3-}}{\overset{k_3}{\rightleftarrows}} \mathrm{P_{ns}} \cdot \mathrm{D}$$

as $V \to \infty$. That is,

$$\mathscr{K}_{\mathrm{ns}} = \frac{k_3}{\kappa_{3-}}. \qquad [\mathbf{S12}]$$

Comparing Eqs. S**10** and S**12**, we obtain

$$\kappa_3 = k_3 \frac{1}{V}, \qquad [\mathbf{S13}]$$

in which $1/V$ is effectively the DNA concentration (or protein concentration, since there is one molecule of each species in the same volume). Using Eqs. S**10** and S**13** in Eq. S**6c**, we have

$$t_{\mathrm{s}} = \frac{V}{k_3} + \left(1 + \frac{V}{\mathscr{K}_{\mathrm{ns}}}\right)\frac{1}{\kappa_1}. \qquad [\mathbf{S14}]$$

In analogy to Eq. S**13** we may define the bimolecular rate constant k_{a} via

$$\frac{1}{t_{\mathrm{s}}} = k_{\mathrm{a}} \frac{1}{V}. \qquad [\mathbf{S15}]$$

Hence

$$\frac{V}{k_{\mathrm{a}}} = \frac{V}{k_3} + \left(1 + \frac{V}{\mathscr{K}_{\mathrm{ns}}}\right)\frac{1}{\kappa_1}. \qquad [\mathbf{S16a}]$$

Taking the $V \to \infty$ limit, we find

$$\frac{1}{k_{\mathrm{a}}} = \frac{1}{k_3} + \frac{1}{\mathscr{K}_{\mathrm{ns}}\kappa_1}. \qquad [\mathbf{S16b}]$$

Using Eqs. S**3** and S**12**, we can rewrite the last result as

$$k_{\mathrm{a}} = k_3 \eta. \qquad [\mathbf{S17}]$$

Reduction Factor α for a Finite-Length DNA. For the system of Fig. 1*A*, the reduction factor α is determined by the following implicit relations (2):

$$\alpha = 1 - \int_0^\infty d\xi \frac{g(\xi)\sin\xi_1/\xi_1}{\xi K_1(\xi)/K_0(\xi) + \xi^2\Gamma}, \qquad [\mathbf{S18a}]$$

$$g(\xi) = (2\Gamma/\pi)\xi\left[q(L)\sin(\xi L/R) + (\xi/R)\int_L^\infty dx\cos(\xi x/R)q(x)\right], \qquad [\mathbf{S18b}]$$

$$q(x) = \alpha q^\infty(x) + \int_0^\infty d\xi \frac{g(\xi)\cos(\xi x/R)}{\xi K_1(\xi)/K_0(\xi) + \xi^2\Gamma}, \qquad [\mathbf{S18c}]$$

where $\xi_1 = \xi h/R$ and

$$q^\infty(x) = \frac{k_a^\infty}{2\pi^2 DR}\int_0^\infty d\xi \frac{\cos(\xi x/R)\sin\xi_1/\xi_1}{\xi K_1(\xi)/K_0(\xi) + \xi^2\Gamma}. \qquad [\mathbf{S19}]$$

Derivation of Eq. 11a. The probability $\eta(\mathbf{r})$ that the protein started at $\mathbf{r}$ will reach the specific site instead of escape to infinity satisfies the backward Smoluchowski equation

$$e^{\beta U(\mathbf{r})}\nabla\cdot\mathscr{D}(\mathbf{r})\cdot e^{-\beta U(\mathbf{r})}\nabla\eta(\mathbf{r}) = 0 \qquad [\mathbf{S20}]$$

with the outer boundary condition

$$\eta(\mathbf{r}) = 0, \qquad r = \infty. \qquad [\mathbf{S21}]$$

By comparing the equations and boundary conditions satisfied by $\eta(\mathbf{r})$ and the pair distribution function $P(\mathbf{r})$, it can be verified that

$$\eta(\mathbf{r}) = 1 - e^{-\beta U(\mathbf{r})}P(\mathbf{r}). \qquad [\mathbf{S22}]$$

In terms of $\eta(\mathbf{r})$, the rate constant k_a can be written as

$$k_a = -\oint_{r=\infty} d\mathbf{s}\cdot\mathscr{D}(\mathbf{r})\cdot\nabla\eta(\mathbf{r}). \qquad [\mathbf{S23}]$$

Let $P_3(\mathbf{r})$ be the pair distribution function for the problem in which the whole DNA exterior surface is absorbing. It satisfies the Smoluchowski equation

$$\nabla\cdot\mathbf{J}_3(\mathbf{r}) = 0, \qquad [\mathbf{S24a}]$$

$$\mathbf{J}_3(\mathbf{r}) = -\mathscr{D}(\mathbf{r})\cdot e^{-\beta U(\mathbf{r})}\nabla e^{\beta U(\mathbf{r})}P_3(\mathbf{r}) \qquad [\mathbf{S24b}]$$

with the inner boundary condition

$$P_3(\mathbf{r}) = 0\,, \quad \mathbf{r} \in \text{DNA exterior surface}, \qquad [\mathbf{S25a}]$$

and the outer boundary condition

$$P_3(\mathbf{r}) = 1\,, \quad r = \infty. \qquad [\mathbf{S25b}]$$

Using Eqs. S**20** and S**24**, it can be easily shown that

$$\nabla \cdot \left[\mathscr{D}(\mathbf{r}) \cdot P_3(\mathbf{r}) \nabla \eta(\mathbf{r}) + \mathbf{J}_3(\mathbf{r}) \eta(\mathbf{r}) \right] = 0. \qquad [\mathbf{S26}]$$

Consequently

$$-\oint_{r=\infty} d\mathbf{s}\mathbf{n} \cdot \left[\mathscr{D}(\mathbf{r}) \cdot P_3(\mathbf{r}) \nabla \eta(\mathbf{r}) + \mathbf{J}_3(\mathbf{r}) \eta(\mathbf{r}) \right]$$

$$= - \oint_{\text{any DNA-enclosing surface}} d\mathbf{s}\mathbf{n} \cdot \left[\mathscr{D}(\mathbf{r}) \cdot P_3(\mathbf{r}) \nabla \eta(\mathbf{r}) + \mathbf{J}_3(\mathbf{r}) \eta(\mathbf{r}) \right].$$

Using the boundary conditions of Eqs. S**21** and S**25b**, the left-hand side becomes

$$-\oint_{r=\infty} d\mathbf{s}\mathbf{n} \cdot \mathscr{D}(\mathbf{r}) \cdot \nabla \eta(\mathbf{r}),$$

which according to Eq. S**23** is just k_a. For the right-hand side, we specialize the DNA-enclosing surface to the DNA exterior surface and use the boundary condition of Eq. S**25a**. We thus derive

$$k_a = - \oint_{\text{DNA exterior surface}} d\mathbf{s}\mathbf{n} \cdot \mathbf{J}_3(\mathbf{r}) \eta(\mathbf{r}), \qquad [\mathbf{S27}]$$

which is Eq. **11a** of the main text. A special case of Eq. S**27** has been derived previously (3), where $P_3(\mathbf{r})$ satisfies the absorbing boundary condition on a spherical surface, $U(\mathbf{r})$ is spherically symmetric outside this surface, and the diffusion tensor is reduced to a scalar.

Derivation of $\bar{\eta}$ for an *x*-Dependent Surface Potential. Here we consider a protein, with a fixed conformation, that experiences an *x*-dependent surface potential. Specifically, we consider a surface potential with a step-function form:

$$U(x) = U_0 \quad \text{when } |x| < h;$$

$$= U \quad \text{when } h < |x| < L. \qquad [\mathbf{S28}]$$

Correspondingly the values of $\kappa_{3-}(x)$ in the two intervals will be denoted as κ_{03-} and κ_{3-}, respectively. Because of the symmetry of the problem about $x = 0$, we now restrict to the domain $0 < x < L$. The capture probability, $\eta(x)$, for the nonspecifically bound protein starting at position x is governed by

$$D\frac{d^2\eta(x)}{dx^2} - \kappa_{03-}\eta(x) = 0, \quad \text{when } 0 < x < h; \qquad [\mathbf{S29a}]$$

$$D\frac{d^2\eta(x)}{dx^2} - \kappa_{3-}\eta(x) = 0, \quad \text{when } h < x < L. \qquad [\mathbf{S29b}]$$

The solution has the form

$$\eta(x) = A_1 e^{x/m_0} + A_2 e^{-x/m_0}, \qquad \text{when } 0 < x < h; \qquad [\mathbf{S30a}]$$

$$= B_1 e^{x/m} + B_2 e^{-x/m}, \qquad \text{when } h < x < L. \qquad [\mathbf{S30b}]$$

where $m_0 = (D/\kappa_{03-})^{1/2}$, $m = (D/\kappa_{3-})^{1/2}$, and A_1, A_2, B_1, and B_2 are coefficients to be determined.

The boundary conditions are

$$\eta(0) = 1, \qquad [\mathbf{S31a}]$$

$$\left.\frac{d\eta(x)}{dx}\right|_{x=L} = 0. \qquad [\mathbf{S31b}]$$

In addition, $\eta(x)$ and its derivative satisfy the following continuity conditions at $x = h$:

$$\eta(h^-) = \eta(h^+), \qquad [\mathbf{S32a}]$$

$$e^{-\beta U(x)}\left.\frac{d\eta(x)}{dx}\right|_{x=h^-} = e^{-\beta U(x)}\left.\frac{d\eta(x)}{dx}\right|_{x=h^+}. \qquad [\mathbf{S32b}]$$

Using these conditions, we find

$$A_1 = \frac{e^{-2h/m_0}}{1-e^{-2h/m_0}}\frac{e^{-\beta\Delta U}m/m_0 - \tanh(L_1/m)}{G}, \qquad [\mathbf{S33a}]$$

$$A_2 = \frac{1}{1-e^{-2h/m_0}}\frac{e^{-\beta\Delta U}m/m_0 + \tanh(L_1/m)}{G}, \qquad [\mathbf{S33b}]$$

$$B_1 = \frac{e^{-(2L-h)/m}}{(1+e^{-2L_1/m})\sinh(h/m_0)}\frac{e^{-\beta\Delta U}m/m_0}{G}, \qquad [\mathbf{S33c}]$$

$$B_2 = \frac{e^{h/m}}{(1+e^{-2L_1/m})\sinh(h/m_0)} \frac{e^{-\beta\Delta U} m_0/m_0}{G}, \qquad [S33d]$$

where $\Delta U = U_0 - U$, $L_1 = L - h$, and

$$G = e^{-\beta\Delta U}(m/m_0)\coth(h/m_0) + \tanh(L_1/m). \qquad [S34]$$

The average capture probability of the nonspecifically bound protein is

$$\begin{aligned}\overline{\eta} &= (1/L)\int_0^L dx\eta(x) \\ &= \frac{e^{-\beta\Delta U}(m/m_0)[1+(m/m_0)\tanh(L_1/m)/\sinh(h/m_0)]}{GL/m_0} \\ &+ \frac{\tanh(h/2m_0)\tanh(L_1/m)}{GL/m_0}. \qquad [S35]\end{aligned}$$

This reduces to Eq. **15** of the main text when $\Delta U = 0$ and $\kappa_{03-} = \kappa_{3-}$.

If the local energy well at the specific site is infinitely deep (i.e., $\beta\Delta U \to -\infty$), which is the case for the system of Fig. 1*A*, then $m_0 \to \infty$ and Eq. S**35** becomes

$$\overline{\eta} = \frac{h}{L} + \frac{\tanh(L_1/m)}{L/m}. \qquad [S36]$$

In this limiting situation, the whole interval $|x| < h$ is effectively absorbing, so there $\eta(x) = 1$; in $h < x < L$ $\eta(x)$ can be determined by solving Eq. S**29b** with the boundary condition $\eta(h) = 1$. Averaging over the two intervals of $|x|$ results in Eq. S**36**. The solid curves in Figs. 2*A* and 2*B* display k_a obtained by combining Eq. S**36** with Eqs. **6** and **11b** of the main text.

Capture Probability for a Protein with Conformational Switch. When the nonspecifically bound protein can switch between two conformations, the capture probability, $\eta_g(x)$, starting at position x and conformation g = a or i is governed by Eqs. **18** of the main text. We first note that the detailed balance condition given by Eq. **17** of the main text constraints the relation between $\omega_g(x)$ and $\kappa_{g-}(x)$. According to Eq. **19** of the main text,

$$\frac{\kappa_i(x)}{\kappa_a(x)} = e^{-\beta[U_a(x)-U_i(x)]}. \qquad [S37a]$$

Using this result in Eq. **17** of the main text, we get

$$\frac{\omega_{\mathrm{i}}(x)/\omega_{\mathrm{a}}(x)}{\kappa_{\mathrm{i}}(x)/\kappa_{\mathrm{a}}(x)} = \frac{p_{3\mathrm{a}}}{p_{3\mathrm{i}}}. \qquad [\mathbf{S37b}]$$

Therefore the ratio of $\omega_{\mathrm{i}}(x)/\omega_{\mathrm{a}}(x)$ to $\kappa_{\mathrm{i}}(x)/\kappa_{\mathrm{a}}(x)$ must be x-independent.

If $U_{\mathrm{g}}(x)$, $\omega_{\mathrm{g}}(x)$, and $\kappa_{\mathrm{g}-}(x)$ are all x-independent, then Eqs. **18** become

$$D_{\mathrm{a}}\frac{d^2\eta_{\mathrm{a}}(x)}{dx^2} - \omega_{\mathrm{a}}\eta_{\mathrm{a}}(x) + \omega_{\mathrm{a}}\eta_{\mathrm{i}}(x) - \kappa_{\mathrm{a}-}\eta_{\mathrm{a}}(x) = 0, \qquad [\mathbf{S38a}]$$

$$D_{\mathrm{i}}\frac{d^2\eta_{\mathrm{i}}(x)}{dx^2} + \omega_{\mathrm{i}}\eta_{\mathrm{a}}(x) - \omega_{\mathrm{i}}\eta_{\mathrm{i}}(x) - \kappa_{\mathrm{i}-}\eta_{\mathrm{i}}(x) = 0. \qquad [\mathbf{S38b}]$$

The solution has the form

$$\eta_{\mathrm{a}}(x) = A_1 e^{\lambda_+^{1/2}x} + A_2 e^{-\lambda_+^{1/2}x} + B_1 e^{\lambda_-^{1/2}x} + B_2 e^{-\lambda_-^{1/2}x}, \qquad [\mathbf{S39a}]$$

$$\eta_{\mathrm{i}}(x) = -F_+(A_1 e^{\lambda_+^{1/2}x} + A_2 e^{-\lambda_+^{1/2}x}) - F_-(B_1 e^{\lambda_-^{1/2}x} + B_2 e^{-\lambda_-^{1/2}x}), \qquad [\mathbf{S39b}]$$

where A_1, A_2, B_1, and B_2 are coefficients to be determined, and

$$\lambda_\pm = \frac{D_{\mathrm{a}}(\omega_{\mathrm{i}}+\kappa_{\mathrm{i}-}) + D_{\mathrm{i}}(\omega_{\mathrm{a}}+\kappa_{\mathrm{a}-}) \pm \{[D_{\mathrm{a}}(\omega_{\mathrm{i}}+\kappa_{\mathrm{i}-}) - D_{\mathrm{i}}(\omega_{\mathrm{a}}+\kappa_{\mathrm{a}-})]^2 + 4D_{\mathrm{a}}D_{\mathrm{i}}\omega_{\mathrm{a}}\omega_{\mathrm{i}}\}^{1/2}}{2D_{\mathrm{a}}D_{\mathrm{i}}}, \qquad [\mathbf{S40a}]$$

$$F_\pm = \frac{D_{\mathrm{a}}\lambda_\pm - (\omega_{\mathrm{a}}+\kappa_{\mathrm{a}-})}{\omega_{\mathrm{a}}}. \qquad [\mathbf{S40b}]$$

The boundary conditions of $\eta_{\mathrm{g}}(x)$ are

$$\eta_{\mathrm{a}}(0) = 1; \quad \left.\frac{d\eta_{\mathrm{i}}(x)}{dx}\right|_{x=0} = 0; \qquad [\mathbf{S41a}]$$

$$\left.\frac{d\eta_{\mathrm{a}}(x)}{dx}\right|_{x=L} = 0; \quad \left.\frac{d\eta_{\mathrm{i}}(x)}{dx}\right|_{x=L} = 0. \qquad [\mathbf{S41b}]$$

Using these we find

$$A_1 = \frac{e^{-2\lambda_+^{1/2}L}}{1-e^{-2\lambda_+^{1/2}L}} \frac{1}{\coth(\lambda_+^{1/2}L) + F\coth(\lambda_-^{1/2}L)}, \qquad [\mathbf{S42a}]$$

$$A_2 = \frac{1}{1-e^{-2\lambda_+^{1/2}L}} \frac{1}{\coth(\lambda_+^{1/2}L) + F\coth(\lambda_-^{1/2}L)}, \qquad [\mathbf{S42b}]$$

$$B_1 = F\frac{e^{-2\lambda_-^{1/2}L}}{1-e^{-2\lambda_-^{1/2}L}} \frac{1}{\coth(\lambda_+^{1/2}L) + F\coth(\lambda_-^{1/2}L)}, \qquad [\mathbf{S42c}]$$

$$B_2 = F\frac{1}{1-e^{-2\lambda_-^{1/2}L}}\frac{1}{\coth(\lambda_+^{1/2}L)+F\coth(\lambda_-^{1/2}L)}, \qquad [S\mathbf{42d}]$$

where

$$F = -\frac{\lambda_+^{1/2}F_+}{\lambda_-^{1/2}F_-}. \qquad [S\mathbf{43}]$$

The average capture probability of the nonspecifically bound protein is

$$\begin{aligned}\overline{\eta} &= (p_{3a}/L)\int_0^L dx\eta_a(x) + (p_{3i}/L)\int_0^L dx\eta_i(x) \\ &= \frac{[p_{3a}+p_{3i}\omega_i/(\omega_i+\kappa_{i-})](1/\lambda_+^{1/2}L + F/\lambda_-^{1/2}L)}{\coth(\lambda_+^{1/2}L)+F\coth(\lambda_-^{1/2}L)}. \qquad [S\mathbf{44}]\end{aligned}$$

The dashed curve in Fig. 3 displays k_a obtained by combining Eq. S**44** with Eqs. **6** and **11b** of the main text.

We now consider two opposite limits of Eq. S**44**. When $\omega_g \to 0$, we find

$$\lambda_+ \to \frac{\kappa_{a-}}{D_a} \equiv m_a^{-2},$$

$$\lambda_- \to \frac{\kappa_{i-}}{D_i} \equiv m_i^{-2},$$

$$F \to 0,$$

and consequently the slow conformational-transition limit of $\overline{\eta}$ is

$$\overline{\eta}_s \equiv \lim_{\omega_g\to 0}\overline{\eta} = p_{3a}\frac{\tanh(L/m_a)}{L/m_a}. \qquad [S\mathbf{45}]$$

In the limit $\omega_g \to \infty$ we find

$$\lambda_+ \to \infty,$$

$$\lambda_- \to \frac{\omega_i\kappa_{a-}+\omega_a\kappa_{i-}}{\omega_i D_a+\omega_a D_i} \equiv \overline{m}^{-2},$$

$$F \to \infty,$$

and therefore the fast conformational-transition limit of $\overline{\eta}$ is

$$\overline{\eta}_f \equiv \lim_{\omega_g\to\infty}\overline{\eta} = \frac{\tanh(L/\overline{m})}{L/\overline{m}}. \qquad [S\mathbf{46}]$$

For the potentials shown in Fig. 1*B*, U_i is x-independent but U_a has the step-function form of Eq. S**28**. The U_a well spans the specific site. We will add a "0" in the subscript to denote parameters within the specific site. For example, κ_{0g-} denotes the rate constant for the protein in conformation g to escape to the bulk solution from the specific site whereas κ_{g-} denotes the counterpart when the protein is elsewhere on the DNA surface. Due to symmetry we will only consider $0 < x < L$.

We are particularly interested in the case where the U_a well at the specific site is infinitely deep. Then $\kappa_{0a-} \to 0$ (see Eq. S**37a**) and $\omega_{0a} \to 0$ (see Eq. **17** of the main text). In addition, the continuity condition on $d\eta_a/dx$ at $x = h$ (see Eq. S**32b**) means that $d\eta_a/dx = 0$ at $x = h^-$. Solving Eq. S**38a** under these conditions and $\eta_a(0) = 1$ leads to

$$\eta_a(x) = 1 \quad \text{for } 0 < x < h. \qquad [\text{S}\mathbf{47a}]$$

The interactions that produce the U_a well at the specific site will also affect the inactive-to-active transition rate ω_{0i}. Here we consider the extreme case $\omega_{0i} \to \infty$. Then we must have (see Eq. S**32a**)

$$\eta_i(x) = \eta_a(x) = 1 \quad \text{for } 0 < x < h. \qquad [\text{S}\mathbf{47b}]$$

We can then solve $\eta_g(x)$ in $h < x < L$ using the boundary conditions $\eta_g(h) = 1$ as demanded by Eqs. S**47**. The solution has the form of Eqs. S**39**, but with the coefficients now given by

$$A_1 = \frac{e^{-\lambda_+^{1/2}(2L-h)}}{1+e^{-2\lambda_+^{1/2}L_1}} \frac{1+F_-}{F_- - F_+}, \qquad [\text{S}\mathbf{48a}]$$

$$A_2 = \frac{e^{\lambda_+^{1/2}h}}{1+e^{-2\lambda_+^{1/2}L_1}} \frac{1+F_-}{F_- - F_+}, \qquad [\text{S}\mathbf{48b}]$$

$$B_1 = -\frac{e^{-\lambda_-^{1/2}(2L-h)}}{1+e^{-2\lambda_-^{1/2}L_1}} \frac{1+F_+}{F_- - F_+}, \qquad [\text{S}\mathbf{48c}]$$

$$B_2 = -\frac{e^{\lambda_-^{1/2}h}}{1+e^{-2\lambda_-^{1/2}L_1}} \frac{1+F_+}{F_- - F_+}. \qquad [\text{S}\mathbf{48d}]$$

Averaging over initial position and initial conformation leads to

$$\bar{\eta} = \frac{h}{L} + (p_{3a} - p_{3i}F_+)\frac{\tanh(\lambda_+^{1/2}L_1)}{\lambda_+^{1/2}L}\frac{1+F_-}{F_- - F_+} + (-p_{3a} + p_{3i}F_-)\frac{\tanh(\lambda_-^{1/2}L_1)}{\lambda_-^{1/2}L}\frac{1+F_+}{F_- - F_+}. \quad [S\mathbf{49}]$$

The solid curve in Fig. 3 displays k_a obtained by combining Eq. S**49** with Eqs. **6** and **11b** of the main text.

The slow transition limit of $\bar{\eta}$ is now

$$\bar{\eta}_s = \frac{h}{L} + p_{3a}\frac{\tanh(L_1/m_a)}{L/m_a} + p_{3i}\frac{\tanh(L_1/m_i)}{L/m_i}, \quad [S\mathbf{50a}]$$

where $m_g = (D_g/\kappa_{g-})^{1/2}$. The fast transition limit is

$$\bar{\eta}_f = \frac{h}{L} + \frac{\tanh(L_1/\bar{m})}{L/\bar{m}}. \quad [S\mathbf{50b}]$$

MFPT of a Nonspecifically Bound Protein. The mean-first-passage-time (MFPT), $\tau(x)$, for a protein with a fixed conformation starting at position x is governed by (4)

$$e^{\beta U(x)}\frac{d}{dx}De^{-\beta U(x)}\frac{d\tau(x)}{dx} = -1. \quad [S\mathbf{51}]$$

Applying the absorbing boundary condition at $x = 0$ and the reflecting boundary condition at $x = L$, one finds

$$D\tau(x) = \int_0^x dx_1 e^{\beta U(x)}\int_{x_1}^L dx_2 e^{-\beta U(x)}. \quad [S\mathbf{52a}]$$

For an x-independent potential, the result is

$$D\tau(x) = [L^2 - (L-x)^2]/2. \quad [S\mathbf{52b}]$$

For the step-function potential of Eq. S**28**, we find

$$D\tau(x) = [h^2 - (h-x)^2]/2 + e^{\beta\Delta U}L_1 x \qquad \text{when } 0 < x < h;$$

$$= h^2/2 + e^{\beta\Delta U}L_1 h + [L_1^{\,2} - (L-x)^2]/2 \quad \text{when } h < x < L. \quad [S\mathbf{52c}]$$

We are particularly interested in the case shown in Fig. 1*A*, where the energy well at the specific site is infinitely deep (i.e., $\beta\Delta U \to -\infty$).

What enters the two-state model (see Eq. **10a** of the main text) is the MFPT averaged over initial position. For an x-independent potential, the result is $\bar{\tau} = L^2/3D$ (4). However,

when the potential is x-dependent, there is a question of whether the average should be weighted by the Boltzmann factor $\exp[-\beta U(x)]$. The usual use of MFPT has the Boltzmann weight (4). The Boltzmann-weight average for the step-function potential is $h^2/3D$ when $\beta\Delta U \to -\infty$. That would predict a strong dependence of k_a on h, which contradicts the weak h dependence of the exact result given by Eq. **4** of the main text (see Fig. S2). However, the unweighted average,

$$\overline{\tau} = \frac{h^3 + 3h^2 L_1/2 + L_1^3}{3DL}, \qquad [S\mathbf{53}]$$

seems appropriate for use in the two-state model. Recall that the average of $\eta(x)$ over initial position is also unweighted. When $h \to 0$, this $\overline{\tau}$ properly reduces to $L^2/3D$, the result obtained when the potential is constant. The dashed curves in Figs. 2*A* and 2*B* display k_a obtained by combining Eq. S**53** with Eqs. **6**, **9**, and **10a** of the main text.

Next we consider the case where the nonspecifically bound protein can switch between two conformations. The MFPT $\tau_g(x)$ for the protein starting at position x and conformation g = a or i is governed by

$$e^{\beta U_a(x)} \frac{d}{dx} D_a e^{-\beta U_a(x)} \frac{d\tau_a(x)}{dx} - \omega_a(x)\tau_a(x) + \omega_a(x)\tau_i(x) = -1, \qquad [S\mathbf{54a}]$$

$$e^{\beta U_i(x)} \frac{d}{dx} D_i e^{-\beta U_i(x)} \frac{d\tau_i(x)}{dx} + \omega_i(x)\tau_a(x) - \omega_i(x)\tau_i(x) = -1. \qquad [S\mathbf{54b}]$$

We focus on the case where $U_g(x)$ and $\omega_g(x)$ are x-independent. Hu et al. (5) has studied a similar problem. The two equations for $\tau_g(x)$ can be uncoupled by forming appropriate linear combinations. We find

$$\frac{d^2[p_a D_a \tau_a(x) + p_i D_i \tau_i(x)]}{dx^2} = -1, \qquad [S\mathbf{55a}]$$

$$D_a D_i \frac{d^2[\tau_a(x) - \tau_i(x)]}{dx^2} - \omega\overline{D}[\tau_a(x) - \tau_i(x)] = -\Delta D, \qquad [S\mathbf{55b}]$$

where

$$\omega = \omega_a + \omega_i, \qquad [S\mathbf{56a}]$$

$$p_a = \omega_i / \omega, \quad p_i = \omega_a / \omega, \qquad [S\mathbf{56b}]$$

$$\Delta D = D_i - D_a, \qquad [S\mathbf{56c}]$$

$$\bar{D} = p_a D_a + p_i D_i. \qquad [S\mathbf{56d}]$$

The solution of Eq. S**55a** has the form

$$p_a D_a \tau_a(x) + p_i D_i \tau_i(x) = A_1 + A_2 x - x^2/2, \qquad [S\mathbf{57a}]$$

where A_1 and A_2 are coefficients to be determined. The solution of Eq. S**55b** has the form

$$\tau_a(x) - \tau_i(x) = \Delta D / \omega\bar{D} + B_1 e^{(\omega\bar{D}/D_a D_i)^{1/2} x} + B_2 e^{-(\omega\bar{D}/D_a D_i)^{1/2} x}, \qquad [S\mathbf{57b}]$$

where B_1 and B_2 are to be determined.

To determine A_1, A_2, B_1, and B_2, we apply the boundary conditions at $x = 0$ and $x = L$. These are

$$\tau_a(0) = 0; \quad \left.\frac{d\tau_i(x)}{dx}\right|_{x=0} = 0; \qquad [S\mathbf{58a}]$$

$$\left.\frac{d\tau_a(x)}{dx}\right|_{x=L} = 0; \quad \left.\frac{d\tau_i(x)}{dx}\right|_{x=L} = 0. \qquad [S\mathbf{58b}]$$

The results are

$$A_1 = -\frac{p_i D_i \Delta D}{\omega\bar{D}} + \frac{p_i D_i L \coth[(\omega\bar{D}/D_a D_i)^{1/2} L]}{p_a D_a (\omega\bar{D}/D_a D_i)^{1/2}}, \qquad [S\mathbf{59a}]$$

$$A_2 = L, \qquad [S\mathbf{59b}]$$

$$B_1 = -\frac{L e^{-2(\omega\bar{D}/D_a D_i)^{1/2} L}}{p_a D_a (\omega\bar{D}/D_a D_i)^{1/2} [1 - e^{-2(\omega\bar{D}/D_a D_i)^{1/2} L}]}, \qquad [S\mathbf{59c}]$$

$$B_2 = -\frac{L}{p_a D_a (\omega\bar{D}/D_a D_i)^{1/2} [1 - e^{-2(\omega\bar{D}/D_a D_i)^{1/2} L}]}. \qquad [S\mathbf{59d}]$$

We then want to average $\tau_g(x)$ over initial position and initial conformation. There is uncertainty regarding the weighting factor for averaging over initial conformation. A natural choice is to use p_g as the weighting factor, resulting in

$$\bar{\tau} = (p_a / L)\int_0^L dx \tau_a(x) + (p_i / L)\int_0^L dx \tau_i(x)$$

$$= \frac{L^2}{3\bar{D}} + \frac{p_i D_i L}{p_a D_a \bar{D}(\omega\bar{D}/D_a D_i)^{1/2}} \left\{ \frac{1}{\tanh[(\omega\bar{D}/D_a D_i)^{1/2} L]} - \frac{1}{(\omega\bar{D}/D_a D_i)^{1/2} L} \right\} + \frac{p_i}{\omega_i}. \qquad [\mathbf{60}]$$

It has the following limiting values:

$$\bar{\tau} \to \frac{L^2}{3\bar{D}} \quad \text{as } \omega_g \to \infty; \qquad [\text{S}\mathbf{61a}]$$

$$\bar{\tau} \to \frac{L^2}{3p_a D_a} + \frac{p_i}{\omega_i} \quad \text{as } \omega_g \to 0. \qquad [\text{S}\mathbf{61b}]$$

Note that $\bar{\tau}$ goes to infinity in the latter limit, since when $\omega_i = 0$ the protein started in the inactive conformation stays in that conformation and hence cannot be absorbed. The calculation of $\bar{\eta}$ suggests a different weighting factor, i.e., p_{3g} (see Eq. S**46**), although p_{3g} has so far not been introduced to an MFPT problem. Regardless of the weighting factor, $\bar{\tau}$ would go to infinity in the slow transition limit.

It is unclear how this $\bar{\tau}$ can be used to predict k_a. In the two-state model for the protein with a fixed conformation, $\bar{\tau}$ along with the escape rate κ_{3-} is used in Eq. **10a** of the main text to obtain the average capture probability. However, in the present case where the protein switches between two conformations, there are two distinct escape rates, κ_{a-} and κ_{i-}. Naively we may use a conformationally averaged escape rate, $\bar{\kappa}_{3-} = p_a\kappa_a + p_i\kappa_i$, in place of κ_{3-}. However, other than in the fast transition limit, it is unclear whether this use can be justified. In any event, the infinite value of $\bar{\tau}$ in the slow transition limit would lead to a zero k_a, which cannot be unjustified. So the use of a two-state model to predict k_a for any DNA-binding protein that switches between two conformations is fraught with problems.

Rate Constant k_3 When Nonspecific Binding Is Restricted to a Narrow Angular Range. Here we consider the case where nonspecific binding is restricted to a narrow angular range. The position **r** of the protein relative to the DNA can be specified by the coordinate x along, the distance ρ to, and the rotation angle ϕ around the DNA axis. The steady-state Smoluchowski equation for the pair distribution function $P(\mathbf{r})$ now takes the form

$$D_3\left(\frac{\partial^2}{\partial x^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2}\right)P(\mathbf{r})=0. \qquad [\mathbf{S62}]$$

The nonspecific binding site is specified by $|x| < L$, $|\phi| < \phi_0$, and $\rho = R$, where $2\phi_0$ is the angular range of the binding site. For convenience we will denote a point on this binding site as $\mathbf{r} \in \Xi$. The inner boundary condition is

$$P(\mathbf{r})=0, \qquad \text{when } \mathbf{r}\in\Xi; \qquad [\mathbf{S63a}]$$

$$D_3\frac{\partial P(\mathbf{r})}{\partial\rho}=0, \text{ elsewhere on } \rho=R. \qquad [\mathbf{S63b}]$$

Even though the DNA has a finite length $2L$, following Berg and Ehrenberg (6) we have treated the surface $\rho = R$ beyond the DNA as reflecting.

It is clear that $P(\mathbf{r})$ is an even function of both x and ϕ. We can express $P(\mathbf{r})$ in the following expansion:

$$P(\mathbf{r})=1-\int_0^\infty d\zeta\sum_{j=0}^\infty A_j(\zeta)\cos(\zeta z)\cos(j\phi)K_j(\zeta\rho), \qquad [\mathbf{S64}]$$

where $K_j(x)$ are modified Bessel functions. The coefficients $A_j(\zeta)$ are determined by the boundary condition of Eqs. S**63**. This mixed-type of boundary condition presents significant difficulty. We thus replace Eq. S**63a** by the constant-flux approximation (7),

$$D_3\frac{\partial P(\mathbf{r})}{\partial\rho}=Q, \text{ when } \mathbf{r}\in\Xi, \qquad [\mathbf{S63c}]$$

and determine the constant Q by requiring that the absorbing boundary condition of Eq. S**63a** is satisfied on the average:

$$< P(\mathbf{r}) > \equiv \int_{\mathbf{r}\in\Xi} dsP(\mathbf{r}) / \int_{\mathbf{r}\in\Xi} ds = 0. \qquad [\mathbf{S63d}]$$

Now the flux at $\rho = R$ has value Q when $\mathbf{r} \in \Xi$ and 0 otherwise. This function has the following expansion:

$$D_3\left.\frac{\partial P(\mathbf{r})}{\partial\rho}\right|_{\rho=R}=\frac{4Q}{\pi^2}\left[\frac{\phi_0}{2}+\sum_{j=1}^\infty\frac{\sin(j\phi_0)\cos(j\phi)}{j}\right]\int_0^\infty d\zeta\frac{\sin(\zeta L)\cos(\zeta z)}{\zeta}. \qquad [\mathbf{S65}]$$

Using Eq. S**64** to evaluate the flux and comparing against Eq. S**65**, we find

$$-D_3 A_j(\zeta)\zeta K_j'(\zeta R) = \frac{4Q}{\pi^2}\frac{\phi_0}{2}\frac{\sin(\zeta L)}{\zeta} \qquad \text{when } j = 0; \qquad [\mathbf{S66a}]$$

$$= \frac{4Q}{\pi^2}\frac{\sin(j\phi_0)}{j}\frac{\sin(\zeta L)}{\zeta} \quad \text{when } j > 0, \qquad [\mathbf{S66b}]$$

where we have used a prime to denote derivative. Using Eq. S**63d**, we obtain

$$Q = \frac{\pi^2 D_3 / 2\phi_0 L}{\int_0^\infty d\xi \frac{\sin^2 \xi_2}{{\xi_2}^2}\left[\frac{1}{\xi K_1(\xi)/K_0(\xi)} + \sum_{j=1}^{\infty}\frac{2\sin^2(j\phi_0)/(j\phi_0)^2}{\xi K_{j-1}(\xi)/K_j(\xi) + j}\right]}. \qquad [\mathbf{S67}]$$

where $\xi = \zeta R$ and $\xi_2 = \xi L/R$. Finally the rate constant is

$$k_3 = \int_{\mathbf{r}\in\Xi} da D_3 \frac{\partial P(\mathbf{r})}{\partial \rho} = 4\phi_0 RLQ$$

$$= \frac{2\pi^2 D_3 R}{\int_0^\infty d\xi \frac{\sin^2 \xi_2}{{\xi_2}^2}\left[\frac{1}{\xi K_1(\xi)/K_0(\xi)} + \sum_{j=1}^{\infty}\frac{2\sin^2(j\phi_0)/(j\phi_0)^2}{\xi K_{j-1}(\xi)/K_j(\xi) + j}\right]}. \qquad [\mathbf{S68}]$$

If the nonspecific binding site extends to the full angular range (i.e., $\phi_0 = \pi$), then all the $j > 0$ terms disappear and Eq. S**68** reduces to Eq. **6** of the main text.

1. Kolomeisky AB (2011) Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys Chem Chem Phys* 13:2088-2095.
2. Zhou HX (2005) A model for the mediation of processivity of DNA-targeting proteins by nonspecific binding: dependence on DNA length and presence of obstacles. *Biophys J* 88:1608-1615.
3. Zhou HX (1990) On the Calculation of Diffusive Reaction-Rates Using Brownian Dynamics Simulations. *J Chem Phys* 92:3092-3095.
4. Szabo A, Schulten K, Schulten Z (1980) First passage time approach to diffusion controlled reactions. *J Chem Phys* 72:4350-4357.

5. Hu LH, Grosberg AY, Bruinsma R (2009) First passage time distribution for the 1D diffusion of particles with internal degrees of freedom. *J Phys A-Math Theor* 42:434011.
6. Berg OG, Ehrenberg M (1982) Association kinetics with coupled three- and one-dimensional diffusion. Chain-length dependence of the association rate of specific DNA sites. *Biophys Chem* 15:41-51.
7. Shoup D, Lipari G, Szabo A (1981) Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. *Biophys J* 36:697-714.

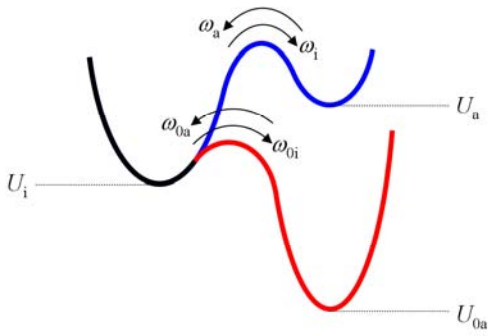

Fig. S1 Conformational transition rates ω_{0a} and ω_{0i} at the cognate site and ω_a and ω_i elsewhere on the DNA surface. The potential energies $U_a(x)$ and $U_i(x)$ at position x along the DNA, shown in Fig. 1*B*, correspond to two energy minima in conformational space. The energy surface in conformational space at two x positions is shown here. The red trace shows the energy surface when x is at the cognate site, and the blue trace shows the energy surface when x is at a noncognate site. The energy minimum corresponding to the inactive conformation has the same value U_i at the two x positions, but the minimum corresponding to the active conformation has a value U_{0a} at the cognate site and U_a at the noncognate site. U_{0a} is much more negative than U_i but U_a is much more positive than U_i. Consequently $\omega_{0a} \ll \omega_{0i}$ but $\omega_a \gg \omega_i$.

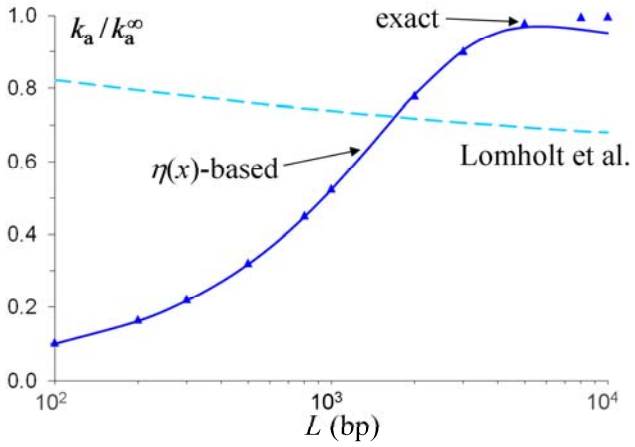

Fig. S2 Comparison of the predicted k_a of Lomholt et al. against the exact solution for the system of Fig. 1*A*. Lomholt et al.'s k_a is

$$k_a = \frac{2\pi^2 D_3 R}{\int_0^\infty d\xi \dfrac{1}{\dfrac{\xi K_1(\xi)/K_0(\xi)}{1+(4\pi D_3 L/k_3)\xi K_1(\xi)/K_0(\xi)} + \xi^2\Gamma}}.$$

The dashed curve display this result at $\Gamma = 10^4$. The exact solution is shown as symbols. We also show the k_a predicted by the $\eta(x)$-based approach as the solid curve. To match the problem solved by Lomholt et al., we took the $h \to 0$ limit in the exact and $\eta(x)$-based solutions. The latter solutions have very weak dependence on h and the $h \to 0$ limits shown here are virtually identical to those shown in Fig. 2*A* for $h = 3$ Å. Note that the exact k_a increases with increasing DNA length L whereas Lomholt et al.'s k_a shows the opposite trend.