

THE STRUCTURE AND FUNCTION OF THE siHF PROTEIN

CHARACTERIZING THE STRUCTURE AND FUNCTION OF A NOVEL NUCLEOID-ASSOCIATED
PROTEIN sIHF

By TAMIZA NANJI, B. Sc. (Hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements
for the Degree Master of Science

McMaster University

© Copyright by Tamiza Nanji, July 2014

MASTER OF SCIENCE (2014)
(Biochemistry and Biomedical Sciences)

McMaster University
Hamilton, Ontario

TITLE: Characterizing the structure and function of a novel nucleoid-associated protein sIHF

AUTHOR: Tamiza Nanji, B. Sc (Hons.) (McMaster University)

SUPERVISOR: Dr. Alba Guarné

NUMBER OF PAGES: xi, 68

Abstract

All living organisms must organize their genome so that it not only fits within the cell, but remains accessible for cellular processes. In bacteria, an arsenal of nucleoid-associated proteins contributes to chromosome condensation. A novel nucleoid-associated protein was recently discovered in actinobacteria, and is essential in *Mycobacterium*. It was classified as an integration host factor protein (IHF); however, it does not share sequence or structural homology with the well characterized *Escherichia coli* IHF. In this study, we characterize the structure and function of *Streptomyces coelicolor* IHF (sIHF). We have used a combination of biochemistry and structural biology to characterize the role of sIHF in DNA binding and DNA topology. We have solved crystal structures of sIHF bound to various double-stranded DNA substrates, and show that sIHF is able to contact DNA at multiple surfaces. Furthermore, sIHF inhibits the activity of TopA, impacting DNA topology *in vitro*. Our work demonstrates that sIHF is a novel nucleoid-associated protein with key roles in condensing DNA. We believe that sIHF performs its function by differentially using multiple nucleic-acid binding surfaces. Further characterization is required to confirm this hypothesis *in vivo*. Given that the *Mycobacterium* homolog of sIHF (mIHF) is essential, our studies lay the foundation to explore novel drug targets for *Mycobacterium tuberculosis* and *Mycobacterium leprae*.

Acknowledgements

I would like to thank my supervisor, Dr. Alba Guarné, for her guidance and support. She has gone above and beyond the role of a supervisor to ensure that my journey through graduate school was successful. I would also like to thank her for all the hard work she has put into this project and for teaching me techniques and skills that I will carry with me for the rest of my scientific career. I would like to thank my committee members; Dr. Marie Elliot for fostering my learning of *Streptomyces* and for making me feel welcome at her lab meetings, and Dr. Yingfu Li for his valuable input and thought provoking discussions.

I would like to thank the past and present members of the Guarné and Elliot labs that have made my experience in the lab more engaging and enjoyable. I owe much of my success to senior members who have not only helped me with laboratory techniques, but have encouraged my growth through stimulating scientific conversations. To Dr. Marie Elliot, Dr. Julia Swiercz, Melanie Gloyd, Sabrina Lue Tam, and Dr. Emma Sherwood, thank you for all your help and contribution to the project.

I would also like to thank my family and friends for their support. To my parents, Shamim and Moazali Nanji, thank you for taking care of me and instilling the values of hard work and discipline which have been instrumental in completing this degree. To my brother, Jamil Nanji, thank you for always making sure I have fun in my life. To Safir Kassam, thank you for your support and encouragement.

Table of Contents

Abstract.....	iii
Acknowledgements.....	iv
List of Figures.....	vii
List of Tables.....	viii
List of Abbreviations and Symbols.....	ix
Declaration of Academic Achievement.....	xi
Chapter 1-Introduction.....	1
1.1 The bacterial nucleoid.....	1
1.2 Nucleoid-associated proteins in <i>E. coli</i>	3
1.2.1 IHF.....	4
1.2.2 HU.....	4
1.2.3 Fis.....	5
1.2.4 H-NS.....	6
1.2.5 MukB.....	8
1.2.6 Lrp.....	8
1.2.7 Dps.....	9
1.3 <i>Streptomyces coelicolor</i> as a model organism to study novel nucleoid-associated proteins in actinobacteria.....	10
1.4 Thesis objective.....	14
Chapter 2-Methods and Materials.....	15
2.1 Overexpression of sIHF.....	15
2.2 Purification of sIHF.....	16
2.3 Forming the sIHF-DNA complex.....	18
2.4 Crystallization of sIHF bound to DNA.....	19
2.5 Data collection and structure determination.....	20
2.6 Cloning of sIHF mutants.....	20
2.7 Dynamic light scattering (DLS).....	23
2.8 Electrophoretic mobility shift assays (EMSAs).....	24
2.9 Topoisomerase assays.....	24
2.10 Data collection and processing for small-angle X-ray scattering (SAXS).....	25
Chapter 3-Results.....	26
3.1 Purification of sIHF.....	26
3.2 Crystallization, data collection and structure of sIHF bound to a 19-bp duplex DNA substrate.....	28
3.3 Crystallization, data collection and structure of sIHF bound to an 8-bp DNA substrate.....	34
3.4 sIHF contacts DNA at three interfaces in the crystal structure of sIHF bound to a hairpin DNA substrate.....	39
3.5 Three interfaces on sIHF contribute to DNA binding <i>in vitro</i>	43
3.6 sIHF does not bridge DNA.....	47

3.7 sIHF affects TopA activity independently of DNA binding	50
Chapter 4-Discussion	54
Chapter 5-Conclusion	62
References	63

List of Figures

Figure 1.1 Factors contributing to forming a condensed bacterial nucleoid.....	2
Figure 1.2 NAPs that bend DNA.....	6
Figure 1.3 <i>Streptomyces</i> life cycle.....	12
Figure 3.1 Purification scheme of untagged sIHF.....	27
Figure 3.2 Crystals of sIHF bound to a 19-bp duplex DNA substrate.....	29
Figure 3.3 Structure of sIHF	32
Figure 3.4 Structure of sIHF bound to TN08-BP1.....	38
Figure 3.5 Structure of sIHF bound to TN23-HP9.....	42
Figure 3.6 EMSAs of sIHF mutants with a radiolabeled 23-bp duplex DNA substrate.....	45
Figure 3.7 Electrostatics of sIHF bound to DNA.....	47
Figure 3.8 sIHF affects the activity of TopA.....	51
Figure 3.9 Gel filtration assay of sIHF with TopA.....	52
Figure 3.10 Topoisomerase assays of sIHF mutants	53
Figure 4.1 Ribbon structure of sIHF bound to TN08-BP1 and possible interactions with the N-terminal helix.....	56
Figure 4.2 Proposed mechanism of action of sIHF.....	59

List of Tables

Table 1.1 Well characterized nucleoid-associated proteins.....	10
Table 2.1 Expression plasmids.....	16
Table 2.2 DNA substrates used for crystallization.....	18
Table 2.3 Primers used to generate SIHF mutants.....	22
Table 2.4 Cloning plasmids.....	23
Table 3.1 X-ray data collection and refinement.....	31
Table 3.2 Base-pair parameters of duplex DNA alone and in complex with SIHF	36
Table 3.3 Small-angle X-ray scattering data collection and analysis.....	49

List of Abbreviations and Symbols

α	alpha
Å	angstrom
aIHF	actinobacterial integration host factor protein
BNL	Brookhaven National Laboratory
bp	base pair
BSA	bovine serum albumin
C	Celsius
COOT	Crystallographic Object-Oriented Toolkit
°	degree
DPS	DNA protection during starvation protein
DLS	dynamic light scattering
DTT	dithiothreitol
DNA	deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
Fis	factor for inversion stimulation
γ	Gamma
GE	General Electric
H2TH	helix-two turns-helix
HU	histone-like protein from <i>E. coli</i> strain U93
IDT	Integrated DNA Technologies
IHF	integration host factor
IPTG	isopropyl-beta-D-thiogalactopyranoside
LB	Luria-Bertani
Lrp	leucine responsive regulatory protein
mIHF	mycobacterial integration host factor
μg	microgram
μL	microliter
μM	micromolar
mL	milliliter
mM	millimolar
MOBIX	McMaster Institute for Molecular Biology and Biotechnology
MR	molecular replacement
MWCO	molecular weight cut-off
NMR	nuclear magnetic resonance
NAP	nucleoid-associated protein
NSLS	National Synchrotron Light Source
OD ₆₀₀	optical density measured at 600 nm
PAGE	polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PCR	polymerase chain reaction

PEG	polyethylene glycol
PHENIX	Python-based Hierarchical Environment for Integrated Xtallography
PMSF	phenylmethanesulfonyl fluoride
RMSD	root-mean-squared deviation
SAD	single-wavelength anomalous dispersion
SAXS	small angle X-ray scattering
<i>S. coelicolor</i>	<i>Streptomyces coelicolor</i>
SDS	sodium dodecyl sulphate
SELEX	systematic evolution of ligands by exponential enrichment
Sel-Met	selenomethionine
sIHF	<i>Streptomyces</i> integration host factor
SMC	structural maintenance of chromosome complex
SOC	super optimal broth with catabolite repression
TAE	Tris base, acetic acid and EDTA
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
WT	wild type

Declaration of Academic Achievement

I devised the protocol for large-scale siHF over-expression and purification. I conducted initial electrophoretic mobility shift assays and determined the optimal DNA substrates for crystallization of the siHF-DNA complexes. I grew all crystals that resulted in the structures presented in this work. Data was processed and refined by Dr. Alba Guarné and myself. I prepared the samples for small-angle X-ray scattering studies and Dr. Alba Guarné conducted the analysis. All siHF mutants were designed by Dr. Alba Guarné and myself; aside from siHF Δ 1-36 and siHF 35-37Gly, which were made with the help of Sabrina Lue Tam, I constructed the siHF variants. Electrophoretic mobility shift assays and topoisomerase activity assays were conducted by Dr. Emma Sherwood and Melanie Gloyd respectively using siHF protein that I purified. Gel filtration assays of siHF with TopA were performed by Melanie Gloyd.

Chapter 1

Introduction

1.1 The bacterial nucleoid

DNA stores essential information required for replication, and for the synthesis of RNA and protein molecules. It is contained within the bacterial genome, which unconstrained exceeds the size of the cell by 1 000- 10 000 times (Gitai et al., 2005). Therefore, chromosome compaction is essential. Even more impressive is that the condensed chromosome remains dynamic and accessible for cellular processes, such as DNA replication, transcription, and segregation, throughout the cell cycle (Thanbichler et al., 2005a; Thanbichler et al., 2005b). Bacteria lack features of eukaryotic cells that aid in chromosome compaction, such as a membrane bound nucleus and histone proteins (Zhu and Wani, 2010). Instead, bacteria compact their DNA into a condensed, yet dynamic, structure called the bacterial nucleoid. Nucleoid organization is mediated through molecular crowding and using multiple proteins. These proteins include topoisomerases that affect the superhelicity of the bacterial chromosome, RNA polymerases and effector proteins that mediate their activity, as well as abundant nucleoid associated proteins (NAPs) (Travers and Muskhelishvili, 2005; Lujsterburg et al., 2008) (Figure 1.1).

In crowded environments, characteristic of the bacterial cell, entropic forces impact the interaction between macromolecules to a greater extent than in diluted environments (Lujsterburg et al., 2008). This causes larger macromolecules to come together allowing smaller molecules to move more freely. This phenomenon is known as attraction and depletion forces

(Marenduzzo et al., 2006; Hancock, 2004) and causes the nucleoid to self-condense (Figure 1.1; blue stars represent crowding from the environment).

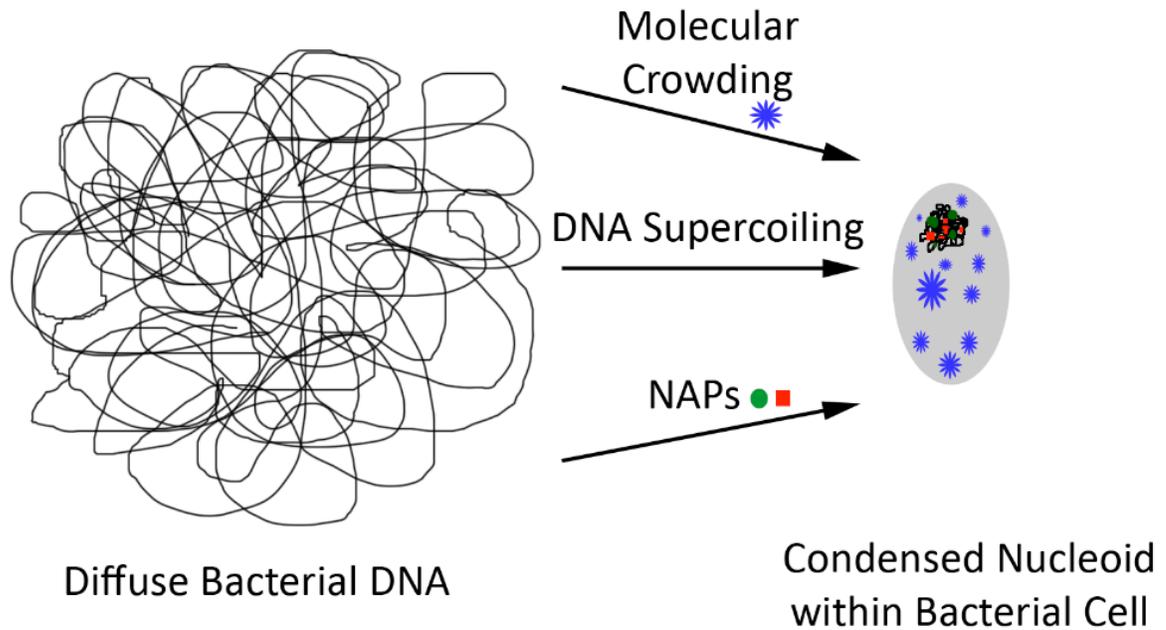


Figure 1.1. Factors contributing to forming a condensed bacterial nucleoid. Unconstrained, bacterial DNA is diffuse, unstructured, and exceeds the size of the bacterial cell. Black lines indicate duplex DNA. Molecular crowding, indicated by the blue stars, acts as a force which condenses bacterial DNA. DNA supercoiling allows the bacterial chromosome to intertwine in an orderly manner to condense the nucleoid, but also allows for DNA dynamics. Nucleoid-associated proteins (NAPs), depicted as green circles and red squares, are a diverse group of bacterial proteins that bind DNA and aid in DNA compaction.

DNA supercoiling allows DNA to pack closer together to form a condensed structure. Supercoils are introduced and removed by topoisomerase enzymes creating a dynamic nucleoid (Champoux, 2001). Once regions of the nucleoid are relaxed, transcription factors and polymerase enzymes can access parts of the genome that were occluded. This modifies levels of gene expression which is necessary for cell cycle progression (Myers et al., 2013; Azam et al., 1999). The topological organization of the bacterial chromosome is not random. It has been shown to be organized into small topological domains (~10 kb) where diffusion of supercoils is

restricted (Postow et al., 2004; Sinden and Pettijohn, 1981). The torsional tension of supercoiled DNA within these domains drives cellular processes such as transcription (Lim et al., 2003), replication (Funnell et al., 1986), and recombination (Nash, 1990). Furthermore, the organization of DNA into smaller domains ensures that a break in one region of the chromosome only affects that topological unit, and does not cause changes to the overall superhelicity of the chromosome which can lead to cell death (Wang, 1996).

Eukaryotes mediate DNA organization using histone proteins that wrap DNA in an orderly fashion, to tightly package the chromosome. Histones can unravel DNA to alter the accessibility of various regions of the genome (Luger, 2006). Bacteria use an arsenal of nucleoid-associated proteins (NAPs) (Figure 1.1; depicted by red squares and green circles) to compact the genome; however, they are not structurally similar to histone proteins. We will look into these proteins in greater detail in the next sections.

1.2 Nucleoid-associated proteins in *E. coli*

NAPs are an abundant, diverse group of proteins that are expressed at various levels throughout the cell cycle (Azam et al., 1999). Most NAPs bind DNA promiscuously to play a significant part in compacting the bacterial nucleoid. Many of these proteins bind DNA with little sequence specificity, but prefer binding DNA sequences that are AT-rich. Promoter sequences are AT-rich; therefore, NAPs have also been characterized as transcription factors as they modulate gene expression (Myers et al., 2013; Dillon and Dorman, 2010; Dorman, 2013). NAPs have numerous roles in the bacterial cell and have pleiotropic effects. They are diverse in structure, DNA binding ability, and their effects on DNA upon binding. They are generally small, basic proteins and have

been shown to compact the bacterial chromosome through bridging, bending, and/or wrapping DNA (Dillon and Dorman, 2010; Rimsky and Travers, 2011). Although many NAPs have been well characterized in *E. coli*, there is less known in other classes of bacteria. We will first explore well characterized NAPs in *E. coli*.

1.2.1 IHF

The *E. coli* integration host factor protein (IHF) was first identified for its role in recombination of bacteriophage lambda (Miller and Friedman, 1980) and was later classified as a NAP (Rice et al., 1996). IHF binds DNA and induces sharp bends which alters the trajectory of the DNA duplex, contributing to a more condensed nucleoid. IHF functions as a heterodimer and is composed of two subunits; an alpha- and beta-subunit. The alpha-subunit is 11 kDa and the beta-subunit is 9.5 kDa (Rice et al., 1996, Luijsterburg et al., 2006). Both subunits are structurally homologous; they consist of a body composed of alpha-helices flanked by flexible beta-hairpins (Figure 1.2A). A conserved proline residue within each beta-hairpin intercalates into the minor groove of DNA to induce a sharp bend (Swinger et al., 2003; Swinger and Rice, 2004). IHF binds double-stranded DNA with a footprint of ~30 base-pairs at a consensus sequence rich in adenines and thymine nucleobases (Goodrich et al., 1990) (Table 1.1). It is maximally expressed during early stationary phase (Azam et al., 1999).

1.2.2 HU

In *E. coli* and other enterobacteriaceae, the histone-like protein first identified in *E. coli* strain U93 (HU) (Oberto et al., 1994) has a role in chromosome condensation. Similar to IHF, it induces bends in the DNA (Figure 1.2B). It also functions as a heterodimer and is composed of an alpha-

and beta-subunit (Dame and Goosen, 2002) that are 9.5 kDa each and share 70% homology (Table 1.1) (Swinger et al., 2003). HU binds DNA with a footprint of ~9 base-pairs and prefers sequences rich in adenine and thymine nucleobases which are inherently distorted. Furthermore, HU prefers binding to curved or distorted regions of DNA (Swinger et al., 2003). HU is structurally homologous to IHF; it consists of a body composed of alpha helices capped by two beta-hairpin arms that intercalate into the minor groove of DNA using a conserved proline residue (Figure 1.2B). The bend induced by HU is less pronounced compared to that of IHF, and HU's minimal DNA binding site (~9 bp) is smaller than that of IHF (~30 bp) illustrating the diversity that exists within NAPs, even when they are structurally homologous (Swinger and Rice, 2004). Moreover, HU induces negative supercoiling directly by bending DNA (Rouvière-Yaniv et al., 1979), and indirectly through stimulating the activity of DNA gyrase (Swinger and Rice, 2004). HU is maximally expressed during logarithmic phase to alter DNA dynamics required during development in this stage of the cell cycle (Azam et al., 1999).

1.2.3 Fis

The factor for inversion stimulation (Fis) protein is a NAP that condenses the chromosome by bending DNA, as well as by aiding in loop formation (Luijsterburg et al., 2006; Schneider et al., 1999). It functions as a 22 kDa homodimer and is composed of four alpha-helices that are connected by beta-turns and two beta-hairpins (Table 1.1). It binds DNA through a common DNA binding fold; a helix-turn-helix motif (Figure 1.2C). Fis prefers binding DNA at consensus nucleotide sequence, but also binds DNA at other sequences with high affinity (Table 1.1). It is the most abundant NAP during early exponential growth; however, it is completely absent during stationary phase (Luijsterburg et al., 2006). Fis has been suggested to mediate DNA

supercoiling by inhibiting the expression of DNA gyrase in a manner dependant on DNA topology (Schneider et al., 1999; Schneider et al., 2001). Fis is a multifaceted NAP that affects chromosome organization via many avenues. This shows that NAPs are complex and can affect multiple cellular processes.

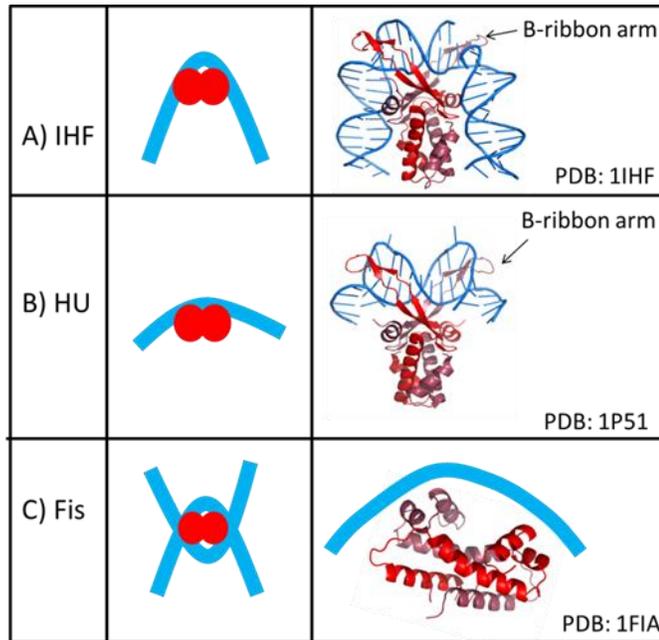


Figure 1.2. NAPs that bend DNA. **(A)** IHF induces sharp bends into duplex DNA by intercalating into the minor groove of DNA using conserved proline residues within the β -hairpins. The protein is indicated in red while duplex DNA is in blue in all panels. **(B)** HU induces bends by intercalating into the minor groove of DNA using conserved proline residues within the β -hairpin arms. **(C)** Fis compacts the bacterial chromosome by bending duplex substrates. The crystal structure was solved without DNA; duplex DNA is modeled in with a blue line illustrating where the protein is thought to contact DNA.

1.2.4 H-NS

The histone-like nucleoid structuring protein (H-NS) plays a key role in compacting and organizing the bacterial nucleoid, and has a role in controlling gene expression. H-NS is a 15.4 kDa protein that functions as a dimer (Table 1.1) (Luijsterburg et al., 2006) and is maximally

expressed during the exponential phase of the cell cycle (Azam et al., 1999). Unlike HU and IHF, H-NS bridges adjacent DNA molecules (Dame et al., 2005; Dame et al., 2000). H-NS binds DNA through its C-terminal DNA binding domain, and forms a homodimer with another H-NS molecule bound to DNA, through its N-terminal dimerization domain. In this way, H-NS is able to bridge distant DNA molecules together to aid in chromosome condensation. The dimerization domain consists of a long alpha-helix and two smaller alpha-helices (Bloch et al., 2003), whereas the DNA binding domain consists of one alpha-helix and two anti-parallel beta-strands (Shindo et al., 1995). H-NS is suggested to bind DNA at the major groove through a positively charged face formed by residues Arg80 to Lys96 on the beta-strand closest to the N-terminus, and residues Thr110 to Ala117 located between the other beta-strand and alpha-helix (Shindo et al., 1995; Shindo et al., 1999).

H-NS does not bind DNA at a specific sequence; however, it prefers AT-rich sequences as they are intrinsically curved (Table 1.1) (Dame et al., 2001). Furthermore, H-NS affects the expression of many genes, mainly through negatively affecting transcription (Atlung and Ingmer, 1997). This is potentially due to its preference for binding AT-rich DNA sequences. AT-rich DNA sequences are normally found at promoters (Newton-Foot and Gey van Pittius, 2013); hence, H-NS represses transcription of a wide variety of genes. In addition, overexpression of H-NS results in more condensed nucleoids compared to wildtype (Spuro et al., 1992), illustrating its function in DNA compaction.

1.2.5 MukB

MukB in *E. coli* and its structural homologue, the structural maintenance of chromosome complex (SMC) in *Bacillus subtilis*, belong to a large class of proteins that have been characterized for their roles in chromosome condensation and chromosome segregation. This group of proteins is conserved from bacteria to humans (Losada and Hirano, 2005). They are larger than other NAPs at 150-200 kDa (Table 1.1) (Luijsterburg et al., 2006). They function as a V-shaped homodimer to gather distant DNA molecules and bring them together; this function is dependent on their ATPase activity (Chen et al., 2008). It is suggested that these proteins form larger oligomers mediated by auxiliary cofactor proteins (MukE and MukF in *E. coli* and ScpA and ScpB in *Bacillus subtilis*) to gather DNA and separate the genetic information into the two poles of replicating cells (Gloyd et al., 2011; Kleine-Borgmann et al., 2013). Hence, these large complexes aid in chromosome condensation, DNA organization and are thought to help separate newly replicated chromosomes (Kleine-Borgmann et al., 2013).

1.2.6 Lrp

The leucine-responsive regulatory protein (Lrp) aids in nucleoid compaction by bridging and wrapping distant DNA molecules. Lrp is a 15 kDa protein that functions as a dimer to bridge DNA, but can also form an octamer to effectively wrap DNA molecules. The protein is composed of an N-terminal DNA binding domain and a C-terminal dimerization domain. The DNA binding domain consists of 3 alpha-helices that adopt a helix-turn-helix motif, which is a common DNA binding fold (Cui et al., 1995). Lrp binds DNA at a consensus sequence but can also bind at multiple suboptimal sequences to modulate gene expression and chromosome organization

(Table 1.1). Lrp is suggested to affect the expression of about 10% of all genes (Luijsterburg et al., 2006; Cui et al., 1995). This NAP functions via multiple mechanisms, illustrating the variability that exists within this group of proteins.

1.2.7 Dps

The DNA protection during starvation protein (Dps) is an important NAP as it protects cells when nutrients are limited and aids in the transition from growth to stationary phase. Unlike most NAPs that affect chromosome structure at the local level, Dps acts globally and affects most of the nucleoid. This 19 kDa protein (Table 1.1) binds DNA non-specifically (based on DNase I footprinting). The mechanism of DNA binding is not well known, although DNA becomes DNase resistant upon binding (Almirón et al., 1992). Electron microscopy has shown that Dps alone forms 6 membered rings, but in complex with DNA it forms a honeycomb like sheet with interconnected rings (Almirón et al., 1992). It also binds DNA to form sheets of condensed dodecamers that significantly compact the nucleoid (Frenkiel-Krispin et al., 2004). This suggests that Dps forms a hexameric structure that is multi-layered. It is thought that Dps interacts with DNA through a positively charged surface composed of three lysine residues (Grant et al., 1998). As Dps is expressed maximally during stationary phase (Azam et al., 1999), it has a significant role in protecting, organizing, and compacting DNA consistent with stationary phase requirements. Furthermore, it alters the expression of many genes which are required during the transition between exponential to stationary phase which manifests in physiological and morphological changes (Almirón et al., 1992). Based on the X-ray crystal structure of Dps, the protein consists of a fold similar to that of ferritin. As such, it is hypothesized that Dps protects cells by sequestering iron ions (Grant et al., 1998).

Table 1.1. Well characterized nucleoid-associated proteins

NAP	Oligomerization state	Molecular mass	Bridge/Bend/ Wrap	DNA binding preference
IHF	Heterodimer	alpha: 11 kDa	Bend	(A/T)ATCAANNNTT(A/G)*
		beta: 9.5 kDa		
HU	Heterodimer	alpha: 9.5 kDa	Bend	distorted DNA substrates
		beta: 9.5 kDa		
Fis	Homodimer	22 kDa	Bend	(G/T)NN(C/T)(A/G)NN(A/T)NN(C/T)(A/G)NN (C/A)*
H-NS	Homodimer	15.4 kDa	Bridge	AT rich sequences
Lsr2	Homodimer	12 kDa	Bridge	AT rich sequences
SMC	Homodimer	150-200 kDa	Alternate	Non-specific
Lrp	Octamer	15 kDa	Wrap/bridge	AGAATTTTATTCT
Dps	Hexamer/ Dodecamer	19 kDa	Alternate	Non-specific

*N is any nucleotide

1.3 *Streptomyces coelicolor* as a model organism to study novel nucleoid-associated proteins in actinobacteria

E. coli has a relatively small genome compared to *Streptomyces coelicolor*. *E. coli* contains one circular chromosome that is 4.6 Mb in size and has a GC content of approximately 50% (Postow et al., 2004); whereas *S. coelicolor* has a linear chromosome of 8.7 Mb (Bentley et al., 2002) and additional plasmids, including SCP1 and SCP2 which are 365 Kb and 31 Kb respectively (Bibb et al., 1981). *Streptomyces* belong to a group of Gram-positive bacteria within the phylum of actinobacteria. *Streptomyces* have a characteristic, complex life cycle (Figure 1.3); they begin as a spore that germinates and develop into vegetative mycelium that forms branches. They then develop aerial hyphae which coincide with the production of secondary metabolites (Figure 1.3, depicted as cyan dots) which include many antibiotic compounds such as actinorhodin and

undecylprodigiosin. These compounds are blue and red pigmented, respectively. *Streptomyces* may require a large genome and accessory plasmids to maintain their complex life cycle and mediate the production of antibiotics. The *Streptomyces* chromosomes remain uncondensed until the reproductive stages of their life cycle, at which point the chromosomes segregate into compartments and the pre-spore chain is formed. This spore-chain goes on to develop into a spore chain. Each spore in the spore chain can separate and the cycle can re-occur. To understand how DNA organization is coordinated with this complex life cycle, their NAPs can be studied.

As previously discussed, NAPs can bind DNA and act as transcription factors to cause changes in gene expression (Dillon and Dorman, 2010). Changes in gene expression can be easily studied in *S. coelicolor* as they produce many secondary metabolites. *S. coelicolor* are a good model organism to study NAPs in actinobacteria, as they allow us to monitor changes in the production of pigmented antibiotics. Furthermore, aerial hyphae have a distinct appearance which allows us to monitor changes in cell cycle progression. Moreover, in a laboratory setting, chromosome compaction and cell division are not essential allowing us to monitor chromosome condensation.

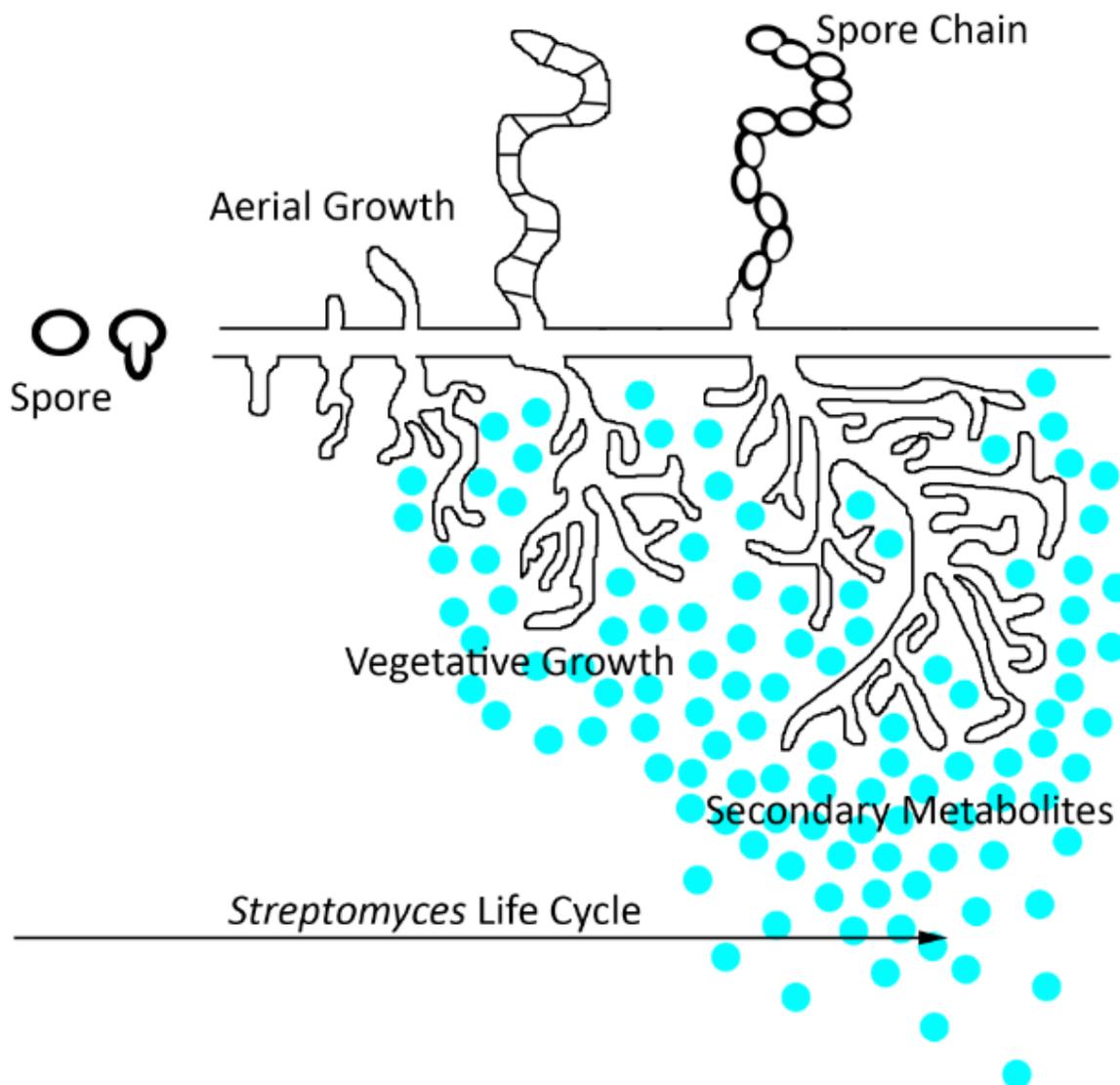


Figure 1.3. *Streptomyces* life cycle. *Streptomyces* germinate from a spore. They form a vegetative mycelium and then raise aerial hyphae at a time that correlates with the production of secondary metabolites, including pigmented antibiotics, which are depicted as cyan dots. As aerial hyphae grow the bacterial chromosome compacts and is segregated into prespore compartments, which go on to develop into spore chains. Each spore in the spore chain can separate and the cycle can continue. (Figure adapted from Swiercz and Elliot, 2012).

Although most NAPs are not structurally conserved, they have been suggested to be functionally conserved across the kingdoms of life (Luijsterburg et al., 2008). This is true for the actinobacterial Lsr2 protein; Lsr2 performs the same function as the *E. coli* H-NS protein but is

not structurally homologous (Chen et al., 2008). Like H-NS, Lsr2 prefers binding AT-rich DNA sequences and is able to bridge disparate DNA molecules together (Gordon et al., 2011; Chen et al., 2008, Qu et al., 2013). Not only do Lsr2 and H-NS have similar properties, Lsr2 has been shown to be functionally analogous to H-NS as per complementation assays and DNA binding assays. *lsr2* is able to complement *hns* null mutants and Lsr2 specifically binds to genes regulated by H-NS (Gordon et al., 2008). This illustrates that multiple NAPs may have synonymous roles across bacterial species even though they are not structurally similar.

A new actinobacterial specific NAP has recently been discovered; the actinobacterial integration host factor (aIHF) protein (Yang et al., 2012). These proteins are conserved among actinobacteria, but share low amino acid similarity to the *E. coli* IHF protein. Synteny suggests that these proteins are homologous as they are found within the same gene cluster (Yang et al., 2012). The aIHF family of proteins are classified as IHF proteins since they were first identified as having a role in phage integration; the protein is required to form the recombinogenic intasome complex. Furthermore, the aIHF in *Mycobacterium* (mIHF) does not bind specifically to the sequence specifying the site of integration (Goosen and van de Putte, 1995), and Δ *mIHF* strains cannot be complemented with *E. coli* IHF or HU to restore recombination (Pedulla and Hatful, 1998). The aIHF proteins are of particular importance as they are essential in many actinobacterial species, including *Mycobacterium tuberculosis* and *M. leprae* (Pedulla and Hatful, 1998; Sassetti et al., 2003), the bacterial infections resulting in tuberculosis and leprosy, making them potential drug targets. Furthermore, as multidrug resistant bacterial strains are becoming more prevalent, novel drugs and novel drug targets are important research topics.

As the mIHF protein is essential in *Mycobacterium*, we decided to study the aIHF protein in the model organism *Streptomyces coelicolor*. *Streptomyces* is a good model organism to study this protein as it shares a genetic core with *Mycobacterium*, which includes many of the same housekeeping and essential genes (Cole et al., 1998; Bentley et al., 2002). The aIHF protein is also well conserved and shares 65% identity and 95% similarity between these two organisms. It is advantageous to study the aIHF protein in *Streptomyces* (sIHF) as we have been successful in generating a viable Δ sIHF strain (Swiercz et al., 2013). The viability of this strain is thought to be due to the complex life cycle of *Streptomyces* described above, whereas *Mycobacterium* divides by binary fission. Using the *S. coelicolor* Δ sIHF strain, we have shown that sIHF associates with the nucleoid, plays a significant role in chromosome compaction, and affects gene expression; *S. coelicolor* lacking sIHF displayed decondensed nucleoids, elongated cells and aberrant production of pigmented antibiotics (Swiercz et al., 2013).

1.4 Thesis objective

My thesis objective was to understand how sIHF interacts with DNA and modulates the function of other proteins important in DNA topology. To this end, I completed three specific aims during my MSc:

- 1) Solve the crystal structure of sIHF bound to duplex DNA,
- 2) Determine the regions of sIHF that are important for DNA binding,
- 3) Explore how sIHF affects the activity of the topoisomerase TopA.

Chapter 2

Methods and Materials

2.1 Overexpression of sIHF

S. coelicolor *sIHF* was cloned into the pET-15b vector (Novagen), which contains a removable N-terminal poly-histidine tag, between restriction sites NdeI and BamHI (pAG 8380, see Table 2.1). This plasmid was incubated with calcium chloride competent *E. coli* BL21(DE3) Rosetta cells (Invitrogen Life Technologies) at 4°C for 1 hour prior to heat shock at 42°C for 45 seconds and subsequent incubation at 4°C for 5 minutes. Super Optimal broth with Catabolite repression (SOC) medium was added to the cells, incubated for 1 hour at 37 °C and plated onto Luria-Bertani (LB) agar plates with 100 µg/mL ampicillin and 25 µg/mL chloramphenicol. Mixed colonies were picked and grown in LB medium at 37°C until an OD₆₀₀ of ~0.7. Protein expression was induced with 1.0 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) and allowed to proceed for 3 hours at 37°C. Cells were harvested by centrifugation for 15 minutes at 3,315 x g. Cell pellets were washed with 1x phosphate buffered saline (BioShop) and spun at 4°C for 10 minutes at 2,930 x g, and stored at -80°C for later use.

Selenomethionine labeled sIHF was produced by transforming pAG 8380 into calcium chloride competent *E. coli* B843 Rosetta cells and grown at 37°C in minimal medium (Sigma) supplemented with 2.0×10^{-4} mM selenomethionine (Sigma), 100 µg/mL ampicillin and 25 µg/mL chloramphenicol to an OD₆₀₀ of ~ 1.0. Cultures were induced, harvested and stored as described above. All sIHF mutants were overexpressed in the same manner as wildtype, using their respective plasmid listed in Table 2.1.

Table 2.1. Expression plasmids

Name	Construct	Vector	Cloning restriction sites	Source
pAG 8380	sIHF full length	pET-15b	NdeI, BamHI	Gift from Elliot M. A
pAG 8775	sIHF RR85AS	pET-15b	NdeI, NheI*, BamHI	This work
pAG 8779	sIHF Δ1-13	pET-15b	NdeI, BamHI	This work
pAG 8780	sIHF NQ93AS	pET-15b	NdeI, NheI*, BamHI	This work
pAG 8845	sIHF S(19)	pET-15b	NdeI, BamHI	This work
pAG 8846	sIHF RR85AS+NQ93AS	pET-15b	NdeI, NheI*, BamHI	This work
pAG 8851	sIHF Δ1-36	pET-15b	NdeI, BamHI	This work
pAG 8860	sIHF S(19)+RR	pET-15b	NdeI, BamHI	This work
pAG 8866	sIHF G66+	pET-15b	NdeI, BamHI	This work
pAG 8867	sIHF 35-37 Gly	pET-15b	NdeI, BamHI	This work

**This enzyme was included to screen for the presence of the point mutation.*

2.2 Purification of sIHF

E. coli BL21(DE3) Rosetta cells overexpressing His-sIHF, were resuspended in 20 mL of buffer A (20 mM Tris pH 8.0, 300 mM NaCl, 1.4 mM beta-mercaptoethanol, 5% (v/v) glycerol) and lysed at 4°C by sonication. Lysis was complete after two 60 second pulses. Protease inhibitors (5 µg/ml leupeptin, 1 mM phenylmethanesulfonyl fluoride (PMSF), 0.7 µg/ml pepstatin A, and 1 mM benzamidine) were added before and immediately after cell lysis to reduce proteolysis and maintain protein integrity. Cellular debris was removed by centrifugation at 39,191 x g for 40 minutes at 4°C.

The supernatant containing the His-sIHF protein was loaded onto a 5 mL HiTrap Chelating HP column (GE Healthcare) using immobilized nickel. The column was washed with 20 column volumes of buffer A to remove unbound protein, 10 column volumes of buffer A supplemented with 7.5 mM imidazole, and 5 column volumes of buffer A with 20 mM imidazole to remove proteins that interact with the column unspecifically. sIHF was eluted with buffer A supplemented with 150 mM imidazole. Fractions containing sIHF were pooled and diluted using

buffer B (20 mM Tris pH 8.0, 1.4 mM beta-mercaptoethanol, 5% (v/v) glycerol) to dilute the imidazole and reduce the salt concentration to 100 mM. The diluted protein was loaded onto a Mono S 10/100 GL column (GE Healthcare). His-sIHF was eluted using a linear gradient from 100 mM to 1000 mM NaCl. His-sIHF eluted off the cation exchange column at 420 mM NaCl. Pooled fractions of the protein, typically at ~ 170 μ M, were diluted in buffer C (20 mM Tris pH 8, 150 mM NaCl, 5 mM CaCl_2 , 5 mM DTT, 5% (v/v) glycerol) to 40 μ M with a final concentration of 5 mM CaCl_2 to ensure optimal digestion of the histidine tag using the thrombin protease.

A small scale thrombin (Sigma) digestion reaction (10 μ L) was conducted from 0.01-0.125 units/ μ L to determine the concentration of enzyme required to fully digest the poly-histidine tag. Nine μ L of diluted sIHF at 40 μ M was incubated with 1 μ L of each thrombin concentration for 1 hour at room temperature. Digestion products were resolved on an 18% sodium dodecyl sulphate-polyacrylamide gel (SDS-PAG) stained with coomassie containing 20% (v/v) acetic acid. The remainder of His-sIHF at 40 μ M was incubated with the optimal concentration of thrombin for 1 hour at room temperature. The reaction was quenched with 1 mM benzamidine and tagless sIHF was purified using a linear gradient from 100 mM to 1000 mM NaCl on a Mono S 5/50 GL column (GE Healthcare). Purified sIHF eluted off the cation exchange column at 420 mM NaCl. Pooled fractions were concentrated in a Vivaspin 2 5,000 molecular weight cut-off (MWCO) centrifugal concentrator (GE Healthcare) and the buffer was exchanged to buffer D (40 mM Tris pH 8.0, 300 mM NaCl, 20 mM MgCl_2 , 2.8 mM beta-mercaptoethanol, 10% (v/v) glycerol). Protein concentrations were determined using the Bradford assay (Bradford, 1976).

2.3 Forming the sIHF-DNA complex

Oligonucleotides were purchased from IDT and suspended in filtered autoclaved water. Complementary oligonucleotides were mixed at equal concentrations and annealed by boiling for 3 minutes prior to cooling overnight. Oligonucleotides containing longer than an 8-bp duplex were cooled to room temperature. Substrates containing 7 or 8 pairing bases were cooled to 4°C, as their melting temperature was approximately ~24°C. See Table 2.2 for oligonucleotide sequences. Equal volumes of sIHF (diluted in buffer D) and dsDNA (diluted in autoclaved water) were mixed. sIHF-DNA complexes were formed in 20 mM Tris pH 8.0, 150 mM NaCl, 10 mM MgCl₂, 1.4 mM beta-mercaptoethanol, and 5% (v/v) glycerol. Protein-DNA complexes of duplex substrates larger than 8-bp were incubated for 10 minutes at room temperature, followed by 30 minutes on ice, while protein-DNA complexes of duplex substrates that are 8-bp were incubated on ice at 4°C overnight.

Table 2.2. DNA substrates used for crystallization

TN22-Top	5' GGG AGT GCG TGG GTC TGA AGC C ^{3'}
TN22-Bottom	5' GGC TTC AGA CCC ACG CAC TCC C ^{3'}
TN200-Top	5' GAG GGA GTG CGT GGG TCT GA ^{3'}
TN200-Bottom	5' CTC AGA CCC ACG CAC TCC CT ^{3'}
TN19O2-Top	5' GAG GGA GTG CGT GGG TCT G ^{3'}
TN19O2-Bottom	5' TCC AGA CCC ACG CAC TCC C ^{3'}
TN15-Top	5' GGG AGT GCG TGG GTC ^{3'}
TN15-Bottom	5' GAC CCA CGC ACT CCC ^{3'}
TN08-BP1	5' CAT GCA TG ^{3'}
TN08O-Top	5' GGG CGC GG ^{3'}
TN08O-Bottom	5' CCC GCG CC ^{3'}
TN23-HP9	5' GTGCGTGGATTTTTCCACGCAC ^{3'}

2.4 Crystallization of sIHF bound to DNA

Crystallization trials of sIHF bound to duplex DNA substrates (Table 2.2) were conducted using sparse matrix screens set by the Phoenix Liquid Handling System (Art Robbins Instruments). Initial crystal hits of sIHF bound to TN200 (Table 2.2) were obtained using the sitting drop method in condition 14 (0.2 M KSCN, 20% (w/v) PEG 3350, pH 7) of the PEG/Ion Screen (Hampton Research). Crystals were optimized by the hanging drop method with the addition of HEPES at pH 7.6 and ethylene glycol. Optimal crystals at 1.0 mM complex grew at a 1:1 ratio in 0.1 M HEPES pH 7.6, 0.21 M KSCN, 19% (w/v) PEG 3350, and 5% (v/v) ethylene glycol. Once crystals reached maximum size, drops were dehydrated against increasing concentrations of KCl (1.0, 1.25, 1.5, and 1.75 M KCl) for 8-12 hours prior to flash freezing in liquid nitrogen.

Crystals of Sel-Met labeled sIHF bound to TN200 were obtained at a 1.4:1 ratio (1.05 mM:0.75 mM) in 0.1 M HEPES pH 7.6, 0.21 M KSCN, 18% (w/v) PEG 3350, and 5% (v/v) ethylene glycol. Crystals of sIHF bound to TN08O (Table 2.2), at 1.0 mM at a 1:1 ratio, grew in 0.1 M HEPES pH 7.6, 0.1 M KSCN, 28% (w/v) PEG 3350, and 5% (v/v) ethylene glycol. Crystals of sIHF bound to TN08-BP1 (Table 2.2) at 1.5 mM at a 1:1 ratio were obtained in 0.1 M HEPES pH 7.6, 0.2 M KSCN, 19% (w/v) PEG 3350, and 5% (v/v) ethylene glycol. Crystals of sIHF bound to TN23-HP9 (Table 2.2) were grown using the streak-seeding method. Crushed crystals obtained at a 1:1.2 ratio (1.0 mM:1.2 mM) in 0.1 M MES pH 5.6, 0.12 M MgCl₂, 16% (w/v) PEG 3350, and 5% (v/v) ethylene glycol were seeded into crystallization drops of sIHF bound to TN23-HP9 at a ratio of 1:1.2 (0.63 mM:0.75 mM) in 0.1 M MES pH 5.6, 0.1 M MgCl₂, 14% (w/v) PEG 3350, and 5% (v/v) ethylene glycol after 1 day using a cat whisker. Crystals were dehydrated and frozen as above.

2.5 Data collection and structure determination

Data were collected using beamline X25 at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL) (Upton, NY). Crystal TN055 (sIHF bound to TN200) diffracted to 2.85 Å, crystal TN051 (sIHF Sel-Met bound to TN200) diffracted to 2.6 Å, crystal TN100 (sIHF bound to TN080) diffracted to 3.00 Å, crystal TN103 (sIHF bound to TN08-BP1) diffracted to 1.66 Å, and crystal TN131 (sIHF bound to TN23-HP9) diffracted to 2.9 Å. Data were indexed, processed, and merged using HKL2000 (Otwinowski and Minor, 1997). Refer to Table 3.1 for full data collection and refinement statistics.

Crystals of Sel-Met sIHF:TN200 were phased by single-wavelength anomalous dispersion (SAD) using SOLVE (Terwilliger, T. C. and Berendzen, J., 1999). The initial model was manually built using COOT and refined using iterative cycles of model building in COOT and refinement in phenix.refine (Afonine et al., 2012; Emsley and Cowtan, 2004). Crystals of sIHF bound to TN080, TN08-BP1 and TN23-HP9 were phased by molecular replacement (MR) using Phaser-MR in Phenix and subsequent model building and refinement was done using standard protocols in phenix.refine and COOT (Afonine et al., 2012; Emsley and Cowtan, 2004).

2.6 Cloning of sIHF mutants

sIHF mutants were produced by either site-directed mutagenesis or overlap polymerase chain reaction (PCR). sIHF RR85AS (pAG 8775) (Table 2.1) was generated by site-directed mutagenesis using the QuikChange II kit (Agilent Technologies) from template pAG 8380 (Table 2.1) with primers, purchased from IDT, AG 1794, and AG 1795 (Table 2.3). These primers were designed to have a NheI restriction site to discriminate between template and product DNA. The site-

directed mutagenesis reaction was incubated with DpnI, for 1 hour in a 37°C water bath to digest template DNA, and transformed into Top10 electrocompetent cells (Life Technologies) by electroporation. Plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit (Thermo Scientific) and the presence of the desired mutation was assessed by analytical restriction digestion with NheI. Presence of the mutation was confirmed by DNA sequencing (MOBIX Lab).

sIHF NQ93AS (pAG 8780), was also generated by site-directed mutagenesis using pAG 8380 (Table 2.1) as the template and primers AG 1825 and AG 1826 (Table 2.3) as described above. Lastly, we combined these mutations to make a quadruple sIHF mutant, sIHF RR85AS+NQ93AS (pAG 8846), using site-directed mutagenesis from pAG 8775 as a template and primers AG 1825, and AG 1826. The reaction was digested by DpnI and transformed, isolated, and sequenced as described above.

N-terminal truncations lacking the first 13 and 36 residues were generated by PCR using primers AG 1796 and AG 1797 as forward primers respectively and AG 1759 as the reverse primer (Table 2.3). Both forward primers contained an NdeI restriction site to be incorporated at the beginning of the gene, and the reverse primer contained a BamHI site, to be incorporated at the end of the gene for cloning purposes. Each PCR reaction product was ligated into the blunt cloning vector pJET1.2 (Thermo Scientific) to yield plasmids pAG 8771 and pAG 8777 (Table 2.4). These plasmids along with a plasmid containing the pET-15b vector (pAG 8160) (Table 2.4) were digested by NdeI for 15 minutes and BamHI for 5 minutes at 37°C, and run on an agarose gel. DNA fragments were extracted using the QIAEX II Gel Extraction Kit (Qiagen) and each sIHF truncation plasmid was ligated with the digested pET-15b vector overnight at 16°C with T4 DNA

Ligase (New England Biolabs) to yield plasmids pAG 8779 and pAG 8851 (Table 2.1). (pAG 8851 was generated with the help of Sabrina Lue Tam).

Table 2.3. Primers used to generate sIHF mutants

Name	Use	Sequence
AG 1759	35-37 Gly, Δ1-13, Δ1-36, G66+	5' AAG GAT CCT CAG CTG CCG GTG CTG CCG AAC TCG C ^{3'}
AG 1794	RR85AS	5' CTG GGC ATC TCC GAG AGC GCT AGC GTG CGC GGT CTC GGG TCC ^{3'}
AG 1795	RR85AS	5' GGA CCC GAG ACC GCG CAC GCT AGC GCT CTC GGA GAT GCC CAG ^{3'}
AG 1796	Δ1-13	5' AAC ATA TGG CGC TCG AAA AGG CCG CCG CGG CTC G ^{3'}
AG 1797	Δ1-36	5' AAC ATA TGG CCT CCC TCC ACG AGG TCA TCA AGC AGG GTC AG ^{3'}
AG 1825	NQ93AS	5' CGC GTG CGC GGT CTC GGG TCC GCT AGC ATC GCG TCC CTG GAG ^{3'}
AG 1826	NQ93AS	5' CTC CAG GGA CGC GAT GCT AGC GGA CCC GAG ACC GCG CAC GCG ^{3'}
AG 1908	S(19)+RR	5' GGC ATC TCC GAG AGC CGC CGG GTG TCC GGT CTC GGG TCC AAC ^{3'}
AG 1909	S(19)+RR	5' GTT GGA CCC GAG ACC GGA CAC CCG GCG GCT CTC GGA GAT GCC ^{3'}
AG 1910	35-37 Gly	5' CGA CTC AAG CAC GGT GGC GGG TCC CTC CAC GAG GTC ATC AAG CAG ^{3'}
AG 1911	35-37 Gly	5' CTC GTG GAG GGA CCC GCC ACC GTG CTT GAG TCG ATT CTT GAC CTC ^{3'}
AG 1912	G66+	5' CTC CGC CCT CCT GGA GTC CCT GCC GGG CGG TGT GGG CAA ^{3'}
AG 1913	G66+	5' CGC GGA CTT TGC CCA CAC CGC CCG GCA GGG ACT CCA GGA GGG CGG AG ^{3'}
AG 1914	35-37 Gly, G66+	5' AAG GCA GCC ATA TGG CTC TTC CGC CCC TTA CCC CTG AAC AG ^{3'}

In addition to truncating the N-terminal helix of sIHF, we also generated an sIHF variant where residues 35-37 were mutated to glycine residues (residue 36 is a glycine in the wildtype protein) through overlap extension PCR using primers AG 1910, AG 1911, AG 1759, and AG 1914 (Table 2.3). The resulting product was ligated into the pJET1.2 cloning vector (Thermo Scientific), digested with NdeI and BamHI, and subsequently ligated into the pET-15b vector as described above. (This sIHF mutant was made by Sabrina Lue Tam).

A sIHF variant containing a glycine residue insertion after position 66 was generated (sIHF G66+). The overexpression plasmid for sIHF G66+ (pAG 8866) was generated using overlap extension PCR using primers AG 1759, AG 1912, AG 1913, and AG 1910, (Table 2.3). The resulting product was ligated into the pJET1.2 cloning vector (Thermo Scientific), digested with NdeI and BamHI and subsequently ligated into the pET-15b vector as described above.

All His-sIHF mutants were overexpressed from *E. coli* BL21(DE3) Rosetta cells and purified as described in section 2.2. The buffer was exchanged to 150 mM NaCl, 20 mM Tris pH 8.0, 1.4 mM beta-mercaptoethanol, 5% (v/v) glycerol while concentrating the proteins. All sIHF constructs except for sIHF Δ 1-36 were supplemented with 25% (v/v) glycerol, at 0.43 mM and stored at -80°C for later use. Assays involving sIHF Δ 1-36 were conducted with freshly purified protein.

Table 2.4. Cloning plasmids

Name	Construct	Vector	Restriction Sites	Source
pAG 8771	sIHF Δ 1-13	pJET1.2	NdeI, BamHI	This work
pAG 8777	sIHF Δ 1-36	pJET1.2	NdeI, BamHI	This work
pAG 8825	sIHF S(19)	pUC57	NdeI, BamHI	GenScript
pAG 8861	sIHF G66+	pJET1.2	NdeI, BamHI	This work
pAG 8865	sIHF 35-37 Gly	pJET1.2	NdeI, BamHI	This work
pAG 8160	Dbf4 residues 1-314	pET-15b	NdeI, BamHI	In house

2.7 Dynamic light scattering (DLS)

DLS was used to assess protein quality for all sIHF constructs using the Zetasizer Nano (Malvern Instruments). Protein samples were analyzed at ~0.43 mM in 150 mM NaCl, 20 mM Tris pH 8.0, 1.4 mM beta-mercaptoethanol, and 5% (v/v) glycerol. Prior to data collection 20 μ L of sample

was centrifuged for 10 minutes at 4°C at 15,700 x g and 15 µL was loaded into a 12 µL quartz cuvette.

2.8 Electrophoretic mobility shift assays (EMSAs)

sIHF variants at increasing concentrations of protein (0-100 µM) were added to 0.02 µM of [γ -³²P]dATP 5'end-labelled duplex DNA (Motif 1: top strand 5'TCGAAAATCGGAATCTGGTGCA; bottom strand 5'TGCACCAGATTCCGATTTTCGA) with 1 mg/mL bovine serum albumin (BSA) and binding buffer (10 mM Tris pH 7.8, 5 mM MgCl₂, 60 mM KCl and 10% (v/v) glycerol). The reaction was incubated at room temperature for 10 minutes, followed by 30 minutes on ice. A glycerol-based loading dye was added and the samples were separated on a 15% native polyacrylamide gel at 100 V for 40 minutes. Gels were first exposed to a phosphor plate for ~30 minutes and visualised using a phosphorimager (Amersham Biosciences Ltd.), then subsequently exposed to Kodak Biomax XAR film for ~1 hour and developed. Assays were conducted in triplicate. (Assays were conducted by Dr. Emma Sherwood).

2.9 Topoisomerase assays

sIHF variants (5 µL at 540 nM and 2160 nM) were incubated with 5 µL pUC19 (64 nM) for 10 minutes at room temperature followed by 30 minutes on ice. One µL of TopA (7730 nM), 0.6 µL BSA (1 mg/mL) and 8.5 µL of reaction buffer (50 mM Tris pH 7.5, 50 mM KCl, 10 mM MgCl₂, 0.1 mM EDTA, 0.5 mM DTT, 0.06 mg/mL BSA) were added. Reactions were incubated at room temperature for 30 minutes and stopped with 5 µL of stop buffer (6% SDS, 30% (v/v) glycerol, 10 mM EDTA and 0.25% bromophenol blue). Samples were separated on a 1% TAE-agarose gel at

45 V for ~16 hours, stained with ethidium bromide and visualized using UV light. Assays were conducted in triplicate. (Assays were conducted by Melanie Gloyd).

2.10 Data collection and processing for small-angle X-ray scattering (SAXS)

sIHF was purified as described in section 2.2 and subsequently resolved using a Superdex-75 (GE Healthcare) size exclusion chromatography column in 40 mM Tris pH 8.0, 200 mM NaCl, 20 mM MgCl₂, 2.8 mM beta-mercaptoethanol, and 10% (v/v) glycerol. Sample homogeneity was assessed by dynamic light scattering (see Section 2.7). The protein-DNA complex was formed as described in section 2.3. Scattering data for His-sIHF bound to duplex DNA (Motif 1: top strand ^{5'}TCGAAAATCGGAATCTGGTGCA; bottom strand ^{5'}TGCACCAGATTCCGATTTTTCGA) at ratios of 1:1, 2:1, and 1:2 (sIHF:DNA) over a range of protein concentrations (0.868-0.109 mM) was collected on a BioSAXS-1000 mounted on a MicroMax-007HF X-ray generator. Data was collected for 120 minutes at each concentration with images refreshing every 20 minutes. Sample scatter curves were generated using Rigaku SAXSLab 3.0.0r1 by subtracting buffer scatter from sample scatter. Data quality was assessed for aggregation using Guinier plots (Guinier and Fournet, 1955), and Kratky plots (Glatter and Kratky, 1982) were used to compare protein concentrations and exposure times. Refer to Table 3.3 for SAXS data collection and analysis.

Chapter 3

Results

3.1 Purification of sIHF

The *sIHF* gene was inserted into the pET-15b overexpression vector, downstream of a 6xhistidine tag and thrombin cleavage site. Cell lysates overexpressing His-sIHF were loaded onto an immobilized metal affinity chromatography column. His-sIHF binds the nickel resin as the histidine tag at the N-terminus of the protein is exposed. Unbound contaminants were washed off the column with buffer A (see Methods), and weakly bound contaminants were eluted using a step gradient of buffer A supplemented with 7.5 and 20 mM imidazole. The protein was eluted from the column using buffer A supplemented with 150 mM imidazole (Figure 3.1A). A step gradient of imidazole was chosen over a linear gradient to enhance protein purity. Before the histidine tag was removed, eluted fractions from the nickel column were further purified using a cation exchange chromatography column to obtain pure His-sIHF (Figure 3.1B).

Eluted fractions containing His-sIHF (indicated by the single asterisks in Figure 3.1B) were pooled and diluted to 40 μ M. A small scale thrombin digestion reaction was conducted using a fraction of the pooled protein at a range of thrombin (Sigma) concentrations (0-0.125 units/ μ L) to determine the optimal concentration of thrombin needed for complete, but not over digestion. The optimal concentration of thrombin was selected based on the lowest concentration necessary to yield full digestion as visualized on an 18% SDS-PAG (Figure 3.1C). This concentration of thrombin was used to remove the histidine tag from the remainder of the pooled protein.

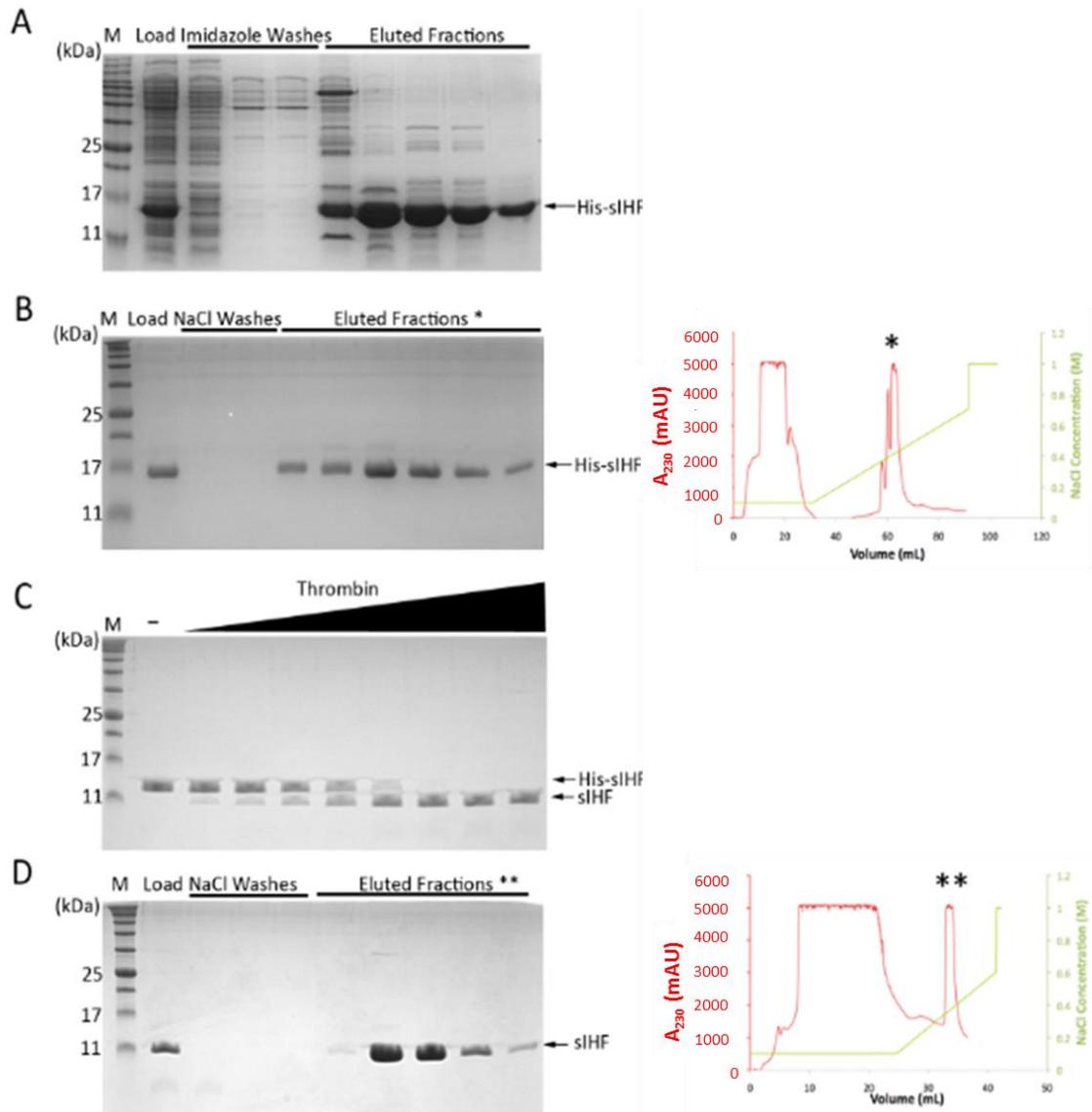


Figure 3.1. Purification scheme of untagged sIHF. 18% SDS-PAGs of fractions collected (left panels) and chromatograms (right panels) of the 4-step purification for the sIHF protein. sIHF bands are indicated by the arrows. **(A)** Cell lysate was loaded onto a HiTrap chelating affinity column and eluted using imidazole. **(B)** Fractions containing the protein were pooled and loaded onto an ion exchange column. Eluted fractions are indicated by the asterisks. **(C)** The 6xHis tag was removed using the thrombin protease. The optimal concentration of thrombin to be used was determined by a small scale thrombin digestion where 40 μ M of sIHF was digested with 0.01-0.125 units/ μ L of thrombin. **(D)** The untagged sIHF protein was further purified over an ion exchange column. Eluted fractions indicated by the double asterisks were pooled and concentrated.

Untagged sIHF was further purified by cation exchange chromatography. The protein was eluted using a linear salt gradient (Figure 3.1D). Eluted fractions containing pure sIHF (indicated by the double asterisks in Figure 3.1D) were concentrated.

3.2 Crystallization, data collection and structure of sIHF bound to a 19-bp duplex DNA substrate

We determined that sIHF binds DNA in a sequence independent manner with a strong preference for double stranded over single stranded substrates (Swiercz et al., 2013). To understand how sIHF interacts with DNA we wanted to solve the crystal structure of the protein bound to DNA. We attempted to crystallize sIHF with various G-C rich duplex DNA substrates, as this protein is from an organism with a high G-C content. DNA lengths ranging from 15- and 22-bp, that were blunt ended as well as contained 1- or 2-bp complementary overhangs (Table 2.2) were tested. We designed the oligonucleotides to have a unique pattern of purines and pyrimidines so that if we were successful in obtaining crystals, and subsequently obtaining an experimental electron density map, we could identify the DNA sequence from the pattern of large and small densities for each base. Furthermore, we tested DNA substrates with complementary overhangs as these overhangs could potentially aid in crystal packing; the oligonucleotides may stack end to end allowing complementary overhangs to base-pair forming a pseudo-continuous helix across the crystal.

We were successful in obtaining crystals of sIHF bound to a 19-bp duplex with a one base-pair overhang (TN200, see Table 2.2), at a 1:1 ratio of protein to DNA, using the PEG/Ion screen (Hampton Research) in the following condition: 20% PEG 3350 (w/v), 0.2 M KSCN, pH 7.

These crystals were optimized using the hanging drop method to favour the growth of large single crystals. We first determined the optimal pH for crystal growth using a pH screen (Hampton Research). We also varied the ratio of protein to DNA and found that a 1:1 ratio yielded the largest single crystals (Figure 3.2).

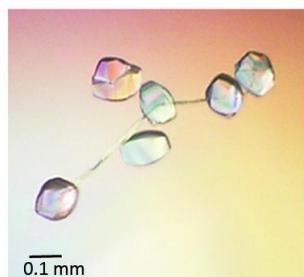


Figure 3.2. Crystals of sIHF bound to a 19-bp duplex DNA substrate. Crystals of sIHF bound to TN200 grown in 18% (w/v) PEG 3350, 0.21 M KSCN, 0.1 M HEPES pH 7.6, and 5% (v/v) ethylene glycol.

To minimize crystal damage by X-ray radiation during data collection, crystals were cryoprotected and data were collected at 100 K. Conventional cryoprotection of crystals is achieved by transferring crystals to a solution with the crystallization condition supplemented with a cryoprotecting agent such as ethylene glycol, polyethylene glycol or glycerol. These methods damaged our crystals. In order to surmount this problem and still achieve good cryoprotection, ethylene glycol was included in the crystallization solution and the crystals were dehydrated over increasing concentrations of potassium chloride (see Methods). A complete data set of the native sIHF-TN200 crystal (TN055, see Table 3.1) was collected at the NSLS (Brookhaven National Laboratory, NY, USA). This crystal diffracted to 2.85 Å.

The structure of this complex could not be determined by molecular replacement because a good structural homologue for sIHF does not exist. Therefore, to solve the phase problem, crystals of selenomethionine substituted protein bound to TN200 were obtained (see Methods). The best crystal (TN051, see Table 3.1) diffracted to 2.6 Å and a complete data set

was collected using the X25 beam line at the NSLS (Brookhaven National Laboratory, NY, USA) at a wavelength of 0.979. This wavelength was chosen as it is the X-ray absorption edge of selenium, allowing for anomalous signal from the selenium atoms to be collected and used to phase the crystal. As this crystal diffracted to a higher resolution than the native crystal, we decided to refine the structure of sIHF bound to TN200 using the Sel-Met data set. Data for this crystal was indexed, processed and merged in the space group $C222_1$, C centered orthorhombic, using HKL2000 (Otwinowski and Minor, 1997). The single-wavelength anomalous dispersion (SAD) method was used to phase the crystals using SOLVE (Terwilliger, T. C. and Berendzen, J., 1999). sIHF contains three methionine residues in the 107 residue protein. This provides enough anomalous signal to phase the structure by SAD. The initial model was built manually in COOT by modeling in the protein sequence into the experimental electron density map using signal from the selenium atoms as landmarks. Residues 14-103 were readily identified and a final model was refined to a resolution of 2.70 Å. Complete statistics for data collection and refinement can be found in Table 3.1.

The refined structure of sIHF bound to TN200 contained 98.84% of residues in the preferred region of the Ramachandran plot, 1.16% in the allowed region and none in the disallowed region. The structure revealed that sIHF is composed of a long protruding N-terminal α -helix (α 1) followed by four shorter α -helices that make up the core of the protein (Figure 3.3A). The sIHF protein contacts DNA at two distinct interfaces. One interface uses the loop region between helices 4 and 5 which spans residues Gly80 to Gly91 and has been termed the lid region (Figure 3.3 A, B in pink). This lid serves two functions; it conceals the hydrophobic core of

the protein as well as provides a positively charged flat surface that is free to interact with the negatively charged phosphate backbone of one strand of the DNA duplex.

Table 3.1. X-ray data collection and refinement					
Crystal	TN055 (native)	TN051 (Sel- Met)	TN100 (native)	TN103 (native)	TN131 (native)
DNA	TN200	TN200	TN08O	TN08-BP1	TN23-HP9
Data Collection					
Space Group	C222	C222 ₁	C222 ₁	C222 ₁	P2 ₁
Unit Cell (Å)	a=41.8, b=72.7, c=102.3	a=43.4, b=71.8, c=102.9	a=45.7, b=71.9, c=100.8	a=42.3, b=72.1, c=103.3	a=67.3, b=40.4, c=89.9; α=90, β=108.6, γ=90
Wavelength (Å)	1.1	0.979	0.979	0.979	1.1
Resolution (Å)*	50-2.85 (2.90-2.85)	50-2.60 (2.64-2.60)	35-3.00 (3.11-3.00)	50-1.66 (1.69-1.66)	35-2.90 (2.95-2.90)
Completeness (%)*	99.7 (98.9)	100 (99.5)	99.8 (99.7)	98.7 (99.8)	94.1 (81.3)
Redundancy*	4.8 (5.8)	6.1 (4.6)	3.4 (3.3)	6.1 (5.9)	3.2 (2.9)
I/σ*	19.2 (1.9)	31.2 (1.4)	23.1 (1.3)	21.1 (3.7)	24.1 (2.7)
Data Refinement					
Resolution (Å)*		29.4-2.7 (3.1-2.7)		39.11-1.60 (1.64-1.60)	34.13-3.00 (3.11- 3.00)
Reflections (work)*		4613 (1349)		42000 (2632)	8132 (1014)
Reflections (test)*		447 (142)		1998 (131)	914 (120)
Atoms refined		1006		2485	2261
Solvent atoms		6		229	6
R _{free} (%)		31.7		29.5	32.68
R _{work} (%)		26.1		26.9	28.6
Rmsd in bonds (Å)		0.003		0.009	0.012
Rmsd in angles (°)		0.789		1.4	2.1
<i>*Data in the highest resolution shell shown in parenthesis.</i>					

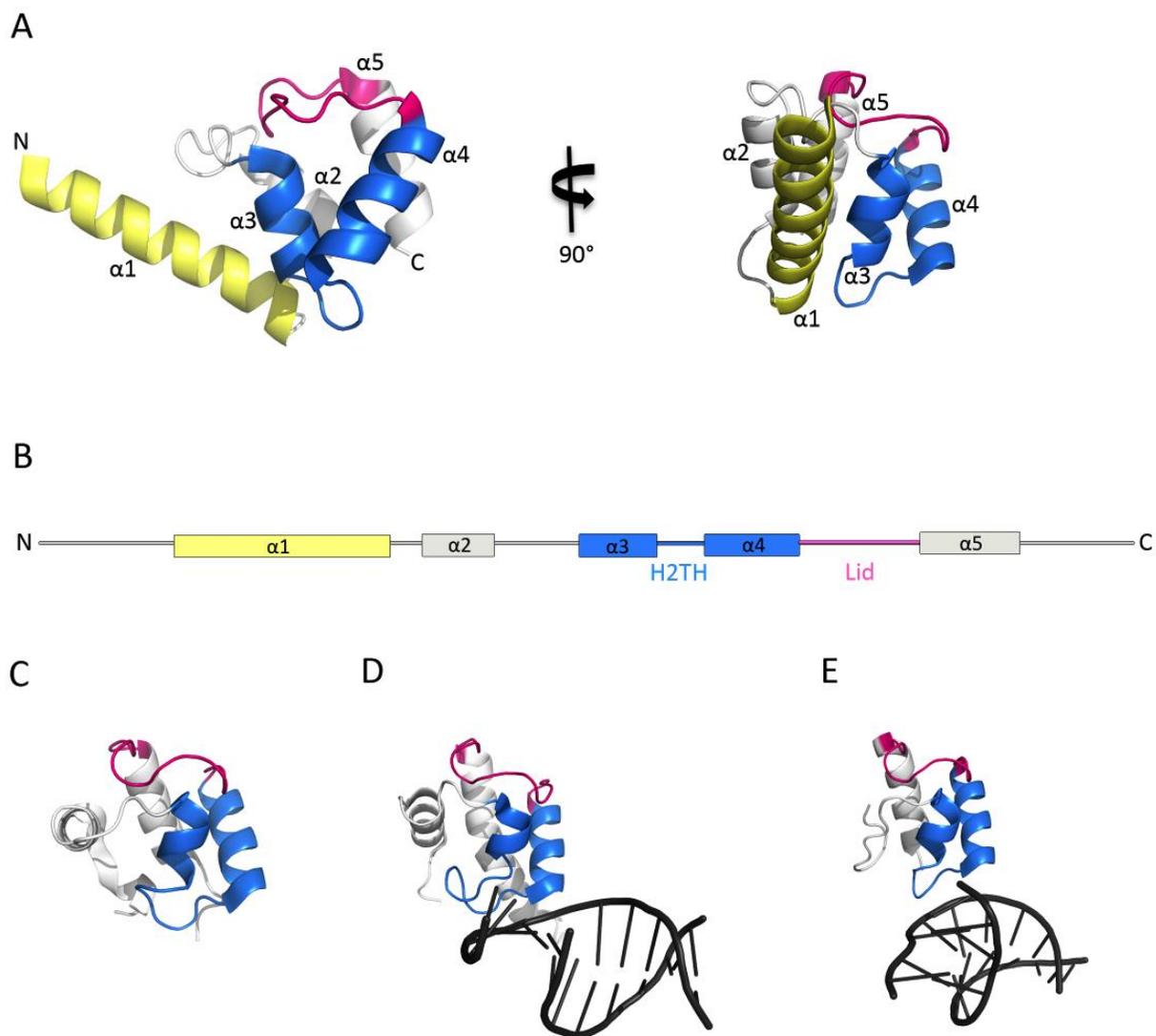


Figure 3.3. Structure of sIHF. **(A)** Ribbon diagram of sIHF with the N-terminal helix shown in yellow, the H2TH motif shown in blue and the lid region shown in pink. The image is rotated 90° in the second panel. **(B)** Secondary structure of sIHF highlighting the N-terminal helix, the H2TH, and the lid regions in yellow, blue, and pink respectively. **(C-E)** Ribbon diagrams of topoisomerase VI (residues Lys230-Phe306, PDB 1MU5), endonuclease VIII (residues Pro132-Gln214, PDB 1K3W) and ribosomal protein S13 (residues Ala1-Phe62, PDB 2GY9). The H2TH motif is displayed in blue, and the lid region in pink.

The second interface of sIHF that contacts DNA uses helices 3 and 4, spans residues Lys56 to Leu79 (Figure 3.3 A, B in blue) and forms a fold classified as a helix-two turns-helix (H2TH) motif. This motif is found in a few other DNA binding proteins including topoisomerase

VI (a type IIb topoisomerase), the endonuclease VIII family of base excision repair enzymes, and the ribosomal protein S13 (Corbett and Berger, 2003; Zharkov et al., 2002; Brodersen et al., 2002) (Figure 3.3 C-E respectively, in blue). The function of this motif is unknown, although it is thought to aid in peripheral nucleic acid binding. Endonuclease VIII and ribosomal protein S13 contact DNA and RNA respectively, through this interface in their crystal structures; however, this motif is embedded within much larger proteins that form extensive contacts with nucleic acids (Zharkov et al., 2002; Brodersen et al., 2002). Interestingly, the lid region of siHF is also found in the three proteins mentioned above and adopts a similar conformation to that found in siHF (Figure 3.3 in pink). In both the monomer and dimer structures of topoisomerase VI the lid region is more exposed than the H2TH motif; however, as this structure has not been solved in complex with DNA, we cannot infer which interface(s) contact DNA (Corbett and Berger, 2003).

The DNA substrate used to form the siHF-DNA crystal was 20-bp long; however, we only identified 8-bp within the asymmetric unit in the experimental electron density map. The electron density, which represents an average of each unit cell within the crystal, was well defined for the DNA phosphate backbone, but not for the nucleobases. Therefore, we modeled in the nucleobases as guanine or cytosine based on the size of electron density, as the substrate used was G-C rich. We believe that the nucleobases were disordered because the crystals packed to form a pseudo-continuous duplex allowing siHF to bind several different sequences along the duplex within the crystal. This reaffirms our observation that siHF does not bind DNA at a specific sequence (Swiercz et al., 2013). We next wanted to obtain a crystal structure of siHF bound to DNA where the nucleobases could be identified. As 8-bp were identified in the asymmetric unit, we decided to pursue the structure of siHF bound to an 8-bp DNA substrate.

3.3 Crystallization, data collection and structure of sIHF bound to an 8-bp

DNA substrate

Crystallization screens of sIHF bound to a 7-bp duplex with 1-bp overhangs on each end (TN080, see Table 2.2) were conducted using the original crystallization condition as a starting point. Optimal crystals were obtained using 1.0 mM of the complex in 0.1 M HEPES pH 7.6, 0.1 M KSCN, 28% (w/v) PEG 3350, and 5% (v/v) ethylene glycol (TN100, see Table 3.1). These crystals diffracted to 3.00 Å. Once again, the nucleobases could not be identified as the DNA bases bound by sIHF were not consistent throughout the crystal. Consequently, we decided to crystallize the protein bound to a blunt ended 8-bp duplex that is a palindrome, and whose structure has previously been solved by solution NMR (PDB 1D18) (Baleja et al., 1990). We decided to work with a palindromic sequence so that regardless of the strand of DNA that sIHF binds; the bases it interacts with will be identical. We were successful in obtaining crystals of sIHF bound to TN08-BP1 at 1.5 mM in 0.1 M HEPES pH 7.6, 0.2 M KSCN, 19% (w/v) PEG 3350, and 5% (v/v) ethylene glycol. A complete data set was collected on beam line X25 at the NSLS (Brookhaven National Laboratory, NY, USA). Molecular replacement was used to phase the crystal with the structure of sIHF bound to TN200 as a model. Data were refined to a resolution of 1.65 Å (Table 3.1) and sIHF residues 15-103 were readily identified in the experimental electron density map. In the structure, we were able to identify the phosphate backbone and the nucleobases of the DNA. In the refined structure, 99.42% of residues were in the preferred region of the Ramachandran plot, and the rest were in the allowed regions. The resolution of this structure was higher than those solved previously, allowing for more detailed analysis of the interaction between sIHF and DNA.

These crystals packed in the same space group as crystal TN051 (Table 3.1) and interact with DNA at the same interfaces. The resolution obtained in this structure is high enough to assess if siHF alters DNA upon binding. We analyzed differences in the structure of the DNA alone (PDB 1D18) and the DNA from the crystal structure. The Web 3DNA software for three-dimensional analysis of nucleic-acid structures (Zheng et al., 2009) was used to compare complementary base-pair and base-pair step parameters. These results are summarized in Table 3.2. We observed that the stagger (vertical displacement between complementary base-pairs) of the DNA alone is greater than that of the DNA in complex with siHF, with greater differences at the center of the substrate. Differences are also seen in the buckle (when both complementary base-pairs fold upwards and are not planar) of the base-pairs with more significant differences at the ends of the substrate. A difference in the angle of propeller twist (when complementary base-pairs twist in opposite directions) is also observed and can be visualized in the crystal structure in the central base-pair.

There is also a difference in the opening (the space between complementary base-pairs) of the base-pairs. The opening between base-pairs for the 8-bp duplex previously solved deviate from standard B-DNA (reported in Table 3.2) more than our structure. This could be because the structure of PDB 1D18 was solved in solution by NMR, whereas the structures of DNA solved in this work as well as that stated for standard B-DNA are from crystal structures that are in a locked conformation. Differences in the tilt and roll DNA structural parameters also vary between DNA alone and in complex with siHF. These parameters have been affiliated with mediating the widths of the major and minor grooves (Oguey et al., 2010).

Table 3.2. Base-pair parameters of duplex DNA alone and in complex with sIHF

	PBD 1D18								Standard B-DNA*	
Parameter	C-G	A-T	T-A	G-C	C-G	A-T	T-A	G-C		
Buckle (°)	8.97	10.61	9.55	3.44	-3.78	-8.84	-10.60	-7.45	0.50	
Propeller (°)	-7.56	-12.48	-8.91	-17.22	-17.23	-7.88	-12.80	-6.34	-11.40	
Opening (°)	-4.43	-6.81	-7.20	-4.87	-4.12	-6.74	-6.65	-4.78	0.60	
Shear (Å)	0.22	-0.44	0.25	-0.40	0.46	-0.25	0.43	-0.18	0.00	
Stretch (Å)	-0.18	-0.21	-0.17	-0.25	-0.24	-0.18	-0.21	-0.19	-0.15	
Stagger (Å)	0.20	0.29	0.42	0.37	0.34	0.47	0.24	0.20	0.09	
Tilt (°)	-0.73	0.15	2.70	0.29	-3.01	0.05	0.76		-0.10	
Roll (°)	8.03	3.72	3.50	-6.07	3.44	4.47	8.74		0.60	
Twist (°)	29.01	32.79	34.71	44.02	33.91	32.60	29.37		36.00	
Shift (Å)	-0.75	0.02	0.44	0.04	-0.45	-0.01	0.70		0.02	
Slide (Å)	-0.18	-0.08	-0.42	-0.37	-0.45	-0.02	-0.09		0.23	
Rise (Å)	3.17	3.41	3.34	3.47	3.30	3.45	3.12		3.32	
	DNA in crystal TN103									
Parameter	C-G	A-T	T-A	G-C	C-G	A-T	T-A	G-C		
Buckle (°)	-6.40	-6.49	-2.62	4.27	-2.66	2.16	5.81	10.72		
Propeller (°)	-10.97	10.91	-7.86	-13.86	-8.46	-9.78	-5.90	-12.47		
Opening (°)	0.00	-0.49	0.19	0.52	-0.35	-0.56	0.74	-0.46		
Shear (Å)	0.03	0.06	-0.06	-0.05	0.18	0.11	0.07	-0.45		
Stretch (Å)	-0.07	-0.15	-0.13	-0.15	-0.21	-0.12	-0.13	-0.15		
Stagger (Å)	0.10	0.02	-0.10	0.27	0.20	-0.13	-0.11	0.21		
Tilt (°)	0.00	-2.11	0.94	-4.27	0.00	1.40	-0.55	-0.36		
Roll (°)	0.00	7.33	-0.57	9.89	0.18	7.34	0.60	13.17		
Twist (°)	0.00	34.95	28.35	37.45	30.24	32.65	30.98	33.76		
Shift (Å)	0.00	-0.49	0.19	0.52	-0.35	-0.56	0.74	-0.46		
Slide (Å)	0.00	0.41	-0.41	-0.19	-0.52	0.24	-0.62	-0.29		
Rise (Å)	0.00	3.36	3.20	3.19	3.53	3.27	3.24	3.19		
	DNA in crystal TN131									
Parameter	G-C	T-A	G-C	C-G	G-C	T-A	G-C	G-C	A-T	T-T
Buckle (°)	3.31	-11.19	-10.98	0.11	-7.60	11.63	1.88	10.86	5.28	-5.63
Propeller (°)	-8.72	-23.67	-2.55	-7.77	-12.68	-10.96	-12.11	-17.50	-34.91	-25.55
Opening (°)	-11.26	4.10	-6.65	6.80	-2.06	-4.49	6.68	12.75	21.09	3.61
Shear (Å)	-0.04	-0.47	-0.83	-0.29	1.04	-0.90	-0.80	-0.58	0.98	1.77
Stretch (Å)	-0.53	0.36	0.05	0.64	-0.18	-0.12	-0.11	0.19	0.05	-2.01
Stagger (Å)	-0.53	0.41	-0.86	-0.19	0.14	0.11	0.56	-0.15	0.15	-0.68
Tilt (°)	-3.22	5.35	-4.13	0.16	1.55	0.74	5.23	1.03	-0.12	
Roll (°)	1.88	18.50	1.95	6.60	3.25	-5.25	9.16	-7.64	-0.38	
Twist (°)	32.80	28.05	38.08	40.95	21.16	42.37	34.92	42.42	32.50	
Shift (Å)	0.21	-0.20	0.17	-0.48	0.17	0.82	-0.12	1.01	0.57	
Slide (Å)	-1.06	-0.60	0.23	0.64	-1.03	1.90	0.32	0.71	-1.02	
Rise (Å)	3.74	3.26	3.06	3.61	2.77	3.66	3.13	3.69	3.80	

*Average values from crystal structures solved at 2 Å resolution or higher (Olson et al 2001, Berman et al., 1992).

This suggests that SIHF induces structural changes to DNA upon binding. In addition, the width of the minor groove of the DNA alone (8.3 Å) is much smaller than the minor groove of the DNA in the crystal structure (13.5 Å). SIHF binds DNA within the minor groove of DNA and may widen the minor groove upon binding to cause changes in complementary base-pair and base-pair step parameters.

The residues of SIHF that interact with DNA could also be analyzed in more detail in this high resolution structure. SIHF contacts DNA at two interfaces (Figure 3.4A). One surface, which from now on will be referred to as Interface I, is mediated by the lid region at four key residues; Arg85, Arg86, Asn93, and Gln94 (Figure 3.4B and C). Both positively charged arginine residues pack well against the negatively charged DNA phosphate backbone. Furthermore, these residues form extensive contacts with the DNA sugar-phosphate backbone. The guanidinium group in the arginine side chain at position 85 contacts the DNA ribose sugar at the oxygen, as well as the carbon atoms at the 1', 2' and 4' positions, and the oxygen atoms forming the phosphodiester linkage. Arg86 does not form direct contacts with DNA; however this residue is positively charged and could aid in DNA binding by neutralizing the negatively charged DNA phosphate backbone, allowing for a stronger interaction between SIHF and DNA. Further contacts were made between the amino group of Asn93 and the 3' carbon of the DNA ribose sugar, while the oxygen atom of Asn93 contacts an ordered water molecule. The amino group of Gln94 makes extensive contacts with the phosphate group of the DNA backbone.

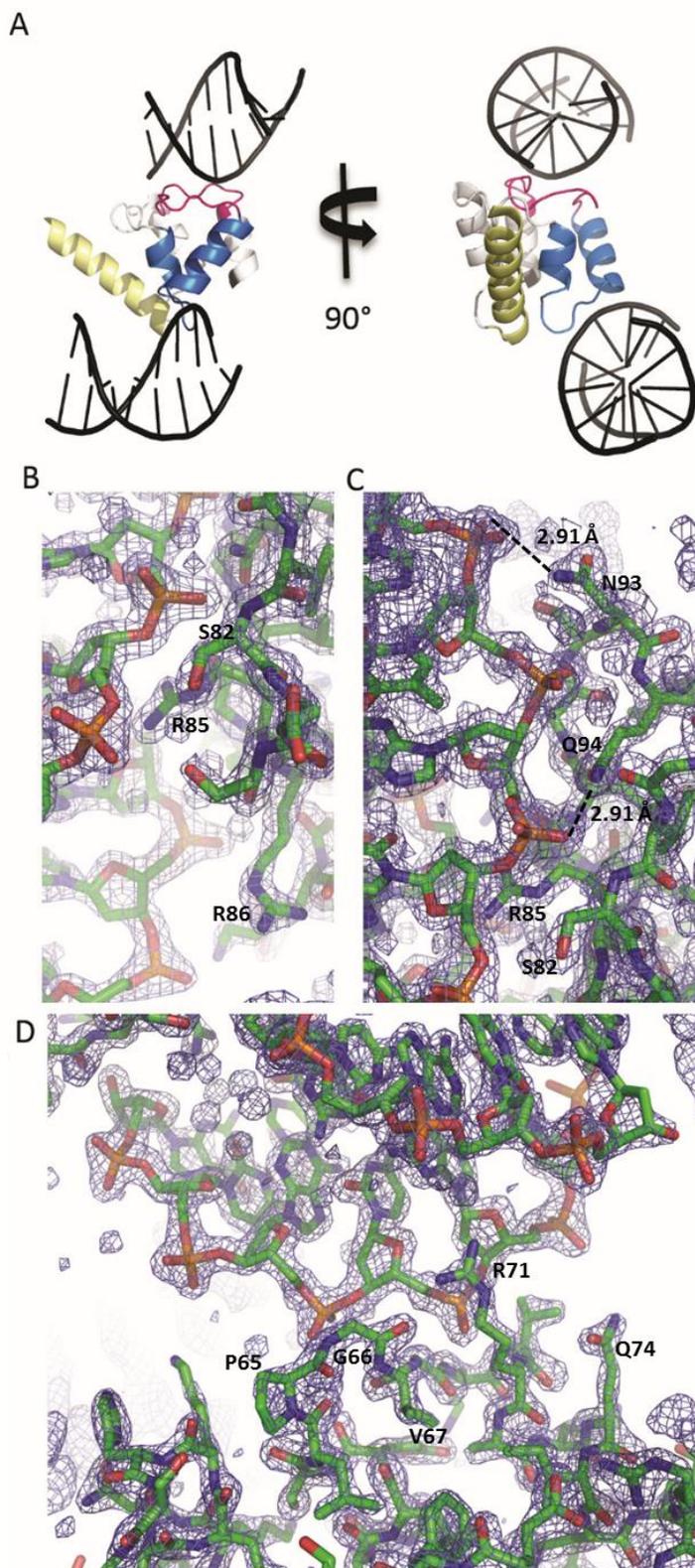


Figure 3.4. Structure of sIHF bound to TN08-BP1. **(A)** Cartoon of sIHF bound to TN08-BP1 with the same color scheme used in Figure 3.3. The image is rotated 90° in the second panel illustrating the interfaces of sIHF that contact DNA. **(B)** 2Fo-Fc electron density map, obtained from crystal TN103, and modeled structure indicating positions of residues R85 and R86 with respect to the DNA phosphate backbone. **(C)** 2Fo-Fc electron density map, obtained from crystal TN103, and modeled structure illustrating hydrogen bonds observed between sIHF residues N93 and Q94 with the DNA phosphate backbone. **(D)** 2Fo-Fc electron density map, obtained from crystal TN103, and modeled structure with focus on interface II. In panels B-D carbon atoms are coloured in green, nitrogen atoms in blue, oxygen atoms in red and phosphorous atoms in orange.

The second interface of sIHF that contacts DNA, which will from now on be referred to as Interface II, is mediated by the H2TH motif, using main chain contacts between the protein peptide backbone and the DNA phosphate backbone (Figure 3.4D). In the two crystal structures mentioned, sIHF contacts DNA at the same two interfaces. In an attempt to determine if these two interfaces are real, or if they are an artifact of crystal packing, we decided to solve the crystal structure of sIHF bound to DNA in a different crystal form, as interaction surfaces caused by crystal contacts, are less likely to be conserved amongst crystals grown in different space groups.

3.4 sIHF contacts DNA at three interfaces in the crystal structure of sIHF bound to a hairpin DNA substrate

To obtain crystals of sIHF bound to DNA in a different space group than the crystals previously obtained, we decided to conduct crystallization screens using a hairpin DNA substrate (TN23-HP9, see Table 2.2). We picked a hairpin substrate as the loop at the end of the hairpin would disrupt the continuity of the pseudo-continuous helix found in crystals TN051 and TN103, forcing the crystals to pack differently. Crystals of sIHF bound to TN23-HP9 were difficult to produce; however, they were successfully grown using the streak seeding method. The best crystal (TN131, see Table 3.1) diffracted to 2.9 Å and was obtained in 0.1 M MES pH 5.6, 0.1 M MgCl₂, 14% (w/v) PEG 3350, and 5% (v/v) ethylene glycol (see Methods).

Data for the best crystal were collected on beamline X25 at the NSLS (Brookhaven National Laboratory, NY, USA). The crystal diffracted weakly and anisotropically, resulting in a data set that was only 94.1% complete (only 81.3% complete in the highest resolution shell) with

low redundancy (see Table 3.1). Since the space group could not be unequivocally determined, we first indexed the data in the space group with the lowest symmetry (P1). Using the structure of sIHF bound to TN08-BP1 as a model, we attempted many algorithms to obtain a correct molecular replacement solution. We altered the number of residues used in our model as well as the number of molecules to search for within the unit cell, and the root-mean-squared deviation (RMSD) from the models. Once we obtained a solution that placed our model well within the experimental electron density, we assessed the validity of this solution by refining the structure using phenix.refine. The R_{work} and R_{free} decreased, suggesting that the solution obtained was correct. We then reprocessed the data in the most probably space group ($P2_1$). We continued to search for molecular replacement solutions using PhaserMR in the Phenix suite using the first solution obtained as a partial solution. We successfully obtained a molecular replacement solution in the $P2_1$ space group. Missing residues were manually built into the experimental electron density map using COOT and the model was refined using iterative cycles of model building in COOT and refinement in phenix.refine (Afonine et al., 2012; Emsley and Cowtan, 2004). The final model includes residues 14-103 of the sIHF protein and the 23 DNA base-pairs of the DNA substrate used to obtain the crystals (Figure 3.5). In the refined structure 90.23% of residues were in the preferred region of the Ramachandran plot, 8.05% of residues were in the allowed region while 1.72% of residues were in the unfavoured region. The three residues that are found in the unfavoured region of the Ramachandran plot are His40, Val42 and Val70.

We designed the DNA substrate to form a 9-bp stem and a loop consisting of 5 thymine residues. In the crystal, the DNA forms a 10-bp stem and a loop with 3 thymine residues. In the

hairpin stem, 9-bp form complementary Watson-Crick base pairs while the last base pair, at the junction of the stem and loop, occurs between two thymine residues. The nucleobases in the loop of the hairpin are exposed and contact the protein at the N-terminal helix.

We can also analyze the structural base-pair and base-pair step parameters of the DNA in this crystal structure. Using the Web 3DNA software for three-dimensional analysis of nucleic-acid structures (Zheng et al., 2009), we observed variations in DNA structural parameters compared to standard B-DNA. Once again we observed that the stagger and buckle base pair parameters differ from standard B-DNA. In addition, there was a pronounced propeller twist at the junction of the stem and hairpin that deviate from that of standard B-DNA. This twist may be necessary for the formation of the hairpin loop (Table 3.2). With respect to the tilt and roll parameters, the structure of the hairpin DNA substrate deviated from standard B-DNA, with greater differences at the stem of the hairpin. Furthermore, the rise between each base-pair step on average was comparable to that of standard B-DNA (Table 3.2). Lastly, the minor groove of the DNA was wide (12.9 Å) compared to standard B-DNA (~6 Å) (Chandrasekaran and Arnott, 1996).

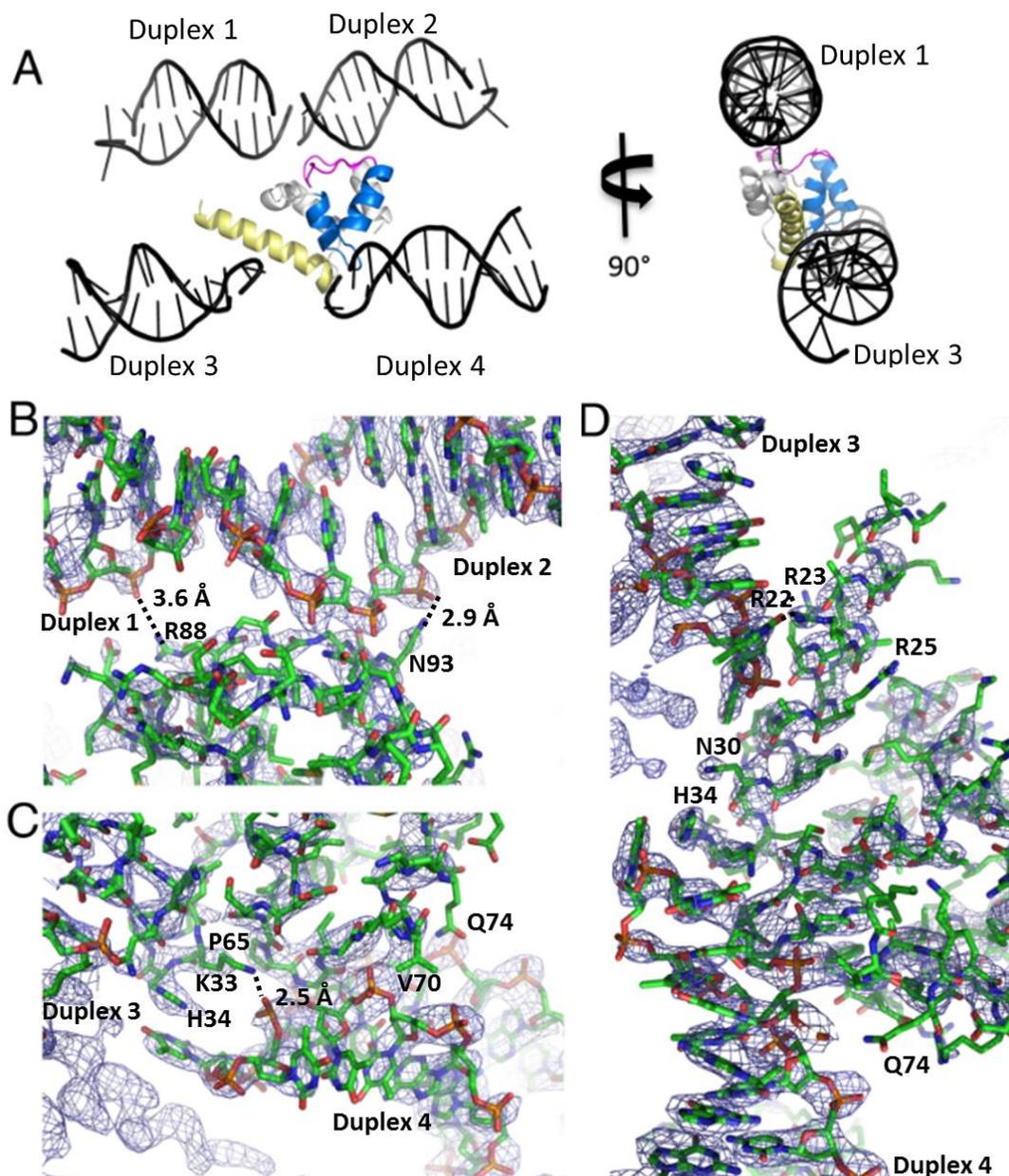


Figure 3.5. Structure of siHF bound to TN23-HP9. **(A)** Ribbon diagram of siHF bound to TN23-HP9 highlighting the 4 duplex substrates that contact siHF. Image is rotated 90° in the second panel. **(B-D)** 2Fo-Fc electron density maps and model of siHF bound to TN23-HP9 determined from crystal TN131 highlighting the interactions with siHF at **(B)** Interface I, **(C)** Interface II, **(D)** and the N-terminal helix and DNA.

In this crystal structure, the DNA contacts the protein at three interfaces (Figure 3.5); Interface I and II, previously discussed, and the N-terminal helix. siHF makes contacts with DNA through several residues. At Interface I, siHF contacts DNA using residues Arg85, Asn93, and Gln94 which contact the DNA phosphate backbone (Figure 3.5B). siHF contacts DNA at Interface II through the peptide backbone (Figure 3.5C). The N-terminal helix of siHF interacts with DNA at residues Arg22, Asn30, and His34. Arg22 contacts an oxygen atom in the phosphate group of the thymine residue located at the intersection of the stem and loop of duplex 3, and the sugar at the loop of the hairpin at C atoms in the 1' and 2' positions (Figure 3.5C). Asn30 contacts the methyl group of the exposed thymine base within the loop of the hairpin in duplex 3 while His34 forms extensive contacts with the exposed thymine base of duplex 4 (Figure 3.5C). In order to assess if these three interfaces are important for DNA binding in solution, the siHF protein was mutated to disrupt these interfaces and DNA binding was probed.

3.5 Three interfaces on siHF contribute to DNA binding *in vitro*

The X-ray crystal structures of siHF bound to DNA substrates TN200 and TN08-BP1 show siHF-DNA interactions at two interfaces: Interface I which is composed of the lid region (Figure 3.4A in pink) of the protein and is mediated by residues Arg85, Arg86, Asn93 and Gln94 (Figure 3.4B and C); and Interface II which is facilitated by the peptide backbone of the H2TH motif (Figure 3.4A in blue). The N-terminal helix was also identified to contact DNA in the crystal structure of siHF bound to DNA substrate TN23-HP9 (Figure 3.5D). To validate that these surfaces are important for DNA binding in solution *in vitro*, electrophoretic mobility shift assays (EMSAs) of siHF mutants were conducted with radiolabeled duplex DNA. siHF was shown to bind the sequence

of duplex DNA chosen for this assay preferentially through systematic evolution of ligands by exponential enrichment (SELEX) (Swiercz, 2013).

To test if sIHF Interface I contributes to DNA binding, sIHF mutants were constructed where residues arginine 85 and arginine 86 were mutated to alanine and serine residues (sIHF-RR85AS); asparagine 93 and glutamine 94 were mutated to alanine and serine residues (sIHF-NQ93AS); and where all four key residues on Interface I were mutated (sIHF-RR85AS+NQ93AS). EMSAs using these sIHF variants were conducted and the results were compared with those done using wildtype sIHF to determine if DNA binding was reduced for these mutants, indicating that the mutated residue is important for DNA binding.

Duplex DNA that was not bound by protein migrated further on a native acrylamide gel (Figure 3.6A indicated by the asterisks) than the sIHF-DNA complex (Figure 3.6A indicated by the arrow). Wildtype sIHF was used as a control to compare DNA binding affinity of the sIHF mutants. Wildtype sIHF bound duplex DNA (0.02 μM) readily at the lowest concentration of protein assayed (3 μM) as visualized by a shift of the radiolabeled DNA to form the sIHF-DNA complex. sIHF-RR85AS displayed a reduction in DNA binding; even at the highest concentration of protein tested (100 μM) sIHF-RR85AS did not fully shift the DNA (migration of the band was smeared). The sIHF-NQ93AS mutant behaved similar to wildtype (Figure 3.5A). The quadruple mutant, sIHF-RR85AS+NQ93AS, displayed a similar phenotype to the sIHF-RR85AS mutant (Figure 3.6B), suggesting that residues Arg85 and/or Arg86 were important for DNA binding, whereas residues Asn93 and Gln94 were dispensable *in vitro*.

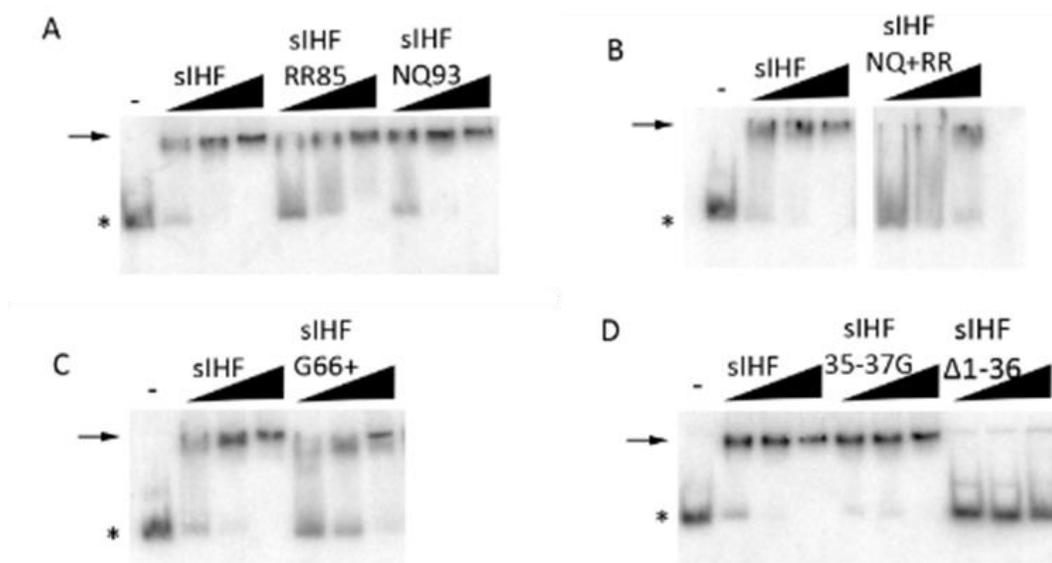


Figure 3.6. EMSAs of sIHF mutants with a radiolabeled 23-bp duplex DNA substrate. **(A, B)** EMSAs of sIHF variants to test Interface I. Wildtype sIHF, sIHF RR85AS (sIHF RR85) and sIHF NQ93AS (sIHF NQ93) with radiolabeled duplex DNA (0.02 μ M) at sIHF concentrations of 0, 3, 10 and 100 μ M. **(C)** EMSA of wildtype sIHF and an sIHF variant to test Interface II. sIHF G66+ with radiolabeled duplex DNA (0.02 μ M) at sIHF concentrations of 0, 3, 10 and 100 μ M. **(D)** EMSA of wildtype sIHF and sIHF variants to test the N-terminal helix. sIHF 35-37Gly (sIHF 35-37G) and sIHF Δ 1-36 with radiolabeled duplex DNA (0.02 μ M) at sIHF concentrations of 0, 3, 10 and 100 μ M. Asterisks indicate the migration of duplex DNA alone while the arrows indicate the migration of the sIHF-DNA complex.

sIHF interacts with DNA through a second interface defined by the loop region of the H2TH motif, solely through the peptide backbone (Figure 3.4D). In order to alter the length and flexibility of the loop at this motif, we decided to insert an extra glycine residue after position 66 (sIHF-G66+). sIHF-G66+ displayed reduced binding compared to wildtype (Figure 3.6C), suggesting that Interface II is also important for DNA binding.

The N-terminal helix of sIHF contacts DNA in the crystal structure of sIHF bound to a hairpin DNA substrate (Figure 3.5D). The N-terminal helix might interact with DNA at this region as it contains many positively charged residues including Arg22, Arg25, Lys29 and Lys33. These

residues are all within the region connecting the N-terminal helix to the 4-helix bundle (Figure 3.7B, indicated by the arrow). Furthermore, this patch of positive charge runs continuously with that of the charge at Interface I to potentially provide a large positively charged surface for binding DNA (Figure 3.7C). To test if the N-terminal helix is important for DNA binding we generated two sIHF mutants. One was designed to unlatch the helix from the helix bundle (sIHF 35-37Gly), by mutating the three residues that connect the N-terminal helix and the rest of the protein to glycine residues. The second sIHF construct lacks the entire N-terminal helix (sIHF Δ 1-36). sIHF 35-37Gly bound DNA similarly to wildtype (Figure 3.5D) suggesting that this helix does not need to be rigid for wildtype binding, or that our mutant did not effectively unlatch the helix. sIHF Δ 1-36 did not bind DNA even at 1000 x excess protein compared to DNA, suggesting that this helix is important for DNA binding or that this mutant was not stable in the conditions tested.

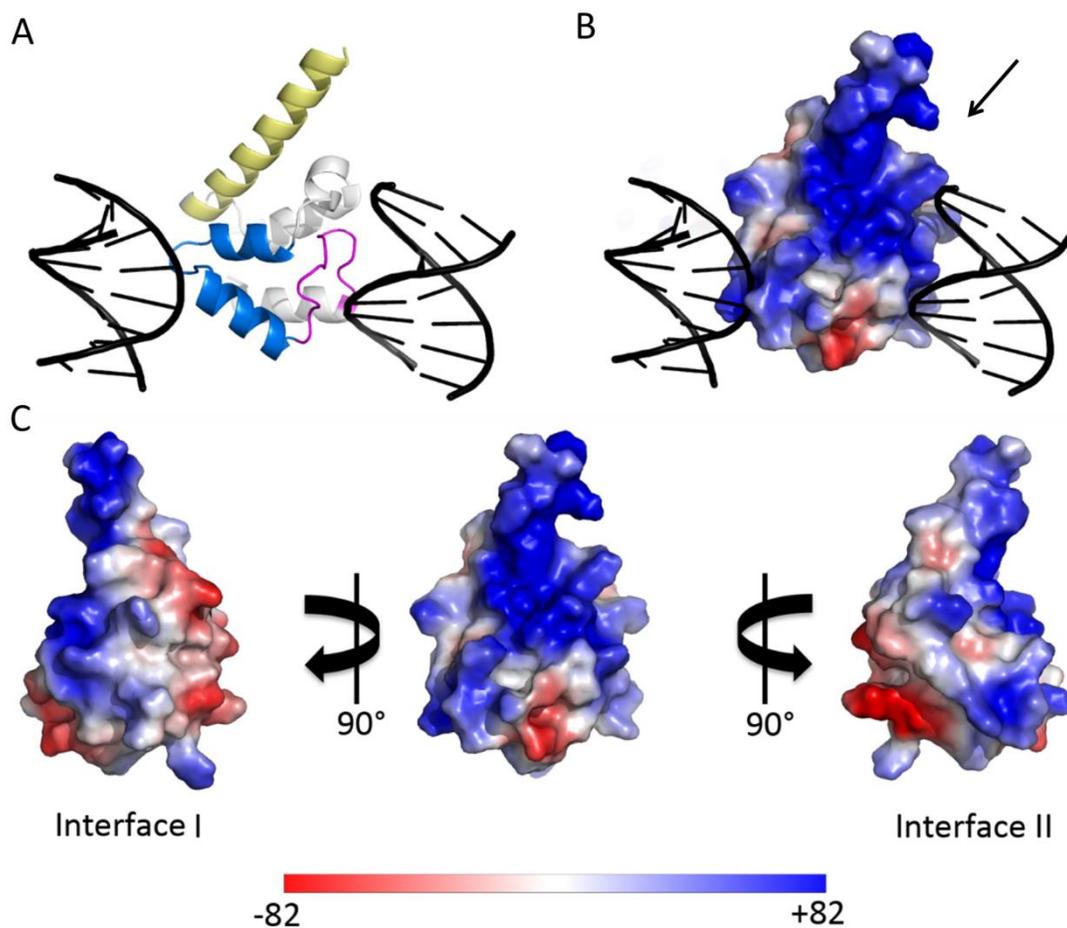


Figure 3.7. Electrostatics of siHF bound to DNA. From crystal structure of siHF bound to TN08-BP1. **(A)** Ribbon diagram and **(B)** electrostatic representation of siHF bound to DNA. (A) and (B) are in the same orientation. The arrow points at the interface between the N-terminal helix and the 4-helix bundle. **(C)** Electrostatic map of siHF indicating charge at Interface I and Interface II of siHF. Center molecule is in the same orientation as (A) and (B). Positive charge is indicated in blue, negative charge in red and neutral charge in white.

3.6 siHF does not bridge DNA

Multiple DNA-binding interfaces identified in the crystal structure were confirmed to bind DNA *in vitro*. One mechanism that NAPs use to condense the bacterial nucleoid is through bridging disparate DNA molecules (Dillon and Dorman, 2010). In order to determine if siHF is able to

bring together two DNA molecules, we employed small-angle X-ray scattering (SAXS). SAXS provides structural information at low resolution of samples in solution, thereby bypassing the need to grow crystals. Although, crystal structures allow for high resolution structural data, they are obtained in an environment where the molecules of interest are highly concentrated, and the conditions are not physiological. This often leads to crystallographic artifacts. Furthermore, the crystal is in a locked conformation prohibiting molecules from moving. SAXS allows us to monitor complexes in dynamic equilibrium, in optimal experimental conditions; samples do not need to be labeled or altered, and data can be collected at low temperatures (Tuukkanen and Svergun, 2014). Collecting data at a low temperature is important to ensure that our DNA duplex remains double stranded.

In our SAXS experiment, we incubated the protein with a 23-bp duplex DNA substrate (with the same sequence used for our DNA binding assays) at different ratios to assess if siHF could simultaneously bind two DNA molecules. We first determined the maximum particle diameter (D_{\max}) and experimental molecular weights of the siHF protein and the DNA duplex individually. The molecular weights determined experimentally are very close to the calculated molecular weights of the protein and DNA alone (Table 3.3).

We first incubated the protein and DNA at a 1:1 ration (siHF:DNA) to ensure that the protein forms a complex with DNA in solution in the experimental conditions used. The D_{\max} for the complex (94.5 Å) is larger than that observed for the protein (66 Å) and DNA (71.5 Å) alone, suggesting that the complex is able to form. Furthermore, the molecular weight determined from the experiment (26.9 kDa) is in good agreement with the estimated molecular weight of

the complex (27.8 kDa), confirming that the complex forms at a 1:1 ratio in this condition (Table 3.3).

Based on the crystal structures determined in this work, we predicted that more than one sIHF molecule could bind the 23-bp DNA substrate. We tested our prediction using SAXS by incubating the protein at a 2:1 ratio (sIHF:DNA) with duplex DNA prior to data collection. The D_{\max} as well as the molecular weights determined increased (103 Å and 46 kDa), indicating that two protein monomers are bound to the 23-bp duplex substrate (Table 3.3).

	sIHF	DNA	(sIHF:DNA) 1:1	(sIHF:DNA) 2:1	(sIHF:DNA) 1:2
Data collection					
Instrument	BioSAXS-1000	BioSAXS-1000	BioSAXS-1000	BioSAXS-1000	BioSAXS-1000
X-ray generator	MicroMax-007HF	MicroMax-007HF	MicroMax-007HF	MicroMax-007HF	MicroMax-007HF
Beam geometry (mm)	0.5	0.5	0.5	0.5	0.5
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418	1.5418
Exposure time (min)	20	180	20	20	20
Concentration (µM)	434	109	109	109	217
Structural parameters					
Io	0.24	0.2	0.6	0.6	0.6
Io/conc	0.5	1.8	1.4	5.6	1.4
Rg (Å)	21.3	21.9	25.7	29.6	24.7
Dmax (Å)	66	71.5	94.5	103	84
MW ^{Experimental} (kDa)	16.4	10.6	26.9	46	20
MW ^{Calculated} (kDa)	13.7	14	27.8	41.4	41.7

In order to determine if sIHF could bridge DNA, we then incubated the protein at a 1:2 ratio (sIHF:DNA). The experimental molecular weight for this sample is smaller than expected (20 kDa compared to 41.7 kDa). Furthermore, the D_{\max} at 84 Å is between that of the DNA alone and the sIHF:DNA complex at a 1:1 ratio. This is also true for the experimental molecular weight;

the experimental molecular weight determined is between that of the DNA alone and that of the 1:1 protein-DNA complex (Table 3.2). This suggests that sIHF cannot bridge two DNA molecules together, and instead the sample collected is a mixture of sIHF:DNA complex at a 1:1 ratio and free DNA (Table 3.3).

3.7 sIHF affects TopA activity independently of DNA binding

Since sIHF does not bend nor bridge DNA; we next explored if it compacts the nucleoid indirectly by modulating the activity of other enzymes that affect chromosome compaction. The H2TH fold has been found in topoisomerase VI and, thus, we thought that sIHF could potentially modulate the activity of one of the topoisomerases found in *S. coelicolor*. We decided to test if sIHF alters the activity of the sole type I topoisomerase expressed in *Streptomyces*; TopA.

We cloned and purified TopA and compared its ability to relax supercoiled plasmids in the absence or presence of sIHF. We found that sIHF inhibits TopA activity *in vitro* as the formation of relaxed plasmids by TopA is reduced in the presence of sIHF (Swiercz et al., 2013). As a control, we performed the same experiment replacing sIHF with a different protein of similar molecular weight (lysozyme) and found that only sIHF was able to inhibit the activity of TopA (Figure 3.8). We later found that this effect is not due to a direct interaction between sIHF and TopA as the two proteins do not interact using gel filtration chromatography (Figure 3.9). If sIHF were to form a complex with TopA we would expect sIHF to co-elute with TopA; however, even when we incubate sIHF with a 10 x excess over TopA, sIHF does not form a complex with TopA (Figure 3.9C). Interestingly, high sIHF concentrations alter the migration of TopA (Figure 3.9B). Since sIHF does not directly interact with TopA, we reasoned that binding of sIHF to DNA

either renders DNA inaccessible to TopA, or causes a conformational change that inhibits TopA activity.

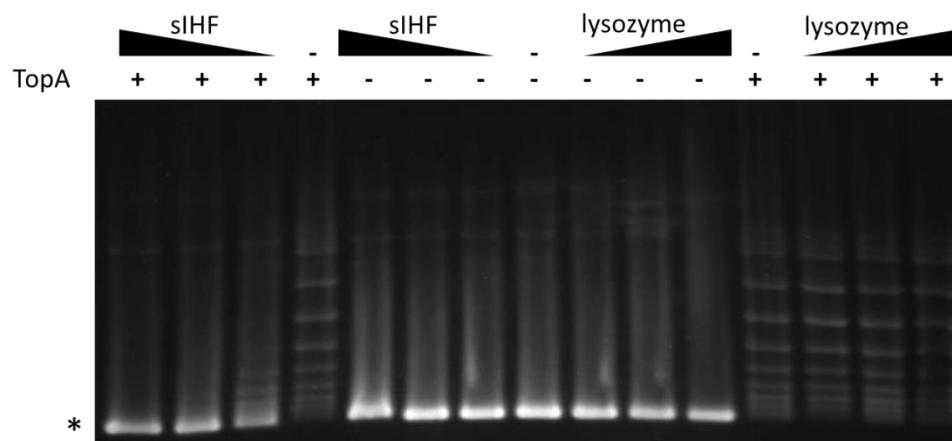


Figure 3.8. sIHF affects the activity of TopA. Agarose gel of sIHF or lysozyme incubated with supercoiled pUC19 plasmid DNA and the *Streptomyces coelicolor* TopA. Protein concentrations tested range from 135-540 molar excess over DNA. The migration of supercoiled plasmid DNA is indicated by the asterisk while bands that migrate slower indicate plasmids relaxed by TopA (Figure adapted from Swiercz et al., 2013).

In order to determine if sIHF is simply coating DNA, we assessed the ability of TopA to relax DNA when incubated with different sIHF variants. We found that sIHF-RR85AS and sIHF-RR85AS+NQ93AS no longer inhibit TopA activity, whereas sIHF-NQ93AS is still able to inhibit TopA activity; albeit, to a lesser extent than wildtype (Figure 3.10A). Therefore, the integrity of Interface I and, in particular residues Arg85 and Arg86 are important for this role of sIHF. Conversely, the sIHF-G66+ variant inhibits TopA activity similar to wildtype sIHF (Figure 3.10B) even though its affinity for DNA binding is reduced compared to wildtype (Figure 3.6C). This illustrates that Interface I is important for sIHF's role in inhibiting TopA activity and DNA binding, whereas Interface II is only implicated in binding DNA.

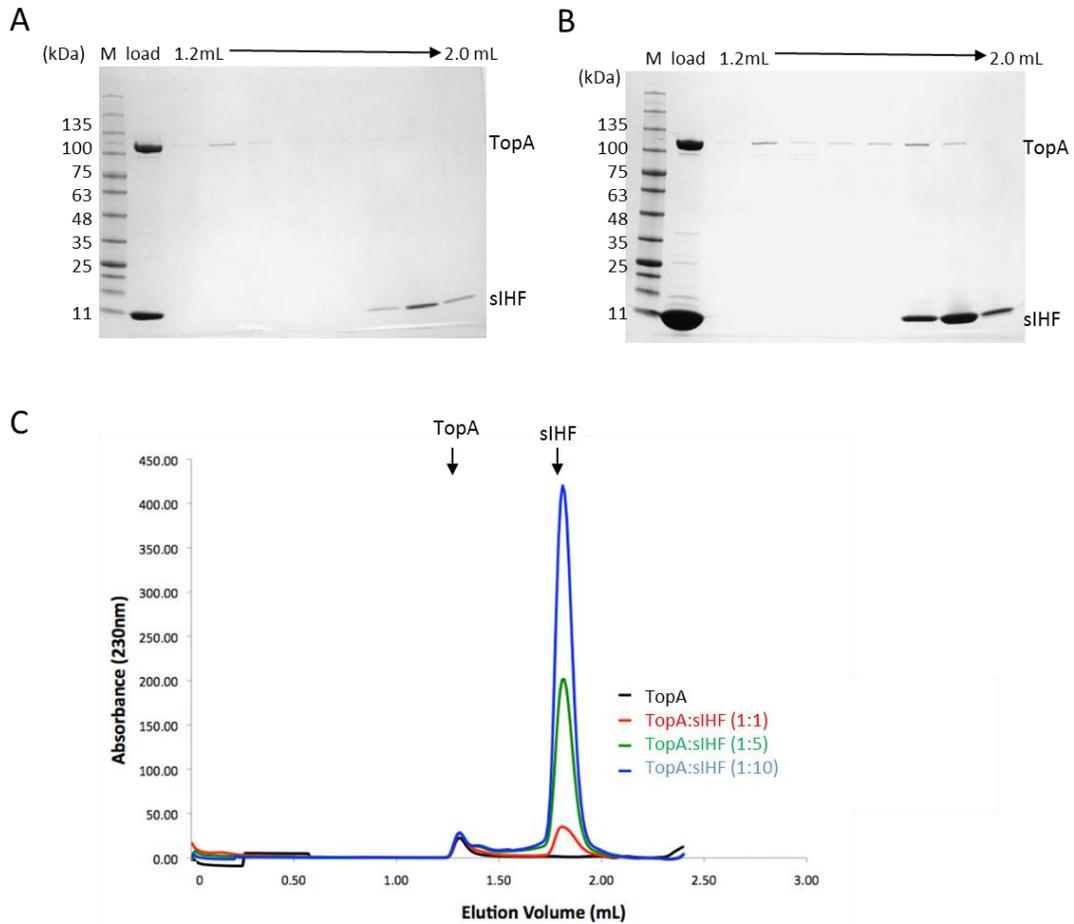


Figure 3.9. Gel filtration assay of sIHF with TopA. SDS-PAGE and chromatogram of fractions eluted from the gel filtration column of sIHF incubated with TopA. 0.1 mL fractions were collected and loaded onto a 4-15% precast gel (Bio-Rad). Migration of TopA and sIHF are as indicated. sIHF and TopA were incubated at various ratios. Panel (A) displays fractions collected with TopA:sIHF at a 1:1 ratio, while panel (B) displays fractions collected from TopA:sIHF at a 1:10 ratio.

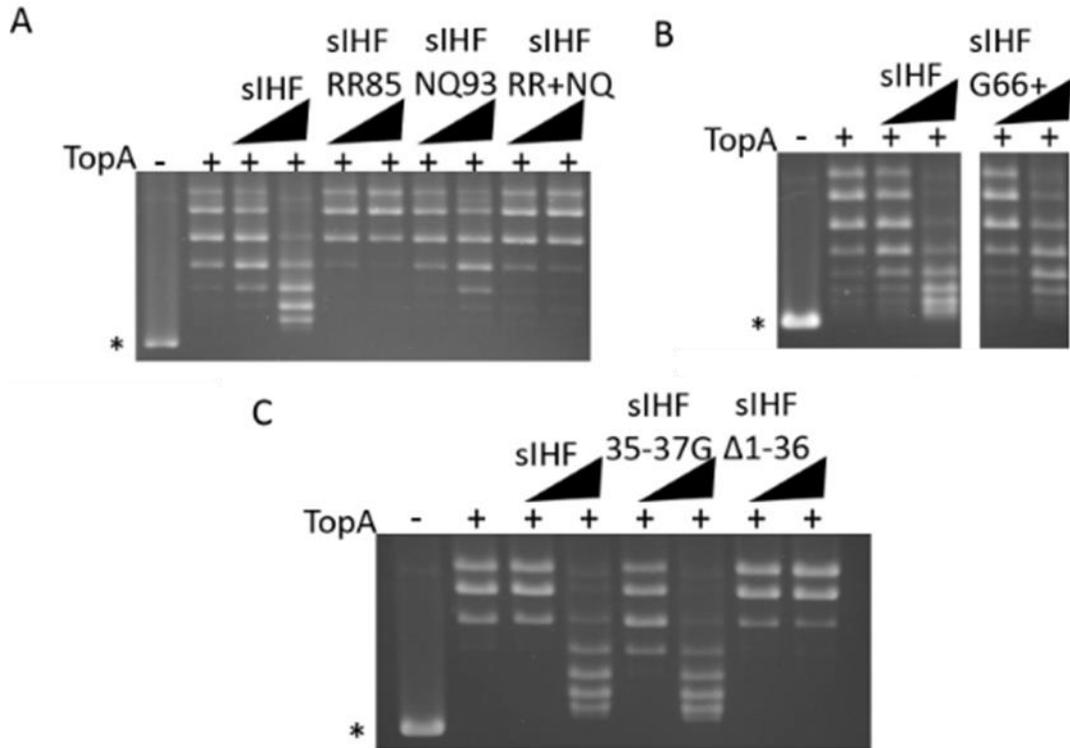


Figure 3.10. Topoisomerase assays of sIHF mutants. Topoisomerase assay, using TopA (7.73 μM) from *S. coelicolor*, in the presence of sIHF variants to test **(A)** Interface I; wildtype sIHF, sIHF RR85AS (sIHF RR85), sIHF NQ93AS (sIHF NQ93) and sIHF RR85AS+NQ93AS (sIHF RR+NQ) **(B)** Interface II; wildtype sIHF and sIHF G66+ **(C)** the N-terminal helix; wildtype sIHF, sIHF 35-37Gly (sIHF 35-37G) and sIHF Δ 1-36 with plasmid DNA (pUC19) at sIHF concentrations of 0, 0.54, and 2.16 μM . The asterisks indicate supercoiled DNA, while DNA bands migrating slower than the supercoiled plasmid indicate relaxed topoisomers.

Next we tested whether the N-terminal helix of sIHF is important to modulate the activity of TopA. We observed that a variant of sIHF lacking the N-terminal helix (sIHF Δ 1-36) was no longer able to inhibit the activity of TopA whereas sIHF 35-37Gly inhibited TopA activity to a similar extent to wildtype (Figure 3.8C). These results reinforce the idea that the 35-37Gly variant of sIHF does not unlatch the N-terminal helix from the H2TH domain, and that the N-terminal helix is relevant for the function of sIHF.

Chapter 4

Discussion

In this work, we have refined three crystal structures of sIHF bound to various duplex DNA substrates, and have identified three distinct surfaces on the protein that contact DNA. These three surfaces have been confirmed to bind DNA in solution *in vitro* (Figure 3.6). The first interface, Interface I, uses the lid region of the protein (Figures 3.3, 3.4A, and 3.5A in pink) and contacts the DNA phosphate backbone through residues Arg85, Asn93, and Gln94 (Figure 3.4C, 3.5B). Arg86 is in close proximity to the DNA duplex, and could potentially aid in DNA binding (Figure 3.4B). Based on *in vitro* DNA binding data, residues Arg85 and/or Arg86 are important for the interaction between sIHF and DNA (Figure 3.6A).

Interface I interacts with DNA within the asymmetric unit of the structures refined from crystals TN051 and TN103; however, sIHF also contacts DNA at a second interface, Interface II, through symmetry. Interface II, contacts DNA solely through the peptide backbone using the H2TH domain (Figures 3.4 and 3.5). This domain is also found in other proteins that interact with nucleic acids (Figure 3.3 C-E in blue) and is thought to aid in peripheral DNA binding. The lid region of Interface I is also found amongst these proteins (Figure 3.3 C-E in pink) and is more exposed than the H2TH domain in their crystal structures (Corbett and Berger, 2003; Zharkov et al., 2002; Brodersen et al., 2002). Therefore, the contribution of these two regions for DNA binding must be explored further.

A third interface of sIHF, the N-terminal helix, has also been shown to contribute to DNA binding (Figures 3.4A and 3.5A in yellow). Residues involved in this interaction, identified from

the crystal structure of sIHF bound to a hairpin DNA substrate (crystal TN131), include Arg22, Asn30, and His34 (Figure 3.5D). Crystals TN051 and TN103 failed to capture the interaction between this interface and DNA. This could be because the DNA in these crystals adopts a pseudo-continuous helix, and favours contacts at Interface I and Interface II. The N-terminal helix is approximately one third of the protein as it contains 34 of 107 residues of the protein. The crystal structures of sIHF bound to TN200 (from crystal TN051) and sIHF bound to TN08-BP1 (from crystal TN103) successfully identified residues 13-103 and 15-103, respectively, in their experimental electron density maps. However, the N-terminal residues are missing from our model. It is common for the ends of proteins and flexible regions to be unidentified in electron density maps as they do not adopt the same conformation in each repeating unit across the crystal. In the crystal, perhaps the N-terminal residues are flexible and interact with duplex DNA on either side, causing these residues to be unidentified in the crystal structure (Figure 4.1). This would explain why these residues are disordered even if they interact with DNA.

To test the relevance of the interaction between the N-terminal helix of sIHF and DNA, we removed the entire N-terminal helix, which may have caused the protein to become unstable. Assessed by dynamic light scattering (DLS), the Δ 1-36 sIHF protein is stable at the concentration tested (see Methods), but the protein may have become unstable upon dilution. Therefore we should repeat our DNA binding and topoisomerase activity assays with a less severe mutation. Instead of removing the entire helix, we could remove the first 13 residues (sIHF Δ 1-13) which are disordered in the crystal structures, or mutate the residues believed to facilitate the interaction between DNA and sIHF at this interface (Arg22, Asn30 and His34).

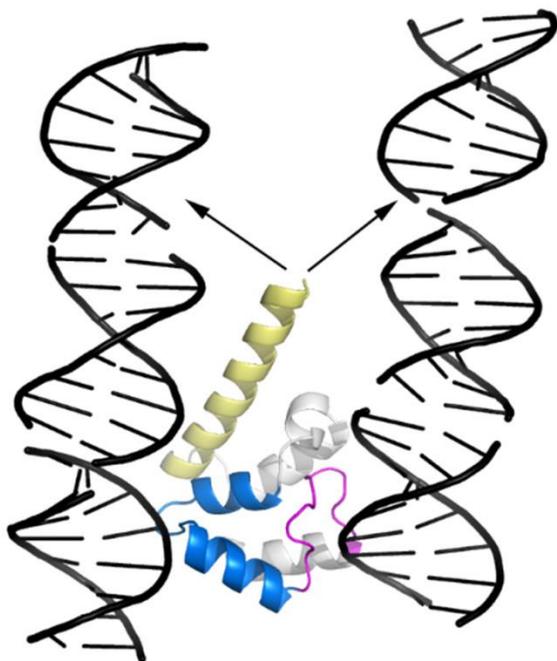


Figure 4.1. Ribbon structure of sIHF bound to TN08-BP1 and possible interactions with the N-terminal helix. sIHF residues 15-103 with TN08-BP1. The 14 residues missing from this model may interact with DNA, but are not identified as they are flexible and may interact with either duplex forming the pseudo-continuous helix (indicated by the arrows).

Once we verified that multiple surfaces of sIHF contact DNA, we were curious to see if sIHF could simultaneously bind DNA at these interfaces to bring together two disparate DNA molecules. Common NAPs that bridge DNA, such as H-NS and Lsr2, are composed of a DNA binding domain and an oligomerization domain. They function as a homodimer and bring together distant DNA molecules by binding DNA through their DNA binding domain, and subsequently forming a homodimer with another H-NS/Lsr2 molecule bound to DNA (Dillon and Dorman, 2010). Conversely, sIHF functions as a monomer and would need to simultaneously bind multiple DNA substrates, with one sIHF molecule, to bridge DNA. Based on SAXS data collected of sIHF incubated with DNA at various ratios, we did not observe bridging of DNA by sIHF (Table 3.3). However, sIHF may still use multiple interfaces to compact DNA and neutralize the charge repulsion between DNA molecules *in vivo*.

To further investigate sIHF's role in chromosome organization, we wanted to explore if sIHF affected the structure of DNA at the local level. We decided to crystallize sIHF in complex with a DNA palindrome of known structure to analyze the conformational changes induced by sIHF upon binding. We found differences in DNA base-pair, and base-pair step parameters between duplex DNA with and without sIHF (Table 3.2). Collectively, this caused the minor groove of the DNA to widen from 8.3 Å to 13.5 Å, determined by comparing the structures of identical DNA sequences alone (by solution NMR) and in complex with sIHF (by X-ray crystallography). Furthermore, when we analyzed the structure of the DNA in the crystal structure of sIHF bound to the hairpin DNA substrate, we observed base-pair and base-pair step parameters that deviated from standard B-DNA (Table 3.2). In addition, the minor groove has a similar width to that in crystal TN103 (sIHF bound to TN08-BP1) at 12.9 Å, illustrating that sIHF may widen the minor groove of DNA upon binding.

The crystal structures presented in this work, show that sIHF interacts with DNA through the minor groove (Figure 3.4 and 3.5). sIHF is a small positively charged protein, and may prefer binding DNA through the minor groove as the negative electrostatic potential is increased due to the relative concentration of negative charge from the phosphate backbone (Rohs et al., 2009). It is also worth noting that SELEX assays conducted by Julia Swiercz (Swiercz, 2013) identified three motifs that sIHF had preferred binding, all of these had under 50% GC content and at least three A-T consecutive base-pairs. sIHF may prefer binding AT-rich sites as they contain narrow minor grooves (Gavathiotis et al., 2000, Oguey et al., 2010) where the negative electrostatic potential is enhanced even more (Rohs et al., 2009). This suggests that perhaps sIHF prefers binding to DNA substrates with narrow minor grooves, and upon binding widens the groove.

Thus, sIHF may prefer binding certain DNA structures (narrow minor grooves), rather than particular sequences.

In addition to altering the conformation of DNA at the local level, sIHF may affect chromosome organization *in vivo* through mediating the activity of other enzymes that alter DNA topology, such as topoisomerases (Dean and Cozarelli, 1985). We show that sIHF inhibits the activity of the sole type I topoisomerase in *S. coelicolor*, ScTopA *in vitro* (Figures 3.8) (Swiercz et al., 2013); where ScTopA has been shown to function as a processive enzyme (Szafran et al., 2014). We have found a separation of function sIHF mutant, whereby Interface II (tested using sIHF-G66+) is important for DNA binding but does not affect sIHF's ability to mediate TopA activity. sIHF variants where Interface I is disrupted have reduced DNA binding and have lost their ability to affect TopA activity (Figures 3.6 and 3.10). Interface I, in conjunction with the N-terminal helix, form a continuous positively charged patch. This network of residues (Arg22, Asn30, Arg85, Arg86, Asn93 and Gln94) would allow the DNA duplex to form extensive contacts with the DNA phosphate backbone. In this way, Interface I and the N-terminal helix may bind DNA over a larger surface area stalling TopA from relaxing supercoiled DNA. Interface II covers less surface area, and sIHF may be displaced from the DNA in the presence of TopA.

Since sIHF can bind DNA through multiple interfaces there is the possibility that each interface binds DNA with a different affinity and stability. We propose that sIHF may bind more stably at Interface I and effectively stall the processive TopA enzyme (Figure 4.2A), while the interaction between DNA at Interface II may be less stable allowing for TopA to progress (Figure 4.2B). The difference in stability for various interfaces of sIHF is consistent with having a dynamic nucleoid. sIHF is expressed throughout the cell cycle (Swiercz et al., 2013); if it were to

bind DNA with high affinity at multiple surfaces, the nucleoid would be constrained and not dynamic. The interaction between sIHF Interface II and DNA being less stable would allow for the nucleoid to remain dynamic. Interface I in combination with the positive patch along the N-terminal helix may form a more stable complex than Interface II, allowing for sIHF complexes that bind DNA through Interface I to effectively stall TopA progression, while sIHF Interface II and DNA may form a more transient complex which will allow TopA to continue to relax supercoiled DNA. In this way, sIHF may use multiple interfaces to compact the bacterial nucleoid and affect gene expression; while only Interface I can stall TopA and inhibit chromosome relaxation.

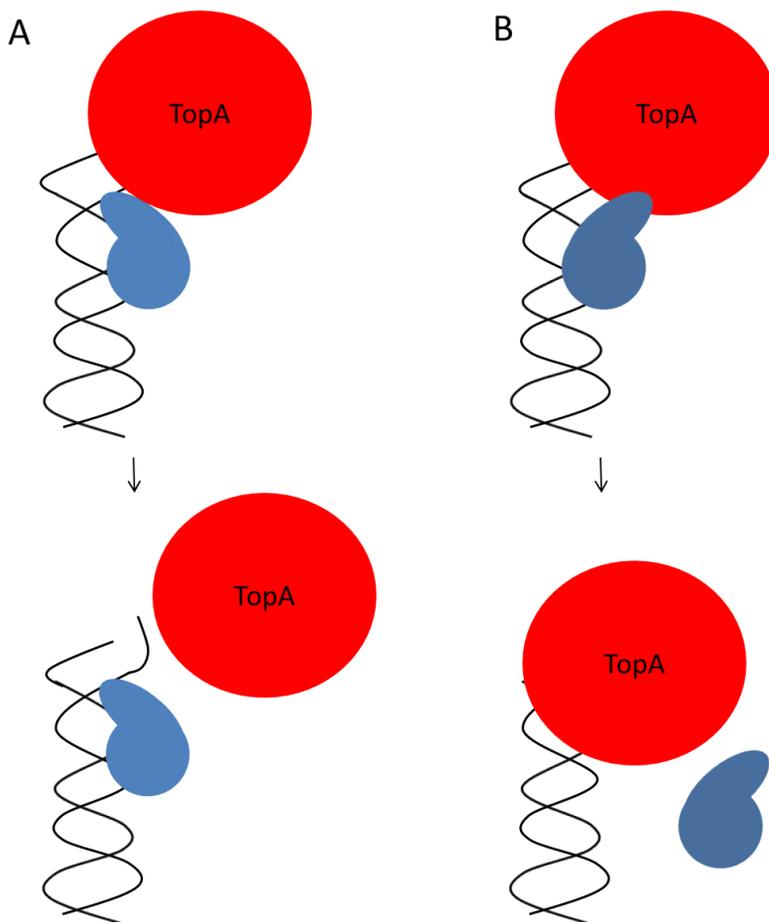


Figure 4.2. Proposed mechanism of action of sIHF. **(A)** sIHF (blue) bound to DNA using Interface I and the N-terminal helix binds strongly to DNA inhibiting the progression of the TopA (red) enzyme from further relaxing duplex DNA. **(B)** sIHF bound to DNA at Interface II is weak; the interaction between sIHF and DNA cannot be maintained in the presence of TopA which is able to progress along the duplex.

In order to confirm our model, follow up studies must be conducted. First, to determine if Interface I of sIHF with the positive patch along the N-terminus forms a more stable complex with DNA than Interface II, we should compare the stability of the complexes of sIHF, sIHF RR85AS and sIHF G66+ with duplex DNA. This can be done using gel filtration chromatography, isothermal titration calorimetry and surface plasmon resonance. If our model is correct, sIHF G66+ should form a more stable complex with DNA than sIHF RR85AS as Interface I is still intact in the former mutant.

Our model is based on the idea that the N-terminal helix and Interface I both simultaneously contribute to DNA binding, allowing for a strong complex to form as the interaction surface area increases when these two surfaces are combined. If this is true, the positively charged patch between the N-terminal helix and 4-helix bundle composed of residues Arg22, Arg25, Lys29 and Lys33 should contribute to DNA binding. In order to test this, we will mutate these residues to serine residues and assess this variant (*sIHF-R22, R25, K29, K33(S)*) for DNA binding. This sIHF variant can also be assessed for its role on TopA activity.

The assays discussed above have been conducted *in vitro*. In order to assess if the interfaces identified in this study are important for wildtype function *in vivo*, with the help of our collaborators, we would like to complement Δ *sIHF* strains with *sIHF-RR85AS*, *sIHF-G66+*, and *sIHF-R22, R25, K29, K33(S)* to observe their phenotype. This will allow us to assess the various roles of each region of the protein on chromosome condensation, gene expression and sporulation.

Lastly, we have obtained several crystal structures of sIHF bound to DNA; however, we believe it would be beneficial to determine the crystal structure of sIHF bound to the DNA

sequence determined using SELEX, at a 2:1 ratio (slHF:DNA). This structure would allow us to investigate whether slHF coats DNA and, if the A-tract and associated narrow minor groove aid in slHF binding. By conducting this experiment at progressively higher protein:DNA ratios, we will be able to assess if slHF binds DNA cooperatively.

Chapter 5

Conclusion

We have shown that sIHF is a novel NAP that is expressed throughout the cell cycle, binds DNA in a sequence independent manner, associates with the nucleoid, plays a role in chromosome condensation, when absent leads to aberrant levels of gene expression, and affects the activity of TopA (Swiercz et al., 2013). Following this work, we have solved crystal structures of sIHF bound to DNA and confirmed that sIHF binds DNA at multiple interfaces, yet does not bridge DNA molecules together. We have also discovered an interface of sIHF that contributes to DNA binding, but does not affect the activity of TopA. With these findings, we propose a model for the mechanism of action of sIHF. sIHF contributes to chromosome organization at two levels. The first uses all DNA binding surfaces of the positively charged sIHF protein, to neutralize the charge repulsion from the negatively charged DNA phosphate backbone. This aids in chromosome condensation. The second uses Interface I and the positively charged patch along the N-terminal helix to stall ScTopA from relaxing DNA supercoils. Further studies must be conducted to confirm this model.

References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica D Biological Crystallography* 68:352-367.
- Almirón, M., Link, A. J., Furlong, D., and Kolter, R. (1992). A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes and Development* 6:2646-2654.
- Atlung, T., and Ingmer, H. (1997). H-NS: a modulator of environmentally regulated gene expression. *Molecular Microbiology* 24:7-17.
- Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *Journal of Bacteriology* 181:6361-6370.
- Baleja J, D., Germann, M.W., van de Sande, J. H., and Sykes, B. D. (1990). Solution conformation of purine-pyrimidine DNA octamers using nuclear magnetic resonance, restrained molecular dynamics and NOE-based refinement. *Journal of Molecular Biology* 215:411-428.
- Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A. Fraser, A. Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A (2002). Complete genome sequence of the model actinomycete *Streptomyces Coelicolor* A3(2). *Nature* 417:141-147.
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S., Srinivasan, R., and Schneider, B. (1992). A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal* 63:751-759.
- Bibb, M. J., Ward, J. M., Kieser, T., Choen, S. N., and Hopwood, D. A. (1981). Excision of chromosomal DNA sequences from *Streptomyces coelicolor* forms a novel family of plasmids detectable in *Streptomyces lividans*. *Molecular Genetics and Genomics* 184:230-240.
- Bloch, V., Yang, Y., Margeat, E., Chavanieu, A., Augé, M. T., Robert, B., Arnold, S., Rimsky, S., and Kochoyan, M. (2003). The H-NS dimerization domain defines a new fold contributing to DNA recognition. *Nature Structural Biology* 10:212-218.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72:248-254.

- Brodersen, D. E., Clemons, W. M. Jr., Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. *Journal of Molecular Biology* 316:725-768.
- Champoux, J.J. (2001). DNA topoisomerases: structure, function and mechanism. *Annual Review of Biochemistry*. 70:369-413.
- Chandrasekaran, R., and Arnott, S. (1996). The structure of B-DNA in oriented fibers. *Journal of Biomolecular Structure and Dynamics* 13:1015-1027.
- Chen, J. M., Ren, H., Shaw, J. E., Wang, Y. J., Li, M., Leung, A. S., Tran, V., Berbenetz, N. M., Kocíncová, D., Yip, C. M., Reyrat, J., and Liu, J. (2008). Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Research* 36:2123-2135.
- Chen, N., Zinchenko, A. A., Yoshikawa, Y., Araki, S., Adachi, S., Yamazoe, M., Hiraga, S., and Yoshikawa, K. (2008). ATP-induced shrinkage of DNA with MukB protein and the MukBEF complex of *Escherichia coli*. *Journal of Bacteriology* 190:731-737.
- Corbette, K. D., and Berger, J. M. (2003). Structure of the topoisomerase VI-B subunit: implications for the type II topoisomerase mechanism and evolution. *The EMBO Journal* 22:151-163.
- Cui, Y., Wang, Q., Stormo, G. D., and Calvo, J.M. (1995). A consensus sequence for binding of Lrp to DNA. *Journal of Bacteriology* 177:4872-4880.
- Dame, R. T., Wyman, C., and Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Research* 28:3504-3510.
- Dame, R. T., Wyman, C., and Goosen, N. (2001). Structural basis for preferential binding of H-NS to curved DNA. *Biochimie* 83:231-234.
- Dame, R. T., and Goosen, N. (2002). HU: promoting or countering DNA compaction? *FEBS Letters* 529:151-156.
- Dame, R. T., Luijsterburg, M. S., Krin, E., Bertin, P. N., Wagner, R., and Wuite, G. J. L. (2005). DNA bridging: a property shared among H-NS-like proteins. *Journal of Bacteriology* 187:1845-1848.
- Dean, F. B., and Cozzarelli, N. R. (1985). Mechanism of strand passage by *Escherichia coli* topoisomerase I. The role of the required nick in catenation and knotting of duplex DNA. *Journal of Biological Chemistry* 260:4984-4994.
- Dillon, S. C., and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology* 8:185-195.

- Dorman, C. J. (2013). Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription. *Biochemical Society Transactions* 41:542-547.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica D Biological Crystallography* 60:2126-2132.
- Frenkiel-Krispin, D., Ben-Avraham, I., Englander, J., Shimoni, E., Wolf, S. G., and Minsky, A. (2004). Nucleoid restructuring in stationary-state bacteria. *Molecular Microbiology* 51:395-405.
- Funnell, B. E., Baker, T. A., and Kornberg, A. (1986). Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. *Journal of Biological Chemistry* 261:5616-5624.
- Gavathiotis, E., Sharman, G. J., and Searle, M. S. (2000). Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d(CTTTTGCAAAG)(2). *Nucleic Acids Research* 28:728-735.
- Gitai, Z., Thanbichler, M., and Shapiro, L. (2005) The choreographed dynamics of bacterial chromosomes. *Trends in Microbiology*. 13:221-228.
- Glatter, O., and Kratky., O. (1982) *Small-angle X-ray Scattering* (London, Academic Press).
- Gloyd, M., Ghirlando, R. and Guarné, A. (2011). The role of MukE in assembling a functional MukBEF Complex. *Journal of Molecular Biology* 412:578-590.
- Goodrich, J. A., Schwartz, M. L., McClure, W. R. (1990). Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). *Nucleic Acids Research* 18:4993-5000.
- Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998). The crystal structure of DNA, a ferritin homologue that binds and protects DNA. *Nature Structure* 5:294-303.
- Guinier, A., and Fournet, G. (1995). *Small-angle Scattering of X-rays* (New York: John Wiley & Sons, Inc).
- Hancock, R. (2004). A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *Journal of Structural Biology* 146:281-290.
- Kleine Borgmann, L. A., Ries, J., Ewers, H., Ulbrich, M. H., and Graumann, P. L. (2013). The bacterial SMC complex displays two distinct modes of interaction with the chromosome. *Cell Reports* 3:1483-1492.
- Lim, H. M., Lewis, D. E., Lee, H. J., Liu, M., and Adhya, S. (2003). Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry* 42:10718-10725.

- Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes and Development* 19:1269-1287.
- Luger, K. (2006). Dynamic nucleosomes. *Chromosome Research* 14:5-16.
- Luijsterburg, M. S., Noon, M. C., Wuite, G. J. L., and Dame, R. T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *Journal of Structural Biology* 156:262-272.
- Luijsterburg, M. S., White, M. F., van Driel, R., and Dame, R. T. (2008). The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology* 43:393-418.
- Marenduzzo, D., Finan, K., and Cook, P. R. (2006). The depletion attraction: an underappreciated force driving cellular organization. *Journal of Cell Biology* 175:681-686.
- Myers, K. S., Yan, H., Ong, I. M., Chung, D., Liang, K., Tran, F., Keleş, S., Landick, R., and Kiley, P. J. (2013). Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. *PLoS Genetics* 9:e1003565.
- Nash, H. A. (1990). Bending and supercoiling of DNA at the attachment site of bacteriophage lambda. *Trends in Biochemical Sciences* 15:222-227.
- Newton-Foot, M., and Gey van Pittius, N. C. (2013). The complex architecture of mycobacterial promoters. *Tuberculosis* 93:60-74.
- Oberto, J., Drlica, K., and Rouvière-Yaniv, J. (1994). Histones, HMG, HU, IHF: Môme combat. *Biochimie* 76:901-908.
- Oguey, C., Foloppe, N., and Hartmann, B. (2010). Understanding the sequence-dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One* 29:e15931.
- Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C., Westhof, E., Wolberger, C., and Berman, H. M. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology* 313:229-237.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* 276:307-326.
- Postow, L., Hardy, C. D., Arsuaga, J., and Cozzarelli, N. R. (2004). Topological domain structure of the *Escherichia coli* chromosome. *Genes and Development* 18:1766-1779.

Rimsky, S., and Travers, A. (2011). Pervasive regulation of nucleoid structure and function by nucleoid-associated proteins. *Current Opinion in Microbiology* 12:136-141.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature* 461: 1248–1253.

Rouvière-Yaniv, J., Yaniv, M., and Germond, J. (1979). *E. coli* DNA binding protein HU forms nucleosome-like structure with circular double-stranded DNA. *Cell* 17:265-274.

Schneider, R., Lurz, R., Lüder, G., Tolksdorf, C., Travers, A., and Muskhelishvili, G. (2001). An architectural role of the *Escherichia coli* chromatin protein FIS in organizing DNA. *Nucleic Acids Research* 29:5107-5114.

Schneider, R., Travers, A., Kutateladze, T., and Muskhelishvili, G. (1999). A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Molecular Microbiology* 34:953-964.

Shindo, H., Iwaki, T., Ieda, R., Kurumizaka, H., Ueguchi, C., Mizuno, T., Morikawa, S., Nakamura, H., and Kuboniwa, H. (1995). Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS from *Escherichia coli*. *FEBS Letters* 360:125-131.

Shindo, H., Ohnuki, A., Ginba, h., Katoh, E., Ueguchi, C. Mizuno, T., and Yamazaki, T. (1999). Identification of the DNA binding surface of H-NS protein from *Escherichia coli* by heteronuclear NMR spectroscopy. *FEBS Letters* 455:63-69.

Sinden, R. R., and Pettijohn, D. E. (1981). Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proceedings of the National Academy of Sciences* 78:224-228.

Spurio, R., Dürrenberger, M., Falconi, M., La Teana, A., Pon, C. L., and Gualerzi, C. O. (1992). Lethal overproduction of the *Escherichia coli* nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Molecular Genetics and Genetics* 231:201-211.

Swiercz, J. P. (2013). sIHF IS A NOVEL NUCLEOID-ASSOCIATED PROTEIN SPECIFIC TO THE ACTINOBACTERIA. *Open Access Dissertations and Theses*. Paper 8293.

Swiercz, J. P., and Elliot, M. A. (2012) *Bacterial Spores: Current Research and Applications* Edited by: Ernesto Abel-Santos pp 39-50.

Swiercz, J. P., Nanji, T., Gloyd, M., Guarné, A., and Elliot, M. A. (2013). A novel nucleoid-associated protein specific to the actinobacteria. *Nucleic Acids Research* 41:4171-4184.

Swinger, K. K., and Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. *Current Opinion in Structural Biology* 14:28-35.

Swinger, K. K., Lemberg, K. M., Zhang, Y., and Rice, P. A. (2003). Flexible DNA bending in HU-DNA cocrystal structures. *The EMBO Journal* 22:3749-3760.

Szafran, M.J., Strick, T., Strzałka, A., Zakrzewska-Czerwińska, J., and Jakimowicz, D. (2014). A highly processive topoisomerase I: studies at the single-molecule level. *Nucleic Acids Research* 42:7935-7946.

Terwilliger, T. C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallographica D* 55:849-861.

Thanbichler, M., Viollier, P. H., and Shapiro, L. (2005a). The structure and function of the bacterial chromosome. *Current Opinion in Genetics and Development* 15:153-162.

Thanbichler, M., Wang, S. C., and Shapiro, L. (2005b). The bacterial nucleoid: a highly organized and dynamic structure. *Journal of Cellular Biochemistry* 96:506-521.

Travers, A., and Muskhelishvili, G. (2005). Bacterial chromatin. *Current Opinion in Genetics and Development* 15:507-514.

Tuukkanen, A. T., and Svergun, D. I. (2014). Weak protein-ligand interactions studied by small-angle X-ray scattering. *FEBS Journal* 281:1974-1987.

Wang, J. C. (1996). DNA topoisomerases. *Annual Review in Biochemistry* 65:635-692.

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P., and Shoham, G. (2002). Structural analysis of *Escherichia coli* endonuclease VIII covalent reaction intermediate. *The EMBO Journal* 21:789-800.

Zheng, G., Lu, X. J., and Olsen, W. K. (2009). Web 3DNA—a web server for the analysis, reconstruction, and visualization of three dimensional nucleic-acid structures. *Nucleic Acids Research* 37:W240-6.

Zhu, Q. and Wani, A. A. (2010). Histone modifications: crucial elements for damage response and chromatin restoration. *Journal of Cellular Physiology* 223:283-288.