Supporting Information
for
Investigating the conformational dynamics of a Y-family DNA polymerase during its folding and binding to DNA and a nucleotide

Xiakun Chu\textsuperscript{1}, Zucai Suo\textsuperscript{2} and Jin Wang\textsuperscript{1,3*}

\textsuperscript{1} Department of Chemistry
State University of New York at Stony Brook, Stony Brook, New York 11794, USA

\textsuperscript{2} Department of Biomedical Sciences, College of Medicine
Florida State University, Tallahassee, Florida 32306, USA

\textsuperscript{3} Department of Physics and Astronomy
State University of New York at Stony Brook, Stony Brook, New York 11794, USA

\* jin.wang.1@stonybrook.edu
Models and Simulation Procedures

Two-bead coarse-grained double-well structure-based model of DPO4

A two-bead coarse-grained double-well structure-based model (SBM) was developed for studying the conformational transition of DPO4 between the DPO4 apo form (PDB: 2RDI [1]) and the DPO4-DNA-nucleotide (PDB: 1JX4 [2]) ternary form. In this SBM, each residue in DPO4 is represented by one backbone bead and one sidechain bead, locating at the $C_\alpha$ position and the centroid of the side chain of the residue, respectively (except for glycine, which is only represented by one $C_\alpha$ bead) (Figure S1A and S1B). The double-well SBM is derived from the conventional single-basin SBM, which has one single structure as the global minimum of the energy landscape [3]. The single-basin SBM has been widely applied to study protein folding and has produced results that are consistent with experiments in many faces [3]. The double-well SBM aims to generate two minima on the energy landscape in order to capture the conformational transition between structurally distinct states (e.g. apo and ternary state of DPO4) related to the functional purposes (DNA binding and nucleotide binding). It was found that the structures of DPO4 in DNA binary state and DNA-nucleotide ternary state are very similar [1]. Here, we used the DPO4 structure in the ternary state to represent DPO4 structure in both binary (DNA binding) and ternary (nucleotide binding) states.
**Figure S1:** All-atom (cartoon) and two-bead (sphere) representations of DPO4 in the (A) apo form and (B) ternary form (without showing DNA and nucleotide). The arrow in (B) indicates the large-scale domain spatial and rotational motions in DPO4 between the apo and ternary forms. This structural arrangement in DPO4 leads to a number of native contacts breaking and forming, partially illustrated by the orange lines. The specific native contacts formed between the T and LF domains in the apo DPO4 structure and formed between the F and LF domains in the ternary DPO4 structure are shown in (A) and (B), respectively. The change of these contacts in DPO4 between the apo and ternary form was used to describe the conformational transition of DPO4 in this study.

The potential of the double-well SBM used in our study has the following expression:

\[
V_{\text{DPO4 SBM}} = \sum_{\text{bonds}} K_r (r - \frac{r_1 + r_2}{2})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_1)^2 (\theta - \theta_2)^2 \\
+ \sum_{\text{dihedrals}} K_{\phi} \frac{F(\phi)}{\text{backbone}} + \sum_{\text{dihedrals}} K_{\phi} \frac{F(\phi)}{\text{sidechain}} + \sum_{\text{dihedrals}} K_{\phi} \frac{F(\phi)}{\text{chirality}} \\
+ \varepsilon_n \left[ \sum_{i < j} \varepsilon_{\text{dihedrals}} V_{LJ}(r_{ij}) + \sum_{i < j} \varepsilon_{\text{apo}} V_{LJ}(r_{ij}) + \sum_{i < j} \varepsilon_{\text{ternary}} V_{LJ}(r_{ij}) \right] \\
+ \varepsilon_{\text{pp}} \sum_{i < j} \left( \frac{\sigma_{\text{pp}}}{r_{ij}} \right)^{12} \\
+ \varepsilon_{\text{Elec}} \sum_{i < j} K_{\text{coulomb}} B(\kappa) \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}},
\]

where

\[
F(\phi) = \left[ \cos (\phi - \frac{\phi_1 + \phi_2}{2}) - \cos (\frac{\phi_1 - \phi_2}{2}) \right]^2,
\]

and

\[
V_{LJ}(r_{ij}) = 5 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{10}.
\]

\(r, \theta, \phi\) are bond, angle and dihedral, with the subscripts “1” and “2”, representing the corresponding values in the apo and ternary structure of DPO4, respectively. The prefactors \(K\) control the strengths of the corresponding local potential terms, including bond stretching, angle bending and dihedral rotation. The dihedral terms are further divided into the backbone, sidechain and chirality terms with the same expression \(F(\phi)\). The non-local potentials are made up of Lenard-Jones-like native contact potential, non-native repulsive potential and electrostatic potential. The prefactors \(\varepsilon\) control the strengths of the corresponding non-local potential terms.
Figure S2: Two-bead coarse-grained double-well SBM for DPO4 conformational dynamics. (A) Bond, (B) angle and (C) dihedral in DPO4 at the apo and ternary structure. (D) Native contacts and contact classification in DPO4 for building the double-well SBM. Native contact maps of DPO4 in the apo and ternary structures were built by the Contacts of Structural Units (CSU) software [4].

To realize the double-well potential in SBM, we used two different strategies at the local and non-local terms. For bond potential, which has little effects on the large-scale protein conformational change (Figure S2A), we simply used the average bond distance between the ones at the apo and ternary DPO4 native structure as the equilibrium position (position with the potential minimum) of the harmonic potential. For both angle and dihedral potentials, we built double-well potentials for the individual interaction terms, which have two potential minima locating at the apo and ternary DPO4 structure, respectively (Figure S2B and S2C). For native contacts, we mixed the native contacts from both the apo and ternary DPO4 structure and further applied criteria to classified these contacts into “Shared”, “Specific apo”, “Specific ternary” contacts (Figure S2D), similar with the strategy developed in our previous studies [5, 6]. \( \sigma_{ij} \) is the minimum distance between the bead \( i \) and \( j \) in the apo and ternary DPO4 structure with expression: \( \sigma_{ij} = \min(r_{ij,1}, r_{ij,2}) \), where \( r_{ij,1} \) and \( r_{ij,2} \) are the native distance between \( i \) and \( j \) in the apo and ternary DPO4 structure, respectively. Therefore, for “Specific apo” contacts, \( \sigma_{ij} = r_{ij,1} \); for “Specific ternary” contacts, \( \sigma_{ij} = r_{ij,2} \). For “Shared” contacts, \( \sigma_{ij} \) corresponds to the minimum of \( r_{ij,1} \) and \( r_{ij,2} \), as this leads to a small repulsive term in
$V_{LJ}(r_{ij})$, which can sample both $r_{ij,1}$ and $r_{ij,2}$ values [7].

![Figure S3](image_url)

**Figure S3:** Debye-Hückel model for electrostatic potentials at different salt concentrations $C_{Salt}$. At $C_{Salt}=0.00$ M, the Debye-Hückel model becomes the Coulomb potential.

In DPO4, only ARG and LYS residues were modeled with one positive charge and GLU and ASP residues were modeled with one negative charge. The charge was placed at the sidechain bead. We used the salt-dependent Debye-Hückel model to describe the electrostatic interaction [8]. $B(\kappa)$ is the salt-dependent coefficient; $q_i$ and $q_j$ are the charges of the beads $i$ and $j$, respectively. $\varepsilon_r$ is dielectric constant. $\kappa^{-1}$ is the Debye screening length, which is determined by the salt concentration ($C_{Salt}$) with relation: $\kappa \approx 3.2 \sqrt{C_{Salt}/C_0}$ nm$^{-1}$, where $C_0$ is reference concentration 1.0 M. To make the balance between the native contact and electrostatic interactions, we modulated the prefactor $\varepsilon_{Elec}$ to make the two negatively charged beads separated by 0.5 nm have the same energy of one native contact ($-\varepsilon_n$) (Figure S3). For salt-bridge native contacts, we further modulated both the $\varepsilon_n$ and $\varepsilon_{Elec}$ to make sure that total potentials maintain the same energy at the native distance and preserve the “steering” characteristic of the electrostatic potential at the long-range distance. Details of electrostatic interaction in SBM can be found in our previous work [9, 10].

All the prefactors of the potentials in $V_{DPO4}^{SBM}$ were determined based on previous work unless specified [3, 11, 8]. Reduced units were used except that the length is in the unit of nm. $K_r=10000.0\varepsilon$, $K_\theta=20.0\varepsilon$, $K_{\phi}^{hh}=1.0\varepsilon$, $K_{\phi}^{sd}=0.5\varepsilon$, $\varepsilon_{PP}=0.7\varepsilon$, $\sigma_{PP} = \sqrt{\sigma_i \sigma_j}$, where $\sigma_i$ and $\sigma_j$ are the diameters of the interacting bead $i$ and $j$, respectively. $\varepsilon$ is the energy unit, so $\varepsilon=1.0$. We set $\sigma_{i(j)}$ to 0.4 nm and 0.2 nm for the backbone and sidechain bead, respectively.

We set a strong value of $K_{\phi}^{chi}=20.0\varepsilon$ to maintain the correct chirality of the improper dihedral. To make the balance between the dihedral and native contact, we followed the suggestion
by Whitford et al. that the total contact to total dihedral energy ratio is 2.0 [11]:

\[ \frac{\sum_{i<j<3} \epsilon_n}{\sum_{\text{backbone dihedrals}} K_{bb}^{\phi} + \sum_{\text{sidechain dihedrals}} K_{sd}^{\phi}} = 2.0 \]

This results in \( \epsilon_n = 0.8 \epsilon \). In electrostatic potential, we set \( K_{\text{coulomb}} = 138.9 \epsilon, B(\kappa) = 1.0, \epsilon_r = 80.0 \).

We used Gromacs (version 4.5.7) [12] with PLUMED (version 2.5.0) [13] to perform the Langevin dynamics simulation with a friction coefficient \( 1.0 \tau^{-1} \), where \( \tau \) is the time unit. The time step was set to be 0.0005 \( \tau \). The non-bonded interactions were cut off at 3.0 nm. Temperature is in the energy unit by multiplying the Boltzmann constant \( (k) \).

To enhance the sampling, we performed Replica-Exchange Molecular Dynamics (REMD) simulations to explore the folding and conformational dynamics of DPO4 [14]. We launched two independent REMD simulations, starting from the DPO4_A and DPO4_T structures, respectively. Each REMD simulation was performed with 28 replicas spanning the temperature from 0.54 to 1.21. Each replica in the REMD simulations ran for \( 5 \times 10^6 \tau \). The exchange attempt between neighboring replicas was set to be every 1.0 \( \tau \) based on the Metropolis criterion. Trajectories were saved at every 10 \( \tau \), leading to \( 5 \times 10^5 \) snapshots in each replica and a total of \( 1.4 \times 10^7 \) snapshots for each REMD simulation. To assess the convergence of the REMD simulations, we calculated the thermodynamic properties of these two REMD simulations, respectively. The calculations were done through Weighted Histogram Analysis Method (WHAM) [15], which provides the free energy landscapes, melting curves, heat capacity curves, etc. In principle, the REMD simulations starting from two different structures, are deemed to be converged if these two sets of REMD simulations produce the same or similar thermodynamic results. We observed very similar thermodynamic results in terms of the free energy landscapes, the heat capacity curves and the melting curves at different salt concentrations (Figure S4). This indicates that the conformational sampling in these two simulations starting from two different DPO4 structures are very similar, thus these two REMD simulations are converged. Therefore, we collected all the data from these two REMD simulations and proceeded them to the WHAM to generate the thermodynamic results.
Figure S4: Thermodynamic results from the REMD simulations, starting from the apo DPO4 and ternary DPO4 structures at different salt concentrations. (A) Free energy landscape of DPO4 conformational dynamics at room temperature $T_r$ (Left), heat capacity curve of DPO4 (Middle) and melting curve of the helical formation of DPO4 (Right) at $C_{Salt}=0.01$ M. (B) Thermodynamic results at $C_{Salt}=0.05$ M. (C) Thermodynamic results at $C_{Salt}=0.15$ M. (D) Thermodynamic results at $C_{Salt}=0.30$ M.

To determine the values of $\varepsilon_{apo}$ and $\varepsilon_{ternary}$ in $V_{SBM}^{DPO4}$, usually one has to obtain the thermodynamics at a range of $\varepsilon_{apo}$ and $\varepsilon_{ternary}$ and make the comparisons to the experimental evidence. Intuitively, it should be done by performing individual simulations with different values of $\varepsilon_{apo}$ and $\varepsilon_{ternary}$. However, the process would be computationally prohibitive. Instead, we
applied a reweighting method using the simulation statistics at one set of parameters ($\varepsilon_{\text{apo}} = 1.0$ and $\varepsilon_{\text{ternary}} = 1.0$) to obtain the thermodynamics at the other sets of parameters [16, 17]. The method is briefly described as follows. The probability distribution $p$ of state having potential energy $E$ and reaction coordinate $r$ with default parameter $\varepsilon_0$ and any parameter $\varepsilon$ can be written as:

$$p(E(\varepsilon_0), r) = n(r)\exp\left[-\frac{E(\varepsilon_0)}{kT}\right]$$

$$p(E(\varepsilon), r) = n(r)\exp\left[-\frac{E(\varepsilon)}{kT}\right],$$

where $n(r)$ is the density of states, intrinsic to the system [18] and $T$ is the simulation temperature. Therefore, we can have distribution $p(E(\varepsilon), r)$ at any parameter $\varepsilon$ by reweighting $p(E(\varepsilon_0), r)$:

$$p(E(\varepsilon), r) = p(E(\varepsilon_0), r)\exp\left[-\frac{E(\varepsilon) - E(\varepsilon_0)}{kT}\right].$$

The parameter $\varepsilon$ can be any parameter in $V_{\text{DPO4}}^{\text{SBM}}$ and was set to $\varepsilon_{\text{apo}}$ and $\varepsilon_{\text{ternary}}$. In order to see the effects of electrostatic interactions on DPO4 conformational dynamics, the parameter $\varepsilon$ was further set to be $C_{\text{Salt}}$. The reweighting method is based on the principles of statistical mechanics and its precision is determined by the sampling sufficiency of the simulations at the default parameter $\varepsilon_0$.

The parameters of the interactions in the SBM are not calibrated in accordance with the realistic interactions, so the temperature in the SBM simulations cannot directly correspond to the experimental temperature. In order to see the effects of electrostatic interactions on DPO4 conformational dynamics, the parameter $\varepsilon$ was further set to be $C_{\text{Salt}}$. The reweighting method is based on the principles of statistical mechanics and its precision is determined by the sampling sufficiency of the simulations at the default parameter $\varepsilon_0$.

The folding temperatures of DPO4 were experimentally determined to be 89.3°C and 102.6°C from the two-step melting curve identified by the experiments [19]. In the SBM simulations, we also observed two peaks on the heat capacity curve (Figure S4), leading to two folding temperatures ($T_f^1$ and $T_f^2$). Then we assumed a linear temperature dependence on the energy, as widely used elsewhere [20, 21], to estimate the correspondence of the experimental temperature $T(\text{Exp})$ to the simulation one $T(\text{Sim})$ via the following expression:

$$T(\text{Sim}) = \frac{T_f^1(\text{Sim}) - T_f^2(\text{Sim})}{T_f^1(\text{Exp}) - T_f^2(\text{Exp})} \cdot [T(\text{Exp}) - T_f^1(\text{Exp})] + T_f^1(\text{Sim}),$$

where $T_f^1(\text{Sim})$ and $T_f^2(\text{Sim})$ are the folding temperatures from the simulation and $T_f^1(\text{Exp})$ and $T_f^2(\text{Exp})$ are the folding temperatures from the experiment. However, we note that due to the solvent effects, there may be a change in the heat capacity of denaturation, which results
in a non-linear temperature dependence on the energy. In this regard, the above linear temperature dependence is regarded as an approximate way to map the simulation temperature to the experimental one.

Figure S5: Probabilities of DPO4 at the DPO4_A, DPO4_I and DPO4_T forms for different sets of parameters $\varepsilon_{apo}$ and $\varepsilon_{ternary}$. (A) Probability distribution at the temperature $T_r$, mimicking room temperature (23°C). (B) Probability distribution at the temperature $T_{NMR}$, mimicking the experimental NMR temperature (50°C) [22]. The criterion for classifying the DPO4 state is based on $Q_{DPO4}(CT)$: DPO4_A with $Q_{DPO4}(CT) < -0.2$, DPO4_I with $-0.2 \leq Q_{DPO4}(CT) \leq 0.2$ and DPO4_T with $Q_{DPO4}(CT) > 0.2$.

Although the recent real-time Förster resonance energy transfer (FRET) method suggested that the dynamics of DPO4 exhibits a conformational equilibrium among the DPO4_A, DPO4_I and DPO4_T form when DNA is absent [23], there is still a lack of precise determination on the probability distribution of these three states. The crystal structure of the isolated DPO4 was solved in the DPO4_A form, implying that the DPO4_A form should be dominant at room temperature. A recent nuclear magnetic resonance (NMR) study, which was performed at an elevated temperature, found that DPO4 in the absence of DNA can form the DPO4_T form, but with a minor population [22]. Based on the above experimental evidence, we finally chose the values of both $\varepsilon_{apo}$ and $\varepsilon_{ternary}$ in $V_{SBM}^{DPO4}$ to be 1.0 (default value), which leads to populations of the DPO4_A, DPO4_I and DPO4_T form being 96.3%, 0.1% and 3.6% at the room temperature and 74.0%, 17.2%, 8.8% at the NMR temperature, respectively (Figure S5).
Figure S6: The free energy landscapes of DPO4 at the NMR experimental temperature. (A) The 2D free energy landscape projected onto $Q_{DPO4}(Rest)$ and $Q_{DPO4}(CT)$. (B) The 1D free energy landscapes projected onto $Q_{DPO4}(CT)$ at $T_r$ and $T_{NMR}$.

It is worth noting that the population of DPO4$_I$ increases more than DPO4$_T$ does when the temperature increases from room temperature to the one where the NMR experiment was conducted (Figure S6). From Figure 1C in the main text and Figure S6, we can see that DPO4 in the DPO4$_I$ breaks the interactions at the interface of the LF domain while the individual domains of DPO4 remain folded and other domain interfaces in DPO4 remain formed, thus the DPO4$_I$ is a high-energy state, which is driven by the entropy. Increasing temperature from room temperature to the NMR experimental temperature increases the proportion of the entropic contribution in the free energy, so the population of the DPO4$_I$ state is increased more significantly than the DPO4$_T$ state. From the experimental aspect, Sherrer et al. found that breaking the interactions involving the linker region is the first step for the DPO4 global unfolding with increasing temperature [19]. Besides, they assumed that DPO4 during thermal unfolding likely forms an intermediate state, where the linker is extended with the LF domain moving freely relative to the other domains. Therefore, our simulation results are consistent with the experiments and underlined the role of temperature in promoting the formation of the DPO4$_I$ state.
**Figure S7:** Folding and conformational dynamics of DPO4 described by RMSDs. (A) Melting curves of DPO4’s RMSD and heat capacity curve. RMSDs are the RMSD to the DPO4 apo structure (RMSD$_A$) and to the DPO4 ternary structure (RMSD$_T$). The three critical temperatures, i.e. room temperature $T_r$, the first folding temperature $T_{f1}$ and the second folding temperature $T_{f2}$, are indicated. (B) The free energy landscape projected onto the RMSD$_A$ and $Q_{DPO4}(CT)$ (Left), RMSD$_T$ and $Q_{DPO4}(CT)$ (Middle), and RMSD$_A$ and RMSD$_T$ (Right) at room temperature $T_r$. (C) and (D) are the same with (B) but for the free energy landscapes at the second folding temperature $T_{f2}$ and the first folding temperature $T_{f1}$, respectively.
Figure S8: The 1D free energy landscapes and structures of DPO4 folding. (A) The 1D free energy landscapes projected onto $Q_{DPO4}(\text{Rest})$ at room temperature $T_r$, the second folding temperature $T_f^2$ and the first folding temperature $T_f^1$. The free energy is divided by $kT$, where $T$ is the corresponding temperature. There are 5 (meta)stable on these three free energy landscapes. (B) The fraction of native contacts on the 5 states for intra- and inter-domain structural formation, represented by $Q_{DPO4}(I)$, where $I$ is the domain or interface. (C) The fraction of native contact matrices for the 5 states at the domain level $Q_{DPO4}(I,J)$, where $I$ and $J$ are the domain or interface. (D) The representative structures of DPO4 in these 5 states. The domains and linker are colored in the same scheme as Figure 1 in the main text.

Figure S9: Folding temperature dependent on the parameters $\epsilon_{apo}$ and $\epsilon_{ternary}$ in $V_{DPO4}^{SBM}$. The change of temperature relative to that with the default parameters ($\epsilon_{apo}$=1.0 and $\epsilon_{ternary}$=1.0) for (A) the second folding temperature $T_f^2$ and (A) the first folding temperature $T_f^1$.

Figure S10: Free energy landscapes of conformational transition of DPO4 at room temperature and melting curves at different parameters $\epsilon_{apo}$ and $\epsilon_{ternary}$. The results were obtained from the reweighting method (lines) and direct REMD simulations (dots) and show high consistency.
Figure S11: Folding and conformational dynamics of DPO4 dependent on electrostatic interactions, which are modulated by the salt concentrations. (A) Folding temperature changes on salt concentration. The REMD simulations at $C_{\text{Salt}}$=0.01 M, 0.15 M and 0.30 M were performed and compared with the results from the reweighting method. (B) Free energy landscapes of DPO4 conformational transition at room temperature $T_r$ among the DPO4_A, DPO4_I and DPO4_T form for different salt concentrations using the reweighting method. Free energy landscapes were also compared with the ones calculated from the direct REMD simulations. (C) Same with (B) but for the temperature at the experimental NMR temperature $T_{\text{NMR}}$.

Coarse-grained DNA model

Figure S12: Three-bead coarse-grained representation of DNA. Each nucleotide is described by three coarse-grained beads: C, N and P.

The DNA molecule in our study was taken from the ternary DPO4-DNA-nucleotide structure (PDB: 1JX4) [2]. Each nucleotide in DNA was coarse-grained into three beads to rep-
resent the base, sugar and phosphate group, respectively (Figure S12). The phosphate group was modeled with one negative charge. DNA was kept frozen throughout the simulations. The potential of the binary DPO4-DNA system is made up of the double-well DPO4 potential, DPO4-DNA native contact potential, DPO4-DNA non-native repulsive potential and DPO4-DNA electrostatic potential. The potential can be expressed as follows:

\[
V_{SBM}^{DPO4-DNA} = V_{SBM}^{DPO4} + V_{native}^{DPO4-DNA} + V_{non-native}^{DPO4-DNA} + V_{Electrostatic}^{DPO4-DNA}
\]

\[
= V_{SBM}^{DPO4} + \varepsilon_{DNA} \sum_{DPO4-DNA} V_{LJ}(r_{ij}) + \varepsilon_{PP} \sum_{DPO4-DNA} \left( \frac{\sigma_{PP}}{r_{ij}} \right)^{12} + \varepsilon_{Elec} \sum_{DPO4-DNA} K_{coulomb}(\kappa) \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}},
\]

where \(\varepsilon_{DNA}\) was adapted to match the experimental binding affinity.

\[
\sigma_{PP} = \sqrt{\sigma_{DPO4} \sigma_{DNA}},
\]

where \(\sigma_{DPO4}\) has two values depending on the bead types (backbone or sidechain) and \(\sigma_{DNA}=0.8\) nm.

**Figure S13:** The probability distribution along with \(Q^#_{DNA}\) during the umbrella sampling simulations.
**Figure S14:** The DPO4-DNA structures in the DNA<sub>US</sub>, DNA<sub>EC1</sub>, DNA<sub>EC2</sub>, DNA<sub>IS</sub> and DNA<sub>BS</sub> states. The figure is enlarged from Figure 2B in the main text.

**Figure S15:** DPO4-DNA binding when the positive charges in the linker region are removed. (A) Free energy landscapes with the presence and absence of the positive charges in the linker. (B) Barrier heights of the DPO4-DNA binding and unbinding with the presence and absence of the positive charges in the linker. (C-E) are the same with Figure 2C-2E in the main text but for the case when the positive charges in the linker are removed.
Figure S16: The conformational distribution of DPO4 during DNA binding. (A) The probability distribution of $Q_{DPO4}(CT)$ along with $Q_{DNA}^\#$. The probability distribution was calculated at each value of the $Q_{DNA}^\#$. The DPO4-DNA free energy landscape is shown as the cyan line. The transition states of the DNA_{EC} to the DNA_{IS} and the DNA_{IS} to DNA_{BS} are indicated by the transparent rectangles. (B) The probability distribution of $Q_{DPO4}(CT)$ at these two transition states. (C) The DPO4-DNA binding process. In each binding state, the most populated form of DPO4 is shown in the parentheses.
Figure S17: Binding free energy landscapes of DPO4 to DNA with different $\varepsilon_{\text{DNA}}$. The landscapes were calculated based on the reweighting method and umbrella sampling simulations at (A) $\varepsilon_{\text{DNA}}=0.70$ and (B) $\varepsilon_{\text{DNA}}=0.85$. The shadow regions are error bars, which were calculated based on three independent umbrella sampling simulations.

Figure S18: Binding free energy landscapes of DPO4 to DNA at different salt concentrations. The error bars were calculated based on three independent umbrella sampling simulations.
Coarse-grained nucleotide model

Figure S19: Five-bead coarse-grained representation of the incoming nucleotide. The adjacent coarse-grained beads are connected by the pseudo bonds. The calcium ion is bonded with two P beads. The beads with charge are indicated.

The nucleotide ddADP in our study was taken from the ternary DPO4-DNA-nucleotide structure (PDB: 1JX4) [2]. The nucleotide was coarse-grained into five beads (including one calcium ion) (Figure S19). The potential of ternary system can be expressed as follows:

\[
V_{DPO4-DNA-NT}^{SBM} = V_{DPO4-DNA}^{SBM} + V_{DPO4-DNA-NT}^{native} + V_{DPO4-DNA-NT}^{non-native} + V_{DPO4-DNA-NT}^{Electrostatic} + V_{NT}^{SBM}
\]

\[
= V_{DPO4-DNA}^{SBM} + \varepsilon_{NT} \sum_{DPO4-DNA-NT} V_{LJ}(r_{ij}) + \varepsilon_{PP} \sum_{DPO4-DNA-NT} \left( \frac{\sigma_{PP}}{r_{ij}} \right)^{12} + V_{DPO4-DNA-NT}^{Electrostatic} + V_{NT}^{SBM} + K_{NT} \left\{ \sum_{bonds} K_{r} (r - r_2)^2 + \sum_{angles} K_{\theta} (\theta - \theta_2)^2 \right. \\
\left. + \sum_{dihedrals} K_{\phi}^{(n)} [1 - \cos(n \times (\phi - \phi_2))] \right\},
\]

where \(\varepsilon_{NT}\) was adapted to match the experimental binding affinity.

\[
\sigma_{PP} = \sqrt{\varepsilon_{DPO4(DNA)} \sigma_{NT}},
\]

where \(\sigma_{NT}=0.8\) nm. \(V_{NT}^{SBM}\) used the local terms in a typical single-basin SBM potential for the nucleotide biasing to the ternary structure [3]. \(K_{r}^{(1)}=1.0\varepsilon\) and \(K_{\phi}^{(3)}=0.5\varepsilon\). The prefactor \(K_{NT}\)
is used for control the rigidity of the ligand and was set to $10.0\varepsilon$ to eliminate the flexibility of the nucleotide.

**Figure S20:** The probability distribution along with $Q^{\#}_{\text{NT}}$ during the umbrella sampling simulations.
**Figure S21:** The DPO4-DNA-Nucleotide structures in the NT\textsubscript{US}, NT\textsubscript{TS1}, NT\textsubscript{TS2} and NT\textsubscript{BS} states. The figure is enlarged from Figure 4B in the main text.

**Figure S22:** Barrier heights for (A) binding and (B) unbinding the nucleotide along with $\epsilon_{NT}$. 

---

23
Figure S23: Binding free energy landscapes of nucleotide to the DPO4-DNA complex with different $\epsilon_{NT}$. The landscapes were calculated based on the reweighting method and umbrella sampling simulations at (A) $\epsilon_{NT}=1.13$ and (B) $\epsilon_{NT}=1.30$. The shadow regions are error bars, which were calculated based on three independent umbrella sampling simulations.
Figure S24: The formations of native contacts (A) within each domain in DPO4, (B) at each domain interface in DPO4, (C) involving the linker in DPO4, (D) between each domain in DPO4 and DNA and (E) between the linker in DPO4 and DNA, formed at the NT<sub>US</sub>, NT<sub>TS1</sub>, NT<sub>TS2</sub> and NT<sub>BS</sub> states during the binding of nucleotide to the DPO4-DNA complex. In each panel, the top subfigure shows the probability distribution of $Q$ and the bottom subfigure shows the changes in the individual native contact ($Q_{ij}$, the probability of the individual native contact formed between the bead $i$ and $j$ in the SBM) from the binding state to the bound state (NT<sub>BS</sub>).
For simplicity, we bonded the ion to the nucleotide in the model, which leads to a coupled ion and nucleotide binding process. This is mainly due to the fact that the ion interactions are extremely challenging to be well described or modeled in the classical MD [24, 25]. In order to see the effects of the ion on the nucleotide binding process, we removed the ion and its related interactions in the SBM and performed the umbrella sampling simulations using the parameters calibrated when the ion is present. We also performed the same analyses as we presented before and obtained the results of the free energy landscapes, the native contacts and the interactions (Figure S25).
**Figure S25:** Nucleotide binding to the DPO4-DNA complex when the ion is absent. (A) Free energy landscapes of nucleotide binding to the DPO4-DNA complex projected onto the binding reaction coordinate $Q_{NT}$ at different $\varepsilon_{NT}$. The black line shows the free energy landscape when the ion is present with the parameters calibrated by the binding affinity. The grey line shows the free energy landscape when the ion is absent with the same parameters obtained from the simulations with the ion. Note that in the SBM, the ion forms native contacts with the DPO4-DNA complex in the ternary structure, so the maximum value of $Q_{NT}$ changes and the positions of the NTUS, NTTS and NTBS change, accordingly. (B) The probability distribution of the fraction of native contacts formed by the individual domains and linker in DPO4 and DNA with the nucleotide. (B) The probability distribution of the interaction energy (C) between DPO4 and the nucleotide and (D) between DNA and the nucleotide.

**Frustration calculations**

Frustration analyses were undertaken by Protein frustratometer server [26]. The core of the server is to calculate the frustration index, which measures how much a residue or residue pair contributes to the energy in a given structure compared to the statistics of the energies that would be found by placing different residues in the same native location or by creating a different environment for the interacting pair [27, 28]. If there is a strong energy stabilization for a native pair, the contact is assigned as “minimally frustrated”. If the stabilization lies in the middle of the distribution of alternatives, the contact is assigned as “neutral”. If the native pair energy shows strong destabilization, the contact is assigned as “highly frustrated”. The energy function of the server is based on the Associative memory, Water-mediated, Structure and Energy Model (AWSEM) [29], which was recently improved to include the electrostatic interactions [30]. The details of the method can be found here [27, 28] and at the server web page (http://frustratometer.qb.fcen.uba.ar/) [26].
Figure S26: Localized frustration and minimally frustrated networks in DPO4 at the native apo (2RDI [1]), DNA binary (2RDJ [1]) and DNA-nucleotide ternary (1JX4 [2]) states. (A) DPO4 at the apo state is shown in cartoon, the interactions are shown with solid lines. Minimally frustrated interactions are shown in green, highly frustrated interactions are shown in red, and neutral contacts are not drawn. At right, the contact map of DPO4 is plotted with color, indicating the frustration index. Red and blue rectangle regions indicate the major conformational differences of DPO4 during the functional “open-to-closed” conformational transition. (B) and (C) are the same with (A), but for DPO4 at the binary and ternary state.

The high-fidelity DNA polymerase we focused on is a DNA polymerase I large fragment
from a thermostable strain of *Bacillus stearothermophilus* (Bacillus fragment, BF). BF is a typical A-family DNA polymerase. The structure of BF shows that it includes three DNA-binding domains, the F, T and P domains, resembling a right-hand architecture [31]. In addition, BF possesses an exonuclease (ExoN) domain in the crystal structures [31, 32, 33], different from DPO4. The apo, binary and ternary structures of BF are very similar (Figure S27). The RMSDs between the apo and binary structures and between the binary and ternary structures are 0.13 nm and 0.21 nm, respectively. The observation indicates that there are no significant conformational changes in BF during substrate binding.
**Figure S27:** The crystal structures of BF in apo (PDB: 1XWL [31]), binary (PDB: 2BDP [32]) and ternary (PDB: 1LV5 [33]) states. (A) The superimposition of the BF apo and binary structures. The BF apo structure is shown in transparency. (B) The superimposition of the BF binary and ternary structures. The BF binary structure is shown in transparency. The F domain (residues 655-818), P domain (residues 617-655 and residues 830-869), T domain (residues 496-595) and ExoN domain of BF (residues 297-468) are colored blue, red, green and gold, respectively. The other regions of BF are colored grey.

**Figure S28:** Localized frustration in BF. (A) The number of highly frustrated interactions in the vicinity of each residue in the apo (PDB: 1XWL), binary (PDB: 2BDP) and ternary (PDB: 1LV5) states (*Top*). The differences between the apo and binary states, the binary and ternary states are respectively shown at *middle* and *bottom* in (A). The x-axis is colored according to the domain index in BF, same to that in Figure S27. (B) and (C) are BF structures colored according to the differences in contacts shown in (A). (D) The differences of frustration index of contacts in BF between the apo and binary states, the binary and ternary states, calculated based on the intra-domain (*Top*) and inter-domain (*Bottom*) interactions.
References

[1] Jimson H Wong, Kevin A Fiala, Zucai Suo, and Hong Ling. Snapshots of a y-family dna polymerase in replication: substrate-induced conformational transitions and implications for fidelity of dpo4. *Journal of molecular biology*, 379(2):317–330, 2008.

[2] Hong Ling, Françoise Boudsocq, Roger Woodgate, and Wei Yang. Crystal structure of a y-family dna polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell*, 107(1):91–102, 2001.

[3] Cecilia Clementi, Hugh Nymeyer, and José Nelson Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. *Journal of molecular biology*, 298(5):937–953, 2000.

[4] Vladimir Sobolev, Anatoli Sorokine, Jaime Prilusky, Enrique E Abola, and Marvin Edelman. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15(4):327–332, 1999.

[5] Yong Wang, Chun Tang, Erkang Wang, and Jin Wang. Exploration of multi-state conformational dynamics and underlying global functional landscape of maltose binding protein. *PLoS computational biology*, 8(4), 2012.

[6] Xiakun Chu, Fei Liu, Brian A Maxwell, Yong Wang, Zucai Suo, Haijun Wang, Wei Han, and Jin Wang. Dynamic conformational change regulates the protein-dna recognition: An investigation on binding of a y-family polymerase to its target dna. *PLOS Comput Biol*, 10(9):e1003804, 2014.

[7] Heiko Lammert, Alexander Schug, and José N Onuchic. Robustness and generalization of structure-based models for protein folding and function. *Proteins: Structure, Function, and Bioinformatics*, 77(4):881–891, 2009.

[8] Ariel Azia and Yaakov Levy. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *Journal of molecular biology*, 393(2):527–542, 2009.

[9] Xiakun Chu, Yong Wang, Linfeng Gan, Yawen Bai, Wei Han, Erkang Wang, and Jin Wang. Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone chz1 and histone h2a. z-h2b. *PLoS Comput Biol*, 8(7):e1002608, 2012.
[10] Xiakun Chu and Jin Wang. Position-, disorder-, and salt-dependent diffusion in binding-coupled-folding of intrinsically disordered proteins. *Physical Chemistry Chemical Physics*, 21(10):5634–5645, 2019.

[11] Paul C Whitford, Jeffrey K Noel, Shachi Gosavi, Alexander Schug, Kevin Y Sanbonmatsu, and José N Onuchic. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Structure, Function, and Bioinformatics*, 75(2):430–441, 2009.

[12] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[13] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.

[14] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[15] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[16] Huaiqing Cao, Yongqi Huang, and Zhirong Liu. Interplay between binding affinity and kinetics in protein–protein interactions. *Proteins: Structure, Function, and Bioinformatics*, 84(7):920–933, 2016.

[17] Maodong Li, Huaiqing Cao, Luhua Lai, and Zhirong Liu. Disordered linkers in multidomain allosteric proteins: Entropic effect to favor the open state or enhanced local concentration to favor the closed state? *Protein Science*, 27(9):1600–1610, 2018.

[18] Xiakun Chu, Linfeng Gan, Erkang Wang, and Jin Wang. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proceedings of the National Academy of Sciences*, 110(26):E2342–E2351, 2013.

[19] Shanen M Sherrer, Brian A Maxwell, Lindsey R Pack, Kevin A Fiala, Jason D Fowler, Jun Zhang, and Zucai Suo. Identification of an unfolding intermediate for a dna lesion bypass polymerase. *Chemical research in toxicology*, 25(7):1531–1540, 2012.

[20] Thomas Veitshans, Dmitri Klimov, and Devarajan Thirumalai. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding and Design*, 2(1):1–22, 1997.
[21] John Karanicolas and Charles L. Brooks. The origins of asymmetry in the folding transition states of protein l and protein g. *Protein Science*, 11(10):2351–2361, 2002.

[22] Eunjeong Lee, Jason D. Fowler, Zucai Suo, and Zhengrong Wu. Backbone assignment of the binary complex of the full length sulfolobus solfataricus dna polymerase iv and dna. *Biomolecular NMR assignments*, 11(1):39–43, 2017.

[23] Brian A. Maxwell, Cuiling Xu, and Zucai Suo. Conformational dynamics of a y-family dna polymerase during substrate binding and catalysis as revealed by interdomain forster resonance energy transfer. *Biochemistry*, 53(11):1768–1778, 2014.

[24] Saleh Riahi, Benoît Roux, and Christopher N. Rowley. Qm/mm molecular dynamics simulations of the hydration of mg (ii) and zn (ii) ions. *Canadian Journal of Chemistry*, 91(7):552–558, 2013.

[25] Bogdan Lev, Benoît Roux, and Sergei Yu Noskov. Relative free energies for hydration of monovalent ions from qm and qm/mm simulations. *Journal of chemical theory and computation*, 9(9):4165–4175, 2013.

[26] R. Gonzalo Parra, Nicholas P. Schafer, Leandro G. Radusky, Min-Yeh Tsai, A. Brenda Guzovsky, Peter G. Wolynes, and Diego U. Ferreiro. Protein frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic acids research*, 44(W1):W356–W360, 2016.

[27] Diego U. Ferreiro, Joseph A. Hegler, Elizabeth A. Komives, and Peter G. Wolynes. Localizing frustration in native proteins and protein assemblies. *Proceedings of the National Academy of Sciences*, 104(50):19819–19824, 2007.

[28] Diego U. Ferreiro, Joseph A. Hegler, Elizabeth A. Komives, and Peter G. Wolynes. On the role of frustration in the energy landscapes of allosteric proteins. *Proceedings of the National Academy of Sciences*, 108(9):3499–3503, 2011.

[29] Aram Davtyan, Nicholas P. Schafer, Weihua Zheng, Cecilia Clementi, Peter G. Wolynes, and Garegin A. Papoian. Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *The Journal of Physical Chemistry B*, 116(29):8494–8503, 2012.

[30] Min-Yeh Tsai, Weihua Zheng, D. Balamurugan, Nicholas P. Schafer, Bobby L. Kim, Margaret S. Cheung, and Peter G. Wolynes. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. *Protein Science*, 25(1):255–269, 2016.
[31] James R Kiefer, Chen Mao, Connie J Hansen, Scott L Basehore, Holly H Hogrefe, Jeffrey C Braman, and Lorena S Beese. Crystal structure of a thermostable bacillus dna polymerase I large fragment at 2.1 Å resolution. *Structure*, 5(1):95–108, 1997.

[32] James R Kiefer, Chen Mao, Jeffrey C Braman, and Lorena S Beese. Visualizing dna replication in a catalytically active bacillus dna polymerase crystal. *Nature*, 391(6664):304–307, 1998.

[33] Sean J Johnson, Jeffrey S Taylor, and Lorena S Beese. Processive dna synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proceedings of the National Academy of Sciences*, 100(7):3895–3900, 2003.