Characterizing IHF Binding to DNA Four-Way Junctions and Single-Stranded Forks

by

Veronica Birdsall
Class of 2015

A thesis submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the Degree of Bachelor of Arts with Departmental Honors in Molecular Biology and Biochemistry

Middletown, Connecticut April, 2015
Abstract

Integration Host Factor (IHF) is an architectural protein that binds and bends DNA, facilitating the formation of protein-DNA complexes important for gene regulation. IHF binds with high affinity to a specific consensus sequence in duplex DNA and induces a 160° bend upon binding. We have shown that IHF also binds DNA four way junctions (4WJ) that do not contain the consensus sequence with nanomolar affinity and 1:1 stoichiometry for the specific interaction. We have also observed that IHF binds DNA forks with nanomolar affinity. The binding to junctions and forks is in direct contrast to IHF binding to non-consensus linear duplex DNA, which is typically 1000-fold weaker. In this study we investigate whether the presence of the IHF consensus sequence influences IHF binding to DNA junctions and forks, as well as the structural features of these substrates that influence binding. We utilized gel shift and fluorescence binding assays to measure binding affinity and have observed that the high affinity for these non-native structures is independent of the presence of the consensus sequence. We have also identified structural features that appear to influence binding to forks. Through these measurements, we are also exploring whether the mechanism of recognition differs between junctions and forks.
Acknowledgements

I have so many people to thank, the first and most important being by mentor, PI, and basically idol Dr. Ishita Mukerji. Although unbelievably busy with multitudinous important duties, you always made time to help me when I needed it, and you believed in me when I was literally the most incompetent scientist ever. Thank you so much, you are inspiring as a scientist and as a person.

Next I would like to thank my lab mates Sudipta and Li Yan, for answering my 25 billion questions about literally everything, for letting me steal their buffers and autoclaved stuff, for teaching me basically everything I know, and for somehow becoming my good friends in the process of all that doing stuff for me. You guys are the best!

Extra special ultra huge thanks to my forever lab partner in crime Vivian. My experience in lab would never have been so great if you hadn’t been by my side through columns and disasters, protein and E. coli spills, and occasional miraculous successes. You are a phenomenal scientist and a great person, and I am going to miss you (and your cookies!) so much, but also I am not that worried because I plan on seeing you at mad conferences for the rest of our careers.

I would also like to thank my advisor Manju Hingorani and my readers!

Also thanks to my newest lab mates Noah, Fatima, Mariam, and Tamid, you guys are going to rock and I can’t wait to hear about how well you all do!

Moving out of Hall-Atwater thank you so much to my main homies Kevin, Neha, Lewis, Earl, Mansoor, Paul, Rachel, and Alex for being there for me throughout this process and throughout college in general. I literally would not even be a person probably if it weren’t for you guys. Also now you are all obligated to read my thesis because I mentioned you HA!

I would also like to thank my parents for always supporting and believing me, and for spending their life savings on my Wesleyan tuition. Love you guys!

Finally I would like to thank the Wesleyan Summer Science Research program and the REU National Science Foundation (NSF MCB-0843656) that have funded this research.
Table of Contents

Chapter 1 – Introduction........................................................................................................7-20

1.1 Architectural Proteins.....................................................................................................7
1.2 Heat Unstable (HU)........................................................................................................8
1.3 Integration Host Factor (IHF)........................................................................................8
1.4 IHF Sequence Specificity...............................................................................................9
1.5 IHF Bending of Duplex DNA.......................................................................................11
1.6 Holliday Junctions.........................................................................................................13
1.7 DNA Replication Forks..................................................................................................17
1.8 Research Goals: IHF recognition of junctions and forks............................................19

Chapter 2 – Theory and Methods..........................................................................................20-38

2.1 DNA Purification...........................................................................................................20
  2.1.1 Theory of DNA Purification....................................................................................20
  2.1.2 DNA Purification....................................................................................................20
2.2 DNA Annealing............................................................................................................22
  2.2.1 Theory of DNA Annealing....................................................................................22
  2.2.2 DNA Annealing....................................................................................................22
2.3 Protein Purification.......................................................................................................24
  2.3.1 Theory of Protein Purification................................................................................24
  2.3.2 Protein Purification................................................................................................24
2.4 Gel Mobility Shift Assay (GMSA)................................................................................26
  2.4.1 Theory of GMSA....................................................................................................26
  2.4.2 GMSA....................................................................................................................28
2.5 Fluorescence Spectroscopy...........................................................................................29
  2.5.1 Theory of Fluorescence........................................................................................29
  2.5.2 Labeling DNA with Fluorophores..........................................................................31
  2.5.3 Theory of Fluorescent Intensity, Anisotropy, Stoichiometry.................................34
  2.5.4 Binding Assays.......................................................................................................36
2.6 DNA Melting................................................................................................................37
  2.6.1 Theory of DNA Melting........................................................................................37
  2.6.2 DNA Melting Experiments.....................................................................................38

Chapter 3 – Results................................................................................................................38-71

3.1 Purification of IHF.........................................................................................................38
  3.1.1 IHF Activity Check.................................................................................................39
3.2 Binding of IHF to Junction J20....................................................................................40
  3.2.1 GMSA IHF and J20................................................................................................42
  3.2.2 Fluorescence Intensity IHF and J20.......................................................................43
  3.2.3 Stoichiometry of IHF binding to J20.....................................................................46
3.3 DNA Forks.....................................................................................................................47
  3.3.1 Annealing and Characterizing Forks......................................................................48
Chapter 4 – Discussion ................................................................. 71-77

4.1 IHF Binding to Junction J20 ................................................. 71
4.2 IHF Binding to Forks ............................................................ 72
4.3 Structural Parameters of IHF Binding to DNA .................. 75
4.4 Future Directions .............................................................. 77

Chapter 5 – References ............................................................. 78-81

Chapter 6 – Appendices ............................................................ 82-85

A.1 Buffers ................................................................. 82
A.2 Gel Recipes ............................................................... 82
A.3 GMSA Worksheet ...................................................... 83
A.4 Fluorescence Anisotropy Worksheet .................. 84
A.5 Stoichiometry Worksheet .............................................. 85

List of Equations

Equation 1: Determination of the dissociation constant ............... 27
Equation 2: The absorbance value of the DNA at 260 nm............ 33
Equation 3: Overall DNA concentration .................................. 33
Equation 4: The concentration of labeled DNA ......................... 34
Equation 5: Labeling efficiency .................................................. 34
Equation 6: Fluorescence anisotropy ........................................... 35
Equation 7: G factor ............................................................ 35
Equation 8: The equation used by Origin to fit intensity and anisotropy data .... 37

List of Figures

Figure 1: Crystal structure of IHF ................................................. 9
Figure 2: Superimposition of the TTR element ......................... 10
Figure 3: Intercalation of IHF proline 65 ................................... 12
Figure 4: A Holliday junction ..................................................... 14
Figure 5: Holliday’s model of junction formation ...................... 15
Figure 6: Junctions can be open or stacked .............................. 16
Figure 7: A model of the replication fork ................................. 17
Figure 8: Model of the formation of a Holliday junction-like structure .... 18
Figure 9: One form of the Jablonski diagram ......................... 30
Figure 10: Structures of Fluorescein and Rhodamine
Figure 11: An SDS-PAGE gel stained with Coomassie
Figure 12: Polyacrylamide gel showing binding of IHF to consensus duplex
Figure 13: A diagram of Junction J20
Figure 14: Gel mobility shift assay of IHF binding to J20
Figure 15: Fluorescence intensity IHF and J20_TAMRA
Figure 16: Fluorescence intensity IHF and J20_FAM
Figure 17: Junction JH1
Figure 18: Fluorescence intensity IHF and J20 Stoichiometry
Figure 19: Preparation of forks
Figure 20: Fork salt test
Figure 21: Melting curve for JH1_F17A fork
Figure 22: Melting curve for JH1_F17B fork
Figure 23: Melting curve for J34_F17 fork
Figure 24: Native gel checking the annealing efficiency of JH1_F12
Figure 25: JH1_F17A Fork
Figure 26: GMSA showing IHF binding to JH1_F17A fork
Figure 27: Change in fluorescence intensity IHF and JH1_F17A
Figure 28: Change in fluorescence anisotropy IHF and JH1_F17A
Figure 29: JH1_F17B Fork
Figure 30: GMSA showing IHF binding to JH1_F17B fork
Figure 31: Intensity experiment with JH1_F17B and IHF
Figure 32: J34_F17 Fork
Figure 33: GMSA showing IHF binding to J34_F17 fork
Figure 34: Change in fluorescence intensity IHF and J341_F17
Figure 35: JH1_F12 Fork
Figure 36: GMSA with IHF and JH1_F12
Figure 37: GMSA with IHF and JH1_F12
Figure 38: H1_F7 Fork
Figure 39: GMSA with IHF and H1_F7
Figure 40: Change in fluorescence anisotropy IHF and H1_F7
Figure 41: F10XB Fork
Figure 42: GMSA with IHF and F10XB
Figure 43: Model of HU Binding to duplex/single-stranded DNA

List of Tables

Table 1: The DNA constructs used in this study and their sequences
Table 2: The fluorescent dyes used in this study
Table 3: A summary of the dissociation constants for junctions
Table 4: The forks characterized in this study
Chapter 1 – Introduction

1.1 Architectural Proteins

Architectural proteins interact with DNA in many ways to create the three-dimensional structure of the genome. One class of such proteins in prokaryotes is the bacterial type II DNA-binding proteins, which function to condense the bacterial genome by binding and inducing bends in the DNA. They also participate in other processes that require DNA manipulation, such as replication, repair, recombination, and gene regulation. In E. Coli, two of the most abundant proteins of this class are Integration Host Factor (IHF) and Heat Unstable (HU), which are present in concentrations of 20,000 copies per cell and 30,000 copies per cell respectively (Travers A 1997). The proteins function by creating dramatic bends in DNA, and have together been found to be involved with the transcription of over 120 genes, as well as the assembly of several higher order nucleoprotein complexes important for site-specific recombination reactions and the initiation of replication at several sites. These proteins have several eukaryotic analogs, which, although they are not structurally similar, perform functions in eukaryotic cells that are identical to those performed in prokaryotic cells by HU and IHF. These include the HMG-domain proteins, which are involved in maintenance of chromatin structure as well as transcription. Furthermore, IHF, HU, and HMG-domain proteins all induce bends greater than 90° through the partial intercalation of a hydrophobic residue in-between base-pairs on the outside of the bend, thereby widening the minor groove, and minimize electrostatic repulsion in the inside of the bend through charge neutralization (Travers A 1997).
1.2 Heat Unstable (HU)

HU is an architectural protein found in *Escherichia coli* and other prokaryotes. Constituted by two approximately 10kDa subunits, it is homodimeric in most bacteria, but heterodimeric in enterobacteria. In E. Coli it is heterodimeric at the end of the growth phase and during the stationary phase. It is essentially structurally identical to IHF, and Rice et al found via X-ray crystal structures that the bodies of the two proteins can be superimposed with a root-mean-square deviation of 1.0 Å (Rice PA 1996). In addition, HU is 30-40% sequence identical to IHF. As it is basic and contributes to chromosomal compaction and negative supercoiling, HU is frequently referred to as histone-like (Swinger K 2003). Unlike IHF, HU displays little sequence specificity in binding, and appears to bind all un-nicked duplex DNA with similarly low affinity, with a $K_d$ of 200–2500 nM. However, HU binds much more tightly to structurally distorted DNA, often binding nicks, gaps, cruciforms and phased loops with nanomolar affinities (Swinger K 2004).

1.3 Integration Host Factor (IHF)

IHF is an E. Coli architectural protein that was originally discovered as a host factor for the phage λ integrative recombination reaction (Lorenz M 1999). It is small, consisting of only 20kDa, and heterodimeric, with two 10kDa subunits that are approximately 30% sequence identical. Crystallography data has suggested that the two subunits are intertwined to form a compact and largely alpha helical ‘body,’ with one beta ribbon arm extending upwards from each subunit to wrap around the DNA minor groove (Rice PA 1996). Although it is involved with replication via binding to oriC, transcriptional
regulation upstream of a few σ54- and σ70-dependent promoters, and various integration reactions, the primary role of IHF appears to be architectural. It introduces a sharp bend of >160° in DNA, thereby facilitating further DNA-protein interactions to construct the nucleoprotein arrays necessary for these processes.

Figure 1: Crystal structure of IHF (pink and grey) bound to nicked duplex DNA containing the consensus sequence (green) (Rice PA 1996).

1.4 IHF Sequence Specificity
Proteins that bind DNA generally recognize their target sequences through one of two general mechanisms: direct readout, in which the sequence is identified by the unique properties of the DNA base pairs via the major groove, or indirect readout, in which sequence-specific physical distortions of the DNA are recognized, as opposed to the base pairs themselves. In contrast with many other DNA binding proteins that recognize known sequences, IHF utilizes mainly indirect readout for DNA recognition (Lynch T 2003). This is clear from crystal structures of IHF bound to nicked duplex
DNA, which show that IHF makes no contacts with the major groove of the DNA and in the minor groove makes minimal hydrogen bonds with nitrogenous bases as hydrogen bond acceptors. Thus IHF must be using indirect readout, and must be recognizing an element of structure that is specific to the DNA it selectively binds (Swinger K 2004).

Figure 2: Superimposition of the TTR element of the consensus sequence from four IHF/DNA crystal structures. The white DNA is a mutated sequence, the black is wild type. The most drastic structural differences between the two are shown with arrows. Overall, the structure of the DNA in this case is determined by DNA sequence (Swinger K 2004).

Unlike many other bacterial type II DNA-binding proteins, and despite its dependence on indirect readout, IHF recognizes and binds DNA in both a sequence-specific and non-sequence specific manner. A consensus DNA-binding motif has been identified for
IHF through the evaluation of its more than 170 known binding sites. The motif consists of multiple small clusters of conserved bases located mainly in the 3’ end of the binding site. Two elements are most highly conserved, and include the sequence WATCAA, (where W=A/T) which starts near the center of the site, and the sequence TTR (where R=A/G), which starts four base pairs away from WATCAA in the 3’ direction. Some IHF binding sites also contain a poly-A tract of between four and six adenines starting 8-9 base pairs from WATCAA in the 5’ direction.

IHF-DNA contacts extend over approximately 35 base pairs, and a crystal structure of IHF bound to duplex DNA containing the H’ sequence indicates that IHF makes 26 contacts with the phosphate backbone and only 3 hydrogen bonds with the bases, none involving a donor or acceptor specific to a particular base (Aeling K 2006), thus underscoring the theory that IHF uses mainly indirect readout to identify specific binding sites.

1.5 IHF Bending of Duplex DNA

DNA bending is crucial to many biological processes in both prokaryotes and eukaryotes. Protein induced bends are involved in control of access to promoters, initiation of DNA replication, site-specific recombination, DNA repair, and genome organization and compaction. IHF-induced bends in prokaryotic DNA were first found to be involved with site-specific recombination of the bacteriophage λ genome into the E. coli chromosome. This recombination requires the interaction of two disparate regions of DNA, called the attP site and the attB site. IHF binds to three sites on the attP sequence, called H1, H2, and H’, and introduces three sharp bends in the DNA,
allowing for the formation of the three-dimensional DNA-protein complex required for site-specific recombination, called the intasome (Sinden RR 1994).

Several crystal structures from the Rice lab have elucidated the molecular interactions underlying IHF bending of duplex DNA. IHF protects >25 base pairs of DNA, with the consensus sequence predominantly located on the right side of the binding site, and appears to bend the duplex >160°. This bending is largely mediated by the intercalation of a proline at the tip of each of IHF’s arms in-between the DNA bases, which creates two kinks in the DNA 9bp apart. To position these prolines, the IHF arms must curl around the DNA, and while doing so interact substantially with the minor groove (Rice PA 1996).

Figure 3: Intercalation of IHF proline 65 (yellow) at the tip of the α subunit arm in between DNA bases. Consensus sequence is in green (Rice PA 1996).
To stabilize the bend, the phosphate backbone makes contact with 26 positively charged side chains, as well as with N-termini of each of the six helices that form the heterodimer. In addition, the ends of helices one and three form a clamp by binding opposite sides of the minor groove. Portions of the consensus sequence appear to be recognized by two disparate segments of IHF: the arm of IHFα and the body of the protein. Although the arm appears to associate directly with conserved bases using two arginines, this does not fully explain the sequence specificity of IHF, and ultimately it appears to be dependent on the accumulation of several small interactions, most of them indirect in nature. Thus IHF appears to be a minor groove-intercalating DNA bending protein with specificity that relies mostly in indirect readout (Rice PA 1996).

1.6 Holliday junctions

Helical junctions in DNA are crucial intermediates for homologous and site-specific recombination, and can be produced in vivo by strand invasion for the purposes of recombination, double-strand break repair, and fork reversal during replication. These junctions, defined by the exchange of strands between different helical DNA segments such that an intersection or branchpoint is formed in the DNA, rely on the covalent continuity of their strands to maintain their structural integrity (Lilley DMJ 2000). Four-way junctions, or Holliday junctions, are defined by the interchange of four DNA single-strands.
Figure 4: A Holliday junction composed of four partially homologous DNA single strands.

Holliday first described these structures in 1964 as intermediates as part of his model of homologous recombination, which involves the creation of nicks in two homologous chromosomes leading to single strand invasion and branch migration to create four-way junctions. Since then similar mechanisms have been described for DNA replication and repair, manifesting the profound biological relevance of DNA four-way junctions.
Figure 5: Holliday’s model of junction formation via strand invasion and branch migration during homologous recombination (Ho P 2001).

It is possible for Holliday junctions to exist in either of two basic conformations: an open planar structure in which the four arms are at 90° angles with one-another, or a folded structure characterized by the coaxial stacking of the helical arms. In solution DNA four-way junctions transition between these two states, with the presence of positively charged metal ions, which shield portions of the negatively charged DNA
phosphate backbone from one-another, favoring the stacked conformation (Déclais A C 2008). For any given four-way junction there are two different stacked conformers possible, depending on which pairs of helices stack with one another, and stacking is antiparallel, meaning that the continuous strand run in opposite directions on either side of the crossover.

Figure 6: Junctions can be open or stacked, depending on the concentration of divalent ions such as magnesium (a). In the stacked structure the arms can be parallel or antiparallel, but only antiparallel strands have been observed (b) (Lilley DM 2001).

Crystal structures have indicated that the formation of a junction minimally disrupts the structural features and associated hydration patterns of duplex DNA, thus suggesting that the formation of these structures can occur in vivo at relatively low energetic cost (Eichman B F 2002).
1.7 DNA replication forks

In both prokaryotes and eukaryotes, DNA forks are crucial for DNA replication. Forks, which consist of both double-stranded and single-stranded elements, are formed through a processive unwinding of the double helix by helicase proteins, which break the hydrogen bonds between DNA base pairs.

Figure 7: A model of the replication fork in vivo (Hamdan S 2010).

For whole-genome replication to occur, the entire genome must be unwound in this way allowing the resulting single strands to serve as templates for DNA polymerase, which synthesizes new DNA strands in the 5’ to 3’ direction (Hamdan S 2010). Throughout this process, DNA is vulnerable to mutative damage if endogenous or exogenous stress impedes the progression of the replication forks. Thus there are numerous mechanisms in place to preserve fork integrity and thereby prevent
replication errors. Several DNA damage response pathways have been implicated in the regulation of replication initiation, stabilization of stalled forks, promotion of replication restart, and facilitation of fork movement (Jones R 2012), and alterations to these pathways, known as “replication stress,” have been implicated in cancer development (Magdalou I 2014). One mechanism that has been proposed for the restart of replication in forks that have been blocked by DNA damage or other stressors is replication fork reversal, which involves the unwinding of a blocked fork into a four-stranded structure that resembles a Holliday junction. This reversal, also called fork regression, is thought to facilitate repair of single-stranded DNA lesions by relocating them into double-stranded portions as well as allow DNA repair enzymes access to the blocked fork (Atkinson J 2009).
Figure 8: Model of the formation of a Holliday junction-like structure via fork unwinding in the event of blockage of replication fork progress. 3’ end of DNA strands contain arrows (Atkinson J 2009).

Although IHF has long been known to assist in replication initiation by specifically binding and bending the duplex DNA at the E. coli chromosomal origin oriC, the binding of IHF to forks has not been studied, as IHF is thought to dissociate from the DNA immediately following replication initiation (Kazutoshi K 2014). However, this study details that that IHF appears binds tightly to fork constructs that are derived from well-characterized Holliday junctions.

1.8 Research goals: IHF recognition of junction and forks

Previous research in the Mukerji lab, done primarily by Olga Buzovetsky and Suzanne Ho, has elucidated the structural consequences of IHF binding to Holliday junctions, and the details thereof. However, the specific structural parameters that mediate IHF binding to junctions are still unclear. IHF clearly has a strong and specific sequence component to its recognition of duplex DNA, yet it also tightly binds Holliday junctions that do not contain its consensus sequence. Is sequence still a factor in the recognition of such structures as junctions? Or do the unique features of IHF that allow recognition of a sequence entirely through indirect readout also allow for the sequence-independent recognition of nucleic acid structures? Our lab has also determined that IHF has a high affinity for DNA forks. Is this mediated by the same structural features as binding to junctions, or sequence factors as binding to duplex DNA? What are these factors, and what can they tell us about structure-based recognition of specific DNA components by DNA-binding proteins? This thesis will attempt to begin to answer
these questions through the study of IHF binding to junctions and forks with different structural and sequence properties with the goal of parsing out the elements that promote IHF recognition of each class of DNA structure.

Chapter 2 – Theory and Methods

2.1 DNA purification

2.1.1 Theory of DNA Purification

Denaturing polyacrylamide gel electrophoresis can be used to separate out impurities and incorrectly synthesized DNA particles from oligomers of the desired sequence. This method denatures DNA with urea and utilizes the negative charge of DNA to pull the DNA through a gel matrix towards a positive electrode. Larger DNA particles will take longer to migrate through the gel, allowing different species present in a sample to be separated out based on size. Thus if a sample that contains different sized DNA species is run on this type of gel, the different components will be separated by size, allowing for the collection of a pure sample from the gel (Summer H 2009).

2.1.2 DNA Purification

Two gel plates, one 20 cm by 20 cm and one 20 cm by 22 cm were set up with spacer 0.2 cm thick and sealed with petroleum jelly. A 20% urea denaturing gel solution was prepared. A plug consisting of 3 mL of polyacrylamide solution with 15 µL of TEMED was first poured into the gel plates. After this was solidified, the rest of the solution,
along with 60 µL of TEMED, was poured into the plates, after which a two well comb was inserted. The gel was allowed to polymerize for an hour, and was then prerun at 400 V for 30 minutes at room temperature. The buffer used was 1X TBE. A syringe with a bent needle was utilized to remove air bubbles from the area between the gel and buffer, as well as to remove excess urea from the wells before loading the samples. The DNA sample was prepared using an 100 µL solution of 100 nanomolar DNA and 100 µL 2X formamide loading buffer (9:1 ratio of formamide to 10X TBE). This sample was heated for 3 minutes in a 90ºC water bath, spun down, and put on ice. The sample was loaded, and then the gel was run for approximately 5 hours at 400 V at room temperature. When the marker had traveled three fourths of the way down the gel, the gel was stopped and visualized using UV shadowing. The band containing the desired DNA was excised using a razor.

The DNA was separated from the gel matrix by electroelution (Whatman Elutrap). The excised gel containing the DNA was diced into small pieces and placed in the reservoir, and the electroelutor was run at 100 V at room temperature in 1X TBE. A well formed by a BT2 membrane, which contained the DNA, at one end and a 3,500 MWCO membrane, which allowed only DNA to pass through, at the other end, was used to collect the DNA. The position of the positive and negative electrodes was reversed for one minute prior to collecting a fraction to free any DNA that was trapped in the BT2 membrane, and send it into the well to be collected. One sample was taken after every hour of electroelution, five in total. A 3,500 MWCO spin column was used to concentrate the five samples. Residual acrylamide was removed from the samples via buffer exchange and ethanol precipitation.
2.2 DNA annealing

2.2.1 Theory of DNA Annealing

DNA structures for in vitro experiments are annealed through the heating and slow cooling of complementary DNA single strands. Each strand of the desired structure is added in equimolar quantity to prevent unwanted single stranded impurities. The solution is heated to break unwanted hydrogen bonds between base-pairs and denature any unwanted secondary structure in the single strands, and slow cooling allows the free single strands to hydrogen bond perfectly with their complementary strands. Two disparate single strands are used to anneal duplex and fork DNA while four are used to anneal junctions.

2.2.2 DNA Annealing

<table>
<thead>
<tr>
<th>DNA Structure</th>
<th>Sequences 5’ - 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Junction J20 (20bp arms, no consensus sequence)</td>
<td>JB20: CCT AGC AAG GGG CTG CTA CC</td>
</tr>
<tr>
<td></td>
<td>JH20: GGT AGC AGC CTG AGC GGT GG</td>
</tr>
<tr>
<td></td>
<td>JR20: CCA CCG CTC AAC TCA ACT GC</td>
</tr>
<tr>
<td></td>
<td>JX20: GCA GTT GAG TCC TTG CTA GG</td>
</tr>
<tr>
<td>(b) H134 Duplex (34bp, consensus sequence)</td>
<td>H134A: TAT GCA GTC ACT ATG AAT CAA</td>
</tr>
<tr>
<td></td>
<td>CTA CTT AGA TGG T</td>
</tr>
<tr>
<td></td>
<td>H134B: TAC GTC AGT GAT ACT TAG TTG</td>
</tr>
<tr>
<td></td>
<td>ATG AAT CTA CCA</td>
</tr>
<tr>
<td>(c) Fork J20 XB (10bp duplex, 10bp single strand, no consensus sequence)</td>
<td>JB20: CCT AGC AAG GGG CTG CTA CC</td>
</tr>
<tr>
<td></td>
<td>JX20: GCA GTT GAG TCC TTG CTA GG</td>
</tr>
<tr>
<td>(d) Fork JH1 XB (17bp duplex, 17bp)</td>
<td>JBH1: C CTC CGT CCT AAG TAG TGG CTG</td>
</tr>
<tr>
<td>Single Strand, Consensus Sequence</td>
<td>Consensus Sequence</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>CTA CCG GAA GGG</td>
<td><em>JXH1</em>: CCA GAC TGC AGT  <strong>AAT CAA CTA</strong> CTT AGG ACG GAG G</td>
</tr>
<tr>
<td><em>(e)</em> Fork JH1 XR (17bp duplex, 17bp single strand, consensus sequence)</td>
<td><em>JXH1</em>: CCA GAC TGC AGT <strong>AAT CAA CTA</strong> CTT AGG ACG GAG G</td>
</tr>
<tr>
<td><em>(f)</em> Fork J34 XB (17bp duplex, 17bp single strand, no consensus sequence)</td>
<td><em>J34B</em>: CCT CCG TCC TAG CAA GGG GCT GCT ACC GGA AGG G</td>
</tr>
<tr>
<td><em>(g)</em> Fork H134 F27 (27bp duplex, 7bp single strand, consensus sequence)</td>
<td><em>H134A</em>: TAT GCA GTC ACT ATG <strong>AAT CAA CTA CTT AGA TGG T</strong></td>
</tr>
<tr>
<td><em>(h)</em> Fork JH1 F22 (22bp duplex, 12 bp single strand, consensus sequence)</td>
<td><em>JXH1</em>(mod): CCA GAC TGC ATG <strong>AAT CAA CTA CTT AGG ACG GAG G</strong></td>
</tr>
<tr>
<td><em>(g)</em> Fork H134 F27 (27bp duplex, 7bp single strand, consensus sequence)</td>
<td><em>H134B(mod)</em>: C CTT CAA AAG TAG TTG ATT CAT AGT GAC TGC ATA</td>
</tr>
<tr>
<td><em>(h)</em> Fork JH1 F22 (22bp duplex, 12 bp single strand, consensus sequence)</td>
<td><em>JRH1</em>(mod): C CTT CAA CCA CCG TAG TTG ATT CAT GCA GTC TGG</td>
</tr>
</tbody>
</table>

Table 1: The DNA constructs used in this study and their sequences. H1 consensus sequence in red.

Equimolar quantities of each oligo were added to a solution of 20 mM Tris-HCl pH 7.6 and 300 mM NaCl in a 600 µL or 1.5 mL eppendorf tube. The final concentration of DNA was 5 µM. The cap of the tube was closed tightly and sealed with parafilm to prevent evaporation. The tube was placed in a water bath and heated at 70°C for four hours. After four hours, the water bath was turned off and the sample slowly cooled back to room temperature overnight. All of the duplex, junctions, and forks used in this study were annealed in this manner.
2.3 Protein purification

2.3.1 Theory of protein purification

To obtain IHF, E. Coli cells that have been transformed with an IHF overexpression vector must be grown and induced to overproduce IHF. Once the cell culture has reached a maximum optical density (OD), the cells are lysed and IHF is separated from the other cytosolic components through a series of purification steps, which include precipitations with Polymin P to remove IHF from the DNA of the cell, and well as ion exchange column chromatography which separates out different proteins from the cell lysate based on their affinity for a column material in varying concentrations of salt. This process yields 1 mg of pure IHF protein per liter of culture.

2.3.2 Protein purification

IHF was harvested from overproducing E. Coli strain 1084B cells. An LB-Amp plate was first streaked from a deep-freeze (-70 °C) glycerol stock of this strain and incubated overnight at 37 °C. A starter culture was grown from a single colony off of this plate in 100 mL of LB media containing 50 µg/ml Ampicillin and sterilized via autoclave. The starter culture was incubated overnight (fewer than 12 hours) at 37 °C, shaking at 220 RPM. A 1:100 dilution of this starter culture was added to 4 L of LB containing 50 µg/ml Ampicillin. This culture was incubated/shaken at 37 °C for five hours. After five hours, the OD was monitored at 650 nm every hour. When the OD reached 0.9 (which signifies the beginning of the exponential growth stage) the culture was induced through the addition of IPTG to a final concentration of 0.04 mM. After
this the cells continued to grow until they reached an OD of 2.6 (samples must be diluted to be measured at ODs of above 1), or as high an OD as possible in twelve hours. After the cells reached the appropriate OD, they were centrifuged at 8,000 RPM for 30 minutes, decanted, and the cell paste was stored at -70°C. The cell paste was thawed on ice and resuspended in Buffer A (25 mM Tris-HCL pH 7.4, 1 mM EDTA, 50 mM NaCl, 10% glycerol, 3 mM β-mercaptoethanol, Appendix A.1). To Buffer A was added 1/100 volume of 25 mg/mL lysozyme stock, KCl to a final concentration of 1M, and PMSF, and the cells were incubated on ice for 30 minutes, and then lysed using the homogenizer.

The lysate was then ultracentrifuged for one hour at 33,000 RPM, and then dialyzed against 4 L of Buffer A to remove KCl.

Polymin P was added (0.067 ml of 10% per ml of supernatant) slowly over 20 minutes while stirring, followed by an additional 20 minutes of stirring and centrifugation at 10,000 RPM for 30 minutes. The supernatant was saved and the pellet was resuspended in 10 ml Buffer A. More Polymin P was added using the same procedure, for a final concentration of 0.5%, followed by another round of centrifugation at 10,000 RPM for 30 minutes. The supernatants were combined.

Ammonium sulfate was added slowly to the supernatant over 20 minutes to a final ratio of 0.334 g/ml (ammonium sulfate/supernatant) while stirring, followed by an additional 20 minutes of stirring. The lysate was centrifuged at 18,000 RPM for 30 minutes. The supernatant was saved and more ammonium sulfate was added to a final ratio of 0.564 g/ml over 20 minutes of stirring, followed by another 20 minutes of stirring. The solution was centrifuged at 18,000 RPM for 30 minutes, and the pellet saved.
The pellet was resuspended in 20 mL of Buffer A and dialyzed twice against 4 L Buffer A, and centrifuged at 10,000 RPM for 10 minutes. The supernatant was applied to a Heparin Sepharose 6 Fast-Flow column equilibrated and pre-washed with Buffer A. A step gradient of Buffer A going from 0.1 M to 1.7 M NaCl was applied, with monitoring of the fractions at 20 nm to determine when the protein eluted (usually between 0.7 M and 1 M NaCl). SDS-PAGE gels can be run to test samples taken from every stage of the prep as well as from the column fraction to determine which fractions contain protein. These fractions were consolidated and concentrated. IHF can be stored in 50% glycerol in -20 °C.

2.4 Gel Mobility Shift Assay (GMSA)

2.4.1 Theory of GMSA

A 6.5% polyacrylamide “native” gel allows DNA to migrate towards the positive electrode in electrophoresis without denaturing or dissociating from tightly bound protein. This allows for the calculation of a dissociation constant for DNA/protein interactions using gel electrophoresis, for the gel separates free DNA segments of different sizes from each other and from protein/DNA complexes based on charge, weight, and shape. A free DNA band will run predictably lower than a protein/DNA complex band. The changing intensity of this band upon an increased concentration of protein added can be correlated to create a binding curve and calculate a Kd of binding between IHF and a DNA substrate using Equation 1. Syber green dye, a fluorescent dye that tightly binds DNA, is used to visualize the DNA bands.
Equation 1: Determination of the dissociation constant, $K_d$ from the molar fraction of DNA bound, $f$ (2.6a). Substituting 2.4b and 2.5a into 2.2 gives the equation for $K_d$, equation 2.6, which rearranges into 2.6a. A 1:1 stoichiometry of binding is assumed. The subscript T indicates total concentration (Buzovetsky, O., 2010).

2.4.2 Gel Mobility Shift Assay (GMSA)

Binding assays were carried out in 6.5% polyacrylamide gels (Appendix A.2).

Following the addition of 35 µL TEMED, the solution was quickly mixed and poured into an assembled gel setup with plates of 10 cm width and 8 cm height with 1.5 mm
spacers. A 1.5 mm 15 well comb was inserted into the gel, which was kept at room temperature for at least one hour to allow complete polymerization. The gel was prerun in 4 °C for 20 minutes at 150 V in 0.25X TBE. Samples with increasing concentrations of protein-DNA complexes according to a projected Kd were prepared in 600 µL eppendorf tubes. Each tube contained 10 nM DNA, and an increasing amount of protein was added to each successive tube. To calculate the quantities of protein, DNA, and buffer in each sample, a spreadsheet was used (Appendix A.3). All samples were prepared in a buffer of 5 mM Tris- HCl pH 7.5, 0.1 mM EDTA, 70 mM KCl, and 0.5% ethylene glycol (Appendix A.1). To prevent the diffusion of the DNA out of the lanes, 2µL of 20% Ficoll was added to the samples, which also yields tighter bands. A marker was prepared using 14.5 µL ddH2O, 5 µL gel loading dye, and 0.5 µL of DNA size marker. Samples were incubated for one hour at 4 °C before loading onto the gel. The gel was run in 4 °C at 100 V for approximately 1.5 – 2 hours, or until the marker, visible due to the loading dye, ran three-fourths of the way down the gel. The gel was immediately incubated at room temperature with 10 µL Syber Green fluorescent dye in 70 mL water for 20 minutes. After 20 minutes the gel was washed thoroughly with ddH2O and scanned using the Typhoon PhosphorImager 840 with the scanner set to blue (488nm). ImageQuant 5.2 was used to quantify the intensity of the free and bound DNA bands through the measurement of pixel volume. Pixel volume or DNA band intensity values were then plotted against protein concentration in Origin v. 9.0 to generate a binding curve, which was fit with a 1:1 binding model to calculate a dissociation constant.
2.5 Fluorescence spectroscopy

2.5.1 Theory of Fluorescence

When a photon of light interacts with a molecule, valence electrons in that molecule are often promoted from their ground state orbitals ($S_0$) to higher energy level orbitals, such as $S_1$, thereby launching the molecules into an excited state. Once the incident light is turned off, molecules do not remain in this excited state, but rather rapidly relax back to $S_0$. For most molecules this occurs through the transfer of the excess energy to solvent molecules or other molecules in the solution via collision. However, some molecules return to ground state by emitting the excess energy as light (Boyer R, 2000).

Luminescence, or the emission of light by a substance due to electronically excited states, can take the form of phosphorescence or fluorescence. Phosphorescence, or the emission of light from triplet-excited states in which the electron in the excited orbital has the same spin orientation as the ground-state electron, is not used in this study. However, fluorescence, which is due to excited singlet states with electrons in the excited orbital exhibiting the spin orientation opposite of that of the ground state electron, is a useful tool in solution. This is largely due to the fact that fluorescent lifetimes, or the average time between excitation and return to ground state, are typically around 10 ns, which is much shorter than the phosphorescent lifetimes. Phosphorescent lifetimes, which are in the milliseconds to seconds range, are often too slow to compete with nonradiative decay and quenching processes in solution (Lakowicz JR, 1999).
Figure 9: One form of the Jablonski diagram, which illustrates the processes that occur in a fluorescent molecule between the absorption and emission of light (Rice University).

Fluorescence generally occurs in aromatic molecules such as Fluorescein and Rhodamine, as illustrated in Figure 10.
Fluorescence is measured experimentally using a spectrofluorometer, which contains a light source, two monochromators, a sample holder, and a detector. In a typical experiment, one of the monochromators selects the excitation wavelength while the other analyzes the wavelength of the emitted light. The detector, which is placed at a 90° angle to the excitation beam to filter out light transmitted through the sample and scattered light, records the fluorescence emitted by the sample in response to the excitation beam (Boyer R, 2000).

2.5.2 Labeling of DNA with Fluorophores

To perform in-solution fluorescence experiments, DNA oligomers were labeled with succinimidyl-ester probes – either Carboxyfluorescein (FAM) or carboxytetramethylrhodamine (TAMRA) – via C6 amino linkers attached to the 5' end. Experiments were also conducted with intrinsically labeled DNA that contained the fluorescent guanine analog 6-methylisoxanthopterin (6MI), which base pairs normally with cytosine but fluoresces at 430 nm. We use the sequence ATFAA, which exhibits close to monomer-like fluorescence upon duplex formation (Moreno A 2014). 6-MI strands were purchased already labeled, so the following protocol describes only the
labeling with FAM and TAMRA of single strands that were subsequently used to create several different labeled forks and junctions.

<table>
<thead>
<tr>
<th>Dye</th>
<th>Excitation λ (nm)</th>
<th>Emission λ (nm)</th>
<th>Extinction Coefficient ε (1/cm M)</th>
<th>Correction Factor CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAM</td>
<td>494</td>
<td>519</td>
<td>77,000</td>
<td>0.32</td>
</tr>
<tr>
<td>TAMRA</td>
<td>555</td>
<td>580</td>
<td>91,000</td>
<td>0.32</td>
</tr>
<tr>
<td>6-MI</td>
<td>340</td>
<td>430</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 2: The excitation λ, emission λ, extinction coefficient, and correction factor of each of the fluorescent dyes used in this study. 6-MI is an intrinsic guanine analog, and thus has neither an extinction coefficient nor correction factor (Invitrogen 2015).

200 µg FAM and TAMRA were weighed into different 600 µL eppendorf tubes and dissolved in 14 µL dimethylsulfoxide (DMSO). 100 µg DNA, 7 µL ddH2O, and 75 µL 0.1 M (0.038 g/ mL) sodium tetraborate pH 8.5 were then added to each tube. The tubes were gently mixed, spun down, and left rotating covered in aluminum foil overnight. Excess dye was removed via several rounds of ethanol precipitation, in which 100 µL ddH2O and 15 µL 3 M NaCl were added to each reaction mixture, followed by 250 µL 90% EtOH that had been stored at -20 ºC. This mixture was then stored at -20ºC for 1 hour, and centrifuged at room temperature for 5 minutes. The supernatant, containing the free dye, was decanted and the pellet left to air-dry before the procedure was repeated. Rounds of ethanol precipitation continued until the supernatant was both colorless and showed minimal fluorescent activity when checked on the Fluoromax. Excess dye was also sometimes removed via buffer exchange in 3,5000 molecular weight cut-off (MWC) spin columns following ethanol precipitation. The labeling efficiency was then determined by measuring the absorbance of the labeled DNA at both 260 nm (the standard wavelength of absorption of nucleic acids)
and the maximum absorbance wavelength of each dye, which is 494 for FAM and 555 for TAMRA (Table 2). The dyes contribute to the absorbance at 260, thus the correction factor is used to calculate an accurate DNA absorbance (Equation 2).

\[ A_{260\text{nm}} = A_{260\text{nm}} - (A_{\text{max abs} \lambda \text{ dye}} \times CF_{\text{dye}}) \]

Equation 2: The absorbance value of the DNA at 260 nm is corrected for fluorophore contribution using the correction factor, which is calculated by dividing the absorption values of each dye at 260 nm by that at their maximum absorbance wavelength.

The concentration of the labeled DNA (Equation 3) is then calculated using the corrected absorbance value obtained via Equation 2.

\[ [\text{Labeled DNA}] = \frac{A_{260\text{nm}} - (A_{\text{max abs} \lambda \text{ dye}} \times CF_{\text{dye}})}{\varepsilon_{\text{DNA}} \times \% \text{ hypochromicity} \times l} \]

Equation 3: Overall DNA concentration is calculated using the corrected 260 nm absorbance value, the dye CF, the path length of the instrument, and the extinction coefficient and hypochromicity of the DNA (Ho S, 2013).

As poorly labeled DNA can make experiments difficult, the labeling efficiency must be calculated once the DNA concentration is known. The Beer-Lambert law is used to determine the concentration of labeled DNA (Equation 4).

\[ [\text{Labeled DNA}] = \frac{A_{\text{dye}}}{\varepsilon_{\text{dye}} \times l} \]

Equation 4: The concentration of labeled DNA is calculated using the absorbance of the dye at its maximum absorbance wavelength, the extinction coefficient of the dye and the path length of the sample.

Finally, labeling efficiency can be calculated using the total concentration of the DNA and the concentration of labeled DNA (Equation 5).
Labeling Efficiency = ([Labeled DNA]/[Total DNA]) * 100%

Equation 5: Labeling efficiency of a sample is calculated by dividing the concentration of labeled DNA by the total DNA concentration.

2.5.3 Theory of Fluorescence Anisotropy, Intensity, and Stoichiometry

Fluorophores preferentially absorb photons with electric vectors that are parallel to the fluorophore’s transition moment, which has a defined orientation relative to the axes of the molecule. As molecules are randomly oriented in solution, upon excitation with polarized light only those fluorophore molecules whose transition dipole happens to be parallel to the excitation vector will be excited. This phenomenon, called photoselection, results in partially polarized fluorescent emission, which also occurs along a fixed axis in the fluorophore. The relative angle between the axes of polarized excitation and emission will depend on the rotational movement of the fluorophore in solution and is called the maximum measured anisotropy. Fluorescence anisotropy (r) can be measured in solution and is calculated using Equation 6.

\[ r = \frac{I_{VV} - I_{VH}}{I_{VV} + 2I_{VH}} \]

Equation 6: Fluorescence anisotropy is determined by the fluorescence intensities of the vertically and horizontally polarized emission when the sample is excited by vertically polarized light (Lakowicz JR 1991).

As most fluorophores can rotate extensively within their 1-10ns excited-state lifetimes, the emission of polarized light is random leading most fluorophores to display anisotropies near zero. Larger molecules such as DNA oligomers and proteins rotate
more slowly however, and thus have measurable and predictable anisotropies in solution. Factors such as size and shape of the labeled molecule will alter the rotational correlation time of the molecule and thereby change the anisotropy. Thus measuring change in anisotropy is a useful way of measuring any event, such as the binding of a protein, which changes the size or shape of a labeled DNA substrate (Lakowicz JR, 1991). As a protein/DNA complex is larger than DNA alone, it will rotate more slowly in solution, leading to a decrease in rotational correlation time and more polarized light emission and therefore an increase in anisotropy.

As the monochromators do not pass vertically and horizontally polarized light with equal efficiency, a correction factor, called the G-factor, is calculated to accurately determine anisotropy (Equation 7).

\[
G \text{ factor} = \frac{I_{HV}}{I_{HH}}
\]

Equation 7: G factor is the ratio of the instrument’s sensitivity for vertically polarized light to its sensitivity for horizontally polarized light.

An event such as protein binding also often alters the intensity of emission of the fluorophore upon excitation, or the total counts per second of emitted photons. Intensity can either increase or decrease upon protein binding depending on a variety of factors including where the protein binds and the fluorophore used. Total intensity of polarized light is calculated by adding the intensities along the x, y, and z axes (Lakowicz JR 1991).

Finally, stoichiometry experiments are used to measure the number of proteins bound to each DNA molecule. This is done by saturating protein-DNA binding by using a DNA
concentration that is 10-fold higher than the Kd of the interaction and titrating up to a protein:DNA ratio of at least 3:1 (Ho S 2013).

2.5.4 Binding Assays

All fluorescence experiments were performed on a Horiba Jobin Yvon Fluoromax-4 spectrofluorometer. In-solution experiments examining the binding of IHF to fluorescently-labeled DNA four-way junctions and forks were performed in GMSA binding buffer (Appendix A.1) that contained low amounts of ultra-pure Tris and ethylene glycol instead of glycerol to minimize the fluorescence of the buffer. To calculate the amount of protein and DNA to add at each step of the experiment an anisotropy worksheet was used (Appendix A.4), and the DNA concentration was kept constant while an increasing amount of protein was added. Titration experiments using FAM or TAMRA were performed in 3 mm by 3 mm glass cuvettes with a total sample volume of 200 µL. FAM was excited at 494 nm and emission was recorded at 520 nm. TAMRA was excited at 555 nm and emission was recorded at 580 nm. Experiments using 6-MI were performed in 5 mm by 5 mm quartz cuvettes, with a total sample volume of 600 µL. 6-MI was excited at 340 nm and emission was recorded at 430 nm. Fluorometer slits were adjusted to 4 nm and 10 nm for excitation and emission respectively. Samples were incubated at 10 °C for 8 minutes before each recording. The polarizers were set at VH, VV, and VM for each experiment. VH and VV are used to calculate anisotropy, and VM, which is the “magic angle” of 54.7°, is used to measure intensity due to the lack of polarization bias at this angle (Lakowicz R. Joseph, 1999).
Fluorescence anisotropy and intensity values were plotted and analyzed using Origin v. 9.0, which uses the following equation to generate a binding curve to determine the dissociation constant of the protein/DNA complex:

\[
y = A_0 + (A_i + A_0) \times \left( \frac{(D + K + x) - \sqrt{(D + K + x)^2 - (4 \times D \times x)}}{2 \times D} \right)
\]

Equation 8: The equation used by Origin to fit intensity and anisotropy data. Parameters: \(y\) = intensity/anisotropy, \(A_0\) = initial intensity/anisotropy, \(A_i\) = final intensity/anisotropy, \(D\) = DNA concentration, \(x\) = protein concentration, and \(K\) = dissociation constant (Ho S, 2013).

Equation 8 is the same as Equation 1, with the change in intensity/anisotropy (y) subbed in for the change in fraction bound (f) in Equation 1.

2.6 DNA Melting

2.6.1 Theory of DNA Melting

As double-stranded DNA structures are heated, the molecules become more and more energetic until the hydrogen bonds between base pairs break, causing the structure to fall apart. As base stacking in double-stranded DNA decreases the overall absorbance through electronic interactions of the bases, DNA absorbance will increase as structures begin to fall apart. Thus an increase in absorbance can be used to calculate the melting temperature of a given DNA substrate. This is done by calculating the Tm, or the temperature at which half of the DNA is single stranded (Bloomfield A 2000).

2.6.2 DNA Melting Experiments
DNA melting experiments are performed on the Beckman DU 650 Spectrophotometer. Samples are heated from 10 °C to 90 °C, then cooled back down to 10 °C over the course of two hours, and the absorbance of each sample is automatically taken every 2 °C. Melt data are analyzed using MeltWin 3.5 to determine Tm.

Chapter 3 – Results

3.1 Purification of IHF
Growth of and purification of IHF from overexpressing E. Coli cells proceeded as expected based on numerous preps performed following the established protocol described in Chapter 2.1, and results mirrored those described by Olga Buzovetsky (Buzovetsky O 2010). One notable exception is the use of a Heparin-Sepharose Fast Flow 6 column instead of a Heparin-Sepharose CL-6B column, from which the protein eluted at 700 mM NaCl instead of the 1.0 M NaCl as described previously. The pure protein ran on an SDS-PAGE gel as a doublet corresponding to the two IHF subunits slightly above the 10 kD band of the BioRad Precision Plus Protein Standards marker, which is consistent with past results (Figure 11).
Figure 11: An SDS-PAGE gel stained with Coomassie Brilliant Blue showing heterodimeric 22 kD IHF (red arrow) purified using a Heparin-Sepharose Fast Flow 6 column.

3.1.1 IHF Activity Check

IHF has a known Kd of ~30 nM with a well-characterized 34mer duplex (H134) based on the well-known IHF H1 consensus sequence (Wang S 1995). Thus a GMSA with H134 is used to verify the activity of our newly purified IHF. Figure 12 shows that the purified IHF is active, as the GMSA indicates that the new IHF has a Kd of 28 nM +/- 15 nM for H134.
Figure 12: (a) Polyacrylamide gel showing binding of IHF to consensus duplex H134. The lower band is free DNA, the upper band is DNA bound to IHF. Protein concentration is from 0 (first lane) to 900nM (last lane). (b) Binding curve analysis of free DNA decrease in (a) created in Origin v. 9.0, indicating a Kd of 28 +/- 15 nM.

3.2 Binding of IHF to Junction J20

Previous studies of IHF binding to Holliday Junctions in the Mukerji Lab focused on examining the affinity to junctions annealed from 34bp single stranded DNA oligomers, which resulted in junctions with four 17bp arms. Previous studies had examined two junctions of this size, one random-sequence junction called J3, based on the J3 junction originally created by Lilley and co-workers (Duckett D 1988) and a junction containing the H1 consensus sequence (JH1). To further characterize IHF binding to Holliday Junctions, and ascertain if the length of junction arms influenced affinity, we elected to assess the binding affinity of IHF to a four-way junction with
shorter arms. We also considered that a smaller junction would potentially be a better substrate for X-ray crystallography. The J20 junction, annealed from four 20bp oligomers, is a four-way junction with 10bp arms. The sequence of this junction is derived from the stable J3 junction (Duckett D 1988).

Figure 13: A diagram of Junction J20 (Nocka L 2013).
3.2.1 GMSA IHF and J20

To be a useful substrate for co-crystallography, J20 must be tightly bound by IHF. In our initial characterization we used the method of GMSA to determine the dissociation constant (Figure 14).

![Figure 14: Gel mobility shift assay of IHF binding to J20. (a) Polyacrylamide gel, band shift observed upon adding an increasing concentration of IHF illustrates binding. (b) Analysis of free DNA band yields a $K_d$ of 170 ± 40 nM.](image)

As shown in Figure 4, binding is mainly observed at higher concentrations of protein with multiple bands. The GMSA was repeated several times, giving an average $K_d$ of 170 ± 40 nM (Figure 14).
3.2.2 Fluorescence Intensity IHF and J20

We also performed binding experiments in solution by monitoring the fluorescence intensity when IHF binds to a TAMRA end-labeled J20 as shown in Figure 15. Increasing concentrations of protein leads to an increase in fluorescence intensity and a saturable binding curve. The experiment was performed four times under different DNA concentrations ranging from 0.3 nM to 5 nM, yielding an average $K_D$ of 4.4 +/- 0.5 nM.

![Graph showing fluorescence intensity change upon titration of IHF into 0.5 nM of TAMRA-labeled J20. A 300% increase in fluorescence intensity is detected. The experiment was repeated four times with an average $K_D$ of 4.4 ± 0.5 nM. Fit generated in Origin v. 9.0.](image)

Figure 15: Fluorescence intensity change observed upon titrating IHF into 0.5 nM of TAMRA-labeled J20. A 300% increase in fluorescence intensity is detected. The experiment was repeated four times with an average $K_D$ of 4.4 ± 0.5 nM. Fit generated in Origin v. 9.0.
To ensure that the dissociation constant derived from the change in intensity was not fluorophore-dependent, the experiment was repeated again using FAM end-labeled J20 (Figure 16), which yielded a $K_D$ of $4.0 \pm 0.7$ nM, which agrees well with results obtained using TAMRA-labeled DNA.

Figure 16: Fluorescence intensity change upon an increasing titration of IHF into a constant concentration (1nM) of FAM-labeled J20. A 27% decrease in fluorescence is shown. A $K_D$ of 4.0 nM ± 0.7 nM was calculated from fitting the data to a one to one binding interaction as described.

These results are comparable to the binding affinities of IHF to the 17bp-arm junctions J3 (no consensus sequence) and JH1 (with consensus sequence), which have been
investigated extensively in the Mukerji Lab, in particular by Suzanne Ho and Olga Buzovetsky.

![Junction JH1 diagram](image)

Figure 17: Junction JH1. Consensus sequence in red. Junction J3 is identical with the exception of the consensus sequence (Ho, S 2013).

Buzovetsky reported that IHF binds J3 with a $K_D$ of 2.2 +/- 0.6 nM in solution (Buzovetsky O 2010), while Ho determined a similar affinity of IHF for JH1, and determined a $K_D$ of 1.4 +/- 0.8 nM in solution (Ho S 2013).

<table>
<thead>
<tr>
<th>Junction</th>
<th>IHF $K_D$ in solution (nM)</th>
<th>Consensus sequence?</th>
<th>Arm length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J3</td>
<td>2.2 +/- 0.6</td>
<td>No</td>
<td>17</td>
</tr>
<tr>
<td>JH1</td>
<td>1.4 +/- 0.8</td>
<td>Yes</td>
<td>17</td>
</tr>
<tr>
<td>J20</td>
<td>4.4 +/- 0.5</td>
<td>No</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 3: A summary of the dissociation constants found for IHF binding to the three junctions characterized by the Mukerji lab. J3 data is from Olga Buzovetsky’s thesis (Buzovetsky O 2010). JH1 data is from Suzanne Ho’s thesis (Ho S 2013).

3.2.3 Stoichiometry of IHF Binding to J20

To calculate accurate $K_D$ values, the stoichiometry of binding, or how many proteins bind each DNA substrate in this case, should be known. Although a 1:1 stoichiometry had been observed for IHF binding to J3 and JH1, an intensity experiment was performed under stoichiometric conditions to ensure that binding stoichiometry was the same for the IHF-J20 interaction. As shown in Figure 18, the fluorescence intensity increases until approximately 100 nM after which a plateau is observed consistent with a 1:1 binding stoichiometry.

![Figure 18: Fluorescence intensity change upon titrating and increasing concentration of IHF into a constant 100nM concentration of TAMRA-labeled J20. A 1:1 stoichiometry is observed.](image)
3.3 DNA Forks

In addition to binding Holliday junctions, this project demonstrates that IHF binds with high affinity to DNA single-stranded forks, a novel result. These interactions are of interest as the structural features of forks and IHF-fork interactions, in conjunction with the growing body of information on IHF recognition of duplex DNA and DNA four-way junctions, will lend insight into the structural basis of DNA recognition by IHF. This study investigated IHF binding to several different fork substrates (Table 1) to elucidate the influence of different structural features on IHF binding affinity for forks. Particularly, we were interested in how altering the length of the single-stranded and double stranded regions, as well as the overall size of the forks, would affect the binding affinity of IHF to these structures. In order to investigate the possibility of a sequence-based element to IHF recognition of forks, we also compared binding to forks that did and did not contain the H1 consensus sequence. Overall the goal of this project was to begin to delineate the structural features of this novel substrate that mediate and influence IHF binding.

To accomplish this, we investigated IHF binding to six different forks, in which we assessed the importance of consensus sequence location and orientation, length of the duplex region and overall length of the substrate. The sequences are given in Table 1. The specific forks investigated are: a 20bp fork with 10bp in both the single stranded and double stranded portions (F10 XB), and five 34bp forks, three with 17bp in both the single stranded and double stranded portions (JH1_F17A, JH1_F17B, J34_F17), one with 22bp in the double stranded and 12bp in the single stranded portions.
(JH1_F12), and one with 27bp in the double stranded and 7bp in the single stranded portion (H1_F7). IHF binding to these different structures, determined through Kd measurements, was investigated and compared in an attempt to elucidate the structural parameters that mediate tight binding of IHF to DNA substrates in vitro.

3.3.1 Annealing and characterizing forks

Despite multiple repetitions, there was difficulty in getting high-precision data while working with the 17bp forks in solution. As two DNA bands were seen in many of the native gels run after generating the forks, the possibility of the samples containing significant amounts of single stranded DNA was investigated by observing the DNA in a native or non-denaturing gel. Each fork was run next to one of its compositional single strands on the native gel to compare the run distance of the single strands to the lower band seen in the gel (Figure 19).
As the second band in the fork samples appeared to run the same distance as or lower than the single stranded samples, it was hypothesized that the fork samples contained significant quantities of single strand. This would interfere with fluorescence experiments as 50% of the single strand in a labeled fork sample would also be fluorescently labeled, but would not interact with IHF in the same way as the fork would, and thus would lead to noisy data. Having ruled out pipetting error as the cause of such inefficient annealing, we then attempted to improve annealing efficiency by adding large amount of NaCl, which would shield the phosphate backbones of the two different strands from repelling one another. We did this by increasing the NaCl in the DNA annealing buffer from 300mM to 1M, and running a side-by-side comparison on a native gel (Figure 20).

Figure 19: Preparation of forks. Each fork was run on a gel alongside one of its single strands to determine whether the lower band consistently seen with annealed forks was single strand or a different conformation.

1. JH1_F17B
2. JH1B
3. JH1_F17A
4. JH1R
5. J34_F17
6. J34B
The increased salt did not appear to improve the annealing efficiency. Thus we next decided to characterize the stability of the forks themselves by performing a slow melt experiment on JH1_F17A, JH1_F17B, and J34_F17. For each of the forks, the Tm appeared to be enough above room temperature (with 47 °C being the lowest melt temperature) that the forks should not fall apart at room temperature due to Tm. Moreover, as all native gels were run at 4 °C, the heat generated through electrophoresis was also unlikely to make them fall apart (Figures 21, 22, 23).
Figure 21: Melting curve for JH1_F17A fork. Tm of 53.87°C calculated.

Figure 22: Melting curve for JH1_F17B fork. Tm of 47.43°C calculated.
Forks with a shorter single-stranded portion were then annealed and a native gel was run to ensure that there was not excess single strand (Figure 24).

Figure 23: Melting curve for J34_F17 fork. Tm of 51.58°C calculated.

Figure 24: Native gel checking the annealing efficiency of JH1_F12 (lane 1, 14% single strand) and H1_F7 (lane 2, 13% single strand). Single strand percentage calculated using ImageQuant.
3.4 IHF Binding to Forks

3.4.1 *IHF binding to 17bp Forks: the effects of sequence*

The initial discovery of IHF binding to forks occurred with 34bp forks that contained double and single-stranded portions of equal size (17bp each). Upon verifying that IHF appeared to be tightly binding these novel structures, we first investigated whether the presence or position of the H1 consensus sequence in these forks influenced IHF binding. As forks are in some ways a structural intermediate between junction and duplex DNA, we were curious whether recognition of them more closely resembled that of junctions or duplex. As the presence of the consensus sequence drastically alters IHF recognition of duplex but has minimal effects on junction binding, testing its influence on fork binding was a logical first step. We investigated this using three structurally identical forks, two with different locations and orientations of the consensus sequence (JH1_F17A, JH1_F17B) and one without the consensus sequence (J34_F17).

*JH1_F17A*

The JH1_F17A fork, derived from the consensus junction JH1, has 5bp of the H1 consensus sequence in the double-stranded portion, and 8bp in the single-stranded portion (Figure 25). Significantly, the 3’ end of the consensus sequence lies in the single stranded region.
A GMSA was performed with IHF and JH1_F17A to assess binding. As shown IHF binding retards the mobility of the DNA in the gel. Considering the intensity of the bound DNA or the upper band of the gel as a function of IHF concentration generates a binding curve. Analysis of the curve gives a $K_D$ of $33 \pm 16$ nM for IHF and JH1_F17A (Figure 26).
Figure 26: GMSA showing IHF binding to JH1_F17A fork. $K_D$ of 33 +/- 16 nM was calculated via a binding curve generated in Origin v. 9.0 analyzing the increase in bound DNA with an increasing IHF concentration. The lowest band is single-stranded DNA resulting from inefficient annealing.

JH1_F17A with TAMRA attached to the 5’ end of the R strand (which is single-stranded) was then annealed and fluorescence intensity and anisotropy change were measured upon an increasing concentration of IHF (Figure 27, 28). In this case IHF binding leads to a decrease in fluorescence intensity. The decrease as a function of IHF concentration yields a saturable binding curve that gives a $K_D$ of 6.1 +/- 2.1 nM.
Figure 27: Change in fluorescence intensity upon an increasing concentration of IHF in a solution containing TAMRA-labeled JH1_F17A fork. $K_D$ of 6.1 +/- 2.1 nM.
Figure 28: Change in fluorescence anisotropy upon an increasing concentration of IHF in a solution containing TAMRA-labeled JH1_F17A fork. $K_D$ of $\leq 7$ nM.

Although this experiment was repeated four times, the data were relatively sporadic. This is likely due to the high concentration of single strand in the DNA sample, as well as the fact that the TAMRA was on the single-stranded 5’ end on the R strand, which would give it a higher mobility than usual and could potentially influence the fluorescence. Nevertheless, intensity and anisotropy experiments yielded $K_D$ values of $\leq 8.2$ nM and $\leq 7$ nM, respectively, which are in good agreement with each other and both indicating relatively high affinity binding between IHF and JH1_F17A. Thus, the anisotropy experiments verified that the observed change in fluorescence seen upon addition of protein was a result of protein binding.
The JH1_F17B fork, also derived from the consensus junction JH1, differs in that it has 8bp of the H1 consensus sequence in the double-stranded portion, and 5bp in the single-stranded portion (Figure 29). In this fork, the 5' end of the consensus sequence is located in the fork region.

Figure 29: JH1_F17B Fork. Derived from Junction JH1, 34bp total, 17bp duplex, 17bp single strand. Consensus sequence (red): 8bp duplex, 5bp single strand.

A GMSA was performed to estimate the $K_D$ for IHF binding to JH1_F17B (Figure 30). A binding curve generated from the intensity of the free DNA band in Figure 30 estimated the $K_D$ to be $32 \pm 10$ nM, which is very similar to that observed for JH1_F17A, suggesting that the amount of the consensus sequence in the fork region did not drastically alter IHF binding, nor did the difference in the orientation of the
consensus sequence. This suggests that IHF binding to forks is not necessarily
governed by the placement of the consensus sequence.

Figure 30: GMSA showing IHF binding to JH1_F17B fork. K_D of 32nM +/- 9.9nM was
calculated via a binding curve generated in Origin v. 9.0.

Fluorescence intensity experiments were then performed to corroborate the K_D value
obtained from the gel. JH1_F17B with TAMRA attached to the 5’ (duplex) end of the
B strand was annealed and the change in TAMRA fluorescence was correlated to an
increase in IHF in the solution. This experiment was performed three times, with the
average K_D ≤ 5.8 nM. This is again very similar to what was observed for JH1_F17A.
However, some of the data were inconsistent and had a large error, likely due to the
large quantity of single strand in the DNA sample. A representative binding curve is
shown in Figure 31.
Figure 3: Intensity experiment with JH1_F17B and IHF. $K_D$ of $2.4 \pm 1.3$ nM calculated via a binding curve generated in Origin v. 9.0.

**J34_F17**

The J34_F17 fork, derived from the non-consensus junction J3 (or J34), also has a 17bp double-stranded portion, and a 17bp single-stranded portion (Figure 32). It only differs from the JH1-derived forks in its lack of the H1 consensus sequence.
A GMSA was performed to evaluate the binding of IHF to this non-consensus fork, and a $K_D$ of $18 \pm 5.8 \text{ nM}$ was obtained from analysis of the free DNA band intensity (Figure 33), slightly tighter binding than that observed for the 17bp consensus forks. However, given the range of error in the determination of the binding affinities we consider this difference to be within the error range. Interestingly, binding to this fork leads to a number of gel-shifted bands, which either arises from a change in conformation of the fork or binding of more than one IHF proteins to the fork.
Figure 33: GMSA showing IHF binding to J34_F17 fork. $K_D$ of 18 +/- 6 nM was calculated via a binding curve generated from an analysis of the decrease in free DNA band intensity in Origin v. 9.0.

Fluorescence intensity experiments were then performed with the slightly different J34I_F17 fork, which was intrinsically labeled with 6MI. IHF binding to J34I_F17 should be comparable to its binding to J34_F17, as they have the same length of duplex and single stranded strand regions and they are both derived from the same junction sequence. Furthermore neither has any element of the IHF consensus sequence. Correlation of the intensity change of 6MI to IHF concentration is solution yielded a $K_D \leq 7$ nM for IHF binding to 6MI-labeled J34I_F17 (Figure 34), again within the error range of what was observed for the forks containing the H1 consensus sequence. Thus, the presence, location and orientation of the consensus sequence do not appear to drastically affect IHF binding to forks.
Figure 34: Change in fluorescence intensity upon an increasing concentration of IHF in a solution containing 6MI-labeled J34I_F17 fork. $K_D \leq 7 \text{ nM}$ observed.

3.4.2 IHF Binding to Shorter Forks: The effect of single-stranded length

As the consensus sequence did not appear to strongly influence IHF binding to forks, we elected to investigate the effects of other structural features, such as the length of the duplex region. The study of IHF binding to forks with different sized single stranded portions, as well as different ratios of duplex to single stranded portions, would lend interesting insight into the structural features that mediate the tight binding of IHF to forks. We created two forks with shorter single-stranded portions (thus longer duplex portions) to test whether the observed tight binding was dependent on a certain
number of single-stranded bases. Both new forks, called JH1_F12 and H1_F7, are 34bp total and contain the consensus sequence.

**JH1_F12**

JH1_F12 is identical to JH1_F17A except that the sequence of the R strand is altered such that the duplex is 22bp long and the single stranded portion is 12bp long. It contains 10bp of the H1 consensus sequence in the duplex and 3bp of the consensus sequence in the single-stranded portion (Figure 35).

Figure 35: JH1_F12 Fork. Derived from Junction JH1, 34bp total, 22bp duplex, 27bp single strand. Consensus sequence (red): 10bp duplex, 3bp single strand.

An initial GMSA performed with 10 nM DNA and a projected $K_D$ of 50 nM surprisingly showed almost no binding of IHF to JH1_F12 (Figure 36). As this fork is identical to the 17bp fork JH1_F17A, to which there was tight binding, except for the length of the single stranded portion, this much weaker binding indicates that the length of the single stranded segment might be important for high affinity IHF binding to forks.
Figure 36: GMSA with IHF and JH1_F12. DNA concentration is 10 nM, IHF concentration is from 0 nM (first lane) to 900 nM (last lane). Minimal binding is observed.

The GMSA was repeated; this time with a DNA concentration of 20 nM to aid in visualization, and with a projected $K_D$ of 100 nM (Figure 37), and the $K_D$ was calculated to be 130 +/- 100 nM, much higher than what was seen for the 17bp duplex forks in the gel. In addition, there is a great deal of nonspecific binding visible in the gel. Thus the shorter single-stranded portion appears to drastically reduce the tightness of IHF binding to 34bp forks. To further investigate this we then examined binding to a fork with an even shorter single-stranded portion.
Figure 37: GMSA with IHF and JH1_F12. DNA concentration is 20 nM, IHF concentration is from 0µM (first lane) to 1.9µM (last lane). K_D of 130 +/- 100 nM.

H1_F7

H1_F7 is derived from the 34bp consensus duplex used in the Mukerji Lab, H134, and consists of a 27bp duplex portion and a 7bp single stranded portion, thus containing the smallest single-stranded to double-stranded ratio of all of the forks studied. The H1 consensus sequence contains 12bp in the duplex portion and 1bp in the single stranded portion (Figure 38).
A GMSA performed with IHF and H1_F7 suggested a $K_D$ of 490 $\pm$ 130 nM (Figure 39), again much higher than what is seen with forks containing a larger single-stranded portion.

Binding to the H1_F7 fork was measured by a change in fluorescence anisotropy upon increasing the concentration of IHF in solution, and a $K_D$ of 36 $\pm$ 17 nM was
determined for IHF and H1_F7 in solution (Figure 40). This suggests that although IHF interacts with this fork, its binding to it is weaker than that measured for any other fork, junction, or consensus duplex in solution.

Figure 40: Change in fluorescence anisotropy upon an increasing concentration of IHF in a solution containing 6MI-labeled H1_F7 fork. $K_D$ of 36 +/- 17 nM.
3.4.3 IHF binding to 10bp forks: The effect of overall size

Having found one potential structural mediator of IHF binding - single-strand length - we investigated whether overall fork size affects IHF binding. To study this a small fork was annealed from two of the single strands used to anneal J20, creating a fork that is 20bp total, 10bp of duplex and 10bp of single stranded DNA, without the consensus sequence, called F10XB (Figure 41).

![F10XB Fork](image)

Figure 41: F10XB Fork. Derived from junction J20, 20bp total, 10bp duplex, 10bp single strand. No consensus sequence.

A GMSA between IHF and F10XB was run to investigate binding of IHF to this small fork. However, there was virtually no binding visible on the gel (Figure 42). This indicates that there is a lower limit, something larger than 20bp, for the overall size of a fork that IHF will tightly bind, in addition to a requirement for a single stranded portion that appears to be between 12 and 17bp.
Figure 42: GMSA with IHF and F10XB. DNA concentration is 10 nM, IHF concentration is from 0nM (first lane) to 900nM (last lane). Analysis of bound DNA using a Boltzmann fit indicated a $K_D \leq 315$ nM.

These results preliminarily suggest that although the consensus sequence does not appear to drastically mediate IHF binding to forks, there are some structural features, such as single stranded portion length and overall fork size, that influence this binding (Table 4).
<table>
<thead>
<tr>
<th>Fork</th>
<th>Total bp</th>
<th>Single-stranded bp</th>
<th>Consensus sequence?</th>
<th>$K_D$ (GMSA), nM</th>
<th>$K_D$ (solution), nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>JH1_F17A</td>
<td>34</td>
<td>17</td>
<td>Yes</td>
<td>33 +/- 16</td>
<td>$\leq 8.2$</td>
</tr>
<tr>
<td>JH1_F17B</td>
<td>34</td>
<td>17</td>
<td>Yes</td>
<td>32 +/- 10</td>
<td>$\leq 5.8$</td>
</tr>
<tr>
<td>J34_F17</td>
<td>34</td>
<td>17</td>
<td>No</td>
<td>18 +/- 6</td>
<td>N/A</td>
</tr>
<tr>
<td>J34I_F17</td>
<td>34</td>
<td>17</td>
<td>No</td>
<td>N/A</td>
<td>$\leq 7$</td>
</tr>
<tr>
<td>JH1_F12</td>
<td>34</td>
<td>12</td>
<td>Yes</td>
<td>130 +/- 100</td>
<td>N/A</td>
</tr>
<tr>
<td>H1_F7</td>
<td>34</td>
<td>7</td>
<td>Yes</td>
<td>490 +/- 130</td>
<td>36 +/- 17</td>
</tr>
<tr>
<td>F10XB</td>
<td>20</td>
<td>10</td>
<td>No</td>
<td>$\leq 315$</td>
<td>N/A</td>
</tr>
<tr>
<td>H134 duplex</td>
<td>34</td>
<td>N/A</td>
<td>Yes</td>
<td>28 +/- 15</td>
<td>1.04 +/- 0.91 (Ho, S 2013)</td>
</tr>
</tbody>
</table>

Table 4: The forks characterized in this study. $K_D$ values are the average obtained for IHF binding to each substrate. H134 duplex is included for reference (last row).

Chapter 4 – Discussion

4.1 IHF Binding to Junction J20

Our study has suggested that IHF binds the junction J20 with a $K_D$ of 4.4 +/- 0.5 nM, comparable, if slightly higher, to the dissociation constants recorded for IHF binding to the larger junctions J3 and JH1, which were 2.2 +/- 0.6 nM and 1.4 +/- 0.8 nM respectively (Table 3). This indicates that in addition to the presence of the H1 consensus sequence, changes in the arm length of Holliday junctions do not drastically affect tightness of IHF binding. This is important not only because it rules out arm
length as a structural parameter important for IHF binding to junctions, but also because it corroborates the hypothesis previously advanced by the Mukerji lab that IHF binds to the center, not the arms, of Holliday junctions. Buzovetsky initially suggested this upon finding tight binding (2.2 +/- 0.6 nM in solution) to the random-sequence Holliday junction J3, since IHF binding to random-sequence duplex DNA is ~1000 times weaker than binding to the H1 consensus sequence, and generally in the low micromolar K_D range (Wang S 1995). Therefore if IHF was binding to junction arms and functionally recognizing the arms as duplex, binding to the random-sequence junction could not possibly be in the specific nanomolar range that Buzovetsky observed in solution. The tight binding observed between IHF and junction J20 corroborates that result, for the arms of this smaller junction are only 10bp long, much shorter than the 35bp that IHF has been shown to contact in duplex binding (Aeling K 2006), and which are likely necessary for a tight interaction between IHF and duplex DNA. Thus the results shown here provide further evidence that IHF interacts primarily with the center of Holliday junctions, thereby narrowing the scope of further investigations into these interactions.

4.2 IHF Binding to Forks

This thesis presents the novel result that IHF tightly binds certain single-stranded forks. Although there is a great deal still to be learned about this binding, the investigation of IHF binding to various fork structures has preliminarily indicated some of the features of this novel IHF substrate that appear to influence tightness of binding, and some that do not. The first aspect of forks that we investigated was the presence and location of
the IHF H1 consensus sequence. Whether the presence of the consensus sequence in a fork affects IHF binding affinity is interesting due to the contrast in how the consensus sequence influences IHF binding to duplex versus junction DNA. Tight IHF binding to duplex DNA is strongly promoted by the presence of the H1 consensus sequence, yet in the case of IHF binding to DNA four-way junctions, recognition appears to be entirely structural with no sequence component. Therefore, upon finding a new IHF binding substrate the first question to answer is whether binding is sequence dependent as in duplex, or mediated entirely by the overall DNA structure, as in four-way junctions.

We investigated this by studying binding to three forks, two with the H1 consensus sequence (JH1_F17A, JH1_F17B), and one that was identical except that it lacked the consensus sequence (J34_F17). We found very tight binding to all of these forks both in the gel and in solution (Table 4). There was slightly tighter binding to JH1_F17B (≤ 5.8 nM in solution), which contained more of the consensus sequence in the duplex portion, and in which the 5’end of the consensus was located in the fork, than to JH1_F17A (≤ 8.2 nM), which contained more of the consensus sequence in the single-stranded portion, and in which the 3’ end of the consensus sequence was located in the fork, indicating that placement and/or orientation of the consensus sequence may have a slight effect on binding, yet the total lack of consensus sequence did not impede IHF binding to J34_F17.

Having thus determined that sequence is not an important factor in IHF binding to single-stranded forks, we next decided to investigate the impact of structural features, the most obvious being the length of the single-stranded portion of the forks. To do this we examined IHF binding to two forks: JH1_F12, and H1_F7, both of which contained
a higher duplex:single-stranded ratio than the previously studied forks. Here we found a
drastically altered binding affinity, and observed dissociation constants 5 to 15 times
higher than with forks with longer single-strand in the gel, and for one of the forks
(H1_F7) a dissociation constant more than 2 times higher than what was observed for
the longer single-strand forks in solution (Table 4). This indicates that the length of the
single-stranded portion of forks is an important factor in IHF recognition and binding to
these structures.

Finally, we wondered whether overall size of the forks influenced IHF binding. We
investigated binding of IHF to F10XB, a 20bp fork much shorter than the 34bp forks
that we had been studying, and found again much weaker binding to this fork,
indicating that there is a lower limit for the overall size of forks that IHF will bind
tightly to, and that it is something larger than 20bp. This was particularly interesting
since the shorter 20bp arm length in the case of the junction did not reduce binding
affinity.

Although there is still a great deal of work to be done in characterizing IHF binding to
DNA single-stranded forks, this study has indicated that tight IHF binding to these
substrates is largely structure-mediated, dependent on length of the single stranded
portion of the fork (with weaker binding to forks with single stranded portions shorter
than their duplex portions), and contingent on the overall size of the forks, with forks
longer than 20bp total necessary for tight binding.
4.3 Structural Parameters of IHF Binding to DNA

Although there are numerous DNA binding proteins that could be potential objects of biochemical study, we believe that IHF is a unique protein for two reasons. The first is the unusual duality in the way that IHF recognizes DNA. As mentioned, tight IHF binding to duplex DNA is clearly highly sequence-dependent, with the presence of the H1 consensus sequence drastically increasing IHF binding to duplex DNA. This is certainly not unusual for a DNA-binding protein, yet this sequence recognition by IHF appears to be mediated largely by indirect readout. Thus, even in the case of sequence-dependent binding, IHF appears to actually be responding to the ‘micro-structure’ of the DNA created by the base pairs, not the base pairs themselves. Furthermore, IHF recognition of DNA four-way junctions, highly irregular iterations of DNA structure, appears to have no sequence component whatsoever. Clearly the IHF mechanism of DNA recognition is a unique hybrid of sequence-specific and structure-specific elements, and may serve as the key to understanding the enigmatic process of specific structure-based recognition of DNA, a field of study that has important repercussions for understanding protein-DNA interactions.

Moreover, IHF is made even more interesting by its second unique quality, which is the contrast in recognition mechanism between IHF and its structural homolog HU. Despite having an entirely identical structure and a 30% identical sequence to IHF, HU has no observed sequence specificity in binding, and appears to rely entirely on structure. Interestingly, HU is known to tightly bind DNA forks and other replication and recombination intermediates ~1000 fold more tightly than duplex DNA. Kamashev and Rouviere-Yaniv found using GMSA that HU binds a 40bp double-stranded fork (20bp
in the single-stranded portion) similar to those studied here with a $K_D$ of 4 nM (Kamashev D 2000). They also hypothesized that the alpha helical HU body interacts with the single-stranded DNA while the arms intercalate between the double-stranded base pairs (Figure 43). Comparison of between IHF and HU binding of forks could yield valuable information about how these structures are recognized.

Figure 43: Model of HU Binding to duplex/single-stranded DNA (Kamashev D 2000).

Overall, the study of these two proteins is an important avenue into the world of structure-based recognition.
4.4 Further Research

As this study largely consists of preliminary investigations into IHF binding to single-stranded forks, there is still a great deal to be done. Most of the experiments done in this study should be repeated to yield more precise and accurate information about IHF binding to each of the forks studied. We could also investigate IHF binding to forks with single stranded regions that are between 12 and 17 base pairs long, to determine the lower limit for low nanomolar $K_D$ interactions with IHF. Kamashev et al found that HU binds DNA with a 3’ overhang with uniform high specificity as long as the single stranded portion was between 8 and 20bp long (with 20 being the maximum length they investigated). It would be interesting to determine whether IHF has a similar range for forks (Kamashev D 2000). Furthermore, upon discovering that IHF binds single stranded forks, it might be interesting to investigate whether IHF also binds tightly to any of the other structures that Kamashev et al investigated HU binding to, such as three way junctions, duplex DNA with a single stranded overhang, or double-stranded forks.

Next, our understanding of IHF binding and recognition of forks and junctions would greatly benefit from X-ray co-crystal structures of IHF bound to a junction and to a fork. This would yield invaluable information about where IHF is binding these structures and what the interactions look like. A crystal structure would give us a better understanding of how to direct future study, namely which structural features of junctions and forks warrant more investigation, allowing us to design more junctions and forks to hopefully more clearly delineate the binding parameters of IHF to these structures.
Chapter 5 – References


Chapter 6 – Appendix

A.1 Buffers

70 mM FRET Buffer
5mM Ultra-Pure Tris-HCl pH 7.6
0.1mM EDTA pH 8.0
350 ul 2M Stock of KCl
0.5% ethylene glycol

Buffer A – 10x
50mM Tris-HCl pH 7.4
20mM NaCl
1mM EDTA
3mM BME
10% glycercol

A.2 Gels

16% SDS-PAGE Recipe (Resolving) 4% SDS-PAGE Recipe (Stacking)
4.9 ml ddH2O 5.8 ml ddH2O
3.75 ml 1.5M Tris-HCl pH 8.8 0.94 ml 1M Tris-
6 ml 40% Acrylamide 0.76 ml 40%
75 ul 20% SDS 38 ul 20% SDS
150 ul 10% APS 75 ul 10% APS
15 ul TEMED 7.5 ul TEMED

6.5% Polyacrylamide Gel (Native)
19.36 ml ddH2O
3.9 ml 40% Acrylamide
0.6 ml 10x TBE
120 ul 10% APS
30 ul TEMED
### A.3 Spreadsheets

**GMSA Worksheet**

| Sample | Fraction DNA bound | [DNA] | [HIF] | [Protein] | Vol HIF ul | Vol HIF F22 ul | Vol HIF Stock | Vol Buffer | Molecules [HIF] | [DNA] per well | Vol 1:100 Dilution | Vol 1:100 Dilution | Vol 1:100 Dilution | Vol 1:100 Dilution |
|--------|------------------|-------|-------|-----------|------------|--------------|--------------|------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1      | 1.00E-08          | 2.00E-05 | 3.50E-07 | 3.50E-06  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 2.00E+05       | 1.00E+00       | 1.00E+00         | 1.00E+00         | 1.00E+00         | 1.00E+00         |
| 2      | 0.08             | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 3      | 0.15             | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 4      | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 5      | 0.5              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 6      | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 7      | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 8      | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 9      | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 10     | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 11     | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 12     | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 13     | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |
| 14     | 0.4              | 0.00E+00 | 0.00E+00 | 0.00E+00  | 0.00E+00   | 0.00E+00     | 0.00E+00     | 0.00E+00 | 0.00E+00       | 0.00E+00       | 0.00E+00         | 0.00E+00         | 0.00E+00         | 0.00E+00         |

**Note:** kd = Molar
<table>
<thead>
<tr>
<th>Stock A</th>
<th>Stock B</th>
<th>Total DNA</th>
<th>Desired DNA</th>
<th>Original DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'00E-06</td>
<td>3'00E-06</td>
<td>4'00E-06</td>
<td>6'00E-06</td>
<td>4'00E-06</td>
</tr>
<tr>
<td>1'00E-06</td>
<td>2'00E-06</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>3'00E-09</td>
<td>2'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>6'00E-09</td>
<td>3'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>9'00E-09</td>
<td>4'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>1'20E-09</td>
<td>5'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>1'50E-09</td>
<td>6'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>1'80E-09</td>
<td>7'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>2'10E-09</td>
<td>8'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>2'40E-09</td>
<td>9'00E-09</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>2'70E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>3'00E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>3'30E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>3'60E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>3'90E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>4'20E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>4'50E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>4'80E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>5'10E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>5'40E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>5'70E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>6'00E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>6'30E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>6'60E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>6'90E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>7'20E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>7'50E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>7'80E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>8'10E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>8'40E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>8'70E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>9'00E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>9'30E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>9'60E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>9'90E-09</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
<tr>
<td>1'00E-08</td>
<td>1'00E-08</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
<td>0'00E+00</td>
</tr>
</tbody>
</table>
### Stoichiometry Worksheet

#### Stoichiometry IHF + JH1B_Rho 6-24-2014

<table>
<thead>
<tr>
<th>Original [DNA]</th>
<th>4.00E-07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original [IHF]</td>
<td>2.90E-06</td>
</tr>
<tr>
<td>DNA Stock</td>
<td>4.00E-07</td>
</tr>
<tr>
<td>1:10 dilution</td>
<td></td>
</tr>
<tr>
<td>Desired [DNA]</td>
<td>1.00E-07</td>
</tr>
<tr>
<td>( k_d )</td>
<td>5.00E-09</td>
</tr>
<tr>
<td>total vol</td>
<td>2.00E-04</td>
</tr>
</tbody>
</table>

\[
\text{desired [IHF] in stocks with constant [DNA]}
\]

<table>
<thead>
<tr>
<th>NAME</th>
<th>[DNA], M</th>
<th>[P]/[D]</th>
<th>[IHF], M</th>
<th>Volume Needed</th>
<th>Volume, uL</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0</td>
<td>1E-07</td>
<td>0.00E+00</td>
<td>0.00E+00</td>
<td>0.00E+00</td>
<td>0</td>
</tr>
<tr>
<td>S1</td>
<td>1E-07</td>
<td>0.25</td>
<td>2.50E-08</td>
<td>8.62E-06</td>
<td>8.62</td>
</tr>
<tr>
<td>S2</td>
<td>1E-07</td>
<td>0.5</td>
<td>5.00E-08</td>
<td>9.01E-06</td>
<td>9.01</td>
</tr>
<tr>
<td>S3</td>
<td>1E-07</td>
<td>0.75</td>
<td>7.50E-08</td>
<td>9.43E-06</td>
<td>9.43</td>
</tr>
<tr>
<td>S4</td>
<td>1E-07</td>
<td>1</td>
<td>1.00E-07</td>
<td>9.90E-06</td>
<td>9.90</td>
</tr>
</tbody>
</table>

**Total Volume** 3.70E-05

### Stock A

<table>
<thead>
<tr>
<th>NAME</th>
<th>[DNA], M</th>
<th>[P]/[D]</th>
<th>[IHF], M</th>
<th>Volume Needed</th>
<th>Volume, uL</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5</td>
<td>1E-07</td>
<td>1.5</td>
<td>1.50E-07</td>
<td>7.41E-06</td>
<td>7.41</td>
</tr>
<tr>
<td>S6</td>
<td>1E-07</td>
<td>2</td>
<td>2.00E-07</td>
<td>7.69E-06</td>
<td>7.69</td>
</tr>
<tr>
<td>S7</td>
<td>1E-07</td>
<td>2.5</td>
<td>2.50E-07</td>
<td>8.00E-06</td>
<td>8.00</td>
</tr>
<tr>
<td>S8</td>
<td>1E-07</td>
<td>3</td>
<td>3.00E-07</td>
<td>8.33E-06</td>
<td>8.33</td>
</tr>
</tbody>
</table>

**Total Volume** 3.14E-05

### Stock B

<table>
<thead>
<tr>
<th>STOcks</th>
<th>IHF STOCKS</th>
<th>DNA STOCKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>total volume</td>
<td>moles IHF</td>
<td>moles DNA</td>
</tr>
<tr>
<td>A</td>
<td>4.00E-05</td>
<td>2.32E-11</td>
</tr>
<tr>
<td>B</td>
<td>3.50E-05</td>
<td>5.08E-11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>free DNA</th>
<th>Vol DNA</th>
<th>5.00E-05</th>
<th>0.00E+00</th>
<th>1.50E-04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock A final volume</td>
<td>1.00E-05</td>
<td>8.00E-06</td>
<td>2.20E-05</td>
<td></td>
</tr>
<tr>
<td>Stock B final volume</td>
<td>8.75E-06</td>
<td>1.75E-05</td>
<td>8.75E-06</td>
<td></td>
</tr>
</tbody>
</table>