Biochemical, structural and functional analysis of an engineered DNA architectural protein

Bao, Qiuye

2007

Bao, Q. Y. (2007). Biochemical, structural and functional analysis of an engineered DNA architectural protein. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/6565

https://doi.org/10.32657/10356/6565

NANYANG TECHNOLOGICAL UNIVERSITY

Biochemical, Structural and Functional Analysis of an Engineered DNA Architectural Protein

BAO QIUYE

SCHOOL OF BIOLOGICAL SCIENCES

2007

Biochemical, Structural and Functional Analysis of an Engineered DNA Architectural Protein

Bao Qiuye

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2007

Abstract

As a DNA architectural protein from *E. coli*, integration host factor (IHF) can bind and bend DNA in a sequence specific manner. It plays an essential role in a variety of DNA transactions including recombination, transcription and replication. IHF's ability of concerted binding to and bending of DNA is key to its biological function. Here we report the design, characterization, and application of a single polypeptide chain IHF, termed scIHF2. In a novel approach for protein engineering, we inserted almost the entire α subunit of heterodimeric IHF into the β subunit at peptide bond Q39/G40 via two short linkers. DNA-binding and -bending assays revealed that the purified wild-type IHF and scIHF2 behave very similarly. Further, scIHF2 is required for site-specific integrative recombination by phage λ integrase and for pSC101 replication in a ΔIHF *E. coli* host. We also demonstrate that scIHF2 is stably expressed in HeLa cells and localized primarily in the cell nucleus, and that it can render cells sensitive to the chemotherapeutic drug cisplatin. Hence, scIHF2 may be used as a novel regulatory co-factor for recombination or other DNA transactions in mammalian cells, and provides a possible target for studies on anti-cancer reagents.

We have also generated other scIHF variants by shortening one or both linkers in scIHF2. These variants exhibit distinct DNA-binding properties and phenotypes in site-specific integrative and excisive recombination by phage λ integrase *in vitro*, as well as in pSC101 replication in a ΔIHF *E. coli* host. We also introduced a K45E substitution within the α domain of variant scIHF3, thus generating a new variant termed scIHF3E. Based on electrophoretic mobility shift assays, it is suggested that scIHF3E exhibits different DNA-bending properties. In order to analyze this in more detail, two scIHF2 derivatives, scIHF2E and scIHF2ER, were generated by site-directed mutagenesis. Both variants bind less tightly to the "left" arm of a target DNA (the A-tract side) as compared to the parental scIHF2. We found that scIHF2E adopts two stable conformational states in complex with a specific DNA

target. In the so-called open state, the degree of protein-induced DNA bending is significantly reduced compared with the closed state. The conformational switch between these states is controlled by divalent metal binding in two electronegative zones arising from the lysine-to-glutamate substitution in the protein body proximal to the phosphate backbone of one DNA arm. scIHF2E was employed to gain insight into the topography of functional nucleoprotein complexes involved in phage λ site-specific recombination, which has illustrated how functional complex formation can be affected by an intricate relationship between DNA architectural proteins and external factors.

Acknowledgement

I would like to express special gratitude to my supervisor Dr. P. Drőge for his direction, encouragement and help during the process of the project. His forthrightness and serious attitude about science gave me deep impression and had a huge effect on me.

Many thanks go to Dr. C. A. Davey who gave me important directions on the second part of my work and, equally important, contributed the solution of the IHF-DNA co-crystal structures. His knowledge opened another avenue of sciences for me. Many thanks go to Drs J. Yan, H. Chen, and Y. Liu who contributed the atomic force microscopy analysis of IHF-DNA complexes.

Thanks also go to Drs. Christ and Schwartz. In an earlier collaboration with my supervisor, Dr. P. Drőge, they laid the foundation for most of the work on scIHF described here.

Special thanks go to Dr. S. D. Goodman who provided IHF antibodies and plasmid SG86. He also suggested the use of the minimal pSC101 origin of replication. Thanks also go to Dr. A. Segall who provided purified Int and Xis protein.

I also extend my thanks to my colleagues who provided assistance, convenience and a friendly atmosphere. Without them, my project couldn't be finished so smoothly. Many thanks go to Chew Yuan Yuan, Sabrina d/o Peter, Dr. Klaus Neef, Li Heng, Cao Shenglan and Tai Seng Thong. They gave me comments and suggestions on the writing of the thesis.

Finally, I will give sincere thanks to my mother, my sister and my best friends forever. They gave me selfless care and encouragement to finish my Ph.D. study.

I am grateful for the financial support of predoctoral scholarship from Nanyang Technology University during my study.

Abbreviations

aa	amino acids
AFM	atomic force microscopy
*att*B	attachment site, bacterium
*att*P	attachment site, phage
bp	base pair
BSA	bovine serum albumin
Cox	control of excision
DMEM	Dulbecco's Modified Eagle's Medium
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid
EMSA	electrophoretic mobility shift assay
FACS	fluorescent activated cell sorting
Fis	factor for inversion stimulation
GFP	green fluorescent protein
HMG	high mobility group
IHF	integration host factor
IPTG	isopropyl-beta-D-thiogalactopyranoside
Int	integrase
kDa	kilodaltons
Kb	kilobases
LB	Luria-Bertani
NCP	nucleosome core particle
NER	nucleotide excision repair
NLS	nuclear localization signal
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG5K-MME	polyethylene glycol 5000 monomethy ether
PI	propidium iodine
PVDF	polyvinylidnene fluoride

SDS	sodium dodecyl sulfate
Snups	specialized nucleoprotein structures
TBP	TATA-box bindng protein
TEMED	N, N, N', N'-tetramethylethylenediamine
Tris	Tris (hydroxmethyl)-aminomethane
v/v	volume per volume
wt	wild type
w/v	weight per volume
Xis	excisionase

Table of Contents

A. Introduction

A.1. Overview

Integration host factor (IHF) was identified as an essential cofactor for phage λ site-specific integration into the genome of *E. coli* (Thompson et al., 1986). Subsequently, IHF was found in a number of Gram-negative bacteria (Calb et al., 1996). IHF is involved in the control of chromosomal DNA replication, in transcriptional regulation and, in more general terms, facilitates the formation of higher order nucleoprotein complexes (Friedman, 1988). As a DNA architectural protein, IHF binds to DNA in a sequence-specific manner and strongly bends DNA, as deduced from the co-crystal structure of an IHF-H' complex (Ellenberger and Landy, 1997; Rice et al., 1996).

Integration host factor is a heterodimer composed of two subunits, IHF α (11.3 kDa) and IHF β (10.6 kDa), which are encoded by *ihf* A and *ihf* D, respectively. The two subunits share about 30% homology at the level of amino acids identity (Engelhorn and Geiselmann, 1998). These two subunits form a compact protein body with two long β ribbon arms extending. On top of each arm, there is one conserved proline which intercalates into the DNA target, causing one kink per subunit 9bp apart. This intercalation interrupts the base pair stacking and widens the minor groove of DNA, making it open on either side of the center of the bend to facilitate the tight wrapping of DNA around IHF (Ellenberger and Landy, 1997). The electrostatic interactions between the body of IHF and DNA stabilize the bent conformation.

The binding affinity of IHF is relatively high for example, the K_d for the IHF-H' complex is below 10^{-9} M (Goodman et al., 1999; Yang and Nash, 1995).

To bind and bend DNA in a sequence specific manner is essential for IHF in order to achieve its diverse biological functions. In addition to Int, IHF is the only factor needed for site-specific integration of phage λ into the *E. coli* genome. On the phage λ recombination site *att*P, there are three IHF binding sites, named H', H1 and H2. The binding of IHF to *att*P generates DNA bends to assist *att*P to adopt a compact structure that is competent for recombination (Robertson and Nash, 1988). The integration site (*att*B) on the *E. coli* genome comprising only 21 bp and lacks additional cofactors binding sites (Christ et al., 2002; Nash, 1981). For site-specific excisive recombination at recombinant sites *att*L and *att*R, an additional cofactor, Xis, is needed. For replication of low copy number plasmid pSC101, the dnaA initiator protein encoded by *E. coli* and a plasmid-encoded initiator protein repA are obligatory requirements (Vocke and Bastia, 1983). The initiation of pSC101 plasmid replication requires the interaction between dnaA and repA, which is facilitated by IHF.

A.2. The Structure of IHF

DNA recognition

There are two different kinds of protein-DNA recognition mechanisms: direct readout and indirect readout. Direct readout implies that the protein recognizes unique functional groups of a DNA sequence. Indirect readout includes recognition

of sequence-dependent structural features of DNA, such as backbone conformations of the major and minor grooves. IHF is one example of proteins that relies mostly on indirect readout during DNA recognition. (Lynch et al., 2003)

Cognate sequences

IHF recognizes and binds to a limited consensus sequence: WATCAANNNNTTR, where W is A or T and R is A or G. Many IHF binding sites contain a 4-6 bp dA/dT-rich element located one helical turn 5' to the start of the conserved sequence element (Engelhorn and Geiselmann, 1998). Rice and co-workers first reported the crystal structure of IHF-DNA complex in 1996. The DNA substrate that they used for IHF-DNA co-crystallization was a 35 bp DNA fragment containing the H' site of phage λ, which was regarded as one of the best characterized IHF binding sites (Yang and Nash, 1995).

Overall protein structure

In the IHF-DNA complex, the α and β subunits of IHF intertwine to form a compact core with two extending β ribbon arms. The two helices of each subunit form the bottom of the body. And a third short α helix as well as two antiparallel β-sheets from each subunit make up the upper part of the body. The two arms extending from the body are comprised of two antiparallel β-sheets. The final DNA path forms a U-turn and wraps around IHF (Ellenberger and Landy, 1997). The center of the U turn is positioned at the 5'-end of the TATCAA conserved site (Fig.A.1). The bending angle was most prominent at two kinks, which are 9 bp

apart and caused by intercalation of the two hydrophobic residues, Pro65α and Pro64β at the tip of each arm. This structure is further stabilized by electrostatic interactions between the main body and the DNA (Bewley et al., 1998; Holbrook et al., 2001).

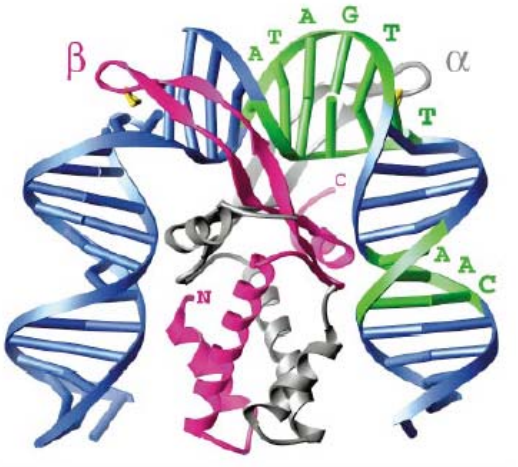


Fig.A.1. Complex of IHF with H'. The α subunit is shown in silver and the β subunit in pink. The consensus sequence is highlighted in green and interacts mainly with the arm of α and the body of β (Reproduced from Rice, PA, 1996)

Clamps of IHF

The interwined α and β subunits of IHF contact the bent DNA in a roughly symmetrical manner. However, the clamps of IHF contact poly(dA) in the left and the conserved bases in the right half of binding sites asymmetrically (Lorenz et al., 1999). IHF clamps are composed of helix 1 of one subunit joining with the β1β2 loop and the N-terminus of helix 3 of the second subunit. At the center of each clamp is the turn of β1β2 strands which lies between two phosphates on opposite sides of the minor groove and is hydrogen bonded to both via successive amide nitrogens (Rice et al., 1996). The N-terminals of helices 1 and 3, separated by β1β2 loop, also contact phosphates backbone of DNA in the minor groove (Fig.A.2A). In

the right side, the conserved sequence element TTG which forms a narrow minor groove is held by the clamp. At this position, βR46 extending from β1β2 loop contacts the edges of conserved bases (Fig.A.2B). In the left side, the clamp interacts with a short poly(dA) via direct and water mediated contact with the DNA phosphate but without direct contact with the bases. A tracts have several features: they tend to be rigid, straight, have a narrow minor groove and a high propeller twist (DiGabriele et al., 1989; DiGabriele and Steitz, 1993). The conformation of the oligo(dA) with a narrow minor goove fits better into the protein clamps and favors IHF binding (Rice et al., 1996). (Fig.A.2)

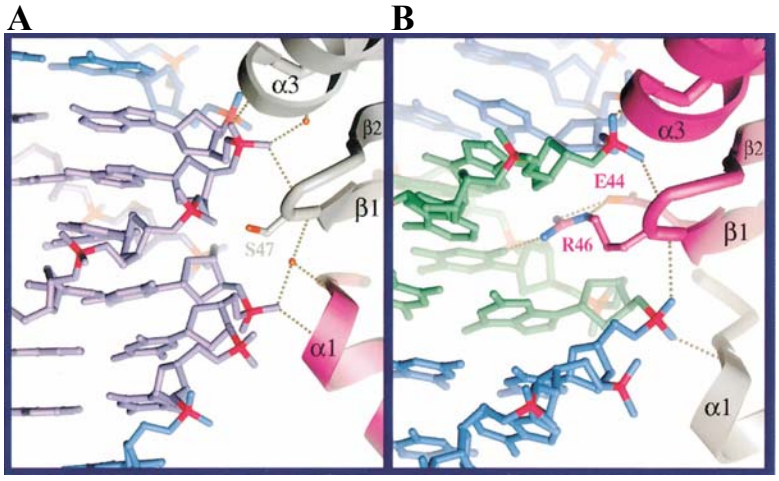


Fig.A.2. The interaction between protein clamps and the minor groove. (A) The left side: β1, β2 strands and α helix 3 of α subunit are shown in silver; α1 of β subunit is in pink. The six A tract is lavender with phosphate atoms in red. Two water molecules involved in hydrogen bonding scheme are in red. (B) The right site: The contacts between IHF body and conserved element TTG. The same as (A), IHFα is in silver; β in pink. (Reproduced from Rice, P. A., et al., 1996)

Direct interaction with DNA bases via hydrogen bonds

In the crystal structure of IHF-H', only three protein side-chains were found to form hydrogen bonds with the DNA bases which locate in the minor groove at

positions where all four bases display similar hydrogen bond acceptors (Lynch et al., 2003). Two arginines at position 60 and 63 in the α subunit reach the minor groove and contact the conserved bases in the A/TATCAA element (Fig.A.3), while the arginine at position 46 in the β subunit makes a direct contact to the second T of the conserved element TTR (Rice et al., 1996). The salt bridge chain between βR46-βE44-βR42 positions the guanidinium of βR46 centered in the minor groove. (Fig.A.2B)

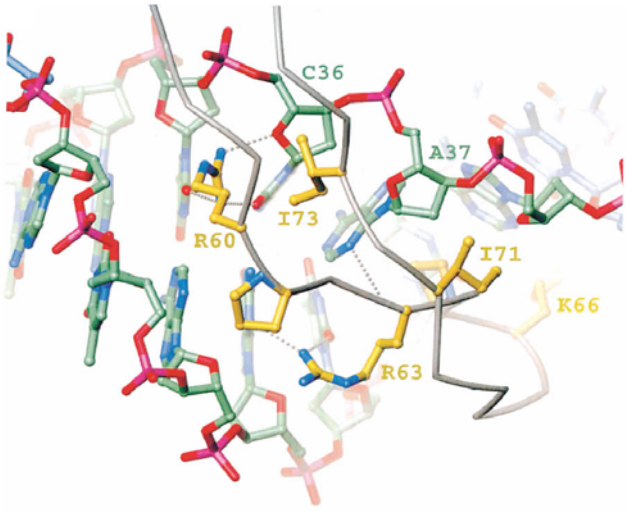


Fig. A.3. The direct contact of IHFα to DNA bases. The protein C*a* trace is shown in grey, with side-chains that interact directly with the DNA in yellow. Carbons in the consensus sequence bases are green; others are blue. (Reproduced from Rice et al., 1996).

Hydrophobic interaction between IHF and DNA

Compared with other minor groove binding architecture proteins, such as TBP and HMG proteins, IHF has less hydrophobic contacts with DNA. In addition to the interaction between the conserved amino acid- proline and DNA bases, others involve contacts between the pseudosymmetrically displaced Ile71α/Val70β , Ile73α/Leu72 β, and deoxyribose rings nearby the kinks, as well as contacts from

Pro61α to the sugar of T-34 (Bewley et al., 1998)

Mechanism used for DNA bending

In order to achieve DNA-bending, IHF applies 26 positively charged side chains and the N-terminal of all six helices to contact the phosphate backbone with positive surface of IHF body inside the U-turn in order to counteract the symmetric repulsion of the double helix (Strauss and Maher, 1994). Among 26 positively charged side chains, there are 23 lysine, arginine and histitine cationic groups within 6 Å of anionic DNA phosphate oxygens (Holbrook et al., 2001; Saecker and Record, 2002). To counteract the favorable energy of base stacking during DNA bending, the β arms of IHF wrap around the opposite face of the DNA and hydrophobic residues-prolines intercalate into the minor groove forming two kinks. Compared to other amino acids, proline has predominance to facilitate IHF to bend DNA. Its width allows extensive hydrophobic contact with the DNA bases, and its shortness and width allow formation of a hydrogen bond between the peptide backbone and N3 of adenine immediately 5' to each kink (Rice et al., 1996).

A.3. Biological properties of IHF

A.3.1. IHF is a cofactor for site-specific recombination

Conservative site specific recombination

Recombination is the process of exchanging DNA segments, leading to intermolecular or intramolecular reorganization. There are three different types of recombination mechanisms: homologous recombination; conservative site-specific

recombination and transpositional recombination. The last two recombination belongs to site-specific recombination, a reaction in which DNA strands are broken and exchanged at precise positions of two target DNA loci to achieve determined biological function (Hallet and Sherratt, 1997). Both types of recombination are represented by diverse genetic systems which generally encode their own recombination enzymes and differ dramatically in their genetic consequences and reaction mechanisms (Craig, 1988).

Unlike homologous recombination and transpositional recombination, conservative site-specific recombination involves the production of a very short homoduplex joint, and, therefore, requires a short DNA sequence that is the same on both donor and recipient DNA molecules. The exchange mechanism involves exact cleavage and joining without DNA synthesis or loss.

Recombinases

Based on their chemical mechanism and evolutionary relationships, recombinases catalyzing the conservative site-specific recombination are divided into two major families: serine recombinase (resolvases/invertase) and tyrosine recombinase, (Groth and Calos, 2004). Members of serine recombinase family have a conserved serine located near to the N-terminal and a catalytic domain containing a conserved Arg-Ser-Arg triad (Smith and Thorpe, 2002). The conserved serine is responsible for DNA cleavage in catalysis. Serine family members include the resolvases γδ and Tn3, the invertase Gin and the phage integrases ΦC31, R4, and TP901-

1(Breuner et al., 2001; Kahmann et al., 1984; Kuhstoss and Rao, 1991; Matsuura et al., 1996). Members of tyrosine recombinases on the other hand are less conserved by sequence but share, in addition to the catalytic tyrosine, five amino acids forming the RKHRH pentad of the active site, which has been shown to be involved in catalysis (Nunes-Duby et al., 1998).

Tyrosine recombinases

The members of tyrosine recombinases share little sequence identity except for several highly conserved amino acids in the C-terminal region (Argos et al., 1986; Nunes-Duby et al., 1998). The tyrosine recombinases are structurally diverse and functionally versatile. They include integrases, resolvases, invertases and transposases (Smith and Thorpe, 2002). With over 50 years of genetic, structural and biochemical characterizations of the site-specific recombination system of the prototypical member phage λ, the tyrosine family is also commonly referred to as the λ Int family (Grainge and Jayaram, 1999; Smith and Thorpe, 2002). Other well known members of this family include Cre recombinase from phage P1, Flp encoded by Saccharomyces cerevisiae 2 micron plasmid (2μm) and the bacterial protein XerC and XerD (Blakely et al., 2000; Chen et al., 1991; Petyuk et al., 2004). Crystal structures of members of this family have been solved in the last decade: Cre complexed with or without DNA (Chen et al., 2000; Guo et al., 1997; Guo et al., 1999); Flp recombinase bound to DNA (Chen et al., 2000); intact *E. coli* XerD (Subramanya et al., 1997); catalytic domain of HP1 phage integrase (Hickman et al., 1997); the full length and the core domain of λ integrase (Biswas et al., 2005;

Kwon et al., 1997).

The following reaction pathway for tyrosine recombinases have been proposed (Fig.A.4): The top strands of two aligned recombination sites are cleaved and joined by recombinase tetramer, forming an intermediate Holliday junction structure (HJ). After isomerization of HJ, the bottom strands are cleaved and ligated. The four recombinase subunits work together, but only two of them are positioned (active) for cleavage of the DNA at a time. During recombination, the scissile phosphate in the DNA backbone is attacked by the active site tyrosine, forming a covalent 3′-phosphotyrosine linkage and leaving a free 5′-hydroxyl group (5′-OH). The 5′-OH then exchanges and performs another nucleophilic attack on the phosphotyrosine linkage at the opposite strand (Chen et al., 2000). Crystal structures of reaction intermediates in different tyrosine recombinases, together with recent biochemical studies in the field, support a "strand swapping" model for recombination that does not require branch migration of the Holliday junction intermediate in order to test homology between recombining sites (Van Duyne, 2001). The strand swapping model involves melting of 2-3 bp in the middle of the overlap region, and an isomerization (swapping) of the HJ that positions the other two Y residues in the previously inactive subunits for cleavage (i.e. close enough to the DNA backbone). The second cleavage and joining reaction completes the reaction (Chen and Rice, 2003; Gopaul and Duyne, 1999).

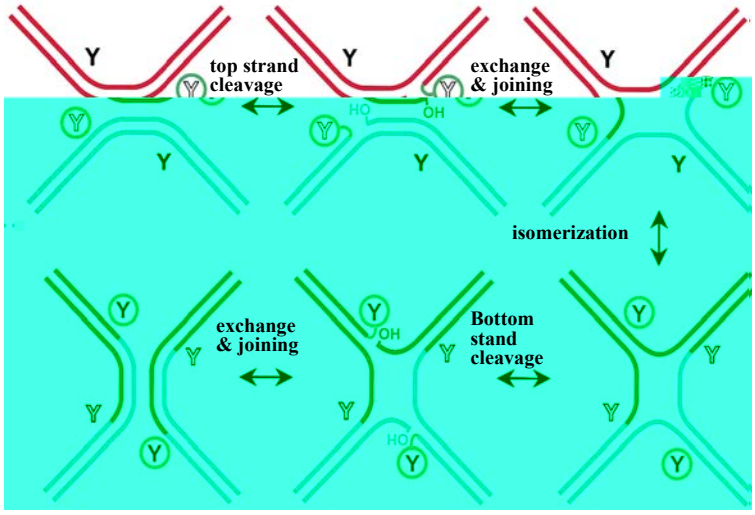


Fig.A.4. Mechanism of site specific recombination catalyzed by tyrosine recombinases. Reombinases catalyze the reaction as a tetramer carrying a tyrosine nucleophile in each active site. Only one pair of molecules is active at a given time, as indicated by the green circles. These tyrosines directly attack the scissile phosphodiester bonds resulting in top strand cleavage, after exchange and joining of the 5'-OH free strands, a Holliday junction intermediate is formed. Isomerization of the complex activates the other pair of tyrosines, initiating a second round of DNA cleavage and strand exchange reactions, which resolves the Holliday junction into recombinant products. (Modified from Chen, Y., et al., 2000)

Cre and Flp

Cre and Flp catalyze recombination without accessory proteins (Gronostajski and Sadowski, 1985; Sauer and McDermott, 2004). Each enzyme consists of a 13 kDa $NH_{(2)}$-terminal domain and a larger COOH-terminal domain that contains the active site of the enzyme as well as the major determinants for the binding specificity of the recombinase to its cognate DNA binding site (Shaikh and Sadowski, 2000).

Cre and Flp catalyze site-specific recombination between two identical sequences which are less than 50 bp (Groth and Calos, 2004). The recognition site of Cre is

loxP with two 13 bp inverted repeats separated by 8 bp overlap region. The natural Flp recombination target (FRT) contains three Flp-binding sites, where the subsite far from the cleavage spacer is unnecessary for recombination *in vivo* and *in vitro* (Chen et al., 2000). For many years, the Cre/lox system has commonly been used for gene targeting, especially in mice (Gelman et al., 2003).

The λ site-specific recombination system

λ integrase is encoded by phage λ and catalyzes the integration and excision of the phage λ genome into and out of the *E. coli* chromosome, respectively. These two recombination pathways employ pairs of different so-called attachment (*att*) DNA sites termed *att*B/*att*P and attL/attR, respectively (reviewed in Weisberg et al., 1983). The bacterial attachment site *att*B is only 21 bp, with two imperfect inverted repeats (B, B'), flanking a 7 bp overlap region (O). The phage attachment site *att*P is more complex than *att*B. In addition to two core binding sites for Int, C and C', which are separated by the overlap region (7bp) that is identical in *att*B, there are five integrase arm binding sites and three integration host factor (IHF) binding sites, as well as Xis/FIS binding sites on the P arm (Rutkai et al., 2003). The integrase arm binding sites consist of two inverted repeats (P1, P2) and three direct repeats (P′1, P′2, P′3). The P1, P′2 and P′3 sequences are required to be bound by Int during integration whereas the P2, P′1 and P′2 sites are occupied during excision (Numrych et al., 1990).

λ Int is a heterobivalent protein with two DNA-binding domains and one catalytic

domain (Christ et al., 2002). The N terminal domain of λ integrase is the arm-DNA binding domain and involved in the interaction with C-terminal tail of Xis which is a cofactor in this system (Sam et al., 2002; Wojciak et al., 2002). It has also been shown to be a context sensitive modulator of integrase functions. The N-terminal domain inhibits core-DNA binding and cleavage. This inhibition is overcome in the presence of arm-type oligonucleotides, which form specific complexes with Int and core-type DNA. Further, when the N-terminal domain is present in *trans*, it has been shown to stimulate core-DNA binding and cleavage (Sarkar et al., 2001). The core binding domain in bacteriophage λ integrase is connected to the arm binding domain by a short linker and shows structural similarities to XerD and Cre recombinases (Swalla et al., 2003). The C-terminal domain is a catalytic domain containing the highly conserved pentad (RKHRH) of the active site and the catalytic Tyrosine 342 (Tirumalai et al., 1997).

Accessory factors in the λ integrase system

In λ integrase system, there are three cofactors: IHF, Xis and Fis. Their expression levels change in response to host physiology. Thus they can control the action of Int and decide the state of phage λ, i.e. integated into or excised from bacterial genome (Radman-Livaja et al., 2006). Fis (factor for inversion stimulation) is a host encoded protein which can bind to a specific region in the lambda *att*P overlapping the Xis (excisionase) binding sites and can bind cooperatively with Xis to these sites, but cannot support excision in the absence of Xis (Thompson et al., 1987). IHF (integration host factor) facilitates Int binding to core binding sites (low

affinity) after the N terminal of Int is bound to arm binding sites (high affinity), thus forming a recombinogenic nucleoprotein complex. The phage-encoded factor Xis is critical required for the excision of phage λ from bacterial chromosome, and functions both as a DNA architectural factor and by cooperatively recruiting integrase to an adjacent binding site specifically required for excision (Sam et al., 2002). Xis can work on its own or in combination with Fis. Overexpression of Xis inhibits integration of lambda and promotes curing of established lysogens, and Xis prevents *att*P and *att*B products of excision from reverting back to *att*L and *att*R (Leffers and Gottesman, 1998; Nash, 1975). Excision is inhibited by high concentrations of IHF (Bushman et al., 1985).

IHF and the λ integrase system

IHF's ability to introduce DNA bending appears important to its biological function. This became clear from early studies of IHF's role as an essential cofactor in phage λ site-specific recombination where the protein serves an architectural role during the assembly of recombinogenic specialized nucleoprotein structures (snups) (reviewed by Azaro and Landy, 2002). The 240 bp comprising phage attachment site (*att*P) is one of the two recombination partner sequences in the integrative pathway. It harbors three specific IHF binding sites, named H1, H2, and H′. Each site must be occupied by IHF for the assembly of a recombinogenic snup. This so-called integrative intasome is composed of *att*P, IHF, and the phage-encoded integrase (Int). The integrative intasome then captures the protein-free 21 bp bacterial attachment site (*att*B) on the *E. coli* genome to form a synaptic complex

in which two successive rounds of DNA strand exchanges are catalyzed by Int.

The excisive pathway which employs hybrid attachment sites *att*L and *att*R as recombination partners also requires IHF. Excision, however, involves only H2 and H′ complexed with and bent by IHF. A second DNA-bending cofactor, the phage-encoded excisionase (Xis) protein, is involved in this pathway and binds specifically at two sites in the vicinity of H2. At limiting concentrations of Xis, the host-encoded factor for inversion stimulation (Fis) is also required for efficient excision. Hence, the exquisitely regulated λ site-specific recombination system employs three different DNA architectural proteins, i.e., IHF, Xis, and Fis. Their main function is to construct high-precision DNA trajectories within snups to allow for the heterobivalent Int monomers to establish contacts between separate DNA sites simultaneously and to form synaptic complexes between snups. (FigA.5, 6)

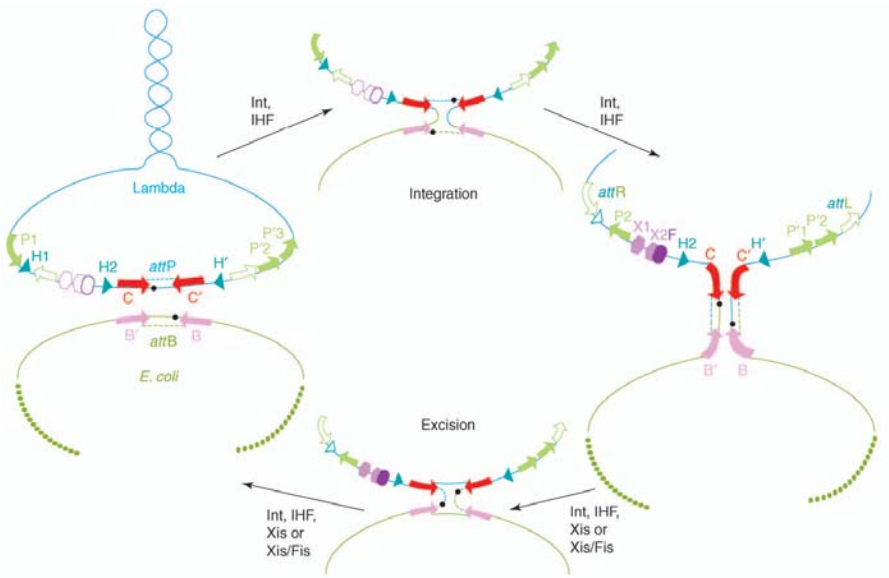


Fig. A.5. Integrative and excisive site-specific recombination. Integrative recombination between the phage *att*P and the bacterial *att*B sites (top) requires supercoiled *att*P DNA, Int and IHF. The Int pair bound to C and B core sites (red and pink arrows, respectively) and P1, P'2, P'3 arm sites

(green filled arrow). All three IHF binding sites (H1, H2, H', blue triangles), are occupied. After the top strand is cleaved and ligated, HJ intermediate is formed. Then the bottom strand is cleaved and ligated, recombinant products *att*R and *att*L are produced. *att*R and *att*L are shown here as the substrates of excisive recombination (bottom), a reaction that additionally requires Xis (and Fis, *in vivo*) (X1 and X2, pink hexagons; F, purple oval). IHF binds to H2 and H', while arm binding sites P2, P'1 and P'2 are bound by Int. Core-type Int-binding sites (C, C', B and B') and the seven base pair overlap region (sequence between cleavage sites, marked by black dots) comprise the 'core region'. (Reproduced from Radman-Livaja, M., et al., 2006)

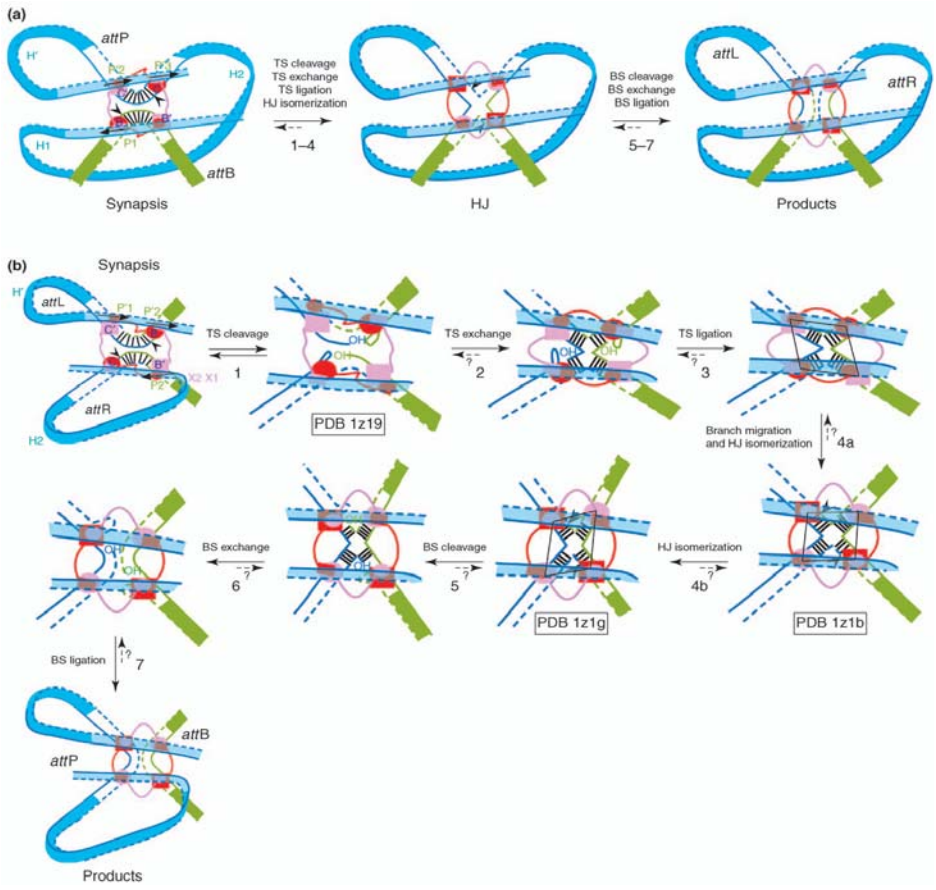


Fig.A.6. the reaction steps of integration and excision. (a) Integration. IHF binds to H1, H2 and H' to facilate int forming intasome. Int acts as tetramer. Active and inactive Int monomers are colored red and pink, respectively. (b) The mechanism of excision. DNA substrates are *att*L and *att*R. H' and H2 are bound by IHF, and Xis is required to bind to its binding sites on *att*R. The top strands are cleaved and ligated forming HJ intermediate. During isomerization, the HJ also performs branch migration of one base pair. Then the bottom stands are cleaved, exchanged and ligated giving excisive recombinant product. (Reproduced from Radman-Livaja, M., et al., 2006)

IHF and other lambdoid phage recombination systems

In addition to λ integrase system, IHF also functions as confactor in some other members of Int family. For example, coliphage HK022 has a similar mechanism of site-specific recombination as phage λ in its lysogenic cycle. Both integration and excision are catalyzed by the phage-encoded Int and both reactions require the presence of the host encoded accessory protein IHF. In addition, excision requires the phage-encoded accessory protein Xis whose activity can be partially replaced by the host-encoded Fis (Kolot et al., 2003). The phage attachment site of HK022 is nearly identical to that of λ integrase (Yagil et al., 1989). In bacteriophage P2 integrase system, *att*P has two Int binding sites on each arm, one IHF binding site located between the left arm binding sites. Cox binding sites are on the right arms between core binding site and arm site (Frumerie et al., 2005). IHF is required for both integrative and excisive recombination (Yu and Haggard-Ljungquist, 1993). The recombination pathway of bacteriophage P22 is similar to λ in its use of a virally encoded Xis protein and a cellularly encoded IHF protein (Smith-Mungo et al., 1994). Different from λ integrase system, IHF has two binding sites flanking the common core on *att*P of P22 and one site located on *att*B (Leong et al., 1985). IHF was also found to stimulate the site-specific recombination reaction between the *att*P site of bacteriophage HP1 and the *att*B site of its host, *Haemophilus influenzae*, *in vitro* and also appear to regulate the expression of HP1 integrase (Goodman and Scocca, 1989). On *att*P of HP1, three IHF binding sites have been found. In *att*B region, two IHF binding sites were identified by using footprinting dialysis, but without integration-related function (Hwang and Scocca, 1990).

Interestingly, unlike integration carried out by phage λ Int, HP1 integrative recombination does not require the presence of IHF, whereas IHF is absolutely required for excision (Esposito et al., 2001).

A.3.2. IHF and DNA replication

E. coli chromosome replication

The DNA replication in bacteria is precisely regulated to ensure synchronous initiation of DNA synthesis from all copies of the chromosomal origin of replication *oriC* (Christensen et al., 1999). A critical first step is unwinding of *oriC* DNA, the chromosomal replication origin, by multiprotein orisome complexes comprising the AAA+ initiator DnaA and modulator proteins that bend DNA such as Fis and IHF (Leonard and Grimwade, 2005). On *oriC*, there are eight DnaA binding sites, which show different binding affinity, three high to medium affinity sites (R1, R4 and R2), and five lower affinity sites (R5M, I2, I3, I1 and R3) (Grimwade et al., 2000). *In vitro* the binding order of *oriC* to these sites is R4≥ R1>R2>R5 (M), I2, I3>I1>R3 (Ryan et al., 2002). In living cells, DnaA interaction with *oriC* is switch-like, with high affinity sites R4, R1 and R2 filled throughout most of the cell cycle and weaker sites R5(M), I1, I2, I3 and R3 only filled immediately before the onset of DNA synthesis (Ryan et al., 2004). IHF binds *E. coli oriC* selectively at the start of chromosome replication *in vivo* and can redistribute DnaA from high affinity (R1 and R4) to lower affinity sites such as R5M, I sites and R3 (Cassler et al., 1995; Grimwade et al., 2000). This IHF-induced redistribution of DnaA causes *oriC* unwinding at DnaA concentrations

below those needed in the absence of IHF. Because ATP-DnaA binding at I2 and I3 is required for origin unwinding, redistribution of DnaA to these sites stimulated by IHF may be the mechanism of IHF's ability to simulate the *oriC* unwinding (Ryan et al., 2004). Contrary to IHF, Fis inhibits the DNA replication from *oriC*, and the extent of inhibition by Fis is modulated by the concentrations of DnaA protein and RNA polymerase. The more limiting the amounts of these proteins, the more severe the inhibition by Fis (Wold et al., 1996). During cell cycle, high levels of initiator DnaA and IHF bind to *oriC* at the time of initiation of DNA replication, while binding of Fis is reduced (Cassler et al., 1999).

Initiation of plasmid pSC101 replication

pSC101 is a small tetracycline resistance gene-bearing plasmid. As an attractive system for the study of replication, plasmid-encoded RepA, host-encoded DnaA and IHF are involved in the initiation of DNA synthesis. On the origin of this plasmid, there is one DnaA box (binding site dnaAS) and RepA-binding iterons separated by a naturally bent AT-rich segment (Fig. A.7). Embedded in the AT-rich region is a binding site for integration host factor (IHF). The interaction of IHF with this cognate site bends DNA further and the interaction is essential for plasmid replication (Datta et al., 1999). DNA-bending induced by IHF promotes the interaction between DnaA protein bound to dnaAs and RepA proteins that are bound to physically separated cognate sites including the weaker $dnaA_w$ sites in the iterons (Datta et al., 1999; Stenzel et al., 1991) (Fig.A.7). DnaA–RepA interactions, which are facilitated by IHF, are critical for unwinding of AT-rich region on

pSC101 *ori*. This initial unwinding probably generates the single-stranded region needed for the entry of the replicative helicase DnaB(Datta et al., 1999).

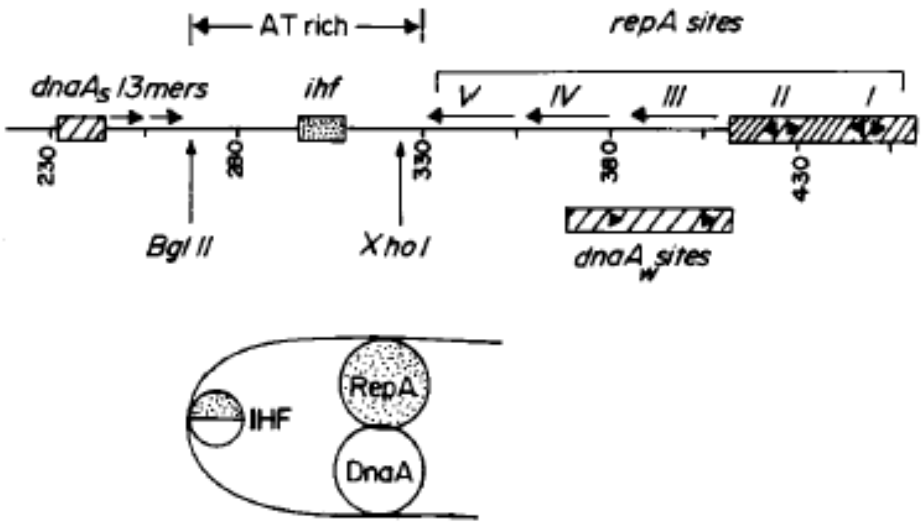


Fig.A.7. Replication origin of pSC101. oriC of pSC101 consists of one strong affinity binding site of DnaA named dnaAs, two 13 mers, one AT rich region with one IHF binding site embedded and RepA binding sites. The dnaAw sites can be bound by DnaA only when IHF bends the DNA. The bending introduced by IHF can facilitate the interaction between RepA and DnaA which have bound to their consensus sequences. (Reproduced from Datta, H. J., et al., 1999)

A.3.3. IHF and the controls of transcription in prokaryotic cells

Transcriptional regulation is a key step in bacterial gene regulation. Transcription is often controlled at the stage of initiation to avoid synthesis of unnecessary transcripts. Activation and repression of transcription is primarily caused by regulatory proteins which act by binding to specific sites on DNA. As a DNA architectural protein, IHF is one of the global regulatory proteins and influences the expression levels of many genes. The mechanism relies on IHF's capability to bend DNA and bring distant sites on the bacterial chromosome together. IHF has been shown to function as a DNA looping protein to facilitate interactions between

regulatory proteins bound at upstream sites and RNA polymerase at core promoter sites (Engelhorn and Geiselmann, 1998). For example, IHF can bend DNA to align the activator with the closed sigma 54-RNA polymerase promoter complex to facilitate interactions that result in open complex formation, thus increasing the probability of transcriptional activation (Claverie-Martin and Magasanik, 1992). Experiments with the *Pseudomonas putida Pu* promoter have shown that IHF can also recruit RNA polymerase holoenzyme to promoter sequences (Muir and Gober, 2005). IHF has been found to stimulate three sigma 70 promoters directly: the phage Mu early Pe promoter, the bacterial ilvP_G promoter of the ilvGMEDA operon and the phage λ early P_L promoter (Giladi et al., 1998). In the Mu Pe promoter, IHF was found to eliminate the repression exerted by H-NS indirectly (van Ulsen et al., 1996), and to increase the initial binding of RNA polymerase to the promoter (KB) directly (van Ulsen et al., 1997). In the ilvPG promoter, IHF activates transcription from the nearby downstream promoter simply by bending the DNA helix in the absence of specific IHF-RNA polymerase or upstream DNA-RNA polymerase interactions (Pagel et al., 1992). In the phage λ early P_L promoter, IHF stimulates transcription by enhancing closed complex formation between RNA polymerase and the promoter (Giladi et al., 1995). The transcription of region 1 of the *E. coli* K5 capsule gene cluster is regulated by IHF because mutations in the *him*A and *him*D genes which encode the subunits of the IHF led to a five-fold reduction in the expression of KpsE at 37 ^{o}C (Simpson et al., 1996).

It has been demonstrated that IHF can inhibit the transition of supercoiling-induced

DNA duplex destabilized (SIDD) sites from a B-form to a partially denatured duplex structure (Sheridan et al., 1999). This results in the translocation of the superhelical energy (negative twist) normally absorbed by the SIDD site to another site in a superhelically constrained DNA domain (Arfin et al., 2000). The supercoiling-dependent, DNA structural transmission mechanism of this type is responsible for the IHF-mediated activation of transcription from the ilvPG promoter of *E. coli* (Sheridan et al., 1999).

In 2000, Arfin and coworkers found IHF affects the expression of one known global regulatory gene (*arcA*), several operon-specific regulatory proteins, and 11 genes encoding putative regulatory proteins. Thus they deduced IHF might indirectly affect the expression of many genes via its effect on the expression of regulatory genes.

A.3.4. IHF and DNA transposition

Transposons are linear double-stranded DNA fragments which "jump" from the bacterial chromosome onto a plasmid, or vice versa; or from one chromosomal location to another, a process named transposition. For this reason, they have been also named "jumping genes".

The ends of transposons which normally consists of inverted repeats (IRs) are important for providing the information of precise translocation within the genome (Weinert et al., 1984). γδ is a member of the Tn3 family of prokayrotic transposons.

It has two IHF binding sites at both ends, adjacent to each IR where it is bound by γδ transposase specifically. It seems likely that IHF plays a role in γδ transposition, either by facilitating binding of transposase or by altering the activity of the transposase-IR complex (Wiater and Grindley, 1988). There is one IHF binding site in the Mu enhancer, and the binding of IHF allows the transposition *in vitro* of Mu from plasmids containing lower levels of supercoiling (Surette et al., 1989).

Another transposon Tn10 comprises two IS10 modules flanking a tetracycline resistance gene. A strong IHF binding site is located near the outside end of Tn10 and IS10, and IHF can modulate Tn10/IS10 activity both positively and negatively, depending on the location of the transposon – in the chromosome or in a multicopy plasmid respectively (Signon and Kleckner, 1995). Another Tn3 family member Tn4651 is also modulated by IHF. There are two IHF binding sites located just near the terminal inverted repeat sequences that are presumable binding sites for the transposase (Teras et al., 2000). IHF may participate in Tn4652 transposition directly, either by modulating the binding of transposase to the ends of the transposon or by influencing the formation of nucleoprotein complexes needed in subsequent transposition reactions. And IHF might be used for conditional activation of Tn4652 in stationary phase bacteria when the concentration of IHF in bacteria increases (Ilves et al., 2004).

A.3.5. Other functions of IHF in bacteria

In addition to functions mentioned above, IHF is also one of the nucleoid

components abundant in the stationary phase of *E. coli* (Ali Azam et al., 1999). The genetic material in bacterial cells is organized in a structure called the nucleoid which consists of a single circular DNA molecule, RNA, and a large variety of bound proteins. IHF, a histone-like protein in *E. coli,* plays an important role to compact DNA by nonspecific low-affinity binding (Ali et al., 2001). In phage λ DNA packaging, IHF plays an additional accessory role by binding to the terminase assembly site forms a binary protein-DNA complex to stimulate terminase-mediated strand separation which is a necessary step before the DNA-enzyme complex is captured by an empty viral prohead to finish packaging (Yang and Catalano, 1997).

A.4. IHF, HU and HMG

There are functional equivalents of IHF – for example the prokaryotic HU and eukaryotic high mobility group (HMG) proteins (Travers, 1997). It has been shown that these proteins, which bind DNA with relaxed sequence specificity, can partially replace IHF during excisive recombination (Segall et al., 1994).

HU

HU is a small protein, only 18 kDa, and exists either as a heterodimer (in *E. coli* and *Salmonella typhimurium*) or a homodimer (in *E. coli* and all other bacteria). The amino acid sequences of the two subunits in heterodimer share 70% homology. HU is one of the most abundant of DNA binding proteins in growing *E. coli* (Azam and Ishihama, 1999) ; around 30,000 copies of per cell (Oberto et al., 1994). There

are 30% - 40% identity between IHF and HU at amino acid level (Swinger et al., 2003). Like IHF, HU can bend DNA. However, unlike IHF, which binds DNA in a sequence specific manner, HU lacks such strong sequence preferences and binds avidly to a variety of structural distortions regardless of sequences (Swinger and Rice, 2004). In the crystal structure of HU-DNA, although the length of DNA sequence contacted is much smaller than IHF, 14 -19 bp and the overall bend in these structures ranges from ~105 to 140°, less than the 160^{o} bending angle introduced by IHF, the overall structure bears many similarities to the IHF-DNA complex structure: the β-ribbon arms lie in the minor groove of the DNA and, at the tip of each arm, the conserved proline residue intercalates between base pairs, creating and/or stabilizing two kinks in the DNA and asymmetric charge neutralization mechanisms of DNA bending (Swinger et al., 2003). (Fig.A.7)

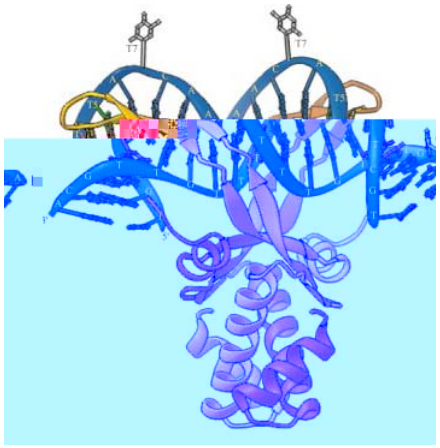


Fig.A.7. The crystal structure of HU-DNA. The homodimer of HU is in yellow. The conserved aa prolines are green. The DNA double helix is in blue. (Reproduced from Swinger, K. K., et al., 2004)

The prominent activity of HU is the wrapping of DNA into nucleosome-like particles. However, HU-DNA particles are less stable, having a dissociation half-life of 0.6 min in 50 mM NaCl. This may explain prior difficulties in detecting prokaryotic nucleosome-like structures (Broyles and Pettijohn, 1986). HU also

plays a role in recombinational DNA repair that is not limited to double strand break repair or daughter strand gap repair (Li and Waters, 1998). Alike IHF, HU can stimulate initiator DnaA-catalyzed unwinding of the chromosomal replication origin, *oriC*, by suppressing binding specifically at I3 (Ryan et al., 2002). At the same time, HU can modulate the binding of IHF to its specific *oriC* site. Depending on the relative concentrations of HU and IHF, HU is able either to activate or to inhibit the binding of IHF to *oriC* (Bonnefoy and Rouviere-Yaniv, 1992). In the transcription control, it has been suggested that HU could introduce structural changes to the DNA which would facilitate or inhibit the binding of regulatory proteins to their specific sites (Preobrajenskaya et al., 1994). During Mu transposition, the formation of precleaved complex in which the Mu ends are held together in a non-covalent protein-DNA complex requires MuA and *E. coli* HU proteins and a high degree of donor DNA supercoiling (Surette et al., 1989). HU is also found active in Tn10 tansposition (Morisato and Kleckner, 1987) and gene inversion (Johnson et al., 1986). In λ site specific recombination, HU can replace IHF by cooperating with integrase to generate a stable and specific complex with electrophoretic mobility and biochemical activity very close to the intersome formed by IHF and integrase (Segall et al., 1994). However, it can't allow integration with wild-type DNA substrates and, to some extent, supports only excisive recombination (Goodman et al., 1992).

High Mobility Group (HMG) architectural proteins

The high mobility group (HMG) chromosomal proteins were named after their

electrophoretic mobility behavior in polyacrylamide gels. The HMG family of proteins comprises members with multiple HMG domains that bind DNA with low sequence specificity, and members with single HMG domains that recognize specific nucleotide sequences (Grosschedl et al., 1994). Based on the characteristic functional sequence motif as well as their substrate binding specificities, HMG nuclear proteins were classified into HMGA, HMGB and HMGN (Bustin, 2001). The HMGA (formerly HMG-I/Y; Mw~10 kDa) family is so called because it preferentially binds to the minor groove of stretches of AT-rich DNA. Three highly conserved regions within each of the known HMGA proteins (HMGA1 and HMGA2), T-P-K-R-P-R-G-R-P-K-K named "AT-hooks", is closely related to the consensus AT-rich sequence (Reeves and Nissen, 1990). Members of the HMGB family (formerly HMG-1 and −2; Mw ~25 kDa) are characterized by the presence of two DNA-binding motifs called "HMG-boxes", A and B. Like the whole protein, which may perform an architectural role in chromatin, the individual boxes bind to minor grove of DNA without sequence specificity, have a preference for distorted or pre-bent DNA, and are able to bend DNA and constrain negative superhelical turns (Broadhurst et al., 1995; Hardman et al., 1995; Weir et al., 1993). Members of the HMGN family (formerly called HMG-14 and −17; Mw ~10–20 kDa) are distinguished by the presence of a highly conserved nucleosome-binding peptide motif allowing them to specifically recognize the genetic structure of the 147 bp NCP (nucleosome core particle) to build block of the chromatin fiber (Reeves and Adair, 2005).

The unifying functional theme of the HMG motifs is their ability to modify the structure of their target site and induce structural changes that facilitate the progression of a wide range of DNA-dependent activities (Bustin, 1999). In HMGA overexpressing cells, 18 genes known to be involved in DNA repair processes were significantly down-regulated. In addition to that, HMGA proteins also inhibit repair as a consequence of direct binding to DNA lesions *in vitro* (Reeves and Adair, 2005). HMGA1 proteins are architectural transcription factors that regulate the transcriptional activity of a large number of mammalian genes by inducing changes in chromatin structure and by controlling the formation of stereo-specific, multiprotein complexes called 'enhancesomes' on the promoter/ enhancer regions of genes (Treff et al., 2004). Overexpression or aberrant expression of the HMGA1 proteins is frequently associated with both neoplastic transformation of cells and metastatic tumor progression (Reeves et al., 2001). Aberrant expression of HMGA2 protein which is highly expressed during embryogenesis and completely repressed in normal adult tissues is observed in some cancer cells. Derepression of HMGA2 gene expression has been detected in retinoblastoma for its function on neoplastic transformation of retina cells (Chau et al., 2003). As scaffold of metaphase chromosomes, HMGA proteins are involved in the formation of the classic chromosomal Q type banding pattern (Saitoh and Laemmli, 1994). The HMGA proteins also function as general host factors with intermolecular integration-complementing activity necessary for the formation of preintegration complexes of human immunodeficiency virus type 1 and Moloney murine leukemia virus (Farnet and Bushman, 1997; Li et al., 1998).

HMGB binding to cisplatin adducts (Ohndorf et al., 1999) inhibits excision repair of 1,2-intrastrand d(GpG) cross-links (Zamble et al., 1996) (Fig.A.8). And HMGB1 has been demonstrated to interact with p53 directly *in vitro* and function as a unique activator of the protein that stimulates its binding to strand breaks and insertion/deletion mismatches and causing arrest in the cell cycle DNA (Boulikas, 1996; Reeves and Adair, 2005). HMGB is also known for the its effect on the binding of steroid receptors to target DNA to enhance transcriptional activity in mammalian cells (Boonyaratanakornkit et al., 1998). In addition, HMGB proteins stimulate overall transcription from chromatin templates by their negatively charged C terminus or by affecting the activity of other transcription factors (Aizawa et al., 1994). In the λ integrase site specific recombination system, the eukaryotic HMGB proteins were found to support excisive recombination by binding to DNA non-specifically and cooperating with integrase to generate a defined structure called intasome (Segall et al., 1994).

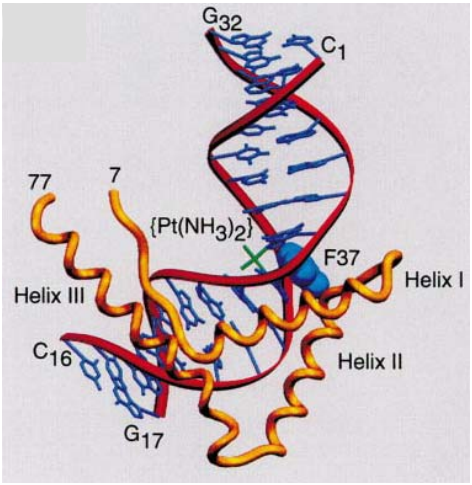


Fig.A.8. **Overall structure of the complex between the nonsequence-specific domain A of HMGB and cisplatin-modified DNA.** The protein backbone is shown in yellow, the intercalating Phe 37 residue as van der Waals spheres, and the DNA in red and blue with the cisplatin intrastrand adduct in green. Numbers indicate the first (N terminus) and last (C terminus) ordered residues in the crystal structure. (Reproduced from Ohndorf, U.M. et al., 1999)

In contrast to the repair inhibitory effects of HMGA and HMGB family members, HMGN proteins were demonstrated to enhance nucleotide excision repair in the context of chromatin (Reeves and Adair, 2005). In addition to that, HMGN proteins enhance transcription and replication by acting as architectural elements that destabilize or unfold the higher-order structure of the chromatin fiber (Bustin et al., 1995). This is mediated by the negatively charged C-terminal domain interacting with histone tails in nucleosomes to disrupt the higher order chromatin structure (Ding et al., 1997; Trieschmann et al., 1995).

A.5. Aim of this project

The mutant lambda integrase (Int h/218) has been introduced into mammalian cells where it promotes site-specific recombination on episomal and genomic substrates at a significant level in the absence of protein co-factors. However, the mutant Int supports not only the integration but also excision in the absence of any other cofactors (Lorbach et al., 2000). Another observation was that wild-type Int-mediated recombination is stimulated by transduction of purified IHF into mammalian cells. This result revealed an interesting possibility, i.e. regaining control over the directionality of recombination in mammalian cells through an employment of protein co-factors (Christ et al., 2002). However, it is known that separate expression of IHF α- and β-subunits in *E. coli* results in unstable polypeptides and insoluble aggregates (Nash et al., 1987; Wang and Chong, 2003). To overcome this problem, a single chain integration host factor (scIHF2) with similar biochemical properties and biological functions as wild type IHF was

constructed. This will circumvent disproportionate expression of the two subunits and may result in a more compatible enzyme for a eukaryotic environment. In addition to the role in recombination, the recombinant protein scIHF2 may also be useful in studies investigating the structure and function of eukaryotic nucleoprotein complexes inside a living cell. And scIHF2 may be employed in biopharmaceutical production techniques by its DNA-bending capacity to keep a promoter free of nucleosomes and, thus, more accessible for the transcriptional machinery. This could boost and/or maintain a desired high level of gene expression.

With wild type DNA substrates, HU or HMG1/2 can not replace IHF to stimulate integrative recombination but both factors support excisive recombination, at least partially (Segall et al., 1994). How critical is the degree of DNA bending in λ site-specific integrative recombination? In order to answer this question, we constructed two mutants based on scIHF2, named scIHF2E and scIHF2ER, which have different modes of protein-DNA interaction than scIHF2. The biochemical and functional analysis as well as structural comparisons may provide us with a possible explanation. The observed divalent metal ion-dependent conformational switch in scIHF2E-DNA complexes could provide a new method to study structure-function relationships in other nucleoprotein assemblies by substituting basic amino residues that are involved in DNA contacts with acidic residues in DNA-binding proteins.

B. Materials and Methods

B.1. Biochemical and functional properties of scIHF2 and its variants

B.1.1. Materials:

B.1.1.1. Plasmids

pET17b	New England Biolabs
pETscIHF1	From Dr. Nicole Christ (Bao et al., 2004)
pETscIHF2	From Dr. Nicole Christ (Corona et al., 2003)
pETscIHF3	This study
pETscIHF3E	From Dr. Nicole Christ (Bao et al., 2004)
pETscIHF4	From Dr. Nicole Christ (Bao et al., 2004)
pTrc99a	From Pharmacia
pTrcscIHF1	From Dr. Nicole Christ (Bao et al., 2004)
pTrcscIHF2	From Dr. Nicole Christ (Corona et al., 2003)
pTrcscIHF3	From Dr. Nicole Christ (Bao et al., 2004)
pTrcscIHF3E	This study
pTrcscIHF4	From Dr. Nicole Christ (Bao et al., 2004)
pλIR	From (Christ et al., 2002)
pλER	From (Christ et al., 2002)
SG86	From Prof. S.D.Goodman (Hashimoto-Gotoh et al., 1981)

B.1.1.2. Bacteria

BMH8117	F^-, Δ(*lac-proAB*) *thi*, *gyrA* (Nal^R), *supE*, λ^- (Kumar et al., 1999)
DH5α	*supE*44 Δ*lacU*169 (φ80 *lacZ* ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1(Hanahan, 1983)
BL21-DE3	F^- *ompT hsdS*($r_B^- m_B^-$)*gal dcm λ(DE3)* (from Strategene)
CSH26-ΔIHF	F^- *ara*Δ *(lac pro) thi himA* Δ*82::Tn10*(Tc^R) *himD* Δ*3::*Cm^R (B. Rak, Freiburg)
CSH 26	F^- *ara*Δ *(lac pro) thi* (Miller, 1972)

B.1.1.3. Media for bacteria culture

LB media	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.4
LB agar plates	LB media plus 1.5% (w/v) bacto-agar.
LB agar plates	LB agar plates with 300 μg/ml ampicillin
LB agar plates	LB agar plates with 300 μg/ml ampicillin, 34 μg/ml Chloramphenicol, 15 μg/ml tetracycline

* Ampicillin was prepared at 100 mg/ml in ddH_2O and sterilized by filtration through 0.22 μm filter.

Chloramphenicol and Tetracycline were prepared at 34 mg/ml and 15 mg/ml in ethanol, respectively.

B.1.1.4. SDS-PAGE and EMSAs

Stock solution for the gel: 30% Acrylamide/Bis solution (29:1); 1.5 M Tris (pH

8.8); 10% SDS; 10% AP; TEMED; 1 M Tris (pH 6.8); 5xTBE (0.45M Tris-boric acid, 0.01 M EDTA pH 8.0)

SDS-PAGE

15% resolving gel	15% (v/v) acrylamide/Bis solution (29:1), 25% (v/v) 1.5 M Tris (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) AP 0.04% TEMED
Stacking gel	5% (v/v) acrylamide/Bis solution (29:1), 12.5% (v/v) 1.0 M Tris (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) AP 0.1% TEMED
2xSDS gel loading buffer	100 mM Tris (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM dithiothreitol (DTT)
5xTris-glycine electrophoresis buffer	25 mM Tris, 250 mM glycine, 0.1% (w/v) SDS, pH 8.3
Staining reagent	2.5% (w/v) coomassie brilliant blue, 45% (v/v) methanol, 10% (v/v) glacial acetic acid
Destaining reagent	45% methanol, 10% glacial acetic acid

Native gel

15% polyacrylamide gel	15% (v/v) acrylamide/Bis solution (29:1), 0.5xTBE 0.7% (w/v) AP; 0.036% TEMED

7.5% polyacrylamide gel	7.5% (v/v) acrylamide/Bis solution (29:1), 0.5xTBE 0.7% (w/v) AP; 0.036% TEMED

B.1.1.5. Enzymes

Restriction enzymes	From NEB
T4 DNA ligase	From Roche
Alkaline phosphatase, Calf intestinal (CIP)	From NEB
Klenow (exo$^-$ 5'-3')	From NEB

B.1.1.6. Kits

PCR purification kit	From Qiagen
Plasmid mini-prep kit	From Qiagen
Gel extraction kit	From Qiagen
DyeEX^{TM} 2.0 spin kit	From Qiagen
Lumi-light^{plus} western blotting kit	From Roche

B.1.2. Methods

B.1.2.1. Competent cells of DH5α, BL21-DE3 and CSH26-ΔIHF

DH5α

DH5α was plated on LB agar and incubated at 37 ^{o}C for 12-14 hrs. Single cloney isolates were inoculated into 50 ml LB media. Cultures were incubated at 37 ^{o}C, 250-300 rpm until the OD600 reached 0.4-0.6. Bacteria were harvested by centrifuging at 4 ^{o}C, 4000xg for 5 min. The pellet was resuspended in 30 ml 0.1 M cold $MgCl_2$. The suspension was again centrifugated at 4 ^{o}C, 4000xg for 5 min,

supernatant was discarded and the bacteria pellet was resuspended in 20 ml 0.1 M cold $CaCl_2$. The cells were kept on ice for half an hour, followed by centrifuging at 4 ^{o}C, 4000xg for 5 min. Finally the pellet was resuspended in 2 ml cold 0.1 M $CaCl_2$ -15% glycerol and aliquoted into 100 µl/tube and stored at -80 ^{o}C.

BL21-DE3

LB agar and LB media used for BL21-DE3 competent cells contain 34 µg/ml Chloramphenicol (CAM), 15 µg/ml tetracycline (TET). The preparation of BL21-DE3 competent cells is the same as above.

CSH26-ΔIHF

LB agar and LB media for CSH26-ΔIHF contained 10 µg/ml CAM and 5 µg/ml TET. The preparation of CSH26-ΔIHF competent cells is the same as above.

B.1.2.2 Molecular modeling

(performed by Prof. Dr. Thomas Schwartz, MIT, USA)

Based on the atomic coordinates of IHF (PDB accession number 1IHF), the β-chain was first opened between residues Q39 and G40. Five residues (GGSGG) were positioned manually between the N-terminus of the α-chain at L3 and β-Q39 (α-A2 was deleted). Similarly, two glycine residues were manually modeled in order to connect β-G40 and α-A94, omitting the terminal three residues of the IHF crystal structure. Modifications involved only protein regions that appear to be flexible in the crystal structure, judged by comparatively high temperature factors.

Further, the modified regions are not part of secondary structure elements. The manually modeled scIHF2 coordinates were then regularized using a simulated annealing regiment performed with the program CNS (Brünger et al., 1998). An ensemble of 10 energetically minimized structures obtained from this procedure superimpose very well. No significant structural changes were detectable in the part of IHF that was not modified.

B.1.2.3. Construction of plasmids

pETscIHF3

pETscIHF3E was first digested with *Eco*RI and *Hin*dIII at 37 ^{o}C for 2 hrs. The 3387bp fragment was purified through gel extraction and used as backbone. The targeting fragment from pETscIHF2 was also digested with *Eco*RI and *Hin*dIII (Fig.B.1.3). The stoichiometry of targeting fragment to backbone was 3:1 in 20 µl ligation reactions with 1 µl T4 DNA ligase (Roche). The reactions were kept at 16 ^{o}C overnight and then mixed with 100 µl BMH8117 competent cells. The mixture was kept at 0 ^{o}C for 30 min, followed by heat shock at 42 ^{o}C for 2 min, then incubated at 0 ^{o}C for 5 min; cells were recovered by adding 880 µl LB media and shaking at 37 ^{o}C, 230 rpm for 40 min. After incubation, bacteria cells were centrifugated at 2000xg for 10 min. Supernatant was discarded and cell pellet was resuspended with 100 µl fresh LB media and plated onto LB agar with 300 µg/ml ampicilin (AMP). LB agar plate was incubated at 37 ^{o}C overnight. Isolated colonies were inoculated into LB media with 300 µg/ml Amp. Cultures were incubated at 37 ^{o}C, shaking 250-300 rpm overnight. The plasmid was purified by

mini-prep kit (Qiagen) and analyzed by digestion and sequencing.

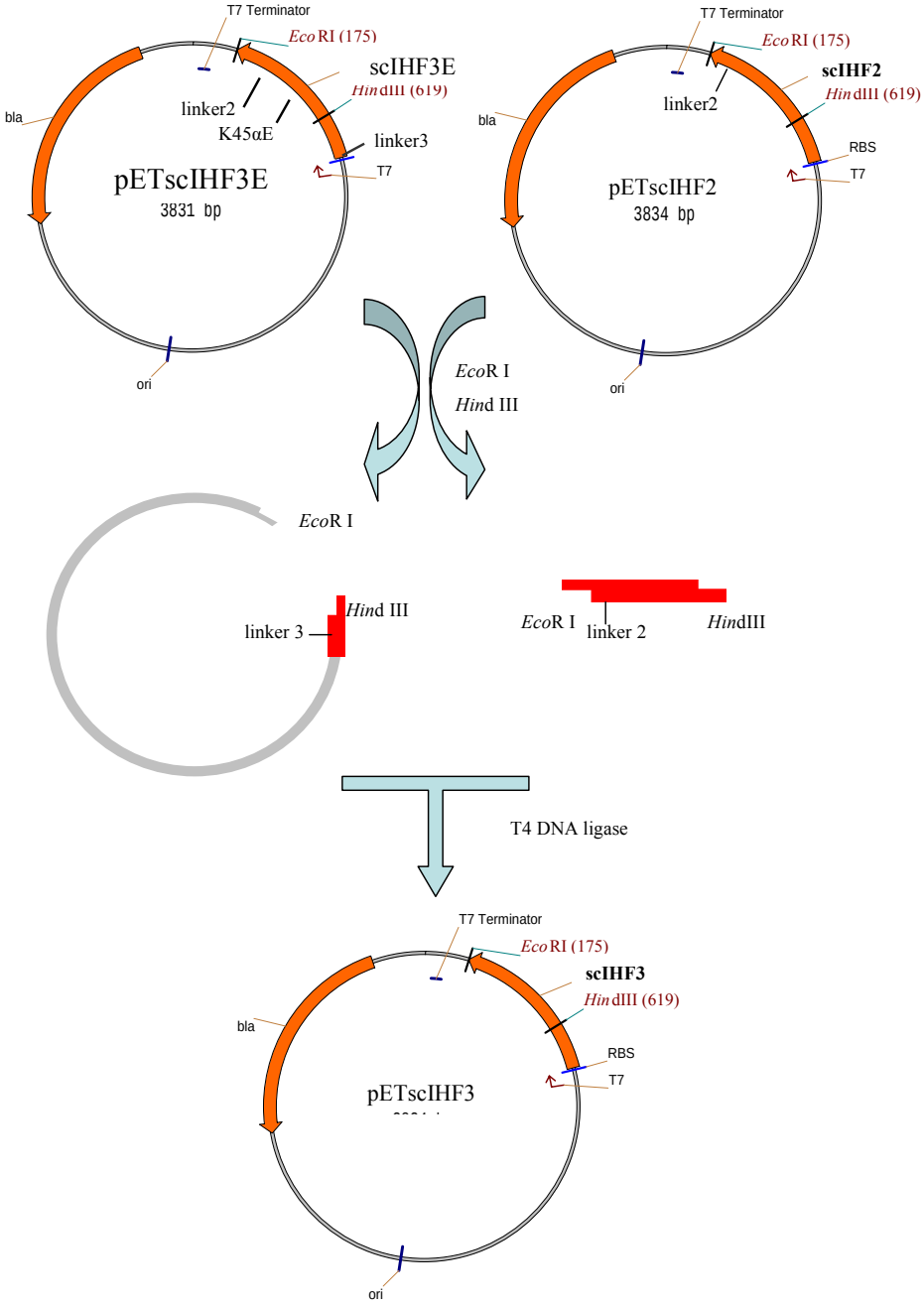


Fig.B.1.1. Construction of pETscIHF3. The targeting fragment was from pETscIHF2 digested with *Eco*RI and *Hin*dIII, because scIHF2 and scIHF3 have the same linker 2. This fragment was ligated with backbone from pETscIHF3E digested with *Eco*RI and *Hin*dIII.

pTrcscIHF3E

*Nco*I site was introduced into pETscIHF3E by the replacement of *Nde*I digested fragment with the DNA fragment from pTrcscIHF3 digested with *Nde*I. The targeting fragment from new plasmid pETscIHF3E(+) was digested with *Asi*SI and *Nco*I and inserted into the backbone of pTrcscIHF3. The plasmid was verified by DNA sequencing.

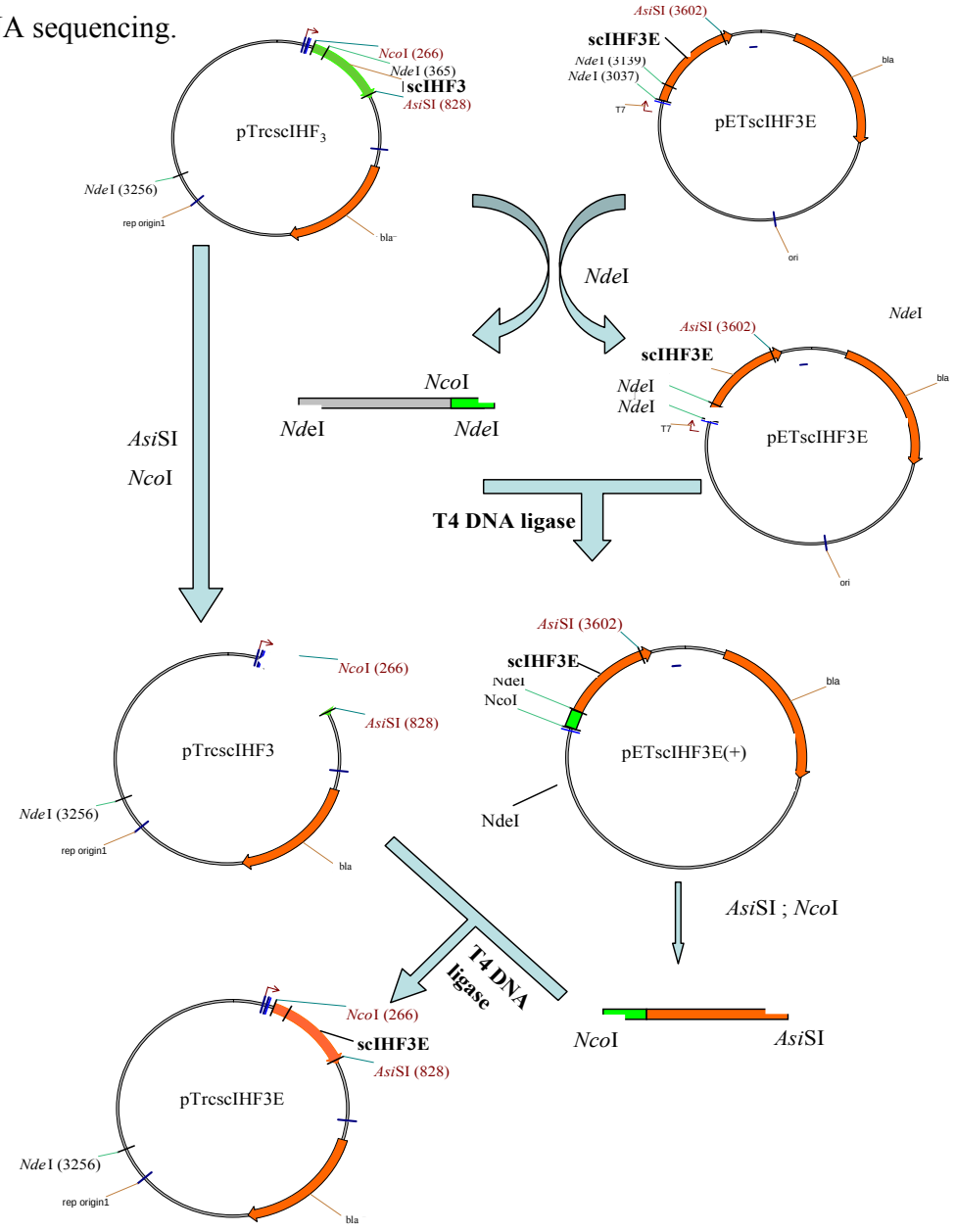
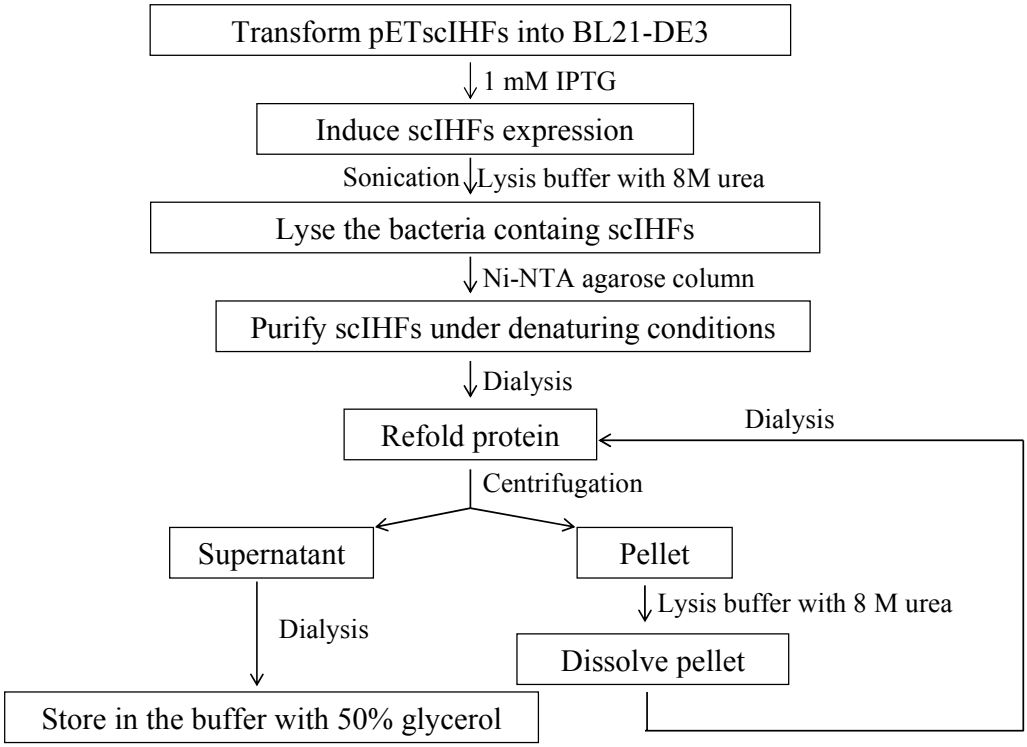


Fig.B.1.2. Construction of pTrcscIHF3E.

B.1.2.4. Purification of scIHFs (eg. scIHF2)

Outline of scIHF purification



B.1.2.4.1. Transformation and induction of protein expression

100 ng pETscIHF2 was added to 100 µl BL21-DE3 competent cells and incubated at 0 ^{o}C for 30 min, followed by heat shock at 42 ^{o}C for 2 min, then 0 ^{o}C for 5 min. Cells were recovered by adding 900 µl LB media. The bacteria cells were incubated with shaking at 37 ^{o}C, 230 rpm for 40 min. 100 µl bacteria were plated on the LB agar with 300 µg/ml AMP, 34 µg/ml CAM and 15 µg/ml TET. The plate was incubated at 37 ^{o}C overnight. A single colony transformant was inoculated into 50 ml LB media (300 µg/ml AMP, 34 µg/ml CAM, 15 µg/ml TET) and cultured at 37 ^{o}C overnight. Subsequently, 20 ml of culture were inoculated into 1L LB media

(300 µg/ml AMP, 34 µg/ml CAM, 15 µg/ml TET) and cultured at 37 ^{o}C until OD600 reached 0.4-0.5. The expression of scIHF2 was induced by adding IPTG to a final concentration of 1 mM. The bacteria were cultured for additional 4hrs and then were harvested by centrifuging at 5500 rpm for 15 min. The pellet was stored at -80 ^{o}C. (1 ml cells were harvested by centrifugation at 16100xg for 1 min, then the pellet was resuspended with 100 µl 1xSDS loading buffer and boiled at 100 ^{o}C for 5 min for checking protein in 15% SDS-PAGE)

B.1.2.4.2. Purification of scIHF2 under denaturing conditions

The cell pellet was frozen and thawed 2-3 times, then resuspended in 20 ml lysis buffer B (100 mM NaH_2PO_4; 10 mM Tris; 8 M urea; pH 8.0), followed by sonication at amplitude 60, pulse 3 s x 20 cycles for 3 times; the extract was cleared by centrifuging at 15000xg for 15 min. The supernatant (lysis I) was decanted into 50 ml falcon tube and 10 µl was aliquoted for protein determination. 5 ml of lysis buffer B was added to resuspend the pellet. After centrifugation at 15000xg for 15 min, the supernatant (lysis II) was also aspirated into the same tube. 10 µl of lysis II was aliquoted for protein determination.

5 ml of Ni-NTA agarose suspension was added into the lysis mixture and mixed thoroughly by rolling at room temperature overnight. The protein-resin complex was packed into a column carefully with a long glass pipette. Then the flow-through was collected, and the column was washed twice with 10 ml lysis buffer B, and five times with 5 ml of washing buffer C (as B, pH 6.3). scIHF2 was finally

eluted with different elution buffers: elution buffer I (as B, pH 5.9), elution buffer II (as B, pH 4.5) and elution buffer III (as B, pH 2.4). Each elution buffer was applied to the column five times, 5 ml each time. 10 µl of aliquots from each elute were used for protein checking. (Fig.B.1.5)

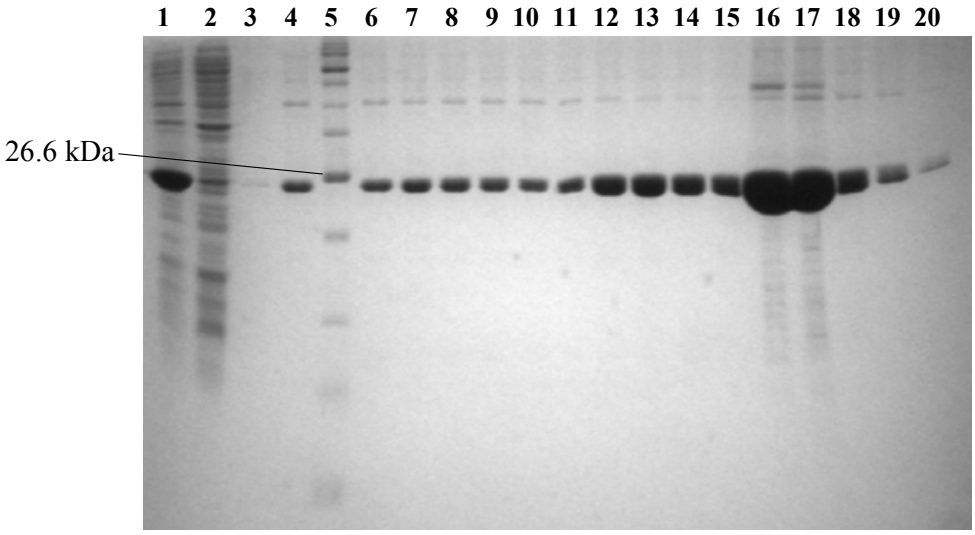


Fig.B.1.3. Purification of scIHF2 under denaturing condition. The supernatant of proteins lysis (lane 1) was incubated with Ni-NTA agarose. The column of protein-resin complex was washed with buffer B containing 100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 8.0 (lane 3) and buffer C (same as B, but pH 6.3) (lane 4). The protein was collected when the column was eluted with different buffer: pH 5.9 (lane6-10), pH 4.5 (lane 11-15), pH 2.4 (lane 16-20). Note protein marker was loaded in lane 5.

B.1.2.4.3. Refolding of scIHF2

Preparation of dialysis tubing

Dialysis tubing (with 10 KD cut off) of desired size (30 cm) was cut and boiled twice in 0.1 M NaH_2CO_3 containing 10 mM EDTA for 20 min/time to remove heavy metal ions. After rinsing with ddH_2O, the tubing was stored at 4 ^oC in the buffer with 10 mM EDTA.

Dialysis

scIHF2 that is eluted with elution buffer II (pH 4.5) was dialyzed against buffer containing 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 50 mM NaCl at 4 ^{o}C overnight. After centrifugation at 4 ^{o}C, 5445 x g for 15 min, the supernatant was aspirated into a /cold 50 ml falcon tube. 20 ml lysis buffer B was added to dissolve the pellet. After incubation at room temperature for 5-6 hrs, the re-dissolved scIHF2 was dialyzed against the same dialysis buffer, followed by centrifuging at 4 ^{o}C, 5445xg for 15 min. The supernatant was combined with the previous one and dialyzed against the same buffer containing 50% glycerol. Purified scIHF2 was stored at -20 ^{o}C.

B.1.2.4.4. Second round of purification

An additional round of purification under native conditions was initiated by dialysis of the stored protein against 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, and pH 8.0. After incubation with Ni-NTA agarose overnight at 4 ^{o}C, scIHF2 was eluted in a buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0. Through dialysis, scIHF2 was kept in the buffer containing 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 50 mM NaCl, and 50% glycerol at -20 ^{o}C. Although the second round of purification appeared not to improve the purity of the protein to a significant extent, all experiments with scIHF2 reported here were performed with this batch. Purified wild-type IHF was a kind gift of D. Esposito. Protein concentrations were determined using the BIO-RAD protein assay kit with IgG as a standard.

Except for scIHF2, all scIHFs were purified by the same procedure, but without the second round of purification.

B.1.2.5. DNA-binding and -bending assays

B.1.2.5.1. Preparation of DNA substrates

The annealing of H' DNA:

In 100 µl reactions, 250 pmol H'-1: 5'-GCCAAAAAAGCATTGCTTATCAATTT GTTGCACC-3' and 250 pmol H'-2: 5'-TTTGGTGCAACAAATTGATAAGCAA T GCTTTTTTGGC-3' were annealed in 50 mM Tris, 2 mM EDTA, 50 mM NaCl, pH 8.0 and incubation at 95 ^{o}C for 10 min. Reactions were cooled to room temperature.

H1, H2 and pSC101 origin were annealed in the same way as H'.

H1-1: 5'-CATATGCAGTCACTATGAATCAACTACTTAGATG-3';

H1-2: 5'-TTTCATCTAAGTAGTTGATTCATAGTGACTGCATATG-3'

H2-1: 5'-ACGTAAAATGATATAAATATCAATATATTAAATT-3';

H2-2: 5'-TTTAATTTAATATATTGATATTTATATCATTTT ACGT-3'

pSC101 origin-1: 5'-GTGTTTTTTTTGTTTATATTCAAGTGGTTATAAT-3'.

pSC101 origin-2: 5'-TTTATTATAACCACTTGAATATAAACAAAAAAAACA C-3'

B.1.2.5.2. Labeling of DNA substrates

1 µl of annealed H' (H1, H2 or pSC101 origin) was incubated with 1 µl [α-^{32}P]

dATP and 1 unit of Klenow (exo$^-$5'-3'; NEB) in 20 µl reactions at 37 ^oC for 20 min. After purification through DyeEX^{TM} 2.0 spin kit (Qiagen), the DNA substrate was diluted 30 times with ddH_2O.

B.1.2.5.3. Binding assays

The K_d of IHF and scIHFs were determined using ^{32}P-labeled 34-mer oligonucleotides comprising H', H1, H2 and pSC101 origin. Binding was performed under two different conditions. "Low salt" buffer (20 µl) contained 0.5xTBE, 17% glycerol, 5 mM EDTA, 200 µg/ml BSA, 0.1 nM labeled DNA, and 640 nM or less protein and "high salt" buffer (20 µl) contained 100 mM NaCl, 20 mM Tris–HCl, pH 7.9, 1 mM EDTA, 200 µg/ml BSA, 17% glycerol, 0.1 nM labeled DNA, and 1280 nM or less protein. Reactions were incubated at room temperature for 40 min and loaded onto 15% (w/v) polyacrylamide gels, which had been pre-run in 0.5xTBE at 4 ^oC and 16 Volts/cm for 1 hr. Electrophoretic mobility shift assays (EMSAs) were quantified using a Bio-Rad Molecular Imager System.

B.1.2.5.4. DNA-bending introduced by scIHFs

H1 as DNA substrate (scIHFs)

A ^{32}P-labeled 34 bp H1 fragment was used as DNA substrate for bending assays. Reactions were performed in 20 µl with 0.1 nM labeled H1 in "low salt" binding buffer without BSA and incubated at room temperature for 40 min; samples were loaded on a 15% polyacrylamide gel. Gels were run in 0.5xTBE at 4 ^oC and 16 volts/cm for 5 hrs. H', H2 and pSC101 origin were also used as DNA substrates.

*Att*L as DNA substrate

*Att*L is a locus containing a single H' site positioned close to its centre. It was acquired from pCMVssattL digested with

gel which was pre-run at 4 °C, 16 volts/cm for 1 hr.

B.1.2.5.6. Stability of IHF-H1 and scIHF2-H1 complexes

The stability of IHF-H1 and scIHF2-H1 complexes was analyzed with unlabeled H1 as specific competitor to compete with binding between IHF (scIHF2) and radiolabeled H1 over time. In 20 µl reactions, there were 0.1 nM radiolabeled H1, 1 nM IHF or 3 nM scIHF2, 80 nM unlabeled H1 in low salt buffer. Reactions were incubated at room temperature for 48 hrs, 32 hrs, 24 hrs, 8 hrs, 6 hrs, 3 hrs and 1 hr before loaded onto 15% polyacrylamide gel. The gels were run at 16 volts/cm, 4 °C for 2 hrs. EMSAs were quantified using a Bio-Rad Molecular Imager System.

B.1.2.5.7. RNA-binding assays

Total RNA from *E. coli* was used as substrate. In 20 µl reactions, there were 1.5 µg RNA, 2 µg or less of scIHF2 in 0.5xTBE, 17% glycerol, 200 ng/µl BSA, 60 mM NaCl and 5 mM EDTA. After incubation at room temperature for 40 min, samples were loaded on 5% polyacrylamide gel which was pre-run at 16 volts/cm, 4 °C for 1 hr. Gels were run under the same conditions for another 3 hrs. After staining in 0.5 µg/ml ethidium bromide solution, the gel was analyzed via Gel Doc system (Bio-Rad).

B.1.2.6. Recombination assays

DNA substrates

Plasmid pλIR contains *att*P and *att*B, and plasmid pλER contains *att*L and *att*R.

Both plasmids have been described (Christ et al., 2002).

Int and Xis

Purified Int and Xis proteins are gifts from A. Segall, San Diego, USA (Cassell and Segall, 2003). Int was diluted 35 fold in a buffer containing 50 mM Tris-HCl pH 8.0, 10% glycerol, 1 mM EDTA, 0.6 M KCl, 2 mg/ml BSA. 1 µl diluted Int was applied in recombinant reactions, while 3 µl of Xis was added in excisive recombination after 10 fold dilution in 50 mM Tris-HCl pH 7.6, 1 mM EDTA pH 8.0, 5% glycerol, 0.5 mg/ml BSA.

***In vitro* integrative and excisive recombination**

Site-specific recombination reactions were performed in 25 µl reactions with 44 nM Tris-HCl (pH 8.0), 60 nM KCl, 0.05 mg/ml BSA, 5 mM spermidine, 1.3 mM EDTA, 42 fmol *att* sites and a 25-, 500-, 75-, 500-, 1000- or 200-fold molar excess of IHF, scIHF1, scIHF2, scIHF3, scIHF3E, or scIHF4, respectively. After incubation at room temperature for 40 min, reactions were stopped by 300 mM KAc, 2.5 volumes absolute ethanol. After incubation at -20 ^{o}C for 1 hr, DNA was precipitated by centrifugation at 16100xg for 15 min. DNA pellet was air-dried at 37 ^{o}C for 10 min and digested by *Nco*I in 20 µl reactions. Samples were loaded onto 0.8% agarose gel after digestion at 37 ^{o}C for 1 hr. The gels were run in 0.5xTBE buffer, at 60 V for 14 hrs. DNA fragments were visualized with ethidium bromide staining.

B.1.2.7. Replication assays

B.1.2.7.1. Replication assay *in vivo*

Plasmid SG86 contains a functional pSC101 origin of replication (Hashimoto-Gotoh et al., 1981). Accordingly, replication and maintenance of SG86 requires IHF. SG86 and pTrcscIHFs were co-transformed into CSH26-ΔIHF cells, then cells were plated on LB-agar containing 30 or 40 μg/ml TET, 50 or 60 μg/ml CAM, 50 μg/ml AMP, and 0.5 mM IPTG and cultured at 37 ^{o}C overnight. Transformants were expanded in liquid culture overnight under the same conditions and plasmids were isolated using affinity chromatography (Qiagen)

B.1.2.7.2. The expression of scIHFs in CSH26-ΔIHF

CSH26-ΔIHF cells transformed with pTrc99a or pTrcscIHFs were cultured at 37 ^{o}C until OD600 reached 1.0. Cells were sonicated and boiled in SDS gel loading buffer for 5 min. Equal amounts of proteins obtained were separated in a denature 12% (w/v) polyacrylamide gel and transferred onto a PVDF membrane (Roche Diagnostics), with 1ng purified scIHF2 as positive control. The expression of the proteins was analyzed by western blot. The membrane was blocked with 5% blocking solution (skim milk) and incubated with rabbit polyclonal antibodies raised against wild-type IHF. Bound antibody was detected with anti-rabbit peroxidase-conjugated secondary antibody (Amersham Biosciences).

B.2. Further elucidation of biological properties

B.2.1. Materials

B.2.1.1. Plasmids

pETscIHF2E	This study
pETscIHF2ER	This study
pETscIHF2mut	This study
pAttPpuro-CMVRed	From Dr. Klaus Neef
pλIR	From (Christ et al., 2002)
pλER	From (Christ et al., 2002)
pWSRGFP	From (Christ et al., 2002)
pCMV*att*L	From (Corona et al., 2003)

B.2.1.2. Stock solution for crystallization

1 M Tris.HCl (pH 7.5); 1 M NaCl; 1 M $MnCl_2$; 50% (v/v) PEG5K-MME (polyethylene glycol 5000 monomethy ether); 100% glycerol. Except for glycerol, all the buffers were sterilized through filtration with 0.22 μM filter.

B.2.1.3. Devices for concentrating protein

Amicon Ultra-4 Centrifugal Filter Devices (Millipore)

B.2.2. Methods

B.2.2.1. Construction of plasmids

pETscIHF2E

K45αE fragment was cut out from pETscIHF3E by the digestion of *Eco*RI and *Hin*dIII. After purification, it was iligated into vector derived from pETscIHF2 digested with the same enzymes. (Fig.B.2.1)

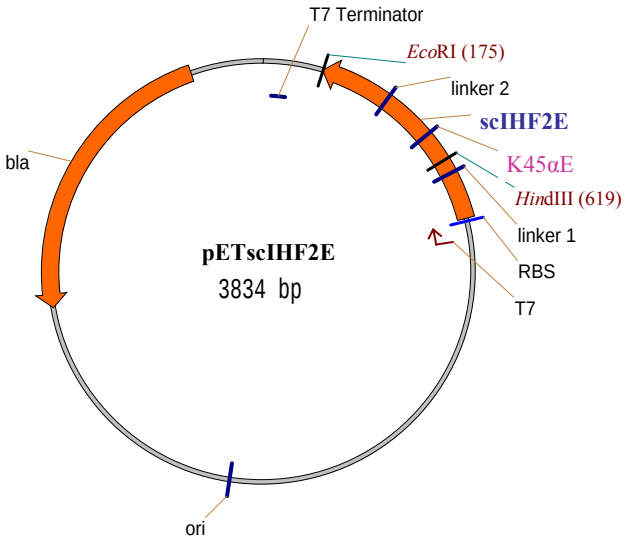


Fig.B.2.1. Map of pETscIHF2E.

pETscIHF2ER

With pETscIHF2E as the template, ER-F1: 5'- GTTTGATAAGCTTGGG CTTA G-3', and ER-R1: 5' -CCAAAACCACGGAGTTCCACCTGTTCGC-3' or ER-F2: 5'-GTGGAACTCCGTGGTTTTGGTAACTTCGAT-3' and ER-R2: 5'-GCAGAA TTCTTATCAGTGG -3' as primers, ER1 and ER2 were amplified by PCR. The fragment consisting K45αE-S47αR was obtained by PCR using ER1 and ER2 as the template, ER-F1 and ER-R2 as primers. After digestion with restriction

enzymes *Eco*RI and *Hin*dIII, this fragment was ligated to the backbone from pETscIHF2 digested with *Eco*RI and *Hin*dIII. (Fig.B.2.2).

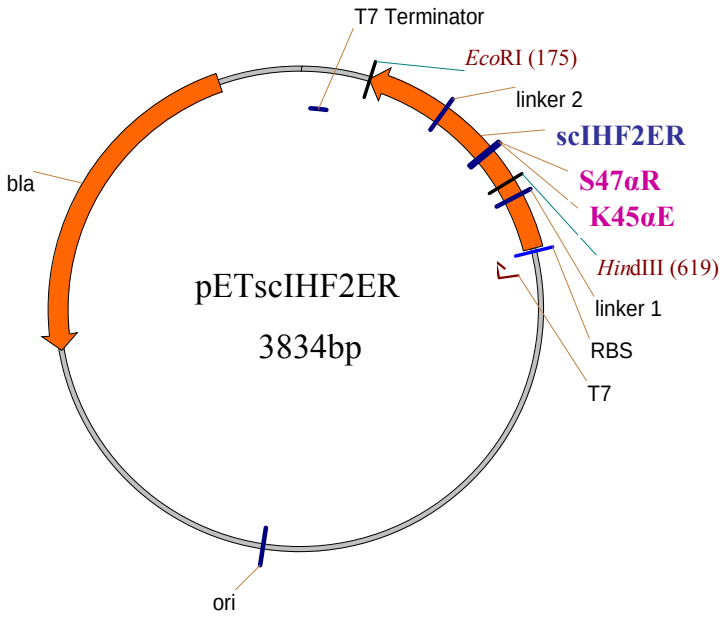


Fig.B.2.2. Map of pETscIHF2ER.

pETscIHF2mut

Fragments M1 and M2 was amplified by PCR with pETscIHF2 as the template, M1-S, M1-AS and M2-S, M2-AS as primers. The DNA polymerase used here was *Pfu*. The target fragment was amplified by assemble PCR with M1 and M2 as templates, M1-S and M2-AS as primers. The primers' sequences are as follows: M1-S: 5'-GGGACGTAACGGGAAAACGGGCGAGG-3'; M1-AS: 5'-TCCCAT TACGTCCGGTACGTGG-3'; M2-S: 5'-CCGGACGTAATGGGAAGACTGGC GATAAA-3'; M2-AS: 5'-TGCAGAATTCTTATCAGTGGTG-3'. The PCR product was digested with *Eco*RI and purified using PCR purification kit (Qiagen). Then it was inserted into a purified vector, pETscIHF2 which had been previously digested with *Eco*RI and *Sma*I. (Fig.B.2.3)

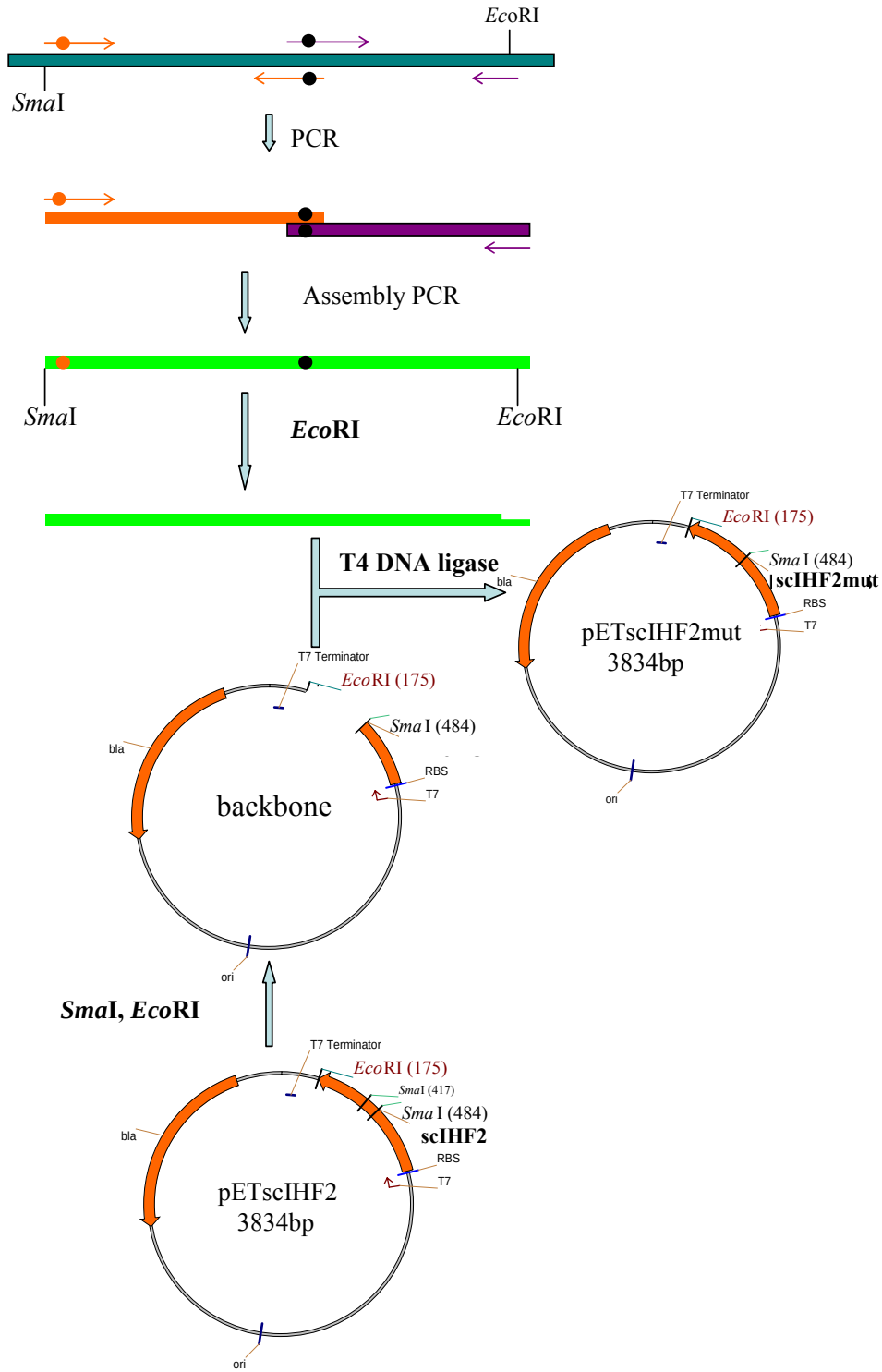


Fig.B.2.3. Construction of pETscIHF2mut. On plasmid pETscIHF2, scIHF2 fragment with *Eco*RI and *Sma*I sites at both ends was replaced by assembly PCR product with pETscIHF2 as template and specific primers for introduction of mutations P65αG-P64βG.

B.2.2.2. Purification of scIHF2E, scIHF2ER and scIHF2mut.

The expression plasmids were transformed into BL21-DE3 and cultured in LB media containing 300 μg/ml AMP, 34 μg/ml CAM and 15 μg/ml TET; 1 mM IPTG was added to induce the expression of protein when the OD600 reached 0.4-0.6. The bacteria were harvested after 4 hrs of induction. The proteins were purified according to the same protocol described in **B.1.** Protein concentrations were determined using the BIO-RAD protein assay kit using IgG as a standard.

B.2.2.3. Biochemical properties of scIHF2E, scIHF2ER and scIHF2mut

Radiolabeled 34-mer oligonucleotides H', H1 or H2 were used to measure the binding affinity of three scIHF2 mutants. In 20 μl reactions, 0.1 nM DNA and 160 nM or less scIHF2E or 80 nM or less scIHF2ER were included in low salt buffer. In high salt buffer, the proteins were increased to 1280 nM or less for scIHF2E, 640 nM or less for scIHF2ER. After incubation at room temperature for 40 min, the protein–DNA complexes were analyzed by EMSA.

The binding affinity of scIHF2mut to IHF consensus sequences was analyzed using H1 as DNA substrate in the low salt buffer with BSA. The molar ratio of scIHF2mut to H1 was 8000 to 1 or less.

For the bending assay, three different DNA substrates H', H1 and H2 were used to react with protein under the same conditions as the binding assay but without BSA

in the binding buffer.

Plasmid pAttPpuro-CMVRed was also used as a DNA substrate in bending assay. In 20 µl reactions, there were 272 fmol supercoiled DNA substrate and scIHF2, scIHF2E, scIHF2ER or scIHF2mut which had 90-fold, 400-fold, 400-fold or 1000 fold molar excess over DNA respectively , 2 µl 10x buffer 3(100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT) and 0.2 µl 100x BSA (NEB). The reaction was incubated at room temperature for 30 min, followed by addition of 1 µl *Pst*I and 1 µl *Eco*RI were added to digest the DNA at 37 ^{o}C for 10 min. The samples were mixed with DNA loading buffer (0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol FF and 5% (v/v) glycerol) and loaded onto a 4% polyacrylamide gel which were prerun at 150 V, room temperature for 1 hr. After 3 hrs, the gel was unfixed and stained in 0.5 µg/ml ethidium bromide.

Plasmid pWSRGFP containing *att*R processing two IHF binding sites spreading on it was also used in this assay. Supercoiled or *Eco*RI precleaved DNA substrate were incubated with scIHF2, scIHF2E, scIHF2ER and scIHF2mut in the NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT) and 100 µg/ml BSA. After reaction at room temperature for 30 min, the DNA was digested with *Hind*III and *Not*I. Then the samples were analyzed on 4% polyacrylamide gels.

B.2.2.4. Recombination assays

Recombination assays with supercoiled pλIR and pλER as DNA substrates

Reactions were performed essentially as described (Corona et al., 2003) in 25 µl buffer containing 84 fmol att sites and a 90-, 400-, 400- or 1000 molar excess of scIHF2, scIHF2E, scIHF2ER or scIHF2mut, respectively, and 1 pmol Int. Excisive recombination assays contained in addition 3 pmol purified Xis. Reactions were kept at room temperature for 40 min, followed by DNA precipitation and digestion with *Nco*I. Restriction fragments were analyzed on a 0.8% agarose gel run in 0.5xTBE buffer and visualized through ethidium bromide staining.

Excisive recombination assays with supercoiled or linear pλER as DNA substrates.

The excisive recombination assay on supercoiled pλER was performed as described above, while the reactions were 20 µl and four different molar ratios of scIHF2E and scIHF2ER to DNA were tested, 400:1, 200:1, 100:1 and 50:1.

With linear DNA as substrate, 43 fmol *Pst*I digested pλER was used in 20 µl reactions containing a 90-, 400-, 400- or 1000-fold molar excess or less of scIHF2, scIHF2E, scIHF2ER or scIHF2mut as well as 1 pmol Int and 3 pmol Xis. After 40 min, reactions were stopped by 1% SDS.

Excisive recombination assays with supercoiled *att*R and linear *att*L or supercoiled *att*L and linear *att*R as DNA substrates

Supercoiled or linear pWSRGFP containing *att*R and in the same molar ratio as supercoiled or linear pCMVattL with *att*L sites are the DNA substrates in this

assay. In each reaction, IHF, scIHF2, scIHF2E, scIHF2ER and scIHF2mut have 30-, 90-, 400-, 400- and 1000- fold molar excess over DNA substrates. The reaction condition is the same as the previous excisive recombination assay. However, these reactions were stopped by 1% SDS after incubation at room temperature for 40 min. The recombination products were analyzed in 0.8% agarose gel electrophoresis.

B.2.2.5. Crystallization of scIHF2-H' and scIHF2E-H' complexes

(in cooperation with and guided by Prof. Dr. Curt A. Davey, NTU, Singapore)

Three HPLC-purified oligonucleotides H'-1: 5'-GGCCAAAAAAGCATT-3'; H'-2: 5'-GCTTATCAAT TTGTTGCACC-3' and H'-3: 5'-CGGTGCAACAAATT GATAAGCAATGCTTTT TTGGC-3' were dissolved in the annealing buffer (50 mM Tris.HCl pH 7.5, 100 mM KCl, 1 mM EDTA pH 8.0). Equimolar amounts of H'-1, H'-2 and H'-3 were mixed and annealed by incubation at 95 ^{o}C for 5 min and cooled down to room temperature slowly. Then the DNA was concentrated and the buffer exchanged with 10 mM Hepes (pH 7.0), 20 mM NaCl, 0.1 mM EDTA (pH 8.0) in 8% glycerol through Amicon Ultra-4 Centrifugal Filter Devices (Millipore). scIHF-DNA complex was formed by mixing purified scIHF protein and H' DNA with 1:1 stoichiometry in a buffer containing 10 mM Hepes (pH 7.0), 0.02 M NaCl, 0.1 mM EDTA, and 8% glycerol to give a final concentration of 14 mg/ml on protein-DNA complex.

Crystallographic analysis

(performed by Prof. Curt A. Davey, NTU, Singapore)

Crystals of scIHF2-DNA complex were grown via vapour diffusion at 18°C in droplets made by mixing 14 mg/ml complex solution with an equal volume of well solution containing 50 mM Tris (pH 7.5), 15% PEG5000-MME, 10% glycerol, 50 mM NaCl, and 30 mM $MnCl_2$. For crystallization of scIHF2E-DNA complex, a modified well solution was used containing 50 mM Tris (pH 7.5), 25% PEG5000-MME, 20% glycerol, 50 mM NaCl, and 7.5 mM $MnCl_2$. Large crystals of the scIHF2-DNA complex and initial crystals of the scIHF2E-DNA complex were obtained by microseeding using scIHF2-DNA crystals.

For stabilization and improvement of diffraction capability, crystals were transferred into a buffer containing 50 mM Tris (pH 7.5), 15% PEG5000-MME, 10% glycerol, 50 mM NaCl, 30 mM $MnCl_2$, and 10% (v/v) 2-methyl-2,4-pentanediol (scIHF2-DNA) or 50 mM Tris (pH 7.5), 15% PEG5000-MME, 10% glycerol, 50 mM NaCl, 7.5 mM $MnCl_2$, and 15% (v/v) 2-methyl-2,4-pentanediol (scIHF2E-DNA). Crystals were flash-cooled in liquid N_2 and transferred into a N_2 gas stream at -170 ^{o}C for collection of X-ray diffraction intensities on a Rigaku MicroMax 007 X-ray diffractometer equipped with a Rigaku Raxis IV++ image plate. Diffraction data were processed using MOSFLM and SCALA from the CCP4 suite. The scIHF2-DNA structure was solved by molecular replacement with routines from the CCP4 suite using the wild-type IHF-DNA model. Structural refinement was carried out with the programs O and CNS. In the final stages of

refinement, specific B-form sugar pucker restraints present in the CNS DNA parameter file were removed. Molecular structure figures were prepared with PyMOL. Atomic coordinates have been deposited in the RCSB Protein Data Bank with accession codes XYZ1 (scIHF2-DNA) and XYZ2 (scIHF2E-DNA).

B.2.2.6. AFM Imaging

(performed by Dr. Hu Chen, Dr. Yingjie Liu and Prof. Jie Yan, NUS, Singapore)

Binding reactions contained 0.246 nM *att*L and 10-, 30-, 100-fold excess of IHF, scIHF2, and scIHF2E, respectively, in 0.5X TBE buffer. Complexes were formed at room temperature for 40 min. Droplets of 30 µl were spotted on AP- or Glu-mica and incubated for 15 min. Samples were washed with deionized water and dried under pure nitrogen. Imaging was done with a Veeco Dimension 3000 AFM and Nanoscope IIIa controller (Digital Instruments, Santa Barbara, CA, U.S.A.). PPP-NCH silicon tips (NANOSENSORS, Switzerland) were used in tapping mode scanning (1.0 Hz). Images were analyzed with the Nanoscope software and free software ImageJ (http://rsb.info.nih.gov/ij/). AP-mica was prepared as described recently (Liu et al., 2005). Glu-mica was prepared by incubating AP-mica in glutaraldehyde aqueous solution (0.1%-5% v/v) for 10 min, and subsequently washed with deionized water and dried under pure nitrogen (Wang et al., 2002). Variation in the concentration of glutaraldehyde was used to optimize adherence of DNA to the mica surface.

B.2.2.7. Reversal of "supershifts" with divalent cations in EMSAs

In 20 µl reactions containing 0.2 nM radiolabeled 34-mer oligonucleotides H', 100-, 200- and 300-fold molar excess of scIHF2, or 200-, 400- and 600-fold molar excess of scIHF2E and scIHF2ER, respectively, were added to 0.5xTBM buffer (45 mM Tris-Boric acid, 0.19 / 38 / 76 µM $MgCl_2$) and incubated at room temperature for 40 min. Samples were then loaded onto 15% polyacrylamide gel which was pre-run in 0.5xTBM buffer at 210 V, 4 ^{o}C for 1 hr. This was followed by running at the same condition for another 8 hrs with buffer constantly exchanging between the upper and lower reservoirs.

B.2.2.8. Increase of integrative recombination efficiency of scIHF2E with divalent cations in recombination assay

In different reaction buffers (44 mM Tris.HCl pH 7.9, 60 mM KCl, 50 µg/ml BSA, 5 mM spermidine, 0 / 0.1 / 0.2 / 40 / 80 mM $MgCl_2$), 84 fmol pλIR and 400- fold molar excess of scIHF2E or 90- fold molar excess of scIHF2 as well as 1 pmol Int were mixed and incubated at room temperature for 40 min. Then DNA was precipitated and digested with *Nco*I. Samples were loaded onto 0.8% agarose gel to analyze recombination products.

B.3. scIHF2 in mammalian cells

B.3.1. Materials

B.3.1.1. Plasmids

pCMVSSeGFP	From (Lorbach et al., 2000)
pCMVssIHF2	From (Corona et al., 2003)
pCMVssIHF2mut	This study
pCMVss-IHF2eGFP	This study
pCMVss-IHF2muteGFP	This study
pCMVss-IHF2eGFPNLS	This study
pCMVss-IHF2muteGFPNLS	This study
poIHF2P	From (Corona et al., 2003)
poIHF2mutP	This study
poIHF2-GFP	From Dr. Nicole Christ
poIHF2mut-GFP	This study
pTrc-IHF2eGFP	This study
pTrc-IHF2muteGFP	This study
poIHF2N	This study
poIHF2mutN	This study

B.3.1.2. Cell lines

Hela	lab stock (Henrietta Lacks, 1951)
NIH/3T3	From ATCC

IHF17/H	From (Corona et al., 2003)
IHF19/H	From (Corona et al., 2003)
IHF2mut1-7/H	This study

B.3.1.3. Media and trypsin-EDTA for cell culture

D3: Dulbecco's modified eagle medium (DMEM), 10% (v/v) fetal calf serum (FBS), 2 mM L-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin

M3: D3 + 5 μg/ml puromycin

*Stock solution of puromycin: 5 mg puromycin was dissolved in 1 ml ddH_2O and sterilized by filtration through 0.22μM filter.

Other solutions: 100x L-glutamine (200 mM), 100x penicillin/streptomycin, 10x trypsin-EDTA; All are from Gibco

B.3.1.4. Cisplatin solution

Cisplatin (sigma) was dissolved in ddH_2O with concentration 5 mM by mixing overnight.

B.3.1.5. Transfection reagent

Lipofectamine 2000: From Invitrogen

B.3.1.6. Synchronization reagent

Thymidine	From Sigma
Colcemid	From Invitrogen

B.3.1.7. Cell staining buffer or Kit

Annexin-V-FLUOS staining kit	From Roche
DAPI	From Sigma

B.3.2. Methods

B.3.2.1. Construction of plasmids

poIHF2mutP

poIHF2P digested with *Asi*SI and *Nco*I was used as the backbone. To introduce a *Nco*I restriction site, the 100bp fragment from pETscIHF2mut having *Nde*I on both ends was replaced by *Nde*I digested fragment from poIHF2P formed a new plasmid pET-poIHF2mut. Fragment from pET-poIHFmut was double digested with *Nco*I and *Asi*SI and inserted into vector by ligation. (Fig.B.3.1).

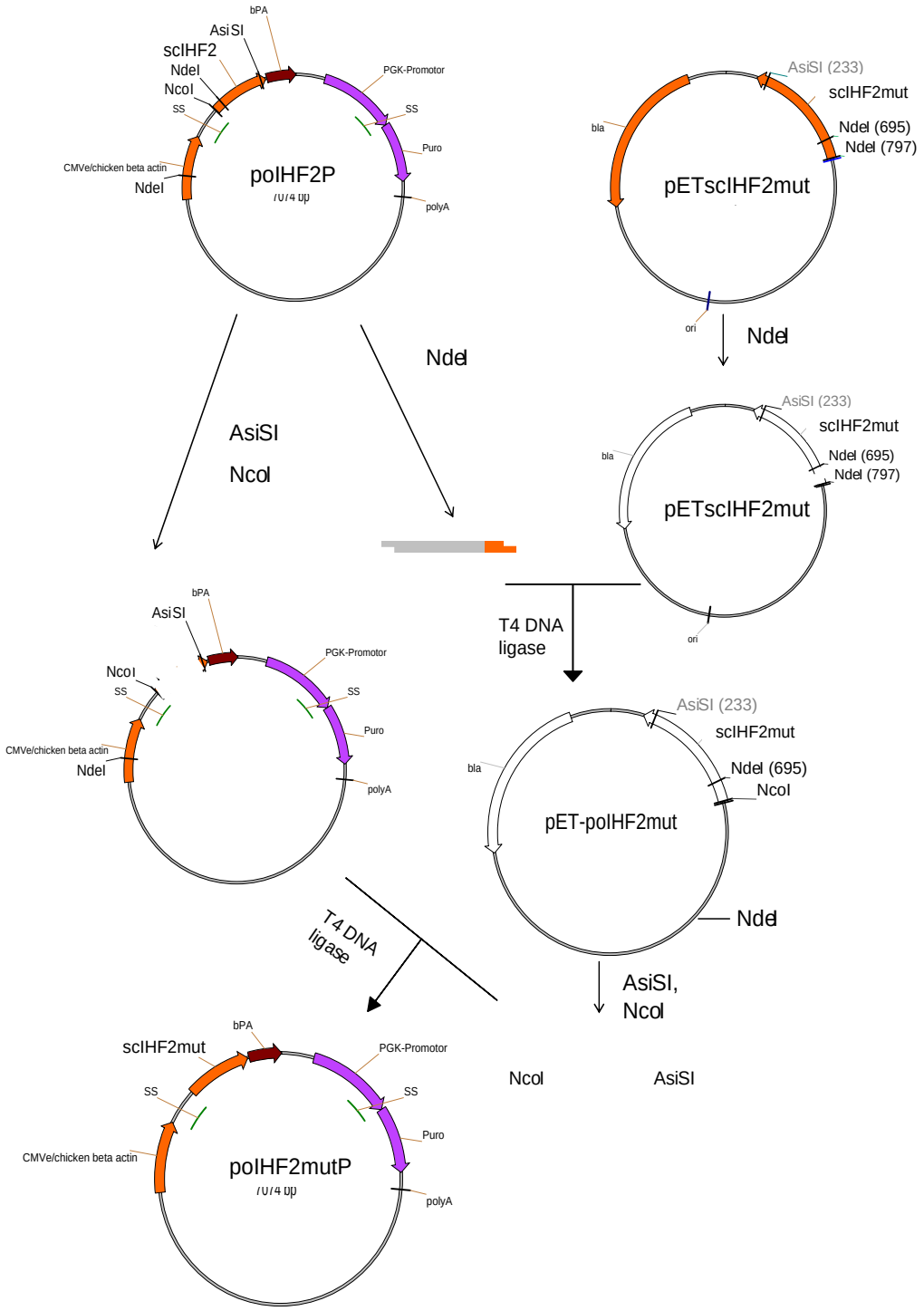


Fig.B.3.1. Construction of poIHF2mutP

pCMVssIHF2mut

Procedure for constructing pCMVssIHF2mut is similar to poIHF2mutP construction, except the restriction enzymes used were *Pst*I and *Asi*SI. The backbone was the 4.2 kb fragment from pCMVssIHF2 digested with *Pst*I and *Asi*SI. The targeting fragment was derived from the same double digested pET-pCMVssIHF2mut which was the ligation product of 750 bp fragment from *Nde*I digested pCMVssIHF2 and pETscIHF2mut which was digested with *Nde*I. (see Fig.B.3.2)

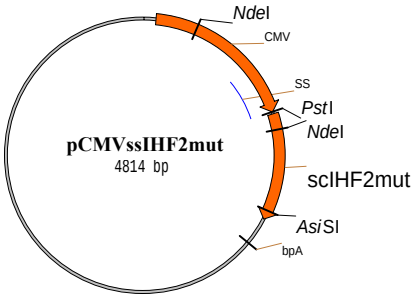


Fig.B.3.2. Map of pCMVssIHF2mut.

poIHF2mut-eGFP

poIHF2mut-eGFP was produced by the replacing of *Pst*I and *AsiS*I digested fragment from poIHF2-eGFP with 570 bp fragment from the same restriction digest of poIHF2mutP.

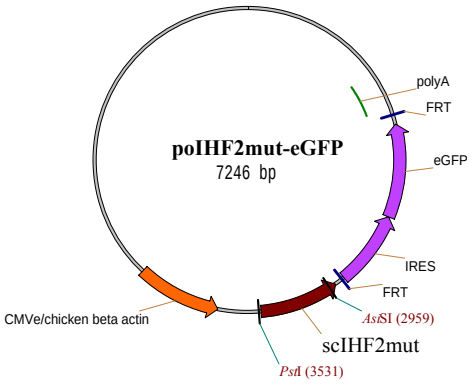


Fig.B.3.3. Map of poIHF2mut-eGFP.

pCMVss-IHF2eGFP and pCMVss-IHF2muteGFP

The first strategy used for the construction of pCMVss-IHF2muteGFP was done as follows: scIHF2mut with an additional six amino acids GGSGGS was amplified by PCR with pCMVssIHF2mut as template and with N: 5'- AAACTGCAGA TGGGC ACCAAGTCAGAAT -3' and C: 5'-CATGCCATGGAACCGCCGCTTCCACCG CCTGATCCACCGT AGAT-3' as primers. PCR product was digested with *Pst*I and *Nco*I and ligated to two fragments from pCMVsseGFP, one was the CMV promoter with *Eco*RI and *Pst*I sites, the other was the eGFP backbone with *Eco*RI and *Nco*I sites**.** The recombinant plasmid was named pCMVss-IHF2muteGFP*. For pCMV**ss**-IHF2eGFP, scIHF2mut from pCMVss-IHF2muteGFP* was replaced by a 600 bp fragment from pCMVssIHF2 digested with *Asi*SI and *Pst*I. pCMVss-IHF2muteGFP with the identical backbone as pCMVss-IHF2eGFP was constructed in the same manner as pCMVss-IHF2eGFP, where the replacement fragment was from pCMVssIHF2mut digested with *Asi*SI and *Pst*I. (see Fig.B.3.4)

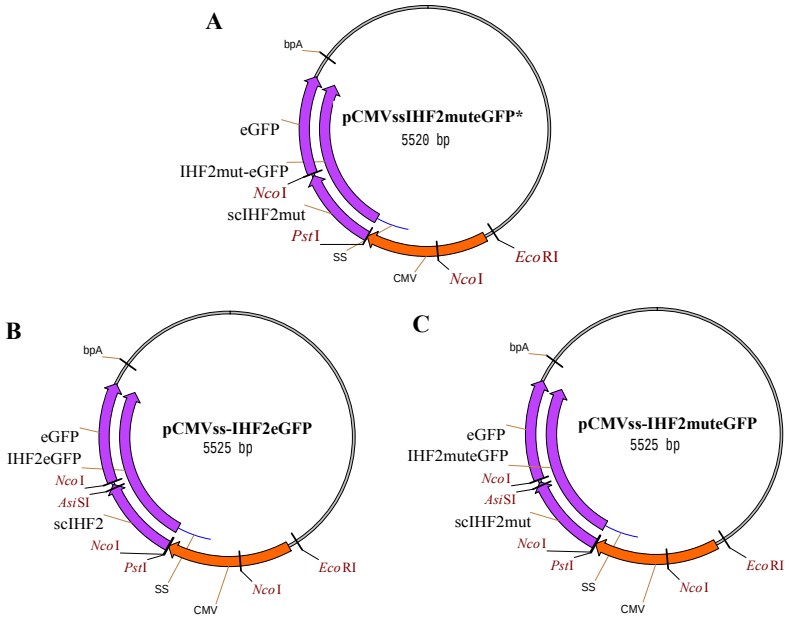


Fig.B.3.4. Maps of pCMVss-IHF2eGFP and pCMVss-IHF2muteGFP. **A**. The first strategy for

the pCMVss-IHF2muteGFP. **B**. Map of pCMVss-IHF2eGFP. **C.** pCMVss-IHF2muteGFP with the identical backbone as pCMVss-IHF2eGFP.

pCMVss-IHF2eGFPNLS and pCMVss-IHF2muteGFPNLS

Nuclear localization signal (NLS) was added to the C-terminal of eGFP by assemble PCR. First, G1 and G2 were amplified with pCMV-IHF2eGFP as the template, G1-F, G1-R and G2-F, G2-R as primers, respectively. Then the final PCR product was amplified with G1 and G2 as templates, G1-F and G2-R as primers. The primers sequences were shown as follows: G1-F (23 mer) 5'- GAAC TGCGCGATCGCGCCAATAT -3'; G1-R (39 mer) 5'- AACCTTCCTCTTCTT CTTAGGCTTGTACAGCTCGTCCAT- 3'; G2-F (46 mer) 5'- CCTAAGAAGAA GAGGAAGGTTTAATAGTCTAGAGCTCGCTGATCAG - 3'; G2-R (27mer) 5'- AGCTCCACCGCGGTGGCGGCCGCTCTA- 3'. The PCR product was digested with *Asi*SI and *Not*I. After purification, this fragment was ligated with the vector backbone from the *Asi*SI and *Not*I double digested pCMVssIHF2 formed pCMVss-IHF2eGFPNLS. pCMVss-IHF2muteGFPNLS was constructed with the same manner, but the vector backbone was from pCMVssIHF2mut. (see Fig.B.3.5)

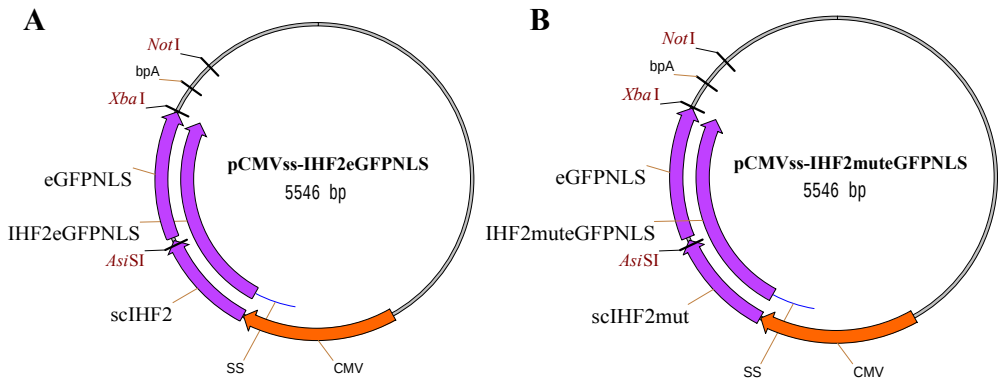


Fig.B.3.5. Maps of pCMVss-IHF2eGFPNLS and pCMVss-IHF2muteGFPNLS. **A.** pCMVss-IHF2eGFPNLS. **B.** pCMVss-IHF2muteGFPNLS

pTrc-IHF2eGFP and pTrc-IHF2muteGFP

The inserted fragment was amplified by PCR using pCMV-IHF2eGFP as the template and L: 5'- AACTGCGCGATCGCGCCAATAT- 3' and R: 5'-GCTCT AGACTATTACTTGTACA GCTCGT- 3' as primers. The fragment was digested with *Asi*SI and *Xba*I and ligated to the vector backbone derived from pTrcscIHF2 digested with the same enzymes. pTrc-IHF2muteGFP is constructed by the replacement of scIHF2 with *Nco*I sites at both ends in pTrc-IHF2eGFP with scIHF2mut from *Nco*I-digested pCMV-IHF2muteGFP. (Fig.B.3.6)

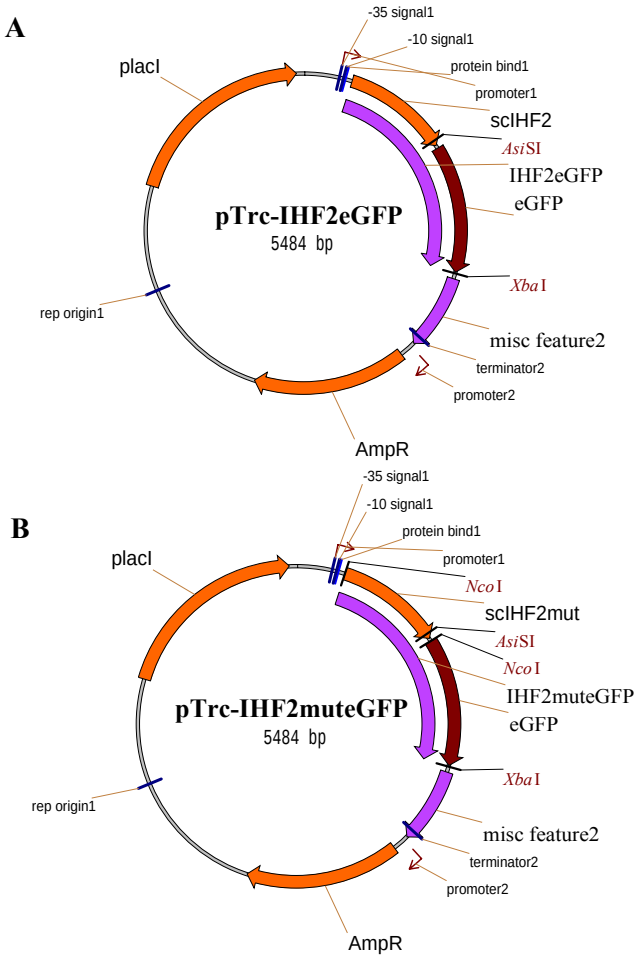


Fig.B.3.6. Maps of pTrc-IHF2eGFP and pTrc-IHF2muteGFP. **A.** pTrc-IHF2eGFP. **B.** pTrc-IHF2muteGFP.

poIHF2N and poIHF2mutN

A fragment containing scIHF2-bpA-pPGK was digested from poIHF2P by *Pst*I and ligated to the *Pst*I site of pgkSSneobpA forming poIHF2N. poIHF2mutN was constructed in the same manner as poIHF2N, except scIHF2mut-bpA-pPGK was digested from poIHF2mutP.

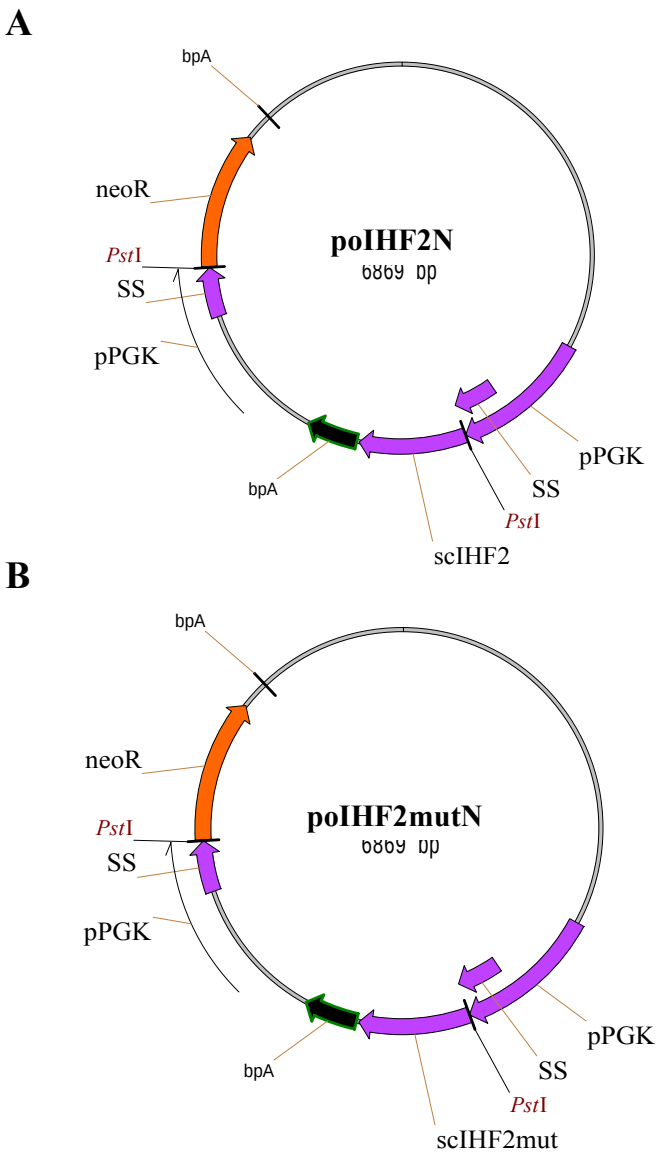


Fig.B.3.7. Maps of poIHF2N and poIHF2mutN. **A.** poIHF2N; **B.** poIHF2mutN

B.3.2.2. Cell culture

Hela cells were cultured in D3 media including Dulbecco's modified eagle medium (DMEM), 10% (v/v) fetal calf serum (FBS), 2 mM L-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin. While H/IHF17 and H/IHF19 were cultured in M3 media (D3 media plus 5 μg/ml puromycin).

Construction of stable cell line H/IHF2mut

Hela cells were passaged in 175 cm^2 flask by treatment with 0.5% trypsin-EDTA one day before transfection. Around 30 μg poIHF2mut was transfected into 10^7 Hela cells by electroporation at 300 V and 950 μF using a Gene Pulser (Bio-Rad) in a 0.2 cm cuvette. All the cells in the cuvette were placed into 10 cm Petri-dish containing 10 ml of D3 media. On day 2, 20 μg/ml puromycin was applied to the transfected cells to select stable cell lines H/IHF2mut. Colonies on the plates were inoculated into 96 wells plate to grow then transferred to 24 wells and flasks for growth. Finally, the cells were harvested and resuspended with storage solution 10% (v/v) DMSO-90% (v/v) FBS ($2x10^6$ cells/ml). The cells were aliquoted and frozen in a Freezing Container (Nalgene) at -80 ^oC for 24 hrs before transferred into liquid nitrogen for long term storage.

B.3.2.3. Western blot analysis

Recovering cell stocks from liquid nitrogen

The frozen vials containing Hela, H/IHF and H/IHF2muts were thawed rapidly by incubation in a 37 ^oC water bath. After 3 min, the cells were dropped into 15 ml

falcon tube containing 9 ml of D3 media by mixing gently. The freezing solution was decanted after centrifugation at 600xg, room temperature for 3 min. Each cell pellet was then resuspended in D3 or M3 media.

Western blot analysis

The expression of scIHF2 or scIHF2mut in stable cell lines was analyzed by Western blot. Cell lysates from $1x10^6$ cells were prepared by sonicating at 2 output Watts and 3 seconds pulse for half minute and boiling in 1 x SDS sample buffer for 5 min (sonication device: ultrasonic processor VCX 130 from Sonics & Materials Inc; the probe: 2 mm microtip) Equal amounts of proteins were separated on a denaturing 12% (w/v) polyacrylamide gel and transferred onto a PVDF membrane (Immobilon P, Millipore) by semi-dry tranferring. The membrane was blocked with 5% blocking solution (skim milk) and incubated with rabbit polyclonal antibodies raised against wild-type IHF. Bound antibody was detected with anti-rabbit peroxidase-conjugated secondary antibody (Amersham Biosciences). Anti-β-actin antibodies (Sigma) were used to control for differences in the amount of protein loaded per well.

B.3.2.4. Analysis of cisplatin tolerance

Cell lines were passaged into 6 well plates with 20% confluence one day before the cisplatin treatment. Cisplatin was added at a final concentration of 0, 10, 20, 40, 80 and 160 μM. After 20 hrs, cells (including dead and live cells) were harvested and washed with PBS by centrifugation at 600xg, room temperature for 3 min. Cell

pellets were resuspended with 100 µl HEPES buffer including 2 µl Annexin-V-Fluos labeling reagent as well as 2 µl Propidium iodide solution (Roche) and stained at room temperature for 15 min. Then 0.5 ml HEPES buffer was added to stop the reaction. Cells were analyzed by FACS.

B.3.2.5. Cellular localization of scIHF2 and scIHF2mut

B.3.2.5.1. DNA-Binding of fusion protein IHF2eGFP and IHF2muteGFP

pCMVSSeGFP**,** pCMVss-IHF2eGFP and pCMVss-IHF2muteGFP were transfected into Hela cells by electroporation seperately. The transfected cells were cultured overnight in 37 ^{o}C, 5% CO_2. Cells were harvested with a rubber policeman and transferred into 1.5 ml eppendorf tubes after cells were washed with PBS twice. Cell pellets were resuspended evenly with 100 µl of cold buffer A containing 10 mM Hepes (pH 7.9), 5 mM $MgCl_2$, 0.1 mM EDTA, 10 mM NaCl, 1 mM DTT, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and 10 mM NaF. Cells were swollen for 5 min and lysed with a Dounce homogenizer (about 20-30 strokes). 100 µl of cold buffer B (buffer A, 830 mM NaCl, 34% glycerol) was added into the lysate and incubated at 4 ^{o}C for 45 min with continuous agitation. The lysate was centrifuged at 10000xg and 4 ^{o}C for 10 min. The supernatant was aspirated into 2 ml tube. Protein extract was aliquoted as 25 µl/tube, stored at -80 ^{o}C.

Prokaryotic vector pTrc99a and pTrc-IHF2eGFP were transformed into *E. coli* strain CSH26-ΔIHF. Protein expression was induced with 1 mM IPTG and cultured

for 5 hrs. Cells were harvested and resuspended in Buffer A and B and sonicated at 40-60 amplitude, 3 seconds pulse for 0.5 min. Lysate was separated by centrifugation at 4 ^{o}C, 10000xg for 10 min. Supernatants were aliquoted and stored at -80 ^{o}C.

Binding assays were performed in 20 µl reactions containing 0.5xTBE, 17% glycerol, 5 mM EDTA, 200 µg/ml BSA, 0.1 nM labeled DNA, 52 nM salmon sperm DNA (average MW 260 kDa) and 1 µg or less protein extracts. Reactions were incubated at room temperature for 50 min and loaded onto a 15% polyacrylamide gel. Gels were run at 4 ^{o}C, 16 volts/cm for 3 hrs and gel images were analyzed using a Bio-Rad molecular Imager System.

B.3.2.5.2. Localization of scIHF2 in mammalian cells

Synchronization of cells

NIH/3T3 cells were passaged into 6-well plates at 20% confluence. After 24 hrs, thymidine was added to a final concentration of 2 mM. 16 hrs later, thymidine containing media was removed from the well and replaced with fresh media. Colcemid at a final concentration of 0.6 µg/ml was added after 5.5 hrs further incubation. The cells were analyzed after 22 hrs.

The synchronized cells were harvested by trypsinization and washed with PBS. Cells were stained with propidium iodine (PI) in 50 mM HEPES buffer (pH 7.4) with 0.3% NP-40 by incubation on ice for 30 min, and then stained with fresh PI

buffer for another 15 min at room temperature. The staining was stopped with 1 ml HEPES buffer. The synchronization of the cells was analyzed by FACS.

Localization of fusion proteins in NIH and Hela cells.

10 μg pCMVSSeGFP**,** pCMVss-IHF2eGFPNLS and pCMVss-IHF2muteGFPNLS were transfected into synchronized NIH cells by lipofectamin 2000 (Invitrogen) in a 6-well plate, respectively. Cells grew on glass coverslips overnight, and were washed with PBS, fixed in 100% ethanol for 30 min, washed again with ddH_2O. Finally 4',6-diamidino-2-phenylindole (DAPI, Sigma) was added to stain cells at room temperature for 1.5 hrs at concentration of 5 μg/ml. The coverslips were mounted on glass slides after drying at room temperature overnight. The location of fusion proteins was analyzed with Laser Scan Microscope (LSM, Zeiss).

The fusion protein scIHF2eGFP without NLS in Hela cells were also analyzed. 30 μg pCMVSSeGFP, pCMVss-IHF2eGFP and pCMVss-IHF2muteGFP were transfected into $1x10^7$ Hela cells by electroporation, respectively. Cells were cultured in the 6-well plates with glass coverslips inside overnight. On the second day, cells were fixed in 2 ml 10% formaldehyde/70% glycerol and stained with DAPI and analyzed with confocal microscope.

C. Results

C.1. Characterization of scIHF2 and its variants

C.1.1. Background - design of a single chain integration host factor

(performed by Dr. Nicole Christ and Prof. Thomas Schwartz; published in Corona et al., 2003)

An inspection of the published co-crystal structure of IHF bound to the H' site of λ *att*L revealed that the N- and C-terminal ends of the α-subunit are located on one side of the complex, while the corresponding termini of the β-subunit are positioned on the opposite side (Fig.C.1.1a). This essentially precluded any possibility of connecting the two subunits via a traditional C-to-N-terminal peptide linker. However, upon closer inspection, we also noticed that the two termini of the α subunit are in rather close proximity to a connecting region (β-Q39 to β-E44) between the second alpha-helix and the first beta-sheet of the β-subunit (Fig.C.1.1a). We therefore decided to insert almost the entire α-subunit into this region of the β-subunit.

First, we modeled a 5 aa linker to connect residue β-Q39 with the N-terminus of the α-subunit at position α-L3 (linker 1), thus bridging the 11Å distance between the two residues in the co-crystal structure (Fig.C.1.1b,d; Fig.C.1.2). The freed β-G40 was then connected with α-A94 at the C-terminus of the α-subunit via a second linker, which bridges about 5Å (Fig.C.1.1b,d; Fig.C.1.2). Importantly, a

superimposition of the IHF structure and the energy-minimized scIHF2 model revealed that no significant perturbations were introduced by the two linkers (Fig.C.1.1c).

In order to explore further the feasibility of our protein engineering approach, we analyzed how the biochemical and functional properties of scIHF2 might be affected through a shortening of linkers by one aa. We thus generated three new scIHF variants and their corresponding aa sequences are depicted in Fig.C.1.2. In scIHF1, both linkers were trimmed. In scIHF3 and 4, only one linker was altered while the other remained unchanged as in scIHF2. During construction of the expression vector for scIHF3, we also generated a scIHF3 variant that has a glutamate at position 45 in the α domain instead of a lysine. The latter residue is highlighted in bold in Fig.C.1.2, and corresponds to residue 89 in scIHF2; (numbering throughout the text refers to aa sequences of IHF subunits). This particular amino acid exchange in scIHF3 seemed interesting to us because a glutamate is in fact present at the corresponding position of the homologous β domain (Ellenberger and Landy, 1997). In order to distinguish it from scIHF3, we named the mutant protein scIHF3E (scIHF3K45αE in Bao, Q., et al., 2004).

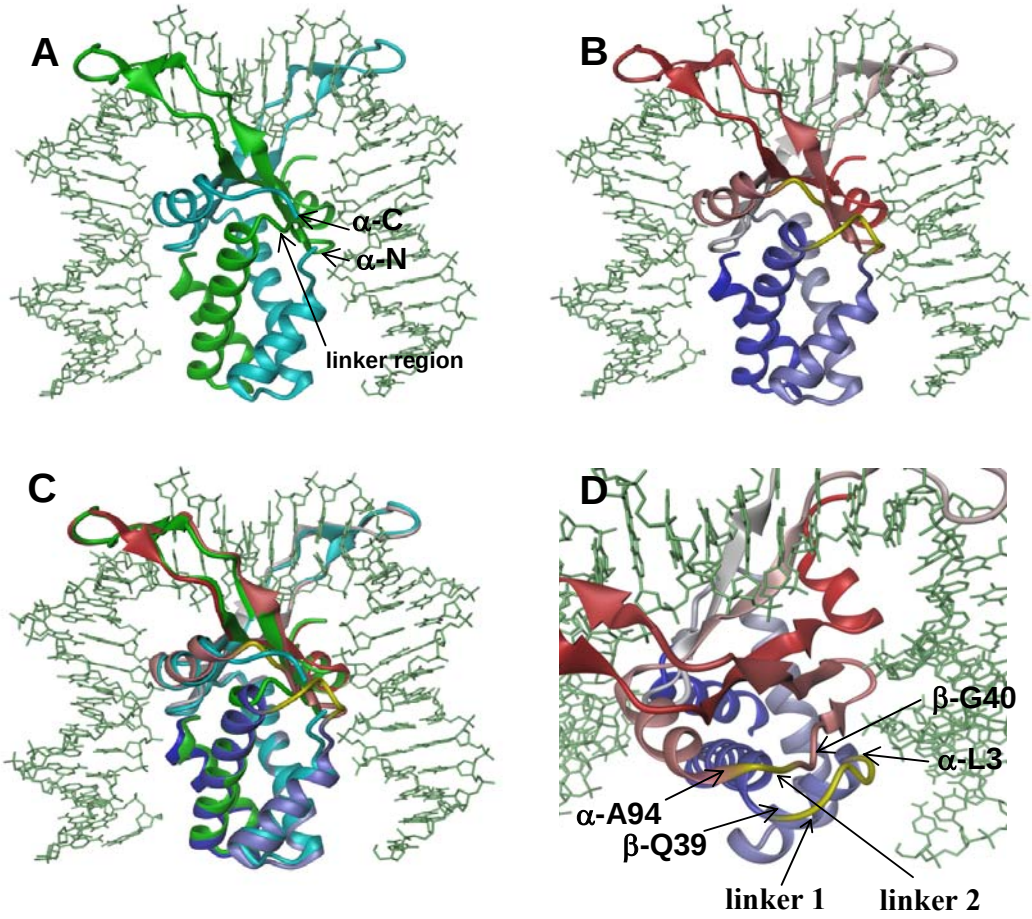
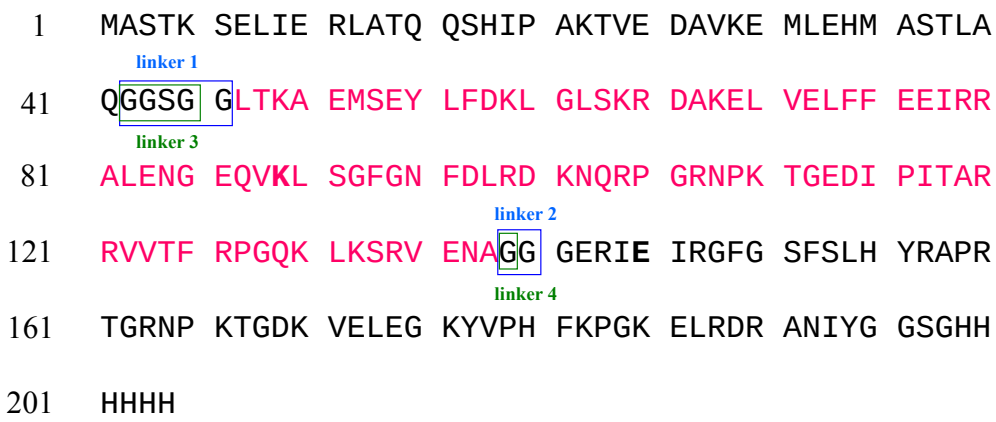


Fig.C.1.1. Structure of IHF and model of scIHF2. (A) Co-crystal structure of IHF and the H' site of *att*L, as published by Rice et al. with α- and β-subunit drawn as ribbons in cyan and green, respectively. The N- and C-terminal ends of the α-subunit are marked, as well as the short linker region in the β-subunit that was chosen as an insertion point for the α-subunit to generate scIHF2. (B) scIHF2-H' model. The two linkers used to connect the two subunits are shown in yellow. The protein is colored in a gradient from blue over white to red following N- to C-terminus. (C) Superimposition of the IHF structure and the scIHF2 model. (D) Same as (B), but zoomed and tilted around the horizontal axis to highlight the two linkers (yellow) and the respective residues that were chosen to connect the subunits. The figure was prepared with the program Dino (http://www.dino3d.org) (Reproduced from Corona, T., et al., 2003)



scIHF1: linker 3 + linker 4

scIHF2: linker 1 + linker 2

scIHF3: linker 3 + linker 2

scIHF4: linker 1 + linker 4

scIHF3E: linker 3 + linker 2

Fig.C.1.2. Amino acid sequence of scIHF variants. The respective peptide linkers are bracketed and labeled, and combinations of linkers in each scIHF are listed below the sequence. The sequence of the α domain within scIHFs is marked in red, and the lysine at position 45 is highlighted in bold and substituted by a glutamate in scIHF3E. The corresponding glutamate at position 44 of the β domain is also highlighted in bold.

C.1.2. Biochemical properties

Purification of scIHFs (Bao et al., 2004)

In order to simplify the purification procedure for scIHFs, we generated an *E. coli* expression vector (pETscIHFs) that contains six histidine codons at the C-terminus of the coding sequence (Fig.C.1.2). We reasoned that the short tag would not significantly interfere with either the DNA-binding or -bending because the C-terminus appears to be at a rather remote location from DNA (Fig.C.1.1b). Two amino acid residues, alanine and serine, were added at the second position for the

using of *Nhe*I in cloning strategy. The calculated molecular weight of the recombinant, His-tagged protein is therefore 22.8 kDa. The tag allowed us to purify 30-50 mg of scIHFs with >95% purity from a one liter culture. We purified all five scIHFs to homogeneity (Fig.C.1.3).

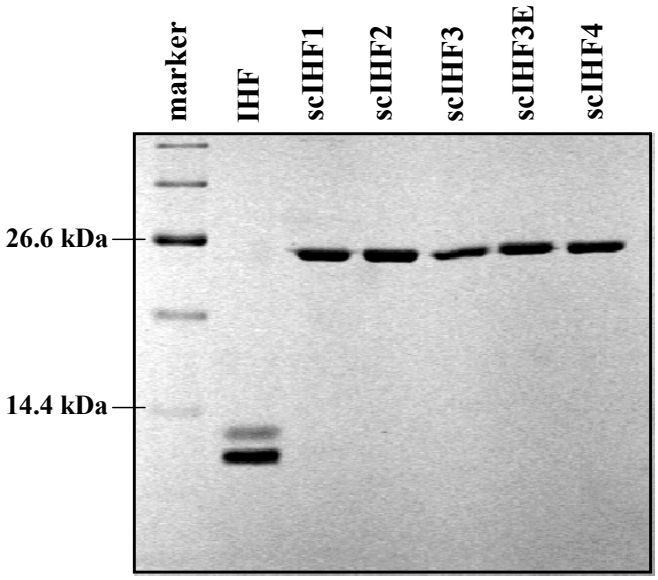


Fig.C.1.3. The purified scIHFs and IHF. 3.5 μg of each protein was analyzed on a 15% SDS-PAGE together with molecular weight markers at the left side of the coomassie-stained gel.

DNA-Binding

To determine whether the single chain integration host factors (scIHFs) still have the binding affinity to IHF consensus sequences, gel shift assays were performed. In this assay, we employed four different radiolabeled oligonucleotides comprising the λ H', H1 and H2 sites as well as the IHF site in the pSC101 origin. The assays were performed under two different conditions. In addition to a low ionic strength (0.5xTBE) binding buffer, we employed also one containing 100 mM NaCl in order to resemble more closely the conditions for protein–DNA interactions in our

functional assays described below. A representative example shown in Fig.C.1.4. is binding of scIHF2 and wild type IHF employing a radiolabeled oligonucleotide comprising the λ H' site in low salt conditions. When the concentration of scIHF2 or IHF was increased, the free labeled H' is reduced, while the IHF (scIHF2)-H' complexes increase (Fig.C.1.4a, b). In contrast, less IHF was needed to fully occupied radiolabeled H' than scIHF2 (Fig.C.1.4c).

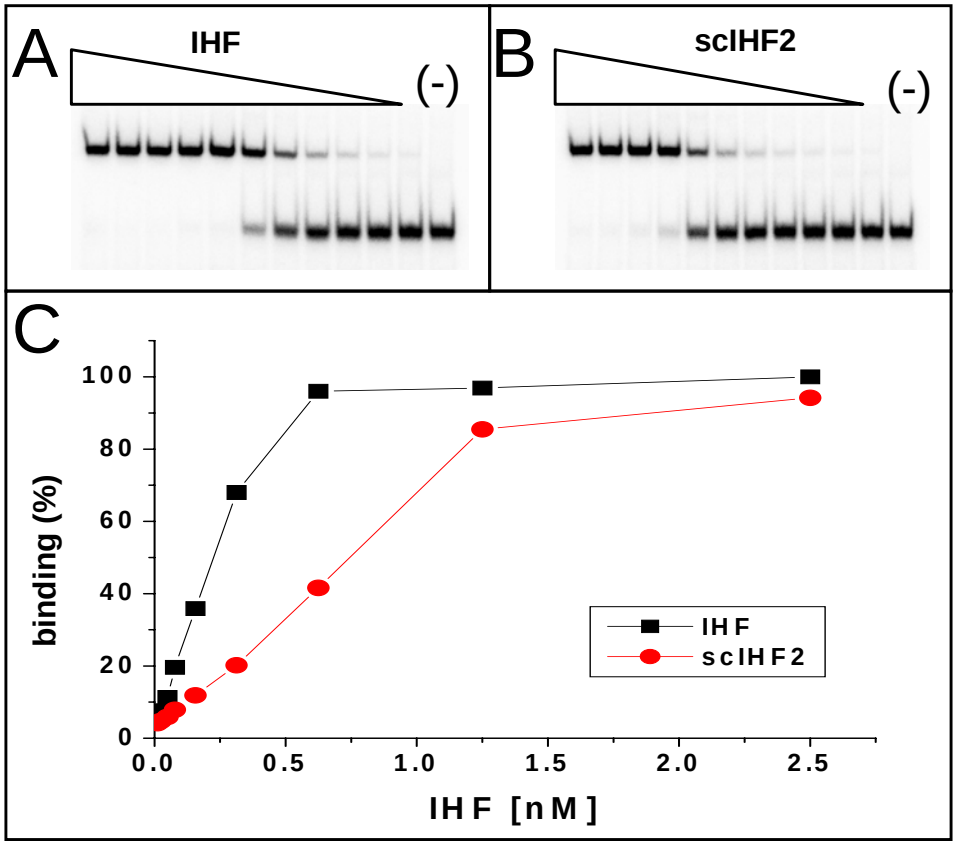


Fig.C.1.4. Binding assays of IHF and scIHF2 with H' as DNA substrate. The concentration of H' is 0.1 nM. **A** shows the binding of IHF-H', the concentration of IHF is 10 nM, 5 nM, 2.5 nM, 1.25 nM, 0.625 nM, 0.3125 nM, 0.156 nM, 0.078 nM, 0.039 nM, 0.02 nM, 0.01 nM and 0 nM. **B** shows the binding of scIHF2-H', the concentration of scIHF2 is 10nM or less. **C.** The binding curves of IHF-H' (black) and scIHF2-H' (red) according to the band densities in **A** and **B.**

K_d determination

According to the previous measurements, the dissociation constant K_d of IHF-H' is around 1 nM (Goodman et al., 1999; Yang and Nash, 1995). The DNA concentration used in the binding assay is 0.1 nM, thus lower than K_d. That is the necessary condition for our K_d calculations. Under this situation, we can define K_d as the concentration of total DNA binding protein when half the sites are occupied at equilibrium. (Rippe, 1997)

In the following assay, we measured the K_d values of scIHF2 and wild type IHF. We did the titration of IHF and scIHF2 in the binding assay. The highest concentration of IHF and scIHF2 made the DNA substrate (radiolabeled H') fully occupied. When the concentration of IHF and scIHF2 is 0 nM (lowest concentration), DNA substrate is completely unbound. The radioactivity in the free and bound forms from each lane was quantified. With these data, we generated binding curves and revealed a K_d of 0.24 nM for IHF, which is in excellent agreement with previous measurements (Goodman et al., 1999; Yang and Nash, 1995).

scIHF2 exhibits a 3-fold higher K_d (Fig.C.1.5). However, a K_d of 0.79 nM for scIHF2 indicates that the protein is still stably bound to H'. This was confirmed by a series of competition experiments using specific and unspecific competitor DNA (Fig.C.1.8; Fig.C.1.9).

As described for scIHF2, we obtained all K_d values of scIHFs with four different DNA substrates. The data presented in Table C.1.1 reveal that each scIHF binds

with high affinity to DNA targets in 0.5xTBE buffer. Only scIHF3E exhibits a somewhat reduced affinity with λ H′. However, DNA binding was in some cases severely affected when we used higher ionic strength in the binding buffer. In particular, scIHF1 and scIHF3E showed a 20- to 70-fold increase in K_d, while scIHF2 and scIHF3 were moderately influenced (<20-fold). It is apparent that the affinity of scIHF3E with the λ H′ site is now in a range of an unspecific DNA-binding protein. Finally, it should be noted that the K_d values for the heterodimeric wild-type IHF were not at all affected by the increase in ionic strength (Table C.1.1). This is in excellent agreement with previous reports (Holbrook et al., 2001).

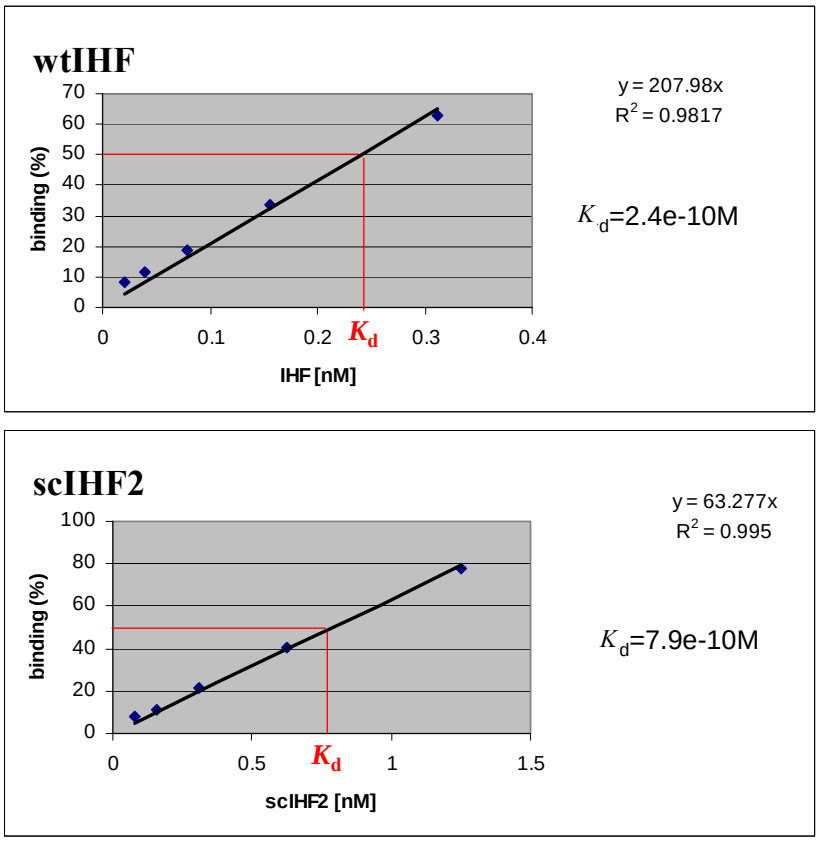


Fig.C.1.5. K_d values measurement of wtIHF-H' and scIHF2-H'. The fitted dots shown in the curves represent the binding percentage under different IHF concentration. The equations of trendlines give the relationship of binding percentage and concentration of IHF. Abscissa (x) is the concentration of IHF. Ordinate (y) is the binding percentage. When y equals to 50%, the value of x is the K_d.

A

H'	5′-GCCAAAAAAGCATTGCTTATCAATTTGTTGCACCAAA-3′
H1	5′-CATATGCAGTCACTATGAATCAACTACTTAGATGAAA-3′
H2	5′-ACGTAAAATGATATAAATATCAATATATTAAATTAAA-3′
pSC101 *ori*	5′-GTGTTTTTTTTGTTTATATTCAAGTGGTTATAATAAA-3′

B

DNA / IHF / K_d (nM)	H'		H1		H2		pSC101 *ori*	
	0.5 X TBE	100 mM NaCl	0.5 X TBE	100 mM NaCl	0.5 X TBE	100 mM NaCl	0.5 X TBE	100 mM NaCl
IHF	0.28	0.19	0.3	0.39	0.17	0.14	0.35	———
scIHF1	2.8	205	9.02	210	3.04	190	2.56	57
scIHF2	0.79	3.8	0.71	7.04	0.51	1.8	0.46	2
scIHF3	1.47	10.1	2.33	41	1.37	8.05	1.42	5.4
scIHF3E	56.2	1000	1.52	61	2.82	110	3.01	120
scIHF4	2.44	49	1.29	46	0.59	28	0.58	11

Table.C.1.1. K_d values of scIHFs. **A** shows the four different DNA substrates for the scIHFs binding assays. H', H1 and H2 are IHF binding sites on attp. pSC101 is the IHF binding site on the origin of pSC101 replication. **B.** K_d values of scIHFs binding with different IHF consensus sequences under 'low salt' and 'high salt' condition. Each binding experiment for the calculation of K_d had been repeated at least 2 times.

Comparing of scIHFs' DNA-bending abilities

Comparing scIHFs' DNA-bending abilities using EMSAs, a representative example is shown in Fig.C.1.6. Using a radiolabeled oligonucleotide comprising the λ H1 site, it is evident that scIHF variants produced band-shifts indistinguishable from that obtained with scIHF2. However, scIHF3E complexed to H1 was significantly more retarded during electrophoresis than the complex

formed between H1 and scIHF3. This was confirmed by mixing preformed H1-IHF and H1-scIHF3E complexes, and analyzing them together. Two distinct band-shifts are clearly detectable (Fig.C.1.6, lane 10). Supershifts (regarded as any scIHF-DNA complex that migrates slower than IHF-DNA complex) were also observed with all other IHF cognate sites examined at this low salt condition or higher ionic strength so far (data not shown).

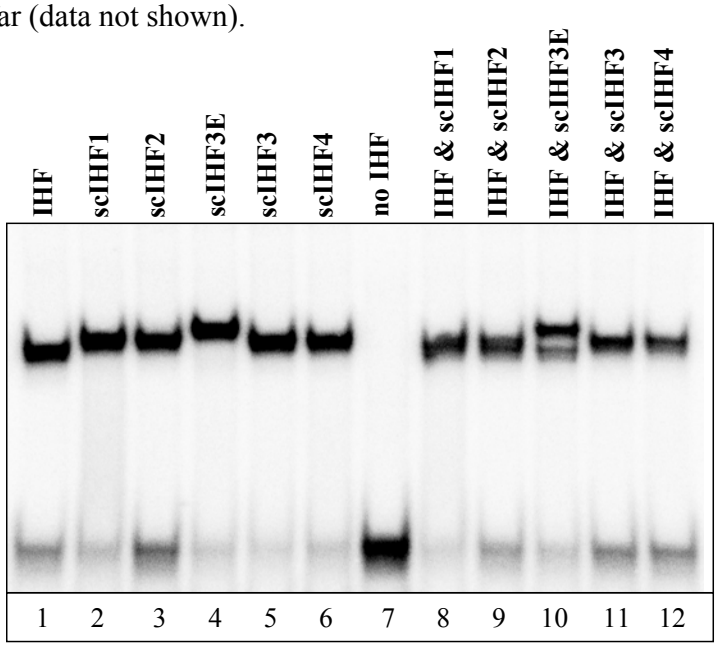


Fig.C.1.6. Analysis of DNA binding and bending. A representative example of DNA-binding and bending analyses is shown with a radiolabeled oligonucleotide comprising the H1 site which was incubated with a 10-fold molar excess of IHF, a 400-fold excess of scIHF1, a 50-fold excess of scIHF2, a 200-fold excess of scIHF3, a 200-fold excess of scIHF3E, and a 100-fold excess of scIHF4 (left). Mixed, preassembled DNA–protein complexes were analyzed to verify changes in electrophoretic mobilities between these complexes (right).

When we analyzed scIHF2's DNA-bending capacity and compared it with that of the parental protein, we employed DNA band-shift assays using one 217 bp radiolabeled fragment comprising *att*L in addition to the shorter H', H1, H2 and

pSC101 sites. Within the limits of resolution of these assays, we could not detect any significant differences in DNA-bending between wild-type IHF and scIHF2 (see Fig.C.1.7).

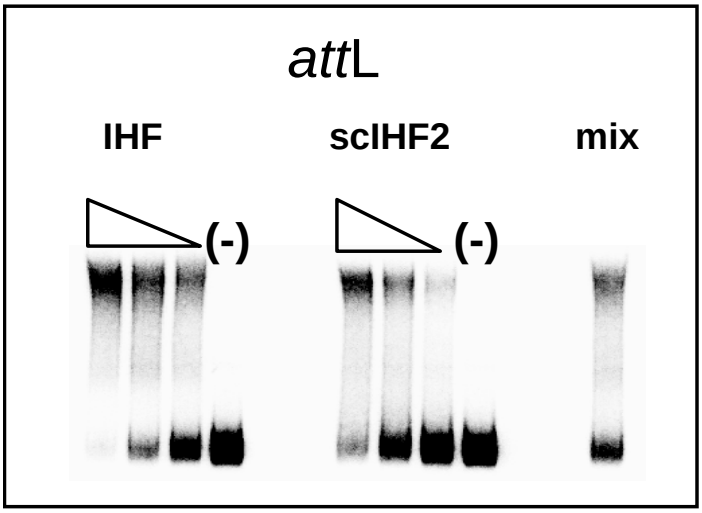


Fig.C.1.7. Analysis of DNA-bending induced by scIHF2. DNA-bending analyzed with a radio-labeled 217 bp *att*L fragment (70 fmol) and either a 20-fold (or less) excess of IHF or a 40-fold (or less) excess of scIHF2, as indicated. Pre-assembled IHF-*att*L and scIHF2-*att*L complexes were also analyzed together (mix).

Competitive DNA-binding

The DNA binding assays showed scIHF2 had quite strong binding affinity with a K_d lower than 1 nM. In order to further elucidate the binding affinity of scIHF2 to IHF cognate sites, specific and unspecific competition were performed with its parental protein IHF as a control. The DNA substrate used in both assays was radiolabeled H'. Unlabeled H' was the competitor in the specific competition assay, while sonicated salmon sperm DNA was a competitor in the nonspecific competition assay. In both assays, scIHF2 or IHF was incubated with H' first.

After a short time (around 3 min) competitor was added to compete with the binding of scIHF2 or IHF to H'. When the concentration of the competitor was increased, the proportion of bound radiolabeled H' decreased. Finally all the radiolabeled H' existed in free state if enough competitor was applied (Fig.C.1.8a,b, Fig.C.1.9a,b). In contrast to IHF, scIHF2 need fewer competitors to decrease the binding percentage to the same extent. However, 100- or 400- fold molar excess of unlabeled H' or salmon sperm DNA over labeled H' were needed to decrease the labeled H' in bound state to 50% (Fig.C.1.8c, Fig.C.1.9c). We can conclude the binding between scIHF2 and H' is strong and stable.

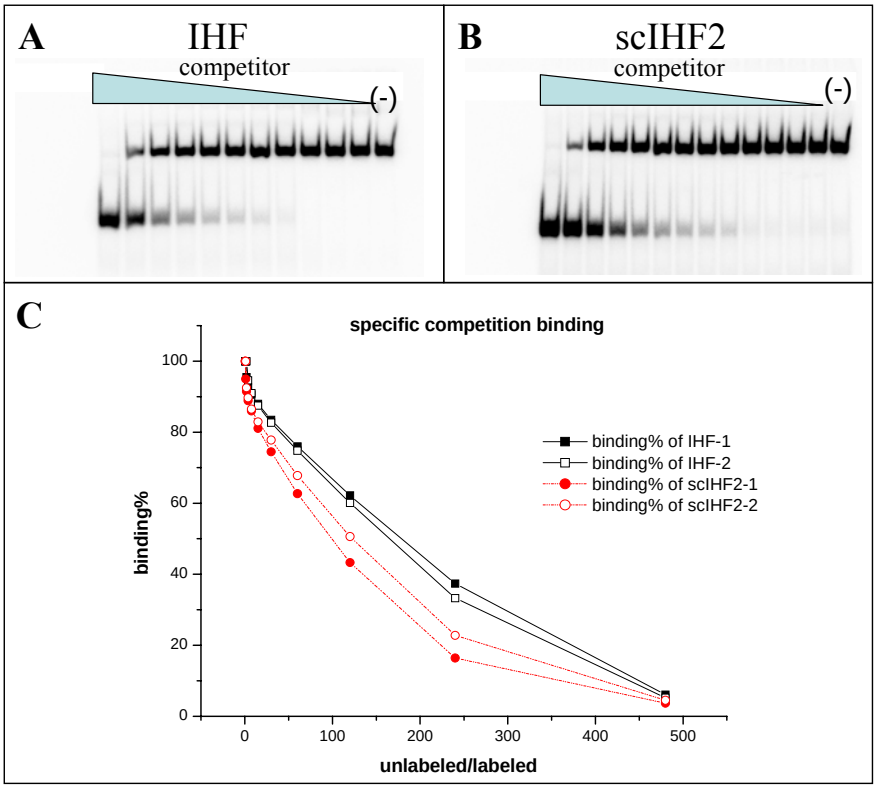


Fig.C.1.8. Specific competition assays with IHF and scIHF2. A. Specific competition binding assay of IHF. The stoichiometry of IHF to labeled H' (1 nM) is 10 to 1. Unlabeled H' has 480 molar excess or less over the labeled H'. **B**. Specific competition binding assay of scIHF2.

scIHF2/labeled H'=10/1. unlabeled H' / labeled H' = 480/1, 240/1, 120/1 …. 0. **C** shows the competition curves of IHF (black) and scIHF2 (red). Each experiment has been repeated twice.

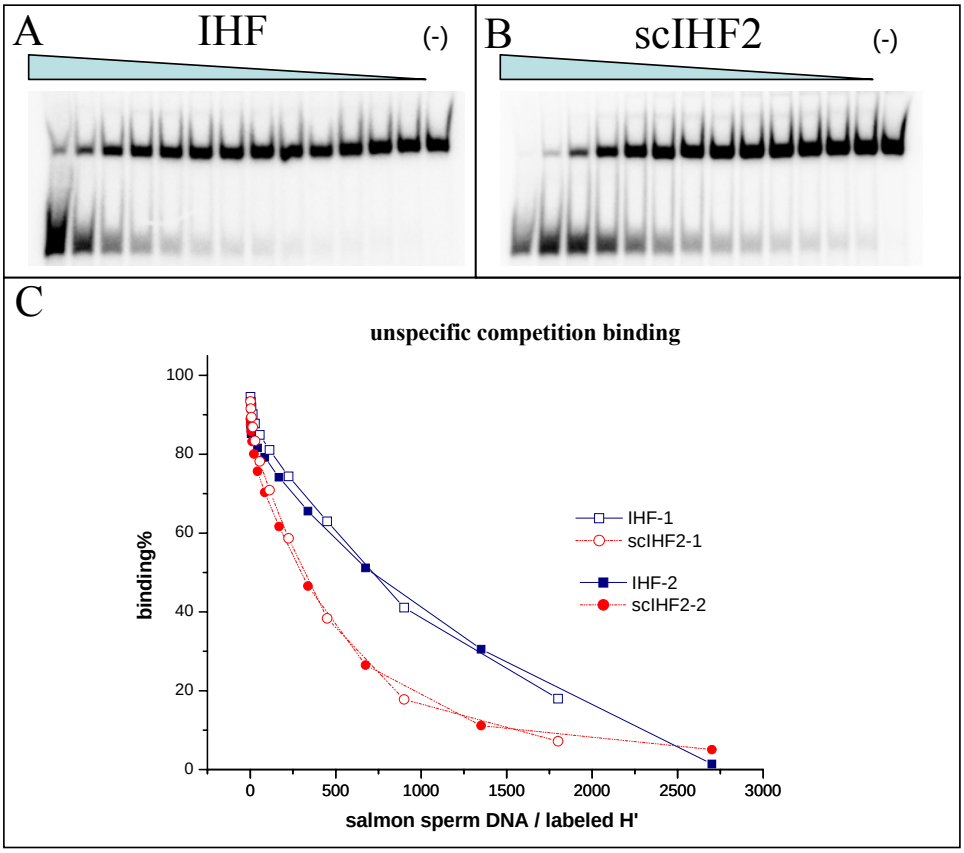


Fig.C.1.9. Unspecific competition assays with IHF and scIHF2. Radiolabeled H' is the DNA substrate (1 nM). The molar ratio of protein to labeled H' is 10 to 1, no matter the protein is IHF (**A**) or scIHF2 (**B**). As an unspecific competitor, salmon sperm DNA has 2700-fold molar excess over the labeled H'. **C** shows competition binding curves according to the results of A and B. Blue curves represent IHF, red ones are scIHF2.

Stability of IHF-H1 and scIHF2-H1 complexes

The K_d values of scIHF2-DNA (H', H1, H2 and pSC101 ori) are 2 or 3-fold higher than those of wild type IHF-DNA, but still less than 1 nM. And our previous results of specific and unspecific competition binding assays showed that much higher concentration of competitor was required to replace the radiolabeled H' bound with IHF and scIHF2. And with very high concentration of competitors, there are still IHF-DNA or scIHF2-DNA complexes after 40 min competition. This suggests that the scIHF2-H' complex is pretty stable, similar as IHF-H'. To further analyze the similarity between scIHF2 and wild type IHF, we designed an experiment to test the stability of IHF-H1 and scIHF2-H1 complexes. In this assay, IHF or scIHF2 and radiolabeled H1 were pre-incubated for 15 min and then specific competitor was added to compete with radiolabeled H1 binding with IHF or scIHF2. The competition time varied from 1 hr to 48 hrs. After electrophoretic mobility shift assay (EMSA), the radioactivity of bound and free H1 of each time point was quantified. The IHF-H1 stability curves generated according to these data showed that only 40% of radiolabeled H1 was replaced by specific competitor H1 which has 800-fold molar excess over the radiolabeled H1 after competition for 48 hrs. The stability of scIHF2-H' is similar as IHF-H' within 30 hrs, then the binding percentage decreased to 35% when the time increased to 48 hrs, 20% less than that of IHF-H'. (Fig.C.1.10)

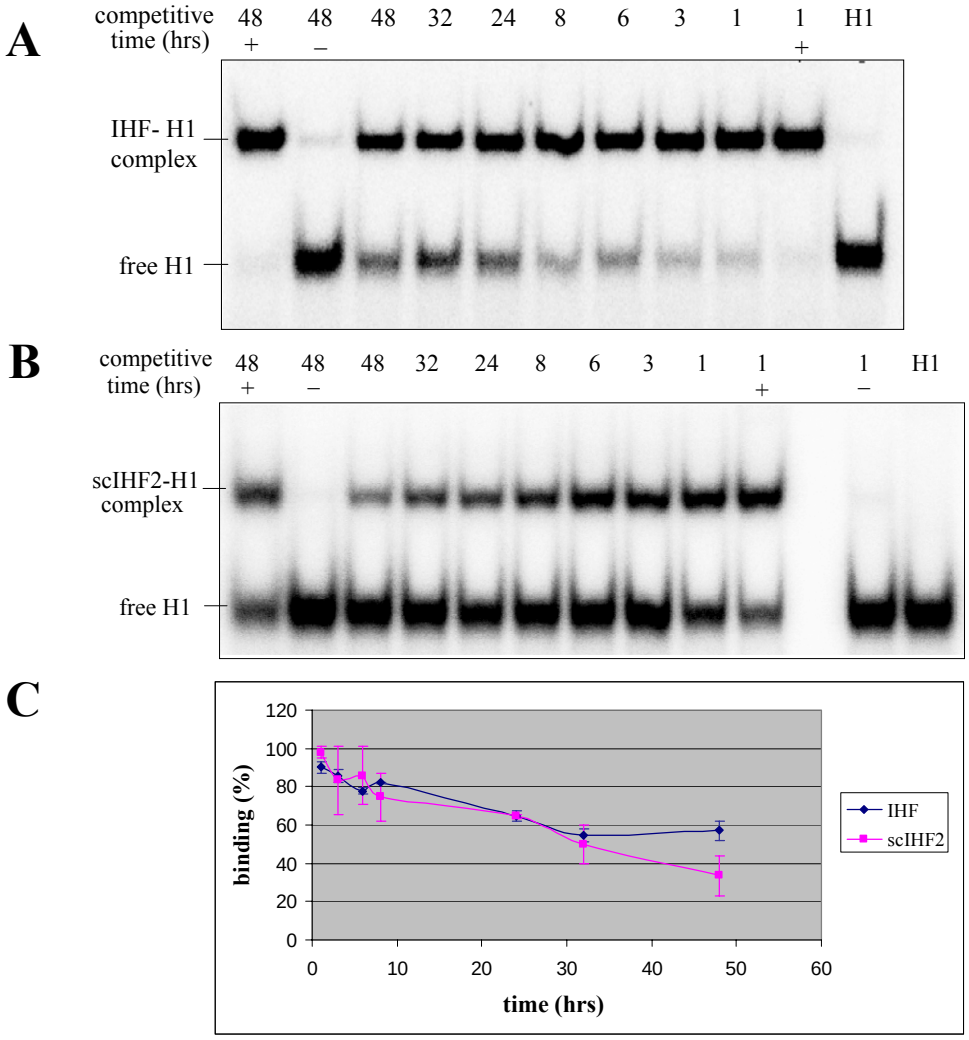


Fig.C.1.10. The stability of IHF-H1 and scIHF2-H1 complexes. **A**. Stability of IHF-H1. IHF has 10 fold molar excess over radiolabeled H1. **B.** Stability scIHF2-H1. scIHF2 had 30 fold molar excess over labeled H1. Unlabeled H1 was competitor. In both cases, the molar ratio of competitor to radiolabeled H' was 800 to 1. The concentration of radiolabeled H' is 0.1 nM. The competition time was 48 hrs or less. In **A** and **B, "+"** indicates samples without competitor, positive control. "-" presented the samples in which competitor and radiolabeled H' were incubated with IHF or scIHF2 at the same time, negative control. **C.** Stability curves of IHF-H1 and scIHF2-H1. The dots in the curves represented the relative binding percentage of each time point with the average binding (%) of positive control done in the beginning and the end (1 hr and 48 hrs) as reference.

The binding between scIHF2 and RNA

There are many IHF binding sites on the genomic DNA of *E. coli*. We were interested to find out whether RNAs secondary structures exist in *E. coli* which can be recognized and bound by IHF. With total RNA solated from *E. coli,* we found that scIHF2 bound to RNA under the condition of 60 mM NaCl. In Fig.C.1.11, there are two, perhaps three, extra bands in lanes 3, 4, 5, 6 compared with lane 12. We deduce that those bands are scIHF2- RNA complexes. Lane 1 and 2 showed the shift of the major RNA bands caused by unspecific binding. However, this interesting initial observation was not investigated further in this work.

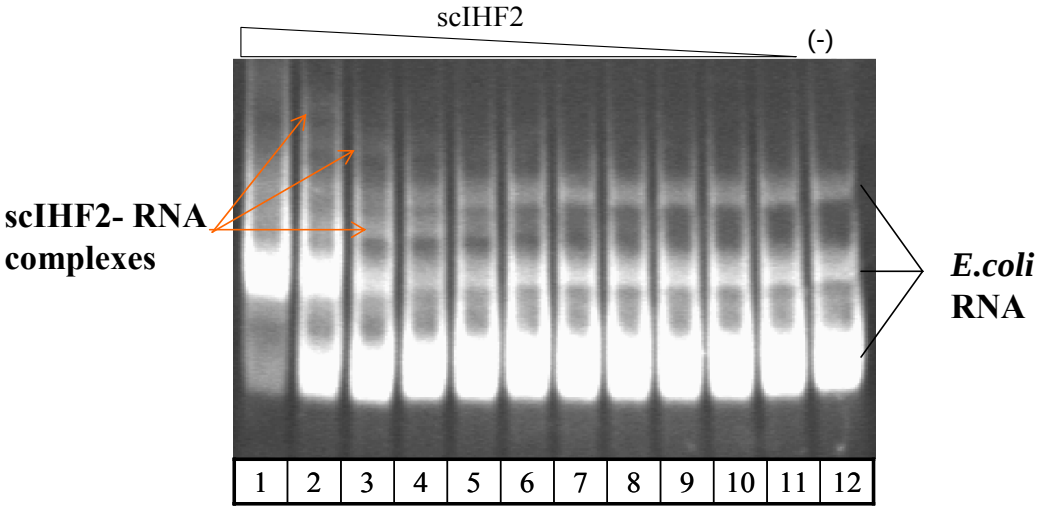


Fig.C.1.11. **scIHF2 binding to RNA from *E. coli*.** The amount of RNA for each reaction is 1.5 µg. 2 µg or less scIHF2 was added into reactions (lane 1-11). Lane 12 showed the naked RNA which has 3 bands. The bands pointed with orange arrows were scIHF2-RNA complexes.

C.1.3. Biological properties of scIHF2 and its variants

C.1.3.1. scIHF variants differ in their ability to promote λ recombination

As shown above, some marked differences in binding affinities between individual scIHFs exist at 100 mM NaCl. In addition, the super band shifts observed with scIHF3E and all four DNA substrates indicated that the DNA geometry within these complexes must be significantly altered. We therefore tested how these scIHFs performed in integrative and excisive recombination *in vitro*. We found that scIHF2, scIHF3, and scIHF4 promoted integrative recombination to an extent comparable with IHF (Fig.C.1.12. lanes 5, 8, 9, and 3, respectively). However, scIHF1 and scIHF3E appeared almost inactive (lanes 4, 6, and 7). With the exception of scIHF3E, we also tested whether the scIHFs promote integrative recombination in *E. coli* strain CSH26ΔIHF as described before (Corona et al., 2003). In excellent agreement with our *in vitro* results, we found that scIHF1 exhibited no detectable activity while the other three scIHF variants promoted integrative recombination efficiently (data not shown, Christ, N., 2002). We also analyzed scIHFs' abilities to promote excisive recombination *in vitro*. As depicted in Fig.C.1.12 (lanes 12 to 20), all of them, including scIHF3E, promoted excisive recombination to a significant extent.

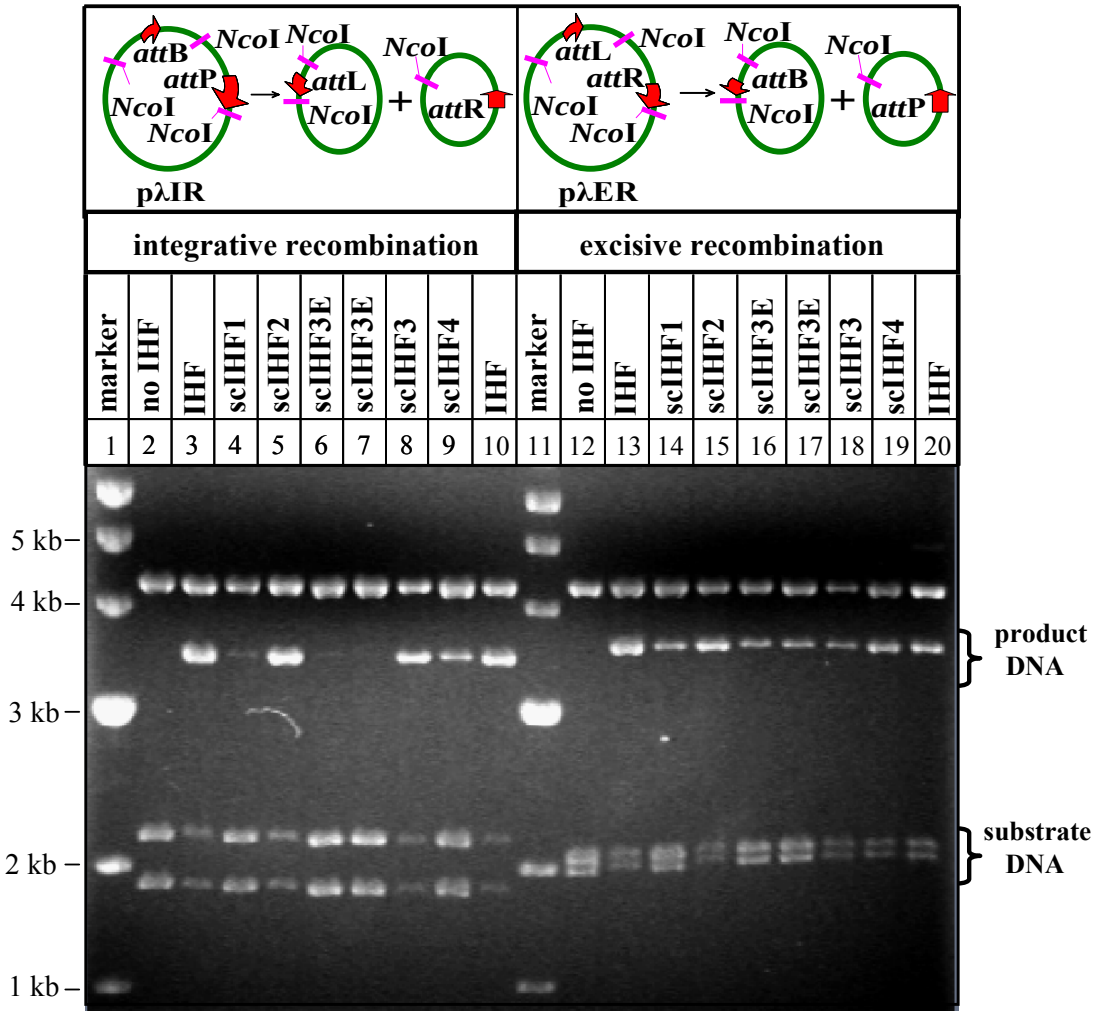


Fig.C.1.12. The recombination *in vitro* supported by scIHFs. The DNA substrate used for integrative recombination is pλIR (lane 2-10), while pλER is the plasmid for the excisive recombination (lane 12-20). Each reaction contains 84 fmol *att* sites and a 25-, 500-, 75-, 500-, 1000- or 200-fold molar excess of IHF, scIHF1, scIHF2, scIHF3, scIHF3E or scIHF4, respectively over substrate DNA and 1 pmol Int (plus 3 pmol Xis in excisive recombination). For scIHF3E, a 1000-fold stoichiometry of protein to DNA was applied in addition to 500 fold (lane 6 and 16). Lane 10 and 20 are the reactions with IHF in the presence of an additional amount of glycerol to match reaction conditions applied for scIHF3E. The recombination products (depicted on the two top panels) can be detected by restriction analysis of *Nco*I. The predicted digested fragment sizes are 4.3 kb, 2.2 kb and 1.9 kb for integration substrate; 4.3 kb, 3.4 kb, 700 bp for integration product; 4.3 kb, 2.1 kb and 2.0 kb for excision substrate; 4.3 kb, 3.5 kb and 600 bp for excision product. Note the top DNA band represents both substrate and product DNA. DNA length marker were run in lane 1 and 11

C.1.3.2. scIHF variants as a cofactor for the initiation of pSC101 replication in *E. coli*

As a DNA architectural protein, IHF plays an essential role in the initiation of DNA replication at the pSC101 origin. *E. coli* cells lacking functional IHF subunits cannot be used for the propagation of plasmids bearing this origin. In order to further analyze the biological activities of scIHF variants, we tested whether they support replication of plasmid SG86 that contains a minimal origin derived from pSC101 (Hashimoto-Gotoh et al., 1981). Three independent experiments employing *E. coli* strain CSH26ΔIHF revealed that co-transformation of SG86 with expression vectors for either scIHF2 or scIHF3 led to significant numbers of stably transformed colonies, and that the plasmid vectors could be recovered from liquid cultures. However, co-transformation with expression vectors for scIHF1 or scIHF4 resulted in only few colonies, and expression of scIHF3E gave no viable transformants at all (Fig.C.1.13). Importantly, all controls were unable to grow under the same conditions (Fig.C.1.13) and Western Blot analysis revealed that all five scIHFs were expressed in CSH26ΔIHF at almost the same level (Fig.C.1.14).

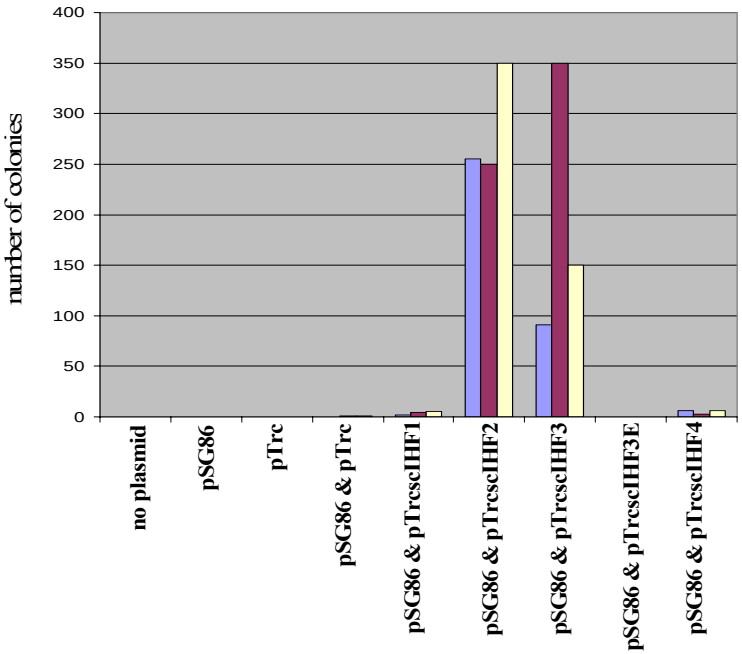


Fig.C.1.13. Replication assay *in vivo*. Plasmid SG86 containing a functional pSC101 origin of replication which has one IHF binding site was co-transformed into CSH26-ΔIHFs with an expression vector for the respective scIHF. Cells were plated on LB-agar containing various concentrations of antibiotics plus IPTG (see materials and methods). Colonies from three different experiments were counted after 24 hrs and the results depicted here as differently colored columns in the graph.

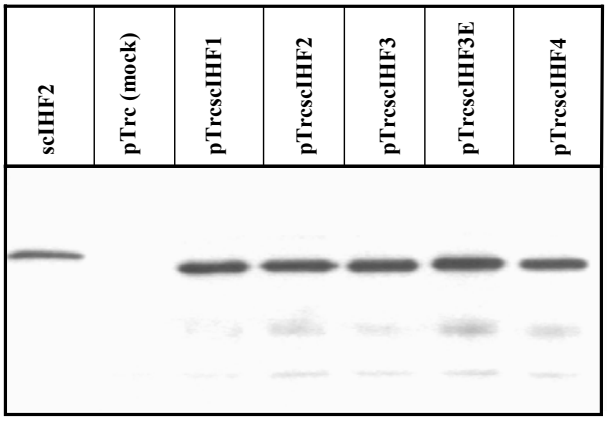


Fig.C.1.14. Expression of scIHFs in CSH26-ΔIHF. Expression vectors for scIHFs (pTrcscIHF) or a mock control (pTrc) were transformed into *E. coli* strain CSH26-ΔIHF; individual colonies were then inoculated into liquid LB media. The crude extracts prepared for western analysis after 24 hrs. Note that equal amounts of total protein from each extract were loaded. Purified scIHF2 (1 ng) served as a control (left lane).

C.2. Further elucidation of biological properties

C.2.1. scIHF2 and its derivatives

scIHF3E showed different biochemical properties and biological functions compared to other scIHFs. For example, scIHF3E has a 1000 fold larger K_d value for H' than scIHF2. scIHF3E showed supershifts in EMSA, extremely weak activities in integrative recombination assays and no activity in replication assays. To make sure these properties of scIHF3E are caused by the mutation of K45E in the α subunit or by a cooperative effect of mutation and different linker length, we introduced the K45αE mutation into scIHF2 to generate scIHF2E. The crystal structure of IHF-H' showed αK45 interacts with the backbone of the DNA through a hydrogen bond (Fig.C.2.1A). Its corresponding amino acid, glutamic acid, in the β subunit at position 44 forms a salt bridge with βR46. This interaction positions βR46 forming hydrogen bond with O2 of T (Fig.C.2.1B). αS47, the corresponding aa of βR46, makes water-mediated contacts to the backbone of the minor groove of the A-tract. If αS47 was mutated to αR47 in scIHF2E, the structure in the mutated area of the α subunit may become similar to the β subunit. In order to investigate this possibility, we also constructed scIHF2ER. The conserved amino acid prolines at position 65 in α subunit and position 64 in β subunit are quite important for the interaction between IHF and DNA consensus sequences. They intercalate into the minor groove of DNA to form two kinks. We mutated these two prolines into glycines to generate scIHF2mut and deduced this mutant protein would lose the biochemical and biological properties of its parental protein, scIHF2. All protein sequences of scIHF2s are shown in Fig.C.2.2.

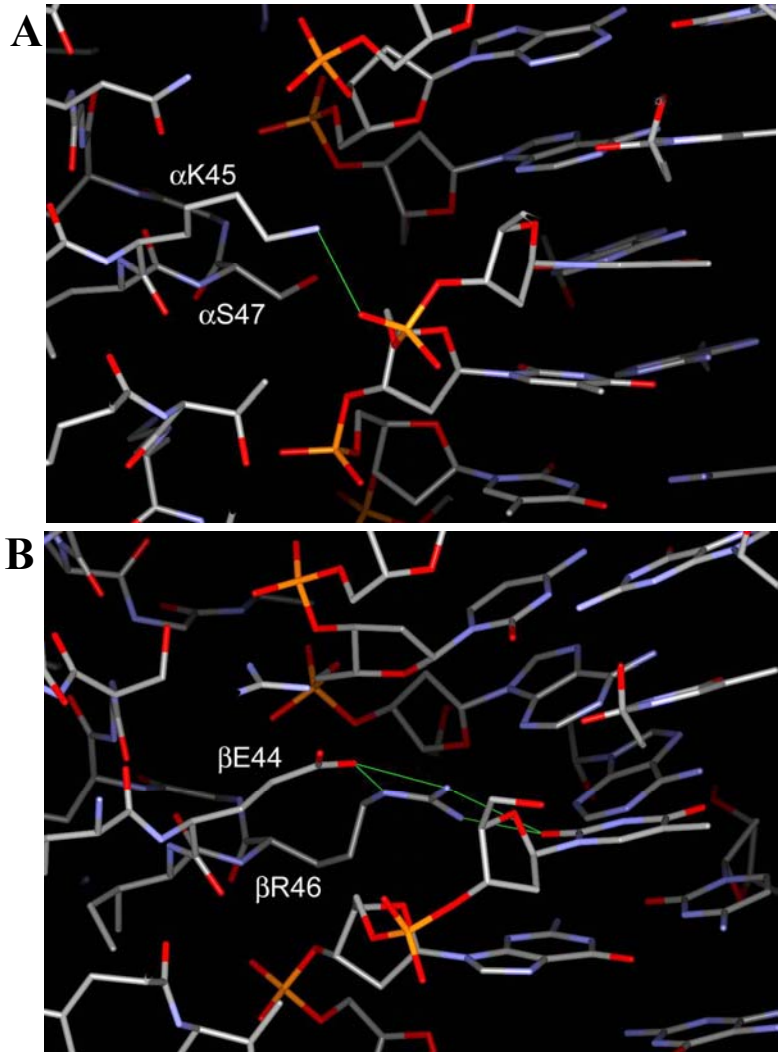


Fig.C.2.1. The interaction between IHF and H' DNA. **A.** The interaction between αK45, αS47 and backbone of H'. αK45 formed hydrogen bond with the DNA backbone. **B.** The corresponding amino acid of αK45 in β subunit βE44 forms salt bridges with βR46 (the corresponding aa of α47R) to position βR46 forming hydrogen bond with the O2 of T (the element of TTR) (**kindly provided by Curt A. Davey**)

```
  1  MASTK SELIE RLATQ QSHIP AKTVE DAVKE MLEHM ASTLA
          linker 1
 41  QGGSG GLTKA EMSEY LFDKL GLSKR DAKEL VELFF EEIRR
 81  ALENG EQVKL SGFGN FDLRD KNQRP GRNPK TGEDI PITAR
                                  linker 2
121  RVVTF RPGQK LKSRV ENAGG GERIE IRGFG SFSLH YRAPR
161  TGRNP KTGDK VELEG KYVPH FKPGK ELRDR ANIYG GSGHH
201  HHHH
```

scIHF2: linker 1 + linker 2

scIHF2E: **K**45α ⟶ **E**

scIHF2ER:**K**45α ⟶ **E** ; S47α ⟶ **R**

scIHF2mut: ***P***65α ⟶ G ; ***P***64β ⟶ G

Fig.C.2.2. Amino acid sequence of scIHF2 derivatives. The respective peptide linkers are bracketed and labeled. The sequence of the α domain within scIHF2s is marked in red, and the lysine at position 45 is highlighted in bold and substituted by a glutamate in scIHF2E. The corresponding glutamate at position 44 of the β domain is also highlighted in bold. In addition to K →E, another mutation in scIHF2ER is serine → arginine. The serine in the α subunit at position 47 and its corresponding aa in β subunit arginine were highlighted with underline and bold. The prolines at position 65 of α subunit and position 64 of β subunit were mutated into glycines in scIHF2mut and highlighted with italic as well as bold. The mutations on scIHF2 were listed below the sequence.

C.2.2. Purification and biochemical properties of scIHF2E, scIHF2ER and scIHF2mut

Purification and DNA binding

The His-tag of scIHF2s simplifies purification (Bao et al., 2004). The purity of each protein reaches more than 95% after affinity purification (Fig.C.2.3). Next we measured the K_d values of scIHF2E, scIHF2ER, scIHF2mut with H', H1 and H2 as DNA substrates. This binding assay was performed under low salt and high salt

conditions, as described before. scIHF2E, unlike scIHF3E, did not exhibit the huge difference from scIHF2 under high salt conditions with H' as the DNA substrate. Under low salt condition, both scIHF2E and scIHF2ER have strong binding affinities to DNA consensus sequences with the same K_d values as scIHF2. A summary of all K_d values is presented in Table C.2.1.

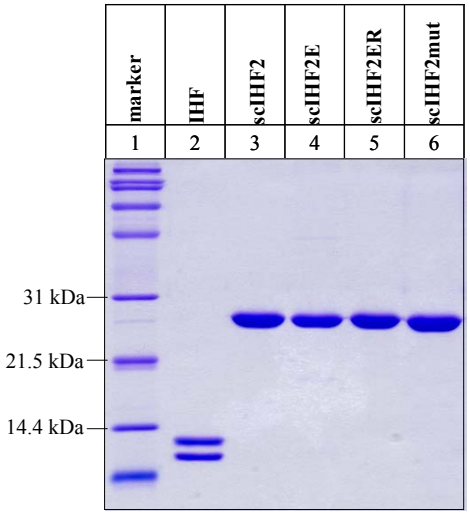


Fig.C.2.3. **The purified scIHF2 and its derivatives**. Purified IHF and scIHF2 derivatives (3 μg of each) were analyzed on 15% SDS-PAGE (lane 2-6). Protein marker was loaded in lane 1. The gel was stained by coomassie bright blue.

DNA / IHF / K_d (nM)	H'		H1		H2	
	0.5 X TBE	100 mM NaCl	0.5 X TBE	100 mM NaCl	0.5 X TBE	100 mM NaCl
scIHF2	0.79	3.8	0.71	7.04	0.51	1.8
scIHF2E	1.6	35	0.38	3	0.71	71
scIHF2ER	0.52	50	0.22	4.3	0.42	40.1

Table.C.2.1. DNA binding constant of scIHF2 and its derivatives with different consensus sequences. H', H1 and H2 are three IHF binding sites. The measurements were done under two different conditions: low salt and high salt. The K_d values are mean values obtained from at least 2 different protein titration experiments

scIHF2mut has very weak binding affinity to DNA substrates under low salt conditions. The examples presented here was binding between scIHF2mut and H1. Fig.C.2.4 shows that scIHF2mut behaves differently from scIHF2 and wild type IHF. scIHF2mut-H1 formed two different complexes, while the density of these two bands does not increase when the ratio of scIHF2mut to DNA substrate is increased. In addition to these two bands, scIHF2mut can bind to H1 forming a complex shift from the free labeled H1 when 800 nM scIHF2mut was added in the reaction.

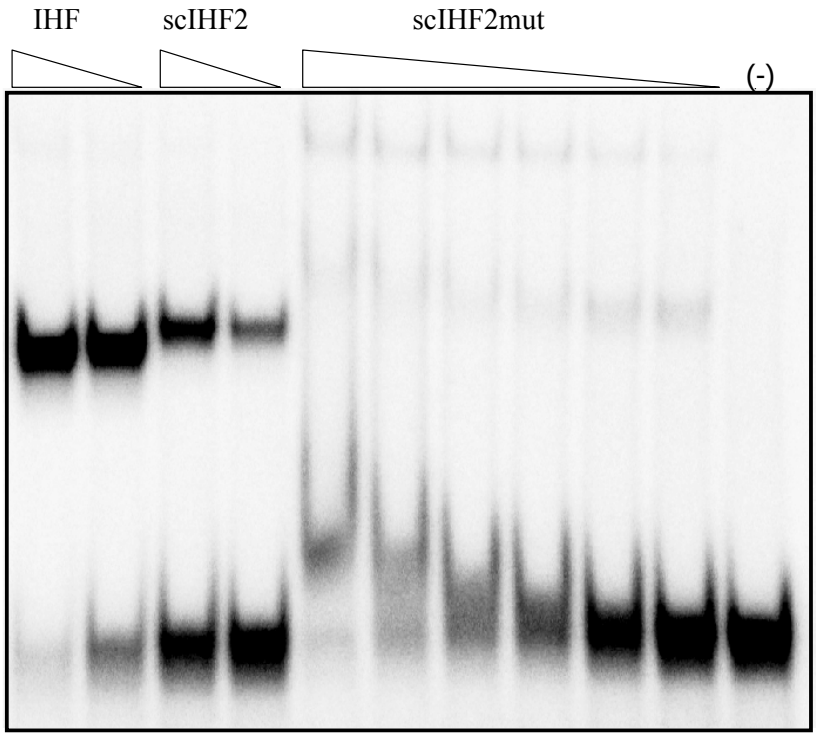


Fig.C.2.4. The binding affinity of scIHF2mut. The DNA substrate was 0.1 nM radiolabeled H1. IHF had 10 fold and 5 fold molar excess over H1. scIHF2 had 30 fold and 15 fold molar excess over H1. The stoichiometry of scIHF2mut to H1 was titrated from 8000 to 250. The last lane is free H1 as negative control.

DNA bending introduced by scIHF2s with H' as DNA substrate

Next we analyzed scIHF2E's and scIHF2ER's DNA-bending capacity and compared them with scIHF2. We employed DNA gel shift assays using radio-labeled H' site under two different conditions: low salt and high salt. Similar to scIHF3E, scIHF2E shows supershifts if compared with scIHF2. The mobility of scIHF2ER-H' complex is between that of scIHF2E-H' and scIHF2-H' complexes (lanes 3, 8 in Fig.C.2.5). The shift of each protein-DNA complex is not affected by salt (left and right panel in Fig.C.2.5). Since the MWs of scIHF2, scIHF2E and scIHF2ER are the same and the length of DNA substrate is short, the mobility difference in the gel shift assay most likely due to different DNA bending angle introduced by scIHF2s. H1 and H2 were also used in this investigation and showed the same result (data not shown).

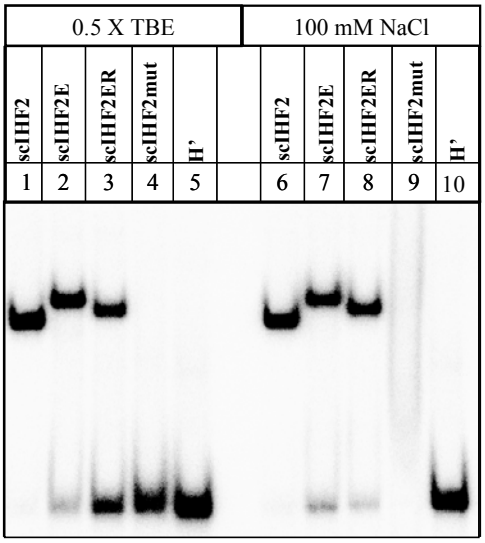


Fig.C.2.5. Analysis of DNA-binding and –bending. The representative DNA substrate here used is a radiolabeled oligonucleotide comprising the H' site (0.1 nM). The reaction was done under different conditions: 0.5xTBE (low salt) and 100 mM NaCl (high salt). The molar excess of scIHF2,

scIHF2E, scIHF2ER and scIHF2mut over H' was 10, 20, 10 and 500 under low salt condition; 50, 500, 600 and 5000 in high salt buffer.

Binding and bending to supercoiled *att*P

pAttPpuro-CMVRed is a plasmid containing *att*P with three IHF binding sites H', H1 and H2. Using this plasmid as DNA substrate, we performed a DNA-binding assay. Plasmid DNA was first reacted with protein and then digested with restriction enzymes which recognize sites that flank *att*P. However, we couldn't exclude the rebinding of scIHF2s to linear fragment. From Fig.C.2.6, one can see that scIHF2 forms three complexes with *att*P (lane 3) when compared with the control (lane 2). This is expected because *att*P contains three specific binding sites for IHF. The most retarded band corresponds to the complex in which all three sites were bound by scIHF2. scIHF2E gave only one clear extra band with increased mobility compared to the correspondent band of scIHF2 (lane 4).

Compared with the negative control, scIHF2ER showed two extra bands which present one and two of three IHF binding sites occupied by the protein. The band which represents one site on *att*P bound by scIHF2ER migrates faster than that formed with scIHF2, but slightly slower than the one of scIHF2E. The second extra band of scIHF2ER migrated also faster than the one of scIHF2. The complex with all three sites occupied by scIHF2ER did not exist as one sharp band but rather as a smear (lane 5). scIHF2mut did not give raise to any shift with *att*P (lane 6).

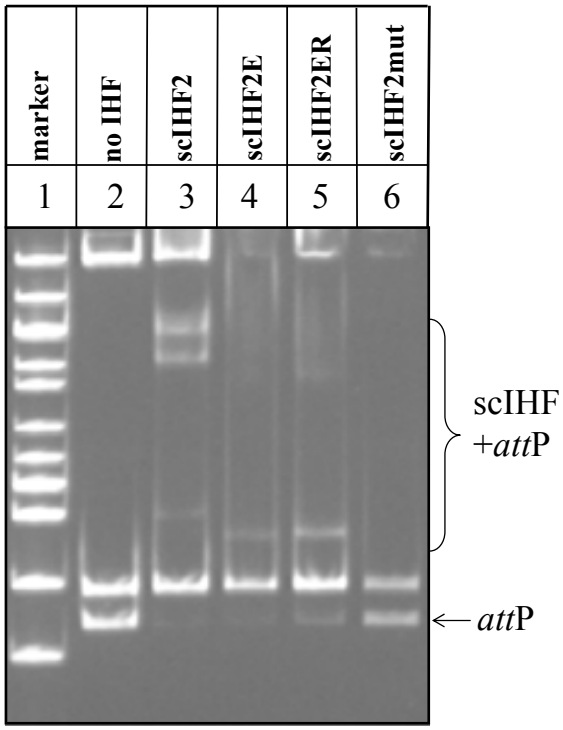


Fig.C.2.6. The bending of supercoiled *att*P introduced by scIHF2 and its derivatives. The *att*p is a 273 bp fragment in plasmid pAttPpuro-CMVRed flanked with *Eco*RI and *Pst*I. The amount of *att*P is 272 fmol. The stoichiometry of scIHF2 to plasmid is 90 to 1 (lane 3). For scIHF2E and scIHF2ER, the ratio is 400 to 1(lane 4 and 5). scIHF2mut has 1000-fold molar excess over substrate DNA (lane 6). Lane 1 is the plasmid pAttPpuro-CMVRed as control. After the plasmid was incubated with scIHF2s 40 min, it was digested by *Eco*RI and *Pst*I for 10 min and analyzed in polyacrylamide gels. Note 1kb DNA ladder was run in lane 1.

The DNA bending assay with supercoiled or linear *att*R as DNA substrate

*Att*R is the attachment site of λ integrase for excicive recombination, containing two IHF consensus sequences H1 and H2 which is separated by P2 (Int binding site), X1 and X2 (Xis binding sites). The binding of scIHF2s retards *att*R in EMSA no matter whether *att*R is post-cleaved (supercoiled) or pre-cleaved (linear) (Fig.C.2.7). Obviously there are two populations of *att*R-scIHF2s complexes: I and

II. And *att*R-scIHF2 complexes move the slowest among *att*R-scIHF2s, while *att*R-scIHF2E complexes move faster than *att*R-scIHF2ER. Further, these differences are detectable regardless of whether scIHF2s is first bound to supercoiled *att*R followed by restriction digest, or vice versa (left and right panel, respectively). These results reveal that there must be significant conformational differences between scIHF2E-*att*R complexes and those formed by scIHF2 and *att*R as well as scIHF2ER-*att*R. And there is no difference between the results with pre-cleaved *att*R as DNA substrate and those with post-cleaved *att*R as DNA substrate.

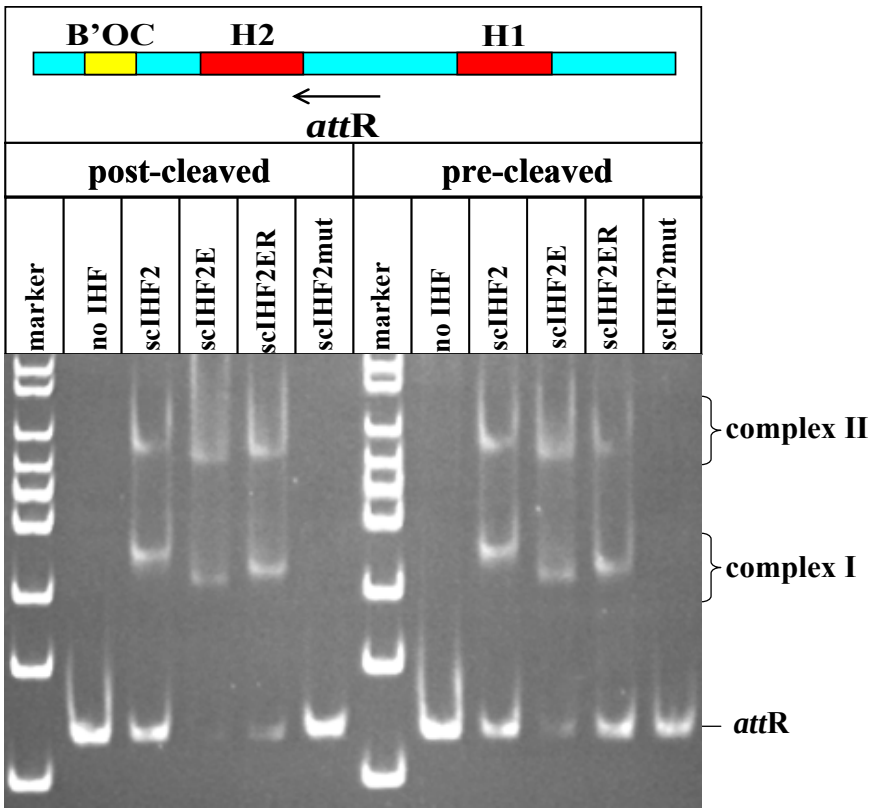


Fig.C.2.7. Analysis of *att*R bending introduced by scIHF2s. Upper box showed the relative positions of H1 and H2 on *att*R. The left part of the gel represents supercoiled *att*R reacted with scIHF2s first, then it was digested before analysis on agarose gel. The right part shows the result of *Eco*RI-precleaved *att*R reacting with scIHF2s then digested with *Hind*III and *Not*I. The

concentration of DNA substrate is 13.5 nM. The stoichiometry of scIHF2 to plasmid is 90 to 1. For scIHF2E and scIHF2ER, the ratio is 400 to 1. scIHF2mut has 1000-fold molar excess over substrate DNA (lane 6). The bottom band presented free *att*R. The upper two bands are *att*R-scIHF2s complexes.

C.2.3. Recombination assays *in vitro*

Recombination assay with supercoiled pλIR and pλER as DNA substrates

As shown above (Table.C.2.1) the DNA binding affinity of scIHF2E and scIHF2ER were ten-fold lower compared with scIHF2 in high salt buffer. scIHF2E showed a supershift with short IHF binding sites as DNA substrates in EMSAs. The shift difference between scIHF2ER and scIHF2 can be detected in the gel shift assay, although it is not as pronounced as that between scIHF2E and scIHF2. *Att*P or *att*R bound with scIHF2, scIHF2E or scIHF2ER moved to different positions in the polyacrylamide gel. These indicated that scIHF2E and scIHF2ER formed different complexes compared with scIHF2.

In the recombination assay, supercoiled pλIR containing *att*P and *att*B was used as substrate for integrative recombination, while supercoiled pλER with *att*L and *att*R was the substrate for excisive recombination. We checked the activities of scIHF2E, scIHF2ER and scIHF2mut. Fig.C.2.8 showed that scIHF2E had only weak activity in integrative recombination (lane 3) but supported excision (lane 9). scIHF2ER supported integration and excision to an extent comparable with scIHF2 (lane 2, 4, 8, 10). As expected, scIHF2mut was inactive in both assays (lane 5 and 11).

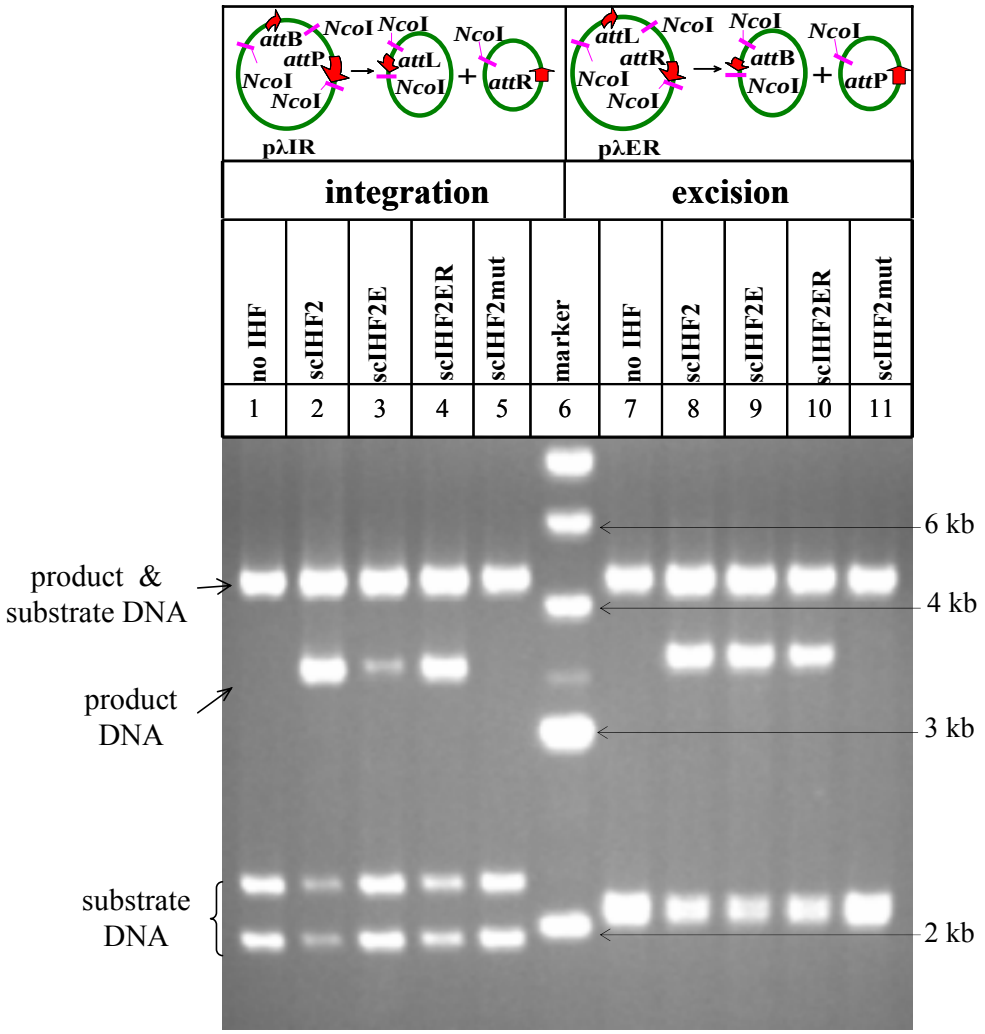


Fig.C.2.8. *In vitro* recombination of scIHF2 derivatives. pλIR is the integrative recombination substrate DNA (lane 1-5), while pλER is for excisive recombination (lane 7-11). In addition to scIHF2s, Int was needed in both recombination cases, while Xis is another cofactor used in excisive recombination. The concentration of DNA substrates is 3.5 nM. The stoichiometry of IHF to DNA substrate is 90 to 1 for scIHF2 (lane 2, 8), 400 to 1 for scIHF2E (lane 3, 9) and scIHF2ER (lane 4, 10), 1000 to 1 for scIHF2mut (lane 3, 11). Recombination product was analyzed on 0.8% agarose gel after *Nco*I digestion. The predicted digested fragment sizes are 4.3 kb, 2.2 kb and 1.9 kb for integration substrate; 4.3 kb, 3.4 kb, 700 bp for integration product; 4.3 kb, 2.1 kb and 2.0 kb for excision substrate; 4.3 kb, 3.5 kb and 600 bp for excision product. Note the reactions without IHF were control (lane 1, 7). 1 kb DNA ladder was loaded in lane 6.

Comparing excisive recombination on supercoiled and linear pλER substrates

The activity of scIHF2E in integrative recombination is much weaker than that of scIHF2ER, although their K_d values are comparable (Table.C.2.1). However, both of them are as active as scIHF2 in excisive recombination. The question then arises, what is the phenotype of scIHF2E and scIHF2ER if the concentrations of these two proteins are decreased in excisive recombination reactions?

Except for scIHF2mut, scIHF2E and scIHF2ER support excisive recombination as scIHF2 with supercoiled pλER as DNA substrate (Fig.C.2.9.A). No significant difference between scIHF2E and scIHF2ER can be detected, even after titration of proteins (lane 5-8, 10-14). Recombination of linear pλER produces a 4859 bp circular plasmid and a 3521 bp linear fragment, which can be detected in recombination reactions employing wild type IHF and scIHF2 (Fig.C.2.9.B). The circular 4.86 kb plasmid is split into at least 6 different topoisomers (lane 2 and 3 in Fig.C.2.9.B). In addition to these two recombination products, there is one extra band. According to the fragment size, we deduce this is an intermolecular excisive recombination product of 13.3 kb. The corresponding other intermolecular recombination product is 3.5 kb fragment, and thus similar to the one produced by the intra-molecular recombination reaction. In any case, scIHF2E, scIHF2ER and scIHF2mut do not give any recombinant product at all with linear DNA substrate (lanes 5-13 in Fig.C.2.9.B).

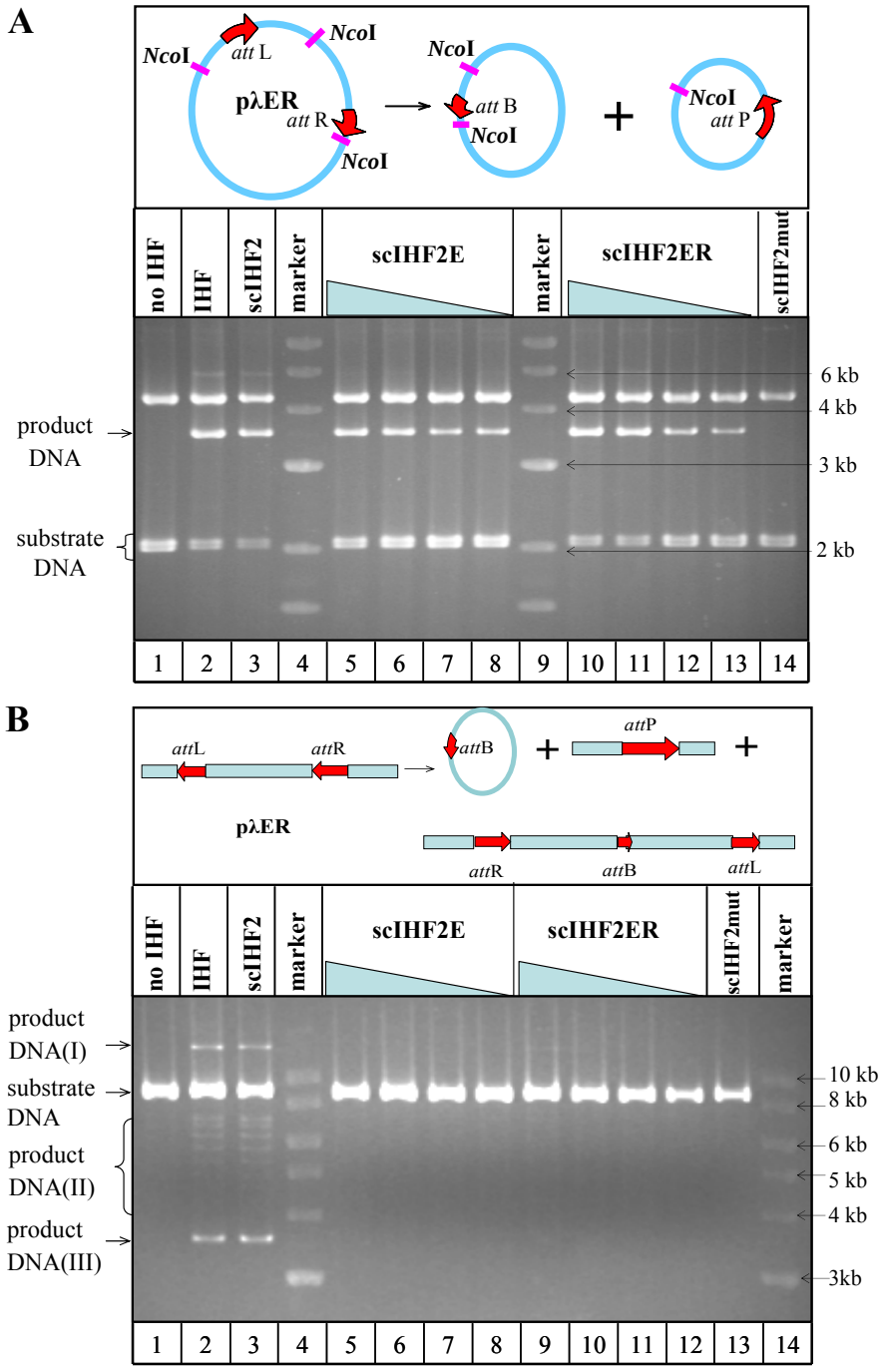


Fig.C.2.9. The excisive recombination *in vitro* supported by scIHF2 derivatives. **A.** The excisive recombination with 4.4 nM supercoiled pλER as DNA substrate. The recombination product was analyzed by the digestion of *Nco*I. The predicted digestion fragments sizes are 4.3 kb, 2.1 kb and 2.0 kb for excision substrate; 4.3 kb, 3.5 kb and 600 bp for excision product. **B.** The excisive recombination with 2.15 nM linear pλER as DNA substrate. The size of linear pλER is 8.4 kb. The recombination products are 4.9 kb circular DNA, 3.5 kb fragment and 13.3 kb fragment.

The top panels in (A) and (B) depict the supercoiled DNA substrate and linear DNA substrate for excisive recombination (pλER) as well as the expected products .In both cases, the reactions without IHF were control (lane 1). The stoichiometry of IHF to DNA substrate was 30 to 1 (lane 2). For scIHF2, the ratio was 90 to 1 (lane 3). For scIHF2E, the ratio was 400 to 1(line 5), 200 to 1 (line 6), 100 to 1 (line 7) and 50 to 1 (line 8). For scIHF2 ER, the ratio was 400:1, 200:1, 100:1 and 50:1 (**A.** lane 10-13; **B.** lane 9-12). For scIHF2mut, the ratio was 1000 to 1 (**A.** lane 14; **B.** lane 13**).** 1 pmol Int and 3 pmol Xis were added in both experiments. Note 1 kb DNA ladder was loaded in Lane 4, 9 (**A)** and lane 4, 14 (**B**).

Excisive recombination assays with supercoiled *att*R and linear *att*L or supercoiled *att*L and linear *att*R as DNA substrates

The results above showed that the activities of scIHF2E and scIHF2ER are supercoil-dependent in excisive recombination. But which one is more important, supercoiled *att*L or supercoiled *att*R? We therefore decided to use intermolecular recombination in order to dissect whether *att*R or *att*L, or both, require superhelical tension with scIHF2E and scIHF2ER. We applied two separate plasmids pWSRGFP containing *att*R and pCMVattL with *att*L as DNA substrates for excisive recombination (Fig.C.2.10A). With linear *att*R and supercoiled *att*L, only IHF and scIHF2 showed some activity producing one 10 kb fragment (Fig.C.2.10B, lane 5, 6). scIHF2E and scIHF2ER were inactive (Fig.C.2.10B, lane 7, 8). With supercoiled *att*R and linear *att*L, all scIHF2s supported excision except scIHF2mut. The recombination product of supercoiled *att*R and linear *att*L was also one 10 kb fragment but in this experiment, there was another 15 kb DNA product because supercoiled *att*R substrate used here was actually a mixture of monomers and dimmers(Fig.C.2.10C). Therefore, the supercoiled *att*R is critical required

compared with supercoiled *att*L in the excisive recombination supported by scIHF2E and scIHF2ER.

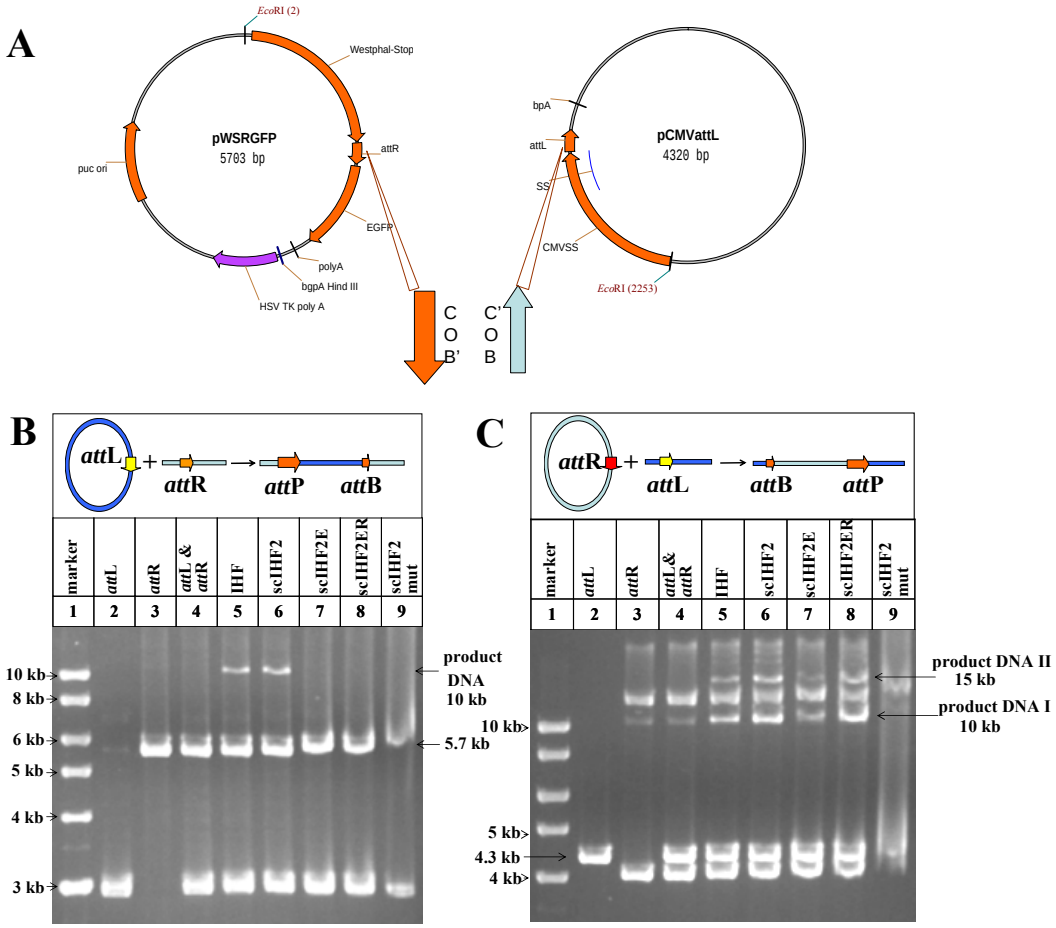


Fig.C.2.10. A. DNA substrates for excisive recombination: pWSRGFP containing *att*R and pCMVattL with *att*L. **B.** Excisive recombination with linear *att*R and supercoiled *att*L as DNA substrates. Supercoiled *att*L (lane 2), linear *att*R (lane 3), linear *att*R and supercoiled *att*L (lane 4) which were not incubated with scIHF2s were negative control. The upper bands in lane 5 and lane 6 were 10 kb recombination product. 1 kb DNA ladder was loaded in lane 1. **C.** Excisive recombination with supercoiled *att*R and linear *att*L as DNA substrates. The same as B, lane 2-4 were DNA substrates without scIHF2s. Linear *att*L was in lane 2. Supercoiled *att*R was in lane 3. Supercoiled *att*R and linear *att*L without scIHF2s were loaded in lane 4. Compared with negative control (lane 4), the extra band and brighter band in lane 5-8 represented recombinant products (10 kb and 15 kb). In both assays, the concentration of DNA substrates is 3.55 nM. The concentration of proteins is 106.5 nM for IHF; 319.5 nM for scIHF2; 1.42 μM for scIHF2E and scIHF2ER; 3.55 μM for scIHF2mut. The recombination products of these two assays are depicted on the top panels of **B** and **C**

C.2.4. Crystal structures of scIHF2-H' and scIHF2E-H' complexes

(contributed by Prof. Curt A. Davey)

The supershift in EMSA, weak activity in integrative recombination as well as a dependence on supercoiled *att*R in excisive recombination revealed that scIHF2E interacts differently with DNA. Because αK45 which interacts with the backbone of DNA in the IHF-DNA complex, we reasoned that the repulsive force between Glutamic acid and phosphate of DNA would push the DNA away from scIHF2E. In order to analyze this possibility, we solved the scIHF2-H' and scIHF2E-H' co-crystal structures.

The construct used to obtain crystals of the wild type IHF-DNA complex was composed of a 35 bp DNA containing a nick in one strand adjacent to the proline intercalation site of the β-subunit (Rice et al., 1996). The presence of the single-strand break was necessary for obtaining well diffracting crystals, consistent with its involvement in an extensive interparticle contact in the crystal. It was shown before, that the single-strand nick at this location does not influence the overall DNA bending angle (Lorenz et al., 1999).

We utilized the same DNA construct and similar crystallization conditions as Rice and co-workers to solve the crystal structure of the scIHF2-DNA complex (Fig.C.2.11A; Table C.2.2). With the exception of amino acids flanking the N- and C-termini and linker regions where IHF and scIHF2 differ in primary sequence, the structures of the two complexes appear to be nearly identical (Fig.C.2.11B). Least-

squares superposition of the IHF-DNA and scIHF2-DNA models (2825 atoms: entire DNA + 173 amino acids) yielded an r.m.s.d. in atomic position of only 0.67Å. Importantly, protein-DNA contacts between the two complexes appear to be nearly the same, consistent with the very similar biochemical and functional properties of IHF and scIHF2.

The temperature (B-) factors in the scIHF2-DNA complex are significantly higher for the flanking regions of the DNA compared to the 13 bp central zone, where extensive protein-base contacts between the sites of intercalation apparently reduces DNA mobility (Table C.2.2; Fig.C.2.11A). The mean B-factor for the 11.5 bp left DNA arm is slightly higher than that of the 10.5 bp right arm, which may relate to the relatively more extensive protein-DNA contacts on the right side. At the left arm, the αK45 side chain takes part in DNA binding by forming direct, as well as water-mediated, hydrogen bonds with two phosphate groups at the mouth of the minor groove (Fig.C.2.11C). This configuration is further supported through stacking of the αK45 side chain with that of αQ43, which is in turn hydrogen-bonded to that of αD53.

Least-squares superposition of the protein and DNA components of the scIHF2-DNA and the scIHF2E-DNA models gave an r.m.s.d. in atomic position of 0.32 Å. Thus, the two complexes are essentially identical, with the exception of local alterations in structure due to the presence of the glutamate in place of the αK45 side chain (Fig.C.2.11D and E). Interestingly, the orientation of the αE45 side

chain is very similar to that of αK45. Strikingly, two divalent metal binding sites appear in the scIHF2E-DNA complex, which are not observed in the scIHF2-DNA complex. A Mn^{2+} ion coordinated to the αE45 carboxylate group and the adjacent DNA phosphate group serves as a cation bridge, which apparently compensates for an otherwise unfavorable electrostatic interaction, allowing a native-like protein-bound DNA configuration to persist.

In contrast to our expectations based on EMSAs, the K45αE substitution has a more pronounced effect on local protein as compared to DNA structure. The presence of the negative-charged αE45 side chain causes repositioning of the adjacent αQ43 side chain, which engages in a hydrogen bond with the more distant peptide backbone NH group of αL54. This reorientation of the αQ43 side chain apparently opens up an additional electronegative pocket, formed by the side chains of αD53 and αE45 and the opposing DNA phosphate groups, which is occupied by a second Mn^{2+} ion. Thus, divalent metal binding to the electronegative zones proximal to the DNA phosphate backbone, which would appear to otherwise disrupt association of the left arm in the scIHF2E mutant, stabilizes DNA binding and bending. At the same time, the B-factor elevation of the left versus the right DNA arms observed in the scIHF2-DNA complex is actually three-fold greater in the scIHF2E-DNA complex (Table C.2.2). Therefore, in spite of the divalent metal-mediated stabilization of left DNA arm association, this region appears to still have increased mobility relative to the scIHF2-DNA complex.

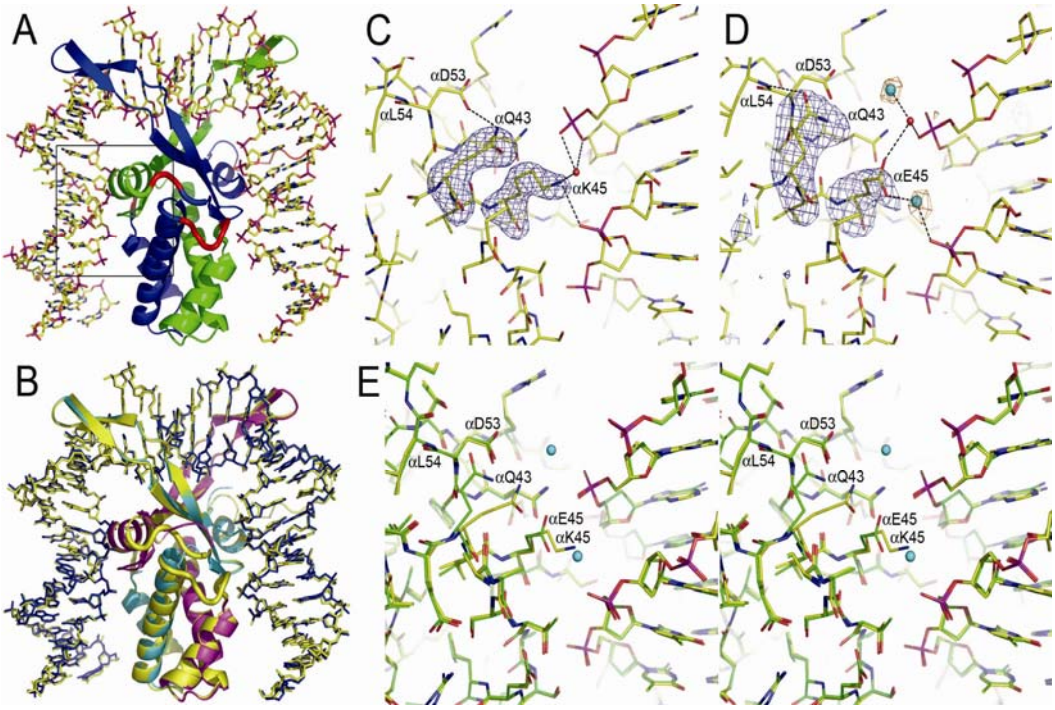


Fig.C.2.11. Protein-DNA interactions in the crystal structures of the scIHF2-DNA and scIHF2E-DNA complexes. **A.** The scIHF2-DNA complex, with protein regions corresponding to the α- and β-subunits of wild-type IHF colored green and blue, respectively. Regions differing in primary sequence with respect to IHF are colored red, including the linker amino acids (thick tubes) used to fuse the α- and β-subunits. The boxed area associated with the left DNA arm corresponds to the section viewed in C, D, and E. **B.** Superposition of the scIHF2-DNA (yellow) and IHF-DNA (α-subunit, magenta; β-subunit, cyan; DNA, blue) models. **C, D.** Simulated annealing omit maps (blue), contoured at 4σ and 3σ, superimposed on the scIHF2-DNA (C) and scIHF2E-DNA (D) models, respectively. Residues αQ43 and αK45 of scIHF2-DNA and αQ43 and αE45 of scIHF2E-DNA were omitted. An anomalous difference map (brown), contoured at 3σ, denotes the positions of Mn2+ ions (cyan spheres) in (D). Water molecules are shown as red spheres, and hydrogen and coordinate bonds are indicated by dashed lines. **E**. Structural comparison of the scIHF2-DNA (yellow) and scIHF2E-DNA (green) complexes in stereo view. Mn^{2+} ions associated with the K45αE mutant are shown as cyan spheres.

	scIHF2	scIHF2E
Data Collection		
Resolution (Å)	44.5-2.41	45.0-2.72
Resolution of last shell (Å)	2.54-2.41	2.87-2.72
Redundancy (last shell)	6.3 (3.9)	7.7 (7.9)
No. unique *hkl*	18,596	12,530
Completion % (last shell)	99.0 (94.5)	97.1 (95.9)
R_{merge} % (last shell)	6.2 (41.6)	6.5 (47.6)
Unit cell	$P2_12_12_1$	$P2_12_12_1$
a (Å)	47.37	47.24
b (Å)	55.31	54.44
c (Å)	177.83	177.87
Refinement		
Resolution (Å)	40.0-2.41	40.0-2.72
Resolution (last shell Å)	2.52-2.41	2.84-2.72
R-factor % (last shell)	23.2 (33.3)	23.7 (35.1)
Free R-factor % (last shell)	26.8 (34.7)	27.4 (39.7)
Reflections work / free	17,961 / 587	12,130 / 381
No. atoms in model		
Total (mean B-factor $Å^2$)	3,318 (48.9)	3,207 (52.5)
Protein (mean B-factor $Å^2$)	1,610 (48.2)	1,610 (53.0)
DNA (mean B-factor $Å^2$)	1,426 (50.4)	1,426 (52.8)
Left arm (mean B-factor $Å^2$)	470 (57.2)	470 (64.5)
Right arm (mean B-factor $Å^2$)	426 (54.6)	426 (56.7)
Solvent (mean B-factor $Å^2$)	282 (45.7)	171 (44.8)
r.m.s.d. from ideality		
Bond lengths (Å)	0.005	0.004
Bond angles (deg.)	1.11	0.85

Table C.2.2. Data collection and refinement statistics

C.2.5. DNA-Bending Determined by Atomic Force Microscopy

(contributed by Dr. Hu Chen, Dr. Yingjie Liu and Prof. Jie Yan)

A 623 bp *att*L-carrying DNA fragment was used for AFM imaging. The fragment harbors the H' site asymmetrically positioned 200 bp from one end. DNA was incubated either with purified wild-type IHF, scIHF2, or scIHF2E, and resulting complexes were adsorbed to mica. Naked DNA served as control. AFM images which showed both a protein signal and DNA bending at the expected DNA region were further analyzed.

None of the more than one hundred images inspected in the control sample with naked DNA show significant DNA bending at a corresponding position (not shown). Our analysis of AFM images from wild-type IHF-DNA and scIHF2-DNA complexes reveal mean bending angles of 117 (+/-19)^{o} and 114 (+/-15)^{o}, respectively (Fig.C.2.12.A and B). These data are in very good agreement with those from a recent AFM study performed with wild-type IHF on a segment from the TOL plasmid (Seong et al., 2002). Our analysis using scIHF2E, however, yield a distribution of bending angles with a significantly smaller mean value of 91 (+/-19)^{o} (Fig.C.2.12.C). Together, these results support our hypothesis that the degree of overall DNA bending in scIHF2E-DNA complexes is significantly reduced, most likely due to the K45αE substitution which weakened protein-DNA interactions with the left DNA arm.

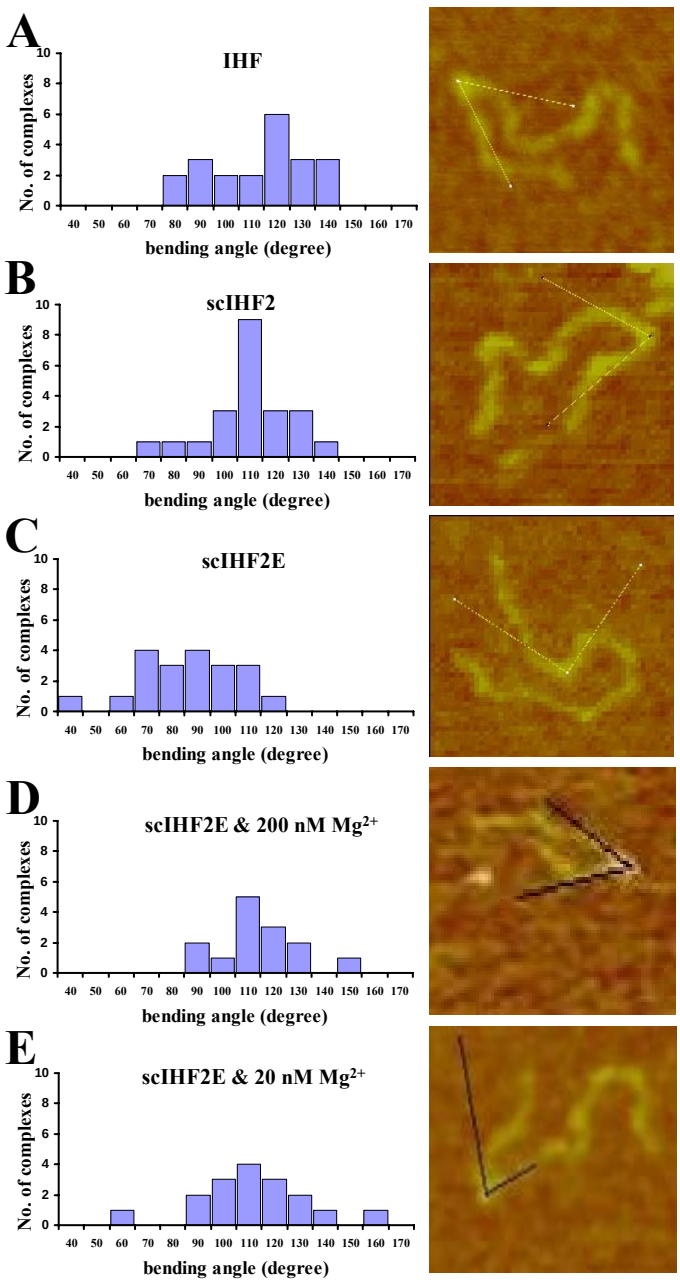


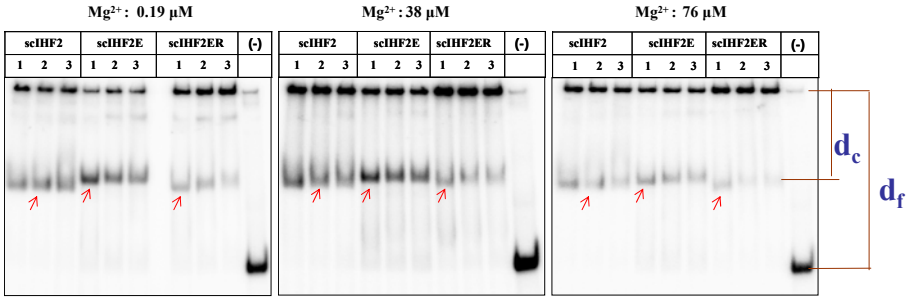
Fig.C.2.12. AFM analysis. About 20 high-resolution images of *att*L-protein complexes obtained with wild-type IHF (A), scIHF2 (B), and scIHF2E (C) were analyzed and the degree of DNA bending determined. In (D) and (E), the concentration of magnesium ions present in the binding reactions with scIHF2E is indicated. One representative example for each protein-*att*L complex is presented with tangents used for the measurements indicated. Images are 100 nm x 100 nm in size.

C.2.6. Reversal of supershift with divalent ions in EMSAs

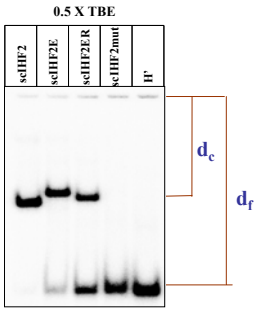
Our structural analysis in conjunction with biochemical data indicates that a scIHF2E-DNA complex can adopt two stable conformational states: One is an "open" state in which the left DNA arm is mostly detached from the protein body, thus leading to a significantly smaller degree of overall DNA bending. In the closed state, two divalent metal ions stabilize interactions between this region and the protein body, which result in the more severe DNA-bending observed in co-crystals. However, the H'-DNA used in our crystallographic studies contained a nick in one phosphodiester backbone near the site of proline intercalation. Therefore, in order to establish that scIHF2E can adopt a closed conformational state also with covalently closed H'-DNA, we incubated scIHF2, scIHF2E and scIHF2ER with H'-DNA in the presence of magnesium ions and analyzed complex formation through EMSA.

The presence of Mg^{2+} reverses the supershifts of scIHF2E as well as scIHF2ER in EMSAs (in TBM buffer) (Fig.C.2.13). With 0.19 μM Mg^{2+}, the mobility of scIHF2ER-DNA was almost the same as scIHF2 complexed with H' DNA and the complex of scIHF2E-H' moved closer to scIHF2-H'. When the concentration of Mg^{2+} ions increased to 38 μM or 76 μM, scIHF2ER-H' ran even faster than scIHF2-H', the position of scIHF2E-H' was closer and closer to that of scIHF2-H' with the relative Rf increasing from 0.908 to 0.981 compared with the shift in TBE buffer. (Fig.C.2.13)

A



B



$Rf = d_c / d_f$

C

IHF-H' / Rf / buffer		scIHF2-H'	scIHF2E-H'	scIHF2ER-H'
0.5 X TBE	exp1	0.545	0.495	0.520
	exp2	0.560	0.501	0.530
0.5 X TBM	Mg^{2+}/ 0.19 µM	0.540	0.509	0.540
	Mg^{2+}/ 38 µM	0.514	0.504	0.522
	Mg^{2+}/ 76 µM	0.518	0.502	0.524

IHF-H' / Relative Rf / buffer		scIHF2-H'	scIHF2E-H'	scIHF2ER-H'
0.5 X TBE	exp1	1	0.908	0.954
	exp2	1	0.895	0.946
0.5 X TBM	Mg^{2+}/ 0.19 µM	1	0.943	1
	Mg^{2+}/ 38 µM	1	0.981	1.016
	Mg^{2+}/ 76 µM	1	0.969	1.012

Fig.C.2.13. The effect of divalent cations on scIHF2E-H' and scIHF2ER-H' complexes. A. Gel shift assay of scIHF2s with Mg^{2+} present in the buffer (TBM). The concentration of H' is 0.1 nM.

The concentration of scIHF2 is 10 nM, 20 nM , and 30 nM. While for scIHF2E and scIHF2ER, the concentration of protein is 20 nM, 40 nM and 60 nM (lane 1, 2, 3). The lanes highlighted with red arrows were chosen for Rf analysis because the same concentration of proteins were used in these samples. **B.** One of two separate experiments of gel shift assay with 0.5xTBE buffer. **C.** Rf values of scIHF2s in TBE and TBM buffer. The upper table showed original Rf data according to the gels in **A** and **B**. The bottom table gave the relative Rf with Rf of scIHF2-H' as standard.

C.2.7. Increase of scIHF2E' activity with Mg^{2+} in the integrative recombination assay

Since Mg^{2+} can reverse the supershift of scIHF2E-H' complexes, we wondered if the existence of Mg^{2+} can promote scIHF2E to support integrative recombination more efficiently. We therefore performed integrative recombination assays with different concentration of Mg^{2+} present in the reaction buffer. For scIHF2E, the recombinant efficiency increased almost two times when the concentration of Mg^{2+} was 0.1 μM (Fig.C.2.14.A, lane 10). The increase of scIHF2E's activity is sensitive to the concentration of Mg^{2+}. When the concentration of Mg^{2+} was increased further, the integration efficiency supported by scIHF2E decreased, but was still higher than that under the condition without Mg^{2+}. As a control, scIHF2 promoted the integration in the same level with different concentration of Mg^{2+} in the buffer (Fig.C.2.14.A. lane 3-8). Interestingly, this metal ion-mediated activation was restricted to integrative recombination since we could not observe an effect of metal ions on excisive recombination with topologically relaxed pλER (data not shown).

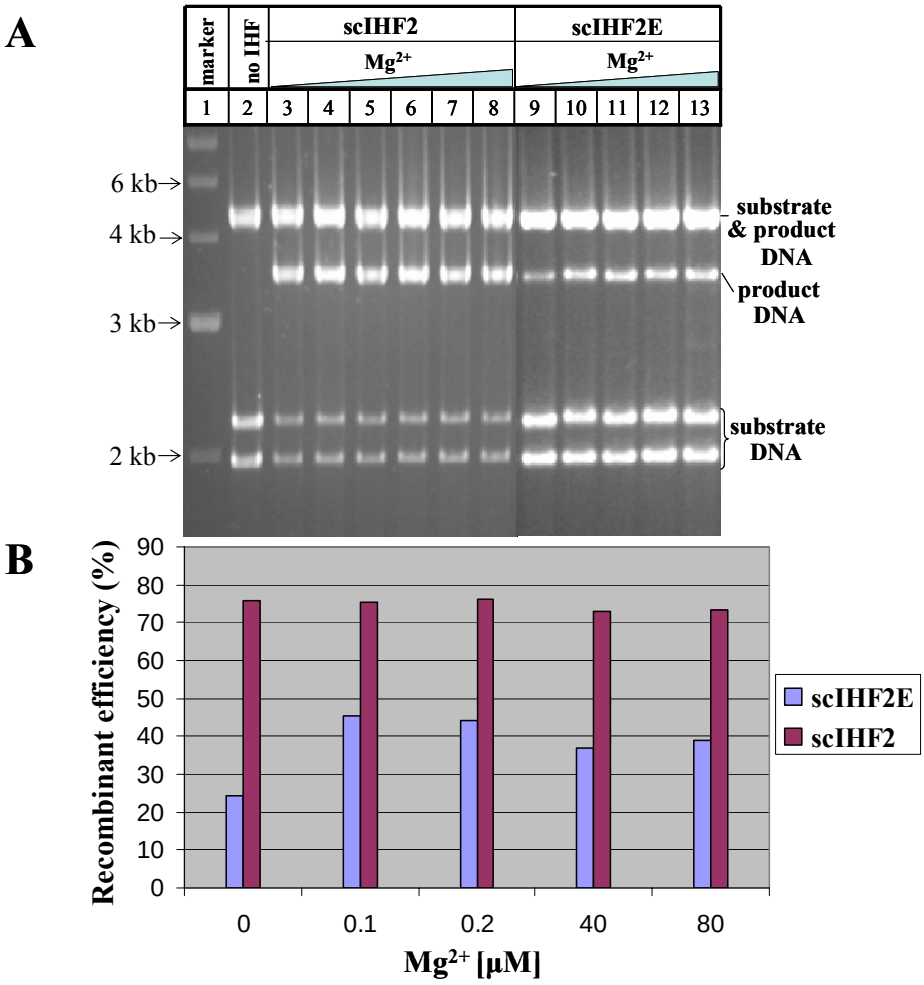


Fig.C.2.14. Integrative recombination assay with Mg^{2+} in the reaction. **A** Integrative recombination assay of scIHF2 and scIHF2E in the buffer with Mg^{2+}. Each reaction contains 0.1 nM DNA substrate (pλIR), 3 nM scIHF2 or 40 nM scIHF2E as well as Int. The concentration of Mg^{2+} was 0, 0.05, 0.1, 0.2, 40, 80 μM from lane 3 to lane 8. Lane 2 contained the negative control (no IHF). For scIHF2E, the concentration of Mg^{2+} varied from 0 to 80 μM, the same as scIHF2 but without 0.05 μM (lane 9-13). The recombination products can be detected by restriction analysis of *Nco*I. The predicted digested fragment sizes are 4.3 kb, 2.2 kb and 1.9 kb for integration substrates; 4.3 kb, 3.4 kb, 700 bp for integration products; **B.** Chart of recombinant efficiency. Recombination efficiency (%) = (the signal of 3.4 kb product DNA/ the signal of 4.3 kb substrate & product DNA) x 100%. The purple columns represented scIHF2. The blue columns were scIHF2E.

C.3. Analysis of scIHF2 in mammalian cells

C.3.1. Difference between scIHF2 and scIHF2mut cell lines

We established so far that scIHF2 is functional as an essential co-factor in two different DNA transactions in *E. coli* and in recombination *in vitro*, and detected that scIHF2mut lost the binding ability to IHF consensus sequences and is thus inactive in recombination assays *in vitro*. We decided to generate stable scIHF2-HeLa transgenic cell lines in preparation for functional tests in mammalian cells with scIHF2mut-HeLa cell lines as control. These cell lines were characterized first by PCR and DNA sequencing (data not shown). Western analysis performed with crude cell extracts prepared from transgenic cells reveals that the expression levels of scIHF2mut in H/IHFmut1, H/IHFmut2, H/IHFmut3 and H/IHFmut7 are the same as that of scIHF2 in H/IHF17 as well as H/IHF19. No major proteolytic degradation products detectable (Fig.C.3.1). These results imply that the recombinant protein appears to be tolerated by mammalian cells kept in culture. However, the high level expression of scIHF2 or scIHF2mut affected the morphology of the cells. The examples shown here are H/IHF17 and H/IHFmut3, which show similar expression levels of scIHF2 and scIHF2mut, respectively. Compared with parental cells Hela, H/IHF17 grows in tight clumps with regular edges and the cells are much smaller. Each cell of H/IHFmut3 is bigger than that of H/IHF17, but smaller than Hela. The growth patten of H/IHFmut3 is also in clumps, but the clump is smaller and looser than that of H/IHF17 (Fig.C.3.2). The other cell lines with the same expression level of scIHF2 or scIHF2mut as H/IHF17 or H/IHFmut3 shows similar morphology as the examples here (data not shown).

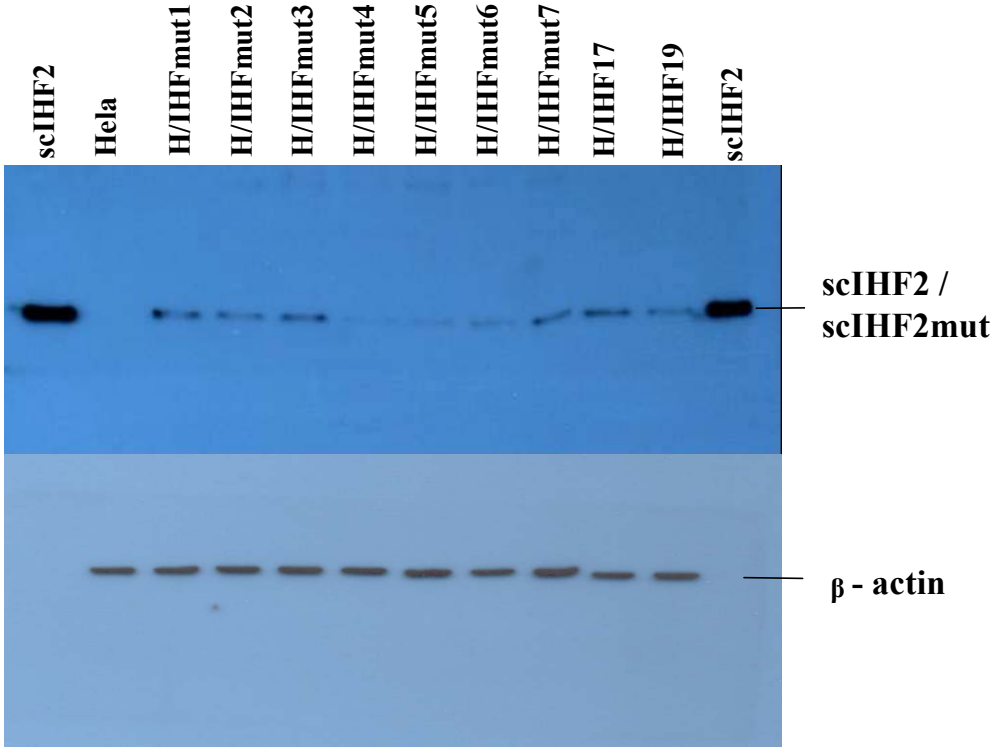


Fig.C.3.1. Western blot analysis of the expression of scIHF2 or scIHF2mut in the stable cell lines. The crude protein extractions were prepared from different scIHF2mut transgenic stable cell lines (H/IHFmut1-7) after the cell lines were generated one month. The H/IHF17 and H/IHF19 had been generated 2 years. The crude cell extracts were analyzed by western blot. We used protein extracts from Hela as negative control, 1 ng purified scIHF2 as positive control. The internal control was detected by the antibodies against β-actin.

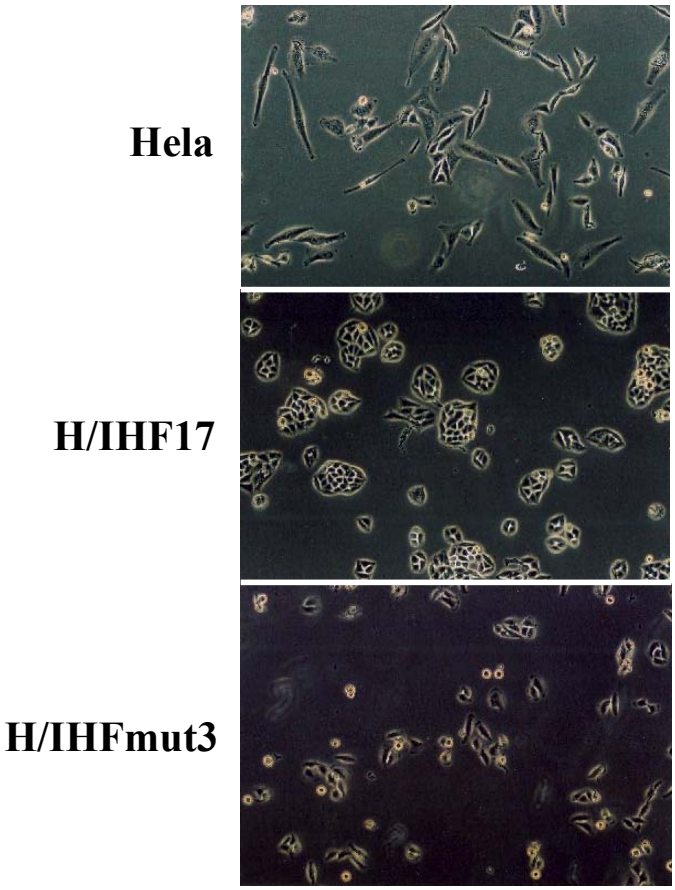


Fig.C.3.2. The morphology of Hela, H/IHF17, H/IHFmut3. The cells from different cell lines were passaged with the same confluence (around 20%), while the culture media was different. For H/IHF17 and H/IHFmut3, they were cultured in the media with 5 μg/ml puromycin, Hela were cultured in the media without puromycin. After one day, pictures were taken of the cells with 100 fold magnification. The morphology of H/IHF19 is the same as H/ IHF17. Other cell lines of H/scIHF2mut with the same expression level of scIHF2mut as H/IHFmut3 had similar morphology showed here.

C.3.2. Tolerance of cisplatin in scIHF2 cell lines

HMGB is a functional orthologue of IHF from mammalian cells. As a nuclear protein, HMGB can bind specifically to DNA isolated from the cells treated with

cisplatin [*cis*-diamminedichloroplatinum(II)] (Chu, 1994). Cisplatin is an anti-cancer chemotherapeutic reagent which can cause intrastrand adducts and interstrand crosslinks which affect DNA structure. The major DNA lesions formed in cells treated with cisplatin (>90%) are 1,2-intrastrand d(GpG) and d(ApG) cross-links (Wozniak and Blasiak, 2002). The repair of 1,2-intrastrand cisplatin adducts occurs primarily by nucleotide excision repair (NER) pathways (Zamble et al., 1996). While HMGB and other B-box containing proteins (e.g., lymphoid enhancer-binding factor LEF-1; the mammalian testis-determining factor SRY) inhibit the NER of cisplatin adducts both *in vivo* and *in vitro* (Jordan and Carmo-Fonseca, 2000) by a " shielding mechanism". This suggests that when HMG-domain proteins bind to cisplatin–DNA adducts, the adducts would not be able to access NER factors, and that DNA repair would be less active than normal (Arioka et al., 1999). As an architectural protein, scIHF2 has similar functions as HMG. We were therefore interested to explore the possibility whether scIHF2 cell lines are more sensitive to cisplatin than normal mammalian cells.

Here we applied cisplatin to the media of three cell lines: Hela, H/IHF17 as well as H/IHFmut3. The results revealed that in H/IHF17 apoptosis was induced with much lower cisplatin concentrations, and reached a maximal effect when the concentration of cisplatin was around 10 μM, while the peak of the apoptotic effect for H/IHFmut3 and Hela occurred when cisplatin increased to 40 μM. The concentration of cisplatin is around 28 μM when 50% of H/IHF17 cells became necrotic, 2.5 times less than that needed for H/IHFmut3 and Hela (Fig.C.3.3). The

apoptosis and necrosis of the cells were distinguished by the staining of Annexin V and PI (Roche). Annexin-V-Fluos binds in a Ca^{2+}-dependent manner to negatively charged phospholipid surfaces and shows high specificity to phosphatidylserine. Therefore, it stains apoptotic as well as necrotic cells. Propidium iodide stains DNA of leaky necrotic cells only.

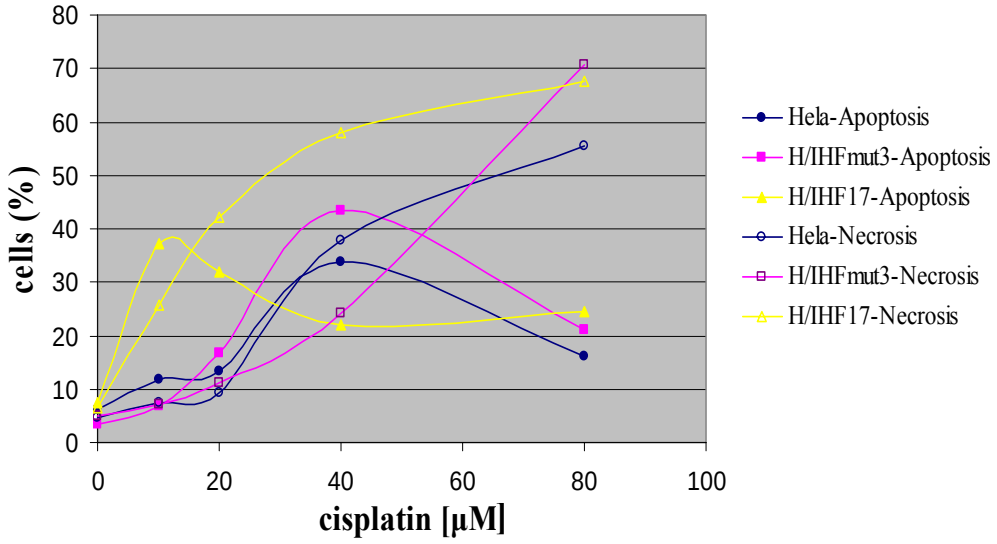


Fig.C.3.3. The apoptosis of scIHF2 expressed cell lines. Hela (blue), H/IHFmut3 (pink) and H/IHF17 (yellow) were passaged into 6-well plates with the same confluence (around 20%) one day before the cisplatin treatment. Different amount of cisplatin were added into wells. The cells were harvested and stained with annexin V- PI after 20 hrs cisplatin treatment. The data were obtained through FACS analysis. The curves with empty symbols represent necrosis. Curves with solid symbols represent apoptosis. Three independent experiments were performed.

C.3.3. Localization of scIHF2 and scIHF2mut in mammalian cells

In order to analyze the localization of scIHF2 and scIHF2mut in the mammalian cells, we constructed fusion proteins scIHF2eGFP and scIHF2muteGFP with linker

GGSGGS to bridge the C terminal of scIHF2/scIHF2mut and the N terminal of eGFP. To make sure the fusion of eGFP would not affect the biochemical function of scIHF2 and scIHF2mut, we measured the DNA binding affinity of fusion proteins in the same way as scIHFs described previously. The proteins used here were crude protein extracts from scIHF2eGFP and scIHF2muteGFP transiently expressed in HeLa cells, so we added salmon sperm DNA to avoid nonspecific binding. Fig.C.3.4 showed that scIHF2eGFP bound to H' specifically no matter whether it was expressed in Hela or *E. coli* cells (lane 5, 6, 8). As expected, scIHF2muteGFP does not show any binding affinity to H' (lane1-4).

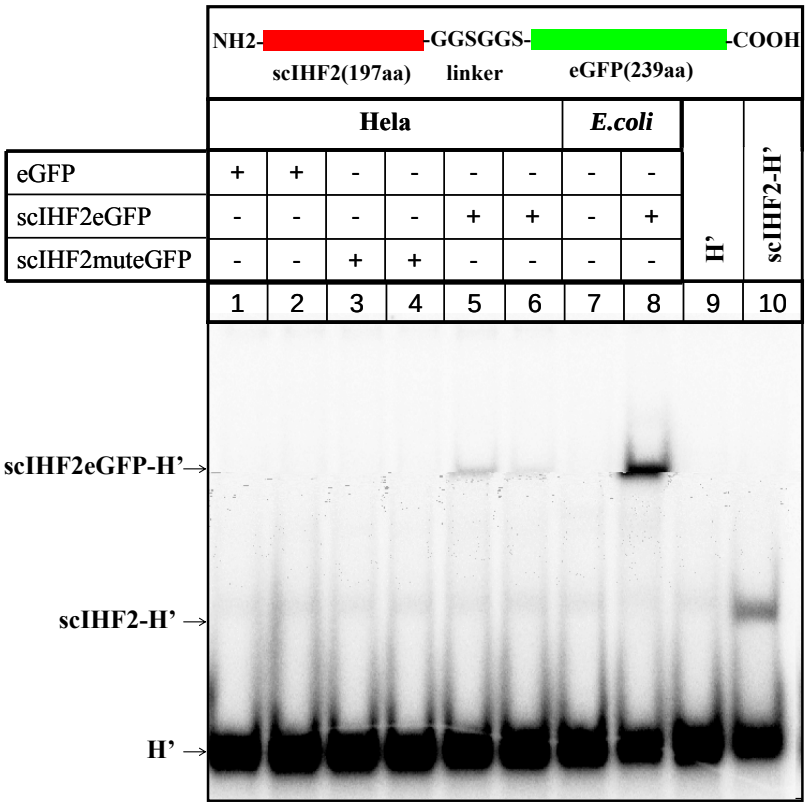


Fig.C.3.4. The binding affinity of fusion protein scIHF2eGFP. 0.1 nM radiolabeled H' was DNA substrate. As nonspecific competitor, salmon sperm DNA was used at a 520-fold molar excess over H'. pCMVSSeGFP (lane 1, 2), pCMVss-IHF2eGFP (lane 3, 4) and pCMVss-IHF2muteGFP (lane 5, 6) were transfected into Hela by electroporation. After one day, the crude protein extracts were

prepared from the transfected cells to react with DNA substrate. For *E. coli,* pTrc99a (lane 7) and pTrc-IHF2eGFP were transformed into CSH26-ΔIHF. After protein expression was induced by IPTG for 5 hrs, the crude protein extracts were prepared to react with H'. The amount of crude protein extracts in each reaction is 1 ug (lanes 1, 3, 5, 7, 8) or 0.5 ug (lanes 2, 4, 6). Note free radiolabeled H' was used as the negative control (lane 9). scIHF2-H' was loaded in lane 10 as positive control. The map of fusion protein scIHF2eGFP is shown in the upper panel.

4'-6-Diamidino-2-phenylindole (DAPI) is known to form fluorescent complexes with natural double-stranded DNA, showing fluorescence specificity for AT, AU and IC clusters. In Hela cells, scIHF2eGFP mainly located in the nucleus region which was highlighted with DAPI staining. In the other parts of the cells, it exists in much reduced amounts. scIHF2muteGFP had the same location as scIHF2eGFP. While eGFP spreaded in the cell evenly. (Fig.C.3.5)

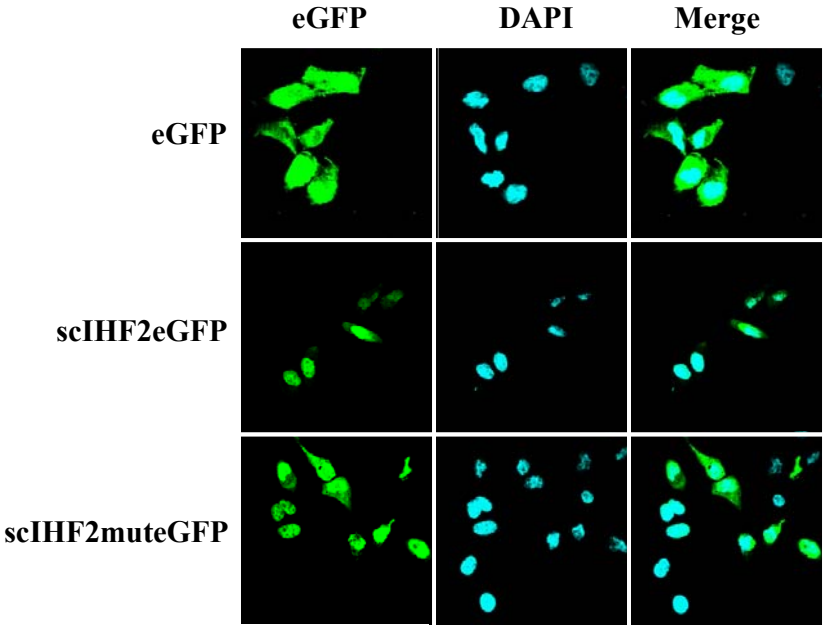


Fig.C.3.5. Localization of scIHF2 in Hela cells. pCMVSSeGFP, pCMVss-IHF2eGFP and pCMVss-IHF2muteGFP were transfected into Hela cells, respectively. One day later, cells were fixed with 4% formaldehyde/PBS and stained with DAPI. The location of fusion proteins was analyzed with Laser Scan Microscope.

As the MW of scIHF2eGFP is 49.2 kDa, we added the nuclear localization signal (NLS) to the C terminal of fusion protein to guide the protein through nucleus membrane. In order to avoid the effect of cell phase, we synchronized NIH cells. After synchronization, DNA content was increased twice (Fig.C.3.6.A), scIHF2eGFPNLS located in nucleus, but does not superimpose perfectly with DNA (Fig.C.3.6.B). The localization of scIHF2muteGFPNLS is the same as scIHF2eGFPNLS, mainly in the nucleus (Fig.C.3.6.B). As a control, eGFP was lost completely after the cells were fixed in pure ethanol (data not shown).

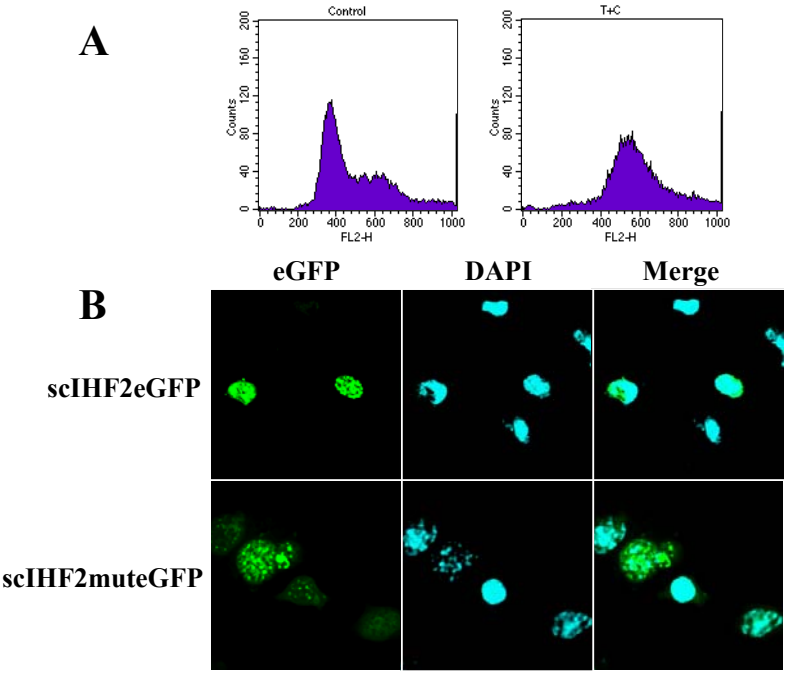


Fig.C.3.6. Localization of scIHF2 in NIH. **A.** Synchronization of NIH cells. NIH cells were treated with 2 mM Thymidine and 0.6 μg/ul Colcemid. Then the same number of NIH cells and synchronized NIH cells were stained with Propidium iodide (PI) and analyzed with FACS. **B.** The localization of scIHF2 in NIH. pCMVSSeGFP**,** pCMVss-IHF2eGFPNLS and pCMVss-IHF2muteGFPNLS were transfected into synchronized NIH cells, respectively. After one day, the cells were fixed with pure ethanol and stained with DAPI.

D. Discussion

scIHF2 is similar to IHF

One aim of this work was to generate a functional single chain integration host factor that can be used in eukaryotic cells. To the best of our knowledge, this is the first example of the design of a recombinant protein in which an entire protein subunit has been inserted into another. We think that it should be possible to apply this strategy also for other heteromeric protein complexes whose three-dimensional structures are known. Our approach might eventually be used to synthesize mutants which could aid in the elucidation of structure–function relationships of heteromeric proteins in general. In addition, our approach could facilitate the production of desired large quantities of functional single polypeptide chain variants of suitable heteromeric complexes, which might be difficult to assemble from purified subunits.

The fact that scIHF2 supports both site-specific integrative recombination and pSC101 replication in an *E. coli* host lacking parental IHF has important implications for the structure of scIHF2. First, integrative as well as excisive recombination requires high precision DNA-bending by IHF in order to generate functional recombinogenic nucleoprotein complexes (Nunes-Duby et al., 1995). Our scIHF2 must therefore be able to facilitate a very similar, if not identical, type of DNA-looping upon binding to IHF sites present in *att*P. Second, the functional IHF site as part of the pSC101 minimal origin differs from the λ sites at permissively variable bases within the consensus element as well as in the

composition of the upstream located poly(dAT) element and its connecting spacer (Stenzel et al., 1987). Hence, scIHF2 recognizes a different biologically significant consensus sequences and performs IHF's function(s) at these sites. Together, this implies that most, if not all, side-chains which interact directly or indirectly with DNA are appropriately positioned in scIHF2. This inference is supported by the crystal structure of scIHF2 in complex with H'.

The crystal structure of scIHF2 in complex with H' is very similar to that of the IHF co-crystal structure (Fig.C.2.11A, B). The positions of the His-tag and second linker of scIHF2 are far from DNA and the their presence does not change the relative positions of other amino acids. Small differences exist in scIHF2 at the N-termini of α and β subunits. We added two amino acids Ala and Ser between βM1 and βT2 in the N terminal of β subunit, and deleted two amino acids in the N-terminal of α subunit for adding the first linker connecting βQ39 with the N-terminus of the α subunit at position αL3 to bridge the 11Å distance between the two residues in the co-crystal structure. As the N-termini of α and β subunits are parts of IHF's clamps structure which interacts with DNA by electrostatic force, the change of those two termini affect scIHF2's binding affinity. As a consequence, the K_d of scIHF2 is around 3 or 10 times larger than that of IHF under low salt or high salt conditions. However, this doesn't affect DNA bending by scIHF2, which is similar to that seen with IHF.

The stability of scIHF2-H' and IHF-H1 complexes

The results of specific and nonspecific DNA competition assays showed that a much higher concentration of competitor is needed to replace the radiolabeled H' binding to IHF. And with high concentrations of competitors, there are still IHF-DNA or scIHF2-DNA complexes left after 40 min of competition. That means the scIHF2-H' complex is quite stable, similar to IHF-H'. The stability experiments showed the half-life of IHF-H' is more than 48 hrs. For scIHF2-H' complexes, the half-life is reduced to around 30 hrs. However, The half-life of IHF complexed to its binding sites was reported previously as being less than 2 minutes (Giladi et al., 1995; Yang and Nash, 1995). In the stability measurement, the concentration of protein is 4-fold over the K_d, and the amount of complex represent around 80% of the initial substrate concentration, more than the protein requirement in the stability experiment (less than 10% of K_d). I think that is the main reason for the longer half life of IHF-DNA in our measurements. The reaction condition for competitive experiments and stability assay is 0.5xTBE buffer (low salt), which can help to stabilize the salt bridges between IHF and DNA. Under high salt condition, the half life of IHF-H' or scIHF2-H' becomes perhaps shorter (Holbrook et al., 2001). Another reason for the longer half-life in our measurements is reaction temperature. The reactions were performed at room temperature (20-25 ^{o}C), lower than the temperature used in previous stability measurement (37 ^{o}C). Because the interaction between IHF and DNA is temperature dependent (Giladi et al., 1995), lower temperatures could increase the half life of complexes.

scIHF2 binds to RNA

In addition to specific DNA substrates, our preliminary data show that scIHF2 binds to RNA. The binding may be specific because the complexes exist as sharp bands in EMSAs. The RNA used in this experiment is from *E. coli.* There are more than 1500 potential IHF binding sites in the genome of *E. coli* (unpublished data from Drőge, P. and Li, J-M). It is possible for RNA to form double stranded secondary structures, which can be recognized and bound by scIHF2. The difference of scIHF2-RNA complexes caused by the length of RNA, the position of IHF binding sites on RNA and the bending introduced by scIHF2 makes it impossible to show only one or two bands in the EMSA. However, only three specific bands were shown on the gel. This may be due to poor resolution of the gel, and the electrophoresis time is not long enough to separate all the bands of scIHF2-RNA complexes. The specific binding between scIHF2 and RNA implies IHF may affect protein expression at translation level. And IHF has extensive similarity to S7 which binds to 16S rRNA (Robert et al., 2000; Wimberly et al., 1997), so scIHFs has high possibility to bind to 16sRNA. To further prove these, we can perform *in vivo* RNA-scIHF2 immunoprecipitation assay and find out potential RNA binding sites of IHF.

The length of linkers affects the biochemical and functional properties of scIHFs

We generated five scIHFs which show similar functions but with different activities. First, while all five scIHFs, like IHF, promote excisive recombination *in*

vitro, both scIHF1 and scIHF3E appear unable to facilitate integrative recombination to a significant extent *in vitro* (Fig. C.1.12) and as shown for scIHF1 *in vivo* (Christ, 2002). A critical factor for this restricted functionality may be the significantly increased K_d for all three λ cognate sites at 100 mM salt (Table C.1.1). A recombinogenic integrative intasome with *att*P may not be stably formed under these conditions to allow synaptic complex formation with the protein-free *att*B. The two functional nucleoprotein complexes involved in excisive recombination, however, may be generated and are stable enough because each snup requires only one occupied IHF cognate site and Xis is also involved as an architectural cofactor. In the absence of structural data on the respective scIHF–DNA complexes, we find it difficult, however, to speculate further on reasons for the observed K_d changes. In case of scIHF3E, a different DNA trajectory within the protein–DNA complex, as discussed below, is almost certainly a contributing factor. Eventually, the observed restricted accessory roles of scIHF1 and scIHF3E could prove useful in conjunction with Int in our attempts to regulate site-specific recombination in mammalian cells (Christ et al., 2002; Corona et al., 2003).

Second, scIHF4 serves as a cofactor for both λ recombination pathways *in vitro* (Fig.C.1.12), but the protein is nearly inactive in pSC101 replication assays *in vivo* (Figs.C.1.13 and C.1.14). We think that a low affinity especially for the pSC101 cognate site is an unlikely explanation for this phenotype (Table C.1.1). Furthermore, as revealed by EMSA (Fig.C.1.6), it is unlikely that a significant deviation from the expected overall bend angle exists within the scIHF4–DNA

complex. We consider here two possibilities that might account for the observed selective biological activity of scIHF4. First, the shortened linker 4 may have repositioned residues 44 to 46 within the β domain of scIHF4. In particular, βR46 is involved in specific DNA recognition and stabilization of the DNA bend around the protein (Rice et al., 1996). It is possible that this repositioning leads to an altered dihedral angle of DNA within the complex which is undetectable by EMSA. The change of DNA path may prevent the establishment of functional interactions between DnaA and RepA proteins at the origin of pSC101 replication (Lu et al., 1998). It may not be detrimental, however, to scIHF4's role in site specific recombination which may allow for more flexible DNA geometries within snups. Second, the formation of hypothetical protein–protein interactions between IHF and DnaA (or RepA) within the initiation complex may be prohibited due to conformational changes in scIHF4.

A third intriguing observation is that scIHF3E shows a super shift with all four cognate sites analyzed so far, indicating that the DNA trajectory within these complexes must deviate significantly from that in complexes with scIHF3 or scIHF2. Hence, this effect cannot be attributed to the altered linker length alone, but must involve the glutamate which substituted the lysine at position 45 of the α domain in scIHF3. It is interesting to note that a glutamate is in fact present at the corresponding position (βE44) in the β domain where it acts as a buttress, holding the arginine 46 in place for specific DNA contacts (Lynch et al., 2003; Rice et al., 1996). Residue αK45 is located on the opposite side of the complex near αS47. The

latter makes water-mediated contacts to the backbone of the minor groove of the A-tract (Rice et al., 1996). We noticed that in the cocrystal structure, αK45 forms a hydrogen bond with a phosphate group of the A-tract DNA. Hence, a substitution with an oppositely charged glutamate will weaken or even destroy protein–DNA interactions within this region of the complex. This explains the observed decrease in affinities for all binding sites tested so far. In addition, if scIHF3E cannot make stable contacts with the A-tract on one side of the protein–DNA complex, the path of the DNA will be altered severely. The further retardation of the complex that we observe in EMSAs has, to our knowledge, not been observed with mutant IHFs before, and is perhaps attributable to a significant decrease in the overall bend angle since one side of the protein body cannot serve as a clamp which stabilizes the strong bending.

An "Open" conformation of DNA in complex with scIHF2E

When we introduced the substitution K45αE from scIHF3E into scIHF2, this resulted in the formation of scIHF2E. We found scIHF2E also showed supershift with three IHF consensus sequences compared with the complex formed by scIHF2 (Fig.C.2.5), and the shift position of scIHF2E-DNA is the same as scIHF3E-DNA (data not shown). In addition to that, complexes of scIHF2E-*att*P (*att*R) are in different positions from scIHF2-*att*P (*att*R) (Fig.C.2.6, C.2.7). This implies that the supershift is due to the substitution of lysine at position 45 in the α subunit by a glutamine acid. It is known that αK45 interacts with the phosphate backbone of DNA (Fig.C.2.1). Hence αE45 should affect the scIHF2E-DNA interactions by

itself or by repositioning of neighbor amino acids. Therefore it would push the DNA away from the original position by repulsive force. Compared with scIHF2 and IHF, scIHF2E-bound DNA has a smaller bending degree which has been shown in AFM (Fig.C.2.12). If we name the conformation of DNA in the complex of scIHF2-DNA as “closed”, then the DNA bound with scIHF2E is in an “open” conformation (Fig.D.1). This “open” conformation is different from the conformation of HU-bound DNA. In HU-DNA complex, except the 9 bp between kinks caused by the intercalation of pralines are tightly bound across the top of HU, but the DNA outside the kinks ‘flaps' against the sides of the protein. This balance could easily be affected by other factors (Swinger et al., 2003). Therefore, HU has a more ‘open’ and flexible nucleoprotein complex. While the open conformation of scIHF2E-bound DNA is only caused by the disruption of left part DNA-protein interaction, the complex of scIHF2E-DNA is more rigid. However, the smaller bending angle will no longer facilitate the interaction between the integrase and core region after the N- terminal domain of integrase bound to arm binding sites on *att*P. scIHF2E, Int and *att*P will form a recombinogenic integrative intasome which is not stable or effective. The facilitation may simply be weak; the opportunities for arm bound Int to make additional contacts with the core ought to be there. Maybe that is why scIHF2E shows weak activity in the integrative recombination assay. In contrast to scIHF2, the weaker binding affinity of scIHF2E to IHF consensus sequences under high salt condition could also result in the marginal activity in the integrative recombination. However, we have attempted to avoid this by adding enough protein to saturate IHF binding sites on *att*P and not to aggregate DNA

(Fig.C.2.6). In integrative recombination, scIHF2E is supercoiled *att*P dependent, as scIHF2 and IHF (data not shown)(Richet et al., 1986). We believe that an increased superhelical density can increase recombination efficiency triggered by scIHF2E because the supercoiling is capable to organize a more stable *att*P intasome.

scIHF2E can support excisive recombination on supercoiled pλER as efficiently as scIHF2 (Fig.C.2.8). The reduced bending introduced by scIHF2E does not seem to affect the formation of the right *att*L and *att*R nucleoprotein complexes. Similar to HU or HMGB can replace IHF in the excisive recombination, not in the integrative recombination with wild type substrates (Segall et al., 1994). However, the supportive activity of scIHF2E in the excisive recombination is dependent on supercoiling of *att*R (Fig.C.2.10), which differs from scIHF2 and IHF which are active in excisive recombination with linear DNA substrates (Fig.C.2.9). Because the *att*R recombinogenic nucleoprotein complex is a more delicate structure than the *att*L recombinogenic nucleoprotein complex (Segall et al., 1994), and the functionally important cognate site for IHF in *att*R is H2, where extreme DNA-bending is needed in formation of *att*R recombinogenic nucleoprotein complex (Radman-Livaja et al., 2006). Supercoiling of *att*R seems important to complement the reduced bending introduced by scIHF2E and thus can overcome a thermodynamic penalty imposed on the formation of a stable closed scIHF2E-H2 complex. Thus scIHF2E, Xis/Fis and Int as well as supercoiled *att*R can form a wrapped structure to perform excisive recombination. We think that this penalty is

due to the combined thermodynamic effects of the stiffness of short DNA segments (Kahn, 1994) and weakened molecular interactions between the left DNA arm and the protein body of scIHF2E.

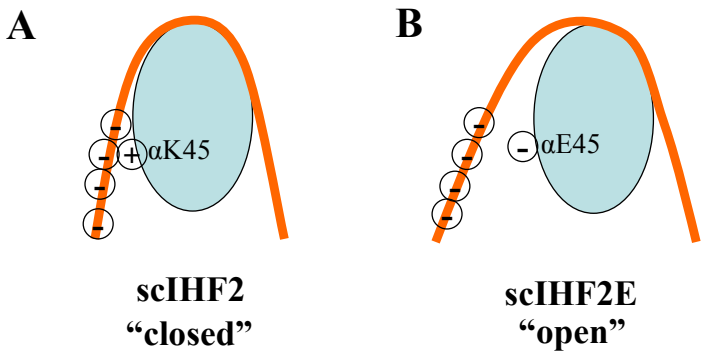


Fig.D.1. Hypothesis of the DNA conformation in the complexes of scIHF2-DNA and scIHF2E-DNA. A. The closed conformation of scIHF2-DNA. **B.** The open conformation of scIHF2E-DNA. The blue oval represents protein scIHF2 or scIHF2E. The orange curves are the consensus DNA sequences of IHF. With a dash in the circle means negative charge. With the plus sign in the circle means positive charge.

A Switch between an "open" and a "Closed" conformation

Surprisingly, the overall crystal structure of scIHF2E-H' is almost identical to that of scIHF2-H' (IHF-H') (Fig.C.2.11). We observed a "closed" conformation, contrary to our hypothesis. When we focus on the mutated region, we found there are two 'new' divalent cation (Mn^{2+}) binding sites stabilizing the closed conformation of scIHF2E-DNA complex. One of them is exactly mediating the interaction between αE45 and DNA. It is possible that the DNA in scIHF2E-H' has different conformations, shifting all the time. In the conditions of EMSAs and recombination assays, the probability of open conformation is larger than the others. While the closed conformation is the major one in the crystallization condition (Fig. D.2). Even for the DNA of scIHF2-H' or IHF-H', it is also the same, but the closed

conformation is always the major conformation no matter whether it is under EMSAs or crystallization conditions. However, the H'-DNA used in our crystallographic studies contained a nick in one phosphodiester backbone near the site of proline intercalation. Therefore, in order to gain evidence that scIHF2E can adopt a closed conformational state also with covalently closed H'-DNA, we incubated either scIHF2 or scIHF2E with H'-DNA in the presence of magnesium ions, and analyzed complex formation through EMSA and AFM. The reversal of supershift with Mg^{2+} in the EMSAs provided evidence that the nick in DNA of the scIHF2E-H' crystal is not necessary for the closed conformation (Fig.C.2.13). The AFM analysis revealed scIHF2E-DNA complex has similar bending angle as scIHF2-DNA in the presence of magnesium ions. And the existence of Mg^{2+} increases the integration efficiency supported by scIHF2E two times compared with the condition without divalent cations (Fig.C.2.14). The concentration range of Mg^{2+} in EMSAs and integrative recombination is limited. Just like the crystallization, the growth of scIHF2E-H' crystals is sensitive to Mn^{2+} (data not shown). Low concentration of Mg^{2+} may not be enough to mediate the contact between αE45 and phosphate backbone of the DNA, while high concentration of Mg^{2+} will bind to DNA to shield the negative charge, competing with the interaction between scIHF2E and DNA.

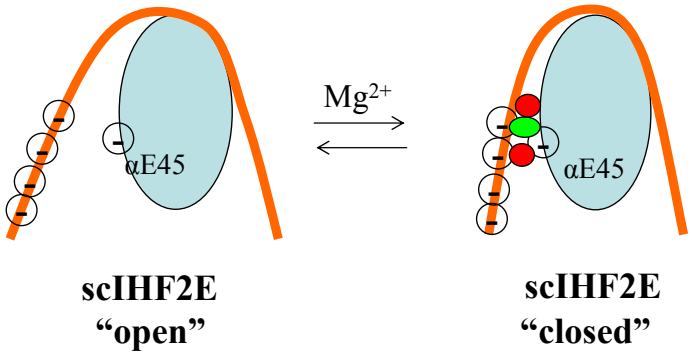


Fig.D.2. The switch of DNA conformation in the complex of scIHF2E-DNA. The left part is the open conformation, which can be changed to close conformation (right part) by adding Mg^{2+}. The blue oval represents protein scIHF2 or scIHF2E. The orange curves are the consensus DNA sequence of IHF, with a dash in the circle indicating negative charge. The black circle with red filling means divalent ion, Mg^{2+}. The black circle with green filling represents water molecule.

Complementary effect of mutation S47αR in scIHF2E

As mentioned before, the corresponding amino acid of αK45 in the β subunit is E44, which forms a salt bridge with βR46 to position the guanidinium of βR46 centered in the minor groove contacting the edges of conserved bases (Rice et al., 1996). When we mutated αS47 into Arg in scIHF2E to form scIHF2ER, the left and right part of scIHF2ER clamp areas should be similar. The effect caused by αE45 may be mitigated by S47αR. In EMSAs, the shift position of scIHF2ER-DNA is between complexes of scIHF2E-DNA and scIHF2-DNA (Fig.C.2.5, C.2.6, C.2.7). That means the bending angle introduced by scIHF2ER may be larger than scIHF2E, but smaller than in the scIHF2-DNA complex. scIHF2ER can support integrative recombination as efficiently as scIHF2. The supercoiling energy in *att*P appears sufficient to facilitate scIHF2ER and Int to form a recombinogenic integrative intasome. While in excisive recombination, scIHF2ER is supercoiled *att*R dependent, the same as scIHF2E, and can not assist Int and Xis to form efficient nucleoprotein complex with the linear DNA substrates as scIHF2. In EMSAs with divalent cations present, scIHF2ER is much more affected than scIHF2E (Fig.C.2.13). One possibility is that in this case, the bending angle of DNA in the scIHF2ER is even larger than that introduced by scIHF2.

Integration host factor in mammalian cells

As an architectural protein, IHF is involved in the control of various replication, transcription, and recombination systems as well as transposition systems in prokaryotic cells. Our biochemical and functional analysis indicate that scIHF2 can replace IHF and perform its roles in *E. coli.* Equivalents of IHF in eukaryotes are HMGB proteins. Some of them show limited sequence-specificity, such as LEF-1 and SRY proteins, while others, more abundant members of this family display no sequence specificity (Giese et al., 1992; Travers, 1997). Like IHF, these proteins are involved in DNA transactions such as transcription and recombination. However, the degree of protein-induced DNA bending is significantly smaller ($<120^{o}$) than that observed with IHF ($>160^{o}$) (Giese et al., 1992). We showed here that scIHF2 and scIHF2mut are stably expressed in transgenic HeLa cells (FigC.3.1), and that the protein is found primarily inside the nucleus (Fig.C.3.5, C.3.6) (Corona et al., 2003). The fusion protein scIHF2eGFP used for localization analysis does not lose the ability to bind specifically to IHF binding sites, while scIHF2muteGFP has lost this ability, which is similar to scIHF2mut (Fig.C.3.4). A conservative estimate of the number of scIHF2 molecules, which is based on several Western blot analyses, indicates that between 20,000 and 200,000 copies must be present per cell. A preliminary bioinformatics approach then identified over 90,000 consensus IHF binding sites scattered more or less randomly throughout the human genome (Li and Dröge, unpublished). About 15,000 of those sites have a suitable poly(dAT) element positioned at an appropriate distance upstream. Hence, those latter DNA sites represent segments which could

potentially be bent strongly upon scIHF2 binding. We think that at least some of those sites must be accessible for scIHF2, thereby contributing to the rather strong nuclear immunostaining that we observed in some cells (Corona et al., 2003). In addition to sequence specific binding, scIHF2(IHF) can bind to DNA nonspecifically and retain its ability to deform DNA (Segall et al., 1994). Thus, scIHF2 might affect the transcription of some genes and cause the down- or up-regulation of protein expression. Here we showed the different morphology of Hela/scIHF2 cell lines compared with Hela/scIHF2mut and parental Hela cells (Fig.C.3.2). The presence of scIHF2 seems to affect the cytoskeleton and cell coat as well as extracellular matrix. Although scIHF2mut does not have the ability to bind and bend IHF cognate sites as scIHF2, it is still able to bind to DNA either nonspecifically or in its own way and even affect the expression of some proteins. This could explain the nuclear localization of scIHF2muteGFP.

scIHF2 stimulates apoptosis induced by cisplatin

Cisplatin [*cis*-diamminedichloroplatinium(II)] is a highly effective ant-cancer drug and exerts its cytotoxic effects by disrupting the DNA structure in cells through the formation of intrastrand adducts and interstrand cross-link (Jones et al., 1991). Major DNA lesions formed in cells treated with cisplatin (>90%) are 1,2-intrastrand d(GpG) and d(ApG) cross-links (Wozniak and Blasiak, 2002). The repair of 1,2-intrastrand cisplatin adducts occurs primarily by nucleotide excision repair pathways (Zamble et al., 1996). The overexpression of NER genes was associated with cisplatin resistance in human ovarian, glioma, bladder, and lung

cancer cells (Wu et al., 2003). While HMGB and other B-box containing proteins (e.g., lymphoid enhancer-binding factor LEF-1; the mammalian testis-determining factor SRY) can inhibit NER of cisplatin adducts both *in vivo* and *in vitro* (Jordan and Carmo-Fonseca, 2000) by a " shielding mechanism" which suggests that when HMGB-domain proteins bind to the cisplatin–DNA adducts, the adducts would not be accessable to NER factors and DNA repair would be slower than normal (Arioka et al., 1999), thus causing apoptosis. Here, we observed that stable cell line Hela/scIHF2 is much more sensitive to cisplatin than Hela/scIHF2mut and Hela (Fig.C.3.3). Therefore scIHF2 could inhibit the nucleotide excision repair of cisplatin-crosslinked DNA by the same shielding mechanism as HMGB. The crystal structure of HMGB binding with cisplatin-modified DNA showed HMGB binds to the minor grove of DNA and Phe37 intercalates into a hydrophobic notch created at the platinum crosslinked d(GpG) site (Fig.A.8). This makes the cisplatin crosslinked site inaccessible for other proteins. Cisplatin adducts can cause two helical segments of DNA to adopt an angle of around 120^{o} with respect to each other (Lilley, 1992). We deduce that scIHF2 can bind to such pre-bent cisplatin adducts as a non-specific DNA binding protein and deform DNA further. And it is possible that cisplatin modified specific binding sites of IHF, and then scIHF2 binds and shields the modified sites. The existence of aberrant DNA will cause apoptosis of the cell mainly due to the inability of the transcription and replication machinery of the cell to use these stretches as template.

Possible applications of scIHF2 in mammalian cells

The λ recombination system is more complex than the Cre/*loxP*, the Flp/*FRT*, or the phiC31 integrase system (Lewis and Hatfull, 2001). This higher level of complexity results in a very stringent control over the directionality of recombination. It also increases the fidelity of the system with respect to the choice of recombination partner sequences and the generation of recombinant products (Radman-Livaja et al., 2003). For example, Int-mediated integration of the phage attachment site *att*P into other genomic sequences than the 21 bp comprising natural partner *att*B is a rare event in *E. coli* cells that express IHF (Miller and Friedman, 1980). However, in an *att*B deletion *E. coli* strain, the secondary *att*B sites become highly efficient targets. We have demonsttated that scIHF2 stimulates integrative recombination by wild-type Int on episomal substtates 4- to 8-fold. The very low level of excisive recombination catalyzed by wild-type Int remained, however, unaffected by the presence of scIHF2 (Corona et al., 2003). This is most likely due to the lack of Xis protein in HeLa cells. Functional recombinogenic nucleoprotein complexes are apparently formed only on *att*P, but not on hybrid sites *att*L and/or *att*R. One potentially important application of the wild-type λ Int system in mammalian cells could thus include scIHF2 for the safe and controlled integration of foreign DNA into one of about 20 suitable yet distinct *att*B-like sequences that were identified in the human genome (Li and Dröge, unpublished results). The high fidelity of the λ system could then limit the frequency of unwanted integration events into other genomic locations as well as illegitimate intragenomic recombination observed with other, more promiscuous site-specific

recombination systems (Schmidt et al., 2000; Thyagarajan et al., 2001). Eventually such a system could become important in safe genome modifications of, for example, human stem cells for gene therapy purposes.

Potential applications for scIHF2 in mammalian cells are, however, not confined to a role in recombination. We speculate that the recombinant protein may also be useful in studies investigating the structure and function of eukaryotic nucleoprotein complexes inside a living cell. In addition, scIHF2 may be employed in biopharmaceutical production techniques where, through its DNA-bending capacity, it may keep a promoter free of nucleosomes and, thus, more accessible for the transcriptional machinery. This could boost and/or maintain a desired high level of gene expression. Because scIHF2-expressing cells are more sensitive for cisplatin-induced apoptosis, they could provide a good target for the identification and study of related anti-cancer reagents.

E. Summary

As a potential tool of gene therapy, the site-specific recombination system of bacteriophage λ can transfer genetic material into eukaryotic cells in a controlled way. Recently, the mutant lambda integrases (Int h/218) has been introduced to mammalian cells where it promotes site-specific recombination at a significant level in the absence of protein co-factors. However, mutant Int supports not only the integration but also excision in the absence of any other cofactors. The project presented in this thesis had as a goal to construct a recombinant integration host factor to control the directionality of recombination in mammalian cells. Single chain integration host factor, scIHF2, was designed by inserting almost the entire α subunit of heterodimeric IHF into the β subunit via two short linkers. The biochemical and functional analysis showed scIHF2 can bind and bend a cognate sequences similar to the parental protein. We showed that scIHF2 could support both site-specific recombination and plasmid replication. The stable expression of scIHF2 in Hela cells implies that the recombinant protein is tolerated by mammalian cells kept in culture. Furthermore, the stable cell line HeLa/scIHF2 is much more sensitive to cisplatin treatment than HeLa/scIHF2mut and HeLa cells, which provides a good target for the identification and study of related anti-cancer reagents. To further study the interaction between scIHF2 and DNA, the other two scIHF2 variants, scIHF2E and scIHF2ER, were constructed. The supershift in 0.5xTBE buffer and its reversal in a buffer with divalent metal ions, as well as the crystal structure analysis revealed that a scIHF2E-DNA complex exists in two different conformations, closed and open, and the switch between these two

conformations is dependent on the presence or absence of divalent metal ions, respectively. Introduction of acidic residues at the protein-DNA interface holds a general potential for the design of metal-mediated switches for the investigation of functional relationships or engineering nano-scale DNA devices.

F. Bibliography

Aizawa, S., Nishino, H., Saito, K., Kimura, K., Shirakawa, H., and Yoshida, M. (1994). Stimulation of transcription in cultured cells by high mobility group protein 1: essential role of the acidic carboxyl-terminal region. Biochemistry *33*, 14690-14695.

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J Bacteriol *181*, 6361-6370.

Ali, B. M., Amit, R., Braslavsky, I., Oppenheim, A. B., Gileadi, O., and Stavans, J. (2001). Compaction of single DNA molecules induced by binding of integration host factor (IHF). Proc Natl Acad Sci U S A *98*, 10658-10663.

Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehle, M. M., Paegle, E. S., and Hatfield, G. W. (2000). Global gene expression profiling in Escherichia coli K12. The effects of integration host factor. J Biol Chem *275*, 29672-29684.

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V., Pierson, L. S., 3rd, and et al. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. Embo J *5*, 433-440.

Arioka, H., Nishio, K., Ishida, T., Fukumoto, H., Fukuoka, K., Nomoto, T., Kurokawa, H., Yokote, H., Abe, S., and Saijo, N. (1999). Enhancement of cisplatin

sensitivity in high mobility group 2 cDNA-transfected human lung cancer cells. Jpn J Cancer Res *90*, 108-115.

Azam, T. A., and Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. J Biol Chem *274*, 33105-33113.

Bao, Q., Christ, N., and Drőge, P. (2004). Single-chain integration host factors as probes for high-precision nucleoprotein complex formation. Gene *343*, 99-106.

Bewley, C. A., Gronenborn, A. M., and Clore, G. M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu Rev Biophys Biomol Struct *27*, 105-131.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by lambda integrase. Nature *435*, 1059-1066.

Blakely, G. W., Davidson, A. O., and Sherratt, D. J. (2000). Sequential strand exchange by XerC and XerD during site-specific recombination at dif. J Biol Chem *275*, 9930-9936.

Bonnefoy, E., and Rouviere-Yaniv, J. (1992). HU, the major histone-like protein of E. coli, modulates the binding of IHF to oriC. Embo J *11*, 4489-4496.

Boonyaratanakornkit, V., Melvin, V., Prendergast, P., Altmann, M., Ronfani, L., Bianchi, M. E., Taraseviciene, L., Nordeen, S. K., Allegretto, E. A., and Edwards,

D. P. (1998). High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. Mol Cell Biol *18*, 4471-4487.

Boulikas, T. (1996). DNA lesion-recognizing proteins and the p53 connection. Anticancer Res *16*, 225-242.

Breuner, A., Brondsted, L., and Hammer, K. (2001). Resolvase-like recombination performed by the TP901-1 integrase. Microbiology *147*, 2051-2063.

Broadhurst, R. W., Hardman, C. H., Thomas, J. O., and Laue, E. D. (1995). Backbone dynamics of the A-domain of HMG1 as studied by 15N NMR spectroscopy. Biochemistry *34*, 16608-16617.

Broyles, S. S., and Pettijohn, D. E. (1986). Interaction of the Escherichia coli HU protein with DNA. Evidence for formation of nucleosome-like structures with altered DNA helical pitch. J Mol Biol *187*, 47-60.

Bushman, W., Thompson, J. F., Vargas, L., and Landy, A. (1985). Control of directionality in lambda site specific recombination. Science *230*, 906-911.

Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. Mol Cell Biol *19*, 5237-5246.

Bustin, M. (2001). Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends Biochem Sci *26*, 152-153.

Bustin, M., Trieschmann, L., and Postnikov, Y. V. (1995). The HMG-14/-17 chromosomal protein family: architectural elements that enhance transcription from chromatin templates. Semin Cell Biol *6*, 247-255.

Calb, R., Davidovitch, A., Koby, S., Giladi, H., Goldenberg, D., Margalit, H., Holtel, A., Timmis, K., Sanchez-Romero, J. M., de Lorenzo, V., and Oppenheim, A. B. (1996). Structure and function of the Pseudomonas putida integration host factor. J Bacteriol *178*, 6319-6326.

Cassell, G. D., and Segall, A. M. (2003). Mechanism of inhibition of site-specific recombination by the Holliday junction-trapping peptide WKHYNY: insights into phage lambda integrase-mediated strand exchange. J Mol Biol *327*, 413-429.

Cassler, M. R., Grimwade, J. E., and Leonard, A. C. (1995). Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. Embo J *14*, 5833-5841.

Cassler, M. R., Grimwade, J. E., McGarry, K. C., Mott, R. T., and Leonard, A. C. (1999). Drunken-cell footprints: nuclease treatment of ethanol-permeabilized bacteria reveals an initiation-like nucleoprotein complex in stationary phase replication origins. Nucleic Acids Res *27*, 4570-4576.

Chau, K. Y., Manfioletti, G., Cheung-Chau, K. W., Fusco, A., Dhomen, N., Sowden, J. C., Sasabe, T., Mukai, S., and Ono, S. J. (2003). Derepression of HMGA2 gene expression in retinoblastoma is associated with cell proliferation. Mol Med *9*, 154-165.

Chen, J. W., Evans, B. R., Zheng, L., and Jayaram, M. (1991). Tyr60 variants of Flp recombinase generate conformationally altered protein-DNA complexes. Differential activity in full-site and half-site recombinations. J Mol Biol *218*, 107-118.

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., and Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Mol Cell *6*, 885-897.

Chen, Y., and Rice, P. A. (2003). New insight into site-specific recombination from Flp recombinase-DNA structures. Annu Rev Biophys Biomol Struct *32*, 135-159.

Christ, N., Corona, T., and Drőge, P. (2002). Site-specific recombination in eukaryotic cells mediated by mutant lambda integrases: implications for synaptic complex formation and the reactivity of episomal DNA segments. J Mol Biol *319*, 305-314.

Christensen, B. B., Atlung, T., and Hansen, F. G. (1999). DnaA boxes are important elements in setting the initiation mass of Escherichia coli. J Bacteriol *181*, 2683-2688.

Chu, G. (1994). Cellular responses to cisplatin. The roles of DNA-binding proteins and DNA repair. J Biol Chem *269*, 787-790.

Claverie-Martin, F., and Magasanik, B. (1992). Positive and negative effects of

DNA bending on activation of transcription from a distant site. J Mol Biol *227*, 996-1008.

Corona, T., Bao, Q., Christ, N., Schwartz, T., Li, J., and Drőge, P. (2003). Activation of site-specific DNA integration in human cells by a single chain integration host factor. Nucleic Acids Res *31*, 5140-5148.

Craig, N. L. (1988). The mechanism of conservative site-specific recombination. Annu Rev Genet *22*, 77-105.

Datta, H. J., Khatri, G. S., and Bastia, D. (1999). Mechanism of recruitment of DnaB helicase to the replication origin of the plasmid pSC101. Proc Natl Acad Sci U S A *96*, 73-78.

DiGabriele, A. D., Sanderson, M. R., and Steitz, T. A. (1989). Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. Proc Natl Acad Sci U S A *86*, 1816-1820.

DiGabriele, A. D., and Steitz, T. A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. J Mol Biol *231*, 1024-1039.

Ding, H. F., Bustin, M., and Hansen, U. (1997). Alleviation of histone H1-mediated transcriptional repression and chromatin compaction by the acidic activation region in chromosomal protein HMG-14. Mol Cell Biol *17*, 5843-5855.

Ellenberger, T., and Landy, A. (1997). A good turn for DNA: the structure of

integration host factor bound to DNA. Structure *5*, 153-157.

Engelhorn, M., and Geiselmann, J. (1998). Maximal transcriptional activation by the IHF protein of Escherichia coli depends on optimal DNA bending by the activator. Mol Microbiol *30*, 431-441.

Esposito, D., Thrower, J. S., and Scocca, J. J. (2001). Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation. Nucleic Acids Res *29*, 3955-3964.

Farnet, C. M., and Bushman, F. D. (1997). HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell *88*, 483-492.

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

Frumerie, C., Sylwan, L., Ahlgren-Berg, A., and Haggard-Ljungquist, E. (2005). Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. Virology *332*, 284-294.

Gelman, D. M., Noain, D., Avale, M. E., Otero, V., Low, M. J., and Rubinstein, M. (2003). Transgenic mice engineered to target Cre/loxP-mediated DNA recombination into catecholaminergic neurons. Genesis *36*, 196-202.

Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein

structures. Cell *69*, 185-195.

Giladi, H., Goldenberg, D., Koby, S., and Oppenheim, A. B. (1995). Enhanced activity of the bacteriophage lambda PL promoter at low temperature. Proc Natl Acad Sci U S A *92*, 2184-2188.

Giladi, H., Koby, S., Prag, G., Engelhorn, M., Geiselmann, J., and Oppenheim, A. B. (1998). Participation of IHF and a distant UP element in the stimulation of the phage lambda PL promoter. Mol Microbiol *30*, 443-451.

Goodman, S. D., Nicholson, S. C., and Nash, H. A. (1992). Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. Proc Natl Acad Sci U S A *89*, 11910-11914.

Goodman, S. D., and Scocca, J. J. (1989). Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of Haemophilus influenzae. J Bacteriol *171*, 4232-4240.

Goodman, S. D., Velten, N. J., Gao, Q., Robinson, S., and Segall, A. M. (1999). In vitro selection of integration host factor binding sites. J Bacteriol *181*, 3246-3255.

Gopaul, D. N., and Duyne, G. D. (1999). Structure and mechanism in site-specific recombination. Curr Opin Struct Biol *9*, 14-20.

Grainge, I., and Jayaram, M. (1999). The integrase family of recombinase: organization and function of the active site. Mol Microbiol *33*, 449-456.

Grimwade, J. E., Ryan, V. T., and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled oriC of Escherichia coli. Mol Microbiol *35*, 835-844.

Gronostajski, R. M., and Sadowski, P. D. (1985). The FLP recombinase of the Saccharomyces cerevisiae 2 microns plasmid attaches covalently to DNA via a phosphotyrosyl linkage. Mol Cell Biol *5*, 3274-3279.

Grosschedl, R., Giese, K., and Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet *10*, 94-100.

Groth, A. C., and Calos, M. P. (2004). Phage integrases: biology and applications. J Mol Biol *335*, 667-678.

Guo, F., Gopaul, D. N., and van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature *389*, 40-46.

Guo, F., Gopaul, D. N., and Van Duyne, G. D. (1999). Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. Proc Natl Acad Sci U S A *96*, 7143-7148.

Hallet, B., and Sherratt, D. J. (1997). Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS Microbiol Rev *21*, 157-178.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol *166*, 557-580.

Hardman, C. H., Broadhurst, R. W., Raine, A. R., Grasser, K. D., Thomas, J. O., and Laue, E. D. (1995). Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry *34*, 16596-16607.

Hashimoto-Gotoh, T., Franklin, F. C., Nordheim, A., and Timmis, K. N. (1981). Specific-purpose plasmid cloning vectors. I. Low copy number, temperature-sensitive, mobilization-defective pSC101-derived containment vectors. Gene *16*, 227-235.

Hickman, A. B., Waninger, S., Scocca, J. J., and Dyda, F. (1997). Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. Cell *89*, 227-237.

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

Hwang, E. S., and Scocca, J. J. (1990). Interaction of integration host factor from Escherichia coli with the integration region of the Haemophilus influenzae bacteriophage HP1. J Bacteriol *172*, 4852-4860.

Ilves, H., Horak, R., Teras, R., and Kivisaar, M. (2004). IHF is the limiting host factor in transposition of Pseudomonas putida transposon Tn4652 in stationary phase. Mol Microbiol *51*, 1773-1785.

Johnson, R. C., Bruist, M. F., and Simon, M. I. (1986). Host protein requirements for in vitro site-specific DNA inversion. Cell *46*, 531-539.

Jones, J. C., Zhen, W. P., Reed, E., Parker, R. J., Sancar, A., and Bohr, V. A. (1991). Gene-specific formation and repair of cisplatin intrastrand adducts and interstrand cross-links in Chinese hamster ovary cells. J Biol Chem *266*, 7101-7107.

Jordan, P., and Carmo-Fonseca, M. (2000). Molecular mechanisms involved in cisplatin cytotoxicity. Cell Mol Life Sci *57*, 1229-1235.

Kahmann, R., Rudt, F., and Mertens, G. (1984). Substrate and enzyme requirements for in vitro site-specific recombination in bacteriophage mu. Cold Spring Harb Symp Quant Biol *49*, 285-294.

Kolot, M., Meroz, A., and Yagil, E. (2003). Site-specific recombination in human cells catalyzed by the wild-type integrase protein of coliphage HK022. Biotechnol Bioeng *84*, 56-60.

Kuhstoss, S., and Rao, R. N. (1991). Analysis of the integration function of the streptomycete bacteriophage phi C31. J Mol Biol *222*, 897-908.

Kumar, A., Galaev, I., and Mattiasson, B. (1999). Purification of Lac repressor protein using polymer displacement and immobilization of the protein.

Bioseparation *8*, 307-316.

Kwon, H. J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. Science *276*, 126-131.

Leffers, G. G., Jr., and Gottesman, S. (1998). Lambda Xis degradation in vivo by Lon and FtsH. J Bacteriol *180*, 1573-1577.

Leonard, A. C., and Grimwade, J. E. (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. Mol Microbiol *55*, 978-985.

Leong, J. M., Nunes-Duby, S., Lesser, C. F., Youderian, P., Susskind, M. M., and Landy, A. (1985). The phi 80 and P22 attachment sites. Primary structure and interaction with Escherichia coli integration host factor. J Biol Chem *260*, 4468-4477.

Lewis, J. A., and Hatfull, G. F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. Nucleic Acids Res *29*, 2205-2216.

Li, L., Farnet, C. M., Anderson, W. F., and Bushman, F. D. (1998). Modulation of activity of Moloney murine leukemia virus preintegration complexes by host factors in vitro. J Virol *72*, 2125-2131.

Li, S., and Waters, R. (1998). Escherichia coli strains lacking protein HU are UV

sensitive due to a role for HU in homologous recombination. J Bacteriol *180*, 3750-3756.

Lilley, D. M. (1992). DNA--protein interactions. HMG has DNA wrapped up. Nature *357*, 282-283.

Liu, Z., Li, Z., Zhou, H., Wei, G., Song, Y., and Wang, L. (2005). Immobilization and condensation of DNA with 3-aminopropyltriethoxysilane studied by atomic force microscopy. J Microsc *218*, 233-239.

Lorbach, E., Christ, N., Schwikardi, M., and Dröge, P. (2000). Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. J Mol Biol *296*, 1175-1181.

Lorenz, M., Hillisch, A., Goodman, S. D., and Diekmann, S. (1999). Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence resonance energy transfer. Nucleic Acids Res *27*, 4619-4625.

Lu, Y. B., Datta, H. J., and Bastia, D. (1998). Mechanistic studies of initiator-initiator interaction and replication initiation. Embo J *17*, 5192-5200.

Lynch, T. W., Read, E. K., Mattis, A. N., Gardner, J. F., and Rice, P. A. (2003). Integration host factor: putting a twist on protein-DNA recognition. J Mol Biol *330*, 493-502.

Matsuura, M., Noguchi, T., Yamaguchi, D., Aida, T., Asayama, M., Takahashi, H.,

and Shirai, M. (1996). The sre gene (ORF469) encodes a site-specific recombinase responsible for integration of the R4 phage genome. J Bacteriol *178*, 3374-3376.

Miller, H. I., and Friedman, D. I. (1980). An E. coli gene product required for lambda site-specific recombination. Cell *20*, 711-719.

Miller, Jeffrey H. (1972). Experiments in Molecular Genetics. *Cold Spring Harbor Laboratory*

Morisato, D., and Kleckner, N. (1987). Tn10 transposition and circle formation in vitro. Cell *51*, 101-111.

Muir, R. E., and Gober, J. W. (2005). Role of integration host factor in the transcriptional activation of flagellar gene expression in Caulobacter crescentus. J Bacteriol *187*, 949-960.

Nash, H. A. (1975). Integrative recombination of bacteriophage lambda DNA in vitro. Proc Natl Acad Sci U S A *72*, 1072-1076.

Nash, H. A. (1981). Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. Annu Rev Genet *15*, 143-167.

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

Numrych, T. E., Gumport, R. I., and Gardner, J. F. (1990). A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage lambda. Nucleic Acids Res *18*, 3953-3959.

Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res *26*, 391-406.

Nunes-Duby, S. E., Smith-Mungo, L. I., and Landy, A. (1995). Single base-pair precision and structural rigidity in a small IHF-induced DNA loop. J Mol Biol *253*, 228-242.

Oberto, J., Drlica, K., and Rouviere-Yaniv, J. (1994). Histones, HMG, HU, IHF: Meme combat. Biochimie *76*, 901-908.

Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O., and Lippard, S. J. (1999). Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. Nature *399*, 708-712.

Pagel, J. M., Winkelman, J. W., Adams, C. W., and Hatfield, G. W. (1992). DNA topology-mediated regulation of transcription initiation from the tandem promoters of the ilvGMEDA operon of Escherichia coli. J Mol Biol *224*, 919-935.

Petyuk, V., McDermott, J., Cook, M., and Sauer, B. (2004). Functional mapping of Cre recombinase by pentapeptide insertional mutagenesis. J Biol Chem *279*,

37040-37048.

Preobrajenskaya, O., Boullard, A., Boubrik, F., Schnarr, M., and Rouviere-Yaniv, J. (1994). The protein HU can displace the LexA repressor from its DNA-binding sites. Mol Microbiol *13*, 459-467.

Radman-Livaja, M., Biswas, T., Ellenberger, T., Landy, A., and Aihara, H. (2006). DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. Curr Opin Struct Biol *16*, 42-50.

Radman-Livaja, M., Shaw, C., Azaro, M., Biswas, T., Ellenberger, T., and Landy, A. (2003). Arm sequences contribute to the architecture and catalytic function of a lambda integrase-Holliday junction complex. Mol Cell *11*, 783-794.

Reeves, R., and Adair, J. E. (2005). Role of high mobility group (HMG) chromatin proteins in DNA repair. DNA Repair (Amst) *4*, 926-938.

Reeves, R., Edberg, D. D., and Li, Y. (2001). Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. Mol Cell Biol *21*, 575-594.

Reeves, R., and Nissen, M. S. (1990). The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J Biol Chem *265*, 8573-8582.

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

Richet, E., Abcarian, P., and Nash, H. A. (1986). The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. Cell *46*, 1011-1021.

Ripper, K. (1997). Analysis of protein-DNA binding at equilibrium. B. I. F. Futura *12*, 20-26.

Robert, F., Gagnon, M., Sans, D., Michnick, S., and Brakier-Gingras, L. (2000). Mapping of the RNA recognition site of Escherichia coli ribosomal protein S7. Rna *6*, 1649-1659.

Robertson, C. A., and Nash, H. A. (1988). Bending of the bacteriophage lambda attachment site by Escherichia coli integration host factor. J Biol Chem *263*, 3554-3557.

Rutkai, E., Dorgai, L., Sirot, R., Yagil, E., and Weisberg, R. A. (2003). Analysis of insertion into secondary attachment sites by phage lambda and by int mutants with altered recombination specificity. J Mol Biol *329*, 983-996.

Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E., and Leonard, A. C. (2004). Escherichia coli prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. Mol Microbiol *51*, 1347-1359.

Ryan, V. T., Grimwade, J. E., Nievera, C. J., and Leonard, A. C. (2002). IHF and HU stimulate assembly of pre-replication complexes at Escherichia coli oriC by two different mechanisms. Mol Microbiol *46*, 113-124.

Saecker, R. M., and Record, M. T., Jr. (2002). Protein surface salt bridges and paths for DNA wrapping. Curr Opin Struct Biol *12*, 311-319.

Saitoh, Y., and Laemmli, U. K. (1994). Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. Cell *76*, 609-622.

Sam, M. D., Papagiannis, C. V., Connolly, K. M., Corselli, L., Iwahara, J., Lee, J., Phillips, M., Wojciak, J. M., Johnson, R. C., and Clubb, R. T. (2002). Regulation of directionality in bacteriophage lambda site-specific recombination: structure of the Xis protein. J Mol Biol *324*, 791-805.

Sarkar, D., Radman-Livaja, M., and Landy, A. (2001). The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. Embo J *20*, 1203-1212.

Sauer, B., and McDermott, J. (2004). DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. Nucleic Acids Res *32*, 6086-6095.

Schmidt, E. E., Taylor, D. S., Prigge, J. R., Barnett, S., and Capecchi, M. R. (2000). Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. Proc Natl Acad Sci U S A *97*, 13702-13707.

Segall, A. M., Goodman, S. D., and Nash, H. A. (1994). Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA

binding proteins. Embo J *13*, 4536-4548.

Seong, G. H., Kobatake, E., Miura, K., Nakazawa, A., and Aizawa, M. (2002). Direct atomic force microscopy visualization of integration host factor-induced DNA bending structure of the promoter regulatory region on the Pseudomonas TOL plasmid. Biochem Biophys Res Commun *291*, 361-366.

Shaikh, A. C., and Sadowski, P. D. (2000). Chimeras of the Flp and Cre recombinases: tests of the mode of cleavage by Flp and Cre. J Mol Biol *302*, 27-48.

Sheridan, S. D., Benham, C. J., and Hatfield, G. W. (1999). Inhibition of DNA supercoiling-dependent transcriptional activation by a distant B-DNA to Z-DNA transition. J Biol Chem *274*, 8169-8174.

Signon, L., and Kleckner, N. (1995). Negative and positive regulation of Tn10/IS10-promoted recombination by IHF: two distinguishable processes inhibit transposition off of multicopy plasmid replicons and activate chromosomal events that favor evolution of new transposons. Genes Dev *9*, 1123-1136.

Simpson, D. A., Hammarton, T. C., and Roberts, I. S. (1996). Transcriptional organization and regulation of expression of region 1 of the Escherichia coli K5 capsule gene cluster. J Bacteriol *178*, 6466-6474.

Smith, M. C., and Thorpe, H. M. (2002). Diversity in the serine recombinases. Mol Microbiol *44*, 299-307.

Smith-Mungo, L., Chan, I. T., and Landy, A. (1994). Structure of the P22 att site.

Conservation and divergence in the lambda motif of recombinogenic complexes. J Biol Chem *269*, 20798-20805.

Stenzel, T. T., MacAllister, T., and Bastia, D. (1991). Cooperativity at a distance promoted by the combined action of two replication initiator proteins and a DNA bending protein at the replication origin of pSC101. Genes Dev *5*, 1453-1463.

Stenzel, T. T., Patel, P., and Bastia, D. (1987). The integration host factor of Escherichia coli binds to bent DNA at the origin of replication of the plasmid pSC101. Cell *49*, 709-717.

Strauss, J. K., and Maher, L. J., 3rd (1994). DNA bending by asymmetric phosphate neutralization. Science *266*, 1829-1834.

Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J., and Wigley, D. B. (1997). Crystal structure of the site-specific recombinase, XerD. Embo J *16*, 5178-5187.

Surette, M. G., Lavoie, B. D., and Chaconas, G. (1989). Action at a distance in Mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). Embo J *8*, 3483-3489.

Swalla, B. M., Gumport, R. I., and Gardner, J. F. (2003). Conservation of structure and function among tyrosine recombinases: homology-based modeling of the lambda integrase core-binding domain. Nucleic Acids Res *31*, 805-818.

Swinger, K. K., Lemberg, K. M., Zhang, Y., and Rice, P. A. (2003). Flexible DNA

bending in HU-DNA cocrystal structures. Embo J *22*, 3749-3760.

Swinger, K. K., and Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. Curr Opin Struct Biol *14*, 28-35.

Teras, R., Horak, R., and Kivisaar, M. (2000). Transcription from fusion promoters generated during transposition of transposon Tn4652 is positively affected by integration host factor in Pseudomonas putida. J Bacteriol *182*, 589-598.

Thompson, J. F., Moitoso de Vargas, L., Koch, C., Kahmann, R., and Landy, A. (1987). Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. Cell *50*, 901-908.

Thompson, J. F., Waechter-Brulla, D., Gumport, R. I., Gardner, J. F., Moitoso de Vargas, L., and Landy, A. (1986). Mutations in an integration host factor-binding site: effect on lambda site-specific recombination and regulatory implications. J Bacteriol *168*, 1343-1351.

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S., and Calos, M. P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol *21*, 3926-3934.

Tirumalai, R. S., Healey, E., and Landy, A. (1997). The catalytic domain of lambda site-specific recombinase. Proc Natl Acad Sci U S A *94*, 6104-6109.

Travers, A. (1997). DNA-protein interactions: IHF--the master bender. Curr Biol *7*,

R252-254.

Treff, N. R., Pouchnik, D., Dement, G. A., Britt, R. L., and Reeves, R. (2004). High-mobility group A1a protein regulates Ras/ERK signaling in MCF-7 human breast cancer cells. Oncogene *23*, 777-785.

Trieschmann, L., Postnikov, Y. V., Rickers, A., and Bustin, M. (1995). Modular structure of chromosomal proteins HMG-14 and HMG-17: definition of a transcriptional enhancement domain distinct from the nucleosomal binding domain. Mol Cell Biol *15*, 6663-6669.

Van Duyne, G. D. (2001). A structural view of cre-loxp site-specific recombination. Annu Rev Biophys Biomol Struct *30*, 87-104.

van Ulsen, P., Hillebrand, M., Zulianello, L., van de Putte, P., and Goosen, N. (1996). Integration host factor alleviates the H-NS-mediated repression of the early promoter of bacteriophage Mu. Mol Microbiol *21*, 567-578.

van Ulsen, P., Hillebrand, M., Zulianello, L., van de Putte, P., and Goosen, N. (1997). The integration host factor-DNA complex upstream of the early promoter of bacteriophage Mu is functionally symmetric. J Bacteriol *179*, 3073-3075.

Vocke, C., and Bastia, D. (1983). Primary structure of the essential replicon of the plasmid pSC101. Proc Natl Acad Sci U S A *80*, 6557-6561.

Wang, H., Bash, R., Yodh, J. G., Hager, G. L., Lohr, D., and Lindsay, S. M. (2002). Glutaraldehyde modified mica: a new surface for atomic force microscopy of

chromatin. Biophys J *83*, 3619-3625.

Wang, H., and Chong, S. (2003). Visualization of coupled protein folding and binding in bacteria and purification of the heterodimeric complex. Proc Natl Acad Sci U S A *100*, 478-483.

Weinert, T. A., Derbyshire, K. M., Hughson, F. M., and Grindley, N. D. (1984). Replicative and conservative transpositional recombination of insertion sequences. Cold Spring Harb Symp Quant Biol *49*, 251-260.

Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R., Laue, E. D., and Thomas, J. O. (1993). Structure of the HMG box motif in the B-domain of HMG1. Embo J *12*, 1311-1319.

Weisberg, R. A., and A. Landy. (1983). Site-specific recombination in bacteriophage λ, p. 211-250. *In* R. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Wiater, L. A., and Grindley, N. D. (1988). Gamma delta transposase and integration host factor bind cooperatively at both ends of gamma delta. Embo J *7*, 1907-1911.

Wimberly, B. T., White, S. W., and Ramakrishnan, V. (1997). The structure of ribosomal protein S7 at 1.9 A resolution reveals a beta-hairpin motif that binds double-stranded nucleic acids. Structure *5*, 1187-1198.

Wojciak, J. M., Sarkar, D., Landy, A., and Clubb, R. T. (2002). Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain. Proc Natl Acad Sci U S A *99*, 3434-3439.

Wold, S., Crooke, E., and Skarstad, K. (1996). The Escherichia coli Fis protein prevents initiation of DNA replication from oriC in vitro. Nucleic Acids Res *24*, 3527-3532.

Wozniak, K., and Blasiak, J. (2002). Recognition and repair of DNA-cisplatin adducts. Acta Biochim Pol *49*, 583-596.

Wu, X., Fan, W., Xu, S., and Zhou, Y. (2003). Sensitization to the cytotoxicity of cisplatin by transfection with nucleotide excision repair gene xeroderma pigmentosun group A antisense RNA in human lung adenocarcinoma cells. Clin Cancer Res *9*, 5874-5879.

Yagil, E., Dolev, S., Oberto, J., Kislev, N., Ramaiah, N., and Weisberg, R. A. (1989). Determinants of site-specific recombination in the lambdoid coliphage HK022. An evolutionary change in specificity. J Mol Biol *207*, 695-717.

Yang, Q., and Catalano, C. E. (1997). Kinetic characterization of the strand separation ("helicase") activity of the DNA packaging enzyme from bacteriophage lambda. Biochemistry *36*, 10638-10645.

Yang, S. W., and Nash, H. A. (1995). Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. Embo J *14*, 6292-6300.

Yu, A., and Haggard-Ljungquist, E. (1993). Characterization of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. J Bacteriol *175*, 1239-1249.

Zamble, D. B., Mu, D., Reardon, J. T., Sancar, A., and Lippard, S. J. (1996). Repair of cisplatin--DNA adducts by the mammalian excision nuclease. Biochemistry *35*, 10004-10013.

Publications

1. Corona, T., Bao, Q., Christ, N., Schwartz, T., Li, J., and Drőge, P. (2003). Activation of site-specific DNA integration in human cells by a single chain integration host factor. Nucleic Acids Res *31*, 5140-5148.

2. Bao, Q., Christ, N., and Drőge, P. (2004). Single-chain integration host factors as probes for high-precision nucleoprotein complex formation. Gene *343*, 99-106.

3. Bao, Q., Chen, H., Liu, Y., Yan, J., Dröge, P., and Davey, C. A. (2006). A Divalent Metal-Mediated Switch Controlling Protein-Induced DNA Bending. J Mol Biol., in press.

Meeting Abstracts

1. Corona et al. (2003). *Use of recombinant "DNA master bender" IHF in eukaryotic cells*. "The Biology of DNA", February 2003, **Cold Spring Harbor Symposium. Poster Presentation**

2. Bao et al. (2004). *Single chain integration host factor variants as probes for high precision nucleoprotein complex formation.* Workshop on site-specific recombination and transposition. **Woods Hole, MA, USA. Poster Presentation**

3. Bao et al. (2006) *To Bend or Not to Bend DNA (Correctly)? A Tale From Engineered Single Chain Integration Host Factors.* Keystone conference on nuclear acid enzymes. **Taos, New Mexico, USA. Poster Presentation**