A nucleoid-associated protein bends and bridges DNA in a multiplicity of topological states with varying specificity

S. B. Yoshua[1†], G. D. Watson[1†], J. A. L. Howard[1], V. Velasco-Berrelleza[1], M. C. Leake[1,2*], A. Noy[1*]

[1] Department of Physics, University of York, York YO8 8FH, United Kingdom.

[2] Department of Biology, University of York, York YO8 8FH, United Kingdom.

[*] Correspondence to: mark.leake@york.ac.uk, agnes.noy@york.ac.uk.

[†] These authors contributed equally to this work.

Abstract

Nucleoid-associated proteins perform crucial roles in compacting prokaryotic DNA, but how they generate higher-order nucleoprotein structures by perturbing DNA topology is unclear. Integration host factor, IHF, is a key nucleoid-associated protein in bacteria, creating sharp bends in DNA. We show that the IHF–DNA complex is more elaborate than a simple two-step model previously suggested and also provide structural insights about its multimodal nature. Using atomic force microscopy and molecular dynamics simulations we find three topological modes in roughly equal proportions: “associated” (73˚), “half-wrapped” (107˚) and “fully-wrapped” (147˚), with only the latter occurring with sequence specificity. DNA bridging is seen not only with high IHF concentrations but also with densely packed binding sites, with simulations showing this occurs through another binding mode that does not depend on sequence. We present a model of these states and propose a crucial biological role for these observed behaviors.

Introduction

Nucleoid-associated proteins (NAPs) are a collection of DNA-interacting proteins that perform crucial roles of organization, packaging and gene regulation in prokaryotic chromosomes (*1*). They often bind non-specifically and create a wide variety of DNA topologies (*2*, *3*). Integration host factor (IHF) is a key NAP in *Escherichia coli* and other Gram-negative bacteria. Its architectural role is thought to involve creating some of the sharpest bends seen in DNA (*4*), in excess of 160˚ (*5*), at around 300 sites containing the consensus sequence WATCARNNNNTTR (W is A or T; R is A or G), thereby facilitating the assembly of higher-order nucleoprotein complexes (*6*) and gene regulation (*7*), such as the CRISPR-Cas system (*8*), the origin of replication (*oriC*) (*9*) and a Holliday junction complex involved in the integration and excision of phage λ DNA (*10*). IHF's large repertoire of roles supports the long-standing view that it has an essential function in the structural organization of DNA in a wide variety of genetic transactions. However, with copy numbers on the order of tens of thousands per cell depending on the growth phase (*11*), non-specific binding must also play a role despite a 1000-fold larger K_d (*12*), and it has been shown to be able to compact DNA (*13*, *14*). This non-specific binding may also play a role *ex vivo*, as IHF has been implicated in biofilm stability. In some cases the removal of IHF caused a 50 ± 6% reduction in biofilm thickness (*15*) and IHF has also been imaged at vertices of an extracellular DNA lattice (*16*).

The crystal structure of IHF is obtained from its conformation bound to DNA at a specific binding site (*5*). The crystallographic structure revealed that IHF is formed by a core of α helices with a pair of extended β−ribbon arms whose tip each contains a conserved proline that intercalates between two base pairs (Figure 1A). These two intercalations stabilize strong bends 9 bp apart and facilitate wrapping of two DNA 'arms' around the protein body , tightened by electrostatic interactions between the phosphate backbone and isolated

cationic amino acids on the protein's surface, resulting in a binding site with a total length of 35 bp and an overall bend angle of 160˚ (see Figure 1B), which has been supported by atomic force microscopy (AFM) (*17*, *18*). IHF's consensus sequence is located on the right side of the binding region and is small compared to the total length of the wrapped DNA. However, many IHF binding sites include an A-tract to the left-hand side (upstream of the specific sequence) that increases affinity (*19*), the degree of bending and the length of the attached DNA site.

Recent studies have shown that IHF can produce DNA bends smaller than the canonical 160˚. IHF can bend DNA flexibly, with a portion of the population that is only partially bent (*20*). Previous work has shown that the IHF/DNA complex is more dynamic than originally thought, with transitions between bending and binding states, with an activation energy of approximately 14 kcal/mol ($24k_BT$ where k_B is the Boltzmann constant, absolute temperature T) (*20*), but the free energy of wrapping is found to be only around 3.6 kcal/mol ($6k_BT$) (*21*). In this study, binding also appeared to occur through a two-step process, a fast ~100 µs step that seems to be sequence non-specific, and a slower millisecond step which was site-specific (*20*, *22*). However, despite these advances, there is still poor understanding of the underlying bending mechanism driven by IHF, with no obvious explanation as to why the apparent free energy of DNA wrapping is so close to the thermal energy scale.

Here, we compared experimental AFM imaging with atomically precise predictions from molecular dynamics simulations for the same DNA sequence of around 300 bp and performed further AFM on three further constructs to investigate the effects of increasing the number of IHF binding sites (Figure 1C). This approach allowed us to demonstrate multiple modes in which IHF can bind to the H2 binding site on DNA. The intermediate strength of this binding site (between those of the strong H' and weak H1 sites) makes it representative of a

generalized NAP. We then formulated a model for multiple IHF bending states encapsulating both specific and non-specific binding. Using advanced sampling methods, we calculate the free energy of each of these conformations and demonstrate that asymmetric behavior between DNA arms. Furthermore, by looking at the protein's non-specific binding properties we see evidence for IHF-mediated bridging between two DNA strands and provide structural insight into how this occurs, which may have implications for the binding modes between DNA and other nucleoid-associated proteins.

Results

Simulation of the two-step binding mechanism

Starting from the intercalated state shown in Figure 1A, the canonical bending can be reproduced, as in Figure 1B. This is true of simulations using both a short 61 bp DNA construct in explicit solvent and a longer 302 bp construct in implicit solvent (see Materials and Methods). In both cases the observed DNA–protein interactions correspond well to those present in a short simulation of the crystal structure in explicit solvent (Figure S1). These interactions can be broadly divided into four regions based on their position relative to the center of the binding site and the protein subunit to which the involved amino acid belongs: on the left-hand side (containing the A-tract), the α subunit is closest to the center and thus constitutes the "near" side, while on the right-hand side (containing the consensus sequence), the β subunit is closer.

While the DNA in the crystal structure is too short to capture some of the more distant interactions, there is generally strong agreement between the crystal structure and simulations starting from unbent DNA. This observation demonstrates the validity of the

molecular dynamics methodology for this system, as well as suggesting a possible explanation for the previously proposed two-step binding mechanism: that the IHF arms bind to DNA first, the proline residues intercalating to induce flexible hinges prior to bending. The two-step mechanism that we have proposed explains both the high activation energy of initial binding (*20*) due to proline intercalation and the smaller free energy of wrapping (*21*).

Interactions across all simulations are dominated by hydrogen bonding between positively charged surface amino acids and the DNA backbone, with almost no interaction between the protein and the nucleobases away from the intercalation sites. The cationic amino acids involved are outliers on an otherwise primarily neutral or negatively charged surface.

Multimodality of IHF binding

To investigate the differences between specific and non-specific binding, two short sequences were amplified from phage λ, one with no IHF consensus sequence (0λ361) and another with a single consensus sequence (1λ302) of lengths 361 bp and 302 bp respectively (Figure 1C). These were then imaged using AFM and compared to molecular dynamics (MD) simulations to determine the different bending behavior.

AFM images were analyzed by skeletonizing the pre-processed images to recover the DNA contour (see Materials and Methods), with qualitatively different behavior observed depending upon the presence of an IHF binding site (Figures 2A–B). Figure 2C (and Figure S2) shows the reduction of the radius of gyration of the DNA if IHF is present, suggesting that the protein bends DNA for the single-binding-site case in this assay, but with no significant change for the 0λ361 sequence. However, as the radius of gyration can only determine that

significant compaction averaged across the entire construct has occurred more precise bending angles analysis was also performed (see Materials and Methods).

When IHF is incubated with DNA before imaging the angular distributions, as expected, contain a large proportion (~60-70% of the total) of the bare DNA angle distributions (Figure S3), which was taken into account. The presence of IHF can be seen via additional peaks in the angular distributions, which for 0λ361 (Figure 2A) and 1λ302 (Figure 2B) show a further 2 and 3 peaks respectively (see Figure S4). Two common peaks are at angles of 73 ± 7˚ and 107 ± 9˚ (mean ± s.d. values from 1λ302). As these are present in both distributions, this finding suggests that these bending modes occur due to non-specific binding, whilst the canonical large bending angle (147 ± 30˚), as seen in the crystal structure (*3*), only appears in 1λ302, that contains one specific IHF binding site. A wide range of angles have previously been seen using AFM imaging, but usually as a single broad distribution centered around 100-130˚ (*17*, *18*) and not a combination of distributions. Our new finding suggests that this broad angular range could likely be due to IHF–DNA complexes being primarily the canonical fully wrapped state where the angle has been underestimated by AFM or by the low angular resolution of AFM. The existence of the other two states can be explained as incompletely bound states, as have been previously proposed (*20*, *23*). However, unlike previous studies we show two possible states before the DNA is fully bound.

The proportions of the associated, half-wrapped and fully wrapped states (~32%, 27% and 41% respectively) show that all three binding modes are present in roughly equal quantities with a single binding site. However, the proportion of the associated states to the wrapped states is also much larger for 0λ361 (~64%), suggesting that this mode is unrelated to the consensus sequence. The 73˚ state is one that has not been observed in AFM before,

possibly as it is within the expected range of angles for bare DNA and could be excluded if not accounting for the bare DNA distribution (Figure S3) or due to the selection of an H2 binding site making the fully wrapped state less probable.

Multiple MD simulations were performed of the 1λ302+IHF construct in implicit solvent, starting from the bound but unbent state shown in Figure 1A. The frames of these simulations were clustered through hierarchical agglomerative clustering and a free-energy landscape was constructed based on separate umbrella-sampling simulations (see below). Three clusters were found with bend angles of 66°, 115°, and 157°, in good agreement with the AFM results; representative structures are shown in Figure 3C. The relative probabilities of the clusters can be determined by integrating over the relevant regions of the free-energy landscape (Figure 3D); this analysis predicts that the canonical fully wrapped (157°) state should be observed around 33% of the time, the half-wrapped (115°) state 27%, and the associated (66°) state 39%. Each cluster corresponds to a minimum or plateau in the free-energy landscape; the locations of these features, along with the observed paths between them, are highlighted in Figure 3E.

The clusters can also be characterized by the hydrogen bonds present between the protein and DNA (Figure 3F). In the fully-wrapped state, both sides of the DNA form hydrogen bonds with both subunits of the protein; in the half-wrapped state, the left-hand side, which contains the AT-tract, forms contacts with both subunits while the right-hand side forms no hydrogen bonds; and in the associated state, the left and right sides of the DNA form hydrogen bonds with only the α and β subunits, respectively. It is notable that the consensus sequence does not interact with IHF in the half-wrapped state.

Free-energy landscape of IHF binding to DNA

The free energies and relative probabilities of each of these states were determined via umbrella sampling simulations (*24*) of a 61bp segment of the construct in explicit solvent (see Materials and Methods). The reaction coordinate (the distance between a chosen protein atom and a chosen DNA atom) on each arm was varied, once with the other arm allowed to wrap and once with the other arm held away from the protein, and the potential of mean force (PMF) for each simulation was calculated using the weighted histogram analysis method (WHAM). By linearly interpolating between the two sets of results for each arm (Figure S5) a two-dimensional free energy landscape can be constructed with the reaction coordinates as orthogonal axes, as in Figure 3C.

It is observed that there is a large free-energy barrier preventing the right-hand side (which contains the consensus sequence) from binding before the left-hand side (which contains the AT-tract). The form of the potentials is also qualitatively different. While the left-hand side presents a relatively deep Lennard-Jones-type potential with a minimum at around 20Å and an additional small plateau at around 35Å, implying that binding is energetically favorable, the right-hand side has a much flatter shape with small minima around 10Å and 35Å, and appears to be dominated by thermal noise.

This free-energy landscape suggests that it is very energetically favorable for the protein to bind the A-tract on its left-hand side, but that it is less inclined to bind the consensus sequence. This is reflected in the observed clusters—there is no cluster in which the left-hand side is unbound, while the position of the right-hand side distinguishes the fully and half-wrapped states. The associated state corresponds to the smaller local plateau or minimum in each arm's energy landscape, as seen when the values of the reaction coordinates are plotted together for each frame of a cluster as in Figure S6.

The lack of interaction of IHF with the consensus bases suggests that the half-wrapped and associated states are forms of non-specific binding, a conclusion supported by the occurrence of bend angles corresponding to these states in AFM observations of the 0λ361 construct (in which the consensus sequence does not appear). The consensus sequence is therefore necessary only for the fully wrapped binding mode, suggesting that its biological function is to moderate bending by - not binding of - IHF.

The free-energy landscape features a significant minimum that appears to correspond to another state, in which the left-hand side is fully bound and the right-hand side is partially bound. However, the frames in this region are difficult to distinguish from those in the fully wrapped state, and the corresponding minima in the free-energy landscape are separated by only a narrow energy barrier of less than 2 k_BT. Since the system can move freely between these two minima under the influence of thermal noise, it is more proper to consider this state an extension of the fully wrapped mode.

DNA aggregation and bridging by IHF

Increasing the concentration of IHF for both 0λ361 and 1λ302 caused clusters of increasing size to be formed (Figures 4A and 4B). This behavior at high concentration (~1 IHF:3bp / 120 nM IHF:1 nM DNA) has previously been seen (*14*) and shows how large cellular concentrations of IHF – such as in the stationary phase – could increase genome compaction.

Two further constructs, with the same three binding sites in close proximity (but of different lengths, 3λ343 and 3λ474, with 343 and 474 bp respectively), behave very differently to the previous constructs with no or one binding sites. Even at low concentrations, from around 5 nM up to 50 nM (for 1 nM DNA, so ~1 IHF: 6 – 60 bp), IHF/DNA clusters could

be seen, larger than those for the other two constructs, especially in the case of 3λ343. This could be due to the shorter length (around ¾ of 3λ474) as most of the sequence is the same, or that it has a much higher AT-content (70% vs 63%). As with 1λ302, there appears to be a large unbound population but also fewer DNA molecules that are bound by IHF. Although this may be due to protein aggregation, the difference between DNA binding constructs suggests that it is not just a simple case of protein-protein binding but that this behavior is mediated by the different sequences of the constructs, and that bridging (when conditions are met) is preferable to bending. The prominence of clustering with three IHF binding sites so close together suggests it has the same behavior to a high local concentration of IHF on the DNA, allowing bridging to occur at more biologically relevant concentrations.

IHF-induced aggregation can be explained by the bridging of two DNA strands by a single IHF, a phenomenon observed in MD simulations. These bridges result from non-specific interactions between the DNA and the protein; some of these hydrogen bonds involve amino acids that also interact in one or more of the wrapped states, while others are only observed in the bridged state. The main DNA strand, into which the prolines intercalate, appears to remain mostly unwrapped, in a conformation closest to the associated state. By performing umbrella-sampling simulations and weighted histogram analysis (see Materials and Methods), it is found that bridging is very energetically favorable, with a free-energy difference on the order of 14 kcal/mol ($24k_BT$) between the bridged and unbridged states.

Discussion

Here we have integrated advanced theory and experiment to explore the physics of life at the single-molecule level (*25*) and enable new insights into the complex interactions between

proteins and DNA. Both atomic force microscopy experiments and molecular dynamics simulations show evidence of a larger number of bending modes of the IHF/DNA complex than previously proposed. The presence of all three states as well as the large proportion of unbound DNA suggests a highly dynamic system, where higher AT-content makes binding more probable (*3*, *6*, *26*) but where a consensus sequence is needed to fully bend DNA. We put forward the following model for the mechanism of IHF binding (Figure 5).

When IHF binds to DNA, it is highly likely that the left-hand side will bind to the DNA first, but not with the consensus sequence. This results in the half-wrapped state in which the left-hand side is fully bound while the right-hand side is free. Alternatively, both sides may make interactions with the nearest subunit of the protein, resulting in a smaller bending angle, that we designate the associated state. Transitions between the associated and half-wrapped states were not observed (Figure S7), and both appear to be long-lived metastable states corresponding to plateaus or local minima in the free-energy landscape (Figure 3D). As the associated and half-wrapped states are experimentally seen in constructs both with and without a consensus sequence, this suggests a non-specific binding mechanism for these states. However, in the presence of a consensus sequence, both of these two initial states lead towards the global minimum, resulting in canonical binding. This does not imply that this is the only mechanism by which IHF binds nonspecifically to DNA; for example, the similar electrostatic profiles of IHF and HU suggest that another nonspecific binding mode similar to that observed by Hammel *et al.* (*27*) is likely, but such a state would require a distinct simulation methodology and any induced bending is likely to be within the range of typical angles for bare DNA, so this possibility was not explored in this work.

This work was performed on sequences containing the intermediate-strength H2 binding site. Meanwhile the H' site—on which most previous work has focused—may behave differently, such as by exhibiting a bias towards the fully wrapped state, this work is likely to be more representative of general nucleoid-associated proteins, which do not typically interact with DNA as strongly as the IHF interacts with the H' site.

Semi-stable states in which IHF remains bound to DNA would allow for IHF to remain associated with the DNA while retaining some flexibility in the region of the IHF - which could be important in the formation of, for example, a transcription complex. Interestingly, the consensus bases are within the region that is less tightly bound, which, as an AT-rich region could demonstrate the need for flexibility downstream of the IHF.

Another interesting behavior observed was the bridging/clustering of DNA by IHF. At high concentrations of IHF (such as in the stationary phase) this can occur non-specifically, giving it a role in compaction. The DNA constructs with multiple binding sites in close proximity seem to preferably select for the bridging behavior over bending, showing how the formation of bridges could be influenced by the DNA sequence. This could be due to steric hindrance or tension due to the presence of other proteins preventing complete wrapping by each IHF, leaving the lower regions of IHF free to bind other DNA. This could explain the role of IHF in biofilms, by cross-bridging extracellular DNA, but undoubtedly shows that the function of IHF is far more multifaceted than bending DNA.

Materials and Methods

Protein amplification and purification

IHF was overexpressed from the *E. coli* strain BL21AI containing the plasmid pRC188 (a kind gift from the Chalmers laboratory, the University of Nottingham, UK). The cells were grown

in 2x 1 L LB + 100 μgml^{-1} carbenicillin at 37°C with shaking at 180 rpm to an OD_{600} ~ 0.6. Overexpression of IHF was induced by the addition of IPTG and Arabinose to final concentrations of 1 mM and 0.2% (w/v) respectively and growth was allowed to proceed for a further 3 hours. Cells were harvested by centrifugation at 4,000 rpm and 4°C in a Sorvall SLC6000 rotor. The pelleted cells were then resuspended in 20 mL of 10 mM Tris pH 7.5, 10% sucrose (w/v) before being flash frozen in liquid nitrogen and stored at -80°C.

For purification stored cell pellets were thawed on ice and the buffer was adjusted to contain 50 mM Tris pH 8.4, 150 mM KCl, 20 mM EDTA, 10 mM DTT and 0.2 mgml^{-1} lysozyme. The resultant suspension was mixed by inversion and left on ice for 15 mins before the addition of brij58 to 0.1% (w/v) and a further 15 minutes on ice. The suspension was then clarified by centrifugation at 4°C and 38,000 rpm in a Beckmann ty70ti rotor for 60 minutes. Polymin P added to 0.075% (w/v) from a 1% stock in a dropwise fashion to the supernatant whilst stirring at 4°C, stirring was continued for 10 mins before centrifugation at 4°C and 16,000 rpm in a Sorvall SS34 rotor for 20 minutes. The supernatant was collected before being subjected to a 50% $AmS0_4$ precipitation followed by an 80% $AmS0_4$ precipitation, in each case the sample was centrifuged as above IHF remained soluble at 50% $AmS0_4$ and precipitated at 80% $AmS0_4$. The precipitated IHF was resuspended in 20 ml buffer A (50 mM Tris·HCl pH 7.5, 2 mM EDTA, 10 mM β-ME, 10% glycerol) such that the conductivity matched that of buffer A + 100 mM KCl. The sample was loaded onto a 10 mL P-11 phosphocellulose column equilibrated with buffer A + 100 mM KCl, the column was washed with 300 ml of buffer A + 100 mM KCl before being developed with a 200 mL gradient of 0.1-1 M KCl in buffer A. Fractions containing IHF were identified using 15% SDS polyacrylamide gel electrophoresis and pooled. The pooled fractions were dialysed against buffer A + 100 mM NaCl before loading onto a 5 mL Hitrap Heparin column (GE Healthcare) equilibrated with the same buffer,

the column was washed with 300 ml of buffer A + 100 mM NaCl before being developed with a 200 mL gradient of 0.1-1 M NaCl in buffer A. Fractions containing IHF were again identified using 15% SDS polyacrylamide gel electrophoresis and pooled. Pooled fractions were aliquoted and flash frozen in liquid nitrogen before storage at -80°C. Protein concentrations were determined using the Bradford Protein Assay (Bio-Rad).

DNA amplification

DNA constructs were chosen from phage λ with different numbers of IHF binding sites, within the range of 300-400 base pairs. All DNA constructs were amplified by PCR using Q5 DNA polymerase (see Table S1 for primers) resulting in three constructs: 0λ361 (361 bp long containing no IHF consensus sequence, 1λ302 (302 bp long with one IHF consensus sequence) and 3λ343/474 (343/474 bp long with three IHF consensus sequences). These were all chosen in the region around the *xis* gene that is known to interact with IHF (*28*).

Atomic force microscopy acquisition and analysis

Mica was first pre-treated by depositing 20 µl 0.01% (w/v) 1-5 kDa poly-L-lysine (Sigma-Aldrich, MO, USA) and left for 5 minutes before washing with 1000 µl filtered milliQ H2O and vacuum drying. The DNA samples were prepared by adding a 20 µl solution of 10 mM Tris, 50 mM KCl with 1 nM DNA (and 5 – 150 nM IHF) to the pre-treated mica and leaving for ~5 min. These were then washed once more with 1000 µl of filtered milliQ H2O and vacuum dried before being imaged.

All samples were imaged on a Bruker Bioscope Resolve (Bruker, CA, USA) in tapping mode using TAP-300AI-G probes (BudgetSensors, Bulgaria). Images were then loaded using pySPM (*29*)and preprocessed by using a 1st order line-by-line flattening and 2nd order fits to

flatten the surface before filtering scars. The images were then analyzed using custom code using the scikit-image package (*30*) to threshold the image and segment it into the individual DNA strands. The individual segments were then skeletonized to recover the DNA contour for further analysis. Angles were measured by first identifying IHF as the peak along the contour, then taking a 4px (16 nm) vector either side of a 4px window around the IHF.

Molecular Dynamics Simulations

Molecular dynamics simulations were performed using AMBER 18 (*31*). Linear B-DNA molecules with sequences identical to the experimental constructs were generated using the NAB utility; IHF and the eleven base pairs at the center of the binding site were extracted from PDB entry 5J0N (*10*), manually mutated where necessary, and inserted at the relevant location; a 61 bp section of the 1λ302+IHF construct, centered on the binding site, was extracted. The full constructs were solvated using the implicit generalized Born model with a salt concentration of 0.2M; the 61 bp constructs were explicitly solvated using a truncated octahedral TIP3B box and neutralized with a 0.2M-equivalent concentration of K and Cl ions. The protein and DNA were represented using the ff14SB (*32*) and BSC1 (*33*) force fields, respectively.

Following minimization, four 50 ns simulations were performed of the 1λ302+IHF construct using the CUDA version of pmemd, starting from the same minimized structure. Following the observation of three populations in the experimental results, the simulation results were clustered using hierarchical agglomerative clustering in cpptraj until three clusters were formed. The DNA helix axis was calculated for each frame of these clusters using WrLINE (*34*); this was projected onto the best-fit plane to approximate the experimental methodology and bend angles were calculated using SerraLINE. The bend angle was defined

as the angle between vectors joining points 30 bp apart along the helix axis, where these vectors were separated by a further 30 bp centered on the binding site.

Hydrogen bonds were determined using cpptraj with a distance cutoff of 3.5Å and an angle cutoff of 120°, and the time-average number of intermolecular hydrogen bonds involving each residue was calculated; in all plots, this value is capped at 1.

Free-energy Calculation

Umbrella sampling simulations were performed on the 61 bp construct in explicit solvent and the Grossfield implementation of the weighted histogram analysis method (*35*) was used to extract the potential of mean force (PMF). The simulations used pmemd's NMR restraints with a spring constant of 2 kcal mol^{-1} $Å^{-2}$ ($3kT$ $Å^{-2}$); the reaction coordinates were reduced over a series of 5ns simulations in 1Å increments from their positions in the minimized structure until the PMF was observed to increase sharply. The reaction coordinates for the left- and right-hand sides were chosen as the distances between backbone atoms that are close together in the crystal structure: A75 P–Pro18β Cα and A13 P–Ser19α Cα, respectively. Generally, 300 histogram bins were used in the WHAM analysis.

For each arm, two sets of umbrella-sampling simulations were performed. In the first, the other arm was unrestrained and allowed to bind to the protein (which it did, in both cases within the first umbrella-sampling window). In the second, the other arm was held away from the protein by a potential with a spring constant of 2 kcal mol^{-1} $Å^{-2}$ if its reaction coordinate fell below 40Å. These two sets of results (which were in generally good agreement) were considered to represent the extremes of the other arm's position, and linear interpolation was performed between them.

Two-dimensional landscapes were then constructed by simple addition of energies and multiplication of probabilities; the relative probabilities of the clusters were calculated by integrating over the probability landscape when computed with 1Å resolution, with the allocation of grid squares to clusters weighted by the proportion of frames falling within the square that belong to each cluster, and the probabilities were normalized to sum to 100%.

To investigate bridge formation, an identical 61bp piece of DNA was pushed towards the IHF-DNA complex by gradually decreasing the distance between the backbone atoms closest to their centers of mass (A55 OP2–Phe81α Cα) using a one-sided harmonic potential with a spring constant of 2 kcal mol^{-1} $Å^{-2}$. Following the formation of a bridge between the two segments of DNA, umbrella sampling as described above was used to pull the second piece of DNA away from the protein, with the reaction coordinate until a plateau was observed in the PMF.

References

1. S. C. Dillon, C. J. Dorman, Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8** (2010), pp. 185–195.
2. R. T. Dame, The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56**, 858–870 (2005).
3. D. C. Grainger, D. Hurd, M. D. Goldberg, S. J. W. Busby, Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Res.* (2006), doi:10.1093/nar/gkl542.
4. N. M. Luscombe, S. E. Austin, H. M. Berman, J. M. Thornton, An overview of the structures of protein-DNA complexes. *Genome Biol.* **1** (2000), doi:10.1186/gb-2000-1-1-reviews001.
5. P. A. Rice, S. W. Yang, K. Mizuuchi, H. A. Nash, Crystal structure of an IHF-DNA complex: A protein-induced DNA U-turn. *Cell*. **87**, 1295–1306 (1996).
6. S. D. Goodman, N. J. Velten, Q. Gao, S. Robinson, A. M. Segall, In vitro selection of integration host factor binding sites. *J. Bacteriol.* **181**, 3246–55 (1999).

7. Y. X. Huo, Y. T. Zhang, Y. Xiao, X. Zhang, M. Buck, A. Kolb, Y. P. Wang, IHF-binding sites inhibit DNA loop formation and transcription initiation. *Nucleic Acids Res.* (2009), doi:10.1093/nar/gkp258.
8. J. K. Nuñez, L. Bai, L. B. Harrington, T. L. Hinder, J. A. Doudna, CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol. Cell*. **62**, 824–833 (2016).
9. D. S. Hwang, A. Kornberg, Opening of the Replication Origin of Escherichia coli by DnaA Protein with Protein HU or IHF. *J. Biol. Chem.* **267**, 23083–23086 (1992).
10. G. Laxmikanthan, C. Xu, A. F. Brilot, D. Warren, L. Steele, N. Seah, W. Tong, N. Grigorieff, A. Landy, G. D. Van Duyne, Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *Elife*. **5**, e14313 (2016).
11. T. A. Azam, A. Iwata, A. Nishimura, S. Ueda, A. Ishihama, Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J. Bacteriol.* **181**, 6361–6370 (1999).
12. S. Wang, R. Cosstick, J. F. Gardner, R. I. Gumport, The Specific Binding of Escherichia coli Integration Host Factor Involves both Major and Minor Grooves of DNA. *Biochemistry*. **34**, 13082–13090 (1995).
13. B. M. J. Ali, R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi, J. Stavans, Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10658–10663 (2001).
14. J. Lin, H. Chen, P. Dröge, J. Yan, Physical Organization of DNA by Multiple Non-Specific DNA-Binding Modes of Integration Host Factor (IHF). *PLoS One*. **7**, e49885 (2012).
15. L. A. Novotny, A. O. Amer, M. E. Brockson, S. D. Goodman, L. O. Bakaletz, Structural Stability of Burkholderia cenocepacia Biofilms Is Reliant on eDNA Structure and Presence of a Bacterial Nucleic Acid Binding Protein. *PLoS One*. **8**, e67629 (2013).
16. J. E. Gustave, J. A. Jurcisek, K. S. McCoy, S. D. Goodman, L. O. Bakaletz, Targeting bacterial integration host factor to disrupt biofilms associated with cystic fibrosis. *J. Cyst. Fibros.* **12**, 384–389 (2013).
17. G. H. Seong, E. Kobatake, K. Miura, A. Nakazawa, M. Aizawa, Direct atomic force microscopy visualization of integration host factor-induced DNA bending structure of the promoter regulatory region on the Pseudomonas TOL plasmid. *Biochem. Biophys. Res. Commun.* **291**, 361–366 (2002).
18. R. T. Dame, J. van Mameren, M. S. Luijsterburg, M. E. Mysiak, A. Janićijević, G.

Pazdzior, P. C. van der Vliet, C. Wyman, G. J. L. Wuite, Analysis of scanning force microscopy images of protein-induced DNA bending using simulations. *Nucleic Acids Res.* **33**, 1–7 (2005).

19. L. M. Hales, R. I. Gumport, J. F. Gardner, Examining the contribution of a dA+dT element to the conformation of Escherichia coli integration host factor-DNA complexes. *Nucleic Acids Res.* (1996), doi:10.1093/nar/24.9.1780.
20. S. V. Kuznetsov, S. Sugimura, P. Vivas, D. M. Crothers, A. Ansari, Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18515–18520 (2006).
21. S. Dixit, M. Singh-Zocchi, J. Hanne, G. Zocchi, Mechanics of binding of a single integration-host-factor protein to DNA. *Phys. Rev. Lett.* **94**, 118101 (2005).
22. Y. Velmurugu, P. Vivas, M. Connolly, S. V Kuznetsov, P. A. Rice, A. Ansari, Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein. *Nucleic Acids Res.* **46**, 1741–1755 (2017).
23. M. Connolly, A. Arra, V. Zvoda, P. J. Steinbach, P. A. Rice, A. Ansari, Static Kinks or Flexible Hinges: Multiple Conformations of Bent DNA Bound to Integration Host Factor Revealed by Fluorescence Lifetime Measurements. *J. Phys. Chem. B*. **122**, 11519–11534 (2018).
24. J. W. Shepherd, R. J. Greenall, M. I. J. Probert, A. Noy, M. C. Leake, The emergence of sequence-dependent structural motifs in stretched, torsionally constrained DNA. *Nucleic Acids Res.* **48**, 1748–1763 (2020).
25. M. C. Leake, The physics of life: One molecule at a time. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120248 (2013).
26. J. A. Fyfe, J. K. Davies, An AT-rich tract containing an integration host factor-binding domain and two UP-like elements enhances transcription from the pilEp1 promoter of Neisseria gonorrhoeae. *J. Bacteriol.* **180**, 2152–9 (1998).
27. M. Hammel, D. Amlanjyoti, F. E. Reyes, J. H. Chen, R. Parpana, H. Y. H. Tang, C. A. Larabell, J. A. Tainer, S. Adhya, HU multimerization shift controls nucleoid compaction. *Sci. Adv.* **2**, e1600650 (2016).
28. S. Yin, W. Bushman, A. Landy, Interaction of the λ site-specific recombination protein Xis with attachment site DNA. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1040–1044 (1985).
29. O. Scholder, pySPM v0.2.20 (2019), , doi:10.5281/ZENODO.2650457.

30. S. Van Der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, Scikit-image: Image processing in python. *PeerJ* (2014), doi:10.7717/peerj.453.
31. D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. . Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, AMBER 18 (2018).
32. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
33. I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, M. Orozco, Parmbsc1: A refined force field for DNA simulations. *Nat. Methods*. **13**, 55–58 (2015).
34. T. Sutthibutpong, S. A. Harris, A. Noy, Comparison of Molecular Contours for Measuring Writhe in Atomistic Supercoiled DNA. *J. Chem. Theory Comput.* **11**, 2768–2775 (2015).
35. A. Grossfield, WHAM: the weighted histogram analysis method, (available at http://membrane.urmc.rochester.edu/wordpress/?page_id=126).

Acknowledgements

We thank the Physics of Life (PoL) Group, University of York for providing pump-priming resources. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [EP/N027639/1, EP/T002166/1, EP/R029407/1 and EP/P020259/1]; Biology and Biotechnology Research Council (BBSRC) [BB/R001235/1]; Leverhulme Trust [RPG-2017-

340]. Calculations were performed on ARCHER, JADE, Cambridge Tier-2 and the local York facilities (Viking and YARCC clusters).

Figures and Tables

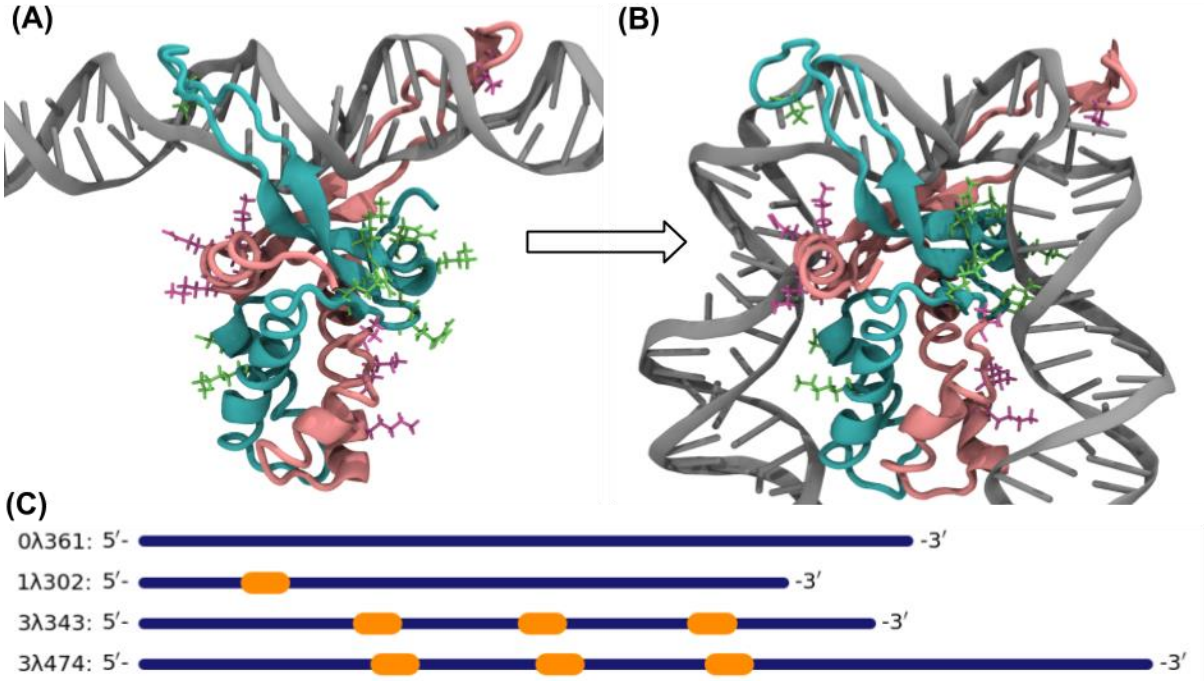


Fig. 1. Molecular structure of IHF and DNA when canonically bound as determined by molecular dynamics (MD) simulations. (A) Initial state for MD simulations, with the α subunit shown in pink and the β subunit in blue; the intercalating prolines are highlighted, and the DNA is shown in gray. **(B)** Final state of a representative canonical simulation. **(C)** DNA constructs used in experiments with binding sites highlighted (1λ302 simulated), shown to relative scale.

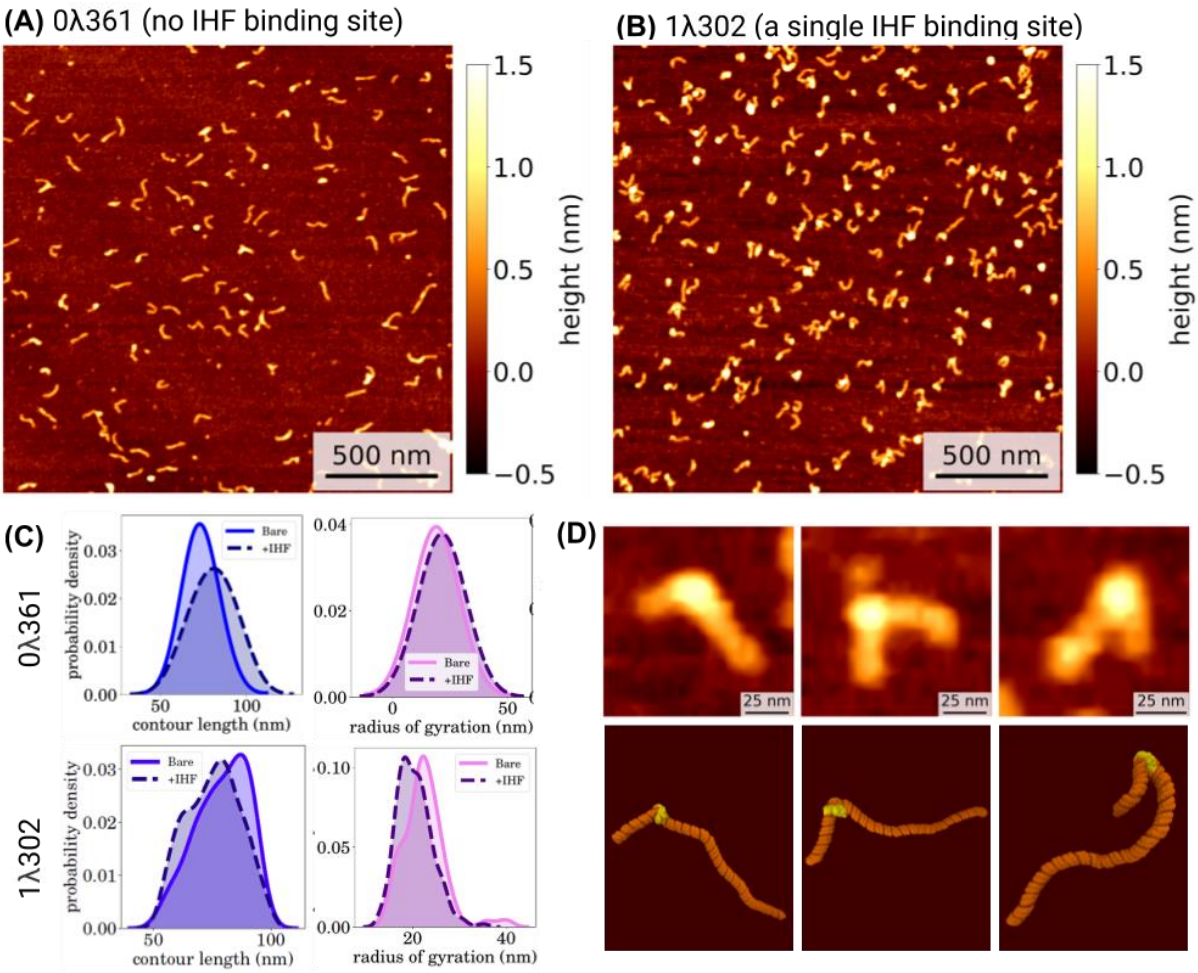


Fig. 2. Atomic force microscopy (AFM) images of IHF+DNA, with and without IHF binding sites. (A) AFM image of DNA without IHF binding site (0λ361), showing minimal binding when compared to **(B)** AFM image of DNA with an IHF binding site (1λ302). **(C)** Kernel density estimates of the contour lengths and radius of gyration of DNA ± IHF, with a significant reduction in the radius of gyration only seen for 1λ302. **(D)** Three representative examples of IHF binding, with a comparison of AFM (top) and molecular dynamics simulations (bottom).

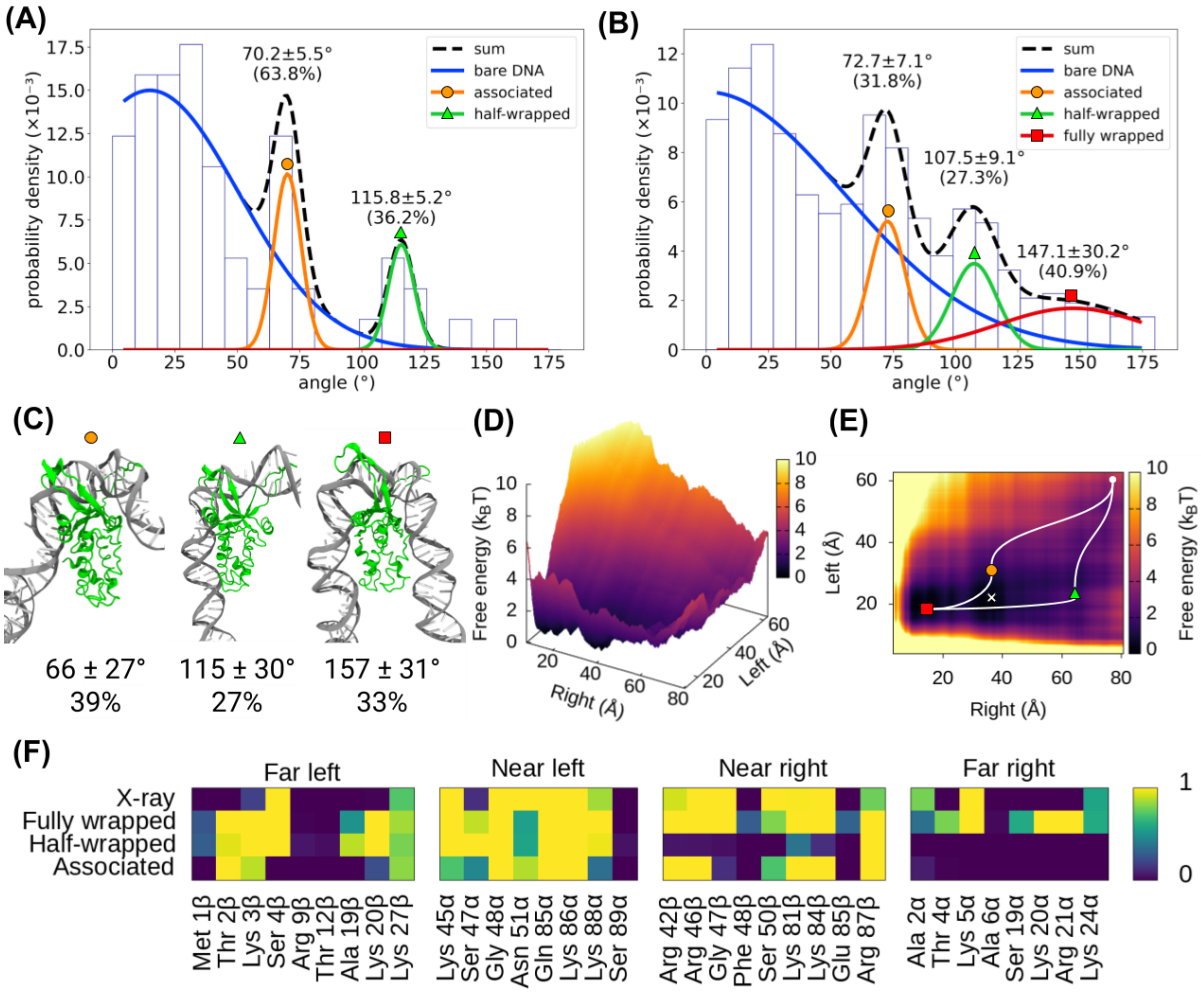


Fig. 3. Analysis of different bending angles of IHF-bound DNA. (A) Bending angle distribution of 0λ361 with IHF. There is a considerable amount (72%) of unbound DNA, with the unbound distribution set by 0λ361 without IHF. The remaining peaks of the distribution are well fitted (as shown by the reduced χ^2 in Figure S4) by two Gaussians, representing three substates with bending angles of ~70˚, ~116˚. **(B)** For 1λ302+IHF, the two IHF-induced substates are again present (~73˚, ~108˚) as well as an additional substate (reduced χ^2 in Figure S4) appears with a more canonical angle of ~147˚. **(C)** Simulations of the same construct can be classified into three clusters with mean bending angles that correspond well with the experimental data; the relative probabilities of these clusters can be determined from weighted-histogram analysis, which results in the free-energy landscape **(D)**. The clusters correspond to minima and plateaus in this free-energy landscape, and the

observed cluster transitions provide a guide to the paths by which the system traverses the landscape **(E)**. The clusters can also be categorized based on the DNA–protein hydrogen bonds present in the structure **(F)**; while the fully wrapped DNA interacts with both subunits of the protein on both sides, the half-wrapped DNA interacts only on the left, and the associated state interacts primarily with the “near” subunit on each side. This right-left asymmetry is reflected in the free-energy landscape, which features a large free-energy change as the left-hand side binds that is not present on the right-hand side. “X-ray” refers to the interactions present in a short explicitly solvated simulation of the 1IHF crystal structure.

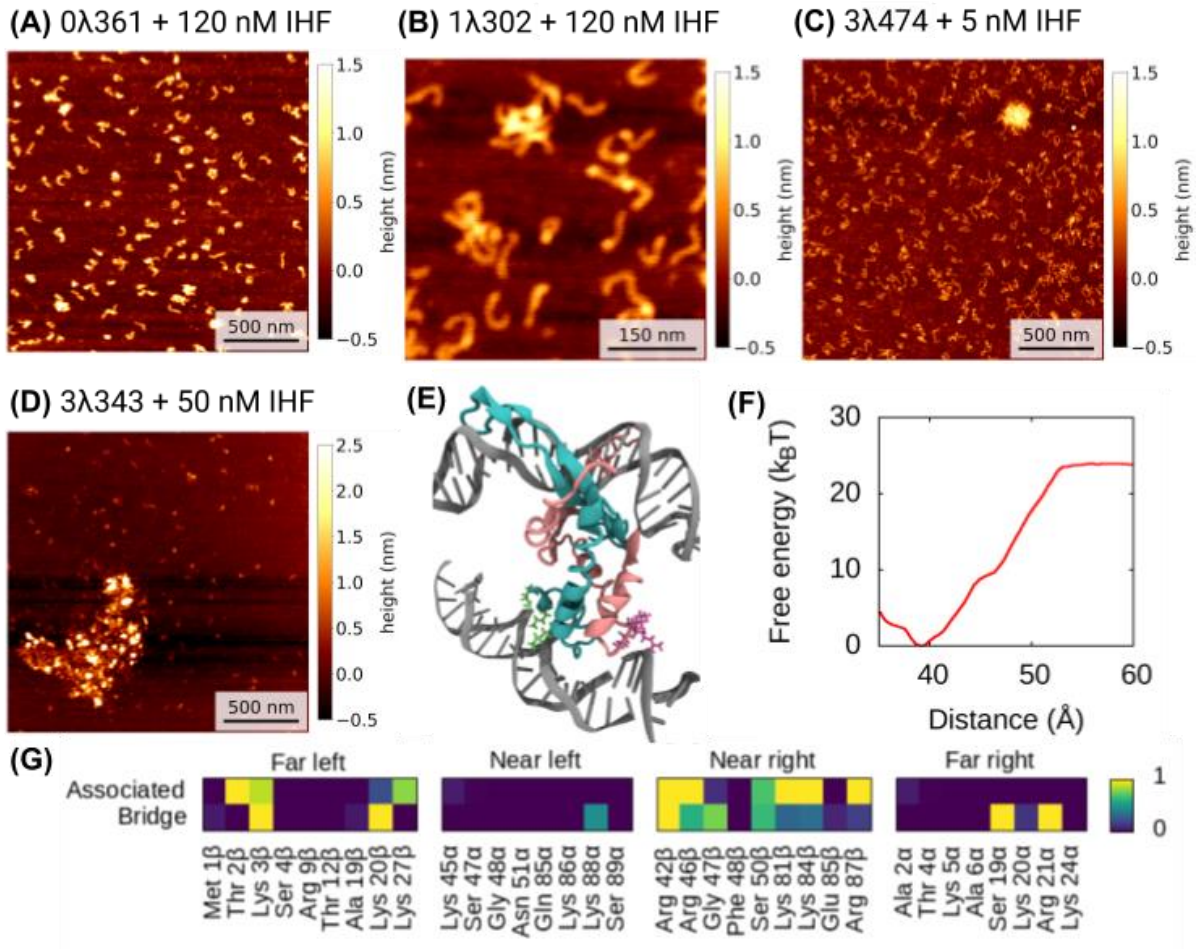


Fig. 4. AFM images of aggregates on different concentrations of IHF and different constructs (at 1 nM). (A) With no binding site a large amount of IHF leads to some aggregation, **(B)** similar to that of DNA with a single IHF binding site. **(C)** DNA constructs with 3 IHF binding sites form clusters larger than that of the previous constructs even at low concentrations of IHF **(D)** and when the concentration of IHF is increased a three-binding-site construct form large extended structures. **(E)** Bridge structure from MD simulation, **(F)** Free-energy landscape, **(G)** Contact maps of bridge.

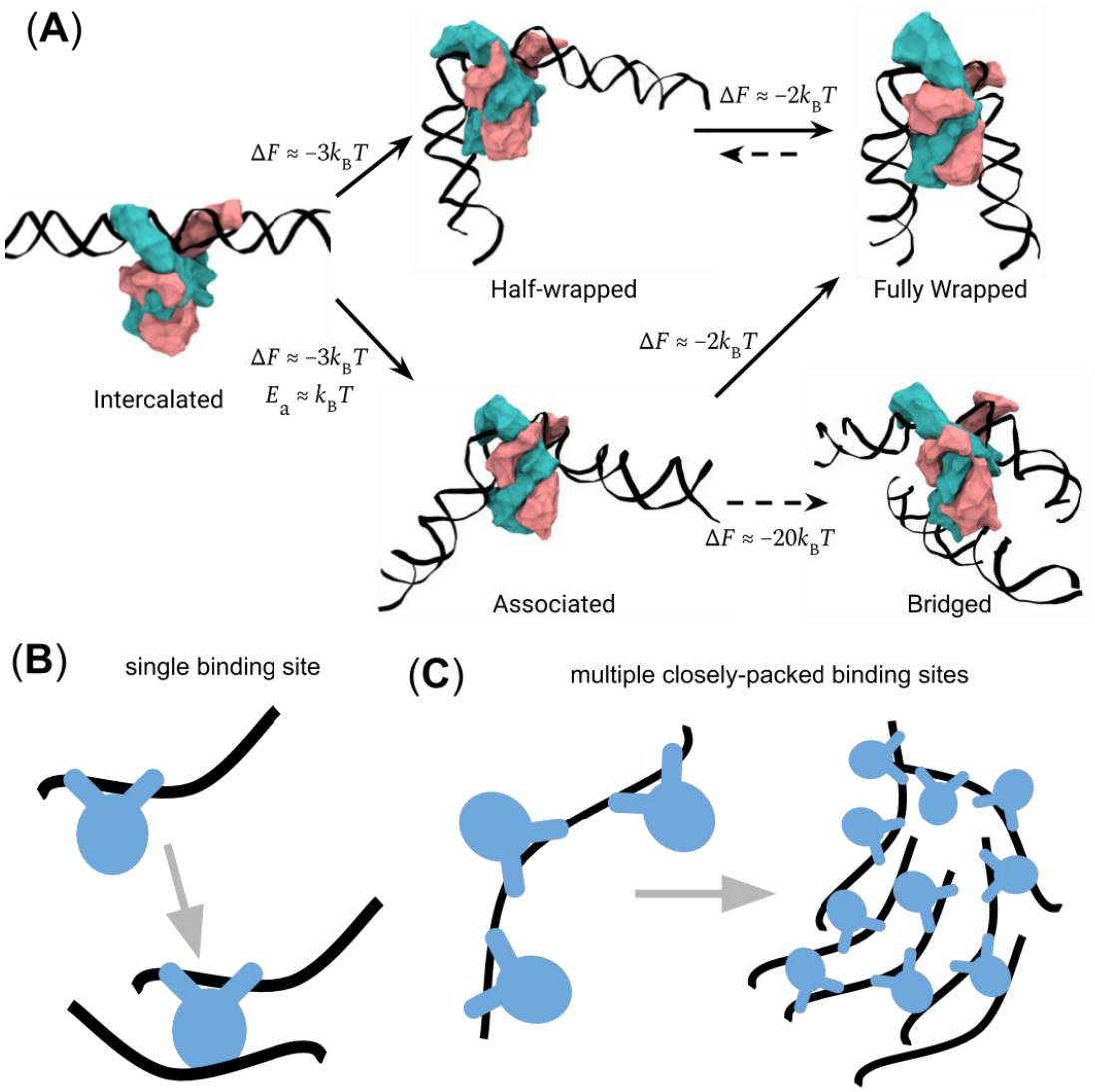


Fig. 5. Model of IHF binding – bending and bridging. (A) IHF first intercalates with the DNA. The first binding step appears to be either a loose association to the DNA (73˚ binding) or a half-wrapped state (120˚). Both of these can then progress to the canonical fully wrapped state. Although not observable by our experiments, MT experiments show reversibility between the PW and AS. states. As the associated state leaves the bottom half of IHF unbound it can also bind to another region/strand of DNA non-specifically resulting in a bridge. **(B)** A single binding site would only form a single bridge whereas **(C)** multiple binding sites or a large concentration of IHF would form multiple bridges causing aggregation.

Supplementary Materials

Table S1. Primers for DNA amplification by PCR of sequences used, amplified from phage λ.

Primer Name	Sequence
0λ361-F	5`- GCATCATCAAGTGCCGGTCG -3`
0λ361-R	5`- TGCTGTTGGTTGCACTGCTG -3`
1λ302-F	5`- CAAGACACCGGATCTGCAC -3`
1λ302-R	5`- GCATATGATGTCTGACGCTGG -3`
3λ343-F	5` - CTTTGTGCTTCTCTGGAGTGCG – 3`
3λ343-R	5` - GGCAGGGAGTGGGACAAAATTG – 3`
3λ474-F	5` - GATTGCGAGGCTTTGTGCTT - 3`
3λ474-R	5` - CTACCTTTCACGAGTTGCGC - 3`

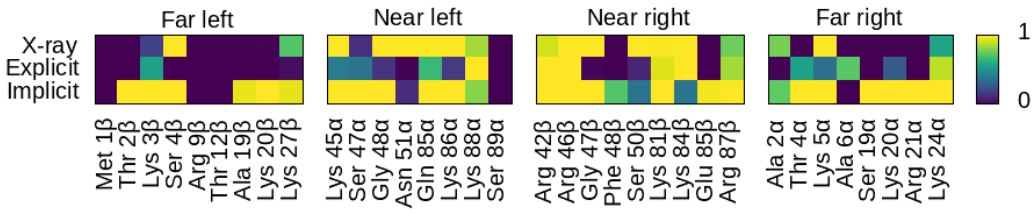


Fig. S1. MD reproduces canonical interactions. The main DNA–protein hydrogen bonds present in a short simulation of the 1IHF crystal structure are also present in an explicit-solvent simulation of a straight 61bp construct and the most canonical replicas of the 302bp construct in implicit solvent. Note that the crystal structure is too short to capture some of the far contacts.

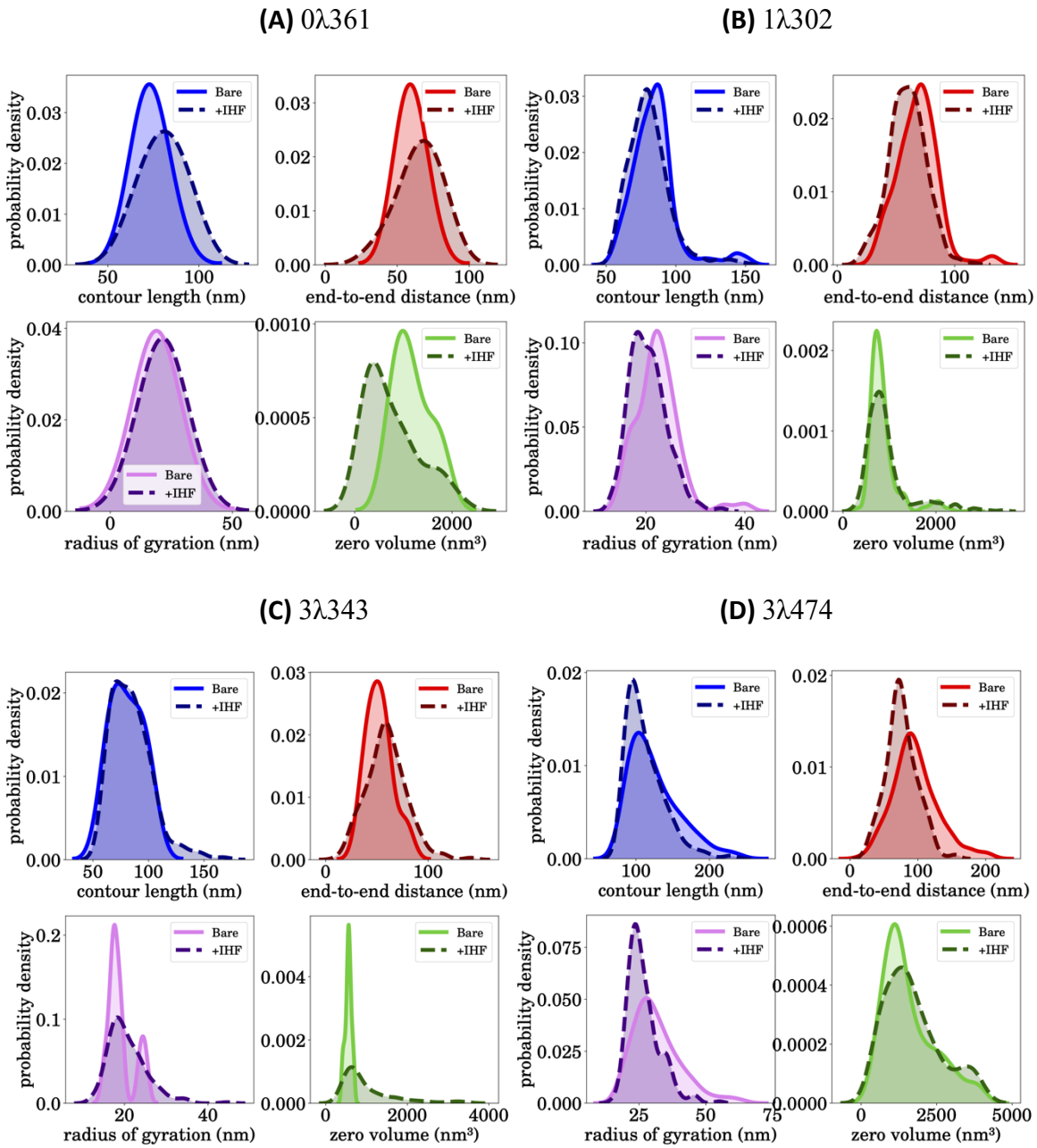
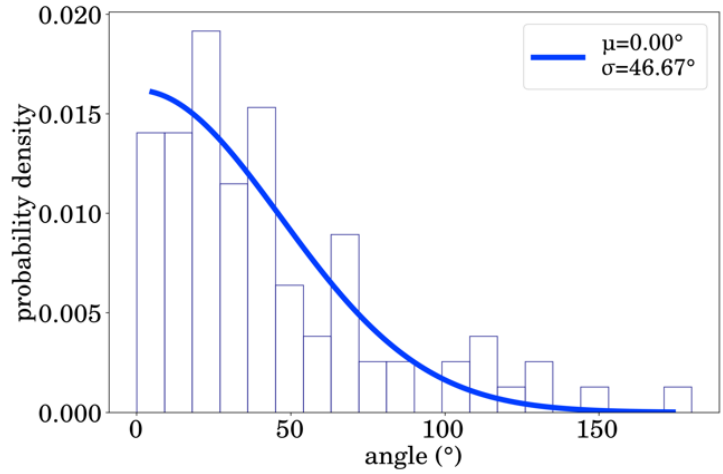


Fig. S2. Kernel density estimates for properties of all DNA constructs used +/- IHF. (A) 0λ361, **(B)** 1λ302, **(C)** 3λ343 and **(D)**/3λ474. For all of these the contour lengths remain the same, but only for 1λ302 and 3λ474 do the radius of gyration and the end-to-end distance reduce upon addition of IHF (3λ343 was too prone to cluster).

(A) 0λ361



(B) 1λ302

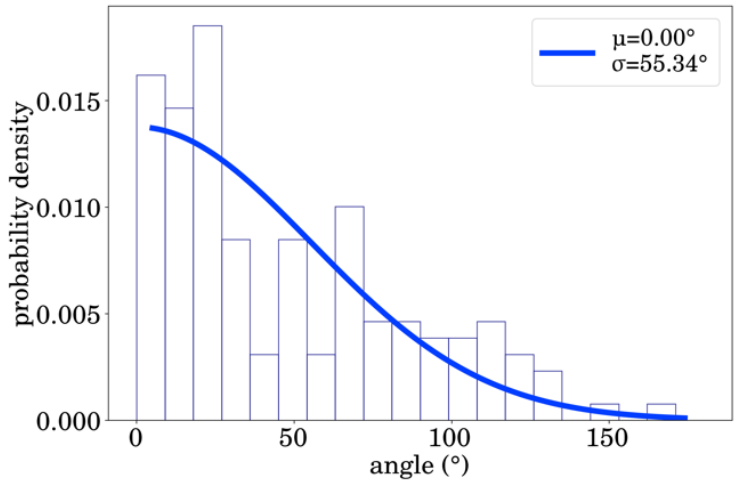
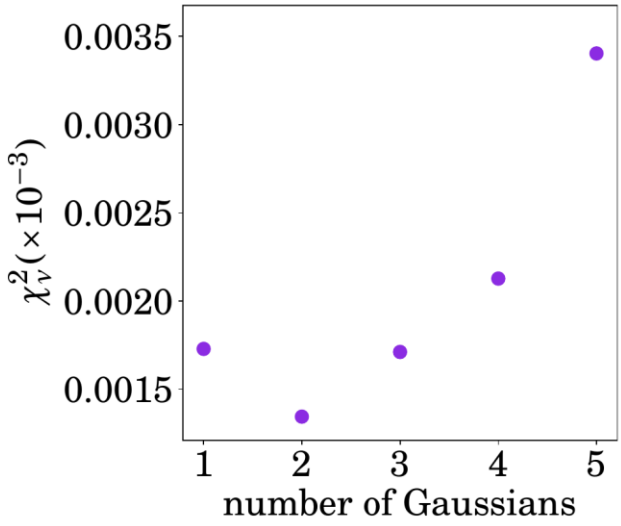


Fig. S3. Angle distributions of 0λ361 and 1λ302 without IHF from AFM images. Angle distributions for **(A)** 0λ361 and **(B)** 1λ302 without IHF, showing a Gaussian centered around zero with a width of 47˚ and 55˚ respectively, suggesting that the 1λ302 construct is slightly more flexible, perhaps due to sequence-specific curvature.

(A) 0λ361



(B) 1λ302

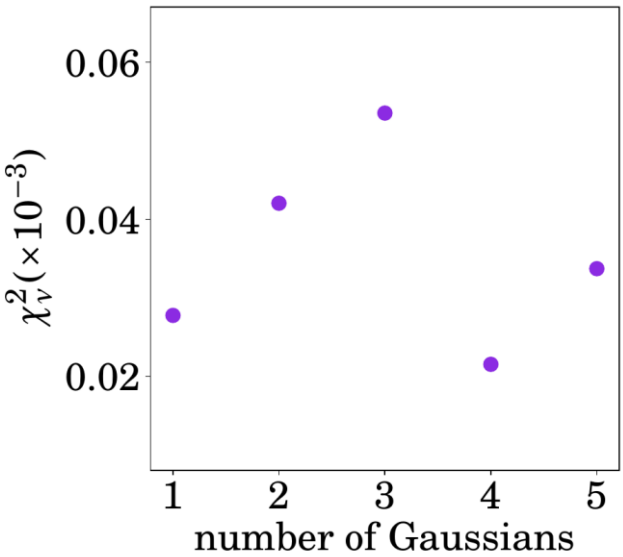
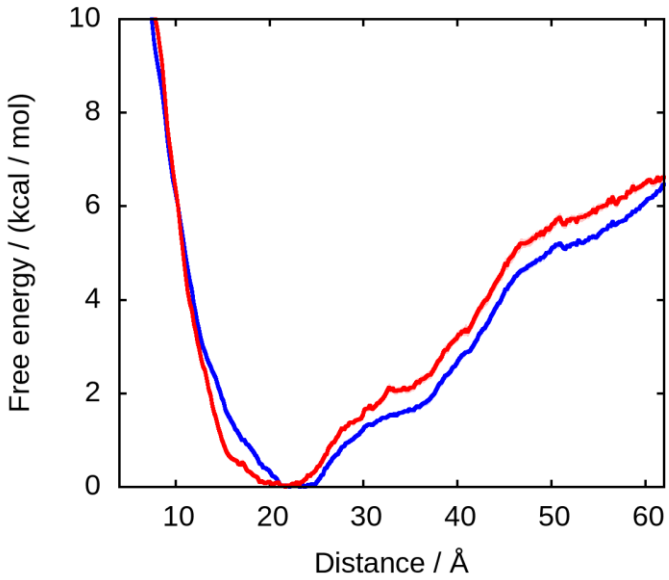


Fig S4. Reduced χ^2 fits for 0λ361 and 1λ302 with IHF. (A) For 0λ361 (no binding site) the data is fit well by either 2 or 3 Gaussians, suggesting an unbound and 1-2 bound states. **(B)** For 1λ302 the reduced χ^2 is lowest for four Gaussians, suggesting that there are four states, the fixed unbound state and three bound states.



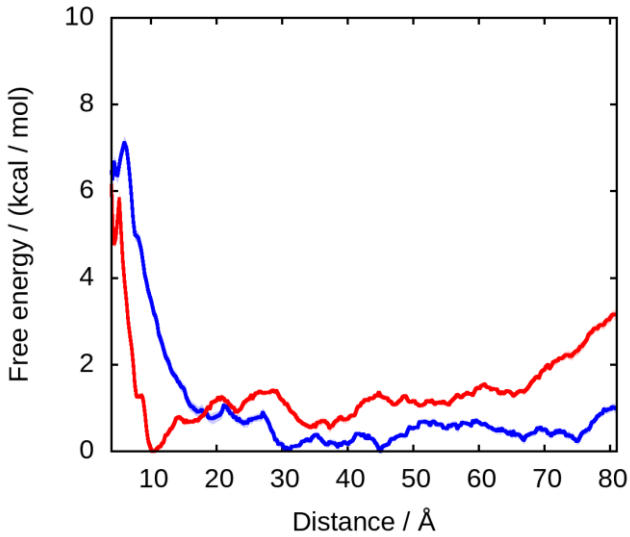


Fig. S5. Demonstration of independence of potentials of mean force for each arm. The forms of the potentials for the left (left) and right (right) arms are similar whether the other arm is free (red) or restrained (blue), suggesting that the reaction coordinates are approximately independent. On the right-hand side, the PMF is dominated by noise in both cases.

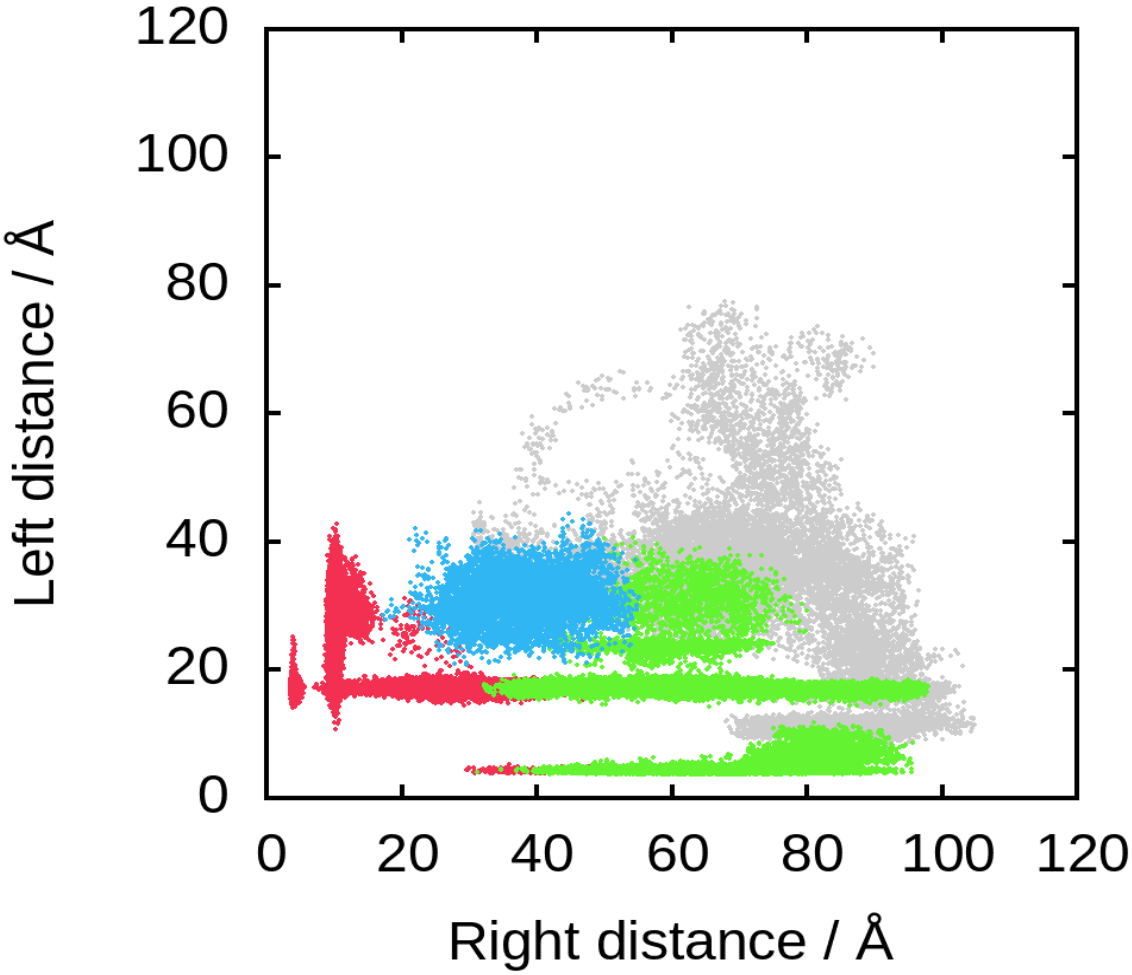


Fig. S6. Values of reaction coordinates by cluster. The frames of the fully wrapped cluster are shown in red, half-wrapped in green, and associated in blue. Gray dots represent states present only in the first 5ns of each replica.

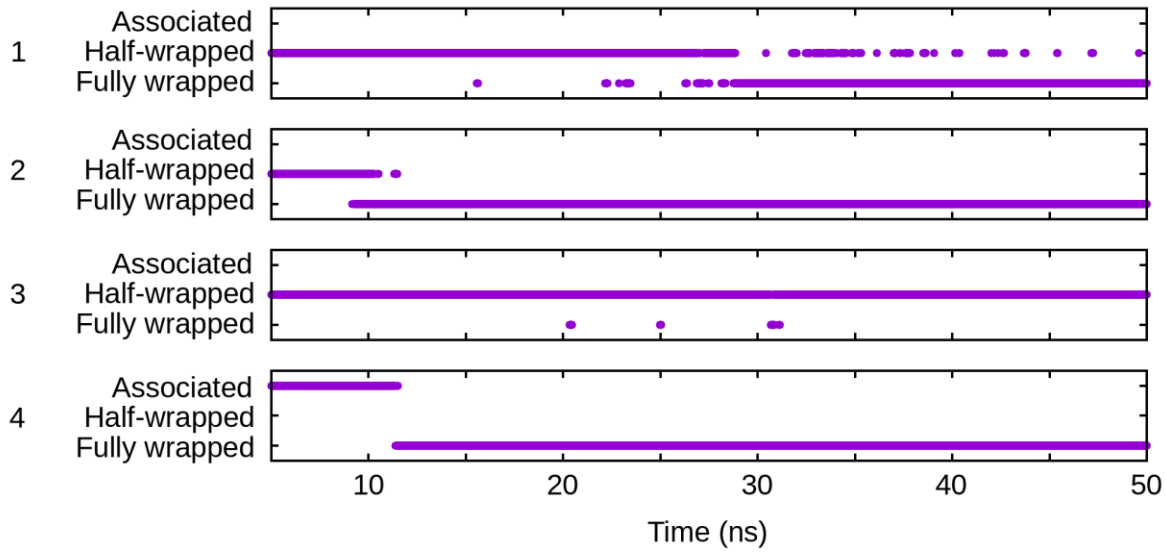


Fig. S7. Cluster-vs-time for each replica. By identifying the cluster to which each frame of a simulation belongs, the possible transitions can be observed. These are half-wrapped to fully wrapped (replicas 1 and 2) and associated to fully wrapped (replica 4). Every replica spends some time in the associated or half-wrapped state before transitioning to the canonical fully wrapped state.