Understanding DNA Deformability by DNA Bending Studies with the IHF/HU Family of Architectural Proteins

By
MITCHELL JEFFREY CONNOLLY
B.S., Southern Illinois University – Carbondale, 2014
M.S., University of Illinois at Chicago, 2016

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Defense Committee:
Anjum Ansari, Chair and Advisor
David Hofman
Phoebe Rice, Biochemistry and Molecular Biology, University of Chicago
Mark Schlossman
Huan-Xiang Zhou

Dedicated to my first teachers, Mom and Dad.

ACKNOWLEDGEMENTS

First, I would first like to thank my advisor, Prof. Anjum Ansari. Her vision and guidance were instrumental to my success. She showed me what it means to be a good scholar, teacher, and leader.

I would also like to thank my thesis committee - Prof. David Hofman, Prof. Phoebe Rice, Prof. Mark Scholssman and Prof. Huan-Xiang Zhou - whose thoughtful critiques and suggestions enriched the quality of this thesis.

I am grateful to have had such wonderful peers and coworkers during my tenure at UIC. Viktoria Zvoda, Sagnik Chakraborty, Saroj Baral, Jorjethe Roca, Timur Ten, Manas Sarangi, and Yogambigai Velmurugu – you made the journey worthwhile. Thank you for all the support over the years. Also, thank you to Aline Arra, Rafid Rahman, and Jacob Bathery who trusted me to guide their research projects.

Thank you to the many exceptional collaborators who contributed to this work - especially, Prof. Phoebe Rice, Dr. Ying Pigli (U. Chicago), Prof. Aseem Ansari (St. Jude Children's Hospital), Prof. Jose Rodruigez-Martinez (U. Puerto Rico - Rio Piedras), Dr. Devesh Bhimsaria (Bioinformaticals, Inc.), Dr. Peter J. Steinbach (National Institute of Health), Prof. Hasheem Al-Hashimi (Duke U.), Prof. Huiqing "Jane" Zhou (Boston College), Prof. Wilma Olson (Rutgers U.), and Robert Young (Rutgers U.).

To those friends who kept me sane during this period of my life – especially, Dan and Anna Ziemniak, Mike Avendano, Alec Mark, Mitch and Katie Catalano, Ariel Davis, Aviva Schwartz, Dhanush Hangal, Brian Good and Rachel Kohl, Eric Mascot, Bilash KC, the Physics United soccer team, the band, and the trivia crew - thank you.

Thanks to my family for their support. To my brother Neil, thanks for fueling my competitive spirit. To my sister Kera, thanks for showing me how to live as my best self. Finally, thank you to my parents, Linda and Jim, for always being there.

Lastly, thank you Rebecca Rich. Your love and endless support carried me through.

CONTRIBUTIONS

Chapter 3. Aline Arra performed IHF-DNA titrations
Viktoriya Zvoda assisted with experiment design
Peter J. Steinbach developed the maximum entropy method program used for data analysis
Phoebe Rice purified the IHF protein

Chapter 4. Jacob Bathery assisted in model development and performed titrations
Phoebe Rice purified the IHF protein

Chapter 5. Rafid Rahman performed IHF-DNA titrations
Huiqing Zhou and Hashim Al-Hashimi co-developed experiment design
Peter J. Steinbach developed maximum entropy method program used for data analysis
Phoebe Rice purified the IHF protein

Chapter 6. Aseem Ansari and Jose Rodriguez-Martinez developed SELEX protocol
Dhevesh Bhimsaria performed DNAshape analysis
Robert T. Young performed elasticity calculations and energy minimization
Wilma Olson oversaw elasticity calculations and energy minimization
Phoebe Rice purified IHF and Hbb proteins

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribunucleic acids
IHF	Integration Host Factor
Hbb	Histone-like protein *Borrelia burgdorferi*
HU	Histone-like protein
dsDNA	Double stranded DNA
ssDNA	Single stranded DNA
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
N	Any DNA base
7dA	7-deaza Adenine
WC	Watson-Crick basepair
HG	Hoogsteen basepair
bp	Basepair
F	Fluorescien-dT
R	TAMRA-dT
Cy3	Cyanine 3
FLT	Fluorescence lifetime
FRET	Forster resonance energy transfer
CD	Circular dichroism
NMR	Nuclear magnetic resonance
MEM	Maximum entropy method
T-jump	Temperature jump
EMSA	Electrophoretic mobility shift assay
SELEX	Systemic enrichment of ligand by exponential enrichment

SUMMARY

DNA is often bent, twisted, or otherwise deformed inside the cell because of interactions with DNA-binding proteins that perform a variety of essential cellular functions such as transcriptional regulation, replication, or repair. The extent of these deformations ranges from localized distortions over a few base pairs to looping or bending over hundreds of base pairs. Structures of several protein-DNA complexes show the DNA sharply kinked at specific target sites, indicating that "site-specific" recognition must rely, at least in part, on the ease with which DNA can be deformed. This is especially the case for those proteins that are unable to directly "read" the sequence at that site and thus rely almost entirely on sensing sequence-dependent DNA shape and deformability for target recognition, also referred to as "indirect readout". However, many of the rules that govern sequence-dependent DNA shape/deformability remain largely untested, especially for DNA deformations of high severity.

In this thesis, I present multiple studies aimed at characterizing severe DNA deformations. These studies were performed with a prokaryotic, nucleoid-associated, architectural protein Integration Host Factor (IHF) that belongs to a family of DNA-bending proteins. IHF has been called a "master-bender" for its remarkable ability to bend specific sites on DNA by nearly 180^{o} degrees over 35 base pairs (bp), which it accomplishes by inducing two sharp kinks into the DNA about 9 bp apart. All contacts that IHF makes with the DNA in the specific complex are either to the minor groove – where the bases are hard to differentiate – or to the sugar phosphate backbone. Thus, IHF serves as an excellent model system for biophysical studies of a DNA-bending protein that exhibits high specificity to certain sequences but relies almost exclusively on indirect readout to recognize these sites.

While crystal structures of IHF bound to one of its target sites, the ~35-bp H' site on λ-phage DNA, showed the DNA bent into a "U-turn", the structure provided only a static picture. Whether the two DNA kinks induced by IHF were rigid or flexible was not known. In Chapter 3, I report our study that addressed this question by using fluorescence lifetime (FLT) studies to measure Forster resonance energy transfer (FRET) efficiency between fluorescent probes attached to the ends of 35-mer DNA substrates of IHF. The FLT-FRET approach enabled, for the first time, observation of additional, partially-bent conformations in various IHF-DNA complexes, with either one side unkinked or the other, thus revealing a surprisingly dynamic specific complex. This conformational flexibility was further modulated by sequence variations in the cognate site. In particular, our study demonstrated how an A-tract on one of the flanking arms of H' on one side of the U-bent binding site is critical in keeping that arm of the DNA held against the protein by forming favorable interactions with that side of the IHF molecule; in the absence of the A-tract, as in another binding site, the H1 site, that side becomes more "flappy" in the complex. Notably, the difference in dynamics between IHF-H' and IHF-H1 reflects the different roles of these complexes in their natural context, in the λ-phage "intasome" – the complex that integrates λ-phage virus into the chromosome of the bacterial host during viral infection. Taken together, these results provide important insights into the finely tuned interactions between IHF and its many biological targets that keep the DNA bent (or not); yield quantitative data on the dynamic equilibrium between different DNA conformations (kinked or not kinked) that depend sensitively on DNA sequence and deformability; and highlight the important implications of different levels of flexibility of the IHF-DNA complexes in their biological role of facilitating the formation of higher-order DNA-protein complexes.

While the specific binding of IHF to DNA has been extensively studied, it also binds nonspecifically to facilitate DNA packaging, and possibly for other cellular functions. The ability of IHF to bind DNA in multiple modes to a single DNA site poses a simple question: what are the relative binding affinities in the two binding modes? Indeed, there is wide disagreement in the literature as to what the binding affinities are even for specific binding, with previously reported affinities for the IHF-H' complex varying by nearly three orders of magnitude. Much less is known about the nonspecific binding affinities. In Chapter 4, I address these issues by measurements of isotherms for binding of IHF to the specific H' substrate as well as to a nonspecific DNA construct of the same length, using (1) fluorescence anisotropy to monitor protein binding stoichiometry and (2) FRET measurements on end-labeled DNA constructs to monitor the extent of DNA bending. I applied a lattice model – where the DNA is treated as a finite lattice to which one IHF protein can bind specifically or multiple proteins can bind nonspecifically on non-overlapping sites – and used global modeling to simultaneously describe the four binding isotherms (from anisotropy and FRET measurements on specific and nonspecific DNA). This "three-state" model provided an excellent description of the four binding isotherms up to [IHF]/[DNA] $\leq$ 200. However, for even higher excess protein conditions, discrepancies between the data and the fits became evident, suggesting that this minimal description, in which nonspecific protein binding displaces a specifically bound protein, may be incomplete. Our data suggest that specific and nonspecific binding need not be mutually exclusive and that a binding site with a specifically bound protein could also accommodate a nonspecifically bound protein. Our study thus elucidates the added complexities of IHF-DNA interactions that were not considered in previous descriptions and lays

the foundation for further examining higher order complexes at high protein-to-DNA ratios that may be present *in vivo*.

In Chapter 5, I investigated if transient Hoogsteen base pairing played a role in the mechanism of site-specific recognition by IHF. Hoogsteen pairing is a thermodynamically less favored yet thermally accessible alternative to Watson-Crick base pairing in duplex DNA; is more prevalent at sites where DNA is more flexible; and involves structural changes to the sugar-phosphate backbone of DNA. These transient changes in shape brought about by Hoogsteen formation are likely sensed by indirect readout proteins as they scan the DNA. Here, we investigated the role transient Hoogsteen base pairing may have in the ease with which IHF recognizes its DNA binding site. We utilized laser temperature-jump (T-jump) spectroscopy to measure changes in the DNA bending/unbending kinetics in the IHF-H' complex when a Hoogsteen inhibiting nucleotide analog, 7-deaza adenine (7dA), was substituted instead of the normal adenine at different positions along the H' substrate. We hypothesized that if transient Hoogsteen formation facilitates DNA kinking then 7dA substitution made at or near a kink site should destabilize the modified IHF-H' complex and slow the rates of DNA bending. We anticipated no change in complex stability and DNA bending rates with 7dA substitutions made at sites that remain B-DNA like in the complex and are at some distance from the kinks.

Our study revealed some unexpected results. First, 7dA substitutions near a kink site had little or no effect on IHF-H' binding affinity while the T-jump measured rates for DNA bending/unbending within the specific complex *increased* by ~2-fold. Second, 7dA substitutions made away from the kink site, and initially expected to have no effect on affinity nor on kinetics, in fact stabilized the complex by ~40-fold and again slightly increased the T-jump measured

rates. We therefore conclude that transient Hoogsteen pairing at potentially weak sites in the DNA does not appear to be an important factor in facilitating DNA kinking or target recognition by IHF. Instead, we attribute the ~40-fold increase in binding affinity for our "control" experiment by recognizing that the 7dA substitution in this case was made in a consensus region of the H' sequence, previously shown to be more readily twisted to facilitate interactions with IHF and to stabilize the bent DNA arm on that side of the complex. Thus, increased stability of the complex with the "control" is attributed to weakened stacking interactions from the 7dA substitution and hence increased DNA "twistability" at that site. While our findings show that Hoogsteen pairing has little importance within the IHF-DNA complex, these studies nonetheless highlight the profound impact that apparently minor modifications in a nucleotide moiety – that still enable Watson-Crick pairing – can have on DNA flexibility/deformability and thereby influence target-site recognition.

In Chapter 6, I present high-throughput "SELEX" studies aimed at identifying DNA sequences of high deformability. SELEX – systemic evolution of ligands by exponential enrichment – is an *in vitro* protocol that separates a pool of random DNA sequences by their affinities to DNA-binding proteins. For these studies we used IHF and a closely related, structurally homologous, nucleoid-associated protein, Hbb, found in *Borrelia burgdorferi*, a spirochete responsible for causing Lyme disease in humans. Both proteins bend the DNA into a U-turn to form very similar protein-DNA complex structures; thus high-affinity sequences are hypothesized to be of higher deformability. Our goal was to identify these high-affinity DNA sequences and to uncover sequence patterns at the sites known to be sharply kinked when in complex with this family of DNA-bending proteins; to examine the intrinsic "kinkability" of

these DNA sequences; and to provide a robust database for further development of base-step and longer range parameters for modeling DNA deformability.

The DNA library for the SELEX studies was based on the H' sequence but with the A-tract on the flanking arm replaced by the sequence from the H1 site that lacks an A-tract, as described above. A 6-bp region at one of the protein-induced kink sites that does not fall within the consensus region for either protein was chosen to be randomized, resulting in a DNA library with $4^6 = 4{,}096$ unique sequences. We reasoned that, in the absence of the stabilizing interactions of the A-tract, sequence selectivity at the randomized site would be determined primarily by the "kinkablity" of that site.

These SELEX studies enabled us to rank DNA sequences by their binding affinities and to separate high affinity "winners" from low affinity "losers" for both IHF and Hbb. The sequences were then analyzed in multiple different ways. First, we identified dinucleotide and tetranucleotide sequences in the randomized region that were more probable among winners or losers. Second, we parameterized the DNA shapes of all the sequences using the DNA-Shape program developed by Prof. Remo Rohs (University of Southern California). Third, we used nearest-neighbor elastic models to compute the energetic cost for distorting B-DNA into the U-turn shape of the protein-bound structures. A few key results stand out: (1) A-tracts of greater than 3 As in a row at or near the randomized kink sites were found to be disfavored by both proteins, indicating that such stretches inhibit the formation of the sharply kinked DNA; (2) the DNA-shape analysis revealed an asymmetry between the two kink sites in that pre-distorted DNA was revealed to be an important feature of the non-randomized (consensus region) kink site and less so for the randomized kink site that lay outside the consensus region; (3) a correlation,

albeit weak, was observed for IHF (though not for Hbb) between binding affinity and ease of deformability at the randomized kink site, as obtained from the elastic modeling. These studies further demonstrated the unique properties of A-tracts that make them highly important sequence elements for genetic regulation. They also underscored the limitations of elastic (harmonic approximation) models for describing sharply kinked DNA. Harmonic models impose a high energy penalty on the 90° bends observed at each kink site in the IHF-H' crystal structure, despite growing evidence that kinks are more probable and less energetically costly than these models predict. These SELEX studies lay a strong foundation for systematically and quantitatively approaching questions surrounding DNA of high deformability and for developing a database of highly deformable sequences that can be used to improve current models of sequence-dependent DNA mechanics and flexibility.

Chapter 1

Introduction

1.1. DNA

Deoxyribonucleic acids (DNA) carry the essential genetic information for all living organisms, information that must be properly stored, organized, and accessed by the cellular machinery with high fidelity. Cells accomplish these takes by employing proteins to properly package the DNA and to access specific sites along the DNA. Protein binding to specific sites is needed to initiate a variety of cellular functions such as gene regulation and DNA repair. Therefore, understanding how proteins bind to the DNA and why some DNA sequences are preferred is needed to understand how the cell operates. Crystal structures of several protein-DNA complexes show that DNA is often severely deformed when site-specifically bound to the protein (as illustrated in Figure 1.1), indicating that sequence-dependent DNA deformability is an important feature of DNA that helps proteins discriminate between sites as they search the genome. This thesis includes several studies that aim to characterize DNA deformability and its contribution to recognition and specificity by DNA-binding proteins.

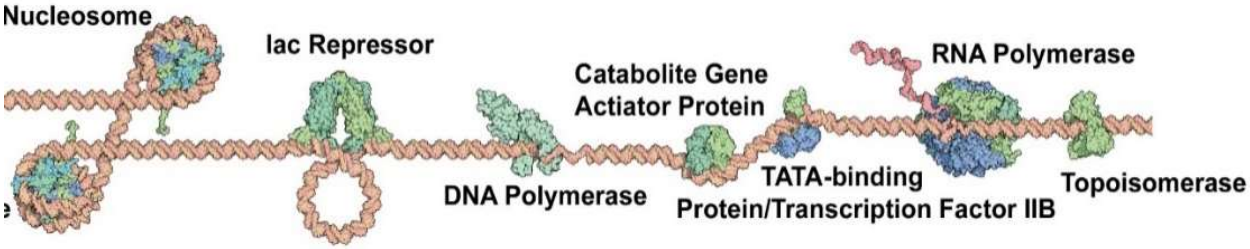


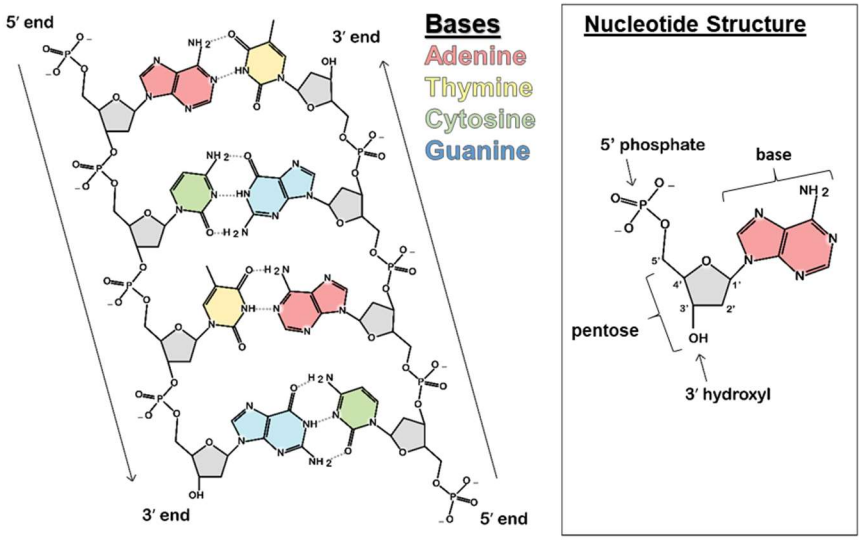
Figure 1.1. Schematic representation of different DNA-bending proteins. Open source images by Dr. David Goodsell (https://pdb101.rcsb.org/).

1.1.1. DNA structure

Duplex DNA is comprised of two polymer strands. Each strand consists of a string of nucleotides. Each nucleotide is made up of a phosphate group, a deoxyribose sugar, and one of four bases: adenine (A), thymine (T), cytosine (C) and guanine (G) (Figure 1.2.). Alternating phosphate groups and sugars, linked by phosphodiester bonds, create the backbone. The orientation of the pentose sugar group gives direction to the DNA strand (ordered 5'→ 3'). One base is attached to each sugar along the backbone. The order of the bases is the DNA sequence.

Two strands are joined to form the duplex DNA by the pairing of complementary nucleotides – As to Ts and Cs to Gs – with hydrogen bonds formed between the pairs; this pairing is referred to as the Watson-Crick pairing (Figure 1.2.). In the cell, duplex DNA is synthesized by proteins that use a single strand as a template to synthesize its complement. The backbone of the complementary strand is ordered the opposite direction, creating a so called anti-parallel structure.

Figure 1.2. Chemical structure of duplex DNA and subunit. Watson-Crick pairs are shown.



Adapted from the Khan Academy. Discovery of the structure of the double helix. Image modified from "DNA chemical structure," by Madeleine Price Ball (CC0/public domain).

1.1.2. B-DNA

The most common structure of duplex DNA is called B-form DNA (Figure 1.3.), normally found under neutral pH and physiological salt conditions. Other forms include A-form, a thicker form of nucleic acid common in RNA-RNA interactions and RNA-DNA interactions, and Z-form DNA, which forms usually under torsional stress, especially in stretches of alternating pyrimidine-purine sequences, and is thought to be important for cellular regulation (1).

Hydrophobic forces drive DNA's duplex structure. Non-polar faces of nucleotides, unfavorable within a polar water environment, are hidden while favorable negatively charged phosphates in the backbone are exposed. In B-form, a right-handed helical ladder twists approximately one full rotation per 10-10.5 base pairs (2). The shape of nucleotides gives asymmetry to the helix, which forms a major and a minor groove. Proteins may interrogate the DNA sequence differently when binding to the major groove or minor groove. While the major groove may offer unique chemical moieties for proteins to recognize, the minor groove presents an orientation of Watson-Crick base pairs in which the hydrogen bonding patterns between A-T and C-G base pairs appear similar, preventing direct recognition of the nucleotides by site-specific proteins (3, 4). Proteins that bind to the minor groove often need to identify their target sites via sequence-dependent differences in shape or deformability of a local DNA sequence.

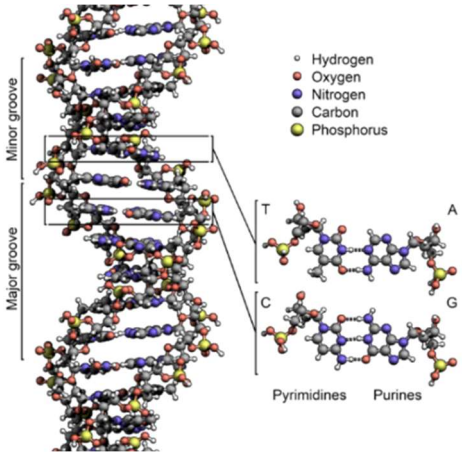


Figure 1.3. Schematic representation of B-DNA structure with major and minor grove and its canonical Watson-Crick base pairing scheme. Open source image. (https://commons.wikimedia.org/wiki/File:DNA_Structure%2BKey%2BLabelled.pn_NoBB.png ?uselang=en)

1.1.3. Semiflexible polymer description of DNA

Many properties of DNA flexibility are well described in terms of an elastic continuum (semiflexible) polymer, the so-called wormlike chain model, and characterized in terms of bending and twisting rigidity. The bending rigidity is quantified by a characteristic persistence length. On length scales much less than the persistence length, the polymer is thought to be rather stiff, while for length scales much longer than the persistence length, it is often treated like a random coil (5). For duplex DNA, this persistence length has been measured experimentally by rotational diffusion measurements on DNA oligomers that probe the shortening of the end-to-end distance in the elastic rod regime (6-8), ligase-catalyzed cyclization of DNA molecules with cohesive ends that probe ring closure probabilities (9-11), and more recently, micromanipulation experiments on long single molecules that probe the entropic spring constant of polymeric DNA under applied force (12, 13). These measurements yielded persistence length values of to be ~50-60 nm (~150-180 bp), with little ionic strength dependence in the range of 10 mM- 1 M Na+ ions (14) as shown in Figure 1.4.

It should be noted that physiological conditions also have some divalent ions that further alter the persistence length.

According to the wormlike chain description, characteristic length scales on which most proteins interact with DNA, usually much shorter than the persistence length of DNA, should be quite stiff. The stiffness of DNA derives mostly (~90%) from base-stacking interactions (a combination of hydrophobic effects to bury non-polar nucleotide components from the environment and π-π electrostatic interactions between the aromatic rings of adjacent nucleotides) (15). Backbone electrostatics also contribute to DNA stiffness, though to a lesser extent, and explains many of the changes to DNA behavior under different ionic conditions. For example, DNA persistence length drops monotonically as the salt concentration of the solution increases (Figure 1.4.). Positively charged "counter ions" in solution coat the negatively charged phosphate group in the DNA backbone, and increasing the ionic strength helps to minimize the repulsion between the phosphates (14).

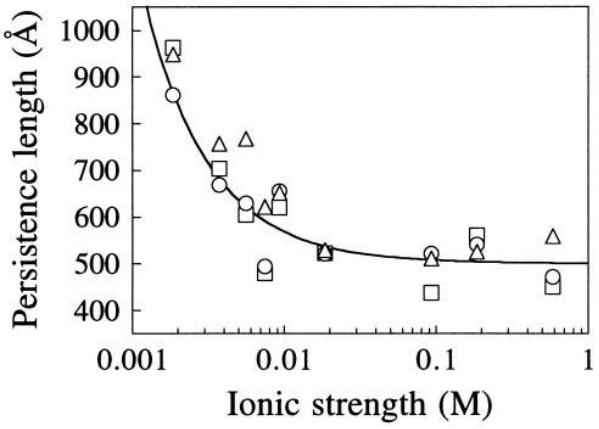


Figure 1.4. Persistence length of DNA in different ionic conditions. Reprinted with permission from Baumen et. al. (14)

1.1.4. Sequence-dependence of DNA flexibility

DNA stiffness relies on stacking interactions that have strong dependence on sequence dependence. Indeed, on the length scales on which proteins bind, which is less than few tens of base pairs, what is critical for site-specific recognition are the local deformations at the base-pair level that are not captured by the continuum description of the DNA polymer. Over the past few decades, there has been a tremendous effort by several groups to characterize the strength of stacking interactions between different groups of nearest neighbor base-pair "stacks" (15-21). Sequence-dependent elastic parameters that can describe the local shapes and flexibility of DNA are best encapsulated in terms of a set of parameters at the di-nucleotide level, as illustrated in Figure 1.5. For the last twenty years, plenty of effort has been directed toward refining elasticity parameters. Wilma Olson and co-workers as well as a consortium of molecular dynamics groups have worked to expand elasticity parameters from the di-nucleotide to the tetra-nucleotide level (22-24). Additionally, work has been done to refine the energy models used when calculating elastic properties of DNA. Current approaches use harmonic approximations to describe DNA deformations, but these estimates are meant to estimate small distortions in DNA, which does not fully capture the severe bending (>30^{o}) often observed in nature (23-27). While small amplitude bending and twisting deformations in DNA are captured well by elastic models, there is a large body of evidence indicating that severe DNA distortions, such as those seen upon protein-binding, need models that go beyond the harmonic approximation.

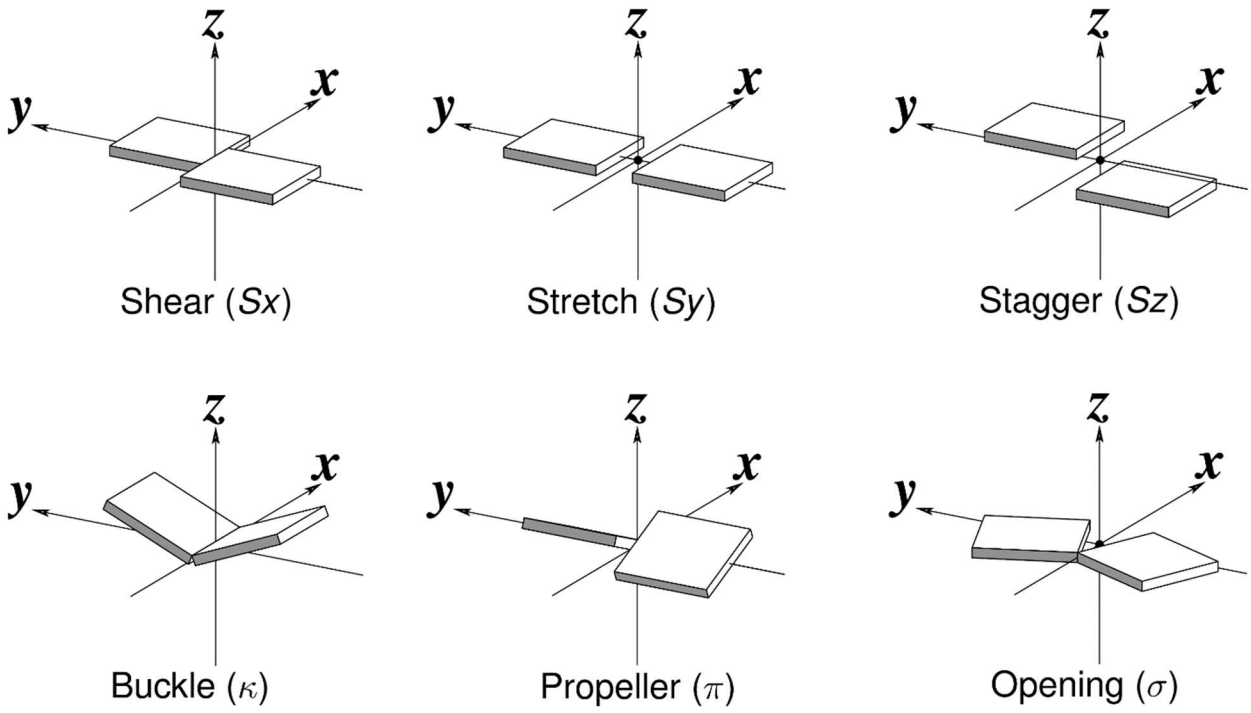


Figure 1.5. Six rigid-body parameters that describe DNA base pairing. The three translations along the *x*-, *y*- and *z*-axes are termed shear, stretch, and stagger, respectively. The three corresponding rotations are called buckle, propeller (twist), and opening. (Open source image, first published in ref. (28))

1.1.5. Beyond the harmonic approximation: deviations from the elastic model

Even in the absence of bound protein, there is evidence of spontaneous fluctuations in DNA that do not always conform to elastic model predictions. Cyclization studies performed by Widom and coworkers revealed that ~100-bp DNA fragments could form circles with roughly 3-5 orders of magnitude higher probabilities than predicted by the worm-like chain model, especially for nucleosome-preferred DNA sequences (29, 30). From these studies, extensions were made to the continuum, elastic, worm-like chain model to include sharp bending or kinking of DNA, either from local melting to produce highly flexible single strand regions (31), or from sharp kinks at sites of low stacking energy (32). Notably, Vologodskii and co-workers revisited these cyclization measurements with 105-130 bp DNA fragments and demonstrated that the wormlike chain model

in fact remains adequate for predicting cyclization probabilities of these fragments (33). Later, Ha and coworkers showed using single molecule FRET (smFRET) cyclization studies that some DNA sequences had faster looping rates than others, indicating that those sequences prefer to be deformed (34). More recently, small-angle x-ray scattering studies showed the sequence-dependence of DNA kinks in the absence of any bound protein (35). These studies underscore the ongoing debate over the validity of the wormlike chain description for sharp DNA bends and highlight the need to further characterize sharply bent DNA and to refine sequence-dependent bendability models beyond the harmonic approximation descriptions.

1.2. Deviations from B-DNA and target site recognition

Upon binding by proteins, DNA can adopt a broad range of conformations that go well beyond its canonical shape (Figure 1.6). Localized distortions such as sharp kinks are seen in DNA when bound to different DNA-bending proteins including Integration Host Factor (IHF), an *E. Coli* architectural protein (36); TATA-binding protein, a eukaryotic transcription factor (37); and MutS, a DNA mismatch repair protein (38). DNA can be looped. One protein that induces looping upon binding is the lac repressor, a gene regulatory protein required in the cellular regulation of lactose metabolism (39). Full wrapping of DNA is observed around the histone protein complex to form the nucleosome, the building block of the human chromosome structure (36, 40, 41). Local regions of DNA can also endure high degrees of overwinding or underwinding due to shearing forces associated with replicating or transcribing proteins (42-44). Lastly, DNA repair proteins recognize damaged DNA sites that exhibit irregular twisting dynamics (45).

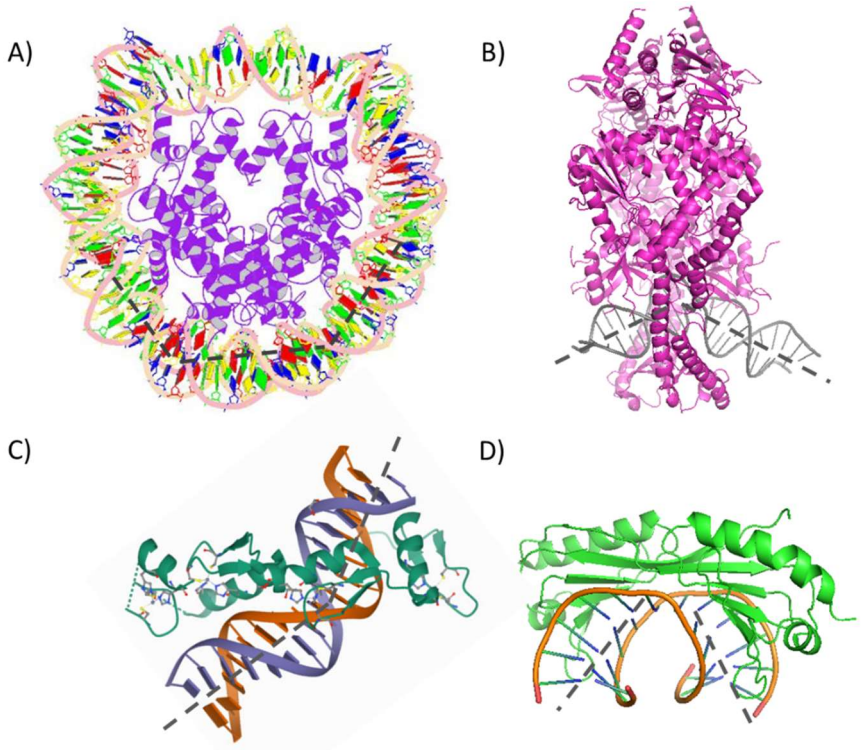


Figure 1.6. Crystal structures of DNA-bending protein complexes: **A)** Nucleosome (3LZ0), **B)** Zinc finger domain of WT1 bound to DNA (6BLW), **C)** MutS-ADPBeF3-DNA complex structure (1NNE), **D)** TATA binding protein complex (1TGH). Dashed grey lines are drawn to help guide the eye (Adapted from crystal structure images solved in refs. (37, 41, 46, 47)).

The ease by which DNA is deformed can be highly important to how efficiently site-specific proteins can recognize their respective target sites that are typically buried amidst billions of nonspecific sites. Site-specific recognition of DNA by proteins is often divided into two primary categories: (1) direct readout, in which proteins recognize target sequence by forming direct contacts with the unique moieties of the DNA bases and (2) indirect readout, in which proteins rely on the flexibility of DNA sequence to recognize their target site. These two forms of readout are not exclusive – many proteins use a combination of them to identify their target site (48), although there are several examples of proteins that rely almost entirely on indirect readout. For indirect readout, DNA sequences that can more readily adopt the contorted conformations needed

for protein binding will present lower free energy barriers to form the complex, and hence DNA deformability is thought to be a critical factor in site-specific recognition for this class of proteins.

The question that is often raised about the indirect readout mechanism for formation of DNA-protein complex is: does the protein induce the DNA bends (protein-induced conformational change) or does the protein recognize a pre-distorted conformation of DNA (conformational capture)? Identifying the mechanism by which proteins recognize their target site can offer a window of insight into how the cell identifies DNA target sites or how to design drugs that can target genes with high specificity. Spontaneous kinking and high deformability observed in DNA at length scales smaller than the persistence length (29-31, 34, 35) suggests that conformational capture plays a role, but more evidence about the role of severe bendability in context of DNA bending proteins requires more rigorous investigation. It has been a challenge to investigate population distributions and dynamics of non-canonical conformations states of DNA, and their role(s) in recognition by DNA-binding proteins.

1.3. Integration Host Factor: member of a family of DNA-bending proteins

In this thesis, I address several questions about DNA deformability in the context of protein binding and target-site recognition. I focus on localized DNA distortions, such as DNA kinking, like those induced by a site-specific architectural protein, *E. coli* Integration Host Factor (IHF) (36). IHF is a member of a family of proteins (IHF/HU family) found in many organisms and all share a similar structures, whether free or in complex with DNA. Notable cousins to IHF are HU, a nonspecific architectural protein in *E. coli* (49-51) and Hbb, a site-specific protein found in a Lyme disease causing spirochete *Borrelia burgdorferi*(52). The structures of some of the IHF/HU family proteins are shown in Figure 1.7.

IHF is a small (22 kDa) heterodimer from *E. coli*. It binds DNA both nonspecifically, in its role as a DNA compaction protein, and specifically, when required for site-specific recombination, DNA replication and transcription. IHF also recognizes and binds to several sites on bacteriophage lambda DNA, which is how the virus co-opts the protein from the *E. coli* host for lysogeny (53, 54). All contacts that IHF makes with the DNA in the specific complex are either to the minor groove (3) or to the sugar phosphate backbone. IHF is a remarkable example of a DNA-bending protein that exhibits high specificity to certain sequences but relies almost exclusively on indirect readout to recognize its target sequence (55).

IHF bends the DNA containing its specific site into a U-turn by wrapping ~35-bp DNA around three sides of the protein, which has earned it the moniker of the "master bender" (56). The crystal structure of IHF bound to one such cognate site, denoted as the H' site (Figure 1.4.), showed that IHF sharply kinks the DNA at two sites spaced ~9 bp apart, explaining how it can bring distal regions of DNA together to facilitate the formation of higher-order nucleoprotein complexes (36). The kinks in the DNA are stabilized by conserved proline residues located on two β-ribbon arms of the protein. The arms are thought to be flexible in the absence of the DNA but wrap around the DNA in the complex and make additional stabilizing contacts in the consensus region between the kink sites. Further stabilization needed to overcome the large energy penalty for the sharply kinked DNA comes from additional contacts between the flanking DNA segments and the core of the protein dimer, as well as from an extensive network of electrostatic interactions with charged residues of the protein and the phosphates in the DNA backbone (57, 58).

The specificity of IHF is determined by a small region within a larger DNA binding site. The footprint of DNA bound to IHF is ~35 base pairs (59). However, the consensus sequence identified on the basis of a large number of cognate sites for IHF consists of two short elements in one half of the footprinted region (Figure 1.5.). The consensus sequence is often described as: WATCARnnnnTTR, where W denotes A or T, R denotes purine, and n refers to any base (59, 60) though some IHF sites also contain an A-tract containing 4-6 adenines and located in the other half of the binding site (61, 62). IHF has proven to be an excellent model system for elucidating sequence-dependent characteristics that favor high deformability (>60° kinks at key sites) and features that contribute to indirect readout in binding specificity.

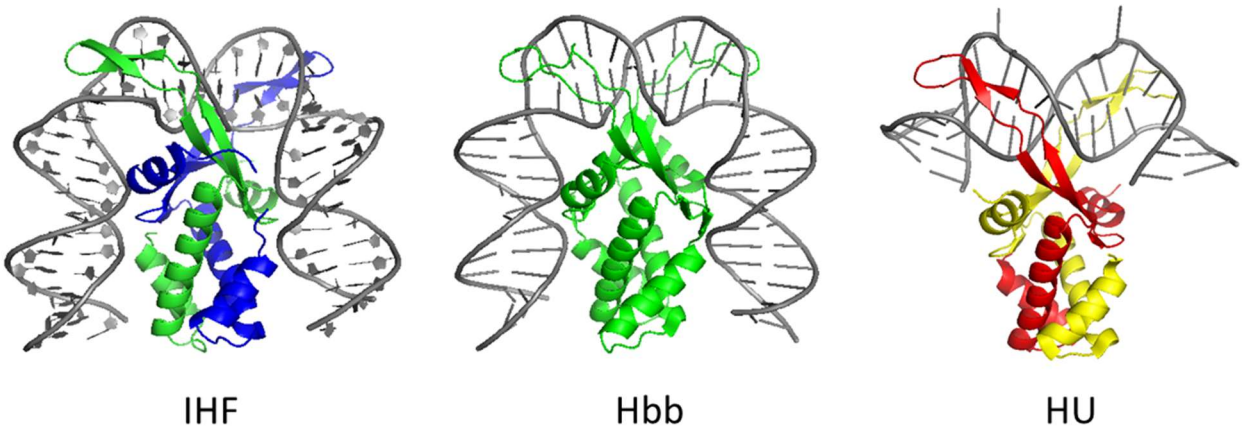


Figure 1.7. Crystal structures of IHF/HU family proteins in complex with DNA. From left to right: IHF (pdb:1IHF), Hbb (pdb:2NP2), and HU (pdb:1P78). Heterodimers are shown with the two proteins domains in two colors (IHF: blue/green and HU: red/yellow) while homodimers are shown with the two protein domains in the same color (Hbb: green). DNA is shown in gray for all figures.

1.4. Thesis Overview

This thesis addresses several different aspects of IHF-DNA binding and bending, using a combination of experimental and theoretical biophysical approaches. The thesis is organized as follows.

Chapter 2 describes the experimental methods and analysis techniques used in these studies. First, several steady-state biophysical measurement techniques are discussed including absorption, Förster resonance energy transfer (FRET), and circular dichroism. Then, time-resolved ensemble FRET techniques of fluorescence lifetime spectroscopy and laser-temperature jump (T-jump) spectroscopy are described. Finally, the protocol for SELEX (systemic enrichment of ligands by exponential enrichment) is described.

In Chapter 3, I investigate if kinked DNA in complex with IHF holds its deformed structure or is free to explore many conformations. The fluorescence lifetime studies performed suggest highly dynamic IHF-DNA complexes that sample multiple conformations with varying degrees of DNA bending. The population distribution of bent DNA states depended on the bound DNA sequence, highlighting the finely tuned interactions between IHF and its many biological targets that keep the DNA bent (or not). I show how the conformational distributions of IHF-DNA complexes differ when an A-tract is included in the flanking arm of the U-bent DNA, as in one specific site H', or is lacking, as in a different specific site, H1. We describe sequences lacking an A-tract to be more "flappy" in complex, because one of the bent arms of DNA is unable to form the favorable interactions with the side of the IHF molecule. These studies have important implications for how the different sites (H' and H1) within λ-phage sequence are leveraged to facilitate the formation of a higher-order DNA-protein complex called the "intasome", which is required for the mechanism by which the viral DNA infects *E. coli* host DNA.

Chapter 4 investigates the competition between nonspecific and specific binding modes of IHF to DNA. Previously reported values of specific binding constants of IHF to one of its cognate sites, the H' site, vary by nearly three orders of magnitude and even less is known about its

nonspecific binding affinities. Throughout *E. coli's* life cycle, IHF concentrations may fluctuate by nearly 5-fold (63), suggesting varying levels of competition between specific versus nonspecific binding during each cell phase. Previous studies used lattice binding models to quantify this competitive binding of IHF to one of its cognate sites, H', as measured by either isothermal titration calorimetry (64) or gel-shift assays (65). These studies restricted the analysis to three major states: free DNA, one protein bound specifically, or multiple proteins bound nonspecifically. Here, I revisited the three-state lattice model applied to binding isotherms on IHF-H', using FRET measurements on end-labeled DNA to monitor DNA bending and fluorescence anisotropy to monitor protein binding stoichiometry. I also monitored IHF binding to a nonspecific DNA. While this three-state model provided an excellent description to the four binding isotherms (FRET and anisotropy for both specific and nonspecific DNA constructs) up to [IHF] ~ 10 μM, for even higher excess protein conditions, where nonspecific binding dominates, discrepancies between the data and the fits became evident. These results suggested that this minimal description, in which nonspecific protein binding displaces a specifically bound protein, may be incomplete. I expanded the model to include a fourth "mixed state", in which one DNA binding site could accommodate both specifically and nonspecifically bound IHF. However, global modeling of all four binding isotherms marginally improved the fit. These results suggest further investigation into the competition between specific and nonspecific binding by IHF is warranted.

Chapter 5 presents experiments that were designed to test whether transient Hoogsteen pairing (an alternate to the canonical Watson-Crick base paring) at or near one of the kink sites in the H' sequence facilitates DNA kinking and formation of the IHF-DNA specific complex. Transient Hoogsteen pairing has been observed in B-DNA in solution from NMR studies, with higher probability of such pairing observed at nearest neighbor steps of highest flexibility (66, 67),

which suggested that DNA flexibility and Hoogsteen base-pairing are related. To examine whether Hoogsteen formation was relevant for DNA kinking in the IHF-H' complexes, T-jump studies were performed on IHF-bound H' sequence variants with 7-deaza adenine (7dA) substitutes at select sites, designed to complement time-resolved NMR studies using the same constructs (68). The 7dA modification disrupts a hydrogen bond needed to stabilize Hoogsteen pairing. Our prediction was that if Hoogsteen pairing enabled DNA kinking, then a 7dA substitution for adenine at a kink site in the DNA should destabilize the complex and slow down the rates for DNA kinking in the complex. However, little change to complex stability or kinetic rates was detected with 7dA insertions at one of the kink sites. Interestingly, a 7dA substitution in the TTG consensus region of the H' binding site, far removed from the kink sites, resulted in an ~40-fold increase in the complex stability. Taken together, the results from this study rule out Hoogsteen pairing as an important factor in enhancing DNA "kinkability" for the IHF substrates; instead, we attribute the dramatic effect of the 7dA substitution at the TTG site as the effect of reduced base-stacking interactions between 7dA and its neighboring nucleotides (69), which likely enhanced local DNA "twistability" previously shown to be an important indirect readout feature of the TTG consensus region (70).

In Chapter 6, we ask: what sequences renders the DNA to be more kinkable? Elucidating the rules that govern DNA sequence-dependent deformability and how these variations in flexibility recruit proteins to binding sites has thus far proven difficult. In this chapter, I present SELEX studies aimed to identify highly deformable DNA sequences using in vitro selection of random DNA sequences that bind with high affinity to DNA bending proteins of the architectural IHF/HU family. This protein family recognizes ~35 bp DNA sites primarily by indirect readout and bends the DNA at these sites into a U-shape, with the DNA severely kinked (~70° roll) at two

sites ~9 bp apart. SELEX studies were performed using two site-specific proteins from the IHF/HU family: IHF and Hbb. The DNA library included a randomized 6-bp region located at one of the kink sites and outside the known consensus region of an IHF positioning sequence. Several rounds of SELEX were carried out to separate the tight-binders ("winners") from the weak-binders ("losers"). Interestingly, little correlation was observed between sequence winners and losers between IHF and Hbb, suggesting subtle differences in the sequence preferences for the two proteins despite the very similar structures of the bound complexes. The winners and losers for both proteins were robustly evaluated to discern patterns in sequence-dependent shape and flexibility. Our studies suggest that several factors contribute to the favorability of DNA kinking by IHF and Hbb. The winners, evaluated using DNAshape by our collaborators Devesh Bhimsaria (BioInformaticals Inc.) and Aseem Ansari (St. Jude Children's Research Hospital), revealed a "two-dip" structure in the minor groove width with narrow minor grooves flanking the kink site, suggesting that the beta arms of the proteins may be recognizing pre-formed shapes in the DNA at the kink sites prior to proline intercalation that stabilizes the kinked conformation. Energy calculations to deform DNA, performed by Robert Young and Wilma Olson (Rutgers University – New Brunswick) based on DNA elasticity functions developed by Olson and coworkers, show a correlation, albeit weak, for IHF (though not for Hbb) between binding affinity and ease of deformability at the randomized kink site. These studies indicate that both DNA deformability and pre-distorted shapes likely contribute to indirect readout by IHF and Hbb.

Collectively, my studies contribute to the understanding of DNA deformability and its role in target-site recognition by site-specific proteins. They showcase a variety of experimental biophysical techniques that could be applied to other protein-DNA systems, thus helping to advance our understanding of DNA-protein interactions that dictate cellular behavior.

References:

1. Hardison RC. The Structure of DNA. Working with Molecular Genetics2008.
2. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737-8.
3. Rice PA, Correll CC. Introduction. In: Rice PA, Correll CC, editors. Protein-Nucleic Acid Interactions. Cambridge: Royal Society of Chemistry; 2008.
4. Seeman NC, Rosenberg JM, Rich A. Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A. 1976;73(3):804-8.
5. Howard JMoMPatCS, Mass: Sinauer Associates, Publishers, 2001. Print.
6. Hogan M, LeGrange J, Austin B. Dependence of DNA helix flexibility on base composition. Nature. 1983;304(5928):752-4.
7. Hogan M, Dattagupta N, Crothers DM. Transient electric dichroism of rod-like DNA molecules. Proc Natl Acad Sci U S A. 1978;75(1):195-9.
8. Hagerman PJ. Investigation of the flexibility of DNA using transient electric birefringence. Biopolymers. 1981;20(7):1503-35.
9. Shore D, Langowski J, Baldwin RL. DNA flexibility studied by covalent closure of short fragments into circles. Proc Natl Acad Sci U S A. 1981;78(8):4833-7.
10. Crothers DM, Drak J, Kahn JD, Levene SD. DNA bending, flexibility, and helical repeat by cyclization kinetics. Methods Enzymol. 1992;212:3-29.
11. Levene SD, Crothers DM. Ring closure probabilities for DNA fragments by Monte Carlo simulation. J Mol Biol. 1986;189(1):61-72.
12. Bustamante C, Marko JF, Siggia ED, Smith S. Entropic elasticity of lambda-phage DNA. Science. 1994;265(5178):1599-600.
13. Bustamante C, Smith SB, Liphardt J, Smith D. Single-molecule studies of DNA mechanics. Curr Opin Struct Biol. 2000;10(3):279-85.
14. Baumann CG, Smith SB, Bloomfield VA, Bustamante C. Ionic effects on the elasticity of single DNA molecules. Proc Natl Acad Sci U S A. 1997;94(12):6185-90.
15. Mills JB, Hagerman PJ. Origin of the intrinsic rigidity of DNA. Nucleic Acids Res. 2004;32(13):4055-9.
16. Frank-Kamenetskii F. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. Biopolymers. 1971;10(12):2623-4.
17. Protozanova E, Yakovchuk P, Frank-Kamenetskii MD. Stacked-unstacked equilibrium at the nick site of DNA. J Mol Biol. 2004;342(3):775-85.
18. Wartell RM, Benight AS. Fluctuational base-pair opening in DNA at temperatures below the helix-coil transition region. Biopolymers. 1982;21(10):2069-81.
19. Coman D, Russu IM. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. Biophys J. 2005;89(5):3285-92.
20. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res. 2006;34(2):564-74.
21. Allawi HT, SantaLucia J, Jr. Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA. Biochemistry. 1998;37(8):2170-9.
22. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A. 1998;95(19):11163-8.
23. Beveridge DL, Barreiro G, Byun KS, Case DA, Cheatham TE, 3rd, Dixit SB, et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. Biophys J. 2004;87(6):3799-813.

24. Pasi M, Maddocks JH, Beveridge D, Bishop TC, Case DA, Cheatham T, 3rd, et al. muABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. Nucleic Acids Res. 2014;42(19):12272-83.
25. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A. 1998;95(19):11163-8.
26. Olson WK, Zhurkin VB. Modeling DNA deformations. Curr Opin Struct Biol. 2000;10(3):286-97.
27. Dixit SB, Beveridge DL, Case DA, Cheatham TE, 3rd, Giudice E, Lankas F, et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. Biophys J. 2005;89(6):3721-40.
28. Zheng G, Colasanti AV, Lu XJ, Olson WK. 3DNALandscapes: a database for exploring the conformational features of DNA. Nucleic Acids Res. 2010;38(Database issue):D267-74.
29. Cloutier TE, Widom J. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. Proc Natl Acad Sci U S A. 2005;102(10):3645-50.
30. Cloutier TE, Widom J. Spontaneous sharp bending of double-stranded DNA. Mol Cell. 2004;14(3):355-62.
31. Yan J, Marko JF. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. Phys Rev Lett. 2004;93(10):108108.
32. Wiggins PA, Phillips R, Nelson PC. Exact theory of kinkable elastic polymers. Phys Rev E Stat Nonlin Soft Matter Phys. 2005;71(2 Pt 1):021909.
33. Du Q, Smith C, Shiffeldrim N, Vologodskaia M, Vologodskii A. Cyclization of short DNA fragments and bending fluctuations of the double helix. Proc Natl Acad Sci U S A. 2005;102(15):5397-402.
34. Vafabakhsh R, Ha T. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. Science. 2012;337(6098):1097-101.
35. Schindler T, Gonzalez A, Boopathi R, Roda MM, Romero-Santacreu L, Wildes A, et al. Kinky DNA in solution: Small-angle-scattering study of a nucleosome positioning sequence. Physical Review E. 2018;98(4).
36. Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell. 1996;87(7):1295-306.
37. Nikolov DB, Chen H, Halay ED, Hoffman A, Roeder RG, Burley SK. Crystal structure of a human TATA box-binding protein/TATA element complex. Proc Natl Acad Sci U S A. 1996;93(10):4862-7.
38. Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK. The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature. 2000;407(6805):711-7.
39. Lewis M. Response: DNA Looping and Lac Repressor--CAP Interaction. Science. 1996;274(5294):1931-2.
40. Garcia HG, Grayson P, Han L, Inamdar M, Kondev J, Nelson PC, et al. Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. Biopolymers. 2007;85(2):115-30.
41. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251-60.
42. Fogg JM, Randall GL, Pettitt BM, Sumners WL, Harris SA, Zechiedrich L. Bullied no more: when and how DNA shoves proteins around. Q Rev Biophys. 2012;45(3):257-99.
43. Kim Y, Eom SH, Wang J, Lee DS, Suh SW, Steitz TA. Crystal structure of Thermus aquaticus DNA polymerase. Nature. 1995;376(6541):612-6.
44. Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature. 2002;417(6890):712-9.
45. Chakraborty S, Steinbach PJ, Paul D, Mu H, Broyde S, Min JH, et al. Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. Nucleic Acids Res. 2018;46(3):1240-55.

46. Wang D, Horton JR, Zheng Y, Blumenthal RM, Zhang X, Cheng X. Role for first zinc finger of WT1 in DNA sequence specificity: Denys-Drash syndrome-associated WT1 mutant in ZF1 enhances affinity for a subset of WT1 binding sites. Nucleic Acids Res. 2018;46(8):3864-77.
47. Alani E, Lee JY, Schofield MJ, Kijas AW, Hsieh P, Yang W. Crystal structure and biochemical analysis of the MutS.ADP.beryllium fluoride complex suggests a conserved mechanism for ATP interactions in mismatch repair. J Biol Chem. 2003;278(18):16088-94.
48. Lawson CLB, H. M., Indirect readout of DNA sequence by proteins. In Protein-Nucleic Acid Interactions, Rice, P. A.; Correll, C. C., Eds. Royal Society of Chemistry: Cambridge, 2008.
49. Swinger KK, Lemberg KM, Zhang Y, Rice PA. Flexible DNA bending in HU-DNA cocrystal structures. EMBO J. 2003;22(14):3749-60.
50. Swinger KK, Rice PA. IHF and HU: flexible architects of bent DNA. Curr Opin Struct Biol. 2004;14(1):28-35.
51. Swinger KK, Rice PA. Structure-based analysis of HU-DNA binding. J Mol Biol. 2007;365(4):1005-16.
52. Mouw KM, Rice PA. Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending Hbb. Molecular Microbiology. 2007;63:1319-30.
53. Craig NL, Nash HA. E. coli integration host factor binds to specific sites in DNA. Cell. 1984;39(3 Pt 2):707-16.
54. Yang CC, Nash HA. The interaction of *E. coli* IHF protein with its specific binding sites. Cell. 1989;57(5):869-80.
55. Lawson CL, Berman HM. Indirect readout of DNA sequence by proteins. In: Rice PA, Correll CC, editors. Protein-Nucleic Acid Interactions. Cambridge: Royal Society of Chemistry; 2008.
56. Travers A. DNA-protein interactions: IHF--the master bender. Curr Biol. 1997;7(4):R252-4.
57. Holbrook JA, Tsodikov OV, Saecker RM, Record MT, Jr. Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. Journal of molecular biology. 2001;310:379-401.
58. Vander Meulen KA, Saecker RM, Record MT, Jr. Formation of a wrapped DNA-protein interface: experimental characterization and analysis of the large contributions of ions and water to the thermodynamics of binding IHF to H' DNA. J Mol Biol. 2008;377(1):9-27.
59. Goodrich JA, Schwartz ML, McClure WR. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). Nucleic Acids Res. 1990;18(17):4993-5000.
60. Engelhorn M, Boccard F, Murtin C, Prentki P, Geiselmann J. In vivo interaction of the Escherichia coli integration host factor with its specific binding sites. Nucleic Acids Res. 1995;23(15):2959-65.
61. Hales LM, Gumport RI, Gardner JF. Determining the DNA sequence elements required for binding integration host factor to two different target sites. J Bacteriol. 1994;176(10):2999-3006.
62. Hales LM, Gumport RI, Gardner JF. Examining the contribution of a dA+dT element to the conformation of Escherichia coli integration host factor-DNA complexes. Nucleic Acids Res. 1996;24(9):1780-6.
63. Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J Bacteriol. 1999;181(20):6361-70.
64. Holbrook JA, Tsodikov OV, Saecker RM, Record MT, Jr. Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. J Mol Biol. 2001;310(2):379-401.
65. Aeling KA, Steffen NR, Johnson M, Hatfield GW, Lathrop RH, Senear DF. DNA deformation energy as an indirect recognition mechanism in protein-DNA interactions. IEEE/ACM Trans Comput Biol Bioinform. 2007;4(1):117-25.

66. Alvey HS, Gottardo FL, Nikolova EN, Al-Hashimi HM. Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. Nat Commun. 2014;5:4786.
67. Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM. Transient Hoogsteen base pairs in canonical duplex DNA. Nature. 2011;470(7335):498-502.
68. Zhou H, Sathyamoorthy B, Stelling A, Xu Y, Xue Y, Pigli YZ, et al. Characterizing Watson-Crick versus Hoogsteen Base Pairing in a DNA-Protein Complex Using Nuclear Magnetic Resonance and Site-Specifically (13)C- and (15)N-Labeled DNA. Biochemistry. 2019;58(15):1963-74.
69. Kowal EA, Ganguly M, Pallan PS, Marky LA, Gold B, Egli M, et al. Altering the electrostatic potential in the major groove: thermodynamic and structural characterization of 7-deaza-2'-deoxyadenosine:dT base pairing in DNA. J Phys Chem B. 2011;115(47):13925-34.
70. Lynch TW, Read EK, Mattis AN, Gardner JF, Rice PA. Integration host factor: putting a twist on protein-DNA recognition. J Mol Biol. 2003;330(3):493-502.

Chapter 2

Materials and Methods

Sections 2.3, 2.3.1. and 2.3.6. of this chapter adapted from [Connolly, M; Arra, A.; Zvoda, Z.; Steinbach, P.J.; Rice, P.A.; and Ansari, A., Static kinks or flexible hinges: multiple conformations of bent DNA bound to Integration Host Factor revealed by fluorescence lifetime measurements, J. Phys. Chem. B 2018, 122, 49, 11519-11534] with permission from American Chemical Society.

Sections 2.2.4., 2.4-2.4.9. of this chapter adapted from [Velmurugu, Y.; Vivas, P.; Connolly, M.; Kuznetsov, S. V.; Rice, P. A.; and Ansari, A., Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein, Nucleic acids research, 2018, 46, 4, 1741–1755] with permission from Oxford University Press.

2.1 Materials

2.1.1. DNA Substrates

DNA sequences used for FRET-based studies were ordered from Keck (with gel purification) or from IDT (with HPLC purification). Two different labeling strategies were used and are shown for the H' sequence in Table 2.1. Design I constructs were end-labeled with FRET-pair fluorescein (F) and TAMRA (R) which were attached to thymidine overhangs at the at 5'-end of the top and bottom strands respectively by six-carbon phosphoradimite linkers. Design II constructs included an internally-labeled fluorescein (F) - attached to a thymidine located 10 nucleotides from the 3'-end of the top strand - and an end-labelled Atto550 (At) attached to the 3' end of the bottom strand by a six-carbon linker to the phosphate group of the DNA backbone.

Table 2.1. DNA Constructs with Labelling Designs. Attached labels are highlighted: Fluorescein (yellow), TAMRA (green) and Atto550 (cyan)

Construct	Sequence
H' (design I)	5' – TGGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC CCGGTTTTTTCGTAACGAATAGTTAAACAACGTGGT – 5'
H' (design II)	5' – GGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC PCCGGTTTTTTCGTAACGAATAGTTAAACAACGTGG – 5'

DNA concentrations were determined by absorbance measurements at 260nm, with extinction coefficients determined by IDT oligo analyzer (https://www.idtdna.com/calc/analyzer/). Labeling efficiencies were determined by simultaneous absorbance measurements of the attached fluorophore. Fluorophore extinction coefficients were obtained from Molecular Probes (www.glenresearch.com/Technical/Extinctions.html) and are summarized in Table 2.2. The percentage of labeled DNA in solution was estimated to be >87% for all donor-labeled strands and >95% for all acceptor-labeled strands.

Duplex DNA was formed by annealing complimentary oligomers. Equimolar concentrations of top and bottom strands were used unless otherwise specified. The annealing buffer used was 20 mM Tris-HCl, pH 8.0, 1 mM ethylenediaminetetraacetic acid (EDTA), and 100 mM KCl. The mixture of oligomers was heated in a water bath at 85 °C for 10 minutes, then allowed to cool slowly at room temperature.

Table 2.2. Fluorophores, symbols, and absorption properties.

Flurophore	**Symbol**	**λ_{max} (nm)**	**Extinction Coefficient at λ_{max} ($M^{-1}cm^{-1}$)**
Fluorescein	F	494	75,000
TAMRA	R	555	91,000
Atto-550	At	560	131,200
Cyanine-3	Cy3	547	136,000

2.1.2. Proteins

IHF and Hbb proteins were generous gifts from the laboratory of Prof. Phoebe Rice of the University of Chicago. Protein purification methods are described previously for IHF (1) and Hbb (2). The IHF protein is stored in a storage buffer (20 mM Tris–HCl (pH 8.0), 1 mM EDTA, 200

mM NaCl, 0.01% NP-40). Hbb storage buffer also included and 5 mM DTT. Droplets of proteins were first flash-frozen in liquid nitrogen prior to storage in cryogenic tubes at −80 °C. Individual frozen droplets were diluted into the binding buffer, as needed. All measurements were performed in binding buffer: 20 mM Tris–HCl (pH 8.0), 1 mM EDTA, 0.01% NP-40 with salt concentrations ranging from 100 to 300 mM KCl and 5 mM DTT for Hbb samples. Protein concentrations were determined by absorbance measurements at 280 nm with an extinction coefficient of 5,800 $M^{-1}cm^{-1}$ and 19,9400 $M^{-1}cm^{-1}$ for IHF and Hbb respectively.

2.2 Steady-state Measurements

The steady-state fluorescence emission spectra and anisotropies were measured on a FluoroMax4 spectrofluorometer (Jobin Yvon, Inc., NJ, USA), with samples loaded in a 100-µL quartz cuvette (Starna 26.100F-Q-10/Z20). Steady-state circular dichroism (CD) measurements were performed using a Jasco J-810.

2.2.1. FRET efficiency determined by donor emission

To measure FRET efficiency, donor emission intensity is measured for two nearly-identical DNA substrates, one that is donor-labeled (DNA_D) and the other that is donor-acceptor labelled (DNA_DA). For both substrates, under identical experimental conditions, the donor is excited, and the donor emission is collected from 500 nm to 700 nm for both substrates. Due to the presence of the acceptor molecule, the emission intensity drops for the donor-acceptor labelled substrate (I_{DA}) relative to the donor-labeled substrate (I_D) corresponding to the amount of energy transferred from donor to acceptor. We quantify the FRET efficiency of this energy transfer by

$$E = 1 - \frac{I_{DA}}{I_D} \qquad \text{Eq. 2.1.}$$

2.2.2. FRET efficiency determined by Acceptor Ratio

Alternatively, we can understand quantify FRET by measuring acceptor ratio. In this method, fluorescence emission spectra is collected for DNA_DA from 500 nm to 700 nm, with excitation of the donor (F) at 485 nm. On the same sample, the emission spectra from direct excitation of the acceptor were also collected: For design I samples, emission was collected 565 nm to 700 nm, with excitation at 555 nm; For design II constructs, emission was collected from 560 nm to 700 nm, with excitation at 555 nm. Acceptor ratio was determined from the measured spectra as follows. A normalized emission spectrum of a DNA_D was subtracted from the DNA_DA emission spectrum to isolate the acceptor emission in the DNA_DA that results from FRET. DNA_D samples were prepared under identical experimental conditions and excited at 485 nm. To calculate acceptor ratio of the ensemble (denoted as r_a), the area under this corrected acceptor-emission spectrum, from 565 nm to 595 nm, was divided by the area under the directly-excited acceptor emission spectrum, also integrated from 565 nm to 595 nm. For each sample, at least two independent measurements were performed at each temperature, and errors were estimated from the standard deviations from the average value.

2.2.3. Steady-State Anisotropy

Steady-state anisotropy measurements were performed on DNA_DA constructs, in the presence and absence of IHF, at 20 °C. Fluorescein donor molecules measurements were made with an excitation of 485 nm and emission collection was performed at 520 nm. For each set of measurements, the emission intensities for the parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) orientations were

collected using an integration time of 0.1 s, and the data were processed using a program provided by Horiba to obtain anisotropy values, defined as:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad \text{Eq 2.2.}$$

Errors in the measured anisotropy were calculated as standard deviations from the average of five independent sets of measurements.

2.2.4. Circular dichroism measurements

The thermal stability of the IHF protein was measured in the absence and presence of H' DNA, using far-UV circular dichroism (CD) measurements at 222 nm, in the temperature range 15-85 °C. Two sets of measurements were carried out for IHF (30 μM protein concentration), IHF-H' (12.5:15 μM protein:DNA concentration), and H' DNA only (15 μM). Under these conditions, the CD signal for the DNA only at 222 nm was 3.2 ± 0.1 mdeg at 20 °C compared with -15.3 ± 0.1 mdeg for the IHF-H' complex and was weakly dependent on temperature, decreasing to 2.6 ± 0.2 mdeg at 85 °C. The DNA contribution was subtracted from the CD signal of the IHF-H' complex prior to further analysis. The CD measurements for the complex were found to be reversible after heating up to at least 60 °C.

The thermal unfolding profiles were analyzed in terms of a two-state van't Hoff transition with linear upper and lower baselines using the following equation:

$$A(T) = A_L(T) + [A_U(T) - A_L(T)]f_U \quad \text{Eq. 2.3.}$$

where $A(T)$ is the CD signal of the IHF-H' complex minus the CD signal of H'-DNA alone, $A_L(T)$ and $A_U(T)$ are the lower and upper baselines, respectively, parameterized as straight lines, and f_u is the fraction of unfolded molecules, written in terms of the van't Hoff parameters as:

$$f_u = \frac{A(T) - A_L(T)}{A_U(T) - A_L(T)} = \left(1 + \exp\left[-\frac{\Delta H_{vH}}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right]\right)^{-1} \quad \text{Eq. 2.4.}$$

In Eq. 2.2., ΔH_{vH} is the van't Hoff enthalpy change for the unfolding transition, and T_m the corresponding melting temperature.

The fraction of unfolded molecules as a function of temperature obtained from the van't Hoff analysis. The uncertainties in the computed fractions, as well as in the reported T_m values, are the standard deviations from fits to two independent sets of CD measurements.

2.2.5. Determination of active protein concentration

Initial estimates of the IHF protein concentration were obtained from absorption measurements at 276 nm, with extinction coefficient 5800 $M^{-1}cm^{-1}$. Lifetime studies (Chapter 3) required accurate determination of the active protein concentration, which were determined by performing equilibrium titration experiments on each day of the experiments. In those experiments, DNA concentrations were held constant at 1 μM while the protein concentration (as estimated by absorption measurements) was varied from 0–3 μM, and the acceptor ratio (described below) on double-labeled (DNA_DA) samples was measured as a function of [IHF]. The protein concentration at which we observed the maximum acceptor ratio was determined to represent a 1:1 complex, as illustrated in Figure 2.1; the "active" protein concentration under these conditions

was assumed to be 1 μM, the same concentration as the DNA in these measurements. All protein concentrations in Chapter 3 of the text are rescaled to refer to this active protein concentration.

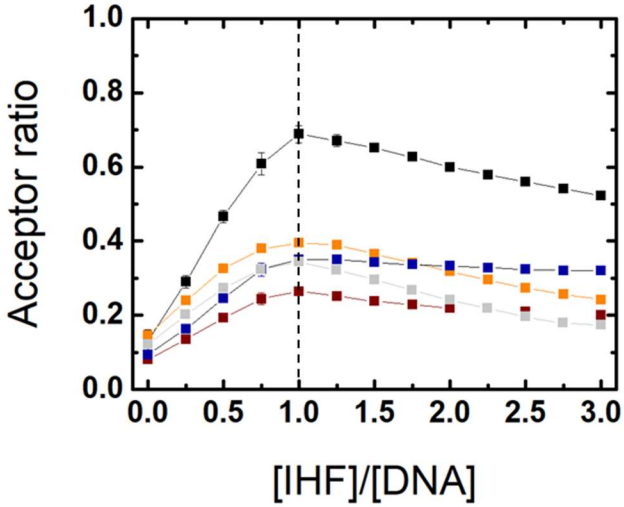


Figure 2.1. Estimating active protein concentration by binding isotherms for different IHF-DNA complexes. Acceptor ratio values measured on 1 μM DNA_DA samples are plotted versus the ratio of IHF/DNA concentrations, for sequences: H' (black), H1 (maroon), H1_CTloop (blue), and H'44A (orange) in 100mM KCl. A dashed line is added at [IHF]/[DNA] = 1 to indicate equimolar concentrations of IHF and DNA. Sequence information can be found in Table 3.1.

2.2.6. Determination of K_d from binding isotherms with 2 states

To determine the effect of the different labeled sequences on the IHF specific binding affinities, we performed equilibrium titration measurements with IHF and different DNA constructs at 20 °C. Comparative binding affinity measurements were done at 300 mM KCl, to bring the dissociation constants of the complexes (K_d) in the > nM range. At 100 mM KCl, where most lifetime and laser temperature-jump measurements were done, the K_d for the IHF-H' complex was previously found to be in the pM range (1, 3-5), such that conventional titration experiments cannot measure these K_d values accurately.

Evidence for pM K_d in 100 mM KCl conditions is provided in Chapters 3, 4 , and 5 where we show and extrapolated the K_d to these ionic based on K_d values measured in the 200 mM – 400 mM range. The extrapolation is a linear fit described by:

$$K_d = K_{d_0}\left(\frac{[KCl]}{[KCl]_0}\right)^{SK_d} \quad \text{Eq. 2.5.}$$

Where K_{d_0} is a reference K_d at salt conditions $[KCl]_0$, and SK_d is the expected slope of $\log(K_d)$ vs. $\log([KCl])$.

Several samples were prepared with a fixed DNA concentration (50 nM) and varying concentrations of IHF, ranging from 1 nM to 100 μM. At each protein concentration, fluorescence emission spectra were collected and acceptor ratio (r_a) was calculated as described above. Binding isotherms were obtained by plotting $r_a$versus IHF concentration for each sample, and the corresponding Kd for the IHF-DNA complex was determined by fitting each r_a versus [IHF] profile to a 1:1 binding isotherm, as described below. Assuming only two states for the protein and DNA, free and bound, the measured or r_a depends on the fraction of DNA in complex (fx) as:

$$r_a = r_a^x(f_x) + r_a^{free}(1 - f_x) \quad \text{Eq. 2.6.}$$

where r_a^x is the acceptor ratio for the complex, r_a^{free} is the acceptor ratio for free DNA, and f_x is obtained from:

$$f_x = \frac{1}{2D_0}\left[(P_0 + D_0 + K_d) - \sqrt{(P_0 + D_0 + K_d)^2 - 4(P_0D_0)}\right] \quad \text{Eq. 2.7.}$$

where P_0 and D_0 are the total IHF and DNA concentrations, respectively. Descriptions of binding isotherms with fits that exceed two states are discussed in Chapter 4.

2.3 Fluorescence Lifetime measurements

Fluorescence decay curves were measured with a PicoMaster fluorescence lifetime spectrometer (HORIBA-PTI, London, Ontario, Canada) equipped with time-correlated single photon counting (TSCPC) electronics (6). For all FRET measurements, decay traces were measured for donor-only duplexes without acceptor, denoted as DNA_D, as well as donor–acceptor-labeled duplexes denoted as DNA_DA. The excitation source was a Fianium Whitelase Supercontinuum laser system (maximum power 4W), which produces ~6 ps broad band pulses. For excitation of fluorescein, the laser pulses were passed through a monochromator set at 485 nm (bandpass 10 nm) followed by a 488 ± 10 nm bandpass filter. The emission from the sample was collected orthogonal to the excitation beam after passing through a 496 nm longpass filter (Semrock BrightLine FF01-496/LP25), followed by another monochromator set at 520 nm (bandpass 10 nm), and detected by a Hamamatsu microchannel plate photomultiplier (MCP-650). The instrument response function (IRF) of the system was measured using a dilute aqueous solution of Ludox (Sigma-Aldrich). The full width at half maximum (fwhm) of the IRF was ~100 ps. Fluorescence decay curves were recorded on a 100 ns timescale, resolved into 4096 channels, to a total of 10,000 counts in the peak channel, with the repetition rate of the laser adjusted to 10 MHz.

2.3.1. Maximum entropy analysis of the fluorescence decay traces

Fluorescence decay curves were analyzed using a maximum entropy method (MEM) in which the effective distribution of log-lifetimes $f(\log \tau)$ was inferred from the decay traces using the program MemExp (available online), and described in detail elsewhere.(7, 8) The signal measured at time t_i was fit by the expression:

$$F_i = D_0 \int_{-\infty}^{\infty} d\log\tau \, f(\log \tau) \int_{t_0}^{\min(t_i, t_f)} dt' R(t' + \delta) \, e^{-(t_i - t')/\tau} \qquad \text{Eq. 2.8.}$$

where R is the measured instrument response function, δ is the zero-time shift and D_0 is a normalization constant. The instrument response is appreciable in the range $[t_0, t_f]$. The zero-time shift was determined using Brent's method of optimization and preliminary MEM calculations, each performed with δ fixed at a different value. A similar estimation of δ has been reported.(9) The Poisson deviance between the fit and data was minimized while maximizing the entropy of the f distribution.

The MEM inverts fluorescence decay traces into lifetime distributions without any *a priori* assumptions about the number of exponential terms. The MEM outputs of the donor-only samples gave single-peaked, narrow distributions, consistent with single-exponential decay. Fits to discrete exponential decays with two or more exponentials yielded less than ~1% amplitude in the additional decay components. Average FRET efficiencies measured on donor-acceptor labeled (DNA_DA) samples were computed from the MEM distributions.

2.3.2. Multi-modal interpretations of MEM distributions

In the presence of the IHF, lifetime decays on some DNA_DA samples yielded bimodal lifetime distributions. In Figure 2.2., we show a comparison of MEM distributions for several IHF-bound substrates to highlight the reproducibility of our methods. Gaussian fitting of lifetime

distributions had been done previously by our group to quantify ensemble populations of mismatched DNA in the presence and absence of mismatch repair protein Rad4/XPC (10). However, here gaussian fitting failed to aptly fit lifetime distributions of IHF-DNA complexes (Figure 2.3.). Thus, to quantify populations of differing FRET states, we partitioned our lifetime distributions as dictated by the local minimum between the adjacent peaks. Within each peak, the average lifetime, denoted as $< \tau_1 >$ or $< \tau_2 >$ was computed as well as the corresponding average FRET $< E_{1,2} > = 1 - \frac{<\tau_{1,2}>}{<\tau_D>}$. The fractional population in each component was obtained from the sum of the amplitudes A_j from that part of the distribution. The uncertainties reported are standard deviations from the average of two independent sets of measurements.

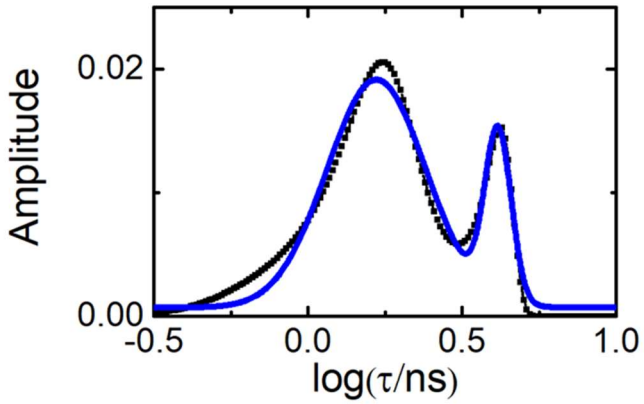


Figure 2.2. Inaccuracy of Gaussian fitting of IHF lifetime data. MEM distribution of IHF-H' (design I) fit by 2-gaussian fitting.

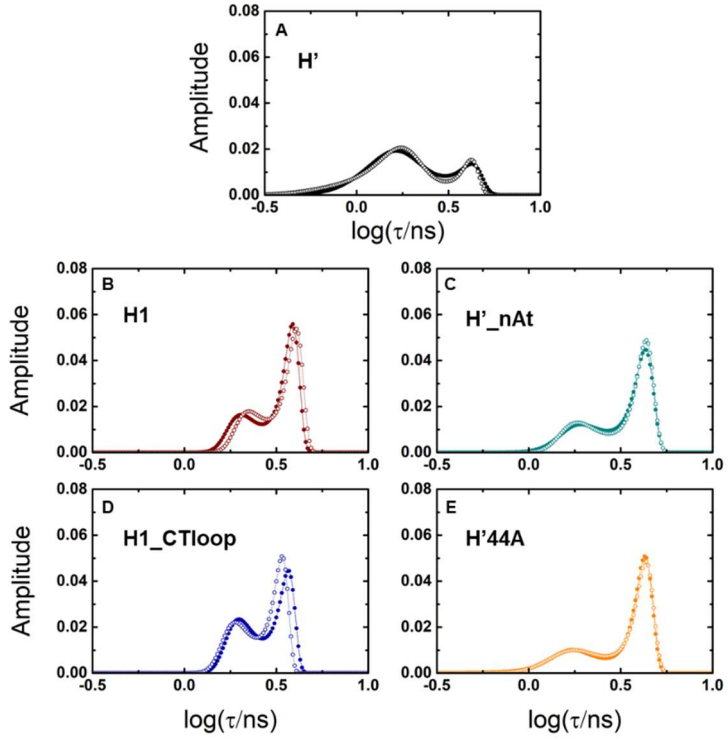


Figure 2.3. Reproducibility of MEM distributions from measurements on IHF-DNA_DA (design I) complexes. MEM outputs from two sets of measurements (closed versus open symbols) are shown for **(A)** IHF-H', **(B)** IHF-H1, **(C)** IHF-H'_nAt, **(D)** IHF-H1_CTloop, and **(E)** IHF-H'44A. See Table 3.1. to see full sequences.

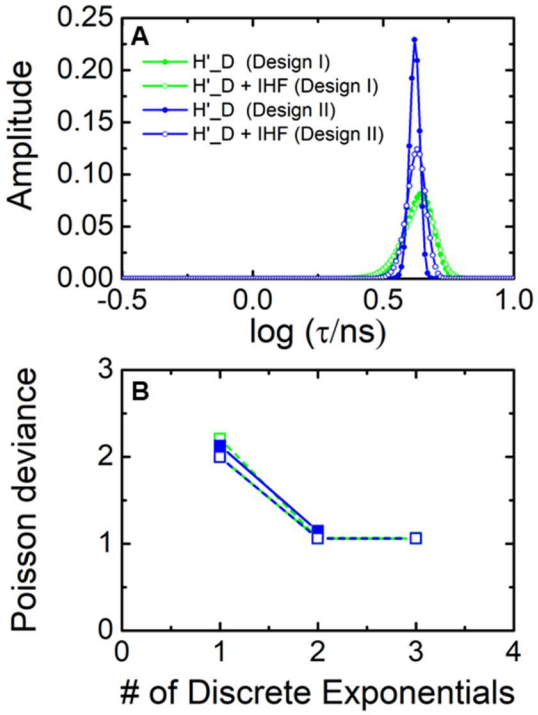
2.3.3. Controls to rule out dye labeling artifacts

DNA labels in fluorescence lifetime provided a single lifetime in the absence of protein which were insensitive to environment (Figure 2.4.). We detail here controls that were performed to rule out possible contributions to the two components observed in our lifetime distributions, and in particular, to the low-FRET state, from artifacts such as (i) incompletely labeled (or incompletely annealed) DNA or (ii) partial stacking of the fluorophores at the ends of the DNA that could result in stacked/unstacked populations with different dye orientations.

(i) Incompletely labeled DNA could result in some fraction of donor-only labeled DNA molecules that are missing an acceptor Our measurements of the percent labeling for each of the oligomers indicated >95% for the acceptor-labeled strands, which cannot account for the 20% population we observe in the low-FRET state.

(ii) To examine whether stacking/unstacking of the fluorophores could be contributing to our measurements, we also carried out lifetime measurements on a shorter, 14-bp DNA duplex, with nucleotides adjacent to the dyes identical to that in H' (Figure 2.5.). While the average FRET measured on this construct was larger than in H', as expected, we observed only a single dominant population in the MEM analysis. Indeed, a fit to a single-exponential decay was adequate to describe the decay curves, indicating that different FRET states attributable to different dye orientations are not resolvable outside the noise of these measurements.

These control experiments demonstrate that when two components are observed in the lifetime distributions, they indeed correspond to distinct protein-bound DNA conformations in our ensemble.



C

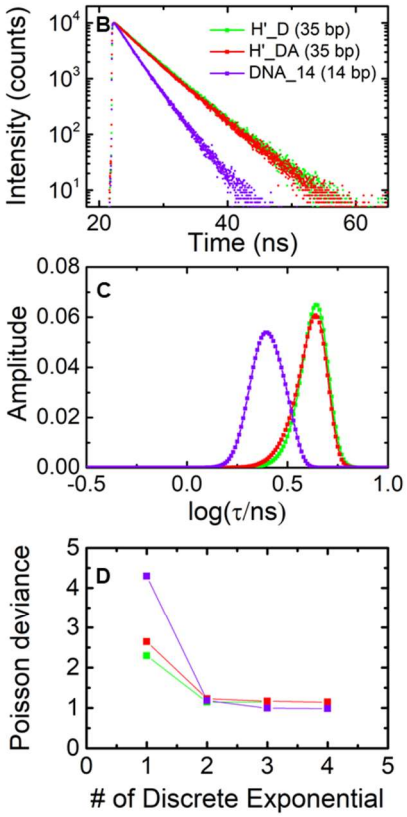
Constructs	$< \tau >$	τ_1	α_1(%)	τ_2	α_2(%)
H'_D (design I)	4.32	4.32	99.0	10.35	1.0
H'_D + IHF (design I)	4.32	4.32	98.8	9.91	1.2
H'_D (design II)	4.17	4.18	99.9	17.95	0.1
H'_D + IHF (design II)	4.22	4.23	99.8	15.35	0.2

Figure 2.4. MEM and discrete exponential analysis on donor-only labeled H' DNA constructs. (**A**) MEM distributions obtained from fluorescence lifetime decays measured on H'_D in the absence (closed) and presence (open) of IHF are shown for design I (green) and design II (blue) constructs. (**B**) The Poisson deviance obtained from a fit to a sum of discrete exponentials versus the number of exponentials are plotted for each of the decay traces. (**C**) The table summarizes the average lifetimes $< \tau_D >$ from the MEM analysis as well as the lifetimes and amplitudes obtained from a two-exponential fit to the decay traces for each of the samples.

A

H′(design I) 5′ – TGGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC
CCGGTTTTTTCGTAACGAATAGTTAAACAACGTGGT – 5′

DNA_14 5′ – TGGCCAGCATTCACC
CCGGTCGTAAGTGGT – 5′



E

Constructs	⟨τ⟩	τ_1	α_1(%)	τ_2	α_2(%)
H'_D (design I)	4.28	4.27	98.7	9.68	1.3
H'_DA (design I)	4.14	4.12	98.0	8.95	2.0
DNA_14	2.56	2.61	99.8	12.22	0.2

Figure 2.5. Fluorescence lifetime measurements on 35 bp (H') and 14 bp (DNA_14) DNA constructs in design I. (**A**) The sequences and labeling schemes for the two different DNA constructs are shown. **(B)** Fluorescence intensity decay traces are shown for donor-acceptor-labeled DNA _14_DA (purple), H'_D (green), and H'_DA (red). **(C)** Corresponding MEM distributions are shown for the decay traces of panel B. (**D**) The Poisson deviance obtained from a fit to a sum of discrete exponentials versus the number of exponentials are plotted for each of the decay traces. (**E**) The table summarizes the lifetimes and amplitudes obtained from a two-exponential fit to the decay traces, as well as the average lifetimes $< \tau_D >$ and $< \tau_{DA} >$ on donor-only and donor-acceptor-labeled constructs, respectively.

2.3.4. Computing distance and FRET from crystal structures

Distance estimates were made from lifetime distributions using PyMOL viewer. Estimates were made between two atoms on nucleotides where fluorophores were attached (see Figure 3.5.). Distances were converted to FRET estimates for Fluorescein/TAMRA FRET pair using a known Forster distance, $R_0 = 50$ Å (https://www.lifetein.com/Peptide-Synthesis-FITC-modification.html).

2.4 Laser Temperature Jump (T-Jump)

2.4.1. Setup

The home-built laser T-jump spectrometer (5, 11, 12) - schematic shown in Figure 2.6. - uses 10-ns IR laser pulses at 1550 nm, generated by Raman shifting the 1064 nm pulses from the output of an Nd:YAG laser, which are focused to ~1 mm spot size onto a 2-mm wide sample cuvette of path length 0.5 mm. Each laser pulse (~40 mJ/pulse at the sample position) yields ~5–10 °C T-jump at the center of the heated volume. A continuous wave (cw) output from a 488 nm diode laser (Newport PC13589) was used as a probe source for exciting the donor (fluorescein), and was focused to a ~100 μm spot in the middle of the heated volume (11). The fluorescence emission intensities of the donor were monitored perpendicular to the excitation direction, using a filter combination to allow only light in the wavelength range of 496-550 nm (Semrock FF01-496/LP and Thorlabs FESH0550), and measured with a Hamamatsu R928 photomultiplier tube. All T-jump measurements carried out in 100 mM KCl were digitized with a 500 MHz transient digitizer (Tektronix DP04054B). For each experiment, 512 kinetic traces were acquired and averaged by the digitizer. The time interval between subsequent laser pulses was kept at 1 s, to

allow sufficient time for the temperature of the heated volume to decay back to that of the surrounding sample holder (maintained at the initial temperature by a water bath circulator) before the arrival of the next laser pulse

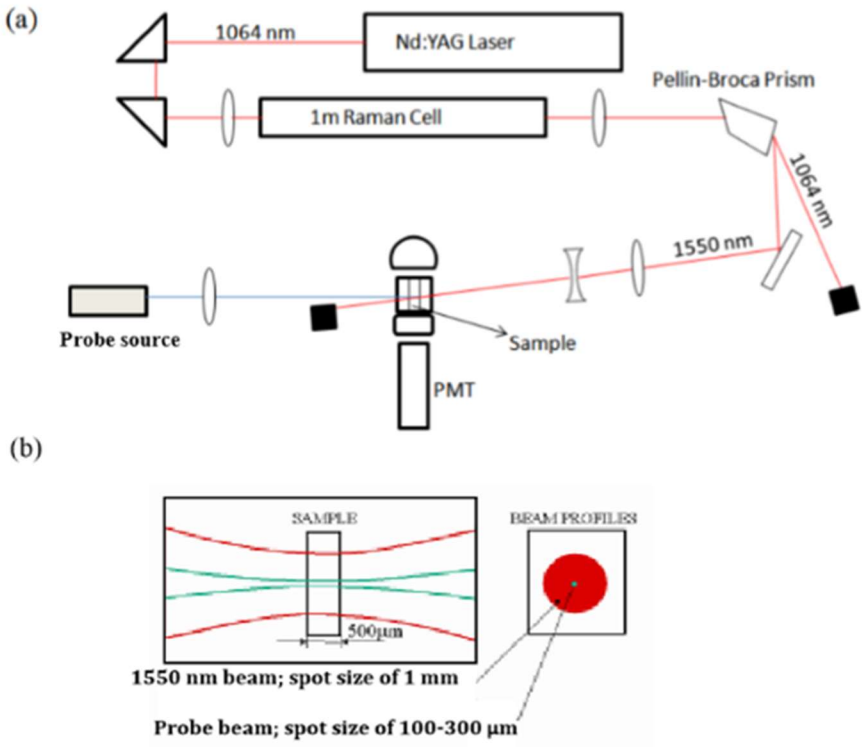


Figure 2.6. (a) Schematic of the Laser Temperature Jump (T-jump) setup. **(b)** IR beam profile (red; spot size 1 mm) and probe beam (green; spot size 100-300 µm) at the sample position. The only probe used in the thesis was a 488 nm continuous wave diode laser (Chapter 5).

2.4.2. Acquisition of T-jump kinetics traces

To acquire data with the highest temporal resolution and be able to span several orders of magnitudes in time scale, we measured T-jump kinetics traces over different time-scales and then combined these traces. We typically acquired the kinetics traces on at least two time-scales: the short time-scale covered kinetics up to 1.6 ms, with a time-resolution of 80 ns, while the longer

time-scale covered kinetics up to 16-32 ms, with a time-resolution of 2 μs. Data collected on each time scale consisted of 1 million points, which were reduced to 10,000 points by averaging 100 points together. After averaging, the time interval between data points was 160 ns and 4 μs for the short and long time-scale data, respectively.

Temporal resolution of the T-jump traces is often limited by optical or electrical artifacts. Unblocked scattered light and cavitation (bubbles that spontaneously form due to rapid changes in temperature and pressure) invade the observed signal at short timescales and prevent accurate interpretation. Additionally, electrical signals may also manifest in collected data, usually as ripples extending from ~300 us preflash to beyond the T-jump due to the firing of flashlamp in by the Nd:YAG necessary to create the IR pulse. Insulation of PMT, wiring and other electrical equipment with Faraday cages reduced the intensity of this artifact. But out of an abundance of caution, data below 20 μs is often discarded. Prior to any further analysis, data acquired below ~20 μs in each trace were discarded because of artifacts (13). Relaxation traces acquired over different timescales were combined as described below.

2.4.3. Matching of kinetics traces measured over different time scales

Relaxation kinetics traces measured over two different time scales were fitted simultaneously to double-exponential decay convoluted with the T-jump recovery kinetics (see Figure 2.7.):

$$I(t) = \big(I(0^+) - I(\infty)\big)[f_1 \exp(-k_1 t) + (1 - f_1)\exp(-k_2 t)] + \big(I(\infty) - I(0^-)\big)\left(\frac{1}{1 + t/\tau_{rec}}\right) + I(0^-) \quad \text{Eq. 2.9.}$$

Here $I(0^-)$ is the intensity prior to the *T*-jump, $I(0^+)$ is the initial intensity measured immediately after *T*-jump, $I(\infty)$ is the intensity expected at the end of the conformational relaxation, prior to the recovery of the T-jump itself (see Figure 2.8.), τ_{rec} is a characteristic time constant for the T-jump recovery, k_1 is the relaxation rate for the fast phase, with fractional amplitude f_1, and k_2 is the relaxation rate for the second (slow) phase, with fractional amplitude $f_2 = 1 - f_1$. The following parameters were varied to obtain the best fit: $I(0^+)$, $I(\infty)$, k_1, k_2, and f_1, with τ_{rec} fixed from measurements on control samples. An additional fitting parameter served as a multiplicative scale factor applied to one of the traces to account for any systematic difference in the measured intensities between the two traces (14, 15). Once appropriately scaled, the multiple data sets were combined into a single kinetic trace that covered the time range from 20 μs to 16-40 ms.

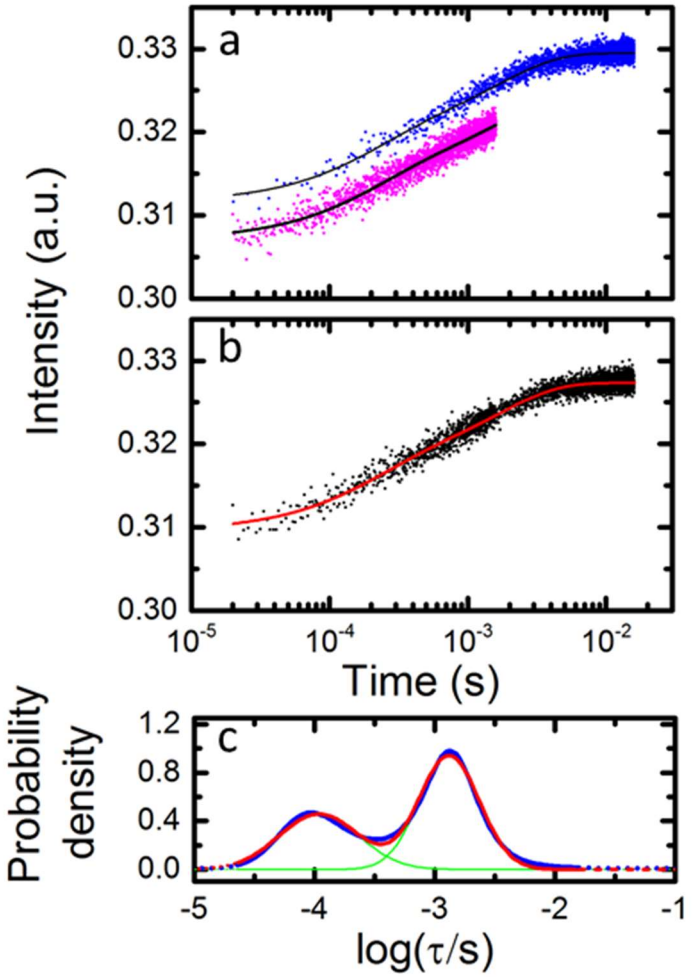


Figure 2.7. Matching kinetic traces measured on two separate time scales and maximum entropy analysis of the combined relaxation trace. (**a**) Kinetics traces taken on two separate time-scales are shown: full time-scale 1.6 ms (magenta) and full time scale 16 ms (blue). The black lines through the kinetics traces are a fit to the data using the two-exponential decay function and a scale factor to account for the difference in intensities between the two traces. (**b**) The kinetics trace obtained after matching and combining the two traces shown in (**a**), as described in 2.2.4. The continuous red line is from a fit to the data using the maximum entropy analysis. (**c**) The distribution of the log of the relaxation times, log(τ), that best describes the relaxation trace in (**b**) is shown (red). The blue line represents a fit to the distribution in terms of two Gaussians (shown in green), with independently varied peak positions and widths. The range of the distribution that falls outside the time-scales of the measured relaxation trace is shown as dashed lines.

2.4.4. Monte-Carlo search in parameter space for discrete exponential and data matching analysis

To examine whether the scaling and matching of the relaxation traces from the two separate time-scales were indeed robustly done, we carried out a Monte-Carlo search in parameter space using a simulated annealing procedure to minimize the residuals, as described previously (16-18). The parameters were randomly chosen from a wide range and the residuals were minimized by simulated annealing. Each set of parameters obtained by this procedure was then used as starting values for minimization using a least-squares non-linear fitting procedure implemented in MATLAB (R2014A 8.3.0.532). This fitting procedure was repeated with 30 independent randomly chosen sets as starting points. All 30 sets converged to the same global minimum after the non-linear least-squares minimization step. The uncertainties σ_{MC} in each of these parameters for each data set are calculated as the weighted standard deviation from the outputs at the end of the simulated annealing procedure, as follows:

$$\sigma_{MC}^2 = \sum_{i=1}^{30} \frac{(p_i - p_{best})^2}{\chi_i^2} \quad \text{Eq. 2.10.}$$

where p_i is the value of a given parameter in the i-th fit, χ_i^2 is the corresponding residual chi-square for that fit, and p_{best} is the best-fit value of that parameter obtained from the non-linear least-squares minimization. The results from the two independent data sets were then averaged together to get the combined best-fit values for each of the parameters, and the corresponding uncertainties in the parameters for the combined data sets σ_c were computed as follows:

$$\sigma_c^2 = \frac{\sigma_{MC,1}^2 + \sigma_{MC,2}^2}{4} \quad \text{Eq. 2.11.}$$

2.4.5. Estimating the size of the T-jump

The magnitude of the T-jump in each relaxation trace was determined by comparing the initial change in donor fluorescence, measured immediately after the T-jump, with equilibrium measurements of the temperature dependence of the donor quantum yield, measured on a reference sample such as donor-only labeled strand of DNA or donor-acceptor-labeled DNA in the absence of IHF, as illustrated in Figure 2.8.

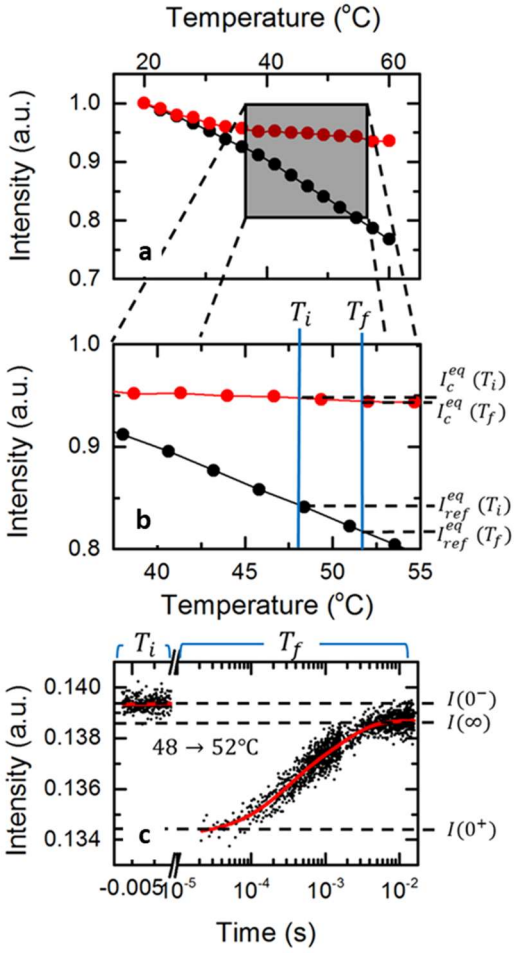


Figure 2.8. Magnitude of T-jump and amplitude of relaxation in the T-jump relaxation traces. (**a**) Equilibrium measurements of the donor fluorescence emission intensity versus temperature are shown for the donor-acceptor labeled IHF-H' complex (red) and for H' DNA only (black), at 100 mM KCl. The intensities in each panel have been normalized to match

at the lowest temperature. The donor intensity in the DNA only samples reflects the change in the donor quantum yield as a function of temperature. (**b**) An enlarged region of interest from (**a**). The vertical lines indicate the initial (T_i) and final (T_f) temperatures corresponding to the respective T-jump relaxation traces shown in (**c**). The horizontal lines indicate the corresponding donor intensities of the complex, denoted by $I_c^{eq}(T_i)$ and $I_c^{eq}(T_f)$, and of the reference (H' DNA only) sample, denoted by $I_{ref}^{eq}(T_i)$ and $I_{ref}^{eq}(T_f)$. (**c**) Representative relaxation kinetics traces in response to a T-jump perturbation for the IHF-H' complex at 100 mM KCl. The horizontal dashed lines indicate the fluorescence intensity levels $I(0^-)$, $I(0^+)$ and $I(\infty)$ as defined in Eq. 2.9. The magnitude of the T-jump in each relaxation trace is estimated by comparing the initial drop in the T-jump traces with the equilibrium temperature dependence of the donor fluorescence of the reference sample. The value of T_f such that $\frac{I(0^+)}{I(0^-)} = \frac{I_{ref}^{eq}(T_f)}{I_{ref}^{eq}(T_i)}$ is assigned as the final temperature after the T-jump for that kinetics trace. The amplitude of the relaxation trace, determined by the increase in the intensity from $I(0^+)$ to $I(\infty)$ (see Eq. 2.9.), is determined by the equilibrium temperature dependence of the donor fluorescence in the complex sample such that $\frac{I(\infty)}{I(0^-)} = \frac{I_c^{eq}(T_f)}{I_c^{eq}(T_i)}$.

We note here that in our T-jump spectrometer, the IR pulse that heats the sample is incident from only one side of our sample cuvette, which results in a nonuniform T-jump across the 0.5 mm pathlength of the cuvette. An average T-jump of 5 °C over a 0.5 mm path length cuvette is estimated to span ~6.5°C T-jump at the near edge of the sample to ~3.5°C at the far edge (19, 20). This range corresponds to ~30% error in the estimated final temperature T_f, which is approximately the size of the symbols used in our Arrhenius plots; we have therefore not included these errors in explicitly in our plots. We further note that, despite this temperature gradient across the sample, our initial T-jump studies on the IHF-H' complexes, that were done using both 0.5 mm and 1 mm pathlength cells, demonstrated good overlap between the relaxation rates measured as a function of T_f, well within the uncertainties in the measured rates from each sample cell (20). These initial studies indicated that the temperature nonuniformity, which is significantly worse in the 1 mm cells and far from ideal, did not severely impact our ability to obtain an average relaxation trace, averaged over the range of T_f, spanning the cuvette.

2.4.6. Control measurements to obtain T-jump recovery kinetics

The characteristic decay curves that best describe the recovery of the T-jump back to the initial temperature were obtained from measurements on control (donor-labeled stand of DNA or free fluorescein) samples. To obtain a complete profile of the T-jump recovery kinetics, measurements were performed over a time-window up to about 400 ms. A typical recovery profile is shown in Figure 2.11., which was fitted to the following "T-jump recovery" function:

$$I(t) = \left(I(0^+) - I(0^-)\right)\left(\frac{1}{1 + t/\tau_{rec}}\right) + I(0^-) \qquad \text{Eq. 2.12.}$$

The parameters varied to obtain the best fit to the T-jump recovery traces are $I(0^+)$ and τ_{rec}. $I(0^-)$ is determined from the average of the measured intensities prior to the arrival of the infrared heating pulse. The scatter in the recovery time constants measured on control samples over a range of initial and final temperatures is shown in Figure 2.9., with an average value of 206 ± 24 ms.

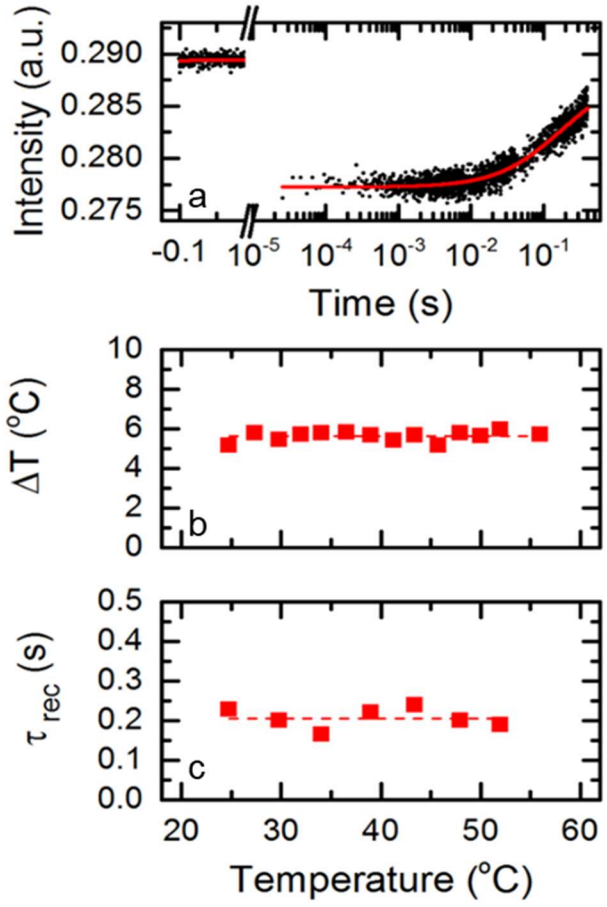


Figure 2.9. T-jump measurements on donor-only labeled ssDNA control sample. (**a**) Representative relaxation trace measured on donor-only labeled ssDNA (5 μM) after a ~4 °C T-jump is shown. The continuous line is a fit to the T-jump recovery function (Eq. 2.12.), which yields the recovery time constant τ_{rec} = 208 ms. (**b**) The magnitude of the T-jump (ΔT) is plotted versus the initial temperature for a series of control measurements on this sample. The average T-jump from this set is 5.6 ± 0.2 °C. (**c**) The recovery time constant τ_{rec} values obtained from a series of control measurements are plotted as a function of initial temperature. The average value for τ_{rec} from this set is 206 ± 24 ms.

2.4.7. Maximum Entropy analysis on T-jump relaxation traces

All relaxation traces were analyzed with the maximum entropy method (MEM), using an algorithm provided to us by Dr. Pete Steinbach of the National Institutes of Health (7, 21, 22). The MEM approach has several advantages over discrete exponential analysis to fit relaxation traces. It provides a more robust way to interpret incomplete and noisy data than discrete exponential analysis, which can result in comparable fits to the relaxation traces even with significantly different parameters (amplitudes and relaxation rates). It also provides a model independent description of the relaxation traces in terms of a distribution of relaxation times without *a priori* assumptions as to the number of discrete exponentials required.

The MEM analysis yields a distribution $f(\log \tau)$, which reflects the probability density in logarithmic scale for a given relaxation time τ, that best fits the relaxation kinetics while maximizing the entropy S, defined as:

$$S(f,F) = \sum_{j=1}^{M} \left[f_j - F_j - f_j \ln(f_j / F_j) \right] \qquad \text{Eq. 2.13.}$$

where f_j are the discretized values of the probability density distribution $f(\log \tau)$ and F is a model distribution that is the default distribution in case of noisy data and is assumed to be uniform and flat.

The MEM analyses on our data typically revealed two distinct peaks, indicating deviations from single-exponential decay (Figure 2.7.c). In cases where the two peaks were reasonably well separated, we computed an average relaxation time $\tau_{ave} = 10^{<\log\tau>}$ for each phase, where $<\log \tau>$ is computed from the distribution of relaxation times within each peak, and defined as:

$$< log\tau >= \frac{\sum_{j=1}^{M} log\tau_j \, f\left(log\tau_j\right) dlog\tau_j}{\sum_{j=1}^{M} f\left(log\tau_j\right) dlog\tau_j} \quad \text{Eq. 2.14.}$$

If there was some overlap between the peaks, the probability density distribution $f(\log \tau)$ was fitted to the sum of two Gaussian distributions, and the average $< \log\tau >$ was obtained from the peak position of each Gaussian distribution. The area under each peak reflects the fractional amplitude in that relaxation phase.

2.4.8. **Arrhenius fits to relaxation rates versus temperature**

Two independent sets of measurements were carried out for each sample, over a range of initial and final temperatures. Before carrying out the Arrhenius fits on the ln(rate) versus inverse temperature plots, the relaxation rates (and amplitudes) obtained for a given final temperature (T_f), from each set, were averaged together. Although the initial temperatures (T_i) can be controlled to be identical (or close) for the two sets of measurements, the final temperatures obtained from both sets need not overlap, in part because the magnitude of the T-jump depends on the spectrometer alignment for that day, and also fluctuates slightly within the day. Therefore, prior to averaging, the rates and amplitudes for each set of measurements were interpolated on to a common grid for T_f, using a linear interpolation on an Arrhenius plot of ln(rates) versus $1/T_f$. The data from the two sets were averaged together on the same Arrhenius scale. The errors in the averaged quantities (rates and amplitudes) are the standard deviations from the two sets.

The averaged ln(rates) versus versus $1/T_f$ thus generated were fitted to the following Arrhenius equation:

$$S(f,F)=\sum_{j=1}^{M}\left[f_j-F_j-f_j\ln(f_j/F_j)\right] \quad \text{Eq. 2.15.}$$

where E_a is the activation enthalpy, and R is the universal gas constant. The errors in the activation energy values reported in the text are standard deviations of the E_a values obtained from each set independently.

2.4.9. Control measurements to rule out contributions to the relaxation kinetics from dye dynamics

A series of control experiments were performed to rule out any contributions from dye interactions and/or dye dynamics to the observed relaxation kinetics. (1) For any FRET-labelled sample, corresponding measurements were done on donor-only labelled samples to ensure that the observed kinetics were from FRET changes and not from interactions of the donor (fluorescein) dye with either the DNA or the protein. No kinetics were observed on any of the donor-only samples other than the slow T-jump recovery kinetics (Figure 2.10.a), similar to those observed on control measurements done using either a fluorescein-labelled single strand of the H' DNA or with free fluorescein dye (see control experiments shown in all T-jump kinetics figures). These measurements also confirmed the absence of any photo bleaching of the fluorescein dye in the window of the T-jump measurements. (2) To rule out any contribution to the observed kinetics from interactions of the acceptor (TAMRA) dye, T-jump experiments were carried out on FRET-labelled IHF-H' complex by directly exciting TAMRA (at 532 nm, where fluorescein does not absorb) and measuring the temporal response of the fluorescence emission of TAMRA. Again, no relaxation kinetics were observed (Figure 2.10.b), thus confirming that interactions of the acceptor

dye with the DNA or the protein do not interfere with the observed relaxation kinetics. (3) To examine whether there were any T-jump induced changes in the relative orientation of the dyes attached to the DNA that were not related to an overall change in the DNA conformation, we carried out T-jump measurements on FRET-labelled DNA only samples. For these measurements we designed a shorter 14 bp DNA oligomer labelled with fluorescein and TAMRA, with the sequence context of each of the dyes identical to that in the longer H' sequence (Figure 2.10.d). Note that in this 14 bp sequence, the four nucleotides at the ends, next to the dyes, are the same as in the original H' sequence. The reason for this shorter sequence is that the end-to-end distance in the 35-bp H' sequence is too long for there to be any significant FRET between the dyes without the bound protein. By design, the FRET value in the 14-bp DNA is ~0.5, close to the FRET of H' when bound to IHF, and in a region where small changes in the distance or relative orientation of the dyes would result in a detectable FRET change. Once again, T-jump measurements on this DNA sample did not reveal kinetics other than the T-jump recovery kinetics (Figure 2.10.c). Thus, we see no evidence of dye reorientation dynamics in DNA only samples contributing to the observed relaxation kinetics (23).

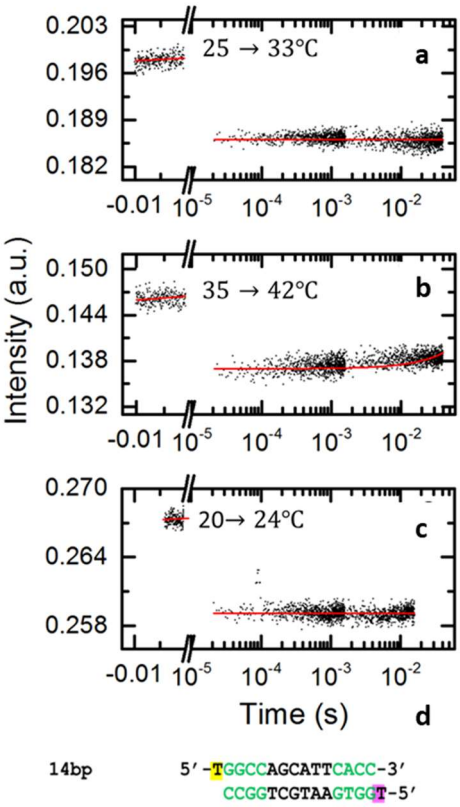


Figure 2.10. Control T-jump experiments to rule dye dynamics as contributing to relaxation kinetics. (**a**) Donor fluorescence emission intensities of donor-only labeled H' samples in the presence of IHF, with excitation of donor at 488 nm, measured in response to a T-jump perturbation, are plotted as a function of time. (**b**) Acceptor fluorescence emission intensities of donor-acceptor labeled H' in complex with IHF, with direct excitation of acceptor at 536 nm, measured in response to a T-jump perturbation. (**c**) Donor fluorescence emission intensities of donor-acceptor labeled 14-mer DNA oligonucleotide, with excitation of donor at 488 nm, measured in response to a T-jump perturbation. (**d**) Sequence of the 14-bp construct; nucleotides indicated in green are identical to the nucleotides next to the dyes in the longer H' sequence. The FRET E measured for this DNA construct was 0.48 ± 0.05, similar to the FRET E of 0.57 ± 0.02 in the IHF-H' complex.

2.5. Systemic enrichment of ligand by exponential enrichment (SELEX)

2.5.1. Oligos

DNA oligos used in SELEX studies were ordered from IDT (with HPLC purification) and are listed in Table 6.1. Labeled oligos included a Cy3-fluorophores attached to thymidine overhangs at the 5'-end, through six-carbon phosphoradimite linkers. DNA concentrations were initially determined by bulk absorbance measurements at 260nm (Shimadzu UV-2700).

dsDNA was synthesized from the ssDNA by PCR. ssDNA was mixed with reverse primer (rPCR) at a 1:10 ratio in EconoTaq Plus 2x master mix (Lucigen, Cat. No. 30035-1). The PCR reaction ran as follows: (1) 94^{o}C for 2 min 30 s, (2) 55^{o}C for 2 min, (3) 72^{o}C for 30 min. Library was purified using QIAquick PCR purification kit (QIAGEN, Cat. No. 28104). Final concentrations were verified by absorption at 260 nm (Nanodrop ND-2000).

2.5.2. Electrophoretic Mobility Shift Assays (EMSA)

PAGE gels were prepared each experiment day (6% 19:1 acrylamide:bis-acrylamide, 0.5x TBE buffer, 3% glycerol, 0.08% Ammonium persulfate, and 0.08% TEMED). DNA and protein were mixed in 1x binding buffer, 10% glycerol, 50 ng/uL poly dI-dC, 50 ug/ml BSA, and 0.025% NP-40, and allowed to incubate for 30 minutes before loading into gels. Gels were run on ice at 170 V with 1x TAE buffer for 2-4 hours. Gels were imaged by fluorescence detection using Cy3 settings with Azure c400 Gel Imager, property of Aldrich lab at University of Illinois at Chicago.

2.5.3. SELEX protocol

SELEX protocol closely resembles the solution-based cognate sequence identification described by Aseem Ansari and Coworkers (24). 25 nM of biotin-labelled dsDNA library was

mixed with either IHF or Hbb at concentrations 250 nM-2μM. EMSA assays were performed as described above. After EMSA, Gel fragments with DNA were excised, mixed with elution buffer (1x TE buffer), and allowed to sit overnight. After elution process, liquid containing of TE buffer and DNA library, with or without protein, was separated from gel fragments. Library was pulled down using Dynabeads M-270 (Invitrogen, Cat. No. 112-05D) according to manufacturers protocol. Protein was removed from bead-bound library by washing three times using 1x PBS + 0.1% BSA solution (100-fold excess buffer). Beads were then transferred to PCR strip tubes with 20 μM fPCR and rPCR primers. DNA was amplified using the following thermal cycle: a) 94 °C for 2 min, b) 94 °C for 30 s, c) 55 °C for 30 s, d) 72 °C for 30 s, e) repeat steps b-e 19 times for a total of 20 cycles. Amplified DNA was purified using QIAquick PCR Purification Kit (QIAGEN, Cat. No. 28104). Purified DNA concentrations were measured. All DNA was saved for subsequent rounds and/or sent for sequencing. Sequencing was performed at the DNA Services facility within the Research Resources Center at the University of Illinois at Chicago.

2.5.4. Evaluating shape parameters of SELEX sequence with DNAshape

All sequences of interest were analyzed in collaboration with Dr. Devesh Bhimsaria (Bio Informaticals), a former postdoc of Prof. Aseem Ansari (St. Jude Children's Research Hospital) using the DNAshape developed by Dr. Remo Rohs from University of Southern California (25). DNAshape provides DNA shape feature profiles for all sequences input into the program. DNAshape derives sequence information from the motif databases JASPAR (26) and UniPROBE (27) and generates DNA shape data for transcription factor binding sites by high-throughput prediction of DNA structural features (25). The approach uses a sliding five-base pair window and collects tables of structural features based on all-atom Monte Carlo simulations for all 512 unique

pentamers. This method has been validated by comparing model predictions with data collected from from X-ray crystallography, NMR spectroscopy and hydroxyl radical cleavage studies, as well as statistical analysis and cross-validation (25). Shape features evaluated here were minor groove width, propeller twist, buckle, electrostatic potential, opening, shear stagger, stretch, helical twist, rise, roll, shift, slide, and tilt. Sequence enrichment scores from SELEX libraries were reassigned on a normalized scale from 0 (indicating lowest sequence population) to 1 (indicating largest sequence population) for presentation purposes. Normalized scores were calculated by dividing each sequence score by the maximum sequence score from the data set. Thus, sequence features important to sequence SELEX performance can be readily identified.

2.5.5. Sequence-dependent DNA elastic energy calculations

Elasticity calculations of DNA were performed by Robert T. Young in the lab of prof. Wilma Olson lab at Rutgers University – New Brunswick. Calculation methods described in ref. (28). We summarize the model here.

To model sequence-dependent elasticity, DNA is treated as a collection of N rigid-body base pairs. For the j-th base pair, where j = 1, 2 ,…, N, the base-step parameters were determined relative to its nearest-neighbor base pair, j+1. (29). These parameters describe the relative orientation between nearest-neighbor base pairs and are written for the i-th step, $p_i = (\theta_i^1, \theta_i^2, \theta_i^3, \rho_i^1, \rho_i^2, \rho_i^3)$ (30) where i = 1, 2,.., N-1. The six parameters are nominally referred to, respectively, as tilt, roll, twist, shift, slide and rise. The first three parameters (θ) correspond to three angles describing the relative orientation of base-pair frames and the last three parameters (ρ) describe orientation to the next step frame.

Elastic deformation energy of a sequence is defined as the sum of the energy of elastic deformation for each step. For the i-th step, this energy is given by the following quadratic form:

$$E_i = \frac{1}{2}(p_i - p_i^{rest})^T F_i (p_i - p_i^{rest}) \qquad \text{Eq. 2.16.}$$

Where p_i^{rest} contains the intrinsic step parameters and describes the reference configuration of the step (i.e., the configuration of zero energy, also called the rest state), and F_i is a 6 × 6 matrix describing nearest-neighbor intrinsic elasticity characterized from a survey of structures in the Protein Data Bank (31). Both the intrinsic step parameters and the force constant matrix can differ for each step depending on the DNA sequence. The elastic energy (ε) for the base-pair collection is then given by:

$$\varepsilon = \sum_{i=1}^{N-1} E_i \qquad \text{Eq. 2.17.}$$

Initial elastic energies are computed via a so-called "threading" approach, where each base-step parameters, p_i, are assigned using base pair orientations of the crystal structure using from 3DNA software (26) and energies are calculated according to Eq. 2.16. and Eq. 2.17.

2.5.6. Elastic energy minimization

Energy minimization were also performed by Robert T. Young in the lab of Prof. Wilma Olson (Rutgers University – New Brunswick), using the procedure described in ref. (28). In this approach, the elastic energy of a base-pair collection given by Eq. 2.17 is minimized under given constraints (i.e. fixed positions of first, last, or kink site base pair). It should be noted that if the *i*-the site is constrained during optimization, its energy contribution E_i is omitted from the total elastic energy. Starting positions for each site p_i were assigned by threading to each known crystal

complexes (2, 32). Minimum elastic energy is when the gradient of elastic energy, $\frac{\partial E_i}{\partial p_i}$, for a single step is 0.

References:

1. Vivas P, Velmurugu Y, Kuznetsov SV, Rice PA, Ansari A. Mapping the transition state for DNA bending by IHF. J Mol Biol. 2012;418(5):300-15.
2. Mouw KM, Rice PA. Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending Hbb. Molecular Microbiology. 2007;63:1319-30.
3. Sugimura S. Kinetic and steady-state studies of binding and bending of lambda phage DNA by Integration Host Factor [Ph.D. thesis]. New Haven, CT: Yale University; 2005.
4. Sugimura S, Crothers DM. Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. Proc Natl Acad Sci USA. 2006;103(49):18510-4.
5. Vivas P, Kuznetsov SV, Ansari A. New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. J Phys Chem B. 2008;112(19):5997-6007.
6. O'Connor DV. Time-correlated single photon counting. London: Academic Press; 1984.
7. Steinbach PJ, Ionescu R, Matthews CR. Analysis of kinetics using a hybrid maximum-entropy/nonlinear-least-squares method: application to protein folding. Biophys J. 2002;82:2244-55.
8. Steinbach PJ. Filtering artifacts from lifetime distributions when maximizing entropy using a bootstrapped model. Anal Biochem. 2012;427(1):102-5.
9. Esposito R, Mensitieri G, de Nicola S. Improved maximum entropy method for the analysis of fluorescence spectroscopy data: evaluating zero-time shift and assessing its effect on the determination of fluorescence lifetimes. Analyst. 2015;140(24):8138-47.
10. Chakraborty S, Steinbach PJ, Paul D, Mu H, Broyde S, Min JH, et al. Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. Nucleic Acids Res. 2018;46(3):1240-55.
11. Kuznetsov SV, Kozlov AG, Lohman TM, Ansari A. Microsecond Dynamics of Protein-DNA Interactions: Direct Observation of the Wrapping/Unwrapping Kinetics of Single-Stranded DNA around the *E. coli* SSB Tetramer. J Mol Biol. 2006;359:55-65.
12. Kuznetsov SV, Sugimura S, Vivas P, Crothers DM, Ansari A. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc Natl Acad Sci USA. 2006;103(49):18515-20.
13. Wray WO, Aida T, Dyer RB. Photoacoustic cavitation and heat transfer effects in the laser-induced temperature jump in water. Appl Phys B. 2002;74:57-66.
14. Velmurugu Y. Dynamics and mechanism of DNA-bending proteins in binding-site recognition. Chicago: University of Illinois at Chicago; 2016.
15. Velmurugu Y, Chen X, Slogoff Sevilla P, Min JH, Ansari A. Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. Proc Natl Acad Sci U S A. 2016;113(16):E2296-305.
16. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21:1087-92.
17. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA. Conformational relaxation and ligand binding in myoglobin. Biochemistry. 1994;33(17):5128-45.
18. Kuznetsov SV, Shen Y, Benight AS, Ansari A. A semiflexible polymer model applied to loop formation in DNA hairpins. Biophys J. 2001;81:2864-75.
19. Kubelka J. Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics. Photochem Photobiol Sci. 2009;8(4):499-512.
20. Vivas P. Mechanism of integration host factor, a DNA-bending protein, probed with laser temperature-jump. Chicago: University of Illinois at Chicago; 2009.

21. Livesey AK, Brochon JC. Analyzing the Distribution of Decay Constants in Pulse-Fluorimetry Using the Maximum Entropy Method. Biophys J. 1987;52(5):693-706.
22. Steinbach PJ. Inferring lifetime distributions from kinetics by maximizing entropy using a bootstrapped model. J Chem Inf Comput Sci. 2002;42(6):1476-8.
23. Kuznetsov SV, Sugimura S, Vivas P, Crothers DM, Ansari A. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc Natl Acad Sci U S A. 2006;103(49):18515-20.
24. Tietjen JR, Donato LJ, Bhimisaria D, Ansari AZ. Sequence-specificity and energy landscapes of DNA-binding molecules. Methods Enzymol. 2011;497:3-30.
25. Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, et al. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic Acids Res. 2013;41(Web Server issue):W56-62.
26. Zheng G, Colasanti AV, Lu XJ, Olson WK. 3DNALandscapes: a database for exploring the conformational features of DNA. Nucleic Acids Res. 2010;38(Database issue):D267-74.
27. Robasky K, Bulyk ML. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. 2011;39(Database issue):D124-8.
28. Clauvelin N, Olson WK. The synergy between protein positioning and DNA elasticity: energy minimization of protein-decorated DNA minicircles. arXiv. 2014.
29. Dickerson RE. Definitions and Nomenclature of Nucleic Acid Structure Parameters. Journal of Biomolecular Structure and Dynamics. 1989;6(4):627-34.
30. Coleman BD, Olson WK, Swigon D. Theory of sequence-dependent DNA elasticity. The Journal of Chemical Physics. 2003;118(15):7127-40.
31. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A. 1998;95(19):11163-8.
32. Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell. 1996;87(7):1295-306.

Chapter 3.

Static Kinks or Flexible Hinges: Multiple Conformations of Bent DNA Bound to Integration Host Factor Revealed by Fluorescence Lifetime Measurements

This chapter adapted from [Connolly, M; Arra, A.; Zvoda, Z.; Steinbach, P.J.; Rice, P.A.; and Ansari, A., Static kinks or flexible hinges: multiple conformations of bent DNA bound to Integration Host Factor revealed by fluorescence lifetime measurements, *J. Phys. Chem. B* 2018, 122, 49, 11519-11534] with permission from American Chemical Society.

3.1. Introduction

Severe distortions to the B-DNA structure are ubiquitous in the cell and are induced not only by nonspecific DNA packaging proteins but also by site-specific proteins when they bind to their target sites on DNA for gene regulation, replication and repair. These distortions often appear as DNA bends or localized kinks that range from ~30° kinks per helical turn when wrapped in the nucleosome,[1-4] to more severe kinks of ~80-90° in several site-specific DNA bending proteins,[5-8] to what may be the most severely bent DNA conformation found in complex with the IHF/HU family of bacterial type II DNA-bending proteins.[9-11] While structural studies on many DNA-bending protein complexes have revealed the nature of these DNA distortions at atomic resolution, the behavior of the distorted, sharply kinked, DNA in solution needs further study. One question that has yet to be fully resolved is whether these distorted DNA structures behave as rigid kinks in solution or as flexible hinges.[12] Precise measurements of the accessible protein-DNA conformations in solution are needed to elucidate the functional roles of these dynamic fluctuations as well as to provide experimental checkpoints in the development and improvement of computational models of sequence-dependent DNA deformability and stabilization of bent DNA by proteins.[13-22]

Several lines of evidence suggest that DNA in nonspecific complexes indeed adopts a range of bent conformations. Evidence for conformational heterogeneity is well documented in the case of the nonspecific architectural histone-like nucleoprotein HU that is known to bend DNA into a U-shape.[10, 23] Indirect evidence for multiple bent states in HU-DNA complexes came from multiple bands revealed in gel shift assays,[24] and from different degrees of bending observed in different crystal structures of these complexes.[10] This heterogeneity in HU-DNA complexes was directly confirmed from atomic force microscopy (AFM) studies that showed a broad distribution of bent conformations with bend angles observed in the entire range from 0 to 180°.[25] It is perhaps not too surprising that a nonspecific architectural DNA bending protein, especially one that has a variety of biological roles, would exhibit such broad conformational heterogeneity, given the lack of specific contacts that would otherwise compensate for the energetic penalty required to severely deform DNA.

However, are specific complexes with DNA-bending proteins also conformationally heterogeneous? AFM studies on some specific complexes have revealed relatively broad distributions of bend angles, although the resolution of these AFM studies primarily revealed single-peaked distributions of varying widths, consistent with thermal fluctuations within a single free energy well in the conformational landscape of the complex.[26-28] Notable exceptions in which two or more distinct bent states of DNA are observed include mismatch repair protein MutS[29] and DNA glycosylase hTDG[30] bound to mismatched DNA. Interestingly, these are examples of complexes with DNA damage recognition proteins in which the specific site is typically a single mismatch or a single chemically modified base, and all other interactions of the protein with the flanking DNA sequences are nonspecific. For site-specific proteins that bind to canonical,

undamaged DNA, and for which the target site extends beyond a few base pairs, data are scarce regarding conformational heterogeneity in such complexes.

Here, we present results that quantify the heterogeneity of bent DNA conformations when bound to the eukaryotic DNA-bending protein, the *Escherichia coli* integration host factor (IHF). We utilized picosecond-resolved fluorescence lifetime measurements to identify multiple FRET states, which reflect the distribution of bent conformations in the IHF-DNA complex. The lifetime approach has a few advantages over single-molecule assays of FRET distributions.[31] Measurements are done under solution conditions and thus avoid many of the issues that may come with immobilization of protein-DNA complexes as needed for AFM studies[32-33] or single-molecule FRET (smFRET) studies,[34-37] although smFRET on freely diffusing molecules do alleviate immobilization issues, as showcased in a series of elegant papers by Eaton and co-workers[38-41] on the dynamics and distributions during protein folding. More important, fluorescence decay curves measured in bulk solution provide significantly better resolution of the shape of the FRET efficiency distributions than can be obtained from single-molecule studies where the number of molecules sampled are in the range of $O(10^2)$ to $O(10^4)$. Moreover, the time resolution of the lifetime studies enables snapshots of the conformational distribution with an effective "shutter speed" that is typically less than tens of nanoseconds (within the excited state lifetime of the fluorophore). In smFRET, the observation time window is few milliseconds or longer, which can obscure conformational heterogeneity if dynamic fluctuations between different conformations are faster than a few milliseconds.[42] Recently, we demonstrated the effectiveness of fluorescence lifetime based FRET measurements in unveiling the conformational heterogeneity in mismatched DNA bound to nucleotide excision repair protein XPC/Rad4, with evidence for two or more distinct DNA conformations in the complex.[43]

Here we focus on IHF, which is a small (22 kDa) bacterial heterodimeric nucleoid-associated protein that is closely related to HU and structurally very similar.[23, 44] Unlike HU, however, IHF binds DNA both nonspecifically, in its role as a DNA compaction protein, and specifically, when required for site-specific recombination, DNA replication and transcription. IHF also recognizes several sites on bacteriophage lambda DNA in its role as a host factor for lysogeny by phage lambda.[45-46] The footprint of DNA bound to IHF exceeds 25 base pairs (bp).[47] However, the consensus sequence identified on the basis of a large number of cognate sites for IHF consists of two short elements in one half of the footprinted region (Figure 3.1.): WATCARnnnnTTR, where W denotes A or T, R denotes purine, and n refers to any base.[47-48] Some IHF sites also contain an A-tract containing 4-6 adenines and located in the other half of the binding site.[49-50]

The crystal structure of IHF bound to one such cognate site, denoted as the H' site (Figure 3.1. and Table 3.1.), showed that IHF sharply kinks the DNA at two sites spaced ~9 bp apart, explaining how it can bring distal regions of DNA together to facilitate the formation of higher-order nucleoprotein complexes.[9] The kinks in the DNA are stabilized by conserved proline residues located on two β-ribbon arms of the protein. The arms are thought to be flexible in the absence of the DNA but wrap around the DNA in the complex and make additional stabilizing contacts in the consensus region between the kink sites (Figure 3.1.). Further stabilization needed to overcome the large energy penalty for the sharply kinked DNA comes from additional contacts between the flanking DNA segments and the core of the protein dimer, as well as from an extensive network of electrostatic interactions with charged residues of the protein and the phosphates in the DNA backbone.[51-52]

All contacts that IHF makes with the DNA in the specific complex are either in the minor groove, where the hydrogen bonding patterns offered by the different bases are very similar,[53] or to the sugar phosphate backbone. Thus, IHF is a remarkable example of a DNA-bending protein that exhibits high specificity to certain sequences but relies almost exclusively on "indirect readout" to recognize its target sequence; i.e. it recognizes sequence-dependent DNA shape, local geometry, and DNA deformability without the need for direct recognition of the specific bases by the protein residues.[54] Several features within the binding site play a role in facilitating this recognition. One may be the ease with which the two sites can be kinked: although only one kink occurs within the consensus sequence, introducing single-T insertions or mismatches to make the DNA more flexible at these sites was shown to increase the binding affinity for IHF as well as for the nonspecific HU.[55-56] Another important feature is the flanking DNA sequences that interact with the sides of the protein, where the minor groove of the DNA is clamped between the N-termini of two alpha helices. On the right side (as shown in Figure 3.1. and Table 3.1.), the TTR of the consensus sequence allows over-twisting of the DNA that facilitates its fit into the protein clamp and the formation of salt bridges with the side chains of IHF.[57] On the left side of H' (but not all IHF binding sites) is an A-tract. The structure shows that the unusually narrow groove naturally adopted by A-tract DNA fits well into the protein clamp, and this may explain why crystals could not be grown with IHF binding sites that did not include the A-tract.[9] This study seeks to elucidate how each of these features help keep the DNA bent and clamped against the protein and to what extent the DNA resists these bending deformations.

The fluorescence lifetime-based FRET measurements presented here demonstrate that IHF bound to a 35-bp substrate containing the cognate H' site, the sequence used in the crystal structure, in fact samples at least two distinct states: a high-FRET state that appears to be fully bent, as in

the crystal structure, and a low-FRET state consistent with partially bent DNA that could include an ensemble of conformations with one side unclamped or the other. Another cognate site that is lacking the A-tract (the H1 site on phage lambda DNA) showed a significantly smaller population in the fully bent state, highlighting the effectiveness of the A-tract to keep that side of the DNA clamped down. The population in the fully bent state in the IHF-H1 complex is partially recovered when mismatches are introduced at the kink site near the missing A-tract, to reduce the energetic cost of kinking the DNA at that site. Finally, a single base modification in the TTR consensus region in the other flanking arm, previously known to destabilize the complex, also resulted in a significantly diminished population in the fully bent state. Taken together, these results provide additional insights into the finely tuned interactions between IHF and its cognate sites that keep the DNA bent (or not) and yield quantitative data on the populations of different DNA conformations (kinked or not kinked) that are sensitively dependent on the DNA sequence and deformability. These results also correlate well with the biological function of the two IHF-binding sites studied. These two sites are found within the large protein-DNA complex responsible for integrating phage lambda DNA into the *E. coli* chromosome, and structural modeling shows that the IHF-induced bend at the H1 site, but not at the bend H' site, must flex during assembly.[58]

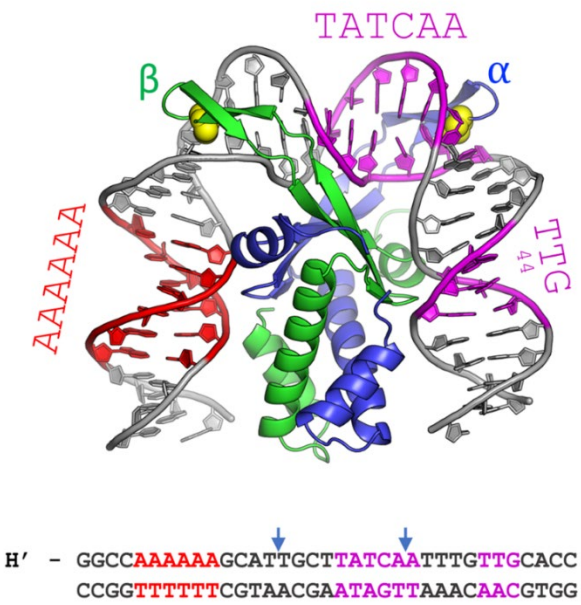


Figure 3.1. The cocrystal structure of IHF in complex with the H' binding site from phage lambda. The α- and β-chains of the IHF protein are shown in blue and green, respectively, with the conserved proline residues shown as yellow spheres. The DNA is shown in gray, with the consensus region highlighted in magenta and the A-tract in red. The sequence shown below is that of the 35-mer that contains the H' binding site, which was used to obtain the cocrystal structure with IHF. In the complex, the DNA is sharply kinked at the two sites indicated by the blue arrows. In the DNA oligomer used for the structural studies, the DNA was nicked at a position shifted 1 bp to the 3'-side of the left blue arrow, to facilitate crystal packing. That nick was "sealed" *in silico* to generate the model shown (starting from PDB ID 1IHF).

Table 3.1.: DNA Constructs Used in This Study

Construct	Sequence[a]	$< FRET\ E >$ at 100 mM KCL	K_d at 100 mM KCL	K_d at 300 mM KCL[b]
H' (design I)	5′ – TGGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC CCGGTTTTTTCGTAACGAATAGTTAAACAACGTGGT – 5′	0.50 ± 0.03	23^{+57}_{-14} pM[b] 27 ± 2 pM[c]; 25 pM[d]	1.9 ± 0.1
H' (design II)	5′ – GGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC PCCGGTTTTTTCGTAACGAATAGTTAAACAACGTGG – 5′	0.41 ± 0.01	----------	1.5 ± 0.2
H1 (design I)	5′ – TGGCCATGCAGTCACTATGAATCAACTACTTAGACC CCGGTACGTCAGTGATACTTAGTTGATGAATCTGGT – 5′	0.25 ± 0.01	20 pM[e]	----------
H1 (design II)	5′ – GGCCATGCAGTCACTATGAATCAACTACTTAGACC PCCGGTACGTCAGTGATACTTAGTTGATGAATCTGG – 5′	0.21± 0.01	----------	2.0 ± 0.1
H'_nAt (design I)	5′ – TGGCCATGCAGGCATTGCTTATCAATTTGTTGCACC CCGGTACGTCCGTAACGAATAGTTAAACAACGTGGT – 5′	0.23 ± 0.01		3.7 ± 0.8
H1_CTloop (design I)	5′ – TGGCCATGCAGTCACTATGAATCAACTACTTAGACC CCGGTACGTCAGTCTTACTTAGTTGATGAATCTGGT – 5′	0.36 ± 0.02	----------	0.06 ± 0.02
H'44A (design I)	5′ – TGGCCAAAAAAGCATTGCTTATCAATTTGTAGCACC CCGGTTTTTTCGTAACGAATAGTTAAACATCGTGGT – 5′	0.23 ± 0.01	7.4 ± 2.8 nM[c]	----------

[a] The blue arrows indicate the position of the kinks in complex with IHF; the yellow- and green-highlighted thymines represent fluorescein- and TAMRA-attachment, respectively, to that base; the blue-highlighted P indicates attachment of Atto550 to the phosphate group; A-tract is shown in red and the consensus sequence in magenta; the position of the mutation in H'44A is indicated in purple; the position of 4-nt loop is H1_CTloop is indicated in brown.

[b-e] K_d values are from [b]this study, [c]ref.[59], [d-e]stopped-flow measurements reported in [d]ref.[60] and [e]ref.[61]

3.2. Results

3.2.1. Fluorescence lifetime measurements provide a snapshot of the distribution of bent conformations in IHF-DNA complexes

In this study, we characterized the equilibrium distribution of DNA conformations when bound to IHF, by measuring the FRET efficiency between a donor-acceptor FRET pair attached to a 35-bp DNA substrate of varying sequences and binding affinities, using fluorescence lifetime studies. The FRET efficiency (*E*) between the donor and acceptor labels depends strongly on the relative distance and orientation between the labels and can be computed from the lifetimes of the excited donor fluorophore in the presence (τ_{DA}) and absence (τ_D) of the acceptor, as $= 1 - \frac{\tau_{DA}}{\tau_D}$. The measured FRET efficiencies can sense changes in the conformations of the DNA to which the labels are attached, for example upon binding of a DNA-bending protein. In the case of IHF-DNA complexes, FRET between labels attached at the ends of 35-mer DNA substrates has proven to be exceptionally useful in measuring not only equilibrium changes from unbent (straight) to bent conformations upon binding of IHF,[62] but also the dynamics of these conformational changes along the recognition trajectory of the IHF binding site.[59, 63-65]

Here, we take advantage of the sub-nanosecond time-resolution of the fluorescence lifetime studies that enables fluorescence decays to be measured over a wide temporal range and with high temporal resolution. A single DNA conformation in the ensemble of molecules, with a well-defined separation and relative orientation between the donor and acceptor of the FRET pair, is characterized by a unique FRET efficiency, and is expected to yield one (faster) exponential decay for the donor fluorescence in the presence of the acceptor for every exponential decay observed in the absence of the acceptor. A distribution of DNA conformations corresponding to different

donor-acceptor distance/orientations would then correspond to a distribution of FRET efficiencies, which should be reflected in a distribution of lifetimes measured in the decay traces. The lifetime distributions were obtained from the measured decay traces by the maximum entropy method (MEM).

The fluorescence lifetime of the donor (fluorescein), measured in DNA constructs in the absence of any acceptor (DNA_D), exhibited close to single-exponential decays with a relatively narrow distribution of lifetimes, and characterized by an effectively unique lifetime $< \tau_D > \approx 4.2 - 4.3$ ns that was found to be mostly insensitive to the position of attachment to the DNA or to the presence of the protein (Figure 2.5.). In the presence of an acceptor fluorophore (DNA_DA), the decay profiles remained largely single-exponential with no protein bound (Figure 3.2.), albeit with a shift in the donor lifetime $< \tau_{DA} >$ relative to $< \tau_D >$ if there was significant FRET (Figure 2.6.). In contrast, for lifetime measurements on DNA_DA in the presence of IHF, not only did the donor lifetime shorten due to FRET, we also detected a broadening of the distribution of lifetimes and appearance of at least two distinct components (Figure 3.2. and Figure 2.4.), which points to direct evidence for multiple DNA conformations in the IHF-bound complexes in the ensemble.

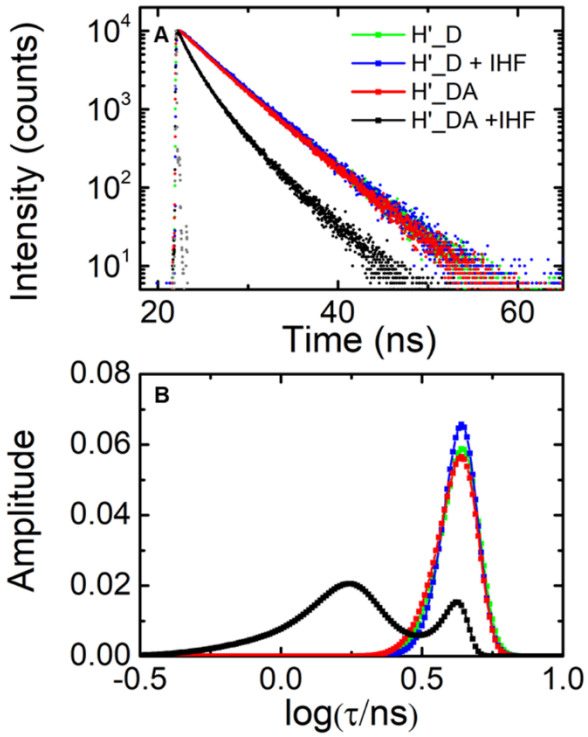


Figure 3.2. Fluorescence lifetime measurements on IHF-H' complex in design I constructs. **(A)** Fluorescence intensity decay traces measured on H' labeled with fluorescein and TAMRA (H'_DA) are shown in the absence (red) and presence (black) of IHF. The corresponding donor-only (H'_D) decay traces are also shown in the absence (green) and presence (blue) of IHF. Measurements were done with 5 μM DNA and 5 μM IHF. The instrument response function (gray) is shown for comparison. **(B)** The MEM lifetime distributions obtained from the fluorescence decay traces are shown. The amplitudes of the distributions are normalized to add up to one.

3.2.2. IHF-H' complex reveals two dominant DNA conformations: fully-bent (high-FRET) and partially-bent/straight (low-FRET)

We first present lifetime measurements carried out on the H' DNA substrate, end-labeled with fluorescein and TAMRA (H' in design I; Table 3.1.), in the presence and absence of IHF. The design of this construct is identical to what we and others have used in previous equilibrium and kinetics studies on IHF-DNA complexes.[60, 63] All lifetime measurements were done with 5 μM DNA and 5 μM IHF. At 100 mM KCl, the average FRET efficiency in the H' DNA construct alone

was found to be 0.022 ± 0.001, which increased to 0.50 ± 0.03 upon binding of IHF under 1:1 binding conditions. These FRET efficiency values are consistent with steady-state values reported previously for these constructs,[60, 63, 65] and reflect the decrease in the end-to-end distance when H' DNA is bent into a U-turn in the complex, as illustrated in the crystal structure of the IHF-H' complex (Figure 3.1.).

MEM analysis on the double-labeled IHF-H' fluorescence decay traces revealed two dominant lifetime populations for these constructs, with 78 ± 3% of the population in a high FRET state (0.65 ± 0.03), and 22 ± 3% in a low FRET state (0.085 ± 0.002). We note here that the low FRET state is close to what we observe in unbound (straight) DNA, and could arise from (i) a fraction of DNA that is unbound or (ii) a fraction that is nonspecifically bound and that competes with specific binding or (iii) a fraction that is specifically bound but only partially bent, for example if the energetic cost of kinking at two sites is not fully compensated for by stabilizing interactions between protein and bent DNA. Other contributions to this low-FRET state could be from potential artifacts, such as incompletely labeled (or incompletely annealed) DNA with some fraction of donor-only labeled DNA molecules lacking an acceptor, or partial stacking of the fluorophores at the ends of the DNA that could result in different dye orientations (stacked or unstacked) and hence two or more distinct FRET states. We ruled out any significant contributions to the observed lifetime distributions from these artifacts, as described in Chapter 2.3.6.

3.2.3. The low-FRET population does not reflect unbound DNA under the 1:1 binding conditions

We first address whether the low-FRET population could have contributions from a significant fraction of unbound DNA, even in our 1:1 binding conditions. We note that under buffer and salt conditions (100 mM KCl) identical to those used for these measurements, the K_d of the IHF-H'

complex was previously determined to be 25 pM, from stopped-flow measurements,[60] and 27 pM, from equilibrium salt-titration measurements.[59] The binding studies on IHF-H' presented here are consistent with the previous studies, with an extrapolated value of $K_d \approx 23$ pM at 100 mM KCl Therefore, at the protein and DNA concentrations used for the lifetime studies (5 μM each) we expect to have >99.8% of the DNA in complex.

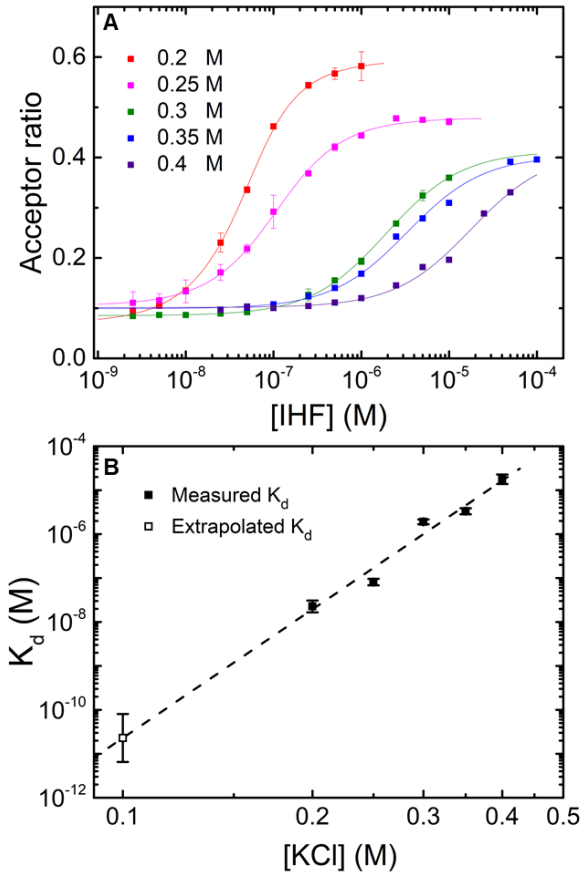


Figure 3.3. Binding affinity measurements on IHF-H' complex at varying salt concentrations. (**A**) Equilibrium binding isotherms for IHF binding to H' are shown from acceptor ratio measurements on 50 nM H'_DA (design I) as a function of [IHF]. The binding curves are for varying KCL concentrations: 200 mM (red), 250 mM (pink), 300 mM (green), 350 mM (blue), and 400 mM (purple). The lines are the best fit to a 1:1 binding isotherm, yielding the dissociation constants K_d shown in panel (B). (**B**) K_d values (filled squares) obtained from the data in panel (A) are plotted as a function of [KCl]. The dashed line is a linear fit to log(K_d) versus log([KCl]), with a slope of 9.7 ± 1.2. The K_d value at 100 mM KCl, extrapolated from this linear fit, is 23^{+57}_{-14} pM (open square).

To confirm 1:1 binding conditions for these complexes, we carried out equilibrium titration measurements at a fixed (1 μM) concentration of H'_DA DNA and varying concentrations of IHF, and simultaneously measured both the acceptor ratio (to measure the extent of bent DNA), and donor anisotropy (to measure the extent of protein binding). These data, shown in Figure 3.4.A-B, exhibited behavior that deviated from what is expected for a 1:1 binding isotherm. The acceptor ratio initially increased with increasing protein concentration, as anticipated, and reached a maximum value at [IHF]:[H'] ≈ 1:1; however, the acceptor ratio began to drop with further excess of protein. In contrast, the anisotropy also initially increased with increasing protein concentration, but then continued to increase with further excess of protein beyond [IHF]:[H'] ≈ 1:1. Taken together, we draw the following conclusions from these data: at [IHF]:[H'] ≈ 1:1, each DNA molecule is bound to a single copy of IHF in a specific complex; below this concentration ratio, there is a mixture of unbound and specifically-bound DNA; above this ratio, there is competition between specific and nonspecific binding, with increasing protein concentrations tilting the equilibrium in favor of a nonspecific binding mode in which more than one copy of IHF is bound to a single 35-mer, resulting in higher molecular weight complexes containing less bent DNA (Figure 3.4.C). Competition between specific and nonspecific binding at high IHF to DNA concentrations has been well documented in previous studies.[51, 66-67]

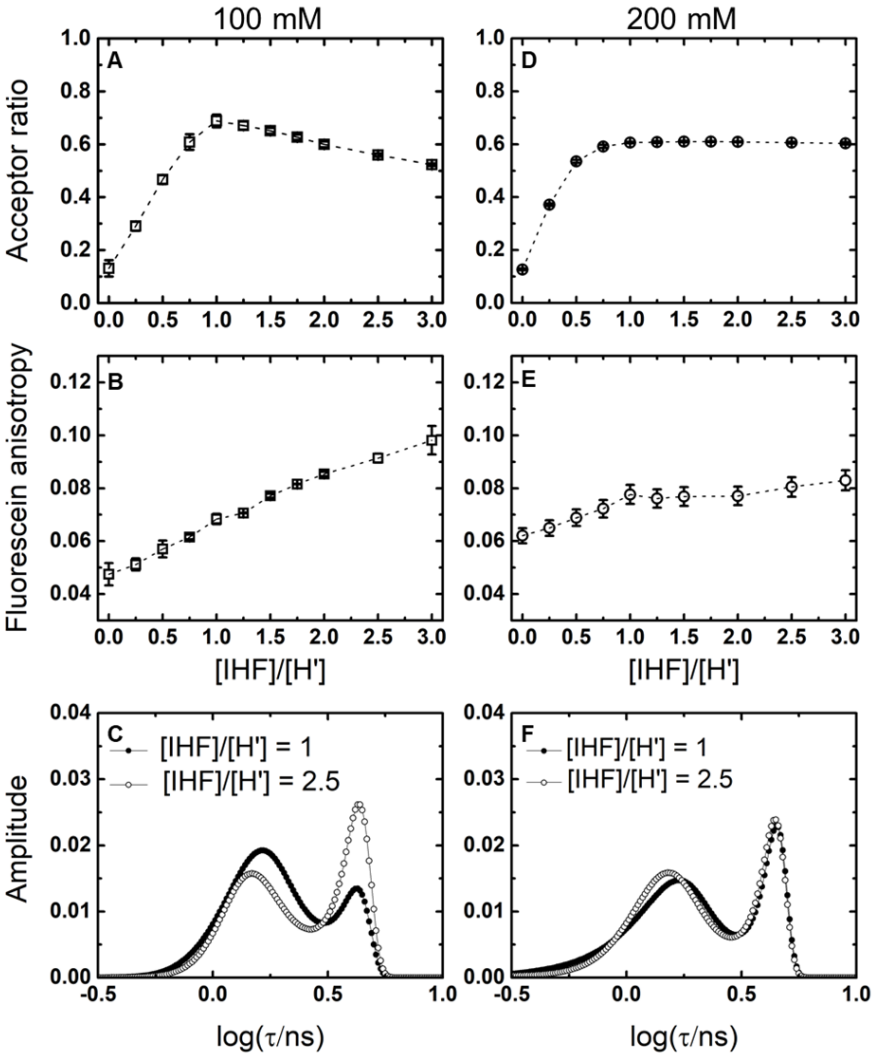


Figure 3.4. Binding isotherms and MEM distributions for the IHF-H' complex at 100 and 200 mM KCl. (**A,D**) Acceptor ratio measurements and (**B,E**) anisotropy measurements are shown for 1uM H'_DA and varying concentration of IHF in 100mM KCl (**A,B**) and 200mM KCl (**D,E**). (**C,F**) The MEM lifetime distributions obtained from fluorescence decay traces measured for IHF-H' are shown for [IHF]/[DNA] = 1 (continuous lines) and [IHF]/[DNA] = 2.5 (dashed lines) in 100 mM KCl (**C**) and 200 mM KCl (**F**). DNA concentrations for the lifetime measurements were 5 μM.

3.2.4. Increasing the salt concentration diminishes contributions from nonspecific binding

To further examine this competition between specific and nonspecific binding, we performed measurements at a higher salt concentration, 200 mM KCl (Figure 3.4.). Equilibrium studies on the ionic strength dependence of protein-DNA complexes have shown that while higher ionic strength conditions disfavor both binding modes, such conditions are typically more disruptive to nonspecific than to specific binding.[51, 68-72] In other words, an increase in salt is expected to shift the equilibrium in favor of specific binding, as has indeed been demonstrated by isothermal titration calorimetry (ITC) studies for IHF-H'[51] as well as for other DNA-bending proteins.[73-76] This behavior reflects the fact that nonspecific interactions are primarily electrostatic interactions between the protein and the DNA phosphate groups, while specific interactions have significant contributions from hydrogen bonds, van der Waals interactions, and water-mediated interactions that are less affected by ionic strength. Therefore, by increasing the salt concentration from 100 to 200 mM KCl, we anticipated that the contributions from nonspecific binding in our IHF-H' complexes should diminish, even when IHF is present in excess over DNA.

We repeated our equilibrium titration measurements in 200 mM KCl conditions, with acceptor ratio and anisotropy as probes of the binding process, as before (Figure 3.4.D-E). Indeed, unlike the 100 mM KCl data, the acceptor ratio versus protein concentration profile in 200 mM KCl exhibited the behavior expected for a 1:1 binding isotherm, with the acceptor ratio initially increasing with protein concentration, and then reaching a plateau with protein in excess. The anisotropy data also resembled a 1:1 binding isotherm and did not show any significant evidence for multiple protein binding. Taken together, we infer that at 200 mM KCl nonspecific binding is sufficiently destabilized such that we observe only specific binding at all protein:DNA concentrations.

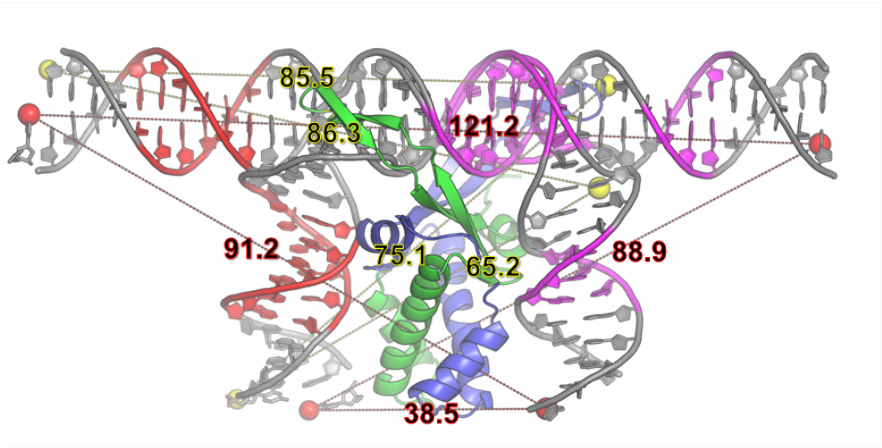
We now return to the question: what is the origin of the low-FRET component in the lifetime distributions measured on IHF-H' in 100 mM KCl under 1:1 conditions? Corresponding lifetime measurements for the 1:1 complex at 200mM KCl also showed two populations in the lifetime distributions (Figure 3.4F), with 66 ± 2% in the high FRET state ($E \approx 0.67 \pm 0.02$) and 34 ± 2% in the low FRET state ($E \approx 0.043 \pm 0.001$). Even with a 2.5-fold increase in the protein concentration (2.5:1 complex) in 200 mM KCl, almost no change was observed in the IHF-H' lifetime distribution, consistent with our acceptor ratio and anisotropy data at 200mM that showed no evidence for nonspecific binding at excess protein conditions. From these data we conclude that the low-FRET state must be a less-bent DNA conformation of a specifically-bound complex, and which could not be detected in previous steady-state measurements or in crystal structures. The population in this less-bent conformation increased from ~22% to ~34% with the increase in salt concentration from 100 to 200 mM KCl.

3.2.5. Measurements with different placement of FRET labels (design II constructs) implicate the low-FRET state as arising from conformations with less bent DNA

The results presented thus far are insufficient to conclude whether the low-FRET component was from specifically-bound but straight DNA or from an ensemble of partially bent conformations, for example with DNA kinked at only one site or the other. With the FRET labels attached at the ends of the 35-mer H' substrate (design I in Table 3.1.), the end-to-end distance for a straight piece of DNA is estimated to be ~121 Å (Figure 3.5.), with near zero FRET efficiency, assuming Forster distance $R_0 \approx 50$ Å for this FRET pair.[62, 77] If we envision a distribution of partially bent conformations with either one side kinked or the other, as illustrated in SI Figure S7, then the end-to-end distance is estimated to be ~90 Å, with FRET $E \approx 0.03$. We note that these FRET estimates,

especially for the partially bent conformations, are only rough approximations based on reasonable guesses of these structures and the expected distances between the atoms where the labels are attached. Furthermore, these FRET estimates do not account for linker lengths used to attach the labels or their dynamics. Despite these caveats, these estimates are in reasonably good agreement with $E \approx 0.083 \pm 0.002$ measured for the low-FRET state in IHF-H' at 100 mM KCl. However, the width in our lifetime distributions makes it difficult to unambiguously separate these low-FRET states from those obtained on straight DNA (Figure 3.2.). To increase the sensitivity of our FRET measurements and more clearly detect FRET changes between straight and partially bent conformations as envisioned above, we altered the labeling strategy to design II (see Table 3.1.), in which the labels are placed closer together along the DNA. In these constructs, we attached the donor (fluorescein) internally, at a thymine located at position 26 of the top strand (counting from the 5' end), and attached Atto550 (At) at the 3' end of the bottom strand. The separation between the attachment points of the labels in these design II constructs is ~86 Å in unbent DNA (Figure 3.5.).

A



B

Structure	Estimated Distance (FRET)	
	Design I	Design II
Free DNA	121.2 Å (0.005)	85.5 Å
Fully bent complex	38.5 Å (0.83)	65.2 Å
A-tract side unkinked	91.2 Å (0.03)	86.3 Å
TTG side unkinked	88.9 Å (0.03)	75.1 Å

Figure 3.5. Computed distances and FRET values from structures. **(A)** Overlay of crystal structures of the fully bent complex with that of free DNA. The red (yellow) spheres indicate the atoms where the labels were attached in design I (design II) constructs. The dotted lines indicate the distances between the attachment points (red pair or yellow pair); four lines for each design correspond to distances computed for four different conformations: free DNA, fully bent complex, and two partially bent conformations, one in which the A-tract side (shown by the red nucleotides) is unkinked and the other in which the TTG side (shown by the magenta nucleotides) is unkinked. **(B)** The distances between the attachment points, computed as described above, are tabulated for the four different conformations in both design I and design II constructs. The FRET E values based on these distances are shown in parenthesis for design I constructs, calculated using $R_0 = 50$ Å; for design II constructs R_0 is not well known and hence no corresponding FRET E values are computed. We caution that these distance/FRET estimates are rough approximations and do not represent the actual distance/FRET between the labels in each of these structures, since they do not take into account the lengths and dynamics of the linkers that are used to attach the labels.

Fluorescence lifetime decay traces on donor-only H' DNA in the design II constructs, performed at 100 mM KCl conditions, still exhibited a predominantly single-exponential decay, both in the absence and presence of IHF (Figure 3.6. and Figure 2.5.). The lifetime decay traces for donor-acceptor-labeled H'_DA remained primarily single-exponential, with FRET *E* of 0.113 ± 0.002. In the presence of IHF, the lifetime distributions on the double-labeled constructs showed two distinct FRET states, with a high-FRET state observed at 0.51 ± 0.01 and a low-FRET state at 0.29 ± 0.01, reaffirming our results from design I that the DNA in this ensemble samples different conformations. More important, in the design II constructs, the FRET efficiency of the low-FRET component in the IHF-DNA complex is well separated from the FRET efficiency of unbound DNA, supporting the conclusion that the low-FRET state corresponds to partially bent DNA. These partially bent complexes could be a mixture of conformations with one side kinked or the other, as explained below, while the high-FRET state likely corresponds to both DNA sites being kinked and the flanking arms of the DNA held against the sides of the protein, as seen in the crystal structures.

We note here that while the FRET values measured in the two constructs (design I and II) cannot be quantitatively compared because the FRET pairs used in these constructs are different and may not have the same Forster distance, our observation that the high-FRET component shifts to lower FRET values while the low-FRET component shifts to higher FRET values when we change from design I to design II is not inconsistent with what we anticipated. Although the separation between the labels is shorter along the DNA in design II, the labels in design I are in fact spatially closer than in design II in the fully-bent state, as illustrated in Figure 3.5..; in the partially bent states, the labels in design I are expected to be further away than in design II. We also note that the population distribution between the high- and low-FRET states in the design II

constructs are split more evenly, with ~50% in each of these conformations, compared with the approximately 80% versus 20% split observed for the complex in the design I construct under identical conditions. The reason for this population shift could be some steric hindrance between the internally located fluorescein label on the DNA near one of the kink sites that could hamper IHF from keeping the flanking arm of the DNA against its side, although the binding affinity measurements at 300 mM KCl showed no significant difference in the K_d values for IHF-H' between the two designs (see Table 3.1.). The remaining results presented below are for the design I constructs.

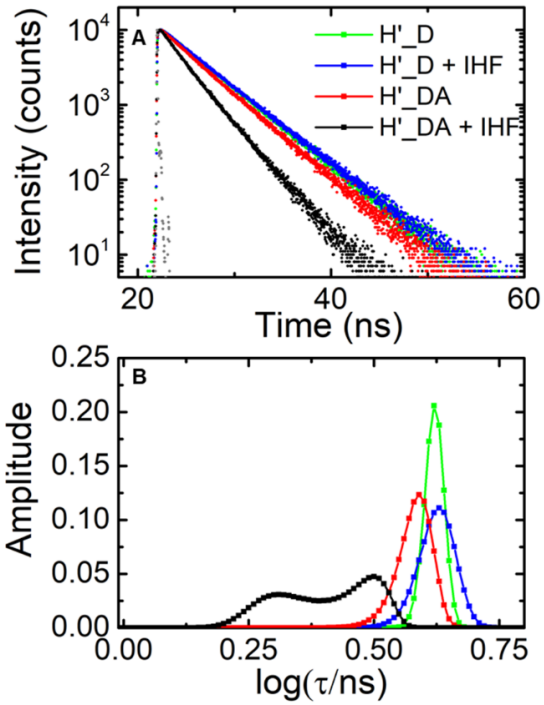


Figure 3.6. Fluorescence lifetime measurements on IHF-H' complex in design II constructs. **(A)** Fluorescence intensity decay traces measured on H' labeled with fluorescein and Atto550 (H'_DA) are shown in the absence (red) and presence (black) of IHF. The corresponding donor-only (H'_D) decay traces are also shown in the absence (green) and presence (blue) of IHF. Measurements were done with 5 μM DNA and 5 μM IHF. The instrument response function (gray) is shown for comparison. **(B)** The MEM lifetime distributions obtained from the fluorescence decay traces are shown. The amplitudes of the distributions are normalized to add up to one.

3.2.6. The H1 site, which lacks the A-tract, favors the less bent state

We next examined the conformational distribution of another specific binding site recognized by IHF, the H1 site on phage lambda DNA. H1 has the same consensus region as H' (the WATCAAnnnnTTR indicated in gray in Table 3.1.) but differs primarily in the other half of the binding site where it lacks the A-tract that is present in the H' sequence. Previous stopped-flow measurements revealed a K_d of ~20 pM for the IHF-H1 complex at 100 mM KCl,[61] very similar to the K_d from stopped-flow on the IHF-H' complex.[60-61] These results indicate that the lack of the A-tract appears to be compensated for by other changes in the H1 sequence compared with H' so as not to significantly perturb the overall binding affinity. However, the average FRET efficiencies measured for the two end-labeled (design I) constructs in complex with IHF are distinctly different. Our lifetime studies (performed at 100 mM KCl and 20 °C) revealed an average FRET efficiency of 0.25 ± 0.01 for the IHF-H1 complex compared with 0.50 ± 0.03 for IHF-H' complex (Table 3.1.), consistent with previous steady-state measurements on these complexes using identical constructs and buffer conditions.[60-61, 63] The decreased FRET efficiency measured in the IHF-H1 complex suggests an inability of IHF to keep the bent arm of the DNA by its side in the absence of the A-tract.

To examine how the absence of the A-tract affects the relative populations of differently bent DNA conformations, we analyzed the fluorescence decay traces measured on the end-labeled IHF-H1 complex with the MEM, as before. The lifetime distributions for IHF-H1 also show two peaks (Figure 3.7. and Figure 2.4.); however, the DNA in the IHF-H1 complex prefers the low-FRET conformation, with only 32 ± 1% in the high FRET state ($E \approx 0.50 \pm 0.03$) and a larger fraction, 68 ± 2%, in the low FRET conformation ($E \approx 0.14 \pm 0.01$). We assert that these remarkable differences in the conformational distributions of IHF-H' and IHF-H1 are primarily

from the presence and absence of the A-tract, although there are some other changes in the two sequences, especially at the kink sites (see Table 3.1.), that could also be affecting the distributions and the binding affinities. For example, we note that even the high FRET conformation in IHF-H1 (FRET $E \approx 0.50$) appears to be not as fully bent as in IHF-H' (FRET $E \approx 0.65$).

To further examine the differences that arise purely from the presence or absence of the A-tract, we also conducted experiments on a modified H' construct, denoted as H'_nAt, which differs from the H' construct only in the A-tract segment, which was swapped by the corresponding segment in the H1 site (see Table 3.1.). Binding affinity measurements with end-labeled H'_nAt showed a ~2-fold increase in K_d compared with H' (measured at 300 mM KCl; Table 3.1.). The distribution of FRET components in IHF-H'_nAt mirrored that observed in IHF-H1, with a lower fraction (33 ± 1%) in the high-FRET state ($E \approx 0.57 \pm 0.03$) and a larger fraction (67 ± 1) in the low-FRET state ($E \approx 0.077 \pm 0.002$) compared with IHF-H'. These results highlight the importance of the A-tract and the contacts that it facilitates with IHF to help the protein clamp the bent arm of the DNA on that side. The binding preference of IHF for the A-tract was well noted when the crystal structure was solved, since the A-tract has a unique structure with a narrow minor groove and a high twist which allows it to fit into the protein clamp without significant additional distortions.[9]

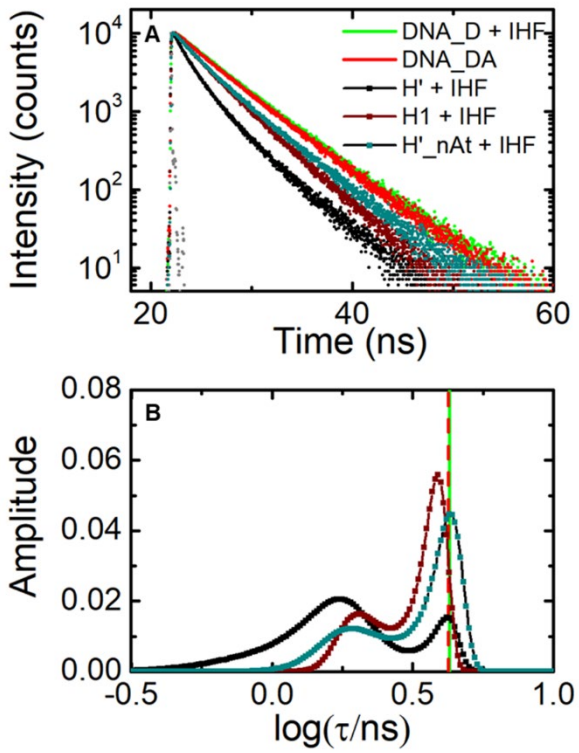


Figure 3.7. Fluorescence lifetime measurements on IHF-H' compared with IHF-H1 (design I constructs). (A) Fluorescence intensity decay traces are shown for H'_DA (black), H1_DA (maroon), and H'_nAt_DA (blue-green), measured in the presence of IHF. Decay traces on DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are shown for comparison. Measurements were done with 5 μM DNA and 5 μM IHF. The instrument response function (gray) is shown for comparison. **(B)** The MEM lifetime distributions obtained from the fluorescence decay traces measured for IHF-H' (black), IHF-H1 (maroon), and IHF-H'_nAt (blue-green) are shown. The amplitudes of the distributions are normalized to add up to one. The average lifetime for the DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are indicated by the vertical dashed lines.

3.2.7. Insertion of a mismatch at a kink site in H1 on the A-tract side helps recover the U-bent conformation

Next, we inserted a 4 nucleotide bubble (CT/TC mismatch) into the H1 sequence at the kink site on the same side as where the A-tract is in H' (see H1_CTloop in Table 3.1.). The 4-nt mismatch bubble is expected to enhance the "kinkability" of the H1 sequence on that side and increase the IHF binding affinity, as was shown previously for similar mismatched sequences in the H'

context.[56, 59, 65] Binding affinity measurements with end-labeled H1_CTloop showed a ~125-fold decrease in K_d compared with matched H1 (measured at 300 mM KCl; Table 3.1.), a result consistent with previous K_d measurements on H' sequence with a TT/TT mismatch introduced at the same kink site.[56, 59, 65] We anticipated that the insertion of the mismatch in H1 to lower the energetic penalty for kinking should help compensate for the lack of stabilizing interactions afforded by the A-tract. In other words, we expected to recover some of the lost high FRET state in IHF-H1 in the presence of the mismatch. Our lifetime studies are consistent with this expectation.

First, the average FRET efficiency measured in IHF-H1 increased from 0.25 ± 0.01 in the matched construct to 0.36 ± 0.02 in the mismatched (H1_CTloop) construct (Figure 3.8. and Figure 2.4.). Second, the population in the high-FRET component increased from 32 ± 1% in the matched IHF-H1 to 43 ± 3% in IHF-HI_CTloop. In addition, we observed a shift in the FRET efficiency values for the two populations, with the high FRET state shifted from 0.50 ± 0.03 to 0.55 ± 0.02, and the low-FRET state shifted from 0.14 ± 0.01 to 0.25 ± 0.01, when comparing matched versus mismatched versions of this complex. These shifts in the FRET efficiencies indicate additional conformational changes in the kinked/straight states of the DNA introduced by the mismatched bubble.

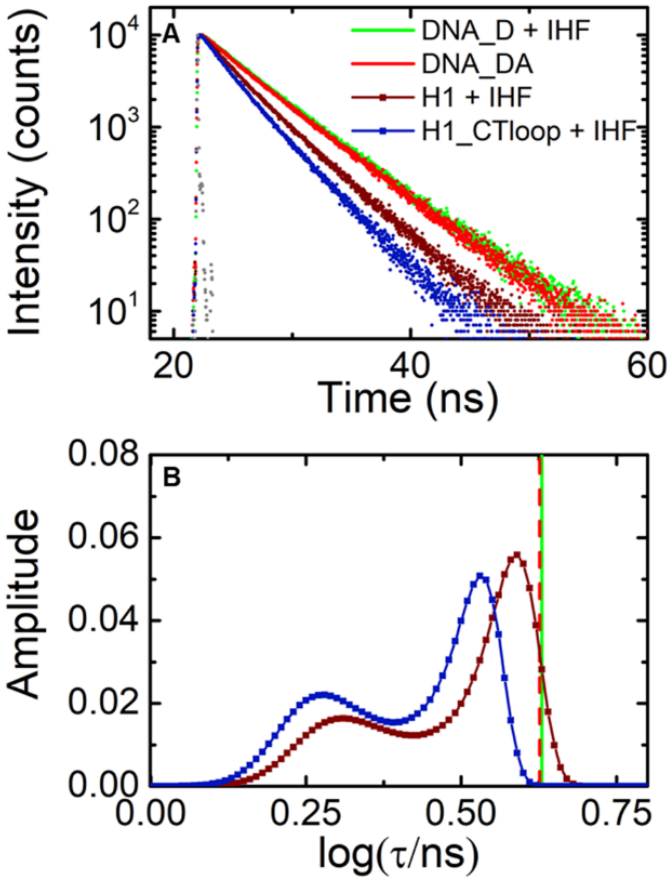


Figure 3.8. Fluorescence lifetime measurements on IHF-H1 compared for matched and mismatched (design I constructs). (A) Fluorescence intensity decay traces measured on H1_DA (maroon) and H1-CTloop_DA (blue), both in the presence of IHF. Decay traces on DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are shown for comparison. Measurements were done with 5 μM DNA and 5 μM IHF. The instrument response function (gray) is shown for comparison. **(B)** The MEM lifetime distributions obtained from the fluorescence decay traces measured for H1_DA (maroon) and H1_CTloop_DA (blue) are shown. The amplitudes of the distributions are normalized to add up to one. The average lifetime for the DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are indicated by the vertical dashed lines.

3.2.8. A destabilizing modification in the TTG consensus region of H' increases the population in (another) low-FRET state

Next, we examined the effect of sequence modifications in the TTR consensus region on the other flanking arm of the DNA (Figure 3.1.). In this consensus site, a single adenine substitution (TTG→TAG) as shown in the sequence H'44A (Table 3.1.), was found to destabilize the binding affinity of IHF by about 100- to 250-fold.[57, 59] The crystal structure for the IHF-H'44A complex revealed that this single nucleotide substitution inhibited the twisting of the DNA that was needed to form a network of salt bridges with IHF that stabilized the bent DNA conformation against that side of the protein.[57] However, the crystal structure of IHF-H'44A did reveal a fully bent conformation very similar to that of IHF-H'. In contrast, steady-state FRET measurements on the IHF-H'44A complex at 20 °C in solution yielded a smaller average FRET efficiency than IHF-H' for the design I constructs under identical binding conditions,[59] indicating some degree of "floppiness" in the bent conformations of IHF-H'44A.

To examine how this T→A mutation in the TTG region affects the populations in the fully versus partially bent conformations of the IHF-bound specific complexes, we carried out lifetime studies on the end-labeled (design I) IHF-H'44A complex (Figure 3.9. and Figure 2.4.). We still observe two populations as for the IHF-H' complex, with a high FRET state at $E \approx 0.62 \pm 0.02$ and a low FRET state at $E \approx 0.076 \pm 0.001$, but with a significant decrease in the high FRET population of 32 ± 1% compared with ~78% in IHF-H'. These results demonstrate the importance of the TTG consensus region in keeping that side of the DNA clamped against the protein and further show that neither of the flanking DNA arms are held rigidly in place. Taken altogether, we conclude that although only two distinct FRET populations are discerned in the MEM lifetime distributions, the combined results from IHF-H', IHF-H'_nAt, and IHF-H'44A strongly suggest

that at least three conformations co-exist in solution for a specifically bound protein complex: a fully bent conformation reflected in the high-FRET component and at least two partially bent conformations with one arm straight or the other, with overlapping FRET levels that are not distinguishable in the low-FRET component (Chapter 2.6.). The results from IHF-H'44A suggest that the population in the low-FRET state increases by 46% as a result of increased population in the second partially bent conformation that affects the TTG side of the complex when that side is compromised.

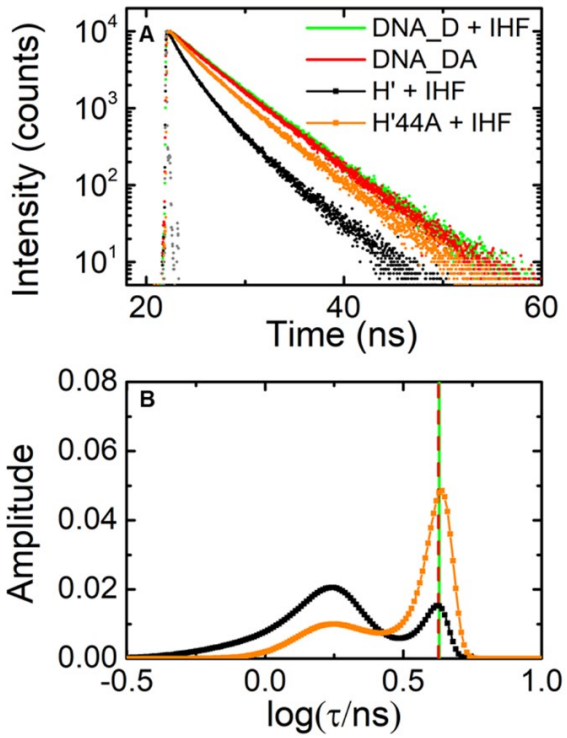


Figure 3.9. Fluorescence lifetime measurements on IHF-H' compared with IHF-H'44A (design I constructs). (A) Fluorescence intensity decay traces measured on H'_DA (black) and H'44A_DA (orange), both in the presence of IHF. Decay traces on DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are shown for comparison. Measurements were done with 5 μM DNA and 5 μM IHF. The instrument response function (gray) is shown for comparison. **(B)** The MEM lifetime distributions obtained from fluorescence decay traces measured for H'_DA (black) and H'44A_DA (orange) in the presence of IHF are shown. The amplitudes of the distributions are normalized to add up to one. The average lifetime for the DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are indicated by the vertical dashed lines.

Table 3.2: Lifetime Measurements on IHF-DNA Complexes

Construct	$\langle \tau_1 \rangle$ (ns)	$\langle E_1 \rangle$	α_1	$\langle \tau_2 \rangle$ (ns)	$\langle E_2 \rangle$	α_2	$\Delta G/k_BT = -\ln\left(\frac{\alpha_1}{\alpha_1}\right)$
IHF-H' (design I)	1.52 ± 0.06[a] (1.41 ± 0.03)[b]	0.65 ± 0.03 (0.67 ± 0.02)	$78 \pm 3\%$ ($66 \pm 2\%$)	3.96 ± 0.01 (4.13 ± 0.02)	0.085 ± 0.002 (0.043 ± 0.001)	$22 \pm 3\%$ ($34 \pm 2\%$)	-1.3 ± 0.2 (-0.66 ± 0.04)
IHF-H' (design II)	2.06 ± 0.02	0.51 ± 0.01	$48 \pm 3\%$	3.00 ± 0.01	0.29 ± 0.01	$52 \pm 3\%$	0.090 ± 0.008
IHF-H1 (design I)	2.14 ± 0.14	0.50 ± 0.03	$32 \pm 1\%$	3.71 ± 0.12	0.14 ± 0.01	$68 \pm 1\%$	0.75 ± 0.03
IHF-H'_nAt (design I)	1.87 ± 0.01	0.57 ± 0.01	$33 \pm 1\%$	3.99 ± 0.04	0.077 ± 0.001	$67 \pm 1\%$	0.70 ± 0.02
IHF-H1_CTloop (design I)	1.96 ± 0.08	0.55 ± 0.02	$43 \pm 3\%$	3.22 ± 0.16	0.25 ± 0.01	$57 \pm 3\%$	0.32 ± 0.03
IHF-H'44A (design I)	1.66 ± 0.04	0.62 ± 0.02	$32 \pm 1\%$	3.99 ± 0.01	0.076 ± 0.001	$68 \pm 1\%$	0.75 ± 0.03

[a] Measurements were done in 100 mM KCl

[b] Values in parenthesis are for measurements done in 200 mM KCl

3.3. Discussion

IHF is a small protein belonging to a class of nucleoid-associated DNA-bending proteins. Apart from its nonspecific biological function in condensing the bacterial nucleoid, it also binds in a sequence-specific manner and serves as an architectural factor in many cellular activities such as site-specific recombination, DNA replication and transcription.[78-79] The ability of IHF to bend the DNA containing its specific site into a U-turn by wrapping ~35-bp DNA around three sides of the protein has earned it the moniker of the "master bender".[80] Remarkably, IHF accomplishes this feat with almost no direct interactions between the protein residues and specific bases, and has thus become an excellent model system for studies of sequence-dependent DNA shape and deformability that underpins binding site recognition by indirect readout.[67, 81-83]

The sharp DNA bends induced by IHF allow for FRET measurements to be sensitive reporters of the extent of DNA bending. Previous studies took advantage of time-resolved FRET to investigate DNA-bending kinetics in IHF-DNA complexes,[60, 63] which demonstrated the step-wise binding-then-bending mechanism for site recognition by IHF.[84] Further kinetics studies resolved a two-step "interrogation-then-recognition" process: nonspecific interrogation on ~100-500 μs timescale prior to recognition on 1-10 ms timescale.[65] However, these "ensemble" approaches could only provide an average picture of the dynamics along the reaction trajectory. What remained elusive was whether multiple conformations of the complex could co-exist in solution, as was recently shown for damaged DNA specifically bound to NER damage-recognition protein XPC/Rad4.[43] Here, we investigated the distribution of bent conformations in IHF-DNA complexes with varying DNA sequence composition and binding affinities. We utilized fluorescence lifetime decay measurements and the maximum entropy method (MEM) to infer

multiple FRET efficiencies that enabled us to visualize and quantitatively characterize the heterogeneity of bent conformations.

Like many DNA-bending proteins that kink DNA at localized sites, IHF concentrates the U-shaped bend in two sharp kinks separated by 9 bp. At the kink sites, a single base step is unstacked and opened towards the minor groove of the DNA, and stabilized by intercalation of conserved proline residues on the β-arms of the protein that wrap around the DNA. The flanking sides of the DNA on the outer sides of the kinks are held against the body of the protein through a myriad of specific and nonspecific interactions. Notably, the consensus DNA-binding motif consists of only a 6-bp stretch (WATCAR) in between the kink sites and another 3-bp stretch (TTR) in the flanking DNA, making it all the more remarkable that IHF is able to overcome the energy penalty needed to severely deform the DNA at its preferred sites and bind with affinities that can exceed 10^3-10^4-fold compared with random sequences.[85-88]

How accurate is this picture of bent DNA rigidly held against the protein, as implied by the static crystal structure of the IHF-H' complex? Incorporation into a crystal can "freeze out" macromolecular dynamics and will tend to select a single conformation from the ensemble that may exist in solution. The results reported here demonstrate that IHF does in fact experience some difficulty in keeping the bent arms of the DNA at its side. For the IHF-H' complex in 100 mM KCl, the ensemble of bent conformations appears as two discernible populations, as inferred from the lifetimes recovered for labeled DNA constructs. For the end-labeled (design I) constructs, the population in the fully-bent high-FRET state ($E \approx 0.65$) is found to be 78%, with the remaining 22% in a low-FRET state ($E \approx 0.085$). Although measurements on the design I constructs could not readily distinguish between partially bent or unbent DNA, FRET measurements with design II constructs, where the FRET labels are placed closer together along the DNA, establish that the

low-FRET state is not from unbent (straight) DNA. The low-FRET state is attributed to a partially bent but still specific complex, as clarified from measurements at 200 mM KCl, where nonspecific binding is disfavored, and yet the amplitude of the low-FRET state increases. From these and further results discussed below, we assign the low-FRET state to an ensemble of specifically-bound conformations with either one or the other kink site unkinked (Figure 3.10.). The free energy of the fully bent IHF-H' conformation is estimated to be 1.3 k_BT lower than the partially bent ensemble in 100 mM KCl and 0.7 k_BT lower in 200 mM KCl (Table 3.2). Notably, the FRET efficiency levels observed in the fully-bent conformation are very similar in both ionic conditions.

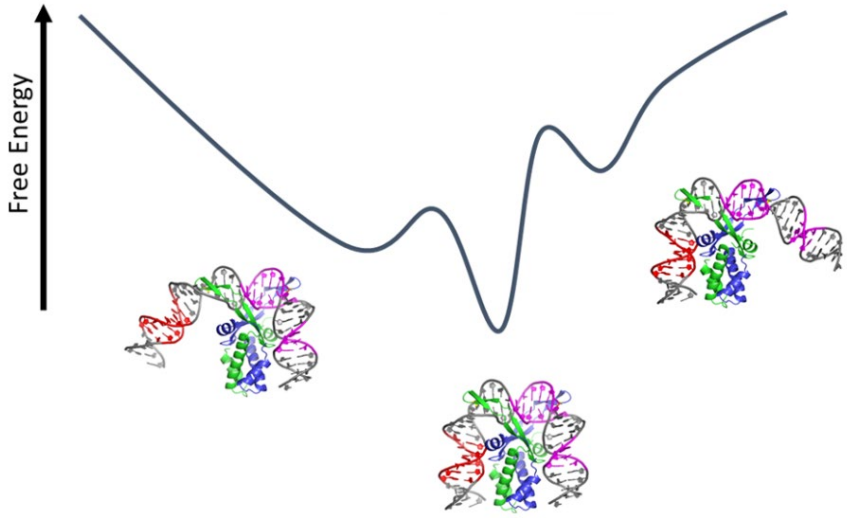


Figure 3.10. Schematic representation of the free energy landscape of the specific IHF-DNA complex, with multiple conformations accessible in solution. These conformations include the most stable complex, with two DNA sites kinked, as well as conformations with one or the other side unkinked. The partially bent conformations shown are models of the low-FRET population and not based on real structural data.

This tug-o-war between protein-induced DNA kinking and the propensity of DNA to retain an unkinked conformation is further illustrated by measurements with the H1 binding site that is missing an A-tract found in the H' site, which is known to help stabilize the bent DNA conformation in the IHF-H' complex.[9, 50] FRET efficiencies obtained for IHF-H1 revealed significantly less population (32%) in the high-FRET conformation, indicating that the free energy of the fully-bent IHF-H1 conformation is 0.75 k_BT higher than the partially bent conformations. Furthermore, the protein-DNA interactions in the fully bent IHF-H1 conformation (FRET $E \approx$ 0.50) appeared to be looser than that in the fully-bent IHF-H' (FRET $E \approx 0.65$). Similar results were obtained on complexes with a modified H' construct (H'_nAt) that was also lacking the A-tract. Insertion of 2-bp mismatches at one of the kink sites in H1 close to the side that was missing the A-tract (H1_CTloop in Table 3.1.), designed to make the DNA more "kinkable", compensated to some extent for the loss of the A-tract; the fully-bent complex in IHF-H1_CTloop was still less favored but now only by 0.32 k_BT (Table 3.2.). Thus, the A-tract helps to maintain a tight fit in the complex and its loss results in significant unkinking on that side. These results are in accord with previous hydroxyl radical footprinting studies that showed less protection in that patch in sequences that were missing the A-tract.[46, 81]

Another important and highly conserved feature common to all known binding sites of IHF is the TTR consensus region on the other flanking side of the DNA. Previous studies identified that a T→A switch at the center position of the TTR element (H'44A) resulted in >100-fold decrease in binding affinity for IHF.[57, 59, 89] Comparison of IHF-H' and IHF-H'44A structures showed that the ability of the TTR site to adopt an unusually highly twisted conformation at the Y-R step when bound to IHF was necessary to facilitate stabilizing salt bridges between key

residues of IHF.[57] These studies exemplified how sequence-dependent DNA deformability was critical to the recognition of that consensus site.

The crystal structures of IHF-H' and IHF-H'44A were otherwise very similar, with approximately the same overall bend in the DNA observed in both structures.[57] In contrast, previous steady-state FRET studies that monitored the average end-to-end distance already revealed a less bent conformation for IHF-H'44A in solution, with FRET efficiency levels nearly half of what was observed in IHF-H'.[59] Our lifetime measurements directly show that indeed the fraction in the fully-bent conformation of IHF-H'44A is only 32%, indicating a penalty ($\Delta\Delta G$) of about 2.1 k_BT between the bent and unbent states that is attributable to the loss of the salt bridge interactions at the TTR site.

It is informative to compare the conformational distributions characterized here for the specific IHF-DNA complexes with the conformational distributions for the structurally similar but nonspecific HU protein observed in AFM studies.[25] HU is known to bind in a sequence-independent manner to DNA with K_d values that range from 200 nM to 2.5 μM,[24, 90] and with much higher affinities to distorted DNA.[55, 90-91] Single-molecule micromanipulation studies of HU-bound DNA showed that at low HU/high monovalent salt concentrations, HU dimers induce very flexible bends that result in DNA compaction and a dramatic decrease in the apparent persistence length of DNA compared with bare DNA.[25, 92] AFM studies under similar conditions revealed a very broad range of bend angles in the DNA at the sites where HU was bound,[25] with a nearly uniform distribution of angles from 0° (unbent) to 180° (bent into a U-shape). Together, these studies revealed a highly compliant and very flexible HU-DNA complex.

Similar conclusions were drawn from force-extension measurements on long DNA with IHF bound nonspecifically that also showed enhanced apparent DNA flexibility with the bound

proteins.[88, 93] AFM studies with specific IHF-DNA complexes revealed single broad distributions peaked at bending angles of ~120-130°, with a range that covered bending angles from ~80° to ~160°.[27, 33] These AFM studies were done with other IHF-DNA complexes; we are unaware of similar studies with IHF-H'. Notably, the range of bent conformations observed in AFM covers what we expect for the low-FRET state (with one site kinked) and for the high-FRET state (with both sites kinked).

However, our studies indicate a somewhat "brittle" complex for IHF bound to its specific site. Rather than describing a broad and continuous range of bent conformations as seen in AFM images of HU-DNA and IHF-DNA complexes, our data support two or likely three preferred conformations, with the populations among these distinct valleys in the free energy landscape modulated by the DNA sequence. While only two distinct FRET states could be discerned in our lifetime distributions on all the complexes, the relative populations in these so-called low- and high-FRET states could be modulated by making modifications in the DNA flanking segments on either side, whether in the A-tract segment on one side or the TTR consensus segment on the other side, with destabilizing modifications on either side resulting in an increase in the low-FRET population at the expense of the high-FRET population. These observations suggested that the low-FRET component likely included contributions from two subpopulations, with either one arm unbent or the other, that had very similar FRET levels and therefore were not distinguishable as two distinct peaks in our lifetime distributions.

In a previous smFRET study on IHF-H' that used a 55-bp DNA construct containing the H' site and end-labeled with a FRET pair,[94] a bimodal FRET distribution was indeed observed, with ~85% in a high-FRET conformation consistent with the crystal structure and ~15% population in what appeared as a "zero-FRET" conformation. Remarkably, a very similar bimodal

distribution was also observed in complexes of HU bound to a 55-mer with two TT mismatches 9 bp apart, but not with HU bound to the 55-mer H' construct, which revealed a broad, feature-less distribution reflecting less severely bent conformations. HU has been shown to bind with very high affinity (K_d in the 4-10 pM range) to DNA substrates with mismatches spaced 9 bp apart and it is not unexpected that HU can induce U-bends in these high affinity sequences similar to IHF-H'. The authors of this study interpreted the zero-FRET component as arising from nonspecifically bound proteins. Indeed, it is not evident that these smFRET measurements could discern distinct populations within the specific complex, if those conformations interconverted on timescales faster than (or comparable to) the 1 ms binning times of this smFRET study.[94] Kinetics measurements on IHF-DNA complexes indicate that DNA bending/unbending dynamics within the complex are indeed fast, on micro-to-millisecond timescale.[59-60, 63, 65]

Another smFRET study,[95] designed to examine rigid versus flexible kinks for another nonspecific DNA-bending protein from the eukaryotic family of HMG box proteins, concluded that the ~60° kinks induced by that protein appeared to be rigid kinks, with the apparent enhancement of DNA flexibility induced by these proteins attributed to binding/unbinding of the protein to induce random and transient kinks. Again, as the authors of that study noted, the ~30 ms binning time of their smFRET measurements could have averaged out any dynamic flexibility. Further studies, including measurements of the kind reported here, are needed to carefully flesh out the dynamics and distributions of these ubiquitous DNA benders.[96]

Our observation of partially bent specific conformations in IHF-DNA complexes also provides structural insights into the underlying mechanism for the "facilitated dissociation" observed for several protein-DNA complexes, whereby dissociation rates of these complexes are found to depend on the protein concentrations.[97-101] As postulated by others,[100, 102-103] for dimeric

DNA-binding proteins, the release of a monomer from a half-site is a plausible mechanism to generate a partially bound intermediate, thereby making room for another protein to bind and eventually displace the first protein. Our measurements provide direct evidence for analogous partially bound structures in IHF-DNA complexes, in which the unkinked DNA arm could interact with a second IHF dimer before the first one fully dissociates.

Finally, we propose that the details of IHF binding sites have evolved to fit their biological roles (Figure 9). In particular, the A-tract that clearly helps to keep the DNA bent in the U-shape is not conserved across the many known IHF sites, H1 being a case in point. Binding of IHF to both the H' and H1 sites is required for integration of phage lambda into the *E. coli* chromosome to establish lysogeny. The H' and H1 sites are found within the "attP" region of phage lambda, which is bound synergistically by 3 copies of IHF and 4 copies of lambda integrase to form a large complex (termed an intasome) that then binds the bacterial insertion site ("attB") and catalyzes a site-specific recombination reaction between attP and attB that results in the integration of the phage DNA into that of the host. Modeling of this intasome shows that, while the IHF-H' complex might need to flex slightly to allow synergetic binding of two domains of integrase to DNA sequences flanking it, it can remain static throughout the integration reaction.[58] However, in the fully bent form of the IHF-H1 complex, the flanking DNA and the copy of integrase bound to it block incorporation of the bacterial attB DNA segment. Significant flexing of the H1-induced DNA bend as shown in Figure 3.11. is therefore required for its biological function.

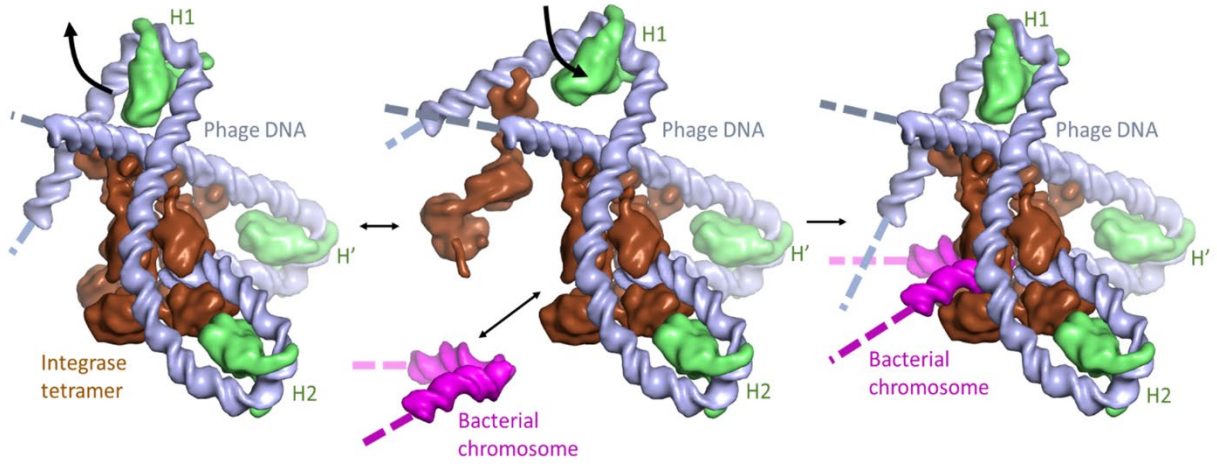


Figure 3.11. Flexible DNA bending at the H1 site may facilitate assembly of the phage lambda integration complex. Proteins and DNA segments in the model are shown as smoothed surfaces. The "intasome" assembles on phage DNA ("attP"; gray), with 3 copies of IHF (green) bending the DNA such that the integrase tetramer (brown) can bridge multiple DNA sites. Transient flapping of the IHF-induced bend at the H1 site (center panel) allows insertion of the bacterial integration site DNA ("attB"; magenta) into the complex, which is then trapped by closure of the H1 bend. Note that in these images the H1 binding site is oriented such that the (missing) A-tract side is on the right, i.e. orientation of IHF relative to the DNA at the H1 site is flipped 180° from that shown in Figure 3.1..

3.4. Conclusion

The present study showcases the power of combining fluorescence lifetime measurements with MEM analysis for investigating conformational flexibility in protein-DNA complexes and establishes conclusively that IHF bound to its specific sites samples two or more distinct conformations. These conformations include a fully-bent conformation such as that observed in the crystal structures of IHF-DNA and competing conformations in which very likely either one kink site or the other is unkinked. The equilibrium distribution between these different conformations depends sensitively on DNA sequence, especially the A-tract on one side of the U-bend and the TTR consensus site on the other side. The "kinkability" at the kink sites also has a measurable effect on the distribution. Further studies of this nature would be very useful in

characterizing DNA sequences that render DNA highly “kinkable” and data such as these could be used to further refine models for sequence-dependent DNA deformability and protein-DNA interactions needed to stabilize distorted DNA conformations in complex with DNA-bending proteins.

References

1. Crick, F. H.; Klug, A., Kinky helix. *Nature* **1975,** *255* (5509), 530-3.
2. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **1997,** *389* (6648), 251-60.
3. Hogan, M. E.; Rooney, T. F.; Austin, R. H., Evidence for kinks in DNA folding in the nucleosome. *Nature* **1987,** *328* (6130), 554-7.
4. Richmond, T. J.; Davey, C. A., The structure of DNA in the nucleosome core. *Nature* **2003,** *423* (6936), 145-50.
5. Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E., Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **1995,** *376* (6543), 791-795.
6. Schultz, S. C.; Shields, G. C.; Steitz, T. A., Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **1991,** *253* (5023), 1001-7.
7. Kim, J. L.; Nikolov, D. B.; Burley, S. K., Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **1993,** *365* (6446), 520-7.
8. Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B., Crystal structure of a yeast TBP/TATA-box complex. *Nature* **1993,** *365* (6446), 512-520.
9. Rice, P. A.; Yang, S.; Mizuuchi, K.; Nash, H. A., Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **1996,** *87* (7), 1295-306.
10. Swinger, K. K.; Lemberg, K. M.; Zhang, Y.; Rice, P. A., Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J.* **2003,** *22* (14), 3749-3760.
11. Mouw, K. M.; Rice, P. A., Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending Hbb. *Molecular Microbiology* **2007,** *63*, 1319-1330.
12. Czapla, L.; Peters, J. P.; Rueter, E. M.; Olson, W. K.; Maher, L. J., 3rd, Understanding apparent DNA flexibility enhancement by HU and HMGB architectural proteins. *J Mol Biol* **2011,** *409* (2), 278-89.
13. Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B., DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U S A* **1998,** *95* (19), 11163-8.
14. Olson, W. K.; Zhurkin, V. B., Modeling DNA deformations. *Curr. Opin. Struct. Biol.* **2000,** *10* (3), 286-97.
15. Morozov, A. V.; Havranek, J. J.; Baker, D.; Siggia, E. D., Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res* **2005,** *33* (18), 5781-98.
16. Morozov, A. V.; Fortney, K.; Gaykalova, D. A.; Studitsky, V. M.; Widom, J.; Siggia, E. D., Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res* **2009,** *37* (14), 4707-22.
17. Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., 3rd; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P., Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys J* **2005,** *89* (6), 3721-40.
18. Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., 3rd; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J., A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res* **2010,** *38* (1), 299-313.
19. Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., 3rd; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; Mitchell, J.; Osman, R.; Orozco, M.; Perez, A.; Petkeviciute, D.; Spackova, N.; Sponer, J.; Zakrzewska, K.; Lavery, R., muABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res* **2014,** *42* (19), 12272-83.

20. Olson, W. K.; Zhurkin, V. B., Working the kinks out of nucleosomal DNA. *Curr Opin Struct Biol* **2011,** *21* (3), 348-57.
21. Lequieu, J.; Schwartz, D. C.; de Pablo, J. J., In silico evidence for sequence-dependent nucleosome sliding. *Proc Natl Acad Sci U S A* **2017,** *114* (44), E9197-E9205.
22. Chakraborty, D.; Hori, N.; Thirumalai, D., Sequence-dependent three interaction site model for single- and double-stranded DNA. *J. Chem. Theory. Comp.* **2018,** *14*, 3763-3779.
23. Swinger, K. K.; Rice, P. A., IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* **2004,** *14* (1), 28-35.
24. Wojtuszewski, K.; Hawkins, M. E.; Cole, J. L.; Mukerji, I., HU binding to DNA: evidence for multiple complex formation and DNA bending. *Biochemistry* **2001,** *40* (8), 2588-2598.
25. van Noort, J.; Verbrugge, S.; Goosen, N.; Dekker, C.; Dame, R. T., Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (18), 6969-74.
26. Erie, D. A.; Yang, G.; Schultz, H. C.; Bustamante, C., DNA bending by Cro protein in specific and nonspecific complexes: implications for protein site recognition and specificity. *Science* **1994,** *266* (5190), 1562-6.
27. Seong, G. H.; Kobatake, E.; Miura, K.; Nakazawa, A.; Aizawa, M., 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* **2002,** *291* (2), 361-6.
28. Kong, M.; Liu, L.; Chen, X.; Driscoll, K. I.; Mao, P.; Bohm, S.; Kad, N. M.; Watkins, S. C.; Bernstein, K. A.; Wyrick, J. J.; Min, J. H.; Van Houten, B., Single-molecule imaging reveals that Rad4 employs a dynamic DNA damage recognition process. *Mol Cell* **2016,** *64* (2), 376-387.
29. Wang, H.; Yang, Y.; Schofield, M. J.; Du, C.; Fridman, Y.; Lee, S. D.; Larson, E. D.; Drummond, J. T.; Alani, E.; Hsieh, P.; Erie, D. A., DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc. Natl. Acad. Sci. U S A* **2003,** *100* (25), 14822-7.
30. Buechner, C. N.; Maiti, A.; Drohat, A. C.; Tessmer, I., Lesion search and recognition by thymine DNA glycosylase revealed by single molecule imaging. *Nucleic Acids Res* **2015,** *43* (5), 2716-29.
31. Klostermeier, D.; Millar, D. P., Time-resolved fluorescence resonance energy transfer: a versatile tool for the analysis of nucleic acids. *Biopolymers* **2001,** *61* (3), 159-79.
32. Lyubchenko, Y. L.; Shlyakhtenko, L. S., Imaging of DNA and protein-DNA complexes with atomic force microscopy. *Crit Rev Eukaryot Gene Expr* **2016,** *26* (1), 63-96.
33. Beckwitt, E. C.; Kong, M.; Van Houten, B., Studying protein-DNA interactions using atomic force microscopy. *Semin Cell Dev Biol* **2018,** *73*, 220-230.
34. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat Methods* **2008,** *5* (6), 507-16.
35. Sass, L. E.; Lanyi, C.; Weninger, K.; Erie, D. A., Single-molecule FRET TACKLE reveals highly dynamic mismatched DNA-MutS complexes. *Biochemistry* **2010,** *49* (14), 3174-90.
36. Vafabakhsh, R.; Ha, T., Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **2012,** *337* (6098), 1097-101.
37. Singh, D.; Wang, Y.; Mallon, J.; Yang, O.; Fei, J.; Poddar, A.; Ceylan, D.; Bailey, S.; Ha, T., Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat Struct Mol Biol* **2018,** *25* (4), 347-354.
38. Schuler, B., Lipman, E.A., Eaton, W.A., Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **2002,** *419*, 743-747.
39. Lipman, E. A.; Schuler, B.; Bakajin, O.; Eaton, W. A., Single-molecule measurement of protein folding kinetics. *Science* **2003,** *301* (5637), 1233-5.
40. Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A., Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci U S A* **2007,** *104* (5), 1528-33.

41. Schuler, B.; Eaton, W. A., Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* **2008,** *18* (1), 16-26.
42. Gopich, I. V.; Szabo, A., Single-molecule FRET with diffusion and conformational dynamics. *J Phys Chem B* **2007,** *111* (44), 12925-32.
43. Chakraborty, S.; Steinbach, P. J.; Paul, D.; Mu, H.; Broyde, S.; Min, J. H.; Ansari, A., Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. *Nucleic Acids Res* **2018,** *46* (3), 1240-1255.
44. Nash, H. A., *Regulation of Gene Expression in E. coli*. Chapman and Hall: New York, 1996.
45. Craig, N. L.; Nash, H. A., E. coli integration host factor binds to specific sites in DNA. *Cell* **1984,** *39* (3 Pt 2), 707-16.
46. Yang, C. C.; Nash, H. A., The interaction of *E. coli* IHF protein with its specific binding sites. *Cell* **1989,** *57* (5), 869-880.
47. Goodrich, J. A.; Schwartz, M. L.; McClure, W. R., Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). *Nucleic Acids Res.* **1990,** *18* (17), 4993-5000.
48. Engelhorn, M.; Boccard, F.; Murtin, C.; Prentki, P.; Geiselmann, J., In vivo interaction of the Escherichia coli integration host factor with its specific binding sites. *Nucleic Acids Res* **1995,** *23* (15), 2959-65.
49. Hales, L. M.; Gumport, R. I.; Gardner, J. F., Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J. Bacteriol.* **1994,** *176* (10), 2999-3006.
50. Hales, L. M.; Gumport, R. I.; Gardner, J. F., Examining the contribution of a dA+dT element to the conformation of Escherichia coli integration host factor-DNA complexes. *Nucleic Acids Res.* **1996,** *24* (9), 1780-6.
51. Holbrook, J. A.; Tsodikov, O. V.; Saecker, R. M.; Record, M. T., Jr., Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *Journal of molecular biology.* **2001,** *310*, 379-401.
52. Vander Meulen, K. A.; Saecker, R. M.; Record, M. T., Jr., Formation of a wrapped DNA-protein interface: experimental characterization and analysis of the large contributions of ions and water to the thermodynamics of binding IHF to H' DNA. *J. Mol. Biol.* **2008,** *377* (1), 9-27.
53. Rice, P. A.; Correll, C. C., Introduction. In *Protein-Nucleic Acid Interactions*, Rice, P. A.; Correll, C. C., Eds. Royal Society of Chemistry: Cambridge, 2008.
54. Lawson, C. L.; Berman, H. M., Indirect readout of DNA sequence by proteins. In *Protein-Nucleic Acid Interactions*, Rice, P. A.; Correll, C. C., Eds. Royal Society of Chemistry: Cambridge, 2008.
55. Swinger, K. K.; Rice, P. A., Structure-based analysis of HU-DNA binding. *J. Mol. Biol.* **2007,** *365* (4), 1005-16.
56. Grove, A.; Galeone, A.; Mayol, L.; Geiduschek, E. P., Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *Journal of molecular biology.* **1996,** *260* (2), 120-125.
57. Lynch, T. W.; Read, E. K.; Mattis, A. N.; Gardner, J. F.; Rice, P. A., Integration host factor: putting a twist on protein-DNA recognition. *Journal of molecular biology.* **2003,** *330* (3), 493-502.
58. Laxmikanthan, G.; Xu, C.; Brilot, A. F.; Warren, D.; Steele, L.; Seah, N.; Tong, W.; Grigorieff, N.; Landy, A.; Van Duyne, G. D., Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *Elife* **2016,** *5*.
59. Vivas, P.; Velmurugu, Y.; Kuznetsov, S. V.; Rice, P. A.; Ansari, A., Mapping the transition state for DNA bending by IHF. *J Mol Biol* **2012,** *418* (5), 300-15.
60. Sugimura, S.; Crothers, D. M., Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (49), 18510-4.

61. Sugimura, S. Kinetic and steady-state studies of binding and bending of lambda phage DNA by Integration Host Factor. Ph.D. thesis, Yale University, New Haven, CT, 2005.
62. Lorenz, M.; Hillisch, A.; Goodman, S. D.; Diekmann, S., Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence resonance energy transfer. *Nucleic Acids Res.* **1999,** *27* (23), 4619-4625.
63. Kuznetsov, S. V.; Sugimura, S.; Vivas, P.; Crothers, D. M.; Ansari, A., Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (49), 18515-20.
64. Vivas, P.; Kuznetsov, S. V.; Ansari, A., New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. *J Phys Chem B* **2008,** *112* (19), 5997-6007.
65. Velmurugu, Y.; Vivas, P.; Connolly, M.; Kuznetsov, S. V.; Rice, P. A.; Ansari, A., Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein. *Nucleic Acids Res* **2018,** *46* (4), 1741-1755.
66. Tsodikov, O. V.; Holbrook, J. A.; Shkel, I. A.; Record, M. T., Jr., Analytic binding isotherms describing competitive interactions of a protein ligand with specific and nonspecific sites on the same DNA oligomer. *Biophys J* **2001,** *81* (4), 1960-9.
67. Aeling, K. A.; Opel, M. L.; Steffen, N. R.; Tretyachenko-Ladokhina, V.; Hatfield, G. W.; Lathrop, R. H.; Senear, D. F., Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy. *J. Biol. Chem.* **2006,** *281* (51), 39236-48.
68. Lohman, T. M.; Mascotti, D. P., Thermodynamics of ligand-nucleic acid interactions. *Methods Enzymol.* **1992,** *212*, 400-24.
69. Record, M. T., Jr.; Anderson, C. F.; Lohman, T. M., Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* **1978,** *11* (2), 103-78.
70. Manning, G. S., The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **1978,** *11* (2), 179-246.
71. Jen-Jacobson, L.; Jacobson, L. A., Role of water and effects of small ions in site-specific protein-DNA interactions. In *Protein-Nucleic Acid Interactions*, Rice, P. A.; Cornell, C. C., Eds. Royal Society of Chemistry: Cambridge, 2008; pp 13-46.
72. Vivas, P.; Velmurugu, Y.; Kuznetsov, S. V.; Rice, P. A.; Ansari, A., Global analysis of ion dependence unveils hidden steps in DNA binding and bending by integration host factor. *J Chem Phys* **2013,** *139* (12), 121927.
73. Engler, L. E.; Sapienza, P.; Dorner, L. F.; Kucera, R.; Schildkraut, I.; Jen-Jacobson, L., The energetics of the interaction of BamHI endonuclease with its recognition site GGATCC. *J Mol Biol* **2001,** *307* (2), 619-36.
74. Engler, L. E.; Welch, K. K.; Jen-Jacobson, L., Specific binding by EcoRV endonuclease to its DNA recognition site GATATC. *J Mol Biol* **1997,** *269* (1), 82-101.
75. Lesser, D. R.; Kurpiewski, M. R.; Jen-Jacobson, L., The energetic basis of specificity in the Eco RI endonuclease--DNA interaction. *Science* **1990,** *250* (4982), 776-86.
76. Record, M. T., Jr.; Anderson, C. F.; Mills, P.; Mossing, M.; Roe, J. H., Ions as regulators of protein-nucleic acid interactions in vitro and in vivo. *Adv Biophys* **1985,** *20*, 109-35.
77. Lorenz, M.; Hillisch, A.; Payet, D.; Buttinelli, M.; Travers, A.; Diekmann, S., DNA bending induced by high mobility group proteins studied by fluorescence resonance energy transfer. *Biochemistry* **1999,** *38* (37), 12150-8.
78. Friedman, D. I., Integration host factor: a protein for all reasons. *Cell* **1988,** *55* (4), 545-54.
79. Landy, A., Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **1989,** *58*, 913-49.

80. Travers, A., DNA-protein interactions: IHF--the master bender. *Curr. Biol.* **1997,** *7* (4), R252-4.
81. Goodman, S. D.; Velten, N. J.; Gao, Q.; Robinson, S.; Segall, A. M., In vitro selection of integration host factor binding sites. *J Bacteriol* **1999,** *181* (10), 3246-55.
82. Steffen, N. R.; Murphy, S. D.; Tolleri, L.; Hatfield, G. W.; Lathrop, R. H., DNA sequence and structure: direct and indirect recognition in protein-DNA binding. *Bioinformatics* **2002,** *18 Suppl 1*, S22-30.
83. Aeling, K. A.; Steffen, N. R.; Johnson, M.; Hatfield, G. W.; Lathrop, R. H.; Senear, D. F., DNA deformation energy as an indirect recognition mechanism in protein-DNA interactions. *IEEE/ACM Trans Comput Biol Bioinform* **2007,** *4* (1), 117-25.
84. Khrapunov, S.; Brenowitz, M.; Rice, P. A.; Catalano, C. E., Binding then bending: A mechanism for wrapping DNA. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (51), 19217-8.
85. Mengeritsky, G.; Goldenberg, D.; Mendelson, I.; Giladi, H.; Oppenheim, A. B., Genetic and biochemical analysis of the integration host factor of Escherichia coli. *J Mol Biol* **1993,** *231* (3), 646-57.
86. Wang, S.; Cosstick, R.; Gardner, J. F.; Gumport, R. I., The specific binding of Escherichia coli integration host factor involves both major and minor grooves of DNA. *Biochemistry* **1995,** *34* (40), 13082-90.
87. Yang, S. W.; Nash, H. A., Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. *Embo. J.* **1995,** *14* (24), 6292-300.
88. Ali, B. M.; Amit, R.; Braslavsky, I.; Oppenheim, A. B.; Gileadi, O.; Stavans, J., Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc. Natl. Acad. Sci. U S A* **2001,** *98* (19), 10658-10663.
89. Lee, E. C.; MacWilliams, M. P.; Gumport, R. I.; Gardner, J. F., Genetic analysis of Escherichia coli integration host factor interactions with its bacteriophage lambda H' recognition site. *J Bacteriol* **1991,** *173* (2), 609-17.
90. Pinson, V.; Takahashi, M.; Rouviere-Yaniv, J., Differential binding of the *Escherichia coli* HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. *J. Mol. Biol.* **1999,** *287* (3), 485-497.
91. Castaing, B.; Zelwer, C.; Laval, J.; Boiteux, S., HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J. Biol. Chem.* **1995,** *270* (17), 10291-10296.
92. Xiao, B.; Johnson, R. C.; Marko, J. F., Modulation of HU-DNA interactions by salt concentration and applied force. *Nucleic Acids Res* **2010,** *38* (18), 6176-85.
93. Lin, J.; Chen, H.; Droge, P.; Yan, J., Physical organization of DNA by multiple non-specific DNA-binding modes of integration host factor (IHF). *PLoS One* **2012,** *7* (11), e49885.
94. Sagi, D.; Friedman, N.; Vorgias, C.; Oppenheim, A. B.; Stavans, J., Modulation of DNA conformations through the formation of alternative high-order HU-DNA complexes. *J. Mol. Biol.* **2004,** *341* (2), 419-28.
95. Coats, J. E.; Lin, Y.; Rueter, E.; Maher, L. J., 3rd; Rasnik, I., Single-molecule FRET analysis of DNA binding and bending by yeast HMGB protein Nhp6A. *Nucleic Acids Res* **2013,** *41* (2), 1372-81.
96. Murugesapillai, D.; McCauley, M. J.; Maher, L. J., 3rd; Williams, M. C., Single-molecule studies of high-mobility group B architectural DNA bending proteins. *Biophys Rev* **2017,** *9* (1), 17-40.
97. Graham, J. S.; Johnson, R. C.; Marko, J. F., Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic Acids Res* **2011,** *39* (6), 2249-59.
98. Joshi, C. P.; Panda, D.; Martell, D. J.; Andoy, N. M.; Chen, T. Y.; Gaballa, A.; Helmann, J. D.; Chen, P., Direct substitution and assisted dissociation pathways for turning off transcription by a MerR-family metalloregulator. *Proc Natl Acad Sci U S A* **2012,** *109* (38), 15121-6.
99. Loparo, J. J.; Kulczyk, A. W.; Richardson, C. C.; van Oijen, A. M., Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proc Natl Acad Sci U S A* **2011,** *108* (9), 3584-9.

100. Giuntoli, R. D.; Linzer, N. B.; Banigan, E. J.; Sing, C. E.; de la Cruz, M. O.; Graham, J. S.; Johnson, R. C.; Marko, J. F., DNA-segment-facilitated dissociation of Fis and Nhp6A from DNA detected via single-molecule mechanical response. *J Mol Biol* **2015,** *427* (19), 3123-36.
101. Hadizadeh, N.; Johnson, R. C.; Marko, J. F., Facilitated dissociation of a nucleoid protein from the bacterial chromosome. *J Bacteriol* **2016,** *198* (12), 1735-42.
102. Sing, C. E.; Olvera de la Cruz, M.; Marko, J. F., Multiple-binding-site mechanism explains concentration-dependent unbinding rates of DNA-binding proteins. *Nucleic Acids Res* **2014,** *42* (6), 3783-91.
103. Cocco, S.; Marko, J. F.; Monasson, R., Stochastic ratchet mechanisms for replacement of proteins bound to DNA. *Phys Rev Lett* **2014,** *112* (23), 238101.

Chapter 4.

Bind… and Bend? Characterization of specific and nonspecific binding by Integration Host Factor in steady-state fluorescence studies with analytical binding model

4.1. Introduction

Integration Host Factor (IHF) is a versatile protein that can fulfill a variety of roles in *E. coli* because of its multiple binding modes. IHF exhibits two primary modes of binding: specific - in which it severely bends specific sites on the DNA into a "U-turn" conformation by introducing two sharp kinks into the DNA, and nonspecific – where the degree of bending is likely much less severe (1). In its nonspecific binding mode, IHF mainly helps to package the DNA inside the cell. In its specific binding mode, it plays an important architectural role in making higher order nucleoprotein complexes needed for site-specific recombination, DNA replication and transcription, and CRISPR genome editing systems (2-12).

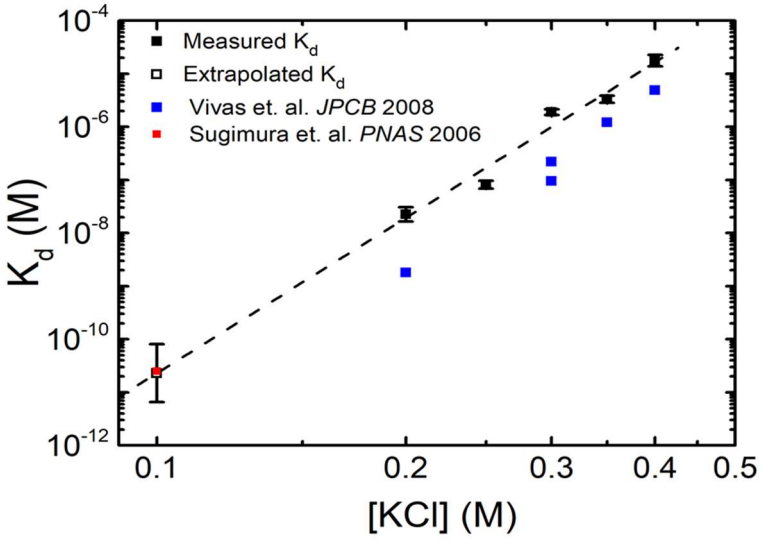
The crystal structure of IHF bound in a specific complex was solved using the 35-bp H' site, which is one of the sites on lambda DNA to which IHF binds (1). The crystal structure of this protein-DNA complex is shown in Figure 1.7. The DNA in the complex is shown to adopt a U-turn conformation, which involves sharp kinks at two sites, spaced 9-bps apart. This severely distorted DNA is stabilized by proline residues that intercalate at the kink sites, and protein-DNA contacts made with the flanking arms of the DNA on either side. The crystal structure provides one snapshot of the IHF-H' specific complex. Fluorescence lifetime measurements made on the IHF-H' complex reveal a very dynamic complex, with more than one conformation accessible to the specific complex at equilibrium (Chapter 3). IHF can also bind nonspecifically to the H' site, with significant competition from nonspecific binding observed at high ratios of protein to DNA concentrations (13, 14). Indeed, recent AFM studies revealed partially bent nonspecific complexes

of IHF-DNA in more than one binding mode and indicated a dynamic equilibrium between specific and nonspecific binding (15). The ability of IHF to bind DNA in multiple modes to a single DNA site poses a simple question: what are the relative binding affinities in the two binding modes?

During the lifetime of *E. coli*, IHF concentrations change dramatically, significantly altering the equilibrium between specific and nonspecific binding. Reports suggest that IHF concentrations may vary from 12-55 μM depending on the phase of the cell's life cycle (16). Indeed, previous IHF binding studies with H' showed that increasing IHF concentration favors nonspecific binding within this range (14, 17, 18). Thus, accurate quantification of the binding affinities of IHF in its specific and nonspecific binding modes and the conditions under which one binding mode dominates or another is needed to know the cellular consequences of IHF concentration fluctuations as well as the effect of other solution conditions such as the ionic strength. Yet, despite numerous binding studies of IHF with specific and nonspecific sites, the binding affinity of IHF for specific versus nonspecific sites remain poorly determined, with wide disagreements in the binding affinities reported by various groups.

The bulk of binding studies on the IHF-H' complex in ~60-100 mM monovalent salt conditions have yielded specific dissociation constants (K_d^{sp}) in the range of 1-100 nM, from measurements of equilibrium binding isotherms using a variety of techniques that include gel-shift assays (19-23), isothermal titration calorimetry (ITC) (13, 14, 17), fluorescence assays (24, 25).

In contrast, Sugimura and Crothers first reported $K_d^{sp} \approx 25$ pM from direct measurements of the association and dissociations rates of the IHF-H' complex from stopped-flow experiments at 100 mM KCl (26). Follow up studies in our group have confirmed the 25 pM value of Sugimura and Crothers (Figure 4.1.) (18, 27, 28) where binding isotherms were measured at higher salt concentrations to bring the K_d^{sp} values in the range $K_d^{sp} \geq$ [DNA], and then extrapolated to 100 mM KCl.



4.1. Binding affinity of IHF-H' complex at 100 mM KCl from extrapolation of high-salt measurements. The measured dissociation constants of the IHF-H' complex are plotted as a function of the KCl concentration; both x- and y-axes are on a logarithmic scale. Measurements made in this study (black) are compared to earlier measurements from (red) Sugimura and Crothers (26) and (blue) Vivas et. al. (28). The most recent data (in black), when extrapolated to 100 mM KCl assuming a linear dependence of log(K_d) versus log([KCl]), overlaps with the measured data from Sugimura and Crothers. Error of extrapolated fit estimated by K_d values at which χ^2 changes by 1 (described in 4.2.3.).

The report by Sugimura and Crothers indicated that the much higher K_d^{sp} values measured in previous studies may be due to the conditions of the experimental studies. For example, DNA concentrations greater than specific binding dissociation constant values, like the >1-10 nM DNA

concentrations needed for many equilibrium binding assays, hamper the ability of these binding assays to accurately measure sub-nanomolar K_d^{sp} values. These techniques include ITC, that require μM concentrations or greater for DNA and IHF (14, 29), and fluorescence-based titration assays that typically require at least nM concentrations of DNA (18, 24, 28, 30). Discrepancies in EMSA gel shift results which can use pM DNA concentrations are more puzzling. Gel shift studies have captured pM binding affinities for HU-DNA complexes, a complex of similar structure and function (31). Gel shift studies by Martin and coworkers using 50 pM DNA yielded IHF-H' K_d^{sp} values to be ~18 nM (32) and studies by Rice and coworkers using 5 pM DNA concentrations yielded IHF-H' binding affinities of ~2 nM (20). Indeed, we performed EMSA studies on the IHF-H' performed which yielded >1 nM binding affinities (Chapter 6). Though not completely understood, K_d values from gel shift assays that differ from other measurement techniques have been reported (33-35).

Another factor that could contribute to the wide variations in the reported values of K_d^{sp} is the contribution from nonspecific binding. While most binding studies on IHF-DNA complexes have been done with DNA oligomers that are approximately the same length as the specific binding site size and analyzed in terms of 1:1 binding isotherms, nonspecific binding can be a significant factor if the specificity ratio is not very large. Indeed, Record and co-workers, in a series of ITC measurements on the IHF-H' complex, clearly demonstrated that increasing concentrations of [IHF]/[DNA] favored multiple protein molecules binding nonspecifically over one protein binding specifically (17, 19, 36). This phenomenon was immediately evident in their so-called "reverse titration" measurements (Figure 4.2.), in which a fixed protein concentration was titrated against varying DNA concentration, starting from low [DNA] and therefore high [IHF]/[DNA] ratio that gradually decreased as the [DNA] increased. The binding isotherm in this case deviated

dramatically from a 1:1 binding profile. In contrast, "forward titration" measurements, in which a fixed DNA concentration was titrated against varying protein concentration, did not show as clear a signature of nonspecific binding (14). Notably, most titration studies, including ours, are done as forward titration, where the binding isotherms appear to follow a 1:1 binding profile even when nonspecific binding is a factor.

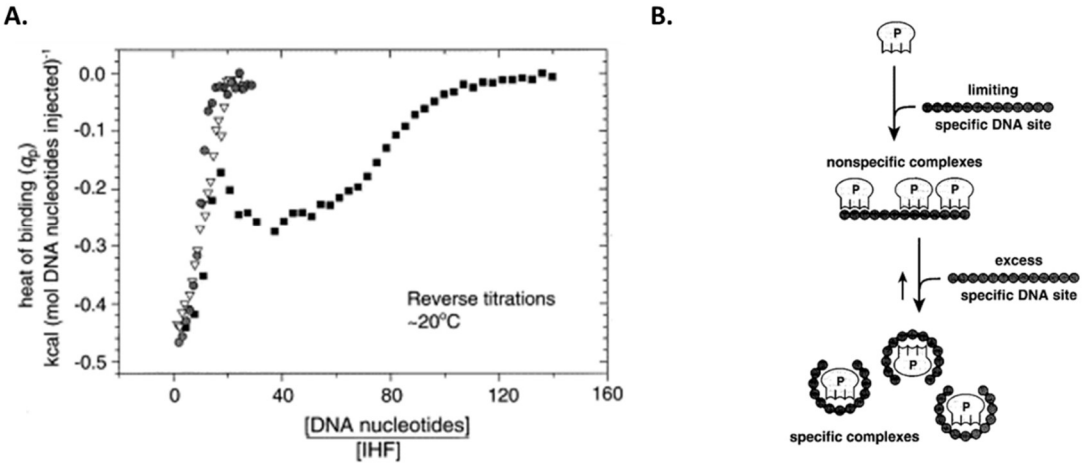


Figure 4.2. Reverse isothermal calorimetry (ITC) titrations and suggested model representation (reproduced from Holobrook et. al. (14), with permission). (A) Heat of binding per mol nucleotide injected at 60 mM K^+, for reverse titrations of 20 mM H′ DNA (nucleotide) into 22 μM IHF at 19.7 °C (■), 3.4 mM 14mer DNA (nucleotide) into 28 μM IHF at 19.8 °C (▿), and 2.8 mM calf thymus DNA (nucleotide) into 15 μM IHF at 19.6 °C (•), all plotted as functions of the ratio of [DNA nucleotides]/[IHF]. (B) IHF protein (P) is sketched with a much smaller site size for non-specific binding than for specific binding. In the initial (DNA limiting) stage of the reverse titration, if the specificity ratio is not very large, multiple proteins bind non-specifically to the specific site. Later in the titration, where DNA is in excess, the majority of proteins are present in specific 1:1 complexes.

To take into account competition between specific and nonspecific binding, Record and co-workers derived analytic expressions for binding isotherms to short DNA oligomers that included both contributions (17), starting from the nonspecific binding model to an infinite one-dimensional lattice developed by McGhee and von Hippel (37). The Record group's model successfully captured their ITC isotherms for both forward and reverse titrations and yielded

specific and nonspecific binding affinities for IHF-H' and the nonspecific binding site size for a range of salt conditions (14). However, their K_d^{sp} from these studies were still high (~70 nM at 100 mM KCl), indicating that these studies also could not capture the pM binding affinity reported by the stopped-flow studies of the Crothers group and "salt-extrapolation" studies from our group.

In this study, we revisit the "three-state" lattice model for competitive binding developed by Record and co-workers and apply it to binding isotherms on IHF-H' measured using two fluorescence-based methods, anisotropy and acceptor ratio (analogous to FRET), that simultaneously monitor the extent of IHF binding and DNA bending, respectively. We performed our measurements at 200 mM KCl, which is sufficiently high ionic strength to bring the K_d^{sp} to a measurable range (~1-35 nM) using equilibrium fluorescence methods, as shown by previous studies in our group (18, 27, 28), while still retaining competition from nonspecific binding as a significant factor at high protein concentrations.

This three-state model, with DNA that is either (i) unbound, (ii) has one protein bound specifically, or (iii) one or more proteins bound nonspecifically, can simultaneously fit four binding isotherms obtained from both acceptor ratio and anisotropy measurements done on specific (H') and nonspecific DNA. However, some discrepancies between the data and the fits at extremely high [IHF]:[DNA] ratios (more than 200:1) suggest that this three-state model description, in which nonspecific protein binding displaces a specifically bound protein, may be incomplete. Our data suggest an additional state, where a specifically bound protein could also accommodate a nonspecifically bound protein. Our study elucidates the added complexities of IHF-DNA interactions that were not considered in previous descriptions and lays the foundation for further examining higher order complexes at high protein-to-DNA ratios that may be present *in vivo*.

4.2. Materials and Methods

4.2.1. Materials

DNA and IHF samples were prepared as described in Chapter 2.1.1. DNA was end-labeled with fluorescein/TAMRA FRET pair (Table 2.1. – Design I). Sequences are shown in Figure 4.3.

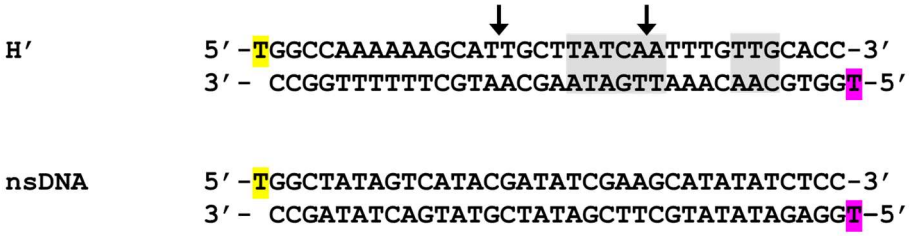


Figure 4.3. Specific and nonspecific DNA sequences. H' DNA which contains IHF consensus sequence (highlighted in gray) is sharply kinked at the two sites indicated by the black arrows. FRET labels fluorescein and TAMRA are highlighted in yellow and magenta, respectively. Nonspecific DNA (nsDNA) contains same nucleotides as H' but the sequence order is scrambled to remove the consensus site.

4.2.2. Equilibrium Binding Isotherms

The steady-state fluorescence emission spectra and anisotropies were measured on a FluoroMax4 spectrofluorometer (Jobin Yvon, Inc., NJ, USA), with samples loaded in a 100-µL quartz cuvette (Starna 26.100F-Q-10/Z20). Equilibrium fluorescence measurements were performed with donor-acceptor labeled DA constructs (denoted as DNA_DA). All measurements were performed using 50 nM [DNA] at 20 °C with protein concentrations ranging from 1 nM – 50 µM. The binding isotherms were measured in IHF binding buffer (Chapter 2.1.) with 200 mM KCl salt conditions.

4.2.3. Monte Carlo search in parameter space for fitting of model to data

Fitting was carried out using Monte-Carlo search in parameter space. Starting parameters for the Monte Carlo search were randomly chosen from a wide range and the residuals were minimized by a simulated annealing procedure (38-40). We performed 100 independent, randomly chosen sets as starting points with 30 rounds of simulated annealing for each set. The optimal set of parameters obtained by this procedure for each independent set of randomly chosen parameter values was then used as starting values for further minimization using a least-squares non-linear fitting procedure implemented in MATLAB (R2015b).

4.2.3. Error estimation of fitting parameters

To estimate uncertainties in the fitting parameters, each fitting parameter was set to fixed amount, one at a time, and refitted the binding isotherms by allowing the rest of the parameters to vary to minimized the chi-square (χ^2) defined as:

$$\chi^2 = \sum_{i}^{n} \frac{(O_i - E_i)^2}{\sigma_i^2} \qquad \text{Eq. 4.1.}$$

Where *i* is a single measurement out of *n* measurements, O_i is the observed value, E_i is the expected value from the fit, and σ_i is the standard deviation in the data. We then generated χ^2 vs parameter value plots, from which the uncertainties in the parameters were identified as the values at which χ^2 changes by 1; this increase, $\Delta\chi^2 = 1$, corresponds to one standard deviation in parameter space (41).

4.3. Lattice model for specific and nonspecific binding of protein to DNA

4.3.1. Case I: Nonspecific binding to a finite lattice

We model nonspecific binding of proteins to DNA using a finite lattice model first developed by Record and coworkers (42) as an extension from the infinite lattice developed by McGhee and von Hipple (MvH) (37). DNA is treated as a linear array of length N, where N is the number of base pairs (bps) in the DNA substrate. For all sequences studied here, N = 35 bp. We model the protein binding site size to be n, described as the number of lattice points occluded by a single nonspecifically bound protein. We force n to be an integer value, and the maximum number of nonspecific proteins to bind is given by the largest integer value less than $P_{max} = N/n$. Nonspecific binding strength, described by the dissociation constant K_d^{ns}, treats all lattice sites to be thermodynamically equivalent and independent of number of previously bound proteins or pattern of binding to the lattice. This model has been previously applied to the IHF-H' binding studies (14, 17, 21).

If each bound protein occupies exactly n consecutive lattice sites, the number of distinct conformational arrangements (or multiplicity) for j nonspecifically bound proteins, Ω_j is:

$$\Omega_j = \frac{(N-n-1)j!}{j!\,(N-nj)!} \qquad \text{Eq. 4.2.}$$

where j can be an integer value up to P_{max}. Thus, we can write the partition function for nonspecific binding to a lattice in the absence of any specific binding to be:

$$Z_{2s} = 1 + \sum_{j=1}^{P_{max}} z_{ns,j} = 1 + \sum_{j=1}^{P_{max}} \frac{\Omega_j [L]}{K_d^{ns}} \qquad \text{Eq. 4.3.}$$

where $z_{ns,j}$ is the statistical weight of DNA that has j ligands bound nonspecifically

and $[L]$ is the ligand (or protein) concentration in solution (See 4.3.3.). With this formalism, the fractions of free DNA (f_{free}) and of DNA with j ligands (proteins) bound nonspecifically ($f_{ns,j}$) are readily written as:

$$f_{free}^{2s} = \frac{1}{Z_{2s}} \qquad \text{Eq. 4.4.}$$

$$f_{ns,j}^{2s} = \frac{z_{ns,j}}{Z_{2s}} \qquad \text{Eq. 4.5.}$$

4.3.2. Case II: Competitive specific and nonspecific binding to a finite lattice

Similar to how Record and co-workers considered the specific binding of IHF to a lattice (Case 1 in (17)), we considered the case of ligand binding to a finite lattice containing one specific site and multiple nonspecific sites in which the specific binding site size (s) is the same as the lattice size (N), while the nonspecific binding site size (n) is much less than N ($n \ll s = N$). Under these conditions, where the DNA can accommodate either one specifically bound protein or one or more nonspecifically bound protein, the partition function for this "three-state" scenario, becomes:

$$Z_{3s} = 1 + z_{sp} + \sum_{j=1}^{P_{max}} z_{ns,j} = 1 + \frac{[L]}{K_d^{sp}} + \sum_{j=1}^{P_{max}} \frac{\Omega_j [L]}{K_d^{ns}} \qquad \text{Eq. 4.6.}$$

Where $z_{sp} = [L]/K_d^{sp}$ is the statistical weight for the specifically bound fraction f_{sp}. The fractional population of free (f_{free}^{3s}), nonspecifically bound with j ligands ($f_{ns,j}^{3s}$), and specifically bound DNA (f_{sp}^{3s}) is given by:

$$f_{free}^{3s} = \frac{1}{Z_{3s}} \quad \text{Eq. 4.7.}$$

$$f_{ns,j}^{3s} = \frac{z_{ns,j}}{Z_{3s}} \quad \text{Eq. 4.8.}$$

$$f_{sp}^{3s} = \frac{z_{sp}}{Z_{3s}} \quad \text{Eq. 4.9.}$$

4.3.3. Relating total protein and DNA concentrations to free ligand concentration

The protein concentration $[L]$ in Eqs. 4.3. and 4.6. is the free protein concentration (also denoted P_{free}), which depends on how much protein is bound to DNA. To obtain the free protein concentration from the known total protein and DNA concentrations at each set of conditions, we used the following closed-form expressions obtained by Record and co-workers (17) that related the total protein concentration P_0, and the concentrations of nonspecifically and specifically bound proteins P_b^{ns} and P_b^{sp}, respectively, for Case II above, to the nonspecific binding density ν_{ns} and conditional probability ff (described below in Eqs. 4.17. and 4.18. respectively), the total DNA concentration D_0, the dissociation constants K_d^{ns} and K_d^{sp}, the nonspecific binding site size n and the length of the DNA oligomer N as:

$$P_0 = \frac{Nv_{ns}K_d^{ns}}{(N-n+1)(ff)^{n-1}(1-nv_{ns})} + ND_0v_{ns}\left(\frac{K_d^{sp}(N-n+1)+K_d^{ns}(ff)^{N-n}}{K_d^{sp}(N-n+1)+Nv_{ns}K_d^{ns}(ff)^{N-n}}\right) \quad \text{Eq. 4.10.}$$

$$P_b^{ns} = \frac{Nv_{ns}D_0}{1+\frac{NK_d^{ns}}{K_d^{sp}(N-n+1)}v_{ns}(ff)^{N-n}} \quad \text{Eq. 4.11.}$$

$$P_b^{sp} = \frac{K_d^{ns}Nv_{ns}D_0\,(ff)^{N-n}}{K_d^{sp}(N-n+1)+NK_d^{ns}v_{ns}(ff)^{N-n}} \quad \text{Eq. 4.12.}$$

The free ligand (protein) concentration for Case II equations was then obtained as

$$[L] = P_{free} = P_0 - P_b^{ns} - P_b^{sp} \quad \text{Eq. 4.13.}$$

The corresponding equations for Case I (no specific binding) were obtained from Eqs. 4.10.-4.12. by making K_d^{sp} infinitely large, which yielded:

$$P_0 = \frac{Nv_{ns}K_d^{ns}}{(N-n+1)(ff)^{n-1}(1-nv_{ns})} + ND_0v_{ns} \quad \text{Eq. 4.14.}$$

$$P_b^{ns} = Nv_{ns}D_0 \quad \text{Eq. 4.15.}$$

$$[L] = P_{free} = P_0 - P_b^{ns} \quad \text{Eq. 4.16.}$$

The nonspecific binding density v_{ns} describes the number of ligands bound per lattice unit, which was first derived by McGee and von Hipple (37) for an infinite lattice and later derived for a finite lattice by Record and coworkers (17). The expression for v_{ns} for the finite lattice is reproduced here as:

$$\frac{v_{ns}K_d^{ns}}{[L]} = (1 - nv_{ns})(ff)^{n-1}\left(\frac{N-n+1}{N}\right) \quad \text{Eq. 4.17.}$$

Where ff is the conditional probability defined as:

$$ff = \frac{1 - nv_{ns}}{1 - (n-1)v_{ns}} \quad \text{Eq. 4.18.}$$

The conditional probability describes the likelihood that the site adjacent to a bound lattice site is free. For a finite lattice, the dependence of this conditional probability on binding density is a function of the lattice size N. However, Record and Coworkers show, to a good approximation show that the infinite-lattice conditional probability $(ff)^{n-1}$ is applicable to all interior units on a finite lattice especially at low enough lattice saturation (v_{ns}) (17). v_{ns} was solved numerically for each total protein concentration, P_0.

4.4 Lattice model applied to IHF-DNA acceptor ratio and anisotropy binding isotherms

4.4.1 IHF binding to nonspecific DNA: two-state description

To describe IHF binding to nonspecific DNA, we separated two primary populations of DNA that were assumed distinguishable with our fluorescent probes: free DNA and nonspecifically-bound complex, with acceptor ratio levels Υ_{free} and Υ_{ns}, respectively, and anisotropy levels r_{free}, and r_{ns}, respectively. We further assumed that the acceptor ratio for the nonspecific complex (Υ_{ns}) was insensitive to the number of bound proteins.

However, the anisotropy for the nonspecific complex was parameterized as increasing with the number of bound proteins, j = 1,2,.., P_{max}, by the Perrin equation (43):

$$r_{ns,j} = \frac{r_0}{1+\frac{\tau RT}{\eta V_j}} \quad \text{Eq. 4.19.}$$

where $r_{ns,j}$ is the anisotropy of the complex with *j* protein molecules bound, r_0 is the anisotropy constant that would be observed in the absence of any depolarizing process (intrinsic to the fluorophore), τ is the lifetime of the fluorophore, R is the molar gas constant, T is the temperature, η is the viscosity of the solvent, and V_j is the molar volume of the complex with *j* proteins bound. If we assume the molecular density, ρ, to be constant for all complexes, then V_j is proportional to the molecular weight of the complex, which is equal to the sum of the molecular weight of 35 bp duplex DNA substrate, MW_D ~22 kDa, and the molecular weight of each bound IHF protein, MW_P ~22 kDa. Thus, we rewrote the Perrin equation as:

$$r_{ns,j} = r_{ns,1}\frac{1+\frac{r_\eta}{(MW_D+MW_P)}}{1+\frac{r_\eta}{(MW_D+jMW_P)}} \quad \text{Eq. 4.20.}$$

where $r_{ns,1}$ is the anisotropy of the complex with only one protein bound and $r_\eta = \frac{\tau\rho RT}{\eta}$ is a free parameter.

The acceptor ratio and anisotropy data measured directly on the IHF-nsDNA nonspecific complex (denoted as Y_{exp}^{ns} and r_{exp}^{ns}, respectively), were described by the following equations:

$$Y_{exp}^{ns} = \Upsilon_{free}(f_{free}) + \Upsilon_{ns}(1 - f_{free}) \quad \text{Eq. 4.21.}$$

$$r_{exp}^{ns} = r_{free}(f_{free}) + \sum_{j=1}^{P_{max}} r_{ns,j}(f_{ns,j}) \qquad \text{Eq. 4.22.}$$

where the fractional populations f_{free} and $f_{ns,j}$ were obtained from Eqs. 4.4. and 4.5.

4.4.2 IHF binding to specific DNA: three-state description

For IHF binding to specific DNA, we now need to consider three primary populations of DNA: free DNA (D_{free}), nonspecifically-bound complex (D_b^{ns}), and specifically bound complex (D_b^{sp}), with acceptor ratio levels Υ_{free}, Υ_{ns}, and Υ_{sp}, respectively, and anisotropy levels r_{free}, r_{ns}, and r_{sp}, respectively. We further assume that the anisotropy of the specific complex, which has only one protein bound, is the same as the anisotropy of the nonspecific complex with only one protein bound; i.e. $r_{sp} = r_{ns,1}$.

The acceptor ratio and anisotropy data measured directly on the IHF-H' specific complex (denoted as Y_{exp}^{sp} and r_{exp}^{sp}) were fitted using the following equations:

$$Y_{exp}^{sp} = \Upsilon_{free}(f_{free}) + \Upsilon_{sp}(f_{sp}) + \Upsilon_{ns}(1 - f_{free} - f_{sp}) \qquad \text{Eq. 4.23.}$$

$$r_{exp}^{sp} = r_{free}(f_{free}) + r_{sp}(f_{sp}) + \sum_{j=1}^{P_{max}} r_{ns,j}(f_{ns,j}) \qquad \text{Eq. 4.24.}$$

where the fractional populations f_{free}, $f_{ns,j}$, and f_{sp} were obtained from Eqs.4.7-4.9.

In the description of the four binding isotherms - Υ_{exp}^{ns}, r_{exp}^{ns}, Υ_{exp}^{sp}, and r_{exp}^{sp} - there were a total of 9 free parameters: the dissociation constants K_d^{ns} and K_d^{sp}, the nonspecific binding site size n, the acceptor ratio parameters, Υ_{free}, Υ_{ns}, Υ_{sp}, and the anisotropy parameters r_{free}, $r_{sp} = r_{ns,1}$, and r_{η}.

4.4.3. Results of simultaneous fitting of IHF binding isotherms measured with nonspecific and specific DNA constructs.

Model parameters were varied to simultaneously fit the four data sets. 100 fits were generated by Monte Carlo search in parameter space and further optimized by nonlinear least-square fitting, as described in Section 4.2.3. A summary of our results is shown in Figure 4.4. Model fits that describe acceptor ratio (Y_{exp}^{sp}) and anisotropy (r_{exp}^{sp}) isotherms for IHF binding to the specific DNA substrate, H', using the three-state model, are shown in Figure 4.4.A and Figure 4.4.C, respectively. Fractional populations of free (f_{free}^{3s}), specifically bound (f_{sp}^{3s}), and non-specifically bound states ($\sum f_{ns,j}^{3s}$) are shown in Figure 4.4.E. Similarly, fits that describe acceptor ratio (Y_{exp}^{ns}) and anisotropy (r_{exp}^{ns}) isotherms for IHF binding to the specific DNA substrate, nsDNA, using the two-state model, are shown in Figure 4.4.B and Figure 4.4.D, respectively, with fractional populations of free (f_{free}^{2s}) and nonspecifically bound ($\sum f_{ns,j}^{2s}$) states shown in Figure 4.4.F.

The best fit parameters are summarized in Table 4.1. The uncertainties in the parameters were identified as the values at which χ^2 changes by 1; this increase, $\Delta\chi^2 = 1$, corresponds to one standard deviation in parameter space (41). The model yields K_d^{sp} in the range from 9 - 22 nM and K_d^{ns} in the range from 4.7-7.8 μM, at 200 mM KCl. Our K_d^{sp} value of 9-22 nM shows good agreement with K_d^{sp} values previously reported by our group using typical two-state (free and specifically bound) binding isotherm analysis, for identical buffer and KCL conditions: 16-30 nM (18), 46 ± 4 nM (27, 30) and ~2 nM (28). Note that these nM binding affinities are for 200 mM KCl, and are not consistent with other reports of nM binding affinities carried out at 100 mM KCl, where the complex is expected to be significantly tighter (14, 44-47).

The nonspecific binding affinity K_d^{ns} determined in this study shows that the nonspecific complex is approximately 200-900 times weaker than the specific complex. Though direct comparisons are not readily available because of differences in the experimental techniques and salt conditions, nonspecific binding for IHF-H' complexes has been previously reported to be two-three orders of magnitude weaker than specific interactions. ITC at 100 mM KCl reported K_d^{ns} to be 120-150 times weaker than K_d^{sp} (14) and EMSA studies, also performed at 100 mM KCl, reported K_d^{ns} to be roughly 700-800 times weaker than K_d^{sp} (19).

Our model yielded the nonspecific binding site size, n, in the range from 9-15 bp, which indicates that no more than 2 - 3 proteins may bind nonspecifically to our 35-bp DNA at the same time. In Figure 4.4.G, we show $\Delta\chi^2$ versus n for several good fits to highlight the broad range of n that can still fit the data. Interestingly, parameters n and Perrin parameter, r_η turned out to be highly correlated, which is not surprising. Models with large n require a higher r_η (i.e. each nonspecific protein, j, contributes a greater fraction to the measured anisotropy, $r_{ns,j}$) to compensate for fewer numbers of binding proteins needed to occupy all lattice sites. The relationship between r_η and n is shown in Figure 4.4.H.

Previous studies on the IHF-H' complex also yielded a wide range of values for the nonspecific binding site size, n, from 4 to 16 bps (14, 17, 19, 21). Taken together, these studies highlight the difficulty of determining n, especially with short DNA substrates used in most IHF studies involving the analytical binding model (34 -50 bp). The best fits (with the lowest χ^2) were for $n = 10\ bp$, which indicates that IHF occludes ~ 34 Å of DNA upon binding nonspecifically. Considering that a specifically-bound IHF kinks DNA at two positions that are 9 bps apart and that the diameter of IHF across the center of the protein – the distance measured between IHF residues

αQ85 and βE85 in the IHF-H' complex crystal structure – is 28.5 Å (1) (see also Figure 3.5), this best-fit estimate of n is reasonable.

On the other hand, the model's determination of r_η indicates more work to be done. r_η is unique because unlike any of the other parameters, its theoretical value may be directly compared to the model output ($r_\eta = \frac{\tau \rho RT}{\eta}$ where the lifetime of fluorescein τ_F~4.1 ns, $T = 293.15\ K$, and viscosity of IHF binding buffer approximated to ~1 mPa·s and ρ is estimated for globular proteins as 1.33 g/cm^3 (43). Estimation of r_η yields ~13.3 when molecular weight is described in kDa. Best fits in which $n = 10\ bp$ determines r_η to be roughly >25-fold higher. Moreover, r_η increases sharply up to the parameter limit of 500,000 even before $n = 13\ bp$ (Figure 4.4H). This greater than expected value for r_η suggests that the Perrin description as written in Eq. 4.20 may be incomplete.

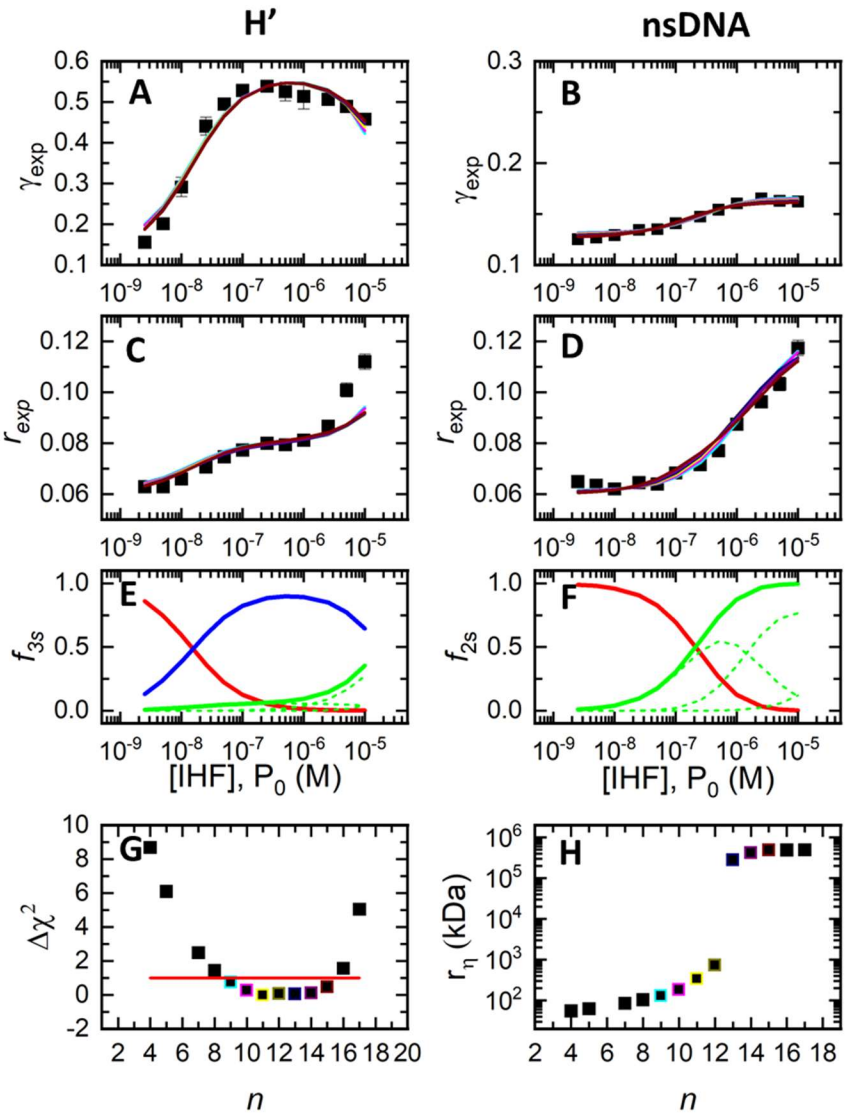


Figure 4.4. Simultaneous fitting of fluorescence-based titrations with two- and three-state models. (A-D) Binding isotherms at 20°C, measured with acceptor ratio shown for specific DNA, H' **(A)**, and nonspecific DNA, nsDNA **(B)**. **(C-D)** Similar measurements made with anisotropy shown for H' (**C**) and nsDNA (**D**). Best fit for each *n* from 9-15 bp shown (*n* = 9, cyan; *n* =10, pink; *n* = 11, yellow; *n* =12, gold; *n* =13, navy; *n* = 14, magenta; *n* = 15, maroon). **(E)** DNA populations of H' titrations: free DNA (red), specifically bound DNA (blue), nonspecifically bound DNA (green) are shown. Subpopulation of nonspecific DNA binding are shown in dashed lines. **(F)** DNA populations of nonspecific DNA titrations: free DNA (red) and nonspecifically bound DNA (green) populations are shown as well as nonspecific subpopulations (dashed green). **(G)** Difference in χ^2 for each *n* best fit from global best fit (*n* =10) which is set to zero. $\Delta\chi^2$=1 line shown in red. For *n* with $\Delta\chi^2$<1, fits are shown in (A-D) and point is outlined by color of fit. **(H)** Preferred r_η determined from the best fit for each n. For *n* where fits are shown in (A-D) and point is outlined by the color of fit. [DNA] (D_0) = 25 nM..

4.4.4. Successes and shortcomings of the three-state lattice model description

This three-state model for the specific IHF-H' binding, together with the corresponding two-state model for the nonspecific IHF-nsDNA binding, does well to simultaneously describe the dependence of the acceptor ratio and anisotropy binding isotherms for the IHF-nsDNA complexes with increasing [IHF] (Figure 4.4.B and D), as well as the acceptor ratio binding isotherm for the IHF-H' complex (Figure 4.4.A) but fails to fully capture the anisotropy increase observed at high protein concentrations ([IHF]>3 μM) (Figure 4.4.C). Interestingly, at the highest IHF concentrations in our measurements, the IHF-H' anisotropy level (Figure 4.4.C) approaches that of the IHF-nsDNA (Figure 4.4.D), which suggests that, at sufficiently high [IHF]/[DNA] (>200), the nonspecific and specific complexes have roughly the same number of molecules bound. In this three-state model, this would occur in the situation where all specifically bound proteins in the IHF-H' samples have been displaced by nonspecifically bound proteins at high enough protein concentrations. However, at these high [IHF] conditions, the IHF-H' acceptor ratio values (Figure 4.4.A), though decreasing as expected from competition by nonspecific binding, are still significantly higher than the IHF-nsDNA acceptor ratio values (Figure 4.4.B), indicating that there is a sufficient population of specifically bound IHF-DNA complexes in the IHF-H' ensemble even at high [IHF] concentrations.

To reconcile the acceptor ratio and the anisotropy data sets, we must release a constraint in the three-state model, whereby nonspecific binding by IHF was assumed to occur only upon complete displacement of the specifically bound IHF. Below we propose an extension to the

Record binding model which considers the possibility that specific and nonspecific binding may also occur simultaneously to the same DNA lattice.

4.5. Expansions to the lattice model: five-state description

4.5.1. IHF can bind nonspecifically on top of a specifically bound complex

Here, we expand our "three-state" description to a "five-state" description by considering an additional partially bent ("open") specific state to which IHF can bind nonspecifically (see cartoon in Figure 4.5.). Lifetime measurements on the IHF-DNA complex ((18); Chapter 3) showed partially-bent conformations of the specific IHF-H' complex with ~34% population under identical conditions (200 mM KCl) as these binding studies. This partially bent DNA would expose DNA sites that would normally be occluded in the specific complex, allowing nonspecific proteins to bind to the specific complex. Thus, we describe the specific DNA that is specifically bound to IHF as a mixture of three primary states, fully-bent specifically bound complex (D_b^{sp}), partially-bent specifically bound complex ($D_b^{sp,open}$), and a mixed state (D_{mixed}) where the partially-bent DNA accepts a nonspecific protein. The acceptor ratio level Υ_{sp} for the D_b^{sp} state is assumed to be different from the acceptor ratio levels for the $D_b^{sp,open}$ and D_{mixed} states, both of which are assigned to be the same as Υ_{ns}, the acceptor ratio level for the nonspecific complexes. The

anisotropy levels for the D_b^{sp} and $D_b^{sp,open}$ are described by r_{sp} while D_{mixed}, with two IHF proteins bound, is described by $r_{ns,j=2}$.

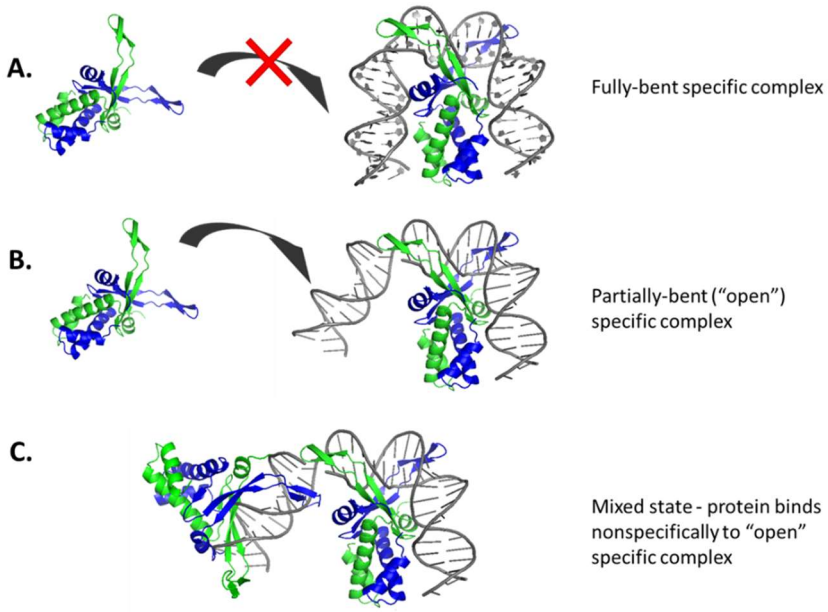


Figure 4.5. Nonspecific binding to specific complexes. (A) We show the crystal structure representation (1IHF) of IHF (blue and green) bound to H' DNA (grey). In the crystal structure, DNA is bound in the fully-bent state where all DNA is occluded (i.e. s ≈ N) which blocks additional proteins from binding to DNA. However, if DNA is in the "open" state **(B)** – such that one of the side arms of the DNA is exposed – an additional protein in solution may bind nonspecifically to form the mixed state **(C)**.

From lifetime studies, we can estimate binding affinity of the partially-bent IHF-H' complex as:

$$K_d^{sp,open} = K_d^{sp} e^{-(\frac{\Delta\Delta G}{RT})} \qquad \text{Eq. 4.25.}$$

where ΔΔG = –0.6 kcal/mol (determined experimentally in (18)), $R = 0.002 \frac{kcal}{mol}$, and $T = 293.15\ K$ for our experimental conditions.

Though we do not know the exact conformation of the partially-bent specific complex, our group has proposed that the partially bent conformations involve the flapping of flanking DNA arms – as illustrated in Figure 4.5 – which would expose ~13 bps of DNA, or roughly the binding site size for one nonspecifically bound protein (9-14 bps as determined by the three-state specific binding model). We therefore assume that only one additional protein will bind to an "open" specific complex to form the D_{mixed} state. Thus, our new "five-state" partition function becomes:

$$Z_{5s} = 1 + z_{sp} + z_{sp}^{open} + z_{mixed} + \sum_{j=1}^{P_{max}} z_{ns,j}$$
$$= 1 + \frac{[L]}{K_d^{sp}} + \frac{[L]^2}{K_d^{sp,open} K_d^{ns}} + \sum_{j=1}^{P_{max}} \frac{\Omega_j [L]}{K_d^{ns}} \qquad \text{Eq. 4.26.}$$

In Eq. 4.26, $z_{sp}^{open} = \frac{[L]}{K_d^{sp,open}}$ is the statistical weight for the specific, partially bent fraction $f_{sp,open}$ and $z_{mixed} = \frac{[L]^2}{K_d^{sp,open} K_d^{ns}}$ is the statistical weight for the mixed state f_{mixed}, where a specific DNA substrate has one protein bound specifically and one protein bound nonspecifically. The fractional populations of DNA are given by:

$$f_{free}^{5s} = \frac{1}{Z_{5s}} \qquad \text{Eq. 4.27.}$$

$$f_{ns,j}^{5s} = \frac{z_{ns,j}}{Z_{5s}} \qquad \text{Eq. 4.28.}$$

$$f_{sp}^{5s} = \frac{z_{sp}}{Z_{5s}} \qquad \text{Eq. 4.29.}$$

$$f_{sp,open}^{5s} = \frac{Z_{sp}^{open}}{Z_{5s}}$$ Eq. 4.30.

$$f_{mixed}^{5s} = \frac{Z_{mixed}}{Z_{5s}}$$ Eq. 4.31.

The acceptor ratio and anisotropy data measured directly on the IHF-H' specific complex (denoted as Y_{exp}^{sp} and r_{exp}^{sp}) were fitted using the following equations:

$$Y_{exp}^{sp} = \Upsilon_{free}\left(f_{free}^{5s}\right) + \Upsilon_{sp}\left(f_{sp}^{5s}\right) + \Upsilon_{ns}\left(f_{sp}^{5s,open} + f_{mixed}^{5s} + f_{ns}^{5s}\right)$$ Eq. 4.32.

$$r_{exp}^{sp} = r_{free}\left(f_{free}^{5s}\right) + r_{sp}\left(f_{sp}^{5s} + f_{sp}^{5s,open}\right) + r_{ns,j=2}\left(f_{mixed}^{5s}\right) + \sum_{j=1}^{P_{max}} r_{ns,j}\left(f_{ns,j}\right)$$ Eq. 4.33.

where the fractional populations were obtained from Eqs.4.27-4.31. It should be noted that for simplicity, we did not change the way that the free ligand concentration $[L] = P_{free}$ was calculated. This simplification affects only a small range of protein concentrations where nonspecific binding starts to contribute but before we reach the range where $P_{free} \approx P_0$.

As described above, this five-state model adds no free parameters to the three-state model. Only 9 free parameters are required again: the dissociation constants K_d^{ns} and K_d^{sp}, the nonspecific binding site size n, the acceptor ratio parameters, Υ_{free}, Υ_{ns}, Υ_{sp}, and the anisotropy parameters r_{free}, r_{sp}, and r_{η}.

4.5.2. Results of simultaneous fitting to IHF binding isotherms using the five-state lattice model.

Model parameters were varied to simultaneously fit the four titration data sets. 100 fits were generated by Monte Carlo search in parameter space followed by least-squares minimization using nonlinear fitting as described earlier. A summary of our results is shown in Figure 4.6. Model fits that describe acceptor ratio (Y_{exp}^{sp}) and anisotropy (r_{exp}^{sp}) binding isotherms for IHF binding to our specific DNA substrate, H', are shown in Figure 4.6.A and Figure 4.6.C, respectively. Fractional populations of free, specifically bound states, including partially bent and mixed states, and non-specifically bound states (f_{free}^{5s}, f_{sp}^{5s} , $f_{sp,open}^{5s}$, $f_{mixed}^{5s}$and $\sum f_{ns,j}$ respectively) are shown in Figure 4.6.E. Similarly, fits to describe acceptor ratio (Y_{exp}^{ns}) and anisotropy (r_{exp}^{ns}) binding isotherms for IHF binding to nonspecific substrate, nsDNA, using the two-state model are shown in Figure 4.6.B and Figure 4.6.D and populations of free and nonspecifically bound states (f_{free}^{2s} and $\sum f_{ns,j}^{2s}$) are shown in Figure 4.6.F.

The best fit parameters are summarized in Table 4.1. The uncertainties in all parameters listed were estimated finding the parameter values where $\Delta\chi^2 = 1$ as described in 4.2.3. (41). The model yields K_d^{sp}values in the range from 3–9 nM and K_d^{ns} values in the range from 10–20 μM. The nonspecific binding site size n, ranged from 10–16 bp which indicates that 3 to 5 proteins may bind nonspecifically to our 35-bp DNA at a time. The $\Delta\chi^2$ from each best fit versus n is shown in Figure 4.6.G to highlight the broad range of n that still fit the data. Again, parameters n and Perrin parameter, r_η were highly correlated. Models with large n require a higher r_η (i.e. each nonspecific protein, j, contributes a greater anisotropy $r_{ns,j}$) to compensate for loss of a binding protein. The relationship between r_η and n is shown in Figure 4.6.H. No r_η value could be determined here because no r_η value increased χ^2 by ≥1.

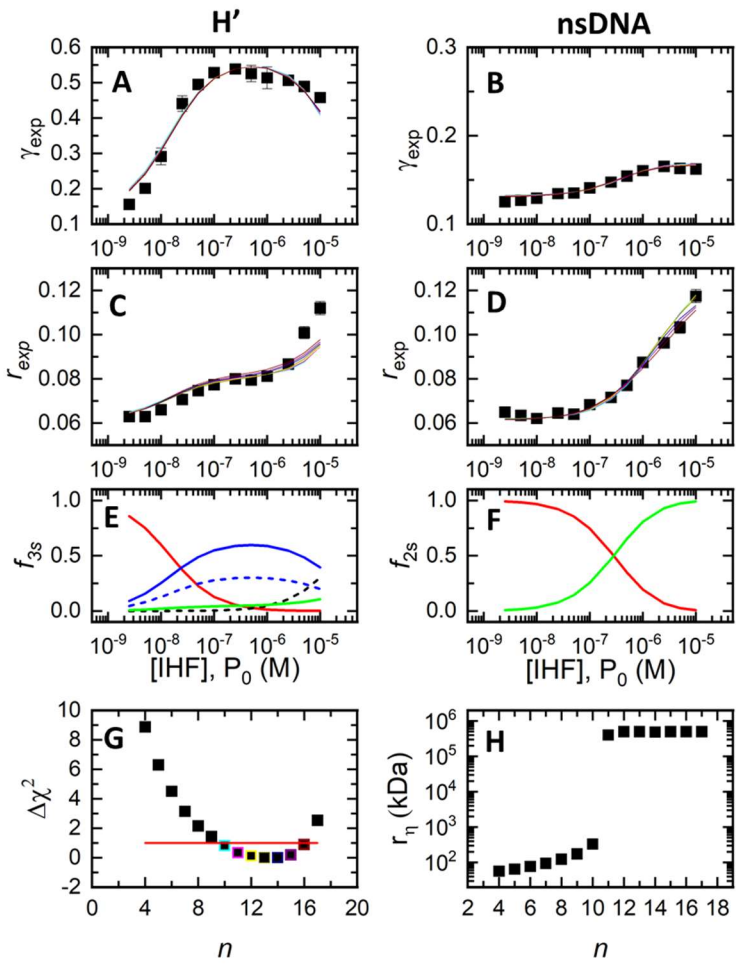


Figure 4.6. Simultaneous fitting of fluorescence-based titrations with two- and five-state models. (A-D) Binding isotherms at 20°C, measured with acceptor ratio shown for specific DNA, H' **(A)**, and nonspecific DNA, nsDNA **(B)**. **(C-D)** Similar measurements made with anisotropy shown for H' (**C**) and nsDNA (**D**). Best fit for each *n* from 9-15 bp shown (*n* = 10, cyan; *n* =11, pink; *n* = 12, yellow; *n* =13, gold; *n* =14, navy; *n* = 15, magenta; *n* = 16, maroon). **(E)** DNA populations of H' titrations. free DNA (red), fully- (solid) and partially- bent (dashed) specifically bound DNA (blue), mixed state DNA (black dashed), nonspecifically bound DNA (green) are shown. **(F)** DNA populations of nonspecific DNA titrations. free DNA (red) and nonspecifically bound DNA (green) populations are shown. **(G)** Difference in χ^2 for each *n* best fit from global best fit (*n* =10) which is set to zero. $\Delta\chi^2$=1 line shown in red. For *n* with $\Delta\chi^2$<1, fits are shown in (A-D) and point is outlined by color of fit. **(H)** Preferred r_η determined from the best fit for each n. For *n* where fits are shown in (A-D) and point is outlined by color of fit. [DNA] (D_0) = 25 nM.

Table 4.1. Fitting parameters for IHF-DNA binding models.

Model	K_d^{sp}(M)	K_d^{ns} (M)	n (bp)	Υ_{free}	Υ_{ns}	Υ_{sp}	r_{free}	r_{sp}	r_η (kDa)	χ_ν^2 (Best fit)†
2-state fitting to specific complex *	$24^{+8}_{-4} \times 10^{-9}$ **	--------	--------	0.100 ± 0.004	--------	0.54 ± 0.04	n.d.	n.d.	--------	--------
Two- and three-state	$15^{+7}_{-6} \times 10^{-9}$ **	$6.0^{+1.8}_{-1.3} \times 10^{-6}$ **	$9-15$	0.130 ± 0.003	0.164 ± 0.004	0.59 ± 0.03	0.061 ± 0.006	0.080 ± 0.007	$338^{+n.d.}_{-296}$ **	6.13
Two- and five-state	$6^{+2}_{-3} \times 10^{-9}$ **	$15^{+5}_{-5} \times 10^{-6}$ **	$7-13$	0.134 ± 0.003	0.182 ± 0.005	0.80 ± 0.04	0.063 ± 0.005	0.084 ± 0.005	n.d.	2.71

* Model used for salt titration studies (see Chapter 3)

** boundary values of parameters asymmetric in linear space are given as superscripts(upper bound) and subscripts (lower bound)

† χ_ν^2 is the reduced χ^2 which equal to the value calculated by Eq. 4.1 divided by the number of points measured (n=12)

4.5.3. Successes and shortcomings of the five-state lattice model description

Here we proposed an extension to the three-state model, a "five-state" model that includes a partially bent specific state and a mixed state, which considers nonspecific binding to this partially bent specific state. The inclusion of these partially-bent specific states and mixed states were motivated by previous observations of IHF-DNA complex heterogeneity from lifetime studies (18) and, more recently, from AFM studies (15).

Our five-state description of the IHF-H' specific complex changed some of the output parameters compared with the three-state description. The two most noticeable changes in parameters were (1) specific binding to the fully bent complex, K_d^{sp}, became stronger (6 nM in the five-state model) compared with the ~15 nM K_d^{sp} in the three-state model, which did not distinguish between fully-bent and partially bent; and (2) the Υ_{sp} increased (~0.59 in the three-state model to ~0.80 in the five-state model). The shifts in both these parameters are to be expected. The three-state model describes the average K_d^{sp} and acceptor ratio of the specific complex, which is a mixture of a high-FRET, fully-bent complex and a low-FRET, partially-bent complex, as shown in our lifetime studies (Chapter 3). The five-state model separates the ~66% population in the fully-bent, high-FRET state from the remaining partially-bent, low-FRET state, and the K_d^{sp} for the fully-bent state is modeled to be stronger (lower value) than the $K_d^{sp,open}$ for the partially-bent specific complex, as described in Eq. 4.25. Therefore K_d^{sp} in the five-state model is expected to be tighter than the K_d^{sp} in the three-state model, and also tighter than previous reports of ~45 nM binding affinity from our previous one-to-one binding isotherm studies at 200 mM KCl (27, 30). Similarly, the increase in Υ_{sp} in the five-state model compared with the three-state model is as expected and consistent with the lifetime studies that showed that the fully-bent conformation

in the specific complex is characterized by a FRET of ~0.67, while the average FRET (averaged over both conformations) is ~0.46, under identical conditions.

In the five-state model, nonspecific binding site size n (10-16 bp) was approximately the same as in the three-state model (9-15 bp). We made the case earlier that our range agrees with what should be expected physically based on previous report of binding site size (14, 17, 19, 21) and evaluation of protein size within the IHF-H' crystal structure.

Overall, the five-state model only marginally improved the fit. Overall goodness-of-fit improved, indicated by a reduction in the overall χ^2 (Table 4.2). In the end, our studies suggest that while our five-state model expanded the description of IHF binding in a direction suggested by other data, the fact that we added no additional parameters even as we expanded the number of accessible states gave us more-or-less an equivalent fit with the parameters appropriately adjusted. Future studies to fully understand the binding behavior of this protein would need to explore how the fits would change if we relaxed some of the constraints such as allowing K_d^{ns} in the mixed state to be an independent parameter and not constraining the acceptor ratio of the mixed state to be identical to that of the nonspecifically bound complexes.

4.6. Discussion and conclusions

In this report, we fit fluorescence-based titration data to multistate analytical binding models for binding by ligands to a finite lattice. This is the first study, to our knowledge, in which this analytical binding model was applied to simultaneously describe binding isotherms obtained from both acceptor ratio (to measure extent of bending) and anisotropy (to measure extent of

binding) for IHF binding to both specific and nonspecific DNA substrates in a self-consistent manner.

While this model could reasonably describe all four binding isotherms up to [IHF]/[DNA] ~ 100, deviations from the model were evident at higher protein/DNA ratios, largely because the measured anisotropy of the IHF-H' complex indicated far more proteins bound to the specific substrate than the model allowed. An extension of the model to allow one additional protein to bind nonspecifically on top of a specifically bound complex did not have a significant effect on the quality of the fit, although we did not explore all the ramifications of this extended model. Furthermore, these studies were carried out only under one salt condition, at 200 mM KCl, and therefore did not directly address why different measurement techniques gave widely varying estimates of specific binding affinity of IHF for H' at ~100 mM salt conditions. Nonetheless, our results here that IHF binds to the specific H' site with nM binding affinity at 200 mM KCl is consistent with our previous studies at 200 mM and higher salt conditions (18, 28, 30), and also consistent with the much tighter binding affinity of 25 pM affinity at 100 mM KCl, as reported by us and by the Crothers group (26). In the future, binding studies performed for a range of salt conditions are needed to provide accurate quantification of both specific and nonspecific binding of IHF under different conditions, including those found *in vivo*.

Our studies are currently limited to a single-lattice model and does not take into consideration one protein binding to two different lattices. Recent AFM studies show IHF-mediated bridging between multiple DNA containing IHF binding sites (e.g. H2 sites that are also naturally occurring binding site on phage λ DNA), when IHF is in large excess of DNA. Bridging is observed when proteins bind to multiple DNA sites to form tight "supercomplexes" consisting of many DNAs and IHF molecules (15). Similar observations were made on a DNA-bending

mitochondrial transcription factor, TFAM, that form nucleoid-like structures at high protein concentrations (48). It is unclear if such supercomplexes are present under the conditions of our titration studies, though that could explain why anisotropy levels are significantly higher at high-[IHF] (>2 μM) than predicted by the Perrin Equation for one DNA molecule binding to multiple IHF molecules (Eq. 4.19). Further studies that take into consideration IHF-mediated bridging as an expansion to the analytical binding model presented here are needed to fully explain our data and to provide a complete picture of the role of nonspecific binding and DNA compaction by IHF *in vivo*.

References:

1. Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell. 1996;87(7):1295-306.
2. Laxmikanthan G, Xu C, Brilot AF, Warren D, Steele L, Seah N, et al. Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. Elife. 2016;5.
3. Grindley ND, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. Annu Rev Biochem. 2006;75:567-605.
4. Khrapunov S, Brenowitz M, Rice PA, Catalano CE. Binding then bending: A mechanism for wrapping DNA. Proc Natl Acad Sci USA. 2006;103(51):19217-8.
5. Ali BM, Amit R, Braslavsky I, Oppenheim AB, Gileadi O, Stavans J. Compaction of single DNA molecules induced by binding of integration host factor (IHF). Proc Natl Acad Sci U S A. 2001;98(19):10658-63.
6. Huo YX, Zhang YT, Xiao Y, Zhang X, Buck M, Kolb A, et al. IHF-binding sites inhibit DNA loop formation and transcription initiation. Nucleic Acids Res. 2009;37(12):3878-86.
7. Nash HA. Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. Annu Rev Genet. 1981;15:143-67.
8. Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat Struct Mol Biol. 2014;21(6):528-34.
9. Polaczek P, Kwan K, Liberies DA, Campbell JL. Role of architectural elements in combinatorial regulation of initiation of DNA replication in Escherichia coli. Mol Microbiol. 1997;26(2):261-75.
10. Winkelman JW, Hatfield GW. Characterization of the integration host factor binding site in the ilvPG1 promoter region of the ilvGMEDA operon of Escherichia coli. J Biol Chem. 1990;265(17):10055-60.
11. Xin W, Feiss M. Function of IHF in lambda DNA packaging. I. Identification of the strong binding site for integration host factor and the locus for intrinsic bending in cosB. J Mol Biol. 1993;230(2):492-504.
12. Wright AV, Liu JJ, Knott GJ, Doxzen KW, Nogales E, Doudna JA. Structures of the CRISPR genome integration complex. Science. 2017;357(6356):1113-8.
13. Vander Meulen KA, Saecker RM, Record MT, Jr. Formation of a wrapped DNA-protein interface: experimental characterization and analysis of the large contributions of ions and water to the thermodynamics of binding IHF to H' DNA. J Mol Biol. 2008;377(1):9-27.
14. Holbrook JA, Tsodikov OV, Saecker RM, Record MT, Jr. Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. J Mol Biol. 2001;310(2):379-401.
15. Yoshua SB, Watson GD, Howard JAL, Velasco-Berrelleza V, Leake MC, Noy A. A nucleoid-associated protein bends and bridges DNA in a multiplicity of topological states with varying specificity. bioRxiv. 2020:2020.04.17.047076.
16. Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J Bacteriol. 1999;181(20):6361-70.
17. Tsodikov OV, Holbrook JA, Shkel IA, Record MT, Jr. Analytic binding isotherms describing competitive interactions of a protein ligand with specific and nonspecific sites on the same DNA oligomer. Biophys J. 2001;81(4):1960-9.
18. Connolly M, Arra A, Zvoda V, Steinbach PJ, Rice PA, Ansari A. Static Kinks or Flexible Hinges: Multiple Conformations of Bent DNA Bound to Integration Host Factor Revealed by Fluorescence Lifetime Measurements. J Phys Chem B. 2018.

19. Aeling KA, Opel ML, Steffen NR, Tretyachenko-Ladokhina V, Hatfield GW, Lathrop RH, et al. Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy. J Biol Chem. 2006;281(51):39236-48.
20. Lynch TW, Read EK, Mattis AN, Gardner JF, Rice PA. Integration host factor: putting a twist on protein-DNA recognition. J Mol Biol. 2003;330(3):493-502.
21. Aeling KA, Steffen NR, Johnson M, Hatfield GW, Lathrop RH, Senear DF. DNA deformation energy as an indirect recognition mechanism in protein-DNA interactions. IEEE/ACM Trans Comput Biol Bioinform. 2007;4(1):117-25.
22. Yang SW, Nash HA. Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. Embo J. 1995;14(24):6292-300.
23. Wang S, Cosstick R, Gardner JF, Gumport RI. The specific binding of Escherichia coli integration host factor involves both major and minor grooves of DNA. Biochemistry. 1995;34(40):13082-90.
24. Zhou H, Sathyamoorthy B, Stelling A, Xu Y, Xue Y, Pigli YZ, et al. Characterizing Watson-Crick versus Hoogsteen Base Pairing in a DNA-Protein Complex Using Nuclear Magnetic Resonance and Site-Specifically (13)C- and (15)N-Labeled DNA. Biochemistry. 2019;58(15):1963-74.
25. Lorenz M, Hillisch A, Goodman SD, Diekmann S. Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence resonance energy transfer. Nucleic Acids Res. 1999;27(23):4619-25.
26. Sugimura S, Crothers DM. Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. Proc Natl Acad Sci USA. 2006;103(49):18510-4.
27. Vivas P, Velmurugu Y, Kuznetsov SV, Rice PA, Ansari A. Mapping the transition state for DNA bending by IHF. J Mol Biol. 2012;418(5):300-15.
28. Vivas P, Kuznetsov SV, Ansari A. New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. J Phys Chem B. 2008;112(19):5997-6007.
29. Freyer MW, Lewis EA. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. Methods Cell Biol. 2008;84:79-113.
30. Velmurugu Y, Vivas P, Connolly M, Kuznetsov SV, Rice PA, Ansari A. Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein. Nucleic Acids Res. 2018;46(4):1741-55.
31. Swinger KK, Lemberg KM, Zhang Y, Rice PA. Flexible DNA bending in HU-DNA cocrystal structures. EMBO J. 2003;22(14):3749-60.
32. Murtin C, Engelhorn M, Geiselmann J, Boccard F. A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: re-evaluation of the effective concentration of IHF in the cell. J Mol Biol. 1998;284(4):949-61.
33. Fried MG, Liu G. Molecular sequestration stabilizes CAP-DNA complexes during polyacrylamide gel electrophoresis. Nucleic Acids Res. 1994;22(23):5054-9.
34. Vossen KM, Fried MG. Sequestration stabilizes lac repressor-DNA complexes during gel electrophoresis. Anal Biochem. 1997;245(1):85-92.
35. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc. 2007;2(8):1849-61.
36. Holbrook JA, Tsodikov OV, Saecker RM, Record MT, Jr. Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. J Mol Biol. 2001;310:379-401.
37. McGhee JD, von Hippel PH. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. J Mol Biol. 1974;86(2):469-89.

38. Kirkpatrick S, Gelatt CD, Jr., Vecchi MP. Optimization by simulated annealing. Science. 1983;220(4598):671-80.
39. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes 3rd Edition: The Art of Scientific Computing: Cambridge University Press; 2007.
40. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA. Conformational relaxation and ligand binding in myoglobin. Biochemistry. 1994;33(17):5128-45.
41. Bevington PR. Data reduction and error analysis for the physical sciences. 3rd ed. ed. Robinson DK, editor. Boston: McGraw-Hill; 2003.
42. Zhang W, Ni H, Capp MW, Anderson CF, Lohman TM, Record MT, Jr. The importance of coulombic end effects: experimental characterization of the effects of oligonucleotide flanking charges on the strength and salt dependence of oligocation (L8+) binding to single-stranded DNA oligomers. Biophys J. 1999;76(2):1008-17.
43. Lakowicz JR. Principles of fluorescence spectroscopy. 3rd ed. ed. New York: Springer; 2006.
44. Lohman TM, Mascotti DP. Thermodynamics of ligand-nucleic acid interactions. Methods Enzymol. 1992;212:400-24.
45. Record MT, Jr., Anderson CF, Lohman TM. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. Q Rev Biophys. 1978;11(2):103-78.
46. Manning GS. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q Rev Biophys. 1978;11(2):179-246.
47. Vivas P, Velmurugu Y, Kuznetsov SV, Rice PA, Ansari A. Global analysis of ion dependence unveils hidden steps in DNA binding and bending by integration host factor. J Chem Phys. 2013;139(12):121927.
48. Kaufman BA, Durisic N, Mativetsky JM, Costantino S, Hancock MA, Grutter P, et al. The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. Molecular biology of the cell. 2007;18(9):3225-36.

Chapter 5.

Does transient Hoogsteen pairing play a role in binding-site recognition by DNA-bending protein IHF?

5.1. **Introduction**

Hoogsteen base-pairs (bps) are energetically available alternate to Watson-Crick bps that are transiently available in B-DNA with lifetimes lasting up to a few ms (1-3). Compared to a Watson-Crick A-T bp, Hoogsteen pairing contains a 180-degree flipped adenine (*syn* rather than an *anti* conformation) to form two different hydrogen bonds. Similarly, guanine can flip in G-C bps (Figure 5.1). Hoogsteen flipping impacts DNA backbone structure. To form Hoogsteen hydrogen bonds, nucleobases are ~2.0-2.5 Å closer than in Watson-Crick bps, changing electrostatics and backbone shape, and in some cases may even introduce kinks into DNA (4, 5). Recent nuclear magnetic resonance (NMR) studies have revealed that Watson-Crick bps exist in dynamic equilibrium with short-lived (ms lifetime), low-frequency (populations <1%) Hoogsteen bps (1). Transient Hoogsteen pairs exist for longer than the so-called 'residence' time of a protein undergoing 1D diffusion on DNA – defined as the length of time that a protein sits on a particular DNA site before diffusing to a neighboring site – which is estimated to be in the range of 500 ns to 500 μs (6-12). Thus, transient Hoogsteen base pairing could play a role in stalling a diffusing protein to give it more time to interrogate a potential binding site. In this chapter, we investigated the impact transient Hoogsteen base pairing may have in the ability of Integration Host Factor (IHF) to recognize one of its specific target sites, the H' site on λ DNA.

Hoogsteen pairing in B-DNA adds diversity to DNA conformations that site-specific proteins may potentially sense, especially proteins like IHF that rely on "indirect readout" to recognize

their target sites. Hoogsteen formation is shown to be more probable and with a greater lifetime in dinucleotide steps that are considered more bendable (2). Hoogsteen bps also appear with higher frequency in damaged DNA sites, where DNA-flexibility is naturally enhanced, suggesting a role for Hoogsteen pairing in DNA damage accommodation, recognition, and repair – cellular processes that use many indirect readout protiens (13-17). Finally, two crystal structures of DNA bound to DNA-bending proteins reveal Hoogsteen bps, Integration Host Factor (IHF) (18) and TATA box binding protein (19). Considering that Hoogsteen bps may form in B-DNA (1, 13, 20) with a preference for sites of major groove kinking (4, 5) it comes as something of a surprise that Hoogsteen bps have not been more widely observed in structures of protein−DNA complexes in which structural distortions could destabilize Watson−Crick bps.

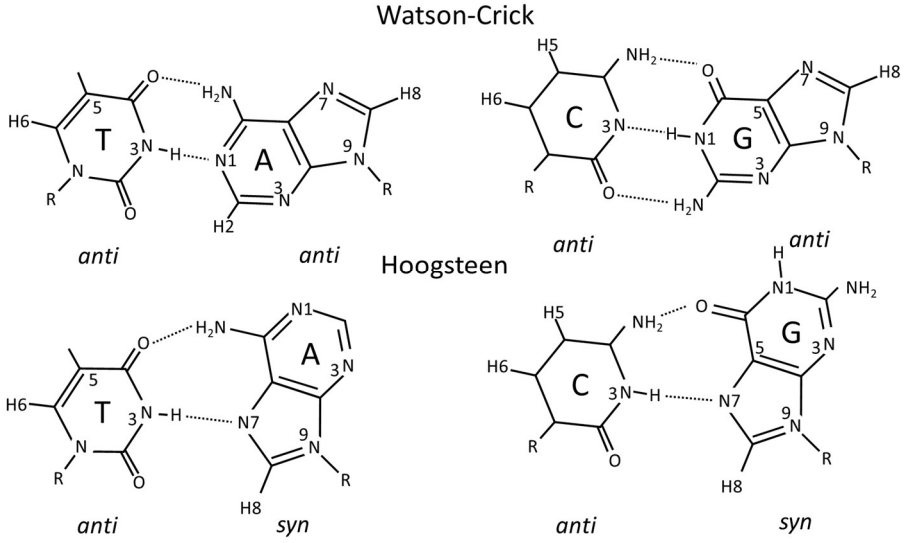


Figure 5.1. Chemical Structures of Watson-Crick and Hoogsteen Base pairs. Differences in hydrogen bonding and purine base orientation (anti vs syn) are shown.

Several features of the IHF-H' complex suggested that Hoogsteen formation may be an important factor in the recognition mechanism of IHF. First, as mentioned above, IHF is an "indirect readout" protein that likely senses sequence-dependent differences in the DNA conformational dynamics and flexibility. Distortions to the DNA backbone by Hoogsteen bps,

especially if they induce a kink, could offer a prebent DNA structure to IHF, thereby stalling the protein at that site and lowering the energy barrier required to kink the DNA to form the stable specific complex. Interestingly, in the H' sequence, which is a naturally occurring specific DNA site that IHF binds to with pM binding affinity, CA/TG steps are found adjacent to two sites that are sharply kinked in the complex that enables IHF to bend the 35-bp H' site into a U-turn. CA/TG steps are where transient Hoogsteen bps were first observed in naked duplex DNA in solution and that show the highest prevalence of Hoogsteen formation (1, 2). Next, although Hoogsteen pairing in B-DNA is transient, the lifetimes overlap with nonspecific interrogation times measured on IHF-DNA complexes (21), indicating that Hoogsteen formation could be relevant for target recognition by IHF. And finally, as mentioned above, IHF-H' crystal structure is one of the few crystal structures in which Hoogsteen bps have been identified (18, 20). Hoogsteen pairing was observed adjacent to one of the DNA kink sites – the site that lies outside of the consensuses sequence. Notably, a nick was introduced adjacent to that kink site to faciliate crystallization of the IHF-H' complex. And while recent NMR studies showed that intact DNA in complex with IHF is predominantly Watson-Crick paired even at the kink sites (22), it does not preclude the possibility that transient Hoogsteen base-pairing could facilitate and accelerate recognition by IHF.

Here, we investigated the role transient Hoogsteen base-pairing may have in the ease with which IHF recognizes its DNA binding-site site. We utilized laser temperature-jump (T-jump) spectroscopy to measure changes in the DNA bending/unbending kinetics in the IHF-H' complex, when a Hoogsteen inhibiting nucleotide analog 7-deaza adenine (7dA) was substituted instead of the normal adenine either near one of the kink sites or at some distance away from the kink site, as a control. 7deaza nucleotide analogs have been used previously to characterize the likelihood of Hoogsteen formation in the absence of any bound proteins (23) and to investigate

the role of Hoogsteen base pairing in DNA-binding by proteins (22, 24). If transient Hoogsteen formation facilitates DNA kinking, then we hypothesized that 7dA substitution made at or near the kink site should destabilize the modified IHF-H' complex and slow the rates of DNA bending. We anticipated no change in complex stability and DNA bending rates with 7dA substitutions made at sites that remain B-DNA like in the complex and are at some distance from the kinks.

Our study revealed some unexpected results. First, 7dA substitutions near one of the kink sites had little or no effect on the binding affinity of IHF for H' and the rates for DNA bending attributed to the formation of the specific complex *increased* by ~2-fold. Second, 7dA substitutions made away from the kink site, and initially expected to have no effect on affinity nor on kinetics, in fact stabilized the complex by ~40-fold and slightly increased the rates of DNA bending/unbending. From these results we conclude that transient Hoogsteen pairing at potentially weak sites in the DNA does not appear to be an important factor in facilitating DNA kinking or target recognition by IHF. Instead, we attribute the ~40-fold increase in binding affinity for our "control" experiment by recognizing that the 7dA substitution in this case was made in the TTG consensus region of the H' sequence. We interpret the increased stability of the "control" to be due to weakened stacking interactions from the 7dA substitute and hence increased DNA "twistability" at the TTG site, previously shown to be critical in facilitating a series of ionic contacts between the protein and the DNA needed to stabilize the bent DNA arm on that side of the complex (21, 25). Our findings support the recent NMR findings on the IHF-H' complex that show Hoogsteen pairing to have little importance within the IHF-DNA complex (22). Instead, these studies highlight the profound impact that apparently minor modifications in

a nucleotide moiety - that still enable Watson-Crick pairing – can have on DNA flexibility/deformability and thereby influencing indirect readout of target sites.

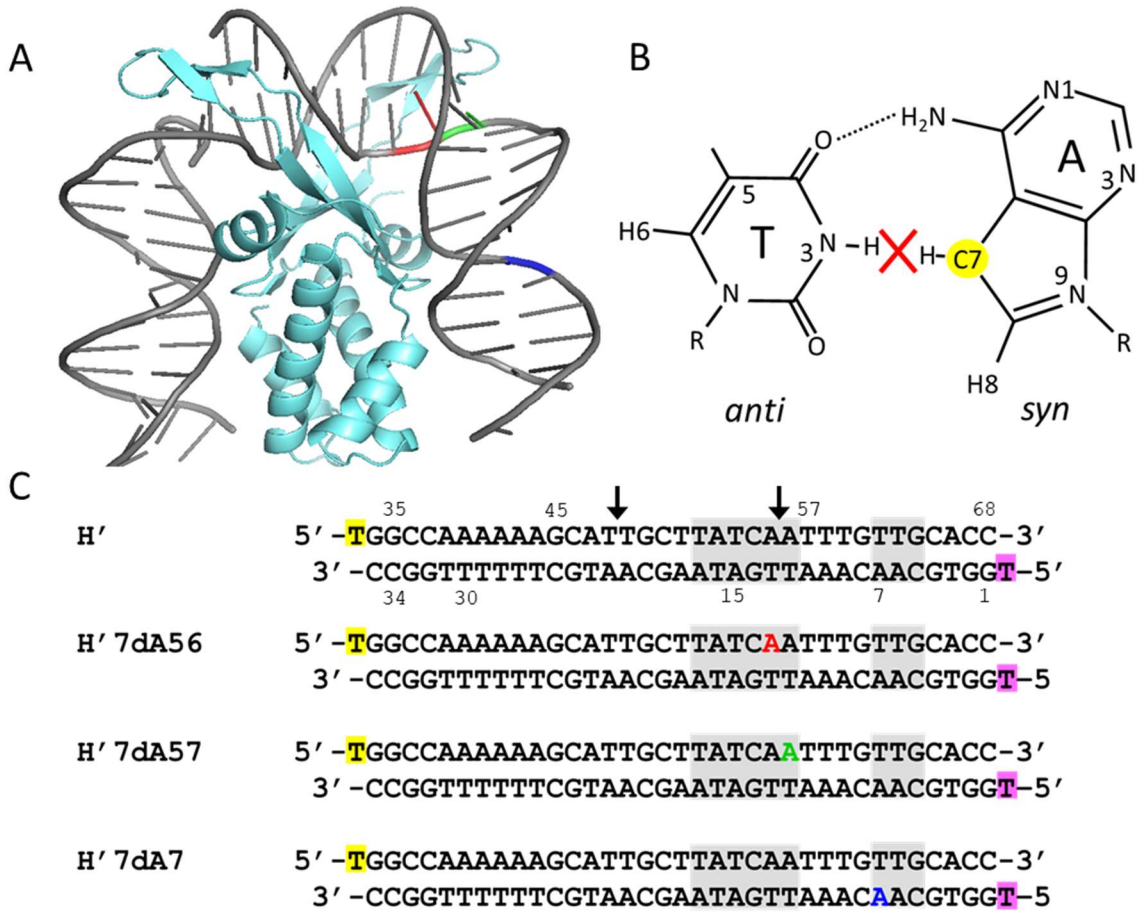


Figure 5.2. IHF-H' crystal structure, 7-deaza modifications, and DNA sequences. (A) Model of the IHF–H′ complex based on the cocrystal structure (PDB code: 1IHF). The DNA is shown in gray, with locations of modifications in the DNA sequences used in this study highlighted in color: 7-deaza adenine (7dA) substitutions are highlighted in red (H' 7dA56), green (7dA57) and blue (7dA7) respectively; The T in the T→A mutation in the H'44A substrate is highlighted in orange. (B) 7dA-T Hoogsteen base pair, where nitrogen at the 7th position in purine ring is modified from a nitrogen to a carbon (highlighted in yellow), blocking a stabilizing hydrogen bond (red 'X'). (C) Sequences of H' substrates and its variants are shown. The 5'-end of the top strand is fluorescein-dT (yellow), and the 5'-end of the bottom (complementary) strand is TAMRA-dT (pink). The consensus region is shown in gray. The black arrows indicate location of the kinks in the DNA when in complex with IHF. 7-deaza adenines were substituted for adenine in H' 7dA56 (red), H' 7dA57 (green) and H' 7dA7 (blue). Numbering scheme shown was set up to match that used by Zhou et. al (22).

5.2. Methods

5.2.1. Materials

The DNA sequences used in this study are shown in Figure 5.2.C. All DNA were end-labeled with Fluorescein and TAMRA (see Design 1 constructs in Chapter 2.1.1.). The IHF protein was prepared as described previously (26). All measurements were performed in binding buffer: 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.01% NP-40, and 100 mM KCl. All details of IHF and DNA sample preparation are described in Chapter 2.1.2.

5.2.2. Equilibrium binding measurements

The steady-state fluorescence emission spectra and anisotropies were measured on a FluoroMax4 spectrofluorimeter (Jobin Yvon, Inc., NJ). FRET efficiencies (FRET E) were obtained from the measured spectra under each condition, as described in Chapter 2.2.3-4. The circular dichroism (CD) measurements were carried out on JASCO J-810 spectropolarometer, as described in Chapter 2.2.2. The dissociation constants (K_d) for all the IHF-DNA complexes at 100 mM KCl were determined from extrapolation of the K_d values obtained at higher salt, in the range of 200-400 mM KCl, from binding isotherms measured at each salt. These binding isotherms were obtained from measurements of acceptor ratio (related to FRET E) versus IHF concentration for fixed DNA concentration, as described in Chapter 2.2.3. The DNA concentrations for binding isotherms were 50 nM. This extrapolation approach from measurements at higher salt concentrations enables accurate determination of sub-nM binding affinities, as detailed in Chapter 2.2.6.3 and Chapter 3 and 4.

5.2.3. Equilibrium FRET measurements

To obtain measures of IHF-H' complex FRET at all equilibrium temperatures measured, FRET value and error were determined by lifetime measurements at 20°C (Lifetime measurements shown in Chapter 3). To obtain FRET values for higher temperatures (E_T), the complex FRET measured at 20° (E_{20}) was multiplied by the ratio of deviations in donor intensity between IHF-H'_DA complex and H'_D described by:

$$E_T = E_{20}\left(\frac{I_D}{I_{DA}}\right) \qquad \text{Eq. 5.1.}$$

where I_D and I_{DA} are intensities of H'_D and IHF-H'_DA respectively. Errors shown for all temperatures (δE_T) are a sum of errors from the three independent measurements ($\delta E_{20}, \delta I_D, \delta I_{DA}$) as described below:

$$\delta E_T = E_T\sqrt{\left(\frac{\delta I_D}{I_D}\right)^2 + \left(\frac{\delta I_{DA}}{I_{DA}}\right)^2 + \left(\frac{\delta E_{20}}{E_{20}}\right)^2} \qquad \text{Eq. 5.2.}$$

This method was preferred to direct temperature-dependent measures of FRET on IHF-H' complexes because it's independent of sample concentration. Direct measures of FRET require measurements to be made on different IHF-DNA complexes samples of identical DNA concentration which often leads to greater sample-to sample variability.

5.2.4. Laser temperature-jump (T-jump) measurements

Rapid T-jump of ~5–10 °C was achieved in sample cuvettes of path length 0.5 mm, and the fluorescence emission intensities of the donor (fluorescein) were measured as a function of time, with excitation at 488 nm, to acquire the relaxation traces (26, 27). To cover the span of over three orders of magnitude in time, we recorded the T-jump kinetics traces over different time-scales and then combined these traces (see Figure 2.8). The combined relaxation traces were analyzed using maximum entropy method (MEM) to obtain a model-independent distribution of relaxation times that best described our relaxation traces. These traces were also analyzed using a sum of discrete exponential decay curves. The details of the laser T-jump spectrometer and the analysis of the relaxation traces are described in Chapter 2.4.

5.3.Results

5.3.1. Fraction of unbent DNA in IHF-DNA complexes, measured by FRET, increase with increasing temperature

The nearly 180° bend in the DNA observed in the IHF-H' crystal structure (Figure 5.2A) shortens the end-to-end distance of 35-bp DNA from ~100 Å to ~50 Å. Therefore, FRET measurements between fluorophores attached to the opposite DNA ends provide a sensitive probe for DNA bending in complex with IHF. In this study, we labeled the 5'-terminus of one DNA strand with fluorescein and the other with TAMRA (Figure 5.2.C). FRET assays were used to measure the end-to-end distance of the DNA, thus the extent of bending within each IHF-DNA ensemble.

Prior to the T-jump studies, we characterized the equilibrium changes in FRET as a function of temperature, in the range from 20 to 60 °C. In the absence of IHF, the distance between the donor and acceptor fluorophores in the double-labeled DNA only construct (H'_DA) is

estimated to be ~120 Å, much larger than the 50 Å Forster distance for this FRET pair (Figure 3.5), and therefore the FRET efficiency between the pair is negligible. In this case, as the temperature is raised, the donor intensity in H'_D decreases monotonically and reflects the temperature-dependent decrease in the quantum yield of the donor fluorophore (Figure 5.3.A). In the presence of IHF, the donor intensity first decreases as in DNA alone, up to ~30 °C, and then starts to increase as a result of unbending of DNA in the IHF-H' complex (26, 27). The FRET E of the complex decreases from 0.50 ± 0.04 at 20 °C to 0.38 ± 0.05 at 60 °C (Figure 5.3.B). Previous studies in our group showed that this temperature-dependent decrease in FRET at 100 mM KCl was independent of the IHF and DNA concentrations, indicating that the FRET changes were monitoring unbending of DNA while still bound to IHF, and not dissociation of the IHF-H' complex (27, 28). Bimolecular dissociation of the complex starts to be significant at salt concentrations greater than about 250 mM KCl (28, 29). Thus, for the 100 mM [KCl] used in this study, we are primarily monitoring unimolecular DNA bending/unbending processes.

5.3.2. IHF bound to DNA is stabilized against thermal denaturation

To ensure that the IHF protein is stable under the conditions where T-jump measurements are done, we carried out thermal denaturation studies of IHF in the presence and absence of H' DNA, using UV circular dichroism (CD) measurements (**Chapter 2.2.2.)**. These studies augment our previous protein-only denaturation studies using near- and far-UV CD, as well as protein and complex denaturation studies using intrinsic Tyr fluorescence (26). Altogether, these studies demonstrate that, while IHF has a melting temperature of ~55-62 °C in the absence of DNA, it is significantly stabilized when DNA is bound, with the melting temperature of IHF in the complex

shifting to ~70 °C (Chapter 2.2.2; Figure 5.4). Therefore, in the ~30-60 °C temperature range where the T-jump measurements are done, the IHF protein in the complex remains stable.

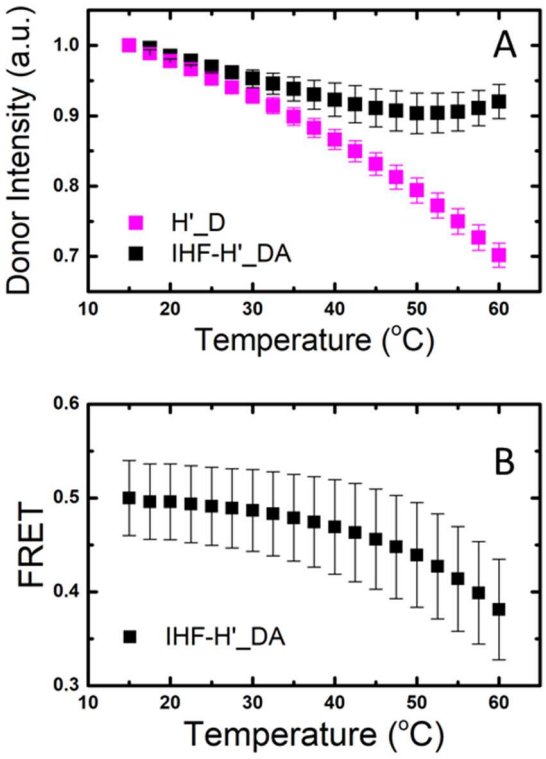


Figure 5.3. Thermal stability of IHF-H' complex. (A) The donor intensity of the H' DNA-only sample (pink), IHF-H' sample (black). The intensities are the area under the measured spectra, integrated from 510-535 nm, and are normalized such that the two data sets match at 15 °C. The error bars are standard deviations from 3 independent sets of measurements. **(B)** FRET of the IHF-H' complex measured over the temperature range 15-60 °C, computed as described in Section 5.2.3.

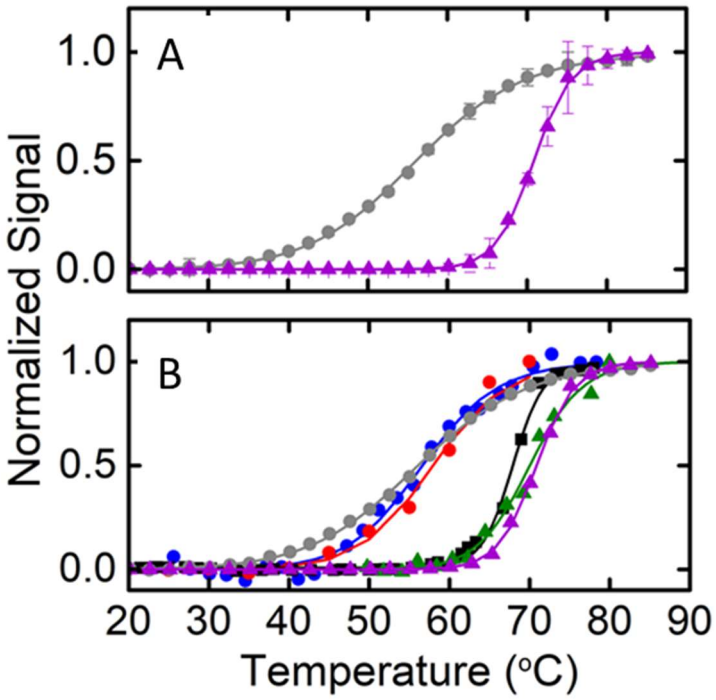


Figure 5.4. Thermal melting profiles of IHF with and without bound H' DNA substrate. (**A**) The normalized change in ellipticity at 222 nm is plotted as a function of temperature for IHF (30 μM; gray circles) and IHF-H' (12.5 μM:15 μM; purple triangles). The continuous lines are a fit to the melting profiles using a two-state van't Hoff relation for fraction unfolded that yield the melting temperatures (T_m) of 55.9 ± 1.5 °C and 70.8 ± 0.8 °C, respectively. (**B**) The data shown in panel (**A**) are plotted together with previous measurements of IHF stability, from Vivas et al. (26): (red circles) ellipticity of IHF (30 μM) at 276 nm ($T_m \approx 58$°C); (blue circles) Tyr fluorescence intensity in IHF (10 μM) at 310 nm, with excitation at 276 nm ($T_m \approx 57$°C); (green triangles) Tyr fluorescence emission intensity in IHF-H' (5 μM:5 μM) complex ($T_m \approx 70$°C); and (black squares) absorbance of H' substrate (4.5 μM) at 266 nm ($T_m \approx 68$°C). All measurements were in 100 mM KCl.

5.3.3. Time-resolved FRET measurements of IHF-DNA conformational dynamics in response to laser T-jump perturbation

Relaxation kinetics in IHF-DNA complexes in response to a laser T-jump perturbation were measured using time-resolved FRET. A detailed description of the T-jump apparatus and approach can be found in Chapter 2.4.7-9. Briefly, ~10 ns IR laser pulses were used to rapidly increase the temperature of a small volume of the sample by 5-10 °C within the duration of each

pulse. The temporal response of the ensemble of molecules as they re-equilibrated from the conformational distribution characteristic of the initial temperature (T_i) to that of the new higher temperature (T_f) was monitored by recording the donor fluorescence emission, as described previously (26, 27). The temperature of the heated volume eventually decayed back to the equilibrium temperature T_i, which was maintained by a water bath.

The results from T-jump studies on the IHF-H' complex in 100 mM KCl salt conditions are summarized in Figure 5.5. A representative kinetic trace measured on the complex (IHF bound to H'_DA) is shown in Figure 5.5.A with a final temperature (T_f) of ~50 °C in response to a ~9 °C T-jump. A control measurement, performed on donor-only H' (H'_D) under similar T-jump conditions, is shown in Figure 5.5.B. The initial rapid drop in intensity in the control sample (H'_D) reflects the decrease in the quantum yield of the donor in response to the T-jump to a higher temperature (Figure 5.5.B), consistent with the decrease in donor fluorescence with increasing temperature observed in equilibrium measurements on these control samples (Figure 5.3.A, pink). This initial drop in donor intensity measured in the T-jump apparatus and the corresponding equilibrium data is used to estimate the size of the T-jump (see Chapter 2.4.5). The relaxation traces measured on the control samples reflect the heat dissipation from the heated volume of the sample and the decay of the temperature back to the equilibrium temperature T_i. These decays are well described by the "T-jump recovery" function (see Eq. 2.12. in Chapter 2.4.5.) and show that the temperature of the heated volume decays to the temperature of the bath with a characteristic time constant of ~206 ± 24 ms. Thus, we can characterize kinetics of conformational changes that fall within the time window of ~20 μs (the dead-time of our instrument) up to at least 50 ms (before the T-jump recovery signal starts to overlap with any relaxation kinetics).

In the case of the IHF-DNA complex, the donor intensity exhibits a rapid drop immediately after the IR pulse as in the control samples as well as the slow decay back to the pre-laser intensity levels. Additionally, these samples also show conformational relaxation kinetics in the time window of ~20 μs – 10 ms; these kinetics reflect the change in population of the complexes from the ensemble of conformations at the initial (low) temperature T_i to that at the final (high) temperature T_f. In IHF-DNA complexes, the DNA unbends when the sample is heated up, and the FRET decreases (Figure 5.3.B.); this decrease in FRET is reflected in an increase in the donor intensity (Figure 5.3.A; black). Therefore, the donor intensities in our T-jump traces for the complexes increase during the conformational relaxation process from a level characteristic of the equilibrium population at T_i to that of T_f, and eventually decay to the pre-laser levels with a time constant of ~210 ms, which is the timescale for the T-jump recovery from temperature re-equilibration kinetics observed in the control experiments. We note there that the characteristic time constant of T-jump recovery does not depend in any systematic way on the initial and final temperature conditions (See Figure 2.10.B.).

5.3.4. Relaxation kinetics in the IHF-H' complex are biphasic.

We collected relaxation kinetic traces over a series of initial temperatures, with final temperatures ranging from 33-60 °C. Single-exponential fits to these relaxation traces show deviations at short times, especially at the higher temperatures, and require a minimum of two exponentials to fit the data. The IHF-H'_DA complex relaxation kinetics traces were fit by both bi-exponential fit convoluted with T-jump recovery (Figure 5.5.A, red) and by Maximum Entropy Method (MEM) (Figure 5.5.A, green), as described in Chapter 2.4.7-8. The MEM distribution

shown in Figure 5.5.C describes the distribution of lifetimes needed to fit the representative trace; the two peaks indicate the biphasic recovery.

Relaxation rates and corresponding amplitudes of both phases, obtained from MEM and discrete exponential analysis, were collected for all final temperatures. The relaxation rates and amplitudes for the two phases, from the MEM analysis, versus inverse (final) temperature are shown in Figures 5.5.D,E. The fast phase spans the timescales of 100-300 μs over the measured temperature range of 33-60 °C and the slow phase spans 1-2 ms (Figure 5.5.D). The amplitude of the slow phase is considerably larger at lower temperatures (>80% at 35 °C) and decreases to ~55% at 58 ºC (Figure 5.5.E).

Our group has previously made T-jump measurements on the IHF-H' complex under identical experimental conditions (21, 26, 30, 31). The most recent measurements with similar instrumentation first revealed biphasic relaxation kinetics for the complex (21). Comparison of the rates measured here with these previous measurements show that, while there is general agreement between the two sets of measurements, there are also some discrepancies (Figure 5.6). For example, the activation energies for the slow rates – characterized by the slopes on these Arrhenius plots – are smaller (weaker dependence on temperature) in the current studies compared with the earlier studies. The agreement between the fast phase kinetics is largely within the signal-to-noise. The amplitudes, however, differ slightly. In our studies, the slow phase is more prominent and contributes ~80% of the amplitude at 35°C compared to previous studies which showed only ~65% amplitude at 35°C. The underlying cause of the disparities is unclear, but may be attributable to differences in the protein storage buffer as well as the length of time for which the protein was stored prior to when the measurements were performed. We note here that no differences were detected in the IHF-H' binding affinity measurements between the earlier and current studies,

indicating that we are picking up subtle changes in the overall protein conformations or differences in the glycerol content of our storage buffer that our T-jump (kinetics) studies can pick up with great sensitivity. Nonetheless, the essential results that the kinetics are biphasic remain unchanged and measurements made here, that used the same batch of protein for all the measurements, are self-consistent.

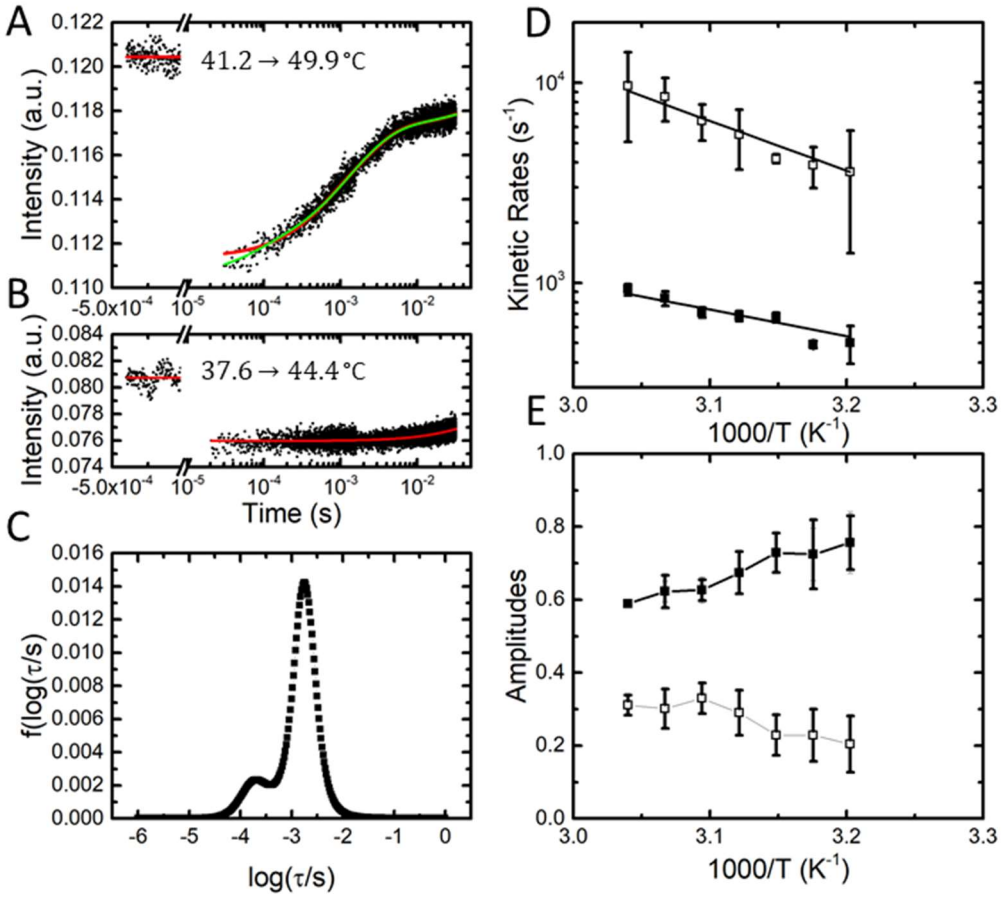


Figure 5.5. T-jump measurements on IHF-H' complex. (**A**) Donor fluorescence emission intensities of end-labeled H' samples (30 μM) in the presence of IHF (40 μM) measured in response to a T-jump perturbation are plotted as a function of time. Relaxation traces show two phases for a T-jump perturbation from 41.2°C to 49.9°C. Relaxation traces were fit by double-exponential with recovery fits (red) and MEM (green). (**B**) Control T-jump measurements on donor-only labeled H'_D is shown. T-jump measurements on these samples do not exhibit any relaxation kinetics, instead they show only the much slower T-jump recovery kinetics. (**C**) Distribution of relaxation times obtained from MEM fit to the T-jump relaxation kinetics shown in panel (A). The relaxation rates for the two phases, k_{fast} and k_{slow}, were obtained from the peak positions, and the amplitudes for the two phases were obtained from the area under the Gaussian curves, normalized to yield a total amplitude of one. For the relaxation trace plotted in (A), the two rates were determined to be 670 ± 37 and 4170 ± 210 s^{-1}, with corresponding amplitudes of 0.77 ± 0.06 and 0.23 ± 0.06. (**D**) Kinetic rates for the two relaxation phases (k_{fast}, open symbols; k_{slow}, filled symbols) in the H'-

IHF complex obtained from MEM analysis are plotted versus inverse temperature for all the measurements. Error bars represent the standard deviation of two independent measurements. The lines are Arrhenius fits to the fast and slow relaxation rates. (**E**) Corresponding amplitudes of the two relaxation phases (fast: open symbols; slow: filled symbols) obtained for the H'-IHF complex are plotted versus inverse temperature; the lines are added to guide the eye.

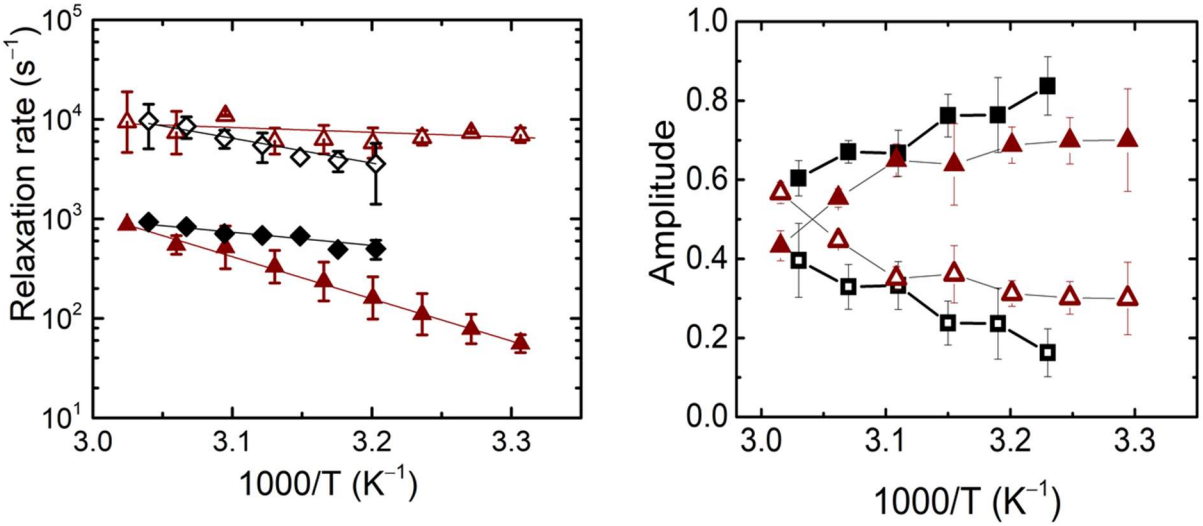


Figure 5.6. **Comparison of new and old T-jump relaxation rates for IHF-H'. (A)** The relaxation rates and **(B)** amplitudes for the fast (open symbols) and slow (filled symbols) phases measured for IHF–H′ in this study (black diamonds), reproduced from **Figure 5.5.D-E**, and from Velmurugu et al. (maroon triangles) (21).

5.3.5. The fast phase is nonspecific "interrogation" and the slow phase is the specific "recognition" step

Previous T-jump studies in our group, on several variants of the H' sequence (summarized in Figure 5.7.) showed that the fast phase rates were unchanged by modifications to the DNA sequence, while the slow phase rates were (a) accelerated by the introduction of mismatches at the kink site to increase DNA flexibility (Figure 5.7; TT8AT) but (b) unaffected by modifications made to DNA away from the kink site (Figure 5.7.; H'44A) (21). Thus, the fast phase was attributed to bending of DNA during nonspecific "interrogation" of the DNA site and the slow

phase was identified as the specific "recognition" step. Comparison of the slow phase rates in the IHF-H' complex with single A:T base pair opening rates (Figure 5.7; red) in DNA by itself showed similar time scales for these processes, suggesting that spontaneous kinking of the DNA at sites with weak stacking interactions could be the rate-limiting step in the recognition process (21, 30). Here, we employ the 7-deaza adenine analog (7dA) substitutions at key sites to investigate the impact of inhibiting Hoogsteen formation on IHF-H' stability and the interrogation/recognition kinetics.

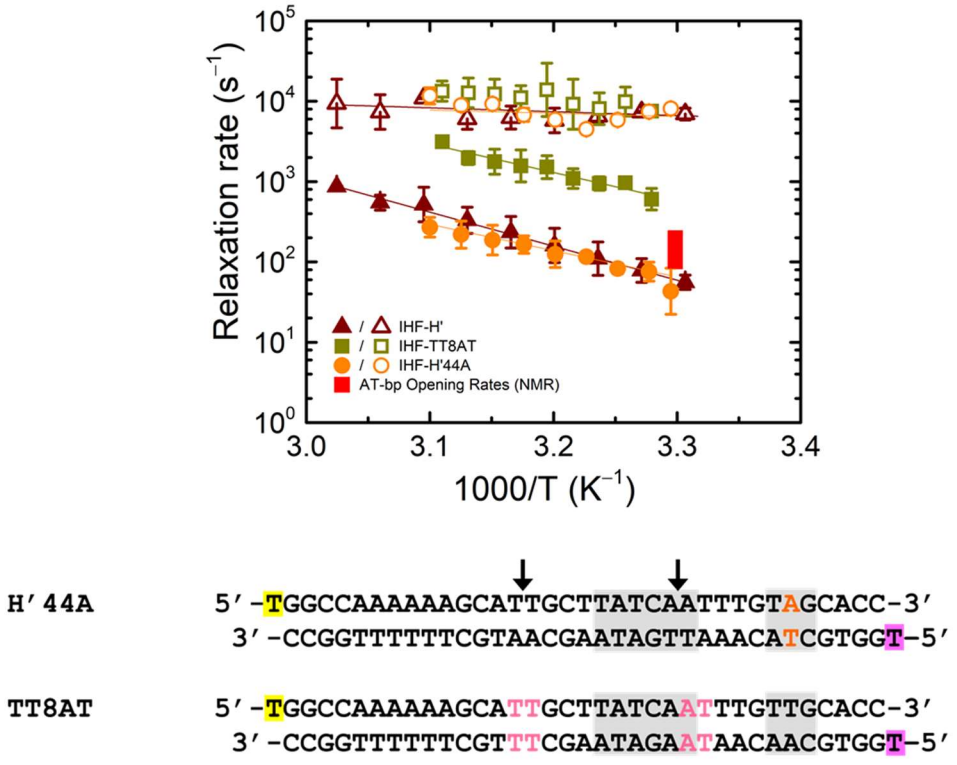


Figure 5.7. Comparison of T-jump relaxation rates for IHF-H' (and H' variants) and base-pair opening rates. The relaxation rates for the fast (open symbols) and slow (filled symbols) phases measured for IHF–H′ (maroon triangles), IHF-TT8AT (olive green squares), and H'44A (orange circles). The corresponding sequences of H' and its variants are shown below. In TT8AT, 2-bp mismatches were introduced at both kink sites at the positions shown in sequences in pink; in H'44A, a T→A mutation was made in the TTG region of the consensus sequence at the position shown in red. The red vertical bar is the A:T base-pair opening rate for the H′ sequence, at the site of one of the kinks, from Dhavan et al. (32), from NMR measurements of imino-proton exchange.

5.3.6. IHF-H' binding affinity unaffected by 7dA substitutions at kink sites but significantly enhanced when in the TTG consensus region

To examine whether transient Hoogsteen base pairing at the DNA kink sites plays a role in IHF binding and kinetics, we made single 7dA substitutions for adenines within the H' sequence. 7dA inhibits a hydrogen bond between adenine and thymine that is needed to stabilize Hoogsteen pairing (Figure 5.2.B) and disfavors the formation of a Hoogsteen pair at that site. Substitutions were made adjacent to one of the kink sites, within the consensus sequence of the H' sequence (as in 7dA56 and 7dA57; Figure 5.2.C) and away from the kink site, within the TTG region of the consensus sequence (as in 7dA7; Figure 5.2.C). Both A56 and A7 substitutions were chosen to fall within CA/TG steps known to exhibit high propensity for Hoogsteen formation (1, 2).

First, we examined the stability of the IHF-H' complexes in the absence and presence of 7dA substitutions. Binding affinity measurements were performed on all complexes as described in Chapter 2, and the results are summarized in Figure 5.8.A. As discussed in Chapters 3 and 4, the sub-nM binding of IHF-H' complex in 100 mM KCl, where the T-jump studies are performed, make it difficult to directly measure the binding affinities from titration experiments that require nM concentrations of DNA (21, 33). To get around this limitation, we performed titration studies at higher salt concentrations, and then extrapolated the specific K_ds measured as a function of [salt] to 100 mM KCl. From here on, because nonspecific binding is not discussed, specific binding affinities will be described use the simplified notation K_d rather than $K_{d,sp}$.

Salt-dependent binding affinity studies revealed essentially overlapping K_d values for IHF-H' and IHF-7dA56 complexes over the entire salt concentration range of these measurements, within the margin of error (Figure 5.8.A and Table 5.1). The IHF-7dA57 complex, which was measured only at 300 mM KCl, also showed no change in binding affinity compared with IHF-H'

and IHF-7dA56. These results were unexpected as we anticipated that inhibiting Hoogsteen formation at the kink site would have some impact on IHF-H' complex stability, if Hoogsteen pairing at that site facilitated kinking. In contrast, the IHF-7dA7 complex revealed a nearly 40-fold decrease in the K_d value reflecting a significant strengthening of the IHF-DNA complex. Again, this result was at first glance unexpected, since this region of the DNA maintains B-DNA conformation in the bound complex and should not be affected by Hoogsteen inhibition. However, this substitution was made in the TTG region of the H' consensus that has been shown previously to be highly sensitive to sequence modifications, as we will discuss below (21, 25, 34, 35).

Next, we examined how these complexes respond to temperature increases. The donor intensity versus temperature profiles for double-labeled DNA (DNA_DA) in complex with IHF are shown for IHF-H', IHF-7dA56 and IHF-7dA7 complexes in Figure 5.8.B, together with the control experiment on the donor-only labeled H' sequence (denoted H'_D). As discussed above (see Figure 5.3), the decrease in donor intensity in H'_D reflect the temperature effects on the quantum yield of the donor. Donor intensities in complexes with DNA_DA initially mirror the decrease in quantum yield seen in the H'_D sample, at low temperatures, but then start to increase at higher temperatures due to unbending of the DNA which is accompanied by a decrease in FRET (Figure 5.8.B.). Complexes IHF-H' and IHF-7dA56 display nearly identical donor intensity and FRET E profiles, indicating that thermal disruption of the bent state in the IHF-7dA56 complex behaves identical to the IHF-H' complex over the temperature range 15-60 °C. In constrast, the donor intensity profile of IHF-7dA7 (blue) falls somewhere in between the IHF-H' data (black) and the H'_D data (pink), indicating a greater resistance to thermal-induced unbending in IHF-7dA7 complex and a correspondingly smaller change in FRET E for this complex with increase in temperature. These temperature profiles are commensurate with our binding affinity data and show

that 7dA7 binds more tightly to IHF and retains a more bent conformation even as the temperature is raised in comparison with IHF-H'.

We note here that these binding studies are consistent with a recent report by the Al-Hashimi and co-workers (22) that also showed no change in stability for the IHF-H' complex when 7dA substitutions were made at positions A56 and A21 (at the other kink site), although the K_d values reported in that study were 1-3 nM for these complexes, measured in 100 mM NaCl, in contrast to the much lower value of ~25 pM that we obtain at 100 mM KCl, when extrapolated from binding isotherms measured at higher salt. Discrepancies in specific binding affinity values obtained from different measurement techniques were thoroughly discussed in Chapter 4.

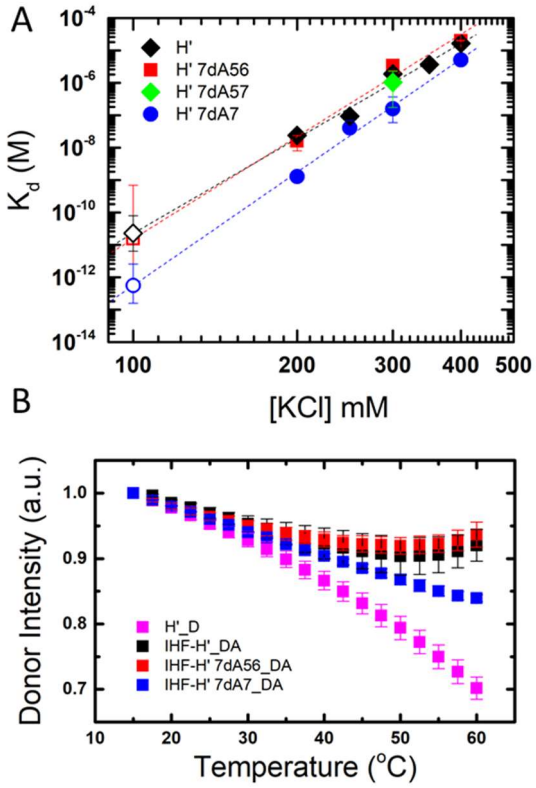


Figure 5.8. Ionic and thermal stability of IHF-H' complex in absence and presence of 7-deaza adenine modifications. (**A**) Dissociation constants (K_d) are plotted versus KCl concentrations for IHF-H' complex (black), IHF-H' 7dA56 (red), IHF-H' 7dA57 (green), and IHF-H' 7dA7 (blue). The filled symbols are K_d values determined by fitting equilibrium binding isotherms (Chapter 3), measured with 50 nM DNA titrated against varying concentrations of IHF. The dashed lines are linear fits to log(K_d) versus log([KCl]) data (for

all samples except IHF-H' 7dA57, which was measured only at 300 mM KCl). The open symbols are the K_d values at 100 mM KCl, extrapolated from this linear fit. The upper and lower limits of the extrapolated point represent the standard deviation determined by varying the K_d value at 100 mM KCL until χ^2 of the minimized fit changes by 1 (36, 37). (**B**) The donor intensity of the H' DNA-only sample (pink), IHF-H' sample (black), IHF-H' 7dA56 sample (red) and IHF-H' 7dA7 sample (blue) versus temperature. The intensities are the area under the measured spectra, integrated from 510-535 nm, and are normalized such that the two data sets match at 15 °C. The error bars are standard deviations from *n* independent sets of measurements [H' DNA-only (*n*=4), IHF-H' (*n*=3), IHF-H' 7dA56 (*n*=3) and IHF-H' 7dA7 (*n*=2)].

Table 5.1. K_d of IHF-DNA complexes under different ionic conditions.

Complex	K_d (300 mM [KCl])	K_d (200 mM [KCl])	K_d (100 mM [KCl])*
IHF-H'	$1.9^{+2.4}_{-1.6}$ µM	24^{+32}_{-18} nM	23^{+31}_{-17} pM
IHF-7dA56	$3.4^{+5}_{-2.4}$ µM	16^{+24}_{-8} nM	15^{+32}_{-7} pM
IHF-7dA57	$1.1^{+2.4}_{-0.3}$ µM	n.d.	n.d.
IHF-7dA7	$0.16^{+0.21}_{-0.1}$ µM	1.3 nM**	$0.6^{+0.8}_{-0.4}$ pM

*Values in this column are extrapolated from a linear fit of log([KCL]) vs. log(K_d) data
**Only one measurement performed

5.3.7. 7dA substitution at kink site slightly increase rate of slow phase

While the binding studies reported above indicate that Hoogsteen formation at the kink site is not a determining factor for the stability of the IHF-H' complex, these equilibrium studies do not provide information on whether transient Hoogsteen formation could be facilitating DNA kinking that IHF could sense during the interrogation and recognition steps. We therefore performed T-jump measurements on IHF-7dA56 to investigate if inhibiting Hoogsteen formation at this kink site had any impact on the conformational relaxation kinetics of the complex. The T-jump results on IHF-7dA56 are summarized in Figure 5.9. Our analysis revealed biphasic kinetics, as in IHF-H' (Figure 5.9.C). The fast phase rates were identical to within the noise for IHF-7dA56 and IHF-H' complexes over the measured temperature range; the slow phase rates increased by ~2-fold for IHF-7dA56 relative to IHF-H' (Figure 5.9.D). The amplitudes in the two phases

remained unaffected by the 7dA modification at this kink site (Figure 5.9.E). From these results we conclude that neither our T-jump nor our binding studies on IHF-H' with 7dA substitutions at the kink site provide evidence to support that transient Hoogsteen pairing at that site facilitates DNA kinking during complex formation. As discussed above, if Hoogsteen pairing played a role, then 7dA modifications at the kink site should have slowed down the observed rates, just as increasing DNA flexibility at the kink sites by introducing mismatches (as in TT8AT; Figure5.7) accelerated the slow phase rates. We attribute the 2-fold increase in the slow phase rates observed in IHF-7dA56 to increased local flexibility at the 7dA site (see Discussion below).

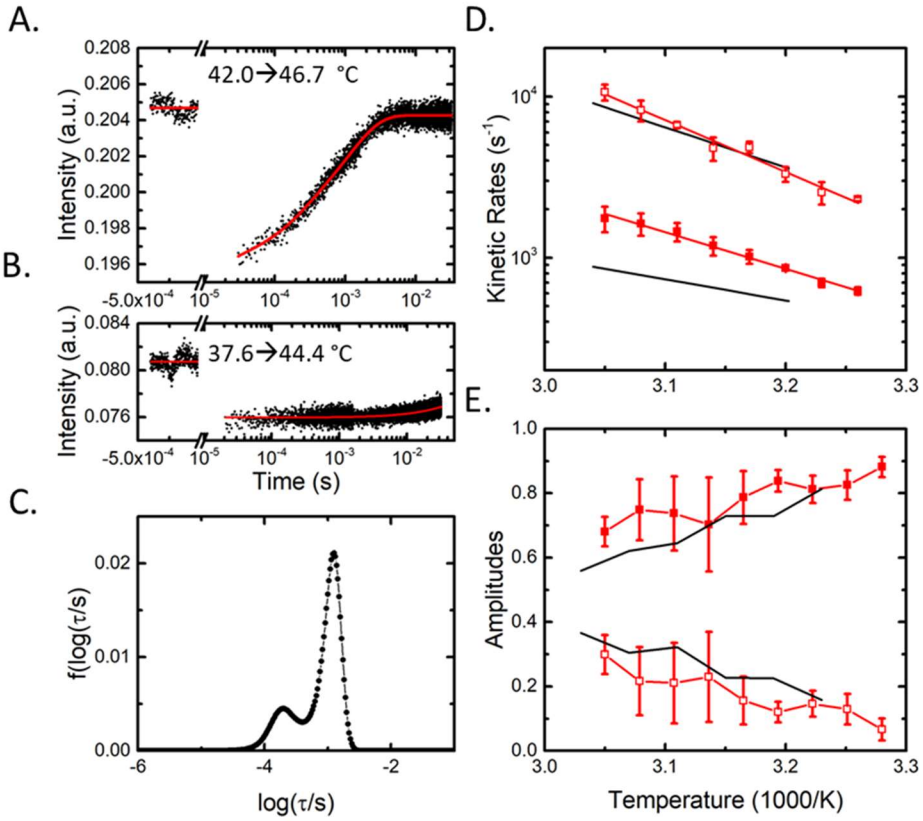


Figure 5.9. T-jump measurements on IHF-7dA56 complex. The data and fits in each panel are as described for Figure 5.5. The T-jump perturbation was from 42.0 to 46.7°C for the complex in (A) and from 37.6 °C to 44.4 °C for the control measurement on H'_D in (B). For the relaxation trace plotted in (A), the MEM analysis yielded two rates at 860 ± 17 s^{-1} and 3310 ± 350 s^{-1}, with corresponding amplitudes of 0.84 ± 0.04 and 0.16 ± 0.04. In panels (D) and (E), the IHF_H' 7dA56 complex data are shown in red; the black lines are for the IHF-H' complex, reproduced from Figure 5.5.D,E, and added here for comparison.

5.3.8. 7dA substitution in TTG region significantly diminishes fast phase amplitudes

Next, T-jump measurements were performed on the IHF-7dA7 complex, with the 7dA modification in the TTG consensus region. The results are summarized in Figures 5.10 and 5.11. In contrast to both IHF-H' and IHF-7dA56, the MEM analyses of the IHF-7dA7 kinetics traces showed predominantly a single peak in the distribution of relaxation times (Figure 5.10.C), indicating that a single kinetic phase was sufficient to describe the IHF-7dA7 kinetics. While a biexponential fit convoluted with the T-jump recovery function could also fit the data, as for the other samples, the relaxation rates and amplitudes of the two components obtained from this analysis showed a significant amount of scatter and lack of reproducibility from one set of measurements to another (Figure 5.11), indicating that the double-exponential fits were over-fitting the data. Therefore, we conclude that the IHF-7dA7 relaxation kinetics are best described as exhibiting a single phase.

The rate constants for this single phase in the IHF-7dA7 complex are similar, albeit slightly accelerated, compared with the slow phase observed in IHF-H' (Figure 5.10.D). We interpret the lack of a fast phase in IHF-7dA7 to be a consequence of the ~40-fold tighter binding in the IHF-7dA7 complex. A tighter IHF-DNA complex favors the specifically bound, fully bent conformation over nonspecifically bound populations. Previous studies showed that destabilizing the complex by weakening the interactions of IHF with the TTG consensus region, as in IHF-H'44A, resulted in an increase in the amplitudes of the fast phase, likely because the balance was tilted in favor of nonspecific binding (21). Here, a tighter complex yields the opposite effect:

nonspecific binding is strongly disfavored in IHF-7dA7, leading to greatly diminished amplitudes in the fast, nonspecific interrogation kinetics.

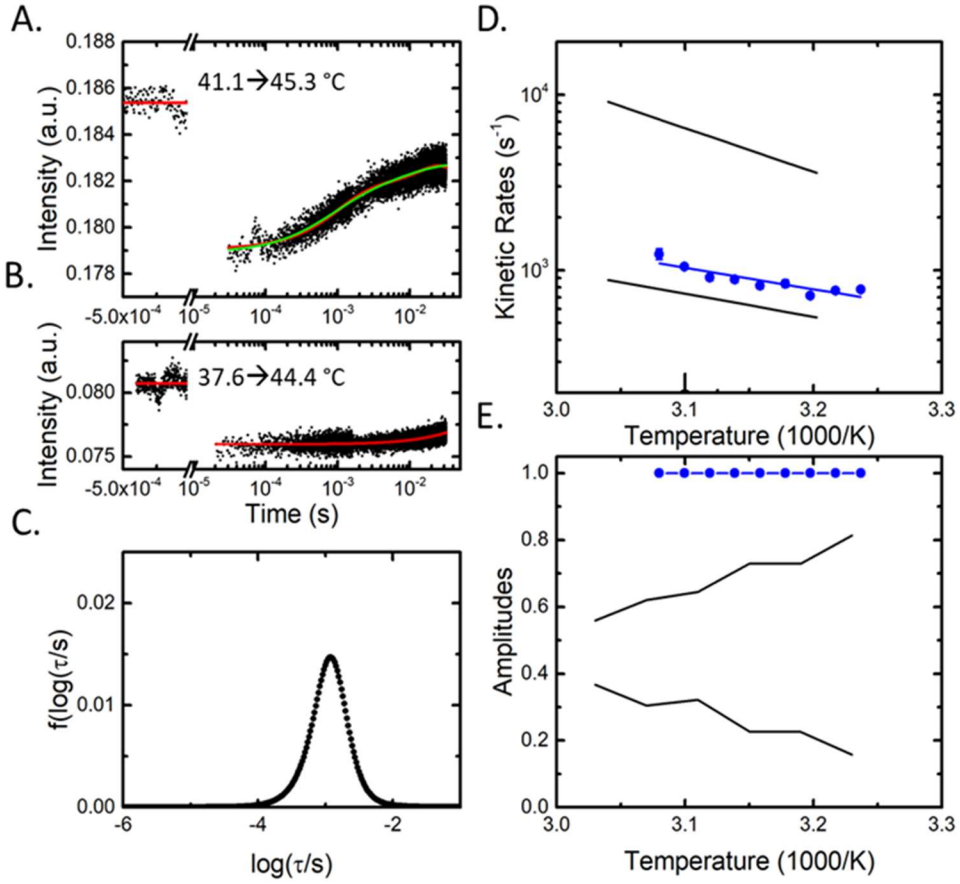


Figure 5.10. T-jump measurements on IHF-7dA7 complex. The data and fits in each panel are as described for Figure 5.5. The T-jump perturbation was from 41.1 to 45.3°C for the complex in (A) and from 37.6 °C to 44.4 °C for the control measurement on H'_D in (B). For the relaxation trace plotted in (A), the MEM analysis yielded a single broad peak (C), with an average relaxation rate of 715 ± 26 s^{-1}. In panels (D) and (E), the rates obtained from the single-peaked MEM distribution for IHF_H' 7dA7 complex data, with 100% amplitude, are shown in blue; the black lines are for the two phases observed in the IHF-H' complex, reproduced from Figure 5.5.D,E, and added here for comparison.

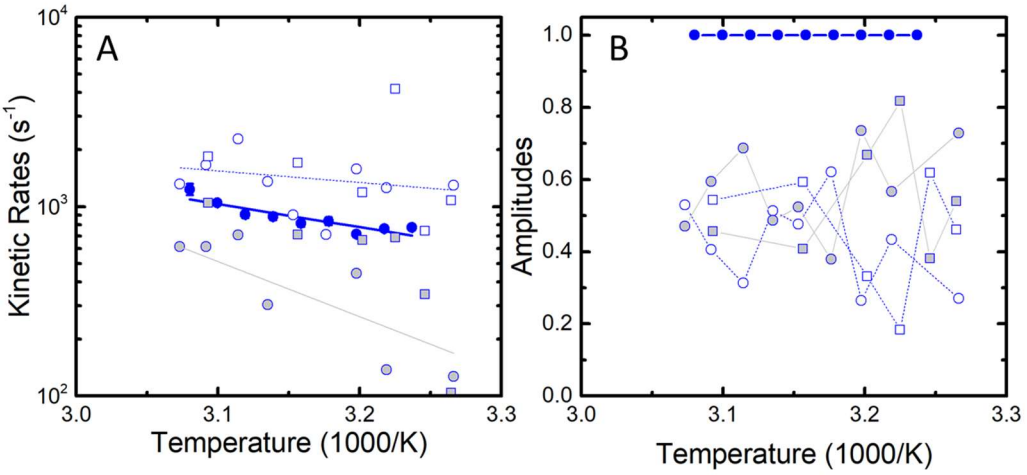


Figure 5.11. Single MEM component versus biexponential fitting of IHF-7dA7 complex. (A) Kinetic rates plotted versus inverse temperature for single component fit from MEM (solid blue circles; averaged over two data sets, reproduced from Figure 5.10.D) and for biexponential fits (blue outline; circles versus squares represent the two independent sets), with fast phase (unfilled symbols) and slow phase (gray filled symbols). Arrhenius fits to the fast (blue dotted line) and slow (gray line) relaxation rates are also shown. (B) The corresponding amplitudes from the biexponential fits are shown; symbols have the same meaning as in panel (A). Lines are drawn to guide the eye.

5.4. Discussion and conclusion

Our studies here were designed to ascertain if transient Hoogsteen base pairing, an alternative to Watson-Crick pairing suggested to be widespread in the genome, plays a role in the ease with which IHF recognizes and binds to one of its target sites, the H' site from λ DNA. 7deaza adenine (7dA) substitutions, designed to inhibit Hoogsteen pairing within DNA, were inserted at several positions in the H' site to examine the impact on IHF binding and DNA bending kinetics. If transient Hoogsteen pairing were to play a significant role, then Hoogsteen inhibition is expected to reduce binding affinity and/or affect the rates of DNA bending within the IHF-bound complex. In none of the 7dA-modified substrates did we see decreased binding affinities or slowing down of the kinetics, as hypothesized. Our equilibrium and T-jump studies therefore provided no evidence to support Hoogsteen bp dynamics as an important contributor to DNA conformational

dynamics in the IHF-H' complex. Instead, we observed no change in stability and a slight increase in the slow phase rates in IHF-7dA56, where the 7dA substitution was made at one of the kink sites. In IHF-7dA7, where the 7dA substitution was made in the TTG consensus region, which is at some distance from the kink site, we observed a ~40-fold strengthening of the IHF-7dA7 complex stability. Our studies therefore revealed that 7dA substitutions designed to inhibit Hoogsteen pairing within DNA had unexpected consequences that we attribute to changes in DNA rigidity at the site of the modifications, as discussed below.

The observed changes to binding affinity and kinetics of the IHF-H' complex are best explained by increased local DNA flexibility as a result of weakened stacking interactions of 7dA with its neighbors. 7deaza modifications have been shown to disrupt DNA rigidity before. NMR studies on DNA constructs including a 7-deaza guanine (7dG) identified weakened base stacking interactions, a key contributor to DNA rigidity, between canonical nucleotides and 7dG relative to canonical guanine (38). The accelerated slow phase rates observed in IHF-7dA56 mirror previous T-jump studies on IHF-TT8AT and other H' variants with mismatches introduced at the kinks – designed to increase DNA flexibility – that also exhibited an increase in the slow phase rates (21). 7dA at this kink site (position 56) in the H' construct appears to similarly increase local flexibility.

The 7dA7 substitution (at position 7) made in the TTG region of the consensus sequence increases binding affinity ~40-fold while diminishing the amplitude of the fast phase in the observed kinetics. In direct contrast, a thymine to adenine substitution within that TTG region (TTG→TAG in H'44A; Figure 5.7) weakens the binding affinity nearly 100-fold and exhibits increased amplitude in the IHF-H'44A fast phase kinetics compared with IHF-H' (21). Crystallography studies on the IHF-H'44A and IHF-H' complexes revealed H'44A to have a different twist distribution within of the TAG region, breaking a series of stabilizing ionic

interactions that the H' sequence would normally make with IHF at that site (25). We therefore postulate that weakened stacking interactions in the TTG region of the 7dA7 construct, from the 7dA substitution, likely increase the "twistability" of the DNA at that site, enabling the stabilizing ionic interactions to be made with greater ease, and thereby favoring the formation of the specific complex.

Our studies complement the growing consensus that subtle modification made to DNA structure can have large effects on DNA flexibility. For example, there is significant evidence from studies involving methylated DNA: GpG methylation may stiffen DNA which can impact binding by EcoRI and cAMP DNA responsive element (39, 40). Similarly, single molecule FRET studies involving numerous epigenetic cytosine modifiers showed a large impact on DNA flexibility and nucleosome binding (41). The 40-fold change in binding affinity seen in the IHF-7dA7 complex underscores how significant the impact a single DNA modified base may have on dictating affinity and position of nucleoid-associated proteins. Finally, our studies point to the difficulties posed when 7dA is used to probe Hoogsteen pairing if the inevitable changes in DNA stacking interactions from the 7dA substitution are not properly accounted for.

References

1. Nikolova, E. N., E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei, and H. M. Al-Hashimi. 2011. Transient Hoogsteen base pairs in canonical duplex DNA. Nature 470(7335):498-502.
2. Alvey, H. S., F. L. Gottardo, E. N. Nikolova, and H. M. Al-Hashimi. 2014. Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. Nat Commun 5:4786.
3. Hoogsteen, K. 1959. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. Acta Crystallographica 12(10):822-823.
4. Sathyamoorthy, B., H. Shi, H. Zhou, Y. Xue, A. Rangadurai, D. K. Merriman, and H. M. Al-Hashimi. 2017. Insights into Watson-Crick/Hoogsteen breathing dynamics and damage repair from the solution structure and dynamic ensemble of DNA duplexes containing m1A. Nucleic Acids Res 45(9):5586-5601.
5. Zhou, H., B. J. Hintze, I. J. Kimsey, B. Sathyamoorthy, S. Yang, J. S. Richardson, and H. M. Al-Hashimi. 2015. New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. Nucleic Acids Res 43(7):3420-3433.
6. Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. K. Moore, J. P. Wang, and J. Widom. 2006. A genomic code for nucleosome positioning. Nature 442(7104):772-778.
7. Blainey, P. C., A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie. 2006. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. Proc. Natl. Acad. Sci. USA 103(15):5752-5757.
8. Gorman, J., A. Chowdhury, J. A. Surtees, J. Shimada, D. R. Reichman, E. Alani, and E. C. Greene. 2007. Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. Mol. Cell 28(3):359-370.
9. Tafvizi, A., F. Huang, J. S. Leith, A. R. Fersht, L. A. Mirny, and A. M. van Oijen. 2008. Tumor suppressor p53 slides on DNA with low friction and high stability. Biophys. J. 95(1):L01-03.
10. Bonnet, I., A. Biebricher, P. L. Porte, C. Loverdo, O. Benichou, R. Voituriez, C. Escude, W. Wende, A. Pingoud, and P. Desbiolles. 2008. Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. Nucleic Acids Res. 36(12):4118-4127.
11. Schonhoft, J. D., and J. T. Stivers. 2012. Timing facilitated site transfer of an enzyme on DNA. Nat Chem Biol 8(2):205-210.
12. Nelson, S. R., A. R. Dunn, S. D. Kathe, D. M. Warshaw, and S. S. Wallace. 2014. Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. Proc Natl Acad Sci U S A 111(20):E2091-2099.
13. Yang, H., Y. Zhan, D. Fenn, L. M. Chi, and S. L. Lam. 2008. Effect of 1-methyladenine on double-helical DNA structures. FEBS Lett 582(11):1629-1633.
14. Yang, H., and S. L. Lam. 2009. Effect of 1-methyladenine on thermodynamic stabilities of double-helical DNA structures. FEBS Lett 583(9):1548-1553.
15. Lu, L., C. Yi, X. Jian, G. Zheng, and C. He. 2010. Structure determination of DNA methylation lesions N1-meA and N3-meC in duplex DNA using a cross-linked protein-DNA system. Nucleic Acids Res 38(13):4415-4425.
16. Freudenthal, B. D., W. A. Beard, L. Perera, D. D. Shock, T. Kim, T. Schlick, and S. H. Wilson. 2015. Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. Nature 517(7536):635-639.
17. Ling, H., F. Boudsocq, B. S. Plosky, R. Woodgate, and W. Yang. 2003. Replication of a cis-syn thymine dimer at atomic resolution. Nature 424(6952):1083-1087.
18. Rice, P. A., S. Yang, K. Mizuuchi, and H. A. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87(7):1295-1306.

19. Patikoglou, G. A., J. L. Kim, L. Sun, S. H. Yang, T. Kodadek, and S. K. Burley. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev 13(24):3217-3230.
20. Nikolova, E. N., H. Zhou, F. L. Gottardo, H. S. Alvey, I. J. Kimsey, and H. M. Al-Hashimi. 2013. A historical account of Hoogsteen base-pairs in duplex DNA. Biopolymers 99(12):955-968.
21. Velmurugu, Y., P. Vivas, M. Connolly, S. V. Kuznetsov, P. A. Rice, and A. Ansari. 2018. Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein. Nucleic Acids Res 46(4):1741-1755.
22. Zhou, H., B. Sathyamoorthy, A. Stelling, Y. Xu, Y. Xue, Y. Z. Pigli, D. A. Case, P. A. Rice, and H. M. Al-Hashimi. 2019. Characterizing Watson-Crick versus Hoogsteen Base Pairing in a DNA-Protein Complex Using Nuclear Magnetic Resonance and Site-Specifically (13)C- and (15)N-Labeled DNA. Biochemistry 58(15):1963-1974.
23. Nikolova, E. N., F. L. Gottardo, and H. M. Al-Hashimi. 2012. Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution. J Am Chem Soc 134(8):3667-3670.
24. Johnson, R. E., L. Prakash, and S. Prakash. 2005. Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase iota. Proc Natl Acad Sci U S A 102(30):10466-10471.
25. Lynch, T. W., E. K. Read, A. N. Mattis, J. F. Gardner, and P. A. Rice. 2003. Integration host factor: putting a twist on protein-DNA recognition. J Mol Biol 330(3):493-502.
26. Vivas, P., Y. Velmurugu, S. V. Kuznetsov, P. A. Rice, and A. Ansari. 2012. Mapping the transition state for DNA bending by IHF. J Mol Biol 418(5):300-315.
27. Kuznetsov, S. V., S. Sugimura, P. Vivas, D. M. Crothers, and A. Ansari. 2006. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc. Natl. Acad. Sci. USA 103(49):18515-18520.
28. Vivas, P., S. V. Kuznetsov, and A. Ansari. 2008. New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. J Phys Chem B 112(19):5997-6007.
29. Vivas, P., Y. Velmurugu, S. V. Kuznetsov, P. A. Rice, and A. Ansari. 2013. Global analysis of ion dependence unveils hidden steps in DNA binding and bending by integration host factor. J Chem Phys 139(12):121927.
30. Kuznetsov, S. V., S. Sugimura, P. Vivas, D. M. Crothers, and A. Ansari. 2006. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc Natl Acad Sci U S A 103(49):18515-18520.
31. Vivas, P., V. Yogambigai, S. V. Kuznetsov, P. A. Rice, and A. Ansari. 2010. Dynamics of DNA-Bending in Binding Site Recognition by IHF. Biophysical Journal 98(3):660a-661a.
32. Dhavan, G. M., J. Lapham, S. Yang, and D. M. Crothers. 1999. Decreased imino proton exchange and base-pair opening in the IHF-DNA complex measured by NMR. J Mol Biol 288(4):659-671.
33. Sugimura, S., and D. M. Crothers. 2006. Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. Proc. Natl. Acad. Sci. USA 103(49):18510-18514.
34. Hales, L. M., R. I. Gumport, and J. F. Gardner. 1994. Determining the DNA sequence elements required for binding integration host factor to two different target sites. J. Bacteriol. 176(10):2999-3006.
35. Lee, E. C., L. M. Hales, R. I. Gumport, and J. F. Gardner. 1992. The isolation and characterization of mutants of the integration host factor (IHF) of Escherichia coli with altered, expanded DNA-binding specificities. EMBO J 11(1):305-313.

36. Roca, J., N. Hori, S. Baral, Y. Velmurugu, R. Narayanan, P. Narayanan, D. Thirumalai, and A. Ansari. 2018. Monovalent ions modulate the flux through multiple folding pathways of an RNA pseudoknot. Proc Natl Acad Sci U S A 115(31):E7313-E7322.
37. Roca, J. 2017. Folding Mechanisms of RNA Pseudoknots Unveiled with Laser Temperature-Jump and Global Modeling Approaches. University of Illinois at Chicago.
38. Ganguly, M., F. Wang, M. Kaushik, M. P. Stone, L. A. Marky, and B. Gold. 2007. A study of 7-deaza-2'-deoxyguanosine 2'-deoxycytidine base pairing in DNA. Nucleic Acids Res 35(18):6181-6195.
39. Nathan, D., and D. M. Crothers. 2002. Bending and flexibility of methylated and unmethylated EcoRI DNA. J Mol Biol 316(1):7-17.
40. Derreumaux, S., M. Chaoui, G. Tevanian, and S. Fermandjian. 2001. Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element. Nucleic Acids Res 29(11):2314-2326.
41. Ngo, T. T., J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, and T. Ha. 2016. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. Nat Commun 7:10813.

Chapter 6.

In search of highly "kinkable" DNA

6.1. Introduction

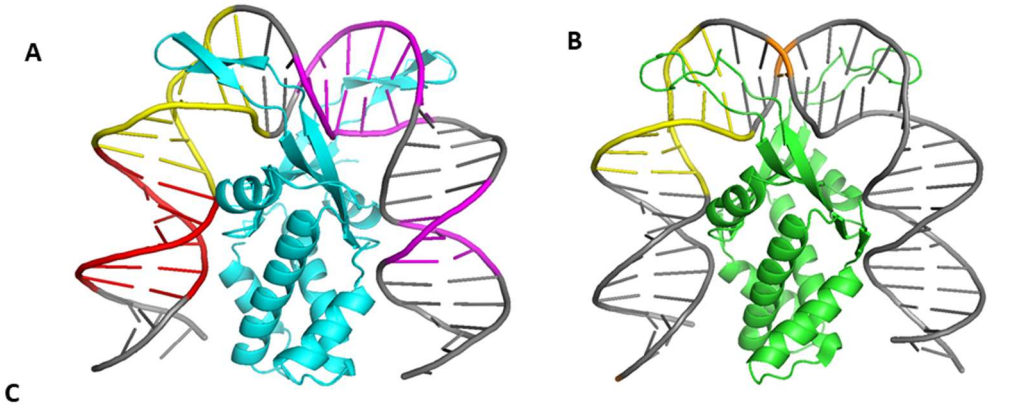
Sequence-dependent deformability is critical to how DNA interacts with proteins that facilitate DNA packaging, transcriptional regulation, replication, and repair. Ease of DNA deformations underpins the indirect readout aspect of site-specific recognition. Even for nonspecific DNA-protein complexes such as the nucleosomes, sequence determines nucleosome positioning within our genomes, thereby modulating gene regulation. The extent of these deformations range from few localized distortions to fully wrapped around proteins (1). However, many of the rules that govern sequence-dependent DNA shape/deformability remain largely untested for DNA deformations of high severity (e.g., >30° kinks).

Several theoretical and experimental developments in the last few decades have tremendously impacted our understanding of the mechanical properties of DNA. These include a continuum wormlike chain description that captures average properties of long DNA molecules, but with increasing evidence of large-amplitude sequence-dependent structural contortions at much smaller length scales relevant for protein binding that are not captured by homogeneous polymer models. Recent advances in elucidating these sequence-dependent properties include knowledge-based energy functions that describe single base-step deformations within harmonic constraints (2-4); the experimental and computational characterization of both synthetic and naturally occurring high affinity nucleosome binding sequences (5-9); analysis of sequence-dependent DNA shapes (10); and all-atom MD simulations by a consortium of computational biophysicists aimed at characterizing DNA fluctuations beyond dinucleotide steps (11-14). Altogether, these studies have contributed greatly to our current understanding of DNA shape and

deformability. However, these studies also underscore the limitations of the current models and highlight the need to develop energy functions that include context dependence and also go beyond harmonic approximations (15-17). The energy costs of the severe DNA deformations seen in many protein-bound structures are not captured accurately by elastic models. Furthermore, large bending fluctuations that deviate from harmonic behavior have been observed even in short DNA oligomers in the absence of bound proteins (18-20).

While nucleosome-preferred sequences have proven extremely valuable in testing models of DNA deformability, the inevitable convolution of DNA shape, intrinsic deformability, and histone-specific interactions, all of which contribute to the sequence-dependent binding energies, complicate the validation of these models. Furthermore, some proteins induce significantly more severe kinks (>50°) than the nucleosome does (21). Therefore an independent set of potentially highly deformable sequences, selected as high-affinity binding sites for a completely different set of architectural proteins that induce severe DNA local distortions would provide a much-needed data set for further testing and developing models of sequence-dependent DNA deformability. To that end, we embarked on a high-throughput "SELEX" (systemic evolution of ligands by exponential enrichment) study coupled with deep sequencing (22-26) to characterize the binding preferences for proteins from the IHF/HU family: IHF and Hbb (Figure 6.1.A,B), starting from a semi-randomized DNA library. Both proteins can severely deform DNA at specific sites and bend the DNA into a "U-turn" shape over 35-bp by kinking the DNA at two locations spaced roughly 9-bps apart (27-29); thus high-affinity sequences are hypothesized to be of higher deformability. Our goal was to identify these high-affinity DNA sequences and to uncover sequence patterns that influence their shapes and deformations away from canonical B-DNA conformations. These studies were inspired by the elegant studies of the late Jonathan Widom and co-workers, who

demonstrated the potential of such SELEX studies in uncovering preferred sequences for binding to histone octamers to form nucleosomes (30-32). Following similar approaches, we aimed to uncover sequence patterns that render the DNA more "kinkable" at sites known to be sharply kinked when in complex with the IHF/HU family of DNA-bending proteins; to examine the intrinsic DNA deformability of these tight binding sequences, particularly at the kink sites; and to provide a robust database for further development of base-step and longer range parameters for modeling DNA deformability.



C

Construct	Sequence	[a][IHF] where binding detected	[a][Hbb] where binding detected	[b]IHF K_d (nM)	[b]Hbb K_d (nM)
H'	GGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC CCGGTTTTTTCGTAACGAATAGTTAAACAACGTGG	≥ 5 nM	---	12 ± 2	145 ± 37
Hbb sequence in crystal structure	TCTTTTAATACTATATGTCATATAGTATTAAAGCT AGAAAATTATGATATACTGTATATCATAATTTCGA	---	---	---	118 ± 27
H'_nAt	GGCCATGCAGGCATTGCTTATCAATTTGTTGCACC CCGGTACGTCCGTAACGAATAGTTAAACAACGTGG	≥ 500 nM	≥1 µM	---	---
DNA library	GGCCATGCAGNNNNNNCTTATCAATTTGTTGCACC CCGGTACGTCNNNNNNGAATAGTTAAACAACGTGG	---	---	---	---

[a] Measurements performed in this study. [b]Dissociation constants reported in Mouw et al. (29)

Figure 6.1. SELEX proteins and library design. Crystal structures are shown for (A) IHF-H' (1IHF) and (B) Hbb-DNA (2NP2). DNA is shown in gray while IHF is shown in cyan (A) and Hbb is shown in green (B). Sequence elements are highlighted to indicate position such as the IHF-consensus sequence (A, yellow), the A-tract (A, red), the randomized region of the library (A/B, yellow), and a T:T mismatch site in the Hbb crystal structure sequence (orange). (C) DNA sequences, protein concentrations at which binding was detected in gel shift in this study, and gel shift binding affinities reported in (29) are shown. Black arrows are used to indicate the kink sites. Sequence elements are colored to match (A) and (B).

IHF was discussed extensively in earlier chapters of this thesis. Hbb is a closely related member of the IHF/HU family. This small nucleoid-associated homodimer is found in *Borrelia burgdorferi*, a spirochete responsible for causing Lyme disease in humans (33, 34). Hbb is the only known member of this family in *B. burgdorferi*, and was originally identified by its ability to complement phage DNA packaging defects in IHF- and HU-deficient *E. coli* (35). Like IHF, Hbb also recognizes a specific DNA binding site by "indirect readout" and bends its specific site by >160° as described in the crystal structure (29), which aids in chromosomal compaction and in the assembly of higher order nucleoprotein complex. Natural IHF-binding binding site, H', has been shown to bind specifically to Hbb with high affinity, suggesting similar binding site preferences and recognition mechanism between the two cousins (29).

These SELEX studies enabled us to separate high-affinity DNA sequences ("winners") from low-affinity sequences ("losers") for both IHF and Hbb. The sequences were then analyzed in multiple different ways. First, we identified dinucleotide steps and tetranucleotide sequences in the randomized region that were more probable among winners or losers. Sequences were also analyzed with current theoretical models such as DNA shape parameters, evaluated by TFBSshape (36, 37), and DNA deformability computed using elastic models for distortions away from B-DNA using nearest-neighbor base-step parameters determined by Dr. Wilma Olson and her group, and referred to here as the Olson-98 parameters (2). A few key results stand out: (1) "A-tracts", poly(dA:dT) stretches of greater than 3 As in a row, were found to be highly prevalent among the losers for both IHF and Hbb, indicating that such stretches inhibit the formation of the sharply kinked DNA; (2) preferred sequences for IHF had distinct minor groove width profile, commensurate with previous studies that indicate "shape readout" (38); (3) a correlation, albeit weak, was observed for IHF (though not for Hbb) between binding affinity and ease of

deformability as obtained from the elastic modeling. Notably, while little sequence correlation was observed among the winners and losers for IHF and Hbb, A-tracts at or near the kink sites were found to be disfavored by both proteins, further supporting the theory that the unique properties of A-tracts make them highly important sequence elements for genetic regulation (39, 40). Furthermore, there were similarities in the shape patterns for the winners/losers for both proteins, indicating that different sequence combinations could yield the same or similar shapes. Interestingly, binding site favorability for IHF showed dependence with computed DNA bending energies when the DNA was forced to adopt the shape seen in the IHF crystal structure. However, this correlation was lacking in the Hbb results. The apparent lack of correlation between binding affinities and ease of DNA kinking underscores the need to further refine our SELEX protocol to reduce the noise, but may also be indicative of the limitations of elastic models based on harmonic approximations. Harmonic models imposes a high energy penalty on the sharp 90° bends observed in the IHF-H' crystal structure, despite the evidence that kinks are more probable and less energetically costly than suggested by a wormlike chain description (19). The results presented here lay a good foundation for systematically and quantitatively approaching questions surrounding DNA of high deformability.

6.2: Overview of the SELEX approach

Here, we describe the design of our "SELEX" study aimed at identifying DNA oligomers that bind with high affinity to IHF or Hbb and to obtain from these selected high-affinity binders sequence patterns that may render high "kinkability" at key sites. SELEX has been widely used to characterize the binding affinities for transcription factors (24, 26, 41-44). We used the solution-based Cognate Site Identifier (CSI) procedure described by Dr. Aseem Ansari and co-workers (42-

44) to interrogate large sequence space to identify protein binding sites, and used Illumina Genome Analyzer to sequence the samples (24, 41, 43). The CSI procedure is illustrated in Figure 6.2.

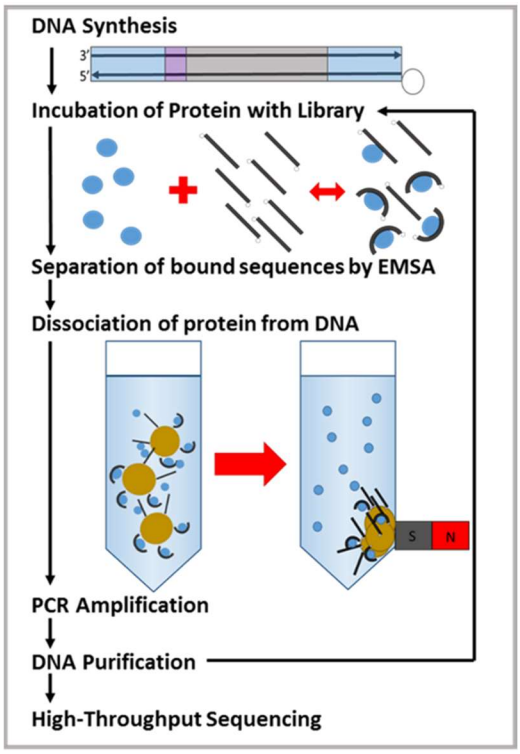


Figure 6.2. Description of the SELEX process. The DNA library is designed with 41-bp sequences flanked by constant PCR-primer sequence (blue). One primer is labeled with biotin (open gray circle). Within the 41-bp region, a 6-bp barcode (purple), unique to a given library, is included to allow multiple libraries to be pooled and sequenced simultaneously, if needed. The remaining 35-bp region (gray) is the protein binding site, within which a 6-nt region is randomized, as shown in Figure 6.1. The library was ordered from IDT. dsDNA library sequences were synthesized by PCR and incubated with the protein; the bound sequences are separated from unbound sequences by electrophoretic mobility gel-shift assays (EMSA); the protein-DNA complexes are extracted from the gel and then transferred to a high-salt buffer to dissociate the protein from the DNA; streptavidin-coated Dynabeads bind to biotin on the released DNA, which are then "pulled-down" by a magnet; the supernatant containing the dissociated protein is removed. After multiple rounds of washing to ensure complete dissociation, the DNA is amplified by PCR and purified to remove the Dynabeads. Bound sequences go through more rounds of selection or sent for sequencing.

The DNA library was based on a 35-bp DNA sequence, H'_nAt (Figure 6.1.C), previously shown to bind specifically to IHF (45) (see also Chapter 3). The sequence of H'_nAt is identical to H', a cognate site for IHF from phage λ DNA and also shown to bind to Hbb (29,

46, 47), but with the A-tract on the flanking arm replaced by the sequence from another binding site of IHF, the H1 site, that lacks an A-tract, as described in Chapter 3. Our rationale for choosing a sequence that lacked the A-tract was to weaken the interactions between the A-tract and the protein that contribute significantly to keeping that flanking arm of the DNA held against the side of the protein in the specific complex, as discussed in Chapter 3 (45). We reasoned that, in the absence of those stabilizing interactions, sequence selectivity at the randomized site would be determined primarily by the "kinkablity" of that site. The DNA library consisted of H'_nAt sequences that included a randomized 6-bp region at one of the protein-induced kink sites that does not fall within the consensus region for either protein (indicated by NNNNNN in Figure 6.1.C). Thus, our DNA library consisted of $4^6 = 4{,}096$ unique sequences in all, with the randomized region located at a site that is expected to be kinked in the specific complex. The nonrandomized region included the consensus region for both IHF and Hbb and was left changed so as to position the protein correctly for site-specific binding and to ensure that the 6-nt randomized region aligned with where the DNA is kinked in the specific complex. The randomized site was separated from the consensus region by 2 nt to minimize the impact of the randomized sequence on the structure and deformability of the consensus region. The drawback of this design is that the randomized region is not symmetric about the kink.

SELEX was performed with IHF and Hbb using the same DNA library. As illustrated in Figure 6.2., the DNA library underwent three rounds of selection by EMSA gel shift to separate sequences that bound to IHF ("binders") from sequences that did not bind ("nonbinders"). In each round, EMSA gel shift assays were performed with six samples including two Cy3-labelled control samples and four DNA library samples. Labeled control DNA was run without (lane 1) and with 2 uM of protein (lane 2). The purpose of the labelled control was to: i) ensure that in

each round SELEX, protein binding occurred and ii) provide guidance for where to cut the unlabeled DNA library (invisible to fluorescence-based gel imaging) from the gel. The separation of binders from non-binders from the DNA library was achieved by performing gel-shift assays in 4 different lanes simultaneously, with 25 nM DNA from the library in each lane and 4 different protein concentrations that extended from the highest concentration of specific binding detected (absent of nonspecific binding) to nearly 10-fold below that concentration to distinguish sequences of higher affinity. This range of protein concentrations allowed for detection of both high, midrange, and low affinity DNA sequences. Both bound and unbound sequences within each lane that had the DNA library were extracted from the gel, amplified with PCR, and purified. The library of winners became the library for the next round of selection. At the end of three rounds of selection, all winners, all losers, and the initial DNA library were collected and sequenced at the UIC sequencing center – RRC, with a depth (i.e. number of sequence reads) of >100,000 sequences .

SELEX performance by each sequence was characterized by "sequence enrichment" defined as:

$$\text{sequence enrichement A} = \frac{\text{fractional population of Sequnce A after SELEX}}{\text{fractional population of Sequnce A in original library}} \qquad \text{Eq. 6.1.}$$

If the logarithm of sequence enrichment is greater than zero, the sequence appears more frequently in the final library, indicating favorability in binding; if less than zero, binding is less favorable. Previous studies have shown that the sequence enrichment correlates strongly with relative binding affinities (22, 23, 26).

An complete overview of SELEX methods is found in Chapter 2.5. Oligos used in this study are listed in Table 6.1.

6.3. Results

6.3.1. Determining protein concentration range for SELEX studies

Before performing SELEX studies, we characterized the binding affinities of IHF and Hbb for a few control sequences using gel mobility shift assays in order to determine the appropriate concentration range for these studies. Both IHF and Hbb bind specifically to the H' site from phage λ DNA (29), a cognate site of IHF (45-48, 56-59). The binding affinity of IHF for the H' site has been measured using several different techniques, as discussed in detail in Chapter 4, with binding affinities that range from ~25 pM to ~100 nM. Therefore, we wanted to ascertain what the measured affinity would be under our gel mobility shift assay conditions. We performed gel shift assays using 25 nM DNA and protein concentrations that ranged from 5 nM to 2 μM. The results for IHF are summarized in Figure 6.3. Consistent with the much higher K_d values (~12 nM) reported by previous gel-shift assays for the IHF-H' complex at similar salt conditions (50 mM NaCl) as used in this study (29), we did not detect much binding for 5 nM [IHF] and approximately half the complex formed at 50 nM [IHF] (Figure 6.3A; right). These results are in good agreement with previous reports (28, 60-63). Similar studies with H'_nAt that lacked the A-tract of H' showed no significant binding until ~ 2 μM [IHF] (Figure 6.3A,B; left), though titrations with greater point density detected trace amounts of binding by 1 μM [IHF] . To verify that nonspecific binding was not contributing to the bound fraction observed in these studies, we additionally performed binding studies using a nonspecific DNA site, H'_nAt_scrambled (Table 6.1). No binding was detected for this construct up to at least ~2 μM [IHF] (Figure 6.3B; right). Thus, SELEX was performed with IHF concentration within the range of 0.25 -2 μM to optimize the range for specific binding and to minimize nonspecific binding. Furthermore, with the consensus region intact, the longer DNA constructs in the DNA

library (see Table 6.1) are optimally positioned to bind to IHF as in the IHF-H' crystal structure, with the randomized region positioned correctly at the kink site. Hbb also binds specifically to the H' site with binding affinities characterized by mobility shift assays to be roughly K_d ~100-200 nM (29). Our gel shift assays showed binding of Hbb to H'_nAt in the range [Hbb] ~1 - 2.5 μM and no nonspecific binding to nsDNA for [Hbb] up to 2 μM (data not shown). Thus, Hbb SELEX studies were performed using the same DNA library, with protein concentration within the same range of 0.25 -2 μM as for the IHF studies.

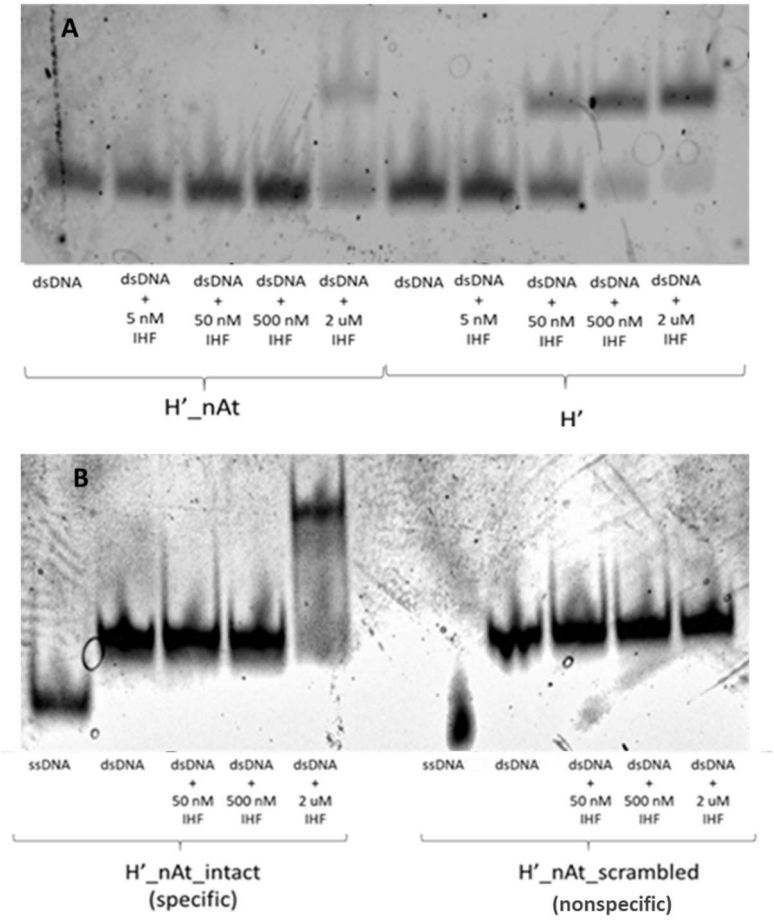


Figure 6.3. EMSA binding studies on specific and nonspecific IHF complexes. (A) H' binding and H'_nAt binding to IHF is shown. (B) H'_nAt (intact) is compared to nonspecific DNA site, H'_nAt_scrambled. Scrambling of nucleotides preserves GC-content but removes consensus sequences.

Table 6.1. Oligo sequences. All sequences are listed 5'→3'. IHF/Hbb binding sites are highlighted in gray. Ns represent any nucleotide (A/T/C/G). Attachments Cy3 (X) and Biotin (**Y**) are shown.

Library	5' - CCT GAT CCT ACC ATC CGT GCT AGC TAC GGC CAT GCA GNN NNN NCT TAT CAA TTT GTT GCA CCG ACA CTT CTG CCC AGG CGA GG - 3'
H'_nAt (control)	5' - X CCT GAT CCT ACC ATC CGT GCT GAC ATC GGC CAT GCA GGC ATT GCT TAT CAA TTT GTT GCA CCG ACA CTT CTG CCC AGG CGA GG - 3'
H'_nAt_2 (control)	5' - X CCT GAT CCT ACC ATC CGT GCT GAC ATC GGC CTA CGT GGC ATT GCT TAT CAA TTT GTT GCA CCG ACA CTT CTG CCC AGG CGA GG - 3'
H'_scrambled	5' - X CCT GAT CCT ACC ATC CGT GCT ATT CGC AAT AAT CGC CTC CTA CGG CTG TGT TGT CCG AAT GTG ACA CTT CTG CCC AGG CGA GG - 3'
fPCR primer	5' - CCT GAT CCT ACC ATC CGT GCT - 3'
rPCR primer	5' - **Y** CCT CGC CTG GGC AGA AGT GTC - 3'

6.3.2. SELEX studies with IHF

The first SELEX study we performed was with IHF, using labeled H'_nAt as our control (Table 6.1). This sequence is identical to the DNA library sequences except for the NNNNNN randomized region. The results after three rounds of SELEX are summarized in Figure 6.4. We present results here for lane 1 ([IHF] at 250 nM). Most sequences were clustered in a range of −1 to +0.5 for the $\log_{10}$(sequence enrichment) score (see Figure 6.4.A), indicating ~30-fold variation in the binding affinities for IHF. A surprising outlier was the biggest winner in this study, with a $\log_{10}$(sequence enrichment) score of 2.75, which turned out to be the sequence of H'_nAt, also our control sequence, and which apparently had a binding affinity 50-fold greater than the next best winner. This result was somewhat unexpected, since the 6-nt sequence in the randomized region of H'_nAT is not a consensus region of the IHF binding sites. Our results showed strong correlations for Lanes 1-3 (Figure 6.4.B)

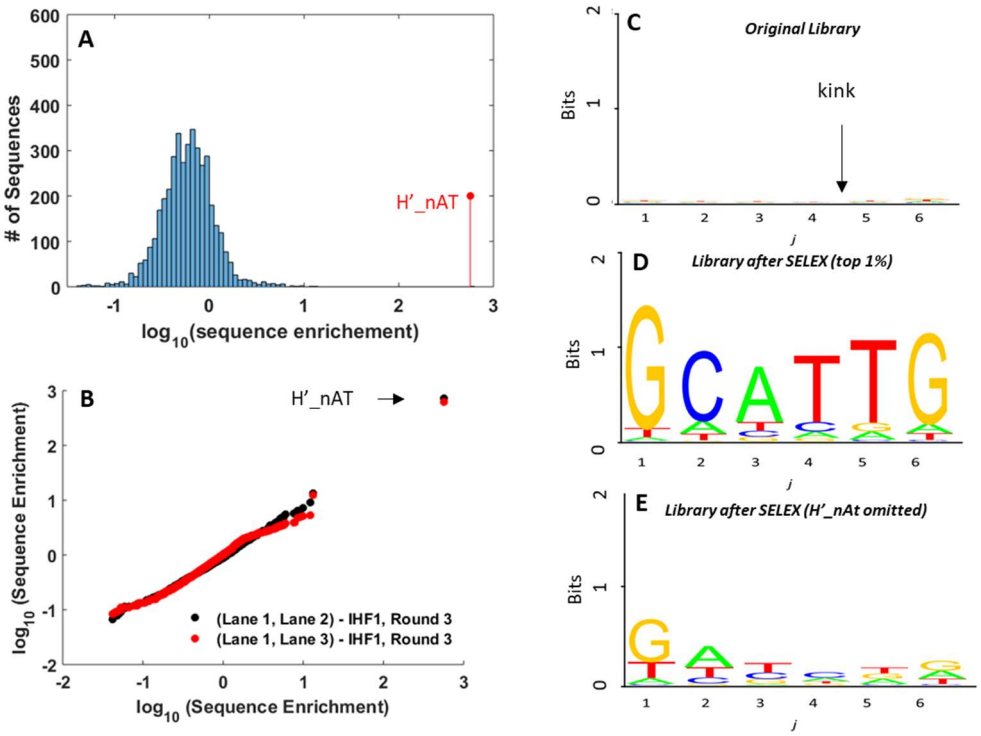


Figure 6.4. Results of SELEX studies from first IHF experiment (denoted as IHF1). (A) Histogram of IHF1 SELEX library sequences after 3-rounds (data from lane 1 with 250 μM [IHF]). The SELEX score for H'_nAt is indicated by the red vertical line. (B) Reproducibility of SELEX scores between different lanes of IHF1 study. SELEX scores for lane 1 (shown in A) are plotted against corresponding scores from lane 2 (black; with 500 nM [IHF]) and lane 3 (red; with 1 uM [IHF]). (C) LOGO of original library. (D) LOGO of IHF1 SELEX library after 3-rounds, using data from lane 1 (with 250 μM [IHF]). (E) LOGO of (D) with H'_nAt removed.

These results are also reflected in the LOGO description of the selected sequences (Figures 6.4.C-E). LOGOs describe the frequency with which a nucleotide appears at a particular DNA site, j. Nucleotide frequency is described by "bits" which scores nucleotides on a scale from 0-2. A score of 2 indicates near certainty that a nucleotide will occur at position j, and a score of 0 indicates no nucleotide favorability. LOGOs were computed using built-in function in MATLAB. LOGO descriptions of DNA library before SELEX (Figure 6.4.C) and after SELEX (Figure 6.4.D-E) indicate a biasing of the library after three rounds, as anticipated. However, the most prevalent sequence at the randomized site after SELEX was GCATTG (Figure 6.4.D), which corresponds to

the sequence of H'_nAt at that site. Removing the H'_nAt sequence from the library yielded the LOGO shown in Figure 6.4.E.

We suspected some contamination of the library by our control DNA sequence as contributing to this surprising result that H'_nAt emerged as an overwhelming winner. The control DNA sequence is key to how we separate bound from unbound DNA in our EMSA gel shifts. The DNA library is unlabeled and hence invisible in our gels. However, a control sequence, which in our case was Cy3-labeled, is in lanes 1 and 2 of our gels, allowing us to image DNA run lengths when bound or unbound. In the IHF1 experiment, we used Cy3-labeled H'_nAt as our control, which had the same sequence as H'_nAt in our library as that of the randomized DNA library. If there was any leakage from the control lanes into the lanes of interest, then the control sequence would be indistinguishable from the H'_nAt sequences from the DNA library, which would yield an apparently large and erroneous sequence enrichment signal.

To eliminate such a "leakage" error, we carried out a second SELEX study (IHF2), in which we changed our control sequence from H'_nAt to H'_nAt_2 (Table 6.1). H'_nAt_2 was designed to have a different barcode than the DNA library, thus enabling us to separate, during the sequencing step, any control sequences that may leak into the lanes of interest and thereby contaminate the DNA library (modifications indicated in bold/italics in Table 6.1). In IHF2, the H'_nAt sequence from the library (identified by the appropriate barcode) was no longer an absolute winner, although it remained in the top 5 sequences. Also, as we suspected, H'_nAt_2 (with the different barcode) was present in 3-fold excess compared with H'_nAt in the samples sent for sequencing. This result underscores a source of error in our protocol in that samples appear to not be restricted to their respective lanes, resulting in some unwanted "smearing". This smearing may also be contributing to some mixing of "winners" and "losers" as a result of leakage from one lane

to another that may explain, in part, the less than ideal correlation between the two sets of independent experiments on IHF, as discussed below.

In Figure 6.5, we compare the histograms of the sequence enrichment scores from the two sets of experiments: IHF1 and IHF2, and a reproducibility scatter plot. Identical studies should produce a scatter plot that falls on a line with a slope of one. Our data are best described by a slope of ~0.3, reflecting the broader distribution of the histogram seen in IHF1 compared to IHF2, indicating a better separation of the winners and losers in IHF1 than in IHF2. We note here that the H'_nAt2 sequence was removed prior to this comparison because of the erroneous score for that sequence in IHF1. Further, the Pearson's correlation score for this comparison yielded r ~ 0.455, indicating considerable irreproducibility in the ranking of the different sequences between the two sets of experiments. This correlation score is low relative to other studies that used similar SELEX technique with the same number of rounds, e.g., r ~ 0.53 – 0.98 reported in ref. (26). These other SELEX studies with similar protocols (26, 42-44) likely benefited from a larger sequence space (>10^7 sequences in library) encompassing the entire binding site of the protein of interest, which allowed for a greater spread in binding affinity between the highest affinity sequence to lowest affinity sequence. Our study focuses on one sequence element that contributes to total biding affinity. Therefore, our study does not have the sensitivity that would exist with large affinity spread. Estimates of affinity range based on our SELEX scores (ratio of highest and lowest sequence enrichment scores) suggest maximum ~300-fold difference in binding affinity between the worst and best DNA sequence. This relatively narrow spread in binding affinities suggests that

our study could benefit from more than three rounds of experiments for better separation of winners and losers.

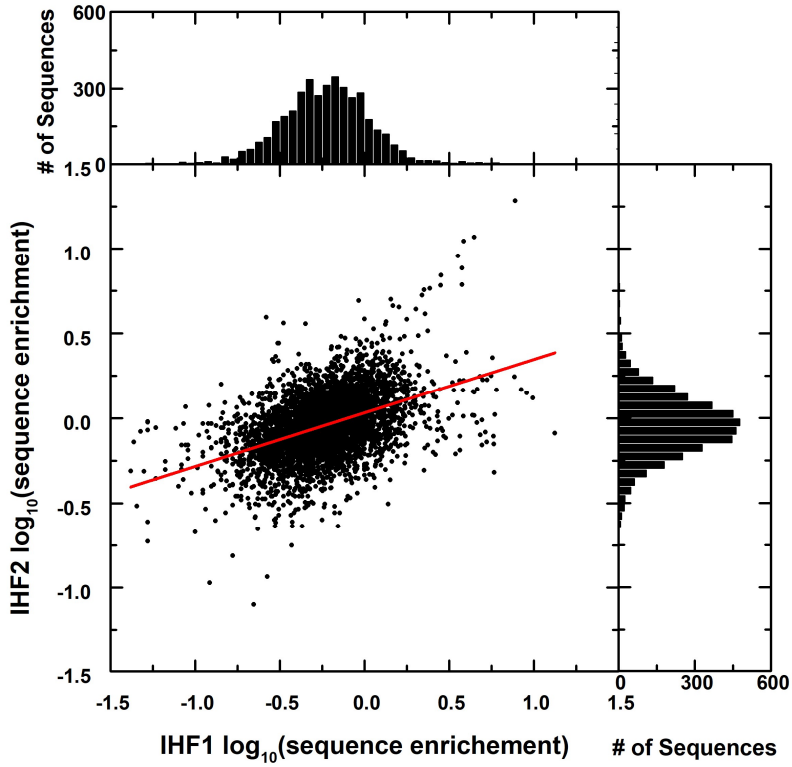


Figure 6.5. Reproducibility between two independent IHF SELEX studies (IHF1 and IHF2). The scatter plot compares the logarithm of sequence enrichment scores from IHF1 study to the corresponding scores from IHF2 study. Data are compared from lane 1 (with 250 μM [IHF]) after three rounds of SELEX in each study. The distribution of $\log_{10}$(sequence enrichment) scores for each study are plotted in the histograms above (for IHF1) and to the right (for IHF2). The red line describes the best linear fit with a slope of 0.315 and a Pearson's r score of 0.449.

6.3.3. What are the DNA features that separate IHF winners and losers?

Indirect readout proteins rely less on the individual position of nucleotides, information that can be wrapped up in a LOGO representation, and more on sequence shape and deformability. We therefore further examined the IHF selected sequences to investigate if any sequence or structural patterns emerged and to identify common features that may render DNA to be more easily deformed. Here we present three approaches by which we characterized the SELEX results.

First, analysis of DNA sequences (omitting H'_nAt) nearest-neighbor base steps and tetranucleotide sequence probabilities within the randomized site revealed that AA dimers and A-tracts were predominant in the "loser" sequences, indicating that these sequence patterns do not readily form the kinked conformations needed to form the specific complex with IHF. Second, DNA shapes computed by TFBSshape revealed that the width of the minor grove may be important to kink site recognition. Third, computation of bending energies from nearest-neighbor elastic models suggest that winners require less energy to adopt the severely kinked conformation. Taken together, these analyses suggest that "kinkability" within the IHF complex is the result of a multitude of physical interactions at play.

6.3.3.1. Dimer and tetramer sequence preferences reveal that IHF rejects an A-tract near the kink site

Sequence flexibility is currently modeled on a dimer (54, 64) and tetramer (12) sequence scales. Therefore, we determined the probability that dimer (dinucleotide steps) and tetramer (tetranucleotide sequences) would appear following SELEX. The probabilities were computed for a subset of the sequences that were deemed winners if their mean $\log_{10}$(sequence enrichment) score from two independent IHF studies, IHF1 and IHF2, was within the top 10% (410/4096) of library sequences (upper right corner in Figure 6.6.A). Similarly, sequences were considered as losers if their mean $\log_{10}$(sequence enrichment) score was in the bottom 10% of library sequences (lower left corner in Figure 6.7.A).

Sequence weight for a dimer or tetramer dimer or tetramer sequence was determined as the number of sequences with that dimer or tetramer sequence times the mean $\log_{10}$(sequence enrichment) score for that sequence normalized by the sum of the mean scores. To normalize, sequence weights were divided by the sum of all resultant scores for the sequences considered.

The results from this analysis are summarized in the 3D bar ("skyline") plots for the winners in Figure 6.6 and for the losers in Figure 6.7. Surprisingly, there is little preference for any dinucleotide step at the kink site (see j = [4-5] in Figure 6.6). Earlier studies that surveyed various IHF binding sites identified TT at this site to be highly preferred to such an extent the authors included as part of the consensus (62). Though TT dimers were some of the most preferred at the kink site, we do not see enough evidence that we would conclude the same. At j = [1-2], GA steps are very preferred; and at j= [5-6], GA and AG steps are the most abundant followed by AA. For the tetranucleotide sequences, 'CTTA' emerged as highly preferred around the kink site (see j =[3-6] in Figure 6.6), present in roughly 1/31 of the winners. Other tetranucleotide sequences that appeared with high abundance at the kink site were 'GGAG' (1/34), 'GGGA'(1/40), and 'CTGA'(1/40). Away from the kink site, GATC stood out as preferred at the j = [1-4] location.

Analysis of the dimers and tetramers for the losers tell a more compelling story. A-tracts are rejected within the randomized region, especially for four subsequent As, "A_4-tracts". Most strikingly, IHF clearly rejects AA dimers if positioned at sites j = [1-2] and [2-3], with the probability of rejection nearly 5-10% greater than any other dimer at those positions., we do not see similar behavior from its complement dimer, TT. At site j= [5-6], CC dimers are the most heavily disfavored. Interestingly Tetramer sequence that are most disfavored are the 4-nt A-tract (AAAA) at j= [1-4] sites (0.0318 for AAAA relative to the next worst tetranucleotide 0.0202 for AAAT). At the j = [3-6] site, several tetramer sequences were roughly equally disfavored, many including AA dimers (0.0293 for AATC compared to 0.0264 each for AACC and 0.0244 for AGTC).

Our results highlight how dimer behavior is affected by its flanking nucleotides. When we expand our analysis to tetramers, we uncover that CTTA as the most preferred kink-site sequence even if TT dimers would not have stood out otherwise. Similarly, our results pull out A-tract is well-known to possess unique mechanical properties on length scales beyond dimer steps (40), our results underscore the importance of considering sequence mechanics beyond nearest-neighbor parametrization.

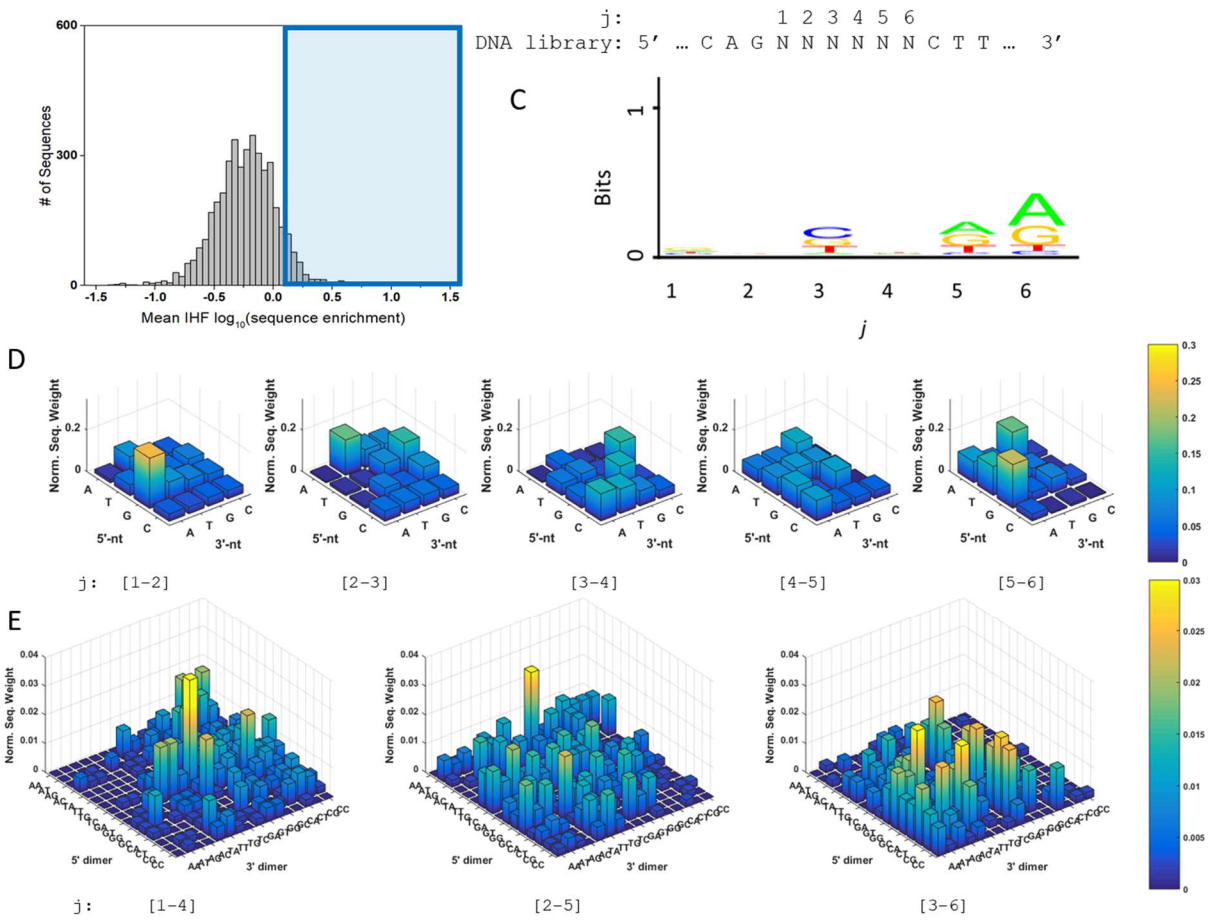


Figure 6.6. Sequence preferences of IHF winners. (A) Subset of winners shown in scatter plot (in blue box). (B) Description of randomized site j. (C) LOGO representation of the subset of winners. (D,E) 3-D skyline plots" represent the probability of (D) dinucleotide steps and (E) tetranucleotide sequence that begins at site j. 5' nucleotides (D) or nucleotide steps (E) are shown on the y-axis, 3' nucleotides (D) or nucleotide steps (E) are shown on the x-axis. Dimer (or tetramer) sequences are created by concatenating the nucleotide (or dimer) sequences from 5' to 3' (i.e. 5' dimer AA with 3' dimer CG would read as 5'-AACG-3').

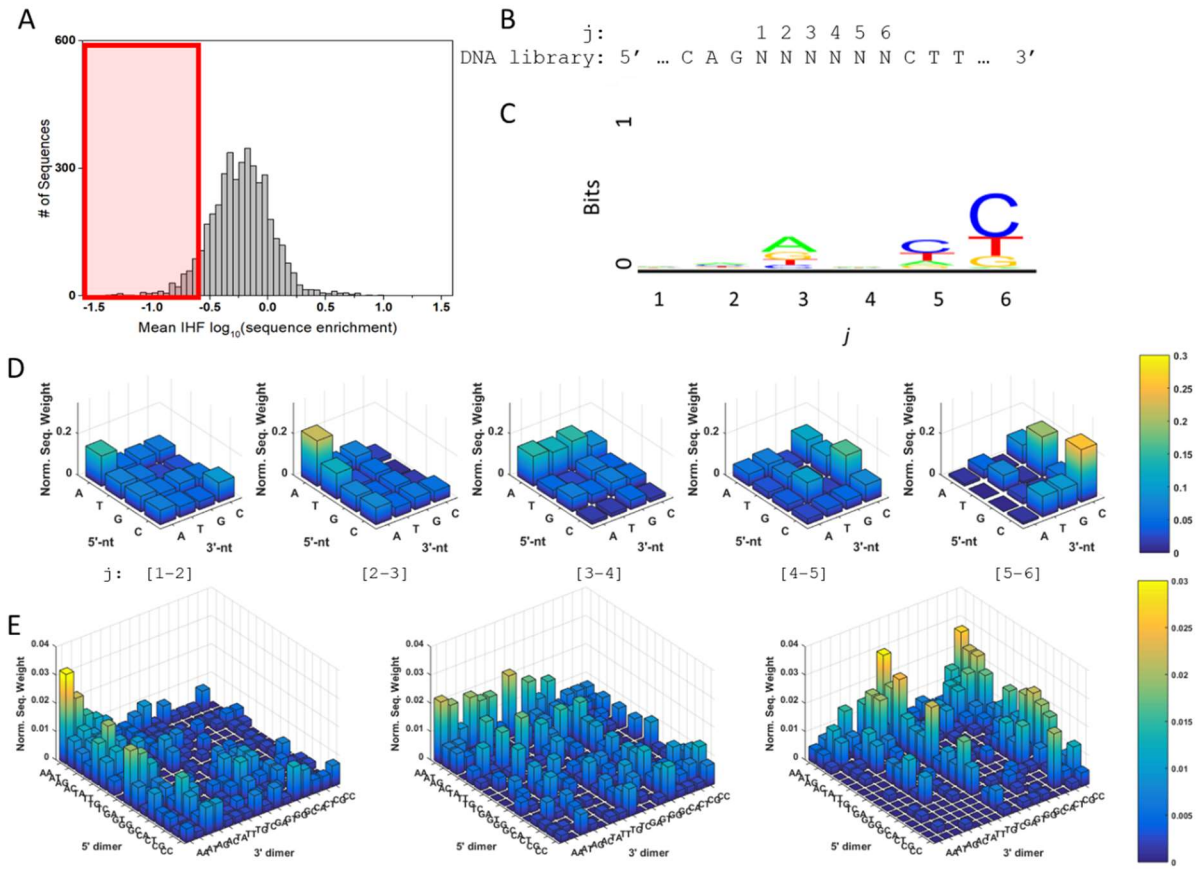


Figure 6.7. Sequence preferences of IHF Losers. (A) Subset of losers shown in scatter plot (in blue box). (B) Description of randomized site j. (C) LOGO representation of the subset of losers. (D,E) 3-D skyline plots" represent the probability of (D) dinucleotide steps and (E) tetranucleotide sequence at each site j. 5' nucleotides (D) or nucleotide steps (E) are shown on the y-axis, 3' nucleotides (D) or nucleotide steps (E) are shown on the x-axis. Dimer (or tetramer) sequences are created by concatenating the nucleotide (or dimer) sequences from 5' to 3' (i.e. 5' dimer AA with 3' dimer CG would read as 5'-AACG-3').

6.3.3.2. "Shape Readout" appears to be key to sequence recognition at the consensus kink site but not at the randomized kink site

Next, we evaluated all DNA sequences based on shapes adopted in the absence of a bound protein to see if we could discern shape patterns that correlated with binding specificity. All sequenced samples were evaluated independently using TBFSshape. More than a dozen shape parameters were evaluated, but only a few parameters showed a significant variation. In Figure 6.8 we show traces of sequence dependent parameters for our library following three rounds of SELEX with 1 μM [IHF]. For visualization of the different shape parameters for different sequences in the randomized region (indicated by the range [37-43] in the DNA library sequence; Figure 6.8.A), enrichment scores were normalized on a scale from zero (lowest score, represented in white) to one (highest score, represented in dark blue) with lighter shades of blue spanning the range of enrichment scores in between. The sequence that lies outside the randomized region is shown in dark blue. Note that the two kink sites are sandwiched between positions 41-42 and 50-51.

The shape analysis of IHF-preferred DNA sequences at the randomized kink site revealed some unexpected results, especially when compared with the corresponding shape parameters for the other (non-randomized) kink site that falls within the consensus region. For example, some of the shape parameters observed for the consensus kink site were markedly different from canonical B-DNA parameters, such as a sharp dip in minor groove width (MGW < 3 Å; Figure 6.8.B), high negative electrostatic potential (EP $\approx -14 k_B T/e$; Figure 6.8.C), a relatively large negative roll ($\sim -8°$; Figure 6.8.D), and high positive twist (~ 36°; Figure 6.8.E). Interestingly,

the roll and twist parameters for the consensus kink site are in good agreement with previous observations made about these DNA parameters from crystal structures of complexes with the

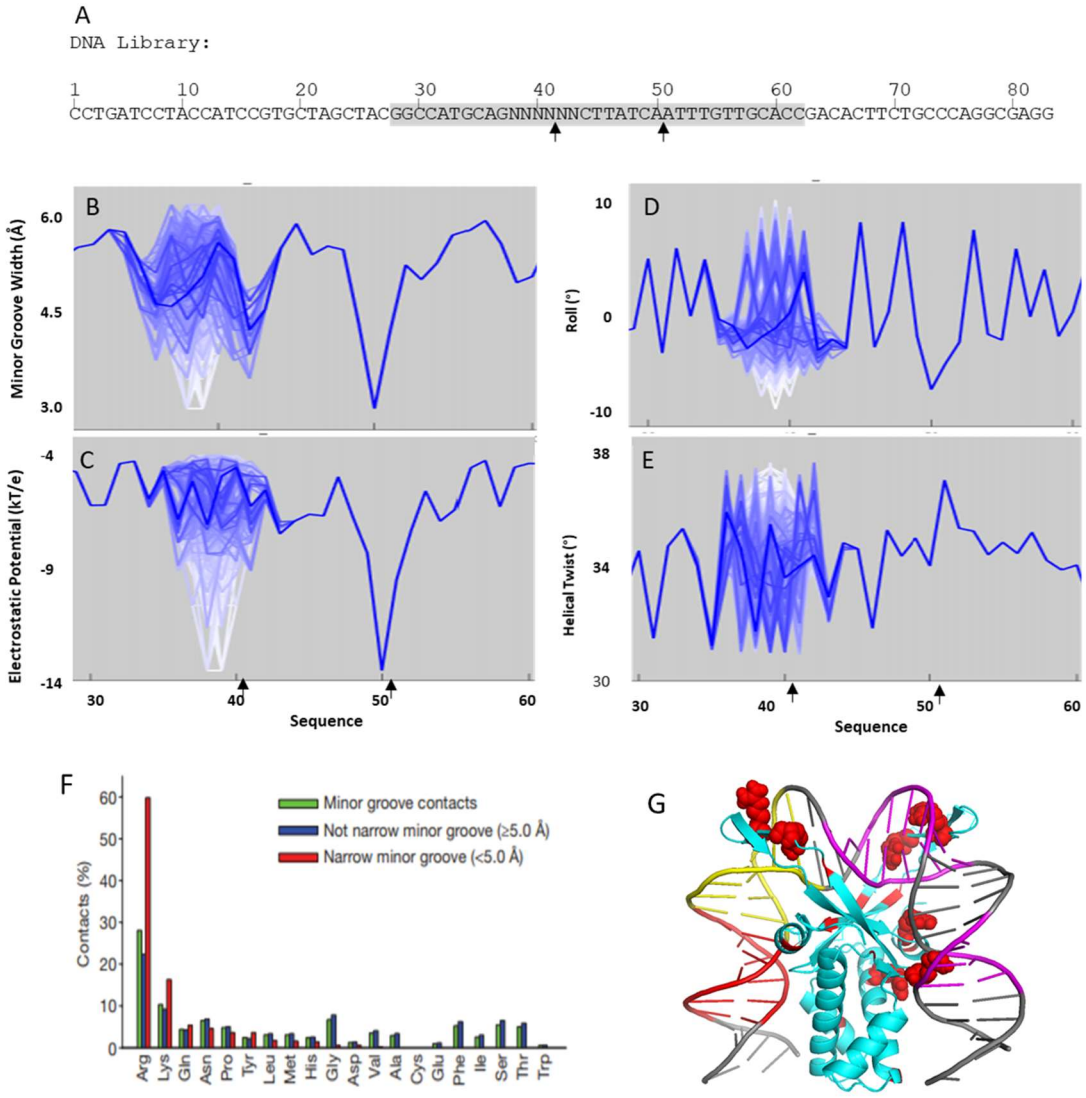


Figure 6.8. TBFSshape parameters for IHF1 Sequences. (A) Description of sequence position number. (B-D) TBFS shape profile of IHF1 SELEX results using data from lane 3 (with 1 µM [IHF]) describing (B) minor groove width, (C) electrostatic potential, (D) roll, and (E) helical twist for each sequence site. Sequence enrichment (in linear scale) is given on a normalized color scale from 0 (white) to 1 (dark blue) where the least enriched sequence is set to 0 and the most enriched is set to 1. Kink sites are indicated by black arrows. (F) Histograms for each amino acid illustrate the frequency with which they are observed in any minor groove (green), in minor grooves with a width of >5.0 A˚ (blue), and in narrow minor grooves of width <5.0 A˚ (red); plot reproduced from (10). (G) IHF-H' crystal structure. All arginine residues in IHF are highlighted in red. Arginine's that interact with the DNA are emphasized by showing them in space filling form. Notable sequence elements are highlighted (consensus – pink; randomized site – yellow; A-tract – red).

IHF/HU family that show high roll and twist at the kink sites (65). Surprisingly, however, none of these shape parameters are mirrored in the shapes at the randomized kink site of the high affinity binders. Instead, we observed a closer resemblance in the shape parameters of the two kink sites for the *low affinity* binders (compare profiles represented by the white lines at the randomized kink site with the profiles at the other kink site). We have, at present, no good explanation for this puzzling result.

The shape parameters that perhaps stand out for the randomized kink site are (i) a preference for a MGW with a two-dip profile (MGW < 5Å near sites 38 and 42) and (ii) a slightly lower than average negative electrostatic potential (EP in the range of -5 to $-8\ k_BT/e$ within sites 37-43). Surprisingly, no dependencies in the randomized region were observed for the roll and twist parameters. Our observations suggest that, while IHF appears to recognize pre-distorted shapes of DNA at the consensus site, for the kink site outside of the consensus region, ease of DNA deformability and not a preformed shape appears to be the important feature for recognition and binding, as discussed in the next section. We therefore conclude that the sequence and shape preferences for the two kink sites have distinct underlying reasons and could be part of the explanation for why only one of the kink site is part of the consensus region, allowing for some variability on the other side. Finally, the heterodimeric nature of IHF may be also be playing a role here.

Width of the minor groove and DNA electrostatics have been suggested as important recognizable features of DNA for proteins that bind to the minor groove (10), like IHF. These DNA features are essential to the so-called "shape readout" mechanism (38). A survey of DNA-binding proteins shows that the primary amino acid that interacts with narrow minor grooves (<5 Å) is Arginine(10, 38). Arginine is unique in its specificity to narrow minor groove and is

significantly more selective for MGW than other positively charged residues like Lysine and Histidine (10), as illustrated in Figure 6.8.F. Arginine is highly abundant on the DNA-interacting surface of IHF (see Figure 6.8.G), and is particularly abundant at sites on IHF where it interacts with regions of the DNA that confer specificity, such as in the β-arms near the intercalating proline and at sites that interact with the consensus TTG region. While the positive charge of Arginine was already known to be favorable to interacting with the negatively charged DNA, Arginine could be playing an additional role in recognition by interacting favorability with DNA sequences that exhibit the appropriate minor groove profile.

6.3.3.3. Ease of DNA deformability at the randomized kink site correlates with SELEX enrichment scores.

Finally, we computed the sequence-dependent energy cost for deforming the DNA into the U-shape resembling the conformation in the IHF-H' crystal structure. We hypothesized that SELEX winners would be more deformable and, therefore, would require less energy to bend. Rigid-body DNA models where base pairs are treated as rigid 3D elements (Figure 6.9.A) were constructed for all sequences in our library. For each DNA sequence, the energies were computed from a set of sequence-dependent base-step parameters (Figure 6.9.B), referred to here as "Olson-98" parameters (2), using a model that uses nearest-neighbor energy functions with harmonic constraints. The knowledge-based parameters in this model were determined by Dr. Wilma Olson and co-workers based on a survey of the protein-DNA crystal structures from the Protein Data Bank that takes into consideration the frequency and extent of distortions away from B-DNA conformations observed in these structures for each base-pair step.

We used several different approaches to compute the energy required to bend the DNA into a U-shape. First, we used the so-called "threading" approach, to determine the extent to

which each base-step is distorted away from B-DNA, were determined by forcing a given DNA sequence to adopt the base-step configurations found in the IHF-H' crystal structure, using a program called 3DNA (49). For the IHF-H' structural studies, a nick in the DNA (adjacent to the non-consensus kink site) was required for crystallization (PDB: 1IHF) (46). For purposes of energy computation, we artificially "sealed" the nick, which modified slightly the DNA structure in the vicinity of that original nick. To verify that this reconstructed structure was reasonable, we compared the "nick-fixed" IHF-H' structure to a newer IHF-H' complex structure that was resolved for intact DNA using cryo-EM in the context of a lambda excision heterojunction complex (PDB: 5J0N) (66). The DNA conformations from the two structures overlapped quite nicely (Figure 6.9 C-D), indicating no significant distortions from our *in-situ* fix. We proceeded to use the "nick- fixed" structure because the original 1IHF crystal structure was characterized with 2.5 Å resolution, significantly higher than the 11.0 Å resolution for the 5J0N cryo-EM structure.

The results of this analysis for both sets of experiments, IHF1 and IHF2, are shown in Figure 6.10, with panels (A-B) showing the correlation between the SELEX score for each experiment and the corresponding computed bending energy for the entire 35-bp IHF binding site. In panel (C), we show the "threaded" nucleotides in the 1IHF crystal structure to give a sense of initial base orientation. These results indicate that there is indeed a correlation between ease of DNA deformability at the kink site (lower bending energy cost) and preference of IHF for that sequence (higher SELEX score), although the cost of bending energy, in the range of $900 - 1500\ k_B T$, is unusually high, as discussed further below.

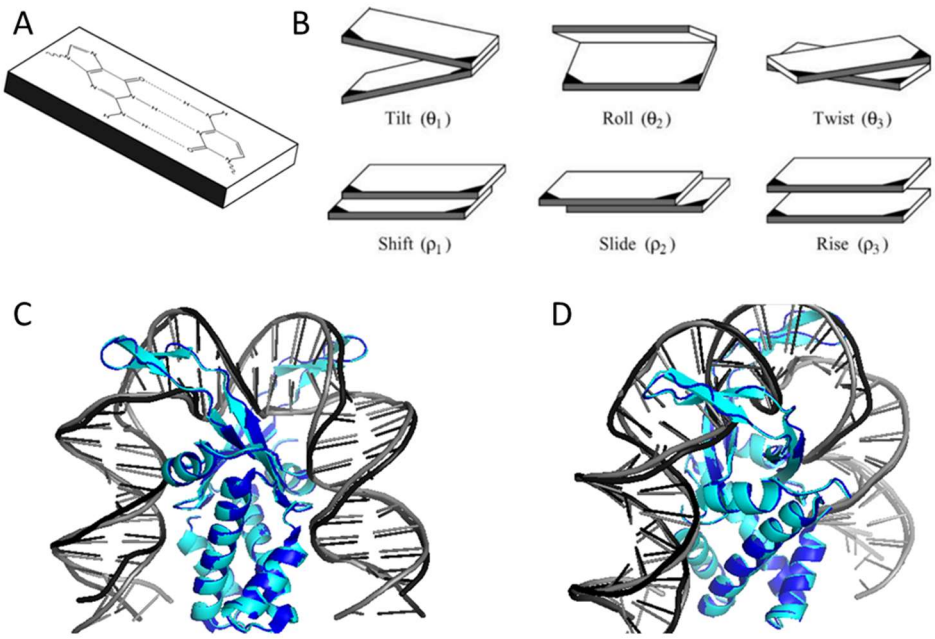


Figure 6.9. DNA model parameters and 1IHF/5J0N structure comparison. (A) Cartoon representation of block description of base-pairs. (B) Descriptions of elasticity model parameters. (C-D) Comparison of IHF-H' structures after *in-situ* sealing of the nick in H' in the x-ray crystal structure (1IHF with sealed nick; protein:cyan, DNA:gray) and from cryoEM on intact H' (5J0N; protein:blue, DNA:black). Two different orientations are shown to facilitate comparison of the two structures.

The second set of analysis we did was to allow the DNA conformation for each library sequence to relax so as to minimize the bending energy, starting from the DNA conformation imposed by the "nick-fixed" 1IHF crystal structure, but with different spatial constraints imposed, as discussed below. The constraints imposed were where the first and last base pair were spatially constrained to be in the same position and orientation as in the 1IHF crystal structure, such that the DNA remains in a U-bent shape (Figure 6.10.D-E); The energy minimization was done as described in (55), with the energy contribution from those base pairs that were constrained omitted from the total energy calculation. Optimized H'_nAt is shown in Figure 6.10.F, in which it displays a smooth bend throughout the optimized structure. The

minimized energy for IHF1 exhibited correlation of bending energy with the SELEX score, but IHF2 showed no correlation.

To summarize the results here, the bending energies when computed for DNA conformations that matched the "U-bent" structure derived from the crystal structure appear to be correlated with IHF-binding preferences (Figure 6.10.A-D). However, when DNA conformations are allowed to relax, without or with constraining the randomized kink site, its possible this correlation be lost, as indicated by IHF2 (Figure 6.10.D-E). Perhaps the latter result is not surprising that, when we either remove the ramdomized kink site from the energy calculation or allow it to relax (and thus not be so severely kinked), the correlation between bending energy and SELEX score is lost. All the selectivity in our SELEX experiments is based on the kink site and high affinity binders are the ones for which that site is more readily kinked. Further studies are needed to elucidate this trend.

We now return to the magnitude of the bending energies computed from the harmonic approximation model. As noted earlier, the $\sim 800\ k_B T$ energy required to kink even the most optimal DNA sequence at two sites to adopt the structure of the IHF-H' complex is more than 20-fold higher than the energy required for that DNA to adopt a bent shape in which the severe distortions at the kink site are allowed to spread over the entire DNA segment ($\sim 30 - 40\ k_B T$ for all sequences examined). The energy cost of a smooth bent can also be readily estimated for a wormlike chain description of an elastic rod of length $L \approx 35$ bp and persistence length $L_0 \approx 150$ bp as

$$E_{bend} = \frac{1}{2} k_B T \left(\frac{L_0}{L}\right) \theta^2 \qquad \text{Eq. 6.5.}$$

For $\theta \approx 180°$, Eq. 6.5 yields $E_{bend} \approx 21\ k_BT$, not too far from the $\sim 30 - 40\ k_BT$ estimated with the nearest-neighbor model with the end base pairs constrained. The 2-fold higher bending energy when the ends are constrained is likely from a torsional strain introduced in the DNA, which is not included in Eq. 6.5.

These results showcase a notable limitation of the elastic model representation in that they significantly overestimate the energy required to sharply kink DNA. Hence, any energy minimization tends to reduce the likelihood of a kink, despite considerable evidence from experiments, supported by theory, that kinks are far more probable than predicted by harmonic approximations (19, 67, 68).

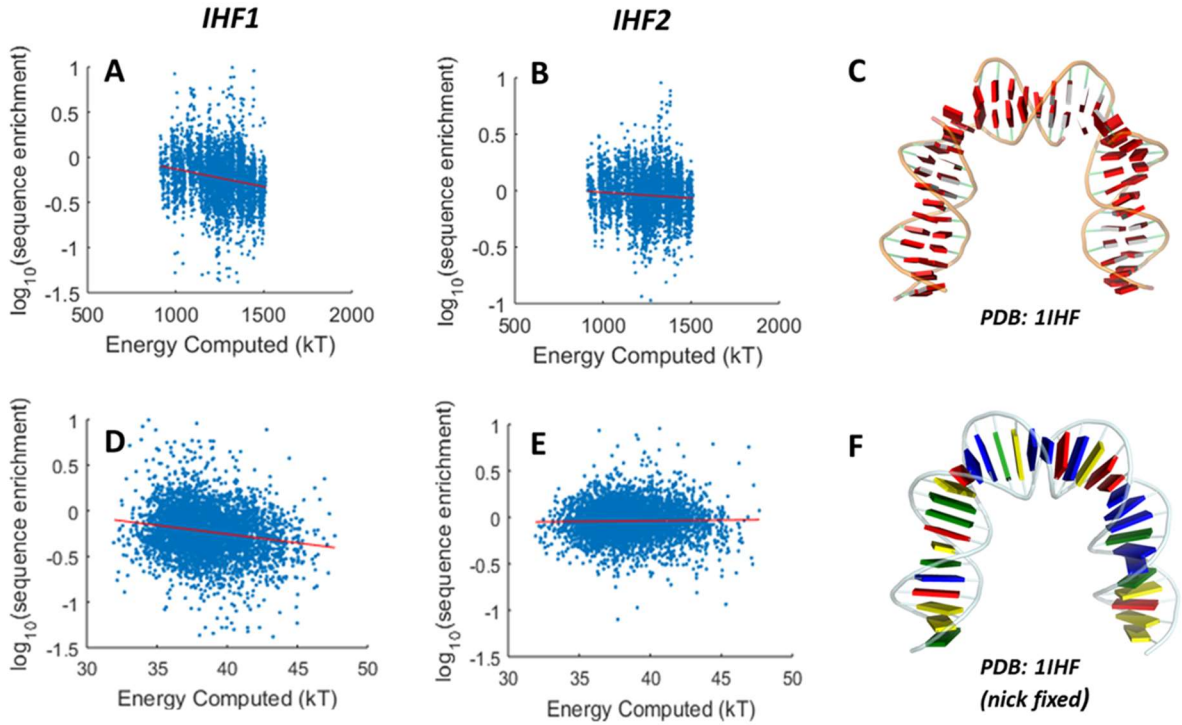


Figure 6.10. Elastic energy calculations compared to SELEX performance. (A-B) Comparison of 'threaded" energies from the IHF-DNA (PDB: 1IHF) structure compared to SELEX performance in IHF1 (B) and IHF2 (B). (C) DNA structure in block form bases shown in red. (D-E) Comparison of 'optimized" energies from the IHF-DNA (PDB: 1IHF with nick fixed) structure compared to SELEX performance in IHF1 (D) and IHF2 (E). (F) DNA structure of H'_nAt in block form bases after optimization.

6.3.4. Shared characteristics of Hbb winners and losers.

Next, we performed SELEX studies using Hbb. We performed only one independent experiment here (named Hbb1) but with four different [Hbb] concentrations: 250 nM, 500 nM, 1 μM, and 2 μM. Surprisingly, the IHF- and Hbb- preferred sequences appear uncorrelated (Pearson's correlation score between IHF studies and Hbb1 were <0.003, Figure 6.11), despite the very similar structures seen for specifically bound DNA complexed with IHF or Hbb. Given that our two IHF experiments were themselves weakly correlated, it is likely that sequence correlation between IHF and Hbb is smeared to the point that we are unable to detect it, at least at the current level of sensitivity of our SELEX studies. However, many of the features that we noted in the IHF winners and losers also show up in corresponding analyses with Hbb sequences, as discussed below.

6.3.4.1. A-tract unfavorable at kink site may be common to IHF/HU family

We evaluated dimer and tetramer sequence probabilities for the Hbb winners and losers, defined here as the top 10% and bottom 10% of sequences, respectively (each 410/4096 sequences). Evaluation of Hbb1 winners and losers are summarized in Figures 6.12. and 6.13. respectively. Similar to the results seen for IHF, a broad distribution of sequences are preferred but AA dimers and A-tracts are highly rejected at the kink site. Thus, the Hbb results add weight to the argument that A-tracts are generally unfavored at sites where DNA needs to be severely kinked.

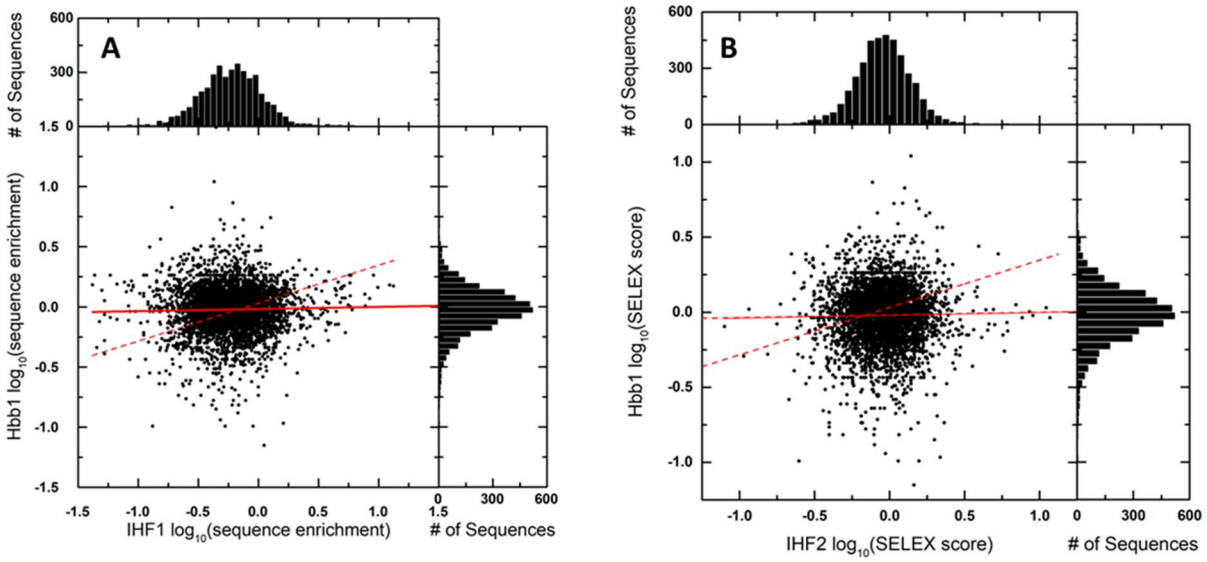


Figure 6.11. Comparison of IHF and Hbb preferred binding sequence. The scatter plots compare each sequences performance in Hbb1 study to IHF1(A) and IHF2 (B) study, as described by logarithm of sequences enrichment scores after three rounds of SELEX at 250 μM [IHF]. The distribution of $\log_{10}$(sequence enrichment) scores for each study are plotted in the histograms above and to the right. The red line describes the best linear fits. Pearson's r is 0.025 in (A) and 0.018 in (B). The red dashed line is the correlation observed when IHF1 was compared with IHF2.

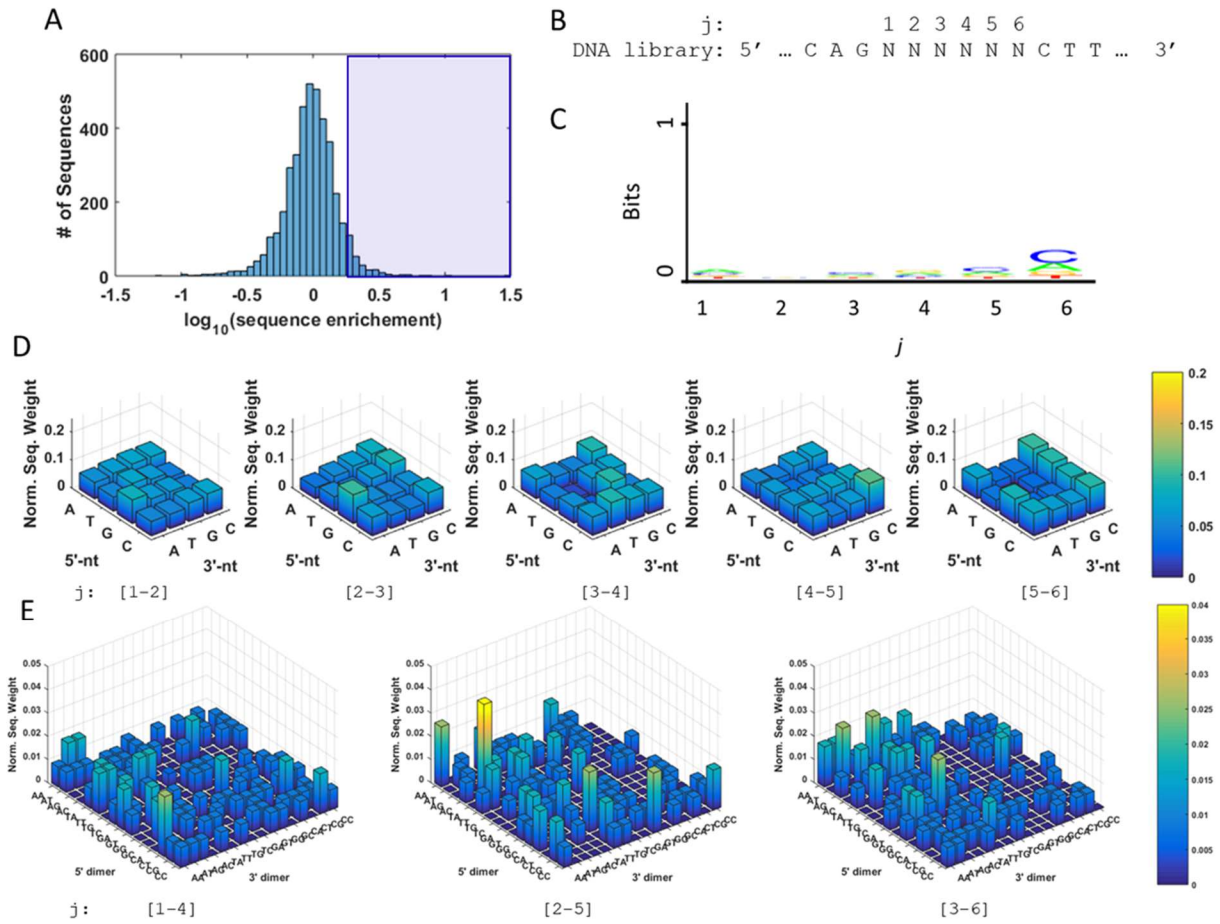


Figure 6.12. Sequence patterns of Hbb1 winners. (A) Subset of winners (top 10%) shown in scatter plot (in blue box). (B) Context and numbering of randomized site j. (C) LOGO representation of the subset of winners. (D,E) 3-D skyline plots" represent the probability of (D) dinucleotide steps and (E) tetranucleotide sequence at each site j. 5' nucleotides (D) or nucleotide steps (E) are shown on the y-axis, 3' nucleotides (D) or nucleotide steps (E) are shown on the x-axis. Dimer (or tetramer) sequences are created by concatenating the nucleotide (or dimer) sequences from 5' to 3' (i.e. 5' dimer AA with 3' dimer CG would read as 5'-AACG-3').

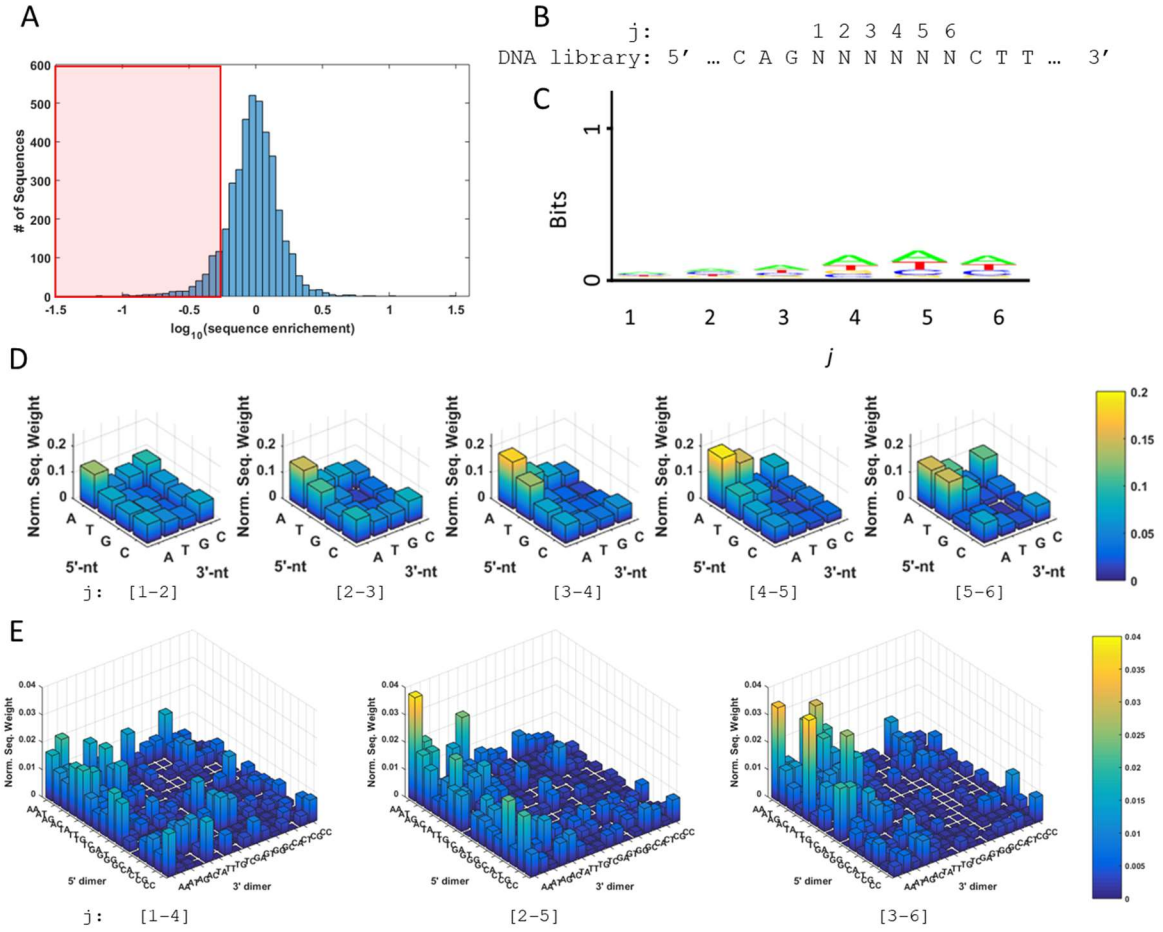


Figure 6.13. Sequences patterns of Hbb1 Losers. (A) Subset of losers (bottom 10%) shown in scatter plot (in blue box). (B) Context and numbering of randomized site j. (C) LOGO representation of the subset of losers. (D,E) 3-D skyline plots" represent the probability of (D) dinucleotide steps and (E) tetranucleotide sequence at each site j. 5' nucleotides (D) or nucleotide steps (E) are shown on the y-axis, 3' nucleotides (D) or nucleotide steps (E) are shown on the x-axis. Dimer (or tetramer) sequences are created by concatenating the nucleotide (or dimer) sequences from 5' to 3' (i.e. 5' dimer AA with 3' dimer CG would read as 5'-AACG-3').

6.3.4.2. Hbb winners show unique "dipped" profile within the minor groove width

Evaluation of sequence shapes by TBFSshape as a function of Hbb SELEX performance revealed similar preferences for minor groove widths and electrostatic potentials as we saw for IHF. Notably, we again observe the curious result that the shape preferences at the two kink

sites are not symmetric and that non-consensus kink-site sequences that are losers appear to resemble shapes seen at the consensus kink site. The origin of these counter-intuitive results merit further investigation.

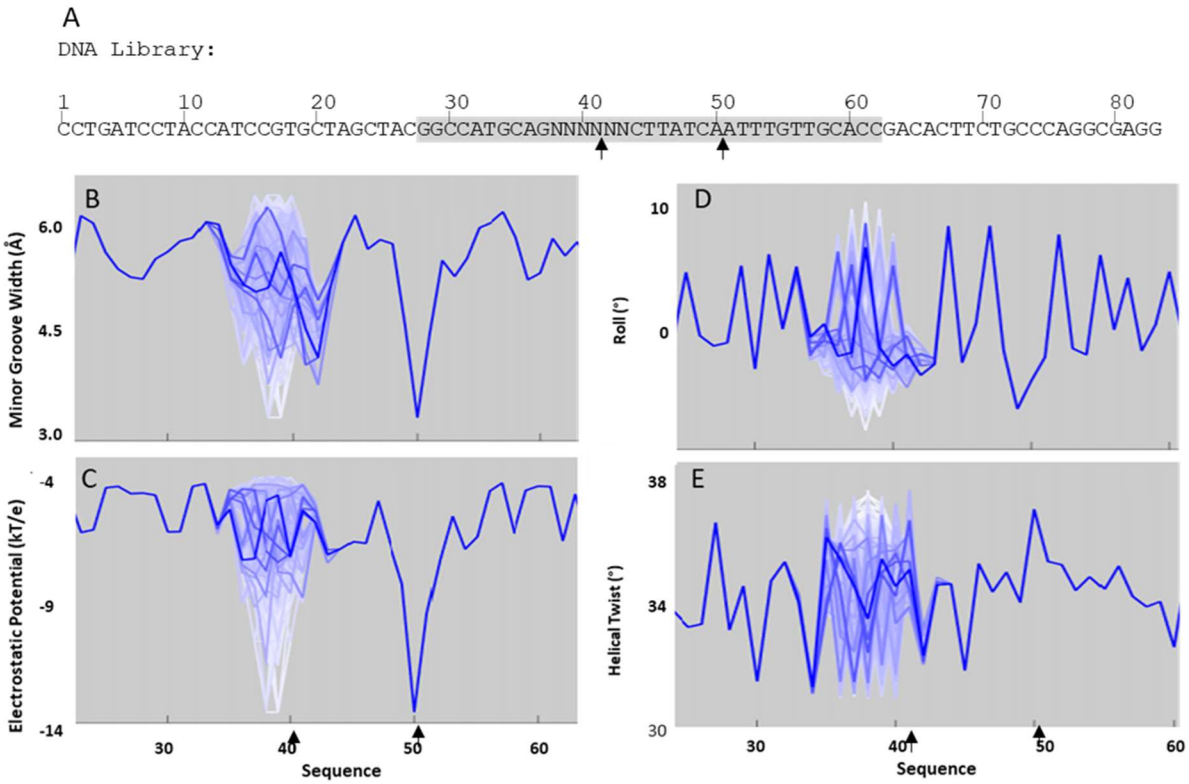


Figure 6.14. TBFSshape parameters for Hbb sequences. (A) Description of sequence position number. (B-D) TBFS shape profile of Hbb SELEX results using 1 μM [IHF] describing (B) minor groove width, (C) electrostatic potential, (D) roll, and (E) helical twist for each sequence site. Sequence enrichment (in linear scale) is given on a normalized color scale from 0 (white) to 1 (blue) where the least enriched sequence is set to 0 and the most enriched is set to 1. Kink sites are indicated by black arrows.

6.3.4.3. Elasticity calculations show no sequence dependence for Hbb

We compared sequence SELEX performance to energy estimates that are sequence and conformationally dependent. Again, rigid-body DNA models were constructed for all library sequences and energies were computed using nearest-neighbor energy functions using Olson-98

elasticity parameters (2). Imposed 3DNA (49) values were based on Hbb-DNA crystal structure (PDB: 2NP2 (46)). DNA in the crystal structure is mostly intact but included a T:T mismatch located at the center of binding site, between the two kinks. Due to its location, the mismatch does not significantly change the orientation of DNA bases, so energy calculations are unaffected by this mismatch. For comparison to IHF structures, 2NP2 was resolved with ~3 Å resolution.

We compare Hbb winners versus elastic energy calculations. Figures are summarized in Figure 6.15. Shockingly, no correlation is observed between elasticity and sequence performance for any of the cases in which electricity was calculated : 1) where DNA structure matched the Hbb "U-bent" structure exactly prior to any energy minimization, 2) where DNA, starting from the "U-bent" conformation, could find the minimum energy while still bending 180° (i.e. first and last base-pair spatially constrained), 3) where DNA found its optimal position while the kink within the consensus sequence, as well as the first and last base pair were fixed. Its hard to fathom that DNA bendability plays no role in Hbb binding preference. These results indicate a need for a new Hbb library, further rounds of SELEX to achieve greater separation between winners and losers, or the need to reconsider elasticity estimations that go beyond harmonic approximations.

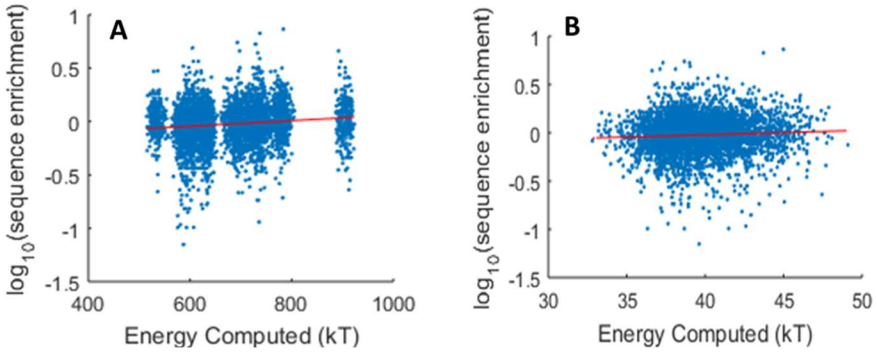


Figure 6.15. Elasticity calculations using the Hbb-DNA structure (PDB: 2np2) versus Hbb SELEX library sequences. (A) "Threaded" energies and (B) "Optimized" are compared to SELEX performance in Hbb1 described by its logarithm of sequence enrichment.

6.4. Discussion and conclusion

We performed SELEX studies to capture binding preferences of two members of the IHF/HU family of DNA-bending proteins, IHF and Hbb, and explored different sequence characteristics such as dimer and tetramer probabilities, DNA shapes, and bending energies that likely contribute to favorable binding. The location of the 6-nt randomized site within a 35-bp binding site was chosen to overlap with one of the kink sites seen in both IHF- and Hbb-DNA specific complexes, and therefore we anticipated that the binding affinities of the selected sequences would correlate with DNA 'kinkability' at that site. Due to the small size of our randomized site (6-nt in length), we were able to evaluate all of sequence space in this study.

Since specific substrates of IHF and Hbb share their consensus sequences and both proteins bend their specific DNA sites into similar U-bent conformations, we expected to see similar rankings for the sequences selected by the two proteins. Our results, however, showed no correlation between sequence preferences for IHF and Hbb, which was surprising. This lack of correlation may, in part, be due to the inherent noise in these SELEX studies, as highlighted by the imperfect correlation between two identical IHF experiments. It may also be a consequence of the fact that our DNA library for both proteins was based on an IHF-specific substrate, which may be introducing unexpected compensating distortions at this kink site when bound to Hbb. Similar SELEX studies using an Hbb-specific DNA library is needed to further investigate these findings.

Despite differences in the sequences selected by IHF and Hbb, common features emerged when we evaluated the propensities for dimer and tetramer sequences at or near the kink sites.

The most notable result here is that both IHF and Hbb selectively reject A-tracts (poly(dA).poly(dT) stretches of three or more nucleotides) near the kink site, consistent with previous observations that nucleosomes that also selectively reject A-tracts. We note here that while IHF and Hbb appear to reject A-tracts at the kink site, this observation is distinct from their strong preference for an A-tract in the flanking arm of the DNA, in the non-consensus side of the asymmetric DNA binding site, as seen in many naturally occurring DNA binding sites of IHF (69), and discussed extensively in Chapter 3. This preference for the A-tract in the flanking arm, as suggested from the crystal structure of IHF-H', is due to the narrow minor groove within A-tract structures that IHF prefers on that side and which allows for easy clamping of the bent arm by IHF (46). The rejection of A-tracts at the kink sites is likely due to the resistance of the DNA to be sharply kinked at that site, as discussed below.

Segal and Widom have reviewed how the A-tract is strongly rejected by the nucleosome, thus playing a major role in chromatin organization and genome expression (39). Genome-wide analyses established that A-tracts are, on average, relatively depleted of nucleosomes *in vivo* (70, 71). Segal and Widom argue that this depletion is because A-tracts directly reduce affinity of histone binding sites (39). *In vitro* binding measurements show a decreased affinity by nucleosomes to A-tract incorporated binding sites (72), and structural studies on a nucleosome containing a poly(dA:dT) element showed a locally distorted DNA structure (73).

What distinguishes A-tracts? Numerous studies reveal that A-tracts adopt distinct conformations that form cooperatively and cannot be purely described as a collection of individual AA steps as assumed by a nearest-neighbor model, reviewed in (40). This non-canonical conformation features a high propeller twist and the narrowing of the minor groove in the 5' to 3' direction (74, 75). A-tract structure tends to resist sharp bending (76). Addition of A-

tracts was shown in a recent study to result in an effective increase in the DNA persistence length (77). A-tracts have also been shown to induce macroscopic curvature when phased in tandem with the helical repeat (78), and this effect is enhanced by placement of CA/TG steps at their 5' junction (79), a feature that is important for DNA looping in transcriptional regulation and chromatin packaging (40). Recent measurements of curvature of phased A-tracts by gel electrophoresis and DNA "nunchucks" report bends of ~17° for an A_6-tract (77, 80). NMR studies on A-tracts of varying length (A_n-tracts, n = 2, 4 and 6) revealed that while the interior of A-tracts is highly rigid, the backbone mobility is increased by the flanking regions that is both dependent on flanking sequence and on A-tract length (81). These observations of enhanced backbone flexibility could explain why the tetrameric sequence CTTA emerged as a favorable sequence at the kink site in our SELEX studies.

A primary motivation for our study was to provide a body of data to aid in the development of improved theoretical models for energy estimation of DNA deformability. Though not completely understood, our evaluation of DNA shape and bendability showed evidence for 'shape readout' (10, 38) at the randomized kink site in both IHF or Hbb. The other kink site within the consensus-region, showed strong preformed shapes that deviated from canonical B-DNA shapes. An evaluation of shape for other known IHF consensus sites may offer new insights about DNA properties that would be rendered highly deformable. We also uncovered some correlation between SELEX enrichment scores and ease of DNA deformability at the kink site, at least for IHF via the "threading" model. These results would indicate that sequences that exhibited high binding affinity were those for which the cost of forming sharp kinks was relatively smaller.

A striking observation from the computed estimates of bending energies was the very high cost of sharply kinking DNA at two sites, to match the distortions seen in the IHF- or Hbb-DNA crystal structures, in comparison with uniformly spreading the distortions over the entire 35-bp binding site (Figures 6.10 and 6.15). In the nearest-neighbor harmonic approximation models widely used to compute DNA deformability, sharp kinks are highly unfavorable compared to small bends and current energy minimization models tend to reduce the occurrence of such sharp bends, despite several protein-DNA crystal structures that show sharply kinked DNA at few sites may be more prevalent than uniformly spread small bends. Experimental studies designed to determine the energy cost for completely breaking a stack show that these energy costs are relatively small, $< 3.0\ k_BT$ for even the most stable base-step stacks and as low as $< 0.5\ k_BT$ for the weakest TA stack (64, 82). These observations accent the need to modify energy approximations and to include anharmonic terms that enable sharp kinks to be modeled accurately.

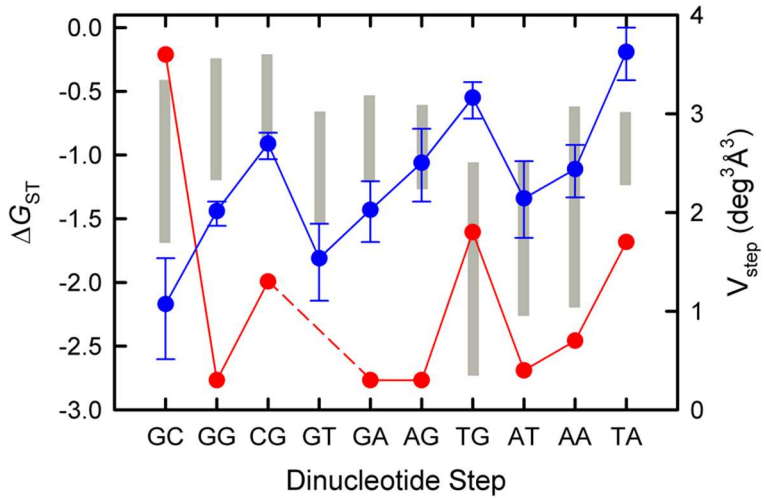


Figure 6.16. Comparison of nearest-neighbor (nn) stacking free energies (ΔG_{ST}) and volumes of conformational space (V_{step}) for the ten different dinucleotide steps of duplex DNA. Figure is reprinted with permission from ref. (). The nn stacking free energy parameters, extracted from thermal denaturation experiments on oligonucleotide duplexes, are shown as vertical gray bars. The length of each vertical bar indicates the range of the stacking parameters from seven independent research groups, obtained under different salt conditions and for varying lengths of

duplex DNA, and unified by SantaLucia in (82). Stacking free energies are from electrophoretic mobility measurements (●) on DNA fragments containing a nick in the sugar-phosphate backbone, between all possible combinations of dinucleotide steps (data from ref. (64)). The volume of thermally accessible conformational space for each dinucleotide step (●), obtained from fluctuations and correlations of base step parameters in DNA-protein crystal complexes, is from ref. (54). The estimates of stacking free energies form the thermal melting and the gel mobility assays, in general, correlate well with the volume of accessible conformational space at each dinucleotide step, with perhaps the exception of the GC step.

Our observations also underscore the need to expand DNA deformability models beyond the nearest-neighbor approximation and the need to for more data which my help serve as a guide for DNA deformation model development. The A-tract may be one of the most extreme examples of how sequence elasticity cannot be captured by dinucleotide steps and how flanking nucleotides play a significant role in the behavior of dimers. MD simulations have been instrumental in this effort (12, 14). Such computational studies combined with a greater number of DNA-protein crystal structures from an updated survey of the protein data bank, to fine-tune the model parameters similar to the approach in ref. (2), would provide new insights into DNA deformability and sequence-context effects on longer sequence scales.

Experimental data such as the kind reported here are critical to help validate theoretical and computational models. Our SELEX studies on IHF and Hbb are a step in this direction and augment insights gained from similar studies with nucleosomes that were the first to take a comprehensive look at sequence-dependent DNA deformability. Signal-to-noise issues in our SELEX studies were a limitation and hampered the reproducibility regarding SELEX scores of the different sequences from two identical IHF experiments, and these issues need to be resolved. In future studies, more rounds of separation are suggested to spread the sequence enrichment scores over a wider range, to better separate the winners from the losers. Extending the

randomized region from the current 6-nt (4,096 sequences) to 8-nt (6.5×10^4 sequence) would still allow for a full evaluation of sequence space and would likely provide a broader range of binding affinities within the DNA library. Furthermore, aligning the randomized region such that it is symmetric about the kink site would allow for sequence context effects to be evaluated on either side of the kink site. Finally, including HU, a third member of the IHF/HU family protein, would be beneficial. Unlike IHF and Hbb, HU has no sequence preference and yet binds DNA in a manner similar to IHF, with sharp kinks introduced 9-bp apart(27, 65, 83). HU has been shown to bind preferentially to DNA that is modified to make the kink sites more flexible, such as mismatches or unpaired nucleotides at that site (65). Furthermore, the lack of a consensus region for HU suggests that both kink sites can be randomized simultaneously, thus providing a more robust data set for testing DNA "kinkability". Similarities and differences in binding preferences for IHF, Hbb, and HU will also help separate the protein-specific interactions from features that are entirely due to intrinsic DNA deformability.

1. Garcia, H. G., P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins. 2007. Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. Biopolymers 85(2):115-130.
2. Olson, W. K., A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. U S A 95(19):11163-11168.
3. Olson, W. K., and V. B. Zhurkin. 2000. Modeling DNA deformations. Curr. Opin. Struct. Biol. 10(3):286-297.
4. Morozov, A. V., J. J. Havranek, D. Baker, and E. D. Siggia. 2005. Protein-DNA binding specificity predictions with structural models. Nucleic Acids Res 33(18):5781-5798.
5. Lowary, P. T., and J. Widom. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276(1):19-42.
6. Morozov, A. V., K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, and E. D. Siggia. 2009. Using DNA mechanics to predict in vitro nucleosome positions and formation energies. Nucleic Acids Res 37(14):4707-4722.
7. Balasubramanian, S., F. Xu, and W. K. Olson. 2009. DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. Biophys J 96(6):2245-2260.
8. Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. K. Moore, J. P. Wang, and J. Widom. 2006. A genomic code for nucleosome positioning. Nature 442(7104):772-778.
9. Lowary, P. T., and J. Widom. 1997. Nucleosome packaging and nucleosome positioning of genomic DNA. Proc Natl Acad Sci U S A 94(4):1183-1188.
10. Rohs, R., S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig. 2009. The role of DNA shape in protein-DNA recognition. Nature 461(7268):1248-1253.
11. Dixit, S. B., D. L. Beveridge, D. A. Case, T. E. Cheatham, 3rd, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer, and P. Varnai. 2005. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. Biophys J 89(6):3721-3740.
12. Pasi, M., J. H. Maddocks, D. Beveridge, T. C. Bishop, D. A. Case, T. Cheatham, 3rd, P. D. Dans, B. Jayaram, F. Lankas, C. Laughton, J. Mitchell, R. Osman, M. Orozco, A. Perez, D. Petkeviciute, N. Spackova, J. Sponer, K. Zakrzewska, and R. Lavery. 2014. muABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. Nucleic Acids Res 42(19):12272-12283.
13. Lavery, R., K. Zakrzewska, D. Beveridge, T. C. Bishop, D. A. Case, T. Cheatham, 3rd, S. Dixit, B. Jayaram, F. Lankas, C. Laughton, J. H. Maddocks, A. Michon, R. Osman, M. Orozco, A. Perez, T. Singh, N. Spackova, and J. Sponer. 2010. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. Nucleic Acids Res 38(1):299-313.
14. Beveridge, D. L., G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, 3rd, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. Biophys J 87(6):3799-3813.
15. Curuksu, J., M. Zacharias, R. Lavery, and K. Zakrzewska. 2009. Local and global effects of strong DNA bending induced during molecular dynamics simulations. Nucleic Acids Res 37(11):3766-3773.

16. Harris, S. A., C. A. Laughton, and T. B. Liverpool. 2008. Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. Nucleic Acids Res 36(1):21-29.
17. Olson, W. K., and V. B. Zhurkin. 2011. Working the kinks out of nucleosomal DNA. Curr Opin Struct Biol 21(3):348-357.
18. Yuan, C., H. Chen, X. W. Lou, and L. A. Archer. 2008. DNA bending stiffness on small length scales. Phys Rev Lett 100(1):018102.
19. Wiggins, P. A., T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson. 2006. High flexibility of DNA on short length scales probed by atomic force microscopy. Nat Nanotechnol 1(2):137-141.
20. Schindler, T., A. Gonzalez, R. Boopathi, M. M. Roda, L. Romero-Santacreu, A. Wildes, L. Porcar, A. Martel, N. Theodorakopoulos, S. Cuesta-Lopez, D. Angelov, T. Unruh, and M. Peyrard. 2018. Kinky DNA in solution: Small-angle-scattering study of a nucleosome positioning sequence. Physical Review E 98(4).
21. 2008. Protein-nucleic acid interactions : structural biology. Royal Society of Chemistry, Cambridge.
22. Slattery, M., T. Riley, P. Liu, N. Abe, P. Gomez-Alcala, I. Dror, T. Zhou, R. Rohs, B. Honig, H. J. Bussemaker, and R. S. Mann. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell 147(6):1270-1282.
23. Zhao, Y., D. Granas, and G. D. Stormo. 2009. Inferring binding energies from selected binding sites. PLoS Comput Biol 5(12):e1000590.
24. Jolma, A., T. Kivioja, J. Toivonen, L. Cheng, G. Wei, M. Enge, M. Taipale, J. M. Vaquerizas, J. Yan, M. J. Sillanpaa, M. Bonke, K. Palin, S. Talukder, T. R. Hughes, N. M. Luscombe, E. Ukkonen, and J. Taipale. 2010. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20(6):861-873.
25. Zykovich, A., I. Korf, and D. J. Segal. 2009. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. Nucleic Acids Res 37(22):e151.
26. Rodriguez-Martinez, J. A., A. W. Reinke, D. Bhimsaria, A. E. Keating, and A. Z. Ansari. 2017. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. Elife 6.
27. Swinger, K. K., and P. A. Rice. 2004. IHF and HU: flexible architects of bent DNA. Curr Opin Struct Biol 14(1):28-35.
28. Lynch, T. W., E. K. Read, A. N. Mattis, J. F. Gardner, and P. A. Rice. 2003. Integration host factor: putting a twist on protein-DNA recognition. J Mol Biol 330(3):493-502.
29. Mouw, K. M., and P. A. Rice. 2007. Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending Hbb. Molecular Microbiology 63:1319-1330.
30. Irvine, D., C. Tuerk, and L. Gold. 1991. SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. J. Mol. Biol. 222(3):739-761.
31. Lowary, P. T., and J. Widom. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. 276(1):19-42.
32. Widlund, H. R., H. Cao, S. Simonsson, E. Magnusson, T. Simonsson, P. E. Nielsen, J. D. Kahn, D. M. Crothers, and M. Kubista. 1997. Identification and characterization of genomic nucleosome-positioning sequences. J. Mol. Biol. 267(4):807-817.
33. Burgdorfer, W., A. G. Barbour, S. F. Hayes, J. L. Benach, E. Grunwaldt, and J. P. Davis. 1982. Lyme disease-a tick-borne spirochetosis? Science 216(4552):1317-1319.
34. Benach, J. L., E. M. Bosler, J. P. Hanrahan, J. L. Coleman, G. S. Habicht, T. F. Bast, D. J. Cameron, J. L. Ziegler, A. G. Barbour, W. Burgdorfer, R. Edelman, and R. A. Kaslow. 1983. Spirochetes isolated from the blood of two patients with Lyme disease. N Engl J Med 308(13):740-742.

35. Tilly, K., J. Fuhrman, J. Campbell, and D. S. Samuels. 1996. Isolation of Borrelia burgdorferi genes encoding homologues of DNA-binding protein HU and ribosomal protein S20. Microbiology (Reading) 142 (Pt 9):2471-2479.
36. Yang, L., T. Zhou, I. Dror, A. Mathelier, W. W. Wasserman, R. Gordan, and R. Rohs. 2014. TFBSshape: a motif database for DNA shape features of transcription factor binding sites. Nucleic Acids Res 42(Database issue):D148-155.
37. Chiu, T. P., B. Xin, N. Markarian, Y. Wang, and R. Rohs. 2020. TFBSshape: an expanded motif database for DNA shape features of transcription factor binding sites. Nucleic Acids Res 48(D1):D246-D255.
38. Rohs, R., X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann. 2010. Origins of specificity in protein-DNA recognition. Annu Rev Biochem 79:233-269.
39. Segal, E., and J. Widom. 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol 19(1):65-71.
40. Haran, T. E., and U. Mohanty. 2009. The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys 42(1):41-81.
41. Stormo, G. D., and Y. Zhao. 2010. Determining the specificity of protein-DNA interactions. Nat Rev Genet 11(11):751-760.
42. Warren, C. L., N. C. Kratochvil, K. E. Hauschild, S. Foister, M. L. Brezinski, P. B. Dervan, G. N. Phillips, Jr., and A. Z. Ansari. 2006. Defining the sequence-recognition profile of DNA-binding molecules. Proc Natl Acad Sci U S A 103(4):867-872.
43. Tietjen, J. R., L. J. Donato, D. Bhimisaria, and A. Z. Ansari. 2011. Sequence-specificity and energy landscapes of DNA-binding molecules. Methods Enzymol 497:3-30.
44. Carlson, C. D., C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria, Y. Lee, F. Cerrina, and A. Z. Ansari. 2010. Specificity landscapes of DNA binding molecules elucidate biological function. Proc Natl Acad Sci U S A 107(10):4544-4549.
45. Connolly, M., A. Arra, V. Zvoda, P. J. Steinbach, P. A. Rice, and A. Ansari. 2018. Static Kinks or Flexible Hinges: Multiple Conformations of Bent DNA Bound to Integration Host Factor Revealed by Fluorescence Lifetime Measurements. J Phys Chem B.
46. Rice, P. A., S. Yang, K. Mizuuchi, and H. A. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87(7):1295-1306.
47. Sugimura, S., and D. M. Crothers. 2006. Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. Proc. Natl. Acad. Sci. USA 103(49):18510-18514.
48. Vivas, P., Y. Velmurugu, S. V. Kuznetsov, P. A. Rice, and A. Ansari. 2012. Mapping the transition state for DNA bending by IHF. J Mol Biol 418(5):300-315.
49. Zheng, G., A. V. Colasanti, X. J. Lu, and W. K. Olson. 2010. 3DNALandscapes: a database for exploring the conformational features of DNA. Nucleic Acids Res 38(Database issue):D267-274.
50. Robasky, K., and M. L. Bulyk. 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res 39(Database issue):D124-128.
51. Zhou, T., L. Yang, Y. Lu, I. Dror, A. C. Dantas Machado, T. Ghane, R. Di Felice, and R. Rohs. 2013. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic Acids Res 41(Web Server issue):W56-62.
52. Dickerson, R. E. 1989. Definitions and Nomenclature of Nucleic Acid Structure Parameters. Journal of Biomolecular Structure and Dynamics 6(4):627-634.
53. Coleman, B. D., W. K. Olson, and D. Swigon. 2003. Theory of sequence-dependent DNA elasticity. The Journal of Chemical Physics 118(15):7127-7140.

54. Olson, W. K., A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A 95(19):11163-11168.
55. Clauvelin, N., and W. K. Olson. 2014. The synergy between protein positioning and DNA elasticity: energy minimization of protein-decorated DNA minicircles. arXiv.
56. Kuznetsov, S. V., S. Sugimura, P. Vivas, D. M. Crothers, and A. Ansari. 2006. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc Natl Acad Sci U S A 103(49):18515-18520.
57. Vivas, P., S. V. Kuznetsov, and A. Ansari. 2008. New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. J Phys Chem B 112(19):5997-6007.
58. Velmurugu, Y., P. Vivas, M. Connolly, S. V. Kuznetsov, P. A. Rice, and A. Ansari. 2018. Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein. Nucleic Acids Res 46(4):1741-1755.
59. Holbrook, J. A., O. V. Tsodikov, R. M. Saecker, and M. T. Record, Jr. 2001. Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. J Mol Biol 310(2):379-401.
60. Yang, S. W., and H. A. Nash. 1995. Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. Embo. J. 14(24):6292-6300.
61. Aeling, K. A., N. R. Steffen, M. Johnson, G. W. Hatfield, R. H. Lathrop, and D. F. Senear. 2007. DNA deformation energy as an indirect recognition mechanism in protein-DNA interactions. IEEE/ACM Trans Comput Biol Bioinform 4(1):117-125.
62. Aeling, K. A., M. L. Opel, N. R. Steffen, V. Tretyachenko-Ladokhina, G. W. Hatfield, R. H. Lathrop, and D. F. Senear. 2006. Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy. J. Biol. Chem. 281(51):39236-39248.
63. Wang, S., R. Cosstick, J. F. Gardner, and R. I. Gumport. 1995. The specific binding of Escherichia coli integration host factor involves both major and minor grooves of DNA. Biochemistry 34(40):13082-13090.
64. Yakovchuk, P., E. Protozanova, and M. D. Frank-Kamenetskii. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res 34(2):564-574.
65. Swinger, K. K., K. M. Lemberg, Y. Zhang, and P. A. Rice. 2003. Flexible DNA bending in HU-DNA cocrystal structures. EMBO J 22(14):3749-3760.
66. Laxmikanthan, G., C. Xu, A. F. Brilot, D. Warren, L. Steele, N. Seah, W. Tong, N. Grigorieff, A. Landy, and G. D. Van Duyne. 2016. Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. Elife 5.
67. Vafabakhsh, R., and T. Ha. 2012. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. Science 337(6098):1097-1101.
68. Yan, J., and J. F. Marko. 2004. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. Phys Rev Lett 93(10):108108.
69. Yang, C. C., and H. A. Nash. 1989. The interaction of *E. coli* IHF protein with its specific binding sites. Cell 57(5):869-880.
70. Field, Y., N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol 4(11):e1000216.

71. Yuan, G. C., Y. J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309(5734):626-630.
72. Anderson, J. D., and J. Widom. 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Mol Cell Biol 21(11):3830-3839.
73. Bao, Y., C. L. White, and K. Luger. 2006. Nucleosome core particles containing a poly(dA.dT) sequence element exhibit a locally distorted DNA structure. J Mol Biol 361(4):617-624.
74. MacDonald, D., K. Herbert, X. Zhang, T. Pologruto, and P. Lu. 2001. Solution structure of an A-tract DNA bend. J Mol Biol 306(5):1081-1098.
75. Stefl, R., H. Wu, S. Ravindranathan, V. Sklenar, and J. Feigon. 2004. DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. Proc Natl Acad Sci U S A 101(5):1177-1182.
76. Hogan, M., J. LeGrange, and B. Austin. 1983. Dependence of DNA helix flexibility on base composition. Nature 304(5928):752-754.
77. Cai, X., D. S. Arias, L. R. Velazquez, S. Vexler, A. L. Bevier, and D. K. Fygenson. 2020. DNA Nunchucks: Nanoinstrumentation for Single-Molecule Measurement of Stiffness and Bending. Nano Lett 20(2):1388-1395.
78. Hagerman, P. J. 1985. Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. Biochemistry 24(25):7033-7037.
79. Nagaich, A. K., D. Bhattacharyya, S. K. Brahmachari, and M. Bansal. 1994. CA/TG sequence at the 5' end of oligo(A)-tracts strongly modulates DNA curvature. J Biol Chem 269(10):7824-7833.
80. Stellwagen, E., J. P. Peters, L. J. Maher, 3rd, and N. C. Stellwagen. 2013. DNA A-tracts are not curved in solutions containing high concentrations of monovalent cations. Biochemistry 52(24):4138-4148.
81. Nikolova, E. N., G. D. Bascom, I. Andricioaei, and H. M. Al-Hashimi. 2012. Probing sequence-specific DNA flexibility in a-tracts and pyrimidine-purine steps by nuclear magnetic resonance (13)C relaxation and molecular dynamics simulations. Biochemistry 51(43):8654-8664.
82. Allawi, H. T., and J. SantaLucia, Jr. 1998. Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA. Biochemistry 37(8):2170-2179.
83. Swinger, K. K., and P. A. Rice. 2007. Structure-based analysis of HU-DNA binding. J. Mol. Biol. 365(4):1005-1016.

Independent quantification binding affinity and population distributions for few IHF-SELEX DNA sequences

This chapter is intended to be an Appendix for the results presented in Chapter 6 of this thesis.

A1. Introduction:

In Chapter 6, sequences in a DNA library were separated by binding affinity to DNA-bending protein IHF using a process called SELEX (systemic evolution of ligands by exponential enrichment). The DNA library was designed by randomizing a 6-nt region (indicated in red in the sequences shown in Table A1) at a location where the DNA is expected to be kinked when bound to IHF to form the specific complex. The 6-nt randomized region was chosen such that it falls outside of the consensus regions of the binding sequence (indicated in gray in Table A1). The rest of the sequence in the library was identical to the sequence shown for H'_nAt, which itself is a modification of an IHF consensus sequence H'. The only difference between H' and H'_nAt is that the latter lacks the A-tract of the H' sequence. Previous studies showed that DNA sequence scores (enrichment scores) obtained from SELEX were correlated with DNA-protein binding affinity (1-6). Our SELEX studies were designed to examine if the enrichment scores also correlated with sequence-dependent DNA bendability. In this appendix, we outline the steps we took to (i) independently verify that enrichment in SELEX studies corresponded to higher binding affinity, (ii) investigate if higher affinity binding sequences in the SELEX library were more kinkable, and (iii) examine how we could further refine our SELEX protocol to improve the signal-to-noise in these studies.

A2. Methods

The relevant methods used are briefly summarized here. All details are provided in Chapter 2.

A2.1. Materials

The DNA sequences used in this study are shown in Table A1. All labeled and unlabeled DNA oligomers were ordered from Keck (with gel purification). For FRET (or acceptor ratio) studies, the DNA oligomers were labeled at the 5'-ends with fluorescein (F; top strand) and TAMRA (R; bottom strand).

Duplex DNA was formed by annealing complimentary oligomers of equimolar concentrations. The annealing buffer used was 20 mM Tris-HCl, pH 8.0, 1 mM EDTA, with salt concentrations ranging from 100 to 300 mM KCl. The mixture of oligomers was heated in a water bath at 90 °C for 10 minutes, then allowed to cool slowly at room temperature.

Table A1: Sequences and SELEX scores used in this study. IHF consensus sequence is highlighted in gray. Region of DNA that was randomized for SELEX is in red font. Fluorescent label attachments, Fluorescein (yellow) and TAMRA (pink), are indicated.

Name	*Sequence*	SELEX score (IHF1) [rank/4096]	SELEX Score (IHF2) [rank/4096]
H'	5′ – TGGCCAAAAAACGATTGCTTATCAATTTGTTGGCACC CCGGTTTTTTGCTAACGAATAGTTAAACAACCGTGGT – 5′	NA	NA
H'_nAt	5′ – TGGCCATGCAGCGATTGCTTATCAATTTGTTGGCACC CCGGTACGTCGCTAACGAATAGTTAAACAACCGTGGT – 5′	568.50 [1]*	8.58 [4]
GACAGA	5′ – TGGCCATGCAGGACAGACTTATCAATTTGTTGGCACC CCGGTACGTCCTGTCTGAATAGTTAAACAACCGTGGT – 5′	4.99 [19]	1.70 [255]
CTAACC	5′ – TGGCCATGCAGCTAACCCTTATCAATTTGTTGGCACC CCGGTACGTCGATTGGGAATAGTTAAACAACCGTGGT – 5′	0.077 [4086]	0.867 [2238]

A2.2. Steady-state fluorescence (acceptor ratio and anisotropy) measurements

The steady-state fluorescence emission spectra and anisotropies were measured on a FluoroMax4 spectrofluorometer (Jobin Yvon, Inc., NJ, USA), with samples loaded in a 100-μL quartz cuvette (Starna 26.100F-Q-10/Z20).

A2.3. Competition assays to determine relative binding affinities between two DNA sequences.

To estimate the binding affinity and specificity of IHF (P) for different DNA sequences in solution, competition binding assays were performed. Two DNA sequences – a reference sequence labeled with fluorescein/TAMRA (D) and an unlabeled competitor (I) – compete to form a complex with IHF (P). First, IHF-DNA complexes using labeled DNA (PD) were formed at 1 μM [P] and 1 μM [D] concentrations. Unlabeled competitor was then added over a concentration range of 1-200 μM. Measurements of ensemble acceptor ratio (measured as described in Chapter 2) yield the fraction of protein molecules bound to the labeled reference DNA, as described below.

We define r_{PD} as the acceptor ratio value corresponding to labeled DNA complexed with a protein and r_{free} as the corresponding acceptor ratio value for labeled DNA that is free. In the absence of a competitor, and at protein and DNA concentrations used in our study, we start with all protein molecules bound to our reference labeled DNA in a 1:1 complex. As we add increasing amounts of the competitor unlabeled DNA (I), the protein molecules redistribute between D and I, and the experimentally measured acceptor ratio value (r_{exp}) shifts from r_{PD} to r_{free}, as the fraction of IHF bound to labelled DNA, x_{PD}, decreases and the fraction of IHF bound to unlabelled DNA, x_{PI}, increases. In this case, r_{exp} is given by the following relation:

$$r_{exp} = x_{PD} r_{PD} + x_{PI} r_{free} \quad \text{A.1.}$$

This model assumes an insignificant fraction of free IHF molecules present in solution at protein concentrations used (a reasonable assumption since all reference sequences used had K_d estimates < 10 nM, much smaller than the > 1 μM concentrations of P, D, and I).

We define $r_{50\%}$ as the acceptor ratio value halfway between the fully complexed r_{PD} and the completely free r_{free}

$$r_{50\%} = \frac{(r_{PD} + r_{free})}{2} \quad \text{A.2.}$$

The concentration of the competitor for which we have $r_{exp} = r_{50\%}$ occurs when half the protein molecules are bound to labeled DNA and half to unlabeled DNA, i.e. $x_{PD} = x_{PI}$ or [PD] = [PI]. The binding affinity of IHF for the unlabeled competitor ($K_{d,I}$) can be obtained in relation to the binding affinity of IHF for the labelled reference DNA ($K_{d,D}$) as described below:

$K_{d,D}$ = [P][D]/[PD] and $K_{d,I}$ = [P][I]/[PI] A.3.

$K_{d,I}$ / $K_{d,D}$ = ([I][PD])/([D][PI]) A.4.

At $r_{50\%}$, when [PD]/[PI] = 1, the ratio of binding affinities is equal to the ratio of the concentrations of the competing substrates, as summarized below:

$K_{d,I}$ / $K_{d,D}$ = [I]/[D] A.5.

We note here that Eqs. 2.3 and 2.4 are approximations of more accurate relations needed to describe the competition assays and the complete binding/competing profiles, as discussed in refs. (7-11). However, our approximate approach based on the measurement of the $r_{50\%}$ level is sufficient to obtain the ratio of the Kd,s as described above.

A2.4. Fluorescence lifetime spectroscopy

Fluorescence decay curves were measured with a PicoMaster fluorescence lifetime instrument (HORIBA-PTI, London, Ontario, Canada) equipped with time-correlated single photon counting (TSCPC) electronics (52). For all FRET measurements, decay traces were measured for donor-only duplexes without acceptor, denoted as DNA_D, as well as donor–acceptor-labeled duplexes, denoted as DNA_DA.

A2.5. Maximum entropy analysis of the fluorescence decay traces

Fluorescence decay curves were analyzed using a maximum entropy method (MEM) in which the effective distribution of log-lifetimes f(log τ) was inferred from the decay traces using the program MemExp (available online), and described in detail elsewhere (53-55).

The average FRET efficiencies were computed from the MEM distributions obtained from fluorescence decay curves measured on donor-acceptor labeled (DNA_DA) samples, as described in Chapter 2.

A3. Results:

A3.1 Binding affinities versus SELEX scores:

SELEX studies were performed to separate DNA sequences in a library by their respective binding affinities, as discussed in Chapter 6. The results of the first SELEX study performed with

IHF, which we refer to here as IHF1, are shown in Figure A1.A. These show a spread of sequence enrichment scores, which quantifies each unique sequence's population in a DNA library after 3 rounds of selection by IHF at 250 nM [IHF] concentration compared to its population in the original DNA library. Sequences with stronger binding affinity to IHF are expected to have a greater enrichment score. Previous SELEX studies demonstrated that binding affinity for a given sequence scaled proportionally to the sequence enrichement score (2-4, 6). Here, we selected a few sequences from the results of our IHF1 study, to directly measure their binding affinities and to examine how well they correlated with their SELEX scores.

Binding affinities were measured for three library sequences, which were selected based on their performance in the first IHF SELEX study IHF1, with SELEX scores as summarized in Table A1. In terms of the 6-nt sequence at the randomized site, the sequences selected were the most enriched sequence GCATTG (which is also H'_nAT), GACAGA (one of the top-20 winners), and CTAACC (one of the bottom-20 losers). The enrichment scores of these sequences are indicated in green, blue, and red vertical lines, respectively, in Figure A1. As discussed extensively in Chapter 6, H'_nAt (GCATTG) showed significant enrichment relative to all other sequences in the IHF1 SELEX study, which we determined was a result of "leakage" of the H'_nAt sequence from a reference lane in gel-shift assays that are an integral part of the SELEX protocol. This leakage gave H'_nAt in IHF1 an erroneously high score in SELEX (see Table A1). We overcame this issue by changing our reference sequence in the second study IHF2 (Figure A2), which brought down the SELEX score for H'_nAt to a more reasonable range. Even with the correction, IHF displayed high affinity for H'_nAt.

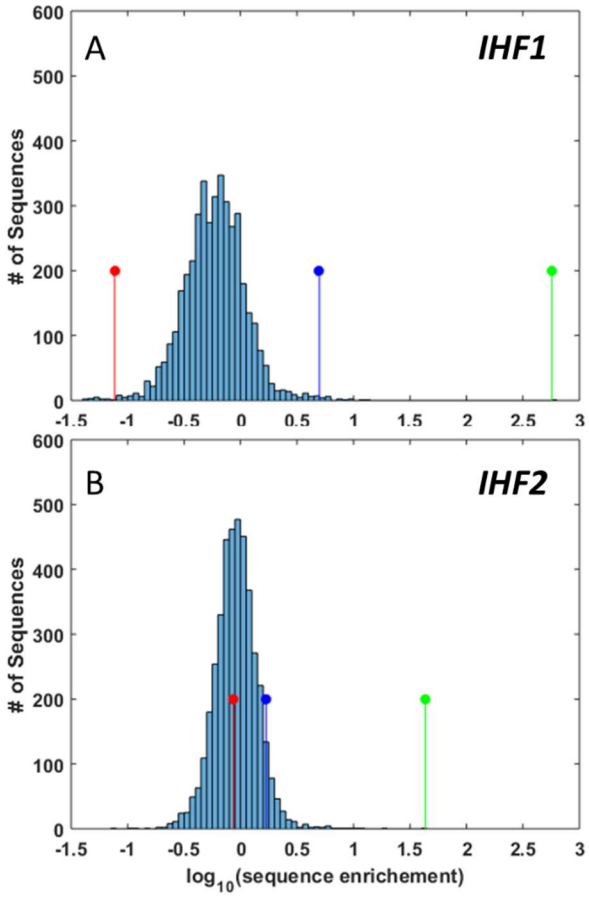


Figure A1: Distribution of SELEX scores following 3 rounds of selection at 250 nM [IHF] in IHF study 1, *IHF1* (A) and study 2, *IHF2* (B). Sequence enrichment is defined as $\frac{sequence\ fractional\ population\ measured\ after\ SELEX}{sequence\ fractional\ population\ measured\ in\ DNA\ library}$. SELEX scores are indicated for measured sequences H'_nAt (green pin), GACAGA (blue pin), and CTAACC (red pin).

Binding affinity measurements for the selected sequences were carried out using competition assays, as described in Section A2.3 and summarized in Figure A2. First, we measured the binding affinity of H'_nAt relative to the reference H' DNA, for which the binding affinity was previously determined to be ~25 pM in 100 mM KCl, as described in Chapter 3. All measurements reported here were carried out under identical buffer conditions as these previous studies. The reference H' DNA was labelled with fluorescein/TAMRA FRET pair and a sample was prepared with 1uM [DNA] mixed with 1uM [IHF]. The sample was then titrated with

unlabeled H'_nAt competitor in the concentration range that was 1- to 200-fold that of the reference (H') sequence. The concentration of the unlabeled competitor at which half the IHF molecules were displaced from the labeled reference was measured from the acceptor ratio measurements on the labeled DNA, which yielded the ratio of the binding affinities between the two DNA sequences (see Eq. 2.5). The results for H'_nAt competing against H' are shown in Figure A2.A. Two sets of titrations were performed; each titration dataset is plotted either blue or red to separate the two experiments and the acceptor ratio value at which 50% of the reference sequence is displaced is plotted as a blue or red horizontal line, respectively. The data show that it takes ~50-fold excess of [H'_nAT] to displace 50% of the IHF protein from the IHF-H' complex, indicating that the K_d of H'_nAt is ~50-fold higher than H'. Thus, we estimate that the binding affinity of IHF for H'_nAt is ~1.3 nM.

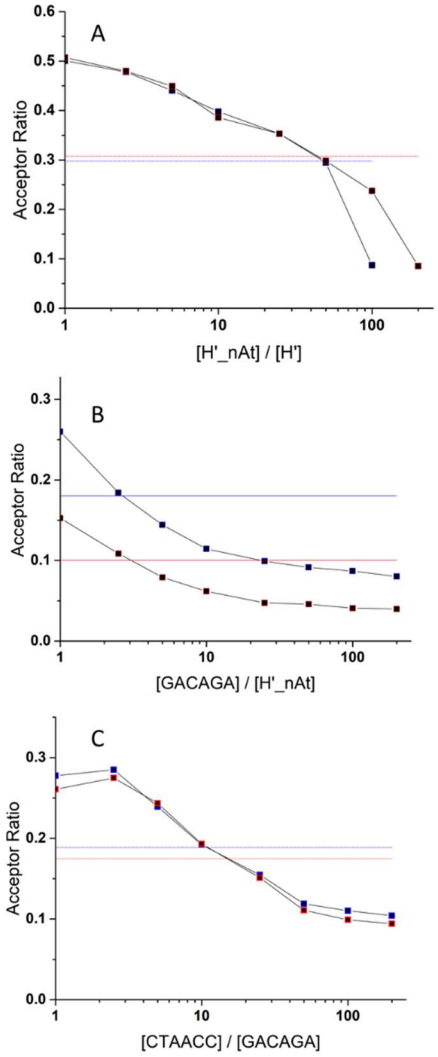


Figure A2: Competition assays to estimate relative binding affinities. (A) IHF-H' complexes are formed between for sequences labeled with donor and acceptor (H'_DA) at a 1uM : 1uM ratio. Unlabeled competitor H'_nAt was added on a range from 0-200 uM. Two independent titrations are shown (blue/black and red/black markers). For each titration, a horizontal line (blue or red) indicates the acceptor ratio at which 50% of the labeled DNA sequence is displaced by competitor. Similar measurements are shown in (B) where IHF-H'_nAt_DA is titrated by unlabeled competitor GACAGA, and (C) where IHF-GACAGA_DA is titrated by unlabeled competitor, CTAACC.

Next, we started with IHF bound to labeled H'_nAt and titrated the sample with unlabeled GACAGA, one of the IHF1 SELEX winners (Figure A2.B). While the two titrations differ in the absolute acceptor ratio values, both titrations indicate that H'_nAt binding affinity is 2-3-fold stronger than GACAGA. Thus we estimate IHF-GACAGA K_d as being ~2-4 nM. Lastly, we picked the GACAGA sequence as the reference (labeled) sample bound to IHF and titrated the sample with unlabeled CTAACC, one of the loser sequences (Figure A2.C). These studies indicate a 10- to 15-fold higher K_d for CTAACC compared with GACAGA and hence in the range from ~20-60 nM.

In Table A2, we compare the ratios of the binding affinities for the different sequences obtained from the competition assays with the ratios of their SELEX scores. The results show that the trends are in the right direction. For example, H'_nAt (GCATTG) has a ~5-fold higher SELEX score than GACAGA (from the "leakage" corrected IHF2 experiment) and a ~2-3-fold higher binding affinity. Similary, GACAGA shows ~2–65-fold higher SELEX score than CTAACC and a ~10- to 15-fold higher binding affinity. However, more quantitative agreement between relative SELEX scores and binding affinities could not be achieved primarily due to the range of error in the SELEX scores, highlighted by the imperfect correlation between IHF1 and IHF2 experiments, as will be discussed further below.

Table A2: Estimated binding affinities difference from SELEX and Competition assays. Relative binding affinities were determined from SELEX by Kd-B/ Kd-A = SELEX score A/ SELEX score B. Binding affinity estimates for competition assays are estimated from the [Competitor]/[Labelled sequence] at which 50% of the labelled sequence is displaced.

Sequence 1	Relative binding affinities ($^{K_d(Sequence\ 1)}/_{K_d(Sequence\ 2)}$) measured by:			Sequence 2
	SELEX (IHF1)	SELEX (IHF2)	Competition Assays	
GCATTG binds	*Not Determined,*	~5-fold,	~3-fold,	stronger than GACAGA
GACAGA binds	~65-fold,	~2-fold,	~10-fold,	stronger than CTAACC
GCATTG binds	*Not Determined,*	~10-fold,	*Not measured,*	stronger than CTAACC

A3.2 Improving signal-to-noise in SELEX studies:

A comparison of results from IHF1 and IHF2 studies illustrate the level of noise present in our measurements. Figure A3 compares the SELEX scores obtained from these two sets of measurements (after 3 rounds at 250 nM). A perfect correlation between two identical experiments would yield a plot in which all points fall on a single line with a slope equal to one. The correlation between the two data sets is characterized by a Pearson's correlation score of $r = 0.449$ and a slope of 0.315, which shows that while the two data sets are indeed correlated, there is significant room for improvement in the signal-to-noise. To examine whether this correlation score could potentially be improved by increasing the number of rounds, we examined what the correlation score was after one or two rounds of selection. Indeed, the Pearson's correlation score increased from 0.078 to 0.101 to 0.449 as we increased the rounds of selection from 1 to 2 to 3, as shown in Figure A4. Thus, increasing the number of rounds even further would be a step in the right direction to achieve higher signal-to-noise and hence an improved estimate of the SELEX scores.

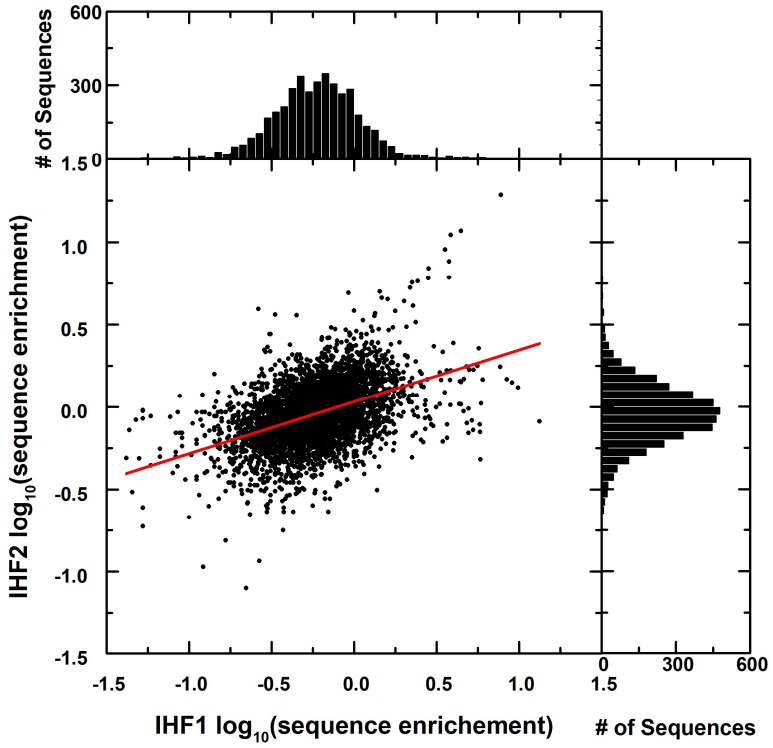


Figure A3: Reproducibility between independent IHF SELEX studies. The scatter plot compares each sequences performance in IHF1 study to its performance IHF2 study, described by logarithm of each sequences enrichment scores after three rounds of SELEX at 250 μM [IHF]. The distribution of $\log_{10}$(Sequence enrichment) scores for each study are plotted in the histograms above and to the right. The red line describes the best linear fit with a slope of 0.315 and a Pearson's r score of 0.449.

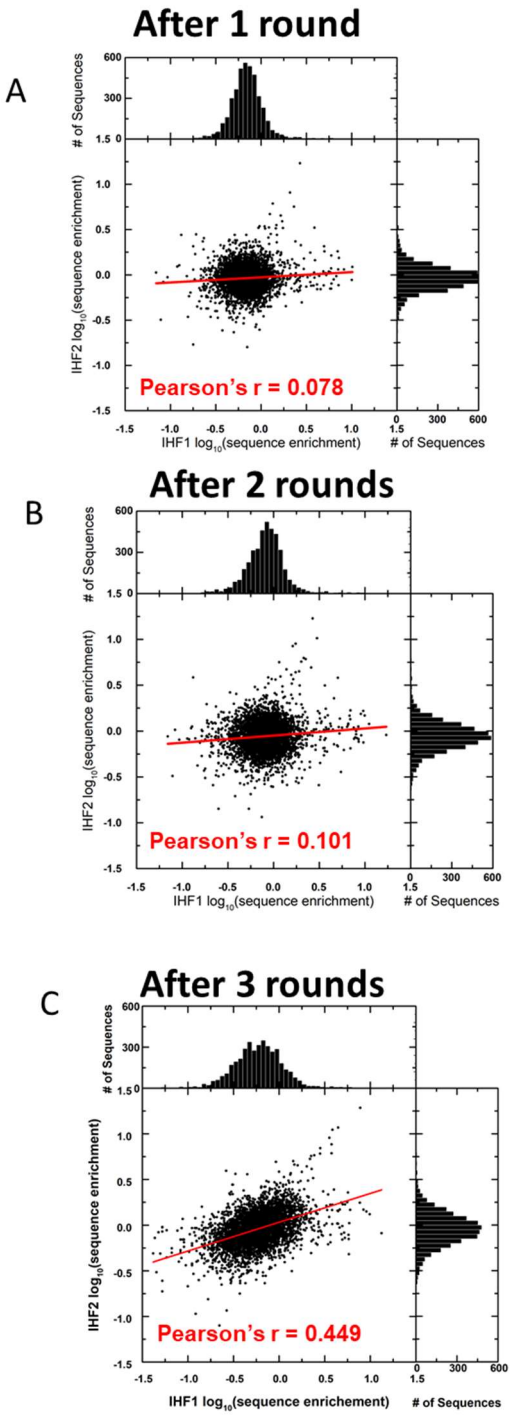


Figure A4: Enriching reproducibility between independent IHF SELEX studies with increased rounds. The scatter plot compares each sequences performance in IHF1 study to its performance IHF2 study, described by logarithm of each sequences enrichment scores after (A) one, (B) two, or (C) three rounds of SELEX at 250 μM [IHF]. The distribution of $\log_{10}$(Sequence enrichment) scores for each study are plotted in the histograms above and to the right. The red line describes the best linear fit and the corresponding Pearson's r score is shown.

A3.3 DNA bendability versus SELEX scores:

Lastly, we measured the extent of DNA bending for each of the three selected sequences in complex with IHF using fluorescence-lifetime-based FRET studies and compared their DNA bendability with their IHF binding affinities. The fluorescence lifetime studies on IHF-H' complexes are discussed extensively in Chapter 3. Briefly, in those studies, IHF-DNA complexes revealed two dominant populations when in the 1:1 specifically bound complex, a fully-bent and partially-bent DNA populations (12). IHF-DNA complexes with sequences that had greater flexibility at the kink site exhibited a larger fraction of population in the fully-bent conformations, indicating that fluorescence lifetime studies could help distinguish less flexible from more flexible sequences at the kink site of DNA. The three sequences studied here, H'_nAt, GACAGA, and CTAACG, showed significant enough changes in binding affinity that they could be used as example sequences to compare extent of bending with corresponding binding affinity.

While these fluorescent lifetime studies did not show a perfect correlation between binding affinity and ease to populate the fully-bent conformation, they did reveal that the winning sequence GACAGA had a higher population in the fully bent conformation compared to the loser sequence CTAACC (~40% versus ~22%; Figure A5). Surprisingly, H'_nAt, with the highest binding affinity among these three sequences, showed intermediate (~33%) population in the fully bent conformation. This could be an indication that H'_nAt has strong favorability for reasons other than strong kinkability. These studies suggest one method to detect DNA bendability, but also underscore the need to perform this study with more than 3 sequences.

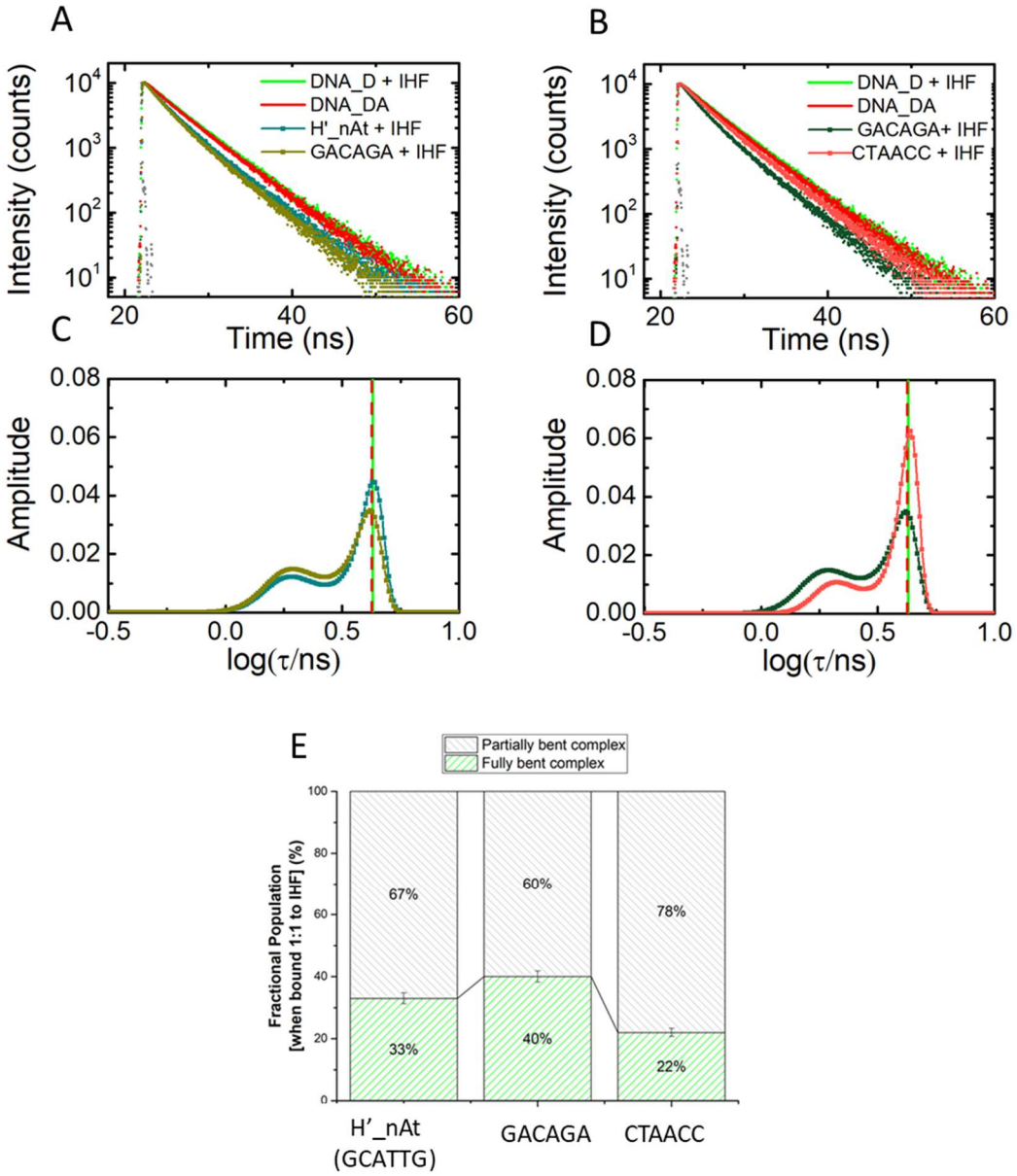


Figure A5: Fluorescence lifetime measurements on IHF-DNA complexes with SELEX library sequences. (A/C) Fluorescence intensity decay traces are shown for H'_nAt_DA (blue-green) and CACAGA_DA (gold), and measured in the presence of IHF. Decay traces on DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are shown for comparison. Measurements were done with 5 μM DNA and 5 μM IHF. The instrument response function (gray) is shown for comparison.(B/D) The MEM lifetime distributions obtained from the fluorescence decay traces measured for H'_nAt_DA (blue-green) and CACAGA_DA (gold) are shown. The amplitudes of the distributions are normalized to add up to one. The average lifetime for the DNA_DA in the absence of IHF (red) and DNA_D in presence of IHF (green) are indicated by the vertical dashed lines. (E) Quantification of partially-bent (grey) and fully-bent (green) obtained from MEM lifetime distributions for H'_nAt_DA, GACAGA_DA, and CTAACC_DA in the presence of IHF. Total populations sum to 100%.

A4. Discussion:

Here, we present results that show that SELEX studies are indeed able to separate DNA sequence by high- or low- affinity to IHF and that ease of DNA bending may be an important factor for high-affinity sequences, as discussed extensively in Chapter 6. From measurements on a few sequences we showed that a SELEX winner has a higher population in the fully bent DNA complex than a SELEX loser, suggesting that the winners have higher affinities because they are more bendable than the losers. Thus, these measurements add support to our primary hypothesis that SELEX studies involving IHF provide a good way to distinguish DNA sequences with varying DNA bendabilities.

However, the extent of noise in these SELEX studies impaired our ability to make rigorous comparisons between SELEX scores, binding affinities, and DNA bendability. On a more promising note, we show that increasing the number of rounds of SELEX could help mitigate some of these experimental uncertainties.

Finally, an independent evaluation of intrinsic DNA bendability of sequences preferred by IHF would be valuable. For example, incorporation of DNA segments into small (<100 bp) minicircles have been shown to induce kinking (13). It would be informative to examine if such kinking is more readily induced in sequences with high SELEX scores. Alternatively, capturing severe bendability due to intrinsic thermal fluctuation in linear DNA may also be possible given recent advancements in DNA imaging technologies with DNA nunchucks (14) and small angle x-ray and neutron scattering (SAXS and SANS) (15).

References:

1. Jolma, A., T. Kivioja, J. Toivonen, L. Cheng, G. Wei, M. Enge, M. Taipale, J. M. Vaquerizas, J. Yan, M. J. Sillanpaa, M. Bonke, K. Palin, S. Talukder, T. R. Hughes, N. M. Luscombe, E. Ukkonen, and J. Taipale. 2010. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20(6):861-873.
2. Stormo, G. D., and Y. Zhao. 2010. Determining the specificity of protein-DNA interactions. Nat Rev Genet 11(11):751-760.
3. Rodriguez-Martinez, J. A., A. W. Reinke, D. Bhimsaria, A. E. Keating, and A. Z. Ansari. 2017. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. Elife 6.
4. Warren, C. L., N. C. Kratochvil, K. E. Hauschild, S. Foister, M. L. Brezinski, P. B. Dervan, G. N. Phillips, Jr., and A. Z. Ansari. 2006. Defining the sequence-recognition profile of DNA-binding molecules. Proc Natl Acad Sci U S A 103(4):867-872.
5. Tietjen, J. R., L. J. Donato, D. Bhimisaria, and A. Z. Ansari. 2011. Sequence-specificity and energy landscapes of DNA-binding molecules. Methods Enzymol 497:3-30.
6. Carlson, C. D., C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria, Y. Lee, F. Cerrina, and A. Z. Ansari. 2010. Specificity landscapes of DNA binding molecules elucidate biological function. Proc Natl Acad Sci U S A 107(10):4544-4549.
7. Lepre, C. A., J. M. Moore, and J. W. Peng. 2004. Theory and applications of NMR-based screening in pharmaceutical research. Chem Rev 104(8):3641-3676.
8. Wang, Z. X. 1995. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. FEBS Lett 360(2):111-114.
9. Sigurskjold, B. W. 2000. Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. Anal Biochem 277(2):260-266.
10. Chakraborty, S., P. J. Steinbach, D. Paul, H. Mu, S. Broyde, J. H. Min, and A. Ansari. 2018. Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. Nucleic Acids Res 46(3):1240-1255.
11. Chakraborty, S. 2018. Visualizing spontaneous DNA dynamics and its role in mismatch recognition by damage repair protein Rad4. University of Illinois at Chicago.
12. Connolly, M., A. Arra, V. Zvoda, P. J. Steinbach, P. A. Rice, and A. Ansari. 2018. Static Kinks or Flexible Hinges: Multiple Conformations of Bent DNA Bound to Integration Host Factor Revealed by Fluorescence Lifetime Measurements. J Phys Chem B.
13. Du, Q., A. Kotlyar, and A. Vologodskii. 2008. Kinking the double helix by bending deformation. Nucleic Acids Res 36(4):1120-1128.
14. Cai, X., D. S. Arias, L. R. Velazquez, S. Vexler, A. L. Bevier, and D. K. Fygenson. 2020. DNA Nunchucks: Nanoinstrumentation for Single-Molecule Measurement of Stiffness and Bending. Nano Lett 20(2):1388-1395.
15. Schindler, T., A. Gonzalez, R. Boopathi, M. M. Roda, L. Romero-Santacreu, A. Wildes, L. Porcar, A. Martel, N. Theodorakopoulos, S. Cuesta-Lopez, D. Angelov, T. Unruh, and M. Peyrard. 2018. Kinky DNA in solution: Small-angle-scattering study of a nucleosome positioning sequence. Physical Review E 98(4).

Rights and Permissions

parameters used to describe the geometry of complementary (or non-complementary) base pairs and sequential base pair steps (19). **The base pair reference frame (lower left) is constructed such that the x-axis points away from the (shaded) minor groove edge of a base or base pair and the y-axis points toward the sequence strand (I)**. The relative position and orientation of successive base pair planes are described with respect to both a dimer reference frame (upper right) and a local helical frame (lower right). Images illustrate positive values of the designated parameters. For illustration purposes, helical twist (Ωh) is the same as Twist (ω), formerly denoted by Ω (19,20) and helical rise (h) is the same as Rise (Dz).

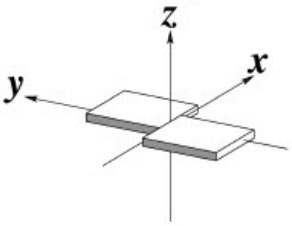
I recall spending around two weeks to produce the above figure. Content-wise, the figure was constructed in only a short while; it was the little details that took me most of the time.

Over time, I've witnessed numerous versions of such schematic images in publications related to DNA/RNA structures. While looking similar, the schematics differ subtly in the magnitude, orientation and relative scale of illustrated parameters. To the best of my knowledge, only 3DNA provides a pragmatic approach to generate the base-pair schematic diagrams consistently.

To make the schematics more readily accessible, I've reproduced a high resolution image (in png format) for each of the 14 parameters shown above. You are welcome to pick and match the diagrams as necessary. If you use any of them in your publications, please cite the 3DNA NAR03 [http://nar.oxfordjournals.org/content/31/17/5108.full] and/or NP08 [http://www.nature.com/nprot/journal/v3/n7/abs/nprot.2008.104.html] paper(s).

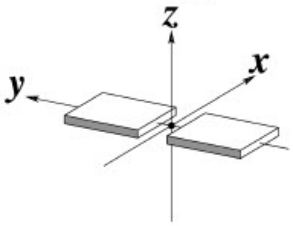
Note that in the schematic diagrams below, the shaded edge (facing the viewer) denotes the **minor-groove side** of a base or base pair.

Shear (*Sx*)



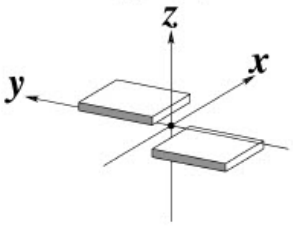
[http://home.x3dna.org/images/x3dna-shear.png]

Stretch (*Sy*)



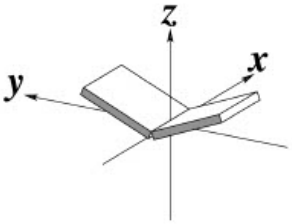
[http://home.x3dna.org/images/x3dna-stretch.png]

Stagger (*Sz*)



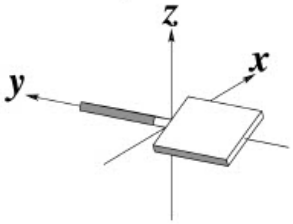
[http://home.x3dna.org/images/x3dna-stagger.png]

Buckle (κ)



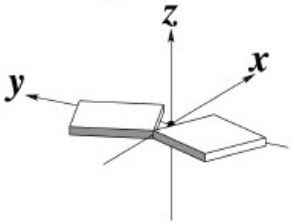
[http://home.x3dna.org/images/x3dna-buckle.png]

Propeller (π)



[http://home.x3dna.org/images/x3dna-propeller.png]

Opening (σ)



[http://home.x3dna.org/images/x3dna-opening.png]

Shift (*Dx*) **Slide (*Dy*)** **Rise (*Dz*)**

RCSB PDB-101 Educational portal of RCSB PDB PROTEIN DATA BANK

How to Cite

Materials and images are free for use. Please cite PDB-101 (PDB101.rcsb.org).

The reference for PDB-101 and RCSB PDB is:
The Protein Data Bank H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) *Nucleic Acids Research*, **28**: 235-242. doi:10.1093/nar/28.1.235

The Molecule of the Month series can be referenced using:
The RCSB PDB "Molecule of the Month": Inspiring a Molecular View of Biology D.S. Goodsell, S. Dutta, C. Zardecki, M. Voigt, H.M. Berman, S.K. Burley (2015) *PLoS Biol* **13**(5): e1002140. doi: 10.1371/journal.pbio.1002140

Individual Molecule of the Month features may be referenced using the author/s of the article and the Digital Object Identifier (DOI). The DOIs are listed on each article, and follow the format: 10.2210/rcsb_pdb/mom_YYYY_MM, where YYYY is year and MM the number of the month (one or two digits).

Visit the RCSB PDB for more information about our Policies and References.

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mitch Connolly

SPRINGER NATURE

The role of DNA shape in protein–DNA recognition

Author: Remo Rohs et al
Publication: Nature
Publisher: Springer Nature
Date: Oct 29, 2009

Review Order

Please review the order details and the associated terms and conditions.

No royalties will be charged for this reuse request although you are required to obtain a license and comply with the license terms and conditions. To obtain the license, click the Accept button below.

Licensed Content

Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	The role of DNA shape in protein–DNA recognition
Licensed Content Author	Remo Rohs et al
Licensed Content Date	Oct 29, 2009

Order Details

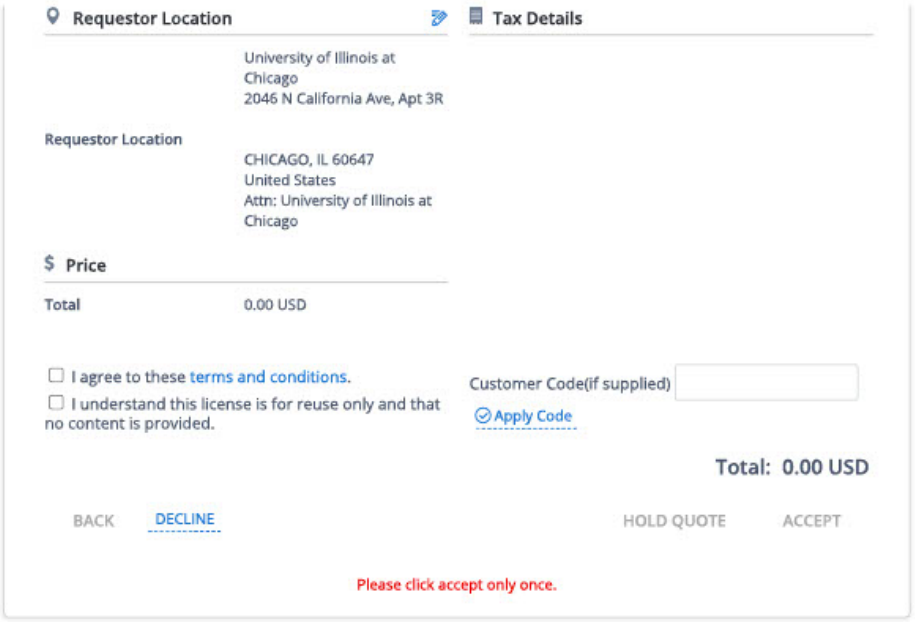
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

About Your Work

Title	Understanding DNA deformability by DNA bending studies with the IHF/HU family of architectural proteins
Institution name	University of Illinois at Chicago
Expected presentation date	May 2021

Additional Data

Portions	Figure 1



Requestor Location

University of Illinois at Chicago
2046 N California Ave, Apt 3R

Requestor Location

CHICAGO, IL 60647
United States
Attn: University of Illinois at Chicago

Tax Details

$ Price

Total 0.00 USD

I agree to these terms and conditions.
I understand this license is for reuse only and that no content is provided.

Customer Code(if supplied)

Apply Code

Total: 0.00 USD

BACK DECLINE HOLD QUOTE ACCEPT

Please click accept only once.

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mitch Connolly

SPRINGER NATURE

Dynamics and Mechanism of DNA-Bending Proteins in Binding Site Recognition

Author: Anjum Ansari, Serguei V. Kuznetsov
Publication: Springer eBook
Publisher: Springer Nature
Date: Jan 1, 2010

Order Completed

Thank you for your order.

This Agreement between University of Illinois at Chicago -- Mitch Connolly ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5037281209707
License date	Mar 27, 2021

Printable Details

Licensed Content

Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Dynamics and Mechanism of DNA-Bending Proteins in Binding Site Recognition
Licensed Content Author	Anjum Ansari, Serguei V. Kuznetsov
Licensed Content Date	Jan 1, 2010

Order Details

Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

About Your Work

Title	Understanding DNA deformability by DNA bending studies with the IHF/HU family of architectural proteins
Institution name	University of Illinois at Chicago
Expected presentation date	May 2021

Additional Data

Portions	Figure 6.1. - that compares nearest neighbor stacking energies with various methods

Requestor Location

University of Illinois at Chicago
2046 N California Ave, Apt 3R

Requestor Location

CHICAGO, IL 60647
United States
Attn: University of Illinois at Chicago

Tax Details

$ Price

Total	0.00 USD

Total: 0.00 USD

CLOSE WINDOW ORDER MORE

Requestor Location

University of Illinois at Chicago
2046 N California Ave, Apt 3R

Requestor Location

CHICAGO, IL 60647
United States
Attn: University of Illinois at Chicago

Tax Details

$ Price

Total	0.00 USD

Total: 0.00 USD

CLOSE WINDOW ORDER MORE

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mitch Connolly

Static Kinks or Flexible Hinges: Multiple Conformations of Bent DNA Bound to Integration Host Factor Revealed by Fluorescence Lifetime Measurements

ACS Publications
Most Trusted. Most Cited. Most Read.

Author: Mitchell Connolly, Aline Arra, Viktoriya Zvoda, et al
Publication: The Journal of Physical Chemistry B
Publisher: American Chemical Society
Date: Dec 1, 2018

Quick Price Estimate

This service provides permission for reuse only. If you do not have a copy of the portion you are using, you may copy and paste the content and reuse according to the terms of your agreement. Please be advised that obtaining the content you license is a separate transaction not involving RightsLink.

Permission for this particular request is granted for print and electronic formats, and translations, at no charge. Figures and tables may be modified. Appropriate credit should be given. Please print this page for your records and provide a copy to your publisher. Requests for up to 4 figures require only this record. Five or more figures will generate a printout of additional terms and conditions. Appropriate credit should read: "Reprinted with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.

I would like to...	reuse in a Thesis/Dissertatio	Format	Print
Requestor Type	Author (original work)	Select your currency	USD - $
Portion	make a selection	Quick Price	Click Quick Price

QUICK PRICE CONTINUE

To request permission for a type of use not listed, please contact the publisher directly.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Nov 15, 2020

This Agreement between University of Illinois at Chicago -- Mitch Connolly ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4950310814692
License date	Nov 15, 2020
Licensed content publisher	Oxford University Press
Licensed content publication	Nucleic Acids Research
Licensed content title	Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein
Licensed content author	Velmurugu, Yogambigai; Vivas, Paula
Licensed content date	Dec 18, 2017
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Understanding DNA deformability by DNA bending studies with the IHF/HU family of architectural proteins

Publisher of your work	University of Illinois at Chicago
Expected publication date	Feb 2021
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Understanding DNA deformability by DNA bending studies with the IHF/HU family of architectural proteins
Institution name	University of Illinois at Chicago
Expected presentation date	Feb 2021
Portions	Supplamentary Figures S1-S4, S7
Requestor Location	University of Illinois at Chicago 2046 N California Ave, Apt 3R CHICAGO, IL 60647 United States Attn: University of Illinois at Chicago
Publisher Tax ID	GB125506730
Total	0.00 USD

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing

Mitchell Connolly

EDUCATION

PhD **University of Illinois at Chicago**, Biophysics 2021
Thesis: "*In search of highly kinkable DNA sequences*"
Committee: Anjum Ansari (Chair), Phoebe Rice, Huan-Xiang Zhou, Mark Schlossman, David Hofman

MS **University of Illinois at Chicago**, Physics 2016
GPA: 3.53/4.00
Relevant Coursework: Written Communication in STEM, Molecular Biophysics of the Cell, Proteins and Proteomics

BS **Southern Illinois University – Carbondale**, Biomedical Physics 2014
GPA: 3.83/4.00
Honors Diploma, Chemistry minor
Provost's Scholar

RESEARCH HISTORY

Exicure, Inc.
Chicago, Illinois

Analyst, Platform Research 2020- Present

- Developed an internal software that automates several routine analyses and tasks for the research and development team.
- Identified motifs in antisense oligonucleotide sequences that exhibit higher therapeutic potential across several genes and tens-of-thousands targeted binding sites.

University of Illinois at Chicago
Chicago, Illinois

Research Assistant 2015- Present

- Designed and executed in vitro fluorescence-based assays to identify methods of target-site recognition by DNA-bending proteins.
- Modeled molecular interactions using MATLAB-based Monte Carlo simulations.
- Extracted DNA sequences of highest flexibility from random library by performing high-throughput selection paired with bioinformatical tools.
- Regularly read and evaluated new journal articles for information and implications for the field.
- Developed strong relationships with scientists in the field, resulting in thoughtful discussions and friendly collaborations.
- Published in reputable journals *Journal of Physical Chemistry B* and *Nucleic Acids Research* .
- Mentored three undergraduate researchers as they performed independent research projects. *(2/3 earned independent funding for their research,1/3 received research awards, 1/3 published, 3/3 to be co-authors for work in-preparation)*

Okinawa Institute of Science and Technology
Okinawa, Japan

Research Intern - NanoJapan Program (*NSF/Rice University*) Summer 2013

- 1 of 12 students awarded for this highly selective international research opportunity (<15% applicants admitted)
- Designed and manufactured antennas for broadband Terahertz wave generation. Scaled output of by 5-fold.
- Characterized 2-dimensional materials by Raman Spectroscopy.

- Initiated collaboration between labs in Okinawa and in Carbondale (Sothern Illinois University - Carbondale) leading to 3 joint publications.

PROFFESSIONAL EXPERIENCE

University of Illinois at Chicago
Chicago, Illinois

Course Writer (*Introduction to Physics for the Life Sciences I/II*) 2017-2019

- Led development of course "problem bank", with >300 original biophysics problems.
- Created weekly group problems and homework assignments for Semester I.
- Managed a team tasked with preparing material for Semester II.

Graduate Student Senator (*Executive Committee*) 2018- 2019

- Reviewed University System statutes and bylaws, considering all proposed amendments to these documents.
- Facilitated communication between a variety of stakeholders, from graduate student body to highest levels of university administration
- Contributed to successful senate proposal to increase funding to UIC counseling center by >$400,000 by instating a small ($15) mental health fee per student.
- Advised and critiqued administrative policies and documents, e.g. a section addressing "Active Shooter" instructions to appear on all syllabi.

Executive Board Member – Club Sports 2018- 2019

- Oversaw operations of >25 university sport club, with operating budget >$20,000 per academic year.
- Prepared monthly reports detailing activities, program information, and achievements of club sports.
- Led workshops communicating important institutional processes and deadlines to club leaders.

Co-founder, Treasurer – Fencing Club 2017-2018

- Created a thriving club for >15 members in the first year.
- Generated additional ~$2,000 from dues, fundraisers, university resources, and donations.
- Coordinated clubs first three competitions.

Representative – Graduate Student Council 2017-2018

- Reviewed policies and grant proposals, disseminating >$4,000 in project awards.
- Communicated university information and relevant events to department.

Teaching Assistant 2014-2017

- Assisted teaching of "Introductory Physics I", "Molecular Biophysics of the Cell", and "Quantum Mechanics."
- Led weekly laboratories and/or discussion sections for class sizes of 20-25 students for Introductory Physics I.
- Received excellent teaching evaluations, averaging 4.9/5.0.

Museum of Science and Industry
Chicago, Illinois

Volunteer 2015-2017

- Performed demonstrations introducing concept of gravity to museumgoers
- Tailored each demonstration to match audience age range and size

PUBLICATIONS

M. Connolly, J. Bathery, R. Rahman, P. Rice, A. Ansari. "Bound and bent? Specific and non-specific binding by Integration Host Factor characterized by steady-state fluorescence" (Manuscript in preparation)

M. Connolly, A. Arra, V. Zvoda, P. Steinbach, P. Rice, A. Ansari. "Static kinks or flexible hinges: conformational distributions of bent DNA bound to Integration Host Factor mapped by fluorescence lifetime measurements" *J. Phys. Chem. B*., 122(49), 11519-11534, (2018)

Y. Velmurugu, P. Vivas, **M. Connolly**, S.V. Kuznetzov, P. Rice, A. Ansari. "Two-step interrogation then recognition of DNA binding site by Integration Host Factor: an architectural DNA-bending protein" *Nucleic Acids Research*, 46(4), 1741-1755, (2018)

RESEARCH INTERESTS

DNA structure and dynamics: understanding sequence-, DNA damage-, and supercoiling-driven dynamics

DNA-protein interactions: especially curious about nucleosome and chromatin transcription regulation

Antisense and other oligonucleotide-based therepeutics

GRANTS AND AWARDS

James Kouvel Fellowship Award for Academic Excellence, University of Illinois at Chicago	2019
Chancellor's Student Service Award, University of Illinois at Chicago	2018,19
Award for Graduate Research, University of Illinois at Chicago	2018
Performer of the Year, Department of Theater, Southern Illinois University – Carbondale	2014
NanoJapan Fellowship for International Undergraduate Research, NSF/Rice University	2013
Provost Award for Academic Excellence, Southern Illinois University – Carbondale	2010

INVITED PRESENTATIONS

Department of Chemistry, Swiss Federal Institute at Lausanne	2019
Department of Physics, Southern Illinois University at Carbondale	2019
Annual Biophysical Society Conference, San Francisco, CA	2018
Department of Physics, University of Illinois at Chicago	2017
Department of Chemistry, University of Wisconsin – Madison	2016
St. Louis Area Research Symposium, Carbondale, IL	2014
Tokyo Terahertz Research Symposium, Tokyo, Japan	2013

ADDITIONAL TRAINING

"Beyond BioSaxs", Argonne National Laboratory	2019
"Physics of DNA", UIUC - Center for Physics of Living Cells Summer School	2017

WORKSHOPS

Chair, *A helpful timeline to get you through your 1st year of Grad School*	2018
Chair, *Qualifying Exams 101*	2016,17

PROFFESSIONAL AFFILITATIONS

Biophysical Society

Gibbs Society of Biological Thermodynamics

American Physical Society

Sigma Pi Sigma, Society of Physics Students Honors Society (Inducted 2014)