Recognition of DNA by IHF: Sequence Specificity Mediated by Residues That Do Not Contact DNA

Erik K. Read[1], Eun Hee Cho[2], and Jeffrey F. Gardner[3]

[1]Wadsworth Anaerobe Laboratory 11301 Wilshire Ave. Los Angeles, CA 90073 USA
[2]Department of Science Education Chosun University Kwangju 501-759 KOREA
[3]Department of Microbiology University of Illinois. Urbana, Il 60801USA

Summary

The Integration Host factor (IHF) of *Escherichia coli* is a small, basic protein that is required for a variety of functions including site-specific recombination, transposition, gene regulation, plasmid replication, and DNA packaging. It is composed of two subunits that are encoded by the *ihfA* (α-subunit) and *ihfB* (β-subunit) genes. IHF binding sites are composed of three elements called the WATCAR, TTG, and poly (dAT) elements. We have characterized IHF binding to the H site of bacteriophage λ. We have isolated suppressors that bind to altered H' sites using a challenge phage selection. Two different suppressors were isolated that changed the adjacent αP64 and αK65 residues. The suppressors recognized both the wild-type site and a site with a change in the WATCAR element. Three suppressors were isolated at β-E44. These suppressors bound the wild-type and a mutant site with a T:A to A:T change (H44A) in the middle of the TTR element. Site-directed mutagenesis was used to make several additional changes at βE44. The wild-type and βE44D mutant could not bind the wild-type site but were able to bind the H44A mutant site. Other mutants with neutral, polar, or a positive charge at β-E44 were able to repress both the wild-type and H44A sites. Examination of the IHF crystal structure suggests that the ability of the wild-type and βE44D proteins to discriminate between the T:A and A:T basepairs is due to indirect interactions. The β-E44 residue does not contact the DNA directly. It imposes binding specificity indirectly by interactions with residues that contact the DNA. Details of the proposed interactions are discussed.

Introduction

Integration Host Factor (IHF) of *Escherichia coli* was discovered as a protein required for phage λ integration [1]. In the following years, IHF was shown to participate in a variety of cellular processes such as control of gene expression, plasmid replication, transposition and chromosome compaction [2, 3]. IHF plays an architectural role in these processes by introducing bends in DNA that facilitate the formation of specialized higher-order nucleoprotein structures.

IHF is a member of the HU-like family of DNA binding proteins. It is a small (20 kDa) basic protein composed of α (Mr = 11.200) and β (Mr = 10,600) heterodimers. HU-like proteins have two domains: a helix-turn-helix domain involved in dimerization and an arm that is composed of two antiparallel β-sheets and a C-terminal α–helix that bind to and bend DNA [4, 5]. IHF is a sequence-specific DNA binding protein [6]. IHF binding sites have three conserved elements [7, 8]. The WATCAR (where A, C, G and T are the standard bases and W is A or T, and R is A or G) element is positioned in the

middle of the site. The TTR element, positioned 3 to the WATCAR element, is separated from The WATCAR element by 4 weakly-conserved basepairs. A third element, the poly (dAT) tract is composed of 4 to 6 bases. It is located several base-pairs 5 of the WATCAR element and is present in some sites.

In a previous genetic study [8] substitution mutations that disrupted IHF binding were isolated in the phage λ H IHF site. Mutations were found in all three of the sequence elements indicating that they are important for IHF interactions with DNA. Footprinting, methylation protection and analog studies indicated that IHF interacts primarily with the minor groove of DNA [6, 9, 10]. A crystal structure for IHF bound to the H site was recently reported [4, 11]. The DNA is wrapped tightly around the protein and the β-ribbon arms wrap around the DNA. The DNA is bent about 180°. Interestingly, contacts between IHF and DNA only occur in the minor groove of the WATCAR and TTR elements. The binding specificity of IHF is probably achieved by a combination of direct protein-DNA contacts to bases in the minor groove as well as a large number of small interactions that depend on the sequence-dependent flexibility of DNA [11]. This work presents genetic evidence which, in conjunction with the crystal structure, supports an indirect readout mechanism for specific protein-DNA contacts. A glutamic acid residue in the β-subunit (β-Glu44) buttresses two arginine residues (β-Arg42 and β-Arg46) at positions where they can interact with DNA.

Altered Binding Specificity Mutants of IHF

A common genetic approach towards analyzing protein-DNA interactions is to isolate suppressor mutants that have altered or expanded DNA binding abilities. The rationale behind this approach is that mutant proteins that show altered or expanded binding specificities are likely to contain amino acid substitutions at residues that are close to or contact DNA. Such information could complement physical and biochemical studies on the molecular interactions between a protein and its DNA binding site. Challenge phages containing individual substitution mutations in the λ H IHF binding site (Fig. 1) were used to select for IHF mutants that could bind to a specific mutant site. One challenge phage contained an A-T to C-G change at position 38 of the H site. Two IHF mutants repressed this challenge phage [8]. One mutant changed the proline at position 64 of the α-subunit to leucine (αP64L) and the other changed lysine 65 to serine (αK65S). Three mutants repressed a second challenge phage that contained

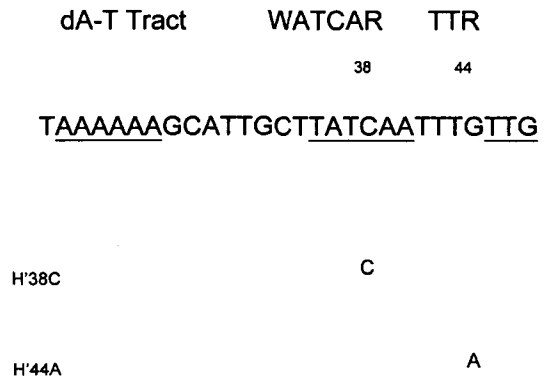

Fig. 1. The sequence of the phage λ H' site and the mutations in the challenge phages are shown. The d(AT) tract and the WATCAR and TTR elements are underlined. Position 38 contains an A:T base-pair that was mutated to C:G in P22-H'38C. Position 44 contains a T:A that was mutated to an A:T in P22-H44.

a T-A to A-T transversion at position 44 in the TTR element of the H site. Sequencing of the mutants revealed that glutamic acid 44 of the β-subunit was changed to lysine (βE44K), valine (βE44V) or glycine (βE44G).

In order to examine the in vivo binding specificity of each mutant, challenge phage assays using phages with H site mutants were performed. The proteins with the substitutions at αP64 and αK65 repressed only the wild-type and H38C (P22 H38C) sites efficiently. The mutant proteins with the substitutions at βE44 repressed the wild-type and H44A (P22 H44A) sites efficiently [8],Data not shown). Thus, the mutant proteins have expanded binding specificity because they bind the mutant site used in the original selection as well as the wild-type H site. In addition, the results indicate that the αP64, αK65, and βE44 residues play important roles in recognition of DNA by IHF.

The αP64 residue is conserved in all known members of the IHF/HU family indicating that it plays an important role in binding DNA. The crystal structure of IHF bound to the H site showed that αP64 intercalates into DNA and introduces a kink [11]. Examination of the structure does not suggest a simple mechanism for how the αP64L and αK65S proteins recognize the wild type and H38C sites. Thus, the mechanism of DNA binding awaits future X-ray crystal analysis. Further genetic analysis of mutants of the βE44 residue in combination with the X-ray structure suggests a mechanism for how the residue at β44 influences DNA binding.

The Role of βE44 in DNA Recognition

The βE44 residue was substituted with various amino acids that were not isolated in the suppressor analysis described above and the mutant proteins tested for DNA binding in challenge phage assays [12]. Challenge phages containing the wild-type site and the P22-H'44A phage, which contains an A:T base-pair instead of a T:A base-pair at position 44 of the TTR element, were used to measure the

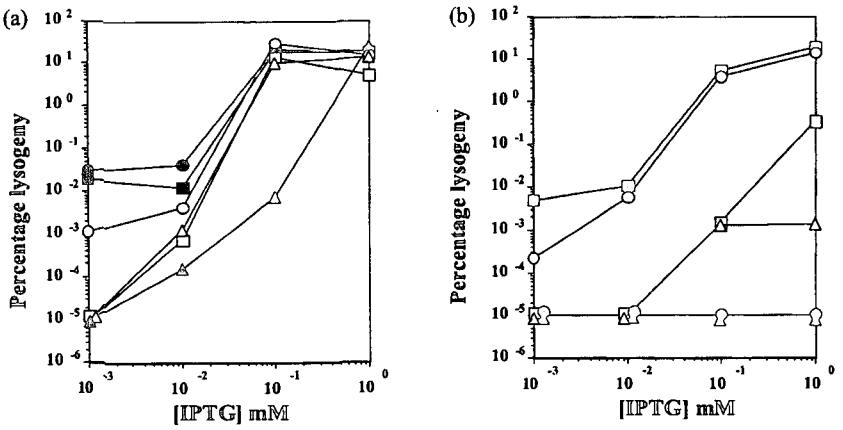

Fig. 2. Results of challenge-phage assays performed with (a) the wild-type H site and (b) the H'44A mutant site. IHF was expressed from a plasmid where expression of IHF was under control of the P*tac* promoter. Expression of IHF was induced by exposing cells to different concentrations of IPTG. Binding of IHF to the wild-type or H'44A sites results in repression of lethal phage-encoded antirepressor protein. This allows formation of kanamycin resistant lysogens. The frequency of kanamycin resistant lysogens is proportional to the ability of IHF to bind to the H site. (For a review on the challenge phage system see Lee et al [13] or Maloy and Youderian [14]. Wild -type IHF (-●-) or mutant IHF where βE44 was replaced by the following amino acids; alanine (-▣ -), glycine (-○-), aspartic (-△ -),Asparagine (-□-), glutamine (-△-). Adapted from Read et al [12]

relative binding affinities of the mutant proteins. The results are shown in Fig. 2. Mutants with substitutions of lysine (βE44K), valine (βE44V), alanine (βE44A), glycine (βE44G), asparagine (βE44N), or glutamine (βE44Q) were able to repress both the wild-type H site and the H'44A site. However, the mutant with the conservative aspartic acid change (βE44D) was unable to repress the challenge phage containing the H'44A site. This result suggests that proteins with an acidic acid residue at position 44 can discriminate between a T:A and an A:T base-pair because neither protein represses the altered site with the A:T transversion. However, the mutants with other substitutions cannot discriminate between sites with a T:A or an A:T at position 44. It appears that residues that lack a negative charge can bind to both sites.

How do proteins with negatively charged residues discriminate between a T:A and an A:T base pair at position H44? Examination of the crystal structure of Rice et al [11] provides a possible mechanism. In the crystal structure of IHF bound to the H site, βE44 does not contact the DNA directly. One of the oxygens of the carboxyl group interacts with the charged ureido group of βR42. The other oxygen of βE44 makes hydrogen bonds to both ureido groups of βR46 (Fig. 3A). The βR42 and βR46 residues make additional contacts with DNA including a salt bridge, a hydrogen bond and a van der Waals interaction not shown in the figure. The series of bonds between βR42, βE44, and βE46 form part of a "clamp" that allows βR46 to hydrogen bond with thymine 44. The hydrogen bonds between βE44 and βR46 holds the latter residue in a position where it cannot adapt to the disruptive interactions imposed by the adenine residue in the H'A44 base. The network of bonds holds βR46 in a position where it cannot move to accommodate an adenine at position H'44. Computer modeling indicates that substitution of the aspartic acid at position 44 could still allow a hydrogen bond to form between the

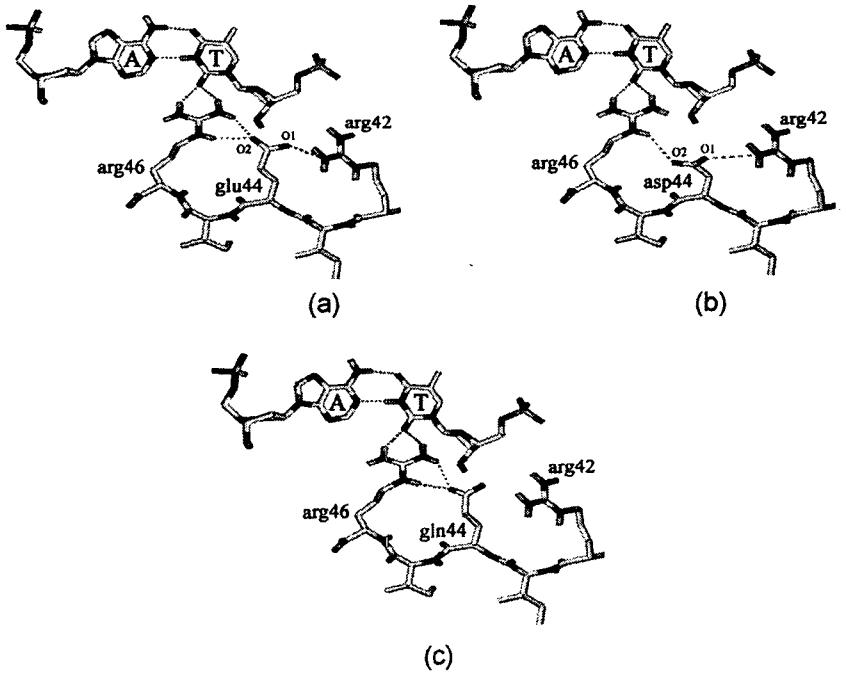

Fig. 3. Basepair 44 of the wild-type H' site and β1-sheet of the IHF β-subunit depicted with the wild-type glutamic acid at position 44 (a), or substituted with aspartic acid (b), or glutamine (c). Dotted green lines indicate hydrogen bonds. The coordinates were modeled with the program Swiss-pdviewer PPC version 3.5 and amino acid substitutions were modeled with the mutate function of the programs. Adapted from Read et al [12]

carboxyl group of βD44 and the ureido group of βR42 and one hydrogen bond between βD44 and a ureido group of βR46 (Fig. 3B). The other amino acid substitutions would not allow a hydrogen bond to form between the residue at position β44 and βR42 (Fig. 3C). We propose that the hydrogen bond between βE44 and βR42 controls part of the specificity of binding to the TTR element. In cases where the bond between the residue at position β44 and βR42 cannot form, the βR42 and βR46 residues have the conformational freedom to bind the H44A site without steric clashes. Thus, all three residues are required for full discrimination against the P22-H44A phage.

In summary, the wild-type βE44 residue forms hydrogen bonds with βR42 and βR46 and provide binding specificity indirectly by positioning the βR46 residue so that it cannot accommodate the steric clash with the H44A base. This interaction is interesting because it shows how interactions away from the protein-DNA interface can confer sequence specificity. We are currently attempting to perform X-ray crystal analyses of the mutant proteins bound to DNA.

References

1. Nash, H.A. and C.A. Robertson, (1981). Purification and properties of the *Escherichia coli* protein factor required for lambda integrative recombination. J Biol Chem. 256: 9246-53.
2. Friedman, D.I., (1988). Integration host factor: a protein for all reasons. Cell. 55: 545-54.
3. McLeod, S.M. and R.C. Johnson, (2001). Control of transcription by nucleoid proteins. Curr Opin Microbiol. 4: 152-159.
4. Rice, P.A., (1997). Making DNA do a U-turn: IHF and related proteins. Curr Opin Struct Biol. 7:86-93.
5. Tanaka, I., K. Appelt, J. Dijk, S.W. White, and K.S. Wilson, (1984). 3-A resolution structure of a protein with histone-like properties in prokaryotes. Nature. 310: 376-81.
6. Craig, N.L. and H.A. Nash, (1984). *E. coli* integration host factor binds to specific sites in DNA. Cell. 39: 707-16.
7. Goodrich, J.A., M.L. Schwartz, and W.R. McClure, (1990). Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). Nucleic Acids Res. **18**: 4993-5000.
8. Lee, E.C., L.M. Hales, R.I. Gumport, and J.F. Gardner, (1992). The isolation and characterization of mutants of the integration host factor (IHF) of *Escherichia coli* with altered, expanded DNA-binding specificities. Embo J. 11: 305-13.
9. Yang, C.C. and H.A. Nash, (1989). The interaction of *E. coli* IHF protein with its specific binding sites. Cell. **57**:869-80.
10. Wang, S., R. Cosstick, J.F. Gardner, and R.I. Gumport, (1995). The specific binding of *Escherichia coli* integration host factor involves both major and minor grooves of DNA. Biochemistry. 34: 13082-90.
11. Rice, P.A., S. Yang, K. Mizuuchi, and H.A. Nash, (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell. **87**: 1295-306.
12. Read, E.K., R.I. Gumport, and J.F. Gardner, (2000). Specific recognition of DNA by integration host factor. Glutamic acid 44 of the beta-subunit specifies the discrimination of a T:A from an A:T base pair without directly contacting the DNA. J Biol Chem. 275: 33759-64.
13. Lee, E.C., M.P. MacWilliams, R.I. Gumport, and J.F. Gardner, (1991). Genetic analysis of *Escherichia coli* integration host factor interactions with its bacteriophage lambda H' recognition site. J Bacteriol. 173: 609-17.
14. Maloy, S.R. and J.F. Gardner, Use of P22 challenge phage to identify protein-nucleic acid binding sites, . (1998).