# Long Time-Scale Atomistic Simulations of the Structure and Dynamics of Transcription Factor-DNA Recognition

*Published as part of The Journal of Physical Chemistry virtual special issue "Young Scientists".*

Qinghua Liao,[†] Malin Lüking,[†] Dennis M. Krüger,[‡,§] Sebastian Deindl,[‡] Johan Elf,[‡] Peter M. Kasson,[‡] and Shina Caroline Lynn Kamerlin*[,†]

[†]Science for Life Laboratory, Department of Chemistry-BMC, Uppsala University, BMC Box 576, S-751 24 Uppsala, Sweden

[‡]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 23 Uppsala, Sweden

[§]Department for Epigenetics and Systems Medicine in Neurodegenerative Diseases, Bioinformatics Unit, German Center for Neurodegenerative Diseases, Göttingen, von Siebold Strasse 3A, 37075 Göttingen, Germany

**S** *Supporting Information*

**ABSTRACT:** Recent years have witnessed an explosion of interest in computational studies of DNA binding proteins, including both coarse-grained and atomistic simulations of transcription factor-DNA recognition, to understand how these transcription factors recognize their binding sites on the DNA with such exquisite specificity. The present study performs microsecond time scale all-atom simulations of the dimeric form of the lactose repressor (LacI), both in the absence of any DNA and in the presence of both specific and nonspecific complexes, considering three different DNA sequences. We examine, specifically, the conformational differences between specific and nonspecific protein−DNA interactions, as well as the behavior of the helix-turn-helix motif of LacI when interacting with the DNA. Our simulations suggest that stable LacI binding occurs primarily to bent A-form DNA, with a loss of LacI conformational entropy and optimization of correlated conformational equilibria across the protein. In addition, binding to the specific operator sequence involves a slightly larger number of stabilizing DNA−protein hydrogen bonds (in comparison to nonspecific complexes), which may account for the experimentally observed specificity for this operator. In doing so, our simulations provide a detailed atomistic description of potential structural drivers for LacI selectivity.
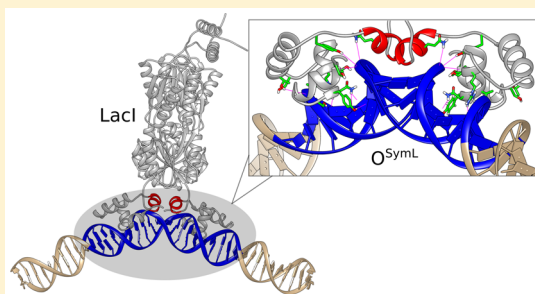
## INTRODUCTION

Gene regulation, which is controlled by DNA-binding proteins such as transcription factors (TFs),[1] is essential to cellular function. Thus, understanding protein−DNA interaction kinetics is of great biological and biophysical interest. Specifically, to perform their in vivo function, DNA-binding proteins must rapidly find their specific binding sites from a remarkable amount of nonspecific DNA.[2] Recent years have seen substantial progress in addressing the molecular description of dynamic target searching on DNA sequences by proteins, leading to a current consensus view that can be summarized as follows:[2−4] (1) The DNA-binding protein first binds nonspecifically to the DNA and then starts quickly searching the DNA for its target, through one-dimensional sliding. (2) During sliding, when the protein encounters a specific site, a tighter complex will be formed to recognize the target; otherwise, (3) the protein continues one-dimensional sliding or dissociates from the DNA at some point and then moves in the three-dimensional space of the solution, until it encounters a DNA segment.[2,5] Intersegmental transfer has also been suggested to be involved in the search process.[2] However,

as it is not possible to directly observe the search process at sufficiently high spatiotemporal resolution using experimental approaches, it is still unknown (in molecular detail) how these TFs recognize their binding sites and how they switch from nonspecific to specific binding. In addition, both specific and nonspecific binding events are considered to be key factors in the search process.[5]

To this end, molecular structures of protein−DNA complexes involving both specific and nonspecific binding have been solved by NMR spectroscopy or X-ray crystallography for EcoRV,[6] *Bam*HI,[7] the γ-repressor,[8] and the lactose repressor.[9−13] In the bacterium *Escherichia coli*, the lactose repressor (LacI) regulates the expression of a set of genes that are involved in lactose metabolism.[14] LacI was discovered as the first gene-regulatory DNA-binding protein, and it is still one of the most-studied model systems for exploring protein−DNA recognition.[9,10,15−23] Structurally, the LacI protein is a

homotetramer, with four identical subunits that are assembled into two dimers connected by four α-helices at the C-terminus of each monomer (Figure 1).[9,24] Each monomer consists in turn of 360 amino acids, which can be divided into four main regions: an N-terminal DNA-binding domain (residues 1−50), a regulatory domain (core domain, residues 61−340), a hinge region connecting the DNA-binding domain and the core domain (residues 51−60), and a C-terminal tetramerization region (residues 341−360) that joins the four monomers into an α-helix bundle.[9,25] The DNA-binding domain is, in turn, characterized by the presence of a helix-turn-helix (HTH) motif, with two symmetrically monomeric DNA binding domains forming a dimeric unit that binds a single operator site.[9,25]

The LacI dimer has different conformations for nonspecific and specific binding.[11−13] In the specific complex, residues 51−60 from each monomer form a well-ordered α-helix, referred to as a "hinge helix" (highlighted in red in Figure 1). The two hinge helices in the LacI dimer interact with each other, binding to the minor groove of the DNA. The insertion of the hinge helices into the minor groove leads to the DNA molecule bending by 36° in the specific complex, thus forming a bent structure.[11] In contrast, when LacI is bound to a nonspecific site, the hinge region has been argued based on NMR studies to be more disordered and in a random coil form, resulting in a more flexible and mobile HTH domain and a DNA molecule that is not bent,[12] although in this study the system carries a mutation in a hinge helix residue, V52C, to promote the formation of a disulfide bond. This could be interfering with the structural integrity of the hinge helices. High-resolution structures of nonspecific LacI-DNA complexes have not been reported as of yet, except for one solution-NMR structure of the LacI DNA binding domain in complex with nonspecific DNA.[12]

In addition, LacI has been shown to bind a range of operator sequences.[23,29−32] The main DNA operator for LacI, O[1], is essential for the function of the *lac* operon.[30] There are two more operators, namely, O[2] and O[3], located 401 base pairs downstream of O[1] and 92 base pairs upstream of O[1], respectively.[31] It was reported that LacI binds to O[2] with twofold lower binding affinity than with O[1], while it shows 200-fold lower affinity for the O[3] operator.[23] Furthermore, a perfectly symmetric operator, O[SymL] (5′-AATTGT-GAGCGCTCACAATT-3′) has been constructed that has been reported to bind LacI 10-fold more tightly than the other operators.[29] In the case of the LacI dimer, specific binding to the O[1] operator is in the range from $1 \times 10^{-8}$ to $1 \times 10^{-9}$ M[33] compared to $1 \times 10^{-3}$ M for nonspecific binding,[34] corresponding to at least $1 \times 10^{5}$ fold difference between specific and nonspecific binding. This is significant, as, in thermodynamic terms, this would be a difference of ∼7 kcal mol$^{-1}$, although the energetic differences for specific and nonspecific binding to other DNA binding proteins are small, on the range of ∼2−3 kcal mol$^{-1}$.[35−39]

There exists abundant structural[9,10,13,20,22,23,25] and functional[15−19,21,40] data to describe LacI, making it an ideal model system for characterizing protein−DNA recognition processes at the atomic level using simulation techniques, and there have been several interesting studies of this system. Given the extremely large (from a simulation perspective) size of ∼310 000 atoms of the LacI-DNA complex (for the full tetramer in complex with a rod model of DNA[26,41]), several of these have been performed using coarse-grained models, to explore the conformational changes of the complexes,[42] as well as more recent atomistic simulations.[41,43−48] Apart from the short time scale all-atom study of tetrameric LacI by Villa et al.,[41] other atomistic studies have included only the binding domain of either monomeric or dimeric LacI in the simulations. There is a risk that not including the core domain and C-terminal regions of LacI leads in turn to missing the collective motions of the binding domain and hinge regions relative to the core region, which is part of the LacI-DNA recognition mechanism. In addition, while a very powerful tool, coarse-grained simulations do not provide atomistic insight into the details driving the differences between specific and nonspecific protein−DNA interactions.

Here we present microsecond time scale all-atom simulations of the free LacI dimer as well as the dimer in specific and nonspecific complexes with three different DNA sequences. We focus on the LacI dimer both to make the simulation size more tractable and due to the extensive experimental data on this model system (5, 9, 10, 13, 15−23, 25, 33, 34, 40, 46, 49, 50), in contrast to the paucity of structural information on the tetramer. We focus, in particular, on understanding the dynamical differences between specific and nonspecific interactions, the behavior of the "clamping" helices of the protein when interacting with DNA. We show that there are sequence-dependent differences in the overall dynamics of LacI in complex with different DNA sequences, which could help explain the differences in selectivity of this transcription factor. In doing so, we provide a detailed atomistic description of the structural drivers for LacI selectivity.
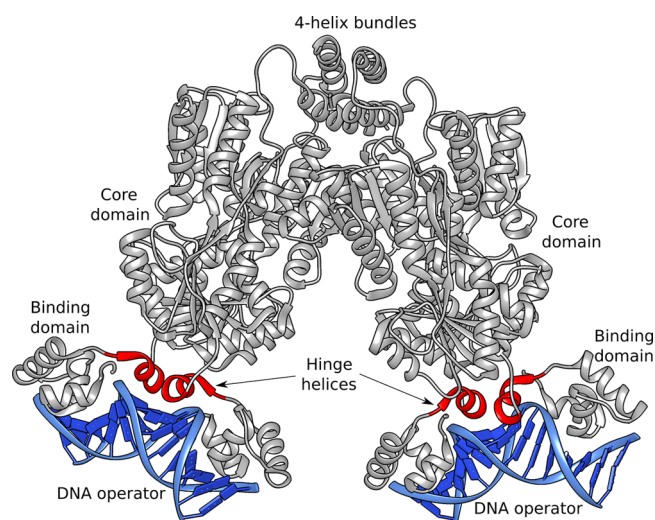


**Figure 1.** Structure of tetrameric LacI in complex with two DNA operators (PDB ID: 1Z04[26,27]). Here, the hinge helices are highlighted in red. These helices are disordered when LacI is bound to nonspecific DNA (see, e.g., PDB ID: 1OSL[12,27]), but they become well-ordered in α-helical form, when LacI is bound to specific DNA, such as the example shown here. This figure was made with Chimera v1.11.2.[28]

## ■ METHODOLOGY

**System Setup.** Our starting point for all simulations in this work is a 2.6 Å crystal structure of a LacI dimer in complex with a symmetric highly specific (consensus) operator sequence, O[SymL] (5′-AATTGTGAGCGCTCACAATT-3′),

and the anti-inducer orthonitrophenylfucoside (ONPF) (protein data bank (PDB) identification (ID): 1EFA[10,27]). This structure contains an A109T mutation; however, as this is a conservative mutation that is located over 20 Å away from the binding domain, we did not revert this mutation in our simulations but rather retained the original crystal structure. The ONPF molecules were removed from the system, and chains A, B, D, and E, comprising the LacI dimer and the DNA molecule, were retained for further simulations. The missing C-terminal tetramerization helices were added with Chimera[28] by taking coordinates from a crystal structure of the *lac* repressor in complex with a 21 base pair symmetric operator DNA and the gratuitous inducer isopropyl-$\beta$-D-1-thiogalacto-side (IPTG) (PDB ID: 1LBG[9,27]). Because of the common problem[51] of fraying of the base pairs at the two ends of the DNA strand during the simulations, we added 10 base pairs to each end of the operator sequence using the 3DNA web server,[52] to create a longer 40 base pair sequence centered on O$^{SymL}$ (5′-GCGTACAAGG-**AATTGTGAGC-GCTCACAATT**-GTGTAGGCGC-3′, Table S1). The resulting complex, which is shown in Figure 2, then served as a baseline for all subsequent simulations.

The simulations performed in this work are summarized in Table S2. In brief, we performed simulations of (1) free DNA, (2) free LacI, (3) LacI in complex with bent DNA (based on PDB ID: 1EFA[10,27]), and (4) LacI in complex with straight DNA (based on complexes constructed from snapshots of our simulations of free DNA and free LacI, respectively). As mentioned above, all simulations were performed using the LacI dimer. In addition, all simulations of free DNA or LacI-DNA complexes were performed using both the O$^{SymL}$ operator sequence described above, as well as two mutated sequences, DNA$^{mut1}$ and DNA$^{mut2}$, which were generated using an in-house script available for download from Zenodo (DOI: 10.5281/zenodo.14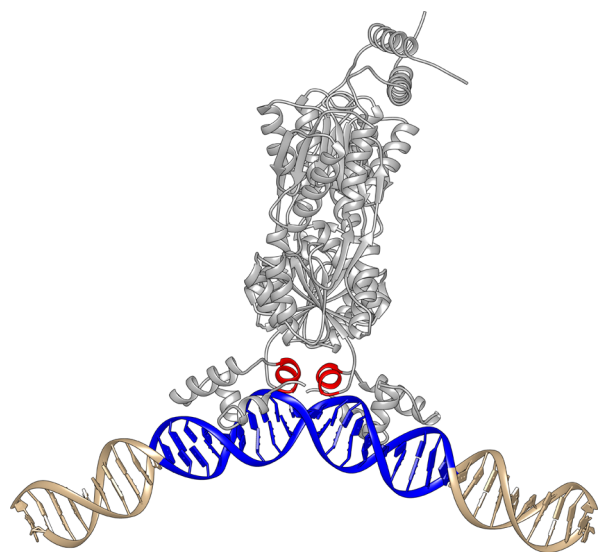94296). The mutations were generated by mutating either every other base of the operator sequence from A ↔ C and T ↔ G (DNA$^{mut1}$, 5′-GCGTACAAGG-**ACTGGGGCGAGATAAAACTG**-GTGTAGGCGC-3′, purine ↔ pyrimidine and pyrimidine ↔ purine), leading to a nonspecific complex with less chemical similarity to the specific complex, or by mutating every other base of the operator sequence from A ↔ G and T ↔ C (DNA$^{mut2}$, 5′-GCGTACAAGG-**AGTCGCGGGTGTTTATAGTC**-GTGTAGGCGC-3′, purine ↔ purine and pyrimidine ↔ pyrimidine), leading to a nonspecific complex with more structural similarity to the specific complex. Thus, in each case, three independent simulations were performed for each of O$^{SymL}$, DNA$^{mut1}$, and DNA$^{mut2}$, either as free DNA or in complex with LacI. In the case of the simulations of free DNA or LacI-DNA complexes with bent DNA, this was achieved simply by mutating every other base-pair directly in the crystallographic conformation of O$^{SymL}$ (from PDB ID: 1EFA[10,27]). In the case of the complexes of LacI and straight DNA, these simulations were performed by generating six artificial complexes based on snapshots extracted from simulations of free DNA and either three snapshots extracted from simulations of free LacI or three complexes built based on the crystal structure. These were then positioned relative to each other such that the protein and DNA were, on average, 17.9 Å apart (with a standard deviation of 1.4 Å), based on calculating the distance between the centers of mass of the binding domain of LacI and the binding site on the DNA sequences (Figure S1 and Table S3). For reference, the corresponding distance in PDB ID: 1EFA[10,27] is 10.6 Å, and therefore this corresponds to an ~7 Å displacement of the protein relative to the DNA.

Each system was then placed into a rectangular box filled with TIP3P water molecules,[53] with a distance of at least 8 Å from the solute to the surface of the box in each direction. The necessary number of Na$^+$ and Cl$^-$ counterions was then added to first neutralize the system and then achieve a 0.15 M salt concentration, in a random scheme using addIonsRand from the LEaP module as implemented in AMBER16[54] (for simulation specifics per system, see Table S3). All simulations were performed using the AMBER 16 simulation package,[54] with the protein described using the AMBER ff14SB force field[55] and the DNA described using the Parmbsc1 force field.[56] Finally, the LEaP module of AMBER16[54] was used to generate the topology and coordinate files for subsequent MD simulations.

**Simulation Details.** All molecular dynamics (MD) simulations were performed for each system (LacI-DNA complexes, free DNA or free LacI) using the same protocol, and using the CUDA version of the PMEMD module[57−59] of the AMBER16 simulation package.[54] Each solvated system was first subjected to a 10 000-step steepest descent minimization, with harmonic positional restraints applied to all atoms of the solute, using a 25 kcal mol$^{-1}$ Å$^{-2}$ force constant. Subsequently, two more 10 000-step steepest descent minimizations were performed for the system, using weaker harmonic positional restraints with a force constant of 5 kcal mol$^{-1}$ Å$^{-2}$. The minimized systems were then gradually heated from 100 to 300 K and equilibrated for 500 ps of simulation time, coupled by the Berendsen thermostat[60] with a time constant of 0.5 ps and 5 kcal mol$^{-1}$ Å$^{-2}$ harmonic positional restraints on all atoms. The system was then further optimized for 1000 ps in an NPT ensemble (300 K, 1 atm), controlled again by the Berendsen thermostat, and the Berendsen barostat[60] using a 1



**Figure 2.** Starting structure of the LacI-O$^{SymL}$ complex used in our simulations, based on PDB structure 1EFA,[10,27] with the DNA strand extended as described in the Methodology section. LacI is shown in gray, and the hinge helices are shown in red. The 20 base pair O$^{SymL}$ sequence is shown in blue, and the extended base pairs (see the Methodology section) are shown in tan. This figure was made with Chimera v1.13.1.[28]

ps time constant. Five further 100 ps equilibration steps were performed in an NPT ensemble, during which the 5 kcal mol$^{-1}$ Å$^{-2}$ harmonic positional restraints were decreased, until no positional restraint was left on the system, and a final 100 ps simulation (again NPT ensemble) was performed with no positional restraints on the system. Finally, production runs of varying lengths (at least 0.5 $\mu$s per trajectory) were performed for each system, as summarized in Table S2. Production simulations were performed at constant temperature (300 K) and constant pressure (1 atm), coupled by the Berendsen thermostat with a 1 ps coupling time, and the Monte Carlo barostat with a 1 ps time constant.[61,62] Each system was simulated using between 3 and 9 different replicas (Table S2), obtained by assigning different initial random velocities to each replica. All simulations were performed using a 2 fs time step, and snapshots were saved from the simulation for analysis every 10 ps. The SHAKE algorithm[63] was applied to constrain all bonds involving hydrogen atoms. An 8 Å cutoff was applied to all nonbonded interactions, with the long-range electrostatic interactions being described using the particle mesh Ewald (PME) approach.[64] A detailed summary of all the MD simulations performed in this work can be found in Table S2.

**PCA Analysis.** Principal component analysis (PCA) was performed using the *gmx covar* and *gmx anaeig* tools from GROMACS v. 2018.1.[65−67] To better quantify LacI-DNA recognition, we additionally calculated native contacts,[68] distances, and electrostatic interactions (approximated by the Debye−Hückel equation[69]) between the binding domain and the DNA binding sites, calculated using PLUMED v2.4.[70]

**Native Contacts Analysis.** Native contacts ($Q$) provide an excellent collective variable with which to define natively bound and unbound states of LacI; they have previously been effectively used in, for example, studies of peptide folding,[71,72] peptide binding,[73,74] and of loop motion.[75] The native contacts were calculated using eq 1:[68]

$$Q(X) = \frac{1}{S} \sum_{(i,j)\in} \frac{1}{S1 + \exp[\beta(r_{ij}(X) - \lambda r_{ij}^0)]} \tag{1}$$

where $X$ is a conformation along the reaction coordinate, $r_{ij}(X)$ is the distance between atoms $i$ and $j$ in conformation $X$, $r_{ij}^0$ is the distance between atoms $i$ and $j$ in the reference conformation, $S$ is the number of all pairs of heavy atoms ($i$, $j$), $\beta$ is set as 5 Å$^{-1}$, and $\lambda$ is chosen as 1.8, based on previous studies.[68]

The distance between LacI and DNA was, in turn, simply calculated based on the center of mass of all heavy atoms of the binding domain (Lys2-Ser61), and that of the central 20 base pairs (nucleotide ID 11−30 and 51−70), using eq 2:

$$d_{AB} = |\mathbf{R}_{BA}| = \left| \frac{1}{\sum_{i\in A} m_i} \sum_{i\in A} m_i \mathbf{r}_i - \frac{1}{\sum_{j\in B} m_j} \sum_{i\in B} m_j \mathbf{r}_j \right| \tag{2}$$

Here, A and B are the two groups of atoms, $\mathbf{R}_{BA}$ is the vector of the center of mass of groups A and B, $m_i$ and $m_j$ are the masses of atoms $i$ and $j$ belonging to groups A and B, respectively, while $\mathbf{r}_i$ and $\mathbf{r}_j$ are the coordinates. Finally, in the case of the electrostatic interaction energy between the binding domain of LacI and the DNA, this was approximated by the Debye− Hückel (DH) equation:[69]

$$E_{DH} = \frac{1}{4\pi\varepsilon_r\varepsilon_0} \sum_{i\in A} \sum_{j\in B} q_i q_j \frac{e^{-\kappa|r_{ij}|}}{|r_{ij}|} \tag{3}$$

where A (B) is the set of atoms of the first (second) group, $i$ and $j$ are the atom indexes in the two sets A and B, $|r_{ij}| = |r_i - r_j|$ denotes the distance between atoms $i$ and $j$, and $\kappa$ is the usual Debye−Hückel parameter.[69]

All native contacts analyses on the LacI-DNA simulations were performed using PLUMED v2.4,[70] where the native contacts $Q$ are defined by taking into account heavy-atom contacts between LacI and DNA, with a cutoff of 4.5 Å from the crystal structure (PDB ID: 1EFA[10,27]).

**Secondary Structure Analysis.** The GROMACS *do_dssp* interface to DSSP[76,77] was used to monitor the secondary structure changes of the hinge helices. In the crystal structure (PDB ID: 1EFA[10,27]), there are six $\alpha$-helical residues (Arg51-Leu56) for each monomer, and thus the hinge helicity is calculated based on the number of residues forming a helix divided by the total number of residues, 12, in $\alpha$-helical form, as in the starting structure. The angle between the two hinge helices is calculated from the dot product between two vectors along the two hinge helices, respectively. For each of the two vectors, it is a sum of the vectors $\overrightarrow{C \to O}$ and $\overrightarrow{H \to N}$ of the six residues in $\alpha$-helical form in each monomer. The angle calculated by Chimera v1.13.1[28] is 125.6° in the crystal structure (PDB ID: 1EFA[10,27]), which is very close to what we obtain using the dot product (127.5°).

**Analysis of DNA Bending.** An algorithm proposed by Curuksu et al.[78] was used in this study to characterize the degree of DNA bending. The orientation of one nucleotide can be defined by a reference axis system with unit vectors ($\overrightarrow{e_1}$, $\overrightarrow{e_2}$, and $\overrightarrow{e_3}$), where $\overrightarrow{e_1}$ is vector $\overrightarrow{N_9 \to C_8}$, and $\overrightarrow{e_2}$ is vector $\overrightarrow{N_9 \to C_1'} \times \overrightarrow{N_9 \to C_8}$ for the nucleotide at strand 1, while it is $\overrightarrow{N_9 \to C_8} \times \overrightarrow{N_9 \to C_1'}$ for the nucleotide at strand 2, and $\overrightarrow{e_3}$ is $\overrightarrow{e_1} \times \overrightarrow{e_2}$. As show in Figure S2, $\overrightarrow{e_1}$ and $\overrightarrow{e_3}$ are parallel to the base plane, while $\overrightarrow{e_2}$ is perpendicular to the base plane. The "arm" vector at each end of the DNA is the sum of the four $\overrightarrow{e_2}$ on the two base pairs at the two ends of the DNA binding sites (nucleotide ID: 9−10 and 71−72 at one end, and 31−32 and 49−50 at the other end), referred to as $\overrightarrow{H_1}$ and $\overrightarrow{H_2}$. The bending angle of the DNA is then calculated based on the dot product between the two arm vectors.

**Scoring A and B Forms of DNA.** To better quantitatively characterize the DNA conformations that may fall intermediate to the A and B forms of DNA, we introduce here a single scalar along the A-B continuum, the A-B index (ABI). A linear combination of four structural parameters ($S$) was used to construct the ABI, following previous studies[79] in which two structural parameters were used. The parameters used in this case are the backbone torsion ($\delta$), the puckering torsions ($\nu_1$ and $\nu_3$), and the glycosidic torsion $\chi$. The four parameters are then normalized relative to their differences between A-DNA and B-DNA (the values of the parameters in "typical" A-DNA and B-DNA conformations of O$^{SymL}$ are shown in Table 1).

On the basis of these values, the ABI of nucleotide $x$ can be defined using

$$ABI = \frac{1}{4}(\Delta\delta + \Delta\nu_1 + \Delta\nu_3 + \Delta\chi) \tag{4}$$

where the difference between the different parameters, $\Delta S$ ($\Delta\delta$, $\Delta\nu_1$, $\Delta\nu_3$, and $\Delta\chi$), is calculated by

**Table 1. Baseline Torsion Values Used to Describe A-DNA and B-DNA in Our A-B Transition Index Calculations[a]**

| torsions (deg) | A-DNA | B-DNA |
|---|---|---|
| $\delta$ | 79.1 | 143.4 |
| $\nu_1$ | −33.5 | 44.7 |
| $\nu_3$ | −40.2 | 23.1 |
| $\chi$ | −157.2 | −98.0 |

[a]These values are average values over all 80 nucleotides based on the standard A-DNA and B-DNA forms of the $O^{SymL}$ operator sequence, as generated by the 3DNA web server.[52] For the corresponding per-nucleotide values, see Figure S3, and for representative DNA conformations see Figure S4.

$$\Delta S(x) = \begin{cases} 1.0 & S(x) < S(A) \\ \dfrac{S(B) - S(x)}{S(B) - S(A)} & S(A) \leq S(x) \leq S(B) \\ 0.0 & S(x) > S(B) \end{cases} \tag{5}$$

Here, the $S(A)$ and $S(B)$ for the four parameters are listed in Table 1. With this definition, the DNA is in an A-like conformation when the ABI value is close to 1.0, while it is a B-like conformation when the ABI value is close to 0.0.

**Analyzing LacI Motion on the DNA Strand.** To describe diffusive motion of LacI relative to the DNA binding sites, the distance between the center of mass of the binding domain of LacI and the DNA binding sites was projected onto the DNA axial vector, which is a sum of the vector $\overrightarrow{e_2}$ on the central four base pairs (nucleotide IDs 20−21, 60−61). The projected distance can be both positive and negative, which indicates that LacI sits in one (5′-direction) or the other (3′-direction) half of the DNA binding sites.

**Statistical Tests.** Statistical differences in number of hydrogen bonds and Debye−Hükel energies were estimated by calculating these quantities on the molecular dynamics trajectories and sampling at statistically uncorrelated intervals. The decorrelation time scales were estimated by plotting time autocorrelation functions and performing double-exponential fits ± a constant. The longest decorrelation time scales thus determined were less than 10 ns, so a 10 ns sampling interval was selected.

**Other Analysis.** All other analyses were performed using CPPTRAJ[80] from AmberTools 16,[54] the Visual Molecular Dynamics (VMD)[81] package, and the MDTraj 1.8.0 library.[82] All structure figures were prepared using Chimera.[28]

## ■ RESULTS AND DISCUSSION

**Probing the Intrinsic Flexibility of the Binding Domain in LacI.** To obtain insight into the intrinsic flexibility of the DNA binding domain, we performed as our starting point $9 \times 1.0$ $\mu$s MD simulations of free LacI starting from the $O^{SymL}$-bound structure (PDB ID: 1EFA[10,27]) in the absence of DNA. Table S2 contains a summary of all simulations performed in this work. As shown in Figure S5A, the backbone root-mean-square deviation (RMSD) of LacI from the reference structure ranges between 2 and 9 Å over these simulations. This is due to the flexibility of the binding domain; the backbone RMSD of LacI with the DNA-binding domains omitted is consistently 2−4 Å during the same simulations (Figure S5B).

Per-residue root-mean-square fluctuation (RMSF) plots (Figure 3A) further show that DNA-binding domain flexibility
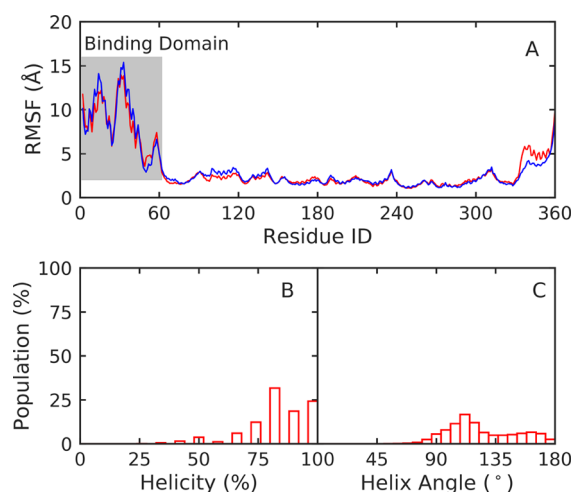


**Figure 3.** (A) The RMSF (Å) of the $C_\alpha$ atoms of the free LacI dimer, calculated over nine independent 1.0 $\mu$s MD simulations (Table S2). The blue and red lines indicate the two monomeric subunits in the dimer, respectively. (B) Population distribution of the % helicity of the hinge helices, calculated based on the number of residues involved in forming an $\alpha$-helix or other forms of helices ($\pi$- and $3_{10}$-helix), over nine independent 1.0 $\mu$s MD simulations of free LacI. A helicity of 100% is equivalent to $2 \times 6$ residues (Arg51-Leu56) in helical form, as in the crystal structure of LacI in complex with $O^{SymL}$ (PDB ID: 1EFA[10,27]). The secondary structure composition was calculated using DSSP.[76,77] (C) The corresponding population distribution of the angle between the hinge helices during the same simulations. The process by which the angle was calculated is described in the Methodology section.

dominates the motions of free LacI. The core domains are more stable, with the C-terminal helices showing intermediate flexibility. We note that experiments suggest that the hinge helices are unfolded when not in a specific LacI-DNA complex.[12] As can be seen in Figure 3B, we observe a general overall loss of helicity in the hinge helices (assessed via DSSP[76,77]) during our simulations, in good agreement with these experimental observations.[83] In addition, complete loss of helicity was observed in 0.02% of simulation snapshots. Accompanying the overall loss of helicity, the angle between the two helices also varies substantially, ranging from 50° to 180° (Figure 3C), indicating similar changes in helical orientation. For reference, the angle between the hinge helices is 127.5° in the crystal complex of LacI-$O^{SymL}$ (PDB ID: 1EFA[10,27]).

The observed changes in the hinge helices in the simulations of free LacI are of interest, as the HTH motif observed in LacI (Figure 1) is a common feature of many DNA-binding proteins,[84,85] and the formation of the hinge helices has been argued to be an important feature for DNA bending in various *lac* repressor-operator complexes.[83,86] It has also been argued that the protein−DNA interface of the nonspecific complex is flexible and that this assists in rapid and efficient search for the target site on the DNA.[12] In addition, a recent study has shown that the LacI hinge domain independently binds DNA,[87] with the hinge region mediating both specific and nonspecific binding,[87] with only small differences in affinity between the two. Following from this, computational studies have suggested that the hinge region plays an important contribution to the electrostatic energy of the LacI-DNA complex, the salt dependence of this electrostatic energy, and the number of salt ions excluded as the LacI-DNA complex is
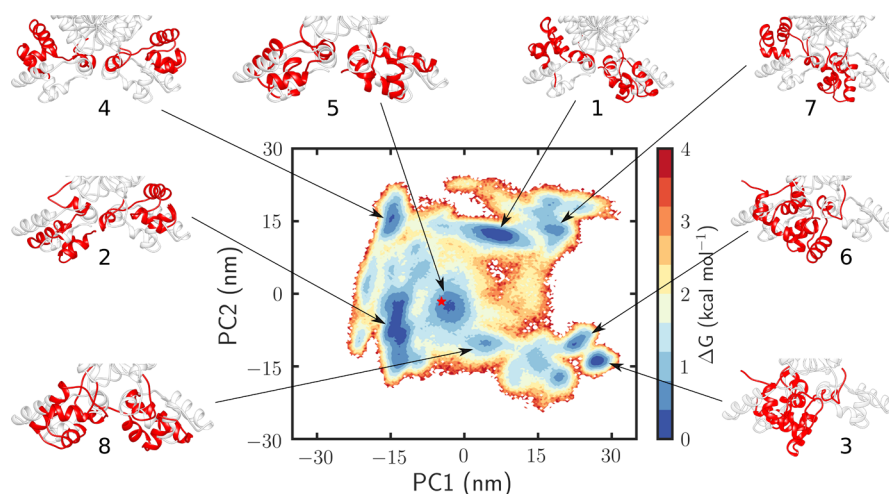
**Figure 4.** Visualization of binding-domain conformational equilibria via projection onto the first two principal components of the binding-domain position. Representative snapshots corresponding to the minima from the estimated free energy landscape, numbered sequentially from highest to lowest occupancy, are also shown here in red to illustrate the different conformations taken by the binding domains during our simulations. The crystal conformation of LacI (from PDB ID: 1EFA[10,27]) is highlighted on the plot as a red star for comparison.

formed[48] and that the binding of the hinge region of the *lac* repressor to DNA perturbs networks of correlated motions in the protein that in turn regulates its DNA binding affinity.[88]

Finally, we estimated prevalent binding-domain conformations in free LacI as follows. We aligned all simulation snapshots to the backbone atoms of the core residues (62−332) in the starting LacI crystal structure and calculated positional covariance of all backbone atoms in the binding domain (residues 2−61). We then projected state density onto the first two principal components of this covariance matrix to estimate the free energy landscape of the binding domain in a low-dimensional visualization (Figure 4). This analysis shows that the binding domain is indeed flexible, with a variety of conformations sampled during the simulation. In addition, in agreement with previous studies,[87] it appears that the DNA binding helices are capable of moving independently of each other, which likely affects the search process on the DNA. This is particularly interesting, because the hinge sequence itself may also contribute substantially to the binding affinity for both specific and nonspecific DNA, while also playing an important role in the transition from nonspecific to specific binding.[86,87]

**Simulations of Free DNA in the Absence of LacI.** A second natural point to consider is the extent to which the intrinsic behavior of the DNA itself determines differences in LacI-DNA recognition. We therefore performed an additional $9 \times 1.0~\mu s$ MD simulations for each of three DNA sequences considered in this work (summarized in Table S2). In these simulations, the differences between the average behaviors of the three DNA sequences are minor (analyzed via Curves+,[89] shown in Figures S6−S15, and Tables S4−S13) and therefore unlikely to determine the observed differences in specificity. We note that, in these simulations, the terminal five base pairs on each end of the DNA strand were not analyzed to avoid end-effects such as fraying.

The bent DNA structures observed in LacI-DNA complexes (PDB ID: 1EFA[10,27]), however, suggest that formation of a bent DNA conformation may be related to LacI discrimination between operator sequences. We first analyze this bending propensity as a static property and then as a dynamic one. Comparison of our simulations of free DNA with the LacI-

$O^{SymL}$ crystal structure shows that the DNA sequence stays in a bent A-form when it is bound to LacI and in a straight B-form in the absence of protein (Figures S4 and S16). These bent A-like DNA conformations have been previously suggested to be important to protein−DNA recognition.[11,90−95] We therefore devised a metric for scoring DNA structures along a continuum between A-form and B-form DNA (see the Methodology section), which we term the A-B index (ABI). The ABI of every nucleotide in the specific and nonspecific complexes is shown in Figures 5 (data from our simulations) and S16 (data from experimental structures). It can be seen that, in the specific complex (PDB ID: 1EFA[10,27]), the DNA is largely in an A-like conformation with ABI values greater than
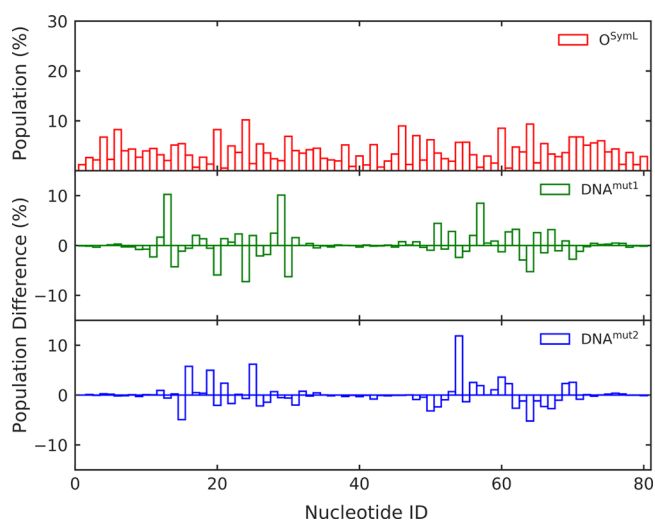


**Figure 5.** Fraction of time each DNA base pair spends in the A form. The percentage time each base pair in simulated DNA containing the $O^{SymL}$ sequence, DNA$^{mut1}$, or DNA$^{mut2}$ spent in A-DNA conformations was calculated based on $9 \times 1.0~\mu s$ simulations of free DNA for each system using the classification scheme described in the Methodology. (top) Averages for each nucleotide across simulations of the DNA containing the $O^{SymL}$ sequence. (bottom) Differences between the DNA containing the $O^{SymL}$ sequence and DNA$^{mut1}$ and DNA$^{mut2}$, respectively.

0.6, whereas in the nonspecific complex (PDB ID: 1OSL[12,27]), the DNA is largely in a B-like conformation, with ABI values less than 0.1 (Figure S16). For comparison, the free-DNA simulations show that these DNA sequences are more likely to stay in the B-form (a probability of ~10% for A-form DNA in the case of the DNA containing the $O^{SymL}$ operator sequence), with some sequence-dependent differences between the three sequences (Figure 5). This is, in turn, in good agreement with the observations of ref 96.

We note here additionally as an aside that it has been shown experimentally that DNA conformation is greatly sensitive to hydration and that dehydration is necessary for the observation of the A-form conformation of DNA, as free A-form DNA is only adopted under conditions of relatively low water content (see, e.g., refs 97−101 and references cited therein) or in certain protein complexes (as in the case of LacI in complex with DNA). In agreement with experimental observations, it has been shown that, with current DNA/RNA force fields, A-form DNA is not stable in water, even with the newly developed Parmbsc1 force field[56] (although it can be maintained for several hundred nanoseconds in organic solvent[56]). Therefore, the fact that we only observe significant presence of A-form DNA in simulations of DNA in complex with LacI and not in simulations of free DNA in aqueous solution is also in excellent agreement with both previous experimental results that show that B-form DNA is the preferred conformation in aqueous solution[97−101] and computational results that show that changes in local microenvironment are necessary to achieve the B- to A-DNA form transition.[101,102]

Surprisingly, when the operator sequence was radically mutated ($DNA^{mut1}$ and $DNA^{mut2}$), the ABI scores for the central 20 nucleotides of the operator where LacI binds showed large changes, while the terminal 20 nucleotides were relatively unchanged. Thus, the mutations do appear to influence both the motion and flexibility of the DNA binding sites. In addition, fraying of the terminal base pairs was observed in all the simulations, as shown in Figure S17.

To analyze the flexibility of the unbound DNA sequences, we examined the RMSF of all heavy atoms of each nucleotide of the three DNA sequences (Figure 6). We note that, while there are some minor differences between the sequences, the DNA containing the $O^{SymL}$ sequences is intermediate in flexibility to $DNA^{mut1}$ and $DNA^{mut2}$, and there appears to be no trend between flexibility and selectivity. We note that previous

studies[103−105] have argued that DNA flexibility is quite sequence-dependent and have suggested that increased flexibility can assist DNA recognition by DNA binding proteins. Our data, however, suggest only minor differences between the three DNA sequences (Figure 6) on the time scales sampled in our simulations.

**Modeling Specific and Non-Specific LacI-DNA Interactions on Straight DNA.** Having confirmed that the binding domain of LacI is indeed highly flexible and can take on multiple configurations even in the absence of DNA, we next tested if we can observe "natural" clamping of the LacI binding domain onto straight DNA using unrestrained conventional molecular dynamics simulations. This is important, due to the lack of detailed experimentally derived structural information on the binding of LacI to a nonspecific DNA sequence: is the bending an inherent property of LacI binding to DNA, or is it unique to formation of the specific complex? We therefore constructed and assessed simulations of three random conformations from simulations of free LacI positioned within a feasible binding range of three random snapshots each from simulations of specific and nonspecific DNA sequences. We created an additional three complexes of the LacI crystal structure (PDB ID: 1EFA[10,27]) positioned within binding range of three random snapshots from simulations of specific and nonspecific DNA sequences, leading to six starting complexes per DNA sequence simulated (see Tables S2 and S3).

We simulated three DNA sequences: $O^{SymL}$ and two mutated DNA sequences, $DNA^{mut1}$ and $DNA^{mut2}$, which contain either a more drastic mutation of every A ↔ C and T ↔ G ($DNA^{mut1}$) or a more conservative mutation of every A ↔ G and T ↔ C ($DNA^{mut2}$). These substantial changes were selected because differences in binding affinity between specific and nonspecific DNA sequences are very small,[35−39] so substantial structural perturbations improve our ability to detect selectivity between different DNA sequences.

Starting complexes were generated by taking the conformational snapshots mentioned above and bringing them together at various distances, 18.0 ± 1.6 Å, from the DNA, defined based on distance between center of mass of the binding domain and the DNA, as described in the Methodology section (for comparison, the corresponding distance in the starting crystal structure is 10.6 Å, so this corresponds to displacing the protein ~7.4 Å from the DNA). For examples of such complexes, see Figure 7. These complexes were then used as starting points for subsequent molecular dynamics simulations. The high shape complementarity between the LacI dimer and the bent DNA structure in the specific complex (PDB ID: 1EFA[10,27]) made it impossible to directly juxtapose LacI and straight DNA with the LacI binding clamps pointing toward the DNA, due to steric clashes formed, in particular, between the DNA and the side chains Tyr7, Tyr17, and His29. This affected also the choice of starting complex, as we needed to start from complexes with reasonable shape complementarity (considering also different conformations of the binding clamps).

From these simulations, we make two key observations: (1) even when LacI starts as close as geometrically possible to straight DNA, we do not observe clamping of the binding helices onto DNA, presumably due to the aforementioned steric clash between LacI and the straight DNA. Native-contacts analysis confirmed this, showing that the fraction of native contacts in the loose complexes ranged from 0.0 to 0.4
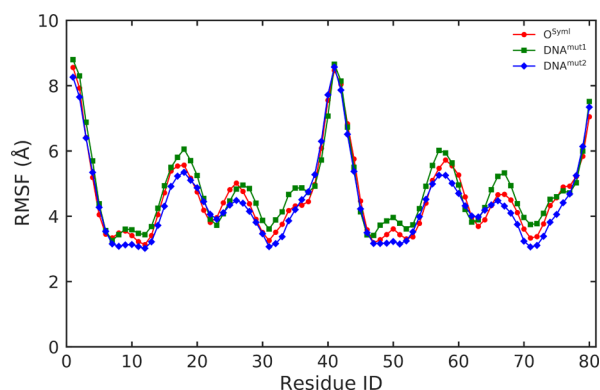


**Figure 6.** RMSF (Å) of all heavy atoms in each nucleotide during the same simulations of free DNA containing the $O^{SymL}$ (red), $DNA^{mut1}$ (green), and $DNA^{mut2}$ (blue) operator sequences, respectively.
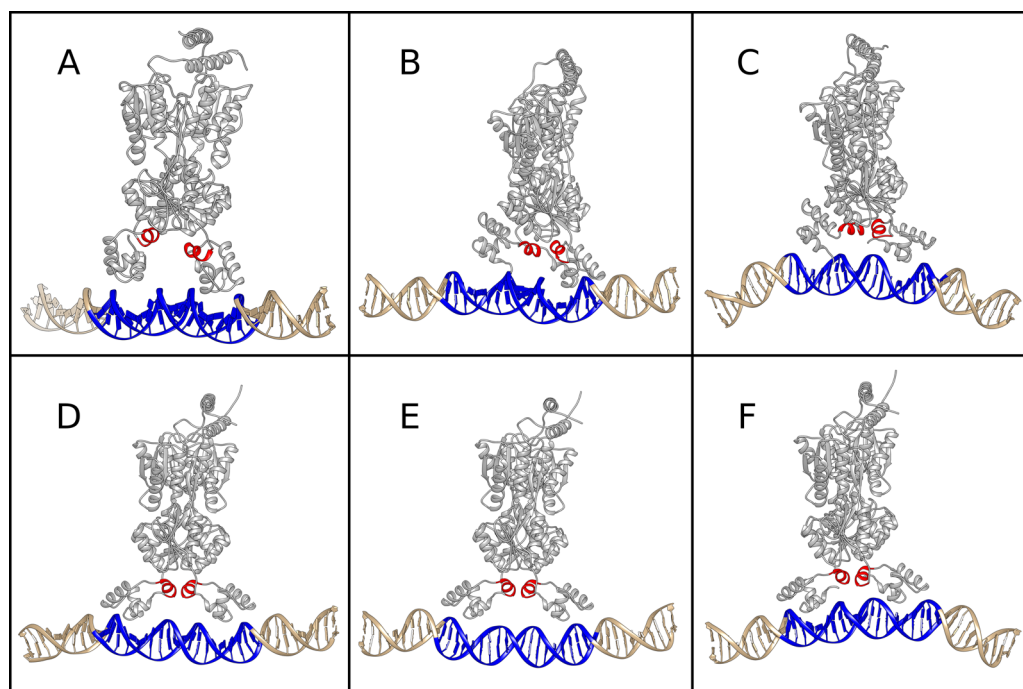
**Figure 7.** Structures of six juxtaposed LacI-DNA "encounter complexes" used as starting conformations for MD simulations of LacI on straight DNA. The central 20 base pairs (corresponding to $O^{SymL}$, $DNA^{mut1}$, and $DNA^{mut2}$) are shown in blue, and the hinge helices of LacI are shown in red. Structures are presented in the same order as in Table S3.

(on a range from 0 to 1, Figure 8) for all systems, relative to the native contacts formed in the initial crystal structure. (2) When the DNA is not bent, we observe what appears to be a natural searching motion along the DNA of 1.0 ± 1.0, 1.6 ± 1.1, and 1.2 ± 0.9 base pairs for the $O^{SymL}$, $DNA^{mut1}$, and $DNA^{mut2}$ systems, respectively, during 6 × 2.0 μs simulations per system (a reference base pair distance of 3.4 Å was used, Figure 9). Taken together, these data strongly suggest that,
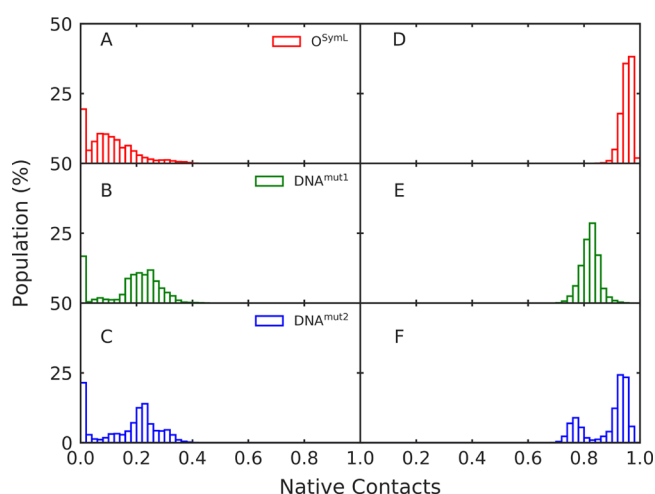


**Figure 8.** Distributions of native contacts during simulations of (A−C) LacI in a loose encounter complex with straight DNA and (D−F) LacI in a tight complex with bent DNA, for the DNA sequences studied in this work. Starting structures were generated as described in the Methodology section. The data shown here for the loose encounter complexes with straight DNA are based on 6 × 2.0 μs simulations of each LacI-DNA complex, while the data shown here for simulations of LacI in complex with bent DNA are based on 3 × 2.5 μs simulations of each LacI-DNA complex (see Table S2).
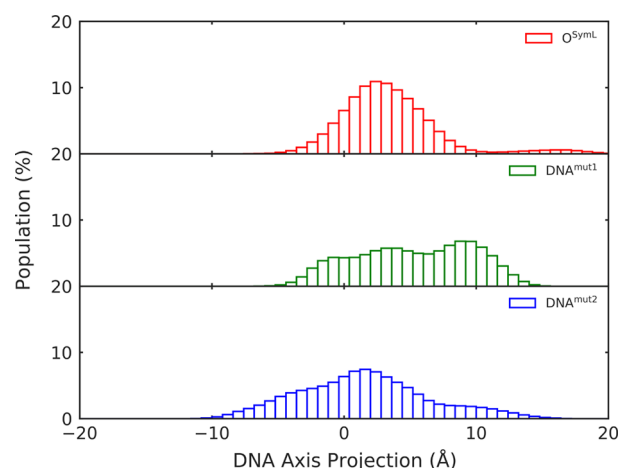


**Figure 9.** Distributions of the projected distance between the center of mass of the binding domain and the DNA binding site along the DNA axial vector, calculated from simulations of LacI in a loose encounter complex with straight DNA, as described in the Methodology section. The data shown here are based on 6 × 2.0 μs simulations of each LacI-DNA complex (see Table S2).

while the search process itself likely occurs on straight DNA, the bending of the DNA itself is an important part of forming a tight-bound specific (and possibly also nonspecific) LacI-DNA complex.

**Modeling Specific and Non-Specific LacI-DNA Interactions on Bent DNA.** As our simulations do not show tight binding between LacI and straight DNA on the time scales sampled, we also performed 3 × 2.5 μs MD simulations per sequence of complexes between LacI and bent DNA, simulating both a specific complex of LacI and $O^{SymL}$ and nonspecific complexes between LacI and $DNA^{mut1}$ and $DNA^{mut2}$. These were again generated by simply mutating

the specific DNA sequence, as described in the Methodology section. As can be seen in Figure 8, the fraction of native contacts during the simulations of LacI in complex with $O^{SymL}$ are almost all higher than 0.9 for all the simulation time, which is quite close to the natively bound state. However, the highly mutated complex (DNA$^{mut1}$) has fractions of native contacts less than 0.9 for 7.4 $\mu$s of the total 7.5 $\mu$s of simulation time, and the protein starts dissociating from the DNA. These shifts of fractional native contacts, while subtle, bring the LacI-DNA$^{mut1}$ complex completely out of the stability range explored by the native LacI-$O^{SymL}$ complex in our simulations. In contrast, the simulations with the sequence with the less radical perturbation, DNA$^{mut2}$, samples fewer conformations with $Q$ values of less than 0.9 than the simulations with DNA$^{mut1}$. In addition, as seen in Figure S18, a small dissociation tendency was observed in the first 0.3 $\mu$s of simulations for all three replicas of LacI-DNA complexes with the DNA$^{mut1}$ sequences, while the native contacts were well-maintained in two of three replicas for DNA$^{mut2}$, with a radical drop in native contacts in the third replica. In all cases, the overall bent shape of the DNA is maintained throughout our simulations, although the hinge helices lose some of their secondary structure in simulations of LacI in complex with DNA$^{mut1}$ and DNA$^{mut2}$ (Figure S19).

One caveat that should be taken into account here is that it is likely that current force fields both over stabilize and over-rigidify "canonical" B-form DNA. This should be mitigated to some extent by the improvements implemented into the bsc1 force field[56] used in this work. It is possible that different results would have been obtained had we, for example, used the OL15 force field[106] instead. However, a comparative study by Galindo-Murillo and co-workers[107] has suggested that both bsc1 and OL15 provide significant advances compared to the bsc0 force field,[108] and they performed similarly in simulations of the Drew-Dickerson dodecamer[109] (DDD), making them both excellent force fields for simulating double-stranded DNA. We note, however, that this study did not consider how well these force fields perform when faced with noncanonical DNA. There are some indications of deformations and also noncanonical $\beta$-angles in the X-ray structures of the transcription factor-DNA complexes. This is, however, challenging to benchmark for in simulations, as the experimental structures may be affected by inaccuracies that stem from data and refinement errors or ambiguities. Despite these caveats, however, our data suggest that LacI spontaneously loosens contact with noncognate DNA sequences on a sub-microsecond time scale. Although it does not fully transition to the "encounter complex" conformation, this loosening does demonstrate DNA sequence-specificity and suggest that simulated $k_{off}$ values will be highly sequence-specific.

The changes in native contacts are coupled with associated changes in energetics, with the average total electrostatic energies for the interaction between the binding domain of LacI and DNA (evaluated using the Debye−Hückel equation as described in the Methodology section) ranging from −11.3 ± 1.9, −11.3 ± 2.3, and −10.8 ± 2.1 kcal mol$^{-1}$ for simulations of the LacI in complex with straight $O^{SymL}$, DNA$^{mut1}$, and DNA$^{mut2}$, and −17.9 ± 1.2, −16.7 ± 1.1, and −17.7 ± 1.2 kcal mol$^{-1}$ for simulations of the LacI in complex with bent $O^{SymL}$, DNA$^{mut1}$, and DNA$^{mut2}$. These differences are slight but statistically significant between all three pairs: $p < 1 \times 10^{-13}$ for $O^{SymL}$ versus DNA$^{mut1}$, $p < 1 \times 10^{-4}$ for $O^{SymL}$ versus DNA$^{mut2}$, and $p < 1 \times 10^{-13}$ for DNA$^{mut1}$ versus DNA$^{mut2}$, via two-sample

$t$-tests with unequal variance and Bonferroni multiple hypothesis corrections. Therefore, the overall average electrostatic energies become more favorable when LacI is in complex with bent DNA compared to straight DNA, and there are some small sequence-specific differences in these values (although within standard deviation), in agreement with the differences in degree of native contacts maintained over the different simulations.

Since our simulations suggest that forming stable LacI-DNA complexes requires highly bent DNA conformations, we tested the propensity of different DNA sequences to bend in such a manner starting from (A) free B-form DNA, (B) loose complexes between LacI and B-form DNA, and (C) LacI-A-form DNA tight complexes. In the case of the free DNA (Figure 10A−C), we obtain similar average values (22.6 ± 11.9° for $O^{sym}$, 22.9 ± 12.0° for DNA$^{mut1}$, and 23.3 ± 12.3° for DNA$^{mut2}$) and distributions for the three DNA sequences, with results that agree very well with the studies of Xiao et al.[105] and Curuksu et al.[78] When the DNA is complexed with LacI, however, the bending angle goes as high as ∼70° (Figure 10D−F). We also sampled some conformations of the DNA (∼0.1%) at high bending angles (>70°) in the simulations of free DNA, but they are not stable, as shown in Curuksu et al.,[78] which could also be a reason that we did not sample enough binding contacts during the simulations of LacI on straight DNA. Finally, as can be seen from this figure, the bent shape of the DNA is maintained slightly better in the $O^{SymL}$ Lac-DNA complex than in the two mutated complexes, based on average bending angles of 69.9 ± 13.5°, 68.6 ± 12.8°, and 59.6 ± 13.8° for $O^{SymL}$, DNA$^{mut1}$, and DNA$^{mut2}$, respectively. These angles are somewhat different from some experimental reports, but the methods of computing angles in those reports were either not reported or differ from our method, so the values are not directly comparable, and if one normalizes to the value calculated for PDB ID: 1L1M,[11,27] the relative changes are similar.[11]

As a final piece of secondary structure analysis, we also calculated the helicity of the hinge helices and the angle between the two hinge helices. As can be seen from Figure
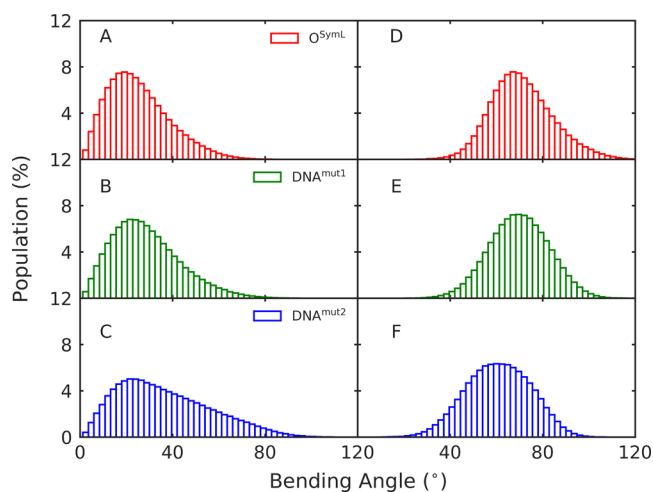


**Figure 10.** Population of the bending angle obtained from (A−C) 9 × 1.0 $\mu$s simulations of free $O^{SymL}$, DNA$^{mut1}$, and DNA$^{mut2}$ and (D−F) 3 × 2.5 $\mu$s simulations of LacI in complex with the same DNA sequences. The bending angle was calculated as described in the Methodology section.
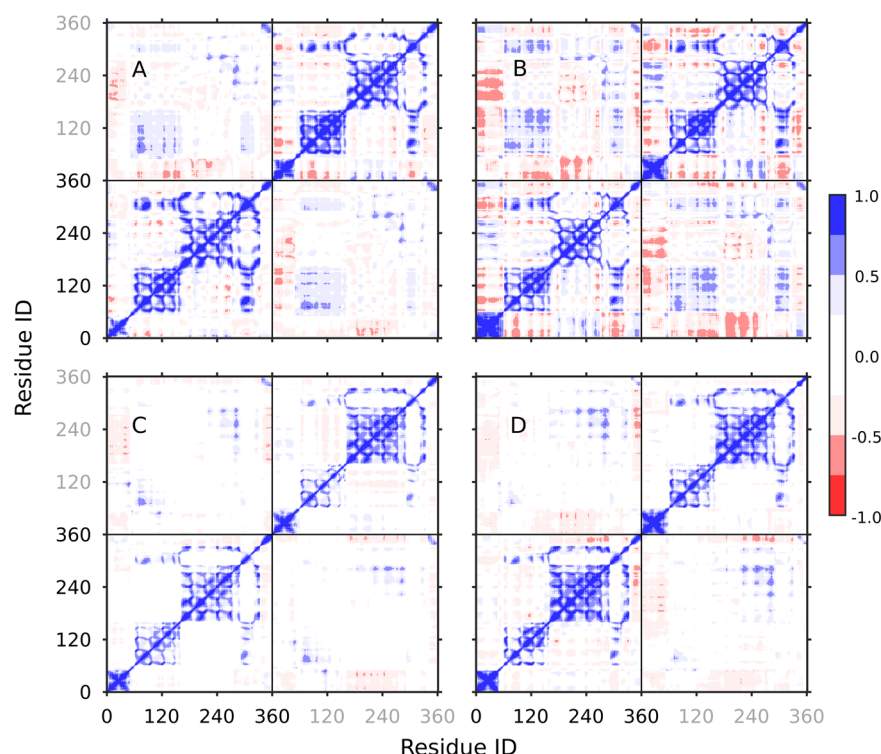
**Figure 11.** Calculated DCCM[110] plots for (A) free LacI (without DNA bound), as well as LacI in complex with bent DNA containing the (B) O$^{SymL}$, (C) DNA$^{mut1}$, or (D) DNA$^{mut2}$ sequences. The DCCM plots were calculated with Bio3D[111] by considering only C$_\alpha$ atoms based on the simulations of $9 \times 1.0$ $\mu s$ of free LacI and $3 \times 2.5$ $\mu s$ of LacI-O$^{SymL}$, LacI-DNA$^{mut1}$, and LacI-DNA$^{mut2}$, respectively. The residue ID for the first monomer are colored in black, and those for the second monomer are colored in gray. The bottom left and top right squares of each DCCM plot show correlated motions within each individual monomer of the dimer, and the top left and bottom right squares of each DCCM plot show correlated motions between the two monomers of the dimer.
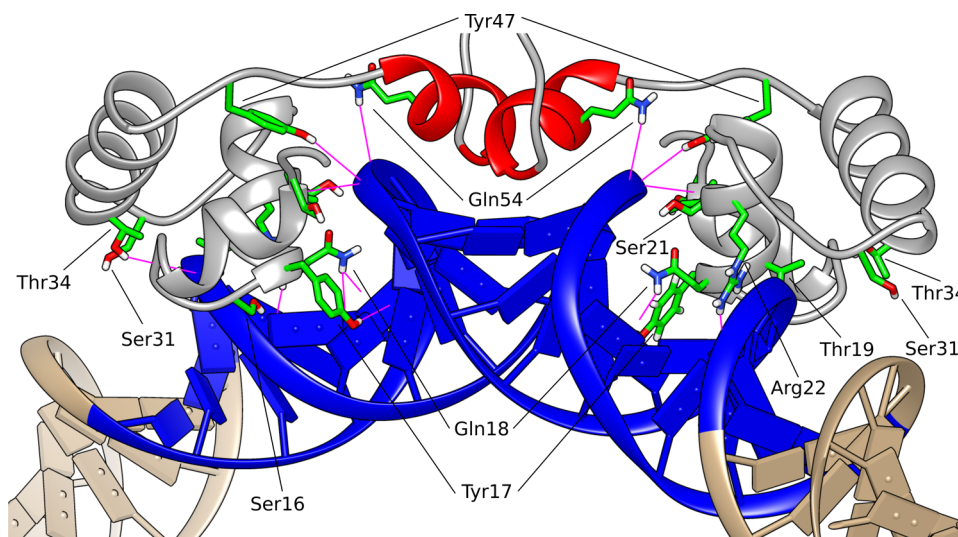


**Figure 12.** Key residues of the binding domain of LacI that form hydrogen bonds with the O$^{SymL}$ operator sequence, based on PDB ID: 1EFA.[10,27] The operator sequence was extended to a 40 base-pair DNA sequence as described in the Methodology section. The original O$^{SymL}$ operator sequence is shown in blue, and the extension of the DNA is shown in tan.

S19, the loss of the hinge helicity in all three complexed systems is smaller than in free LacI, as presented in Figure 3B, and the hinge helicity was better preserved in the LacI-O$^{SymL}$ system than in the two mutated complexes, as shown in the panels A−C of Figure S19.

Following from this, we also created dynamic cross-correlation maps[110] (DCCM) to examine correlated motions both within the two individual LacI monomers and between the two monomers. These data are shown in Figure 11, which shows correlated motions in (A) free LacI (with no DNA bound) and LacI in complex with DNA containing the (B) O$^{SymL}$, (C) DNA$^{mut1}$, and (D) DNA$^{mut2}$ sequences. The bottom left and top right squares of each DCCM plot show correlated motions within each individual monomer of the dimer, and the top left and bottom right squares of each DCCM plot show correlated motions between the two

monomers of the dimer. From these data, it can be seen that LacI behaves similarly in the absence of DNA and when complexed with $DNA^{mut1}$ or $DNA^{mut2}$, with little presence of either correlated or anticorrelated motions. In contrast, these motions become highly correlated (or anticorrelated) when LacI is in complex with DNA containing the $O^{SymL}$ operator sequence. This suggests sequence-specific changes in LacI dynamics,[88] which could (at least partially) account for the differences in DNA recognition by LacI.

Finally, we analyzed differences in hydrogen-bonding interactions between the binding domain of LacI and the DNA, using a donor–acceptor distance cutoff of 3.0 Å and a donor-hydrogen-acceptor angle cutoff of 135°. Hydrogen bonds that occurred in less than 30% of the simulation frames were discarded from further analysis. The position of the key residues forming strong hydrogen bonds during our simulations of LacI in complex with $O^{SymL}$ (specifically, Leu6, Ser16, Thr19, Thr34, and Tyr47) are shown in Figure 12. On the basis of this analysis, the average total number of hydrogen bonds over our simulations drops from 28.8 ± 3.3, in the case of the complex with $O^{SymL}$, to 26.0 ± 3.1 in the case of the complex with $DNA^{mut1}$ and to 26.3 ± 3.3 in the case of the complex with $DNA^{mut2}$. As with the previous case of the electrostatic energies, these differences are again slight but statistically significant: $p < 1 \times 10^{-13}$ for $O^{SymL}$ versus $DNA^{mut1}$, $p < 1 \times 10^{-6}$ for $O^{SymL}$ versus $DNA^{mut2}$, while $DNA^{mut1}$ and $DNA^{mut2}$ are not significantly different, via two-sample $t$-tests with unequal variance and Bonferroni multiple hypothesis corrections. As mentioned in the Introduction, the difference between specific and nonspecific binding in LacI[33,34] can be as large as at least 7 kcal $mol^{-1}$ (although the differences in other DNA-binding proteins are much smaller, in the range of 2–3 kcal $mol^{-1}$).[35–39] Here, the consistent loss or gain of up to two hydrogen bonds as an average property over the simulation time can be sufficient to account for these differences.

We also examined the changes in average occupancy of these key hydrogen bonds during our simulations (Table S14), as well as the time-evolution of interaction distances at the protein–DNA interface involving these key residues (Figures S20–S24), following the detailed analysis performed by Krepl et al. in the analogous case of protein-RNA interactions.[112] These data indicate that the interface is highly dynamic, with significant differences in H-bond occupancies not just between different sequences but also between different monomers of the LacI dimer. The same holds for the distance analysis, where we also see differences between individual replicas (including in terms of comparison to the initial crystal structure). Of particular interest is the behavior of two key residues: Tyr17 and Gln18. As shown in Figure 12, these residues sit in the major groove of the DNA once the LacI-DNA complex is formed, and these interactions seem to be significantly affected in the simulations with the mutated sequences compared to the simulations with the $O^{SymL}$ operator sequence. It is steric clashing with the Tyr17 side chain that is partly responsible for the inability of LacI to form a "clamping" interaction with straight DNA, and we hypothesize that these residues (and, in particular, Tyr17) are critical for allowing the DNA bending that appears to be necessary for forming a tight-bound LacI-DNA complex. Tying in with this, as can be seen from Figure S24, the distance between the Tyr47 and Gln54 side chains is very stable in simulations of complexes of LacI with $O^{SymL}$, but the Gln54–DNA interaction appears to be weakened in

simulations with the mutated DNA sequences. This residue sits on the hinge helices (Figure 12), and therefore the weakening of this interaction could be related to the associated reduction in hinge helicity during simulations of the mutated sequences, shown in Figure S19.

## ■ CONCLUSIONS

Despite extensive prior studies (5, 9, 14, 24, 29, 32, 34, 41, 42, 46–50, 73, 87, 113–115), the precise molecular determinants of LacI binding selectivity, as well as the kinetic factors that allow LacI to rapidly search for cognate DNA sequences, remain incompletely understood. In the present work, we have studied the binding of the LacI dimer to the $O^{SymL}$ operator sequence, as well as two heavily mutated DNA sequences ($DNA^{mut1}$ and $DNA^{mut2}$). Our data yield key differences in the conformational equilibria of LacI that may help explain these properties. Specifically, we observe that the DNA-binding domains of LacI are substantially more conformationally flexible in the absence than in the presence of DNA and that these domains are also oriented differently with respect to each other. The precise functional consequence of these changes for LacI-DNA association kinetics, however, remains unknown. In addition, it appears that unbound DNA is predominantly in a straight B-form conformation, while LacI binding is only stable to bent A-form DNA in our simulations (in good agreement with available LacI-DNA structural data). Two alternate hypotheses that explain this finding are that spontaneous conformational fluctuations of DNA sequences are required for LacI binding, or that some (as yet unidentified) external factor drives this change. However, our simulations suggest that such a conformational change is indeed required for DNA binding.

Our simulations also suggest differences in the conformational equilibria and binding stability of LacI to bent DNA of either the $O^{SymL}$ or the mutated sequences that may account for the experimentally observed selective binding toward $O^{SymL}$. We note that the selectivity of DNA binding proteins for specific versus nonspecific binding is on the order of 2–3 kcal $mol^{-1}$,[35–39] although a larger number in the range of 7 kcal $mol^{-1}$ is expected for the LacI dimer.[33,34] Our simulations suggest that the LacI-$O^{SymL}$ complex has on average 1–2 more hydrogen bonds between LacI and the DNA on our simulation time scales, which would be sufficient to explain even the up to ~7 kcal $mol^{-1}$ differences in binding selectivity toward different operator sequences versus nonspecific binding.[23,33,34] Furthermore, LacI shows substantially different conformational equilibria when bound to the $O^{SymL}$ sequence than in either the DNA free form or in complex with either $DNA^{mut1}$ or $DNA^{mut2}$ (which in turn resemble free LacI rather than LacI bound to an operator sequence). This suggests that the difference in conformational equilibria observed (quantified as correlated positional displacement across residues) can at least report on selective binding and potentially illuminate the mechanistic changes involved in selectivity.

In summary, this study thus yields the following hypotheses regarding LacI binding: binding occurs primarily to bent A-form DNA, involves a loss of LacI conformational entropy, and increases anticorrelated conformational deviations across the protein, reflecting a concerted change rather than several independent ones. Simultaneously, binding to the $O^{SymL}$ operator sequence involves a slightly larger number of stabilizing DNA–protein hydrogen bonds compared to binding to $DNA^{mut1}$ or $DNA^{mut2}$, which may account for the experimentally observed specificity for this operator sequence.

These concrete hypotheses are experimentally testable, and they may help provoke further simulation and experimental studies of the binding kinetics of LacI to partner DNA, a process which still remains out of the reach of unbiased atomistic molecular dynamics simulation.

## ■ ASSOCIATED CONTENT

**ⓢ Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b12363.

Additional structural and dynamical analysis of our simulations (PDF)

## ■ AUTHOR INFORMATION

**Corresponding Author**

*E-mail: lynn.kamerlin@kemi.uu.se.

**ORCID** ⓞ

Shina Caroline Lynn Kamerlin: 0000-0002-3190-1173

**Notes**

The authors declare no competing financial interest.

**Biography**

Lynn Kamerlin is currently a Professor of Structural Biology at Uppsala University, where she has been a member of the faculty since 2011. She received her doctoral training at the University of Birmingham (2002−2005), under the supervision of Dr. John Wilkie, and was subsequently a postdoctoral researcher in the laboratories of Stefan Boresch (University of Vienna, 2005−2007) and Arieh Warshel (University of Southern California, 2007−2010). She is a Fellow of the Royal Society of Chemistry and a Wallenberg Academy Fellow, as well as having been a recipient of an ERC Starting Independent Researcher Grant, as well as Chair of the Young Academy of Europe (2014−2015). She also serves on the Editorial Advisory Boards of the American Chemical Society journals *ACS Catalysis*, *The Journal of Physical Chemistry*, *Journal of Chemical Information and Modeling*, and *ACS Omega*, as well as the Editorial Board of *Electronic Structure* (IOP Publishing).

## ■ ACKNOWLEDGMENTS

## ■ REFERENCES

(1) Latchman, D. S. Transcription Factors: An Overview. *Int. J. Exp. Pathol.* **1993**, *74*, 417−422.

(2) Halford, S. E.; Marko, J. F. How do Site-Specific DNA-Binding Proteins Find Their Targets? *Nucleic Acids Res.* **2004**, *32*, 3040−3052.

(3) Halford, S. E. An End to 40 years of Mistakes in DNA−protein Association Kinetics? *Biochem. Soc. Trans.* **2009**, *37*, 343−348.

(4) Zakrzewska, K.; Lavery, R. Towards a Molecular View of Transcriptional Control. *Curr. Opin. Struct. Biol.* **2012**, *22*, 160−167.

(5) Mahmutovic, A.; Berg, O. G.; Elf, J. What Matters for Lac Repressor Search *in vivo* - Sliding, Hopping, Intersegment Transfer, Crowding on DNA or Recognition? *Nucleic Acids Res.* **2015**, *43*, 3454−3464.

(6) Winkler, F. K.; Banner, D. W.; Oefner, C.; Tsernoglou, D.; Brown, R. S.; Heathman, S. P.; Bryan, R. K.; Martin, P. D.; Petratos, K.; Wilson, K. S. The Crystal Structure of EcoRV Endonuclease and of Its Complexes with Cognate and Non-cognate DNA Fragments. *EMBO J.* **1993**, *12*, 1781−1795.

(7) Viadiu, H.; Aggarwal, A. K. Structure of BamHI Bound to Nonspecific DNA: A Model for DNA Sliding. *Mol. Cell* **2000**, *5*, 889−895.

(8) Stayrook, S.; Jaru-Ampornpan, P.; Ni, J.; Hochschild, A.; Lewis, M. Crystal Structure of the $\lambda$ Repressor and a Model for Pairwise Cooperative Operator Binding. *Nature* **2008**, *452*, 1022−1025.

(9) Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P. Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. *Science* **1996**, *271*, 1247−1254.

(10) Bell, C. E.; Lewis, M. A Closer View of the Conformation of the Lac Repressor Bound to Operator. *Nat. Struct. Biol.* **2000**, *7*, 209−214.

(11) Kalodimos, C. G.; Bonvin, A. M. J. J.; Salinas, R. K.; Wechselberger, R.; Boelens, R.; Kaptein, R. Plasticity in Protein−DNA Recognition: Lac Repressor Interacts with Its Natural Operator O1 Through Alternative Conformations of its DNA-Binding Domain. *EMBO J.* **2002**, *21*, 2866−2876.

(12) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *305*, 386−389.

(13) Kalodimos, C. G.; Boelens, R.; Kaptein, R. Toward an Integrated Model of Protein−DNA Recognition as Inferred from NMR Studies on the Lac Repressor System. *Chem. Rev.* **2004**, *104*, 3567−3586.

(14) Wilson, C. J.; Zhan, H.; Swint-Kruse, L.; Matthews, K. S. The Lactose Repressor System: Paradigms for Regulation, Allosteric Behavior and Protein Folding. *Cell. Mol. Life Sci.* **2007**, *64*, 3−16.

(15) Dunaway, M.; Matthews, K. S. Hybrid Tetramers of Native and Core Lactose Repressor Protein. Assessment of Operator and Nonspecific DNA Binding Parameters and Their Relationship. *J. Biol. Chem.* **1980**, *255*, 10120−10127.

(16) Winter, R. B.; Berg, O. G.; von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 3. The *Escherichia coli lac* Repressor–Operator Interaction: Kinetic Measurements and Conclusions. *Biochemistry* **1981**, *20*, 6961−6977.

(17) Winter, R. B.; von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The *Escherichia coli* Repressor-Operator Interaction: Equilibrium Measurements. *Biochemistry* **1981**, *20*, 6948−6960.

(18) Kleina, L. G.; Miller, J. H. Genetic Studies of the lac Repressor. XIII. Extensive Amino Acid Replacements Generated by the Use of Natural and Synthetic Nonsense Suppressors. *J. Mol. Biol.* **1990**, *212*, 295−318.

(19) Markiewicz, P.; Kleina, L. G.; Cruz, C.; Ehret, S.; Miller, J. H. Genetic Studies of the *lac* Repressor. XIV. Analysis of 4000 Altered *Escherichia coli lac* Repressors Reveals Essential and Non-Essential Residues, as well as "Spacers" Which do not Require a Specific Sequence. *J. Mol. Biol.* **1994**, *240*, 421−433.

(20) Friedman, A. M.; Fischmann, T. O.; Steitz, T. A. Crystal Structure of *Lac* Repressor Core Tetramer and its Implications for DNA Looping. *Science* **1995**, *268*, 1721−1727.

(21) Gilbert, W.; Müller-Hill, B. Isolation of the *lac* Repressor. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *56*, 1891−1898.

(22) Daber, R.; Stayrook, S.; Rosenberg, A.; Lewis, M. Structural Analysis of *lac* Repressor Bound to Allosteric Effectors. *J. Mol. Biol.* **2007**, *370*, 609−619.

(23) Romanuka, J.; Folkers, G. E.; Biris, N.; Tishchenko, E.; Wienk, H.; Bonvin, A. M. J. J.; Kaptein, R.; Boelens, R. Specificity and Affinity of Lac Repressor for the Auxiliary Operators O2 and O3 Are Explained by the Structures of Their Protein - DNA Complexes. *J. Mol. Biol.* **2009**, *390*, 478−489.

(24) Swint-Kruse, L.; Matthews, K. S. Allostery in the LacI/GalR Family: Variations on a Theme. *Curr. Opin. Microbiol.* **2009**, *12*, 129−137.

(25) Lewis, M. The Lac Repressor. *C. R. Biol.* **2005**, *328*, 521−548.

(26) Balaeff, A.; Mahadevan, L.; Schulten, K. Structural Basis for Cooperative DNA Binding by CAP and Lac Repressor. *Structure* **2004**, *12*, 123−132.

(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(28) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(29) Sadler, J. R.; Sasmor, H.; Betz, J. L. A Perfectly Symmetric Lac Operator Binds the Lac Repressor Very Tightly. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 6785−6789.

(30) Straney, S. B.; Crothers, D. M. Lac Repressor Is a Transient Gene - Activating Protein. *Cell* **1987**, *51*, 699−707.

(31) Oehler, S.; Eismann, E. R.; Krämer, H.; Müller-Hill, B. The Three Operators of the Lac Operon Cooperate in Repression. *EMBO J.* **1990**, *9*, 973−979.

(32) Oehler, S. Feedback Regulation of LacI Repressor Expression in *Escherichia coli*. *J. Bacteriol.* **2009**, *191*, 5301−5303.

(33) Hammar, P.; Leroy, P.; Mahmutovic, A.; Marklund, E. G.; Berg, O. G.; Elf, J. The *lac* Repressor Displays Facilitated Diffusion in Living Cells. *Science* **2012**, *336*, 1595−1598.

(34) Elf, J.; Li, G.-W.; Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecular Level in a Living Cell. *Science* **2007**, *316*, 1191−1194.

(35) Oda, M.; Furukawa, K.; Ogata, K.; Sarai, A.; Nakamura, H. Thermodynamics of Specific and Non-Specific DNA Binding by the c-myb DNA-Binding Domain. *J. Mol. Biol.* **1998**, *276*, 571−590.

(36) Holbrook, J. A.; Tsodikov, O. V.; Saecker, R. M.; Record, M. T. Specific and Non-Specific Interactions of Integration Host Factor with DNA: Thermodynamic Evidence for Disruption of Multiple IHF Surface Salt-bridges Coupled to DNA Binding. *J. Mol. Biol.* **2001**, *310*, 379−401.

(37) Zhang, Y.; Larsen, C. A.; Stadler, H. S.; Ames, J. B. Structural Basis for Sequence Specific DNA Binding and Protein Dimerization of HOXA13. *PLoS One* **2011**, *6*, e23069.

(38) Pinto, U. M.; Flores-Mireles, A. L.; Costa, E. D.; Winans, S. C. RepC Protein of the Octopine-Type Ti Plasmid Binds to the Probable Origin of Replication within repC and Functions Only in cis. *Mol. Microbiol.* **2011**, *81*, 1593−1606.

(39) Wieczór, M.; Czub, J. How Proteins Bind to DNA: Target Discrmination and Dynamic Sequence Search by the Telomeric Protein TRF1. *Nucleic Acids Res.* **2017**, *45*, 7643−7654.

(40) Oehler, S.; Amouyal, M.; Kolkhof, P.; von Wilcken-Bergmann, B.; Müller-Hill, B. Quality and Position of the Three Lac Operators of *E. coli* Define Efficiency of Repression. *EMBO J.* **1994**, *13*, 3348−3355.

(41) Villa, E.; Balaeff, A.; Schulten, K. Structural Dynamics of the *lac* Repressor−DNA Complex Revealed by a Multiscale Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6783−6788.

(42) Machado, M. R.; Pantano, S. Exploring LacI−DNA Dynamics by Multiscale Simulations Using the SIRAH Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 5012−5023.

(43) Furini, S.; Domene, C.; Cavalcanti, S. Insights into the Sliding Movement of the Lac Repressor Nonspecifically Bound to DNA. *J. Phys. Chem. B* **2010**, *114*, 2238−2245.

(44) Barr, D.; van der Vaart, A. The Natural DNA Bending Angle in the Lac Repressor Headpiece−O1 Operator Complex is Determined by Protein−DNA Contacts and Water Release. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2070−2077.

(45) Yonetani, Y.; Kono, H. Dissociation Free-Energy Profiles of Specific and Nonspecific DNA−Protein Complexes. *J. Phys. Chem. B* **2013**, *117*, 7535−7545.

(46) Marklund, E. G.; Mahmutovic, A.; Berg, O. G.; Hammar, P.; van der Spoel, D.; Fange, D.; Elf, J. Transcription-Factor Binding and Sliding on DNA Studied Using Micro- and Macroscopic Models. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 19796−19801.

(47) Furini, S.; Barbini, P.; Domene, C. DNA-Recognition Process Described by MD Simulations of the Lactose Repressor Protein on a Specific and a Non-Specific DNA Sequence. *Nucleic Acids Res.* **2013**, *41*, 3963−3972.

(48) Sun, L.; Tabaka, M.; Hou, S.; Li, L.; Burdzy, K.; Aksimentiev, A.; Maffeo, C.; Zhang, X.; Holyst, R. The Hinge Region Strengthens the Nonspecific Interaction between Lac-Repressor and DNA: A Computer Simulation Study. *PLoS One* **2016**, *11*, e0152002.

(49) Hsieh, M.; Brenowitz, M. Comparison of the DNA Association Kinetics of the Lac Repressor Tetramer, Its Dimeric Mutant LacI$^{adi}$, and the Native Dimeric Cal Repressor. *J. Biol. Chem.* **1997**, *272*, 22092−22096.

(50) Swint-Kruse, L.; Elam, C. R.; Lin, J. W.; Wycuff, D. R.; Shive Matthews, K. Plasticity of Quaternary Structure: Twenty-Two Ways to Form a LacI Dimer. *Protein Sci.* **2001**, *10*, 262−276.

(51) Zgarbová, M.; Otyepka, M.; Šponer, J.; Lankaš, F.; Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177−3189.

(52) Zheng, G.; Lu, X.-J.; Olson, W. K. Web 3DNA - A Web Server for the Analysis, Reconstruction, and Visualization of Three-dimensional Nucleic-Acid Structures. *Nucleic Acids Res.* **2009**, *37*, W240−W246.

(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(54) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; et al. *AMBER 2016;* University of California: San Francisco, CA, 2016.

(55) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(56) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13*, 55−58.

(57) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.

(58) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise - A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.

(59) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(60) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(61) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules

with a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384*, 288−294.

(62) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press, 1989.

(63) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(64) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An *N*·log(*N*) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(65) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(66) van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(67) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(68) Best, R. B.; Hummer, G.; Eaton, W. A. Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.

(69) Do, T. N.; Carloni, P.; Varani, G.; Bussi, G. RNA/Peptide Binding Driven by Electrostatics—Insight from Bidirectional Pulling Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1720−1730.

(70) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(71) Zhou, C.-Y.; Jiang, F.; Wu, Y.-D. Folding Thermodynamics and Mechanism of Five Trp-Cage Variants from Replica-Exchange MD Simulations with RSFF2 Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 5473−5480.

(72) Meshkin, H.; Zhu, F. Thermodynamics of Protein Folding Studied by Umbrella Sampling along a Reaction Coordinate of Native Contacts. *J. Chem. Theory Comput.* **2017**, *13*, 2086−2097.

(73) Wang, J.; Wang, Y.; Chu, X.; Hagen, S. J.; Han, W.; Wang, E. Multi-Scaled Explorations of Binding-Induced Folding of Intrinsically Disordered Protein Inhibitor IA3 to its Target Enzyme. *PLoS Comput. Biol.* **2011**, *7*, e1001118.

(74) Ithuralde, R. E.; Roitberg, A. E.; Turjanski, A. G. Structured and Unstructured Binding of an Intrinsically Disordered Protein as Revealed by Atomistic Simulations. *J. Am. Chem. Soc.* **2016**, *138*, 8742−8751.

(75) Liao, Q.; Kulkarni, Y.; Sengupta, U.; Petrović, D.; Mulholland, A. J.; van der Kamp, M. W.; Strodel, B.; Kamerlin, S. C. L. Loop Motion in Triosephosphate Isomerase Is Not a Simple Open and Shut Case. *J. Am. Chem. Soc.* **2018**, *140*, 15889−15903.

(76) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.

(77) Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G. A Series of PDB-related Databanks for Everyday Needs. *Nucleic Acids Res.* **2015**, *43*, D364−D368.

(78) Curuksu, J.; Zakrzewska, K.; Zacharias, M. Magnitude and Direction of DNA Bending Induced by Screw-axis Orientation: Influence of Sequence, Mismatches and Abasic Sites. *Nucleic Acids Res.* **2008**, *36*, 2268−2283.

(79) Waters, J. T.; Lu, X.-J.; Galindo-Murillo, R.; Gumbart, J. C.; Kim, H. D.; Cheatham, T. E.; Harvey, S. C. Transitions of Double-Stranded DNA Between the A- and B-Forms. *J. Phys. Chem. B* **2016**, *120*, 8449−8456.

(80) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(81) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(82) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.

(83) Spronk, C. A. E. M.; Folkers, G. E.; Noordman, A. M. G. W.; Wechselberger, R.; van den Brink, N.; Boelens, R.; Kaptein, R. Hinge-helix Formation and DNA Bending in Various Lac Repressor-operator Complexes. *EMBO J.* **1999**, *18*, 6472−6480.

(84) Harrison, S. C. A. Structural Taxonomy of DNA-binding Domains. *Nature* **1991**, *353*, 715−719.

(85) Brennan, R. J.; Matthews, B. W. The Helix-Turn-Helix DNA Binding Motif. *J. Biol. Chem.* **1989**, *264*, 1903−1906.

(86) Sauer, R. T. Lac Repressor at Last. *Structure* **1996**, *4*, 219−222.

(87) Xu, J. S.; Hewitt, M. N.; Gulati, J. S.; Cruz, M. A.; Zhan, H.; Liu, S.; Matthews, K. S. Lactose Repressor Hinge Domain Independently Binds DNA. *Protein Sci.* **2018**, *27*, 839−847.

(88) Xu, L.; Ye, W.; Jiang, C.; Yang, J.; Zhang, J.; Feng, Y.; Luo, R.; Chen, H.-F. Recognition Mechanism between Lac Repressor and DNA with Correlation Network Analysis. *J. Phys. Chem. B* **2015**, *119*, 2844−2856.

(89) Blanchet, C.; Pasi, M.; Zakrzewska, K.; Lavery, R. CURVES+ Web Server for Analyzing and Visualizing the Helical, Backbone and Groove Parameters of Nucleic Acid Structures. *Nucleic Acids Res.* **2011**, *39*, W68−W73.

(90) Hegde, R. S.; Grossman, S. R.; Laimins, L. A.; Sigler, P. B. Crystal Structure at 1.7 Å of the Bovine Papillomavirus-1 E2 DNA-Binding Domain Bound to its DNA Target. *Nature* **1992**, *359*, 505−512.

(91) Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. Crystal Structure of a Yeast TBP/TATA-box Complex. *Nature* **1993**, *365*, 512−520.

(92) Kim, J. L.; Nikolov, D. B.; Burley, S. K. Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element. *Nature* **1993**, *365*, 520−527.

(93) Garvie, C. W.; Wolberger, C. Recognition of Specific DNA Sequences. *Mol. Cell* **2001**, *8*, 937−946.

(94) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.

(95) Harteis, S.; Schneider, S. Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions. *Int. J. Mol. Sci.* **2014**, *15*, 12335−12363.

(96) Zgarbová, M.; Jurečka, P.; Sponer, J.; Otyepka, M. A- to B-DNA Transition in AMBER Force Fields and Its Coupling to Sugar Pucker. *J. Chem. Theory Comput.* **2018**, *14*, 319−328.

(97) Wood, B. R. The Importance of Hydration and DNA Conformation in Interpreting Infrared Spectra of Cells and Tissues. *Chem. Soc. Rev.* **2016**, *45*, 1980−1998.

(98) Ivanov, V. I.; Minchenkova, L. E.; Minyat, E. E.; Frank-Kamenetskii, M. D.; Schyolkina, A. K. The B to A Transition of DNA in Solution. *J. Mol. Biol.* **1974**, *87*, 817−833.

(99) Egli, M.; Tereshko, V.; Teplova, M.; Minasov, G.; Joachimiak, A.; Sanishvili, R.; Weeks, C. M.; Miller, R.; Maier, M. A.; An, H.; et al. X-ray Crystallographic Analysis of the Hydration of A- and B-form DNA at Atomic Resolution. *Biopolymers* **1998**, *48*, 234−252.

(100) Jose, D.; Porschke, D. The Dynamics of the B−A Transition of Natural DNA Double Helices. *J. Am. Chem. Soc.* **2005**, *127*, 16120−16128.

(101) Zhang, H.; Fu, H.; Shao, X.; Dehez, F.; Chipot, C.; Cai, W. Changes in Microenvironment Modulate the B- to A-DNA Transition. *J. Chem. Inf. Model.* **2019**. In Press. DOI: 10.1021/acs.jcim.8b00885.

(102) Banavali, N. K.; Roux, B. Free Energy Landscape of A-DNA to B-DNA Conversion in Aqueous Solution. *J. Am. Chem. Soc.* **2005**, *127*, 6866−6876.

(103) Grove, A.; Galeone, A.; Mayol, L.; Geiduschek, P. E. Localized DNA Flexibility Contributes to Target Site Selection by DNA-Bending Proteins. *J. Mol. Biol.* **1996**, *260*, 120−125.

(104) Okonogi, T. M.; Alley, S. C.; Reese, A. W.; Hopkins, P. B.; Robinson, B. H. Sequence-Dependent Dynamics in Duplex DNA. *Biophys. J.* **2000**, *78*, 2560−2571.

(105) Xiao, S.; Zhu, H.; Wang, L.; Liang, H. DNA Conformational Flexibility Study Using Phosphate Backbone Neutralization Model. *Soft Matter* **2014**, *10*, 1045−1055.

(106) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar−Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723−5736.

(107) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114−4127.

(108) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of $\alpha/\gamma$ Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(109) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179−2183.

(110) Lange, O. F.; Grubmüller, H. Generalized Correlation for Biomolecular Dynamics. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 1053−1061.

(111) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3D: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695−2696.

(112) Krepl, M.; Havrila, M.; Stadlbauer, P.; Banas, P.; Otyepka, M.; Pasulka, J.; Stefl, R.; Sponer, J. Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein−RNA Complexes? *J. Chem. Theory Comput.* **2015**, *11*, 1220−1243.

(113) von Hippel, P. H.; Revzin, A.; Gross, C. A.; Wang, A. C. Non-specific DNA Binding of Genome Regulating Proteins as a Biological Control Mechanism: I. The Lac Operon: Equilibrium Aspects. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4808−4812.

(114) Furini, S.; Domene, C. DNA Recognition Process of the Lactose Repressor Protein Studied via Metadynamics and Umbrella Sampling Simulations. *J. Phys. Chem. B* **2014**, *118*, 13059−13065.

(115) Kipper, K.; Eremina, N.; Marklund, E.; Tubasum, S.; Mao, G.; Lehmann, L. C.; Elf, J.; Deindl, S. Structure-Guided Approach to Site-Specific Fluorophore Labeling of the *lac* Repressor LacI. *PLoS One* **2018**, *13*, e0198416.