Cooperative Assembly of Terminase and Integration Host Factor at the Packaging Initiation Site of Bacteriophage Lambda

Saurarshi J. Sanyal

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Carlos E. Catalano, Chair

Kelly K. Lee

William M. Atkins

Program Authorized to Offer Degree:

Medicinal Chemistry

University of Washington

Abstract

Cooperative Assembly of Terminase and Integration Host Factor at the Packaging Initiation Site of Bacteriophage Lambda

Saurarshi J. Sanyal

Chair of the Supervisory Committee:

Carlos E. Catalano, Professor

Department of Medicinal Chemistry

Packaging of viral genomes into procapsids by terminase enzymes is conserved in many DNA viruses. Terminases bind to linear concatemers of replicated viral genomes and concomitantly excise (mature) and package a single genome per procapsid. In this thesis, I interrogate the role of *E. coli* integration host factor (IHF) in mediating the site-specific assembly of bacteriophage lambda terminase at its cognate DNA site, *cos*, which serves as the packaging initiation site. IHF binds to an I-element within *cos* and introduces a strong bend in the duplex. It was previously demonstrated that the small terminase subunit could stabilize an IHF-induced bend at *cos*. I hypothesized that terminase holoenzyme and IHF cooperatively assemble at *cos* and wrap the duplex into a compact nucleoprotein complex. Rigorous analysis of

this cooperative assembly is complex due to the multiple terminase and IHF binding elements within *cos*. Therefore, I dissected the *cos* site into individual specific and nonspecific IHF binding sequences, and analyzed the relevant protein affinities for these subsites as well as for (1) the full-length *cos* site and (2) a random nonspecific (NS) sequence of equivalent length. Analytical ultracentrifugation and electrophoretic mobility shift studies show that IHF and terminase only modestly discriminate between *cos* and NS-DNA substrates; however, the two proteins cooperatively bind to *cos*-DNA. The data suggest that IHF confers site-specificity of binding to terminase. Also evident is significant nonspecific DNA binding concurrent with specific interactions, even on specific DNA substrates. IHF likely facilitates the high-affinity cooperative assembly of a relevant nucleoprotein complex at the *cos* site despite significant nonspecific binding of both proteins to DNA, with the functional significance of nonspecific DNA binding being the enhancement of protein-DNA interactions. Furthermore, sedimentation equilibrium studies demonstrate that while terminase assembles in the absence of IHF as a dimer on a 274 bp DNA substrate inclusive of the entire *cos* site, in the presence of IHF a nucleoprotein complex of mass consistent with five terminase protomers and two IHF molecules results. This finding further implicates IHF in the cooperative assembly of a specific ternary IHF-DNA-terminase complex at the packaging initiation site of bacteriophage lambda. A terminase packaging enzyme that both (1) site-specifically matures DNA and (2) packages DNA in a sequence-independent manner must be capable of both specific and nonspecific DNA binding. This work furthers the understanding of (1) one of the factors (IHF) involved in the site-

specific assembly of a nucleoprotein complex required for the initiation of viral packaging, and (2) the nature of the specific nucleoprotein complex assembled at the *cos* site prior to DNA maturation and packaging.

Contents

Acknowledgements

I would like to thank my thesis advisor Carlos E. Catalano for taking me on as a graduate student, and for his guidance throughout my graduate work. I consider myself fortunate to have worked with one of the most understanding mentors one could ask for, providing me with tremendous support on both scientific and personal fronts. Carlos is never short of pithy quotes, and, as an experimental scientist, one that I know will stay with me forever goes as follows: “There’s never time to do it right; there’s always time to do it over again.”

I would also like to thank my dissertation reading committee, consisting of Kelly K. Lee and William M. Atkins, as well as my graduate school representative, Charles L. Asbury. Each individual on my committee provided insights into making my thesis, and my scientific training, stronger.

Life in the wet lab can be trying, and I would like to thank all my friends and colleagues, both past and present, in the Department of Medicinal Chemistry. In particular, there was never a dearth of fun and intelligent discourse throughout my graduate career due to the following folks: Benjamin Andrews, Nick Au, Angelo Condulle, Kip Conner, Matthew Honaker, Abhinav Nath, Mimi Nguyen, John Sumida, Caleb Woods, and Jay Yang. Outside of the lab, I have had a family of friends that have made my time in graduate school delightful. In particular, Jonah Cohen, Zaara Kittenchops, Abhishek Kulkarni, Neeti Nundy, Kayur Patel, and Ilya Rosenberg were

unrelenting in their friendship, and irreplaceably influential in the pursuit of my passions.

Finally, graduate school can often be a time of formidable personal growth. I am lucky to have had the unwavering support of my wonderful family, including my father Gautam Sanyal, my mother Kumkum Sanyal, and my sister Sohini Sanyal. My family has always provided me with a strong foundation from which I could explore my individual interests, visions, and goals.

List of Figures

List of Tables

IV: The Assembly State of Terminase at the Packaging Initiation Site of Bacteriophage λ is Affected by Integration Host Factor.

Common Abbreviations Used:

6-FAM: 6-carboxyfluorescein

AUC: Analytical ultracentrifugation

bp: (DNA) base pair

cos: The cohesive end site of the bacteriophage lambda genome. Also serves as the packaging initiation site.

Da: Daltons

kDa: Kilodaltons

PCR: Polymerase chain reaction

Tris: Tris(hydroxymethyl)aminomethane

EDTA: Ethylenediaminetetraacetic acid

NaCl: Sodium chloride

β-me: β-mercaptoethanol

IHF: E. coli integration host factor protein

E. coli*: Escherichia coli*

SE: Sedimentation equilibrium

SV: Sedimentation velocity

Terminase protomer: the baceteriophage lambda terminase complex composed of two gpNu1 (small) and one gpA (large) subunits in a ratio of $gpNu1_2$:gpA_1.

"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

--Sir Arthur Conan Doyle

The Adventures of Sherlock Holmes

"If you don't build your dreams, someone will hire you to help build theirs."

--Tony Gaskin

I. Introduction

Complex double-stranded DNA (dsDNA) viruses, such as adenoviruses, herpesviruses, and poxviruses, as well as the tailed bacteriophages, though different from one another in many respects, share a common assembly pathway aimed at producing new infectious viral particles. In this pathway, single viral genomes are packaged into preformed protein shells known as procapsids. The process of packaging is an essential step in viral replication, and is performed by a carefully tailored packaging machine evolved to recognize and interact with both viral DNA and an empty procapsid. Bacteriophages, that is, any of a number of viruses that infect bacteria, provide a model systems for the study of such viral genome packaging machines. Many dsDNA bacteriophages share a similar genome packaging pathway in which terminase enzymes recognize linear concatemers of viral genomes linked end-to-end and package the genetic material into preformed empty procapsids, cutting the concatemer as necessary to processively package only single genomes into any one procapsid[1]. Since DNA packaging is an essential step in viral replication, there is considerable interest in fully elucidating the mechanism of packaging enzymes and targeting key portions of these pathways in the development of therapies.

Terminases are enzymes involved in translocating viral DNA into procapsids as a part of the packaging process. As such, they are required to specifically recognize viral DNA, potentially amongst a pool of host DNA. Additionally, they possess ATPase activity to power DNA translocation[2]. These two features of terminase

enzymes are, in most cases, attributable to two separate subunits of a hetero-oligomeric complex. A small subunit typically confers DNA recognition capabilities to the terminase enzyme, whereas a larger subunit possesses both translocation ATPase activity as well as endonuclease activity (for phages utilizing concatemeric packaging systems). The larger subunit typically also contains motifs used in the recognition of the portal complex, a doughnut-shaped ring that resides at a unique vertex of the procapsid and serves as the interface between the packaging motor and the procapsid[3]. Given this capability of terminase enzymes to link the products of viral genome replication and procapsid synthesis, there is considerable interest in understanding how these enzymes mediate the specificity of their interactions and activities.

My studies utilize bacteriophage λ as a model system for the study of terminase enzymes. A brief introduction to the λ infection cycle will aid in subsequent discussion of the specific pathways.

The λ Viral Packaging Pathway

A λ virion contains a 48.5 kilobase (kb) linear dsDNA genome packed to liquid crystalline density inside of an icosahedral protein shell. A tail attached to this shell aids in the delivery of λ genome to target *Escherichia coli* (*E. coli*) cells via binding of said tail proteins to surface *E. coli* maltodextrin porin proteins. Upon injection of the genome into the bacterial cell, the linear genome quickly circularizes via 12-base "sticky", or complementary, ends that reside on either side of the genome. Host DNA

ligase seals the nick in the circular duplex, resulting in an intact subsite that is known as the *co*hesive end *s*ite (*cos*) in the λ genome. This particular site becomes important in subsequent discussions of the packaging process, as it functions uniquely as the packaging initiation site.

As a temperate phage, λ may establish lysogeny, a state in which the viral genome integrates into the host chromosome via site-specific recombination. This pathway will not be discussed here. Alternatively, the infecting virus may enter the lytic developmental pathway. Upon formation of a circularized λ genome, as well as expression of λ *O* and *P* genes involved in viral DNA replication[4], host DNA polymerase III initially replicates viral DNA via a bidirectional (θ) mechanism. Bidirectional replication results in daughter circles that undergo further replication or serve as templates for the production of viral proteins such as capsid and tail proteins. At a later stage of infection, rolling circle (σ) replication affords linear concatemers of up to 10 unit length λ genomes linked head-to-tail[5]. The express purpose of this linear concatemer is to serve as a substrate from which monomeric virion DNA may be generated via site-specific endonuclease activity by the packaging machinery; the end-to-end linkage of many such λ genomes within one concatemer ensures that the packaging machinery can proceed in a processive manner, filling many procapsids with individual genomes per one DNA binding event.

Given that the *cos* site is regenerated by circularization of the linear genome upon infection, the linear concatemer that results from σ replication of this circularized genome contains multiple *cos* sites. More specifically, these *cos* sites reside at the junction between λ genomes joined together end-to-end within the concatemer. The job of the packaging machine, then, is (1) to recognize the packaging initiation *cos* site on a linear concatemer and make a site-specific cut to yield a sticky-end at this *cos* site (a process known as 'DNA maturation'; see Figure 1A), (2) to translocate into a procapsid the region of DNA between this *cos* site and the next (one unit length linear chromosome), and, finally, (3) make another site-specific endonuclease cut at the following *cos* site to generate the second, complementary, sticky-end (steps 2 and 3 comprising 'DNA packaging'; see Figure 1B). DNA packaging affords one capsid containing one linear genome with sticky-ends (mature viral DNA). The packaging enzyme, or λ terminase, remains bound to the linear concatemer, ready to processively fill more procapsids.

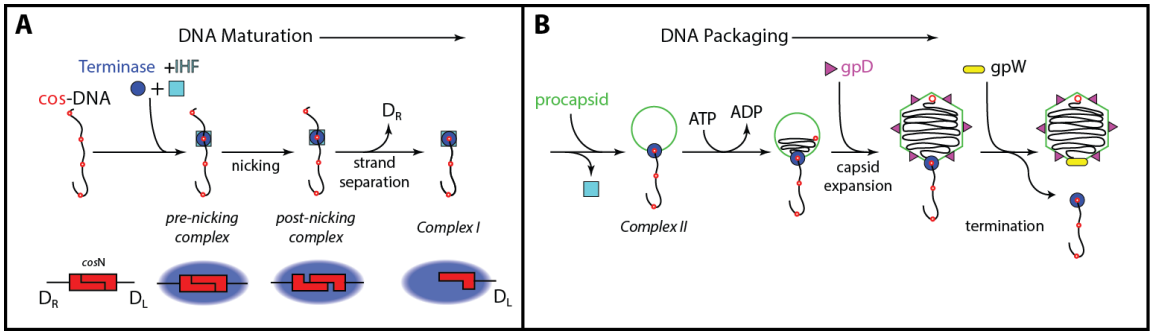


Figure 1. DNA maturation and DNA packaging. *Panel A.* Assembly of a nucleoprotein complex capable of DNA maturation at the packaging initiation site *cos*. *Panel B.* Assembly of a packaging-competent motor complex at the mature *cos* site, progression of the packaging process, and regeneration of a complex I-like nucleoprotein complex.

Terminase Subunit Composition and Interactions at cos

Terminase was named thusly as a result of its presumed role in the 'ter' reaction, an endonucleolytic event, predicted by Mousset and Thomas in 1969, that generates the 12 nucleotide (nt) single-stranded overhangs at the ends of the mature λ genome[6]. Early genetic studies indicated that the gene product of the 'A' gene (gpA) of bacteriophage λ was involved in generating the cohesive ends of mature λ DNA[7]. Combined with evidence that isolated terminase enzyme was able to complement λ(A-) phage-infected cell extracts for packaging λ DNA in *in vitro* experiments, there was strong evidence that gpA itself was, or was a constituent of, the terminase enzyme[8]. The precise composition of the terminase enzyme was not elucidated until Becker and Gold showed that two distinct polypeptides, purified from *E. coli* cells containing modified λ lysogens, could be correlated with the biological activity associated with the 'ter' reaction[8]. The larger 72 kilodaltons (kDa) protein was gpA, and the smaller 20.4 kDa protein was thought to be, and later confirmed to be, gpNu1. Further sedimentation analysis by sucrose gradients and gel filtration analyses led to the determination of a sedimentation coefficient and Stokes radius, respectively, essentially yielding a mass estimate of 117 kDa for the native terminase protomer. This mass was consistent with a composition of one gpA protein and two gpNu1 proteins. A hetero-oligomer of one gpA and two gpNu1 subunits was therefore taken to be the smallest unit of functional terminase and is referred to as *protomer* (Figure 2A). In retrospect, such a hetero-oligomeric composition was to be expected since most packaging terminases studied to date, like λ terminase, herein referred to simply as terminase, are composed of a small

and large subunit that form a hetero-oligomeric holoenzyme complex. The larger (gpA for λ) subunit typically possesses all of the catalytic activities necessary to mature and package a viral genome, including nuclease, helicase, and DNA translocation activities[1]. However, it exhibits largely non-specific DNA binding. The smaller (gpNu1 for λ) subunit confers site-specificity to terminase, not unlike the role transcription factors play in allowing for site-specific binding of RNA polymerase II in eukaryotic transcription. The site-specificity conferred to the large subunit by the small subunit allows the packaging holoenzyme to assemble at the appropriate location on a genomic concatemer in order to initiate packaging (*cos* site for λ).

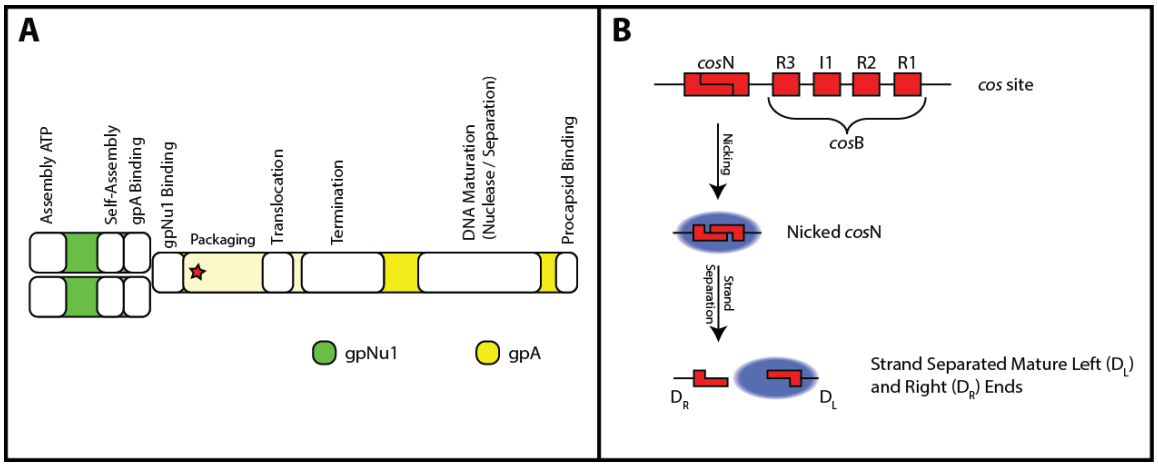


Figure 2. Terminase composition and a summary of subsites within the *cos* site. *Panel A:* The terminase enzyme is minimally composed of a heterotrimer of one gpA, which retains most catalytic activities of the enzyme, and two gpNu1 proteins, which confer site-specificity of binding to the enzyme. Catalytic activities specific to each subunit are indicated. *Panel B: cosB* refers to a terminase binding site that confers efficiency and accuracy of duplex nicking at *cosN*. The *I1* element introduces a bend into the duplex and is specifically recognized by *E. coli* integration host factor (IHF). The R elements are 16bp regions specifically recognized by gpNu1. The process of *cos*-cleavage, in which *cosN* is nicked and strand separated, generates D_R and D_L, the right and left sticky ends representing the end of a previous genome, and the start of a new genome, respectively.

The complex *cos* site and its subsites, which will be referred to extensively below, are enumerated and examined in Figure 2B. Also shown is a schematic of what the *cos* site looks like pre- and post- endonuclease cleavage.

Genetic experiments by Feiss and co-workers implicated that gpNu1 and gpA interact via domains in the C-terminal 91 amino acids of gpNu1 and the N-terminal 48 amino acids of gpA[9,10]. Previous footprinting studies have shown that the isolated large subunit of terminase, gpA, does not specifically protect only *cos*-containing DNA in DNase protection assays[11], instead conferring protection from upstream of *cosQ* to downstream of R1. Furthermore, gpA showed non-specific DNA-binding activity in gel mobility shift assays[12]. However, in the presence of gpNu1, gpA exhibits *cos*-specific binding[12]. Further footprinting studies showed that gpA DNA-binding is promoted by gpNu1 binding to *R3*. Thus it became clear the large and small subunits interact to mediate specific binding of the holoenzyme at *cos*. How gpNu1 mediated site-specific binding, however, was not clear. DNase protection assays[13] showed that gpNu1 assembles at *cosB*, showing cooperative binding to the *R2* and *R3* elements, with binding to the *R1* element likely requiring cooperative interactions across all three R elements (Figure 2B). Its interaction is limited, though, and no protection is seen at the upstream *cosQ* site. gpNu1 was thought to bind as a dimer to DNA, since deletion mutants of gpNu1 lacking the C-terminal self-association and gpA association domains, but retaining the DNA binding domain (DBD), behaved as soluble homodimers at even high concentrations[14]. In order for the gpNu1 DBD to bind the spatially separated *R2* and *R3* elements, it was suggested

that a bend must be induced in the *cos* DNA substrate. This suggestion was strengthened by the observed stimulation of virus yields *in vivo*[15], as well as DNA packaging *in vitro*[15], by the *E. coli* integration host factor (IHF), a DNA bending protein. Additionally, DNase protection assays directly showed that IHF promoted gpNu1 binding to the R elements[16]; however, it wasn't until NMR structural data of the gpNu1 DBD was obtained that a model for the IHF stimulated binding of gpNu1 to *cos*-containing DNA was put forth[17]. Briefly, based off of the crystal structure of an IHF-DNA complex[18] and NMR structural data for the gpNu1 DBD, de Beer *et al.* were able to construct a molecular model of the interaction of gpNu1 DBD dimer with the *R2* and *R3* elements (Figure 3).

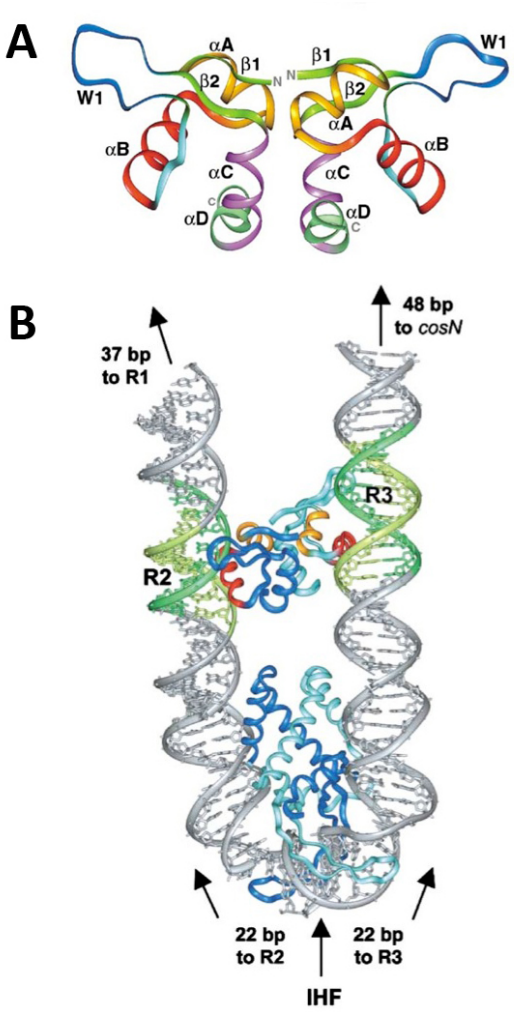


Figure 3. Model of gpNu1 DBD dimer alone and in complex with *cos*-containing DNA. *Panel A:* Ribbon representation of gpNu1 DBD (first 68 N-terminal amino acids of gpNu1) from an ensemble of 20 NMR structures. The wHTH motif is made up of αA and αB helices with a three-residue turn, and the W1 wing, pictured here. *Panel B:* Molecular model of the interaction of the two subunits of IHF (colored in different shades of blue) at *I1* and the gpNu1 DBD dimer at the *R2* and *R3* elements. The HTH motifs of gpNu1 DBD are colored as in (A) to highlight their interactions with the R elements while the rest of the monomers are depicted in shades of blue. Putative interactions of the W1 wings of each monomer with the minor groove of *R2* and *R3* are illustrated. Image taken from de Beer *et al*.[17], with permission.

Isotope-assisted NMR spectroscopy and primary sequence analysis indicated that the gpNu1 DBD had a winged helix-turn-helix (wHTH) motif, an important DNA recognition motif found in a wide array of proteins. The residues of the gpNu1 HTH motif implicated in electrostatic interactions with DNA were found to closely match the DNA interaction profiles of *lac* and *cro* repressor HTH motifs, and so superposition of the gpNu1 HTH motif and that of *lac* repressor complexed with DNA led to a model in which the first few turns of the two N-terminal helices (αA and αB) in the gpNu1 DBD contain residues predominantly involved in major groove DNA binding. This model placed the winged loop (W1) near the minor groove of the DNA; taken together with NMR titration data that showed a dramatic change in signal intensity of residues in the W1 loop upon titration of the *R3* element, as well as the finding that the mutation of a lysine to an aspartic acid within W1 abrogated the ability of gpNu1 to discriminate between *cos*-containing and nonspecific DNA[19], these findings suggested that the wing played a role in mediating specific DNA binding interactions. In order to accommodate residues of both the wing and the HTH binding to an R element, while taking into account the location of the dimerization interface of gpNu1 that places the wHTH motifs on opposite sides of the dimer (Figure 3A), de Beer *et al.* constructed the model shown in Figure 3B where each gpNu1 DBD of a dimer bind to separate R elements on opposite faces of the dimer. A severe bend of the DNA in between *R2* and *R3*, and the inversion of R sequences in *cosB*, could ensure presentation of similar binding surfaces to the recognition helices and wing of each gpNu1 DBD monomer. The severe bend between *R2* and *R3* elements is facilitated by IHF binding to the intrinsic bend

already present in the *I1* IHF recognition sequence between these elements. Indeed, more recently, Ortega and Catalano demonstrated that gpNu1 and IHF do in fact *cooperatively* bind and bend *cos* DNA, showing, via circular permutation assay, that the bend angle between the *R2* and *R3* elements increases from 120°±5° in the presence of IHF to 137°±4° in the presence of IHF and gpNu1[20]. To date there has been no evidence for this cooperativity arising from the direct interaction of IHF with gpNu1 (or terminase holoenzyme, for that matter), suggesting that cooperativity arises from a bent duplex architecture.

With a reasonable model for the site-specific association of gpNu1 with *cos*-containing DNA, extensive effort has been expended of recent in building a model for the interaction of the terminase holoenzyme with the *cos* site to initiate the packaging process. Crucial to the elucidation of how the terminase holoenzyme assembles at the *cos* site is an understanding of the higher-order oligomerization of large and small units to form the biologically relevant packaging enzyme. Efforts in our lab have been aimed at precisely these questions. Catalano *et al.* have extended the findings of Becker and Gold[8] regarding the specific composition of the terminase protein complex that assembles at the packaging initiation site by demonstrating, via both gel filtration and ultracentrifugation techniques[21,22], that terminase exists predominantly as a stable heterotrimeric complex composed of one large gpA and two small gpNu1 proteins (gpA_1•$gpNu1_2$). This minimal terminase unit that demonstrates enzymatic activity, or *protomer*, was shown to possess *cos*-cleavage endonuclease (previously referred to as the 'ter' reaction), DNA packaging, and virus

assembly activities that were strongly stimulated by IHF. This IHF dependence was to be expected given the cooperative binding of IHF and gpNu1 to *cos*-containing DNA. Catalano *et al.* surmised that the protomer likely oligomerizes to higher-order structures in forming catalytically-competent complexes, as evidenced by sedimentation velocity and equilibrium studies of purified terminase that show the presence of both a 5.1S and a 13.3S species with corresponding masses of 116 kDa and 528 kDa, respectively[22]. The latter species exhibited heterogeneity in sedimentation analyses, but dissociated to the homogenous 5.1S species over time. It was later found that the small species could be concentrated, under high salt conditions, to form a more homogenous larger species with an $s_{20,w}$ of ~14S[23]. For the sake of simplicity, we will only be concerned with the 5.1S and 14S species. The mass of the smaller species approximates well the mass of one gpA and two gpNu1 subunits (that is, protomer), while the mass of the 14S species, determined to be 459 kDa, approximates well an oligomer of four protomers, or a tetramer of protomers. Both species exhibit a gpNu1:gpA ratio of approximately 2:1, as confirmed by denaturing gel electrophoresis. Both species can additionally interconvert, with low concentration and low salt conditions favoring the protomer while high concentration and high salt conditions favor the tetramer. The tetramer exhibits the full complement of catalytic activities displayed by the IHF-stimulated protomer, but is able to do so in the absence of IHF. These findings suggested that while IHF introduces a bend in *cos*-containing DNA that facilitates the assembly of protomer onto the *cos* site, the tetramer is a preassembled complex that can directly interact with *cos* in the absence of IHF. Electron micrographs of a ring species

showed a hole at the center of the ring with an average diameter of 3 nanometers (nm), enough to accommodate duplex DNA[23], adding credence to the notion that the tetramer ring is the biologically relevant packaging motor *in vivo*.

Insight into the precise arrangement of protomers in the tetramer is facilitated by an analysis of the higher-order assembly capabilities of each subunit. Primary sequence analysis has identified both hydrophobic self-association domains and coiled-coil protein dimerization motifs within the C-terminal half of gpNu1[24]. Primary sequence analysis of gpA suggests a region near the C-terminal end with homology to the basic leucine zipper motif (bZIP) common to eukaryotic transcription factors[25]. Considering the symmetric nature of *cosN*[26], it was suggested that gpA binds as a head-to-head dimer, facilitated by the bZIP motif, with each gpA monomer binding to each *cosN* half-site[27]. Such dimerization is not uncommon in restriction endonucleases, especially Type II endonucleases that bind as a dimer to palindromic sequences where each monomer binds to a palindromic half-site. Given these association motifs, the precedent for a dimeric endonuclease complex, and the finding that protomer spontaneously self-assemble into tetramers under conducive conditions, Maluf *et al.* put forth a model suggesting that protomers associate with each other via the bZIP motifs of gpA and the coiled-coil dimerization motifs within the hydrophobic self-association domains of gpNu1[23]. The proposed model is shown in Figure 4.

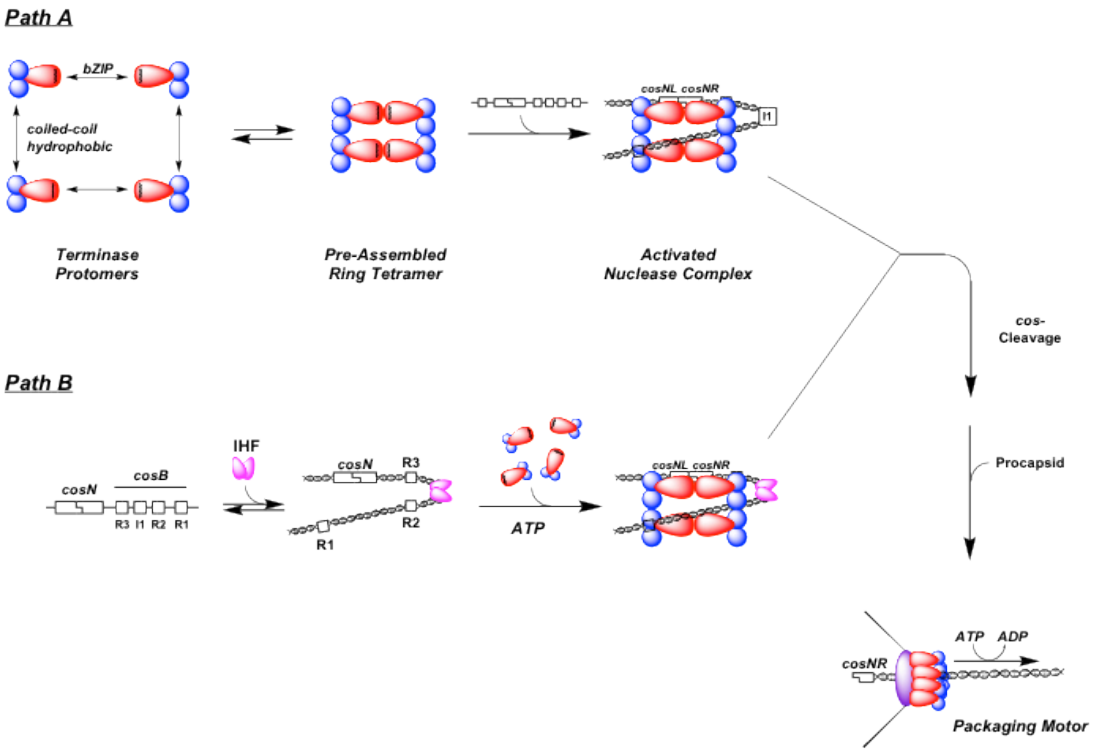


Figure 4. Proposed model for the assembly of λ terminase protomers into a functional DNA packaging motor on *cos*. (A) The terminase protomer preassembles via bZIP dimerization motifs in gpA (red) and coiled-coil dimerization motifs within the hydrophobic self-association domains of gpNu1 (blue). Such preassembly has been shown to occur *in vitro* under conditions of high salt and high protomer concentration. (B) IHF binds to the *I1* site located within the *cos* site, bending the duplex into a conformation conducive to the assembly of terminase protomers into a tetrameric ring on the DNA. This binding mode may explain the IHF dependence of protomer activity observed *in vitro*. Image taken from Maluf *et al.*[23], with permission.

Given the above model, it is now appropriate to consider the specific functions of terminase in the packaging process. At the initiation of packaging, we presume that two or more protomers assemble at *cos* in an IHF dependent manner to form a multi-protein *pre-nicking complex*; the requirement of IHF is evidenced by the 60-fold decrease in phage yield in IHF^-, HU^- double-mutant bacterial strains (HU is an accessory *E. coli* host protein, much like IHF, involved in DNA binding and

bending)[28]. At least two protomers seem to be a likely requirement for endonuclease activity (*vide supra*). Upon the addition of Mg^{2+}, the large gpA subunit of terminase makes staggered nicks at *cosN* 12 bp apart (Figure 2B). The resultant nucleoprotein entity is called the *post-nicking complex*. Upon strand separation, the mature right end of the genome D_R is released (and likely degraded *in vivo*), and a stable nucleoprotein complex known as *complex I* results (Figure 1A). Complex I was first identified *in vivo*[29], and later found to be composed of terminase bound to the mature left end, or D_L, of the λ genome (Figure 1A). Complex I subsequently binds to empty procapsid to form a packaging intermediate known as the pre-translocation complex, likely related to *complex II* (Figure 1B), also identified *in vivo*[30]. This triggers the release of terminase from *cos* and the concomitant transition from a stable nucleoprotein complex to a highly mobile packaging machine. For this to happen, the packaging ATPase of gpA (red star, Figure 2A) must be activated, and the binding mode of terminase must switch from a site-specific high-affinity interaction to a non-specific one of formidable affinity.

What exactly effects this '*cos* clearance' remains to be elucidated, but it is likely that *cos* clearance involves the interaction of complex I with portal proteins as well as potential effects mediated by nucleotide binding and/or hydrolysis by terminase. Essential to our understanding of this transition from a stable to mobile complex is a consideration of the origins of assembly, stability, and the exact nature of, the *pre-nicking*, *post-nicking*, and *strand-separated* (*complex I*) complexes (collectively referred to as the *pre-packaging* complexes, and individually depicted in Figure 1A).

Both the *pre-nicking complex* as well as *complex I* have been shown to be very stable, with half-lives of 12 and 8 hours, respectively[12]. In an effort to address one of the major components of the assembly of the *pre-packaging* complexes, the research currently presented examines the role of IHF in the assembly of terminase at *cos* in the *pre-nicking* complex (Chapters II and III). Additionally, the nature of terminase, including the stoichiometry of subunits, on DNA at the various stages of packaging is not well understood. We do not presume that the oligomeric state of terminase remains the same throughout the packaging process; in fact, in the above model, some conformational rearrangement most likely *must* occur for terminase to release the D_R end of the genome after endonuclease cleavage. In contrast, while the precedent for a dimeric endonuclease is a compelling one (*vide supra*), a hexameric ring structure for the packaging motor is also a pleasing one, since it can accommodate symmetric interactions with a dodecameric portal complex, and shows mechanistic similarity to the translocating hexameric helicase enzymes[31]. Yet a terminase complex with a stoichiometry of four protomers has been identified in our lab (*vide supra*). It is certainly conceivable that the stoichiometry of terminase at *cos* is different at the various packaging stages; ergo the precise nature (stoichiometry) of terminase on DNA in the *pre-packaging* packaging intermediates is a major focus of the research detailed here. More specifically, Chapter IV aims to elucidate the stoichiometry of terminase on DNA in the pre-nicking complex in the presence, and absence, of IHF.

II. Integration Host Factor Interactions at the Bacteriophage Lambda Packaging Initiation Site

Introduction

Bacteriophage λ terminase and *Escherichia coli* integration host factor (IHF) are thought to cooperatively assemble at the packaging initiation *cos* site of a λ genome to yield a catalytically competent nucleoprotein complex that matures the genome end and readies it for packaging. Given the complexity of this assembly process, an investigation of cooperative assembly of terminase and IHF *cos* first requires an understanding of the interactions between IHF and *cos*-DNA alone. IHF is an *Escherichia coli* protein originally discovered as a host protein required by bacteriophage λ for lysogeny[32]. IHF is a basic heterodimeric 22 kDa DNA binding protein belonging to a class of histone-like proteins capable of wrapping DNA into condensed structures. One of its biological functions is thought to be the condensation of bacterial nucleoid and, as such, displays sequence-independent (nonspecific) binding. IHF is rather unique, however, in that it has also evolved site-specific, high-affinity, binding. This sequence-dependent binding capability of IHF allows it to regulate a number of cellular processes, including site-specific recombination and replication[33]. In fact, sequence-specific IHF binding to the *attP* site of the λ genome is required for integration during lysogeny[34].

IHF has also been shown to bend DNA[35], and bending likely plays an architectural role in chromosome condensation[36]. It was recently shown that for its cognate, specific, H' binding site, IHF sequentially binds and then bends DNA, with bending becoming the rate-limiting step at high concentrations of IHF[37]. It is generally accepted that the specific complex (IHF bound to a specific DNA site) typically displays extensive bent architecture (for example, ~180° at its cognate site), while the nonspecific complex (IHF bound to a nonspecific DNA sequence) is not necessarily bent to such a degree, if at all[38,39]. The bent architecture is stabilized by multiple interactions between IHF and features of DNA (Figure 1), particularly between the β-ribbon arms of the protein and the minor groove of the DNA[38]. However, a widespread study of IHF binding affinity to a number of sequences matching the IHF consensus sequence to varying degrees revealed that specific IHF binding sequences can be, in some cases, replaced by flexible, 'nonspecific', sequences while still maintaining moderate to formidable affinities[40]. Indeed, force-extension measurements of the compaction of single DNA molecules upon IHF binding revealed efficient compaction of nonspecific sequences, with low-affinity binding sites likely playing the largest role in compaction[36]. Of course, this comes as no surprise, as histone-like proteins compact stretches of DNA in a sequence-independent fashion; however, it is interesting to consider that there may be a spectrum of unbent, bent, partially bent, and wrapped complexes that form between both specific and nonspecific sequences. The nonspecific interaction of IHF with DNA may then be a search mechanism aimed at finding either (1) specific or (2)

sufficiently flexible regions of DNA to bend and, therefore, bind to with tighter affinity.

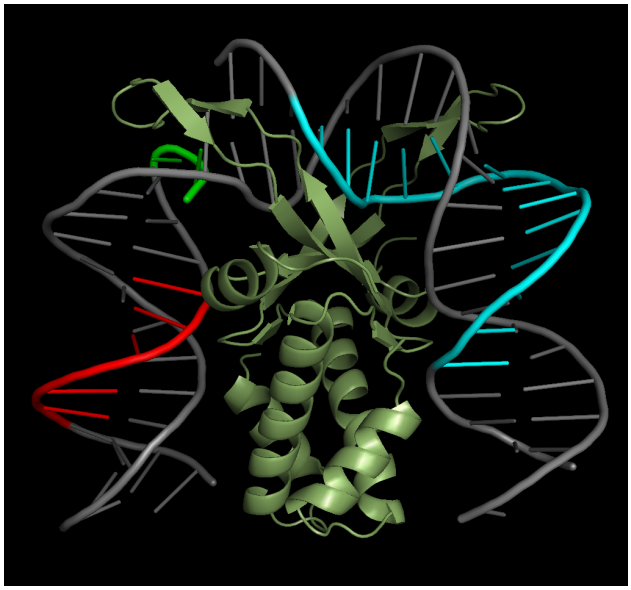


Figure 1. Crystal structure of IHF bound to its cognate H' site. *Pymol* was used to visualize the solved crystal structure[18] of IHF bound to a nicked variant of its cognate H' site (RCSB Protein Data Bank entry '1IHF'; DOI: 10.2210/pdb1ihf/pdb). The β-ribbon arms of the protein interact with the minor groove of the DNA, and a >160° bend is introduced into the duplex. Consensus sequence elements (see text) are indicated by color-coding of the DNA cartoon as follows: AAAAAAnnnTTnnnWATCAA nnnnTTRnn

Xin and Feiss identified a number of potentially specific IHF binding sequences within *cos*[41], based on the IHF consensus sequence originally defined by Craig and Nash[42]. Of these, the *I1* site (Figure 2) was shown to be well protected from DNase I attack in footprinting studies, while the *I2* site showed moderate to weak protection[41]. The *I1* site itself consists of two overlapping subsites (I1A and I1B) matching the IHF consensus sequence. The alignment of all these sequences, concordant with Xin and Feiss, is shown in Table 1.

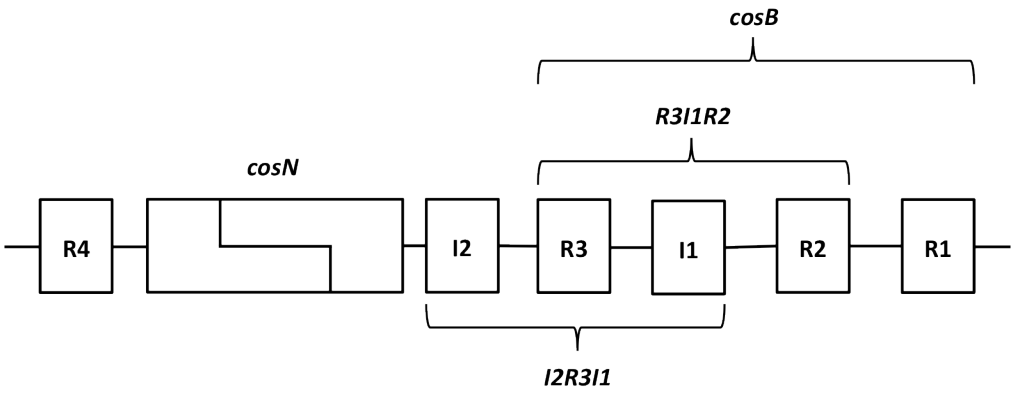


Figure 2. Subsites within the *cos* site of λ DNA. *I1* and *I2* are potentially specific IHF binding sequences. *R3I1R2* and *I2R3I1* indicate the extents of 75 bp substrates containing the *R3* gpNu1 binding consensus sequence plus either the *I1* site alone or with *I2*, respectively. *cosB* and its R-elements have been shown to bind the terminase gpNu1 subunit, but not IHF.

Original IHF Consensus	YNYAA nnnnTTGATW	**Hits**
I1A	CATAA CTTAATGTTT	**8**
I1B	CTTAA TGTTTTTATT	**9**
I1	CATAA CTTAATGTTTTTATTTAAAATA	**8**
I2	TTTTCGCTATT TATGAA AATTTTCCGG	**4**
Extended IHF Consensus	AAAAAAnnnTTnnn WATCAA nnnnTTRnn	

Table 1. Match of specific IHF binding sequences within *cos* to the original IHF consensus sequence defined by Craig & Nash[41,42]. W denotes A or T, Y denotes pyrimidine, R denotes purine, and n denotes any base. With respect to the original consensus sequence, base matches to n are greyed out, mismatches are highlighted in red, and matches are indicated in black. Matches to anything but n are counted in 'hits'. A more comprehensive, 'extended IHF consensus' sequence (see text below), is shown at the bottom for comparison.

In recent years, a more comprehensive IHF consensus sequence has been proposed based on a variety of studies[33,40,43-45]. The consensus sequence was extended to include the WATCAANNNNTTR sequence and poly-A, and possibly poly-T, tracts

upstream of WATCAA (these consensus elements are color-coded into the DNA in the crystal structure shown in Figure 1). For the purposes of this discussion, I will henceforth refer to this consensus as the 'extended IHF consensus'. To my knowledge, the *cos* site has not been thoroughly investigated in relation to this extended consensus sequence and the binding affinity that IHF displays for sequences within *cos*. Therefore, I decided to: (1) determine the affinity of IHF for the entire *cos* sequence, as well as a nonspecific DNA substrate of equivalent length, and (2) dissect the *cos* site into a number of core sequences to help define the site or sites leading to high-affinity IHF binding to *cos*.

Materials & Methods

DNA sequences

The following sequences, along with their complements where necessary, were synthesized, and HPLC purified, by Integrated DNA Technologies (Coralville, IA):

Forward primer for PCR of *cos274*: CCGGAATTCGCATGCCTGCAGGTCTAA

Reverse primer for PCR of *cos274*: CGCGAATTCCATTGTTCATTCCAC

Forward primer for PCR of *ns274*: GGGGAATTCACTGGCCGTCGTTTTA

Reverse primer for PCR of *ns274*: GGGGAATTCACTTAACTATGCGGCATCAGAGCA

I1: CATAACTTAATGTTTTTATTTAAAATA

I2: TTTTCGCTATTTATGAAAATTTTCCGG

R3: CGGTTTAAGGCGTTTCCGTTCTTCTTC

I2R3I1:

CTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGT

TTTTATTTAAAATAC

R3I1R2:

AAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAA

AGAAAGGAAACGACAG

Oligonucleotides were synthesized with fluorescent tags as needed. For electrophoretic mobility shift (EMS) studies, 5' ends of each strand of the duplex were labeled with IR700 dye[46] for subsequent detection using a Licor Odyssey scanner[47]. For sedimentation velocity (SV) studies employing a fluorescence

detection system, the 5' ends of the *I1* or *R3* oligonucleotides listed above, or the forward primer for the *cos274* and *ns274* substrates, were labeled with 6-carboxyfluorescein (6-FAM)[48].

Molecular weights of DNA substrates were calculated as needed, based on their sequences, using an online calculator[49]. A molecular weight of 537.5 Da was added, as necessary, to account for the weight of a 6-FAM label.

DNA sequence alignments and rationale

I utilized a minimal *cos* substrate (*cos274*) 274 bp in length containing the entire *cosB* and *cosN* sequences, diagrammed in Figure 2. A duplex of equivalent length but of random sequence served as a nonspecific DNA substrate, termed *ns274*. To further dissect the *cos* site and investigate the elements that confer high-affinity specific binding of IHF to *cos*, I also utilized 27 bp duplex substrates containing the sequences *I1*, *I2*, or *R3* in isolation (shown in boxes, Figure 2). The R-elements of *cos* are specifically recognized by the terminase small subunit (gpNu1), but are not recognized by IHF. Thus they provide a random DNA sequence with respect to IHF binding interactions. In order to determine if there is cooperative binding of IHF at the *I1* and *I2* sites, I constructed a 75 bp substrate containing both *I1* and *I2* ('*I2R3I1*', Figure 2). Finally, I determined the binding affinity of IHF for a 75 bp substrate containing *I1* alone ('*R3I1R2*', Figure 2) to probe for length effects in DNA recognition by IHF. The sequences of these duplex substrates are aligned against the extended IHF consensus in Table 2.

Extended IHF Consensus	AAAAAAnnnTTnnn WATCAA nnnnTTRnn	Hits
I1	CATAA CTTAATGTTTTTATTTAAAATA	4
I1 3′→5′	TATTTTA AATAAA AACATTAAGTTATG	10
I2	TTTTCGCTATT TATGAA AATTTTCCGG	8
R3I1R2/I2R3I1	...TCCGTTCTTCTTCG TCATAA CTTAATGTT...	6
R3I1R2/I2R3I1 3′→5′	...CAGAGGGTATTTTA AATAAA AACATTAAG...	12
R3I1R2/I2R3I1 3′→5′ #2	...TAAAAACATTAAGT TATGAC GAAGAAGAA...	11

Table 2. Match of specific IHF binding sequences within *cos* to the extended IHF consensus. W denotes A or T, Y denotes pyrimidine, R denotes purine, and n denotes any base. Base matches to n are greyed out, mismatches are highlighted in red, and matches are indicated in black. Only matches to anything but n are counted in 'hits'. I show potential alignments against the reverse complement strands of relevant sequences as well, indicated by 3'→5'.

Note that the first alignment shown for each individual substrate preserves the original alignment against the older IHF consensus[41], simply taking into account the alignment of the older IHF consensus against the extended consensus. Interestingly, the *I1* sequence, as previously defined based on sequence analysis[50] and footprinting studies[41], does not align very well to the extended IHF consensus, showing only 4 matches, or 'hits', to the consensus (Table 2). However, the complementary strand shows 10 matches, albeit at a different location, to the extended IHF consensus (*I1* 3'→5', Table 2). The *I2* sequence shows 8 matches to the extended consensus; the reverse complement, however, does not yield a better alignment. Both 75 bp substrates (*R3I1R2* and *I2R3I1*), when aligned so as to

preserve the original alignment of their *I1* elements against the older consensus, only show a modest increase of 2 matches (for a total of 6 matches, Table 2). Two potential alignments of the reverse complement strands of these 75 bp substrates are shown at the bottom of Table 2. Since I did not perform footprinting studies to directly measure the points of contact between IHF and these substrates, I cannot say which alignments are likely to be involved in specific IHF binding. I simply provide these alignments here for further discussion of my results.

IHF expression and purification

Integration Host Factor (IHF) was purified from HN880 heat-inducible IHF overproducing strain (a kind gift of Howard Nash, National Institutes of Health, Bethesda, MD) using a protocol modified from Filutowicz *et al*[51]. Briefly, 2 L of HN880 cells in Terrific Broth (TB) medium with 50 µg/ml ampicillin were grown to an optical density of 0.6 at 32°C. IHF was induced by the addition of an equal volume of 70°C TB medium. After a 15 minute incubation at 46°C, expression was allowed to continue at 42°C for a period of 3 hrs. Cells were harvested and then brought up in TEBG buffer (50mM Tris pH 7.0, 1mM EDTA, 7mM ß-mercaptoethanol, 10% glycerol v/v) containing 20mM KCl & Thermo Scientific 'Halt™' Protease Inhibitor Single-Use Cocktail. Cells were lysed using a French Press and insoluble cellular debris was removed by centrifugation (8000 x *g* for 30 minutes). Solid ammonium sulfate was added to the supernatant to 45% saturation, and after an additional 30min of stirring over ice, insoluble material was removed by centrifugation (8000 x *g* for 30 minutes). Solid ammonium sulfate was further added to the isolated supernatant to

85% saturation, and after an additional 30min of stirring over ice, insoluble material, containing most of the IHF, was isolated by centrifugation (10,000 x *g* for 30 minutes). IHF was resolubilized by bringing up the pellet in a suitable volume of TEBG (+20 mM KCl), and the ammonium sulfate was removed by overnight dialysis against TEBG (+ 20 mM KCl). The sample was loaded onto a 1 ml Q anion exchange column equilibrated with TEBG (+20 mM KCl), and the flow through was collected. This flow-through was applied to a 5 ml Heparin HP column equilibrated with TEBG (+20 mM KCl). A 20 mM – 1.5 M KCl gradient was applied to the column and IHF was eluted at approximately 870 mM KCl. Collected IHF was dialyzed into TEBG containing 870 mM KCl and 20% glycerol (v/v), and stored at -20°C. IHF heterodimer concentration was determined spectrally (ε_{276} = 5800 M^{-1} cm^{-1}).

A separate growth following the same protocols outlined above was carried out under identical condition save for the final storage conditions. The elution off the 5 ml Heparin HP column was dialyzed into TEBG containing 200mM KCl and 50% glycerol and stored at -80°C. This prep allows for one free-thaw cycle (from -80°C) to yield protein that will stay unfrozen if placed in -20°C after thawing. This latter batch of proteins was used for biophysical characterization studies that required formidable amounts of protein concentrations.

IHF molecular weight calculations

The molecular weight of IHF was calculated by summing the molecular weights of the individual α and β subunits. This resulted in a molecular weight of 22,005 Da for

IHF, based on 11,354 Da for the α subunit and 10,651 Da for the β subunit, as calculated by SEDNTERP[52] using the amino acid sequences for the α[53] and β[54] subunits. I also used SEDNTERP to calculate a partial specific volume of 0.736 for IHF based on the amino acid sequence. This information was used in conjunction with sedimentation velocity studies to determine the purity of my IHF prep.

Preparation of 6-FAM tagged specific and nonspecific duplexes used in sedimentation velocity (SV) studies

The 274 bp specific and nonspecific DNA substrates used in SV studies were prepared by large-scale preparative PCR. pAFP1 plasmid was used as a template in the PCR reactions, and was provided as a generous gift from Dr. Michael Feiss (University of Iowa). This plasmid contains the entire *cos* sequence cloned into a pUC19 vector, and was purified from *E. coli* JM107[pAFP1] cells using a Qiagen Plasmid *Plus* Maxi kit. PCR primers were designed to amplify (1) the entire *cos* sequence, inclusive of the R-elements that bind the gpNu1 small subunit of terminase and the *I1* element recognized by IHF, and (2) an equivalent-length (274 bp) of random sequence (*ns274*). Primer sequences are provided under '*DNA Sequences*', above; the forward primers for each substrate were 5'-labeled with 6-FAM[48] to aid in fluorescent detection of DNA. PCR reaction mixtures (100 µl per tube) contained each primer at a concentration of 0.5 µM, 10 ng of pAFP1 template. *Taq* DNA polymerase, deoxynucleotides, and required buffer components were included as specified by the manufacturer's protocol for the *Taq* PCR kit (New England Biolabs). Total reactions volumes were 1 ml. PCR cycles were as follows:

(94°C for 1 minute, 58°C for 1 minute, 72°C for 2 minutes) x 60 cycles. PCR reaction mixtures were subsequently pooled and purified using a Promega Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison WI). Purity of amplified products was verified by agarose gel analysis, and the concentration of DNA was determined via spectroscopic methods. The sequences of the specific (*cos274*) and nonspecific (*ns274*) PCR products are enumerated below.

cos274:

CCGGAATTCGCATGCCTGCAGGTCTAATCATTATCACTTTACGGGTCCTTTCCGGTGATC
CGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTT
AAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAA
AGAAAGGAAACGACAGGTGCTGAAAGCGAGGCTTTTTGGCCTCTGTCGTTTCCTTTCTCT
GTTTTTGTCCGTGGAATGAACAATGGAATTCGCG

***ns274*:**

GGGGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCA
ACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTA
TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAAT
CTGCTCTGATGCCGCATAGTTAAGTGAATTCCCC

Molecular weights of DNA substrates were calculated as needed, based on their sequences, using an online calculator[49]. Substrates with a 6-FAM[48] label had 537.5 Da added to their molecular weight.

Preparation of IR700-labeled 274 bp specific and nonspecific duplexes used in electrophoretic mobility shift (EMS) studies

The 274 bp specific (*cos274*) and nonspecific (*ns274*) DNA substrates used in EMS studies were prepared by large-scale preparative PCR, as described above in the immediately preceding section, with the following modifications: (1) both forward and reverse primers were 5'-labeled with IR700 dye[46]; and (2) the annealing temperature was raised from 58°C to 61°C to minimize a second, larger PCR product that appeared to more readily form with the IR700-labeled primers. PCR products were purified using PCR Kleen Spin Columns (Bio-Rad, Hercules CA), following the manufacturer's protocol.

Electrophoretic mobility shift (EMS) studies

Equilibrium binding experiments were performed in buffer containing 20mM Tris, pH 8, 1mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 10% glycerol. DNA was included at a final concentration of 2 nM, and IHF was added to the binding mixtures to appropriate concentrations. Mixtures were incubated at room temperature for 20 minutes, and then loaded onto an 8% polyacrylamide gel. I used an acrylamide/bis-acrylamide ratio of 80:1 for the 274bp DNA substrates so as to facilitate entry of the large complexes into the gel; for the smaller DNA substrates, I utilized a higher acrylamide/bis-acrylamide ratio of 29:1 for better resolution. Gels were run at 15V/cm in 0.5x TBE at 4°C for 1 hour, and then scanned on a LI-COR Odyssey scanner using appropriate integration times so as to maximize signal:noise

ratio of bands without clipping any signals. Bands were quantified using *ImageQuant TL* software (GE Healthcare Life Sciences).

The fraction of bound DNA was calculated using:

$$Fraction\ Bound = \frac{counts\ in\ retarded\ band}{counts\ in\ retarded\ band + counds\ in\ free\ DNA\ band}$$

The raw binding data were fit to a phenomenological Hill model, according to:

$$Fraction\ Bound = (m - b) * \left[\frac{[P]^n}{{K_d}^n + [P]^n}\right] + b$$

where [P] is the protein concentration, K_d is the apparent dissociation constant, *n* is the Hill coefficient, b is the baseline offset, and m is the fraction of DNA bound at saturation. Experimental data were fit using Igor Pro (Wave Metrics, Lake Oswego, OR).

Sedimentation velocity experiments

Unless otherwise noted, all sedimentation velocity (SV) experiments were conducted in buffer containing 20 mM Tris, pH 8, 1 mM EDTA, 7 mM β-me, 2 mM spermidine, 55 mM NaCl, and 5% glycerol. Sedimentation velocity experiments

were carried out on either a Beckman Coulter XLA ultracentrifuge, or XLI ultracentrifuge retrofitted with an Aviv Biomedical fluorescence detection system (FDS). For experiments conducted on the XLA absorbance system, data were collected at 280 nm, every 0.003 cm with 2 averages in continuous scan mode. The FDS system on the retrofitted XLI uses a continuous wave solid-state laser to excite samples at 488 nm, and uses a band-pass filter to detect any and all emission falling between 505 and 565 nm. I used this selective excitation to monitor only fluorescein-labeled DNA substrates and any complexes formed with them. Data for these experiments were collected at 20 μm intervals, with 5 averages per position. For all experiments, 400 μl of sample were loaded into preassembled analytical cells containing 12mm Epon charcoal-filled double sector centerpieces. Seven analytical cells were loaded into an An-50 Ti titanium rotor, along with one reference cell ('calibration center piece') containing a solution of fluorescein used by the FDS for both angle and radial calibration (or, alternatively, a Beckman Coulter counterweight for XLA experiments). Experiments were performed at 7°C, at 42K RPM. The photomultiplier tube gain on the FDS system was adjusted optimally so as to attain the highest signal-to-noise ratio for my samples without clipping the signal for the calibration center piece. The latter can be of concern because (a) every cell is scanned during each revolution of the rotor, and (b) the electronics in the FDS have a finite recovery time and can, therefore, be susceptible to not recovering from a saturated signal in time for the next cell being scanned. The finite response time of the electronics is also the reason for the recommendation that all cells be scanned at the same gain setting (personal communication with Aviv Biomedical).

The sedimentation velocity data were analyzed with SEDFIT[55] using a continuous c(s) approach. SEDFIT allows for the estimation of both the sedimentation coefficient s (from the rate of sedimentation of the sample, or, u, the velocity of the boundary) and the diffusion coefficient D (from the rate of the spreading of the boundary during sedimentation). The sedimentation coefficient s is related to the molecular weight M and the frictional coefficient f of the sedimenting species by the following relation (Svedberg equation):

$$s = \frac{u}{\omega^2 r} = \frac{MD(1-\bar{v}\rho)}{RT} = \frac{M(1-\bar{v}\rho)}{Nf}$$

In the above equation, ω is the angular velocity of the rotor, r is the radial distance for any given measurement, R is the gas constant, T is the absolute temperature, N is Avogadro's number, $\bar{v}$ is the partial specific volume, and ρ is the buffer density. The buffer density was calculated based on buffer components using SEDNTERP[52].

Since the FDS allowed me to monitor only the fluorescein-tagged DNA (and any nucleoprotein complex formed with the labeled DNA), the sedimentation coefficient was used as a measure of the fraction of bound DNA. Binding of protein to DNA typically leads to a change in s due to either a mass increase, a change in frictional coefficient, or both.

Furthermore, for simple, homogenous species, the fit s and f values can be used to determine the molecular weight M (if $\bar{v}$ is known), as the fit frictional coefficient is well-representative of the single species present. Hence, it was straightforward to calculate M for my purified IHF using the XLA. Determining the molecular weight of complexes in protein-DNA binding studies, however, is not straightforward. In complicated systems where multiple species exist, the determination of either M or $\bar{v}$ is rendered difficult by the fact that the fit f is an average of the frictional coefficients of all present species. Complicating such an analysis even further is the potential for different partial specific volumes of the various species formed between proteins and DNA. Therefore, without prior knowledge of the exact nucleoprotein species present in any binding mixture, it was not possible for me to rigorously assign a mass to any complexes formed in my protein-DNA SV binding studies. In the absence of a rigorous determination of the masses of complexes formed between protein and DNA, I was unable to estimate protein stoichiometries on DNA. Hence, the SV data were fit to a simple Hill binding model, where:

$$sedimentation\ coefficient = (s_{sat} - s_{ini}) * \frac{E^n}{{K_d}^n + E^n} + s_{ini}$$

In the above equation, s_{ini} is the sedimentation coefficient of free DNA, s_{sat} is the sedimentation coefficient of the saturated nucleoprotein species, K_d is the apparent dissociation constant, *and n* is the Hill coefficient. Experimental data were fit using Igor Pro (Wave Metrics, Lake Oswego, OR).

Since the weight average partial specific volume of any nucleoprotein species requires knowledge of the stoichiometry of the complex, and since it was not possible to accurately determine the mass, and therefore the stoichiometry, of the formed nucleoprotein complexes, I chose, for the sake of simplicity, to use the partial specific volume of the much larger duplex DNA (~170 kDa, vs. 22 kDa for IHF) for all conversions of the fit sedimentation coefficients to sedimentation coefficients under standard conditions (water, 20°C), or $s_{20,w}$.

Results

Purified IHF exists as a homogenous heterodimer in solution.

I first determined the characteristics of my IHF preparation in solution using analytical ultracentrifugation approaches as described in Materials and Methods. Sedimentation velocity (SV) experiments of 80μM IHF in 50mM Tris, pH 7.4, 1mM EDTA, 7mM β-me, 870mM KCl, and 10% glycerol, was carried out at 42K RPM. SEDFIT analysis of the raw SV data yielded an experimental s of 1.141 S, and a $s_{20,w}$ of 1.957 S (Figure 3). The fit yielded a molecular weight of 22.2 kDa, a frictional ratio of 1.42, and a Stokes radius of 2.64 nm. The experimental molecular weight is very close to the theoretical IHF molecular weight of 22.0 kDa (based on the amino acid sequence), and provides confidence that my protein was (1) pure, and (2) well behaved. Importantly, the AUC data showed no evidence of self-association or aggregation even at high concentrations.

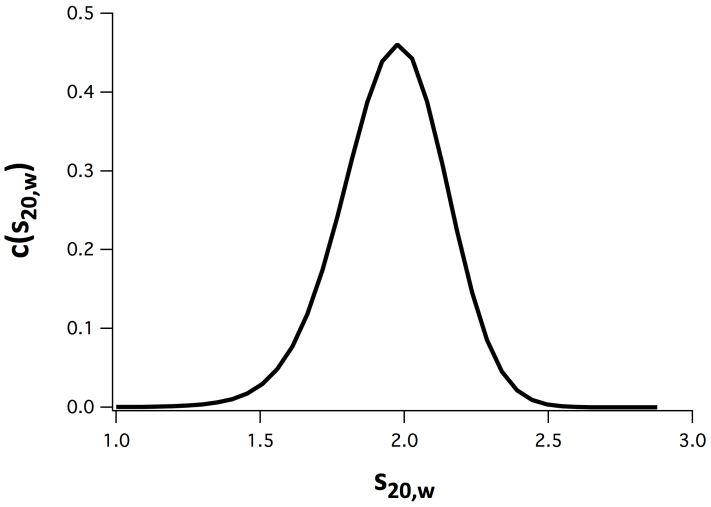


Figure 3. Sedimentation velocity analysis of purified IHF. Sedimentation velocity of IHF indicates a single homogenous species in solution with a molecular weight of 22.2 kDa. No evidence of self-association or aggregation is observed.

IHF binds to the full-length cos sequence with high-affinity to assemble a specific nucleoprotein complex, but does not afford a discrete complex in association with nonspecific DNA.

I next examined the interaction of purified IHF with a "specific" duplex substrate that models the full-length *cos* sequence found in concatemeric λ DNA. These studies utilized *cos274*, a fluorescently-tagged 274bp duplex containing the full *cos* sequence, in electrophoretic mobility shift (EMS) studies as described in Materials and Methods. IHF binds to this duplex to form a specific, gel-retarded, nucleoprotein complex (Figure 4A). The unusually large decreased mobility of the IHF-*cos274* complex in the gel is likely due to a bend introduced into the duplex upon specific binding by IHF, a phenomenon reported previously[38]. In contrast, IHF binds *ns274*, a non-specific 274bp duplex, to form diffuse bands that migrate as a smear in the gel (Figure 4B). I interpret the lack of a tight, gel-retarded, band to indicate the lack of a severe bend introduced into the duplex. The smeared bands are progressively retarded with increasing concentrations of IHF, indicating either (1) higher-order complexes of varying IHF:DNA ratios, with higher stoichiometry complexes running slower in the gel, (2) weaker affinity complexes that fall apart on the gel, or (3) some combination thereof. I note that at concentrations of IHF greater than 1 μM, the specfic IHF-*cos274* complex similarly begins to smear upward in the gel, likely due to nonspecific binding at such high concentrations (data not shown).

Analysis of the EMS data shows moderately higher affinity of IHF for *cos*-containing substrates in comparison to nonspecific DNA substrates of equal length. A roughly five-fold increase in IHF concentration (30 nM as opposed to 6 nM) is needed for the full disappearance of the *ns274* band in comparison to the *cos274* band (indicated by * in panels A and B of Figure 4). This indicates a moderate preference of IHF for the specific *cos274* substrate over the nonspecific *ns274* substrate. Quantitation and fitting of the *cos274* binding data (done in triplicate, over a wider concentration range than the limited range shown in Figure 4A) to a simple Hill model afforded a $K_{d,app}$ of 1.81±.08 nM, with n = 2.2. The quantitated data and fit are presented in Figure 4C. I did not find it appropriate to quantitate the *ns274* binding data, as a method to determine free vs. bound DNA from the gels was not obvious.

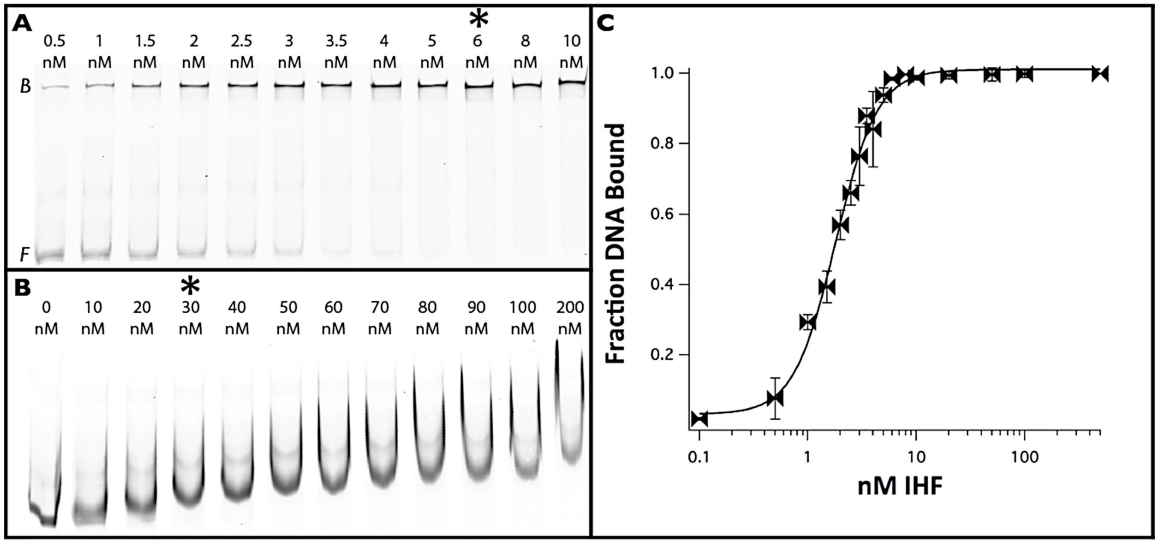


Figure 4. EMS studies of IHF binding to specific (*cos274*) and non-specific (*ns274*) DNA substrates. *Panel A*. IHF titration (0.5 - 10 nM) against the specific *cos274* DNA substrate. Positions of IHF-bound and free DNA are indicated by *B* and *F*, respectively, at left. *Panel B*. IHF titration (0 - 200 nM) against the non-specific *ns274* DNA substrate. A general, upward-moving, smear is observed with increasing IHF concentration, so positions of free and bound DNA are not indicated. The asterisks (*) in panels A and B indicate the concentrations of IHF that showed >98% *cos274* bound, or complete disappearance of the *ns274* band (6 nM and 30 nM IHF, respectively). *Panel C*. Quantitation of IHF-*cos274* EMS data over a broad concentration range in triplicate. Data were fit to a Hill model. The *ns274* binding data were not quantified due to the absence of a well-defined retarded complex in the gel.

Sedimentation velocity shows that multiple IHF proteins bind to both specific and nonspecific DNA substrates to afford higher-order nucleoprotein complexes.

My EMS studies did not allow me to accurately determine the binding affinity for the IHF-*ns274* interaction. Hence, I turned to analytical ultracentrifugation (AUC) as a rigorous, solution-based, method for determining the affinity of IHF for both *cos274* and *ns274* substrates. Sedimentation velocity (SV) experiments were conducted using fluorescein-labeled DNA substrates as described in Materials and Methods. An

IHF titration from 0 to 2 μM IHF was carried out in the presence of either 4 nM *cos274* or 4 nM *ns274.* Raw data from these experiments were analyzed in SEDFIT using a continuous c(s) approach, and the normalized c($s_{20,w}$) distributions from these analyses are shown in Figure 5 (*cos274*) and Figure 6 (*ns274*).

I first note that a fit of the raw data for free *cos274* yields an experimental molecular weight of 166.3 kDa, with a fit frictional ratio of 3.424. This is consistent with the theoretical molecular weight of 169.9 kDa for 6-FAM-labeled *cos274*. A fit of the raw data for *ns274* yields an experimental molecular weight of 172.7 kDa, with a fit frictional ratio of 3.396. This is also consistent with the theoretical molecular weight of 169.9 kDa for 6-FAM-labeled *ns274*. The theoretical vs. experimental agreement afford confidence in the experimental methodology.

Interestingly, with SV-AUC, I do not observe the high affinity IHF-*cos274* binding event evident in EMS studies (see Figures 4A, 4C). There is no discernible change in the $s_{20,w}$ sedimentation coefficient upon addition of 10 nM IHF to 4 nM DNA, despite the observation that the $K_{d,app}$ of the IHF-*cos274* interaction derived from EMSA studies indicate that the DNA should be fully bound at this concentration of IHF. I hypothesize this is due to either (1) compensatory mass and size/shape effects that mask the specific IHF binding event, (2) insufficient resolution of this technique, under these conditions, for the detection of the binding of the small protein to duplex DNA nearly eight times its mass, or (3) some combination thereof. Recall that the sedimentation coefficient from a SV experiment depends on both the mass *and*

the frictional ratio (a measure of the effective size of the species) of the complex; this, in turn, could explain the potential compensatory mass and size/shape effects. For the sedimentation coefficient to remain the same upon the mass increase associated with IHF binding to DNA, the diffusion coefficient would have to decrease (that is, the frictional ratio would have to increase) in a commensurate manner (sedimentation coefficient is directly proportional to both mass and diffusion coefficient). While this is technically possible due to size increases of complexes upon protein binding, I find it highly unlikely for the following reasons: (1) the 274bp duplex DNA itself has a high frictional ratio due to its rod-like structure, and it is unlikely that the binding of a small protein such as IHF would increase its frictional ratio further without some sort of structural alteration to the DNA; and (2) if there were any structural alteration to the DNA upon IHF binding, I would expect it to be a bend at *cos* which would, in turn, *decrease* the frictional ratio of the DNA. Since the fit frictional ratio in SEDFIT is only representative of the ensemble of species present, and since I had no knowledge of the homogeneity of nucleoprotein complexes for any given IHF loading concentration (in fact, I expected a good deal of heterogeneity), it was difficult to draw any significant conclusions from an analysis of the frictional ratios. That said, higher concentrations of IHF did lead to a discernible, saturable, shift to higher sedimentation coefficients, with the saturated species displaying a $s_{20,w}$ of 15 S (Figure 5, bottom). Given the magnitude of the sedimentation coefficient of the saturated species, and considering the high concentration of IHF required to reach saturation, I suspect my experiment to be monitoring the binding of multiple IHFs to *cos274*.

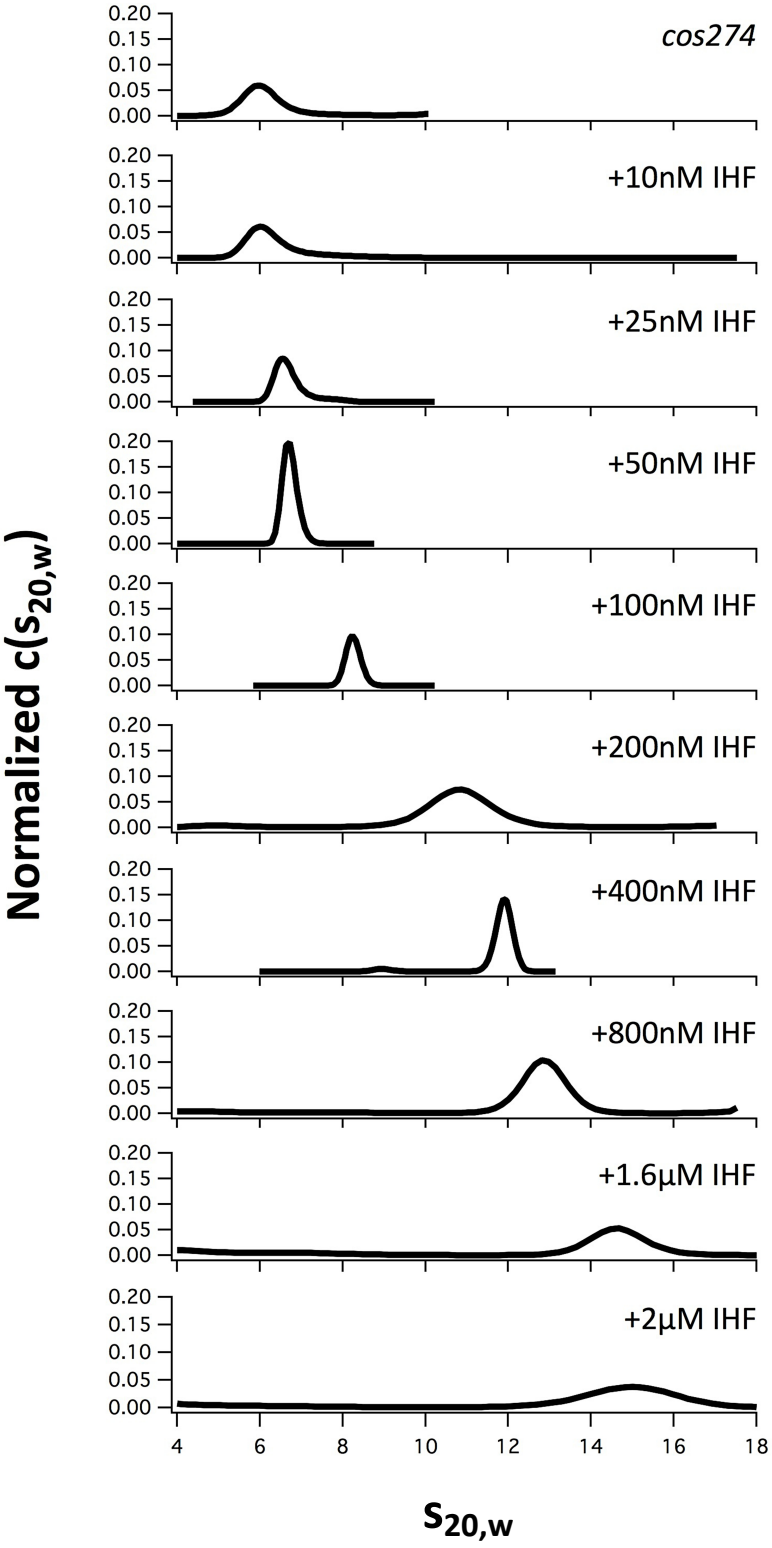


Figure 5. Sedimentation velocity analysis of IHF binding to *cos274*. Sedimentation velocity (SV) of 4 nM *cos274* in the presence of increasing concentrations of IHF was performed at 42K RPM. DNA was excited at 488 nm and all emission between 505 and 565 nm was collected. The high affinity binding event observed in EMS studies (Figure 4A, 4C) was not detected in this SV study. Lower affinity binding events, however, are evident in the evolution of slower sedimenting, higher $s_{20,w}$, species. I note a 15 S species at saturation.

A titration of IHF on *ns274* showed a concentration dependent increase in sedimentation coefficient of the fluorescently-tagged DNA species (Figure 6) similar to that observed with *cos274*. In contrast to my *cos274* data, however, my *ns274* experiment showed the evolution of not only a higher $s_{20,w}$ species with increasing IHF concentration, but also a very slowly sedimenting ~4.2 S species at concentrations of IHF of 500 nM and above. I suspect that in the absence of a specific IHF binding and bending event on DNA, higher concentrations of IHF lead to extensive nonspecific DNA binding that lead to the wrapping of DNA into a slowly sedimenting, low diffusion coefficient (or high frictional ratio), nucleoprotein complex. Perhaps extensive nonspecific binding of DNA at high concentrations of IHF makes it behave as a rigid rod in order to accommodate the maximum number of IHF molecules. The presence of a significant bend in the specific complex with IHF might preclude this from happening with *cos274* (assuming nonspecific binding does not outcompete specific binding), which would explain the lack of the 4.2 S species in SV experiments with *cos274*.

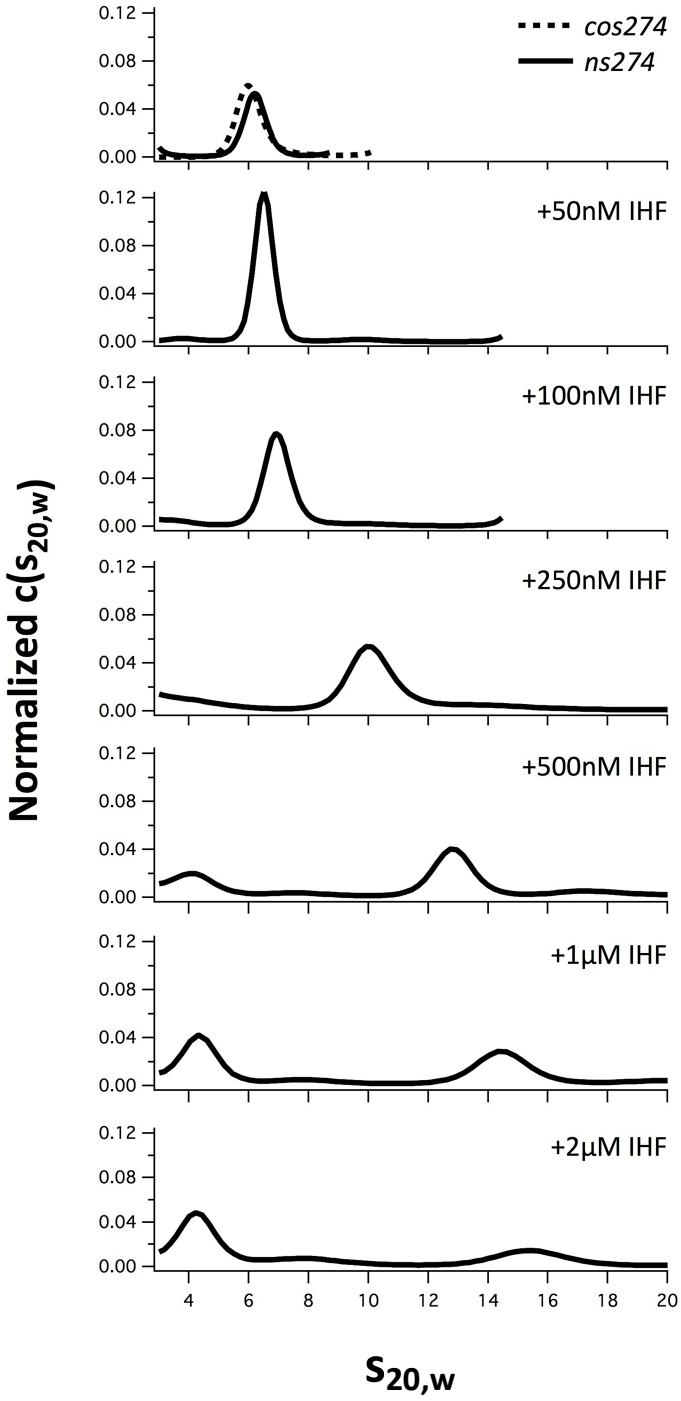


Figure 6. Sedimentation velocity analysis of IHF binding to *ns274*. Sedimentation velocity (SV) of 4 nM *ns274* in the presence of increasing concentrations of IHF was performed at 42K RPM. DNA was excited at 488 nm and all emission between 505 and 565 nm was collected. Addition of IHF to *ns274* leads to a concentration dependent, saturable, increase in sedimentation coefficient of fluorescently-tagged DNA. The magnitude of the sedimentation coefficient at saturation ($s_{20,w}$ = ~15 S) indicates the high likelihood of higher order nucleoprotein complexes of IHF:DNA stoichiometry > 1. Interestingly, I also see the evolution of a species of very low sedimentation coefficient ($s_{20,w}$ = ~4.2 S) at very high IHF concentrations.

I plotted the $s_{20,w}$ sedimentation coefficients of the nucleoprotein complexes formed between IHF and both *cos274* and *ns274* as a function of IHF concentration (Figure 7). Data were fit to a Hill model as described in Materials and Methods, and these fits

are shown in Figure 7. The fit to my *cos274* data afforded a $K_{d,app}$ of 202 ± 33 nM, with n = 1.28. The fit to my *ns274* data afforded a $K_{d,app}$ of 310 ± 16 nM, with n = 1.82.

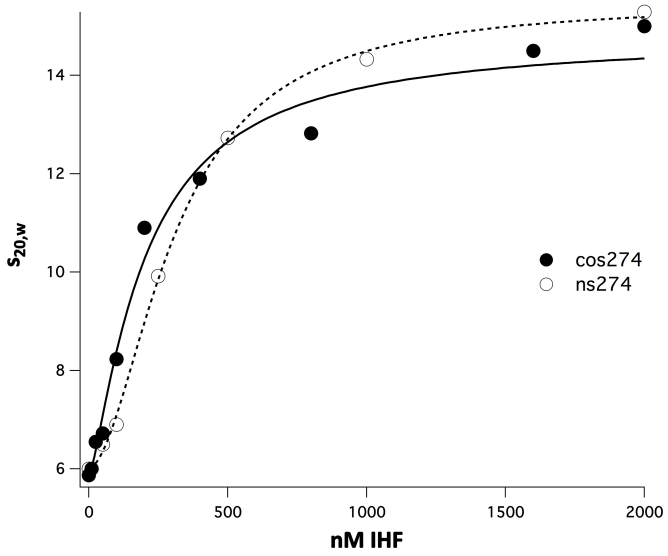


Figure 7. SV-AUC sedimentation coefficents of IHF:DNA complexes formed with *cos274* and *ns274*. $s_{20,w}$ sedimentation coefficients were plotted as a function of IHF concentration for both *cos274* (●) and *ns274* (○). Data were fit to a phenomenological Hill model, which yielded a $K_{d,app}$ of 202 nM and 310 nM for *cos274* and *ns274*, respectively.

In summary, I found that SV-AUC was unable to distinguish a significant difference in complexes formed between IHF and *cos274* or *ns274*, save for the appearance of a slow sedimenting nucleoprotein species formed between IHF and *ns274* at high protein concentrations. Furthermore, saturated complexes for both *cos274* and *ns274* yielded $s_{20,w}$ values of ~15 S, indicating rather large nucleoprotein species and, therefore, higher order complexes of multiple IHF stoichiometries. Given the relatively low discrimination (less than an order of magnitude) IHF shows for specific and nonspecific sequences in EMS studies (Figure 4A, 4B), I postulate that the larger $s_{20,w}$ species in the SV c($s_{20,w}$) distributions for both *cos274* and *ns274* represent weight averages of multiple IHF-DNA species of varied stoichiometry. I

interpret the continuous progression of sedimentation coefficient as a function of IHF concentration to indicate that the IHF-DNA association is under rapid equilibrium on the time-scale of the experiment (~6 hours).

IHF binds short DNA duplexes containing matches to the IHF consensus sequence with moderate to high affinity to yield discrete, gel-retarded, complexes.

My SV experiments on *cos274* and *ns274* suggested multiple stoichiometries of complexes formed between IHF and both specific and nonspecific DNA duplexes (*vide supra*); this complicates a straightforward analysis of the binding affinity of IHF for my specific and nonspecific substrates. Therefore, I decided to dissect the complex *cos* site by addressing the affinity of IHF for minimal substrates representing both minimal specific and nonspecific binding sequences found within *cos*. Please refer to Chapter II: Introduction and Materials and Methods for a treatment on the subsites of *cos* used in the following studies.

The shortest of the specific substrates (*I1*) is a 27 bp fragment inclusive of the *I1A* and *I1B* sequences matching the IHF binding consensus sequence previously identified in *cos*[41,42]. *I1* exhibits a discrete, gel-retarded band in complex with IHF in EMS studies (Fig. 8A). Analysis of the data as described in Materials and Methods yields a $K_{d,app}$ of 20.0 ± 1.28 nM (Figure 10), with n=1.94, which is roughly an order of magnitude greater than that of the full-length *cos274-IHF* complex ($K_{d,app}$ = 1.81

nM, Figure 4C). The reason for this dramatic difference in binding affinity was next examined.

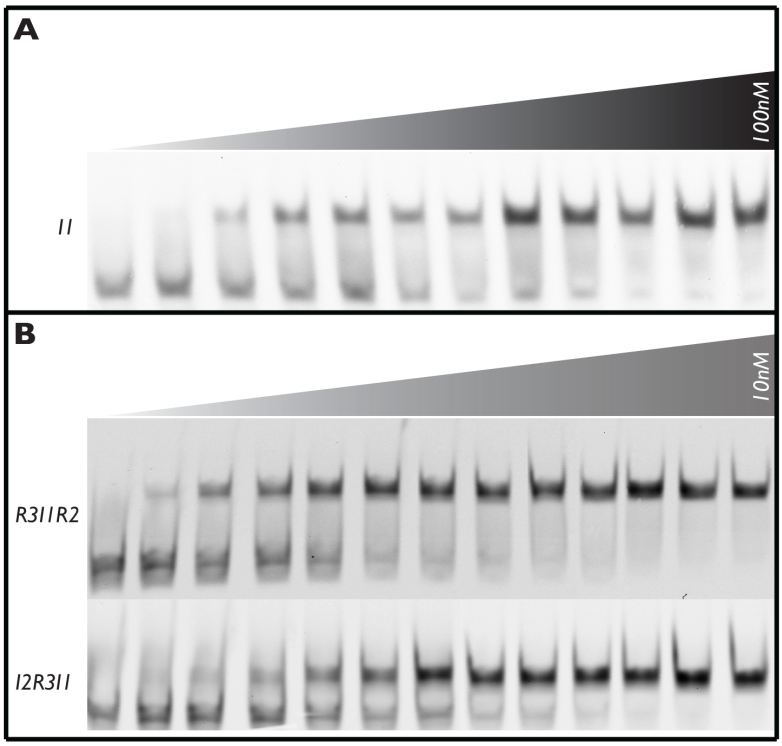


Figure 8. IHF binds specific, short duplexes with moderate to high affinity to form discrete complexes in EMS studies. *Panel A:* EMS study of (0 - 100 nM) IHF binding to the 27 bp isolated *I1* subsite of *cos* shows a discrete, albeit lower affinity in comparison to *cos274*, complex. *Panel B:* EMS study of (0 - 10 nM) IHF binding to the 75 bp *R3I1R2* and *I2R3I1* subsites of *cos* show discrete nucleoprotein complexes formed with similar affinity to that of the IHF•*cos* interaction.

Since a second sequence (*I2*) within *cos* was previously reported to exhibit moderate protection against DNase I attack in the presence of IHF[41], I questioned whether it was the presence of *I2* in conjunction with *I1* (i.e., cooperative binding), an additional as yet unidentified element within *cos,* or simply extra nonspecific interactions (i.e., duplex length) that conferred higher binding affinity to *cos274*. To this end, I performed EMS studies using (1) an *I2R3I1* substrate containing both *I1*

and *I2* elements and (2) a *R3I1R2* substrate of equal length cut containing only the *I1* binding site (R-elements show nonspecific IHF binding, discussed further below). Both 75 bp substrates yielded discrete, gel-retarded, IHF-DNA complexes in EMS studies (Fig. 8B), as is observed with *cos274* and *I1* substrates. This suggests that IHF is binding specifically to the *I1* element in both duplexes. It is noteworth that the IHF-*R3I1R2* complex displays a notably more gel-retarded band in comparison to the IHF-*I2R3I1* complex (Figure 9). I presume this is due to the fact that the bend introduced by IHF at or near the specific *I1* element is closer to the center of the *R3I1R2* sequence than the *I2R3I1* sequence. A bent duplex with symmetrically disposed duplex arms will encounter more resistance to gel migration than a similarly sized bent duplex with asymmetrically disposed duplex arms and will, thus, migrate more slowly.

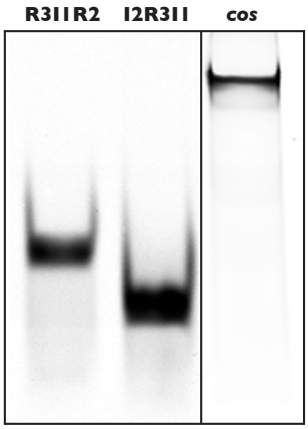


Figure 9. Relative mobility shifts of IHF-DNA specific complexes. Shown here are the gel-retarded complexes formed between IHF and (1) 75 bp *R3I1R2*, (2) 75 bp *I2R3I1*, and (3) 274 bp *cos*. I suspect *R3I1R2* exhibits a slower migration in the gel than *I2R3I1* due to a more centrally located specific IHF binding site and, therefore, a more centrally located bend in the DNA duplex.

Gels from EMS studies of all specific duplexes were quantitated as described in Materials and Methods, and were fit to a Hill model (Figure 10).

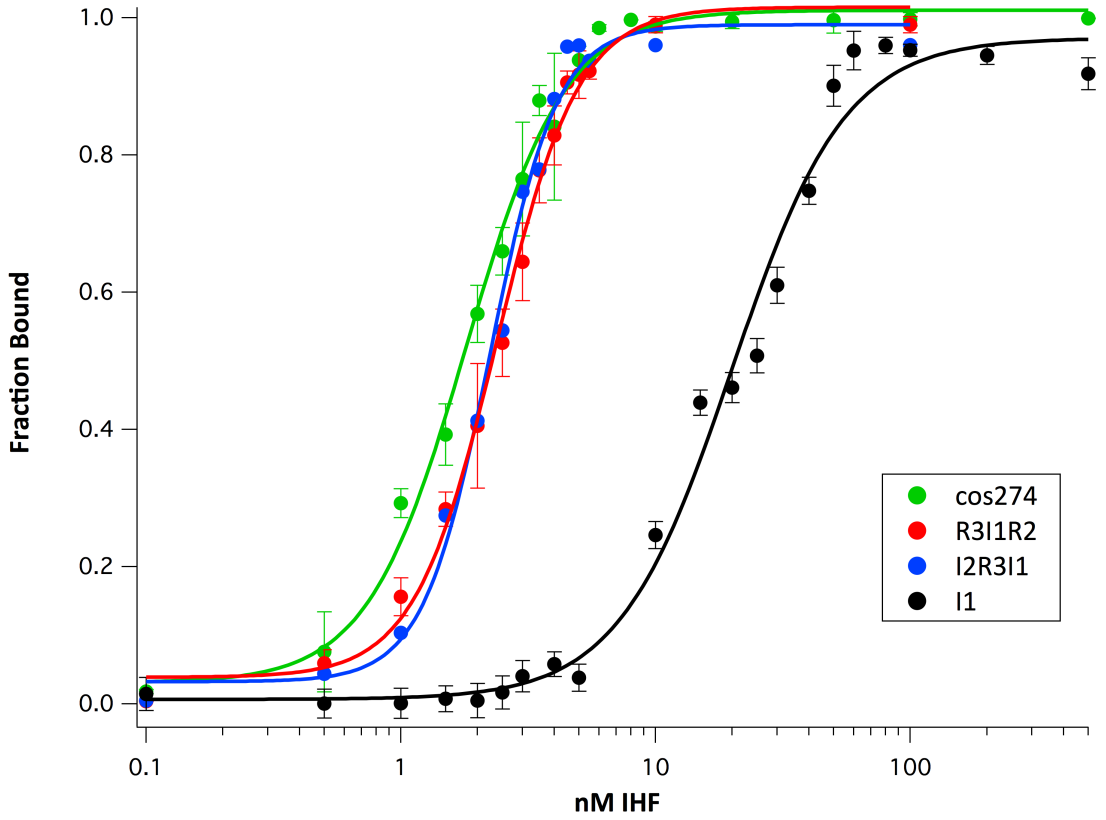


Figure 10. Quantitation of EMS data for IHF binding to specific substrates within *cos*. EMS data were quantitated using *ImageQuant TL* for all IHF interactions with *cos* and specific subsites of *cos*. Data were fit to a phenomenological Hill model to derive apparent K_d values. In summary, the 75 bp *R3I1R2* and *I2R3I1* DNA substrates show specific complexes with IHF of similar affinity to those formed between IHF and *cos*. The 27 bp *I1* substrate forms a specific complex with IHF with an affinity that is an order of magnitude lower than that of the IHF-*cos* complex.

A quantitative analysis of my EMS studies on specific subsites within *cos* indicated the following: (1) the 27 bp *I1* substrate shows an order of magnitude lower affinity for IHF in comparison to *cos*, (2) both 75 bp *R3I1R2* and *I2R3I1* substrates show almost equivalent affinities for IHF, on the order of the affinity of IHF for the full *cos* sequence, and (3) there does not appear to be any cooperativity between the *I1* and *I2* subsites, since the affinities and Hill coefficients for *I2R3I1* and *R3I1R2* are similar. The lack of any observed cooperativity between *I1* and *I2* was not surprising

in light of the binding affinity of IHF for *I2* (discussed below). A summary of the binding data is presented in Table 3.

Substrate	$K_{d,app}$	*n*
cos	1.8 ± 0.08 nM	2.2 ± 0.2
R3I1R2	2.4 ± 0.08 nM	2.7 ± 0.2
I2R3I1	2.2 ± 0.08 nM	3.3 ± 0.3
I1	20.0 ± 1.28 nM	2.0 ± 0.2

Table 3. Summary of binding constants derived from EMS studies of IHF binding to specific sites within *cos*. Data shown from EMS studies in Figure 10 were fit to a simple Hill model to extract binding constants for the interactions of IHF with specific sites within *cos*. The fit binding constants are summarized here.

IHF binds shorter nonspecific DNA duplexes to yield lower affinity, diffuse smears in EMS studies.

Since a subset of my specific substrates contained R-elements, and given the enhanced affinity of IHF for R-element containing 75 bp substrates (*R3I1R2* and *I2R3I1*) showed over *I1*, I attempted to determine the binding affinity of a representative R-element sequence for IHF in order to assess its contributions, if any, to the specific IHF complexes observed. I used the *R3* sequence as a model nonspecific substrate (*R3*, 27bp), because it is the dominant element associated with terminase binding to *cos* and because as it showed little similarity to the IHF consensus sequence. Additionally, a 27bp duplex (*I2*) containing only the *I2* sequence was examined to determine if it formed a specific complex with IHF and

whether its presence could explain the enhanced affinity of *I2R3I1* over *I1*. EMS studies showed that both *R3* and *I2* formed lower affinity, diffusely smeared, complexes in gels (Figure 11). This behavior suggests that the interaction affords a nonspecific, non-bent complex, not unlike those resulting from IHF-*ns274* interactions (Figure 4B). The affinity of IHF for *I2* is similar to that for *R3*, which initially came as a surprise given that *I2* only displays two mismatches to the original IHF consensus sequence (Table 1). On the other hand, this finding corroborated my observation of no cooperativity between the *I1* and *I2* sequences in IHF binding in my studies with the *I2R3I1* substrate. As I was not able to accurately quantify free vs. bound DNA from these EMS studies, I did not attempt to quantitate the data to extract binding constants.

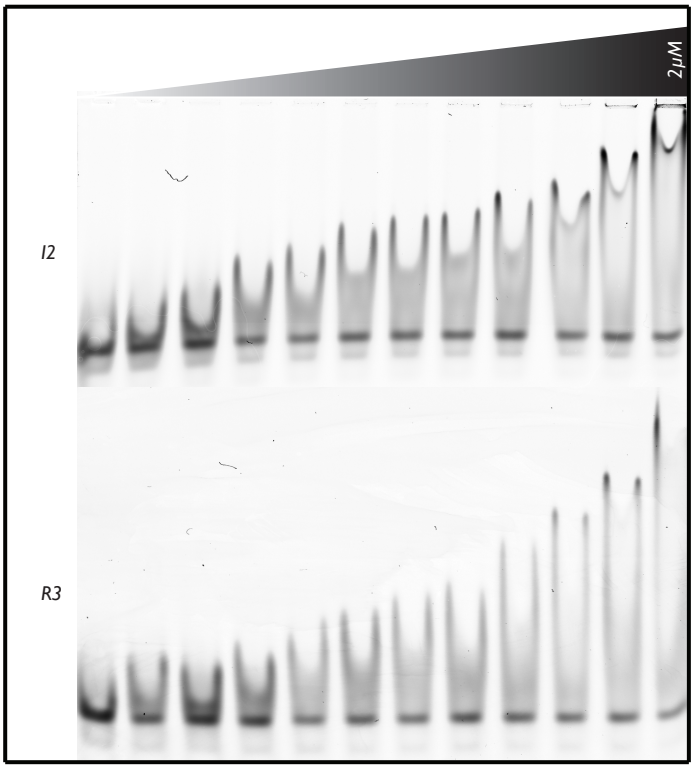


Figure 11. IHF binds representative non-specific, short substrates with moderate to low affinity to yield diffuse, non-discrete complexes in EMS studies. IHF (0 - 2μM) binds to 27bp non-specific substrates with lower affinity than observed for specific substrates of equal length. Notably, the complexes formed between IHF and non-specific substrates qualitatively differ from those formed with specific substrates.

Sedimentation velocity analysis does not differentiate between specific vs nonspecific interactions of IHF with minimal specific and nonspecific substrates of cos.

As was observed with the *ns274* duplex, IHF binds to minimal nonspecific 27 bp duplexes to yield a diffuse smear, which precludes rigorous quantification of binding affinities. Therefore, I turned to SV-AUC to rigorously determine binding affinities for both the specific and nonspecific minimal substrates of *cos* (*I1* and *R3*, respectively). SV experiments were conducted using fluorescein-labeled DNA substrates as described in Materials and Methods. An IHF titration from 0 to 2 μM IHF was carried out in the presence of either 4 nM *I1* or *R3*, and the raw data from these experiments were analyzed with SEDFIT using a continuous c(s) approach. c($s_{20,w}$) distributions from these analyses are shown in Figure 12 (*I1*) and Figure 13 (*R3*).

I first note that a fit to the raw free DNA data yielded a mass of 17.8 kDa for *I1*, which is consistent with the theoretical *I1* molecular weight of 17.1 kDa. Based on the affinity of IHF for *I1* from EMS studies ($K_{d,app}$ = 20.0 nM) I expected to see a concentration-dependent increase in sedimentation coefficient at these low concentrations. While I did see a slight increase in sedimentation coefficient at 50 nM IHF, larger, concentration-dependent increases in sedimentation coefficient were observed at higher IHF concentrations to 500 nM IHF, followed by more modest effects at higher concentrations. A similar trend was observed with the nonspecific *R3* substrate, albeit with a seemingly lower initial IHF concentration-

dependent increase in sedimentation coefficient. In both cases, I believe the reactions are at or near saturation of binding by roughly 2 μM IHF. A plot of $s_{20,w}$ vs. IHF for both *I1* and *R3* allows for easier visualization of this observation (Figure 14).

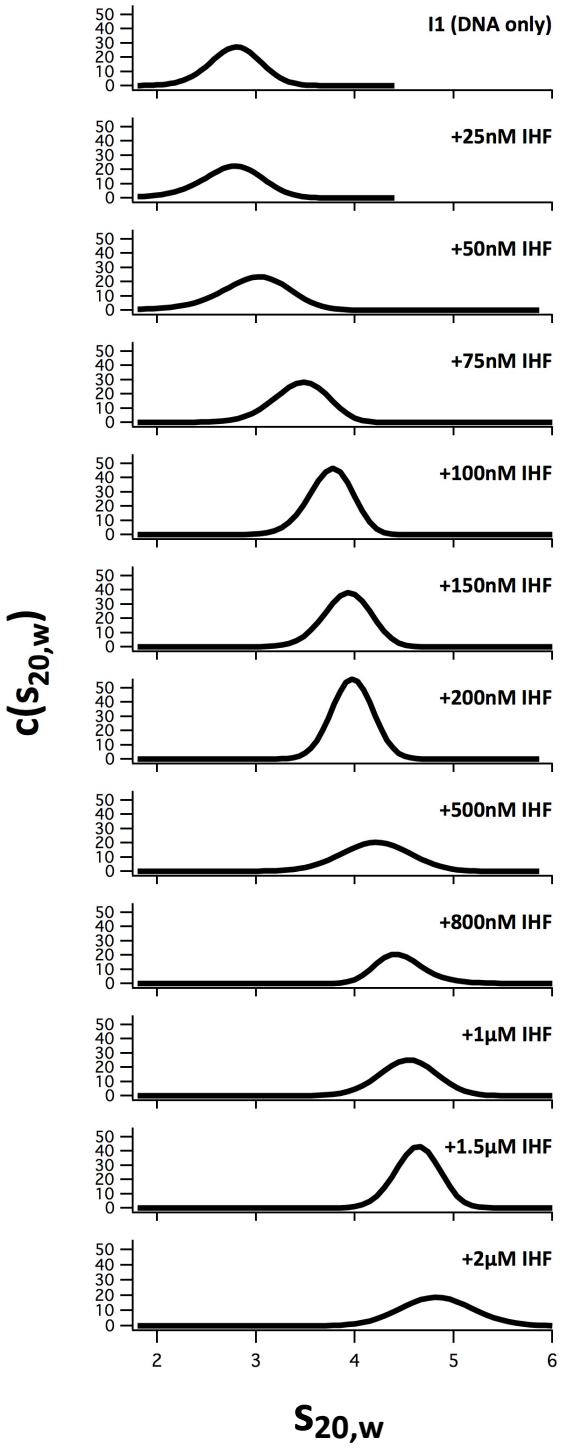


Figure 12. Sedimentation velocity (SV) analysis of IHF binding to *I1*. SV of fluorescently-tagged 4 nM *I1* in the presence of increasing concentrations of IHF was performed at 42K RPM. DNA was excited at 488 nm and all emission between 505 and 565 nm was collected. This data show an IHF concentration-dependent increase in sedimentation coefficient of the DNA species. The moderate affinity (an order of magnitude lower than that observed for IHF and *cos* in EMS studies) binding event observed in EMS studies (Figure 8A) does not appear to be resolved particularly well from presumably lower affinity binding events leading to the final, I believe saturated, 4.8 S, species.

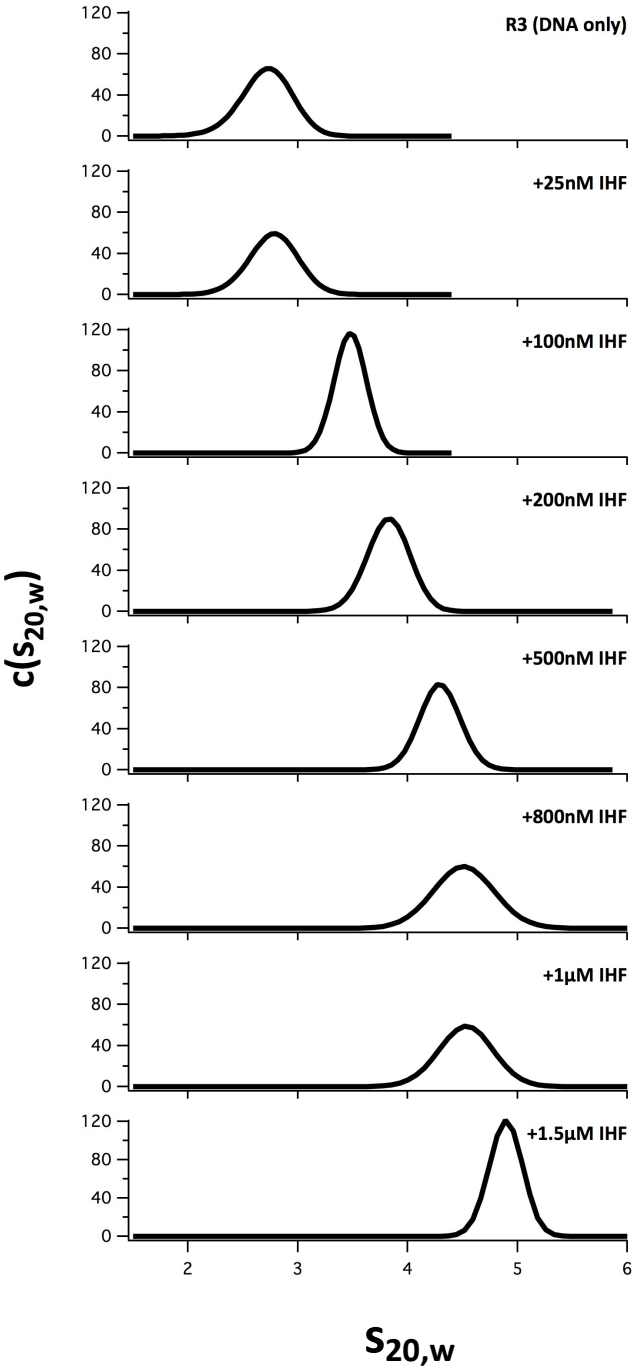


Figure 13. Sedimentation velocity (SV) analysis of IHF binding to *R3*. SV of fluorescently-tagged 4 nM *R3* in the presence of increasing concentrations of IHF was performed at 42K RPM. DNA was excited at 488 nm and all emission between 505 and 565 nm was collected. This data show an IHF concentration-dependent increase in sedimentation coefficient of the DNA species that is, interestingly, not very different from the IHF titration against the equivalent length, but specific, *I1* substrate. The putative saturated species has a sedimentation coefficient of 4.9 S.

In summary, SV profiles of nucleoprotein complexes formed between IHF and either *I1* or *R3* show a saturable, IHF concentration-dependent, increase in sedimentation coefficient. I plotted the fit $s_{20,w}$ sedimentation coefficients of the nucleoprotein

complexes formed between IHF and both *I1* and *R3* as a function of IHF concentration (Figure 14). $s_{20,w}$ values as a function of IHF concentration were fit to a simple Hill model as described in Materials and Methods. The fit to my *I1* data yielded a $K_{d,app}$ of 117 ± 25 nM, with n = 1.28. The fit to my *R3* data yielded a $K_{d,app}$ of 253 ± 108 nM, with n = 0.964.

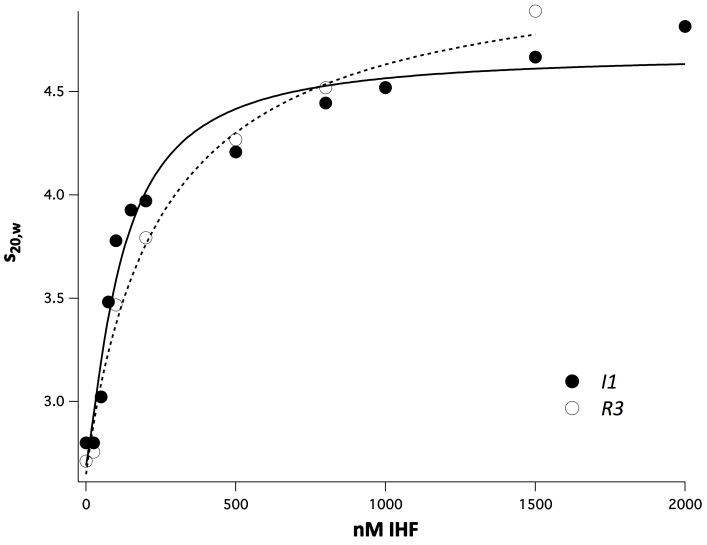


Figure 14. SV-AUC sedimentation coefficients of IHF:DNA complexes formed with *I1* and *R3*. $s_{20,w}$ sedimentation coefficients were plotted as a function of IHF concentration for both *I1* (●) and *R3* (○). Data were fit to a phenomenological Hill model, which yielded a $K_{d,app}$ of 117 nM and 253 nM for *I1* and *R3*, respectively.

The overall IHF concentration-dependent increases in sedimentation coefficients of *I1* and *R3* are not terribly different, with only a modest increase in $K_{d,app}$ observed for the nonspecific *R3* substrate. This data, in conjunction with observations from my EMS studies, suggest that, overall, I am monitoring relatively similar, nonspecific, binding events and the formation of similar (nonspecific) complexes; however, I am tempted to speculate that the slightly steeper rise for *I1*, as opposed to *R3*, in $s_{20,w}$ as a function of IHF concentration at lower IHF concentrations might indicate an initial higher-affinity, specific, binding event.

Discussion

In the present work, I have attempted to dissect both the specific and nonspecific interactions of the DNA binding and bending protein IHF with the bacteriophage λ packaging initiation site, or *cos*. IHF has been shown to bind to, and introduce a significant bend into specific sequences with matches to the IHF consensus sequence, including those found in *cos*[20]. IHF is also known to bind, but not necessarily bend, nonspecific sequences[38]. These activities of IHF are consistent with the fact that it belongs to a class of histone-like proteins, which are known to both nonspecifically and specifically bind DNA in a variety of structural, as well as regulatory, roles[56]. The complex *cos* site contains a number of elements with putative matches to the IHF consensus sequence, as well as significant stretches of duplex to which IHF can potentially bind to in a nonspecific manner. Using complementary biophysical techniques, I have attempted to expand our understanding of how IHF interacts with the *cos* site.

My EMS studies of 'full-length' duplexes showed a gel-retarded, high affinity ($K_{d,app}$ = 1.8 nM), complex between IHF and *cos274*. This duplex is a 274 bp DNA substrate containing both the *cosB* and *cosN* subsites, which are necessary for genome maturation and the initiation of packaging in bacteriophage λ (see Chapter I). IHF bound with lower affinity to a nonspecific duplex substrate of equivalent length (*ns274*). Approximately five times as much IHF is required for the complete disappearance of the free *ns274* DNA band in comparison to the free *cos274* band.

Importantly, nonspecific IHF-*ns274* complexes smeared diffusely in gels (with an IHF concentration-dependent upward movement of these smears), whereas the specific IHF-*cos274* complexes migrated as tight, significantly gel-retarded, bands which were not affected by increasing concentrations of IHF (save for extremely elevated concentrations). The latter is consistent with a bend introduced into the duplex, where the bent architecture largely determines the rate of migration of the bound DNA species. This behavior allows accurate determination of free vs. bound DNA and rigorous analysis of the data. The specific IHF-*cos274* band persisted at concentrations of up to 1 μM IHF, above which it also smeared upwards in the gel. Interestingly, sedimentation velocity (SV) analysis of these same substrates did not resolve this high-affinity binding interaction well. Instead, an IHF concentration-dependent increase in sedimentation coefficient of the complexes was observed for both *cos274* and *ns274* DNA substrates. While this result essentially mirrored what I observed in EMS studies for *ns274* - a continuous, IHF concentration-dependent, change in observed species - it was dramatically different from what appeared to be a single, high-affinity, binding and bending event between IHF and *cos274* in the gel. Given the large sedimentation coefficient for the "saturated" IHF-*cos274* species in SV studies, I posit that there are multiple IHF dimers binding, via a nonspecific binding mode, to both *ns274* and *cos274* at saturation. In fact, given the rather poor fits of my data to 1:1 stoichiometry binding models, even without the "free ligand approximation" (equation 7[57]), I suspect that all my techniques in the experiments presented here are measuring the binding of more than one IHF to any given substrate.

While SV is able to resolve the higher-order binding events, it either (1) does not have enough resolution, under these conditions, to resolve the high-affinity IHF-*cos274* interaction which likely involves only one or two IHF proteins, or (2) was not able to discern the specific binding event because of compensatory mass and size/shape changes upon IHF binding to, and bending of, the *cos274* duplex. I favor the first explanation. Conversely, I believe my EMS studies only resolve the first one or two IHF binding events of IHF to *cos274* that are required to fully bend the duplex. Once the duplex is fully bent, its architecture dominates the migratory properties of the complex in the gel; hence, higher-order complexes with increasing IHF concentration appear to migrate as the same tight, gel-retarded band (until extremely high concentrations of IHF). Since a bend is not introduced into *ns274* at lower IHF, its migration through the gel is dominated by charge and mass of the nucleoprotein complex formed; hence, I see a continuous upward movement of the diffuse smears that result from multiple IHF binding events.

In summary, EMS is able to discern the high-affinity specific complex formed between IHF and *cos274*, but not any subsequent nonspecific IHF binding events unless extremely elevated concentrations are used. SV, on the other hand, was not able to significantly discern the high-affinity specific binding event between IHF and *cos274*, but is able to detect multiple, higher-order, nonspecific complexes. Thus, EMS studies of *cos274* gives us an idea of the affinity of the specific IHF binding event, while SV studies of *cos274* gives us information on the macroscopic affinity of nonspecific IHF binding events (Figure 15).

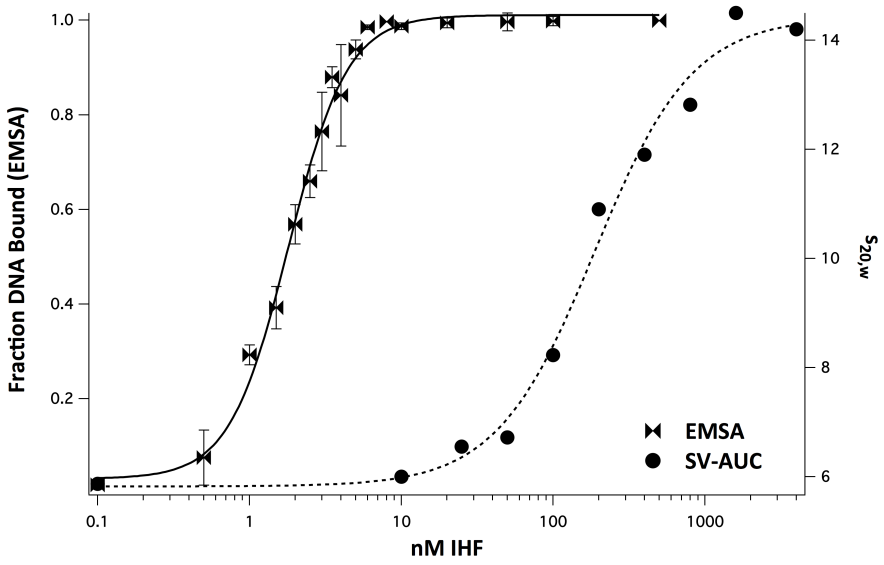


Figure 15. EMSA vs. SV studies of IHF binding to *cos274*. EMS studies (⧓) show us the specific IHF-*cos274* binding event, whereas SV (●) shows us the ensemble of nonspecific binding events.

To dissect these interactions further, I examined IHF binding to minimal duplex substrates. EMS studies of the minimal 27 bp *I1* duplex showed a gel-retarded, moderate affinity ($K_{d,app}$ = 20 nM) complex; I note that this sequence contains two overlapping sequences that match the original IHF consensus sequence defined by Craig and Nash[42] and efficiently binds IHF based on footprinting studies[41]. I interpret the tight, gel-retarded, nature of the IHF-*I1* band to indicate that a bend has been introduced into the duplex and, by definition, the formation of a specific complex. Surprisingly, *I2*, which shows considerable homology to the original IHF consensus and afforded moderate to weak protection against DNase I in footprinting studies, did not show the formation of a specific complex with IHF in my EMS studies. Instead, IHF-*I2* complexes produced a diffuse smear in the gel, similar to IHF-*ns274* complexes. My 27 bp nonspecific control, *R3*, also produced diffuse smears in the presence of IHF, as expected.

Similar findings are observed in my SV studies when comparing full-length specific and nonspecific substrates. SV does not resolve the specific IHF-*I1* binding event very well, as the sedimentation coefficients of complexes formed between IHF and *I1* or *R3* showed very similar IHF concentration dependencies. It appears, once more, that EMS is able to resolve the high-affinity, specific, IHF binding event due to its sensitivity to the bend introduced into the duplex upon specific binding; in contrast, SV was not able to significantly distinguish specific IHF binding to *I1* from nonspecific binding to *R3*. Rather, my SV results likely show a macroscopic binding curve and indicate that more than one IHF binds to the 27 bp minimal substrates. Since the terminal species has a sedimentation coefficient well below that observed at saturation with my full-length substrates, I presume fewer IHF molecules are bound at saturation; this is not surprising given that the minimal substrates are one-tenth the length of the full-length substrates. Interestingly, though, saturation of binding appears to occur at nearly the same concentration of IHF for both minimal and full-length substrates. One explanation for this observation is that the higher concentrations of nonspecific sites in the full-length duplexes simply drives nonspecific interactions with IHF; hence, higher concentrations of IHF are needed to reach the same stoichiometry of IHF:DNA on smaller substrates. This is supported by the observation that cooperativity is not suggested in my IHF-*ns274* binding data ($n = 0.96$); thus, it is feasible that there are multiple nonspecific binding sites in nonspecific duplexes, each with essentially the same affinity for IHF. Longer duplexes, then, simply present a higher concentration of DNA binding sites for IHF, shifting the binding isotherm to the left (such a model for nonspecific IHF binding

has been recently proposed[39]). The higher concentration of available nonspecific DNA binding sites would explain my observation that the longer *ns274* DNA substrate is completely bound at concentrations of IHF lower than that needed to see even 50% of the nonspecific, shorter, 27 bp duplexes bound (Figure 4B vs. Figure 11).

Although the general observation that SV did not significantly distinguish between specific and nonspecific complexes formed between IHF and my DNA substrates holds true, I would like comment further on the previously mentioned observation of a slightly faster rise in sedimentation coefficient, as a function of IHF concentration, for *I1* over *R3* (something that is also slightly evident in binding isotherms of IHF and *cos274* vs. *ns274*). Since nucleoprotein complexes between IHF and *I1* or *R3* at concentrations of IHF 500 nM and greater lead to sedimentation coefficients that are almost superimposable, it is interesting to fit the *I1* data to a Hill model using only the initial IHF concentrations of 0 – 200 nM (Figure 16). Using this approach, the fit $K_{d,app}$ = 69.4 nM, which is commensurate with the $K_{d,app}$ of 20.0 nM from EMS studies. Thus, I interpret the SV binding curve of *R3* as representative of underlying nonspecific interactions and the *I1* binding curve as a combination of this nonspecific binding interaction superimposed upon the higher affinity specific interaction with the *I1* consensus sequence. This initial high-affinity binding event is quickly masked by the continued progression of nonspecific binding at higher IHF concentrations; hence, I observe the IHF-*I1* nucleoprotein complexes exhibiting very similar sedimentation coefficients as the IHF-*R3* complexes at IHF concentrations of

500 nM and above. Further investigation will be required to determine the validity of this approach, but it is interesting nonetheless.

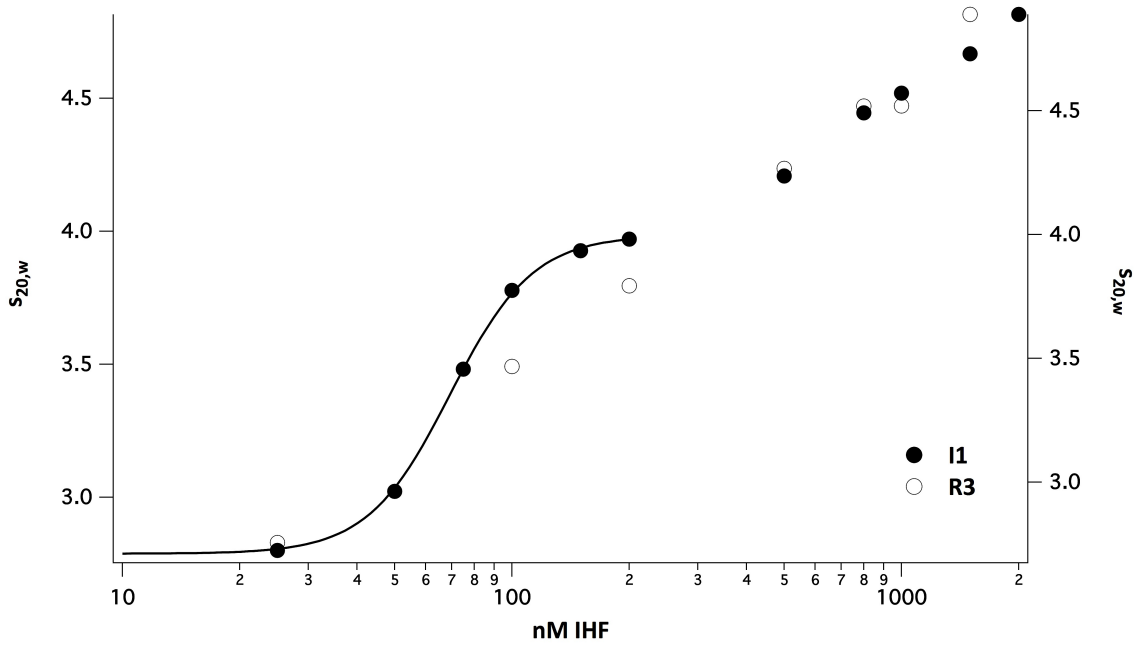


Figure 16. SV-AUC sedimentation coefficients of IHF:DNA complexes formed with *I1* and *R3*. $s_{20,w}$ sedimentation coefficients were plotted as a function of IHF concentration for both *I1* (●) and *R3* (○). Sedimentation coefficients of IHF-*I1* complexes formed between 0 – 200 nM IHF were fit to a Hill model, which yielded a $K_{d,app}$ of 69.4 nM.

While my EMS studies showed a specific complex formed between IHF and *I1*, the affinity of this complex (20 nM) appeared to be an order of magnitude lower than the affinity of IHF for *cos274* (1.8 nM). I questioned if this were due to (1) the presence of a second IHF binding site in *I2* that allowed for a high-affinity, cooperative interaction between IHF and *cos274*; (2) extra sequence on either side of my *I1* substrate that strengthens IHF binding; or (3) a length effect. Indeed, if one looks at the alignment of *I1* against the extended IHF consensus (Table 2), the match

is questionable at best unless you align the reverse complementary strand, denoted as: I1 3'→5'. Even *I2*, which binds IHF nonspecifically in my studies, appears to align better than *I1* to the extended IHF consensus. Hence, I used EMS to assay the binding affinity of IHF for two 75 bp substrates: *R3I1R2* and *I2R3I1*. Both 75 duplexes yielded high-affinity, gel-retarded, bands in complex with IHF, with affinities for IHF on the order of that displayed by *cos274*. I interpret this finding as follows: (1) there is no evidence for cooperativity between *I1* and *I2*, and so cooperativity between these sequences is not the explanation for the observed increased affinity of IHF for *cos274*; (2) at 75 bp, the length of the duplex does not play a large role in determining affinity for specific complexes, since both 75 bp specific duplexes display the same affinity for IHF as the 274 bp *cos274* duplex; and (3) there likely exists some sequence outside of my *I1* sequence that confers higher IHF binding affinity. Given the potential alignments in Table 2, and given that both *R3I1R2* and *I2R3I1* exhibit higher affinity than *I1*, I posit that the 'missing' sequence conferring higher IHF binding affinity to *I1* is upstream (in the 5' direction) of *I1* (that is, between *R3* and *I1*).

The high-affinity interactions of IHF with specific DNA duplexes as resolved by EMS studies, and the nonspecific interactions of IHF with both specific and nonspecific DNA duplexes as resolved by SV studies, are summarized in Table 4.

Substrate	EMS $K_{d,app}$	SV $K_{d,app}$
cos274	1.81 ± 0.08 nM	202 ± 33 nM
ns274	--	310 ± 16 nM
R3I1R2	2.4 ± 0.08 nM	--
I2R3I1	2.2 ± 0.08 nM	--
I1	20.0 ± 1.28 nM	117 ± 25 nM
R3	--	253 ± 108 nM

Table 4. Summary of binding constants derived from EMS and SV-AUC studies of IHF binding to specific and nonspecific sites within *cos*. Data from all EMS and sedimentation velocity (SV-AUC) studies in the present work were fit to a simple Hill model to extract binding constants for the interactions of IHF with specific and nonspecific sites within *cos*. The fit $K_{d,app}$ binding constants from these studies are summarized here.

In conclusion, the data presented in this chapter identify functional IHF binding elements within *cos* and define binding affinities of IHF for these specific sites vs. nonspecific DNA sequences. Within this context, Record *et al* put forth a model, based on their binding studies between IHF and its H' cognate site, suggesting that for low *specificity ratios* (K_{sp}/K_{ns}, where a ratio of 130 was considered to be very small), nonspecific IHF binding, where multiple IHFs assemble along the DNA without bending it, each with a smaller individual site size than what is observed for specific binding, dominates over specific binding in the presence of excess IHF [39]. However, given the persistence of the presumably bent, gel-retarded, complex between IHF and *cos* even at significantly high concentrations of IHF (IHF:DNA = 100:1) in my studies, I feel confident that the specific binding of IHF to *cos* is not easily disrupted by excess IHF. The absence of the slower sedimenting (~4 S) species in complex with IHF (at high protein concentrations) for *cos274*, as opposed to *ns274*, lends further credence to this hypothesis. In conjunction with the model

presented by Record *et. al* that suggests specific complexes are more likely to form when IHF is limiting, these observations suggest that carefully selecting concentrations of DNA and IHF will allow us to form specific IHF-*cos274* complexes upon which to titrate terminase on to in order to determine the assembly state of terminase in the ternary IHF•DNA•terminase complex. This work sets the stage for rational design of experiments that seek to directly characterize the physical nature of the IHF•DNA•terminase ternary complex as described in Chapter IV of this thesis.

III. Cooperative Assembly of Integration Host Factor and Terminase at the Bacteriophage λ Packaging Initiation Site

Introduction

Terminase, the packaging enzyme of bacteriophage λ, assembles in a site-specific manner at the packaging initiation site *cos* in order to mature and subsequently package viral DNA into preformed procapsids (see Chapter I). Crucial to this assembly process is the site-specificity of binding at *cos*, conferred by the gpNu1 small subunit of terminase. de Beer and coworkers used NMR structural data, along with a previously solved crystal structure of IHF bound to its cognate (specific) DNA site, to construct a model for the association of the gpNu1 DNA-binding domain with the R-elements at *cos* (Chapter I, Figure 3); these elements are specifically recognized by gpNu1[17]. The model predicts that gpNu1 binding introduces a severe bend at the *cos* site midway between the *R2* and *R3* elements. As it turns out, there exists an integration host factor (IHF) binding site, *I1*, precisely at this location. IHF is an *Escherichia coli* DNA binding protein known to introduce significant bends upon binding to specific sequences (see Chapter II: Introduction). IHF binds to the *I1* element to yield strong protection against DNase I in footprinting studies[41]. It was thus proposed that IHF might help stabilize a bent DNA structure to facilitate the binding of the small subunit of terminase at *cos*. Indeed, previous studies in our lab demonstrated that the small subunit of terminase and IHF do, in fact, cooperatively

bind and bend *cos* DNA, with up to a roughly five-fold increase in binding affinity of either protein when the other protein was present[20].

Having demonstrated the cooperativity of IHF and the small subunit of terminase, it became of interest to demonstrate this cooperativity with fully-functional terminase holoenzyme. Prior attempts to answer these questions have been frustrated by heterogeneous terminase preparations that potentially complicate rigorous thermodynamic analyses of binding affinity and specificity. More recently, our lab has established a purification protocol that yields homogenous, pure, terminase protomer as determined by analytical ultracentrifugation, gel filtration, and denaturing protein gel electrophoresis. Purification and biochemical characterization of relevant activities of this smallest functional unit of terminase have recently been published[58]. In the current chapter, I provide empirical evidence for the cooperative assembly of IHF and terminase protomer at *cos*.

Materials & Methods

IHF expression and purification

IHF was purified as described previously (Chapter II: Materials and Methods).

Expression and purification of terminase protomer

Terminase protomer was expressed from OR1265[pQH101] cells as described previously[59]. Expression of this heat-inducible vector yields full-length, native gpNu1 protein, as well as the native gpA protein with six histidines attached to the C-terminal end. Briefly, 20 ml of an overnight (32°C) culture of OR1265[pQH101] cells was added to 1 L of Terrific Broth (TB) medium, and grown at 32°C until an optical density (O.D.) of 0.6 at 600 nm was reached. At this time, an equal volume (1 L) of hot medium at 65°C was added to bring the cells to 45°C. After 15 minutes at 45°C, cells were allowed to grow for 1 additional hour at 42°C before being harvested. Note that all cultures contained 50 μg/ml ampicillin to maintain selection.

The cell pellet was resuspended in a suitable volume (~25-50 ml) of buffer H (20 mM Tris at pH 8, 7mM β-me, 500 mM NaCl, 10% glycerol, 20 mM imidazole) containing Protease Inhibitor Cocktail (Sigma-Aldrich). Cells were lysed using a French Press operated at 10,000 PSI. Insoluble cellular debris was removed by centrifugation (6000 x *g*, 40 minutes). The clarified supernatant was loaded onto a 5 ml HisTrap FF column, and terminase was eluted using a 20 column volume linear

gradient to buffer H with 500 mM imidazole. Eluted terminase (~150-200 mM imidazole) was diluted five-fold in buffer H, and dialyzed overnight at 4°C against buffer Q (20 mM sodium phosphate at pH 6.8, 1 mM EDTA, 7mM β-me, 100 mM NaCl, 10% glycerol). Insoluble material in the dialysate was removed by centrifugation (12,000 x *g*, 30 minutes), and the clarified supernatant was loaded onto a 1 ml HiTrap Q column. Bound terminase was eluted using a 20 column volume linear gradient to buffer Q containing 1 M NaCl. Fractions containing heterogeneous terminase (~400-450 mM NaCl) were pooled and stored at -80°C.

Terminase eluted off of the HiTrap Q column contained both homogenous protomer as well as a heterogeneous higher-order species. To isolate homogenous protomer, 1 ml of HiTrap Q column-purified terminase was applied to a 120 ml HiPrep S-300 HR gel filtration column previously equilibrated overnight with 2.4 column volumes of buffer Q. The S-300 column was run at 0.5 ml per minute over 1.2 column volumes, with homogenous terminase protomer eluting at ~65 ml. Fractions containing terminase protomer were pooled, aliquoted, and stored at -80°C. Terminase protomer concentration was determined spectrally (ε_{280} = 1.57 x 10^5 M^{-1} cm^{-1}).

Preparation of specific and nonspecific DNA duplexes

The 274 bp specific and nonspecific DNA substrates used in the present studies were prepared by large-scale preparative PCR. pAFP1 plasmid was used as a template in the PCR reactions, and was provided as a generous gift from Dr. Michael

Feiss (University of Iowa). This plasmid contains the entire *cos* sequence cloned into a pUC19 vector, and was purified from *E. coli* JM107[pAFP1] cells using a Qiagen Plasmid *Plus* Maxi kit. PCR primers were designed to amplify (1) the entire *cos* sequence, inclusive of the R-elements that bind the gpNu1 small subunit of terminase and the *I1* element recognized by IHF, and (2) an equivalent-length (274 bp) of random sequence (*ns274*). Primers were synthesized, and HPLC purified, by Integrated DNA Technologies (Coralville, IA); sequences are provided below.

Forward primer for PCR of *cos274*: CCGGAATTCGCATGCCTGCAGGTCTAA

Reverse primer for PCR of *cos274*: CGCGAATTCCATTGTTCATTCCAC

Forward primer for PCR of *ns274*: GGGGAATTCACTGGCCGTCGTTTTA

Reverse primer for PCR of *ns274*: GGGGAATTCACTTAACTATGCGGCATCAGAGCA

Primers were synthesized with fluorescent labels as appropriate when labeled substrates were required. PCR reaction mixtures (100 µl per tube) contained each primer at a concentration of 0.5 µM, 10 ng of pAFP1 template. *Taq* DNA polymerase, deoxynucleotides, and required buffer components were included as specified by the manufacturer's protocol for the *Taq* PCR kit (New England Biolabs). Total reaction volumes were 1 ml. PCR cycles were as follows: (94°C for 1 minute, 61°C for 1 minute, 72°C for 2 minutes) x 60 cycles. PCR reaction mixtures were subsequently pooled and purified using either: (1) PCR Kleen Spin Columns (Bio-Rad, Hercules CA) for IR700-labeled PCR products, or (2) a Promega Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison WI) for all other

products. Purity of amplified products was verified by agarose gel analysis, and the concentration of DNA was determined via spectroscopic methods. The sequences of the specific (*cos274*) and nonspecific (*ns274*) PCR products are enumerated below.

cos274:

CCGGAATTCGCATGCCTGCAGGTCTAATCATTATCACTTTACGGGTCCTTTCCGGTGATC
CGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTT
AAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAA
AGAAAGGAAACGACAGGTGCTGAAAGCGAGGCTTTTTGGCCTCTGTCGTTTCCTTTCTCT
GTTTTTGTCCGTGGAATGAACAATGGAATTCGCG

***ns274*:**

GGGGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCA
ACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTA
TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAAT
CTGCTCTGATGCCGCATAGTTAAGTGAATTCCCC

The majority of the studies presented monitored the binding of protein to relevant DNA species to measure the affinities of protein-DNA interactions of interest. I chose to monitor the DNA species, and any changes to it (e.g. increase in mass) upon protein binding, as a method for determining the fraction of bound DNA at any given loading protein concentration. Given the high affinities of these interactions, it was necessary to label the DNA substrates to allow detection of very small quantities.

Therefore, electrophoretic mobility shift (EMS) studies employing nanomolar concentrations of DNA were conducted using IR700[46]-labeled duplexes, and subsequently detected using a LI-COR Odyssey scanner[47]. Sedimentation velocity (SV) studies employed nanomolar concentrations of 6-FAM[48]-labeled DNA, which could be detected using the Aviv Biomedical fluorescence detection system. When EMS studies required picomolar concentrations of DNA, radiolabeled DNA substrate was used and detected via exposure of a storage phosphor screen that could be scanned on a Molecular Dynamics Storm system. Briefly, purified PCR products were digested with *Eco*R1 restriction endonuclease, and the digested product was purified using a Bio-Spin 6 Spin Column (Bio-Rad, Hercules CA). The digested substrate was then 3'-end-labeled on both ends using α^{32}P-dATP and the Klenow fragment as per the manufacturer's protocol.

Molecular weights of DNA substrates were calculated as needed, based on their sequences, using an online calculator[49]. Substrates with a 6-FAM[48] label had 537.5 Da added to their molecular weight.

Electrophoretic mobility shift (EMS) studies

Equilibrium binding experiments were performed in buffer containing 20mM Tris, pH 8, 1mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 10% glycerol. DNA was included at a final concentration of either 2 or 4 nM, and protein was added to the binding mixtures to appropriate concentrations. Mixtures were incubated at room temperature for 20 minutes, and then loaded, unless otherwise noted, onto an

8% polyacrylamide gel. Unless otherwise noted, an acrylamide/bis-acrylamide ratio of 80:1 was used so as to facilitate entry of the large complexes into the gel. Gels were run at 15V/cm in 0.5x TBE at 4°C for 1 hour. Gels in EMS studies employing IR700-labeled DNA were scanned on a LI-COR Odyssey scanner using appropriate integration times so as to maximize signal:noise ratio of bands without clipping any signals. Gels in EMS studies employing radiolabeled DNA were dried *in vacuo* on Whatman 3MM filter paper, and subsequently used to expose a storage phosphor screen. After an appropriate length of exposure, storage phosphor screens were scanned on a Molecular Dynamics Storm system.

For all EMS studies, bands were quantified using *ImageQuant TL* software (GE Healthcare Life Sciences).

The fraction of bound DNA was calculated using:

$$Fraction\ Bound = \frac{counts\ in\ retarded\ band}{counts\ in\ retarded\ band + counds\ in\ free\ DNA\ band}$$

The raw binding data were fit to a simple Hill model, according to:

$$Fraction\ Bound = (m - b) * \left[\frac{[P]^n}{{K_d}^n + [P]^n}\right] + b$$

where [P] is the protein concentration, K_d is the apparent dissociation constant, *n* is the Hill coefficient, b is the baseline offset, and m is the fraction of DNA bound at saturation. Experimental data were fit using Igor Pro (Wave Metrics, Lake Oswego, OR).

Sedimentation velocity experiments

Binding studies using sedimentation velocity (SV) were carried out in buffer containing 20 mM Tris at pH 8, 10 mM sodium phosphate, 1 mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 5% glycerol. Experiments were conducted on a Beckman Coulter XLI ultracentrifuge retrofitted with an Aviv Biomedical fluorescence detection system (FDS). The FDS uses a continuous wave solid-state laser to excite samples at 488 nm, and uses a band-pass filter to detect any and all emission falling between 505 and 565 nm. Data were collected at 20 μm intervals, with 5 averages per position. 400 μl of sample were loaded into preassembled analytical cells containing 12mm Epon charcoal-filled double sector centerpieces. Seven analytical cells were loaded into an An-50 Ti titanium rotor, along with one reference cell ('calibration center piece') containing a solution of fluorescein used by the FDS for both angle and radial calibration (similar to the function of the Beckman Coulter counterweight). Experiments were performed at 7°C, at 42K RPM. The photomultiplier tube gain was adjusted optimally so as to attain the highest signal-to-noise ratio for my samples without clipping the signal for the calibration center piece. The latter can be of concern because (a) every cell is scanned during each revolution of the rotor, and (b) the electronics in the FDS have a finite recovery time

and can, therefore, be susceptible to not recovering from a saturated signal in time for the next cell being scanned. The finite response time of the electronics is also the reason for the recommendation that all cells be scanned at the same gain setting (personal communication with Aviv Biomedical).

The sedimentation velocity data were analyzed with SEDFIT[55] using a continuous c(s) approach. SEDFIT allows for the estimation of both the sedimentation coefficient *s* (from the rate of sedimentation of the sample, or, *u*, the velocity of the boundary) and the diffusion coefficient *D* (from the rate of the spreading of the boundary during sedimentation). The sedimentation coefficient *s* is related to the molecular weight *M* and the frictional coefficient *f* of the sedimenting species by the following relation (Svedberg equation):

$$s = \frac{u}{\omega^2 r} = \frac{MD(1-\bar{v}\rho)}{RT} = \frac{M(1-\bar{v}\rho)}{Nf}$$

In the above equation, ω is the angular velocity of the rotor, *r* is the radial distance for any given measurement, *R* is the gas constant, *T* is the absolute temperature, *N* is Avogadro's number, $\bar{v}$ is the partial specific volume, and ρ is the buffer density. The buffer density was calculated based on buffer components using SEDNTERP[52].

Since the FDS allowed me to monitor only the fluorescein-tagged DNA (and any nucleoprotein complex formed with the labeled DNA), the sedimentation coefficient

was used as a measure of the fraction of bound DNA. Binding of protein to DNA typically leads to a change in s due to either a mass increase, a change in frictional coefficient, or both.

Furthermore, for simple, homogenous species, the fit s and f values can be used to determine the molecular weight M (if $\bar{v}$ is known), or the partial specific volume $\bar{v}$ (if M is known). Determining these values in protein-DNA binding studies, however, is not straightforward. In complicated systems where multiple species exist, the determination of either M or $\bar{v}$ is rendered difficult by the fact that the fit f is an average of the frictional coefficients of all present species. Complicating such an analysis even further is the potential for different partial specific volumes of the various species formed between proteins and DNA. Therefore, without prior knowledge of the exact nucleoprotein species present in any binding mixture, it was not possible for me to rigorously assign a mass to any complexes formed in my protein-DNA SV binding studies. In the absence of a rigorous determination of the masses of complexes formed between protein and DNA, I was unable to estimate protein stoichiometries on DNA. Hence, the SV data were fit to a simple Hill binding model:

$$sedimentation\ coefficient = (s_{sat} - s_{ini}) * \frac{E^n}{{K_d}^n + E^n} + s_{ini}$$

where s_{ini} is the sedimentation coefficient of free DNA, s_{sat} is the sedimentation coefficient of the saturated nucleoprotein species, K_d is the apparent dissociation constant, *and n* is the Hill coefficient. Experimental data were fit using Igor Pro (Wave Metrics, Lake Oswego, OR).

I chose, for the sake of simplicity, to not correct the fit sedimentation coefficients to standard conditions (water, 20°C), as this calculation requires prior knowledge of the partial specific volume of the sedimenting species. For the various nucleoprotein complexes that result in a protein-DNA binding study, a calculation of the weight-average partial specific volume requires prior knowledge of the stoichiometry of the nucleoprotein complex. Since stoichiometries were not directly measured in my studies, I simply report the sedimentation coefficients of the major species.

Sedimentation equilibrium experiments

Binding studies using sedimentation equilibrium (SE) were carried out in buffer containing 20 mM Tris at pH 8, 10 mM sodium phosphate, 1 mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 5% glycerol. Experiments were conducted on a Beckman Coulter XLI ultracentrifuge retrofitted with an Aviv Biomedical fluorescence detection system (FDS), as described above (see *Sedimentation velocity experiments*), with the following differences: 100-120 µl of sample were loaded into preassembled analytical cells containing Epon charcoal-filled six-channel centerpieces (with external fill). Experiments were performed at 7°C, with a rotor speed of 6K RPM. This speed was chosen so as to achieve a σ value (see equation

below relating σ to angular velocity of the rotor, ω) of 2 for the suspected nucleoprotein complex formed in the binding studies. Samples were periodically checked for approach to equilibrium using the program WINMATCH (David Yphantis, University of Connecticut; Michael Johnson, University of Virginia; Jeff Lary, National Analytical Ultracentrifuge Center, Storrs CT). Data from samples having reached equilibrium were analyzed using the nonlinear least-squares analysis program WINNONLIN (authors of WINMATCH) to a single species model, given by:

$$F_r = F_{ref} * e^{\sigma z}, where\ z = \frac{r^2 - {r_{ref}}^2}{2}\ and\ \sigma = \frac{M(1-\bar{\upsilon}\rho)}{RT} * \omega^2$$

where F_r is the total fluorescence counts of the sample at radial position r, F_{ref} is the fluorescence counts of the sample at the reference radial position r_{ref}, M and $\bar{\upsilon}$ are the molecular weight and partial specific volume of the monitored species, respectively, ρ is the buffer density, R is the gas constant, T is the absolute temperature, and ω is the angular velocity of the rotor. The density and viscosity for buffer compositions were calculated using SEDNTERP[52], yielding 1.01849 and 0.01683, respectively. Due to an unexplained systemic error leading to severe underestimation of molecular weights of species from the fit σ in my experiments, I do not report molecular weights of species in binding studies monitored via SE. Instead, experimentally fit σ values were used as a measure of extents of binding.

Results

EMS studies indicate that terminase protomer forms similar, nonspecific complexes with both specific and nonspecific sequences of equal length.

I first show, using an electrophoretic mobility shift (EMS), assay, that terminase protomer binds 40 pM radiolabeled *cos274* (a "specific" 274 bp duplex substrate that models the full-length *cos* sequence found in concatemeric λ DNA), as well as 40 pM radiolabeled nonspecific DNA duplex of equal length (*ns274*), to afford complexes that run as diffuse smears in the gel (Figure 1). The mobility of these smears is inversely related to the loading concentration of terminase.

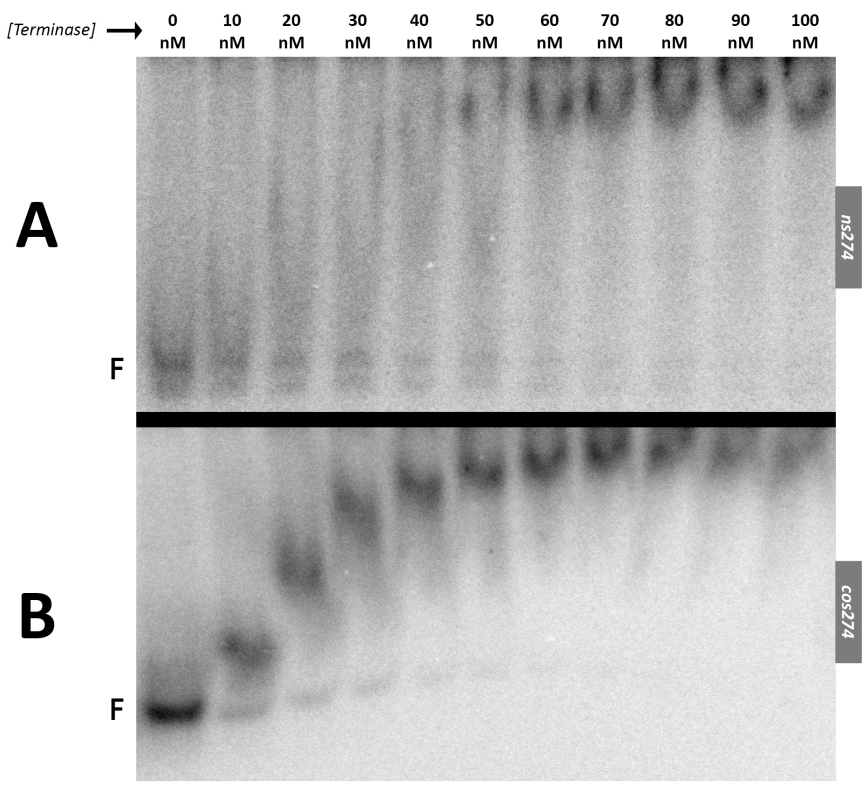


Figure 1. Terminase forms similar, nonspecific complexes with both specific and nonspecific sequences of equal length. *Panel A:* Terminase protomer titration against 40 pM nonspecific *ns274* DNA substrate. *Panel B:* Terminase protomer titration against 40 pM specific *cos274* DNA substrate. Terminase titration range: 0 - 100 nM.

The diffuse smears could be indicative of either (1) nucleoprotein complexes of varying stoichiometry, or (2) weak nucleoprotein complexes that fall apart in the gel, with increasing concentrations of protein simply pushing the binding equilibrium to the right. I favor the second explanation, as the first explanation would require that each discrete loading concentration of terminase leads to a complex of different stoichiometry, a possibility I find to be too coincidental.

Given the difficulty of quantifying free vs. bound DNA from these gels, I decided not to attempt to derive a binding curve from this data. Hence, I could not derive binding affinities of terminase for *cos274* and *ns274* from these experiments.

EMS studies indicate that IHF increases the binding affinity of terminase for cos-containing DNA

Terminase protomer was titrated onto 40 pM radiolabeled *cos274* in the presence of either 0.2 nM, 1.25 nM, or 2 nM IHF. The representative gels from these studies are presented in Figure 2. It is clear from these studies that the presence of IHF substantially increases the binding affinity of terminase for *cos274*. This effect was not observed for the nonspecific duplex of equivalent length, *ns274* (data not shown). It should be noted that full terminase binding curves were difficult to generate for lower IHF concentrations because the binding interaction did not afford a discrete band, but rather diffuse, smeared bands. I interpret this to indicate

that specific binding of terminase at the *cos* site (represented by a tight, gel-retarded band) competes with nonspecific duplex binding (leading to an upward moving smear of the free DNA band). In other words, terminase does not efficiently discriminate between non-specific and *cos* DNA, and the ratio of specific to nonspecific binding of terminase to *cos274* depended heavily on the amount of IHF present. This is not surprising, given that in the complete absence of IHF, terminase forms only weak, nonspecific smears (Figure 1).

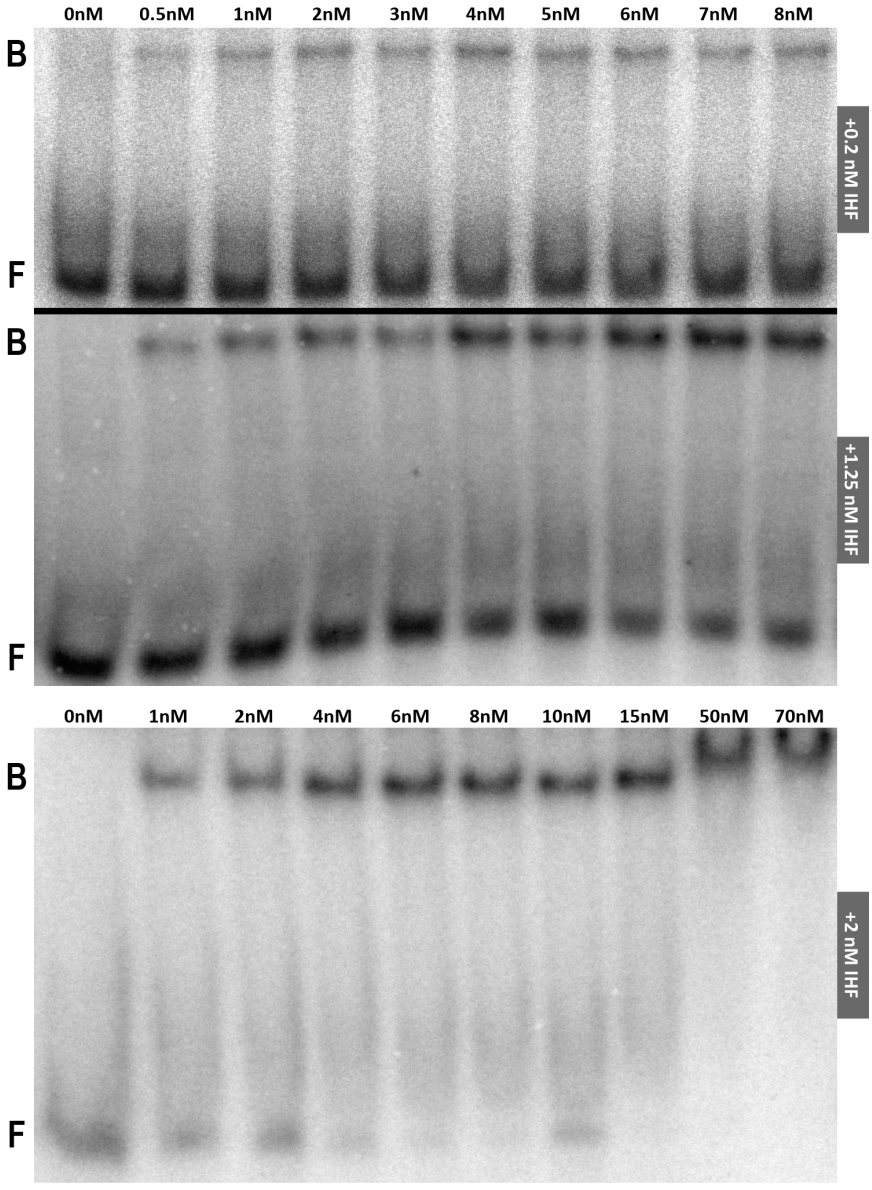


Figure 2. Terminase titrations on 40 pM radiolabeled *cos274*. *Top panel:* Terminase was titrated (0 - 8 nM) in the presence of 0.2 nM IHF. *Middle panel:* Terminase was titrated (0 - 8 nM) in the presence of 1.25 nM IHF. *Bottom panel:* Terminase was titrated (0 - 70 nM) in the presence of 2 nM IHF.

Of note, while a discrete IHF•DNA•terminase complex is observed in the presence of 0.2 nM IHF, higher concentrations of terminase afford an additional, diffuse band directly above the free DNA band. At these latter higher concentrations of terminase, I believe there was no longer enough IHF to form a specific ternary complex with terminase and *cos274*. This could occur for a few of reasons: (1) at such low loading concentrations as 200 pM, it is feasible that I did not actually have 200 pM IHF available for binding in the reaction tube; (2) since the activity of the particular prep of IHF used in these studies was not known, it is feasible that I did not, in fact, have 200 pM active protein; and (3) even in the presence of terminase, the K_d of IHF for *cos274* may not have dropped low enough to see saturation of IHF-*cos274* binding. Any single or combination of these possibilities might explain the lack in significant increase of the bound DNA band as a function of terminase for the 0.2 nM IHF experiment (○).

The EMS data were quantified and fit to a Hill model as described in Materials and Methods (Figure 3).

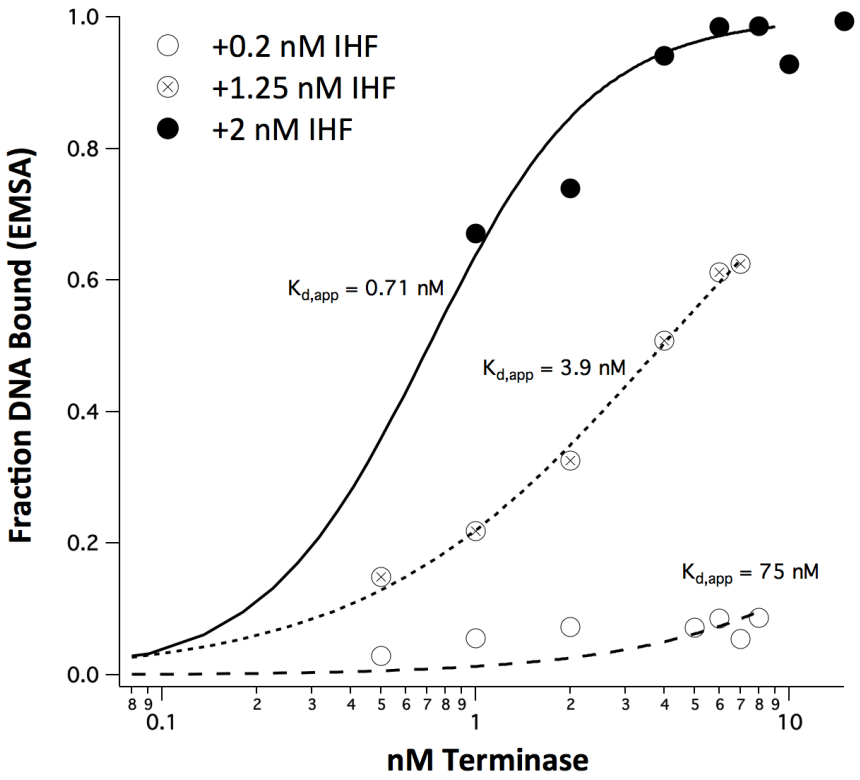


Figure 3. Quantitation of EMS studies of terminase binding to *cos274* in the presence of increasing concentrations of IHF. EMS studies in Figure 2 were quantitated and fit to a Hill model, as described in Materials and Methods. The terminase-*cos274* interaction displays a $K_{d,app}$ of 75 nM in the presence of 0.2 nM IHF (○), a $K_{d,app}$ of 3.9 nM in the presence of 1.25 nM IHF (⊗), and a $K_{d,app}$ of 0.71 nM in the presence of 2 nM IHF (●).

Despite the incomplete titrations for experiments in the presence of 0.2 nM and 1.25 nM IHF, I can fit the data to get an idea of a potential binding affinity for terminase at *cos*. The fit of the 0.2 nM IHF experiment yielded a $K_{d,app}$ of 75 ± 14 nM, with n = 1. The fit of the 1.25 nM IHF experiment yielded a $K_{d,app}$ of 3.9 ± 0.13 nM, with n = 0.93. Finally, the fit of the more complete titration in the presence of 2 nM IHF yielded a $K_{d,app}$ of 0.71 ± 0.08 nM, with n = 1.65. For ease of visualization, $K_{d,app}$ binding constants for terminase on *cos274* as a function of IHF concentration are plotted in

Figure 4. The ratio of the $K_{d,app}$ for the terminase-*cos274* interaction in the presence of 0.2 nM IHF vs. 2 nM IHF is $\frac{75\ nM}{0.71\ nM}$, or 106. The ratio of the $K_{d,app}$ for the terminase-*cos274* interaction in the presence of 1.25 nM IHF vs. 2 nM IHF is $\frac{3.9\ nM}{0.71\ nM}$, or 5.5. Although the 0.2 nM IHF binding curve is likely the most representative of specific terminase binding *in the complete absence of IHF* to *cos274*, given the severity of incompleteness of my data for this experiment, I choose not to jump to the conclusion that IHF increases the binding affinity of terminase for *cos274* two orders of magnitude; the fit $K_{d,app}$ is only a rough estimate (due to the lack of a proper binding curve, and the potential caveats of the 0.2 nM IHF experiment enumerated above). However, the roughly five-fold increase in terminase binding affinity for *cos274* in the presence of 2 nM IHF as opposed to 1.25 nM IHF is not an unreasonable finding. Previous studies in our lab found that doubling the IHF concentration from 0.1 to 0.2 nM increased the binding affinity of the small subunit of terminase for *cos* approximately two-and-a-half-fold[20].

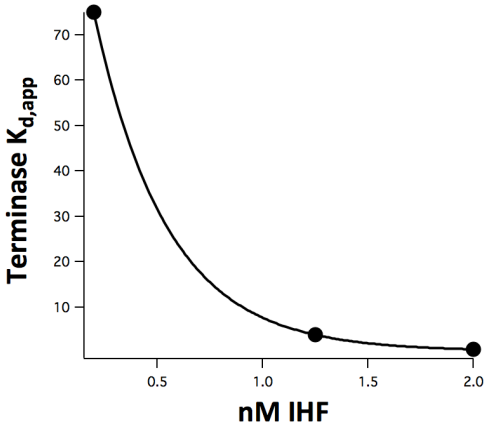


Figure 4. Apparent dissociation constants for terminase on *cos274* DNA as a function of concentration of IHF present. Fit $K_{d,app}$ binding constants from Figure 3 were plotted as a function of IHF present in the binding reactions. This plot gives a very rough idea of the effect IHF has on terminase binding at *cos*. Briefly, IHF appears to increase the binding affinity of terminase for *cos*-containing DNA, alluding to the cooperativity of terminase and IHF assembly at *cos*.

Of note is the fact that I was not able to directly measure the affinity of terminase for *cos274* in the complete absence of IHF, due to the nonspecific, diffuse smears that terminase-*cos274* complexes displayed (see Figure 1B).

EMS studies of terminase protomer interactions with the specific cos274 and the nonspecific ns274 DNA substrates indicate a modest preference for the specific substrate.

While my previous studies afforded me estimates of the binding affinity of terminase for *cos274* in the presence of IHF, I was not able to determine terminase affinities in the absence of IHF due to the diffuse smearing of bands resulting from terminase complexes with both specific and nonspecific substrates (Figure 1). I reasoned that a gel with higher crosslinking ratio that held any formed nucleoproteins up in the well might aid in quantitation and, therefore, determination of binding affinities of terminase for these substrates. Furthermore, given my desire to transition to a more rigorous, solution-based technique for the measurement of binding affinities, I designed conditions of the following experiment to be commensurate with the analytical ultracentrifugation (AUC) studies detailed in subsequent sections. An 8% native polyacrylamide gel with a 29:1 acrylamide:bis-acrylamide ratio was used to hold terminase•DNA nucleoprotein complexes in the wells. Terminase was titrated (0 - 250 nM) against either 4 nM fluorescently-tagged *cos274* or *ns274* (Figure 5A and 5B, respectively), as 4 nM DNA

was the lower limit of substrate concentration affording me a reasonable signal in latter AUC studies. Free vs. bound DNA were quantitated to generate binding isotherms, and data were fit to a Hill model, as described in Materials and Methods. Terminase shows a moderate preference for *cos274* over *ns274*, evidenced by the full disappearance of free *cos274* at concentrations where free *ns274* still remains (Figure 5A, 5B). The fits generated from the binding isotherms yielded a $K_{d,app}$ of 56.4 ± 2.6 nM (n = 2.2) and 212 ± 88 nM (n = 1.4) for *cos274* and *ns274*, respectively (Fig. 5C).

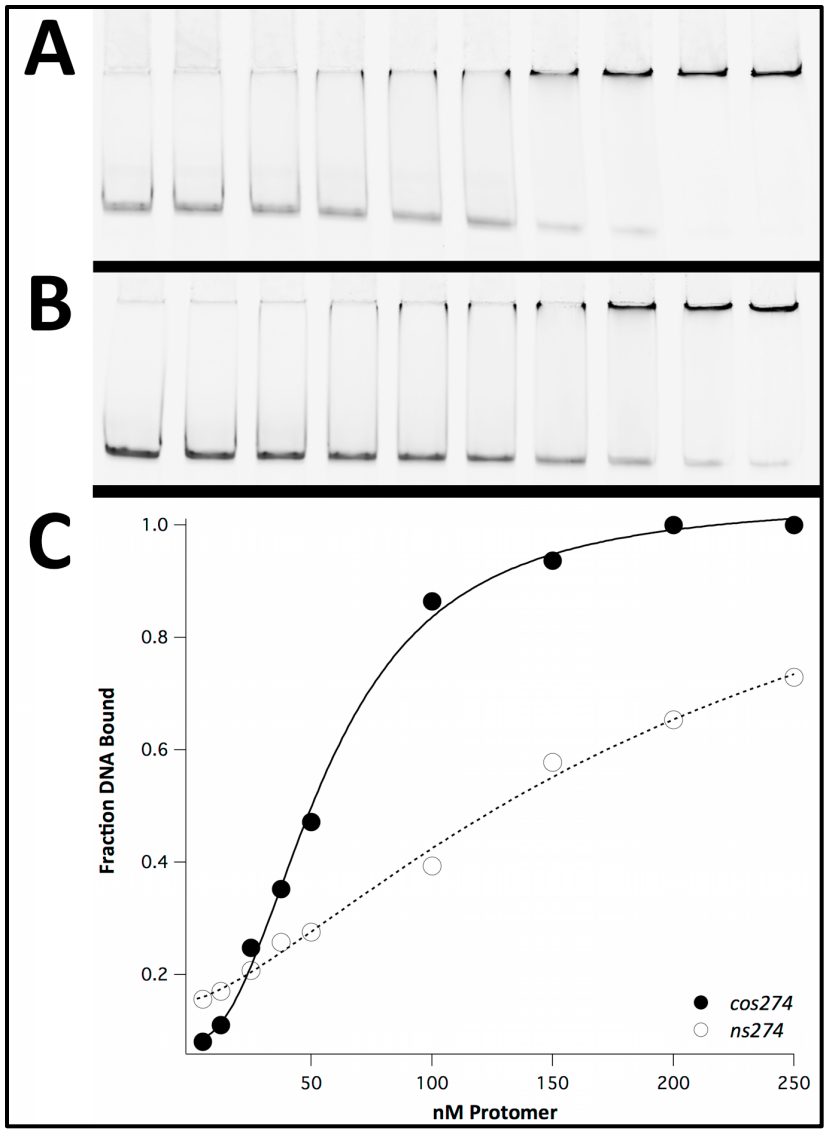


Figure 5. EMS studies of terminase binding to *cos274* and *ns274* in gels of high crosslinking ratio. *Panel A:* Terminase protomer titration against 4 nM specific *cos274* DNA substrate. *Panel B:* Terminase protomer titration against 4 nM nonspecific *ns274* DNA substrate. Terminase was titrated from 0 to 250 nM in both experiments. *Panel C:* The gels were quantitated and fit to a Hill model as described in Materials and Methods. The terminase-*cos274* (●) interaction yielded a $K_{d,app}$ of 56.4 nM, while the nonspecific interaction (○) yielded a $K_{d,app}$ of 212 nM.

EMS studies indicate that the presence of terminase increases the binding affinity of IHF for cos-containing DNA.

Having previously shown evidence for the cooperativity between IHF and terminase in the assembly of a ternary IHF•DNA•terminase complex (Figures 2 and 3), I performed an EMS study of an inverted scenario to further demonstrate the effect. Here, I titrate IHF onto *cos274* in the presence of a fixed concentration of terminase that, by itself, shows little (and, importantly, nonspecific) binding to *cos274*.

I examined the interaction of IHF, in both the presence and absence of terminase protomer, with the specific *cos274* DNA substrate. These studies utilized 4 nM fluorescently-tagged *cos274* in electrophoretic mobility shift (EMS) studies as described in Materials and Methods (NativePAGE™ Novex® 3-12% Bis-Tris gels were used for these studies). IHF binds to this duplex to form a tight, specific, gel-retarded band in a concentration-dependent fashion (Figure 6A). In the presence of 50 nM terminase protomer, IHF similarly binds to *cos274* to form a specific, gel-retarded, nucleoprotein complex that migrates in a distinctly different fashion (discrete, but more diffuse) than the IHF-*cos274* complex (compare Figures 6A and 6B).

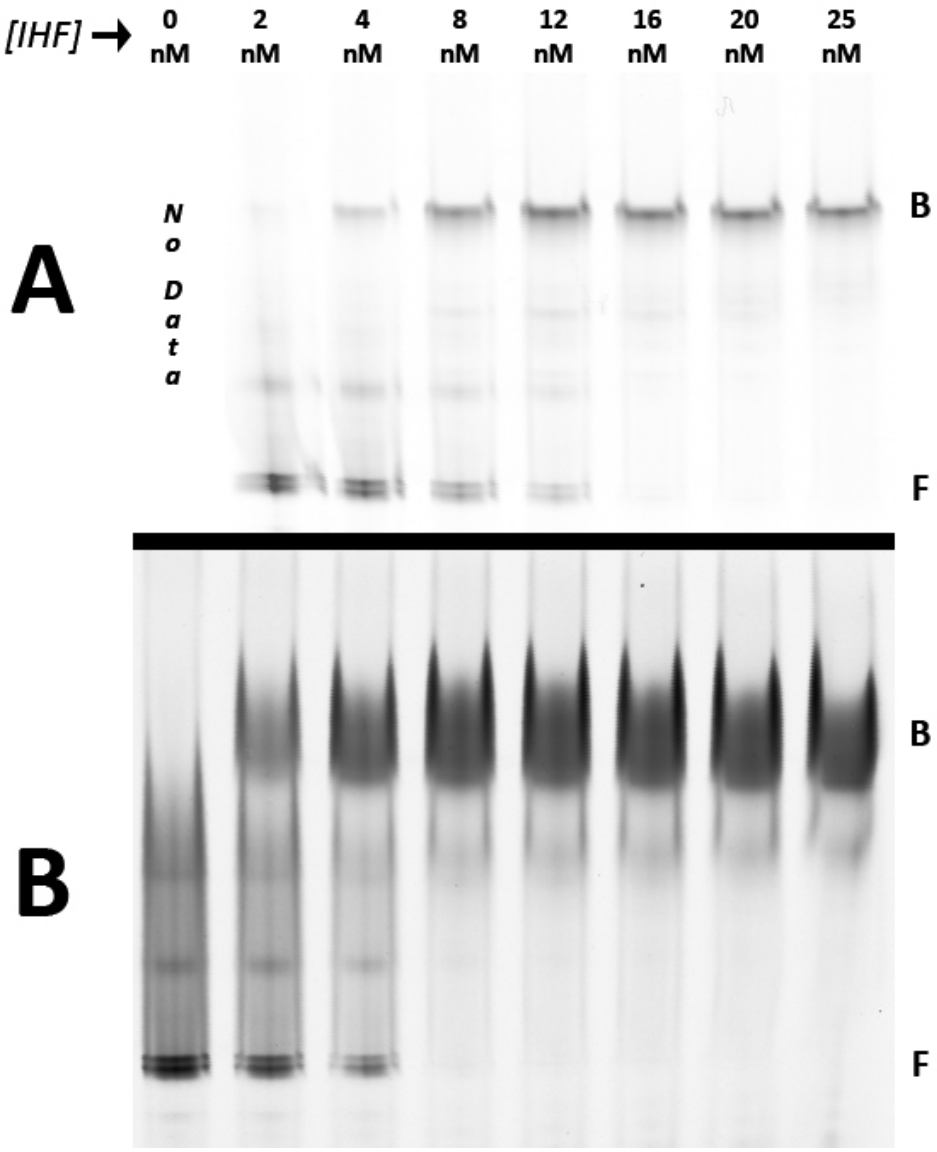


Figure 6. EMS studies of IHF-*cos274* interactions in the absence and presence of terminase. *Panel A*: IHF binds to *cos274* in the absence of terminase to form a tight, gel-retarded, band in a concentration-dependent manner. *Panel B*: IHF binds to *cos274* in the presence of terminase to form a diffuse, yet discrete, gel-retarded band in a concentration-dependent manner.

First, I posit that the tight, gel-retarded band in Figure 6A is a specific IHF-*cos274* binary complex, based on my studies presented in Chapter II as well as previous studies in our lab[20]. Second, I posit that the more diffuse, yet still discrete, gel-

retarded band near the top of the gel in Figure 6B is an IHF•DNA•terminase ternary nucleoprotein complex. This is supported by the following observations: (1) the diffuse band has an upper limit of mobility matching the mobility of the IHF-*cos274* band in Figure 6A; and (2) the diffuse smear lower in the gel, directly above free DNA, in the first lane of Figure 6B, which presumably represents weak interactions between terminase and DNA in the absence of IHF, disappears in an IHF concentration-dependent manner (contributing to the more discrete, yet still diffuse, band near the top of the gel).

Given these assumptions, the presence of 50 nM terminase affords a fraction of bound DNA at any given IHF concentration that is consistently greater than in its absence. I interpret this finding to indicate that lower concentrations of IHF are needed to form an amount of ternary IHF•DNA•terminase nucleoprotein complex equivalent to the amount of binary IHF•DNA complex formed at higher concentrations of IHF. Put another way, the amount of ternary IHF•DNA•terminase at any given concentration of IHF is higher than the amount of binary IHF•DNA complex formed at that IHF concentration. This indicates cooperativity between the two proteins on DNA.

Furthermore, note the diffuse, yet discrete, gel-retarded band at 2 nM IHF in Panel B of Figure 6, but the lack thereof in the corresponding lane in Panel A. The only difference between these binding mixtures is the presence of 50 nM terminase in the former. Hence, it is evident from this comparison that IHF increases the *specific*

binding affinity of terminase for *cos274*. Important in the drawing of this conclusion is the absence of this band in the absence of IHF (lane 1, Figure 6B). It appears that in the absence of IHF, 50 nM terminase binds *cos274* weakly and nonspecifically to afford a smear in the gel (lane 1, Figure 6B); in contrast, the presence of as little as 2 nM IHF affords a significant amount of a gel-retarded (albeit somewhat diffuse) band (lane 2, Figure 6B) that is not present at this concentration of IHF in the absence of terminase (lane 2, Figure 6A).

In order to more rigorously compare the effect of terminase on the apparent binding affinity of IHF on *cos274*, I quantitated and fit the EMS data to a Hill model as described in Materials and Methods. However, a special note needs to be made for the data presented in Figure 6. There are, in fact, two bands both running near the proper position for a 274 bp duplex substrate (as determined by molecular weight standards, data not shown). Since both of these substrates appear to disappear at the same rate with increasing protein concentration, I made the reasonable assumption that both duplexes represented by these bands contain the full *cos* sequence necessary for IHF and terminase binding and were simply two very similar PCR products. Note that the present studies were carried out in gradient gels capable of resolving the presumably minor differences between these two duplexes; all other gel studies in non-gradient 8% polyacrylamide gels showed the co-migration of these two duplexes. For my analysis, then, I chose to quantify free DNA as the summation of both bands (marked 'F' in Figure 6). IHF-containing nucleoprotein complex (representative of 'bound DNA') was considered to be either

the tight band, or the diffuse but distinct band, labeled 'B' in Figures 6A and 6B, respectively. The quantitated EMS data from Figure 6 and generated fits are shown in Figure 7.

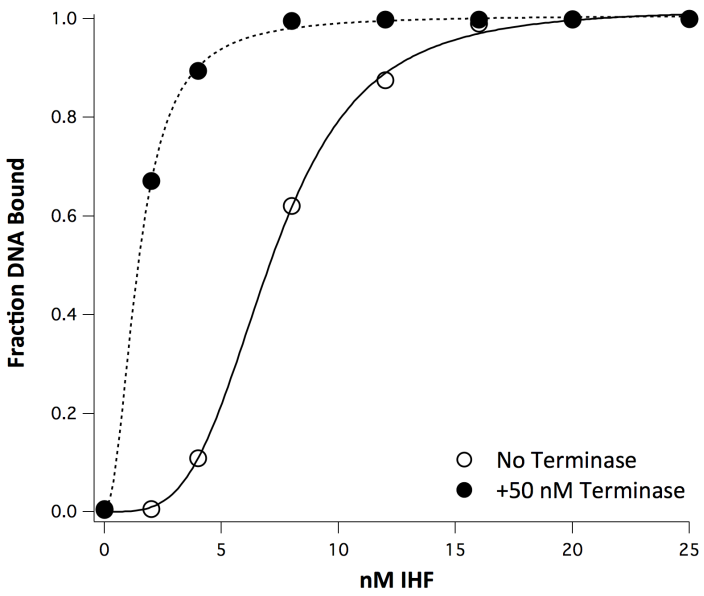


Figure 7. Quantitation of EMS studies for IHF binding to *cos274* in the absence and presence of terminase. EMS studies were quantitated and fit to a Hill model, as described in Materials and Methods. The IHF-*cos274* interaction displays a $K_{d,app}$ of 7.12 nM in the absence of terminase (○), and a $K_{d,app}$ of 1.45 nM in the presence of 50 nM terminase (●). Thus, I note a 5-fold increase in binding affinity of IHF for *cos274* in the presence of 50 nM terminase.

In the absence of terminase, the fit of my EMS data to a Hill model yielded a $K_{d,app}$ of 7.12 ± 0.124 nM[1], with n = 3.69. In the presence of 50 nM terminase protomer, the fit yielded a $K_{d,app}$ of 1.45 ± 0.05 nM, with n = 2.11. The data suggest a nearly five-fold increase in binding affinity of IHF for *cos274* in the presence of 50 nM terminase protomer. Notably, this is commensurate with the five-fold increase in binding

[1] The reason this $K_{d,app}$ is larger than the $K_{d,app}$ of 1.8 nM I obtained for the same DNA substrate in Chapter II (Fig. 4C) is due to the increased DNA concentration. In these studies I used 4 nM DNA in order to match the lowest concentration of DNA we could use in subsequent sedimentation velocity experiments. A more appropriate fit to the Hill model without the use of the "free ligand approximation" yields sub-nanomolar $K_{d,app}$ values for IHF interactions at *cos* that are consistent across all EMS studies. For simplicity, I have not included those fits in the present studies, as they are irrelevant for the comparative analyses of binding constants made here (since comparisons are only drawn between experiments utilizing 4 nM DNA substrate, consistently).

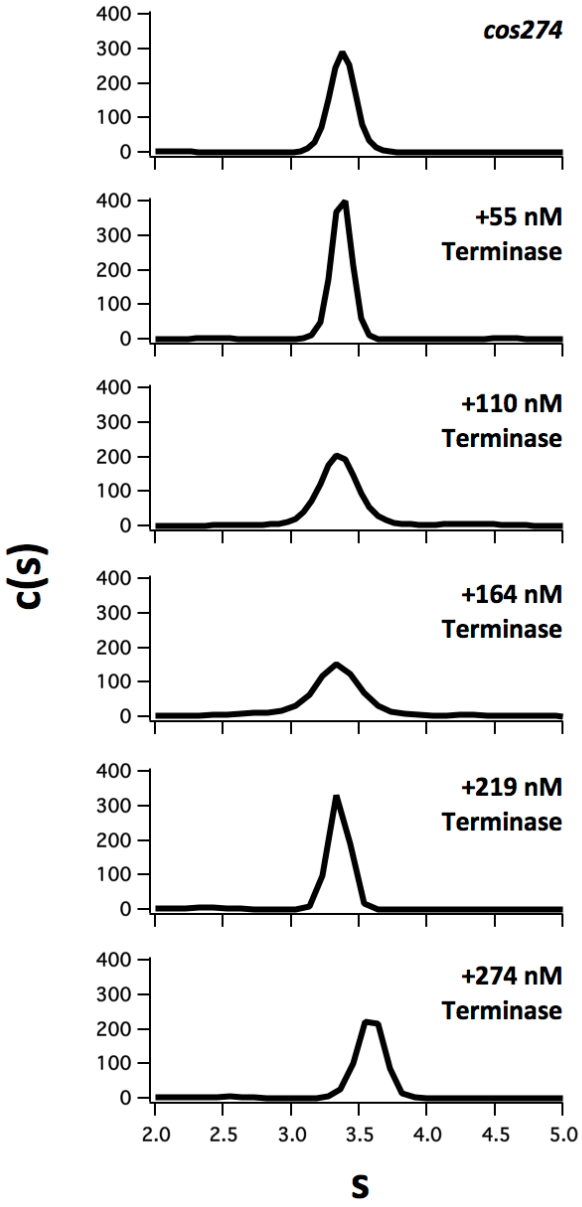
affinity of IHF for the *cos* sequence in the presence of 30 nM gpNu1 previously demonstrated[20]; recall that the gpNu1 terminase subunit confers site-specificity of DNA binding.

Sedimentation velocity shows that IHF strongly stimulates terminase binding at the cos site.

Given the diffuse smears resulting from terminase interactions at *cos* in EMS studies (Figure 1), and, furthermore, the lack of clarity over what effect the gel had on the formed nucleoprotein complexes, I chose to turn to analytical ultracentrifugation (AUC) as a different, solution-based, technique for more rigorous analyses.

Sedimentation velocity (SV) experiments were conducted using fluorescein-labeled DNA substrates as described in Materials and Methods. Increasing amounts of terminase protomer was added to 4 nM *cos274* in either the presence or absence of 50 nM IHF, and the raw data from these experiments were analyzed with SEDFIT using a continuous c(s) approach. The c(s) distributions from these analyses are shown in Figure 8 (no IHF) and Figure 9 (in the presence of 50 nM IHF).

Figure 8. Sedimentation velocity (SV) c(s) distributions of terminase binding to 4 nM *cos274* in the absence of IHF. SV of fluorescently-tagged 4 nM *cos274* in the presence of increasing concentrations of terminase was performed at 42K RPM. The data show virtually no increase in sedimentation coefficient of DNA with increasing concentrations of terminase, a finding I interpret as consistent with the lack of any significant terminase-*cos274* complex formation up to ~200 nM terminase.



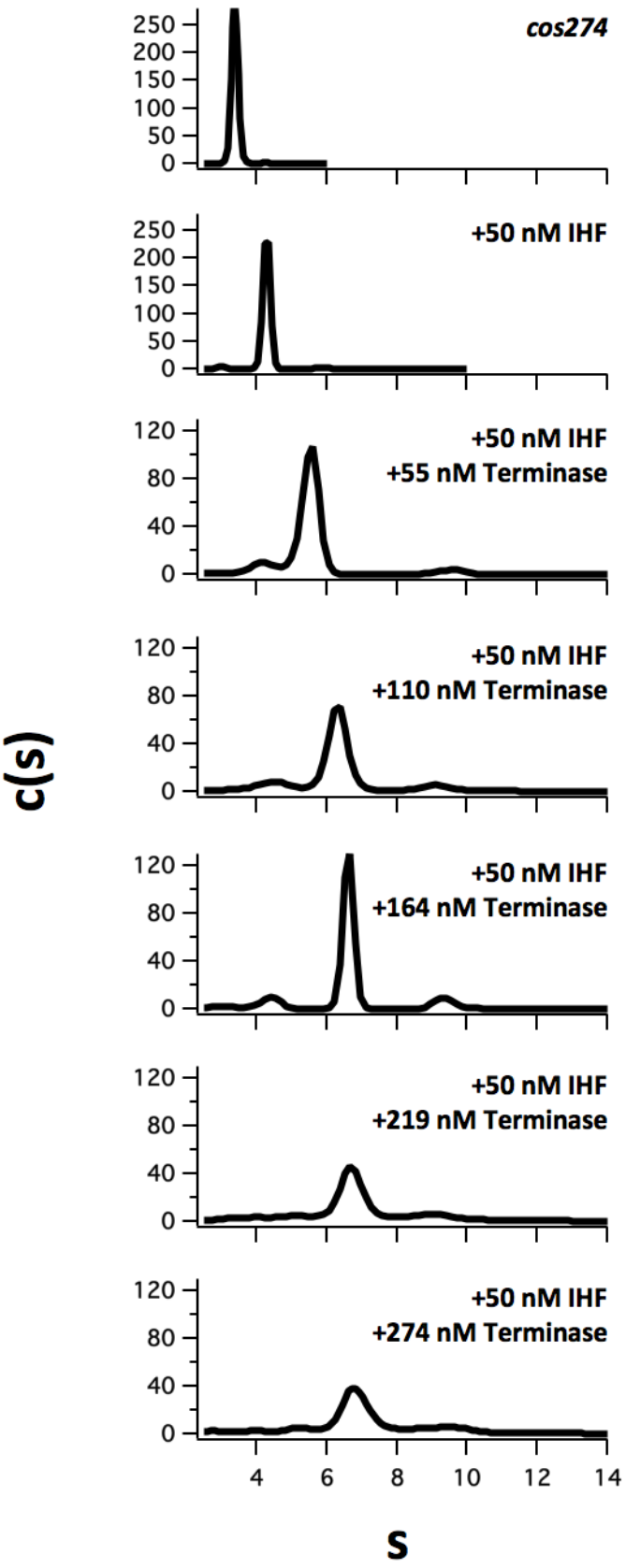


Figure 9. Sedimentation velocity (SV) c(s) distributions of terminase binding to 4 nM *cos274* in the presence of IHF. SV of fluorescently-tagged 4 nM *cos274* in the presence of 50 nM IHF and increasing concentrations of terminase was performed at 42K RPM. First, the data show an increase in sedimentation coefficient of DNA upon the addition of 50 nM IHF, indicating the formation of IHF-*cos274* nucleoprotein complexes. Second, the data show a terminase concentration-dependent, saturable, increase in sedimentation coefficient of DNA. The latter, in conjunction with the lack of any complex formation between terminase and *cos274* in the absence of IHF (Figure 8), indicate the formation of a ternary IHF•DNA•terminase complex in these studies.

I first note that a fit of my free DNA species yielded a 3.38 S (or a $s_{20,w}$, corrected for standard conditions at 20°C in water, of 5.82 S) species with molecular weight 166.3 kDa, which, being very close to the theoretical molecular weight of 169.9 kDa, gave me confidence in the technique. Of note is the fit frictional ratio, related to the diffusion coefficient (which can be measured from the spreading of the boundary in SV experiments), of 3.42, indicating a non-spherical, oblique species. This is unsurprising for duplex DNA of this length.

Second, it is clear that in the absence of IHF, terminase does not cause a concentration-dependent increase in sedimentation coefficient of *cos274* DNA (Figure 8). This is consistent with either (1) binding of terminase at *cos* but compensatory mass and size/shape effects that mask the binding event, or (2) no significant binding of terminase to *cos274* at these concentrations of terminase. Recall that the sedimentation coefficient from a SV experiment depends on both the mass *and* the frictional ratio (a measure of the effective size of the species) of the complex; this, in turn, could explain the potential compensatory mass and size/shape effects. However, I believe such a compensatory effect to be highly unlikely. First, EMS studies suggest that nucleoprotein complexes formed between terminase and either *cos274* or *ns274* are weak and nonspecific, which favors the explanation of no significant terminase binding at these concentrations. Second, the first explanation would require a phenomenal coincidence. Finally, in order for the sedimentation coefficient to remain the same despite a mass increase in the detected species, the diffusion coefficient of said species would concomitantly have

to decrease (frictional ratio would have to increase) upon the binding of terminase, as sedimentation coefficient is directly proprotional to mass and diffusion coefficient. In other words, a shape change, leading to a change in the effective size of the species, would have to occur to compensate for the mass change. However, the frictional ratios of the complexes remain largely similar (Table 1) with increasing concentrations of terminase, indicating a lack of any change in effective size of DNA or potential complexes with DNA. This gives further credence to the position that no significant binding is occurring between terminase and *cos274* under these conditions.

nM Terminase	Sedimentation Coefficient	Frictional Ratio
0	3.38	3.4
55	3.37	3.3
110	3.34	3.7
164	3.35	3.7
219	3.35	3.2
274	3.58	3.4

Table 1. Sedimentation coefficients and frictional ratios of fluorescently-tagged *cos274* species at various concentrations of terminase for the SV experiment of terminase-*cos274* interactions in the absence of IHF.

Despite the lack of any appreciable binding of terminase to *cos274* in the absence of IHF, I do note the slight increase in sedimentation coefficient of *cos274*, at the end-point of my titration, in the presence of 274 nM terminase (Figure 8, bottom). It is feasible this increase is indicative of a small fraction of terminase-bound DNA. Therefore, I note the possibility of a very low affinity complex under these conditions.

Third, 50 nM IHF increases the sedimentation coefficient of *cos274* from 3.38 S to 4.29 S (compare top two panels in Figure 9), indicating that a nucleoprotein complex is being formed between IHF and *cos274*. Interestingly, the frictional ratio of the sedimenting species drops from 3.42 (free DNA) to 2.23 (in complex with 50 nM IHF), indicating a more compact, spherical shape. This is not surprising, given that IHF introduces a bend into the duplex.

Finally, it is clear that in the presence of IHF, terminase causes a concentration-dependent, saturable, increase in sedimentation coefficient of *cos274* DNA (Figure 9). This is consistent with the binding of terminase to *cos274* in the presence of IHF. The sedimentation coefficients of the major peaks from Figures 8 and 9 were plotted as a function of terminase concentrations, and the data were fit to a Hill model as described in Materials and Methods (Figure 10). I do note the presence of other peaks in the c(s) distributions in Figure 9; however, given that the most significant of these peaks still contributed to less than 10% of the total signal, I do not make any significant inferences from their presence.

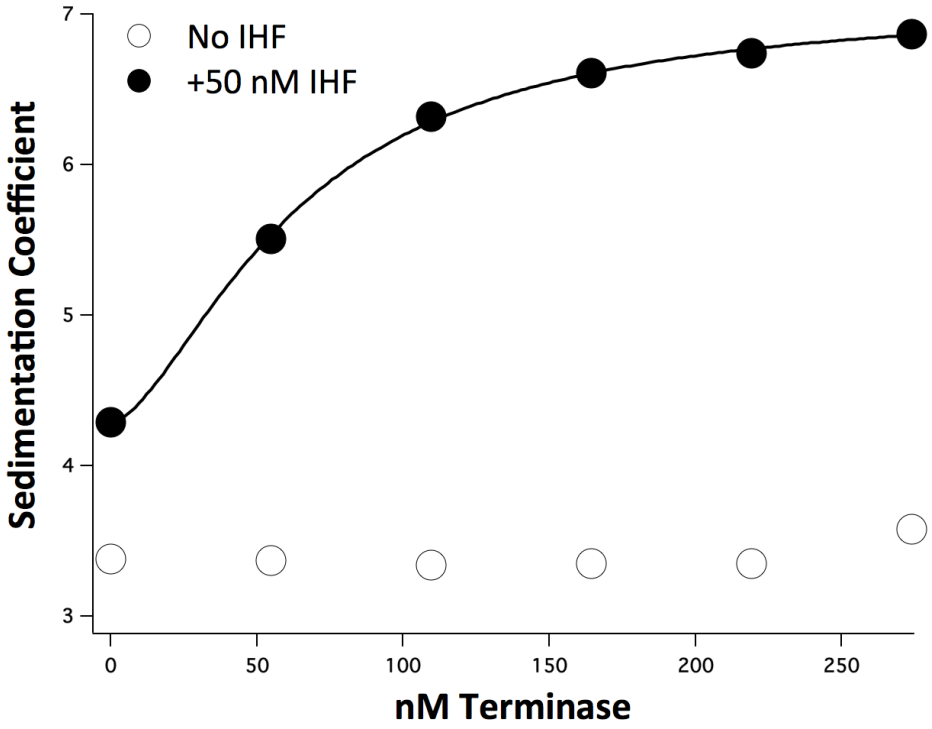


Figure 10. SV sedimentation coefficients of complexes formed between terminase and *cos274* in the presence (●) and absence (○) of IHF. Sedimentation coefficients of the major peaks in Figures 8 and 9 were plotted as a function of terminase concentration. In the absence of IHF, terminase shows little to no binding to *cos274*. In the presence of 50 nM IHF, a terminase concentration-dependent rise in sedimentation coefficient of fluorescently-tagged *cos274* is observed. Data were fit to a Hill model, yielding a $K_{d,app}$ of 63 nM for the terminase-*cos274* interaction in the presence of IHF.

A fit of the SV binding data for terminase-*cos274* interactions in the presence of 50 nM IHF afforded a $K_{d,app}$ of 63 ± 2.2 nM, with n = 1.62.

Interestingly, the frictional ratios (which SEDFIT can directly fit from the spread of the boundary in the SV data) of complexes formed between terminase and *cos274* in the presence of IHF are significantly lower than observed in the similar experiment without IHF (compare Tables 1 and 2).

nM Terminase	Sedimentation Coefficient	Frictional Ratio
0	4.29	2.2
55	5.34	1.6
110	6.33	1.9
164	6.57	1.3
219	6.76	2.1
274	7.11	2.5

Table 2. Sedimentation coefficients and frictional ratios of fluorescently-tagged *cos274* species at various concentrations of terminase for the SV experiment of terminase-*cos274* interactions in the presence of IHF.

Overall, I interpret the decrease in frictional ratios of complexes formed between terminase, IHF, and *cos274* in comparison to *cos274* alone to be indicative of the formation of specific ternary complexes with a severe bend at *cos*. More specifically, the binding of IHF at *cos* appears to decrease the frictional ratio from 3.4 (free *cos274*) to 2.2, indicating a compaction of the DNA species. This bend may be stabilized or enhanced with the addition of terminase, leading to the further decrease of frictional ratio to below 2 for terminase concentrations of 55 - 164 nM. Subsequent binding events increasing the stoichiometry of terminase at *cos* may no longer do anything further for the bend at *cos*, but might lead to a slight recovery (increase) of the frictional ratio due to the simple increase in size of the complex from additional protomer subunits. Much of this is speculation, however, and I wish

to stress that the take-home point from this data is that there is a definite drop in frictional ratio of complexes formed between terminase and *cos274* in the presence of IHF (indicating an effective decrease in size of DNA, consistent with a bend introduced upon binding) that is not observed for the similar binding experiment in the absence of IHF (indicating no effective size change at all, consistent with no significant binding). This finding leads further credence to the observation that in my SV experiments, IHF appears to be required for high affinity terminase binding to *cos274*.

As described in Materials and Methods, molecular weights of complexes formed in SV studies are not reported, due to the required prior knowledge of the partial specific volumes of the formed complexes.

Sedimentation equilibrium reveals only a slight increase in terminase binding affinity for cos274 in the presence, as opposed to the absence, of IHF.

In an attempt to determine the masses of the putative saturated species formed between terminase and *cos274* in the presence and absence of IHF, sedimentation equilibrium (SE) was performed, as outlined in Materials and Methods, on equivalent binding mixtures as those described in the SV experiments above. Briefly, terminase was titrated (0 – 274 nM) against 4 nM fluorescently-tagged *cos274* DNA in either the presence, or absence, of 50 nM IHF. Although SE is capable of

accurately determining the masses of homogenous species in solution, I was unable to obtain a reasonable molecular weight for my free DNA species (the reasons for which are beyond the scope of the current work). Therefore, any estimates of the molecular weights of nucleoprotein species formed in my titrations were suspect and, hence, not reported. That said, a terminase concentration-dependent increase in σ (an experimentally measured value directly proportional to the molecular weight of the observed species, see Materials and Methods) was nonetheless observed. As this increase in σ was additionally saturable, I interpreted it as indicative of a binding event between terminase and *cos274*. The experimentally fit σ values are plotted as a function of terminase concentration, for both terminase titrations in the presence and absence of IHF, in Figure 11.

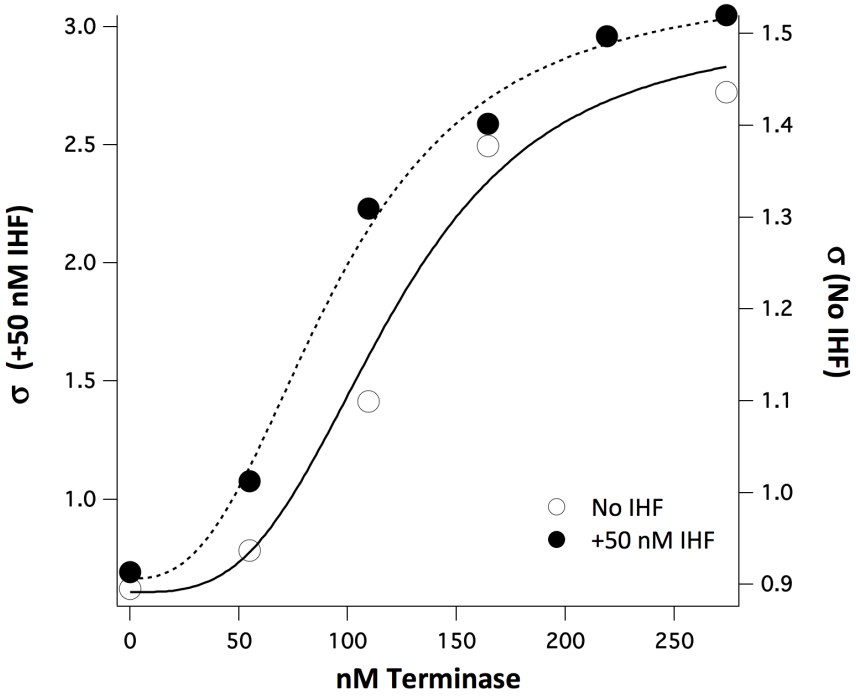


Figure 11. Sedimentation equilibrium of terminase binding to *cos274* in the presence (●) and absence (○) of IHF. The experimentally derived σ values for the nucleoprotein complexes formed in terminase binding reactions outlined in the accompanying text are plotted as a function of terminase concentration.

Of interest is the observation that in SE, I see the binding of terminase to *cos274* in the absence of IHF. This is distinctly different from my SV results, which show that terminase does not significantly bind *cos274* in the absence of IHF (Figure 10). The only difference in experimental conditions between my SE and SV run was the rotor speeds, which were 6K RPM and 42K RPM, respectively. It is appropriate to consider that higher centrifugation speeds lead to higher hydrostatic pressures proportional to the square of the rotor speed, with almost an approximately linear dependence on radial distance. The ability of hydrostatic pressure to dissociate macromolecular complexes has been reported[60-62]. Infante *et al* reported the disruption of the high affinity (pM dissociation constant) interactions between the 56S and 35S subunits of the 75S ribosome from unfertilized eggs of sea urchin as rotor speeds were increased from 24K RPM to 41K RPM[60]. This dissociation was salt-dependent as well as concentration-dependent, the latter due simply to the principle of mass action. Mechanistically, hydrostatic pressure can induce dissociation of molecular interactions typically if there is a positive volume change upon association. When the volume change is positive, there needs to be a re-arrangement of solution molecules that itself is sensitive to hydrostatic pressure. Furthermore, this positive volume change need not be a literal increase in molecular volume, as, for example, exchange of electrostatically bound partners upon association might affect the electrostriction of the aqueous solution in which the macromolecular association occurs; in turn, the resultant volume change might be pressure-sensitive. In fact, there are a number of potential sources of volume change in macromolecular association/dissociation reactions, including the uptake or release of counter-ions

and the effect this has on electrostriction of water, the ability of water to fill accessible volumes, conformational rearrangements that affect the hydration of polar and non-polar groups, etc. Therefore, it is not unreasonable to assume that there is some volume change in a terminase-*cos274* binding event that might be sensitive to hydrostatic pressure.

Although I am not aware of extensive, general studies regarding pressure-induced dissociation of protein-DNA complexes in particular (most information surrounding pressure-induced dissociation comes from protein-protein studies), recent studies on pressure-induced dissociation of IHF•DNA complexes by Senear *et al* indicate that high-affinity specific complexes between IHF and DNA dissociate between 500 and 1000 bar at 10 to 25 nM concentrations of binding partners[63]. A SEDFIT estimation of pressures at the bottom and the middle of my cells are as follows: 152 and 73 bar in SV experiments (42K RPM), and 4.3 and 2.0 bar in SE experiments (6K RPM). Thus, one can reason that even the pressures in my SV experiments are not enough to disrupt specific IHF•DNA complexes; consistent with this reasoning, I do indeed see binding of IHF to *cos274* in my SV experiments at 42K RPM. However, it may be feasible that terminase interactions at *cos274*, which I have shown to be relatively weak and nonspecific via EMS studies (Figure 1), are sensitive to the mild, yet 35-fold higher in comparison to my SE experiments, hydrostatic pressures in my SV runs. This would explain the discrepancy between my SV and SE experiments monitoring the binding of terminase to *cos274* in the absence of IHF. I also note that the increase in sedimentation coefficient observed for terminase-*cos274*

interactions in the presence of IHF, and the lack thereof in the absence of IHF, has been observed on over five separate occasions in the course of my SV experiments in aggregate.

With this context, I posit that SE detects both the (1) weak, nonspecific complexes formed between terminase and *cos274* in the absence of IHF (indicated by open circles [○] in Figure 11), and (2) the high-affinity, specific complexes formed between terminase and *cos274* in the presence of IHF (indicated by filled circles [●] in Figure 11). A fit of this data to a Hill model as described in Materials and Methods afforded a $K_{d,app}$ of 121 ± 16 nM, with n = 3.2, in the absence of IHF, and a $K_{d,app}$ of 95.8 ± 6.4 nM, with n = 3.2, in the presence of IHF. While this indicates only a minor increase in the affinity of terminase for *cos274* in the presence of IHF, I suspect that a direct comparison of these $K_{d,app}$ values is somewhat misleading. The complexes formed in the absence of IHF are likely weak and nonspecific (consistent with their sensitivity to high centrifugation speeds in SV and the lack of tight, gel-retarded complexes in EMS studies) and, therefore, represent a different binding mode of terminase to DNA than that observed for the formation of specific complexes. Meanwhile, the specific binding mode of terminase is aided by the IHF•DNA complexes presenting a more favorable architecture for the formation of a specific ternary complex. Therefore, the respective $K_{d,app}$ constants derived in the absence and presence of IHF are likely indicative of binding affinities of the *nonspecific* and *specific* binding modes of terminase, respectively. Hence, they may not be directly

comparable in the context of assessing the increase in *specific* (and *not* nonspecific) binding affinity of terminase for *cos* in the presence of IHF.

Discussion

I used a variety of biophysical techniques to look at the role of IHF in terminase assembly at *cos*. Particularly, I was interested in the cooperativity between IHF and terminase in the formation of a nucleoprotein complex at *cos*. The overarching conclusions to be drawn from the aggregate data presented in the current chapter are: (1) IHF appears to increase the *specific* binding affinity of terminase for *cos*, (2) terminase appears to increase the binding affinity of IHF for *cos*, and (3) that IHF and terminase cooperatively assemble at *cos*, the packaging initiation site of bacteriophage λ.

Given the variety of biophysical techniques used in these studies, it is worthwhile to discuss their individual contributions in arriving at these conclusions. It is not very surprising that some techniques (sedimentation velocity, or SV, and electrophoretic mobility shift, or EMS) show a rather dramatic increase in apparent binding affinity of terminase in the presence of IHF, while another (sedimentation equilibrium, or SE) shows an extremely modest increase, if any. First, recall that in EMS studies, terminase does bind to both *cos274* and *ns274* at low to moderate concentrations; however, diffuse smears result from this interaction that exhibit mobility in the gel inversely proportional to the loading concentration of terminase (Figure 1). This behavior prohibited me from proper quantitation, and limited my ability to create binding curves for terminase on DNA in the absence of IHF. However, it is evident from the slight upward shift of the free DNA band in Figure 1B that even at 10 nM

terminase, most, if not all, of the *cos274* DNA is associated with terminase. In Figure 2A, in the presence of 0.2 nM IHF, 8 nM terminase appears to be just beginning to shift the free DNA band upward. My method of quantitation, however, where I compute the ratio of the significantly gel-retarded band (top of gels, Figure 2) to the overall DNA signal, does not factor in binding events that lead to the diffuse smears terminase creates when binding to DNA in the absence of IHF. Hence, my EMS studies, first, do not allow me to quantitate nonspecific binding of terminase to DNA at all and, secondly, therefore underestimate the presumably nonspecific binding of terminase to DNA that only results in slight, diffuse, shifts in the DNA band. SV was even more incapable (likely due to pressure-induced dissociation of weak complexes) than EMS studies at detecting terminase binding to *cos274* in the absence of IHF, and so appeared to show an incredible amount of cooperativity between IHF and terminase on DNA (almost a *requirement* of IHF for terminase binding to *cos274*). A more sensitive technique capable of resolving (1) free DNA, (2) terminase nonspecifically bound to DNA, and (3) terminase and IHF specifically bound to DNA, would allow us to better quantitate the level of cooperativity between terminase and IHF on DNA.

SE, unlike the EMS studies discussed above, appears to be capable of observing and measuring the presumably nonspecific complexes between terminase and *cos274* in the absence of IHF. Hence I am able to generate, and compare, binding curves for terminase on *cos274* both in the absence and presence of IHF. However, it is very likely that I am comparing two different end products (namely, the products of

specific vs. nonspecific terminase binding) in Figure 11, and so using the sedimentation equilibrium results as a measure of cooperativity between terminase and IHF on DNA is not without its share of caveats.

In summary, I suspect that my EMS and SV studies tend to *overestimate* the observed cooperativity of IHF and terminase on DNA by not accounting for and measuring the nonspecific, weak complexes formed between terminase and DNA in the absence of IHF. In contrast, I suspect my SE studies tend to *underestimate* the cooperativity of IHF and terminase on DNA due to its inability to resolve the differences between the specific IHF•DNA•terminase and nonspecific terminase•DNA complexes. If, however, I choose to treat the presumably nonspecific, weak, complexes formed between terminase and *cos274* as being irrelevant in the context of a biologically functioning entity, then the extreme cooperativity between IHF and terminase for the assembly of the *specific* nucleoprotein complex (represented by the severely gel-retarded band in EMS studies, and by the 6.5 S species in SV studies) shown in EMS and SV studies may well be believed.

It is worthwhile to speculate on the complexes formed between terminase and *cos274* in the absence of IHF. In EMS studies, these complexes show up as diffuse smears in the gel that show a concentration-dependent shift to lower mobility. In SV studies, under the concentrations tested, these complexes do not appear to be complexes at all, instead displaying a sedimentation coefficient equivalent to free

DNA. In SE studies, however, these complexes appear to form in a terminase concentration-dependent, saturable, manner. I take these observations in aggregate to indicate that complexes formed between terminase protomer and *cos274* in the absence of IHF, at the concentrations of terminase tested in the current studies, are representative of weak, nonspecific interactions between terminase and DNA. It is possible that in the absence of IHF, terminase associates with DNA but only weakly. One could imagine terminase binding and unbinding the DNA, all the while moving around on it while bound (a similar 'DNA scanning' model was recently proposed by Andrews et al[58]). Consider now that in terminase catalytic activity assays in our lab, IHF dependence can be overcome by increasing concentrations of terminase (*unpublished studies*) or nucleotide[64]. This observation might indicate that in the stochastic assembly of terminase subunits to form a productive maturation complex at *cos*, increasing the concentration of terminase protomer (or the addition of nucleotide) might increase the likelihood of productive nucleoprotein complex formation. On the other hand, the presence of IHF might alter the structure of the DNA such that a local "energy sink" is created at *cosB*; when terminase either (1) directly binds to, or (2) binds elsewhere and moves along the DNA to find this structurally altered (bent) *cos* site, it forms a high-affinity complex with a low off-rate. This hypothesis is supported by the 12-hour half-life observed for the *pre-nicking* complex (which should be equivalent to the complex I am forming in these experiments) previously in our lab[12]. The role of IHF *in vivo*, then, may be to help the initiation of packaging by allowing the specific assembly of terminase at the *cos* site

at concentrations of terminase much lower than that needed to overcome the IHF dependence of activity assays in our lab.

IV: The Assembly State of Terminase at the Packaging Initiation Site of Bacteriophage λ is Affected by Integration Host Factor.

Introduction

The packaging initiation site, or *cos*, of bacteriophage λ is a complex site containing recognition sequences that both the λ packaging enzyme, terminase, and the *Escherichia coli* DNA binding protein integration host factor (IHF) recognize. Terminase is involved in both *DNA maturation* as well as *DNA packaging* (Chapter I, Figure 1); as such, it exhibits multiple catalytic activities, including endonuclease, strand separation, and DNA translocation activities. The terminase enzyme itself is composed of multiple subunits of the basic functional unit known as a protomer; this protomer is itself a heterotrimer of a large 72 kDa gpA subunit, that confers catalytic activities to the enzyme, and two smaller 20.4 kDa gpNu1 subunits, conferring site-specificity of binding to terminase. Previous studies have demonstrated that functional protomer can exist in solution at concentrations up to roughly 1 μM, beyond which the protomer self-assembles into a ring tetramer[23]. Prior to this thesis, however, the assembly state of terminase at the *cos* site has not been rigorously studied.

Given the multiple catalytic activities of terminase, it is not unreasonable to imagine that the assembly state of terminase on DNA might be dynamic. In other words, the

stoichiometry of protomers making up terminase in the active nucleoprotein complex may not be constant throughout the packaging process. A dimer is likely required for endonuclease activity, but, following *cos* cleavage, strand-separation and release of one half-site of *cosN* may conceivably require the disruption of gpA dimer interface (Chapter 1, Figure 2B). One could conceive of a rearrangement whereby a different number of protomers may be recruited for assembly of a catalytically-competent packaging motor in concert with portal and procapsid binding. In this regard, a hexameric terminase ring structure is attractive, as it could accommodate symmetric interactions with the dodecameric portal complex, and it shows mechanistic similarity to translocating helicase enzymes[31]. Meanwhile, a ring tetramer has been identified at high concentrations of terminase protomer (*vide supra*).

In an effort to begin to understand the assembly states of terminase on DNA, the research presented in the current chapter aims to study the stoichiometry of terminase on a model *cos* substrate containing both the *cosB* (binding) and *cosN* (site of endonuclease cleavage) sites. Particularly, I aim to show the effect IHF has on the stoichiometry of terminase assembled at *cos*.

Materials and Methods

IHF expression and purification

IHF was expressed and purified as described previously (Chapter II: Materials and Methods).

Expression and purification of terminase protomer

Terminase protomer was expressed and purified as described previously (Chapter III: Materials and Methods).

Preparation of DNA substrates

A minimal nonspecific DNA duplex (*R3-10*) with a 5'-Cy3[65] label on each strand was used for IHF activity studies. The following sequence, and its complement, was synthesized and HPLC purified by Integrated DNA Technologies (Coralville, IA).

R3-10: 5'-Cy3-GCGTTTCCGT

The 274 bp model *cos* substrate used in the present terminase-DNA binding studies was prepared by large-scale preparative PCR. pAFP1 plasmid was used as a template in the PCR reactions, and was provided as a generous gift from Dr. Michael Feiss (University of Iowa). This plasmid contains the entire *cos* sequence cloned into a pUC19 vector, and was purified from *E. coli* JM107[pAFP1] cells using a Qiagen Plasmid *Plus* Maxi kit. PCR primers were designed to amplify the entire *cos*

sequence, inclusive of the R-elements that bind the gpNu1 small subunit of terminase and the *I1* element recognized by IHF. Primers were synthesized, and HPLC purified, by Integrated DNA Technologies (Coralville, IA); sequences are provided below.

Forward primer for PCR of *cos274*: 5'-Cy3-CCGGAATTCGCATGCCTGCAGGTCTAA

Reverse primer for PCR of *cos274*: 5'-Cy3-CGCGAATTCCATTGTTCATTCCAC

Primers were synthesized with fluorescent Cy3[65] labels for detection of the DNA and any nucleoprotein complexes formed with DNA. PCR reaction mixtures (100 µl per tube) contained each primer at a concentration of 0.5 µM and 10 ng of pAFP1 template. *Taq* DNA polymerase, deoxynucleotides, and required buffer components were included as specified by the manufacturer's protocol for the *Taq* PCR kit (New England Biolabs). Total reactions volumes were 1 ml. PCR cycles were as follows: (94°C for 1 minute, 61°C for 1 minute, 72°C for 2 minutes) x 60 cycles. PCR reaction mixtures were subsequently pooled and purified using a Promega Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison WI). Purity of the amplified products was verified by agarose gel analysis, and the concentration of DNA was determined via spectroscopic methods. The sequence of the specific (*cos274*) DNA substrate is given below.

cos274:

CCGGAATTCGCATGCCTGCAGGTCTAATCATTATCACTTTACGGGTCCTTTCCGGTGATC

CGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTT

AAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAA

AGAAAGGAAACGACAGGTGCTGAAAGCGAGGCTTTTTGGCCTCTGTCGTTTCCTTTCTCT

GTTTTTGTCCGTGGAATGAACAATGGAATTCGCG

The molecular weights of DNA substrates were calculated based on their sequences using an online calculator[49]. Cy3-labeled *R3-10* and *cos274* had molecular weights of 7.23 and 170.34 kDa, respectively.

Electrophoretic mobility shift (EMS) studies

EMS binding studies were performed in buffer containing 20mM Tris, pH 8, 1mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 10% glycerol. DNA was included at a final concentration of 200 nM, and IHF was added to the binding mixtures to appropriate concentrations. Mixtures were incubated at room temperature for 20 minutes, and then loaded onto a NativePAGE™ Novex® 3-12% Bis-Tris gel (Invitrogen). Gels were run at 15V/cm in 0.5x TBE at 4°C for 1 hour, and visualized by ethidium bromide staining.

Bands were quantified using *ImageQuant TL* software (GE Healthcare Life Sciences). The fraction of bound DNA was calculated using:

$$Fraction\ Bound = \frac{counts\ in\ retarded\ band}{counts\ in\ retarded\ band + counds\ in\ free\ DNA\ band}$$

The raw binding data were fit to a simple Hill model, according to:

$$Fraction\ Bound = (m - b) * \left[\frac{[P]^n}{{K_d}^n + [P]^n}\right] + b$$

where [P] is the protein concentration, K_d is the apparent dissociation constant, *n* is the Hill coefficient, b is the baseline offset, and m is the fraction of DNA bound at saturation. Experimental data were fit using Igor Pro (Wave Metrics, Lake Oswego, OR).

Limiting nonspecific IHF binding in the ternary IHF•DNA•terminase

Previous data presented in Chapter II (Figure 7) indicated that there is not a dramatic distinction between specific and nonspecific binding of IHF to *cos274* (a "specific" 274 bp duplex substrate that models the full-length *cos* sequence found in concatemeric λ DNA). This was of concern to me, as I desired to limit the amount of nonspecific IHF binding to *cos274* in the presence of terminase so that I could accurately assess the stoichiometry of terminase on *cos274* in the presence of IHF. The binding of multiple IHF heterodimers to *cos274* would affect the mass of the observed nucleoprotein complex and, therefore, the determination of the stoichiometry of terminase on the DNA. I reasoned that I could estimate the minimum amount of IHF needed to form a specific, bent *cos274* duplex by (1)

measuring the DNA binding activity of my IHF prep via sedimentation equilibrium (SE), and (2) performing a stoichiometric electrophoretic mobility shift (EMS) binding study of IHF on *cos274*.

Sedimentation equilibrium experiments

Unless otherwise noted, binding studies using sedimentation equilibrium (SE) were carried out in buffer containing 20 mM Tris at pH 8, 10 mM sodium phosphate, 1 mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 5% glycerol. Experiments were conducted on a Beckman Coulter XLA ultracentrifuge, and data were collected at either 543 nm (maximum absorbance of the Cy3 dye in our XLA detection system) or 495 nm (to decrease measured absorbance when measurements well exceeded an O.D. of 1.2), monitoring only the Cy3-labeled DNA and any nucleoprotein complexes formed with the DNA. Scans were taken every 0.003 cm with 2 averages in continuous scan mode. 100-120 µl of sample were loaded into preassembled analytical cells containing Epon charcoal-filled six-channel centerpieces (with external fill). Typically, three analytical cells were loaded into an An-50 Ti titanium rotor, along with a counterweight. Experiments were performed at 7°C unless otherwise noted. Rotor speeds were chosen so as to achieve a σ value (see equation below relating σ to angular velocity of the rotor, ω) of 2 for the suspected nucleoprotein complex formed in the binding studies. Data at two additional rotor speeds were collected for global analysis; these speeds were chosen such that the ratios of the squares of the speeds were 1.5 and 2.25. Hence, speeds of 22K, 27K, and

33K RPM were used for the IHF activity assay, and speeds of 6K, 7.3K, and 9K were used for all terminase binding studies.

Samples were periodically checked for approach to equilibrium using the program WINMATCH (David Yphantis, University of Connecticut; Michael Johnson, University of Virginia; Jeff Lary, National Analytical Ultracentrifuge Center, Storrs CT). Data from samples having reached equilibrium were analyzed using the nonlinear least-squares analysis program WINNONLIN (authors of WINMATCH) to a single species model, given by:

$$A_r = A_{ref} * e^{\sigma z}, where\ z = \frac{r^2 - {r_{ref}}^2}{2}\ and\ \sigma = \frac{M(1-\bar{v}\rho)}{RT} * \omega^2$$

where A_r is the total absorbance of the sample at radial position r, A_{ref} is the absorbance of the sample at the reference radial position r_{ref}, M and $\bar{v}$ are the molecular weight and partial specific volume of the monitored species, respectively, ρ is the buffer density, R is the gas constant, T is the absolute temperature, and ω is the angular velocity of the rotor. The density and viscosity for buffer compositions in each experiment were calculated using SEDNTERP[52]. The fit σ values were used to determine buoyant molecular weights for the analyses of protein stoichiometries on DNA.

DNA molecular weight calculations

The molecular weights of all DNA substrates used were calculated, based on their sequence, using an online calculator[49]. Molecular weights of fluorophores (available at www.idtdna.com) were added to DNA molecular weights as appropriate.

A buoyant molecular weight (M_b) for each DNA substrate was calculated for each sedimentation equilibrium experiment using the relation $M_b = M(1-\bar{\upsilon}\rho)$.

Protein molecular weight calculations

The molecular weight of IHF was calculated by summing the molecular weights of the individual α and β subunits. This resulted in a molecular weight of 22,005 Da for IHF, based on 11,354 Da for the α subunit and 10,651 Da for the β subunit, as calculated by SEDNTERP[52] using the amino acid sequences for the α[53] and β[54] subunits. I also used SEDNTERP to calculate a partial specific volume of 0.736 for IHF based on the amino acid sequence.

The molecular weight (114.2 kDa) and partial specific volume (0.733) of terminase protomer were calculated in SEDNTERP[52] using the amino acid sequences of gpA and gpNu1, assuming a ratio of 1:2 gpA:gpNu1.

A buoyant molecular weight (M_b) for each protein was calculated for each sedimentation equilibrium experiment using the relation $M_b = M(1-\bar{\upsilon}\rho)$.

Calculation of protein stoichiometry from sedimentation equilibrium (SE)

The following relation was used to determine a buoyant molecular weight of any sample, based upon the fit σ from raw experimental data:

$$M_b = \frac{\sigma RT}{\omega^2(1-\bar{v}\rho)}$$

For IHF activity studies and terminase binding studies in the absence of IHF, the stoichiometry of protein, $n_{protein}$, for any loading protein:DNA ratio was determined as follows:

$$n_{protein} = \frac{M_b - M_{b,DNA}}{M_{b,protein}}$$

where M_b is the buoyant molecular weight of the sample resulting from the fit σ, $M_{b,DNA}$ is the *experimental* buoyant molecular weight of free DNA, and $M_{b,protein}$ is the *theoretical* buoyant molecular weight of protein (either IHF or terminase) under the experimental conditions.

For terminase binding studies in the presence of IHF, the stoichiometry of IHF on DNA was determined using the above relation for $n_{protein}$; however, the stoichiometry of terminase, $n_{terminase}$, for any loading terminase:DNA ratio was determined as follows:

$$n_{terminase} = \frac{M_b - M_{b,DNA} - 2 * M_{b,IHF}}{M_{b,terminase}}$$

where M_b is the buoyant molecular weight of the sample resulting from the fit σ, $M_{b,DNA}$ is the *experimental* buoyant molecular weight of free DNA, $M_{b,IHF}$ is the *theoretical* buoyant molecular weight of IHF under the experimental conditions, and $M_{b,terminase}$ is the *theoretical* buoyant molecular weight of terminase under the experimental conditions. The rationale for these particular calculations is elaborated upon in Results.

Calculation of partial specific volume of free DNA from sedimentation equilibrium (SE)

The following relation was used to experimentally determine partial specific volumes of free DNA based upon the fit σ from raw experimental SE data:

$$\bar{v} = \frac{1}{\rho}(1 - \frac{\sigma RT}{M})$$

Results

Sedimentation equilibrium of IHF binding to a minimal nonspecific substrate indicates at least 80% binding activity of my IHF prep.

Sedimentation equilibrium (SE) was used to estimate the nonspecific DNA binding activity of my IHF prep by measuring (1) the loading concentration of IHF needed to saturate a minimal nonspecific DNA substrate, and (2) the mass, and therefore stoichiometry of IHF, of the resultant complex. Although SE is capable of accurately measuring the mass of observed species, the analysis is considerably simplified and constrained if only one homogenous species exists in solution. One way to ensure this is to achieve saturation of protein binding on a minimal DNA substrate. Record and coworkers previously identified a minimal nonspecific site size of 10 bp for IHF binding under conditions similar to those used here[39]. I reasoned therefore that a 10 bp region within the *R3* element of the *cos* site ('*R3-10*'), an element I already showed only interacts nonspecifically with IHF (Chapter II, Figure 11), would serve as a model nonspecific substrate to which only one IHF heterodimer would bind. In buffer containing 20 mM Tris at pH 8, 1 mM EDTA, 7mM β-me, 2mM spermidine, 55mM NaCl, and 10% glycerol, IHF was titrated to 30 μM against 4μM Cy3-labeled *R3-10* duplex DNA. DNA-containing species were monitored by the absorbance of the Cy3 fluorophore. The fit buoyant molecular weights of resulting complexes (see Materials and Methods) were used to derive an IHF protein stoichiometry on DNA for any loading IHF:DNA ratio (Figure 1 and Table 1).

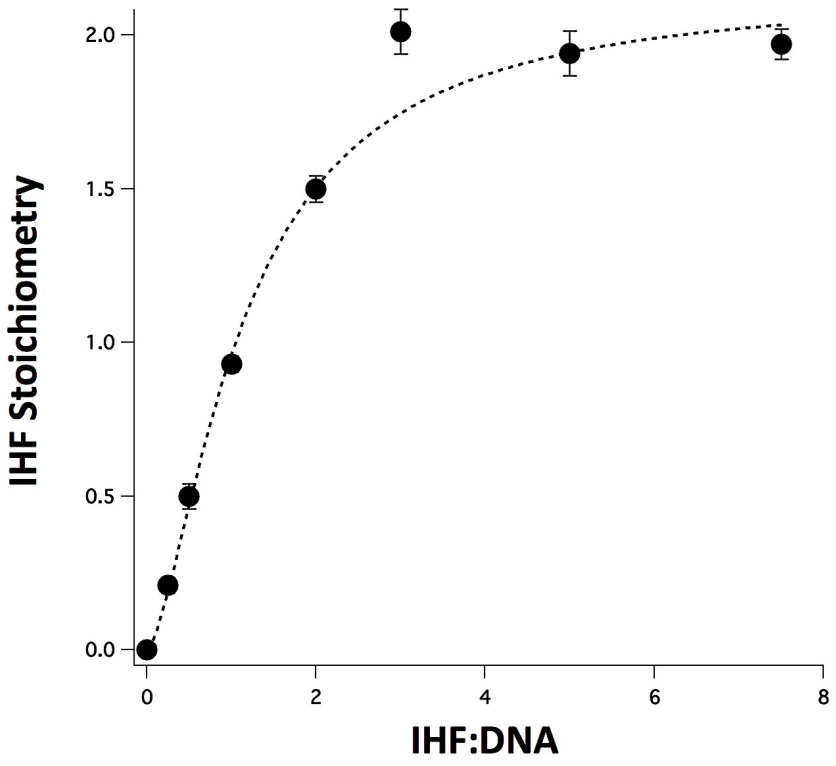


Figure 1. IHF titration against 4 µM Cy3-labeled *R3-10*, monitored by sedimentation equilibrium (SE). SE data at three rotor speeds were globally analyzed to monitor the binding of IHF to a minimal (10bp) nonspecific substrate. The analysis revealed a stoichiometry of two IHF heterodimers per DNA duplex, and a protein activity of at least 80%.

Table 1 summarizes the data presented in Figure 1.

Loading IHF:DNA	M_b (kDa)	IHF Stoichiometry	Standard Deviation
0	2.77	0	0.010
0.25	3.88	0.21	0.026
0.5	5.44	0.50	0.041
1	7.71	0.93	0.026
2	10.69	1.50	0.043
3	13.39	**2.01**	0.074
5	13.01	**1.94**	0.073
7.5	13.21	**1.97**	0.049

Table 1. Experimental buoyant molecular weights and resulting stoichiometries of IHF on *R3-10* from SE studies. The number of IHF heterodimers (~2) on DNA at saturation is indicated in **bold**. Calculations are described in Materials and Methods.

The buoyant molecular weight of free DNA was used to calculate an experimental partial specific volume for my DNA substrate of 0.598, which is slightly overestimated in comparison to the generally accepted average partial specific volume of 0.55 for linear duplex DNA. Assuming a partial specific volume of 0.55, the theoretical M_b is 3.12 kDa. The underestimation evident in the experimental M_b of 2.77 kDa could be due to non-ideality at these concentrations of the highly negatively charged DNA.

Given my experimental conditions, a M_b of 5.29 kDa was calculated for IHF. This results in a saturated IHF•DNA complex with a ~13.2 kDa M_b consistent with two IHF heterodimers bound to the *R3-10* substrate (Table 1). A fit of the data to a stoichiometric binding model indicates that a ratio of 2.5:1 IHF:DNA is required to form a 2:1 IHF•DNA complex; therefore, I estimate a protein activity of at least 80%. Interestingly, a fit (dotted line, Figure 1) of the SE data in Figure 2 to a Hill model (as described in Materials and Methods) yields a $K_{d,app}$ of 4.6 ± 0.7 μM, with n = 1.5. I note that this is very close to the experimentally derived K_d of 10.8 μM for the nonspecific binding of IHF to duplex DNA reported via isothermal calorimetry studies previously[39]. Furthermore, since this $K_{d,app}$ is very close to the substrate concentration in these experiments, a stoichiometric fit to the data may not be appropriate; therefore, I note that the derived 80% protein activity is a *lower limit*.

A two-fold stoichiometric excess of IHF is required to fully saturate a model cos-containing DNA substrate under stoichiometric binding conditions in EMS studies.

SE studies detailed above indicated at least ~80% nonspecific DNA binding activity for my IHF prep (Figure 1). With this information, and the assumption that nonspecific DNA binding activity is equivalent to specific DNA binding activity for my IHF prep, I could have assumed that I need a 1.25-fold stoichiometric excess of IHF to fully bind *cos274* in SE studies aimed to determine the stoichiometry of terminase on the specific IHF•DNA complex. However, I could not rule out the possibility that (1) nonspecific DNA binding activity is not equivalent to specific binding activity, or (2) more than one IHF heterodimer is required to form a specific complex with *cos274*. Therefore, I performed an electrophoretic mobility shift (EMS) assay under stoichiometric binding conditions to determine the minimal concentration of IHF to bind 200 nM Cy3-labeled *cos274* (the exact substrate and concentrations used in the sedimentation equilibrium studies detailed below). Since EMS studies in Chapter II indicated a $K_{d,app}$ of 1.8 nM for my *cos274* substrate, I was confident that the present studies, utilizing a substrate concentration two orders of magnitude higher than the $K_{d,app}$ of the interaction, would be under a stoichiometric binding regime. In other words, any IHF added should bind available DNA, as the equilibrium is pushed toward complex formation at these elevated concentrations of substrate. The gel and quantified data from this experiment are shown below (Figure 2).

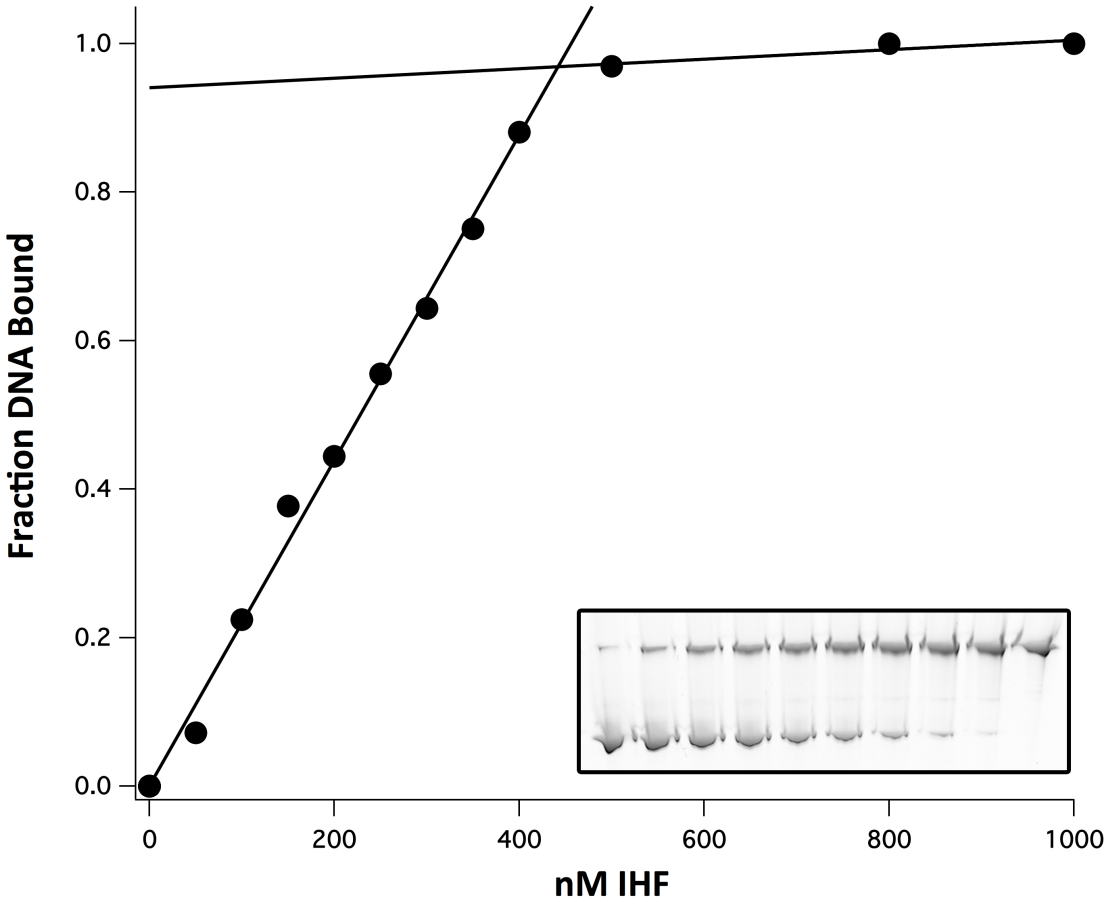


Figure 2. EMS study of stoichiometric IHF binding at *cos274.* IHF was titrated (0 – 1 μM) against 200 nM Cy3-labeled (to mimic the substrate used in sedimentation equilibrium studies, *vide infra*) cos274. The linear increase in fraction of DNA bound is indicative of the fact that substrate concentrations are well above the K_d of the IHF-*cos274* interaction. *Inset*: gel representative of the data presented here.

I note that roughly two stoichiometric equivalents of IHF (~400-450 nM) are required to fully saturate 200 nM *cos274*, as indicated by the breakpoint, or intersection of the two lines, in Figure 2. If I assume that specific and nonspecific DNA binding activity (~80%, determined from SE studies, *vide supra*) are equivalent for my IHF prep, this data suggest that two IHF heterodimers bind to form the specific, bent duplex at the *cos* site on *cos274*. That said, since the equivalency of specific and nonspecific DNA binding activity for my IHF prep has not been

experimentally verified, I do not rule out the possibility that the specific binding activity of my IHF prep is, in fact, ~40%. This lower activity would necessitate a 2.5-fold stoichiometric excess of IHF to form a 1:1 IHF•DNA complex, as would a requirement of two IHF heterodimers to form the presumed (from the data presented in Figure 2) 2:1 IHF•DNA specific complex. Therefore, SE studies below utilized a 2.5-fold stoichiometric excess of IHF to ensure that 100% specific IHF•DNA complexes were formed before any terminase was titrated to generate the ternary species.

Terminase assembles on cos-containing DNA as a pentamer in the presence of IHF, and a dimer in the absence of IHF.

Sedimentation equilibrium studies were performed to measure the masses of nucleoprotein complexes formed between terminase and Cy3-labeled *cos274* in both the presence, and absence[2], of IHF. As described in Materials and Methods, monitoring only the fluorophore label on *cos274* ensured that I monitored only complexes formed with DNA. Briefly, terminase was titrated against 400 nM *cos274* in the absence of IHF (at 4°C), or against 200 nM *cos274* in the presence of 500 nM IHF (at 7°C). Based on a lower limit of activity of my IHF prep of 80% (*vide supra*), and the observation that two stoichiometric equivalents of IHF are needed to fully bind and bend *cos274* (Figure 1), a concentration of 500 nM IHF was used to ensure

[2] SE studies of terminase binding to *cos274* in the *absence* of IHF were performed in collaboration with Teng-Chieh Yang, a postdoctoral research associate in the lab of Carlos E. Catalano.

enough IHF was available to fully form a specific complex with *cos274*. The fit buoyant molecular weights (M_b, see Materials and Methods) were used to derive a terminase stoichiometry on *cos274* for any loading terminase:DNA ratio either in the presence (●) or absence (○) of IHF (Figure 3).

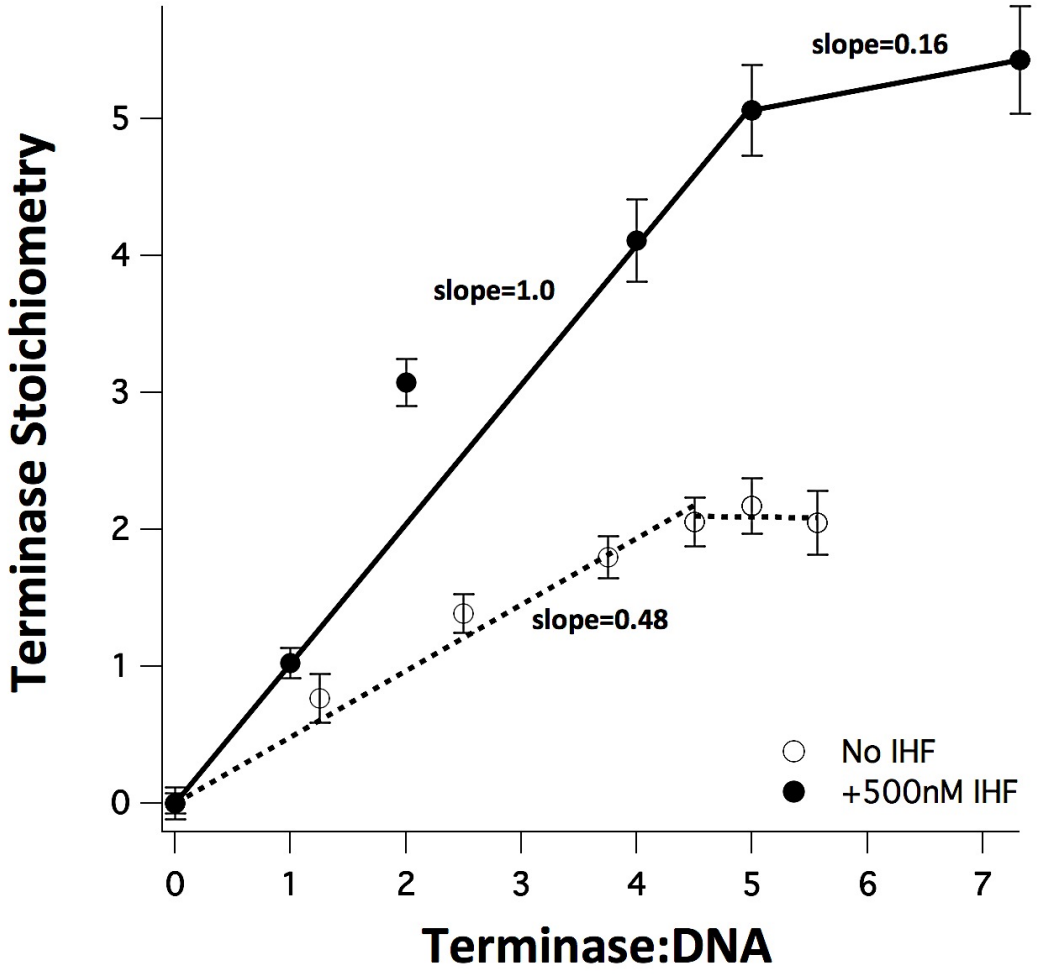


Figure 3. Terminase stoichiometries on *cos274* in either the presence (●) or absence (○) of IHF, monitored by sedimentation equilibrium (SE). SE data at three rotor speeds were globally analyzed, as described in Materials and Methods, to determine buoyant molecular weights and stoichiometries of complexes formed between terminase and *cos274* in the presence or absence of IHF. In the absence of IHF, terminase assembles as a dimer on *cos274* (plateau at a stoichiometry of 2, ○). In the presence of IHF, terminase assembles as a pentamer on *cos274* (plateau at a stoichiometry of 5, ●). The slight rise beyond a stoichiometry of 5 in the presence of IHF may be explained by a fraction of complexes populated with a sixth, nonspecifically bound, terminase.

The data indicate that there are clear differences in stoichiometries of complexes formed between terminase and *cos274* in the presence, as opposed to the absence, of IHF. Furthermore, while the titration in the absence of IHF appears to have definitively reached saturation (plateau at a terminase stoichiometry of 2, Figure 3), the masses of the nucleoprotein complexes appear to still be increasing, albeit slightly, at the endpoint of the titration in the presence of IHF. Unfortunately, it was not possible to attempt higher concentrations of terminase due to experimental and terminase prep constraints. Given the significant decrease in the slope of the data above a stoichiometry of 5 terminase protomers, however, I believe I am approaching saturation of binding.

A closer look at the data reveals that for any given terminase:DNA ratio, a higher M_b was obtained for the terminase-*cos274* complex in the presence of IHF as opposed to the absence of IHF. These differences cannot be trivially explained by simple mass contributions from IHF alone. Table 2 summarizes the experimentally fit buoyant molecular weights, and resulting terminase stoichiometries, for the complexes formed between terminase and *cos274* in the *absence* of IHF. Table 3 summarizes the experimentally fit buoyant molecular weights, and resulting terminase stoichiometries assuming two IHF molecules are bound on *cos274*, for the complexes formed between terminase and *cos274* in the *presence* of IHF. The implications of the fit buoyant molecular weights and terminase stoichiometries for both titrations (*absence* and *presence* of IHF) are discussed further below.

Loading Terminase:DNA	M_b (kDa)	Terminase Stoichiometry	Standard Deviation
0	75.51	0	0.07
1.25	97.96	0.77	0.18
2.5	116.17	1.39	0.14
3.75	128.21	1.80	0.15
4.5	135.76	**2.06**	0.18
5	139.18	**2.17**	0.20
5.56	135.61	**2.05**	0.24

Table 2. Experimental buoyant molecular weights and resulting stoichiometries of terminase on *cos274* in the absence of IHF. The number of terminase protomers (~2) on DNA at saturation is indicated in **bold**. Calculations are explained in Materials and Methods.

The experimentally fit M_b for free DNA (75.5 kDa, Table 2) in the absence of IHF is consistent with the theoretical M_b of 75.04 kDa. The experimental M_b was used to calculate a partial specific volume for free DNA, as described in Materials and Methods. The experimentally derived partial specific volume of 0.547 is commensurate with the generally accepted range of 0.55 - 0.59 for duplex DNA. The terminase titration in the presence of *cos274* yields a complex with M_b of approximately 136 kDa at saturation, and this is consistent with a stoichiometry of two terminase protomers (M_b = 29.3 kDa, per protomer, under my experimental conditions) bound to *cos274* (M_b = 75.5 kDa under my experimental conditions). A stoichiometric binding fit to my data yields a DNA binding activity of 48% (indicated by a slope of 0.48, Figure 3), and this is patently evident from the observation that an approximately 4:1 terminase:DNA loading ratio is required to form the 2:1 terminase•DNA complex at saturation (Table 2).

Loading Terminase:DNA	M_b (kDa)	Terminase Stoichiometry	Standard Deviation
0	69.59 (-IHF)	0	--
0	73.75 (+IHF)	0	--
1	110.47	1.02	0.11
2	170.40	3.08	0.17
4	200.61	4.11	0.30
5	228.44	**5.06**	0.33
7.32	239.20	5.43	0.39

Table 3. Experimental buoyant molecular weights and resulting stoichiometries of terminase on *cos274* in the presence of IHF. The number of terminase protomers (~5) on DNA at saturation is indicated in **bold**. Calculations are explained in Materials and Methods. Note that for a loading terminase:DNA of 0, the M_b is indicative of IHF-bound *cos274* and not free DNA.

The experimentally fit M_b for free *cos274* (69.6 kDa) in the *+IHF* experiment was underestimated in comparison to the theoretical M_b of 74.9 kDa (under these experimental conditions). The experimental M_b was used to calculate a partial specific volume for free DNA, as described in Materials and Methods. The experimental partial specific volume of 0.581 is within the generally accepted range of 0.55 – 0.59 for duplex DNA, although it is overestimated in comparison to the experimentally derived partial specific volume for the same DNA substrate in the previous *–IHF* experiment. It is feasible that the presence of 2 mM spermidine affected the partial specific volume of DNA in the *+IHF* experiment, as the *–IHF* experiment did not contain spermidine. However, without further experimentation, I can only speculate as to the cause of the lower experimental M_b.

The increase in M_b of *cos274* upon the addition of 500 nM IHF was calculated to be 4.16 kDa (73.8 kDa vs. 69.6 kDa, Table 3), which falls slightly short of the M_b of one IHF heterodimer under these experimental conditions (5.51 kDa). I find the data suggestive of a binary 1:1 IHF•DNA specific complex formed at this concentration of IHF, despite the slight underestimation in the theoretical increase in M_b that should be observed upon the addition of one IHF heterodimer to *cos274.* The 25% underestimation in M_b could be indicative of only 75% of the DNA bound by one IHF heterodimer; however, I do note that such a rationalization is in disagreement with previous EMS studies indicating 100% DNA bound at equivalent concentrations of *cos274* and IHF (Figure 2).

Addition of terminase to the IHF•DNA complex should yield a theoretical rise in M_b of 29.2 kDa per protomer bound. The addition of one stoichiometric equivalent of terminase to *cos274* yielded an increase in M_b of 36.7 kDa (final M_b = 110.5 kDa, Table 3), which is 7.5 kDa larger than the theoretical increase in M_b that would result if 100% of the added terminase bound *cos274*. Interestingly, subtracting 29.2 kDa (the theoretical M_b of one terminase protomer) from 110.5 kDa yields 81.3 kDa, which is consistent with the 80.6 kDa buoyant molecular weight resulting from the addition of two IHF heterodimers (theoretical M_b = 5.51 kDa per IHF heterodimer) to the experimentally derived M_b of *cos274* (M_b = 69.6 kDa). Hence, I assumed that the presence of terminase increased the stoichiometry of IHF on *cos274* to two; this is not an unreasonable assumption due to the observations that (1) terminase increases the affinity of IHF for *cos274* in EMS studies (Chapter III, Figure 6), and (2)

EMS studies suggest that two IHF heterodimers might bind to *cos274* (Figure 2) considering that the DNA binding activity of my IHF prep is at least 80% (as determined by SE studies, *vide supra*).

Assuming a stoichiometry of two IHF heterodimers on *cos274* in the presence of terminase, I calculated a stoichiometry of terminase on the binary IHF•DNA complexes for each terminase:DNA loading ratio (Table 3). For terminase•DNA loading ratios of 1:1, 4:1, and 5:1, terminase stoichiometries of 1.02, 4.11, and 5.06 resulted. This is consistent with a stoichiometric binding regime where all titrated protein binds the DNA substrate (that is, the concentrations of binding partners are likely well in excess of the K_d of the binding interaction). Furthermore, the specific DNA binding activity of terminase from this experiment appears to be 100% (slope=1.0, Figure 3). A terminase:DNA loading ratio of 2:1 led to a terminase stoichiometry of 3.08, which is not possible given the limiting concentrations of terminase at this loading ratio. Given that the raw data for this particular loading ratio was especially noisy, I chose to consider the resultant M_b an outlier and, therefore, did not include it in the stoichiometric binding fit pictured in Figure 3.

Further titration of terminase to a loading ratio of 7.3 led to a more modest increase in stoichiometry of terminase on *cos274*, yielding complexes with an average M_b indicating 5.4 terminase protomers bound. I interpret this intermediate stoichiometry to be indicative of only a fraction of the complexes populated with a sixth, nonspecifically bound, terminase protomer. The decrease in slope

(slope=0.16, Figure 3) at the upper end of the terminase titration is indicative of (1) an approach to saturation of binding, and (2) a lower affinity binding event such that all available terminase does *not* bind the *cos274* substrate (if all available terminase bound, the linear rise in stoichiometry would have continued, yielding a complex with terminase stoichiometry of 7.3).

The data in aggregate suggest that a high-affinity interaction between terminase and *cos274*, in the presence of IHF, affords a pentamer of terminase protomers assembled on the model *cos* substrate. Higher concentrations of terminase likely lead to the further addition of a sixth terminase protomer on *cos274*; however, given the seemingly low affinity with which this sixth protomer is added, I suspect it to be a nonspecific binding event and, therefore, not indicative of the stoichiometry of the specific complex assembled by the cooperative interactions of terminase and IHF on *cos*. In the absence of IHF, two terminase protomers assemble on *cos274*. In summary, IHF appears to facilitate the high-affinity assembly of terminase protomers into a specific nucleoprotein complex of five terminase protomers on DNA.

Discussion

The data presented in the current chapter further our understanding of both IHF and terminase interactions at *cos*.

Collaborative sedimentation equilibrium (SE) studies demonstrated that two terminase protomers assemble on *cos274* (a 274 bp "specific" duplex modeling the full-length *cos* site in concatemeric λ DNA) in the absence of IHF. Whether a specific or nonspecific complex is formed in this assembly could not be directly determined from these studies; however, in light of my previous EMS and sedimentation velocity (SV) studies demonstrating nonspecific, weak complexes between terminase and *cos274* in the absence of IHF, it is not unreasonable to assume that nonspecific complexes were formed in the SE studies *sans* IHF presented currently. I do note the caveat that my previous EMS and SV studies may not be *directly* comparable to the current SE studies given the elevated concentrations of DNA (200-400 nM) used in the current studies (previously, 4 nM DNA was used in all EMS and SV studies). The saturable nature of binding, however, calls into question the nonspecific interpretation, unless two to three terminase protomers are all that a 274 bp duplex can accommodate for the nonspecific site size (currently unknown) of terminase•DNA interactions.

Next, I demonstrated via SE that the presence of IHF leads to nucleoprotein complexes between terminase and *cos274* of dramatically higher mass than in the absence of IHF. The mass increases could not be trivially explained by the mass contributed by IHF, as IHF was purposefully stoichiometrically limited in my binding reactions so that, at most, only an average of two IHF molecules could bind the *cos274* DNA substrate.

More specifically, for any given loading terminase:DNA ratio, the fit buoyant molecular weights of resulting complexes in the presence of IHF were significantly higher than in the absence of IHF. At saturation, I note a mass consistent with five terminase protomers and two IHF heterodimers. Although the masses of complexes continue to rise with the addition of more terminase, the increase in average M_b per unit of loading terminase:DNA decreases significantly beyond a loading ratio of 5:1. A stoichiometric binding (linear) fit to the data suggests that all added terminase binds *cos274* up until a pentameric terminase complex is formed on the DNA (slope=1.0, Figure 3); however, beyond a loading ratio of 5:1, terminase binds to IHF•DNA•terminase complexes with lower affinity (slope=0.16, Figure 3). More specifically, a buoyant molecular weight consistent with a terminase stoichiometry of 5.4 is obtained for a loading terminase:DNA ratio of 7.3:1, indicating that a fraction of the complexes are likely populated with a sixth, nonspecifically-bound (due to the lower affinity observed) terminase protomer. This is not surprising, as my previous EMS studies (Chapter III, Figure 5) show significant binding of terminase to nonspecific sequences. At such high concentrations of terminase (1.46

μM terminase for a 7.3:1 terminase:DNA loading ratio), I am likely forcing a nonspecific binding event of lower affinity atop the specific binding that results in a pentameric complex. Further investigation comparing specific and nonspecific DNA substrates may provide further insights into the natures of the complexes formed in the studies presented here.

Since the calculations of terminase stoichiometries on *cos274* in the presence of IHF are contingent upon an assumed mass contribution from bound IHF, I would like to elaborate on the assumption that two IHF heterodimers bind *cos274* in the presence of terminase. This assumption yielded a more reasonable explanation for the observed SE data; without this assumption, terminase stoichiometries were consistently ~0.3 (or M_b of 8.76 kDa) higher than possible given loading terminase:DNA ratios. In other words, terminase stoichiometries of ~1.3, 4.3, and 5.3 resulted from terminase:DNA loading ratios of 1, 4, and 5. It is difficult to rationalize this inconsistency as arising from an inaccurate terminase stock concentration, as any such errors would propagate as a percentage of the loading ratio. Instead, I observe a consistent increase in M_b leading to an overestimation of terminase stoichiometries by 0.3. Hence, barring some other unforeseen systematic error leading to this consistent overestimation, I found the most reasonable explanation for the additional mass to be attributable to two IHF heterodimers bound on *cos274*. Since the addition of 500 nM IHF to 200 nM *cos274* yielded a M_b increase of 4.16 kDa (top two rows, Table 3), and since the addition of two IHF heterodimers should result in a theoretical M_b increase of 11.02 kDa, assuming two

IHF heterodimer are bound to *cos274* in the presence of terminase yields a M_b increase of 6.86 (11.02-4.16) kDa over the M_b measured for *cos274* in the presence of 500 nM IHF. Considering the error in my measurements of M_b (~3.2 kDa for free DNA), 6.86 kDa is very close to the 8.76 kDa overestimation in terminase stoichiometries when it is assumed that all the IHF that is going to bind *cos274* has already done so in the absence of terminase (resulting in a M_b increase of 4.16, consistent with only 75% of the DNA populated with one IHF). Hence, I assumed two IHF heterodimers bound on *cos274* for the analysis of terminase stoichiometries in the *+IHF* experiment.

A comment should also be made regarding the observed increased DNA binding activity of terminase in the *+IHF* experiment in comparison to the *–IHF* experiment (100% vs. 48%, respectively, Figure 3). This increased activity can be explained by either of the following: (1) differential activities for specific and nonspecific binding modes of terminase; or (2) different activity of the separate terminase preps used in the *–IHF* and *+IHF* experiments.

Finally, it is noteworthy that the SE data are most consistent with stoichiometric binding. This observation suggests a K_d well below 200 nM for the binding interaction between terminase and *cos274* in the presence of IHF, and a K_d well below 400 nM for the binding interaction between terminase and *cos274* in the absence of IHF. 200 nM and 400 nM are the respective DNA substrate concentrations in the *+IHF* and *–IHF* experiments.

In conclusion, two terminase protomers bind *cos274* in the absence of IHF. The presence of a 2.5-fold stoichiometric excess of IHF (compared to DNA) facilitates the assembly of a terminase pentamer on *cos274*. Additional, lower affinity, terminase binding appears to occur at terminase:DNA ratios higher than 5:1, potentially implicating nonspecific terminase interactions with *cos274* in addition to the presumed specific interactions leading to the pentamer. Whether or not additional nonspecific interactions occur between terminase and *cos274* at terminase:DNA ratios below 5:1 cannot be directly determined from these studies. Therefore, I cannot rule out the possibility of nonspecific interactions overestimating the stoichiometry of five terminase protomers in the specific complex at *cos*. That said, the data is most suggestive of a pentamer of terminase protomers on DNA due to the high affinity (with a K_d well below the substrate concentration of 200 nM, as indicated by the stoichiometric, or linear, fit to the data in Figure 3) addition of the first five protomers.

Further experiments, particularly at differing DNA and protein concentrations, should be used to constrain these analyses in an attempt to obtain more rigorous estimates of the natures of the complexes formed between terminase and *cos*-containing DNA. Furthermore, equivalent studies with nonspecific DNA will yield further insights into the specificity of the complexes formed.

V. Conclusions and Future Directions

Double-stranded DNA viruses package their viral genomes into procapsids using terminase enzymes. Terminases bind to linear concatemers of replicated viral genomes and concomitantly excise (mature) and package a single genome per procapsid. These activities of terminases require that the enzymes be capable of both specific and nonspecific DNA binding in order to facilitate (1) site-specific DNA maturation, and (2) sequence-independent DNA packaging, respectively.

In this thesis, I demonstrate the significant role of *E. coli* integration host factor (IHF) in mediating the site-specific assembly of λ terminase at its cognate DNA site, *cos*, which serves as the packaging initiation site. IHF binds with high affinity to an I-element within *cos* and introduces a strong bend in the duplex. The presence of IHF facilitates the formation of a high-affinity nucleoprotein complex between terminase and *cos*-containing DNA. In fact, the presence of either protein increases the apparent binding affinity of the other protein for the *cos* site. The bent architecture of the IHF•DNA complex likely plays a role in this cooperative assembly of IHF and terminase at *cos*. Each protein alone, however, only modestly discriminates between *cos* and nonspecific DNA substrates. Although IHF is able to form a specific complex with *cos*-containing DNA in the absence of terminase, the data suggest that terminase protomer, at these concentrations, is unable to do so in the absence of IHF. Therefore, it is strongly evident that IHF confers site-specificity of binding to terminase protomer. Furthermore, both electrophoretic mobility shift (EMS) and

sedimentation velocity (SV) studies suggest that complexes formed between terminase and *cos*-containing DNA are weak in the absence of IHF, leading to their disruption in both gel studies and under the hydrostatic pressures experienced in SV. Notably, this is not the case when IHF is present in binding mixtures.

A challenge in almost all of the studies outlined here was the conflation of specific and nonspecific binding, and the inability to decouple the two easily in the techniques presented here. The significant amount of nonspecific protein binding of DNA, even on specific substrates, is not surprising for the following reasons: (1) nonspecific protein-DNA interactions enhance the likelihood of productive (where binding occurs) protein-DNA encounters, and (2) both proteins have been implicated in biological functions requiring sequence-independent interactions with DNA. Terminase, in particular, must nonspecifically bind DNA with formidable affinity during the translocation of viral genomes into procapsids. Nevertheless, EMS studies were able to distinguish a dramatic difference between specific (gel-shifted) complexes and nonspecific (smeared) complexes, allowing estimations of specific binding affinities of terminase and IHF for the DNA substrates studied. However, EMS studies simply could not distinguish higher-order protein-DNA binding interactions beyond those required to afford the bent, shifted complex; therefore, they did not allow me to determine the nature of the assembled nucleoprotein complexes. In contrast, AUC was able to resolve the higher-order, presumably nonspecific, protein-DNA interactions between IHF, terminase, and DNA. In fact, both SV and sedimentation equilibrium (SE) were able to monitor

nonspecific protein-DNA binding so comprehensively that it was difficult, in certain cases, to discern specific binding events from nonspecific ones. That said, SE was able to distinguish a dramatic increase in the masses of nucleoprotein complexes formed between terminase and *cos*-containing DNA in the presence, as opposed to the absence, of IHF. A mass consistent with a pentamer of terminase protomers on a *cos*-containing 274 bp duplex resulted in the presence of IHF. Since there is no evidence in my EMS studies that IHF increases the nonspecific binding affinity of terminase for DNA, the mass increase is likely attributable to the formation of a specific, bent nucleoprotein complex at *cos*. The decrease in the frictional ratio, in SV studies, upon terminase binding to the IHF•DNA complex further supports this interpretation. The data in aggregate suggest, then, that a pentamer of terminase protomers assemble at *cos* to form the *pre-nicking complex* (Chapter I, Figure 1A) that leads to viral DNA maturation prior to packaging. I cannot, however, definitively rule out the possibility of a tetramer of terminase protomers (which has shown to take the shape of a ring with a hole large enough to accommodate duplex DNA[23]) assembled in the *pre-nicking complex*, with a fifth nonspecifically bound protomer due to the elevated protomer concentrations. I find this latter explanation unlikely, though, as the lower affinity of the nonspecific terminase-DNA interaction would likely have resulted in a decrease in slope of the fit SE data beyond a terminase:DNA loading ratio of 4, as opposed to the observed decrease in slope after a terminase:DNA loading ratio of 5 (Chapter IV, Figure 3).

Further binding studies comparing the affinities and natures of nucleoprotein complexes formed with specific vs. nonspecific DNA substrates may help elucidate the differences between specific and nonspecific complexes. Such studies could further the analysis of the stoichiometry of terminase specific *pre-nicking complex*. More specifically, since salt tends to disrupt nonspecific interactions more so than specific interactions, extensive salt studies could allow one to identify conditions that better resolve specific from nonspecific binding. Ideally, conditions that vastly favor the specific binding of terminase to *cos* may allow for a more rigorous determination of the nature of the specific complexes, simply by discouraging nonspecific binding sufficiently.

Furthermore, it is important to consider that the work currently presented only addresses the nature of the *pre-nicking complex*. A pentamer of terminase protomers assembled in this complex does not preclude the possibility of different protein stoichiometries in the various other *pre-packaging* complexes (namely, the *post-nicking complex* and *complex I*; see Chapter I, Figure 1A) and, potentially, the relevance of the ring tetramer observed at high terminase protomer concentrations. In fact, changes in the DNA due to the catalytic activities of terminase itself could drive conformational rearrangements and changes in stoichiometries of the nucleoprotein complexes that drive the packaging pathway (Chapter I, Figure 1B) forward. To this end, future studies on the assembly of terminase (and IHF) on pre-assembled substrates modeling the various *pre-packaging* complexes would be informative.

Finally, it is worthwhile to speculate on the biological significance of the observed cooperativity between IHF and terminase at the *cos* site. Recent studies in our lab have shown that low concentrations of nucleotide (ATP/ADP) can overcome the IHF dependence of catalytic activity observed for the terminase protomer[64]. Previous studies in our lab suggested that the ring tetramer can perform endonuclease cleavage in the absence of IHF[23]. The ensemble of data suggest, then, that perhaps IHF and nucleotide stimulate the catalytic activities of terminase simply by facilitating the assembly of the competent, specific nucleoprotein complex at *cos*. IHF binds and bends the *cos* site to provide an architecturally favorable binding surface for terminase, while nucleotide perhaps helps nucleate assembly of a specific terminase-DNA complex by binding to protomer subunit interfaces. Terminase may have evolved to use IHF or nucleotide to more efficiently 'find' the correct viral DNA sequence (*cos*) amidst a pool of genetic material available within the cell. One could speculate that a 'less-evolved' version of terminase did not significantly differentiate bent from unbent DNA. A mutation may have arisen that allowed terminase to bind with significantly higher affinity to a duplex bent by IHF at *cos*. This 'mutant terminase' would likely be more efficient, and successful, at 'finding' and binding with high affinity to the *cos* site amongst a pool of genomic material to mature the viral DNA. This would afford an evolutionary advantage to the 'mutant terminase', which may have evolved into the lambda terminase present today.

In conclusion, this work furthers the understanding of (1) one of the factors (IHF) involved in the site-specific assembly of a nucleoprotein complex required for the initiation of viral packaging, and (2) the nature of the specific nucleoprotein complex assembled at the *cos* site prior to DNA maturation and packaging.

VI. References

1. Feiss, M. & Catalano, C. E. in *Viral Genome Packaging: Genetics, Structure, and Mechanism* (Springer, 2010).
2. Rao, V. B. & Feiss, M. The bacteriophage DNA packaging motor. *Annu. Rev. Genet.* **42,** 647–681 (2008).
3. Black, L. W. DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* **43,** 267–292 (1989).
4. Klinkert, J. & Klein, A. Roles of bacteriophage lambda gene products O and P during early and late phases of infection cycle. *J. Virol.* **25,** 730–737 (1978).
5. Skalka, A., Poonian, M. & Bartl, P. Concatemers in DNA replication: electron microscopic studies of partially denatured intracellular lambda DNA. *J. Mol. Biol.* **64,** 541–550 (1972).
6. Mousset, S. & Thomas, R. Ter, a function which generates the ends of the mature lambda chromosome. *Nature* **221,** 242–244 (1969).
7. Wang, J. C. & Kaiser, A. D. Evidence that the cohesive ends of mature lambda DNA are generated by the gene A product. *Nature New Biol.* **241,** 16–17 (1973).
8. Gold, M. & Becker, A. The bacteriophage lambda terminase. Partial purification and preliminary characterization of properties. *J. Biol. Chem.* **258,** 14619–14625 (1983).

9. Frackman, S., Siegele, D. A. & Feiss, M. The terminase of bacteriophage lambda. Functional domains for cosB binding and multimer assembly. *J. Mol. Biol.* **183,** 225–238 (1985).

10. Wu, W. F., Christiansen, S. & Feiss, M. Domains for protein-protein interactions at the N and C termini of the large subunit of bacteriophage lambda terminase. *Genetics* **119,** 477–484 (1988).

11. Parris, W., Rubinchik, S., Yang, Y. C. & Gold, M. A new procedure for the purification of the bacteriophage lambda terminase enzyme and its subunits. Properties of gene product A, the large subunit. *J. Biol. Chem.* **269,** 13564–13574 (1994).

12. Yang, Q., Hanagan, A. & Catalano, C. E. Assembly of a nucleoprotein complex required for DNA packaging by bacteriophage lambda. *Biochemistry* **36,** 2744–2752 (1997).

13. Shinder, G. & Gold, M. The Nul subunit of bacteriophage lambda terminase binds to specific sites in cos DNA. *J. Virol.* **62,** 387–392 (1988).

14. Yang, Q. *et al.* Cloning, expression, and characterization of a DNA binding domain of gpNu1, a phage lambda DNA packaging protein. *Biochemistry* **38,** 465–477 (1999).

15. Kosturko, L. D., Daub, E. & Murialdo, H. The interaction of E. coli integration host factor and lambda cos DNA: multiple complex formation and protein-induced bending. *Nucleic Acids Res.* **17,** 317–334 (1989).

16. Shinder, G. & Gold, M. Integration host factor (IHF) stimulates binding of the gpNu1 subunit of lambda terminase to cos DNA. *Nucleic Acids Res.* **17,** 2005–2022 (1989).

17. de Beer, T. *et al.* Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Mol. Cell* **9,** 981–991 (2002).

18. Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87,** 1295–1306 (1996).

19. Hwang, Y. & Feiss, M. Mutations affecting lysine-35 of gpNu1, the small subunit of bacteriophage lambda terminase, alter the strength and specificity of holoterminase interactions with DNA. *Virology* **231,** 218–230 (1997).

20. Ortega, M. E. & Catalano, C. E. Bacteriophage lambda gpNu1 and Escherichia coli IHF proteins cooperatively bind and bend viral DNA: implications for the assembly of a genome-packaging motor. *Biochemistry* **45,** 5180–5189 (2006).

21. Tomka, M. A. & Catalano, C. E. Physical and kinetic characterization of the DNA packaging enzyme from bacteriophage lambda. *J. Biol. Chem.* **268,** 3056–3065 (1993).

22. Maluf, N. K., Yang, Q. & Catalano, C. E. Self-association properties of the bacteriophage lambda terminase holoenzyme: implications for the DNA packaging motor. *J. Mol. Biol.* **347,** 523–542 (2005).

23. Maluf, N. K., Gaussier, H., Bogner, E., Feiss, M. & Catalano, C. E. Assembly of bacteriophage lambda terminase into a viral DNA maturation and packaging machine. *Biochemistry* **45,** 15259–15268 (2006).

24. Kondabagil, K. R. & Rao, V. B. A critical coiled coil motif in the small terminase, gp16, from bacteriophage T4: insights into DNA packaging initiation and assembly of packaging motor. *J. Mol. Biol.* **358,** 67–82 (2006).

25. Pu, W. T. & Struhl, K. The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 6901–6905 (1991).

26. Catalano, C. E. The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. *Cell. Mol. Life Sci.* **57,** 128–148 (2000).

27. Davidson, A., Yau, P., Murialdo, H. & Gold, M. Isolation and characterization of mutations in the bacteriophage lambda terminase genes. *J. Bacteriol.* **173,** 5086–5096 (1991).

28. Mendelson, I., Gottesman, M. & Oppenheim, A. B. HU and integration host factor function as auxiliary proteins in cleavage of phage lambda cohesive ends by terminase. *J. Bacteriol.* **173,** 1670–1676 (1991).

29. Becker, A., Marko, M. & Gold, M. Early events in the in vitro packaging of bacteriophage lambda DNA. *Virology* **78,** 291–305 (1977).

30. Becker, A., Murialdo, H. & Gold, M. Studies on an in vitro system for the packaging and maturation of phage lambda DNA. *Virology* **78,** 277–290 (1977).

31. Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69,** 651–697 (2000).

32. Hendrix, R. W. *Lambda II*. (1983).

33. Friedman, D. I. Integration host factor: a protein for all reasons. *Cell* **55,** 545–554 (1988).

34. Nash, H. A., Robertson, C. A., Flamm, E., Weisberg, R. A. & Miller, H. I. Overproduction of Escherichia coli integration host factor, a protein with nonidentical subunits. *J. Bacteriol.* **169,** 4124–4127 (1987).

35. Khrapunov, S., Brenowitz, M., Rice, P. A. & Catalano, C. E. Binding then bending: a mechanism for wrapping DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 19217–19218 (2006).

36. Ali, B. M. *et al.* Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10658–10663 (2001).

37. Kuznetsov, S. V., Sugimura, S., Vivas, P., Crothers, D. M. & Ansari, A. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 18515–18520 (2006).

38. Vivas, P., Kuznetsov, S. V. & Ansari, A. New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. *J Phys Chem B* **112,** 5997–6007 (2008).

39. Holbrook, J. A., Tsodikov, O. V., Saecker, R. M. & Record, M. T. Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J. Mol. Biol.* **310,** 379–401 (2001).

40. Aeling, K. A. *et al.* Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy. *J. Biol. Chem.* **281,** 39236–39248 (2006).

41. Xin, W. N. & Feiss, M. The interaction of Escherichia coli integration host factor with the cohesive end sites of phages lambda and 21. *Nucleic Acids Res.* **16,** 2015–2030 (1988).

42. Craig, N. L. & Nash, H. A. E. coli integration host factor binds to specific sites in DNA. *Cell* **39,** 707–716 (1984).

43. Hales, L. M., Gumport, R. I. & Gardner, J. F. Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J. Bacteriol.* **176,** 2999–3006 (1994).

44. Hales, L. M., Gumport, R. I. & Gardner, J. F. Examining the contribution of a dA+dT element to the conformation of Escherichia coli integration host factor-DNA complexes. *Nucleic Acids Res.* **24,** 1780–1786 (1996).

45. Goodrich, J. A., Schwartz, M. L. & McClure, W. R. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). *Nucleic Acids Res.* **18,** 4993–5000 (1990).

46. *IR700 Dye (idtdna.com)*. <https://www.idtdna.com/catalog/Modifications/Modifications.aspx?ProductID=1723>

47. *Licor Odyssey Scanner*. <http://www.licor.com/bio/products/imaging_systems/odyssey_family.jsp>

48. *6-FAM dye (idtdna.com)*. <https://www.idtdna.com/catalog/Modifications/Modifications.aspx?ProductID=1108>

49. *DNA Molecular Weight Calculator*. <http://www.bioinformatics.org/sms2/dna_mw.html>

50. Bear, S. E., Court, D. L. & Friedman, D. I. An accessory role for Escherichia coli integration host factor: characterization of a lambda mutant dependent upon integration host factor for DNA packaging. *J. Virol.* **52,** 966–972 (1984).

51. Filutowicz, M., Grimek, H. & Appelt, K. Purification of the Escherichia coli integration host factor (IHF) in one chromatographic step. *Gene* **147,** 149–150 (1994).

52. *SEDNTERP*. <http://sednterp.unh.edu/>

53. *IHF α subunit sequence*. <http://www.ncbi.nlm.nih.gov/protein/YP_489974.1>

54. *IHF β subunit sequence*. <http://www.ncbi.nlm.nih.gov/protein/YP_489184.1>

55. *SEDFIT*. at <http://www.analyticalultracentrifugation.com/download.htm>

56. Anuchin, A. M., Goncharenko, A. V., Demidenok, O. I. & Kaprel'iants, A. S. [Histone-like proteins of bacteria (review)]. *Prikl. Biokhim. Mikrobiol.* **47,** 635–641 (2011).

57. Anderson, B. J., Larkin, C., Guja, K. & Schildbach, J. F. Using fluorophore-labeled oligonucleotides to measure affinities of protein-DNA interactions. *Meth. Enzymol.* **450,** 253–272 (2008).

58. Andrews, B. T. & Catalano, C. E. The enzymology of a viral genome packaging motor is influenced by the assembly state of the motor subunits. *Biochemistry* **51,** 9342–9353 (2012).

59. Hang, Q., Woods, L., Feiss, M. & Catalano, C. E. Cloning, expression, and biochemical characterization of hexahistidine-tagged terminase proteins. *J. Biol. Chem.* **274,** 15305–15314 (1999).

60. Infante, A. A. & Baierlein, R. Pressure-induced dissociation of sedimenting ribosomes: effect on sedimentation patterns. *Proc. Natl. Acad. Sci. U.S.A.* **68,** 1780–1785 (1971).

61. Silva, J. L., Foguel, D. & Royer, C. A. Pressure provides new insights into protein folding, dynamics and structure. *Trends Biochem. Sci.* **26,** 612–618 (2001).

62. Heremans, K. High pressure effects on proteins and other biomolecules. *Annu. Rev. Biophys. Bioeng.* **11,** 1–21 (1982).

63. Senear, D. F. *et al.* Pressure dissociation of integration host factor-DNA complexes reveals flexibility-dependent structural variation at the protein-DNA interface. *Nucleic Acids Res.* **35,** 1761–1772 (2007).

64. Chang, J. R., Andrews, B. T. & Catalano, C. E. Energy-independent helicase activity of a viral genome packaging motor. *Biochemistry* **51,** 391–400 (2012).

65. *Cy3 dye (idtdna.com)*. <https://www.idtdna.com/catalog/Modifications/Modifications.aspx?ProductID=1106>