

# Indirect Recognition in Sequence-specific DNA Binding by *Escherichia coli* Integration Host Factor

## THE ROLE OF DNA DEFORMATION ENERGY\*

Received for publication, July 5, 2006, and in revised form, October 10, 2006 Published, JBC Papers in Press, October 11, 2006, DOI 10.1074/jbc.M606363200

Kimberly A. Aeling<sup>‡§1</sup>, Michael L. Opel<sup>‡§2</sup>, Nicholas R. Steffen<sup>‡¶1,2</sup>, Vira Tretyachenko-Ladokhina<sup>||</sup>,  
G. Wesley Hatfield<sup>‡§</sup>, Richard H. Lathrop<sup>‡¶</sup>, and Donald F. Seneor<sup>‡||3</sup>

From the <sup>‡</sup>Institute for Genomics and Bioinformatics, <sup>§</sup>Department of Microbiology and Molecular Genetics, School of Medicine, the <sup>¶</sup>Department of Computer Science, School of Information and Computer Sciences, and the <sup>||</sup>Department of Molecular Biology and Biochemistry, School of Biological Sciences, University of California, Irvine, California 92697

Integration host factor (IHF) is a bacterial histone-like protein whose primary biological role is to condense the bacterial nucleoid and to constrain DNA supercoils. It does so by binding in a sequence-independent manner throughout the genome. However, unlike other structurally related bacterial histone-like proteins, IHF has evolved a sequence-dependent, high affinity DNA-binding motif. The high affinity binding sites are important for the regulation of a wide range of cellular processes. A remarkable feature of IHF is that it employs an indirect readout mechanism to bind and wrap DNA at both the nonspecific and high affinity (sequence-dependent) DNA sites. In this study we assessed the contributions of pre-formed and protein-induced DNA conformations to the energetics of IHF binding. Binding energies determined experimentally were compared with energies predicted for the IHF-induced deformation of the DNA helix (DNA deformation energy) in the IHF-DNA complex. Combinatorial sets of *de novo* DNA sequences were designed to systematically evaluate the influence of sequence-dependent structural characteristics of the conserved IHF recognition elements of the consensus DNA sequence. We show that IHF recognizes pre-formed conformational characteristics of the consensus DNA sequence at high affinity sites, whereas at all other sites relative affinity is determined by the deformational energy required for nearest-neighbor base pairs to adopt the DNA structure of the bound DNA-IHF complex.

Site-specific DNA binding by regulatory proteins is a feature of the regulatory processes that maintain, expand, and express genetic information such as replication, recombination, transposition, and transcription. The chemical and physical mechanisms that underlie sequence-specific recognition of regulatory elements by cognate DNA-binding proteins are typically clas-

sified as direct *versus* indirect readout. The former refers primarily to hydrogen bonds between proteins and the unique extra-cyclic substituents at C-4 of pyrimidines, C-6 of purines, and N-7 of purines. These groups provide a base pair-specific pattern of hydrogen bond donors and acceptors in the major groove of DNA that can be directly read by a complementary pattern of amino acid side chain donors and acceptors. Indirect readout refers to recognition of aspects of DNA structure such as intrinsic curvature, topology of major and minor grooves, ordered water structures, local geometry of backbone phosphates, and flexibility or deformability. Because both the local DNA structure and energy to deform DNA are themselves intrinsic sequence-dependent properties, the conserved sequences that distinguish binding sites necessarily include contributions from both direct and indirect mechanisms. Consequently, although the contribution from indirect mechanisms is expected to be significant in protein-DNA complexes that feature substantial DNA deformation, it has proven difficult to evaluate these contributions quantitatively. A protein that relies exclusively, or primarily, on indirect readout would clearly be advantageous for this purpose.

The *Escherichia coli* integration host factor (IHF)<sup>4</sup> is one such example of a DNA-binding protein that relies on indirect readout for sequence-specific recognition. IHF is a small, basic (22 kDa) heterodimeric protein that belongs to a general class of histone-like DNA minor groove-binding proteins present in both prokaryotes and eukaryotes (1, 2). Like other members of its class, IHF forms a complex in which the DNA is wrapped around the protein, producing a bend of 180°. Bending plays an architectural role in the primary role for IHF in chromosome condensation (3). However, unlike other class members that exhibit no sequence specificity in DNA binding, IHF also binds in a sequence-specific manner to sites at which bending aids in the formation of higher order structures required for a variety of cellular functions such as site-specific recombination (4), transposition of mobile genetic elements (5), gene regulation (6), and DNA replication (7).

Evaluation of known IHF-binding sites, of which over 170 have been identified (8), has revealed a consensus DNA-binding motif consisting of several small clusters of conserved bases (9–11). These are located primarily in the 3'-half of the site.

\* This work was supported in part by Grant NIH68903 from the National Institutes of Health (to G. W. H.) and Grant MCB-9728186 from the National Science Foundation (to D. F. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> Recipient of a predoctoral fellowship.

<sup>2</sup> Recipient of postdoctoral fellowship from the Biomedical Informatics Training Program of the University of California, Irvine, Institute for Genomics and Bioinformatics.

<sup>3</sup> To whom correspondence should be addressed. Tel.: 949-824-8014; Fax: 949-824-8551; E-mail: dfseneor@uci.edu.

<sup>4</sup> The abbreviations used are: IHF, *E. coli* integration host factor; EMSA, electrophoretic mobility shift assay.

The two most highly conserved elements are the sequence WATCAA starting near the center of the site and the second sequence TTR located 4 bp in the 3' direction from WATCAA (9). Some IHF sites also contain a poly(A)-tract containing 4–6 adenines and located 8–9 bp in the 5' direction from the WATCAA element (12, 13). Despite a nonrandom distribution of bases throughout the sites, sequence consensus is limited to just these few elements. Sites that contain all of these elements bind IHF with affinities on the order 1 nM and are preferred over random sequences by a factor of ~2,500-fold (see Refs. 14–16 and see below).

A crystallographic model of IHF bound to a 34-bp DNA fragment containing the H' site of bacteriophage  $\lambda$  shows that IHF contacts DNA over the entire 34 bp because of the wrapping of the DNA (17). This structure highlights contacts with 26 backbone phosphate groups and extensive interactions in the minor groove. There are direct hydrogen bonds made to only 3 bp (indicated in red in Fig. 1); none involves a donor or acceptor that is unique to a particular base. Even recognition of the conserved elements can be explained based on sequence-dependent structural characteristics rather than as a consequence of direct readout interactions (17, 18). The dearth of base pair-specific interactions in the major groove supports the widely held view that sequence recognition by IHF is entirely via indirect readout.

Much of the bend is centered on two positions 9 bp apart at which proline side chains at the tip of the  $\beta$ -ribbon arm of each subunit are intercalated between a pair of bases in the minor groove of the DNA. These steps show a strong preference for the sequence ApA. Proline intercalation wedges these bases apart to generate sharp kinks at these individual steps. These steps can be considered to include a fourth consensus element. Combined, these consensus elements include only 15–16 bp. Thus, although a slight majority of the base pairs in IHF sites are not conserved, the entire sequence must conform to the wrapped topology of the DNA in the bound complex. The non-conserved sequence accounts for a still quite substantial DNA bend. A recent investigation relied on systematic evolution of ligands by exponential enrichment to identify preferred sequences in the 5'-half of the site, which is more lacking in consensus elements (19). Although this failed to define new preferred sequence elements, it did find 100-fold effects of sequence on affinity. These considerations suggest that flexibility, a whole system property, contributes to indirect recognition in IHF-DNA binding.

Previously, we applied a bioinformatics approach to explore the relationship between DNA flexibility and IHF binding. DNA sequences were threaded onto the structure of the H' site in the bound complex (17), and the energy difference between these DNA sequences in the bound conformation and the same sequences in their native, unbound conformations was estimated using the harmonic conformational potentials of Olson *et al.* (20). Results indicated lower average deformation energy and a narrower distribution for known IHF sites as compared with either sequences selected at random from the *E. coli* genome, or generated as random sequences matching the *E. coli* base composition (21). Deformation energy calculated in this manner was used to seed classifiers that could be trained to

identify IHF-binding sites (22). Subsequently, this result was extended to four other DNA-binding proteins, which feature highly degenerate consensus DNA-binding sequences and substantial DNA deformation in the bound complexes, and for which high resolution structures are available (23).

In this work, we have analyzed experimentally the connection between the free energy change for specific IHF binding and the deformation energy. The goal was to assess quantitatively the contribution from DNA flexibility to indirect readout recognition of specific IHF-binding sites. Initial analysis of data taken from the literature for 32 IHF sites that have been carefully characterized quantitatively suggested both the need and an approach to separate effects of DNA flexibility from sequence-specific, albeit still indirect, recognition mechanisms that pertain to the conserved elements. A subsequent analysis was carried out on *de novo* sequences designed either to maintain particular combinations of consensus elements while varying the remaining sequence to generate the widest possible range of deformation energies, or designed as control sequences in which consensus elements and nonconsensus sequences were varied interchangeably. An additional set of *E. coli* genomic sequences was also analyzed. These sequences were selected as putative, but as yet uncharacterized, IHF regulatory sites based on classifiers described previously (21, 22, 24). Like the control sequences and other natural IHF regulatory sites, these exhibit variation in both consensus elements and nonconserved sequences. Results obtained from this analysis detail a significant effect of DNA flexibility that is of increasing significance when fewer consensus elements are conserved.

## EXPERIMENTAL PROCEDURES

**Materials**—Buffer components and reagents were electrophoresis grade if available and reagent grade otherwise.

**IHF Purification and Activity**—IHF was purified according to Nash *et al.* (25). Aliquots were stored at  $-70^\circ\text{C}$ . The specific DNA binding activity was determined by conducting site titrations of IHF binding to a high affinity IHF-binding site ( $K_d \approx 1 \times 10^{-9}$  M). Binding assays were conducted using the electrophoretic mobility shift assay (EMSA) as described below. Separate reaction mixtures were prepared that contained  $1 \times 10^{-7}$  M  $^{32}\text{P}$ -end-labeled duplex DNA oligomer and either no IHF or aliquots from serial dilutions of an IHF stock solution. The volume integrated intensities of the bands corresponding to bound and free DNA were quantitated by PhosphorImager analysis as described below. IHF activity was assessed from the concentration necessary to shift 50% of the DNA to the bound complex. The more common approach is to assess activity from the lowest concentration that shifts 100% of the DNA to the bound complex. However, significant nonspecific binding by IHF at this higher concentration complicates the analysis (26).

**IHF-binding DNA Fragments**—Double-stranded oligonucleotides used in IHF binding assays were generated by annealing complementary single strand oligonucleotides. The double strand DNA oligomers were end-labeled with  $^{32}\text{P}$  at both 5' ends using T4 polynucleotide kinase. DNA sequences were 50 bp in length and conformed to the sequence template,

## Role of Deformation Energy in IHF Binding

5'-GTTGGCAT( $X_{34}$ )GAACAGGT-3'.  $X_{34}$  denotes a variable sequence corresponding to either different natural IHF-binding sites or selected based on adherence to different combinations of IHF consensus elements (11) and to otherwise maximize variation in deformation energy. The 8-bp flanking sequences are from the  $\lambda H'$  IHF-binding site; these were held constant. Only the central 34 bp were considered in calculation of deformation energy, as described by Steffan *et al.* (21, 22).

**IHF Binding Assays**—Equilibrium binding of IHF was monitored by using the EMSA or gel shift assay. Reaction mixtures were prepared containing  $1 \times 10^{-10}$  M duplex DNA oligomer and either no IHF or concentrations ranging from  $4 \times 10^{-11}$  to  $3 \times 10^{-7}$  M in binding buffer consisting of 40 mM Tris-HCl, pH 8.0, 4 mM MgCl<sub>2</sub>, 70 mM KCl, 0.1 mM dithiothreitol, and 2  $\mu$ g/ml bovine serum albumin. After 10 min of incubation at room temperature, the full reaction mixture volumes (20  $\mu$ l) were loaded with the current running on  $16 \times 20$  cm, 8% polyacrylamide gels (7.73% acrylamide, 0.26% *N,N'*-methylene bisacrylamide) in TAE buffer (0.04 M Tris acetate, pH 8.0, 0.001 M EDTA). Bound and free DNA species were separated by electrophoresis at constant 140 V for 2.5 h at room temperature. Typically, four independent titrations were conducted for each DNA sequence. The *de novo* sequences were investigated in groups of several at a time, together with the  $\lambda H'$  sequence, which was used as a reference for these studies.

Dried gels were used to expose PhosphorImager screens, which were subsequently scanned using a Bio-Rad Molecular Imager FX. Fractions of bound and free DNA were determined by volume integration of the corresponding electrophoretic bands using the Bio-Rad Quantity One software, version 4.2.1. DNA present in each lane was divided into three fractions. Specifically bound DNA was identified as the distinct mobility-shifted band. A smear of DNA present only at IHF high concentration was interpreted as nonspecifically bound (26). Background exposure was determined by analyzing the regions in the lane with no IHF that would correspond to where bound species would migrate. These background values were subtracted from the pixel values used in the volume integration.

**Numerical Analysis**—IHF binding was analyzed according to the competitive specific and nonspecific finite lattice model developed for IHF binding by Tsodikov *et al.* (27). Application of this model to mobility shift assays, for which the experimental observables are fractions of DNA free, of DNA bound specifically, and of DNA bound nonspecifically, yields the partition function given by Equation 1,

$$Z = 1 + \exp\left(\frac{\Delta G_{\text{spec}}}{RT}\right) + \sum_{j=1}^{N/n} \left( \prod_{i=1}^j \bar{s}_i \cdot \exp\left(\frac{\Delta G_{\text{non-spec}}}{RT}\right) \right) \quad (\text{Eq. 1})$$

where  $\Delta G_{\text{spec}}$  and  $\Delta G_{\text{non-spec}}$  are the free energy changes for intrinsic specific and nonspecific IHF binding, respectively. In

the third term, which accounts for nonspecifically bound species,  $N$  is the size of the oligonucleotide in base pairs (here 50),  $n$  is the number of base pairs occluded by a nonspecifically bound IHF dimer, and  $j$  indexes species by stoichiometry of bound IHF. The average number of available nonspecific binding sites for the  $i$ th binding event is given by Equation 2.

$$\bar{s}_i = N(1 - n\bar{v}) \left( \frac{N - n + 1}{N} \right) f^{n-1} \quad (\text{Eq. 2})$$

The last two terms in the product give the conditional probability of  $n-1$  empty lattice positions adjacent to any particular empty position in an infinite lattice, as formulated originally by McGhee and von Hippel (28) with a correction to account for end effects in a finite lattice (27);  $\bar{v}$  is the binding density in IHF per base pair.  $\Delta G_{\text{spec}}$  and  $\Delta G_{\text{non-spec}}$  values were obtained by analyzing the fractions of unbound and of specifically bound DNA according to Equations 3 and 4.

$$\theta_o = 1/Z \quad (\text{Eq. 3})$$

$$\theta_{\text{spec}} = \exp(\text{Ln}[IHF] - (\Delta G_{\text{spec}}/RT)) / Z \quad (\text{Eq. 4})$$

An issue in applying Equations 3 and 4 is the value of  $n$ , the site size. Both forward, or ligand to lattice, and also reverse, or lattice to ligand, titrations are needed to evaluate this term empirically. The mobility shift assay provides only the former. Holbrook *et al.* (26) report  $n = 9$  and  $n = 16$ , respectively, at 60 and at 100 mM KCl, ionic strengths that bracket our conditions. We found fitted values of  $\Delta G_{\text{spec}}$  and  $\Delta G_{\text{non-spec}}$  to be insensitive to the value of  $n$  over this range. Consequently, the data were analyzed with an intermediate value, fixed as  $n = 12$ .

The closed form expression for the conditional probability embedded in Equation 2 is an approximation of the exact numerical finite lattice isotherm derived by Epstein (29). Tsodikov *et al.* (27) find the difference to be within the limits of reasonable experimental uncertainty when the lattice size is at least severalfold larger than the site size, particularly at the low binding densities to which the mobility shift assay is sensitive. Nevertheless, we also analyzed the mobility shift data using a second formulation for nonspecific binding to assess whether this approximation in the nonspecific binding isotherm might affect fitted values of  $\Delta G_{\text{spec}}$ . For this purpose, we applied a Hill model as a purely phenomenological description of nonspecific binding, leading to Equations 5–8.

$$\theta_o = 1/Z_1 \quad (\text{Eq. 5})$$

$$\theta_1 = \exp(\text{Ln}[IHF] - (\Delta G_{\text{spec}}/RT)) / Z_1 \quad (\text{Eq. 6})$$

$$\theta_{2-n} = \exp\left(\frac{N_{\text{ave}} \cdot \text{Ln}[IHF] - (\Delta G_{\text{non-spec}}/RT)}{RT}\right) / Z_1 \quad (\text{Eq. 7})$$

$$Z_1 = 1 + \exp\left(\frac{\ln[IHF] - (\Delta G_{\text{spec}}/RT)}{RT}\right) + \exp\left(\frac{N_{\text{ave}} \cdot \ln[IHF] - (\Delta G_{\text{non-spec}}/RT)}{RT}\right) \quad (\text{Eq. 8})$$

In this model,  $N_{\text{ave}}$  accounts for both the average number of IHF dimers bound in the mobility-shifted bands obtained in the gel shifts and the lattice binding statistics.  $N_{\text{ave}}$  is not restricted to an integer value. Although incorrect chemically and so of little value to explain actual molecular behavior, this model does yield a reasonable phenomenological alternative to the competitive specific and nonspecific finite lattice model, *i.e.* one that conforms to the nonspecific binding phase well, and so yields a precise estimate of  $\Delta G_{\text{spec}}$ , the parameter of interest.

Nonlinear least squares analysis was conducted using the Origin 7 software (OriginLab Corp.). The Origin software estimates parameter values corresponding to a minimum in the variance. Joint confidence limits that account for correlation between parameters are estimated by adjusting each parameter individually while refitting the others to search for a variance ratio as predicted by the  $F$  statistic (30). Confidence limits reported correspond to the 95% interval. When global analysis of multiple experiments was conducted, normalized weights were calculated for the individual data from the square roots of the variances of separate fits to each of the individual experiments.

## RESULTS

*Analysis of Naturally Occurring IHF-binding Sites*—Fig. 1 shows an atomic model based on the crystallographic structure of IHF bound to the 34-bp DNA fragment containing the H' site of bacteriophage  $\lambda$  (17). Consensus elements are color-coded as described in the figure legend for ease of identification and discussion. The dominant features contributing to the DNA deformation are the sharp kinks at the ApA steps of the two proline intercalation sites, as shown in orange. Although the DNA flanking these sites returns to a canonical B-form structure, within half a helical turn in either direction, the different base pair steps are deformed to varying extents; bending is anisotropic both with respect to degree and direction. Approaches based on general DNA flexibility, even at individual steps, are inadequate to describe the energetic consequences of this situation. For this reason, we have applied the harmonic spring model of Olson *et al.* (20) to capture the molecular level detail of the structural model in estimating deformation energies. The reference structure for these calculations is the structure shown in Fig. 1. All other sequences were threaded onto this structure.

As a starting point for the analysis, we compared the free energy change for IHF binding to the deformation energy in a series of naturally occurring, specific binding sites. The sites included in this analysis meet three criteria. First, each site is involved in a specific regulatory process, such as described in the Introduction, thereby confirming the location-specific nature of binding. Second, the evidence for DNA binding was derived from direct DNA binding assays. Third, titrations were

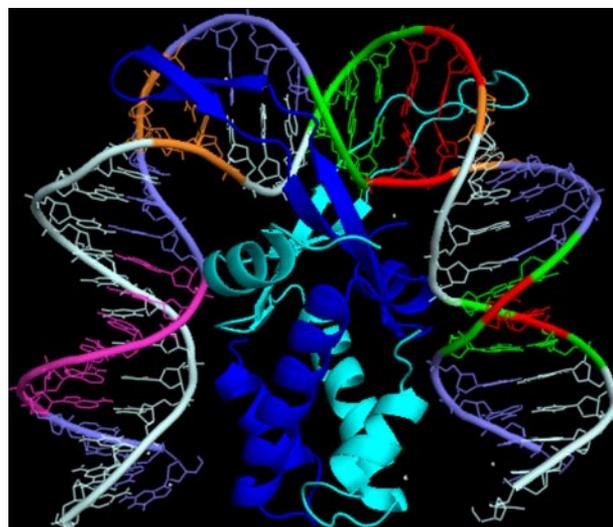


FIGURE 1. Model of IHF bound to form specific complex with  $\lambda$  H' site. The coordinate file for this structure, in which a crystallographic nick in the DNA (17) has been repaired, was supplied by Phoebe Rice (personal communication). The DNA backbone ribbon and nucleotides are color-coded to delineate consensus recognition elements as follows: magenta, A-tract; orange, ApA steps at proline intercalation sites; red, direct interaction; and green, remaining base pairs of consensus, WATCAAnnnTTR motif. The  $\alpha$  and  $\beta$  subunits of IHF are shown in cyan and blue, respectively.

conducted as necessary to provide quantitative estimation of the binding free energy. Table 1 presents the results compiled for 32 IHF-binding sites, the complete list for which reliable binding energies were obtained according to these criteria.

A concern in comparing literature values is the effect of different reaction conditions that were employed in the various studies. In general, variations in monovalent and divalent salt concentration, temperature, and pH would be expected to contribute significantly to apparent differences in DNA binding affinity. However, these IHF binding affinities are particularly advantageous in this regard. First, the majority of binding experiments were conducted within fairly narrow ranges of KCl concentration (50–70 mM),  $\text{MgCl}_2$  concentration (4–6 mM), temperature (20–25 °C), and pH (7.5–8.0), although some of the reaction conditions do fall outside these ranges. Second, IHF is unusual among DNA-binding proteins for its lack of salt dependence over the range of concentrations used in all of the experiments reported (26).<sup>5</sup> Third, where hard data exist, reported values were corrected to a standard set of reaction conditions.  $\Delta G^0$  values calculated from  $K_d$  values reported were corrected to a standard temperature (20 °C) using the standard thermodynamic relationship as shown in Equation 9,

$$\begin{aligned} \ln K_{a,\text{ref}} = \ln K_{a,T} + \frac{\Delta H_{\text{ref}}^0}{R} \left( \frac{1}{T} - \frac{1}{T_{\text{ref}}} \right) \\ - \frac{\Delta C_p^0}{R} \left[ \left( \frac{T_{\text{ref}} - T}{T} \right) + \ln \left( \frac{T}{T_{\text{ref}}} \right) \right] \quad (\text{Eq. 9}) \end{aligned}$$

where  $K_{a,\text{ref}}$  and  $K_{a,T}$  refer to the association equilibrium constants at the reference (293 K) and experimental temperatures,

<sup>5</sup> M. L. Opel, V. Tretyachenko-Ladokhina, L. M. Franklin, G. W. Hatfield, D. F. Senear, and J. B. A. Ross, submitted for publication.

## Role of Deformation Energy in IHF Binding

**TABLE 1**

Binding and deformation energies for IHF binding to natural sites in *E. coli*

IHF binding to regulatory sequences are as shown in *E. coli* genes. The consensus elements are shown in the first line for reference. W denotes A or T, and R denotes purine. Consensus elements are colored to match the structural model shown in Fig. 1. Values of the equilibrium dissociation constant ( $K_d$ ) are as reported by the authors when available or as estimated from the half-titration points of complete titrations. The free energy change for the binding reaction ( $\Delta G_b$ ) was calculated from  $\Delta G_b = RT \ln K_d$ .

Gene	Sequence	$K_d$ (nM)	$\Delta G_b$ (kcal/mol)	Def. Energy (kcal/mol)	Reference
consensus	nnnAAAAAn nnTTnnnWAT CAAnnnnTTR nnnn				
<i>yjbE</i>	ATTTAAATAA CAATAAAAT CATAAACCTTA AAAT	0.3	-12.8	112.6	(46)
<i>b3260-I</i>	CGATAATTAC TCATAAAAT CATCATATTA GAAA	0.9	-12.1	104.5	(47)
<i>ssuEADCB</i>	ATAGTCATTA GATGTTGAAT CAATGAATTA CTCG	0.9	-12.1	117.5	(48)
<i>glnHp2</i>	TTTGCCGCAT CTCGAAAAT CAAGGAGTTG CAAA	1.0	-12.1	114.4	(6,49)
$\lambda H'$	GCCAAAAAG CATTGCTTAT CAATTTGTTG CACC	1.4	-11.9	91.7	this work
<i>ilvGMEDA-1</i>	ACTATTTTAT TTTTAAAATA CAACAATTTA TATT	2.7	-11.5	100.5	(50)
<i>pspA</i>	TCAATCAGAT CTTTATAAAT CAAAAGATA AAAA	3.1	-11.4	114.7	(51-53)
<i>prf-A</i>	TGGTATTTTA CATGCACCTTA CAATTGATTA AAGA	3.4	-11.4	103.3	(54)
<i>tdcA</i>	AATTAATAAC CTTTAAATAT CAACAAGTTA AAGT	3.5	-11.3	115.2	(55)
<i>fimA-II</i>	AATAGAAATA ATTTACATAT CAACAGTTA GATG	3.7	-11.3	120.7	(56,57)
<i>cysJIH</i>	GGCTAAATTC ATTTGTTTTT CATTAGGTTG GTTA	3.8	-11.3	109.4	(58)
<i>gltBDF</i>	GATAAAATCC ATTTTAATTT CAGTCATTTA ATAA	4.2	-11.2	134.4	(59)
<i>ndh</i>	TTAACATAGC TCCTAATTA CAACAGGTTA CAGC	4.5	-11.2	101.0	(60)
<i>ndh II</i>	ATACAAAAT TGATGTTAAT CAAGTTTGT TGCT	4.5	-11.2	125.8	(60)
<i>narG</i>	GTATAAACG TCCTAATTTA CAGTCTGTTA TGTG	5.2	-11.1	119.0	(61,62)
<i>fimA-I</i>	TGTTTTTTTA ACTTTATAT CAATAAGTTA AATC	5.4	-11.1	121.4	(56,57)
<i>focAp1</i>	TTTATTTTTTA TTTGGATAAT CAAATATTTA CTCC	5.5	-11.1	99.9	(63)
<i>ompF-1</i>	TAATTTAAAT ATAAGGAAT CATATAAATA GATT	7.6	-10.9	125.9	(64)
<i>ilvGMEDA-2</i>	GACGAGGAAA GGCCTTAGAGA TCAAGCCTTA ACGA	8	-10.9	99.5	(65)
<i>b3260-II</i>	ACCAATCCTC TGAGATGGAT CAAAGAATTG CATT	9.4	-10.8	117.8	(47)
<i>hycA</i>	CTTAAAAATC TCCTTAATA CAATAAATTA AAAG	15	-10.5	108.8	(66)
<i>ompB-C</i>	TTGTTACAAA GCATATTTAA CAGCAGCTTA AGTA	15	-10.5	116.1	(67)
<i>ompCp1</i>	TAATAATTCA AGGTTAAAT CAATAACTTA TTCT	15	-10.5	128.8	(68)
<i>himAII</i>	GTAAGAAAGT TGGCGTAAAT CAGGTAGTTG GCGT	17	-10.4	114.7	(69)
<i>himD-II</i>	GAATAAAAAAT CCCGAAGAGT CAGAGAATTA CTCG	17	-10.4	120.6	(69)
<i>dps</i>	AAACAAAAT AATAAAGAT CAATGAGTTA GATA	18	-10.4	107.3	(70)
<i>nuoA</i>	ATGTGAAGAA ACTTTTGTTA AAGTTGACAA AAGG	23	-10.2	119.1	(71,72)
<i>ompB-A</i>	ATCAAACAGA CAAAGGGAAT CAACGAGATG AAAA	30	-10.1	116.4	(67)
<i>ompF-2</i>	ATATTTTTTC TTTTGAAC CAAATCTTTA TCTT	38	-10.0	128.3	(64)
<i>dmsABC</i>	CTACCTATCT CTTTGATTTT CAAATTATTC GATG	44	-9.9	115.9	(73)
<i>ompB-B</i>	TGATCATATC AACAGAATCA ATAATGTTTC GCCG	46	-9.8	131.8	(67)
<i>pstS</i>	TGCTTCTCGT AGCCAACAAA CAATGCTTTA TGAA	140	-9.2	132.8	(31)

and  $R$  is the gas constant. Record and co-workers (26) used isothermal titration calorimetry to determine  $\Delta H^0$  and  $\Delta C_p^0$  values for IHF binding to the  $\lambda H'$  sequence at different KCl concentrations. Interpolation was used to calculate values for the particular KCl concentrations used for the determination of each  $K_d$  value. We did not attempt to correct to a standard pH because  $d \ln K / d \ln [H^+]$  is not known for most of the applicable pH range.

Only a weak correlation between these calculated standard binding energies ( $\Delta G^0$ ) and deformation energy is evident in a plot of these data (Fig. 2). Linear regression (excluding one obvious outlier) yields a value of the correlation coefficient,  $R = 0.38$ . The positive correlation reflects the direction of the equilibria compared, *i.e.* in the direction of association *versus* energy to distort canonical *B*-form DNA. Thus, the correlation is in the expected direction that more deformable DNA provides higher

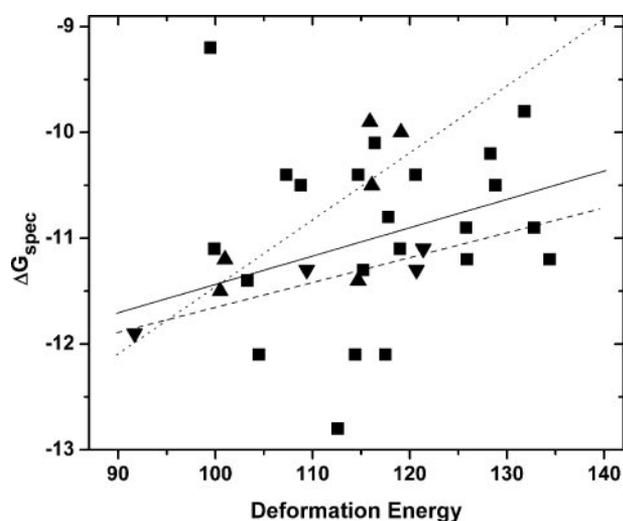


FIGURE 2. Standard free energy change ( $\Delta G_{\text{bind}}^0$ ) for IHF binding to naturally occurring specific sites in *E. coli*, plotted as a function of deformation energy. Values of  $\Delta G_{\text{bind}}^0$  were calculated from literature values of  $K_d$  listed in Table 1. Symbols and curves are as follows: downward triangles and dotted curve, sites that match all 9 bp of the consensus motif and contain ApA steps at both proline intercalation sites, and linear regression thereof; upward triangles and solid curve, sites that contain either 1- or 2-bp mismatches to the motif and contain ApA steps at both proline intercalation sites, and linear regression thereof; squares, all remaining sites; dashed curve, linear regression model for all 32 sites. Details of the linear regression analyses are given in the text.

affinity sites. However, the value of  $R^2$  suggests that only  $\sim 15\%$  of the variation in binding energy can be accounted on this basis. The remaining differences in reaction conditions could be a significant contributor to the variation in binding energy. In consideration, the  $R^2$  value probably understates the contribution of deformation energy to the variation in binding energy. Nevertheless, the relatively small  $R^2$  value suggests that variable composition of consensus elements in the different sites might play a dominant role in binding affinity.

In accordance with this expectation, a different picture emerges when variation in the complement of consensus elements is controlled. For example, among the 32 sites are 4 that contain all three direct interaction nucleotides, the ApA steps at both sites of proline intercalation and match the sequence-defined consensus at all 9-bp positions, differing only at degenerate positions ( $\lambda H'$ , *tdcA*, *fimA-II*, and *fimA-I* in Table 1).  $\Delta G^0$  for IHF binding to these sites varies over a range of 0.8 kcal/mol, or only 3-fold in affinity. Regression analysis of these sites yields  $R = 0.95$ , indicating that deformation energy accounts for 90% of this variation (Fig. 2) despite the small variation in  $\Delta G^0$ . The sequences of six additional sites contain either 1 or 2 mismatches to the sequence-defined consensus (*pspA*, *ilvGMEDA-1*, *ndh*, *hycA*, *ompF-2*, and *dmsABC* in Table 1). In general, the affinity for IHF binding to these sites is lower than when the consensus is matched completely, as might be expected, and the range of variation in  $\Delta G^0$  for IHF binding is much greater. Because both the position(s) of the mismatch(es) and the particular base(s) vary, a weaker correlation would also be anticipated. However, regression analysis yields  $R = 0.72$  indicating that deformation energy accounts for 52% of the variation in this case (Fig. 2). In contrast to these findings, sequences that vary more widely in sequence within the con-

sensus elements, or that are missing elements altogether, do not show similar correlation.

**Analysis of de Novo IHF Sites**—These results from analysis of natural IHF sites suggested a general approach to the analysis of the contribution from DNA flexibility in which recognition of specific sequences would be controlled by fixing specific combinations of consensus binding elements, whereas other sequence positions would be varied to yield the widest possible range of deformation energy. An advantage is that this yields a single set of internally consistent results at a standard condition, thus removing a large source of expected variation. In addition, it addresses an inherent limitation of the natural sequences, which is a relatively narrow range of deformation energy as compared with the entire *E. coli* genome (21). Sequences designed in this manner were compared with a set of control sequences that feature variation both in the conservation of consensus elements and also in nonconserved regions to generate a broad range of deformation energies. Finally, *E. coli* sites predicted to be high affinity IHF-binding sites by algorithms based in part on deformation energy (24) were also evaluated. Sequences and results for 49 sequences analyzed are presented in Table 2. These results represent the analysis of over 200 separate binding experiments.

Representative gel mobility shift data for two sites are shown in Fig. 3. A distinct band of lower mobility shown in Fig. 3A contains the specific IHF-DNA complex with the DNA wrapped around the protein as shown in Fig. 1. We confirmed this interpretation by conducting gel shift experiments as site titrations of IHF binding to DNA in which the 5'-phosphates at the two ends were conjugated with fluorescein and with tetramethylrhodamine to yield a donor-acceptor pair for Forster resonance energy transfer. DNA bending in the specifically bound complex brings the ends of the DNA to within about 55 Å resulting in efficient fluorescence resonance energy transfer that is absent in free DNA (32). Additional bands with successively decreasing mobility become evident at higher IHF concentrations. These are poorly resolved, thus generating a smear. These necessarily represent higher order complexes with more than one IHF dimer bound. A limiting band was observed at the highest IHF concentrations, indicating that saturation had been achieved.

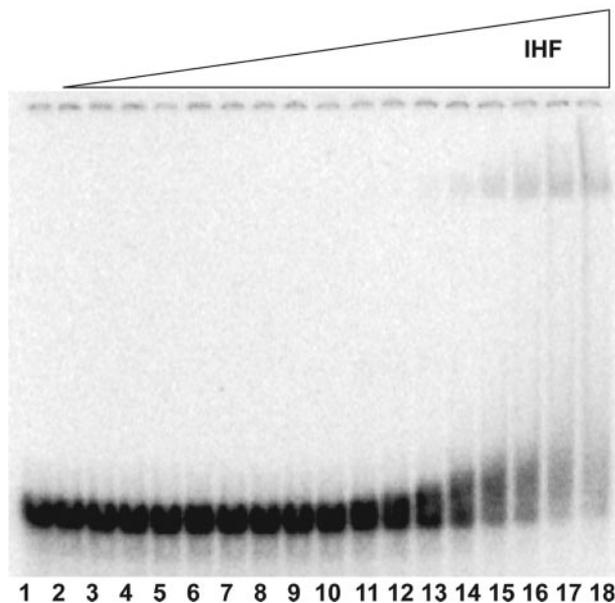
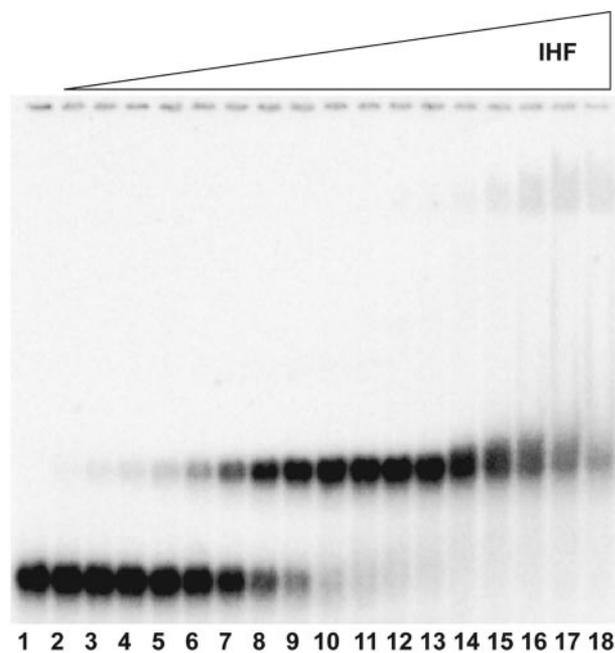
Record and co-workers (26) have described distinct specific and nonspecific IHF binding modes and used isothermal titration calorimetry to investigate the thermodynamics of both binding modes. Nonspecific binding does not produce the large DNA bend that is characteristic of the specific complex, thus yielding a much smaller length of contact with DNA and providing for simultaneous binding of multiple IHF dimers on a 50-bp oligonucleotide. At high IHF concentrations where IHF is in great molar excess over DNA, the nonspecific binding mode can compete successfully with the specific binding mode because of its higher stoichiometry, despite lower intrinsic affinity. Thus, the additional lower mobility bands in our experiments represent successive nonspecific associations by IHF. Different mobility bands might also represent different locations of bound IHF along the 50-bp DNA fragment. The concentration ranges over which IHF binds in specific and nonspecific modes overlap, particularly for the lower affinity specific

TABLE 2

Binding and deformation energies for IHF binding to *de novo* synthetic sites

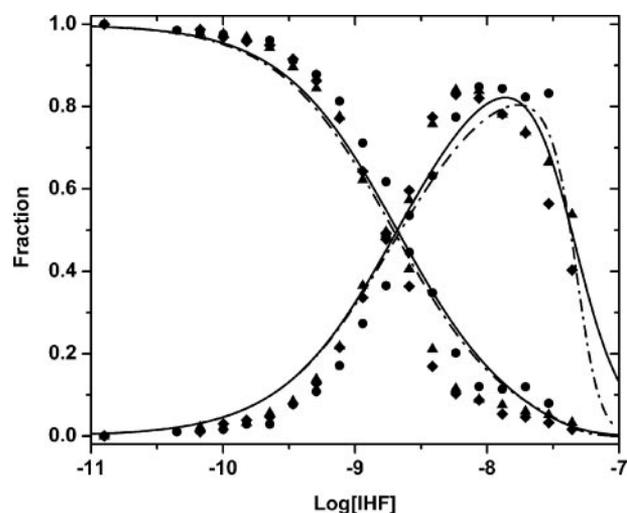
IHF binding to synthetic sequences designed to vary the deformation and to newly predicted sites in *E. coli*. Sequences within each series, A–G, maintain a particular complement of consensus sequence elements as described in the text, while varying the remaining sequence to generate the widest possible range of deformation energy. Consensus elements are color-coded as described in Table 1. Bases shown in lowercase were varied within a series. The consensus elements defined by the  $\lambda H'$  site used as a reference for these studies are shown in the first line for reference. W denotes A or T, and R denotes purine. Consensus elements are colored to match the structural model shown in Fig. 1. Values of the free energy change for specific ( $\Delta G_{spec}$ ) and nonspecific lattice binding ( $\Delta G_{non-spec}$ ) of IHF along with the 95% confidence intervals were obtained by global analysis of mobility shift titrations using Equations 3 and 4 in the text.  $s$  is the square root of the variance of the fitted curves. NS is nonspecific.

Site	Sequence	$\Delta G_{spec}$ (kcal/mol)	$\Delta G_{non-spec}$ (kcal/mol)	$s$	Def. Energy (kcal/mol)
consensus	NNNAAAAAN NN <sup>TTNCT</sup> TAT CAA <sup>NNNN</sup> TTG NNNN				
<b>Series A: Sequences contain all recognition elements</b>					
A.1	ccgAAAAA <sup>c</sup> ca <sup>TT</sup> gct <sup>TAT</sup> CAA <sup>tttg</sup> TTG gcgc	-11.64 ± 0.044	-7.82 ± 0.042	0.004	82.9
A.2	ccgAAAAA <sup>c</sup> tg <sup>TT</sup> act <sup>TAT</sup> CAA <sup>cgcg</sup> TTG cacc	-12.20 ± 0.063	-7.89 ± 0.055	0.013	86.7
A.3	gccAAAAA <sup>g</sup> ca <sup>TT</sup> gct <sup>TAT</sup> CAA <sup>tttg</sup> TTG cacc	-11.71 ± 0.040	-7.70 ± 0.067	0.012	91.7
A.4	gagAAAAA <sup>c</sup> aa <sup>TT</sup> act <sup>TAT</sup> CAA <sup>gggc</sup> TTG accg	-12.08 ± 0.060	-7.73 ± 0.115	0.009	110.1
A.5	aaaAAAAA <sup>g</sup> at <sup>TT</sup> tac <sup>TAT</sup> CAA <sup>tttg</sup> TTG atct	-11.40 ± 0.102	-7.46 ± 0.142	0.015	133.7
A.6	cgtAAAAA <sup>g</sup> at <sup>TT</sup> tct <sup>TAT</sup> CAA <sup>gtac</sup> TTG atgg	-11.60 ± 0.072	-7.31 ± 0.113	0.019	141.9
<b>Series B: Sequences lack poly-A tract</b>					
B.1	ggcgtttg <sup>tt</sup> tg <sup>TT</sup> aa <sup>TAT</sup> CAA <sup>tttg</sup> TTG tggc	-11.11 ± 0.067	-7.57 ± 0.125	0.014	90.2
B.2	aaatatgagg <sup>at</sup> TT <sup>tac</sup> TAT CAA <sup>tttg</sup> TTG atct	-10.56 ± 0.061	-7.11 ± 0.054	0.072	144.7
<b>Series C: Sequences lack poly-A tract and 5' proline intercalation site</b>					
C.1	tcagagatca <sup>tt</sup> agga <sup>TAT</sup> CAA <sup>TTTG</sup> TTG CAAC	-11.66 ± 0.124	-7.41 ± 0.110	0.022	98.7
C.2	at <sup>ttg</sup> tagtag <sup>gc</sup> ctta <sup>TAT</sup> CAA <sup>TTTG</sup> TTG CAAC	-10.26 ± 0.148	-7.54 ± 0.157	0.019	112.8
<b>Series D: Sequences contain alternative consensus and lack poly-A tract</b>					
D.1	atctgt <sup>tac</sup> ggcca <sup>ag</sup> AAT CAA <sup>tttg</sup> TTG caac	-10.96 ± 0.197	-7.83 ± 0.247	0.018	97.4
D.2	gattog <sup>tct</sup> gtccca <sup>ag</sup> AAT CAA <sup>tttg</sup> TTG caac	-10.60 ± 0.122	-7.51 ± 0.142	0.009	104.1
D.3	ggatct <sup>g</sup> ctc <sup>g</sup> aatcc <sup>ag</sup> AAT CAA <sup>gcgc</sup> TTG gcgc	-10.55 ± 0.279	-7.45 ± 0.366	0.027	142.4
<b>Series E &amp; F: Sequences replace consensus with either variable (E) or fixed (F) alternate sequence</b>					
E.2	gccAAAAA <sup>g</sup> ca <sup>TT</sup> gacc <sup>ct</sup> CAA <sup>tttg</sup> gc <sup>Tt</sup> cacc	-10.08 ± 0.139	-7.13 ± 0.121	0.012	110.5
E.1/F.1	gccAAAAA <sup>g</sup> ca <sup>TT</sup> gca <sup>acg</sup> CAA <sup>tttg</sup> t <sup>g</sup> cacc	-10.77 ± 0.135	-7.39 ± 0.153	0.011	82.6
F.2	gggAAAAA <sup>c</sup> tt <sup>TT</sup> tca <sup>ACG</sup> CAA <sup>acga</sup> TTG gagc	-9.55 ± 0.185	-7.17 ± 0.147	0.006	107.5
F.3	cgtAAAAA <sup>g</sup> at <sup>TT</sup> tca <sup>ACG</sup> CAA <sup>gtac</sup> TTG atgg	-9.43 ± 0.122	-6.88 ± 0.112	0.008	132.8
<b>Series G: Sequences lack poly-A tract and replace consensus with fixed alternate sequence</b>					
G.1	ggcgaaccat <sup>cg</sup> TT <sup>gca</sup> ACG CAA <sup>cgcg</sup> TTG gcgc	-10.32 ± 0.127	-7.24 ± 0.132	0.017	77.2
G.2	aaatagg <sup>tccg</sup> ac <sup>TT</sup> tca <sup>ACG</sup> CAA <sup>ccag</sup> TTG gcct	NS			114.3
G.3	aaatatgagg <sup>at</sup> TT <sup>tca</sup> ACG CAA <sup>gcgc</sup> TTG atct	NS			151.3
<b>Series H: Control sequences vary both complement of recognition elements and non-conserved bases</b>					
H.1	ccgAAAAA <sup>c</sup> ca <sup>TT</sup> gctc <sup>cg</sup> CAA <sup>cgca</sup> Tc gcgc	-10.10 ± 0.218	-6.95 ± 0.205	0.012	75.0
H.2	cccg <sup>ttt</sup> gtc <sup>ca</sup> TT <sup>gctc</sup> cg CAA <sup>cgca</sup> Tc gcgc	-9.98 ± 0.129	-7.17 ± 0.122	0.005	76.1
H.3	ccgAAAAA <sup>c</sup> cacgg <sup>ct</sup> TAT CAA <sup>tttg</sup> TTG gcgc	-11.30 ± 0.059	-7.30 ± 0.055	0.008	85.7
H.4	attc <sup>cccc</sup> ca <sup>cg</sup> goc <sup>ta</sup> ac gggc <sup>gc</sup> gga <sup>gg</sup> gggg	-9.57 ± 0.109	-6.75 ± 0.102	0.007	88.3
H.5	gtat <sup>aaa</sup> cg <sup>tc</sup> TT <sup>aatt</sup> ta CAG <sup>tctg</sup> TTA tgtg	-11.20 ± 0.100	-7.64 ± 0.102	0.018	99.9
H.6	gtaaa <sup>acc</sup> gt <sup>tact</sup> ct <sup>tc</sup> acgt <sup>cgc</sup> TTA taaa	-12.72 ± 0.147	-7.81 ± 0.161	0.026	101.4
H.7	cgcc <sup>cg</sup> gcgc <sup>gg</sup> gcgc <sup>TAT</sup> CAA <sup>gggt</sup> TTG gtgg	NS			111.9
H.8	agtca <sup>att</sup> tc <sup>cc</sup> TT <sup>cct</sup> TAT ta <sup>gcgc</sup> TTA cgga	-11.91 ± 0.137	-7.71 ± 0.140	0.024	112.5
H.9	ttgata <sup>agcc</sup> at <sup>tt</sup> agt <sup>cga</sup> tg <sup>gcgcgc</sup> gc <sup>ct</sup> tact	NS			123.1
H.10	atagaa <sup>attg</sup> aa <sup>caag</sup> tt <sup>ct</sup> gct <sup>ctct</sup> tg <sup>ct</sup> gc <sup>tg</sup>	-10.36 ± 0.165	-7.56 ± 0.152	0.010	126.8
H.11	ccgagga <sup>aa</sup> g <sup>gc</sup> tt <sup>aga</sup> GAT CAA <sup>gcct</sup> TAA cgaa	-11.23 ± 0.133	-7.66 ± 0.186	0.019	127.9
H.12	tatt <sup>taa</sup> ata <sup>tt</sup> cat <sup>ga</sup> AAT CTA <sup>taaa</sup> TTA aaga	NS			136.0
H.13	tcagt <sup>acca</sup> ca <sup>act</sup> tt <sup>ct</sup> gg <sup>at</sup> ct <sup>ctgc</sup> cc <sup>ct</sup> ct <sup>tg</sup>	-9.38 ± 0.089	-6.68 ± 0.072	0.002	166.3
H.14	aatAAAAA <sup>g</sup> at <sup>TT</sup> tac <sup>ccc</sup> CAA <sup>gcgc</sup> TA <sup>t</sup> atct	-10.10 ± 0.110	-6.97 ± 0.095	0.003	166.9
H.15	aaatct <sup>gtg</sup> g <sup>at</sup> TT <sup>tacccc</sup> CAA <sup>gcgc</sup> TA <sup>t</sup> atct	-10.18 ± 0.179	-7.31 ± 0.178	0.008	173.9
<b>Putative E. coli regulatory sites</b>					
lysU	ACGATTA <sup>ACT</sup> GTCGTA <sup>AAA</sup> CAA <sup>TTGG</sup> TTA TGAA	-12.28 ± 0.148	-7.49 ± 0.120	0.018	109.3
cysJ	GGCT <sup>AAAT</sup> TC AT <sup>TT</sup> GT <sup>TTT</sup> CATTAG <sup>TTG</sup> GTTA	-11.66 ± 0.178	-7.44 ± 0.161	0.025	115.2
nuvA	ACGTAT <sup>ATT</sup> GTATCT <sup>ATT</sup> CAA <sup>TGGA</sup> ATG ACAA	-11.38 ± 0.154	-7.40 ± 0.201	0.020	117.8
cadB	AGAT <sup>AA</sup> CAT TTAAT <sup>GT</sup> TTA CAA <sup>TGGA</sup> TTG CGTG	-11.47 ± 0.236	-7.79 ± 0.221	0.016	120.4
katG	TTGATCTACA TCTCT <sup>TTAA</sup> C CAA <sup>CAAT</sup> ATG TAAG	NS			120.7
yhgL	AGCAGCT <sup>TAA</sup> TT <sup>TT</sup> TTA <sup>ATT</sup> CATTG <sup>TT</sup> TTA AAAA	-10.87 ± 0.081	-7.43 ± 0.075	0.006	122.3
araE	TGAATT <sup>ATTG</sup> AGATT <sup>TAT</sup> TAA <sup>CCAC</sup> CTA ATTT	NS			125.4
dnaK	TCATC <sup>AAAA</sup> ATTGTA <sup>TAA</sup> AAA <sup>GAAA</sup> AGA TTAC	-12.24 ± 0.133	-7.24 ± 0.105	0.015	131.2
tnaL	GTAAT <sup>TTATA</sup> ATCT <sup>TTA</sup> AAA AA <sup>AGCA</sup> T <sup>TTA</sup> ATAT	NS			132.7
glbB	GAT <sup>AAA</sup> TCC AT <sup>TT</sup> TA <sup>ATT</sup> CAGTCAT <sup>TTA</sup> ATAA	-12.32 ± 0.132	-7.73 ± 0.094	0.014	134.4
narG	GTAT <sup>AAA</sup> CG TC <sup>TT</sup> AA <sup>TTA</sup> CAGTCT <sup>GTTA</sup> TGTG	-12.65 ± 0.137	-7.65 ± 0.118	0.021	99.9
cysP	GTA <sup>AA</sup> CCGT TACTC <sup>TTT</sup> ACGTC <sup>CGTTA</sup> TAAA	NS			101.4
osmY	AGTCAAT <sup>TT</sup> C CC <sup>TT</sup> CCT <sup>TAT</sup> TAGCC <sup>CGTTA</sup> CGGA	-11.73 ± 0.153	-7.45 ± 0.169	0.023	112.5
ssuE-II	TATTT <sup>AAA</sup> TA TTCAT <sup>GAA</sup> T CTATA <sup>AA</sup> TTA AAGA	-10.63 ± 0.099	-7.52 ± 0.108	0.011	136.0



**FIGURE 3. EMSA of IHF binding to 50-bp synthetic DNA sequences.** *Top panel*, IHF binding to the  $\lambda$   $H'$  site (fragment A.1). IHF concentrations are as follows: *lane 1*, no IHF; *lane 2*,  $4.8 \times 10^{-11}$  M; *lane 3*,  $8.3 \times 10^{-11}$  M; *lane 4*,  $1.4 \times 10^{-10}$  M; *lane 5*,  $2.5 \times 10^{-10}$  M; *lane 6*,  $4.4 \times 10^{-10}$  M; *lane 7*,  $7.6 \times 10^{-10}$  M; *lane 8*,  $1.3 \times 10^{-9}$  M; *lane 9*,  $2.3 \times 10^{-9}$  M; *lane 10*,  $4.0 \times 10^{-9}$  M; *lane 11*,  $6.9 \times 10^{-9}$  M; *lane 12*,  $1.2 \times 10^{-8}$  M; *lane 13*,  $2.1 \times 10^{-8}$  M; *lane 14*,  $3.7 \times 10^{-8}$  M; *lane 15*,  $6.4 \times 10^{-8}$  M; *lane 16*,  $1.1 \times 10^{-7}$  M; *lane 17*,  $1.9 \times 10^{-7}$  M; *lane 18*,  $3.3 \times 10^{-7}$  M. *Bottom panel*, fragment G.2, a sequence to which IHF binds only in the nonspecific mode. IHF concentrations are as follows: *lane 1*, no IHF; *lane 2*,  $9.4 \times 10^{-11}$  M; *lane 3*,  $1.6 \times 10^{-10}$  M; *lane 4*,  $2.6 \times 10^{-10}$  M; *lane 5*,  $4.4 \times 10^{-10}$  M; *lane 6*,  $7.3 \times 10^{-10}$  M; *lane 7*,  $1.2 \times 10^{-9}$  M; *lane 8*,  $2.0 \times 10^{-9}$  M; *lane 9*,  $3.4 \times 10^{-9}$  M; *lane 10*,  $5.6 \times 10^{-9}$  M; *lane 11*,  $9.4 \times 10^{-9}$  M; *lane 12*,  $1.6 \times 10^{-8}$  M; *lane 13*,  $2.6 \times 10^{-8}$  M; *lane 14*,  $4.3 \times 10^{-8}$  M; *lane 15*,  $7.2 \times 10^{-8}$  M; *lane 16*,  $1.2 \times 10^{-7}$  M; *lane 17*,  $2.0 \times 10^{-7}$  M; *lane 18*,  $3.3 \times 10^{-7}$  M. The prominent mobility-shifted band in the *top panel* contains the specific complex with wrapped DNA. In comparison, mobility-shifted bands corresponding to nonspecific binding are evident in the *bottom panel*. Sequences and results are given in Table 2.

IHF sites. Consequently, it was necessary to analyze the binding resulting from both mechanisms to obtain accurate quantitative values for specific binding. Because the individual bands



**FIGURE 4. Analysis of EMSA data for IHF binding to fragment A.1** (see Table 2). Different symbols represent three independent titration experiments. The decreasing fraction represents the unliganded DNA band. The second fraction (peak) represents DNA with IHF specifically bound (e.g. mobility-shifted band in Fig. 3A). *Solid and broken curves* show the results of global analysis of these data according to the competitive specific and nonspecific finite lattice model and the phenomenological Hill model as described in the text ("Experimental Procedures"). These yielded values equal to  $-11.64 \pm 0.044$  and  $-11.65 \pm 0.020$  kcal/mol, respectively. Complete results from the analysis according to the competitive specific and nonspecific finite lattice model are listed in Table 2.

produced by nonspecific binding are poorly resolved electrophoretically and their identity is difficult or impossible to interpret as reflecting particular molecular species, these bands were combined into a single fraction comprising all nonspecifically bound DNA, without regard to number or location of bound IHF dimers. Thus, our analysis of the gel images generated three fractions, corresponding to unliganded oligonucleotides, the 1:1 specifically liganded complex, and nonspecifically liganded complexes.

Numerical analysis of the gel shift data applied the competitive specific and nonspecific finite lattice model (27) that is implemented in Equations 3 and 4 as described under "Experimental Procedures." Fig. 4 presents data from three independent titrations of one IHF binding sequence analyzed globally using this model. Fig. 4 also shows the results of analysis by the phenomenological Hill model described by Equations 5–8 ("Experimental Procedures"). The fitting results yield essentially identical transitions for specific binding and indistinguishable transitions for nonspecific binding. The former was found to be the case for all 49 DNA sequences evaluated. In all cases, the difference between the estimates of  $\Delta G_{\text{spec}}$  obtained from the two models was within the confidence intervals obtained by either model alone. Results for all 49 sequences are listed in Table 2.

It is significant that a tight distribution of nonspecific affinities was obtained for these sequences ( $\Delta G_{\text{nonspec}} = -7.4 \pm 0.3$  kcal/mol; see Table 2), a finding that is nicely consistent with the concept of nonspecific binding and therefore strongly supportive of the model. Equally significant, this value is exactly intermediate between those obtained previously from isothermal titration calorimetry experiments for salt conditions that flank what was used in these experiments (26). This favorable

## Role of Deformation Energy in IHF Binding

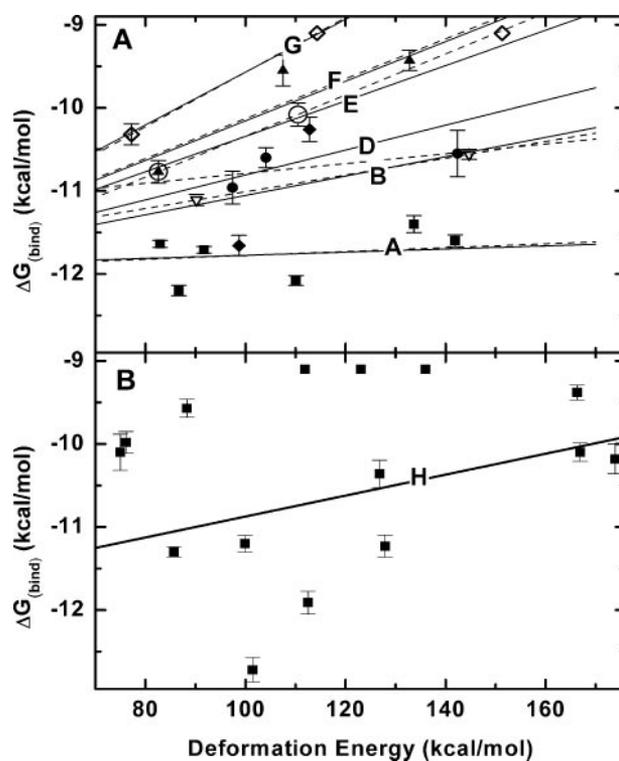
comparison between our results and those obtained using a thermodynamically rigorous method attests to the accuracy of the EMSA results we report.

The range of deformation energies exhibited by these sequences expands that of the natural IHF sites by more than 2-fold. These synthetic sites mimic the full range of deformation energies represented by the *E. coli* genome. Specific binding affinities vary by over 500-fold for these sequences. In nine cases, the specific complex band was not observed, indicating insufficient specific binding affinity to compete with nonspecific binding. Analysis of weakly specific binding sequences indicates that  $\Delta G_{\text{spec}}$  less than approximately  $-9.2$  kcal/mol can be resolved. On this basis we have assigned a value,  $-9.1$  kcal/mol, to represent the upper limit to the specific affinity for these sites. The actual specific binding affinity could be substantially weaker than this value indicates.

Taken as a whole, the data for all 49 sequences exhibit a very weak correlation between binding energy and deformation energy. The correlation coefficient is only 0.23, the same as for the natural sites, suggesting only 5–6% of the overall variation can be accounted as due to deformation energy. However, this value is adversely affected by the significant number of sequences (20%) for which low affinity specific binding is obscured by nonspecific binding. These sequences are skewed toward higher than average deformation energy, so that the limit to measurable specific affinity tends to decrease the apparent correlation between affinity and deformation energy. This effect is apparent in plots of the *de novo* sequences shown in Fig. 5. In support of this point, it is interesting to note that no high affinity sequence was found whose deformation energy exceeds the largest found among the natural IHF sites, *i.e.* greater than 135 kcal/mol. The absence of points in Fig. 5, *lower right quadrant*, contributes significantly to the correlation observed.

The first 20 sequences (labeled A–G) listed in Table 2 includes seven distinct series of related sequences. The sequences within each series are identical with respect to consensus elements but are variable otherwise. Thus, for example, sequences A.1–A.6 contain all consensus elements, including a 6-bp A-tract and ApA steps at both sites of proline intercalation. Sequences designated B–D have each relaxed the requirement for one or more consensus element or have substituted a different base at one of the degenerate positions in the consensus sequence. Series E also has relaxed the requirement to match the consensus WATCAAnnnTTR element, other than the three directly contacted bases. The sequences in series F and G each contain the sequence from E.1 that was substituted for the WATCAAnnnTTR consensus element. This was treated as constituting a modified consensus sequence. In series F, this was held constant, along with all other consensus elements, and the remainder of the sequence varied, just as in series A. In series G, the poly(A)-tract was also allowed to vary, just as in series B.

Each of these series A–G exhibits a distinct correlation between  $\Delta G_{\text{spec}}$  and deformation energy as shown in Fig. 5A. A particularly notable feature is that the slopes of the linear regression lines for the individual series become systematically greater as binding affinity becomes weaker because of the removal of consensus elements. Thus, for example,  $\Delta G_{\text{spec}}$  is



**FIGURE 5. Standard free energy change for specific IHF binding to synthetic DNA sequences ( $\Delta G_{\text{spec}}$ ) plotted as a function of deformation energy.** Values of  $\Delta G_{\text{spec}}$  were obtained from global analysis of gel mobility shift data such as shown in Fig. 4, using the competitive specific and nonspecific finite lattice model as described in the text ("Experimental Procedures"). When no specific binding was observed, sequences were assigned the value,  $\Delta G_{\text{spec}} = -9.1$  kcal/mol, to set an upper limit to the specific binding affinity as described in the text. These points can be identified by the absence of error bars. A, shows binding to selected series of sequences. Symbols are as follows: series A (solid squares), all consensus elements matching  $\lambda H'$  site; series B (open downward triangles) removes A-tract; series C (solid diamonds) removes A-tract and 5' proline intercalation ApA step; series D (solid circles) substitutes A for T in WATCAA element, removes A-tract and 5' proline intercalation ApA step; series E (open circles) retains consensus only for directly contacted bases and proline intercalation ApA step; series F (solid triangles) substitutes specific low deformation energy sequence for variable consensus in series E; series G (open diamonds) removes A-tract from series F. Dashed lines are from linear regression of individual series A–G. Solid lines are from global regression model described in text. Only one of the two points for which no specific binding was observed was included in the analysis of series G data, *i.e.* that with the lower deformation energy. The value plotted represents an upper limit to the specific binding affinity. Consequently, the line drawn connecting only two points in this series represents a lower limit to the slope of the actual trend. B, control sequences in Table 2 (series H). Solid line is from linear regression.

nearly constant for the sequences in series A (Table 3), which contain all of the elements of the reference  $\lambda H'$  site. In contrast,  $\Delta G_{\text{spec}}$  varies by  $\sim 1.4$  kcal/mol as a function of deformation energy in series E and F, in which conservation of the central WATCAAnnnTTR element is relaxed. The slopes of these lines are  $\sim 30$ -fold greater than the series A line (Table 3).

The individual regression lines form a fan-like pattern as if radiating from a single point at a low value of the deformation energy. To assess this possibility, the data analyzed globally were subject to the constraint that the regression lines intersect at a single point. This generated a family of regression lines that are similar to, or indistinguishable from, the individual regression lines, *e.g.* most notably for the limiting two series that have the flattest and steepest trend lines, respectively (Fig. 5A). One

**TABLE 3**  
Global versus local regression analysis for effect of deformation energy on IHF binding

Slopes of regression lines for dependence of IHF binding on deformation energy in individual series of *de novo* DNA sequences listed in Table 2. The global analysis imposes the constraint that all lines intersect at a common point. Values of the slopes along with the 95% confidence intervals were obtained by linear regression of each individual series or by global analysis as described in the text.

Sequence series	Regression slope	
	Local <sup>a</sup>	Global
A	0.0062 ± 0.0052	0.0019 ± 0.0022
B	0.0101	0.0117 ± 0.0021
D	0.0064 ± 0.0066	0.0150 ± 0.0038
E	0.0247	0.0213 ± 0.0054
F	0.0266 ± 0.0127	0.0239 ± 0.0046
G	0.0329	0.0318 ± 0.0074

<sup>a</sup> Absence of confidence limits indicates a slope based on only two points in the series.

way to compare the pairs of lines derived from local and global regression models is by their slopes, because these indicate the correlation between  $\Delta G_{\text{spec}}$  and deformation energy. As shown in Table 3, the slopes are not distinguishable for any pair of lines. The correlation coefficient from the global fit is 0.96, suggesting that most of the variation observed can be accounted by this model. The intersection of the curves occurs at deformation energy of 17 kcal/mol (confidence limits of 0 and 45 kcal/mol), corresponding to a relatively small contribution from DNA flexibility.  $\Delta G_{\text{spec}}$  is  $-12.0 \pm 0.6$  kcal/mol at the point of intersection. At this point, IHF binding becomes insensitive to DNA flexibility suggesting an upper limit to the affinity achievable by IHF of  $\sim 1$  nm under these conditions.

Contrasting results were obtained for two sets of control sequences. The first control set (series H, Table 2) contains 15 synthetic sequences in which both the complement of conserved elements and all nonconserved positions were varied to adjust the deformation energy over the widest possible range. These sequences cover very similar ranges of both deformation energy and  $\Delta G_{\text{spec}}$  as the sequences in Fig. 5A, but the data show only a weak correlation between the two (Fig. 5B). The slope of the linear regression line is similar to what was found for the natural operators in Table 1 and is approximately the average of the slopes for the individual series plotted in Fig. 5A. The correlation coefficient is  $R = 0.26$ , or very similar, as for the natural IHF sites. We conclude that these sequences give a reasonable representation of the characteristics of natural IHF sites but over an expanded range of deformation energy. The contrast between the result obtained from these sequences and the result from the separate series of related sequences is quite distinct.

The second control consists of a set of 14 putative IHF regulatory sites selected from the *E. coli* genome. The purpose of this analysis was to complement the analysis of natural IHF regulatory sites listed in Table 1 but with self-consistent quantitative assays of IHF binding. These are previously uncharacterized sites that were selected based on the proximity to *E. coli* promoters and their location in supercoiling-induced DNA deformation or SIDD loci (33, 34). SIDD loci contain DNA sequences with the greatest propensity to denature or undergo B-Z transitions under conditions of high torsional strain resulting from DNA supercoiling. These sites have been implicated in the mechanism of global transcriptional activation mediated by

IHF (35, 36). The first 10 sequences (named in Table 2 by proximity to a particular *E. coli* gene) were also selected by one or more classifier algorithms trained to recognize aspects of IHF site sequence and structure (full description of selection given in Ref. 24). The last four sequences were scored as less likely to be high affinity IHF sites, based on these same selection criteria. Ten of these 14 sites are found to be high affinity, site-specific sites for IHF; four display only the nonspecific binding mode. These sequences share several characteristics of natural IHF sites, including highly variable complement of consensus elements, a similar narrow range of deformation energies ( $\sim 100$ – $135$  kcal/mol), and a wide range of specific binding affinity. It is not surprising based on these findings that they also show weak correlation between  $\Delta G_{\text{spec}}$  and deformation energy, similar as for other natural IHF regulatory sites and in contrast to the results for sequences in series A–H, above.

## DISCUSSION

DNA-binding proteins recognize specific sequences of base pairs both via direct contacts between amino acid side chains and the purine and pyrimidine bases, and also indirectly by recognizing elements of DNA structure. Moreover, indirect readout mechanisms include not only recognition of preformed features of DNA such as bends, groove topology, local geometry of base pair steps, and geometry of backbone phosphates but also binding-induced distortion of DNA or flexibility. The latter is particularly important to DNA scaffolding proteins such as are required both for chromosome organization and also for regulation of a variety of processes that physically manipulate DNA such as recombination, transposition, replication, and transcription. Because all of these recognition mechanisms rely on intrinsic properties of the particular sequence of bases, it poses a difficult problem to assess quantitatively the relative contributions from each. Because it relies entirely on indirect readout mechanisms, IHF presents an excellent opportunity to assess the relative contributions from preformed structure and from DNA flexibility to indirect readout. Moreover, because IHF functions both as a chromatin organizing protein and also as a specific regulator of genomic processes, this issue is directly relevant to its biological function.

To assess the contribution from DNA flexibility, it is first necessary to evaluate the energy required to deform a particular DNA sequence from its preferred unbound conformation to the conformation in the bound complex. Olson *et al.* (20) first realized that a local sequence-dependent potential could be calculated by analyzing the conformational ensembles of individual base pair steps. Harmonic potentials of mean force were derived from analysis of the structures of protein-DNA complexes along the coordinates of a simplified conformational model that used six parameters to describe the translational and rotational orientation of adjacent base pairs. Sarai and co-workers (37, 38) developed a similar formalism independently and used it to calculate the interaction energies and relative contribution to specificities of direct and indirect readouts in many protein-DNA complexes. More recently, molecular dynamics simulations of all 136 unique tetranucleotides have been used to obtain ensembles of conformations of individual base pair steps (39). Their harmonic potentials of mean force

## Role of Deformation Energy in IHF Binding

yielded similar results as those obtained by Olson *et al.* (20). When we applied this model to assess the change in conformational energy for wrapping of known IHF-binding sites, we found a moderate negative correlation between the free energy change for binding and the deformation energy. This is as expected if favorable protein-DNA contacts are used to drive the (necessarily) unfavorable change in DNA conformation.

Nevertheless, the correlation coefficient suggests that the majority of the variation in binding affinity must be accounted for by another mechanism. The probable reason for this is the strong influence of pre-existing structural characteristics of the conserved IHF recognition elements (17, 40). For example, the consensus hexamer, WATCAA, in the center of the site features a number of unusual base stacking and pairing geometries that facilitate van der Waals packing of protein side chains in the minor groove and optimize H-bonds (17). The narrow minor groove in this region is recognized specifically by a protein clamp that makes salt bridges to phosphates flanking the groove. The A-tract also features a narrow minor groove and a characteristic spine of hydration that provides a regular array of water-mediated H-bond partners. The preference for ApA steps at both sites of proline intercalation might be explained as due to its energetic preference, relative to other base pair steps, to form the particular kinked conformation found in the complex (41). Recent structural and biochemical analysis of a reduced specificity IHF mutant indicates how the unique, sequence-dependent structure of the TTR element contributes to specificity (42). Thus, pre-formed, sequence-dependent structural characteristics constitute a strong component of recognition for each of the conserved elements (17, 18). It is not surprising that variable combinations of these recognition elements, as found in the natural binding site sequences, would tend to obscure a contribution from flexibility.

This led to our strategy of assessing the effect of deformation energy within individual groups of related sequences, where each group contains a fixed configuration of consensus elements. Preliminary analysis of such groups as afforded by natural IHF sites indicates high correlation, suggesting that deformation energy accounts for the majority of the variation. A more systematic analysis of designed sequences confirms this result. These sequences expand the range of deformation energies by about 2-fold as compared with natural IHF sites, approximating the distribution of *E. coli* sequences chosen at random (22). The correlation between binding and deformation energy is also similar as for IHF sites, both for the control sequences plotted in Fig. 5B and for all 49 sequences listed in Table 2 considered as a whole. In contrast, very strong correlation is observed for each of the individual series plotted in Fig. 5A.

The most significant feature apparent in Fig. 5A is the effect when consensus elements are removed and replaced by DNA sequences that expand the range of deformation energy. In every case, *i.e.* removal of the A-tract (compare series B to A and series G to F), removal of a proline intercalation ApA step (compare series D to B), and alteration of the WATCAAnnnTTR consensus sequence (compare series F to A), the result is decreased affinity. This result is consistent with what has been demonstrated previously in a systematic manner for the A-tract

(13). In addition, the slopes of the regression lines are steeper in direct proportion to the effect on affinity. Thus, substitution for consensus sequences both lowers affinity and increases sensitivity to DNA flexibility.

Three of the seven series considered contain only two members. Because these lines have no degrees of freedom, their slopes are of course quite sensitive to experimental uncertainty or secondary effects of sequence of either of the two experimental points. However, two of these three cases conform well to the overall pattern. The one case that does not (series C) is the only one to replace two consensus elements together.

Caution should also be exercised when interpreting the absolute values of the slopes. The magnitude of the calculated deformation energies is large compared with experimental values of the free energy to bend DNA, *e.g.* as determined from circular permutation assays. In applying these potentials to analyze contributions to binding affinity in a number of systems, it has been found necessary to fit scaling factors to bring the calculated values in range of experimentally observed values (*cf.* Morozov *et al.*, (43)). Although the scale of the independent variable may not accurately reflect the translation of conformational energy into binding energy, the relative slopes are significant. These vary by over 30-fold in a systematic manner. These data indicate that deformation can contribute up to 1.4 kcal/mol, or an order of magnitude in affinity, over the observable range of deformation energy and suggest a much larger potential contribution, were it not obscured in our assays because of competition from the nonspecific IHF binding mode.

The fan-shaped family of regression lines that these individual sequence series define suggests a point of convergence at low deformation energy. This is confirmed by the results of a global regression analysis that was conducted to define this point. The analysis yields lines that are essentially indistinguishable from the separate regression lines in half of the cases, and within the confidence limits for the rest. The two significant observations are as follows: first, the point of convergence is at a finite, albeit low, value of deformation energy; and second, the slope for series A, sequences that include perfect matches to all consensus elements, is essentially zero ( $0.0012 \pm 0.0030$ ). The first observation defines a maximum binding energy and perhaps a common penalty paid by all sequences to conform to the structure of the bound complex. The second suggests that IHF is insensitive to the flexibility of sequences outside of the known consensus elements.

Thus, although sequences outside the consensus elements contribute to nonspecific interactions such as the salt bridges and charge neutralization thought to dominate the thermodynamics of IHF binding (26), interactions that confer specificity are restricted to just the consensus elements. Although the confidence limits on the value of the deformation energy at the convergence point are broad, the value is clearly low. This points to two conclusions. First, Saecker and Record (44) have described a pattern of surface salt bridges and cationic side chains that provide a scaffold to dictate the topology of DNA wrapping. Evidently, these interactions are enormously efficient at compensating for the intrinsic energy necessary to bend DNA. Second, it confirms the dominant influence of pre-existing structural characteristics of the conserved IHF recognition

elements as opposed to flexibility to absorb binding-dependent deformations.

What then accounts for the effect of deformation energy when consensus elements are removed? This must reflect the deformation energy of the sequences replacing those elements. Thus, although specificity appears largely because of recognition of pre-existing conformational features, flexibility is able to compensate for the lack of these features to a remarkable extent. Although this analysis does not identify any consensus element as mandatory to high affinity and specificity, it does indicate that all do contribute. It should also be noted that complete compensation for loss of preformed structural elements is achievable only at a theoretical limit of nearly zero deformation energy. Even the lowest deformation energy that is actually attainable does not completely compensate for loss of even one of the consensus elements. As a result, when all consensus elements are absent, even the lowest deformation energy attainable fails to yield a high affinity IHF site. In this regard, affinity and specificity are very much whole system properties rather than a sum of individual contributions.

This ability of DNA flexibility to compensate for specific recognition sequences constitutes an intrinsic property of indirect readout. On this basis, indirect readout mechanisms can provide for substantial degeneracy of the DNA sequence with only minimal decrease in affinity. On the other hand, direct readout relies on critical interactions to yield specific recognition, but at the cost of prohibiting significant degeneracy in binding site recognition. In the case of IHF binding, these properties of indirect readout support a continuous range of affinities that vary over nearly 4 orders of magnitude.

This broad range of specificities is appropriate to the multiple biological roles of this highly abundant, DNA architectural/global regulatory protein. Because IHF is structurally related to the bacterial histone-like HU proteins (45), its primary biological role is to bind to many sequence-independent DNA sites to condense the bacterial nucleoid and to constrain DNA supercoils (2, 35). However, unlike the HU proteins, IHF has evolved a sequence-dependent, high affinity, DNA-binding motif. The total intracellular concentration of IHF is maintained at a level of greater than 1  $\mu\text{M}$ , but its free concentration is estimated to be between 0.7 and 5 nM (14, 46). Therefore, because the  $K_d$  value for IHF binding to its highest affinity, sequence-dependent DNA-binding sites is about  $1 \times 10^{-9}\text{M}$ , these high affinity sites are occupied most of the time and function as constitutive architectural nucleoprotein structures on the bacterial chromosome. Many of these structures have been shown to influence important biological functions such as transcription, recombination, and DNA replication (35). Thus, IHF serves two roles in the cell. It serves a global DNA architectural role by binding in a sequence-independent manner to many DNA sites over a broad range of affinities, and it serves local, biologically important, roles at many highly degenerate but sequence-specific high affinity DNA-binding sites. The remarkable thing about IHF is that it employs an indirect readout mechanism to bind and wrap DNA at both of these nonspecific and high affinity DNA sites. In this study, we have shown that IHF recognizes pre-formed conformational characteristics of the consensus DNA sequence. At all other sites, relative affinity

is defined by the conformational freedom of individual base pair steps to adopt the DNA structure of the bound IHF-DNA complex (21, 22).

*Acknowledgment*—We thank Phoebe Rice for permission to use the coordinate file for this IHF- $\lambda$  H' DNA structure, in which the crystallographic nick in the DNA was repaired.

## REFERENCES

- Bewley, C. A., Gronenborn, A. M., and Clore, G. M. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27**, 105–131
- Friedman, D. I. (1988) *Cell* **55**, 545–554
- Ali, B. M., Amit, R., Braslavsky, I., Oppenheim, A. B., Gileadi, O., and Stavans, J. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10658–10663
- Gardner, J. F., and Nash, H. A. (1986) *J. Mol. Biol.* **191**, 181–189
- Sewitz, S., Crellin, P., and Chalmers, R. (2003) *Nucleic Acids Res.* **31**, 5868–5876
- Freundlich, M., Ramani, N., Mathew, E., Sirko, A., and Tsui, P. (1992) *Mol. Microbiol.* **6**, 2557–2563
- Hwang, D. S., and Kornberg, A. (1992) *J. Biol. Chem.* **267**, 23083–23086
- Toller, L. (2002) *An Interdisciplinary Approach Employing Computational, Biochemical, and Genomic Methods to Examine the Effects of Chromosome Organization on the Regulation of Gene Expression*. Ph.D. dissertation, Università degli Studi di Pavia e Frenze, Florence, Italy
- Goodrich, J. A., Schwartz, M. L., and McClure, W. R. (1990) *Nucleic Acids Res.* **18**, 4993–5000
- Engelhorn, M., Boccard, F., Murtin, C., Prentki, P., and Geiselmann, J. (1995) *Nucleic Acids Res.* **23**, 2959–2965
- Ussery, D., Larsen, T. S., Wilkes, K. T., Friis, C., Worning, P., Krogh, A., and Brunak, S. (2001) *Biochimie (Paris)* **83**, 201–212
- Hales, L. M., Gumpert, R. I., and Gardner, J. F. (1994) *J. Bacteriol.* **176**, 2999–3006
- Hales, L. M., Gumpert, R. I., and Gardner, J. F. (1996) *Nucleic Acids Res.* **24**, 1780–1786
- Yang, S. W., and Nash, H. A. (1995) *EMBO J.* **14**, 6292–6300
- Wang, S., Cosstick, R., Gardner, J. F., and Gumpert, R. I. (1995) *Biochemistry* **34**, 13082–13090
- Mengeritsky, G., Goldenberg, D., Mendelson, I., Giladi, H., and Oppenheim, A. B. (1993) *J. Mol. Biol.* **231**, 646–657
- Rice, P. A., Yang, S., Mizuuchi, K., and Nash, H. A. (1996) *Cell* **87**, 1295–1306
- Travers, A. (1997) *Curr. Biol.* **7**, R252–R254
- Goodman, S. D., Velten, N. J., Gao, Q., Robinson, S., and Segall, A. M. (1999) *J. Bacteriol.* **181**, 3246–3255
- Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11163–11168
- Steffen, N. R., Murphy, S. D., Toller, L., Hatfield, G. W., and Lathrop, R. H. (2002) *Bioinformatics (Oxf.)* **18**, Suppl. 1, 22–30
- Steffen, N. R., Murphy, S. D., Lathrop, R. H., Opel, M. L., Toller, L., and Hatfield, G. W. (2002) *Genome Inform.* **13**, 153–162
- Aeling, K. A., Steffen, N. R., Johnson, M., Hatfield, G. W., Lathrop, R. H., and Senear, D. F. (2006) *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, in press
- Steffen, N. R. (2003) *Computational Methods Using Structure and Sequence for Identifying Protein-binding Sites in DNA*, Ph.D. dissertation, Department of Computer Science, University of California, Irvine
- Nash, H. A., Robertson, C. A., Flamm, E., Weisberg, R. A., and Miller, H. I. (1987) *J. Bacteriol.* **169**, 4124–4127
- Holbrook, J. A., Tsodikov, O. V., Saecker, R. M., and Record, M. T., Jr. (2001) *J. Mol. Biol.* **310**, 379–401
- Tsodikov, O. V., Holbrook, J. A., Shkel, I. A., and Record, M. T., Jr. (2001) *Biophys. J.* **81**, 1960–1969
- McGhee, J. D., and Von Hippel, P. H. (1974) *J. Mol. Biol.* **86**, 469–489
- Epstein, I. R. (1978) *Biophys. Chem.* **8**, 327–339
- Box, G. D. P. (1960) *Ann. N. Y. Acad. Sci.* **86**, 792–816
- Spira, B., and Yagil, E. (1999) *Curr. Microbiol.* **38**, 80–85

## Role of Deformation Energy in IHF Binding

32. Lorenz, M., Hillisch, A., Goodman, S. D., and Diekmann, S. (1999) *Nucleic Acids Res.* **27**, 4619–4625
33. Sheridan, S. D., Benham, C. J., and Hatfield, G. W. (1999) *J. Biol. Chem.* **274**, 8169–8174
34. Sheridan, S. D., Benham, C. J., and Hatfield, G. W. (1998) *J. Biol. Chem.* **273**, 21298–21308
35. Hatfield, G. W., and Benham, C. J. (2002) *Annu. Rev. Genet.* **36**, 175–203
36. Opel, M. L., Aeling, K. A., Holmes, W. M., Johnson, R. C., Benham, C. J., and Hatfield, G. W. (2004) *Mol. Microbiol.* **53**, 665–674
37. Gromiha, M. M., Siebers, J. G., Selvaraj, S., Kono, H., and Sarai, A. (2004) *J. Mol. Biol.* **337**, 285–294
38. Gromiha, M. M., Siebers, J. G., Selvaraj, S., Kono, H., and Sarai, A. (2005) *Gene (Amst.)* **364**, 108–113
39. Arauzo-Bravo, M. J., Fujii, S., Kono, H., Ahmad, S., and Sarai, A. (2005) *J. Am. Chem. Soc.* **127**, 16074–16089
40. Rice, P. A. (1997) *Curr. Opin. Struct. Biol.* **7**, 86–93
41. Bosch, D., Campillo, M., and Pardo, L. (2003) *J. Comput. Chem.* **24**, 682–691
42. Lynch, T. W., Read, E. K., Mattis, A. N., Gardner, J. F., and Rice, P. A. (2003) *J. Mol. Biol.* **330**, 493–502
43. Morozov, A. V., Havranek, J. J., Baker, D., and Siggia, E. D. (2005) *Nucleic Acids Res.* **33**, 5781–5798
44. Saecker, R. M., and Record, M. T., Jr. (2002) *Curr. Opin. Struct. Biol.* **12**, 311–319
45. Drlica, K., and Rouviere-Yaniv, J. (1987) *Microbiol. Rev.* **51**, 301–319
46. Murtin, C., Engelhorn, M., Geiselmann, J., and Boccard, F. (1998) *J. Mol. Biol.* **284**, 949–961
47. Pratt, T. S., Steiner, T., Feldman, L. S., Walker, K. A., and Osuna, R. (1997) *J. Bacteriol.* **179**, 6367–6377
48. Van Der Ploeg, J. R., Iwanicka-Nowicka, R., Bykowski, T., Hryniewicz, M. M., and Leisinger, T. (1999) *J. Biol. Chem.* **274**, 29358–29365
49. Claverie-Martin, F., and Magasanik, B. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1631–1635
50. Tsui, P., and Freundlich, M. (1988) *J. Mol. Biol.* **203**, 817–820
51. Dworkin, J., Ninfa, A. J., and Model, P. (1998) *Genes Dev.* **12**, 894–900
52. Jovanovic, G., and Model, P. (1997) *Mol. Microbiol.* **25**, 473–481
53. Weiner, L., Brissette, J. L., Ramani, N., and Model, P. (1995) *Nucleic Acids Res.* **23**, 2030–2036
54. Browning, D. F., Beatty, C. M., Wolfe, A. J., Cole, J. A., and Busby, S. J. (2002) *Mol. Microbiol.* **43**, 687–701
55. Wu, Y., Patil, R. V., and Datta, P. (1992) *J. Bacteriol.* **174**, 6918–6927
56. Blomfield, I. C., Kulasekara, D. H., and Eisenstein, B. I. (1997) *Mol. Microbiol.* **23**, 705–717
57. Kulasekara, H. D., and Blomfield, I. C. (1999) *Mol. Microbiol.* **31**, 1171–1181
58. Sirko, A., Weglenska, A., and Hulanicka, D. (1998) *Mol. Gen. Genet.* **258**, 174–177
59. Paul, L., Blumenthal, R. M., and Matthews, R. G. (2001) *J. Bacteriol.* **183**, 3910–3918
60. Green, J., Anjum, M. F., and Guest, J. R. (1997) *Microbiology* **143**, 2865–2875
61. Rabin, R. S., Collins, L. A., and Stewart, V. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8701–8705
62. Zhang, X., and Demoss, J. A. (1996) *J. Bacteriol.* **178**, 3971–3973
63. Sirko, A., Zehelein, E., Freundlich, M., and Sawers, G. (1993) *J. Bacteriol.* **175**, 5769–5777
64. Ramani, N., Huang, L., and Freundlich, M. (1992) *Mol. Gen. Genet.* **231**, 248–255
65. Parekh, B. S., and Hatfield, G. W. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1173–1177
66. Hopper, S., Babst, M., Schlensog, V., Fischer, H. M., Hennecke, H., and Bock, A. (1994) *J. Biol. Chem.* **269**, 19597–19604
67. Tsui, P., Huang, L., and Freundlich, M. (1991) *J. Bacteriol.* **173**, 5800–5807
68. Huang, L., Tsui, P., and Freundlich, M. (1990) *J. Bacteriol.* **172**, 5293–5298
69. Bykowski, T., and Sirko, A. (1998) *Biochimie (Paris)* **80**, 987–1001
70. Altuvia, S., Almiron, M., Huisman, G., Kolter, R., and Storz, G. (1994) *Mol. Microbiol.* **13**, 265–272
71. Bongaerts, J., Zoske, S., Weidner, U., and Udden, G. (1995) *Mol. Microbiol.* **16**, 521–534
72. Wackwitz, B., Bongaerts, J., Goodman, S. D., and Udden, G. (1999) *Mol. Gen. Genet.* **262**, 876–883
73. McNicholas, P. M., Chiang, R. C., and Gunsalus, R. P. (1998) *Mol. Microbiol.* **27**, 197–208