Supporting Information

Supplementary Materials and Methods

Preparation of the Minimal DNA Substrates. Oligonucleotide strands were synthesized with either an IRDye® 700 or 6-carboxyfluoresceinlabel (6-FAM) label at their 5' end for use in the electrophoretic mobility shift (EMS) and sedimentation velocity (SV) studies, respectively. The 27 bp and 75 bp duplexes used in these studies (**Figure 1B**) were assembled from individually synthesized oligonucleotide strands by mixing equivalent concentrations of the complementary strands. The sequence of the top strand for each duplex is shown below with the *I1* and *I2* elements indicated in bold/italic type. An asterisk indicates the position of the dye in each strand.

I1 substrate: $^{5'}$*-***CATAACTTAATGTTTTTATT***TAAAATA

I2 substrate: $^{5'}$*-TTTTCGCTATT***TATGAAAATTTTCCG***G

R3 substrate: $^{5'}$*-CGGTTTAAGGCGTTTCCGTTCTTCTTC

The R3 substrate does not contain an IHF consensus sequence and serves as a minimal non-specific duplex.

I2-R3-I1 substrate: $^{5'*-}$CTATT***TATGAAAATTTTCCG***GTTTAAGGCGTTTCCGTT
CTTCTTC GT***CATAACTTAATGTTTTTATT***TAAAATAC

R3-I1-R2 substrate: $^{5'*-}$AAGGCGTTTCCGTTCTTCTTCGT***CATAACTTAATGTT***
TTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAG

Preparation of the Full-Length cos DNA Substrates. A 274 bp duplex that encompasses the entire *cos* sequence (***cos*274**; **Figure 1B**) was prepared by large-scale preparative PCR using a *Taq* PCR kit (New England Biolabs). The template for the reactions was pAFP1, a plasmid that contains the entire *cos* sequence cloned into a pUC19 vector. The plasmid was purified from *E. coli* JM107[pAFP1] cells (a kind gift of Michael Feiss, University of Iowa) using a Qiagen Plasmid *Plus* Maxi kit. The PCR primers were as follows:

cos*274 forward primer:** $^{5'}$**-CCGGAATTCGCATGCCTGCAGGT

cos*274 reverse primer:** $^{5'}$**-CGCGAATTCCATTGTTCATTCCAC

where the asterisk indicates a 5' IRDye® 700 incorporated into both forward and reverse primers for use in the EMS studies. This amplifies bases 18 to 278 in pAFP1, which yields a 274 bp duplex containing the entire *cos* sequence (**Figure 1B**). PCR reaction mixtures (100 µL per tube, 1 mL total) contained each primer at a concentration of 0.5 µM and 10 ng of pAFP1 template. *Taq* DNA polymerase and deoxynucleotides were included per manufacturer's protocol and the PCR cycles were as follows: (94°C for 1 minute, 61°C for 1 minute, 72°C for 2 minutes) x 60 cycles. The PCR reaction mixtures were pooled and the DNA products were purified using a Promega Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison WI). The purity of isolated products was verified by agarose gel analysis. We note that the purified IRDye-labeled *cos*274 duplex contains a small amount of a contaminating non-specific PCR product (see Figure 2A). This band is unperturbed by IHF at concentrations less than 10 nM and was therefore not considered in the quantitation of F_{bound} in this IHF concentration range.

A 274 bp duplex of non-specific sequence (***ns*274**) was similarly prepared using the following primers:

ns*274 forward primer:** $^{5'}$--**GGGGAATTCACTGGCCGTCGTTTTA

ns*274 reverse primer:** $^{5'}$--**GGGGAATTCACTTAACTATGCGGCATCAGAGCA

which amplifies bases 272 to 533 in the pAFP1. These residues originate from the vector and provide a 274 bp duplex of non-specific sequence. PCR amplification and purification of ns274 was as described above for the cos274 duplex.

An identical protocol was used for the preparation of 6-FAM labeled *cos*274 and ns274 duplexes for use in the analytical ultracentrifugation studies, except that only the forward primer contained the 6-FAM dye and the annealing temperature was reduced to 58˚C. The final sequences of the *cos*274 "specific" and *ns*274 "non-specific" duplexes are as follows,

cos274: $^{5'}$***-**CCGGAATTCGCATGCCTGCAGGTCTAATCATTATCACTTTACGGGTCCTTTCCGG
TGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATT***TATGAAAAT***
TTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGT***CATAACTTAATGTTTTTATT***TAAAAT
ACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGGCTTTTTGGCCTCT
GTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAATTCGCG

ns274: $^{5'}$***-**GGGGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTT
ACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGA
AGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGG
CGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATG
GTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGTGAATTCCCC

Analysis of the EMS Binding Data Using a Hill Model. Where indicated, the EMS binding data were analyzed according to a phenomenological Hill model,

$$F_{bound} = (m - b) * \left[\frac{[protein]^n}{(K_{D,app})^n + [protein]^n} \right] + b \qquad \textbf{(S1)}$$

where *[protein]* is the concentration of free IHF in the reaction mixture (calculated by mass conservation), $K_{D,app}$ is the apparent dissociation constant, n is the Hill coefficient, b is the baseline offset, and m is the fraction of DNA bound at saturation. The experimental data were fit to **Equation S1** using the Igor Pro Data Analysis program (Wave Metrics, Lake Oswego, OR).

Mathematical Model for a Nucleoid to Linear Rod Transition. A simple model to describe the ensemble of nucleoprotein complexes present in solution during an IHF titration experiment is presented in **Scheme 1**. This model assumes that there are three DNA species in solution during an IHF titration study; (i) Free/unbound DNA (***DNA$_F$***), (ii) compacted nucleoid IHF•DNA complexes (***DNA$_N$***), and (iii) the Extended Rod conformation (***DNA$_E$***). Specific, high affinity binding of one IHF dimer to *cos*274 affords the strongly bent complex that is not observed in the AUC data. This is likely due to a relatively small addition of mass onto *cos*274 DNA. As the protein concentration is increased, IHF dimers cooperatively assemble on both duplexes in a non-specific binding mode, which yields an ensemble of condensed, bent IHF•DNA intermediates, <DNA_N>. To accommodate a full complement of IHF dimers, the condensed duplex must be “unbent”, a transition that is described by an apparent equilibrium constant (K_L) driven by the binding of “m” IHF dimers (a Hill binding model). Importantly, a complex

that contains a high-affinity specifically bound IHF dimer is energetically more difficult to unbend than a weakly bent, non-specific nucleoid complex (see **Figure 6C**). Thus, the apparent equilibrium constants $K_{L,ns}$ and K_{Lsp} must reflect the inherent stabilities of the DNA_N complexes formed between specific and non-specific DNA substrates, respectively, as depicted in **Scheme 1**.

To derive a mathematical description of this model, we consider that the weight average sedimentation coefficient observed in an AUC titration experiment reflects the ensemble of all the DNA complexes present in solution. Within the context of the model presented in **Scheme 1**, this can be expressed as,

$$\langle s^* \rangle = s^*_{free} * F_{free} + \langle s^*_{nucleoid} \rangle * F_{nucleoid} + \langle s^*_{extended} \rangle * F_{extended} \quad \textbf{(S2)}$$

where F_{free}, $F_{nucleoid}$, and $F_{extended}$ are the fractions of the DNA complexes present in solution during a titration experiment, s^*_{free} is the sedimentation coefficient for free DNA, $\langle s^*_{nucleoid} \rangle$ represents the weight average sedimentation coefficient for the ensemble of condensed complexes in solution, and $\langle s^*_{extended} \rangle$ represents the weight average sedimentation coefficient for the ensemble of extended complexes in solution. The initial complex with *specific* DNA is severely bent, but we assume that additional IHF dimers interact with all DNA species in the same way, i.e. non-specific binding with mild positive cooperativity. This is consistent with our EMS analysis. The total fraction of DNA *bound* by IHF is described by,

$$F_{bound} = F_{nucleoid} + F_{extended} \quad \textbf{(S3)}$$

and

$$F_{bound} = (1 - F_{free}) \quad \textbf{(S4)}$$

According to **Scheme 1**, the fraction of condensed and extended species in solution are described as follows:

$$F_{nucleoid} = \left(\frac{1}{1+K_L*[IHF]^m}\right) * \left(1 - F_{free}\right) \quad \textbf{(S5)}$$

$$F_{extended} = \left(\frac{K_L*[IHF]^m}{1+K_L*[IHF]^m}\right) * \left(1 - F_{free}\right) \quad \textbf{(S6)}$$

where *[IHF]* is the concentration of IHF added to the reaction mixture, K_L is the equilibrium association constant for the transition, and *m* represents the number of IHF dimers required to drive the transition (i.e., the "Hill coefficient"). F_{free} can be obtained from **Equation S4** and substitution of F_{free}, $F_{nucleoid}$ and $F_{extended}$ into **Equation S2** yields an expression that can be utilized to evaluate the AUC data.

Analysis of the AUC Binding Data According to the Unbending Model. The ensemble of AUC data for IHF binding to ns274 and *cos*274 DNA were globally fit to this model using NLLS approach. In the analysis *n,* ω, and K_{ns} were fixed as global constants at their experimentally determined values obtained in the EMS studies (**Table 1**), s^*_{free} was fixed as a global constant at its experimentally determined values from the AUC studies (4.2 S), $K_{L,ns}$ and $K_{L,sp}$ were allowed to float as a local variable in the ns274 and *cos*274 data, respectively; $\langle s^*_{nucleoid} \rangle$, $\langle s^*_{extended} \rangle$, and *m* were allowed to float to their best values by non-linear least squares analytical methods using the Scientist® data package (Micromath Scientific Software). The best fit of the data is shown as solid lines in **Figure 6C** and the derived parameters are presented in **Table S2**.

*Hydrodynamic Modeling of the IHF•DNA Nucleoprotein Complexes***.** A structural model for the *specific* IHF-I1 nucleoprotein complex was constructed starting with the crystal structure of IHF bound to the 35 bp H' element of *attP* (RCSB #1OWF) by deletion of four bp from each end of the duplex (**Figure S1A**). The resulting structural model was used to calculate the theoretical hydrodynamic properties of the IHF-I1 complex using WinHydroPro, Ver. 1[(1)] with the following settings: Shell model from atomic level coordinates, complex molecular weight= 39,610 Da, $\bar{\nu}$= 0.662, based on the weight average of protein ($\bar{\nu}$= 0.7336) and DNA ($\bar{\nu}$= 0.58) in the complex. This returned a sedimentation coefficient of s_{sp}= 2.36.

Simple structural models for non-specific IHF•R3 nucleoprotein complexes were constructed starting with the sequence of the minimal R3 duplex; the DNA structure was constructed *de novo* using the DNA Sequence to Structure tool, Supercomputing Facility for Bioinformatics and Computational Biology, IIT, Delhi (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp). The structual coordinates for the IHF dimer were extracted from the published crystal structure (RCSB #1OWF) and nucleoprotein models were constructed by manual docking of two (**Figure S1B**) or three (**Figure S1C**) IHF dimers on the duplex. The resulting models were used to calculate the theoretical hydrodynamic properties of the complexes using WinHydroPro, Ver. 1[(1)] with the following settings: Shell model from atomic level coordinates; complex molecular weight= 61,602 Da and 83,594 Da for the dimer and trimer, respectively; $\bar{\nu}$= 0.6824 and 0.6952 for the dimer and trimer, respectively, based on the weight average of protein ($\bar{\nu}$= 0.7336) and DNA ($\bar{\nu}$= 0.58) in the complex. This returned sedimentation coefficients of s_{ns}= 2.84 S and 3.52 S for the dimer and trimer complexes, respectively.

The analysis suggests that the non-specific 27 bp duplex can accommodate three IHF dimers, which is consistent with the site size previously determined (n=8 bp)[2].

Table S1 – Analysis of EMS Binding Data According to a Hill Model

Duplex Substrate [§]	K_{app}	*n*
*cos*274	(1.2 ± 0.4) x 10^9 M^{-1}	2 ± 1
I1	(9.6 ± 0.5) x 10^7 M^{-1}	2 ± 1
[R3-I1-R2]	(9.2 ± 0.4) x 10^8 M^{-1}	3 ± 1
[I2-R3-I1]	(1.0 ± 0.9) x 10^9 M^{-1}	5 ± 1

The EMS data presented in **Figure 4** were analyzed according to a Hill model as described in Materials and Methods.

[§.] The duplex substrates are as illustrated in **Figure 1B**.

Table S2 – Analysis of the ns274 and *cos*274 AUC Data to Approximate Nucleoid Unbending Energy.

	m	**<s*>**	***Adair Constant***	***Apparent $K_{L,app}$*** ‡	**ΔG** § (kcal/mol)
ns274	4	(10.0 ± 0.4) S	$K_{L,ns}$=(1.2 ± 0.8) x 10^{36} M^{-4}	$K_{L,app-ns}$= (1.0 ± 0.2) x 10^{9} M^{-1}	-11.5
*cos*274	4	(3.7 ± 0.4) S	$K_{L,sp}$= 3.7 x 10^{31} M^{-4}	$K_{L,app-sp}$ ≤ 7.8 x 10^{7} M^{-1}	≥ -10.1

The SV-AUC binding data for the *cos*274 and ns274 duplexes presented in **Figure 6C** were simultaneously analyzed (globally fit) according to the DNA Unbending Model (**Scheme 1**) as described above. The best fit of the ensemble of binding data are displayed as solid lines in **Figure 6C**.

‡. $K_{L,app}$ values were calculated from $\left(K_L\right)^{1/m}$.

§. The free energy change was calculated from $\Delta G = -RT \ln(K_{L,app})$.

Note: The $\left(\frac{K_{L,app-ns}}{K_{L,app-sp}}\right)$ ratio indicates that the presence of a specific IHF complex imparts an approximately ~13-fold increase in the stability of the nucleoid complexes towards unbending. This corresponds to a $\Delta\Delta G$ ~ 1.4 kcal/mol.

Figure S1. Structural Models Used in Hydrodynamic Modeling

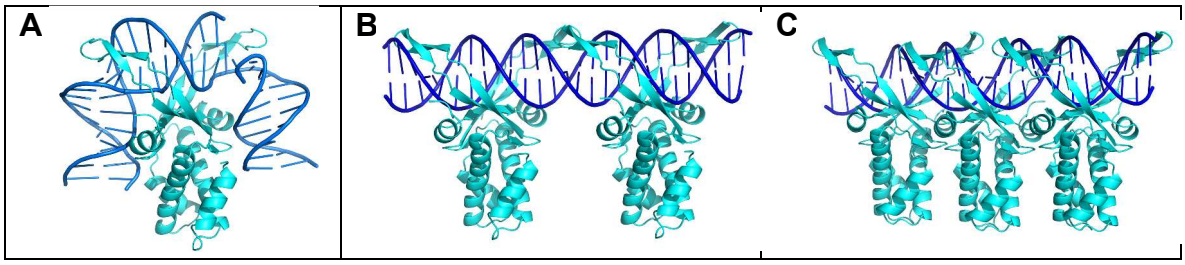
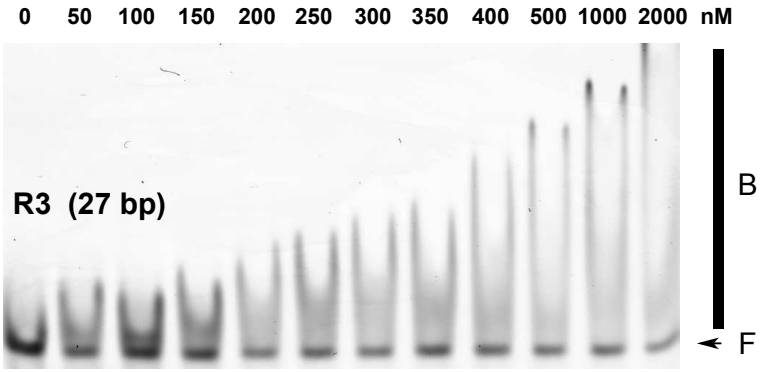


Figure S2 - EMS studies of IHF binding to the *R3* Duplex Substrate.



Representative agarose gel showing IHF binding to the minimal 27 bp R3 non-specific substrate. This duplex comprises the *R3* sequence, the dominant element associated with *terminase* binding to *cosB* (see **Figure 1B**). The element shows no sequence similarity to the IHF consensus sequence and serves as a minimal sized duplex of non-specific sequence. The positions of free (F) DNA and the bound (B) DNA complexes are indicated at the right of the gel image with an arrow and bar, respectively.

Literature Cited

1. Ortega, A., Amorós, D., and García de la Torre, J. (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models, *Biophysical Journal 101*, 892-898.

2. Holbrook, J. A., Tsodikov, O. V., Saecker, R. M., and Record Jr, M. T. (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, *Journal of Molecular Biology 310*, 379-401.