PROTEINS: Structure, Function, and Genetics 5:281–288 (1989)

# A Protein Structural Motif That Bends DNA

**Stephen W. White,[1] Krzysztof Appelt[2], Keith S. Wilson,[3] and Isao Tanaka[4]**

*[1]Center for Structural Biology, Biology Department, Brookhaven National Laboratory, Upton, New York 11973, [2]Agouron Pharmaceuticals Inc., La Jolla, California 92037, [3]European Molecular Biology Laboratory, Deutsches Elektronen Synchrotron, D-2000 Hamburg 52, Federal Republic of Germany, and [4]Department of Polymer Science, Faculty of Science, Hokkaido University, Sapporo, Japan*

***ABSTRACT*** **The prokaryotic protein HU, integration host factor (IHF) from *Escherichia coli*, and transcription factor 1 (TF1) from bacteriophage SPO1 are closely related molecules. Biochemical results suggest that the role of these proteins is to bind and bend DNA. From the high-resolution structure of HU, we propose a model for this interaction with DNA. Crucial amino acid differences between the proteins can be rationalized in terms of their different specific functions.**

**Key words: Hu protein, integration host factor, transcription factor 1, DNA bending protein, protein structure**

## INTRODUCTION

It is becoming increasingly apparent that bending or kinking is a biologically important deformation of the DNA helix. Although it can be induced by certain sequences such as the adenine-rich tracts in kinetoplast DNA,[1] it more usually occurs in response to protein binding. Complexes such as nucleosomes,[2] and gyrasomes[3] are good examples. The possible structures of such DNA deformations have been discussed,[4,5] and there now exists an elegant method for their detection based on gel electrophoresis.[6]

Several groups have recently demonstrated that the protein integration host factor (IHF) can bend DNA.[7–9] IHF was first identified as a factor necessary for the stable integration of the bacteriophage lambda genome into *Escherichia coli*.[10] It is now clear that it is a host-encoded protein, binds specifically to DNA, and affects gene expression and DNA replication (for a review see 11). It is a heterodimer of α- and β-subunits[12] and the genes for both (*himA* and *hip*, respectively) have been cloned and sequenced.[13,14] The molecular weight of each subunit is close to 11,000. Chemical protection experiments (ref. 15, and H.A. Nash and C.C. Yang, personal communication) strongly suggest that it interacts with the minor groove.

Amino acid sequence data reveal that IHF is a member of a family of three small, basic DNA binding proteins[11]; the other two are the ubiquitous prokaryotic protein HU and the SPO1-encoded transcription factor 1 (TF1).

Protein HU binds nonspecifically to DNA and wraps it into nucleoprotein structures similar to the eukaryotic nucleosomes.[11] It has a molecular weight of 9500, and occurs in most organisms as a dimer with identical subunits. In *E. coli*, it somewhat anomalously occurs as a heterotypic dimer of α- and β-subunits,[16,17] and the genes for each (*hupA* and *hupB*) have been cloned.[11,18] The sequences of several HU proteins from different organisms have now been determined and show high homology.[11] Its DNA wrapping and condensing properties suggest that it may function as a prokaryotic histone protein.[19] However, it has been estimated[11] that there are a sufficient number of molecules to cover only one-sixth of the *E. coli* genome (some 20,000 dimers), and cytological studies have shown it to be concentrated on the periphery of the nucleoid.[20] These data, together with demonstrations that HU can affect transcription rates, site-specific recombination, and the initiation of DNA replication,[11] suggest a more subtle function for the protein.

TF1 is encoded by bacteriophage SPO1 and preferentially binds SPO1 DNA, which contains 5′-hydroxymethyluracil in place of thymine.[21] As yet, no clear physiological role for TF1 has been demonstrated, but it is known to selectively inhibit the in vitro transcription of SPO1 DNA by bacterial RNA polymerases,[22] and is essential for viral multiplication.[23] An SPO1 infected cell contains some $5 \times 10^4$ molecules of TF1, but none is present in the mature virion particle.[24] The molecule is a homotypic dimer and the gene has been cloned and sequenced.[25]

We have reported the crystallization[26] and structure determination[27,28] of HU from *Bacillus stearothermophilus* at 3 Å resolution, and suggested in a general way how it might interact with double-stranded DNA. We have now refined the structure to a resolution of 2.1 Å and are in a position to formulate a more detailed model for its interaction with

Received February 16, 1989; accepted March 24, 1989.

Address reprint requests to Stephen W. White, Center for Structural Biology, Biology Department, Brookhaven National Laboratory, Upton, NY 11973.

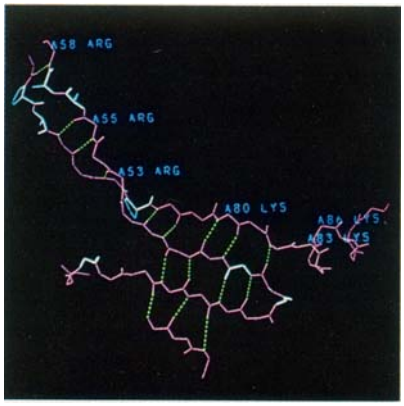


Fig. 1. The peptide backbone of the β-sheet region from one-half of the HU molecule. All of this region belongs to one monomer apart from the short outer strand which is the N-terminus of the other monomer. The positions of the proposed DNA binding, basic residues are labeled. The glycines at the turns and the distortion in the β-ribbon structure of the arm are shown as white/light blue. Hydrogen bonds are green.

DNA. The high sequence homology between HU, IHF, and TF1 gives us confidence that the important features of the model are applicable to all three proteins. These features are consistent with the notion that the principal function of this family of proteins is to bend DNA.

## THE STRUCTURE OF THE COMMON FOLD

### General Description

Based on the refined structure at 2.1 Å resolution, a detailed analysis has been carried out of the HU protein from *B. stearothermophilus*. Details of the refinement and the analysis will be reported elsewhere. The structurally important amino acid residues, such as those that create the hydrophobic core and the several glycines that are crucially placed at tight turns in the structure, are highly conserved in all HU proteins, IHF and TF1.[11] It is clear that the three types of proteins share the same basic structure.

The general features of the structure have been reported previously[28] and will only briefly be described here. Starting from the N-terminus, each monomer consists of two helices arranged in a V-shaped motif, a three-stranded antiparallel β-pleated sheet, and a short C-terminal helix. In the dimer, the four helices form a wedge-shaped base and the top is covered by a "tent" created by the two sheets which remain separate. The second and third strands of the sheet are connected by a long arm of some 25 amino acids. The sheet region from one half of the dimer is shown in Figure 1. The distal part of the arm is flexible and, although the refinement has enabled the visualization of several more residues, 8 remain invisible (60–67). The structure of the proximal half is a simple β-ribbon extension of the sheet. The hydrogen-bonding pattern and strand twist are maintained, and the $\phi/\varphi$ angles map at the correct region of the Ramachandran plot.[29] The hydrogen-bonding pattern stops at residues 56,57 and 72,73; the possible significance of this is discussed later. The monomers are wrapped around each other, and the N-terminus of one is hydrogen-bonded to the outer strand of the sheet of the other (Fig. 1). It is clear that the monomer cannot fold into a stable structure, and any functional reference to it should be avoided.

### The Structure of the DNA Binding Domain

The most likely DNA-binding region of HU has been determined from the 3 Å structure and consists of the two arms and the intervening strands of the sheet. These secondary structural elements describe a right-handed helix that exactly complements that of double-stranded DNA (Fig. 3). It also contains several highly conserved lysines and arginines, and such residues have been shown to be involved in the interaction with DNA.[30] The amino acids that we consider to be important in the DNA binding are shown in Figure 1. A model for the interaction will be described below; it is only necessary to note here that the residues are conserved in IHF and TF1 and that the general mode of binding is, therefore, likely to be the same.

The distribution of amino acids in the DNA-binding region is dictated by the geometrical properties of the β-pleated sheet,[29] its structural role as a cover of the helical base, and its required interaction with DNA. The bottom surface of the sheet is populated with the hydrophobic residues that contribute to the molecular core, while the upper exposed surface contains mostly hydrophilic residues including those that interact with the DNA. This sidedness extends to the arms where the side chains of arginines 53 and 55 point in the same direction. Finally, the natural twist of the sheet is such that the outward strand of the arm and the third strand of the sheet are on the inner side of the helical motif, and these are the secondary structure elements that contain the important lysine and arginine residues (Fig. 1).

A particularly striking feature of the DNA-binding region is that it is supported by a rigid base consisting of eight highly conserved phenylalanines (residues 29, 47, 50, and 79 from each monomer) that fill the space under the tent-like sheet region (Fig. 3a). This situation is similar to that found in other proteins that have relatively small DNA binding motifs.[31] In the DNA-binding regulatory proteins that contain the so-called helix-turn-helix motif,[32] the recognition helix is supported by hydrophobic interactions to a second "platform" helix that straddles the major groove. In proteins such as TFIIIA,[33] a tetrahedrally coordinated zinc atom stabilizes the fingerlike domains that are thought to

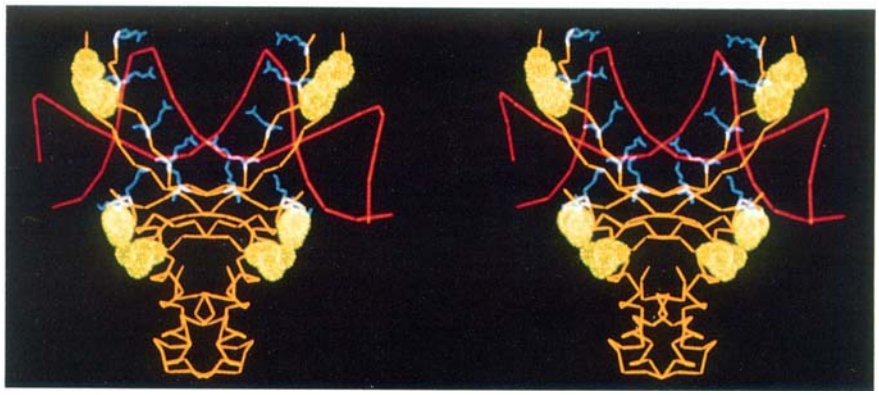

Fig. 2. A stereo view of a model for the nonspecific binding of protein HU to bent or kinked DNA. For clarity, the backbones of the dimeric protein and DNA are represented only by α carbons (yellow) and phosphates (red). Conserved arginine and lysine residues (blue) are well placed to interact with the DNA backbone. Arginine 61 at the end of the arm is not visible in the structure and has been modeled. Note that lysine 83 at the periphery of the molecule will interact only with bent DNA. The green stippled regions represent van der Waals surfaces of conserved, exposed hydrophobic residues. Those in the arm are residues 69 and 71 that project into the expanded minor groove. The other region is comprised of residues 45, 84, and 88. These create a self-complementary hydrophobic patch that could interact with a neighbor in a circular nucleoprotein complex in which the DNA is wrapped around the outside.

interact with the DNA. It has been pointed out[34] that small DNA-binding proteins would require a rigidly supported motif that can interact extensively with the DNA helix whereas this is not required by the large proteins such as *Eco*R1[34] and DNA polymerase I.[35]

## A MODEL FOR THE HU–DNA INTERACTION

The refined structure of HU together with the recent data from IHF described above has allowed a more detailed modeling of the HU–DNA interaction than that described previously.[28]

The matching helical parameters of the DNA binding region and B-DNA allowed a natural docking of the molecules in which the arms are positioned over the minor groove as indicated by the IHF protection experiments. Five conserved positively charged residues (arginines 53, 55, and 58, and lysines 80 and 86) can easily be oriented so as to make good ionic interactions with five successive phosphates on one strand of the DNA. Although it is not possible to see arginine 61, a model can be built in which it forms an ionic interaction with a sixth phosphate group.

In the HU arm, residues 57 and 72 together with 56 and 73 create a distorted β-ribbon structure shortly before the electron density disappears (Fig. 1). A visible hydrogen bond immediately distal to this region suggests a continuation of the normal β-ribbon, but we prefer not to speculate on the conformation of this highly conserved and obviously important region of the protein. The various HU sequences show that the residues in the distorted region have a strong preference to be either proline or alanine.

We suggest two possible functions for the distorted region of the arm. First, it may act as a hinge to allow the DNA easy access to DNA-binding elements in the body of the protein before the end of the arm swings into position. Second, it could act as a spacing element for side chains 58 and 61. Their required spacial positions are dictated by the locations of the phosphates with which they interact on the surface of the cylindrical DNA molecule. However, in a perfect β-ribbon in which side chains alternate in direction,[29] their orientations might not all be correct. This is clearly the case for arginine 58 that makes a good contact with the DNA backbone in our model but that would be pointing in the opposite direction were it not for the distortion.

Thus far, no attempt has been made to distort the DNA away from its classical B conformation. However, there is clearly considerable steric interference between the return side of the arm and the DNA backbone on the other side of the minor groove. This immediately suggests that the correct model involves bent or kinked DNA since an opening of the minor groove in this region would relieve the steric interference. Furthermore, the totally conserved lysine 83 is unable to contact the DNA in the unbent case but makes excellent contact after bending. We make no predictions concerning the possible DNA structure at the kink, but a model which involves

the expansion of the minor groove has been described.[5] Figure 2 shows a model for the HU–DNA interaction that incorporates the features described above.

## SPECIFIC STRUCTURE–FUNCTION RELATIONSHIPS

Having identified the structural and functional features that are common to HU, IHF, and TF1, and shown how these create a structure that can bind kinked or bent DNA, we are now in a position to rationalize how the differences between the molecules may explain their particular functions.

### Protein HU

Although there is disagreement concerning the precise function of HU, there is no doubt that it has the ability to form nucleosome-like particles in which the DNA is negatively supercoiled. These particles must involve a degree of protein–protein interaction, and based on the 3 Å structure,[28] we suggested that several wedge-shaped HU molecules could easily close into a circular object. The model for DNA binding and bending described above has allowed a more detailed investigation of this proposal.

HU does not form stable higher order structures in the absence of DNA,[11] and the particles are rather unstable.[36] Therefore, although the relevant surfaces should reflect the protein–protein contact, strong interactions are not to be expected. From the HU–DNA model, it can be seen where an adjacent HU molecule would be located in the particle, and several features are immediately obvious.

Glycines 46, 48, and 82 are all highly conserved, map close to the contact region, and create a depression on the surface that the C-terminal helix from an adjacent molecule is well situated to occupy. Glycines 46 and 48 do have a definite structural role in the protein since they occur in a tight turn between strands 1 and 2, and a side chain at position 48 would experience steric conflict with phenylalanine 47. Nevertheless, their proximity to glycine 82 is suggestive. Also, it is perhaps significant the the helix residues 84 and 88 that would mediate the contact tend to be small hydrophobic alanines. Another distinctive feature in this area is the conserved, totally exposed hydrophobic residue 45. This would be inserted into a space beneath the C-terminal helix of the adjacent molecule that is lined by the hydrophobic residues 84, 85, and 88 (Fig. 2). Thus, an extensive hydrophobic interface would be created between two molecules by these reciprocating interactions. The possible contacts between the helical bases are less obvious, but the region contains several charged residues and ionic interactions can be constructed between Glu 5 and Arg 37, and Lys 3 and Asp 30. All four are relatively well conserved.[11] Finally, as we suggested previously,[28] there is the possibility that the arms from adjacent molecules contact each other at the rear of the DNA.

### Integration Host Factor

The major functional difference between HU and IHF is that the former binds DNA nonspecifically whereas the latter binds specifically. When the amino acid sequences of the putative DNA binding regions of IHF and HU are compared, several differences are obvious, and these may constitute the recognition elements.

The major difference occurs at the distortion in the arm: the usual alanine and proline residues of HU are replaced by Asn-56, Gln-57, and Thr-72 on subunit α and Glu-72 on subunit β. All four residues could form specific hydrogen bonded interactions with DNA bases. When bound to DNA, this region of the molecule may be close to the bases. In HU, small side chains would be favored whereas in IHF, these positions would be ideal for the residues involved in "reading" the DNA sequence. It should also be noted that arginine 53 is replaced by a histidine in the β-subunit and lysine 83 by a glutamine in the α-subunit. Finally, residue 74, which points directly into the DNA-binding domain, is an arginine and lysine in the α- and β-subunits, respectively. All these differences occur in regions of the molecule that are clearly visible in our 2.1 Å structure and that are therefore relatively rigid. A model of the IHF DNA-binding domain that displays these features is shown in Figure 3b.

It should be noted that the IHF subunits have C-terminal extensions compared to HU and these may also have a role in DNA recognition. However, in our model (Fig. 2), this part of the protein is directly over the major groove, which does not appear to be protected by DNA binding.

Many IHF binding sites from a variety of systems[11,15,37,38] have now been analyzed. The most recent consensus sequence[38] is YAANNNNTT-GATW; this shows that four central bases have little or no role in determining the specificity. The importance of the flanking bases has been confirmed by experiments in which the bases are specifically altered.[39] A single molecule of IHF bound at the center of the sequence appears to be an ideal model since the potential recognition residues described above would then be close to the crucial DNA bases (compare Figs. 2 and 3b). Recent measurements of the stoichiometry of the IHF/DNA complex have confirmed that one protein molecule binds to the recognition sequence (H.A. Nash, personal communication). This explains the lack of any obvious surface on the molecule that could mediate a protein–protein interaction. The exposed hydrophobic residues 45, 84, and 88 of HU are not present and, indeed, it has been shown that the HU–DNA supercoiling experiments cannot be easily repro-

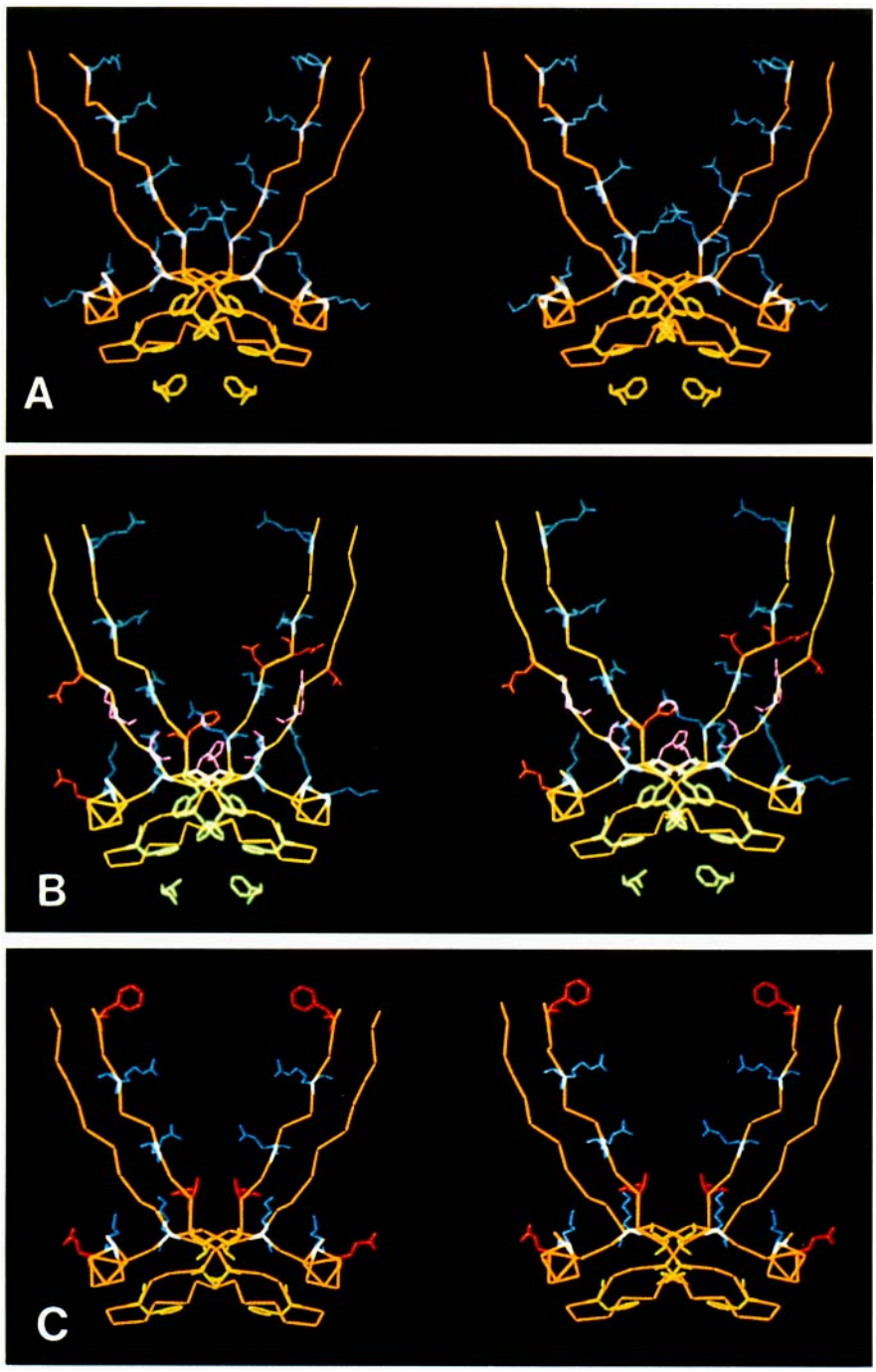


Fig. 3. Stereo views of the DNA binding regions of HU, IHF, and TF1. **a:** (Top) The DNA binding region of HU. The region is a tent-like structure consisting of two symmetry-related antiparallel β-pleated sheets that fits like a lid over the helical base of the molecule (not shown). The two DNA binding arms extend away from the inner strands of the sheet and the C-terminal helices are on either side. Note that the region is supported by a cluster of 8 phenylalanines. The arms create a helical motif and the conserved arginines and lysines (blue) line its inner surface. Arginine 61 at the end of the arm is not visible in the structure and has been modeled. **b:** (Middle) A model of the specific DNA binding region of IHF. This was produced by combining the sequences of the IHF monomers with the coordinates of HU. The residues shown in red are those that differ from conserved residues in HU and may function as DNA recognition elements. Residues in violet are others that may also be involved. For reasons of clarity these are not labeled, but details are in the text. **c:** (Bottom) The same as (**b**) for TF1. The residues that may confer the partial DNA specificity on TF1 are shown in red. Note phenylalanine 61 at the end of the arm, which is proposed to intercalate the DNA bases.

duced with IHF (H.A. Nash, personal communication).

A novel feature of this model is the recognition of the DNA sequence via the minor groove. This is unusual because it is known from both theoretical considerations[40] and crystallographic data[31] that this normally occurs via the major groove. However, if our model of DNA bending is correct, the minor groove would be considerably expanded and the bases more accessible.

### Transcription Factor 1

Geiduschek and co-workers have compared the sequences of TF1 and HU and suggested how TF1 might preferentially bind DNA which contains hydroxymethyluracil.[41-43] Although based on our previous 3 Å structure, these ideas are entirely consistent with the model described above and will be summarized for completeness.

The major differences between HU and TF1 are the loss of three of the positively charged residues that we suggest interact with DNA. These are at positions 53, 61, and 83, and are replaced by a valine, a phenylalanine, and a glutamic acid, respectively (Fig. 3c). One would predict that this would result in a severe reduction in the ability of the molecule to bind normal DNA, and this has been shown to be the case.[43] It is suggested that this is compensated for by the ability of TF1 to recognize a particular structural feature of SPO1 DNA. Recent work has shown that certain sites in the SPO1 DNA show an enhanced affinity for TF1,[41] but there is no obvious sequence homology between them.[42] Therefore, the partial specificity is unlikely to involve a direct interaction with the hydroxyl group of the modified base. One proposal is that the relative instability of SPO1 DNA[44] would allow the phenylalanine at position 61 to intercalate between the bases.

TF1 also has a nine residue C-terminal extension compared to HU[25] and this has been shown to be essential for the binding.[23] A model based on fluorescence studies[45] places this in the major groove. The model does not incorporate the proposed bending of the DNA, and we point out that this would result in a more favorable contact.

## DISCUSSION

It has been shown that IHF actually induces the bound DNA to adopt a bent or kinked conformation rather than stabilizing an already bent region of DNA.[9] Recent crystallographic studies confirm that the B-DNA helix can be readily deformed in protein–DNA complexes.[31] Is it possible with this class of protein to identify the mechanism by which the DNA is bent?

Mirzabekov and Rich,[46] based on the histone–DNA contacts in nucleosome core particles, have suggested that the driving force for DNA bending might result from a protein that binds and neutralizes the negative charges on one side of the DNA. This region would thereby be converted from a stiff extended molecule to one that experiences interphosphate repulsions on one side. A similar, but more complicated situation is present in our model. It can be seen in Figure 2 that the central DNA phosphates are neutralized, whereas those at the back and at the periphery are not. In a complex with adjacently bound proteins, such as in a nucleosome-like particle (see below), the phosphates between protein dimers and on the back side of the DNA would all repel each other. The resulting deformation might generate the negative DNA supercoils that are frequently associated with this type of protein.[11,19,36,42,47]

An additional driving force may be provided by key amino acids on the return side of the arm. Residues 69 and 71 are hydrophobic and totally conserved in all proteins of this type.[11] In our model (Fig. 2), both point directly at the highly charged DNA backbone, thereby causing maximal steric interference and an energetically unstable situation. Also, residues 67, 68, and 70 have a tendency to be negatively charged. Thus, the outgoing strand of the arm can be imagined to firmly bind the sugar-phosphate backbone on one side of the DNA minor groove while the return strand repels the DNA backbone on the other side of the groove and promotes the bending.

There has been much discussion concerning the protein–DNA stoichiometry within the HU nucleosome-like particle. Rouviere-Yaniv and co-workers[19] demonstrated that each particle is associated with 8–10 HU dimers and that one negative supercoil (or a linking number change of $-1$) occurs every 275 base pairs. These parameters were largely confirmed by the later work of Broyles and Pettijohn[36] [19($\pm$1) monomers/linking number change, 1 supercoil/290 base pairs]. However, they also showed evidence for an alteration of the DNA helical twist to 8.5 base pairs per turn when the protein is bound, and failed to detect any long stretches of unbound (linking) DNA. A model has been proposed for the HU nucleosome that attempts to satisfy these various parameters.[11] It consists of 290 base pairs wrapped 7.5 times around 5 tetramers of HU. This number of turns is required to compensate for the apparent overwinding of the DNA helix.

This model is clearly different from the one we have proposed[28] in which the DNA makes one supercoil turn around a core of 8–10 HU dimers. We feel that the general shape of the dimer together with the models of the HU–DNA and HU–HU interactions described above are consistent with our simple model. We can find no structural evidence that the HU dimer might itself dimerize to form the tetrameric object required in the alternative model, and consider it unlikely that 40 base pairs could associate with such an object to form one turn of a

supercoil. A key parameter of the alternative model is the 8.5 base pairs per turn of the bound DNA. This was based on the size of fragments produced by the digestion of the DNA when bound in the HU complex. We point out that the conformation of the DNA would be severely distorted if our bent model is correct, and suggest that the value of 8.5 represents the base pair distance between successive kinks.

What is the function of IHF? It has been shown to affect a wide variety of cellular activities,[11] and a search of the available *E. coli* DNA sequences reveals many potential binding sites.[48] It has been suggested[7–9,11] that its role is to introduce bends or kinks into precise locations on the DNA in order to promote the formation of larger complexes. It might even help to create the looped structures that have been proposed in some types of gene regulation.[49] Bacteriophages such as λ, ϕ80, and P22 apparently exploit this property of IHF to form the nucleosome-like structure that is central to the mechanism of integrative recombination.[37,47]

Recent structural results have clearly shown that the DNA conformation in protein–DNA complexes can be very different from the classical B-DNA molecule.[31] We therefore emphasize that our models should not be regarded as being accurate to atomic resolution, but as useful and reasonable starting points to explain the various properties of the proteins. The recent cloning of the genes that code for the HU, IHF, and TF1[11,13,14,18,25] will permit site-directed mutagenesis experiments to test the various ideas that have been proposed above.

## ACKNOWLEDGMENTS

The authors particularly wish to thank Dr. Howard Nash for invaluable discussions and for providing unpublished data. We also thank Drs. R. Weisberg, K. Mizuuchi, and E.P. Geiduschek for helpful comments and Dr. Sanford Lacks for critically reading the manuscript. S.W. White was supported by the Office of Health and Environmental Research of the United States Department of Energy.

## REFERENCES

1. Marini, J., Levene, S., Crothers, D.M., Englund, P.T. Bent helical structure in kinetoplast DNA. Proc. Natl. Acad. Sci. USA 79:7664–7668, 1982.
2. Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., Klug, A. Structure of the nucleosome core particle at 7Å resolution. Nature (London) 311:532–537, 1984.
3. Kirchhausen, T., Wang, J.C., Harrison, S.C. DNA Gyrase and its complexes with DNA: Direct observation by electron microscopy. Cell 41:933–943, 1985.
4. Crick, F.H.C., Klug, A. Kinky helix. Nature (London) 255: 530–533, 1975.
5. Sobell, H.M., Tsai, C.-C., Jain, S.C., Gilbert, S.G. Visualization of drug-nucleic acid interactions at atomic resolution. J. Mol. Biol. 114:333–365, 1977.
6. Wu, H.-M., Crothers, D.M. The locus of sequence-directed and protein-induced DNA bending. Nature (London) 308: 509–513, 1984.
7. Stenzel, T.T., Patel, P., Bastia, D. The integration host factor of *Escherichia coli* binds to bent DNA at the origin of replication of the plasmid pSC101. Cell 49:709–717, 1987.
8. Prentki, P., Chandler, M., Galas, D.J. *Escherichia coli* integration host factor bends the DNA at the ends of IS1 and in an insertion hotspot with multiple IHF binding sites. EMBO J. 6:2479–2487, 1987.
9. Robertson, C.A., Nash, H.A. Bending of the bacteriophage lambda attachment site by *Escherichia coli* integration host factor. J. Biol. Chem. 263:3554–3557, 1988.
10. Nash, H.A., Robertson, C.A. Purification and properties of the *Escherichia coli* protein factor required for lambda integrative recombination. J. Biol. Chem. 256:9246–9253, 1981.
11. Drlica, K., Rouviere-Yaniv, J. Histonelike proteins of bacteria. Microbiol. Rev. 51:301–319, 1987.
12. Nash, H.A., Robertson, C.A., Flamm, E., Weisberg, R.A., Miller, H.I. Overproduction of *Escherichia coli* integration host factor, a protein with nonidentical subunits. J. Bacteriol. 169:4124–4127, 1987.
13. Miller, H.I. Primary structure of the *himA* gene of *Escherichia coli:* Homology with the DNA-binding protein HU and association with the phenylalanyl-tRNA synthetase operon. Cold Spring Harbor Symp. Quant. Biol. 49:691–698, 1984.
14. Flamm, E.L., Weisberg, R.A. Primary structure of the *hip* gene of *Escherichia coli* and of its product, the beta subunit of integration host factor. J. Mol. Biol. 183:117–128, 1985.
15. Craig, N.L., Nash, H.A. *E. coli* integration host factor binds to specific sites in DNA. Cell 39:707–716, 1984.
16. Mende, L., Timm, B., Subramanian, A.R. Primary structures of two homologous ribosome-associated DNA-binding proteins of *Escherichia coli.* FEBS Lett. 96:395–398, 1978.
17. Rouviere-Yaniv, J., Kjeldgaard, N. Native *Escherichia coli* HU protein is a heterotypic dimer. FEBS Lett. 106:297–300, 1979.
18. Kano, Y., Yoshino, S., Wada, M., Yokoyama, K., Nobuhara, M., Imamoto, F. Molecular cloning and nucleotide sequence of the HU-1 gene of *Escherichia coli.* Mol. Gen. Genet. 201:360–362, 1985.
19. Rouviere-Yaniv, J., Germond, J., Yaniv, M. *Escherichia coli* DNA binding protein HU forms nucleosome-like structure with circular double-stranded DNA. Cell 17:265–274, 1979.
20. Bjornsti, M.A., Hobot, J.A., Kelus, A.S., Villiger, W., Kellenberger, E. New electron microscopy data on the structure of the nucleoid and their functional consequences. In: "Bacterial Chromatin," Gualerzi C.O. and Pon C.L., eds. Heidelberg: Springer-Verlag, 1986, 64–81.
21. Johnson, G.G., Geiduschek, E.P. Specificity of the weak binding between the phage SPO1 trasncription-inhibitory protein, TF1, and SPO1 DNA. Biochemistry 16:1473–1485, 1977.
22. Wilson, D.L., Geiduschek, E.P. A template-selective inhibitor of in vitro transcription. Proc. Natl. Acad. Sci. U.S.A. 62:514–520, 1969.
23. Sayre, M.H., Geiduschek, E.P. TF1, the bacteriophage SPO1-encoded type II DNA-binding protein, is essential for viral multiplication. J. Virol. 62:3455–3462, 1988.
24. Geiduschek, E.P., Ito, J. Regulatory mechanisms in the development of lytic bacteriophages in *Bacillus subtilis.* In: "The Molecular Biology of the Bacilli," Dubnau, D.A., ed., Vol. 1. New York: Academic Press, 1982, 203–245.
25. Greene, J.R., Brennan, S.M., Andrew, D.J., Thompson, C.C., Richards, S.H., Heinrikson, R.L., Geiduschek, E.P. Sequence of the bacteriophage SPO1 gene coding for transcription factor 1, a viral homologue of the bacterial type II DNA-binding proteins. Proc. Natl. Acad. Sci. U.S.A. 81: 7031–7035, 1984.
26. Dijk, J., White, S.W., Wilson, K.S., Appelt, K. On the DNA-binding protein II from *Bacillus stearothermophilus.* Purification, studies in solution and crystallization. J. Biol. Chem. 258:4003–4006, 1983.
27. Tanaka, I., White, S.W., Appelt, K., Wilson, K.S., Dijk, J. The structure of DNA binding protein II at 6Å resolution. FEBS Lett. 165:39–42, 1984.
28. Tanaka, I., Appelt, K., Dijk, J., White, S.W., Wilson, K.S. 3Å resolution structure of a protein with histone-like properties in prokaryotes. Nature (London) 310:376–381, 1984.

29. Creighton, T.E. "Proteins—Structures and Molecular Principles." New York: W.H. Freeman, 1983.
30. Lammi, M., Paci, M., Gualerzi, C.B. Proteins from the prokaryotic nucleoid. The interaction between proteins NS and DNA involves the oligomeric form of the protein and at least one Arg residue. FEBS Lett. 170:99–104, 1984.
31. Ollis, D.L., White, S.W. Structural basis of protein-nucleic acid interactions. Chem. Rev. 87:981–995, 1987.
32. White, S.W. Prediction of DNA-binding regulatory proteins in bacteriophage T7. Protein Engin. 1:373–376, 1987.
33. Miller, J., McLachlan, A.D., Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. 4:1609–1614, 1985.
34. McClarin, J.A., Frederick, C.A., Wang, B.-C., Greene, P., Boyer, H.W., Grable, J., Rosenberg, J.M. Structure of the DNA–*Eco*RI endonuclease recognition complex at 3Å resolution. Science 234:1526–1541, 1986.
35. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., Steitz, T. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature (London) 313:762–766, 1985.
36. Broyles, S., Pettijohn, D.E. Interaction of the *E. coli* HU protein with DNA: Evidence for formation of nucleosome-like structures with altered DNA pitch. J. Mol. Biol. 187:47–60, 1986.
37. Leong, J.M., Nunes-Duby, S., Lesser, C.F., Youderian, P., Susskind, M.M., Landy, A. The ϕ80 and P22 attachment sites. J. Biol. Chem. 260:4468–4477, 1985.
38. Gamas, P., Chandler, M.G., Prentki, P. and Galas, D.J. *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS1 and to its major insertion hot-spot in pBR322. J. Mol. Biol. 195:261–272, 1987.
39. Gardner, J.F., Nash, H.A. Role of *Escherichia coli* IHF protein in lambda site-specific recombination. J. Mol. Biol. 191:181–189, 1986.
40. Seeman, N.C., Rosenberg, J.M., Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. U.S.A. 73:804–808, 1976.
41. Greene, J.R., Geiduscheck, E.P. Site-specific DNA binding by the bacteriophage SPO1-encoded type II DNA-binding protein. EMBO J. 4:1345–1349, 1985.
42. Greene, J.R., Morrissey, L.M., Foster, L.M., Geiduschek, E.P. DNA binding by the bacteriophage SPO1-encoded type II DNA-binding protein, transcription factor I. J. Biol. Chem. 261:12820–12827, 1986.
43. Greene, J.R., Morrissey, L.M., Geiduschek, E.P. DNA binding by the bacteriophage SPO1-encoded type II DNA-binding protein, transcription factor I. J. Biol. Chem. 261: 12828–12833, 1986.
44. Okubo, S., Strauss, B.S., Stodolsky, M. The possible role of recombination in the infection of competent *Bacillus subtilis* by bacterophage deoxyribonucleic acid. Virology 24: 552–562, 1964.
45. Hard, T., Hsu, V., Sayre, M.H., Geiduschek, E.P., Appelt, K., Kearns, D.R. Fluorescence studies of a single tyrosine in a type II deoxyribonucleic acid binding protein. Biochemistry 28:396–406, 1989.
46. Mirzabekov, A.D., Rich, A. Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. Proc. Natl. Acad. Sci. U.S.A. 76:1118–1121, 1979.
47. Richet, E., Abcarian, P., Nash, H.A. The interaction of recombination proteins with supercoiled DNA: Defining the role of supercoiling in lambda integrative recombination. Cell 46:1011–1021, 1986.
48. Mechulam, Y., Blanquet, S., Fayat, G. Dual level control of the *Escherichia coli pheST-himA* operon expression. J. Mol. Biol. 197:453–470, 1987.
49. Ptashne, M. Gene regulation by proteins acting nearby and at a distance. Nature (London) 322:697–701, 1986.