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

Samuel B. Yoshua[1†], George D. Watson[1†], Jamieson A. L. Howard[1], Victor Velasco-Berrelleza[1], Mark C. Leake[1,2*], Agnes 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 are crucial in compacting prokaryotic DNA, but how they generate higher-order structures 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 structural multimodality is far more elaborate than previously proposed and provide insights into how this enables mechanical molecular switching. Using single-molecule atomic force microscopy and all-atom molecular dynamics simulations we find three topological modes in roughly equal proportions: “associated” (73˚ of DNA bend), “half-wrapped” (107˚) and “fully-wrapped” (147˚), with only the latter occurring with sequence specificity. DNA bridging is seen in the presence of multiple binding sites, with simulations showing this occurs through another recognition mode that does not depend on sequence and gives IHF a stoichiometry greater than one. 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 observed 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*) such as gene regulatory loops (*7*), 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 IHF alone is 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% 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*). It 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, the degree of bending and the length of the attached DNA site (*19*).

Recent studies have shown that IHF can produce additional DNA bends smaller than the canonical 160˚. IHF can bend DNA flexibly with a portion of the population that is only partially bent (*20*, *21*), demonstrating that the complex is more dynamic than originally thought. Previous work has also shown binding occurs through a two-step process, a fast ~100 µs step that seems to be not sequence-specific, and a slower millisecond step which is site-specific (*22*). This second step seems be associated with an activation energy of approximately 14 kcal/mol (equivalent to $24k_BT$ where k_B is the Boltzmann constant, absolute temperature T) that would be related with proline intercalation (*23*), but the free energy of wrapping is found to be only around 3.6 kcal/mol ($6k_BT$) (*24*). Despite these advances, the underlying bending mechanism driven by IHF is still poorly understood, 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 sequences of around 300 bp (Figure 1D). This approach allowed us to show how a strong binding site, H2 (*25*, *26*), has a multiplicity of binding modes beyond the canonical 160˚ state. We then formulated a model for multiple IHF bending states encapsulating both specific and non-specific binding. Using advanced modeling methods, we calculate the free energy of each of these conformations and demonstrate asymmetric behavior between DNA arms that allows different mechanical switching depending upon the order of DNA arm binding. Furthermore, by looking at the effects of increasing the number of IHF binding sites using AFM, we see evidence for IHF-mediated bridging between two DNA strands at low protein concentrations and provide structural insight into how this occurs. The variety of DNA topological modes caused by IHF – both bending and bridging - may be representative of the behavior between DNA and other nucleoid-associated proteins.

Results

Modeling DNA wrapping around IHF

To describe the bending imposed by IHF, we created an initial structure with the non-curved H2 binding site attached to the protein via only its protruding β−ribbon arms (Figure 1A). We embedded the complex in a short 61 bp DNA construct for simulations in explicit solvent and in a longer 302 bp construct for simulations in implicit solvent (see Materials and Methods).

Starting from the open state shown in Figure 1A, the crystallographic bending can be reproduced, as in Figure 1B. In both cases the observed DNA–protein interactions correspond well to those present in the 1IHF crystal structure (*5*) (Figure 1E), including the insertion of Ser47α and Arg46β residues into the minor grooves of lateral DNA (Figure 1C) and the additional interactions between the DNA backbone and positively-charged and polar amino acids.

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 closer to the center and thus constitutes the "near left" side, while the β subunit is farther and composes the "far left". On the right-hand side (containing the consensus sequence), the α and β subunits are inversely arranged, delimiting the "far right" and "near right" sides, respectively (see Figure 1).

While the DNA in the crystal structure is too short to capture some of the more distant interactions (Figure 1E), there is generally strong agreement with simulations starting from unbent and intercalated DNA. This observation demonstrates the validity of the molecular dynamics methodology for this system, as well as suggesting an explanation for the previously proposed two-step binding mechanism. In this two-step model, the IHF arms would bind to DNA first and then the proline residues would intercalate to induce flexible hinges prior to wrapping. Although our simulations start from straight DNA, it is possible that some bending may occur following the initial recognition that would facilitate the intercalation of prolines and the subsequent strong kink. In any case, the binding mechanism that we propose explains both the high activation energy of initial binding (*23*) due to proline intercalation and the smaller free energy of wrapping (*24*).

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 1D). These were then imaged using AFM and compared to molecular dynamics (MD) simulations to determine the different morphological behaviors.

DNA contours were recovered by skeletonizing pre-processed AFM images (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 S1) shows the reduction of the radius of gyration of the 1λ302 construct if IHF is present, suggesting that the protein bends DNA for the single-binding-site case in this assay. However, as the radius of gyration can only determine overall compaction, more precise bending angle analysis was also performed (see Materials and Methods).

When 50 nM IHF was incubated with 1 nM DNA (1 IHF: 6 bp) before imaging with AFM, a large proportion (~60-70% of the total) of the resultant angle distributions was due to the behavior of bare DNA molecules (Figures S2A-B), which was accounted for in the subsequent analysis. The addition of IHF leads to a further two and three peaks in the angular distributions for 0λ361 (Figure 3A) and 1λ302 (Figure 3B), respectively (see Figures S2C-D). The peaks around 73 ± 7˚ and 107 ± 9˚ (mean ± s.d. values from 1λ302) are common to both constructs, suggesting that they occur due to non-specific binding, whilst the canonical large bending angle (147 ± 30˚), as seen in the crystal structure (*5*), only appears in 1λ302 that contains one specific IHF binding site.

The proportions of the 73˚, 107˚ and 147˚ states (with proportions of approximately 32%, 27% and 41% respectively) show that all three binding modes are present in roughly equal quantities for the construct with a specific binding site (1λ302). However, the proportion of the 73˚ state (~64%) to the 107˚ (~36%) state is much larger for 0λ361, suggesting this mode is unrelated to the consensus sequence. However, this binding could still be dependent upon AT-content as similar sequences motifs to the left part of the specific binding site (an A-tract followed by AT base pairs downstream) are present in the sequence of 0λ361.

Our results are largely consistent with previous AFM studies where a broad distribution of angles was detected between approximately 80-150˚ (*17*, *18*). 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 (Figures S2A-B) or due to the selection of an H2 binding site.

Multiple MD simulations were performed of the 1λ302+IHF construct in implicit solvent (Figure 2D, Supplementary Movies 1-4), starting from the bound but unbent state

shown in Figure 1A. The frames of these simulations were classified into three groups through hierarchical agglomerative clustering, finding groups with bend angles of 66 ± 27°, 115 ± 30°, and 157 ± 31° and radii of gyration of 24.5 ± 0.7 nm, 22.12 ± 0.02 nm, and 20.78 ± 0.02 nm respectively, in good agreement with the AFM results (see Figure 2C and Figure 3C). A representative frame was selected for each group (Figure 3C). The clusters are also characterized by the hydrogen bonds between the protein and DNA (Figure 3D). In the "fully wrapped" state (with a bend angle of 157°), both sides of the DNA form hydrogen bonds with both subunits of the protein; in the "half-wrapped" state (115°), the left-hand side, which contains the A-tract, forms contacts with both subunits while the right-hand side forms no hydrogen bonds; and in the "associated" state, the left and right near sides are the main ones forming hydrogen bonds with the DNA.

The free energies and relative probabilities of each of these states were determined via umbrella sampling simulations (*27*, *28*) of a 61 bp segment of the construct in explicit solvent (see Materials and Methods). The distance between the bottom part of the protein and the interacting DNA on each side was varied, once with the other arm allowed to wrap and once with the other arm held away from the protein. The potential of mean force (PMF) for each simulation was calculated using the weighted histogram analysis method (WHAM) (*29*, *30*). By linearly interpolating between the two sets of results for each arm (Figure S3A–B) a two-dimensional free energy landscape was constructed with the reaction coordinates as orthogonal axes, as in Figure 3E.

The relative probabilities of the clusters were determined by integrating over the relevant regions of the free energy landscape (Figure 3E and Figure S3C). This analysis predicts that the canonical fully wrapped (157°) state should be observed around 32% of the time, the half-wrapped (115°) state 50%, and the associated (66°) state 18%; while these figures do not directly agree with AFM, probably due to the differences between the simulation and experimental conditions, they do predict that all three states should occur in significant proportions. 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 (Figure S3D), are highlighted in Figure 3F.

Asymmetry between DNA arms causes a mechanical switch

As discussed in the previous section, it is notable that the consensus sequence does not interact with IHF in the half-wrapped state and this state is also observed in the absence of a specific binding site. The free energy surface reveals that there is a large activation barrier preventing the right-hand side (which contains the consensus sequence) from binding before the left-hand side (which contains the AT-tract) (Figure 3E). The form of the potentials is also qualitatively different. While the left-hand side presents a relatively deep potential with a minimum at around 34 Å and an additional small plateau at around 40 Å, implying that binding is energetically favorable (Figure S3A), the right-hand side has a much flatter shape with small minima around 31 Å and 43 Å, and appears to be dominated by thermal noise (Figure S3B).

While the free-energy landscape presented by the left-hand side appears to be independent of the position of the right arm (Figure S3A), the right-hand side is prevented from binding fully by a large barrier when the left arm is held away from the protein; this barrier is not present when the left arm is fully bound (Figure S3B). This barrier is associated with a change in the structure of the protein. When the left arm is unbound, the upper subunit of the protein on the right-hand side protrudes, so full binding of the right arm would require extra DNA bending apart from the flexible region. Binding of the left arm pulls the protein body leftwards, removing this physical barrier and allowing the DNA on the right-arm to bind while remaining mostly straight (Figure 3G).

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, and that the A-tract must bind first (Supplementary Movie S5). 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 (Figure 3C). The associated state corresponds to the smaller local plateau or minimum in each arm's energy landscape (Figure 3F, 3SA-B), as seen when the values of the reaction coordinates are plotted together for each frame of a cluster as in Figure S3C.

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 initial binding of IHF to DNA occurs without sequence specificity and a consensus sequence enhances bending.

The free energy landscape features a significant minimum that appears to correspond to another state (represented by a white cross in Figure 3F), in which the left-hand side is fully bound and the right-hand side is partially bound (Figure S3E). However, the frames in this region are difficult to distinguish from those in the fully wrapped state as the bending angles are very similar, and the corresponding minima in the free-energy landscape are separated by only a narrow energy barrier of less than $2k_BT$. Since the system can move freely between these two minima under the influence of thermal noise, it is more appropriate to consider this state an extension of the fully wrapped mode.

The asymmetric cooperative behavior observed between arms in the presence of a specific sequence could serve for reinforcing the prevalence of the fully-wrapped state as it reduces the amount of accessible states that the system would present in the case of both arms moving in a totally uncorrelated manner. This interconnection between the two halves would help the IHF to sharply bend the DNA with a substantial probability in a similar way to a mechanical switch.

DNA aggregation and bridging by IHF

Further increases of the IHF concentration causes the formation of clusters for both 0λ361 and 1λ302 (Figures 4A-C). This behavior at high concentration (~1 IHF:3 bp) has previously been seen (*14*) and shows how large cellular concentrations of IHF – such as in the stationary phase – could also increase genome compaction by bridging.

A construct with three binding sites (3λ474) behaves very differently to the previous constructs that contain one or no binding sites. Even at low concentrations, from around 1 IHF: 10 – 25 bp (Figures 4D and 4E), a greater number of IHF/DNA clusters could be seen, until the IHF was numerous enough to coat the DNA reducing the size of visible clusters (Figure 4F). A shorter form of this construct (3λ343, 343 bp long) (Figure 4G) was seen to show larger aggregates. The reason for the prominence of aggregation on the latter construct is unclear but could be due to its much higher AT-content (70% *vs.* 63%) facilitating non-specific binding without IHF coating the DNA. Although these conglomerates may be due to protein

aggregation, the difference between these and the earlier constructs demonstrates that it is not the case.

IHF-induced aggregation can be explained by the bridging of two DNA strands by a single IHF, a phenomenon observed in our MD simulations (Figure 4H). These bridges result from non-specific interactions involving positively charged and polar amino acids from the far binding sites of the protein and the backbone of the DNA (Figure 4I). Umbrella-sampling simulations and weighted histogram analysis (see Materials and Methods) reveal that bridging is very energetically favorable when a second DNA strand is close enough, with a free energy difference on the order of 14kcal/mol ($24k_BT$) between the bridged and unbridged states (Figure 4J), while the main DNA strand, into which the prolines intercalate, remains mostly unwrapped, probably due to the electrostatic repulsion exerted between the two DNA strands (Figure 4I).

The combination of AFM and MD suggests that the clustering of constructs containing three binding sites is due to the capacity of IHF to bind more than one DNA strand at the same time. We observe that DNA·IHF·DNA bridging is favorable even for low concentrations of IHF. The strong electrostatic repulsion between close DNA strands could prevent the complete wrapping of DNA around the protein, even on sequence-specific sites, facilitating the bridging interaction described above.

Discussion

Here we have integrated advanced theory and experiment to explore the physics of life at the single-molecule level (*31*) to enable new insights into the complex interactions between proteins and DNA. Both AFM experiments and MD simulations show evidence of a larger number of bending modes of the IHF/DNA complex than previously proposed. The presence of three states (fully wrapped, half-wrapped and associated) as well as the large proportion of unbound DNA suggests a multi-modular system, where higher AT-content makes binding more probable (*32*, *33*) but in which a consensus sequence is needed to fully bend DNA. We put forward a model for the mechanism of IHF binding where full wrapping is sequence specific but initial binding is not (Figure 5).

First, we provide structural insights into the two-step binding mechanism previously proposed (*23*), as the crystallographic structure is reached by our simulations when started from a bound but unwrapped DNA (Figure 1). Thus, our results reinforce the idea that the protein first binds DNA non-specifically via its extended arms; the prolines at the tips of the arms would then intercalate with an activation energy around 24k_BT (*23*); from there, DNA wraps around the protein in a relatively downhill process stabilized by up to 15k_BT, a value calculated from our simulations and in broad agreement with single-molecule experiments (*24*).

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 part of the sequence. This results in the half-wrapped state, which is ~13k_BT more stable than the initial structure and 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 (favored by ~9k_BT). Transitions between the associated and half-wrapped states were not observed (Figure S3D), and both appear to be long-lived metastable states corresponding to plateaus or local minima in the free-energy landscape (Figure 3E). 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. We also note that the half-wrapped state is akin to one seen in structurally-similar HU, which does not bind to specific sequences (*34*). However, in the presence of a consensus sequence, both of initial states lead towards the global minimum ($\Delta F \approx -15k_BT$ relative to the initial structure), resulting in canonical binding.

Previous experiments (*14*, *20*, *21*) already indicated that IHF could bind DNA in more than one state, generally in two states which were broadly described as a fully and partially bent. Magnetic tweezers experiments appeared to show an smaller angle state ~50˚ (*14*) (i.e. similar to our the associated state). The lack of detection of a third state could be due to the fact that the force applied on the DNA was greater than the few tenths of piconewtons we predict is needed to overcome the potential barrier between the half-wrapped and fully wrapped states. By using fluorescence-lifetime-based FRET, Ansari and coworkers deduced that in fact the partially-bend form was composed by two sub-states although they did not have the precision to detect the structural properties of the non-canonical conformation. The

proportion of the fully-wrapped state was found to be higher in their experiments than in our results, but this difference might be caused by the use of a different binding site (H' vs H2).

The half-wrapped and associated states may not be the only conformations in which IHF binds non-specifically to DNA. For example, the similar electrostatic profiles of IHF and HU (see Figure S4) suggest that another non-specific binding mode similar to that observed by Hammel *et al.* (*35*) 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.

The existence of distinct states with different degrees of bending, and the dependence of the populations of these states on the position of the left arm following an asymmetric allostery, suggests that the IHF–DNA complex may act as a mechanical switch, its regulatory behavior switched on or off by tension or structural influences upstream of the binding site. The semi-stable states observed would allow IHF to remain associated with the DNA while retaining some flexibility, which could be important in the formation of higher-order nucleoprotein complexes such as a transcription regulatory loop. Interestingly, the consensus bases are within the region that is less tightly bound, which could illustrate 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. On the other hand, the constructs with multiple binding sites seem to preferably select for the bridging behavior over bending when DNA molecules are moderately concentrated. This behavior shows how the formation of bridges could be driven by the screening of electrostatic repulsion between neighboring DNA molecules by architectural proteins (*36*). This phenomenon could also be caused by steric hindrance or tension due to the presence of other proteins preventing complete wrapping by each IHF, leaving the lower regions of the protein free to bind other DNA. Our study not only explains 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. We observe that specific recognition can be simply widened by non-specific electrostatic-driven interactions between the protein and DNA (*37*, *38*), and we anticipate that this could be a general mechanism used by other NAPs to enable a variety of DNA bending and bridging modes.

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 dialyzed 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 (*39*).

Atomic force microscopy acquisition and analysis

Mica was freshly cleaved and then pre-treated by depositing 20 µl 0.01% (w/v) 1-5 kDa poly-L-lysine (Sigma-Aldrich, MO, USA), left for 5 mins before washing with 1000 µl filtered milliQ H2O and finally 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 approximately 5 mins. 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 (*40*) and preprocessed by using a 1st order line-by-line flattening and 2nd order fits to flatten the surface before filtering scars (*41*). The images were then analyzed using custom code using the scikit-image package (*42*) 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, discarding those that were not within 50% of the actual length. Angles were measured by first identifying IHF as the peak along the contour, then taking a 4 px (16 nm) vector either side of a 4 px window around the IHF. Once the angular distribution was determined goodness-of-fit tests were carried out, with $p > 0.99$.

Molecular Dynamics Simulations

All simulations were set up with the AMBER 18 suite of programs and performed using the CUDA implementation of AMBER's pmemd program (*43*). 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. The full constructs were solvated using the implicit generalized Born model (*44*) at a salt concentration of 0.2 M with GBneck2 corrections, mbondi3 Born radii set and no cut-off for a better reproduction of molecular surfaces, salt bridges and solvation forces (*45*, *46*). Langevin dynamics was employed for temperature regulation at 300 K with a collision frequency of 0.01 ps1 which reduces the effective solvent viscosity and, thus, accelerates the exploration of conformational space. The protein and DNA were represented using the ff14SB (*47*) and BSC1 (*48*) force fields, respectively. Prolines were restrained to remain intercalated by using NMR restraints to restrict the distances between key atoms in the proline side chain and the neighboring bases. Following minimization and equilibration, four 50 ns replica simulations were performed of the 1λ302+IHF construct starting from the same minimized structure.

For further comparison, a 61 bp section of the 1λ302+IHF construct, centered on the binding site, was extracted, explicitly solvated using a truncated octahedral TIP3P box and neutralized with a 0.2M-equivalent concentration of K and Cl ions (*49*). Simulations were performed for 150 ns at constant *T* and *P* (300 K and 1 atm) following standard protocols (*50*) and without the restraints on the intercalative prolines used for the simulations in implicit solvent.

Analysis of Simulations

Following the observation of three populations in the experimental results, the simulation results were clustered using hierarchical agglomerative clustering in cpptraj (average-linkage, using the RMSD between frames as the distance metric) until three clusters were formed. The DNA helix axis was calculated for each frame of these clusters using WrLINE (*51*). Then, SerraLINE was used to project WrLINE molecular contours onto the best-fit plane to approximate the experimental methodology and calculate bend angles (*41*). 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. SerraLINE and WrLINE are freely available at the repository github.com/agnesnoy.

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. The hydrogen bonds present in the crystal structure were determined in the same manner based on the minimized 1IHF structure (*5*), so can only take integer values.

Free energy Calculation

Starting from the explicitly solvated system, both arms of DNA were wrapped around the lateral sides of IHF over a series of umbrella sampling trajectories. The Grossfield implementation of the weighted histogram analysis method (*30*) was used to extract the potential of mean force (PMF). The reaction coordinates for the left- and right-hand sides were chosen as the distances between Cα atoms of representative aminoacids from the protein interacting far sites (Pro18β and Ser19α, respectively) and the phosphorus atoms from the closest DNA base in the crystal structure. The reaction coordinates were reduced over a series of 5ns simulations in 2Å increments from their positions in the minimized structure (Figure 1A) until the PMF was observed to increase, resulting in a total simulation length of 150 ns for the left arm and 195 ns for the right. The final frame of each umbrella sampling window was used as the starting structure for the next.

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. 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. To account for shifts in the free energy landscape between simulations due to the flexibility of the two-atom reaction coordinate, a linear fit was performed in gnuplot to translate the original distances to the distance between the centers of mass of the protein and the 10 bp region of DNA in closest contact from each far side (Figure S3F). The free-energy offsets were estimated by comparison of the means in the flat region far from the protein. The results from the restrained and unrestrained runs were considered to represent the two extremes of the other arm's position, and linear interpolation was performed between them.

A two-dimensional landscape was then constructed by simple addition of energies. This was translated into a probability landscape by mapping over it the function

$P(F) = \exp(-F/k_BT)$. The area representative of each cluster was mapped onto the free-energy landscape by locating each snapshot of the implicitly solvated simulations on the two-

dimensional surface using the value of the two reaction coordinates (see Figure S3C). The relative probabilities of the clusters were calculated by weighted integration over the probability landscape when computed with 1 Å resolution, and the probabilities were normalized to sum to 100%.

DNA bridging by IHF

To investigate bridge formation, another 61 bp 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 (an oxygen atom from a phosphate group and 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 increased 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. W. C. Hacker, S. Li, A. H. Elcock, Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome. *Nucleic Acids Res.* **45**, 7541–7554 (2017).
3. R. T. Dame, F. Z. M. Rashid, D. C. Grainger, Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.* **21** (2020), pp. 227–242.
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. J. Wei, L. Czapla, M. A. Grosner, D. Swigon, W. K. Olson, DNA topology confers sequence specificity to nonspecific architectural proteins. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16742–16747 (2014).
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. A. V. Wright, J. J. Liu, G. J. Knott, K. W. Doxzen, E. Nogales, J. A. Doudna, Structures of the CRISPR genome integration complex. *Science (80-.).* **357**, 1113–1118 (2017).
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. 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).
21. S. Le, H. Chen, P. Cong, J. Lin, P. Dröge, J. Yan, Mechanosensing of DNA bending in a single specific protein-DNA complex. *Sci. Rep.* **3**, 1–6 (2013).
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. 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).
24. 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).
25. N. L. Craig, H. A. Nash, E. coli integration host factor binds to specific sites in DNA. *Cell*. **39**, 707–716 (1984).
26. 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* (1995), doi:10.1021/bi00040a020.
27. 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).
28. A. Noy, A. Pé Rez, C. A. Laughton, M. Orozco, Theoretical study of large conformational transitions in DNA: the B$A conformational change in water and ethanol/water. *Nucleic Acids Res.* **35**, 3330–3338 (2007).
29. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).

30. A. Grossfield, WHAM: the weighted histogram analysis method, (available at http://membrane.urmc.rochester.edu/wordpress/?page_id=126).

31. M. C. Leake, The physics of life: One molecule at a time. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120248 (2013).

32. 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).

33. 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).

34. R. T. Dame, The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56**, 858–870 (2005).

35. 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).

36. A. Noy, A. Maxwell, S. A. Harris, Interference between Triplex and Protein Binding to Distal Sites on Supercoiled DNA. *Biophys. J.* **112**, 523–531 (2017).

37. A. Noy, T. Sutthibutpong, S. A. Harris, Protein/DNA interactions in complex DNA topologies: expect the unexpected. *Biophys. Rev.* **8** (2016), pp. 233–243.

38. I. D'annessa, A. Coletta, T. Sutthibutpong, J. Mitchell, G. Chillemi, S. Harris, A. Desideri, Simulations of DNA topoisomerase 1B bound to supercoiled DNA reveal changes in the flexibility pattern of the enzyme and a secondary protein-DNA binding site. *Nucleic Acids Res.* **42**, 9304–9312 (2014).

39. 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).

40. O. Scholder, pySPM v0.2.20 (2019), , doi:10.5281/ZENODO.2650457.

41. A. L. B. Pyne, A. Noy, K. Main, V. Velasco-Berrelleza, M. M. Piperakis, L. A. Mitchenhall, F. M. Cugliandolo, J. G. Beton, C. E. M. Stevenson, B. W. Hoogenboom, A. D. Bates, A. Maxwell, S. A. Harris, Combining high-resolution AFM with MD simulations shows that DNA supercoiling induces kinks and defects that enhance flexibility and recognition. *bioRxiv*, 863423 (2019).

42. 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.

43. D. A. Case, R. C. Walker, T. E. C. III, C. Simmerling, A. Roitberg, K. M. Merz, R. Luo, T. Darden, J. Wang, R. E. Duke, D. R. Roe, S. LeGrand, J. Swails, A. W. Götz, D. Cerutti, T. Luchko, G. Monard, C. Sagui, F. Pan, C. Lin, D. Mermelstein, P. Li, A. Onufriev, S. Izadi, R. M. Wolf, X. Wu, H. Gohlke, N. Homeyer, R. Qi, W. M. Botello-Smith, L. Xiao, D. Greene, T. Giese, T. Lee, D. York, J. Liu, H. Nguyen, I. Omelyan, A. Kovalenko, "Amber 2017 Reference Manual" (2017), (available at http://ambermd.org/contributors.html).
44. J. Srinivasan, M. W. Trevathan, P. Beroza, D. A. Case, Application of a pairwise generalized Born model to proteins and nucleic acids: Inclusion of salt effects. *Theor. Chem. Acc.* **101**, 426–434 (1999).
45. H. Nguyen, A. Pérez, S. Bermeo, C. Simmerling, Refinement of Generalized Born Implicit Solvation Parameters for Nucleic Acids and Their Complexes with Proteins. *J. Chem. Theory Comput.* **11**, 3714–3728 (2015).
46. A. Perez, J. L. MacCallum, E. Brini, C. Simmerling, K. A. Dill, Grid-Based Backbone Correction to the ff12SB Protein Force Field for Implicit-Solvent Simulations. *J. Chem. Theory Comput.* **11**, 4770–4779 (2015).
47. 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).
48. 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).
49. L. X. Dang, Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study. *J. Am. Chem. Soc.* **117**, 6954–6960 (1995).
50. A. Noy, R. Golestanian, The Chirality of DNA: Elasticity Cross-Terms at Base-Pair Level Including A-Tracts and the Influence of Ionic Strength. *J. Phys. Chem. B*. **114**, 8022–8031 (2010).
51. 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).

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). We would also like to thank Alice Pyne for her assistance with the AFM experimental techniques and analysis.

Figures and Tables

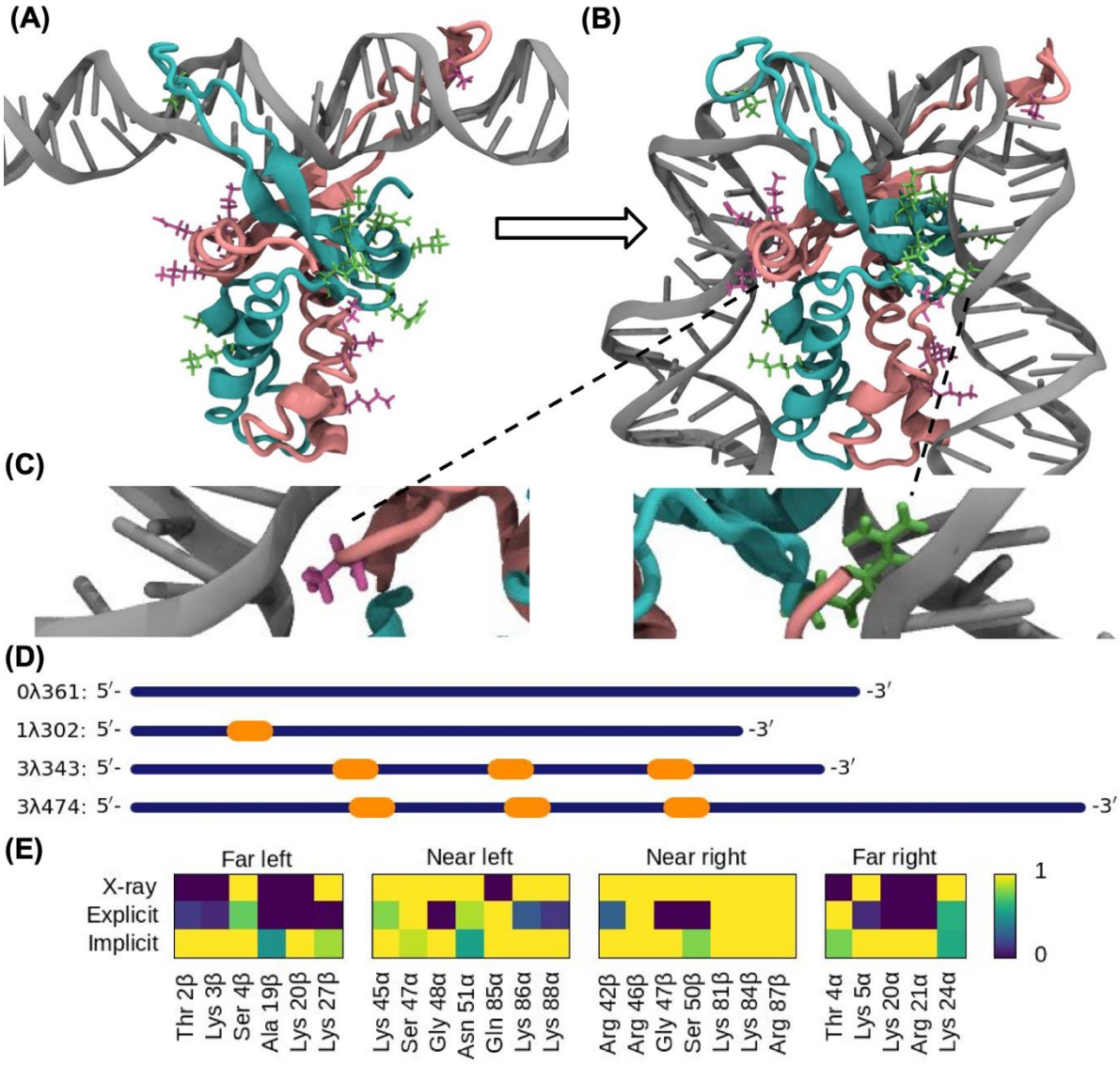

Fig. 1. Molecular structure of IHF and DNA when canonically bound as determined by molecular dynamics (MD) simulations. (A) Initial conformation for MD simulations, with the α subunit shown in pink, the β subunit in blue, and the DNA in gray. The intercalating prolines and amino acids that interact with the lateral DNA arms in the crystal structure are highlighted with atomic representation. **(B)** Final state of a representative canonical simulation in which the nucleoprotein complex ends up in the 'fully wrapped' state close to the crystallographic structure (5). **(C)** Zooms show Ser 47α (pink, left) and Arg 46β (green, right) are both inserted into the minor groove during implicit-solvent simulations as reported on 1HF crystal structure. **(D)** DNA constructs used in experiments with binding sites highlighted, shown to relative scale. **(E)** The main DNA–protein hydrogen bonds present in 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.

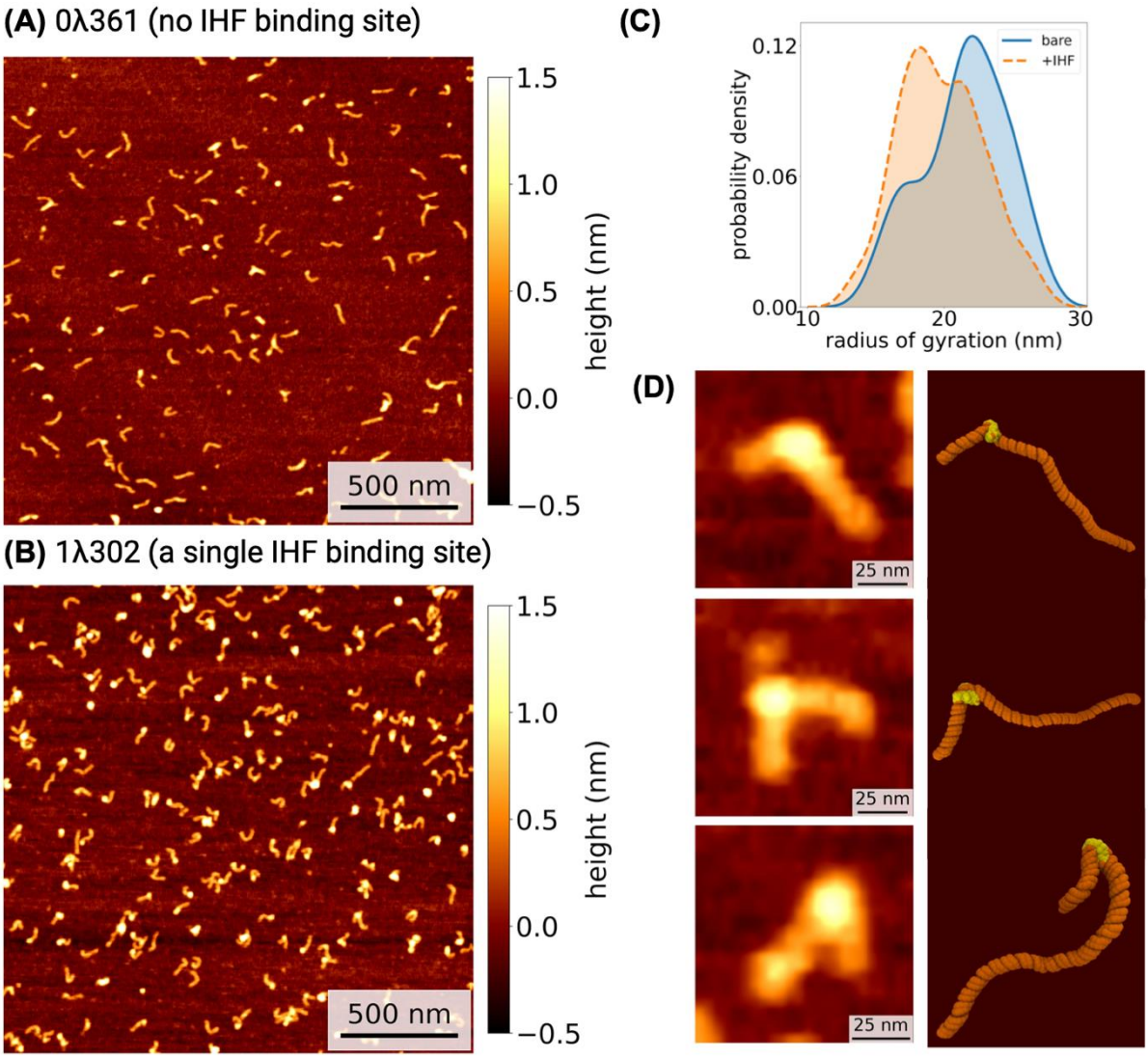

Fig. 2. Atomic force microscopy (AFM) images of IHF+DNA, with and without IHF binding sites. (A) AFM image of DNA without IHF specific binding site (0λ361), showing minimal binding when compared to **(B)** AFM image of DNA with an IHF binding site (1λ302). **(C)** Kernel density estimates of the contour 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 (left) and molecular dynamics simulations (right).

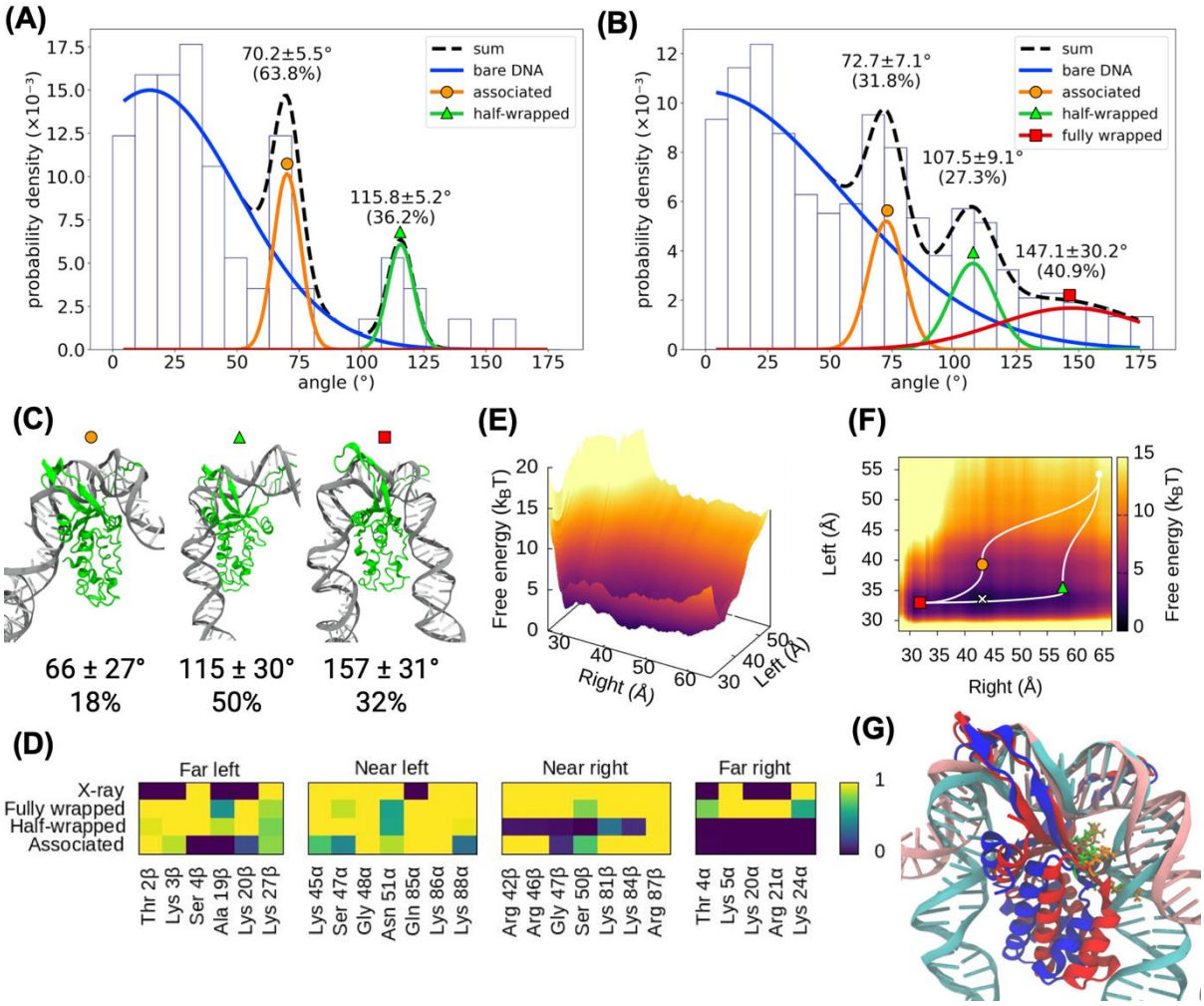

Fig. 3. Analysis of different bending angles of IHF-bound DNA. Bending angle distribution of 0λ361 **(A)** and 1λ302 **(B)** with IHF. There is a considerable amount of unbound DNA, which is disregarded using the distribution observed without IHF (blue line, Figures S2A-B). The remaining peaks are well fitted by Gaussians (as shown by the reduced χ^2 in Figures S2C-D), representing two common states with bending angles around ~70˚ and ~110˚ and an additional state presented only by 1λ302 with an angle around ~147˚, closer to the X-ray structure. **(C)** Simulations of the same construct can be classified into three clusters, the mean bending angles of which correspond well with the experimental data. **(D)** The clusters can be characterized by the DNA–protein hydrogen bonds present in each state; while the fully wrapped DNA interacts with the protein on both sides, the half-wrapped DNA interacts only on the left, and the associated state interacts primarily with the "near" sites. Contact maps show the average number of hydrogen bonds formed by each residue (capped at 1) using all frames belonging to each cluster or from the 1IHF crystal structure. **(E)** The relative probabilities of the clusters are determined from the free-energy landscape. The axes here represent the distance between the centers of mass of the protein and 10 bp interacting with each far side. **(F)** The clusters correspond to minima and plateaus in this landscape, and the observed cluster transitions (Figure S3D) provide a guide to the paths by which the system traverses it. **(G)** When the left arm is unbound (protein red & orange, DNA pink), the upper

subunit of the protein protrudes on the right-hand side, preventing the right arm from fully binding without bending significantly, resulting in a large free energy barrier. When the left arm is fully bound (protein blue & green, DNA cyan), the protein body is pulled to the left, removing this barrier and allowing the DNA to bind fully.

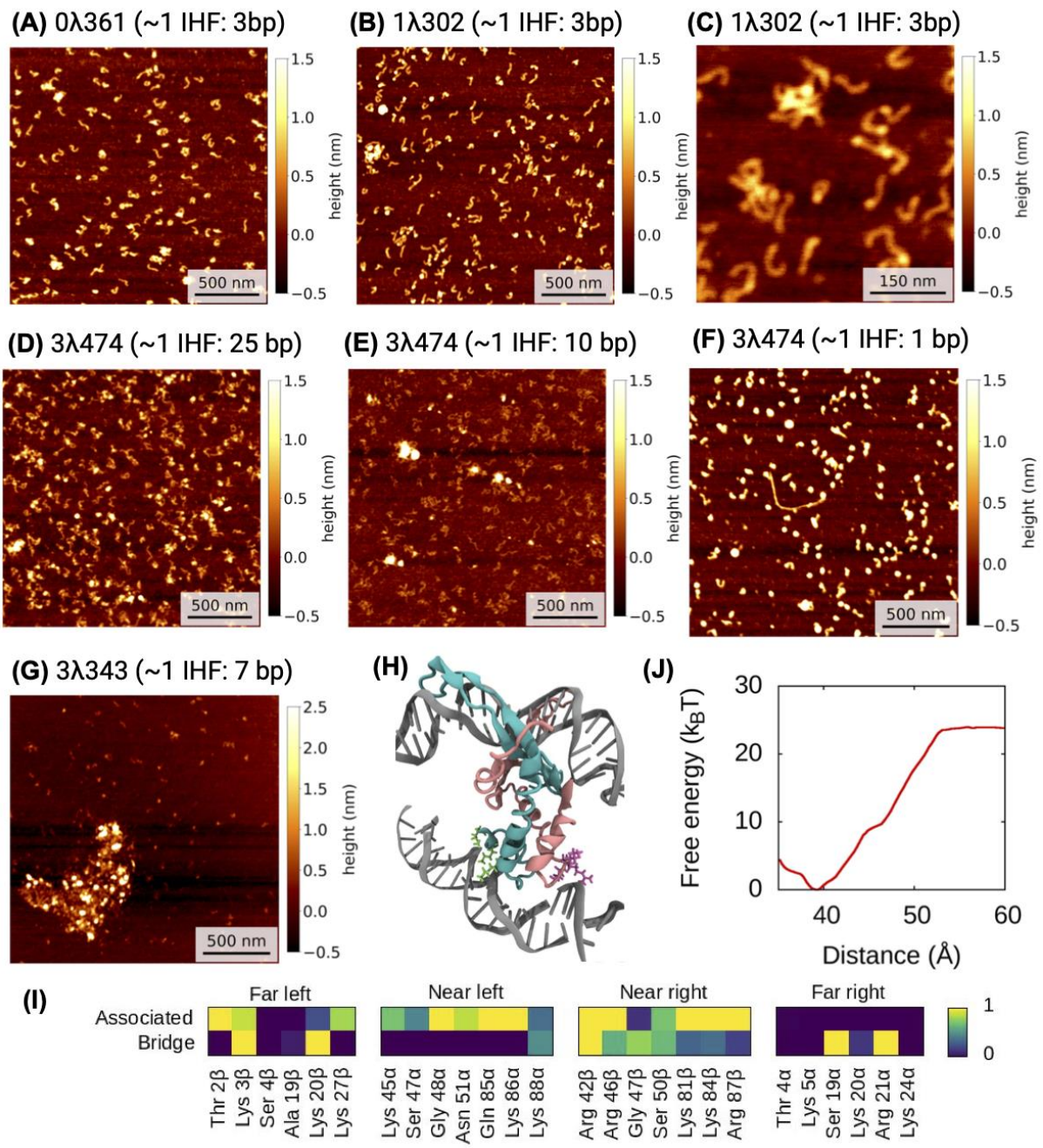

Fig. 4. IHF-bound DNA forms aggregates depending on protein concentration and on the type of DNA constructs. A large amount of IHF leads to some aggregation on 0λ361 **(A)** and 1λ302 **(B-C)**, with no or one binding site respectively. DNA constructs with 3 IHF binding sites (3λ343 and 3λ474) start to form clusters at lower concentrations. **(D-E)** At lower concentration 3λ474 forms more numerous and/or denser clusters, but at saturating concentration IHF coats the DNA **(F).** Sometimes large aggregates could be seen, even at concentrations where IHF does not significantly bind to 1λ302 **(G)**. **(H)** Bridge structure from MD simulations with the interacting amino acids highlighted in atomic representation. **(I)** Contact maps are calculated as in Figures 1 and 3 revealing that the residues from the "far" interacting regions are those which form the bridge. The DNA strand into which IHF intercalates remains mostly unbent, interacting with the protein even less than in the

associated state. **(J)** Free-energy landscape of bridging formation relative to an unwrapped state with a close additional DNA strand

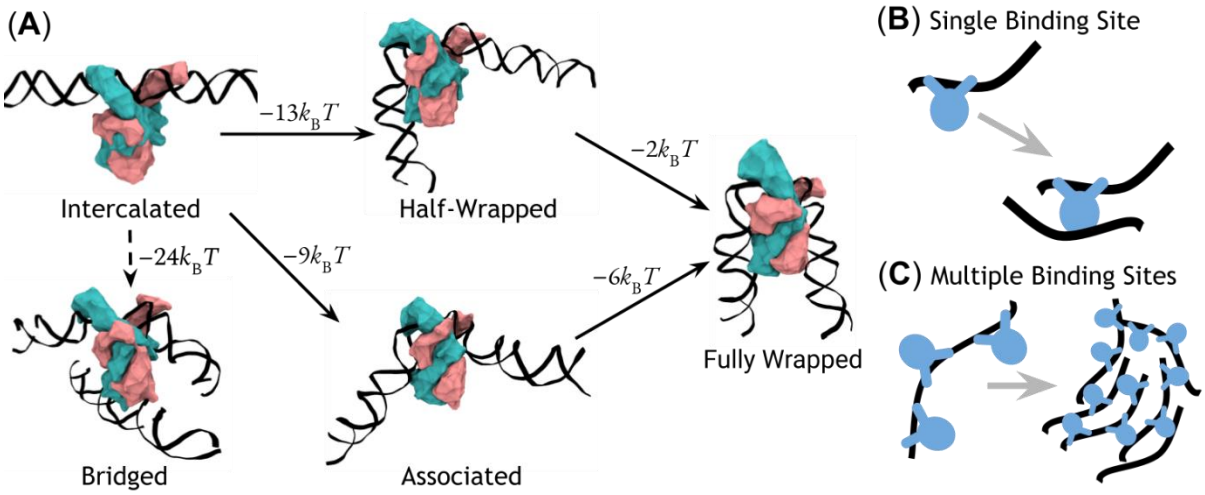

Fig. 5. Model of IHF binding – bending and bridging. (A) IHF first recognizes DNA through its β-arms and the prolines intercalate into a double-helix structure that its straight in our model although it might already contain a certain bent. The first binding step appears to be either a loose association to the DNA (73˚ binding) or a half-wrapped state (120˚). Both can then progress to the canonical fully wrapped state. As the initial state leaves the bottom half of IHF unbound it can also bind to another strand of DNA non-specifically resulting in a bridge (transition conditioned to the second DNA strand being nearby, dashed arrow). Free energy differences between states (ΔF) were estimated via umbrella sampling simulations. **(B)** A construct with a single binding site can only form a single bridge whereas **(C)** multiple binding sites can form multiple bridges, leading to 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`

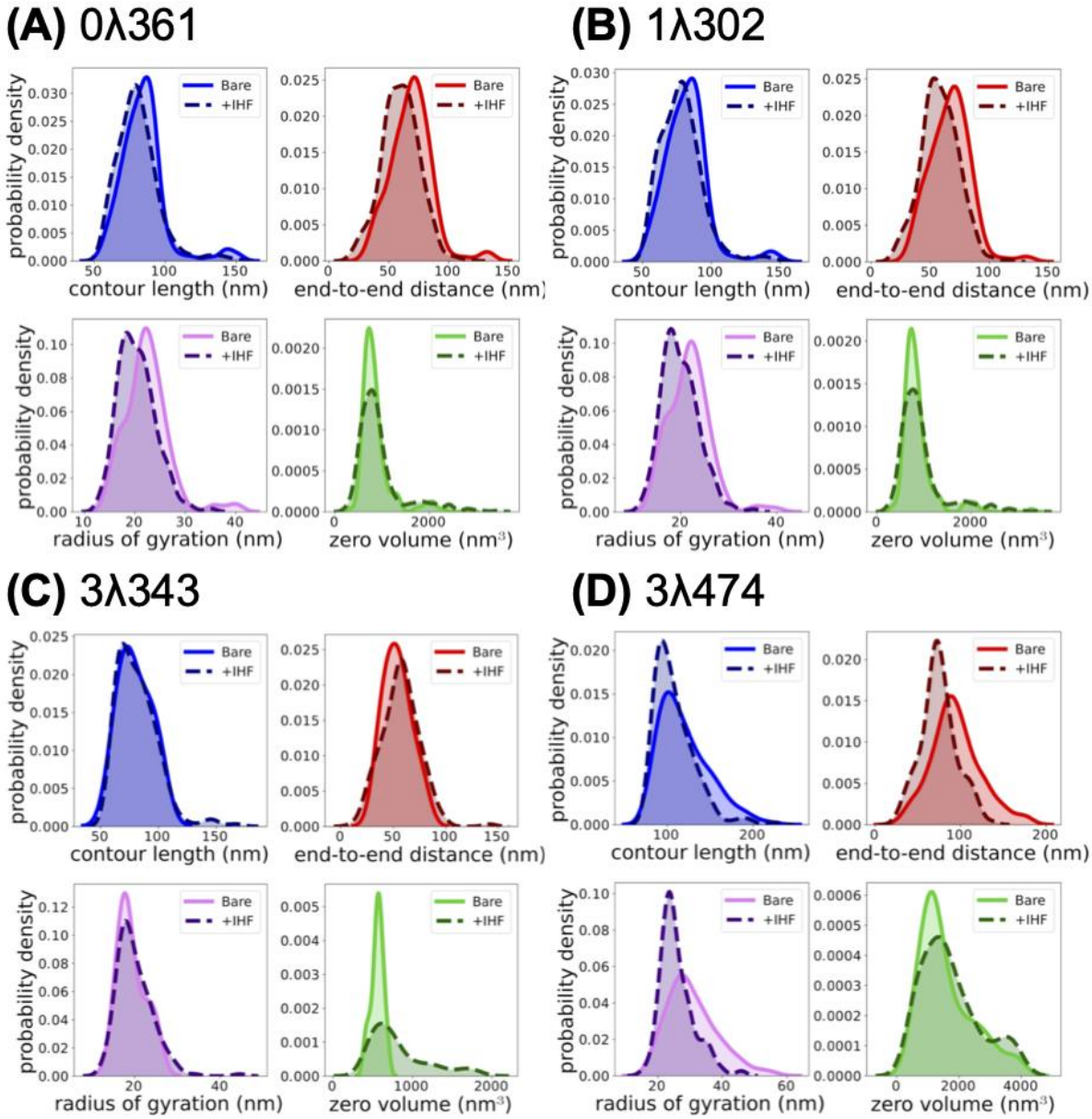

Fig. S1. Additional experimental AFM data.

Kernel density estimates for properties of all DNA constructs used +/- IHF. **(A)** 0λ361, **(B)** 1λ302, **(C)** 3λ343 and **(D)** 3λ474. For all 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).

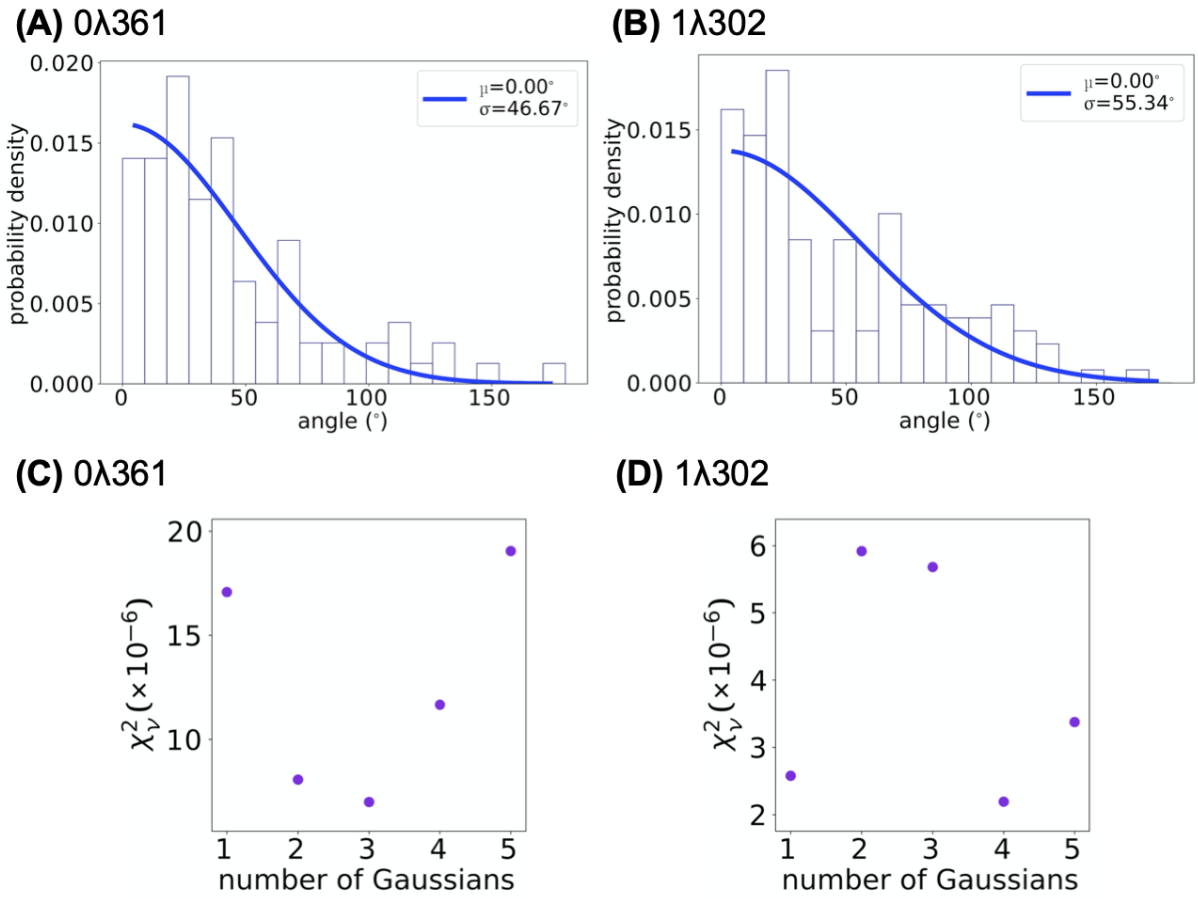

Fig. S2. Additional bending angle analysis

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. Reduced χ^2 fits for 0λ361 and 1λ302 with IHF. **(C)** For 0λ361 (no binding site) the data is fit well by either 2 or 3 Gaussians, suggesting an unbound and 1-2 bound states. **(D)** 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.

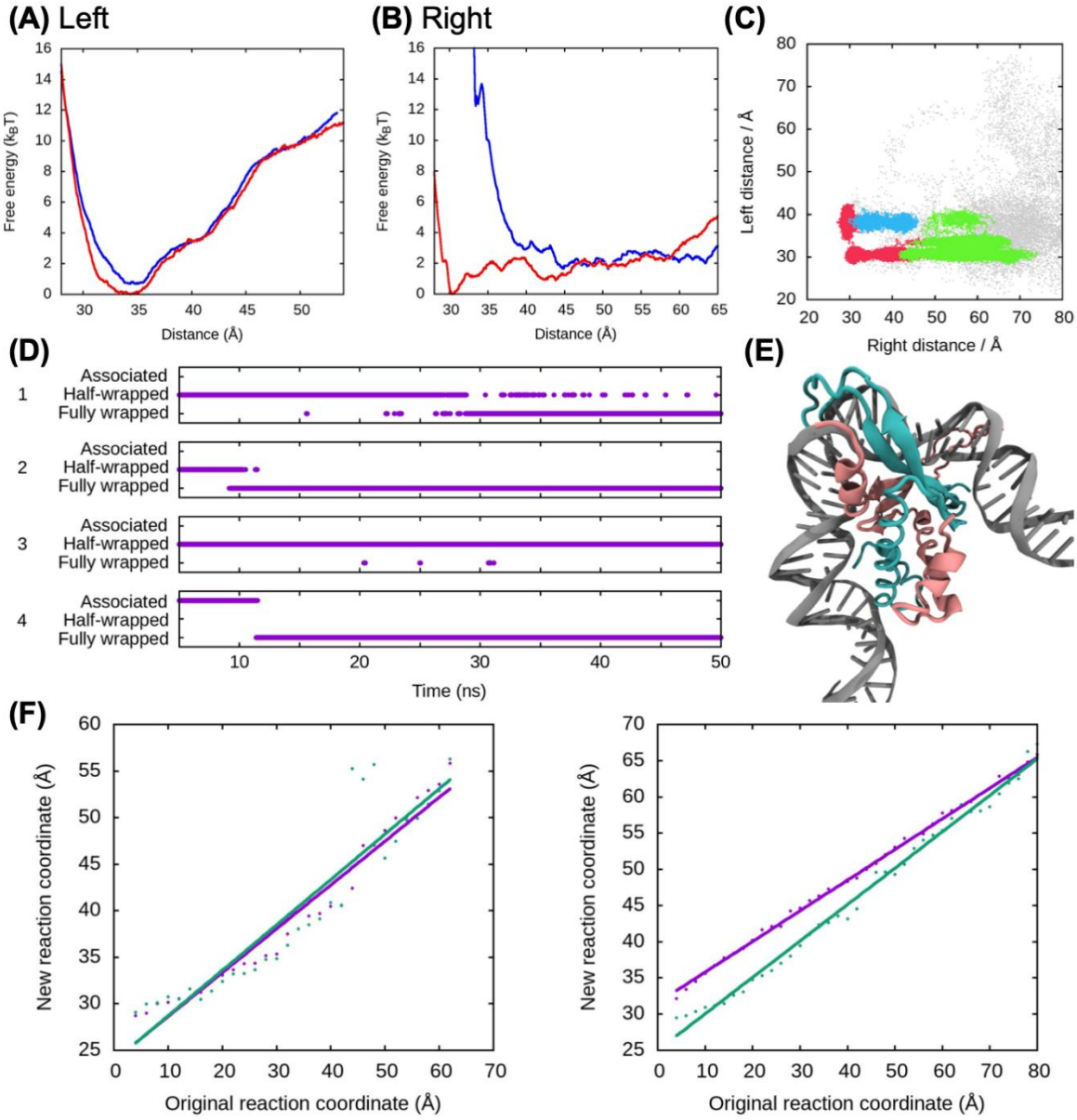

Fig. S3. Additional simulation data.

(A-B) The form of the PMF for the left arm **(A)** is similar whether the other arm is free (red) or restrained (blue). On the right-hand side **(B)**, the PMF is dominated by noise in both cases, but exhibits some significant differences, including a large free-energy barrier preventing binding of the right arm before the left. Standard errors calculated as in (*30*) are already displayed although they are too small to be visible. By considering each of these PMF to be correct at one of the two extremes of the other arm's position and interpolating linearly between them, an approximate 3D landscape can be constructed (Figure 3E) without sampling the entire conformation space. **(C)** By identifying the cluster to which each frame of the simulation belongs to**,** the areas representative of each state are delineated onto the free-energy landscape (fully-wrapped in red, half-wrapped in green, associated in blue, snapshots from the first 5 ns of each replica in grey)**. (D)** The possible transitions can also be observed, with every replica spending some time in the associated or half-wrapped state before transitioning to the canonical fully wrapped state. **(E)** An additional state, in which the left-

hand side is fully bound and the right-hand side is partially bound, is the obvious transition between the half-wrapped/associated states and the fully wrapped state. However, the fully-wrapped cluster encompasses this sub-state as they present similar DNA bending angles and, according to the free energy landscape, the system can move freely between them under the influence of thermal noise. **(F)** The original reaction distance chosen for the calculation of PMFs (as described in Materials and Methods) was converted into to a new reaction distance between the centers-of-mass of the protein and a section of DNA to eliminate distortions caused by local structural fluctuations. The scale factor between the two reaction coordinates is approximately linear (R^2 = 0.84–0.99) on both the left (left) and right (right) sides, whether the other arm is restrained away from the protein (purple) or allowed to move freely (green). The new reaction coordinate is the one used on all free energy plots.

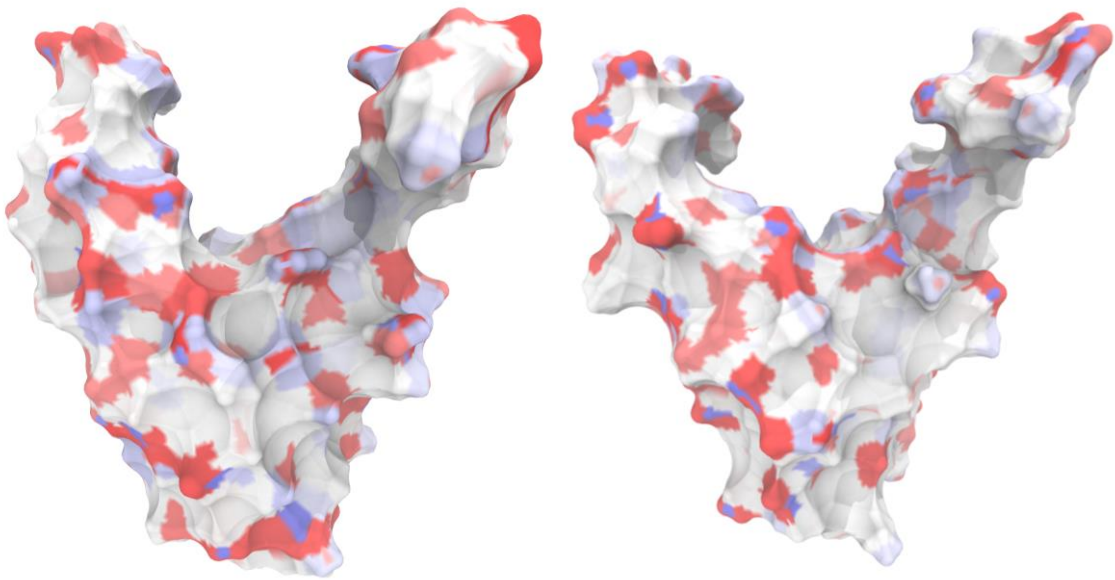

Fig. S4. Surface electrostatic profiles of IHF and HU. The surfaces of IHF (left) and HU (right) are similar from an electrostatic perspective, both consisting of isolated spots of positive charge (blue) on an otherwise neutral (white) or negatively charged (red) surface.

Supplementary Movie 1. Movie showing the first replica of implicitly solvated simulations of the 1λ302+IHF experimental construct where the complex oscillates between the half- and fully-wrapped states (Figure S3D).

Supplementary Movie 2. Movie of the second replica where the complex rapidly reaches the fully-wrapped state (Figure S3D). Because this is the replica with a most canonical behavior, it is selected in Figure 1E for assessing the recovery of crystallographic interactions by means of implicitly solvated simulations.

Supplementary Movie 3. Movie of the third replica where the complex remains in the half-wrapped conformation showing the metastable nature of this state (Figure S3D).

Supplementary Movie 4. Movie of the fourth replica where the complex traverses through the associated state before transitioning to the fully-wrapped state (Figure S3D).

Supplementary Movie 5. Representative movie based on umbrella sampling trajectories of the asymmetric allosteric switch between arms, where the left arm (with the A-tract, in red) binds before the right arm containing the specific recognition sequence (in blue)