SI Appendix

I. Electrostatic Calculations

In order to simulate the translocation process we have to address two challenges. The first is the evaluation of the effective free energy surface and the other is simulating the dynamics on this surface. Obviously a full macroscopic evaluation of the relevant free energy surface is to challenging in part in view of the absence of the full structure of the complex and in part because of expected extreme convergence problems. At present we believe that the most effective strategy is to focus on the electrostatic free energy of the model system and this is done here with the PDLD/S-LRA approach(1).

The PDLD/S-LRA method evaluates the change in electrostatic free energies upon transfer of a given ligand *(l*) from water to the protein by starting with the effective PDLD potentials;

$$\overline{\mathbf{U}}^{\mathbf{p}}_{\mathbf{elec,l}} = \left\{ \left(\Delta\mathbf{G}^{\mathbf{l+p}}_{\mathbf{sol}} - \Delta\mathbf{G}^{\mathbf{l'+p}}_{\mathbf{sol}}\right)\left(\frac{1}{\boldsymbol{\varepsilon}_{\mathbf{p}}} - \frac{1}{\boldsymbol{\varepsilon}_{\mathbf{w}}}\right) + \Delta\mathbf{G}^{\mathbf{l}}_{\mathbf{sol}}\left(1 - \frac{1}{\boldsymbol{\varepsilon}_{\mathbf{p}}}\right) + \frac{\mathbf{U}^{\mathbf{l}}_{\mathbf{q\mu}}}{\boldsymbol{\varepsilon}_{\mathbf{p}}} + \frac{\mathbf{U}^{\mathbf{l}}_{\mathbf{intra}}}{\boldsymbol{\varepsilon}_{\mathbf{p}}} \right\}_{\mathbf{s}} \qquad (1)$$

$$\overline{\mathbf{U}}^{\mathbf{w}}_{\mathbf{elec,l}} = \left\{ \Delta\mathbf{G}^{\mathbf{l}}_{\mathbf{sol}}\left(\frac{1}{\boldsymbol{\varepsilon}_{\mathbf{p}}} - \frac{1}{\boldsymbol{\varepsilon}_{\mathbf{w}}}\right) + \Delta\mathbf{G}^{\mathbf{l}}_{\mathbf{sol}}\left(1 - \frac{1}{\boldsymbol{\varepsilon}_{\mathbf{p}}}\right) + \frac{\mathbf{U}^{\mathbf{l}}_{\mathbf{intra}}}{\boldsymbol{\varepsilon}_{\mathbf{p}}} \right\}_{\mathbf{s'}}$$

where ΔG_{sol} denotes the electrostatic contribution to the solvation free energy of the indicated group in water (e.g., ΔG_{sol}^{l+p} denotes the solvation of the protein-ligand complex in water). To be more precise, ΔG_{sol} should be scaled by 1 / (1 – 1/ε_w) but this small correction is neglected here. The values of the ΔG_{sol}'s are evaluated by the Langevin dipole solvent model. $U_{q\mu}^{l}$ is the electrostatic interaction between the charges of the ligand and the protein dipoles in vacuum (this

is a standard PDLD notation). In the present case $U^{l'}_{q\mu} = 0$. U^{l}_{intra} is the intramolecular electrostatic interaction of the ligand. Now the PDLD/S results obtained with a single protein-ligand configuration cannot capture properly the effect of the protein reorganization (see discussion in ref (1)) and a more consistent treatment should involve the use of the LRA or related approaches (e.g., ref. (1) and (2)). This approach provides a reasonable approximation for the corresponding electrostatic free energies:

$$\Delta \mathbf{G}^{\mathbf{elec}}_{\mathbf{bind}} = \frac{1}{2}\left\{\left[\left\langle \overline{\mathbf{U}}^{\mathbf{p}}_{\mathbf{elec,l}} \right\rangle_{\mathbf{l'}} + \left\langle \overline{\mathbf{U}}^{\mathbf{p}}_{\mathbf{elec,l}} \right\rangle_{\mathbf{l}}\right] - \left[\left\langle \overline{\mathbf{U}}^{\mathbf{w}}_{\mathbf{elec,l}} \right\rangle_{\mathbf{l'}} + \left\langle \overline{\mathbf{U}}^{\mathbf{w}}_{\mathbf{elec,l}} \right\rangle_{\mathbf{l}}\right]\right\} \qquad (2)$$

where the effective potential $\overline{U}$ is defined in Eq. 1 and $\langle\ \rangle_{l}$ and $\langle\ \rangle_{l'}$ designate an MD average over the coordinates of the ligand-complex in their polar and nonpolar forms. It is important to realize that the average of Eq. 2 is always done where both contributions to the relevant $\overline{U}_{elec}$ are evaluated at the same configurations. That is, the PDLD/S energies of the polar and nonpolar states are evaluated at each averaging step by using the same structure. However, we generate two set of structures one from MD runs on the polar state and one from MD runs on the nonpolar state. This is basically the same approach used in the microscopic LRA but now with the effective potential $\overline{U}_{elec}$.

Our initial screening is based on evaluation of electrostatic group contributions. This contribution are defined here as the effect of "mutating" all the residual charges of the given group to zero. In principle, we can perform such mutations and evaluate the PDLD/S-LRA binding energy for the given native and mutant. However, when we are dealing with charged and

polar residues, it is reasonable to start with the faster screening approximation introduced by Muegge et al. (3) This approach evaluates the electrostatic group contributions to the binding energy by looking at the term in Eq.1. This leads to

$$\left(\Delta \mathbf{G}_{\mathbf{bind}}^{\mathbf{elec}}\right)_{\mathbf{i}} \approx \left\langle \frac{\mathbf{U}_{\mathbf{q}\mu}^{\mathbf{i}}}{\boldsymbol{\varepsilon}_{\mathbf{x}}} \right\rangle \qquad (3)$$

where ε_x is taken as $\varepsilon_x \approx 4$ for polar residues and $\varepsilon_x = \varepsilon_{eff} \approx 40$ for ionized residues. This approach was examined in several test cases (e.g. ref (3)) and apparently provide a reasonable result for an initial screening.

The effective electrostatic free energy surface is given in SI-Fig. 1 (A) where we focus on the DNA protein interaction. SI-Figure 1 (B) presents the adjusted surface after taking into account the protein internal energy (see main text).

II. Group contributions to the translocation process

At this stage we may ask what is the basis for the shape of the translocation surface. This is done by evaluating the electrostatic group contribution for the interaction of the DNA with the protein, and taking the corresponding difference between the potential at D and T at R" = 2.4 and the difference between the potential at D and T at R" = 2. Residues that decrease the first difference and increase the second create the special pattern of Fig. 4 of the main text. Similarly we calculated the differences between the potential at R"=2.4 and R"=2.0 at T and at D. Where residues that decrease the first difference and increase the second generate the pattern of Fig 3 of the main text. The calculated results are depicted in SI-Fig. 2, which shows that the overall residue contributions help in establishing the translocation pattern. Mutating residues according to this pattern and examining the resulting translocation will be extremely instructive.

III. Effective free energy surface and Langevin dynamics simulations

The next challenge is to simulate the dynamics in the space defined by the effective coordinates of the system. To address this challenging problem we introduced a Langevin dynamics approach similar to the one used in our studies of proton translocation processes (4). That is, in order to explore the time dependence that coupled protein-DNA motions we approximate the effective surface obtained by the PDLD/S-LRA approach by a multi minima empirical valence bond (EVB)-type potential surface.

In this way the system is represented by mixing potential of the form (see ref. (4) for more details)

$$\mathrm{H}_{\mathrm{lm,lm}} = \boldsymbol{\varepsilon}_{\mathrm{lm}} \approx \frac{\hbar}{2}\boldsymbol{\omega}_{\mathrm{Q}}\left(\mathbf{Q}-\boldsymbol{\delta}_{\mathbf{Q}}^{\mathbf{l}}\right)^2 + \frac{\hbar}{2}\boldsymbol{\omega}_{\mathrm{R}}\left(\mathbf{R}-\boldsymbol{\delta}_{\mathbf{R}}^{\mathbf{m}}\right)^2 + \boldsymbol{\alpha}_{\mathbf{l,m}} \qquad (4)$$

where Q and R are the effective dimensionless coordinates of the protein (solvent) and the DNA, respectively R is related to the dimensional coordinate, R', by $R = \mathrm{R}'(\omega_R M_R/\hbar)^{1/2}$, whereas Q is defined by $Q = -(\varepsilon_{2,m} - \varepsilon_{1,m})_{el}\hbar\omega_Q\delta_Q$. Here l=1, 2, 3 for the ATP, ADP and empty forms, respectively, while m=0,1,2,3,.. for different positions of the DNA. Finally α_i is the difference between the minimums of the diagonal energies. We also use the general index *i = l+3(m-1)*. The effective frequency is evaluated by $\omega = \int \omega P(\omega) d\omega$, in which $P(\omega)$ is the normalized power spectrum of the corresponding contribution to $(\varepsilon_{2,m} - \varepsilon_{1,m})$. We also define reorganization energies

$$\lambda_R = (\hbar/2)\omega_R\delta_R^2 \quad , \qquad (5)$$

$$\lambda_Q = (\hbar/2)\omega_Q\delta_Q^2$$

in addition to the ε_i, we also have off-diagonal element, whose value are chosen to force the barriers between different minima to agree with estimates of the barriers for the chemically driven conformational changes. The actual potential surface is obtained by diagonalizing the system Hamiltonian

$$\mathbf{HC_g} = \mathrm{E_g}\mathbf{C_g} \qquad (6)$$

The surface of Eq. 3 was fitted to the full surface using the proper $\alpha_{l,m}$, with λ_Q=40 kcal/mol and $\mathrm{H_{ij}}$ =4 kcal/mol. (although equal fitting could have obtained with larger λ_Q and larger $\mathrm{H_{ij}}$) to reproduce the chemical barrier. The fitted surface is given in SI-Figure 3.

With the above effective surface it is possible to run Langevin dynamics (LD) simulations and to explore the time dependence of the translocation process.

The corresponding LD equation for the solvent coordinate is now expressed as (4)

$$M_{\mathbf{Q}}\ddot{Q}' = -M_Q\gamma_Q\dot{Q}' - M_{\mathbf{Q}}\omega_Q^2\left[\sum_{\mathbf{i}}\left(C_g^i\right)^2(\mathbf{Q}' - (\delta_{\mathbf{Q}}^{\mathbf{l(i)}})')\right] + A_Q'(t) \qquad (7)$$

where $Q = \left(M_Q\omega_Q/\hbar\right)^{1/2}Q'$, $\boldsymbol{\omega}_Q = \left(M_Q\boldsymbol{\omega}_Q/\hbar\right)^{1/2}\boldsymbol{\delta}_{\mathbf{Q}}'$, while γ_Q and M_Q are the effective friction and effective mass of the solvent and effective friction of the protein coordinate. $A_Q(t)$ here is a random force that satisfies the fluctuation dissipation relationship. The friction term is evaluated (see ref. (5)) from the relationship $\gamma_Q = \omega_Q^2\tau_Q$, where

$$\tau_Q = \int_0^\infty \left(\langle Q(0)Q(t)\rangle / \langle Q(0)Q(0)\rangle\right)dt \qquad (8)$$

The effective mass was estimated by using the relationship $M_Q = k_BT\Big/\left\langle\left(\dot{Q}'\right)^2\right\rangle$.

The equation of motion for the R coordinate is given by

$$M_{\mathbf{R}}\ddot{R}' = -M_R\gamma_R\dot{R}' - M_{\mathbf{R}}\omega_R^2\left[\sum_{\mathbf{i}}\left(C_g^i\right)^2(\mathbf{R}' - (\delta_{\mathbf{R}}^{\mathbf{m(i)}})')\right] + A_R'(t) \qquad (9)$$

Following ref (4) we selected for $\gamma_{\mathbf{Q}}$, ω_Q and M_Q the values 280 ps^{-1}, 40 cm^{-1} and 20 , respectively. A more rigorous treatment can be obtained by evaluating these parameters for the ATP bound system but the results are not expected to be very different. In the present case we estimated using the Einstein formula ($\gamma_R = k_BT/M_RD$). The diffusion constant (D) of a 10-base pair single strand DNA was estimated based on the experimental relationship(6) of DNA molecular length and diffusion constant. Here we obtained that D=9.7 $\mu m^2/s$ for 10 base pairs with m= 3400 amu. ω_R was estimated from the shape of surface and found to be 20cm^{-1} for m_R=3400 amu. Thus we used γ_R=75 ps^{-1}.

As clarified in the text, our simulations were done for different activation barriers in order to allow for interpolation to the actual barrier. The results of this study are summarized in SI-Figure 5. Using the figure to interpolate to a barrier of 18 kcal/mol, gave a translocation time of around 0.004s per nt in a qualitative agreement with the observed trend (0.007 s). Note that the value of

the interpolation curve at zero barriers for the protein conformational transition, still reflects the effect of the barrier for the motion of the DNA in the R direction.

References

1. Lee, F. S., Chu, Z. T. & Warshel, A. (1993) Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. *J Comput Chem* 14**:** 161-185.
2. Warshel, A., Sharma, P. K., Kato, M. & Parson, W. W. (2006) Modeling electrostatic effects in proteins. *Biochimica Biophys Acta Proteins Proteom* 1764**:** 1647-1676.
3. Muegge, I., Tao, H. & Warshel, A. (1997) A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. *Protein Engin* 10**:** 1363-1372.
4. Braun-Sand, S., Strajbl, M. & Warshel, A. (2004) Studies of proton translocations in biological systems: Simulating proton transport in carbonic anhydrase by EVB-based models. *Biophys J* 87**:** 2221-2239.
5. Hwang, J. K., King, G., Creighton, S. & Warshel, A. (1988) Simulation of Free-Energy Relationships and Dynamics of Sn2 Reactions in Aqueous-Solution. *J Am Chem Soc* 110**:** 5297-5311.
6. Robertson, R. M. & Smith, D. E. (2007) Strong effects of molecular topology on diffusion of entangled DNA molecules. *Proc Natl Acad Sci USA* 104**:** 4824-4827.