Supplemental Information

Stopped-flow fluorescence kinetic study of protein sliding and intersegment transfer in the target DNA search process

Alexandre Esadze and Junji Iwahara

1. Derivation of Eqs. 4 and 5 in the main text

The rate equations for the process represented by Scheme 2 in the main text are as follows:

$$\frac{d}{dt}[PD_a] = -k_{f1}[PD_a][D_b] + k_{d1}[D_bPD_a] \quad [S1]$$

$$\frac{d}{dt}[PD_b] = -k_{f2}[PD_b][D_a] + k_{d2}[D_bPD_a] \quad [S2]$$

$$\frac{d}{dt}[D_a] = -k_{f2}[PD_b][D_a] + k_{d2}[D_bPD_a] \quad [S3]$$

$$\frac{d}{dt}[D_b] = -k_{f1}[PD_a][D_b] + k_{d1}[D_bPD_a] \quad [S4]$$

$$\frac{d}{dt}[D_bPD_a] = -(k_{d1}+k_{d2})[D_bPD_a] + k_{f1}[PD_a][D_b] + k_{f2}[PD_b][D_a] \quad [S5]$$

If the DNA-bridging intermediate D_bPD_a is a transient and low-population state, the differential of $[D_bPD_a]$ is far smaller than the differentials of other species, and thereby is negligible, which yields $d[D_bPD_a]/dt \approx 0$ (i.e., a steady-state approximation[1]). Thus, under this condition, $[D_bPD_a]$ can be approximated by:

$$[D_bPD_a] = k_{f1}(k_{d1}+k_{d2})^{-1}[PD_a][D_b] + k_{f2}(k_{d1}+k_{d2})^{-1}[PD_b][D_a] \quad [S6]$$

By plugging this into Eqs. S1-S4, the following equations are obtained:

$$\frac{d}{dt}[PD_a] = -k_{f1}k_{d2}(k_{d1}+k_{d2})^{-1}[PD_a][D_b] + k_{f2}k_{d1}(k_{d1}+k_{d2})^{-1}[PD_b][D_a] \quad [S7]$$

$$\frac{d}{dt}[PD_b] = -k_{f2}k_{d1}(k_{d1}+k_{d2})^{-1}[PD_b][D_a] + k_{f1}k_{d2}(k_{d1}+k_{d2})^{-1}[PD_a][D_b] \quad \text{[S8]}$$

$$\frac{d}{dt}[D_a] = -k_{f2}k_{d1}(k_{d1}+k_{d2})^{-1}[PD_b][D_a] + k_{f1}k_{d2}(k_{d1}+k_{d2})^{-1}[PD_a][D_b] \quad \text{[S9]}$$

$$\frac{d}{dt}[D_b] = -k_{f1}k_{d2}(k_{d1}+k_{d2})^{-1}[PD_a][D_b] + k_{f2}k_{d1}(k_{d1}+k_{d2})^{-1}[PD_b][D_a] \quad \text{[S10]}$$

On the other hand, the rate equations for the process represented by Scheme 3 in the main text are as follows:

$$\frac{d}{dt}[PD_a] = -k_{IT,ab}[PD_a][D_b] + k_{IT,ba}[PD_b][D_a] \quad \text{[S11]}$$

$$\frac{d}{dt}[PD_b] = -k_{IT,ba}[PD_b][D_a] + k_{IT,ab}[PD_a][D_b] \quad \text{[S12]}$$

$$\frac{d}{dt}[D_a] = -k_{IT,ba}[PD_b][D_a] + k_{IT,ab}[PD_a][D_b] \quad \text{[S13]}$$

$$\frac{d}{dt}[D_b] = -k_{IT,ab}[PD_a][D_b] + k_{IT,ba}[PD_b][D_a] \quad \text{[S14]}$$

Eqs. 4 and 5 in the main text are obtained by comparison of Eqs. S11-S14 with Eqs. S7-S10.

2. A complete set of rate equations for the kinetic model shown in Figure 1

Here we provide a complete set of rate equations for the kinetic model illustrated in Figure 1. The rate equations for the probe DNA are as follows:

$$\begin{aligned}\frac{d}{dt}[D_{(1)}P] &= -\frac{d}{dt}[D_{(1)}] \\ &= k_{on,(1)}[P][D_{(1)}] - k_{off,(1)}[D_{(1)}P] + k_{sl,(2)}[D_{(2)}P] - k_{sl,1}[D_{(1)}P] \\ &\quad + \Gamma_{in,(1)}[D_{(1)}] - \Gamma_{out,(1)}[D_{(1)}P]\end{aligned} \quad \text{[S15]},$$

$$\frac{d}{dt}[D_{(h)}P] = -\frac{d}{dt}[D_{(h)}]$$
$$= k_{on,(h)}[P][D_{(h)}] - k_{off,(h)}[D_{(h)}P] + k_{sl,(h-1)}[D_{(h-1)}P] + k_{sl,(h+1)}[D_{(h+1)}P] - 2k_{sl,(h)}[D_{(h)}P]$$
$$+\Gamma_{in,(h)}[D_{(h)}] - \Gamma_{out,(h)}[D_{(h)}P]$$

[S16],

$$\frac{d}{dt}[D_{(L)}P] = -\frac{d}{dt}[D_{(L)}]$$
$$= k_{on,(L)}[P][D_{(L)}] - k_{off,(L)}[D_{(L)}P] + k_{sl,(L-1)}[D_{(L-1)}P] - k_{sl,(L)}[D_{(L)}P] \qquad \text{[S17].}$$
$$+\Gamma_{in,(L)}[D_{(L)}] - \Gamma_{out,(L)}[D_{(L)}P]$$

In these rate equations, an index in parentheses () represents the location of a site on DNA ($1 < h < L$ for Eq. S16); P, protein in the free state; $D_{(j)}$ ($j \neq m$), a protein-free nonspecific site on the probe DNA; $D_{(j)}P$, a protein-bound nonspecific site on the probe DNA; $D_{(m)}$, the target site in the free state; and $D_{(m)}P$, the specific complex with the target. The parameters $\Gamma_{in,(i)}$ and $\Gamma_{out,(i)}$ are given by:

$$\Gamma_{in,(i)} = \sum_{j}^{L} k_{IT,(j)(i)}[D_{(j)}P] + \sum_{n}^{M} k_{IT,(n)(i)}[C_{(n)}P] \qquad \text{[S18]}$$

$$\Gamma_{out,(i)} = \sum_{j}^{L} k_{IT,(i)(j)}[D_{(j)}] + \sum_{n}^{M} k_{IT,(i)(n)}[C_{(n)}], \qquad \text{[S19]}$$

where C represents a nonspecific site on the competitor DNA; CP, a protein-bound nonspecific site on competitor DNA; $k_{it,(j)(i)}$, the second-order rate constant for intersegment transfer from a nonspecific site j to another site i; and $k_{IT,(m)(i)}$, the second-order rate constant for intersegment transfer from the target site. For simplicity's sake, the same rate constants are assumed for all nonspecific sites (i.e., $j \neq m$): $k_{on,(j)} = k_{on,N}$, $k_{off(j)} = k_{off,N}$, $k_{sl,(j)} = k_{sl,N}$. and $k_{it,(j)(i)} = k_{it,N}$. For translocation from the target site, $k_{on,(m)} = k_{on,N}$, $k_{off(m)} = k_{off,S}$, $k_{sl,(m)} = k_{sl,S}$, and $k_{IT,(m)(i)} = k_{IT,S}$. As described in the main text, Eqs. S18 and S19 deal with only intermolecular intersegment transfer between two DNA duplexes, and do not deal with intra-molecular intersegment transfer because

the probe DNA duplexes used in this study are shorter than the persistence length. The rate equations for the competitor DNA are given by:

$$\begin{aligned}\frac{d}{dt}[C_{(1)}P] &= -\frac{d}{dt}[C_{(1)}] \\ &= k_{on,N}[P][C_{(1)}] - k_{off,N}[C_{(1)}P] + k_{sl,N}[C_{(2)}P] - k_{sl,N}[C_{(1)}P] \\ &\quad + \Gamma_{in,(1)}[C_{(1)}] - \Gamma_{out,(1)}[C_{(1)}P]\end{aligned} \tag{S20}$$

$$\begin{aligned}\frac{d}{dt}[C_{(h)}P] &= -\frac{d}{dt}[C_{(h)}] \\ &= k_{on,N}[P][C_{(h)}] - k_{off,N}[C_{(h)}P] + k_{sl,N}[C_{(h-1)}P] + k_{sl,N}[C_{(h+1)}P] - 2k_{sl,N}[C_{(h)}P] \\ &\quad + \Gamma_{in,(h)}[C_{(h)}] - \Gamma_{out,(h)}[C_{(h)}P]\end{aligned} \tag{S21}$$

$$\frac{d}{dt}[C_{(M)}P] = -\frac{d}{dt}[C_{(M)}] = k_{on,N}[P][C_{(M)}] - k_{off,N}[C_{(M)}P] + k_{sl,N}[C_{(M-1)}P] - k_{sl,N}[C_{(M)}P] + \Gamma_{in,(M)}[C_{(M)}] - \Gamma_{out,(M)}[C_{(M)}P] \tag{S22}$$

The index h in Eq. S21 is for a non-edge site on competitor DNA (i.e., $1 < h < M$). The rate equation for the protein in the free state is given by:

$$\begin{aligned}\frac{d}{dt}[P] &= -\sum_{i}^{L}\frac{d}{dt}[D_{(i)}P] - \sum_{i}^{M}\frac{d}{dt}[C_{(i)}P] \\ &= \sum_{i}^{L}\left(k_{off,(i)}[D_{(i)}P] - k_{on,(i)}[D_{(i)}][P]\right) + \sum_{i}^{M}\left(k_{off,N}[C_{(i)}P] - k_{on,N}[C_{(i)}][P]\right)\end{aligned} \tag{S23}$$

Intersegment transfer does not affect the populations of the protein in the free state.

3. ODE-based numerical simulations of target association kinetics

The time courses of the concentrations of individual species in the target association process for the systems involving protein, probe DNA, and competitor DNA were obtained via numerical integration of the rate equations (Eqs S15–S23) by using MATLAB software

(MathWorks, Inc). A standard ODE solver ('ode15s') in MATLAB was used to numerically solve the rate equations via intergration. Due to the principle of detailed balance (or microscopic reversibility),[2] the kinetic rate constants $k_{IT,S}$ and $k_{sl,S}$ were set to:

$$k_{IT,S} = k_{IT,N} K_{d,S} K_{d,N}^{-1} \quad \text{[S24]}$$

$$k_{sl,S} = k_{sl,N} [D_{(j)}]_{eq} K_{d,N}^{-1} [D_{(m)}]_{eq}^{-1} K_{d,S} , \quad \text{[S25]}$$

respectively (m is the index for the target). From the experimental settings, the initial conditions for the rate equations are set to $[P](0) = P_{tot}$; $[D_{(i)}](0) = D_{tot}$; $[C_{(i)}](0) = C_{tot}$; and $[D_{(i)}P](0) = [C_{(i)}P](0) = 0$, where P_{tot}, D_{tot}, and C_{tot} are the total concentrations of the protein, probe DNA, and competitor DNA, respectively. The apparent pseudo-first-order rate constant k_{app} for target association was calculated by mono-exponential fitting to the time course of $[D_{(m)}P](t) / D_{tot}$ from the ODE-based simulation. Independently of the ODE calculations, the equilibrium concentrations of individual species were calculated by solving the following simultaneous equations:

$$K_{d,S} = [D_{(m)}]_{eq} [P]_{eq} / [D_{(m)}P]_{eq} \quad \text{[S26]}$$

$$K_{d,N} = [D_{(j)}]_{eq} [P]_{eq} / [D_{(j)}P]_{eq} \quad (j \neq m) \quad \text{[S27]}$$

$$K_{d,N} = [C_{(i)}]_{eq} [P]_{eq} / [C_{(i)}P]_{eq} \quad \text{[S28]}$$

$$P_{tot} = [P]_{eq} + [D_{(m)}P]_{eq} + (L-1)[D_{(j)}P]_{eq} + M[C_{(i)}P]_{eq} \quad \text{[S29]}$$

$$D_{tot} = [D_{(i)}]_{eq} + [D_{(i)}P]_{eq} \quad \text{[S30]}$$

$$C_{tot} = [C_{(i)}]_{eq} + [C_{(i)}P]_{eq} \quad \text{[S31]}$$

for $[D_{(m)}]_{eq}$, $[D_{(j)}]_{eq}$, $[C_{(i)}]_{eq}$, $[D_{(m)}P]_{eq}$, $[D_{(j)}P]_{eq}$, and $[C_{(i)}P]_{eq}$. For validation, the final concentrations from the ODE-based kinetic simulations were confirmed to agree with the

equilibrium populations calculated from the total concentrations and equilibrium constants by using Eqs. S26-S31.

4. Derivation of Eq. 12 in the main text

For a two-state pseudo-first-order process represented by $S_a \underset{k_{ba}}{\overset{k_{ab}}{\rightleftharpoons}} S_b$, the time courses starting from the initial conditions $[S_a](0) = S_{tot}$ and $[S_b](0) = 0$ are as follows:[3]

$$[S_a](t) = S_{tot}\frac{K + \exp\{-(1+K)k_{ab}t\}}{1+K} \qquad \text{[S32]}$$

$$[S_b](t) = S_{tot}\frac{1 - \exp\{-(1+K)k_{ab}t\}}{1+K}, \qquad \text{[S33]}$$

where K is the equilibrium constant given by k_{ba} / k_{ab} (= $[S_a]_{eq} / [S_b]_{eq}$). Thus, the time course of fluorescence intensity for the ensemble of the states S_a and S_b is given by:

$$F(t) = \frac{1}{1+K}(F_a - F_b)\exp\{-(1+K)k_{ab}t\} + \frac{KF_a + F_b}{1+K} \qquad \text{[S34]}$$

where F_a and F_b correspond to fluorescence from the states S_a and S_b, respectively. Therefore, the apparent rate constant k_{app} from mono-exponential fitting to the fluorescence time-course data is:

$$k_{app} = (1+K)k_{ab} \qquad \text{[S35]}$$

for this pseudo-first-order process.

For systems with the protein, competitor DNA and probe DNA at concentrations satisfying $D_{tot} \ll P_{tot} \ll C_{tot}$, the target site $D_{(m)}$ undergoes a pseudo-first-order process involving the free and protein-bound states (corresponding to S_b and S_a, respectively). Based on Eq. S35, the apparent pseudo-first-order rate constant k_{app} for the target association is given by:

$$k_{app} = \left(1 + \frac{[D_{(m)}]_{eq}}{[D_{(m)}P]_{eq}}\right) k_1 = \left(1 + \frac{K_{d,S}}{f_p P_{tot}}\right) k_1 \qquad \text{[S36]}$$

Due to the inequalities, D_{tot} « P_{tot} « C_{tot}, the association of proteins with the competitor DNA reaches quasi-equilibrium far more rapidly than the association with the target. Therefore, well before the population of protein-bound target significantly increases, the fraction of the protein in the free state instantly become f_P as given by Eq. 13 in the main text. Using the mean search time T_P of the VK model (Eq. 7 in the main text), the initial molar flow rate for the free proteins to reach the target is given by $T_P^{-1} f_P P_{tot}$. This should be equal to the initial molar flow rate for the target to bind to the protein, which is given by $k_1 D_{tot}$. Thus, the pseudo-first-order rate constant k_1 in Eq. S36 is:

$$k_1 = T_P^{-1} f_P P_{tot} D_{tot}^{-1} \qquad \text{[S37]}$$

for systems without the intersegment transfer mechanism. Eq. 12 in the main text is obtained from Eqs. S36 and S37.

5. Consideration on two-orientation systems with $\phi = 2$

For proteins that bind to DNA as a monomer (e.g. Egr-1), two opposite orientations are possible (i.e., $\phi = 2$) at each site due to the structural pseudo-C_2 symmetry of double helical DNA. Because the presence of two possible orientations is equivalent to doubling the number of nonspecific sites, the parameter M is multiplied by ϕ in Eqs. 13, 17, and 20 in the main text. However, unless the target sequence is palindromic, the parameter L in Eqs. 7, 8, 10, and 19 in the main text should not be multiplied by ϕ because sliding with only one of the two possible orientations on the probe DNA can lead to the formation of the specific complex with the target.

The k_{app} constants from Eq. 18 with $\phi = 2$ show excellent agreement with those from ODE-based numerical simulations for two-orientation systems (Figure 3B in the main text).

6. P_{tot}-dependent k_{app} data

We analyzed the protein-concentration dependence of the apparent pseudo-first-order rate constant k_{app} for target association in the presence of competitor DNA at three different concentrations (C_{tot} = 1, 2, and 4 μM). The results are shown in Figure S1. As indicated by the analytical forms (i.e., Eqs. 12, 16, and 18 in the main text), the measured k_{app} constant was proportional to the total protein concentration P_{tot}.

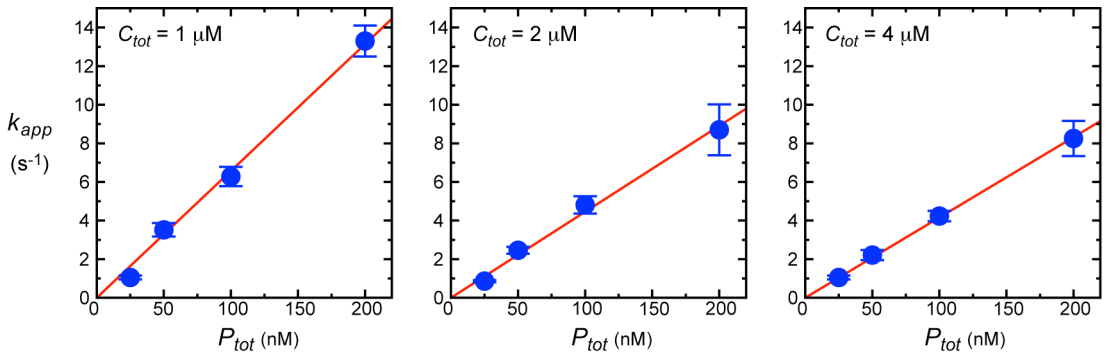

Figure S1. The protein-concentration dependent k_{app} data recorded for the 113-bp probe DNA in the presence of competitor DNA. The experimental conditions were the same as those used for Figure 6. The apparent second-order rate constants k_a for target association measured at C_{tot} = 1, 2, and 4 μM were 6.6×10^7, 4.5×10^7, and 4.2×10^7 M^{-1} s^{-1}, respectively.

References for Supplemental Information

1. Pilling, M. J. & Seakings, P. W. (1995). *Reaction kinetics*, Oxford University Press, Oxford.
2. Hammes, G. G. (2000). *Thermodynamics and kinetics for the biological sciences*. 3 edit, Wiley-Interscience, New York.
3. Atkins, P. & DePaula, J. (2001). *Physical Chemistry*. 7 edit, W. H. Freeman.