DNA Recognition Process of the Lactose Repressor Protein Studied via Metadynamics and Umbrella Sampling Simulations

Simone Furini[a,*], and Carmen Domene[b,c]

[a] Department of Medical Biotechnologies, University of Siena, viale Mario Bracci 16, I-53100, Siena, Italy

[b] Department of Chemistry, King's College London, Britannia House, 7 Trinity Street, London SE1 1DB, UK

[c] Chemistry Research Laboratory, Mansfield Road, University of Oxford, Oxford OX1 3TA, UK

[*]Corresponding author:

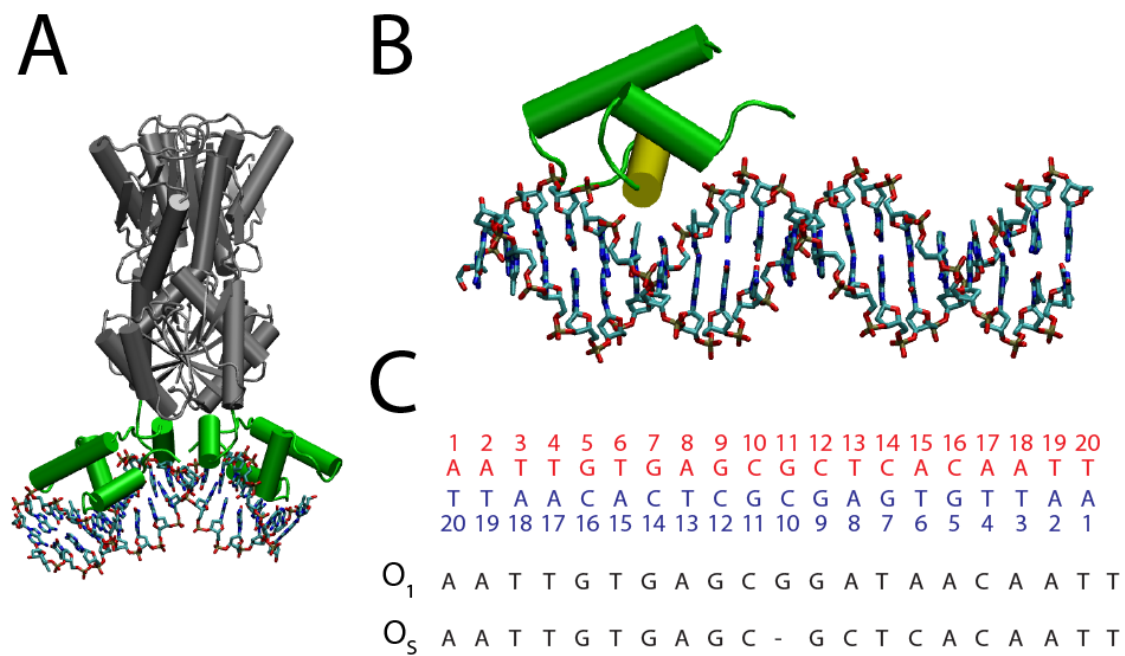Department of Medical Biotechnologies

University of Siena

viale Mario Bracci ,16

Siena, I-53100

Italy

Tel: +390577585730

email: simone.furini@unisi.it

Figure S1. Structure of the protein-DNA complex. (A) X-ray structure [2] (PDB entry 1EFA) of a LacI dimer in complex with DNA at an operator site. Residues 1-62 of the DNA-binding domain are in green. (B) Snapshot of the model system. The recognition helix, residues 16 to 26, is depicted in yellow. (C) DNA sequence of the primary (red) and complementary strand (blue). The sequence of the operator site O_1, and of the totally symmetric operator site O_S are also shown.
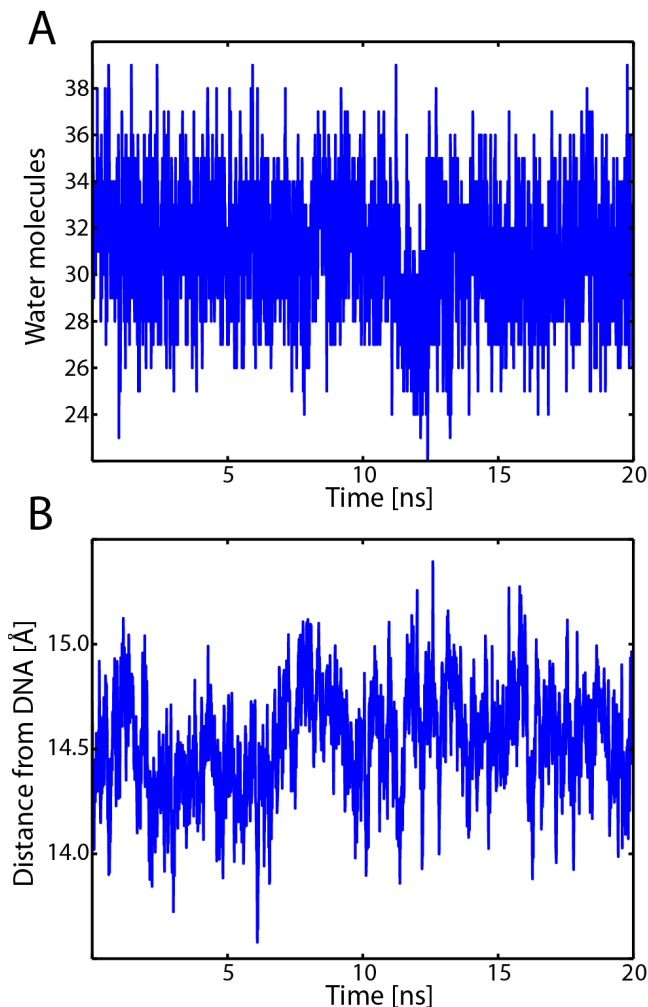
Figure S2. Protein-DNA interface in the equilibration MD trajecotry. (A) Number of water molecules that are closer than 6.0 Å to any atom of the recognition helix (protein residues 16 to 26), and at the same time closer that 6.0 Å to any atoms of the DNA base edges. (B) Distance in the plane orthogonal to the DNA-axis between the center of mass of the protein backbone atoms and the center of mass of the DNA molecule.
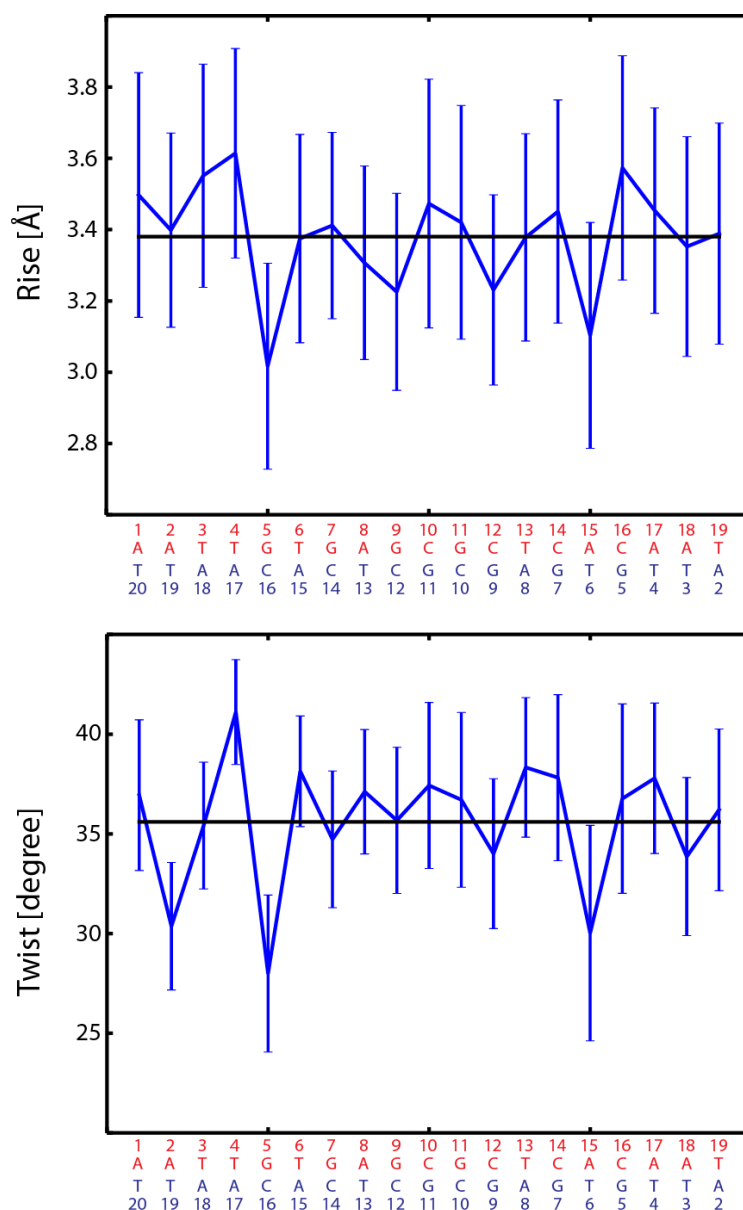
Figure S3. Geometrical parameters of the DNA molecule. (A) Distance between successive base pairs along the DNA axis; (B) Angle between successive base pairs in the plane orthogonal to the DNA-axis. Average values and standard deviations were calculated from a 20 ns unrestrained MD trajectory, taking snapshots every 100 ps. The black line in each graph corresponds to the mean value for the entire DNA fragment of respectively rise and twist. These average values were used to define the pitch of the helical trajectory adopted to analyze protein sliding along DNA in the umbrella sampling simulations.
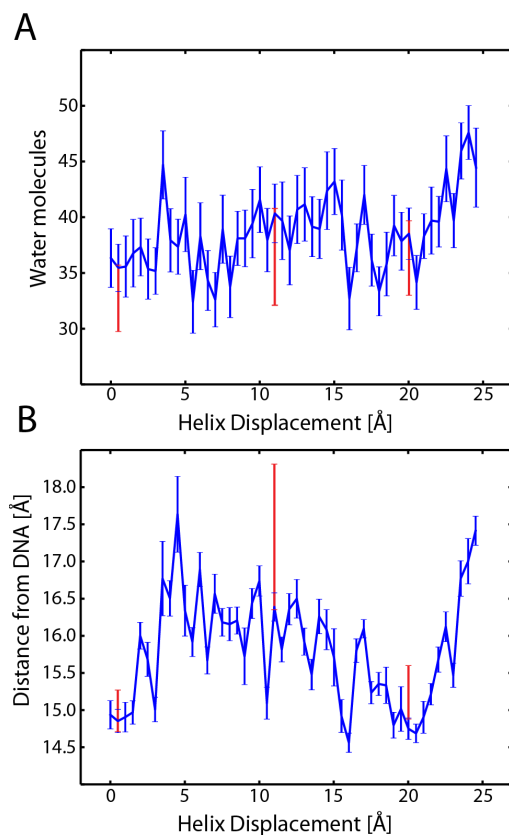
Figure S4. Protein-DNA interface in the umbrella sampling simulations. Average values and standard deviations are shown for each umbrella-sampling trajectory. Values on the x-axis correspond to the center of the harmonic potential of that particular umbrella sampling simulation. (A) Number of water molecules closer than 6.0 Å to any atom of the recognition helix (protein residues 16 to 26), and at the same time closer that 6.0 Å to any atoms of the DNA base edges. (B) Distance between the center of mass of the protein backbone atoms and the center of mass of the DNA molecule in the plane orthogonal to the DNA-axis. The red lines at helical displacements equal to 0.50, 11.01, and 20.02 Å indicate the values calculated from unrestrained MD simulations of 60 ns that started from the last frame of the corresponding umbrella sampling simulations.
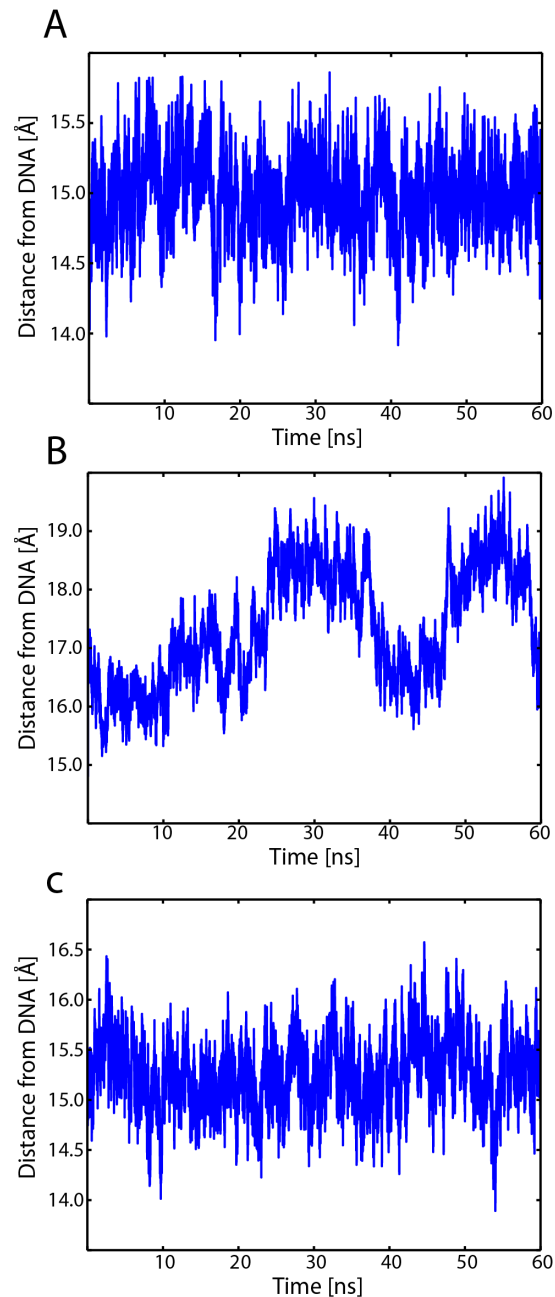
Figure S5. Radial distance between protein and DNA. Distance in the plane orthogonal to the DNA-axis between the center of mass of the protein backbone atoms and the center of mass of the DNA molecule in unrestrained MD simulations that started from the last frame of the umbrella sampling trajectory with center of the harmonic potential closer to the energy minimum O_{1L} (A), $O_{1L}+1$ (B), and $O_{1L}+2$ (C).
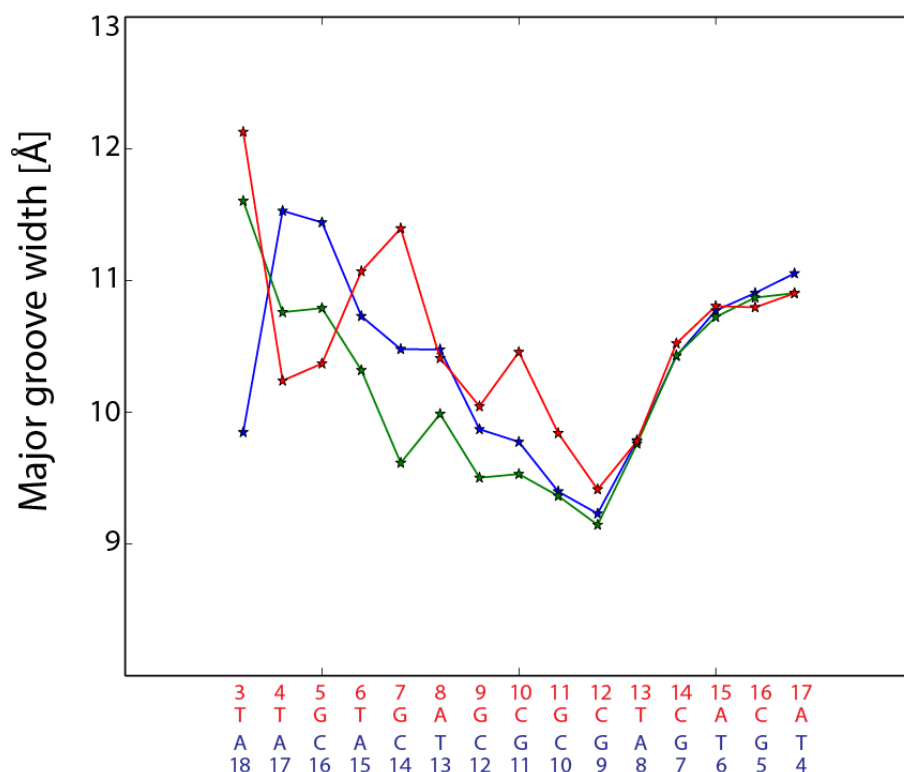
Figure S6. Width of the DNA major groove with protein at different positions along the DNA molecule. The average width values of the major groove were calculated with the software Curves+ using data from unrestrained 60-ns MD simulations that started from the last frame of the umbrella sampling simulations with center of the harmonic potential closer to O_{1L} (blue line), $O_{1L}+1$ (green line), and $O_{1L}+2$ (red line). MD trajectories were sampled every 100 ps.
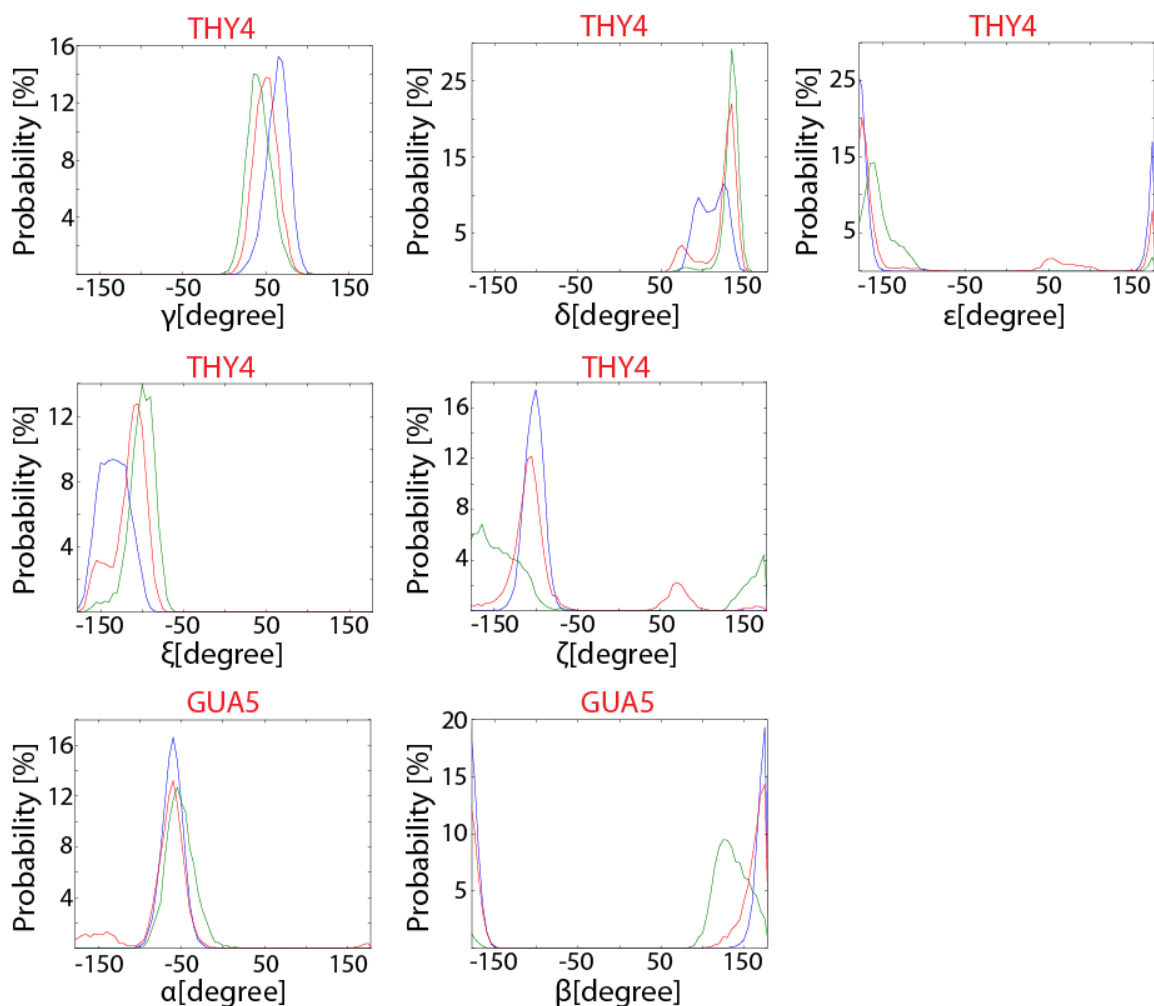
Figure S7. DNA backbone angles when the protein is at different position along the DNA molecule. Data were collected from unrestrained MD simulations that started from the last frame of the umbrella sampling simulations with center of the harmonic potential closer to O_{1L} (blue lines), $O_{1L}+1$ (green lines), and $O_{1L}+2$ (red lines). DNA bases are numbered as defined in Figure S1-C (primary strand).
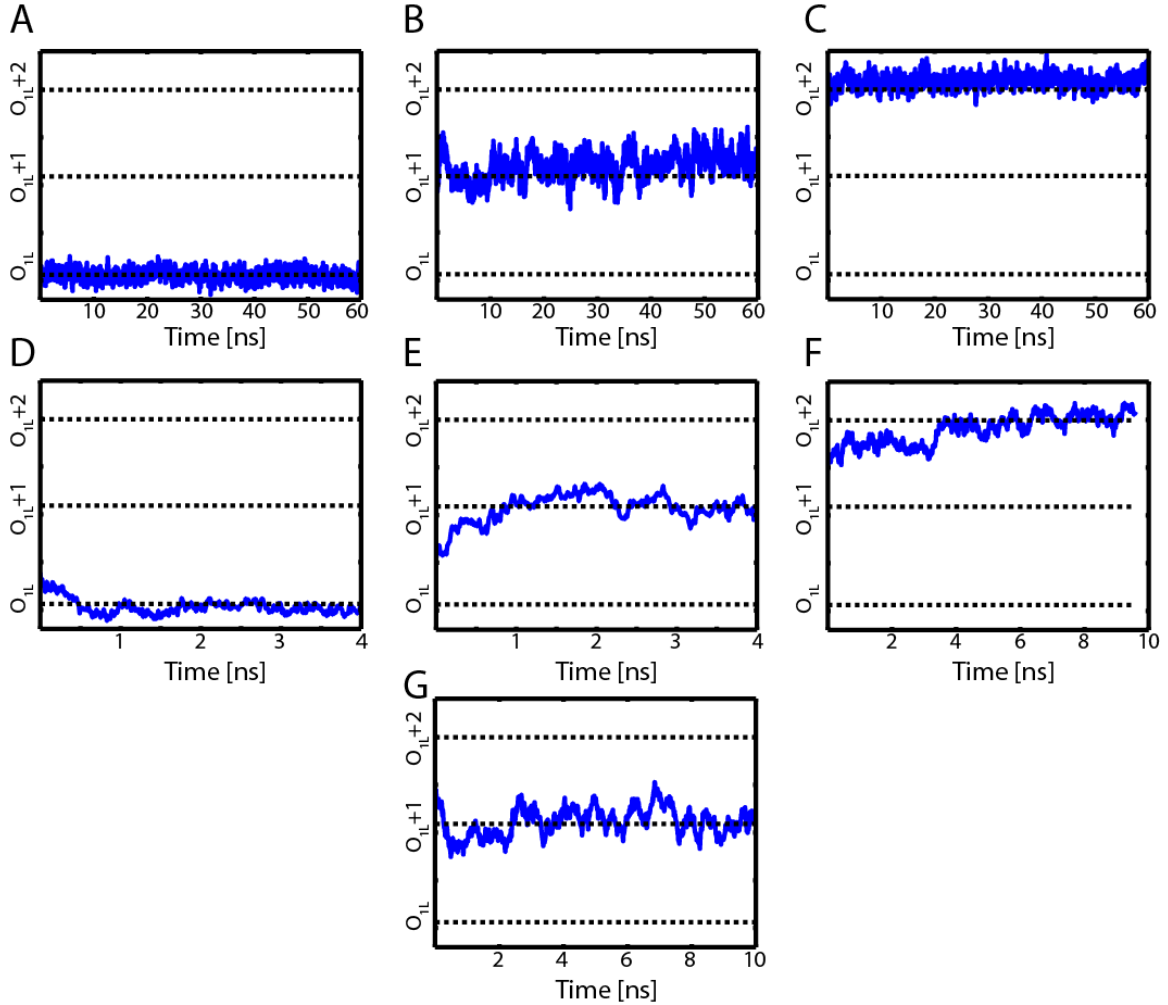
Figure S8. Displacement of the protein along the helical trajectory in unrestrained MD simulations. The displacement along the helical trajectory was calculated with the same method used for the umbrella sampling simulations. Data were collected in unrestrained MD simulations that started from the last frame of the umbrella sampling trajectories with harmonic potential centered at: 0.5 Å (A), 11.01 Å (B), 20.02 Å (C), 4.0 Å (D), 5.51 Å (E), 15.52 Å (F), 14.52 Å (G).
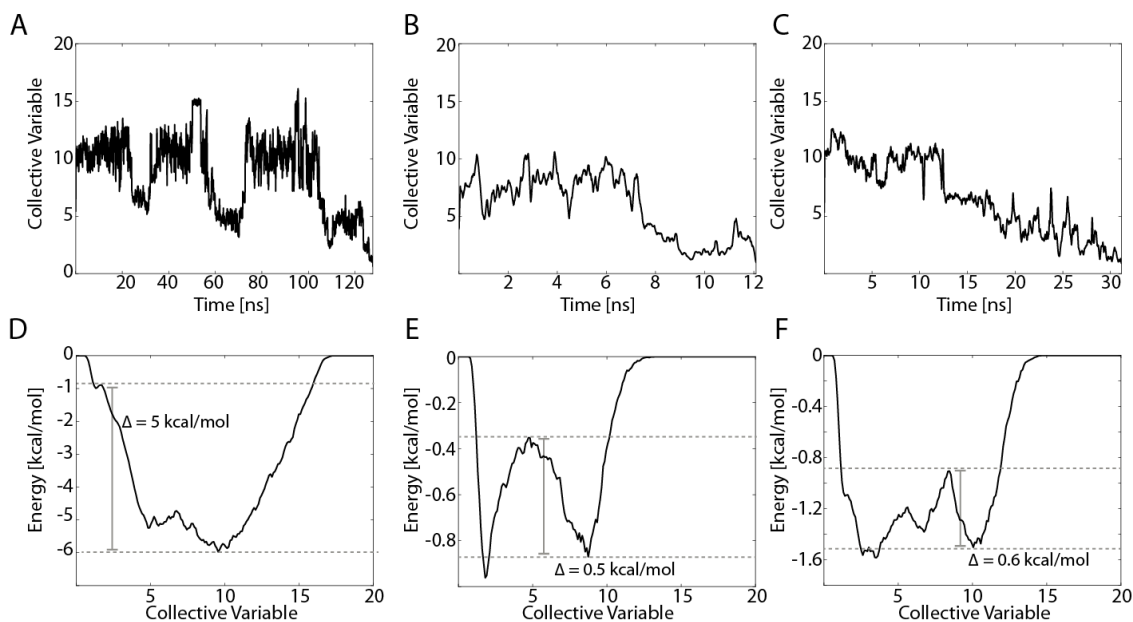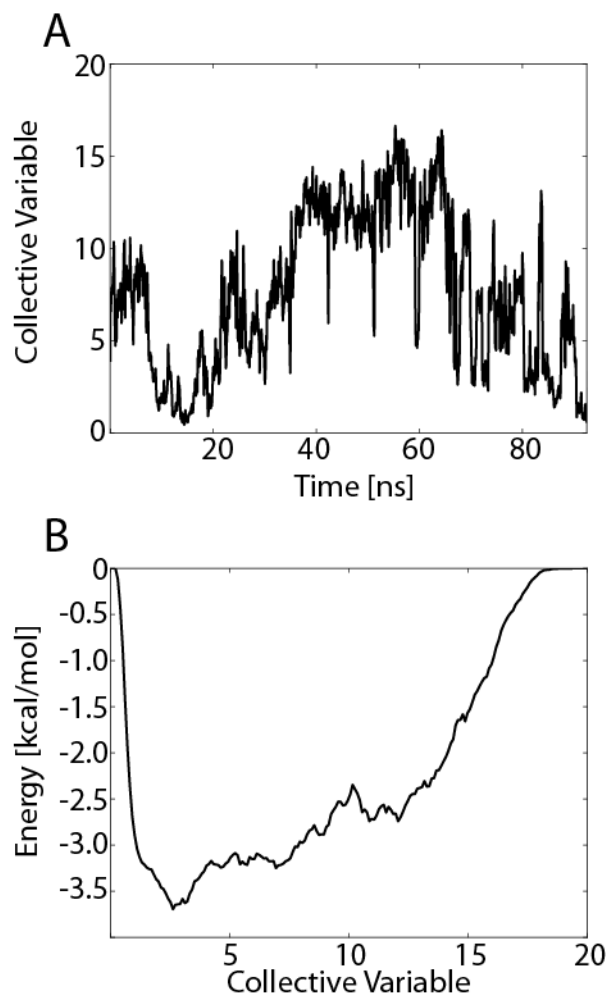
Figure S9. Unbinding of LacI from the DNA sequence O_{1L} (A, D), $O_{1L}+1$ (B, E) and $O_{1L}+2$ (C, F). The DNA sequences $O_{1L}+1$ and $O_{1L}+2$ correspond to the LacI monomer translated respectively 1 and 2 base pairs along the DNA starting from the left half of operator O_1 (O_{1L}). (A-C) Value of the collective variable in Equation 2 as a function of time in the metadynamics simulation. (D-F) Estimated free energy at the first time step with value of the collective variable lower than 1.0 (128 ns in O_{1L}, 12 ns in $O_{1L}+1$, and 31 ns in $O_{1L}+2$).
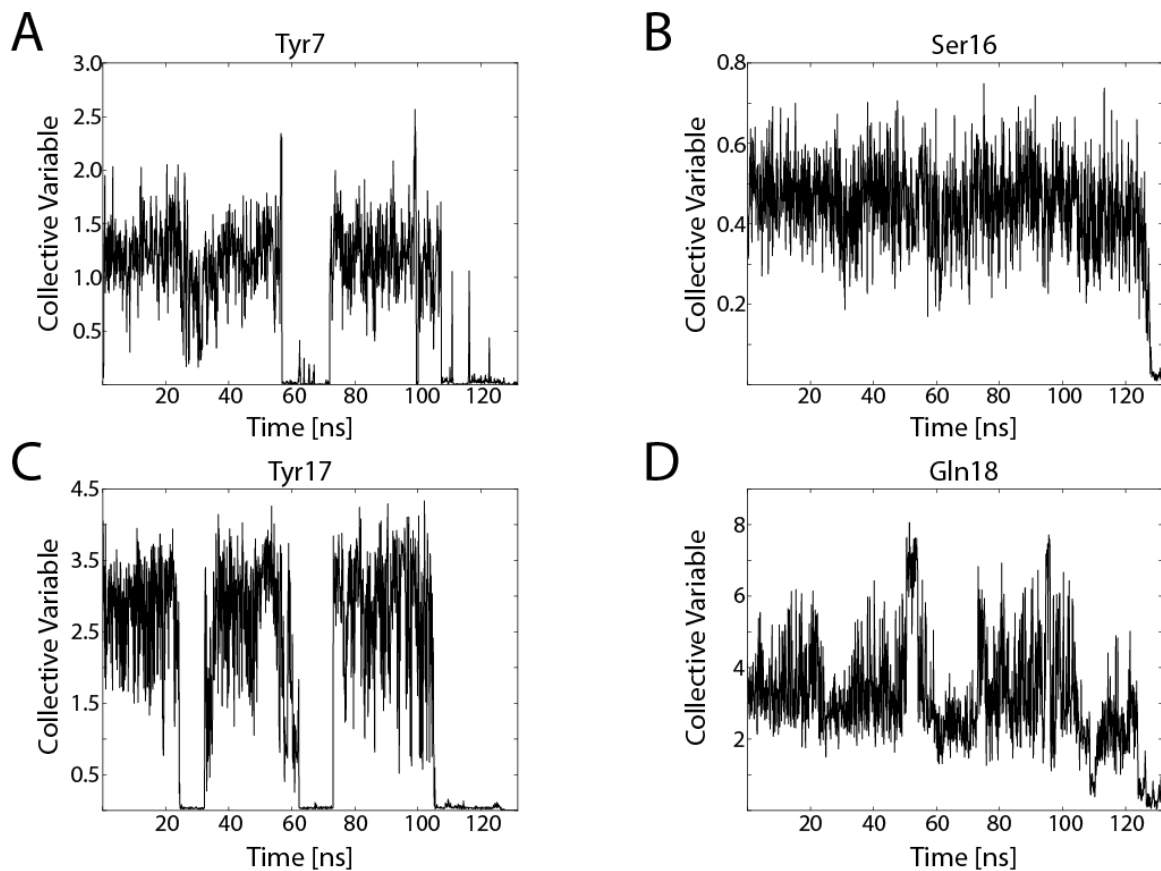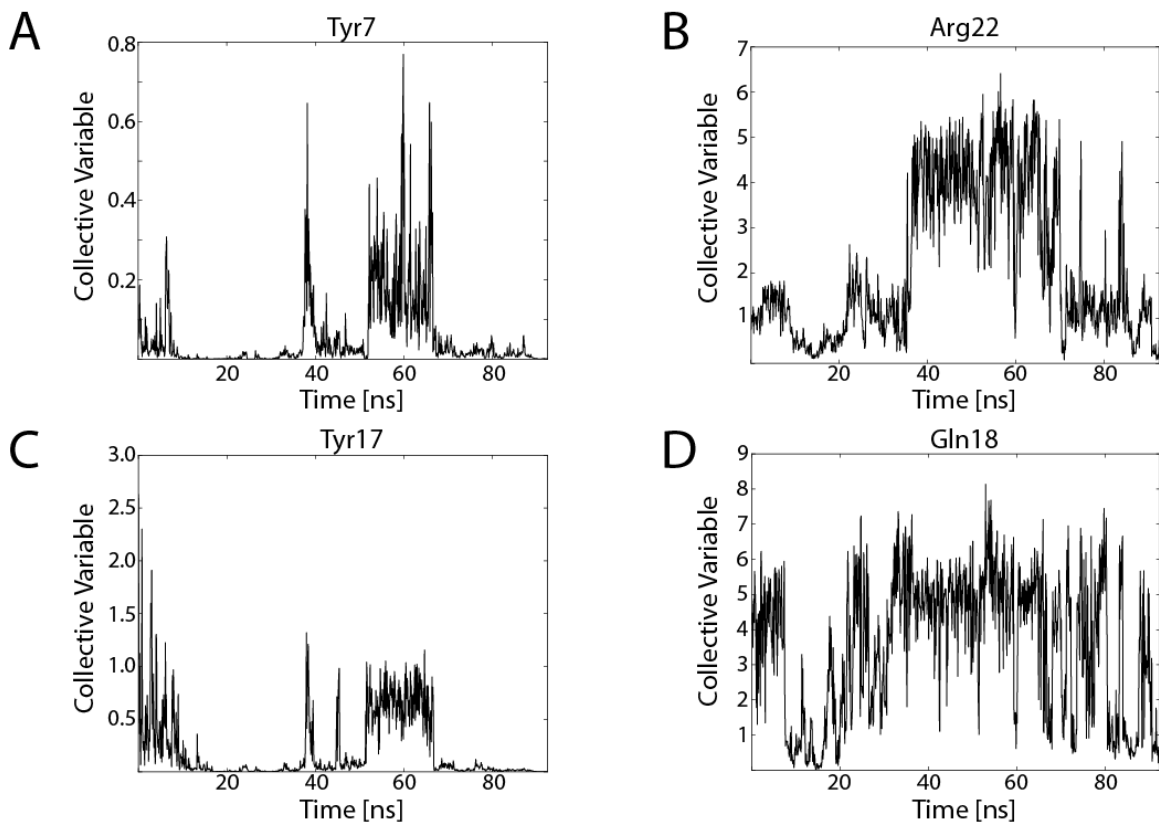
Figure S10. Inter-atomic interactions between protein residues and DNA base edges in the metadynamics simulations starting from configuration O_{1L}. Contribution to the metadynamics collective variable coming from protein-DNA contacts with residue Tyr7 (A), Ser16 (B), Tyr17 (C), and Gln18 (D).
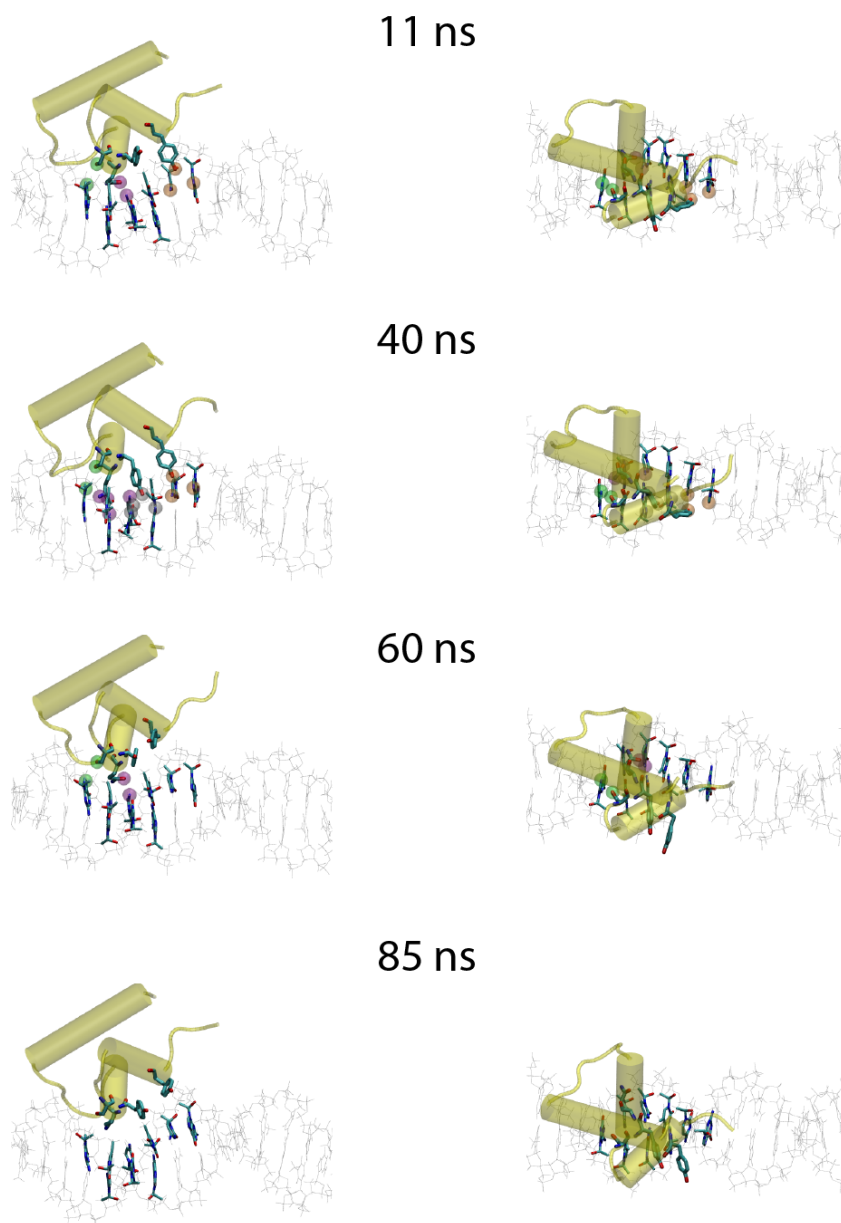
28 ns

53 ns

70 ns

125 ns
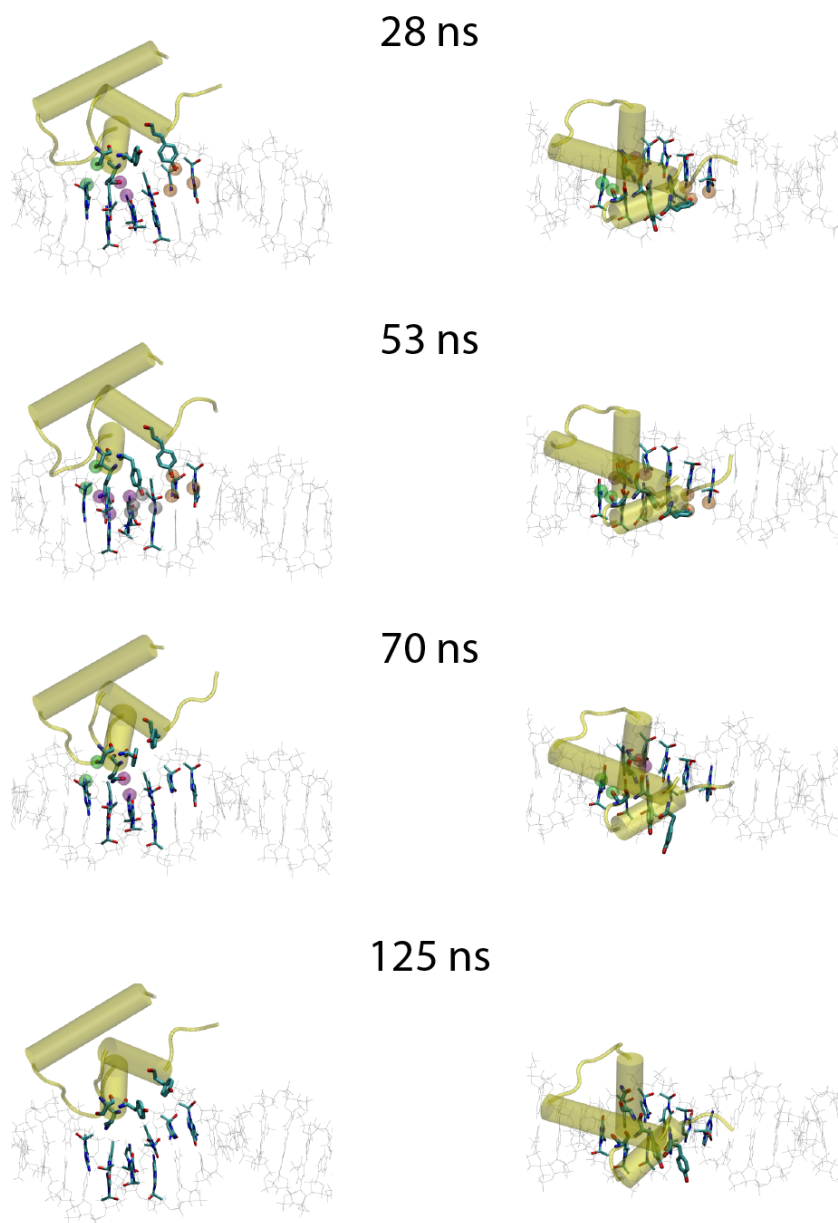
Figure S11. Molecular structures from the metadynamics trajectory O$_{1L}$.
Snapshots from the metadynamics trajectory shown from two views rotated by 90°
along the DNA-axis. DNA bases G5-A8 on the primary strand, and G11-A15 on the
complementary strand are shown in licorice representation. The protein is in yellow.
Protein residues Try17, Try7, Ser16 and Gln18 are shown in licorice representation.
Green, purple, grey and orange spheres correspond to couple of protein-DNA atoms
that contributes to the CV defined by Equation 2 and that are closer than 4 Å from
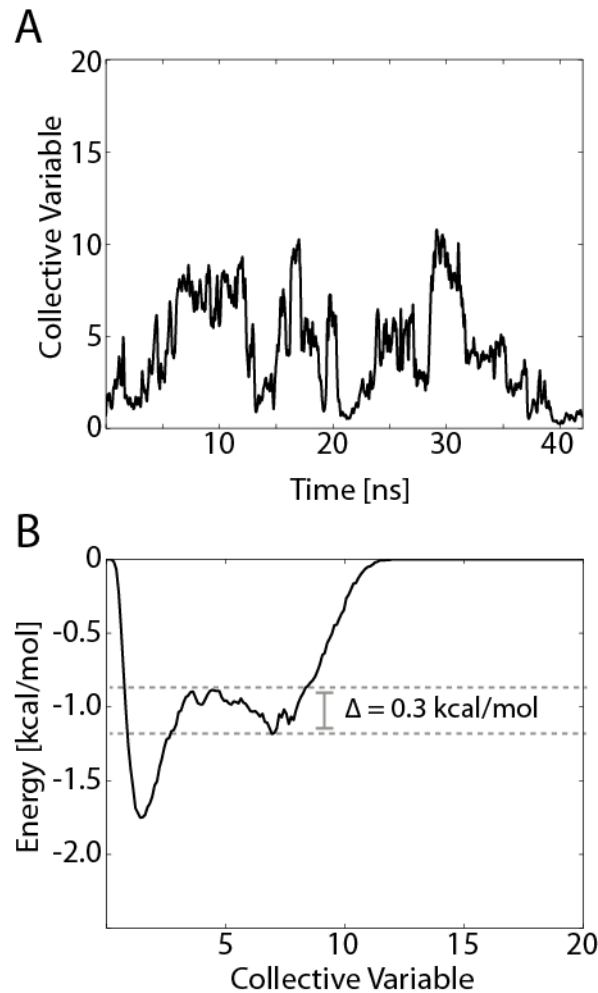each other, respectively for residues Ser16, Gln18, Tyr17, and Tyr7.

Figure S12. Specific/non-specific binding of LacI to the DNA sequence $O_{1L}+1$. (A) Value of the collective variable (Equation 2) as a function of time in the metadynamics trajectory. (B) Estimated free energy at the second transition between a specific and a non-specific binding state (time-step with CV < 0.5 after 40 ns).

Figure S13. Inter-atomic interactions between the protein residues and the DNA base edges in the metadynamics simulations starting from configuration O$_{1L}$+1. Contribution to the collective variable of the metadynamics simulation coming from protein-DNA contacts with residue Tyr7 (A), Arg22 (B), Tyr17 (C), and Gln18 (D).
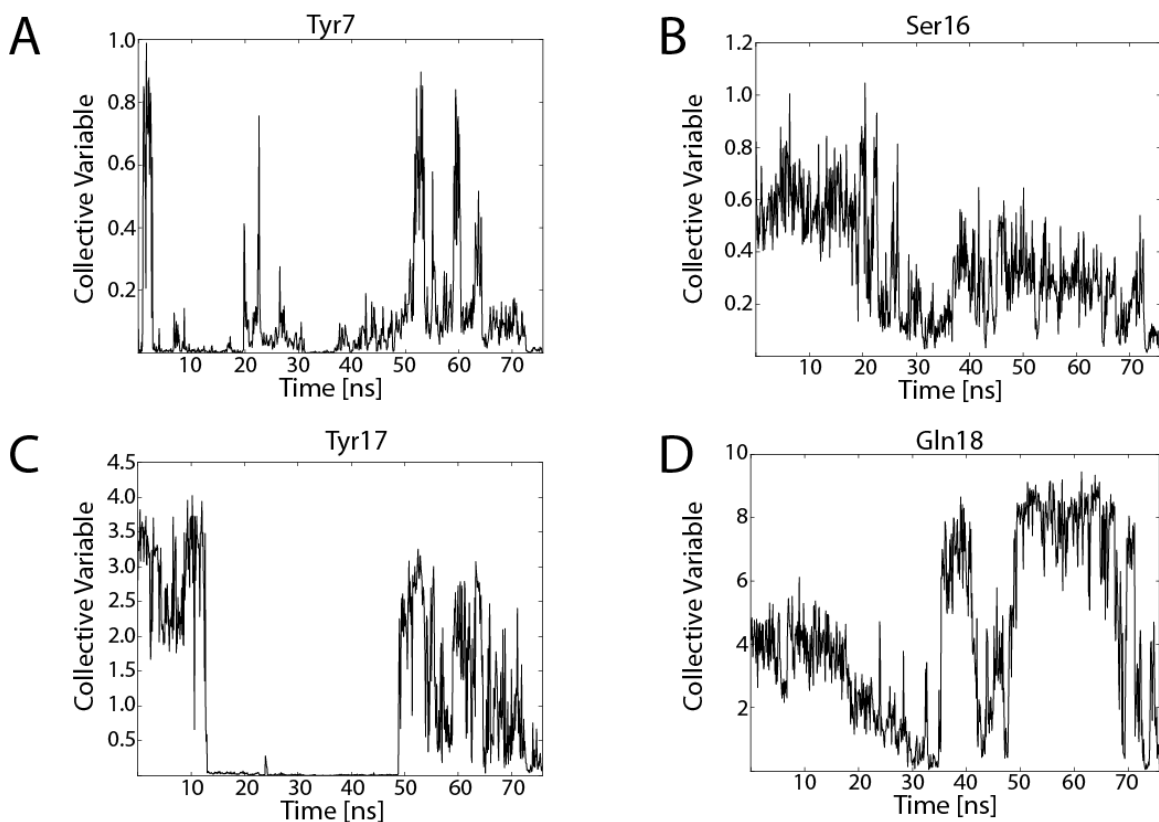
11 ns

40 ns

60 ns

85 ns

Figure S14. Molecular structures from the metadynamics trajectory $O_{1L}+1$. Snapshots from the metadynamics trajectory shown from two views rotated by 90° along the DNA-axis. DNA bases G5-G9 on the primary strand, C10-C14 on the complementary, and protein residues Try17, Try7, Arg22 and Gln18 are shown in licorice representation. Green, purple, and grey spheres correspond to couple of protein-DNA atoms that contributes to the CV defined by Equation 2 and that are closer than 4 Å from each other, respectively for protein residues Arg22, Gln18, and Tyr17.

Figure S15. Specific/non-specific binding of LacI to the DNA sequence $O_{1L}+1$.
(A) Value of the collective variable (Equation 2) as a function of time in the metadynamics trajectory that started from a frame (30 ns) taken from the unrestrained MD simulation with the protein aligned with the DNA sequence $O_{1L}+1$. (B) Estimated free energy at the time-step when the protein switches between a specific and a non-specific binding state (CV < 0.5). The energy barrier for the specific/non-specific transition (Δ) is shown.

Figure S16. Inter-atomic interactions between the protein residues and the DNA base edges in the metadynamics simulations starting from configuration $O_{1L}+2$. Contribution to the collective variable of the metadynamics simulation coming from protein-DNA contacts with residue Tyr7 (A), Ser16 (B), Tyr17 (C), and Gln18 (D).

10 ns

28 ns

59 ns

75 ns
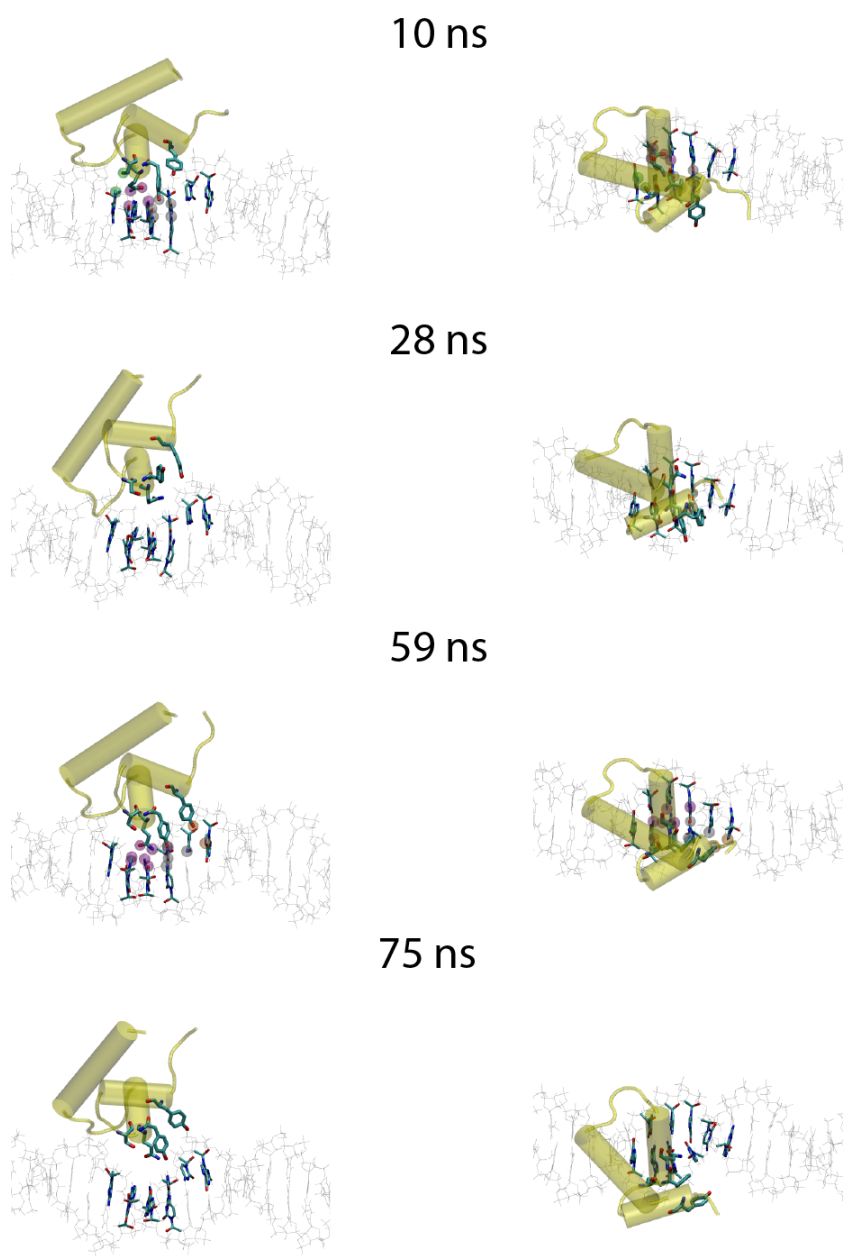
Figure S17. Molecular structures from the metadynamics trajectory $O_{1L}+2$. Snapshots from the metadynamics trajectory shown from two views rotated by 90° along the DNA-axis. DNA bases G7-C10 on the primary strand, G9-T13 on the complementary strand, and protein residues Try17, Try7, Ser16 and Gln18 are shown in licorice representation. The protein is in yellow. Green, purple, grey and orange spheres correspond to couple of protein-DNA atoms that contributes to the CV defined by Equation 2 and that are closer than 4 Å from each other, respectively for protein residues Ser16, Gln18, Tyr17, and Tyr7.