Supplementary Material

1 GSBP setup

Figure S1 shows the GSBP central inner region with radius $18 = 15 + 3$ (Å), where 3 Å is the water exclusion radius around the central sphere, centered around the center of mass of the hybrid. The inner atoms are shown in color with cartoons.

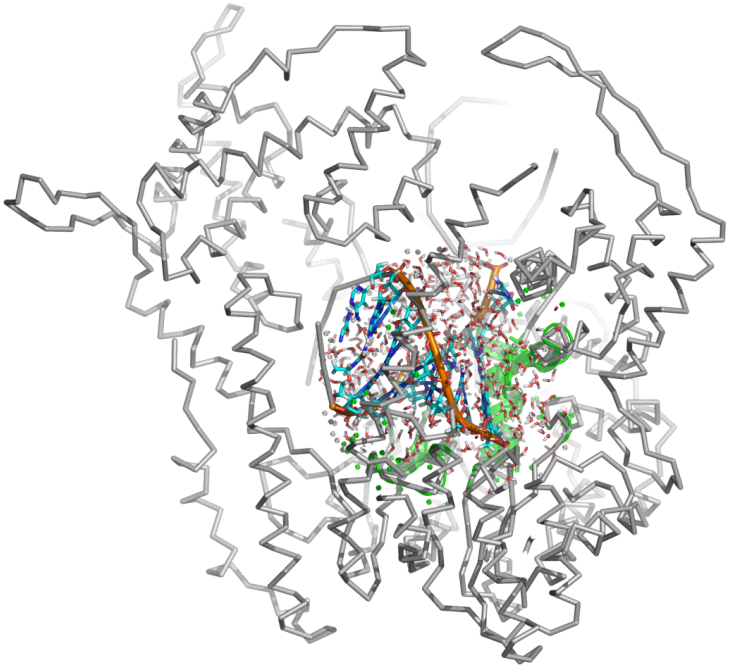

Figure S1: The GSBP sphere for Sampling 4 (cartoon; colored), shown together with the exterior atoms that are fixed during the sampling (ribbons; silver).

2 Umbrella sampling

Figure S2 shows the time series of the sampling sets collected using the full MD with PME (Table 2). Shown in Fig. S2(D) are the results of convergence analysis where trajectories in each sampling set were divided into partially overlapping 1 ns segments and the weighted histogram analysis performed using each segment. Sampling 2, started from the final coordinates

of Sampling 1, shows little sign of drift, with the pre- to post-translocated regime monotonically downhill and the post- to hyper-translocated moderately uphill in free energy. Sampling 3, started from the hyper-translocated state, converges consistently toward the Sampling 2 profile. The free energy profile of the post- to hyper-translocated regime is bracketed by Samplings 2 and 3. Sampling 4, run much longer but with the GSBP that allows less fluctuations, shows a similar profile.

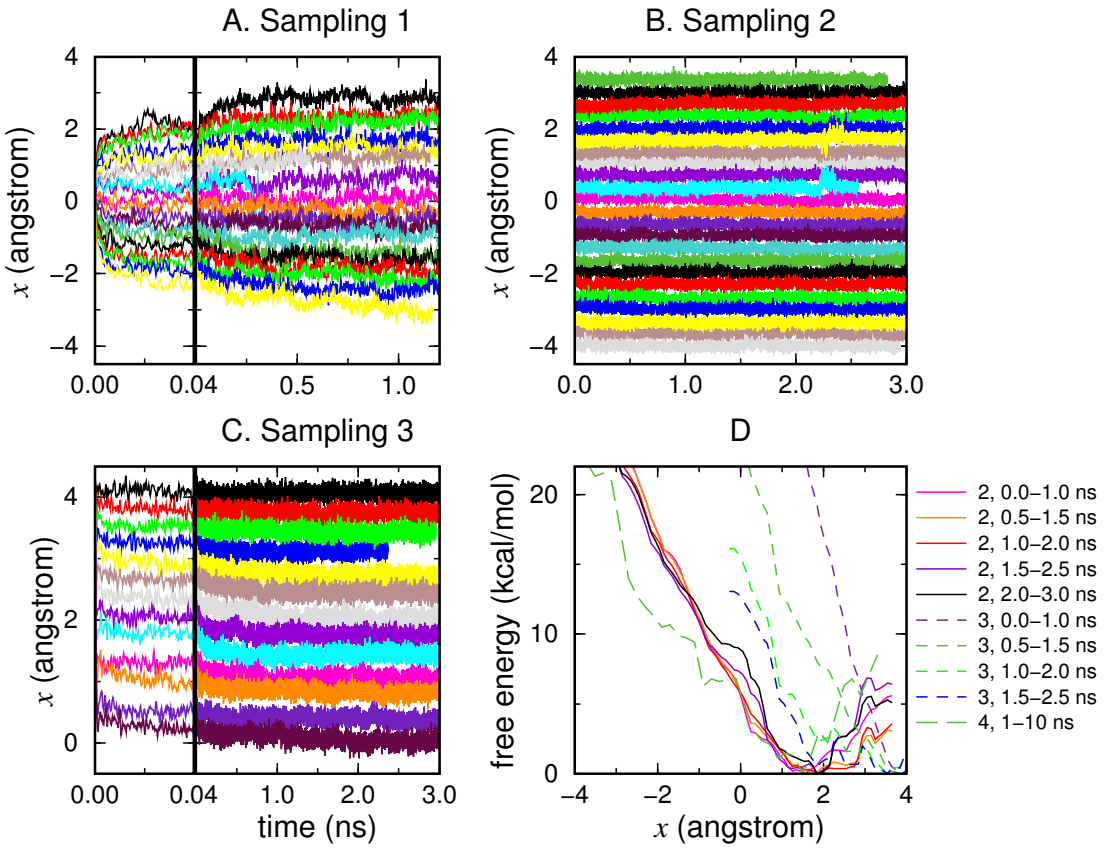

Figure S2: (A)-(C) Time series of translocation reaction coordinates from MD simulations (Table 2). Trajectories of different windows in each sampling are shown with different colors. (D) Convergence properties of free energy profiles; Trajectory segments used in the weighted histogram analysis are indicated in the legends.

3 Movies

Two movies illustrating aspects of the sampled dynamics are included:

1. The equilibration of the constrained dynamics of EC generating the pre-translocated state ("backtracking.avi"): this movie shows one trajectory from Sampling 1, started from the post-translocated initial state ($x = 0$) reaching $x \simeq -3$ Å at 1.2 ns. The constraint is centered at $x_0 = -3.4$ Å. The thumb helix break occurs around ~ 0.7 ns.

2. Close-up views on the active site of the same trajectory, followed by an additional 1 ns

segment from Sampling 2 run with a stronger constraint centered at $x_0 = -3.4$ Å ("backtracking_detail.avi"). The combined trajectories equlibrate a pre-translocated state starting near the post-translocated reference state (silver), corresponding to Fig. 3(B)→(C). The reaction coordinate time series of the two trajectories are shown in yellow in Fig. S2(A) and (B).

4 Translocation reaction coordinates

The translocation reaction coordinates were defined as follows: the contour of the hybrid is represented by the vector field of unit length $\mathbf{v}_i^{(0)}$ of the base pair rise from the i'th to $i+1$'th bases (Fig. 5) given by

$$\mathbf{v}_i^{(0)} = \frac{\mathbf{R}_{i+1}^{(0)} - \mathbf{R}_i^{(0)}}{\left|\mathbf{R}_{i+1}^{(0)} - \mathbf{R}_i^{(0)}\right|}, \tag{1}$$

where $\mathbf{R}_i^{(0)}$ is the center of mass position vector of the i'th base pair in the reference structure. The reaction coordinate x (and analogously for x_2) is then given as

$$x = \frac{1}{N-1} \sum_{i=1}^{N-1} \left(\mathbf{R}_i - \mathsf{U} \cdot \mathbf{R}_i^{(0)}\right) \cdot \mathsf{U} \cdot \mathbf{v}_i^{(0)}, \tag{2}$$

where the summation runs over the group of adjacent neighboring pairs of base pairs chosen for the definition of x (and x_2) with the total number $N-1$ (Figs. 4 and 5), and $\mathbf{R}_i$ is the position vector of the i'th base pair center of mass. The base pair center of mass can be written as

$$\mathbf{R}_i = \frac{1}{M_i} \sum_{j \in i} m_j \mathbf{r}_j, \tag{3}$$

where the summation is over all atoms belonging to the given base pair, m_j is the mass of each atom, and $M_i = \sum_j m_j$. The orthogonal matrix U in Eq. (2) aligns the reference structure of the protein to the given dynamical structure, such that $\mathsf{U} \cdot \mathbf{R}_i^{(0)}$ is the position vector of the base pair i of the reference in the frame aligned with respect to the dynamical structure. Equation (2) therefore averages $N-1$ values of the displacements projected onto the reference base pair rise directions. A double-stranded B-DNA fragment (PDB ID 1BNA[1]) was used to test the reaction coordinate. When the dynamical structure was taken as the reference ideally translocated by one base pair [setting $\mathbf{R}_i = \mathbf{R}_{i+1}^{(0)}$ in Eq. (S2)], the reaction coordinate value was $x = 3.47$ Å, in good agreement with the ideal B-DNA base pair rise distance.[1] The analogous value of $x = 3.82$ Å was calculated for the T7 RNAP EC reference structure also using $\mathbf{R}_i = \mathbf{R}_{i+1}^{(0)}$.

The MD simulation with constraining potential

$$u_n = k(x - x_0)^2 \tag{4}$$

was implemented by modifying the CONS RMSD module of CHARMM as follows: for each dynamical step, the reference structure of the complex (post-translocated crystal structure) is first aligned with respect to the given dynamical structure (Fig. 5). The potential u_n is then added to the total energy (and its derivative to the total force) with a given offset value x_0

and force constant k. From Eq. (2), the force acting on the atom j due to the constraint is given by

$$\mathbf{F}_j = -\frac{\partial u_n}{\partial \mathbf{r}_j} = -2k(x - x_0)\frac{\partial x}{\partial \mathbf{r}_j} \simeq \frac{-2k(x - x_0)m_j}{(N-1)M_i}\mathsf{U} \cdot \mathbf{v}_i^{(0)},$$

where the index i in the last expression corresponds to the base pair group the atom j belongs to, and the approximation $\partial \mathsf{U}/\partial \mathbf{r}_j \simeq 0$ was used (the movement of a single atom does not affect the rotation needed for alignment). This contribution due to constraints is added to the total force on each atom in the constrained dynamics.

References

[1] Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE. Structure of a B-DNA dodecamer: conformation and dynamics. Proc. Natl. Acad. Sci. USA 1981;78:2179–2183.