The Role of Surface-Exposed Lysines in Wrapping DNA about the Bacterial Histone-Like Protein HU†

Anne Grove* and Tatiana C. Saavedra‡

Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803

Received December 27, 2001; Revised Manuscript Received April 21, 2002

ABSTRACT: Several basic proteins, including the ubiquitous HU proteins, serve histone-like functions in prokaryotes. Significant sequence conservation exists between HU homologues; yet binding sites varying from 9 to 37 bp have been reported. TF1, an HU homologue with a 37 bp binding site that is encoded by the *Bacillus subtilis* bacteriophage SPO1, binds with nM affinity to DNA that contains 5-hydroxymethyluracil (hmU) in place of thymine and to T-containing DNA with loops. We evaluated the contribution of three conserved lysines to specifying the length of the binding site and show that Lys3 is critical for maintaining a long binding site in T-containing DNA: A mutant protein in which Lys3 is replaced with Gln(TF1−K3Q) is completely deficient in forming a stable complex. The affinity for 37 bp hmU-containing DNA is also reduced, from ∼3 nM for wild-type TF1 to ∼90 nM for TF1−K3Q. The decrease in affinity of TF1−K3Q for hmU-containing DNA ≥ 25 bp suggests that Lys3 contacts DNA 8−9 bp distal to the sites of kinking. We propose that Lys3 forms an internal saltbridge to Asp26 in HU homologues characterized by shorter binding sites and that its surface exposure, and hence a longer binding site, may correlate with absence of this aspartate.

Prokaryotic cells contain a number of small DNA-binding proteins whose primary function is to compact and organize the genomic DNA. Of these so-called histonelike proteins, the ubiquitous HU protein is the most abundant (*1*−*4*). The most prominent feature of HU is its capacity for wrapping the DNA into nucleosome-like structures, readily visible in electron micrographs (*5*). The structure of HU from *Bacillus stearothermophilus* reveals a compact body generated by the association of two identical monomers from which two β-strands extend into flexible arms that wrap around DNA, generating a significant DNA bend (*6*−*8*). The structure of the HU homologue Integration Host Factor (IHF), whose sequence-specific DNA binding permitted the cocrystallization of a protein-DNA complex, revealed two DNA kinks originating from the intercalation of proline residues from the minor groove, wedging apart two base pair steps separated by 9 bp of duplex and generating an ∼160° DNA bend (*9*). The extensive sequence and structural homology, including the almost universal conservation of the DNA-intercalating proline (*3*, *10*), suggests a similar mode of DNA contact for HU proteins.

For *Escherichia coli* HU, linear DNA is saturated at about one HU dimer per 9 bp of duplex DNA, regardless of sequence (*5*, *11*). Other HU homologues have been reported to prefer longer DNA sites and to bind with some sequence preference (*10*, *12*−*14*). For instance, IHF and another sequence-specific HU homolog, the *B. subtilis* bacteriophage SPO1-encoded TF1, have ∼37 bp binding sites (*9*, *15*, *16*). Sequence-specificity of binding has been proposed to correlate with the short C-terminal extension found in both IHF and TF1 compared to HU proteins (*17*−*20*). However, general determinants of binding site length have not been specified, nor predicted from the sequence of HU proteins.

We have addressed the role of the three most highly conserved, surface-exposed lysines in wrapping of DNA about the body of the HU homolog, TF1, and in determining the length of the binding site. For TF1, which binds DNA as a homodimer, we find that Lys3 is essential for specifying a long, 37 bp DNA site and propose that its surface-exposure in TF1 is dictated by the absence of an otherwise highly conserved aspartate with which it forms a saltbridge in HU homologues characterized by shorter binding sites.

EXPERIMENTAL PROCEDURES

Protein Preparation. TF1 mutant proteins were generated by PCR amplification of plasmid pTF1X (a gift from E. P. Geiduschek), using a forward primer designed to introduce the appropriate substitution at positions 3 or 18 of TF1 and a reverse primer positioned to abut the forward primer. TF1−K3QK86Q was generated similarly, using plasmid encoding TF1−K3Q as the template and a forward primer designed to introduce the mutation at position 86 (primer sequences available on request). The PCR reaction, carried out with a mixture of Taq and Pfu polymerase, generated a full-length plasmid containing the mutated TF1 gene. The original template DNA was removed by Dpn I digestion, and the plasmid harboring the mutated TF1 gene was used to transform *E. coli* BL21(DE3)pLysS. Introduction of the desired mutation was confirmed by sequencing.

† Supported in part by the Louisiana Board of Regents Support Fund (LEQSF(2000-03)-RD-A-07 to A.G.).

* Corresponding author. Tel: (225) 578-5148. Fax: (225) 578-8790. E-mail: agrove@lsu.edu.

‡ Current address: Louisiana State University, School of Medicine, New Orleans, LA 70112.

[1] Abbreviations: hmU, 5-hydroxymethyluracil; IHF, Integration Host Factor.

10.1021/bi016095e CCC: $22.00
Published on Web 05/18/2002

Purified wild-type TF1 was generously given by E. P. Geiduschek. Overexpression and purification of TF1 mutant proteins was accomplished as described for TF1 (*21*). Peak fractions of protein eluted from heparin-Sepharose were judged to be >95% pure by Coomassie blue staining of SDS−polyacrylamide gels. Protein concentrations were determined from the absorbance at 280 nm, using an extinction coefficient of 1200 M^{-1} for the TF1 monomer which contains a single tyrosine.

Preparation and Labeling of DNA Probes. Oligonucleotides with hmU content were provided by L. Mayol (*21*, *22*). T-containing oligonucleotides were purchased and purified by denaturing polyacrylamide gel electrophoresis. The top strand was ^{32}P-labeled at the 5′-end with T4 polynucleotide kinase. Equimolar amounts of complementary oligonucleotides were mixed, heated to 90 °C, and slowly cooled to 4 °C to form duplex DNA.

Electrophoretic Mobility Shift Assay and Quantitation of Protein−DNA complexes. Electrophoretic mobility shift assays were performed as described (*21*). Gels were prerun for 30 min at 20 mA at room temperature before loading the samples with the power on, except for experiments with 19-mer and 15-mer duplexes which were performed at 4 °C. Reaction conditions were as described (*21*, *23*), and each sample contained 50 or 100 fmol DNA in a total reaction volume of 10 μL, except for experiments with wild-type TF1 for which 10 fmol DNA was used. For competition experiments, ^{32}P-labeled 37 bp hmU-containing DNA (50 fmol) and unlabeled competitor DNA was mixed first, followed by addition of protein. After electrophoresis, gels were dried, and protein-DNA complexes were visualized and quantified by phosphoimaging, using software supplied by the manufacturer (ImageQuant 1.1). For experiments shown in Figures 4−6, the concentration of DNA was 5 nM.

Equilibrium dissociation constants, Kd, were determined from the slope of Scatchard plots of DNA titrated with protein (*21*). Assuming binding to a single site and introducing the degree of binding as $[TF1]_B/[DNA]_T$, the Scatchard equation yields $[TF1]_B/([DNA]_T[TF1]_F) = K_a - K_a[TF1]_B/[DNA]_T$, where $[TF1]_B$ and $[TF1]_F$ are the concentrations of bound and free TF1, respectively, $[DNA]_T$ is the total concentration of DNA, and K_a is the equilibrium association constant. This method of determining K_d does not require correction for dissociation of protein−DNA complexes during electrophoresis since such correction would entail adjusting the fractional complex formation $[TF1]_B/[DNA]_T$ by an exponential decay factor representing complex dissociation during electrophoresis, a correction that would not change the slope of the Scatchard plot. All experiments were carried out at least in triplicate. Kd values are reported as the average ± the standard error of the mean.

RESULTS

Design of TF1 Mutants. Like IHF, the homodimeric TF1 kinks its specific DNA target site at two positions separated by 9 bp of duplex (*9*, *21*, *24*). Since substitution of a residue at the ends of the DNA-binding arms significantly affect DNA binding and bending and since compensatory changes can be made to DNA at the sites that are subject to TF1-mediated kinking, TF1 is expected to engage its DNA target similarly to IHF (*9*, *24*, *25*). This notion is supported also by the significant structural homology (9,26). Sequence alignments of HU homologues reveal several positions of significant sequence identity, including three positions where lysine is highly favored (Figure 1; *10*). As illustrated from the solution structure of TF1 (Figure 1; 1, *26*), Lys3 is positioned about two-thirds down the body of the protein, relative to the positively charged cradle from which the DNA-embracing β-arms protrude. In TF1, Lys3 extends directly out from the surface. With 9 bp of duplex engaged by residues lining the top of the β-sheet, followed by proline-mediated kinks at either end, contacts with Lys3 would be predicted to occur with DNA that is appreciably longer than the 11 bp that participate in these interactions. In IHF, the corresponding lysines interact with DNA ~9 bp distal to the sites of kinking (*9*). We note that conservation of Lys3 does not correlate with the reported length of individual HU binding sites. Lys86 is located toward the top of the structure, in a position to interact with DNA flanking the sites of kinking. Indeed, the structure of IHF in complex with DNA shows a lysine equivalent to Lys86 of TF1 interacting with the DNA phosphodiester backbone within the central 9 bp of duplex, immediately before the DNA kink (the corresponding residue within the β subunit, Arg, reaches sideways and contacts DNA distal to the kink; *9*). Lys86 would therefore be expected to be important for contacts to the central part of the DNA site. The third, highly conserved lysine is in position 18, located at the opposite end of the protein relative to the primary region of DNA contact. TF1 is an exception, as it does not feature lysine in this position, but glutamine.

Whereas the structure of TF1 is consistent with the proline-mediated kinking of DNA bordering a central 9 bp segment, no strongly electrostatically favored path for distal DNA along the lateral sides of the protein is evident (*26*). To identify residues involved in wrapping DNA about the body of the protein, we therefore focused our mutational analysis on the conserved lysines. We assessed the role of Lys3 and Lys86 by generating TF1 mutant proteins wherein these residues are replaced with glutamine (TF1−K3Q and TF1−K3QK86Q). The role of lysine in position 18 was evaluated by substituting Gln18 for Lys (TF1−Q18K). TF1 mutant proteins were overexpressed as described for the wild-type protein and purified to apparent homogeneity, as determined by Coomassie Blue staining of SDS−PAGE gels (Figure 2).

Affinity of TF1 Mutants for hmU-Containing DNA. TF1 binds to specific sites within bacteriophage SPO1 genomic DNA and bends its DNA target by an estimated 140° (*25*, *27*, *28*). TF1 also exploits the substitution of all thymine residues with 5-hydroxymethyluracil (hmU) that is a characteristic of SPO1 DNA by binding specifically and with high affinity only to DNA that contains this modification (*21*, *28*). TF1 binds 37 bp T-containing DNA without sequence specificity and with much lower affinity (*21*, *27*). Notably, the ~3 nM affinity of TF1 for 37 bp hmU-containing DNA is reproduced in T-containing DNA that contains a set of 4-nt loops separated by 9 bp of duplex (*21*).

We compared the affinity of TF1 and the mutant proteins to 37 bp hmU-containing DNA that represents the preferred TF1 binding site that overlaps the P_E6 early promoter (Figure 3). For TF1, 37 bp corresponds to the optimal binding site, accommodating exactly one TF1 dimer (*21*). Previous analyses of TF1−DNA complex formation demonstrated that

A

```
                1. . . .10 . .  . .20 . . . .30 . . . .40. . . . 50
SPO1-TF1   ..MNKTELIKAIAQDTE.LTQVSVSKMLASFEKITTETVAKGDKVQLTGFLNI
Bsu        ..MNKTELINAVAEASE.LSKKDATKAVDSVFDTILDALKNGDKIQLIGFGNF
Eco-HUβ    ..MNKTQLIDVIAEKAE.LSKTQAKAALESTLAAITESLKEGDAVQLVGFGTF
Eco-HUα    ..MNKSQLIDKIAAGAD.ISKAAAGRALDAIIASVTESLKEGDDVALVGFGTF
Eco-IHFα   MALTKAEMSEYLFDKLG.LSKRDAKELVELFFEEIRRALENGEQVKLSGFGNF
Eco-IHFβ   ..MTKSELIERLATQQSHIPAKTVEDAVKEMLEHMASTLAQGERIEIRGFGSF

                 . . . .60 . . . .70 . . . .80 . . .  90 . . . .
SPO1-TF1     KPVARQARKGFNPQTQEALEIAPSVGVSVKPGESLKKAAEGLKYEDFAK
Bsu          EVRERSARKGRNPQTGEEIEIPASKVPAFKPGKALKDAVAGK.......
Eco-HUβ      KVNHRAERTGRNPQTGKEIKIAAANVPAFVSGKALKDAVK.........
Eco-HU       AVKERAARTGRNPQTGKEITIAAAKVPSFRAGKALKDAVN.........
Eco-IHFα     DLRDKNQRPGRNPKTGEDIPITARRVVTFRPGQKLKSRVENASPKDE..
Eco-IHFβ     SLHYRAPRTGRNPKTGDKVELEGKYVPHFKPGKELRDRANIYG......
```

B

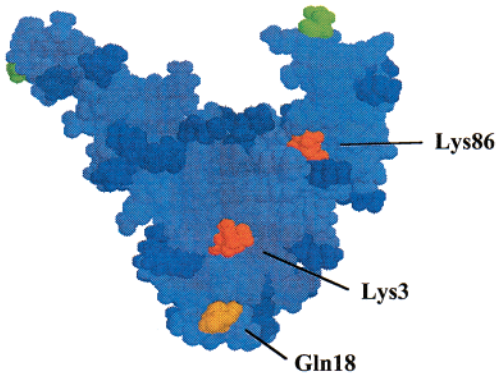

FIGURE 1: Position of conserved lysines. (A) Alignment of TF1; HU from *B. subtilis* and *E. coli* and IHF from *E. coli*. Residues highlighted in blue are >80% conserved among all HU homologues; the most highly conserved lysines are indicated in red (*10*). The prolines responsible for kinking the DNA target are indicated in green. SPO1−TF1: TF1, *B. subtilis* bacteriophage SPO1 (sp|P04445). Bsu: HU, *B. subtilis* (sp|P08821). Eco-HUβ: HU β, *E. coli* (sp|P02341). Eco-HUα: HU α, *E. coli* (sp|P02342). Eco-IHFα: IHF α, *E. coli* (sp|P06984). Eco-IHFβ: IHF β, *E. coli* (sp|P08756). (B) Structure of the dimeric TF1 (*26*; PDB accession number 1WTU) with residues targeted for mutagenesis identified (Q18, yellow; K3 and K86, red). Other lysine residues are highlighted in dark blue, and P63 is shown in green. Illustration generated with RasMol.

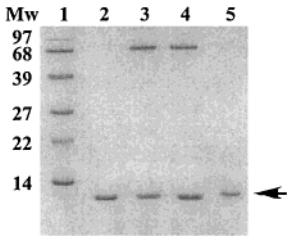

FIGURE 2: Purified TF1 and mutant proteins. Coomassie Blue-stained SDS−PAGE gel showing molecular weight markers in lane 1. Lanes 2−5 contain ~0.5 μg of TF1, TF1−K3Q, TF1−K3QK86Q, and TF1−Q18K, respectively (arrow). Small variations in electrophoretic mobility correlate with the number of surface-exposed lysines. Samples in lanes 3 and 4 contain BSA.

```
h37   5′-CCHAGGCHACACCHACHCHHHGHAAGAAHHAAGCHHC-3′
      3′-GGAHCCGAHGHGGAHGAGAAACAHHCHHAAHHCGAAG-5′

h33     5′-HAGGCHACACCHACHCHHHGHAAGAAHHAAGCH-3′
        3′-AHCCGAHGHGGAHGAGAAACAHHCHHAAHHCGA-5′

h29       5′-GGCHACACCHACHCHHHGHAAGAAHHAAG-3′
          3′-CCGAHGHGGAHGAGAAACAHHCHHAAHHC-5′

h25         5′-CHACACCHACHCHHHGHAAGAAHHA-3′
            3′-GAHGHGGAHGAGAAACAHHCHHAAH-5′

h19           5′-CACCHACHCHHHGHAAGAA-3′
              3′-GHGGAHGAGAAACAHHCHH-5′

h15             5′-CCHACHCHHHGHAAG-3′
                3′-GGAHGAGAAACAHHC-5′
```

FIGURE 3: Sequences of hmU-containing oligonucleotides of decreasing length. DNA constructs were generated by symmetrically shortening h37 at both ends. H stands for hmU. The sequence of h37 corresponds to a preferred binding site for TF1 within the SPO1 genome. Base pair steps that are subject to protein-mediated kinking are underlined (*21*, *24*). The position of a short inverted repeat is indicated by arrows.

binding energy resulted both from interactions with the central region of the 37 bp binding site and with more distal DNA sites (*21*). The affinity of TF1 mutant proteins for shorter DNA probes, generated by symmetrically shortening h37 at both ends, was also measured.

The DNA constructs were used in electrophoretic mobility shift assays to compare the affinity of the TF1 variants. An ~30-fold reduction in affinity for the 37 bp probe is observed for TF1−K3Q, which has an apparent equilibrium dissociation constant Kd ~ 87 nM compared to 2.6 nM for wild-type TF1 (Figure 4 and Table 1). This reduction in affinity is accompanied by a significant increase in complex dissociation during electrophoresis (evidenced by density in the region of the gel between the position of the complex and the free DNA and data not shown). Complex dissociation during electrophoresis was previously analyzed for wild-type TF1, and it does not affect the Scatchard analysis (see Experimental Procedures). A further reduction in DNA-binding affinity is evident for TF1−K3QK86Q, which fails to form complexes that are stable to electrophoresis (Table 1 and data not shown). These observations are consistent with a role for Lys3 and Lys86 in securing essential contacts

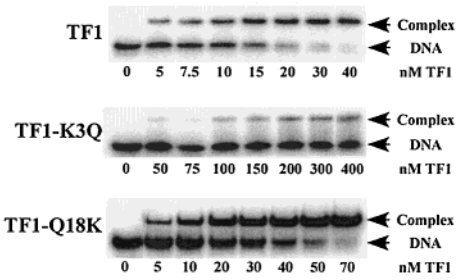

FIGURE 4: Electrophoretic analysis of TF1 variants binding to 37 bp hmU-containing DNA. Proteins are identified at the left, and their concentrations are indicated below each panel.

Table 1: Dissociation Constant Kd (nM) for hmU-Containing DNA of Decreasing Length[a]

	TF1[b]	TF1−Q18K	TF1−K3Q	TF1−K3QK86Q
h37	2.6 ± 0.3	13.1 ± 1.0	87.1 ± 5.6	n/c
h33	5.7 ± 1.0	31.0 ± 4.7	94.6 ± 5.7	n/c
h29	9.5 ± 0.9	71.0 ± 6.6	199.9 ± 16.4	n/c
h25	26.3 ± 3.2	n/c	n/c	n/c
h19	50.1 ± 2.3	n/c	nd	nd
h15	n/c	n/c	nd	nd

[a] The sequences of DNA probes are shown in Figure 3. The notation n/c indicates the absence of a complex that is stable to electrophoresis; nd, not determined. [b] From ref *21*.

to DNA. TF1−Q18K also has reduced affinity for 37 bp DNA (Kd = 13 nM). We surmise that a local redistribution of electrostatic surface potential resulted from introduction of the additional lysine and that this altered surface potential is less favorable for DNA contacts.

In analogy with the IHF-DNA complex, TF1 is expected to recognize a central 9 bp duplex bordered by two sites of enhanced DNA flexibility that serve as target sites for proline intercalation from the minor groove (*9*, *21*, *24*, *26*). DNA extending beyond this central segment makes additional contributions to binding interactions. The expectation would therefore be for the K3Q and Q18K mutations to affect primarily affinity for the longer DNA constructs. Not surprisingly, TF1−K3QK86Q is unable to form stable complex with the shorter duplexes (Table 1 and data not shown). As for TF1, both TF1−K3Q and TF1−Q18K exhibit a diminution of affinity when the DNA length is decreased from 37 to 29 bp (Figure 5 and Table 1). No significant change in affinity is observed for TF1−K3Q upon reducing duplex length from 37 to 33 bp, suggesting that maximal binding interactions are attained with the 33-mer (Figure 5B). However, neither mutant protein is able to form a complex with 25 bp DNA that is stable to electrophoresis. As a decrease in affinity correlates with an increased rate of complex dissociation during electrophoresis (*21*; data not shown), we suspect that any complex that formed with 25 bp DNA would have dissociated during electrophoresis. The reduction in binding energy that accompanies a decrease in duplex length from 29 to 25 bp may be explicitly rationalized in terms of binding interactions between the end of the h29 duplex and sites on the lateral sides of the protein that the shorter duplexes cannot reach. The absence of complex formation between TF1−Q18K and h25 again suggests that introduction of additional surface charge causes a redistribution of electrostatic potential across a more extended region of the protein surface.

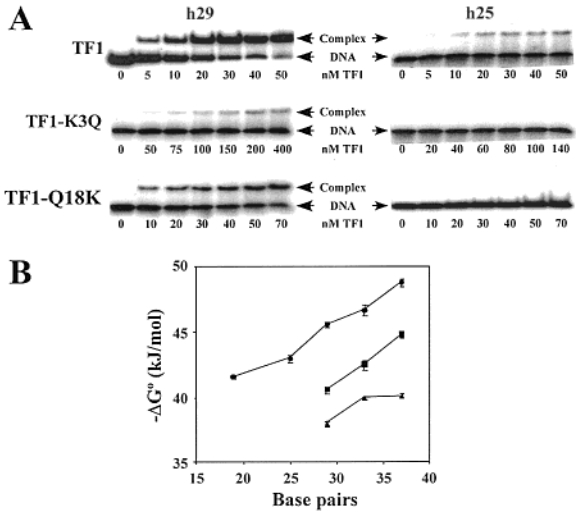

FIGURE 5: Affinity of TF1−K3Q and TF1−Q18K for shorter hmU-containing DNA. (A) Binding of TF1 variants, identified at the left, to 29 or 25 bp hmU-containing DNA. Protein concentrations are indicated below each panel. (B) The standard free energy of association, $-\Delta G°$, determined from Scatchard plots, as a function of duplex length. (●) Wild-type TF1; (■) TF1−Q18K; (▲) TF1−K3Q.

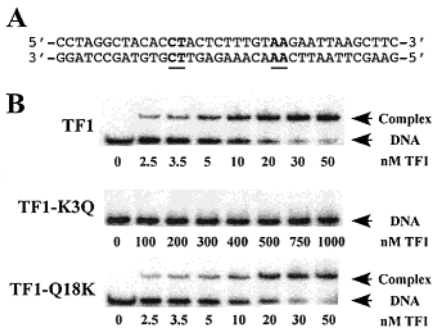

FIGURE 6: Lys3 is critical for interaction with T-containing DNA flanking the sites of protein-mediated kinking. (A) 37 bp loop-containing DNA (T-292). The top strand sequence corresponds to h37, except hmU is replaced with T. Loops are tandem mismatches (in boldface and underlined), generated by altering the sequence of the bottom strand to generate mismatches composed of identical opposing bases. (B) Electrophoretic analysis of TF1 variants binding to T-containing DNA with loops shown in (A). Protein concentrations are indicated below each panel.

Lys3 is Critical for TF1 Binding to T-Containing DNA. All mutant proteins retain a preference for hmU-containing DNA compared to T-containing DNA, forming only barely detectable (TF1−Q18K) or undetectable (TF1−K3Q and TF1−K3QK86Q) complex with the 37 bp fully duplexed T-containing DNA (data not shown). High-affinity binding to T-containing DNA is restored upon introduction of a set of 4-nt loops separated by 9 bp of duplex for both wild-type TF1 (*21*) and TF1−Q18K; however, no complex formation is seen with TF1−K3Q (Figure 6). This observation demonstrates that Lys3 is critical for TF1 binding to T-containing DNA, either because TF1−K3Q no longer binds preferably to DNA with loops or due to a paucity of interaction with DNA flanking the central loop-containing DNA segment. The remarkable difference in TF1−K3Q complex formation with 37 bp hmU-containing DNA and 37 bp T-containing DNA with loops was further explored by assessing complex formation with 37 bp heteroduplex DNA in which one strand has T content and the other contains hmU. Regardless of

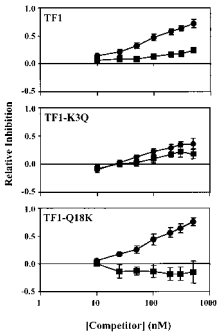

FIGURE 7: Inhibition of complex formation with 37 bp hmU-containing DNA. Relative inhibition is plotted as a function of the concentration of competitor DNA. (●) T-292 (Figure 6A); (■) h25 (Figure 3). Relative inhibition is calculated as C_0-C_i/C_0 where C_0 is complex formation in the absence of competitor and C_i is complex formation in the presence of competitor. The concentration of competitor, indicated below the bottom panel, is identical for all panels. Protein concentrations: TF1, 10 nM; TF1−K3Q, 50 nM; TF1−Q18K, 20 nM.

which strand harbors the hmU-for-T substitutions, no complex was detectable with TF1−K3Q (data not shown). By contrast, wild-type TF1 displays an intermediate affinity for such heteroduplexes between that observed for all-T- or for all-hmU-containing DNA (*21*).

The ability of TF1−K3Q to bind preferentially to loop-containing DNA was also explored by comparing complex formation with heteroduplex DNA and heteroduplex DNA with 4-nt loops. As pointed out above, no complex was seen with heteroduplex DNA; however, TF1−K3Q does form complex with the loop-containing heteroduplex that is stable to electrophoresis, indicating that it retains the ability to bind with increased affinity to loop-containing DNA (data not shown).

The interpretation that Lys3 is important for contacts to T-containing DNA distal to the sites of kinking is valid only if TF1−K3Q makes comparable contacts with hmU-containing DNA and T-containing DNA with loops across the central ~11 bp segment, an explicit inference made with regard to wild-type TF1 (*21*, *24*). Accordingly, TF1−K3Q may exhibit comparable interaction with h25 and the 37 bp T-containing DNA with loops (T-292; Figure 6A), but neither DNA construct forms a complex with TF1−K3Q that is stable to electrophoresis. We therefore assessed the ability of h25 and T-292 to inhibit complex formation with h37 (Figure 7; these experiments were performed with a DNA concentration above the Kd). For wild-type TF1 and TF1−Q18K, complex formation with h37 is effectively inhibited by T-292, as expected. A more modest inhibition of TF1−K3Q−h37 complex formation is seen with T-292, but this DNA construct competes more efficiently for complex formation than h25, consistent with the notion that TF1−K3Q does make contact both to h25 and T-292 across the central segment. An important finding in support of this interpretation is that fully duplex, 37 bp T-containing DNA has no inhibitory effect on complex formation within this concentration range (data not shown). These observations demonstrate that introduction of loops into T-containing DNA significantly increases binding to TF1−K3Q and suggest that the diminished affinity of TF1−K3Q for T-292 compared to h37 is due primarily to a paucity of interaction with distal DNA segments. We also note that the inhibitory effect of T content on TF1−K3Q binding to heteroduplex DNA is not absolute; heteroduplex DNA inhibits complex formation with h37 at a level intermediate between that seen with T-292 and with h25 (Figure 7 and data not shown). This is consistent with the properties of wild-type TF1 which has lower affinity for heteroduplex DNA compared to h25 (*21*).

In contrast, no inhibition of complex formation with TF1−Q18K was observed within this concentration range upon addition of h25, but a modest, yet finite, increase. Since TF1−Q18K can form higher-order complex on the 37 bp DNA (data not shown), we surmise that oligomerization in solution may also occur and that this oligomerization is inhibited by h25, increasing the effective concentration of TF1−Q18K (both T-292 and h25 inhibit formation of higher order complex; data not shown).

DISCUSSION

On the basis of the experimentally determined DNA-binding properties (*17*, *20*, *21*, *23*, *24*, *27*, *28*), analysis of the electrostatic surface potential of TF1 (*26*), and by comparison to the IHF-DNA structure (*9*), the positively charged cradle and the flexible β-ribbon arms of TF1 are expected to engage the DNA target through minor groove interactions, culminating in two prolines (Pro63, Figure 1) wedging apart hmU-A base pair steps separated by 9 bp (Figure 3). Mutagenesis of residues within the DNA binding arms of TF1 confirm that this region is indeed in contact with the DNA and establish that Pro63 is situated at the sites of DNA kinking (*24*, *25*). Correspondingly, conserved arginine residues within the β-ribbon arms of *B. stearothermophilus* HU were shown to play a key role in DNA binding (*29*). TF1 binds sequence-specifically to the DNA probe used in this study, with the two hmU-A base pair steps disposed symmetrically about the center of the probe (underlined in Figure 3) being essential and sufficient for high-affinity binding (*21*, *24*, *25*). As the primary DNA binding region, the β-ribbon arms, is unaltered in the described TF1 mutant proteins, it is therefore a strong expectation that the mutant proteins, like wild-type TF1, engage a central DNA segment via their DNA-binding arms and that they kink their DNA target by means of Pro63 intercalation.

Residues from each IHF subunit, equivalent to Lys86 of TF1 (Figure 1), are seen in the IHF-DNA structure to contact the phosphodiester backbone at either side of the proline-mediated DNA kink (*9*). The inability of TF1−K3QK86Q to engage DNA stably suggests that Lys86 of TF1 likewise contributes important stabilizing interactions by contacting DNA in the vicinity of the DNA kinks. Consistent with this interpretation, we note that a previously reported mutational analysis of *B. subtilis* HU showed that substitution of Lys86 for alanine resulted in a mutant protein that retained only ~20% binding activity (*30*).

A DNA path anchored by Pro63 mediating kinks at the hmU-A base pair steps places Lys3 in a position to interact with a distal DNA region: DNA distal to each kink is expected to return quickly to B-form, to remain largely in a single plane and to follow a relatively straight path, presumably towards positive electrostatic patches at Lys8 or Lys30 which are situated immediately below and to either side of Lys3 (Figure 1; *9*, *26*). The location of Lys3 in the projected

path of distal DNA is consistent with a critical role in specifying a long binding site. Lys3 may contact residues at the end of a 25 bp duplex or immediately beyond, consistent with the diminution of DNA binding that accompanies a decrease in duplex length from 29 to 25 bp with TF1−K3Q (Figure 5). DNA longer than 25 bp may reach Lys8 or Lys30, permitting more stable complex formation with TF1−K3Q. A direct interaction between the two Lys3 residues of the TF1 dimer and central DNA regions—or DNA bordering the sites of kinking—would not be compatible with the physical location of Lys3 at the protein surface (i.e., the central DNA region is much too short to span the two Lys3 residues), nor with the requirement for interaction between sites of DNA kinking and Pro63. In support of these interpretations, we also point out that another TF1 mutant protein, in which two internal, buried residues were substituted, was recently studied by NMR. It was shown that this TF1 mutant protein, whose overall structure is essentially superimposable on the structure of the wild-type protein, exposes additional positive surface potential on each lateral side of the protein dimer that results in an ~40-fold higher affinity, but only for DNA that is longer than 25 bp (*21*, *31*). This is consistent with an important contribution to affinity at a position that corresponds to the location of Lys3.

TF1 identifies its target site largely through recognition of sequence-dependent DNA deformability (*21*, *24*). Although TF1 binds with sequence preference only to DNA that contains hmU in place of thymine, the hmU-requirement can be localized to the two hmU-A base pair steps that are subject to protein-mediated DNA intercalation (Figure 3); no significant difference in binding affinity is observed whether the intervening sequence or distal DNA has T or hmU content (*24*). Since the requirement for hmU−A base pair steps may be bypassed in T-containing DNA that contains two 4-nt loops separated by 9 bp of duplex (construct T-292; Figure 6A), hmU−A base pair steps were inferred to impose local DNA flexure (*21*, *24*). Other DNA-bending proteins, including IHF, have also been shown to bind with higher affinity to the more pliable hmU version of their target sites (*32*, *33*). Our data clearly show that T content is inhibitory to complex formation with TF1−K3Q and that the loops confer increased binding affinity compared to fully duplex T-DNA, as only T-292 and not the fully duplexed T-DNA is capable of inhibiting binding to hmU-DNA (Figure 7). Thus, comparable interactions within the central DNA segment in hmU-containing DNA and in T-containing DNA with loops (T-292) appear to exist both for wild-type TF1 (*21*, *24*) and for TF1−K3Q and TF1−Q18K. The stimulatory effect of 4-nt loops on complex formation with TF1−K3Q is also evident in heteroduplex DNA; TF1−K3Q fails to form stable complex with heteroduplex DNA but binds heteroduplex with loops with a modest increase in affinity (data not shown). We have shown previously that hmU-A base pair steps (underlined in Figure 3) or 4-nt loops (underlined in Figure 6A) are necessary and sufficient for high-affinity binding by wild-type TF1 and that no hmU-requirements exist for flanking DNA (*21*, *24*). As emphasized above, the β-ribbon arms of TF1 that interact with the central DNA segment are unaltered in TF1−K3Q. The simplest interpretation of the current experiments is, therefore, that TF1−K3Q, like wild-type TF1, recognizes the sites of increased flexure in T-292 (i.e., the loops), resulting in enhanced binding. We therefore infer that the inability of TF1−K3Q to form stable complex with T-292 (Figure 6) is a consequence of a paucity of stabilizing interaction with DNA distal to the sites of kinking and that Lys3 is critical for specifying a long binding site in T-containing DNA.

Differences in affinity for h19 between wild-type TF1 and TF1−K3Q exist (Table 1). However, h19 does inhibit h37-TF1−K3Q complex formation at a level equivalent to that seen with T-292 and h25 (Figure 7), consistent with retention of significant interaction between h19 and TF1−K3Q (data not shown). As TF1 exposes several lysines on the surface (Figure 1B), we suggest that a change in electrostatic surface potential may result, and that such a change may modulate DNA binding affinity. Without structural information for the mutant protein, these considerations are speculative, although we submit that a change in electrostatic surface potential would not be an unexpected consequence of making charge-altering mutations to an already highly charged surface. Such considerations apply particularly to the interpretation of binding properties of TF1−Q18K, as additional positive charge was introduced adjacent to the existent Lys23. The rationale for substituting Gln18 was to assess whether a glutamine in this position contributes to the preference for a long binding site. Although the properties of TF1−Q18K are complex, the answer to the question that formed the rationale for its generation does get answered by the available data: wild-type TF1 and TF1−Q18K have comparable affinities for 37 bp DNA, suggesting that Gln18 is not a significant determinant in conferring a long binding site.

Lys3 is known to be surface-exposed from the structure of TF1. In deciding on a suitable residue with which to substitute this lysine, we opted to introduce another polar, but uncharged residue as opposed to the much smaller and less polar alanine. It is possible that Gln3 may interact with the DNA. However, since our primary observation is a diminution in binding affinity for TF1−K3Q, we would argue that it is the absence of lysine that is detrimental to binding. Second, another primary conclusion is that interactions with DNA distal to the sites of bending are critical, as discussed below. These conclusions would remain valid even if glutamine does interact with DNA at this position, as it would have to be argued that it does so in a distinct and less favorable fashion.

These interpretations of course raises the question of why Lys3 is almost universally conserved despite differences in the reported length of HU binding sites. Second, it implies that the basis for the difference in affinity of TF1−K3Q for 37 bp hmU-containing DNA and T-containing DNA with loops resides in differential contacts to DNA distal to the kinks. Examination of structural information for HU homologues offers a rationale for the conservation of Lys3. In TF1 and IHF, Lys3 extends out from the surface, poised for interaction with a DNA target (*9*, *26*). In contrast, Lys3 is buried within the interior of *B. stearothermophilus* HU, held in place through a saltbridge to Asp26 (*6*−*8*). In TF1, Asp26 is substituted by Ala, whereas the IHF β subunit has Lys in this position. We therefore propose that surface exposure of Lys3, which is located in a position where contacts to a longer DNA site evidently are critical, depends on the absence of a negatively charged residue at position 26 and that HU homologues without Asp or Glu in this position

bind a longer target site. The requirement for DNA contact ~9 bp distal to the sites of kinking is also suggested from the structure of HU from *Thermotoga maritima*, which buries Lys3 in the interior, but instead features a unique, surface-exposed Lys4 (PDB Id: 1B8Z; *34*); *T. maritima* HU has also been shown to prefer an ~37 bp target site (*10*).

In its complex with a 44 bp duplex, another DNA-bending protein, CAP, forms DNA contacts 12−14 bp from the center of the binding site that contribute strongly to affinity; contacts to more distal DNA (beyond a 28 bp core) contribute only modestly (*35*). Evidently, TF1 also derives significant binding energy from contacts at the edges of a 29 bp core; the fact that such contacts contribute significantly to affinity was also demonstrated by the abrupt increase in affinity of a TF1 mutant protein upon increasing target size from 25 to 29 bp (*21*). It is reasonable to expect the free energy of binding to reach a local maximum in a region of the DNA that is not bent, since bending requires energy that may be expected to counteract favorable interactions with the protein. In the case of the CAP-DNA complex, it was proposed that the most efficient use of binding energy may be attained when the interactions necessary for bending are located far from the bend center, allowing the protein to anchor the ends of the DNA and pull them toward the center (*35*). Our data suggest that Lys3 may contribute such optimized interactions.

Interactions of TF1−K3Q with the central region of the binding site appear comparable in hmU-containing DNA and T-containing DNA with loops, leading to the conclusion that it is interactions with distal DNA that cause the observed difference in affinity (Figure 7). Since TF1 binding sites show little sequence conservation and since replacement of hmU-containing DNA flanking the sites of kinking with T-containing DNA has no effect on affinity of wild-type TF1 (*21*), we find it less likely that TF1−K3Q engages in base-specific DNA contacts or in recognition of structural features imposed by hmU content. We suggest instead that the removal of a protein-DNA contact that otherwise serves as a critical means of anchoring the ends of bent DNA exposes a differential flexibility of hmU-containing DNA compared to T-containing DNA with loops (or heteroduplex DNA), otherwise masked in its presence. Contacts to DNA too far from the bend center may be inferior in terms of optimizing leverage, thus permitting only stable complex formation with the more pliable hmU-containing DNA.

ACKNOWLEDGMENT

We are grateful to members of the Grove laboratory for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Inhibition by h19 complex formation with h37 DNA. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Drlica, K., and Rouvière-Yaniv, J. (1987) *Microbiol. Rev. 51*, 301−319.
2. Kellenberger, E., and Arnold-Schultz-Gahmen, B. (1992) *FEMS Microbiol. Lett. 100*, 361−370.
3. Oberto, J., Drlica, K., and Rouvière-Yaniv, J. (1994) *Biochimie 76*, 901−908.
4. Azam, T. A., and Ishihama, A. (1999) *J. Biol. Chem. 274*, 33105−33313.
5. Broyles, S. S., and Pettijohn, D. E. (1986) *J. Mol. Biol. 187*, 47−60.
6. Tanaka, I., Appelt, K., Dijk, J, White, S. W., and Wilson, S. (1984) *Nature 310*, 376−381.
7. White S. W., Appelt K., Wilson K. S., and Tanaka I. (1989) *Proteins: Struct. Funct. Genet. 5*, 281−288.
8. Vis, H., Mariani, M., Vorgias, C. E., Wilson, K. S., Kaptein, R., and Boelens R. (1995) *J. Mol. Biol. 254*, 692−703.
9. Rice, P. A., Yang, S. W., Mizuuchi, K., and Nash, H. A. (1996) *Cell 87*, 1295−1306.
10. Grove, A., and Lim, L. (2001) *J. Mol. Biol. 311*, 491−502.
11. Bonnefoy, E., and Rouvière-Yaniv, J. (1991) *EMBO J. 10*, 687−696.
12. Lavoie, B. D., Shaw, G. S., Millner, A., and Chaconas, G. (1996) *Cell 85*, 761−771.
13. Liu, S.-T., Chang, W.-Z., Cao, H.-M., Hu, H.-L., Chen, Z.-H., Ni, F.-D., Lu, H.-F., and Hong, G.-F. (1998) *J. Biol. Chem. 273*, 20568−20574.
14. Kobryn, K., Naigamwall, D. Z., and Chaconas, G. (2000) *Mol. Microbiol. 37*, 145−155.
15. Greene, J. R., and Geiduschek, E. P. (1985) *EMBO J. 4*, 1345−1349.
16. Nash, H. A. (1996) in *Regulation of Gene Expression in* E. coli (Lin, E. C. C., and Lynch, A. S., Eds.) Chapman and Hall, New York.
17. Sayre, M. H., and Geiduschek, E. P. (1988) *J. Virol. 62*, 3455−3462.
18. Mengeritsky, G., Goldenberg, D., Mendelson, I., Giladi, H., and Oppenheim, A. B. (1993*) J. Mol. Biol. 231*, 646−657.
19. Granston, A. E., and Nash, H. A. (1993) *J. Mol. Biol. 234*, 45−59.
20. Andera, A., and Geiduschek, E. P. (1994) *J Bacteriol. 176(5)*, 1364−73.
21. Grove, A., Galeone, A., Mayol, L. and Geiduschek, E. P. (1996) *J. Mol. Biol. 206*, 196−206.
22. Conte, M. R., Galeone, A., Avizonis, D., Hsu, V. L., Mayol, L., and Kearns, D. R. (1992) *Bioorg. Med. Chem. Lett. 2*, 79−82.
23. Andera, L., Spangler, C. J., Galeone, A., Mayol, L., and Geiduschek E. P. (1994) *J. Mol. Biol. 236*, 139−150.
24. Grove, A., Figueiredo, M. L., Galeone, A., Mayol, L., and Geiduschek, E. P. (1997) *J. Biol. Chem. 272*, 13084−13087.
25. Sayre, M. H., and Geiduschek, E. P. (1990) *J. Mol. Biol. 216*, 819−833.
26. Jia, X., Grove, A., Ivancic, M., Hsu, V. L., Geiduschek, E. P., and Kearns, D. R. (1996) *J. Mol. Biol. 263*, 259−268.
27. Greene, J. R., and Geiduschek, E. P. (1985) *Virology 153*, 46−52.
28. Schneider, G. J., Sayre, M. H., and Geiduschek, E. P. (1991) *J. Mol. Biol. 221*, 777−794.
29. Saitoh, F., Kawamura, S., Yamasaki, N., Tanaka, I., and Kimura, M. (1999) *Biosci. Biotechnol. Biochem. 63*, 2232−2235.
30. Köhler, P., and Marahiel, M. A. (1998) *Mol. Gen. Genet. 260*, 487−491.
31. Liu, W, Vu, H. M., Geiduschek, E. P., and Kearns, D. R. (2000) *J. Mol. Biol. 302*, 821−830.
32. Grove, A., Galeone, A., Mayol, L., and Geiduschek, E. P. (1996) *J. Mol. Biol. 206*, 120−125.
33. Grove, A., Galeone, A., Yu, E., Mayol, L., and Geiduschek, E. P. (1998) *J. Mol. Biol. 282*, 731−739.
34. Christodoulou, E., and Vorgias, C. E. (1998) *Acta Crystallogr., Sect. D 54*, 1043−1045.
35. Liu-Johnson, H.-N., Gartenberg, M. R., and Crothers, D. M. (1986) *Cell 47*, 995−1005.

BI016095E