

Symmetry and asymmetry in the function of *Escherichia coli* integration host factor: implications for target identification by DNA-binding proteins

Milton H. Werner, G. Marius Clore, Angela M. Gronenborn
and Howard A. Nash*

Laboratory of Chemical Physics, National Institute of Diabetes, Digestive and Kidney Diseases and *Laboratory of Molecular Biology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, USA.

Background: *Escherichia coli* integration host factor (IHF) is a DNA-binding protein that participates in a wide variety of biochemical functions. In many of its activities, IHF appears to act as an architectural element, dramatically distorting the conformation of bound DNA. IHF is a dimer of non-identical subunits, each about 90 amino acids long. One dimer interacts specifically with a 30 base pair (bp) target, but well-conserved sequences are found in only half of this binding site. Thus, the IHF-DNA system has long been viewed as a paradigm of asymmetry in a protein-DNA interaction.

Results: We have isolated the subunits of IHF and show that either subunit is capable of specifically recognizing natural IHF-binding sites and supporting λ site-specific recombination *in vitro*. Mobility shift and footprinting data indicate that the isolated subunits interact with DNA as homodimers. We also describe the design of symmetric duplexes to which heterodimeric and homodimeric IHFs can

bind by recognizing specific sequences.

Conclusions: Our *in vitro* manipulation of the IHF system demonstrates that binding and bending of target DNA can be accomplished symmetrically. The prevalence of asymmetry found for this system in nature suggests that additional selective forces may operate. We suggest that these follow from the disparity between the size of the DNA that IHF protects (30 bp) and the length of DNA that the protein can initially contact (10 bp). This disparity implies that an IHF target is recognized in stages and may dispose the part of the protein-DNA system used for initial recognition to evolve distinctly from the remainder of the interaction surface. We suggest that a limitation in the length of DNA that can be initially contacted is a general property of DNA-binding proteins. In that case, many proteins can be expected to identify complex targets by step-wise, rather than simultaneous, contact between sequence elements and DNA-binding domains.

Current Biology 1994, 4:477-487

Background

Escherichia coli integration host factor (IHF) participates in a variety of processes, ranging from phage λ site-specific recombination [1] to transcriptional regulation of *E. coli* genes [2]. In many of its activities, IHF functions as part of a multi-protein-nucleic acid complex, within which it apparently acts as an architectural element, bending the bound DNA [3,4]. IHF causes DNA to bend at specific sites, recognition of which is thought to occur via the minor groove [5].

IHF has long been considered a paradigm of asymmetry in a DNA-binding protein. It is a heterodimeric protein with only moderate amino-acid sequence identity (~25%) between its α and β subunits (the *himA* and *hip* gene products, respectively). Moreover, typical IHF-binding sites lack an obvious dyad symmetry and have their most conserved elements clustered in one half of the target sequence [5,6]. We have investigated the role of asymmetry in IHF function by isolation of the IHF subunits and reassembly of the subunits into homodimers. We have also designed symmetric DNA sites

and tested them for sequence-specific binding by homodimeric and heterodimeric IHFs.

Results

Purification and physical properties of IHF subunits

Solutions of denatured IHF can be separated by high performance liquid chromatography (HPLC) into two components (see Fig. 1a). Integration of the absorbance at 214nm from three elution profiles conducted at an analytical scale (data not shown) enables one to calculate that the ratio of these components in the chromatogram is 0.9 ± 0.05 . By pooling selected fractions from the column, we prepared stocks of the two peaks that were more than 98% pure (see Materials and methods). Denaturing SDS-Tricine polyacrylamide gel electrophoresis (PAGE) (Fig. 1b) and mass spectrometry (data not shown) indicate that these isolated component solutions each contain a single peptide corresponding to the α and β subunits of native IHF. Although mass spectrometry has an expected error of ± 5 mass units, the measured masses — equivalent to molecular weights of

Correspondence to: Milton H. Werner.

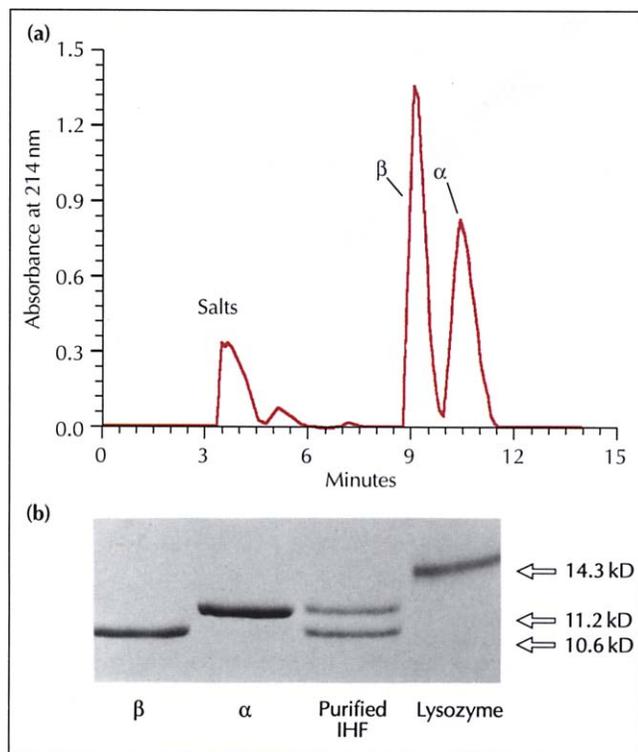


Fig. 1. Purification of IHF subunits. **(a)** Elution profile of IHF upon C18 reversed-phase HPLC and **(b)** denaturing SDS-Tricine gel of purified subunits.

11223 Daltons for α and 10650 Daltons for β — agree remarkably well with those expected from the corrected amino-acid sequence of IHF [7]. The isolated α subunit was soluble in water and buffer solutions, but the isolated β subunit precipitated in buffer solutions containing more than 3 mM added salt. Isolated subunit solutions were therefore stored in deionized water at 4°C.

DNA-binding properties of the isolated subunits

Mobility-shift experiments were carried out to assess the ability of individual IHF subunits to recognize and bind to a typical IHF-binding site, the H' site of bacteriophage

λ [5]. Previous work (S-W. Yang, personal communication) has shown that a 31 base pair (bp) oligonucleotide duplex containing this site binds IHF almost as tightly as does a much longer restriction fragment. We have found that each subunit alone can retard such a duplex (Fig. 2). The mobility of a complex between a protein and a short piece of DNA should be dominated by the mass of the complex and not by its shape. As the mobility of the complexes made by the individual subunits and by IHF are identical, we conclude that, just like IHF [5], the individual subunits bind as dimers to the H' oligonucleotide. Because we are unable to identify a predominant stable homodimer in solution (data not shown), we presume such dimers are assembled on the DNA templates. Comparison of the amount of shifted DNA for a given amount of added protein indicates, however, that the individual subunits bind less tightly to this site than does IHF (Table 1).

We performed two kinds of experiments to assess whether this binding reflects the intrinsic activity of the isolated subunits or is merely the consequence of a trace amount of heterodimer in our preparations. First, we replaced the purified subunits with varying amounts of IHF heterodimer, up to ten-times larger than the maximum amount of heterodimer that could have contaminated our preparations (see Materials and methods). None of these concentrations of IHF produced a detectable bandshift with the H' site (data not shown). To test whether small amounts of IHF are more effective in the presence of an excess of one subunit, we prepared mixtures of 5% (or 10%) of one purified subunit with 95% (or 90%) of the other purified subunit. These mixtures were then used in bandshift experiments with the H' site, and we could linearly extrapolate from the data to the amount of contamination needed to account for the extent of bandshift seen with the isolated subunits alone (data not shown). Even in the worst case, that of the β subunit, we estimate that the degree of contamination would have to be 10% to produce the observed effects. This is at least ten-times higher than the maximum impurity that could have gone undetected.

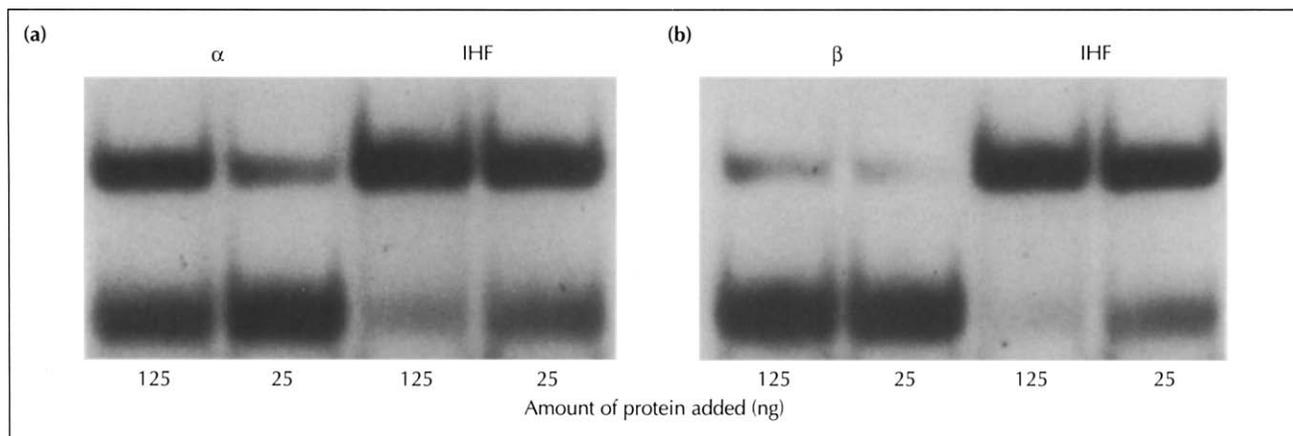


Fig. 2. Binding of IHF subunits to the H' site oligonucleotide. Mobility-shift experiments with 30 bp H' oligonucleotide in the presence of 38 ng salmon sperm DNA. **(a)** Comparison of IHF- α with native IHF. **(b)** Comparison of IHF- β with native IHF.

Table 1. Relative activity of IHF subunits.				
	Binding ^{††}		Recombination ^{††}	
	H'	H1	Integration	Excision
IHF- α	$\geq 50\%$	30–50%	10–30%	10–30%
IHF- β	10–30%	2–10%	$\leq 2\%$	2–10%
IHF- α + IHF- β [§]	100%	100%	100%	100%
HU	0 [¶]	NT	0	$\leq 2\%$

^{*} Values are calculated as percentage of native IHF activity for each experiment; 0 = not detectable, NT = not tested.

[†] Measured as the mobility shift of a 187 bp *attL* fragment (H') or a 450 bp *BstI* fragment (H1) [8] in the presence of 38 ng salmon sperm (ssp) DNA at 125 ng subunit added, relative to the shift in an identical experiment with native IHF.

^{††} Estimated from the amount of subunit that gives maximal recombination relative to the amount of recombinant produced with 12.5 ng of IHF (the empirical maximum with IHF).

[§] A simple mixture of IHF subunits in water without significant incubation was done before dilution in reaction buffer.

[¶] A faint smear is seen with 125 ng HU protein in the presence of 38 ng sspDNA. No shift is seen with 380 ng sspDNA added, in stark contrast to the significant shift seen with IHF- α and IHF- β under the same conditions.

We have also used the H1 binding site of *attP* [5,8] to compare the binding activity of the subunits (Table 1). Whereas IHF- α binds to H1 only slightly more weakly than it binds to H', the affinity of IHF- β for H1 appears to be more than ten-fold weaker than for H'. Accordingly, it is difficult to rule out a role for contaminating heterodimers.

Although individual subunits bind to H' more weakly than IHF, their binding still appears to be specific. Under the conditions of Figure 2, neither of the separated subunits shifted a 187 bp *attL* fragment containing a mutated H' site, the sequence of which differed from that of H' at four critical positions (*attL*-QH')

[9]; data not shown). Furthermore, under these conditions the HU protein failed to shift the H' duplex (data not shown). In order to assess the relative specificity of IHF subunits for H' more quantitatively, we have titrated protein–DNA complexes with sonicated salmon sperm DNA (sspDNA), a non-specific competitor, against protein–DNA complexes formed with a 187 bp *attL* fragment containing a wildtype H' [9]. The individual subunits can bind this DNA despite a significant challenge from sspDNA, albeit less well than the native protein (Fig. 3). The degree of mobility shift for the 187 bp fragment is the same for the subunit complexes and for IHF. Although we believe the mobility shift for the 30 bp oligonucleotide complexes primarily reflects the protein:DNA stoichiometry, the magnitude of the shift for the larger fragment should be more sensitive to the overall shape of the DNA. Thus, we conclude that the degree of deformation of DNA must be very similar for homodimeric and heterodimeric complexes.

We considered the possibility that higher amounts of the individual subunits were needed to bind the H' target DNA because subunit peptides had been modified during preparation. However, quantitative recovery of strong IHF-binding activity was achieved by mixing the separated IHF subunits stoichiometrically in unbuffered water before dilution of the mixture into a DNA binding assay (Fig. 4). Neither lengthy incubation periods nor other special conditions were required to observe this effect.

In vitro recombination experiments

Individual subunits of IHF can support integrative and excisive recombination *in vitro* (Fig. 5). For IHF- α , approximately ten-times as much protein is required to achieve the same quantity of recombinant product as would be produced with IHF. IHF- β is significantly less active, requiring at least fifty-times more protein to produce the same quantity of recombinant as IHF

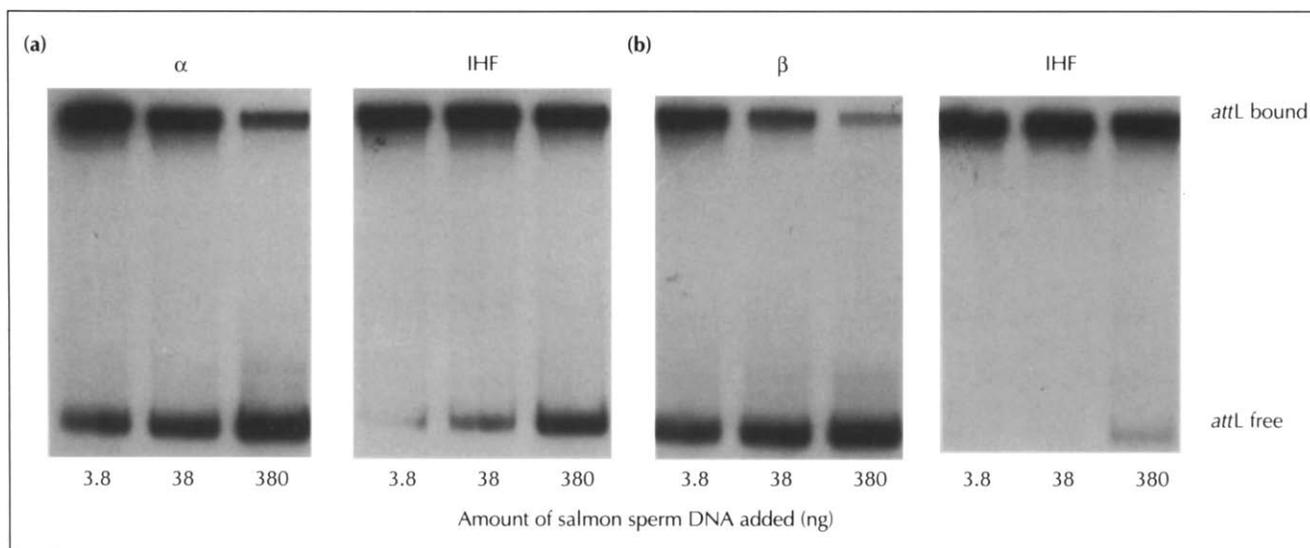


Fig. 3. Relative specificity of IHF subunits binding to the H' site of λ *attP*. (a) 25 ng of α subunit or IHF were added per 20 μ l reaction; (b) 125 ng of β subunit or IHF added per 20 μ l reaction mixture.

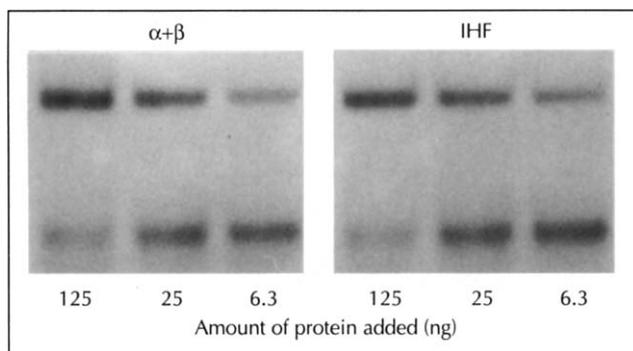


Fig. 4. Mobility shift of reconstituted IHF. Solutions of α and β were mixed stoichiometrically before adding them to DNA-binding mixtures containing the 30 bp H' duplex. The lanes show the amount of subunit or IHF added in the presence of 38 ng salmon sperm DNA.

(Table 1). Stoichiometric mixing of IHF subunits (as in the DNA binding studies) suffices to recover nearly the full recombination activity of native IHF (data not shown). This confirms earlier observations that *in vitro* recombination activity can be recovered by mixing crude extracts derived from *himA*⁻ and *bip*⁻ strains [1,10].

Despite the significantly lower recombination efficiency of the isolated subunits, they are clearly more

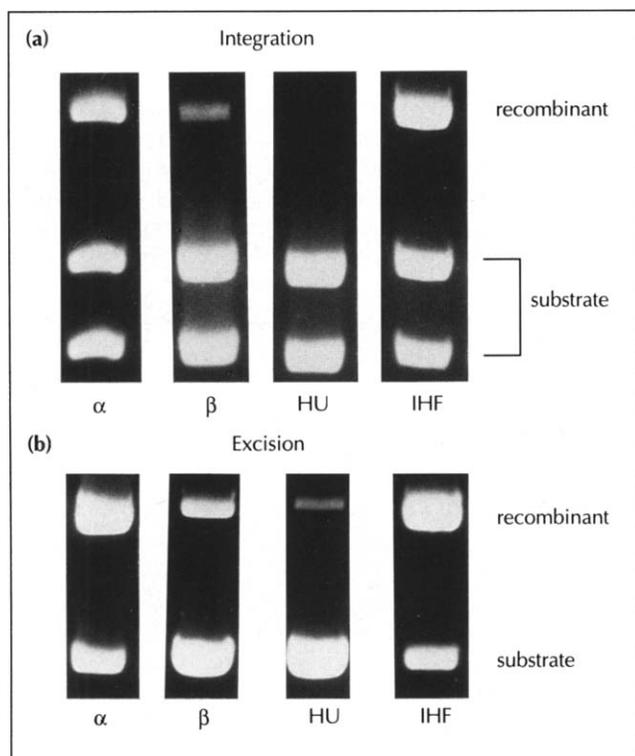


Fig. 5. Comparison of the recombination activity of IHF subunits, HU and IHF. For integration, 250 ng of protein was added when IHF subunits or HU was used. For excision, 125 ng of IHF subunits or HU was added. In both cases, the amount of recombinant is compared to that produced with 12.5 ng native IHF.

active than the homologous, but non-specific, binding protein, HU. In contrast to the readily detectable levels of recombinant product seen when intermolecular integration is carried out with homodimeric complexes of α or β , no recombinant product forms with HU (Fig. 5a). Although HU protein can support excisive recombination *in vitro* [4], the isolated subunits are, again, dramatically more efficient (Fig. 5b; Table 1).

Hydroxyl radical footprinting

We have used hydroxyl radical footprinting to compare the region of DNA contacted by IHF [5] and its subunits (Fig. 6). We find that the α subunit protects the same 30bp segment of a >200bp DNA fragment as does IHF. The footprint of the β subunit could only be observed at very high concentrations of protein, reflecting its very weak interaction with the H' site. Nonetheless, under these conditions, the β subunit did protect approximately the same region of the DNA as did the α subunit and IHF (data not shown). These observations also suggest that the subunits bind as homodimers and contact the DNA in the same way as does IHF.

Imino proton NMR spectra

We assessed the similarities between heterodimeric and homodimeric complexes in more detail by examining ¹H-NMR spectra of the base-paired imino hydrogens of the DNA used to form the complexes. In general, one can observe a single imino-proton signal for nearly every base pair in duplex DNA. For the H' site in the absence of protein, 24 such signals were observed (Fig. 7, bottom); the spectral analysis is complicated by the fact that many A-T base pairs occur in the sequence, resulting in significant overlap in the imino proton NMR spectrum. Nonetheless, the sensitivity of the chemical shift to the local chemical environment permits a qualitative comparison of the DNA conformation in the presence of different protein molecules; differences or similarities in the pattern of imino-proton signals should therefore reflect differences or similarities in the gross DNA conformation. We used this approach to compare the similarity in the DNA-protein complexes of IHF and its isolated subunits.

The DNA-protein complexes formed by IHF and IHF- α give very similar patterns of peaks, with no doubling or extensive line broadening (Fig. 7); these data suggest that the heterodimer and the IHF- α homodimer each form primarily a single species in solution. In contrast, the spectrum of the IHF- β homodimeric complex appears to have features of both the free and bound DNA, but the observed peaks are broad and peaks intermediate in chemical shift between free and bound DNA are not resolved. This suggests that the IHF- β homodimer may undergo intermediate to slow exchange, relative to the NMR timescale, with the DNA. This is consistent with the relatively weak apparent affinity of the IHF- β homodimer for the H' oligonucleotide, as discerned from mobility shift and footprinting experiments.

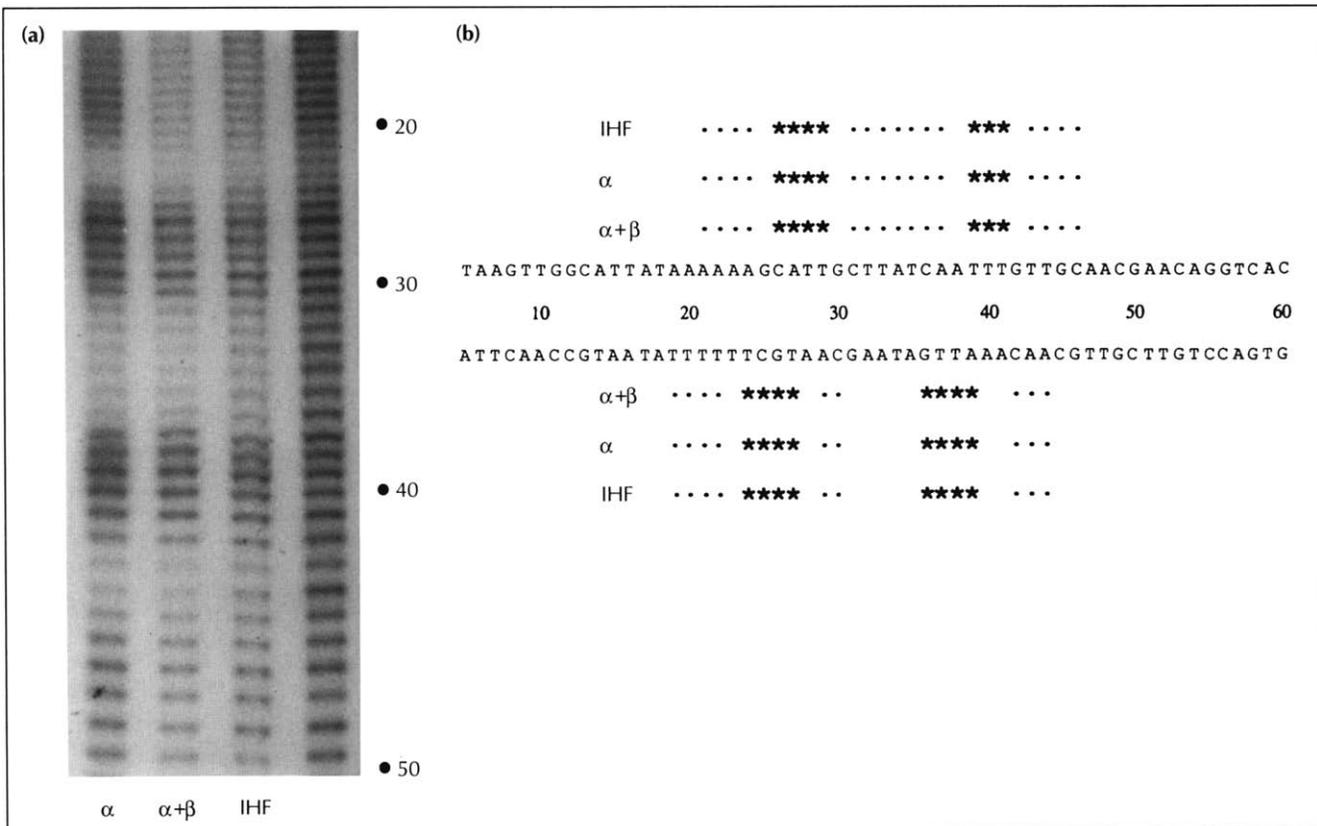


Fig. 6. Hydroxyl radical footprinting of the *attP* H' Site by IHF and its subunits. (a) Representative footprints (performed as described in Materials and methods) for the top strand of the H' site. The lane on the far right contains unprotected DNA. (b) Summary of observed footprints on both strands. Asterisks indicate enhanced cleavage by hydroxyl radical, periods indicate protection from the same. The coordinates refer to the center of the λ attachment site [5].

Homodimeric and heterodimeric complexes with self-complementary oligonucleotides

Despite the large footprint of the homodimeric and heterodimeric complexes, only a few residues are highly conserved between typical IHF-binding sites (the conserved sequence is WATCAAnnnnTTR, where W is A or T, n any base and R a purine). There is also significant information content in regions of IHF-binding sites outside this highly conserved region [6], but no obvious symmetry is apparent between the consensus and non-consensus halves of an IHF site. This raises the possibility that the asymmetry in IHF-binding site sequences is integral to IHF recognition. We have addressed this issue by construction of self-complementary oligonucleotides in which parts of an IHF site are symmetrized.

We have designed two different classes of self-complementary DNA molecules (Table 2). In the first class (H'S1, H'S6, H'S2), we duplicated either the left or right half of the H' binding site. In the second class (H'H7, H'H8), we used the model of Yang and Nash [5], and results from recent mutational studies [10,11], to estimate which region(s) of the H' site should be in contact with either the IHF- α or IHF- β subunit. On the basis of these predictions, we reassembled non-contiguous regions of H' to be adjacent to one another (see Table 2). We found that symmetric sites of both classes

can interact substantially with heterodimeric IHF and with the isolated IHF- α subunit. Not all of our constructs were successful in binding IHF or its subunits, however, showing that symmetry does not override the requirement of this system for a specific site.

Despite the success of several symmetric sites in acting as targets, we have yet to design a site that binds tightly to the isolated IHF- β subunit, and our initial attempts to design a subunit-specific symmetrical site have failed. For example, H'H7 brought together a portion of the consensus sequence (ATCAA, red circle in Table 2) with the A-tract (open square), then duplicated this construction such that a palindrome was formed. These two regions were predicted to contact only the IHF- α subunit [5,10]; H'H7 was therefore expected to bind IHF- α more strongly than IHF- β . H'H8 was designed in a similar fashion from elements thought to be in contact with IHF- β . Neither of these constructs acted in agreement with our predictions. It remains to be seen whether modest refinements in our design will provide any significant improvement.

Discussion

We have tested the ability of symmetric and asymmetric variants of IHF to form specific complexes with both

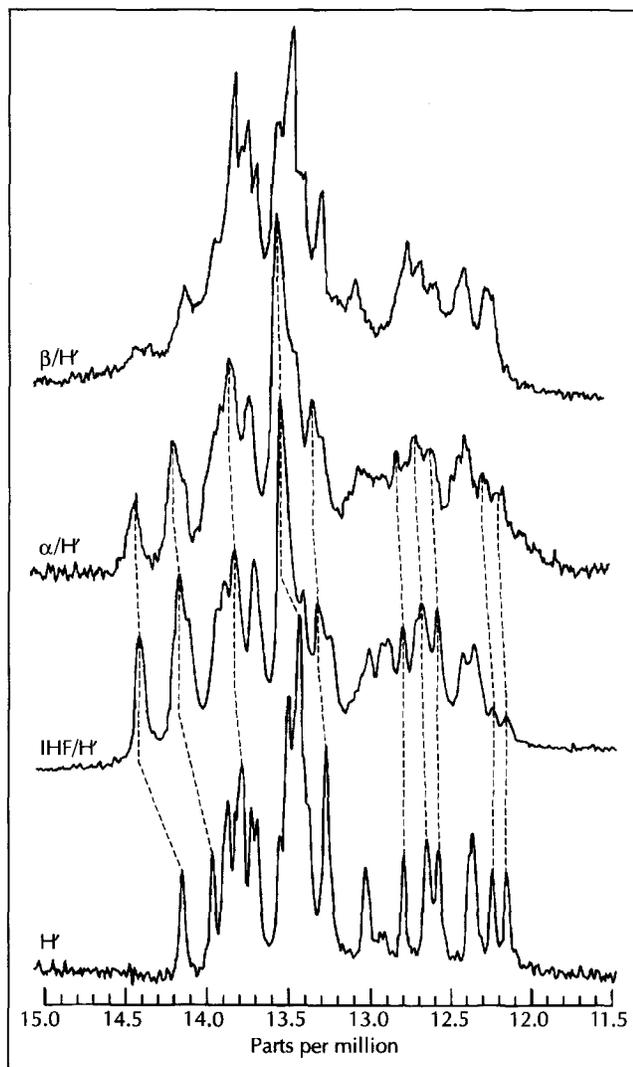


Fig. 7. NMR spectra of heterodimeric and homodimeric IHF complexes: imino proton NMR spectra of IHF and subunit complexes at 30°C. Vertical lines have been drawn to illustrate the correspondence between the free DNA peaks (bottom) and those from the complexes. The difference in signal-to-noise is a reflection of the differences in concentration for each of these samples.

symmetric and asymmetric binding sites. Although a wide variety of previous experiments indicated that the naturally occurring protein is asymmetric in composition and that natural IHF sites are not palindromic, our results demonstrate that asymmetry is not required for specific binding or for the recombination function of IHF. We shall reconsider the possible biological and structural reasons for the observed asymmetry in the IHF system in the light of this insight, and then, using this insight, propose a mechanism for recognition of DNA targets that may be widely applicable.

Asymmetry of *E. coli* IHF protein and its binding sites

Many naturally occurring binding sites for *E. coli* IHF have been characterized biochemically. In all cases, the binding site for IHF appears to be ~30bp in length, with a partially conserved region of sequence in one half of the site. Closer examination of IHF-binding sites

indicates that there is a non-random distribution of nucleotides throughout the 30bp region protected by IHF [6], and that sequence alterations even at poorly conserved positions can have significant effects on IHF binding ([12]; S-W. Yang, H.A. Nash, unpublished data). There is little or no similarity, however, between sequences found in the well-conserved and poorly conserved halves of the 30bp region; thus, natural IHF sites appear to be functionally asymmetric.

Naturally occurring IHF protein is also functionally asymmetric in that the protein seems to require different contributions from two related, but distinct, subunits. Disruption of the genes encoding either IHF- α or IHF- β produces similar phenotypes, and no novel phenotypes have been reported in strains constructed to be simultaneously mutant for both IHF subunits. Bear *et al.* [13] have argued that the IHF requirement for plating of the λ cos154 variant can partially be fulfilled by IHF- α , but not by IHF- β . The efficacy of IHF- α in this assay is small relative to complete IHF, however, and attempts to improve it by overproduction of the IHF- α subunit have not succeeded [14]. Isolated subunits may be subject to intracellular degradation; immunoblotting of cell lysates from *bip* or *bimA* deletion mutants indicates that the amount of the residual gene product is substantially reduced [10,14].

These observations probably reflect inefficient formation of homodimers of either subunit. Although this feature alone could account for the importance of heterodimers, there are also significant differences between the putative DNA-binding surfaces of the two subunits of IHF ([5,10]; S-W. Yang, H.A. Nash, unpublished data). When a mutation is made affecting this surface in one subunit, the resulting phenotype is often different from that induced by the analogous mutation in the other subunit [10,11]. As the DNA-binding surfaces do not impinge on the putative dimer interface of IHF [5,10,11], the functional asymmetry of IHF protein may involve both protein-protein and protein-DNA contacts.

Functional asymmetry in the protein-DNA system

We have argued that the *E. coli* IHF protein and its binding sites are functionally asymmetric *in vivo*, but our results show that neither need be asymmetric for the formation of a stable complex *in vitro*. Binding sites that are artificially constructed to be perfect palindromes bind IHF almost as well as a natural site. Moreover, each isolated subunit of IHF can specifically recognize and bind as a homodimer to an IHF site, a phenomenon that has been independently discovered by Goosen and collaborators [15]. Homodimers of IHF subunits can even specifically bind to palindromic IHF sites. Despite the dispensability of asymmetry in our purified system, IHF-binding sites in every prokaryote that has been tested resemble those of *E. coli* in their lack of symmetry, and proteins analogous to IHF appear also to be dimers of non-identical subunits (for examples, see [7,16-19]). Although our work raises the possibility that a prokaryote will be discovered in which a homodimeric IHF recognizes a symmetric sequence,

Table 2. Binding of heterodimeric and homodimeric proteins to self-complementary oligonucleotides.

	Duplex DNA sequence*	Construction [†]	Strength of binding [‡]		
			α	β	IHF
H'	ATAAAAAAGCATTGC■TTATCAATTTGTTGC		>30%	<10%	100%
H'H7:	AGAAAAATGCTTGATA●TATCAAGCATTTTTC		>30%	<10%	>30%
H'S1:	AGCAACAAATTGATAA●TTATCAATTTGTTGC		10–30%	<10%	>30%
H'S6:	AGCAACAAATTGATAAG●CTTATCAATTTGTTGC		<10%	0	>30%
H'S2:	AGTAAAAAGCATTGC●GCAATGCTTTTTTAC		0	0	0
H'H8:	AACGTTCAAAAACGAA●TTGCTTTTTGAACGT		0	0	0

*For H', only the top strand is shown. The physical center of the 30 bp H' site is indicated by a black square. The center of symmetry in each symmetric sequence is indicated by a black dot.
[†]Circles and squares indicate regions of the H' site that have been shown to be in contact with the DNA based on footprinting data of Yang and Nash [5]. Red circles refer to the ATCAA portion of the consensus, red squares refer to the TTG portion of the consensus. The open square and circle refer to the corresponding regions of sequence in the left half of the H' site; these are not well conserved between IHF sites and are not obviously related by symmetry to the right half of the H' site.
[‡]Strength of binding in mobility-shift assays relative to IHF binding to the H' site. 0 = not detectable.

the prevalence of asymmetry in the known systems suggests that this feature has been selected.

One possibility for such selection is that IHF might need to orient its faces in a defined way. For example, if IHF formed contacts with another protein, such contacts could be restricted to a subset of targets by suitable orientation of an asymmetric binding site. Given a heterodimeric protein, such asymmetry would direct the contact surface of IHF either towards or away from its putative partner. Although this idea is attractive, we can find no circumstance where this possibility appears to be exploited in nature. That is, in cases in which IHF is presumably used to accomplish the same task in different organizational contexts, there appears to be no correlation between the IHF-site orientation and either the phasing or spacing of this site and the other components of the system [2,16,17,20]. Although it remains possible that another use of IHF will be discovered that exploits the capacity to select a specific orientation, the available evidence provides no encouragement for this possibility.

A more attractive possibility for the widespread asymmetry in the IHF system is that this feature reflects the mechanism by which the protein recognizes its DNA target. Suppose that the recognition of a 30 bp IHF site occurs in stages, with one particular segment being the focus of the initial search. The evolutionary constraints on this part of the protein–DNA system might then be very different from those on the remainder of the system. For example, by analogy with HU protein, each subunit of IHF contributes an 'arm', a two-stranded β -ribbon [5], that is thought to be important in binding to DNA. If only one of these arms were used during initial recognition of the site, one would expect that it and its cognate target would co-evolve to maximize the speed and fidelity of the initial interaction. In contrast, the remaining arm would merely be involved in stabilizing the initial complex and would be expected to

co-evolve with its target region according to different principles. Even though the two arms may both be wrapped around DNA similarly in the final complex, their differing roles in the formation of the complex could result in selection for different sequences in their DNA-binding surface and in the DNA targets that they contact.

Target identification by DNA-binding proteins

Our observations on asymmetry in the IHF system have prompted us to consider the possibility that IHF sites are recognized in a step-wise fashion. We now speculate that this notion may have widespread application. Recognition sequences of DNA-binding proteins typically comprise one or more short clusters of information. In many cases, the entire recognition sequence consists of either a single asymmetric segment, or a bipartite and symmetrical segment in which two half-sites are apposed. In these cases, the target comprises sequence elements that are clustered within ten or fewer base pairs, and most or all of these elements are probably recognized simultaneously. We suggest that a protein would have difficulty in initially recognizing a DNA-binding site much larger than this.

Thus, for example, if elements for the initial recognition of a target were to span twenty or thirty base pairs, a DNA-binding protein would either have to establish a very large contact surface or, at the least, would have to locate two or more smaller contact surfaces that were perhaps distributed on one face of the target DNA. The former would require the formation of a complementary surface that follows the contours of several turns of helix. The latter would require a simultaneous fit of several contact surfaces in order to form an overall complementary protein–DNA interface. In either case, it is hard to imagine that such molecular choreography could occur on a timescale that is consistent with the observation that many protein–DNA interactions occur at a rate that approaches the diffusion limit [21,22]. It is

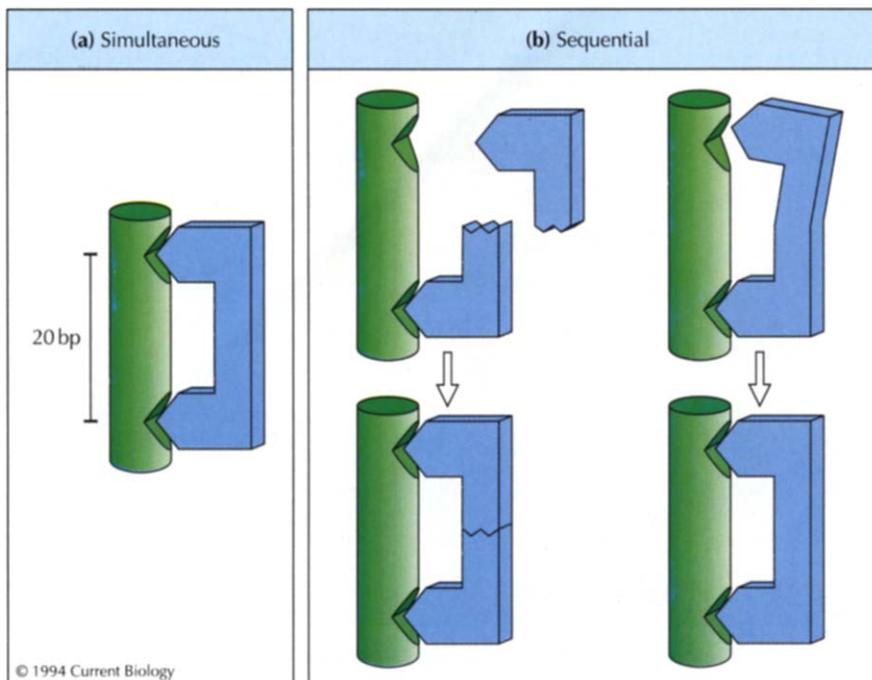


Fig. 8. Alternative pathways for initial recognition of complex targets by DNA-binding proteins. A typical target, composed of two half-sites separated by more than 10bp, is shown as a cylinder with two notches. A protein that binds to this target is shown as a bipartite structure with two surfaces that are complementary to these notches. The figure presents different proposals for the way that the protein could make initial contact with this target. **(a)** DNA sequence elements that are important for initial recognition of the target are distributed over both half-sites. In order to find this target, the protein must contact all of these elements simultaneously. **(b)** The elements important for initial recognition are clustered in one half-site of the target. This DNA segment is contacted by one part of the DNA-binding protein; this protein can either be an isolated subunit (upper left) or one domain of a stable assembly (upper right). Subsequent steps (lower left and right) are needed to convert this initial recognition complex into a final, more stable complex.

much easier to imagine that a few critically important contacts are made in one clustered segment first, thereby localizing the protein search for its target to a small region of sequence.

According to this view, many protein–DNA systems may have evolved to perform the initial recognition over only part of the overall site (Fig. 8). Consider, for example, systems in which oligomeric proteins sense two recognition elements arranged as widely spaced inverted repeats. In one class of these systems, epitomized by LexA protein, recognition of each half-site is clearly sequential. In this system, kinetic and equilibrium studies have demonstrated that the pathway for binding involves recognition of a single half-site by one monomer, followed by the cooperative binding of a second monomer to the remaining half-site [23]. A similar result has been observed with Arc repressor [24], except that in this case the protomer that recognizes the half-site is itself a dimer.

For targets of this type, one might expect that one of the two half-sites is consistently singled out for initial recognition and is thus subject to a different selection pressure from the other half-site. Indeed, an analysis of 19 LexA sites shows that the two half-sites have significantly different matches to the consensus sequence [25]. Cooperative binding to half-sites has been observed in several other cases, for example the DNA-binding domain of the glucocorticoid receptor [26]. In this case, structural studies provide a rationale for sequential binding of protomers [27]. To our knowledge, aside from these three examples (LexA, Arc and the glucocorticoid receptor), it is not known if the cooperation reflects enhanced binding of a second monomer to a filled half-site, rather than enhanced binding of a preformed oligomer.

In our view, proteins like LexA and Arc represent one logical solution to the problem of recognizing large sites: the stable form of the protein (or protein domain) only spans a half-site. However, there are many examples of proteins that recognize widely spaced bipartite sites with a single stable dimer. In at least one case, that of *E. coli* CAP protein [28], a detailed examination strongly suggests that, even here, half-site recognition is operative. No known natural CAP site contains two perfect TGTGA consensus sequences [28] and, in most cases, one half-site matches the consensus much more closely than the other [29]. Furthermore, although CAP is a homodimer under physiological conditions, it is likely that only one subunit contains bound cAMP [30]. CAP may therefore function as an asymmetric dimer that recognizes an asymmetric target. The evidence for the recognition mechanism that we propose is less clear in many other systems (λ repressor, for example [31]). Perhaps a competing selective force(s) that has yet to be discovered has prevented the evolution of asymmetry in such systems. Alternatively, the demands of deforming DNA may have led IHF and CAP to fine-tune the initial steps of binding more stringently than is generally required. A time-resolved method for examining the development of a stable protein–DNA interaction is necessary for further analysis. We predict that, for IHF and many other systems that bind to long stretches of DNA, specific parts of the target will be consistently protected before the remainder.

Conclusions

Our data show that not only can native heterodimeric IHF bind to symmetric and asymmetric IHF-binding sites, but homodimers assembled from either separated

subunit can also bind to these sites. Moreover, homodimeric and heterodimeric IHFs can support *in vitro* site-specific recombination. In nature, however, the IHF system is asymmetric. This dilemma has led us to speculate about the way proteins recognize DNA targets that span several turns of the double helix. Our suggestion that such recognition is step-wise provides a potential explanation for the evolution of asymmetry in the IHF system. Moreover, although IHF binds to DNA as a pre-assembled dimer, our suggestion also places in context the growing number of cases in which the stable DNA-binding species of a protein only spans a half-site.

We emphasize that at equilibrium, a stable protein–DNA complex is expected to use all the information content of a full site. Thus, both halves of a binding site will contribute to the overall binding affinity, including sequences that lie outside the consensus. However, our analysis of IHF leads us to propose that the primary determinant in specific recognition is restricted to approximately 10bp or one helical turn. Rather than reflecting an incidental detail of the protein–DNA interaction, we believe this insight reveals a fundamental limitation in the way proteins find their targets. Although this insight may perhaps seem self-evident, we are unaware of its articulation in previous theoretical or experimental analyses concerning protein–DNA interactions. The speed and accuracy with which the search for a target is accomplished is likely to be vital in certain biological settings, for example the recovery of a repressed system from induction or the changes in gene expression that occur as new gene regulators are made during embryogenesis. The limitation in the mechanism of recognition that we have inferred from our study of IHF is likely to be a critical factor in these processes.

Materials and methods

Purification of IHF and isolation of subunits

IHF was purified by a modification of the method of Nash *et al.* [32], with the phosphocellulose column followed by chromatography on Mono-S (Pharmacia/LKB) with a salt gradient of 0–0.5M NaCl in the presence of 20mM sodium phosphate, pH6.8. The Mono-S fractions containing purified IHF were subsequently dialyzed against 20mM sodium phosphate, 200mM NaCl and lyophilized. 1–20mg of purified, lyophilized IHF was then dissolved in 10ml of 8M guanidine hydrochloride (GuCl) in the presence of 0.1M glacial acetic acid (GAA) or 50mM Tris buffer with 1mM Na₂EDTA. The solution was allowed to stand for 5–24h at room temperature and filtered over 0.2µm Millex-GV filters (Milipore, Inc.).

A Dynamax-300A C18 column (2.5cm × 25cm, 12µm, Rainin Instruments) was washed with 100ml of 0.1% trifluoroacetic acid (TFA, v/v) and equilibrated with a mixture containing 65% Buffer A (0.1% TFA) and 35% Buffer B (90% acetonitrile (v/v), 10% of 0.1% TFA) at 10mlmin⁻¹. 0.1–2ml of the filtered, denatured protein solution was loaded onto the column and eluted at 10mlmin⁻¹ with a gradient of 35% to 52% Buffer B over 40min. The elution was monitored at

214nm and 274nm using a Pharmacia/LKB 2141 variable wavelength flow detector.

Preparative quantities of individual subunits were obtained from fractions of the column that showed little or no cross-contamination with each other. The isolated subunits were lyophilized and redissolved in water then stored either at 4°C or –20°C, without cryoprotectants or buffering salts. Purity and homogeneity were determined by SDS–Tricine polyacrylamide gel electrophoresis (PAGE) [33] (12% (w/v), 0.75mm), electrospray mass spectrometry on a Jeol JMS-SX102 and analytical high performance liquid chromatography (HPLC) on a Dynamax-300A column (0.5cm × 25cm) using the protocol described above, but at a flow rate of 1mlmin⁻¹. The best quantitative estimate of the degree of purity came from experiments in which small amounts of one purified subunit were mixed with a large excess of the other and the resulting mixture resolved on acetic acid–Triton–urea PAGE [34]; as adapted by L. Huang, R. McMacken, personal communication). In this way, we determined that our stocks of IHF-α and IHF-β were at least 98% and 99% homogeneous, respectively. Protein concentrations were determined using the method of Bradford [35], and by spectrophotometry of subunit solutions in the presence of 8M GuCl using a calculated extinction coefficient at 274nm of 1200M⁻¹cm⁻¹ for IHF-α and 3600M⁻¹cm⁻¹ for IHF-β.

Mobility shift assays

Stock solutions of IHF or its subunits were serially diluted in reaction buffer (36mM Tris, 14% glycerol (v/v), 7.5mM Na₂EDTA, 140µgml⁻¹ BSA (bovine serum albumin), 43mM KCl) from 2–800-fold. A 30bp synthetic duplex comprising the λ *attP* H' binding site for IHF (λ coordinates 27747 to 27777), a 187bp fragment of λ *attL* containing a single H' site [9], and a 450bp *BstI* fragment containing the λ H1 site [8] were used to assess the DNA-binding properties of IHF and its subunits. A typical reaction consisted of 1–1.5µl of diluted protein, 1µl of 5'-radiolabelled DNA and reaction buffer to a final volume of 20µl. DNA concentrations varied between 0.25 and 0.75nM in a given experiment. 38ng of sonicated salmon sperm DNA (sspDNA) were also added to the reaction mixture as indicated. The reactions were allowed to incubate for 5–15min at room temperature, then 2µl of 1% (w/v) xylene cyanol dye in reaction buffer was added and the reaction mixture loaded onto an 8% (w/v), 1× TBE polyacrylamide gel run at constant 300V or 150V for one or two hours, respectively. The gel was dried and autoradiographed for 12–24h and/or counted on a Betagen Betascope 603 Blot Analyzer.

In vitro recombination

Intermolecular recombination experiments were performed *in vitro* using 200ng each of two supercoiled, CsCl purified plasmid DNAs. For integration, the plasmids contained either an *attP* site (pHN894, [36]) or an *attB* site (pBB105, [37]); for excision, the plasmids contained either an *attL* site (pHN872, [36]) or an *attR* site (pHN868, [36]). Reactions were set up to contain 50mM Tris, 1mM Na₂EDTA, 0.5mgml⁻¹ BSA, 5mM spermidine and 6–625ng of IHF or its subunits in a final volume of 20µl. Integration reactions were initiated by the addition of 30ng of purified λ Integrase (Int); excision reactions were initiated by adding 30ng of Int and 20ng of purified λ Xis protein premixed with one another just before the experiment. After 60min, recombination reactions were supplemented with MgCl₂ (to 10mM) and restriction enzymes (*Hind*III and *Mlu*I for integration; *Xba*I and *Mlu*I for excision).

Hydroxyl radical footprinting

The *EagI*-*NcoI* *attL* fragment (202 bp, -69 to +136 *attL* coordinates) was used for bottom-strand footprinting, while the *EagI*-*KpnI* *attL* fragment (276 bp, -167 to +110 *attL* coordinates) was used for top-strand footprinting. The fragments (derived from plasmid pHN1681 [9]) were labelled by overnight digestion of plasmid with *EagI* followed by 3' end-labelling using the enzyme Sequenase 2.0 (United States Biochemical). Labelled plasmid DNA was subsequently digested with *NcoI* and *KpnI*.

For footprinting, complexes were formed by the addition of 12.5–125 ng of IHF or IHF- α (1.25–2.5 μ g for IHF- β), 38 ng sspDNA and 3'-[³²P]-labelled DNA in 10 mM Tris/1 mM Na₂EDTA buffer in a final volume of 100 μ l. Typical DNA concentrations were 0.5–2 nM. The footprinting reactions were performed as described [38] for 2 min (1 min for the IHF- β complex).

Duplex oligonucleotides

For non-self-complementary DNAs, the two gel-purified strands were mixed stoichiometrically, heated to 90 °C and slowly cooled to room temperature over several hours. Duplex DNA was separated from single-strand DNA using a Mono-Q HR 5/5 column (Pharmacia, Inc.) with a linear gradient 0.25–1 M NaCl in 20 mM sodium phosphate buffer containing 1 mM Na₂EDTA, pH 5.5, flowing at 1 ml min⁻¹. For self-complementary oligonucleotides, the hairpin form predominated over duplex at concentrations below 20 nM. Thus, 0.2 picomoles of 5'-[³²P]-labelled single strands were mixed with 200–400 nM unlabelled DNA and then annealed in the conditions described above. This procedure was successful in producing mostly duplex DNA for use in mobility-shift assays, but necessitated the use of 20 nM DNA per mobility-shift reaction.

NMR spectroscopy

NMR spectra were collected and analyzed at 30 °C on a Bruker AM-500 spectrometer. For the complex with IHF, 10 mg of purified IHF were dissolved in 25 ml of 20 mM sodium phosphate, 35 mM NaCl, 5 mM NaN₃, pH 6.8. An equimolar quantity of a 30 bp oligonucleotide comprising the H' site was titrated into this solution with gentle mixing. The volume was subsequently reduced to 1–2 ml using a collodion membrane apparatus (Schleicher and Schuell) with a 10000 molecular weight cut-off. For the homodimeric complexes, 5 mg of purified subunit was dissolved in 12–15 ml of the above buffer and a half equimolar quantity of oligonucleotide was titrated into the solution with gentle mixing and subsequently concentrated using a collodion membrane. Despite the poor solubility of free IHF- β in salt solutions, the formed complex was reasonably soluble, with only moderate precipitation occurring during the concentration process using the methodology described above. To remove unbound protein or DNA, the samples were gel-filtered using a 2.6 \times 100 cm column of Sephacryl S-100 (Pharmacia) equilibrated with the above buffer and flowing at 2 ml min⁻¹. The eluted samples were concentrated in a collodion membrane to 0.5 ml and the NMR spectrum collected.

Acknowledgements: The authors wish to acknowledge Will McClure, Robert Weisberg, Alex Burgin and Anca Segall for critical reading of the manuscript, Alex Burgin for oligonucleotide synthesis, Lewis Pannell for collection and analysis of mass spectrometry data, Anca Segall for preparation of plasmid pHN1681 and other DNA fragments of *attL*, Carol Robertson for purified Int and Xis proteins and Andrew Granston and Steven Goodman for technical advice and instruction during the course of

this work. This work was supported in part by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health (to G.M.C. and A.M.G.)

References

- MILLER HI, KIKUCHI A, NASH HA, WEISBERG R A, FRIEDMAN DI: Site-specific recombination of bacteriophage λ : the role of host gene products. *Cold Spring Harbor Symp Quant Biol* 1979, 48:1121–1126.
- FREUNDLICH M, RAMANI N, MATHEW E, SIRKO A, TSUI, P: The role of integration host factor in gene expression in *E. coli*. *Mol Microbiol* 1992, 6:2557–2563.
- NASH HA: Bending and supercoiling of DNA at the attachment site of bacteriophage λ . *Trends Biochem Sci* 1990, 15:222–227.
- GOODMAN SD, NICHOLSON SC, NASH HA: Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. *Proc Natl Acad Sci U S A* 1992, 89:11910–11914.
- YANG C-C, NASH HA: The interaction of *E. coli* IHF protein with its specific binding sites. *Cell* 1989, 57:869–880.
- GOODRICH JA, SCHWARTZ ML, MCCLURE WR: Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *E. coli* integration host factor. *Nucleic Acids Res* 1990, 18:4993–5000.
- HALUZI H, GOITEIN D, KOPY S, MENDELSON I, TEFF D, MENERITSKY G, GILADI H, OPPENHEIM AB: Genes coding for integration host factor are conserved in Gram-negative bacteria. *J Bacteriol* 1991, 173:6297–6299.
- ROBERTSON CA, NASH HA: Bending of the bacteriophage λ attachment site by *E. coli* integration host factor. *J Biol Chem* 1989, 263:3554–3557.
- SEGALL A, NASH HA: Synaptic intermediate in bacteriophage λ site-specific recombination: integrase can align pairs of attachment sites. *EMBO J* 1993, 12:4567–4576.
- GRANSTON AE, NASH HA: Characterization of a set of integration host factor mutants deficient for DNA binding. *J Mol Biol* 1993, 234:45–59.
- LEE EC, HALES LM, GUMPORT RI, GARDNER JF: The isolation and characterization of mutants of the integration host factor (IHF) of *E. coli* with altered, expanded DNA-binding specificities. *EMBO J* 1992, 11:305–311.
- LEE EC, MACWILLIAMS MP, GUMPORT RI, GARDNER JF: Genetic analysis of *E. coli* integration host factor interactions with its bacteriophage λ H' recognition site. *J Bacteriol* 1991, 173:609–617.
- BEAR SE, COURT DL, FRIEDMAN DI: An accessory role for *E. coli* integration host factor: characterization of a lambda mutant dependent upon integration host factor for DNA packaging. *J Virol* 1984, 52:966–972.
- DITTO MD, ROBERTS D, WEISBERG RA: Growth phase variation of integration host factor level in *E. coli*. *J Bacteriol* 1994, in press.
- ZULIANELLO L, DE LA GORGUE DE ROSNY E, VAN ULSEN P, VAN DE PUTTE P, GOOSEN N: The *hlmA* and *hlmD* subunits of integration host factor can specifically bind to DNA as homodimers. *EMBO J* 1994, 13:1534–1540.
- HOOPER TR, SANTERO E, PORTER S, KUSTU S: The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. *Cell* 1990, 63:11–22.
- BOCCARD F, PRENTKI P: Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units. *EMBO J* 1993, 12:5019–5027.
- GOBER JW, SHAPIRO L: A developmentally regulated *Caulobacter* flagellar promoter is activated by 3' enhancer and IHF binding elements. *Mol Biol Cell* 1992, 3:913–926.
- TOUSSAINT B, DELICATTREE I, D'ASPRESMONT R DE S, DAVID L, VINÇON M, VIGNAIS PM: Purification of the integration host factor homolog of *Rhodobacter capsulatus*: cloning and sequencing of the *hlp* gene, which encodes the β subunit. *J Bacteriol* 1993, 175:6499–6504.

20. OPPENHEIM AB, RUDD KE, MENDELSON I, TEFF D: Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *E. coli*. *Mol Microbiol* 1993, 10:113-122.
21. WINTER RB, BERG OG, VON HIPPEL PH: Diffusion-driven mechanisms of protein translocation on nucleic acids III. The *E. coli lac* repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry* 1981, 20:6961-6977.
22. MAZUR SJ, RECORD MT JR: Association kinetics of site-specific protein-DNA interactions: roles of nonspecific DNA sites and of the molecular location of the specific site. *Biopolymers* 1989, 28:929-953.
23. KIM B, LITTLE JW: Dimerization of a specific DNA-binding protein on the DNA. *Science* 1992, 255:203-206.
24. BROWN BM, SAUER RT: Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers. *Biochemistry* 1993, 32:1354-1363.
25. BERG O: Selection of DNA binding sites by regulatory proteins: the LexA protein and the arginine repressor use different strategies for functional specificity. *Nucleic Acids Res* 1988, 16:5089-5105.
26. DAHLMAN-WRIGHT K, SILTALA-ROOS H, CARLSTEDT-DUKE J, GUSTAFSSON J-A: Protein-protein interactions facilitate DNA binding by the glucocorticoid receptor DNA-binding domain. *J Biol Chem* 1990, 265:14030-14035.
27. LUISI BF, XU WX, OTWINOWSKI Z, FREEDMAN LP, YAMAMOTO KR, SIGLER PB: Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 1991, 352:497-505.
28. CROTHERS DM, STEITZ TA: Transcriptional Activation by *E. coli* CAP Protein. In *Transcriptional Regulation*. Edited by McKnight SL, Yamamoto K. Plainview, NY: Cold Spring Harbor Laboratory Press; 1992:501-534.
29. BERG O, VON HIPPEL P: Selection of DNA binding sites by regulatory proteins II: The binding specificity of cyclic AMP receptor protein to recognition sites. *J Mol Biol* 1988, 200:709-723.
30. HEYDUK T, LEE JC: *E. coli* cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. *Biochemistry* 1989, 28:6914-6924.
31. HARRISON SC, AGGARWAL AK: DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem* 1990, 59:933-969.
32. NASH HA, ROBERTSON CA, FLAMM E, WEISBERG RA, MILLER HI: Overproduction of *E. coli* integration host factor, a protein with non-identical subunits. *J Bacteriol* 1987, 169:4124-4127.
33. SCHÄGGER H, VON JAGOW G: Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 1987, 166:368-379.
34. LENNOX RW, COHEN LH: Analysis of histone subtypes and their modified forms by PAGE. *Methods Enzymol* 1989 170:532-549.
35. BRADFORD MA: Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal Biochem* 1976, 72:248-254.
36. KITTS PA, NASH HA: Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J Mol Biol* 1988, 204:95-107.
37. RICHT E, ABCARIAN P, NASH HA: The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* 46:1011-1021.
38. DIXON WJ, HAYES JJ, LEVIN JR, WEIDNER MF, DOMBROSKI BA, TULLIUS TD: Hydroxyl radical footprinting. *Methods Enzymol* 1991, 208:380-413.

Received: 10 February 1994; revised 5 April 1994
Accepted: 14 April 1994.