On the Connection Between Inherent DNA Flexure and Preferred Binding of Hydroxymethyluracil- containing DNA by the Type II DNA-binding Protein TF1

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · AUGUST 1996

Impact Factor: 4.33 · DOI: 10.1006/jmbi.1996.0392 · Source: PubMed

4 AUTHORS, INCLUDING:

Anne Grove
Louisiana State University
72 PUBLICATIONS **1,423** CITATIONS
SEE PROFILE

Aldo Galeone
University of Naples Federico II
114 PUBLICATIONS **1,384** CITATIONS
SEE PROFILE

J. Mol. Biol. (1996) **260**, 196–206

JMB

On the Connection Between Inherent DNA Flexure and Preferred Binding of Hydroxymethyluracil-containing DNA by the Type II DNA-binding Protein TF1

Anne Grove[1], Aldo Galeone[2], Luciano Mayol[2] and E. Peter Geiduschek[1]

[1]*Department of Biology and Center for Molecular Genetics, University of California, San Diego 9500 Gilman Drive, La Jolla CA 92093-0634, USA*

[2]*Dipartimento di Chimica delle Sostanze Naturali Università degli Studi di Napoli Federico II Via Domenico Montesano 49, 80131 Napoli, Italy*

TF1 is a member of the family of type II DNA-binding proteins, which also includes the bacterial HU proteins and the *Escherichia coli* integration host factor (IHF). Distinctive to TF1, which is encoded by the *Bacillus subtilis* bacteriophage SPO1, is its preferential binding to DNA in which thymine is replaced by 5-hydroxymethyluracil (hmU), as it is in the phage genome. TF1 binds to preferred sites within the phage genome and generates pronounced DNA bending. The extent to which DNA flexibility contributes to the sequence-specific binding of TF1, and the connection between hmU preference and DNA flexibility has been examined. Model flexible sites, consisting of consecutive mismatches, increase the affinity of thymine-containing DNA for TF1. In particular, tandem mismatches separated by nine base-pairs generate an increase, by orders of magnitude, in the affinity of TF1 for T-containing DNA with the sequence of a preferred TF1 binding site, and fully match the affinity of TF1 for this cognate site in hmU-containing DNA ($K_d \sim 3$ nM). Other placements of loops generate suboptimal binding. This is consistent with a significant contribution of site-specific DNA flexibility to complex formation. Analysis of complexes with hmU-DNA of decreasing length shows that a major part of the binding affinity is generated within a central 19 bp segment ($\Delta G° = 41.7$ kJ mol^{-1}) with more-distal DNA contributing modestly to the affinity ($\Delta\Delta G = -0.42$ kJ mol^{-1} bp^{-1} on increasing duplex length to 37 bp). However, a previously characterised thermostable and more tightly binding mutant TF1, TF1(E15G/T32I), derives most of its extra affinity from interaction with flanking DNA. We propose that inherent but sequence-dependent deformability of hmU-containing DNA underlies the preferential binding of TF1 and that TF1-induced DNA bendings is a result of distortions at two distinct sites separated by 9 bp of duplex DNA.

Keywords: DNA–protein interactions; transcription factor 1; hydroxymethyluracil; DNA flexibility; phage SPO1

Introduction

The conformation of the phosphodiester backbone is the most variable part of nucleic acid structure. Significant local heterogeneity exists, making sequence-specific conformational variations of double-stranded DNA an integral component of site-specific recognition by DNA-binding proteins. For complexes in which DNA is wrapped around a protein, DNA flexibility is of particular importance in the selection of binding sites (Travers & Klug, 1990). The purest manifestation of this mode of interaction is the formation of nucleosome core particles; the preference for nucleosome occupancy at certain sites speaks to substantial site-selectivity (Satchwell & Travers, 1989; Shrader & Crothers, 1990), although much less than would

Abbreviations used: IHF, integration host factor; hmU, 5-hydroxymethyl-uracil.

0022–2836/96/270196–11 $18.00/0

be afforded through direct readout of base-sequence by amino acid side-chains (Pabo & Sauer, 1992).

The term ''bacterial chromatin'' has been applied, somewhat loosely, to the complex of bacterial chromosomes with numerous basic proteins (Gualerzi & Pon, 1986; Drlica & Rouvière-Yaniv, 1987; Kellenberger & Arnold-Schulz-Gahmen, 1992). In *Escherichia coli*, the most abundant of these proteins is HU, a member of the family of type II DNA-binding proteins. HU proteins are ubiquitous in the Eubacteria; closely related proteins include the *E. coli* integration host factor, IHF, and the *Bacillus subtilis* bacteriophage SPO1-encoded TF1. All the type II DNA-binding proteins significantly bend DNA; the very abundant HU proteins bind DNA non-specifically, resulting in DNA compaction (hence the analogy to eukaryotic nucleoprotein structures). In the case of IHF, DNA bending appears to complement sequence-specific contacts between protein side-chains and DNA bases (Nash, 1996). The relatively large segment of DNA that is protected from DNase I in 1:1 complexes with IHF or TF1, and sharp DNA bending, suggest that DNA wraps around these proteins. Progressive wrapping around laterally associating protein would then be responsible for the DNA-condensing properties associated with TF1 and HU (Johnson & Geiduschek, 1977; Drlica & Rouvière-Yaniv, 1987).

With few exceptions, the HU proteins are 90 to 92-amino acid homodimers. The crystal structure of the *B. stearothermophilus* HU reveals a compact protein whose core is generated by two monomers wrapped around each other. Each monomer consists of two helices arranged in a V-shaped motif, followed by a three-stranded β-sheet. Two β-strands from each monomer extend into loops, or ''arms'', that can embrace a DNA helix resting on top of the β-strands (Tanaka *et al*., 1984; White *et al*., 1989). The high sequence homology between HU, IHF, and TF1 gives credence to the notion that these structural features apply to all three proteins (White *et al*., 1989) and this proves to be the case (Reisman *et al*., 1993; Jia *et al*., 1994; Nash, 1996). TF1 and IHF, both of which bind DNA sequence-selectively, differ from the HU proteins in having carboxy-terminal extensions: TF1 is a homodimer of 99 amino acid subunits, whereas IHF is a heterodimer of 98 and 94 amino acid subunits. These C-terminal ''tails'' are important for DNA recognition (Sayre & Geiduschek, 1988; Granston & Nash, 1993; Mengeritsky *et al*., 1993; Andera & Geiduschek, 1994). Accordingly, sequence-specific recognition by TF1 and IHF has been proposed to be a functional correlate of these C-terminal extensions.

In evaluating the binding preferences of TF1, it was established that it binds avidly to selected sites in a phage genome that is distinguished by the total replacement of thymine (T) with 5-hydroxymethyluracil (hmU) (Johnson & Geiduschek, 1977; Greene & Geiduschek, 1885). TF1 has a markedly reduced affinity for the T-containing version of a preferred binding site in hmU-containing DNA, and this lower affinity is correlated with reduced DNA bending (Schneider *et al*., 1991). These observations suggest that differential DNA flexibility may contribute to complex formation. Our approach to evaluating such contributions to TF1 binding takes advantage of the reduced affinity of the protein for T-containing DNA. Using synthetic oligonucleotides representing a cognate site within the phage genome, we have compared the affinity of TF1 for hmU and T-containing DNA. Model flexible loci, consisting of consecutive mismatches, have been incorporated (see Kahn *et al*., 1994); we show that introduction of such loops at specific positions fully restores the affinity of TF1 for its native binding site within hmU-containing DNA and essentially eliminates the discrimination between T and hmU-containing DNA, consistent with a significant contribution of DNA flexibility to complex formation. We propose that TF1 identifies target sites through recognition of sequence-dependent DNA bendability.

Results

Affinity of TF1 for a 37-mer duplex with hmU or thymine content

In designing suitable DNA probes for the measurement of affinity, we had to take into account several features of TF1 binding. Preferred binding sites within the SPO1 phage genome have been identified (Greene & Geiduschek, 1985), but their sequence does not reveal a simple consensus (Schneider *et al*., 1991). Certain common features can be distinguished, however, including the presence of short inverted repeats. For the experiments that are presented here, we selected the TF1 binding site overlapping the SPO1 early promoter P_E6. The ability of TF1 to form nested complexes on longer DNA fragments (Greene & Geiduschek, 1985) necessitated consideration of a minimum binding site that would accommodate not more than one TF1 dimer. The size of the binding site for TF1 determined by footprinting suggests that such conditions would be satisfied by a 37-mer duplex (Schneider *et al*., 1991). Figure 1 shows 37-mer oligonucleotides with the sequence of the selected TF1 binding site (in (a)). Oligonucleotides, referred to as top and bottom strand, respectively, were synthesised with either T or hmU content. We refer to duplexes composed of two hmU-containing oligonucleotides as hmU/hmU, duplexes in which the bottom strand is replaced by a T-containing oligonucleotide as hmU/T, etc.

These 37-mer duplexes were used in electrophoretic mobility shift assays to compare the affinity of TF1 for hmU/hmU duplex with its affinity for DNA in which one or both strands contain thymine. As shown by Sayre & Geiduschek (1990), the affinity of TF1 is greatly reduced when both strands

(a)	**No loop**	5′-CCTAGGCTACACCTACTCTTTGTAAGAATTAAGCTTC-3′ (* *)
		3′-GGATCCGATGTGGATGAGAAACATTCTTAATTCGAAG-5′
(b)	**6-nt loops**	
	6nt-central	3′-GGATCCGATGTGGATGA**CTT**ACATTCTTAATTCGAAG-5′
	6nt-spacing 7	3′-GGATCCGATGTG**CTA**GAGAAAC**TAA**CTTAATTCGAAG-5′
	6nt-spacing 11	3′-GGATCCGATG**ACC**ATGAGAAACAT**AGA**TAATTCGAAG-5′
	6nt-spacing 17	3′-GGATCCG**TAC**TGGATGAGAAACATTCT**ATT**TTCGAAG-5′
(c)	**4-nt loops**	
	4nt-spacing 7	3′-GGATCCGATGTGG**TA**GAGAAAC**TA**TCTTAATTCGAAG-5′
	4nt-spacing 8	3′-GGATCCGATGTG**CT**TGAGAAAC**TA**TCTTAATTCGAAG-5′
	4nt-spacing 9	3′-GGATCCGATGTG**CT**TGAGAAACA**AA**CTTAATTCGAAG-5′
	4nt-spacing 10[1]	3′-GGATCCGATGT**CC**ATGAGAAACA**AA**CTTAATTCGAAG-5′
	4nt-spacing 11[1]	3′-GGATCCGATGT**CC**ATGAGAAACAT**AG**TTAATTCGAAG-5′
(d)	**(3+2)-nt constructs**	
	(3+2)nt-spacing 7(a)	3′-GGATCCGATGTG**CT-**GAGAAAC**-AA**CTTAATTCGAAG-5′
	(3+2)nt-spacing 7(b)	3′-GGATCCGATGTG**C-A**GAGAAAC**T-A**CTTAATTCGAAG-5′
	(3+2)nt-spacing 7(c)	3′-GGATCCGATGTG**-TA**GAGAAAC**TA-**CTTAATTCGAAG-5′
(e)	**(3+4)-nt constructs**	
	(3+4)nt-spacing 7(a)	3′-GGATCCGATGTG**CTcA**GAGAAAC**TcAA**CTTAATTCGAAG-5′
	(3+4)nt-spacing 7(b)	3′-GGATCCGATGTG**CcTA**GAGAAAC**TAcA**CTTAATTCGAAG-5′

Figure 1. Sequences of 37-mer oligonucleotides. The sequence of the top and bottom strands corresponding to a preferred binding site for TF1 within the SPO1 genome is shown in (a), with the position of a short inverted repeat indicated by arrows, and two T-A steps 9 bp apart noted by asterisks. For loop-containing duplexes, the sequence of the top strand is substituted with inverted polarity at the corresponding site of the bottom strand, yielding two (4-nt loops, (c)) or three (6-nt loops, (b)) consecutive mismatches of identical nucleotides in the top and bottom strand; (3 + 2)-nt and (3 + 4)-nt constructs are derived from the 6nt-spacing 7 loop construct by deleting or inserting one nucleotide in the bottom strand of each loop. For (3 + 2)-nt constructs, the position of the removed nucleotide is indicated by a dash (in (d)); note that (3 + 2)nt-spacing 7(a) can form a (2 + 1)-nt bulge-loop with a spacing of 9 bp. For (3 + 4) -nt constructs, a cytosine is inserted at different positions (lower case c, panel (e)). Oligonucleotides of variant sequence are designated by the number of nucleotides within each loop, followed by the spacing between loops. Sequences generating loops or bulges are in boldface. Oligonucleotides were prepared as written, and also with hmU entirely replacing thymine; except for those marked [(1)], which were made with T-content only.

contain T (Table 1A). Not unexpectedly, hybrid duplexes have intermediate affinities. The affinity for the hmU/hmU duplex (K_d = 2.6 nM) is tenfold greater than for the hmU/T heteroduplex (K_d = 27 nM) and approximately 40-fold greater than for the T/hmU duplex (K_d = 101 nM). For the

Table 1. Dissociation constant, K_d (nM), for 37-mer DNA

		hmU/hmU	hmU/T	T/hmU	T/T
A	No loop	2.6 ± 0.3	26.5 ± 5.9	101.0 ± 9.8	≫
B	6nt-central	17.2 ± 0.2	16.6 ± 0.4	19.7 ± 2.0	19.4 ± 1.2
	6nt-spacing 7	4.8 ± 0.8	4.5 ± 0.4	10.3 ± 1.0	7.1 ± 0.2
	6nt-spacing 11	16.6 ± 1.2	18.4 ± 1.6	44.2 ± 4.9	29.3 ± 2.9
	6nt-spacing 17	24.1 ± 3.1	26.6 ± 2.9	46.0 ± 4.6	29.0 ± 4.7
C	4nt-spacing 7	7.7 ± 2.0	7.6 ± 1.7	18.8 ± 1.6	30.6 ± 3.7
	4nt-spacing 8	6.1 ± 0.7	6.7 ± 1.0	7.1 ± 1.0	7.5 ± 1.0
	4nt-spacing 9	2.2 ± 0.5	2.7 ± 0.6	3.1 ± 0.2	3.4 ± 0.5
	4nt-spacing 10	nd	7.2 ± 1.5	nd	8.6 ± 1.5
	4nt-spacing 11	nd	15.2 ± 1.7	nd	34.0 ± 3.5
D	(3 + 2)nt-spacing 7(a)	2.6 ± 0.5	2.3 ± 0.5	4.8 ± 1.0	5.7 ± 0.7
	(3 + 2)nt-spacing 7(b)	8.2 ± 1.0	7.5 ± 0.2	13.5 ± 1.2	8.4 ± 1.2
	(3 + 2)nt-spacing 7(c)	13.8 ± 1.9	9.5 ± 0.8	14.0 ± 1.7	12.3 ± 2.3
E	(3 + 4)nt-spacing 7(a)	14.4 ± 2.4	15.8 ± 1.3	16.4 ± 0.7	17.6 ± 2.3
	(3 + 4)nt-spacing 7(b)	14.6 ± 1.8	9.6 ± 1.1	12.6 ± 1.6	17.8 ± 4.3

hmU or thymine contents of strands (top/bottom) are listed across the top of the Table. Loop-containing DNA is identified (at the left) by the number of nucleotides within each loop, followed by the number of base-pairs separating the loops (sequences in Figure 1). The symbol ≫ denotes that K_d is too large, and a too rapid dissociation of protein–DNA complexes occurs during electrophoresis to allow a reliable determination.

nd, not determined.

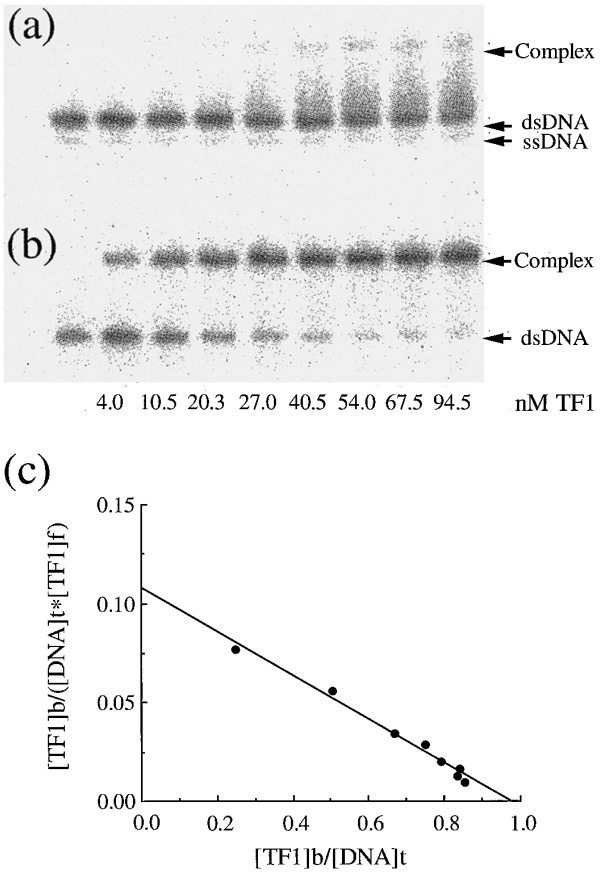

Figure 2. Enhanced binding of TF1 to thymine-containing DNA with loops. Electrophoretic analysis of: (a) perfect T/T duplex and (b) duplex with 6-nt loops at a spacing of 7 bp titrated with TF1. Protein concentrations indicated below (b) are identical for both panels. (c) Scatchard plot of the titration shown in (b).

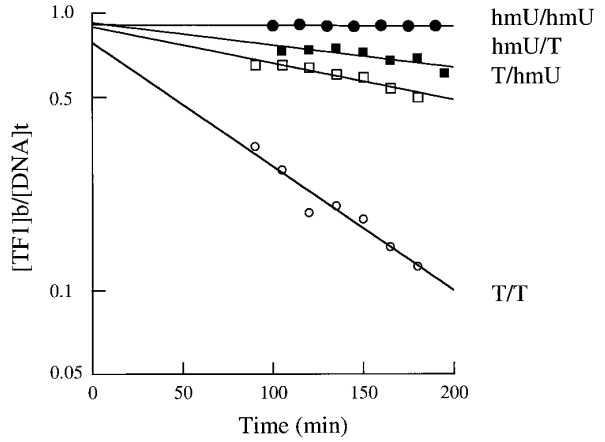

Figure 3. Rates of complex dissociation during electrophoresis. The fraction of complex remaining as a function of electrophoresis time is shown for 37-mer duplexes. The derived half-times, $t_{1/2}$ ($\pm$SEM), are 103($\pm$14) minutes for the complex with T/T duplex (○), 284($\pm$18) minutes for T/hmU duplex (□), and 485($\pm$132) minutes for hmU/T duplex (■); complexes with hmU/hmU duplex are stable (●). Protein concentrations were chosen to allow 100% complex formation at zero time of electrophoresis.

T/T duplex, the affinity is even lower and complexes rapidly dissociate during electrophoresis, as evidenced by the extensive smearing seen in Figure 2(a), precluding a reliable determination of K_d. Further, previous DNase I footprinting of T-containing DNA indicated that binding of TF1 is primarily non-specific (Sayre & Geiduschek, 1990). This loss of site-selectivity in T-containing DNA is substantiated by analyzing TF1 binding to DNA that does not contain sequence elements representing a preferred TF1 binding site; electrophoretic mobility shift experiments with different T-containing probes of equal length are indistinguishable (data not shown). Non-specific binding may still yield discrete bands of retarded DNA, provided that the mobility of complexes is independent of the position of protein relative to the DNA ends (a definite expectation for short DNA probes for which electrophoretic mobility would be primarily determined by the molecular mass of the complex rather than the shape of the bound DNA). Electrophoretic mobility shift experiments therefore provide information on only the aggregate affinity for the DNA probe.

It is also necessary to consider the consequences of complex dissociation during electrophoresis for measuring K_d. Dissociation of protein–DNA complexes during electrophoresis, evident for the T/T duplex in Figure 2(a), was quantified (Hoopes *et al.*, 1992). Figure 3 illustrates the fraction of complex remaining as a function of electrophoresis time for TF1 bound to 37-mer DNA. Data are fitted to a single exponential, yielding a first-order rate constant for dissociation in the gel. Extrapolation to zero time of electrophoresis indicates 100% complex formation in the sample that is loaded onto the gel, consistent with all the dissociation occurring during electrophoresis. For the hmU/hmU duplex, there is no evidence of complex dissociation during three hours of electrophoresis. For T/T DNA, the half-time of complex dissociation, $t_{1/2}$, is 103 minutes and intermediate stabilities are observed for complexes with hmU/T and T/hmU DNA ($t_{1/2}$ = 485 minutes and 284 minutes, respectively): the rate of dissociation during electrophoresis correlates with the equilibrium dissociation constant (Table 1A).

Equilibrium dissociation constants, K_d, are determined from the slopes of Scatchard plots. Correction for complex dissociation in the gel would require adjusting the fractional complex formation, $[TF1]_B/[DNA]_T$, by an exponential decay factor, $\exp(-k_{diss}*t)$, where t is the electrophoresis time and k_{diss} is the decay constant, but this would not change the slope of the Scatchard plot. Accordingly, no corrections were made.

Effects of model flexible sites on TF1 binding

Initial effort toward assessing the effect of model flexible loci on the affinity of TF1 was directed at positional preferences. Three consecutive mismatches generating 6-nt loops were placed at the center of the 37-mer (central to the core binding site) or symmetrically about the center with spacings of 7, 11, and 17 bp (Figure 1(b)). Mismatches were created by substituting the reversed sequence of the top strand into the bottom strand, generating mismatches of identical opposing bases. Oligonucleotides designed to generate these mismatches were synthesised as the T or hmU-containing versions.

A comparison of the binding of TF1 to the T/T duplex without loops with binding to T/T DNA containing two 6-nt loops with a spacing of 7 bp is shown in Figure 2(a) and (b). While the complex of TF1 with the T/T duplex dissociates during electrophoresis, as pointed out above, complexes involving the duplex with loops are clearly much more stable: their half-time of dissociation during electrophoresis of is ~2400 minutes, compared to 103 minutes for the perfect T/T duplex, and no discernible dissociation for the hmU/hmU duplex. The affinity increases correspondingly, to K_d = 7.1 nM for the two-loop duplex (Figure 2(c)).

Six-nucleotide loops with a spacing of 7 bp afford the largest increase in affinity compared to the corresponding duplex DNA (Table 1B). Whereas introduction of one central loop, or of two loops with spacings of 11 or 17 bp, does enhance binding of TF1 relative to that of complete T-containing duplex, the improvement is less pronounced. It was also noted that increased affinity for loop-containing DNA correlates with complex stability during electrophoresis (data not shown). Conspicuously, optimal positioning of loops is the same, irrespective of which strand contains thymine, although TF1 has a higher affinity for hmU/T DNA than for T/hmU DNA with equivalent loop placement (Table 1B). The striking increase in affinity of TF1 for T-containing DNA with two 6-nt loops separated by the optimal 7 bp indicates that introduction of localised flexibility enhances complex formation and confers an affinity approaching that of the hmU/hmU duplex.

In contrast, the introduction of loops into hmU/hmU DNA reduces affinity relative to completely duplex DNA. For the otherwise optimal loop spacing of 7 bp identified for T-containing DNA the effect is rather modest, a less than twofold reduction in affinity. It is conceivable that the diminished affinity of TF1 for hmU/hmU DNA with loops is a reflection of sequence-specific binding or that the introduction of mismatches induces structural changes that compromise complex formation. The different mobilities of hmU-containing DNA are consistent with the latter interpretation; whereas all T/hmU and T/T duplexes migrate similarly irrespective of loop content or position, hmU/hmU DNA with 6-nt loops has a reduced mobility. A marginal reduction in mobility is also discerned for hmU/T duplexes with 6-nt loops, particularly at a spacing of 11 bp (data not shown). Introduction of local flexibility would not be expected to reduce mobility significantly compared with that of canonical DNA (Kahn *et al*., 1994). In contrast, static bends result in significant retardation in polyacrylamide gels, particularly for oligomers longer than about 50 bp (Koo *et al*., 1986). Anomalous migration of DNA fragments has been observed, not only for curved molecules, but also for other structural variations. It is conceivable, therefore, that introduction of three consecutive mismatches into hmU or T-containing DNA gives rise to non-equivalent structural changes.

Three contiguous mismatches function as a site of dynamic DNA bending, as judged by an increased rate of DNA cyclization (Kahn *et al*., 1994). The preceding data demonstrate the merit of such constructs in assessing the effect of DNA flexibility on complex formation. On the other hand, protein–DNA complexes in which the DNA double helix is bent or kinked frequently localise these distortions to specific base-steps (e.g. see J. Kim *et al*., 1993; Y. Kim *et al*., 1993; Werner *et al*., 1995). Furthermore, it would be advantageous to minimize the loss of potential base-specific interactions in loop-containing DNA. For these reasons, 4-nt loops formed by tandem mismatches and separated by 7 to 11 bp were evaluated for their ability to mimic localised flexibility (Figure 1(c)). TF1 binds with highest affinity to DNA with two 4-nt loops separated by 9 bp, irrespective of thymine content (Table 1C). At spacings of 8 or 10 bp, the affinity is reduced about twofold and a further reduction in affinity is observed for DNA with 4-nt loops spaced apart by 7 or 11 bp. These results confirm the contribution of DNA bendability to complex formation. Remarkably, the affinity of TF1 for hmU/hmU duplex is fully matched by incorporating two tandem mismatches at a spacing of 9 bp into T-containing DNA, indicating that affinity is optimised when sites of DNA flexibility are appropriately and precisely positioned.

Design of the DNA probes described above generates a set of 4-nt loop constructs with different spacing (Figure 1). However, the base composition of loops also varies and differences in affinity could be a result of such variations; tandem mismatches of different sequence may confer different structural changes on otherwise duplex DNA. Additional T/T duplexes were therefore generated to maintain a loop spacing of 9 bp, while changing the sequence within 4-nt loops (Table 2). The affinity of TF1 for these constructs varies only modestly, with no more than twofold changes in the K_d. We also note that TF1 has a higher affinity for duplexes referred to as TA-9-TA and CT-9-TA (K_d = 4.5 nM and 2.3 nM; Table 2) than for duplexes with the same 4-nt loops but spacings of 7 and 8 bp (K_d = 30.6 nM and 7.5 nM; Table 1C). Although

Table 2. K_d for T/T DNA with sequence variations in 4-nt loops, spacing 9

Loop sequence	K_d (nM)
CT-9-AA	3.4 ± 0.5
TA-9-TA	4.5 ± 0.4
CT-9-TA	2.3 ± 0.3
CT-9-CT	4.0 ± 0.1
CT-9-CC	5.5 ± 0.2

The sequence of duplex regions is shown in Figure 1 (4nt-spacing 9). The sequence of each loop is indicated in the left-hand column, CT-9-AA corresponding to the sequence shown in Figure 1 and Table 1 (4nt-spacing 9). Within each loop, the sequence of the top and bottom strands is reversed.

sequence context is also likely to affect the structure of tandem mismatches (and in turn the affinity of TF1) these results support the contention that differences in affinity of TF1 for duplexes with loops at different spacings are primarily due to positional preference rather than sequence-dependent structural variations of loops.

It is remarkable also that only 4-nt loops spaced apart by 9 bp leave the affinity of TF1 for hmU/hmU DNA undiminished. For all other spacings, a distinct reduction in affinity is noted (Table 1C). Perhaps this is a consequence of unfavorable structural changes or the loss of base-specific contacts, as already stated. The inability to improve binding of TF1 to hmU-containing DNA by incorporation of tandem mismatches implies that considerable deformability is built into the hmU/hmU duplex.

The interaction of TF1 with a set of duplexes incorporating bulge-loops was also examined. These DNA probes were generated by either deleting or inserting one nucleotide in the bottom strand to create (3 + 2)-nt and (3 + 4)-nt constructs with spacings of 7 bp (Figure 1(d) and (e)). As summarised in Table 1(D and E), the affinity of TF1 for these constructs spans the range of 2 to 18 nM, essentially independent of thymine content. The highest affinity is for the construct named (3 + 2)nt-spacing 7(a), which allows a sequence alignment generating a central 9 bp region bordered by (2 + 1)-nt loops, consistent with the preferred spacing of 4-nt loops (Figure 1(c)).

To summarise, we find that the presence of 6-nt loops in T-containing DNA, structures previously demonstrated to mimic local DNA flexibility, allows TF1 to bind with greatly increased affinity. Four-nucleotide loops appear to have similar effects on protein binding, consistent with both 6-nt and 4-nt loops functioning as model flexible loci. The presence of 4-nt loops at an optimal spacing of 9 bp generates two substantial effects: an increase by orders of magnitude in the affinity of TF1 for T/T DNA and a restoration of the affinity of TF1 for hmU/hmU DNA to the level of the perfect duplex. We propose that hmU-containing DNA possesses an inherent flexibility that is emulated by tandem mismatches, and that this increased bendability underlies the preferred binding by TF1 relative to

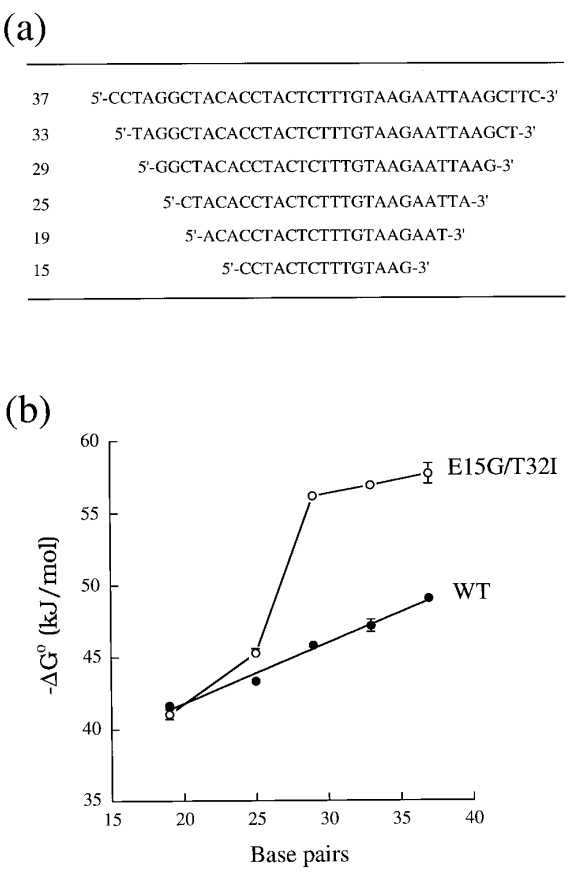

Figure 4. Size of the TF1 binding site. (a) Sequences of oligonucleotides of decreasing length, generated by symmetrically shortening the 37-mer at both ends (top strand is shown). Only hmU-containing DNA was analyzed (i.e. T stands for hmU). (b) Affinity as a function of DNA length. The standard free energy of association, $-\Delta G°$, determined from Scatchard analyses, is plotted as a function of the number of base-pairs for wild-type TF1 (●); $\Delta\Delta G = -0.42$ kJ mol^{-1} bp^{-1}, and for the mutant TF1 (E15G/T32I) (○).

T-containing DNA. TF1 also exhibits site-selectivity in its binding to hmU-containing DNA. Whether this is a result of direct readout of sequence through specific contacts with DNA bases or a manifestation of sequence-dependent local variations in structure or flexibility has not yet been unequivocally established. However, the explicit requirement for loop position in optimising complex formation emphasises the role of sequence-dependent, localised DNA flexibility in identifying preferred TF1 binding sites.

Size of the TF1 binding site

The large size of the TF1 footprint and pronounced DNA bending by TF1 suggest that DNA wraps around the body of the protein (Schneider *et al*., 1991). The experiments described above implicate two sites of enhanced DNA flexibility bordering a 9 bp duplex region in DNA wrapping, and suggest that DNA extending beyond this central segment contributes to binding

Table 3. Affinity of TF1(E15G/T32I) for T/T DNA with loops

	K_d (nM)
No loop	10.7 ± 2.6
4nt-spacing 7	0.86 ± 0.13
4nt-spacing 8	0.20 ± 0.03
4nt-spacing 9	0.10 ± 0.02
4nt-spacing 10	0.52 ± 0.10
4nt-spacing 11	1.66 ± 0.45

The sequences of DNA probes are shown in Figure 1.

Table 4. K_d (nM) for hmU/hmU DNA of decreasing length

		Duplex with overhangs	
n	Complex duplex	37/*n*	29/*n*
37	2.6 ± 0.3	n/a	n/a
33	5.7 ± 1.0	3.3 ± 0.3	n/a
29	9.5 ± 0.9	3.9 ± 0.4	n/a
25	26.3 ± 3.2	5.9 ± 0.9	16.6 ± 2.6
19	50.1 ± 2.3	¶	33.5 ± 2.2
15	n/c	¶	≫

The far left column indicates number of base-pairs (*n*) for duplexes of decreasing length (sequences in Figure 4(a)). For duplexes with overhangs, the 37-mer or 29-mer top strand is annealed to a bottom strand of decreasing length resulting in a central duplex region (*n* base-pairs). The symbol ≫ indicates that K_d is too large (and correspondingly too rapid dissociation of complexes during electrophoresis) to allow determination; n/a, not applicable; n/c, no complex; ¶, self-complementary sequence in single-stranded regions generates multimeric structures that render a determination of K_d unreliable.

interactions. Binding of TF1 to hmU-containing duplexes of decreasing length was therefore evaluated. Complementary oligonucleotides were generated by symmetrically shortening the original 37-mer duplex at both ends (Figure 4(a)). Figure 4(b) shows the standard free energy of binding, $\Delta G°$, as a function of DNA length. From an initial value of –49 kJ mol^{-1} for the 37-mer duplex, $-\Delta G°$ decreases linearly as duplex length is reduced to 19 bp; $\Delta\Delta G = -0.42$ kJ mol^{-1} bp^{-1}. Maximal binding interactions are assumed for the 37-mer: K_d for 44-mer hmU-containing DNA containing the same TF1 binding site was previously determined to be 2.5 nM (Andera *et al*., 1994). No complex formation could be detected for the 15-mer. Although this may genuinely reflect lack of complex formation in solution, it is also likely to have resulted from too-rapid complex dissociation during electrophoresis. Significant dissociation was observed for both the 25-mer and, in particular, the 19-mer for which complexes were barely discernible following electrophoresis (data not shown). Extrapolating the dissociation rate in the gel to complexes involving a 15-mer duplex (anticipated $t_{1/2}$ less than ten minutes) leads us to expect that, were any complex to form, it would dissociate during electrophoresis. Nonetheless, these results establish the existence of protein–DNA interactions beyond a central core region of the DNA, while the relatively modest value of $\Delta\Delta G$/bp indicates that significant interactions are provided within the central region of the binding site.

This analysis was extended to the TF1 double mutant protein, TF1(E15G/T32I), which binds DNA more tightly and is more thermostable, although its amino acid substitutions are in the core of the protein rather than in regions thought to engage DNA directly. K_d for an all-hmU-containing 44-mer duplex is 40-fold higher for TF1(E15G/T32I) than for wild-type TF1 (Andera *et al*., 1994). As shown in Figure 4(b), the standard free energy of binding of the mutant protein does not significantly differ from that of wild-type TF1 when comparing the shorter DNA probes (including the absence of complexes with 15-mer DNA). However, a steep rise in $-\Delta G°$ accompanies an increase in duplex length from 25 to 29 bp. Evidently, interactions outside a central 25 bp duplex region are responsible for the increased affinity of TF1(E15G/T32I). We also find that 4-nt loops in all-T-DNA have similar effects on affinity of the mutant protein and of wild-type TF1 (Table 3). The affinity for 37-mer perfect hmU/hmU duplex ($K_d = 8.6 \times 10^{-11}$ M) is essentially matched in T/T DNA by 4-nt loops separated by 9 bp ($K_d = 10.2 \times 10^{-11}$ M). The affinity for perfect T/T duplex (K_d = 10.7 nM) and for other loop separations is significantly lower.

Single-stranded ends partially substitute for extensions of duplex DNA in interactions with TF1 (Table 4). Annealing the 37-mer top strand to bottom strands of decreasing length results in improved binding compared to the shorter duplex. For the 25-mer duplex, the K_d of 26 nM represents a tenfold reduction in affinity compared to 37-mer duplex; however, annealing the 25-mer bottom strand to the 37-mer top strand yields a K_d of 5.9 nM, only a twofold reduction compared to completely duplexed 37-mer. A modest increase in affinity is achieved when annealing the 25-mer bottom strand to a 29-mer top strand (K_d = 16.6 nM; Table 4). Apparently, required interactions beyond the central core sequence are largely satisfied by single-stranded DNA.

Discussion

For complexes that require DNA to wrap around the body of a protein, the sequence-dependent flexibility of DNA must contribute to the specificity of interaction. Since DNA does not behave as an isotropic rod, bending more readily in one plane than another (Calladine & Drew, 1986; Wolffe & Drew, 1996), anisotropic bendability must also be a determinant of target site selection by DNA-binding and -bending proteins. Predicting sequence-dependent DNA flexure remains elusive, however, largely due to the considerable influence of sequence context on the local structure of individual base-steps, which is a critical determinant of DNA flexibility (Quintana *et al*., 1992).

The expectation that sequence-dependent DNA bendability contributes to target site selection by TF1, and the possibility that differences in affinity

for hmU or T-containing DNA of otherwise similar sequence might be at least partly due to differential flexibility, motivated these studies. Two prior findings appear to counterindicate such expectations: (1) the structure of an A·hmU base-pair (flanked by G·C base-pairs) as determined by NMR spectroscopy reveals no significant deviations from classical *B*-form DNA (Mellac *et al*., 1993); (2) the torsional rigidity of hmU-containing SPO1 DNA determined by time-resolved and steady-state fluorescence polarization anisotropy of intercalated ethidium is the same as that separately reported for T-containing DNA (Härd & Kearns, 1990; Millar *et al*., 1982). However, the latter experiments measure long-range cooperative motions rather than the existence of localised points of flexure (Millar *et al*., 1982). Moreover, even multidimensional NMR techniques are inadequate for detecting subtle changes in local structural variables. One difficulty is the lower density of protons in nucleic acids; as most protons are on sugar rings, bases and phosphates are less well defined (unless one resorts to ^{31}P NMR; Gorenstein, 1994). Our findings, therefore, are not inconsistent with prior work. Our inferences about increased flexure of hmU-containing DNA, which are based on a series of correlations, strongly suggest the need for structural analysis and detailed calculations of the conformational energetics of hmU-DNA, and motivate direct analysis of the internal motions by ^{13}C or ^{31}P NMR relaxation experiments (Borer *et al*., 1994; V. Hsu, personal communication).

The reported structures of duplexes containing single-base loops or mismatches indicate a propensity for stacking the extra base within the helix, as well as compensatory torsions of the glycosidic bond that only slightly disrupt the overall conformation of the duplex (Gao & Patel, 1988). Despite being conformationally indistinguishable, however, DNA with such lesions is thermally destabilised (Plum & Breslauer, 1994) and may exhibit substantially different energetics of deformation. Even if loop-regions were partly single-stranded, this would not be anticipated to increase the affinity of TF1, which has significantly lower affinity for single-stranded than duplex DNA (cf. the lower affinity for duplex with overhangs, Table 4, and lower affinity for constructs with 6-nt loops compared to 4-nt loops, Table 1). We therefore suggest that increased binding of TF1 to loop-containing duplexes is due to recognition based on ease of deformation and not due to an improved fit to a statically deformed DNA target or to nucleation around potentially single-stranded regions.

The requirement for specific loop-positioning in securing optimal binding interactions points to DNA flexibility as an integral part of complex formation and implies that TF1 must recognise sequence-dependent variations in DNA flexibility. The observation that TF1 binding to hmU-DNA is not enhanced by incorporation of loops may be due to substantial inherent flexibility that diminishes the effect of loop-imposed local bendability, losses of base-specific contacts in the looped DNA constructs, or the energetic cost of restraining excessive flexibility in forming a close-packed complex with loop-containing hmU-DNA. Thus, we suggest that inherent flexibility of hmU-containing DNA underlies the differential binding preference of TF1 and further suggest that 4-nt loops in T-DNA may mimic sequence-dependent conformation dynamics of hmU-DNA. In the accompanying Communication (Grove *et al*., 1996), we show that other DNA-bending proteins, including HU and HMG1, also bind preferentially to hmU-containing DNA, and that IHF binds to hmU-DNA lacking a specific IHF binding site more tightly than to the corresponding T-DNA. These findings are consistent with our conclusions regarding the inherent flexibility of hmU-DNA, as is the evidence that HU bends hmU-containing DNA more than T-DNA of identical sequence (Schneider *et al*., 1991). The preferred binding of HU protein to DNA whose flexibility is greater than average (or DNA that is pre-bent) may also be the basis for its reported preference for DNA with single-strand nicks or gaps (Castaing *et al*., 1995) and for cruciform structures (Pontiggia *et al*., 1993; Bonnefoy *et al*., 1994; see also Grove *et al*., 1996, accompanying Communication).

Discrimination between hmU and T-containing DNA by phage-encoded regulatory proteins is important during multiplication of hmU phages (Lee *et al*., 1980). The phage SPO1 middle promoters are selectively recognised by the host RNA polymerase core bearing the phage-encoded σ family protein, gp28 (reviewed by Losick & Pero, 1981). Although correctly recognised by the RNA polymerase and initiating at the proper sites, middle promoters are extremely weak in T-containing DNA and at intermediate strengths in T/hmU hybrid DNA (Choy *et al*., 1986; Romeo *et al*., 1986). The observations that are reported here suggest that hmU selectivity at SPO1 middle promoters could be at least partly due to the relative ease of DNA bending that must occur within the 70 bp RNA polymerase-binding site rather than to direct recognition of hmU in the –10 and –35 promoter elements by SPO1 gp28. If the gp28 σ-family protein is relatively deficient in the steps of promoter complex isomerisation or promoter opening, it might require help at these steps of transcriptional initiation (see Ryu *et al*., 1994; Jordi *et al*., 1995) from the more pliant hmU-DNA. Why, in contrast, does discrimination between C-DNA and hmC-DNA not feature in transcriptional regulation of T-even phages, whose genomes contain hydroxymethylcytosine (hmC) in place of cytosine (Brody *et al*., 1995)? We suggest that this dichotomy reflects mechanical differences in hmU-DNA and T-DNA that are exploited in the design and mechanism of action of regulatory proteins, and that C and hmC offer no comparable opportunities.

In view of the significant role of sequence-dependent DNA deformability in TF1-binding, it is not obvious whether direct base-contacts contribute to the selection of preferred binding sites in hmU-DNA. Interactions with the phosphodiester backbone are likely to make a substantial contribution to site-selectivity (reviewed by Pabo & Sauer, 1992). However, two arguments, the first specific and the second more general, can be offered in favor of direct sequence recognition. (1) TF1 does bring a capacity for selecting preferred binding sites in hmU-DNA that HU lacks (Sayre & Geiduschek, 1990). (2) Any structure that brings a more or less fixed array of amino acid side-chains into close proximity with DNA base-pairs creates a potential for direct contacts with a preferred DNA sequence; TF1 does bring its arm and tail into close proximity with DNA (Härd *et al*., 1989a,b; Andera & Geiduschek, 1994).

Although TF1 derives significant binding energy from interactions with the central portion of the target site, distal DNA regions contribute to overall binding. In the interaction of another DNA-bending protein, CAP, with a 44-mer duplex (Liu-Johnson *et al*., 1986), contacts with DNA 12 to 14 bp from the dyad axis (i.e. two 3-bp patches at the edges of a 28 bp core) contribute strongly to the affinity, with $\Delta\Delta G = -2.1$ to -5.4 kJ mol^{-1} bp^{-1}. More distally placed DNA makes only small and varying contributions to affinity. In comparison, the affinity of TF1 increases linearly with duplex length between 19 and 37 bp ($\Delta\Delta G = -0.42$ kJ mol^{-1} bp^{-1}; Figure 4); thus, most of the binding energy originates from contacts with a much shorter DNA segment. The increased affinity of TF1(E15G/T32I) is evidently contributed by localised interactions on the lateral surfaces of the protein with sites on the DNA that are also (perhaps by pure coincidence) located 13 or 14 bp from the center of the binding site. However, the interactions of wild-type TF1 and of TF1(E15G/T32I) with the central DNA segment, including presumptive sites of DNA bending (Figure 4 and Table 3), appear identical.

The preferred position of flexible sites in T-DNA, 4-nt loops separated by 9 bp of duplex, suggests the probable locations of DNA bends in the complex of TF1 with hmU-DNA. The central 9 bp of our reference binding site are bounded by hmU-A steps (or T-A steps as the case may be). The structures of T-A steps in DNA oligomers are especially diverse, and it has been argued that this is in part due to poor stacking and overlap of the pyrimidine-purine step (Quintana *et al*., 1992; Goodsell *et al*., 1994). T-A steps also have a general propensity for a bend that tends to compress the major groove. Diversity of structure and tendency to bend should go hand in hand with mobility of conformation; that may rationalise the frequent occurrence of T-A steps in promoters and origins of replication, and their involvement in protein-induced DNA bending. Whether hmU-A steps share the unique properties of T-A steps remains to be determined. However, twin hmU-A steps, located exactly at the optimal position of 4-nt loops within the TF1 binding sequence (asterisks in Figure 1(a)) make likely targets for specific distortion by TF1.

Materials and Methods

Protein purification

TF1 was overexpressed in *E. coli* DL39 transformed with a plasmid harboring the TF1 gene (Härd *et al*., 1989a). TF1(E15G/T32I) was generated as described (Andera *et al*., 1994) and overexpressed in *E. coli* BL21. TF1-bearing cells were stored at −70°C and all the steps of TF1 purification were carried out at 0 to 4°C by a method that deviates in detail from previously published procedures (Härd *et al*., 1989a; Sayre & Geiduschek, 1990). Cells were incubated in lysis buffer containing 50 mM Tris-HCl (pH 8.0), 0.25 M NaCl, 5 mM Na_2EDTA, 5% (v/v) glycerol, 5 mM 2-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.5 mg/ml lysozyme for one hour and lysed by addition of 10% Triton X-100 to 0.5% (v/v). Polymin P (BASF) was added dropwise from a 13% (v/v) solution to a final concentration of 0.5%. Precipitated protein and nucleic acids were removed and proteins in the supernatant were fractionated by ammonium sulfate precipitation (Härd *et al*., 1989a); the supernatant was slowly adjusted to 50% saturation with $(NH_4)_2SO_4$ (2.37 M at 24°C), stirred for 15 minutes, centrifuged at, 15,000 ***g*** and the sediment discarded. Solid $(NH_4)_2SO_4$ was added to 75% saturation and the precipitate forming during 30 minutes of stirring was collected, dissolved in buffer A (20 mM Tris-HCl (pH 7.5), 50 mM KCl, 5% (v/v) glycerol, 1 mM EDTA, 0.2 mM PMSF, 3.5 mM 2-mercaptoethanol), dialyzed against buffer A for at least two hours, and applied to DEAE-cellulose and heparin-Sepharose columns attached in tandem and equilibrated in buffer A. The protein was eluted from the heparin-Sepharose column with a linear gradient of 50 mM to 1 M KCl in buffer A. Peak fractions were adjusted to 50% saturation with $(NH_4)_2SO_4$ while stirring and applied to a phenyl-Sepharose column equilibrated in buffer A containing 50% saturated $(NH_4)_2SO_4$. TF1 was eluted with a linear gradient of 50% to 0% saturated $(NH_4)_2SO_4$ and dialyzed extensively against a buffer composed of 50 mM sodium phosphate (pH 6.8), 0.25 M NaCl and 10 μM EDTA. Purity was ascertained by Coomassie blue staining of SDS-polyacrylamide gels. Protein concentration was determined from the absorbance at 280 nm, using an extinction coefficient of 1200 M^{-1} cm^{-1} for the monomer, which contains a single tyrosine, and correcting for contributions of light scattering by subtracting the absorbance at 330 nm.

Isolation and labeling of DNA probes

Oligonucleotides with hmU content were synthesised as described (Conte *et al*., 1992) and purified by HPLC. T-containing oligonucleotides were purchased and purified by denaturing polyacrylamide gel electrophoresis. The top strand, shared among loop-containing duplexes (Figure 1), was ^{32}P-labeled at the 5'-end with bacteriophage T4 polynucleotide kinase. Equimolar mixtures of complementary oligonucleotides were heated to 90°C and slowly cooled to 4°C over several hours to form duplex DNA.

Gel electrophoresis and quantification of protein–DNA complexes

Polyacrylamide gels for electrophoretic mobility shift experiments were 10% (w/v) polyacrylamide (39:1 acrylamide:bis-acrylamide) in TBE (45 mM Tris-borate (pH 8.0), 1 mM EDTA) with 10 mM NaCl. Experiments were carried out at 4°C and gels were prerun in TBE with 10 mM NaCl for one hour at 20 mA before loading samples with the power on. Reaction conditions were as described (Andera *et al*., 1994) and each sample contained 10 fmol 5'-end-labeled DNA in a reaction volume of 10 µl. Dilutions of protein (made immediately prior to use) and binding reactions required the use of siliconised microfuge tubes to minimise protein adhering to surfaces. After electrophoresis, gels were dried and protein–DNA complexes were quantified using a phosphorimager and software supplied by the manufacturer (MacBAS 2.0). The region on the gels between bands corresponding to free and complexed DNA was considered as free DNA.

Equilibrium dissociation constants, K_d, were determined from the slope of Scatchard plots of DNA titrated with TF1. Assuming binding of TF1 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. When using purified protein of known concentration, this method of determining K_d does not require correction for dissociation of protein–DNA complexes during electrophoresis.

For measurements of the rate of dissociation of protein–DNA complexes, samples containing sufficient TF1 to allow complete complex formation were loaded at 15 minute intervals onto polyacrylamide gels with the power on. The fraction of complex remaining ($[TF1]_B / [DNA]_T$) was determined for electrophoresis times of 60 to 180 minutes and used to calculate the first-order rate constant, k_{diss}, for dissociation during electrophoresis (Hoopes *et al*., 1992).

All experiments were carried out at least three times. Values of K_d are reported as the average ± the standard error of the mean.

Acknowledgements

We thank V. L. Hsu for helpful discussion, and gratefully acknowledge support of this research at UCSD by a grant from the NIGMS, and at the University of Naples by the MURST and CNR (Italy).

References

Andera, L. & Geiduschek, E. P. (1994). Determinants of affinity and mode of DNA binding at the carboxy terminus of the bacteriophage SPO1-encoded type II DNA-binding protein, TF1. *J. Bacteriol.* **176**, 1364–1373.

Andera, L., Spangler, C. J., Galeone, A., Mayol, L. & Geiduschek, E. P. (1994). Interrelations of secondary structure stability and DNA-binding affinity in the bacteriophage SPO1-encoded type II DNA-binding protein TF1. *J. Mol. Biol.* **236**, 139–150.

Bonnefoy, E., Takahashi, M. & Rouvière-Yaniv, J. (1994). DNA-binding parameters of the HU protein of *Escherichia coli* to cruciform DNA. *J. Mol. Biol.* **242**, 116–129.

Borer, P. N., LaPlante, S. R., Kumar, A., Zanatta, N., Martin, A., Hakkinen, A. & Levy, G. C. (1994). ^{13}C-NMR relaxation in three DNA oligonucleotide duplexes: model-free analysis of internal and overall motion. *Biochemistry*, **33**, 2441–2450.

Brody, E. M., Kassavetis, G. A., Ouhammouch, M., Sanders, G. M., Tinker, R. L. & Geiduschek, E. P. (1995). Old phage, new insights: two recently recognised mechanisms of transcriptional regulation in bacteriophage T4 development. *FEMS Microbiol. Letters*, **128**, 1–8.

Calladine, C. R. & Drew, H. R. (1986). Principles of sequence-dependent flexure of DNA. *J. Mol. Biol.* **192**, 907–918.

Castaing, B., Zelwer, C., Laval, J. & Boiteux, S. (1995). HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J. Biol. Chem.* **270**, 10291–10296.

Choy, H. A., Romeo, J. M. & Geiduschek, E. P. (1986). Activity of a phage-modified RNA polymerase at hybrid promoters: effects of substituting thymine for hydroxymethyluracil in a phage SPO1 middle promoter. *J. Mol. Biol.* **191**, 59–73.

Conte, M. R., Galeone, A., Avizonis, D., Hsu, V. L., Mayol, L. & Kearns, D. R. (1992). Solid phase synthesis of 5-hydroxymethyluracil containing DNA. *Bioorg. Med. Chem. Letters*, **2**, 79–82.

Drlica, K. & Rouvière-Yaniv, J. (1987). Histonelike proteins of bacteria. *Microbiol. Rev.* **51**, 301–319.

Gao, X. & Patel, D. J. (1988). G(syn)·A(anti) mismatch formation in DNA dodecamers at acidic pH: pH-dependent conformational transition of G·A mispairs detected by proton NMR. *J. Am. Chem. Soc.* **110**, 5178–5182.

Goodsell, D. S., Kaczor-Grzeskoviak, M. & Dickerson, R. E. (1994). The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of *B*-DNA at T-A steps. *J. Mol. Biol.* **239**, 79–96.

Gorenstein, D. G. (1994). Conformation and dynamics of DNA and protein-DNA complexes by ^{31}P NMR. *Chem. Rev.* **94**, 1315–1338.

Granston, A. E. & Nash, H. A. (1993). Characterization of a set of integration host factor mutants deficient for DNA binding. *J. Mol. Biol.* **234**, 45–59.

Greene, J. R. & Geiduschek, E. P. (1985). Site-specific DNA binding by the bacteriophage SPO1-encoded type II DNA-binding protein. *EMBO J.* **4**, 1345–1349.

Grove, A., Galeone, A., Mayol, L. & Geiduschek, E. P. (1996). Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *J. Mol. Biol.* **260**, 120–125.

Gualerzi, C. O. & Pon, C. L. (1986). Editors of *Bacterial Chromatin*, Springer-Verlag, Berlin.

Härd, T. & Kearns, D. R. (1990). Reduced DNA flexibility in complexes with a type II DNA binding protein. *Biochemistry*, **29**, 959–965.

Härd, T., Hsu, V., Geiduschek, E. P., Appelt, K. & Kearns, D. (1989a). Fluorescence studies of a single tyrosine in a type II DNA binding protein. *Biochemistry*, **28**, 396–406.

Härd, T., Sayre, M. H., Geiduschek, E. P. & Kearns, D. R. (1989b). A type II DNA-binding protein genetically engineered for fluorescence spectroscopy: the ''arm'' of Transcription Factor 1 binds in the DNA grooves. *Biochemistry*, **28**, 2813–2819.

Hoopes, B. C., LeBlanc, J. F. & Hawley, D. K. (1992). Kinetic analysis of yeast TFIID-TATA box complex

formation sugests a multi-step pathway. *J. Biol. Chem.* **267**, 11539–11547.
Jia, X., Reisman, J. M., Hsu, V. L., Geiduschek, E. P., Parello, J. & Kearns, D. R. (1994). Proton and nitrogen NMR sequence-specific assignments and secondary structure determination of the *Bacillus subtilis* SPO1-encoded transcription factor 1. *Biochemistry*, **33**, 8842–8852.
Johnson, G. G. & Geiduschek, E. P. (1977). Specificity of the weak binding between the phage SPO1 transcription-inhibitory protein, TF1, and SPO1 DNA. *Biochemistry*, **16**, 1473–1485.
Jordi, B. J. A. M., Owen-Hughes, T., Hulton, C. S. J. & Higgins, C. F. (1995). DNA twist, flexibility and transcription of the osmoregulated *proU* promoter of *Salmonella typhimurium*. *EMBO J.* **14**, 5690–5700.
Kahn, J. D., Yun, E. & Crothers, D. M. (1994). Detection of localized DNA flexibility. *Nature*, **368**, 163–166.
Kellenberger, E. & Arnold-Schulz-Gahmen, B. (1992). Chromatins of low-protein content: special features of their compaction and condensation. *FEMS Microbiol. Letters*, **100**, 361–370.
Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
Koo, H. S., Wu, H. M. & Crothers, D. M. (1986). DNA bending at adenine·thymine tracts. *Nature*, **320**, 501–506.
Lee, G., Hannett, N. M., Korman, A. & Pero, J. (1980). Transcription of cloned DNA from *Bacillus subtilis* phage SPO1. Requirement for hydroxymethyluracil-containing DNA by phage-modified RNA polymerase. *J. Mol. Biol.* **139**, 407–422.
Liu-Johnson, H.-N., Gartenberg, M. R. & Crothers, D. M. (1986). The DNA binding domain and bending angle of *E. coli* CAP protein. *Cell*, **47**, 995–1005.
Losick, R. & Pero, J. (1981). Cascades of sigma factors. *Cell*, **25**, 581–584.
Mellac, S., Fazakerley, G. V. & Sowers, L. C. (1993). Structures of base pairs with 5-(hydroxymethyl)-2'-deoxyuridine in DNA determined by NMR spectroscopy. *Biochemistry*, **32**, 7779–7786.
Mengeritsky, G., Goldenberg, D., Mendelson, I., Giladi, H. & Oppenheim, A. B. (1993). Genetic and biochemical analysis of the integration host factor of *Escherichia coli*. *J. Mol. Biol.* **231**, 646–657.
Millar, D. P., Robbins, R. J. & Zewail, A. H. (1982). Torsion and bending of nucleic acids studied by subnanosecond time-resolved fluorescence depolariztion of intercalated dyes. *J. Chem. Phys.* **76**, 2080–2094.
Nash, H. A. (1996). The *E. coli* HU and IHF proteins: accessory factors for complex protein-DNA assemblies. In *Regulation of Gene Expression in* Escherichia coli (Lin, E. C. C. & Lynch, A. S., eds), in the press.
Pabo, C. O. & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.
Plum, G. E. & Breslauer, K. J. (1994). DNA lesions. A thermodynamic perspective. *Ann. N.Y. Acad. Sci.* **726**, 45–57.
Pontiggia, A., Negri, A., Beltrame, M. & Blanchi, M. E. (1993). Protein HU binds specifically to kinked DNA. *Mol. Microbiol.* **7(3)**, 343–350.
Quintana, J. R., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. (1992). Structure of a *B*-DNA decamer with a central T-A step: C-G-A-T-T-A-A-T-C-G. *J. Mol. Biol.* **225**, 379–395.
Reisman, J. M., Hsu, V. L., Jariel-Encontre, I., Lecou, C., Sayre, M. H., Kearns, D. R. & Parello, J. (1993). A ^{1}H-NMR study of the transcription factor 1 from *Bacillus subtilis* phage SPO1 by selective ^{2}H-labeling. *Eur. J. Biochem.* **213**, 865–873.
Romeo, J. M., Greene, J. R., Richards, S. H. & Geiduschek, E. P. (1986). The phage SPO1-specific RNA polymerase, E.gp28, recognizes its cognate promoters in thymine-containing DNA. *Virology*, **153**, 46–52.
Ryu, S., Garges, S. & Adhya, S. (1994). An arcane role of DNA in transcription activation. *Proc. Natl Acad. Sci. USA*, **91**, 8582–8586.
Satchwell, S. C. & Travers, A. A. (1989). Asymmetry and polarity of nucleosomes in chicken erythrocyte chromatin. *EMBO J.* **8**, 229–238.
Sayre, M. H. & Geiduschek, E. P. (1988). TF1, the bacteriophage SPO1-encoded type II DNA-binding protein, is essential for viral multiplication. *J. Virol.* **62**, 3455–3462.
Sayre, M. H. & Geiduschek, E. P. (1990). Effects of mutations at amino acid 61 in the arm of TF1 on its DNA-binding properties. *J. Mol. Biol.* **216**, 819–833.
Schneider, G. J., Sayre, M. H. & Geiduschek, E. P. (1991) DNA-bending properties of TF1. *J. Mol. Biol.* **221**, 777–794.
Shrader, T. E. & Crothers, D. M. (1990). Effects of DNA sequence and histone-histone interactions on nucleosome placement. *J. Mol. Biol.* **216**, 69–84.
Tanaka, I., Appelt, K., Dijk, J., White, S. W. & Wilson, K. S. (1984). 3-Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature*, **310**, 376–381.
Travers, A. A. & Klug, A. (1990). Bending of DNA in nucleoprotein complexes. In *DNA Topology and its Biological Effects* (Cozzarelli, N. R. & Wang, J. C., eds), pp. 57–106, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J. & Gronenborn, A. M. (1995). The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell*, **83**, 761–771.
White, S. W., Appelt, K., Wilson, K. S. & Tanaka, I. (1989). A protein structural motif that bends DNA. *Proteins: Struct. Funct. Genet.* **5**, 281–288.
Wolffe, A. P. & Drew, H. R. (1996). DNA structure implications for chromatin structure and function. In *Frontiers in Molecular Biology*, IRL Press, in the press.

Edited by R. Schleif

(Received 29 January 1996; received in revised form 3 April 1996; accepted 15 April 1996)