The Specific Binding of *Escherichia coli* Integration Host Factor Involves both Major and Minor Grooves of DNA†

Siqun Wang,‡ Richard Cosstick,§ Jeffrey F. Gardner,‡ and Richard I. Gumport*,‖,⊥

Departments of Microbiology and Biochemistry and College of Medicine, University of Illinois, Urbana, Illinois 61801, and Department of Chemistry, University of Liverpool, Liverpool L69 3BX, United Kingdom

Received May 1, 1995; Revised Manuscript Received August 8, 1995⊗

ABSTRACT: The integration host factor (IHF) of *Escherichia coli* is a small, sequence-specific DNA-binding protein. The specific and nonspecific binding constants of IHF were estimated by gel-retardation assays. The equilibrium association constant of IHF for the H′ site in λ *attP* is 6.8×10^8 M^{-1} (K_d = 1.5 nM), and the nonspecific binding constant is 5.8×10^5 M^{-1} (K_d = 1.7 μM), giving a selectivity of approximately 1000-fold for a specific site over random sequences. To study the molecular determinants specifying IHF binding, we used a series of 41 oligonucleotides containing adenine analogues that modified the surfaces of the major and minor grooves of the DNA. Many of the analogue substitutions within the previously defined consensus region caused decreased binding. Replacement with various analogues outside the consensus domain had little effect. Quantifying the binding constants for those sites with reduced affinities indicated an interaction with the minor groove within the consensus sequence. The binding constants of sites with 2-aminopurine and an inosine substitution within the same region suggest that IHF may also interact with the major groove. Thus, the specific interaction of IHF with its H′ site likely involves interactions with both the minor and major grooves of the DNA.

Proteins that interact with discrete sequences in DNA show recognition specificities ranging from stringent to relaxed. Those recognizing particular sequences often have high DNA-binding specificities, whereas the ones recognizing multiple sequences often display lower specificities. Restriction endonucleases and methyltransferases that must select their target sequences from the millions of potential nonspecific sites on the genome exemplify the former class. Other DNA-binding proteins like structure-specific nucleases and histones represent the latter class. Integration host factor (IHF)[1] is in this latter class and recognizes sequences that contain both conserved and nonconserved elements. The core consensus for the conserved element of IHF-binding sites comprises a degenerate family of sequences, WATCARNNNNTTR (Friedman, 1988). A less-conserved, dA+dT-rich element, which is upstream on the 5′ side of the conserved element, is also important for IHF binding (Goodrich et al., 1990; Hales et al., 1994a,b; Lee et al., 1991).

IHF is a small (M_r = 20 000) DNA-binding protein composed of two related but nonidentical subunits—α, encoded by *himA*, and β, encoded by *hip/himD*. IHF participates in many cellular processes, *e.g.*, site-specific recombination, transposition and inversion, phage chromosome packaging, DNA replication, and gene expression (Freundlich et al., 1992; Friedman, 1988). These roles appear to be effected by the ability of IHF to bend DNA upon binding (Goodman & Nash, 1989). IHF is a member of the bacterial histone-like protein family that includes HU, TF1, and H-NS (Drlica & Rouvier-Yaniv, 1987). Sequence similarities exist among the polypeptides of these DNA-binding proteins, suggesting that they share similar overall structures, although their binding specificities vary greatly. For example, IHF shows sequence-specific recognition while HU displays little specificity.

Footprinting studies *in vitro* suggest that IHF interacts with the minor groove of DNA (Craig & Nash, 1984; Yang & Nash, 1989). This evidence and similarities between IHF and HU led Yang and Nash (1989) to propose a model of DNA−IHF interaction based on the structural model of HU obtained from X-ray diffraction analysis (Tanaka et al., 1984). In the Yang−Nash model, the two subunits are intertwined to form a body from which two arms or loops, each composed of anti-parallel β-sheets, extend from the body and wrap around the DNA into the minor groove at two sites. The bound DNA bends around IHF by about 140°, allowing extensive contacts at the interface between the protein and DNA. A genetic study of IHF-DNA binding led to the isolation of IHF mutants with altered binding specificities (Lee et al., 1992). The amino acid changes were in the loop of the α subunit and body of the β subunit and enabled IHF to recognize binding sites with changes in the WATCAR and TTR regions of the consensus sequence, respectively. These results oriented the heterodimer with respect to the site and are consistent with the structural model of Yang and Nash in that the mutated bases and the corresponding amino acid changes in IHF are in proximity. Other genetic studies with IHF are also consistent with the model since they indicate that amino acid residues in the

† This work is supported by the NIH Grant GM28717.

* To whom correspondence should be addressed at the Department of Biochemistry, 523 Medical Sciences Building, 506 S. Mathews Ave., Urbana, IL 61801. Phone: 217-333-2852. Fax: 217-333-6461. E-mail: gumport@uiuc.edu.

‡ Department of Microbiology, University of Illinois.

§ Department of Chemistry, University of Liverpool.

‖ Department of Biochemistry, University of Illinois.

⊥ College of Medicine, University of Illinois.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1995.

[1] Abbreviations: 2-AP, 2-aminopurine; 3-DA, 3-deazaadenine; DAP, diaminopurine; I, inosine; IHF, integration host factor; HPLC, high-performance liquid chromatography; BSA, bovine serum albumin; TBE, Tris−borate−EDTA; DMS, dimethyl sulfate.

0006-2960/95/0434-13082$09.00/0

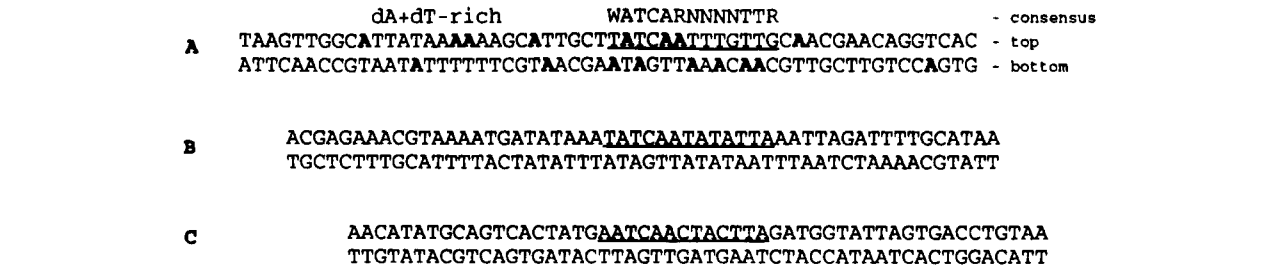

FIGURE 1: Nucleotide sequences for the IHF-binding sites in the *attP* site of bacteriophage λ. Parts A, B, and C are sequences for H′, H2, and H1 sites, respectively. In panel A, the positions substituted with analogues are in boldface. The proposed consensus sequences (Friedman, 1988) are underlined, and the dA+dT-rich sequence is indicated.

loops of the α subunit and the C-terminal portion of the β subunit are important for the DNA sequence recognition of IHF (Granston & Nash, 1993; Mengeritsky et al., 1993). In addition, recent results of photo-cross-linking of IHF to its DNA site also support this structural model (Yang & Nash, 1994).

In this study, we measured the equilibrium binding constant of IHF with its specific and nonspecific sites. The sequence determinants for IHF binding were explored by making adenine analogue substitutions at various, unique positions within the target sequence. The results suggest that IHF binds to its target sites with about a thousand-fold higher affinity than to a nonspecific site and that the recognition is accomplished through interactions with both grooves of the DNA.

MATERIALS AND METHODS

Oligodeoxyribonucleotides. The duplex IHF-binding sites and their analogue-containing variants used in this study were all synthesized as 56mers (Figure 1). The synthesis of the oligodeoxyribonucleotides was carried out by standard phosphoramidite chemistry (Gait, 1984) except for those containing 2-aminopurine. The oligomers containing 2-AP were synthesized starting with the 5′-*O*-dimethoxytrityl-N^2-isobutyryl-2-aminopurine-*p*-chlorophenyl phosphate, barium salt, obtained from Pharmacia, Inc. Briefly, two Applied Biosystems 380B DNA synthesizers were used. One was employed to synthesize the oligomer containing the usual bases by phosphoramidite chemistry and the other to insert 2-AP by phosphotriester chemistry. After a standard synthesis to the required addition step, the column was moved to the second machine where the triester, which had been converted from the barium salt to the triethylammonium salt by chromatography on Dowex 50W-X8, triethylammonium form, was activated with 1-(mesitylenesulfonyl)-3-nitro-1,2,4-triazole (Sigma, Inc.) and 1-methylimidazole (Aldrich) in anhydrous pyridine (Aldrich) (Gait, 1984). The activation and coupling steps took 1 min and 15 min, respectively. After addition of the analogue and washing of the column with acetonitrile, it was returned to the first machine for the completion of the synthesis with regular phosphoramidite nucleotides. The full-length oligomer was cleaved from the column by concentrated ammonium hydroxide by standard procedures. The solution was then dried *in vacuo* and the residue treated with 200 μL of a solution containing 70 mg of *syn*-2-nitrobenzaldoxime and 50 μL of 1,1,3,3-tetramethylguanidine in 1 mL of a 50% (v/v) dioxane/water solution for 16 h at room temperature to remove the 2-chlorophenyl group. This solution was then dried *in vacuo*, dissolved in ammonium hydroxide, and heated at 55 °C for 12 h to remove the remaining protecting groups. The protected phosphoramidite, 3-deaza-2′-deoxyadenosine [2-*N*,*N*-dibenzoyl-5′-*O*-(4,4-dimethyltrityl)-3-deaza-2′-deoxyadenosine 3′-*O*-(2-cyanoethyl)-*N*,*N*-diisopropylphosphoramidite], was synthesized as described (Cosstick et al., 1990). The protected 2,6-diaminopurine phosphoramidite was purchased from American Bionetics, Hayward, CA. All syntheses were performed at the University of Illinois Genetic Engineering Facility. The oligonucleotides were purified with an oligonucleotide purification cartridge (OPC, from Applied Biosystems, Inc.) or by reverse-phase HPLC. The purified oligomers were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Unreacted ATP was removed by chromatography on DE52 (diethylaminoethyl)cellulose (Whatman). The columns were washed 3 times with low-salt buffer (0.15 M NaCl, 0.1 M EDTA, and 20 mM Tris-HCl, pH 8.0) and the oligonucleotides eluted by high-salt buffer (1.0 M NaCl, 0.1 M EDTA, and 20 mM Tris-HCl, pH 8.0). The oligonucleotides were precipitated with ethanol and washed with 70% ethanol, and the pellets were dried and resuspended in H_2O. The purified oligonucleotides appeared as a single band after electrophoresis on acrylamide gels under denaturing conditions. The specific activities of the oligonucleotides were determined by scintillation counting using a Beckman Model LS 1801 scintillation counter and ultraviolet spectroscopy at 260 nm using either an extinction coefficient assuming 1 OD unit equals 33 μg/mL single-strand DNA or a molar extinction coefficient obtained by summing the molar extinction coefficients of individual deoxyribonucleoside monophosphates (Maniatis et al., 1982). For the analogue-containing oligonucleotides, the same values were assumed (Aiken et al., 1991). Equal molar amounts of complementary DNA strands were annealed in a thermocycler. The oligonucleotides were heated for 3 min at 90 °C and cooled to room temperature at a rate of 1 °C/min. The annealed products were more than 98% duplex as indicated by nondenaturing gel electrophoresis of the ^{32}P-labeled oligonucleotides. A summary of the analogues used in this study is shown in Figure 2. When incorporated into the DNA, the analogues modify the exposed surfaces of the major and minor grooves.

Electrophoretic Mobility-Shift Assays. The IHF-binding conditions were derived from those described previously (Lee et al., 1991). Purified *E. coli* IHF (1.1 mg/mL, 55 μM) was a gift from Dr. Howard Nash. Diluted IHF (2 μL) at the desired concentration in dilution buffer (50 mM Tris-HCl, pH 7.4, 20 mg/mL BSA, 240 mM KCl, and 10% glycerol) was incubated at room temperature for 20 min with ^{32}P-labeled DNA in a total volume of 20 μL (50 mM Tris-HCl, pH 8.0, 70 mM KCl, 0.3 mg/mL BSA, 5% glycerol, and

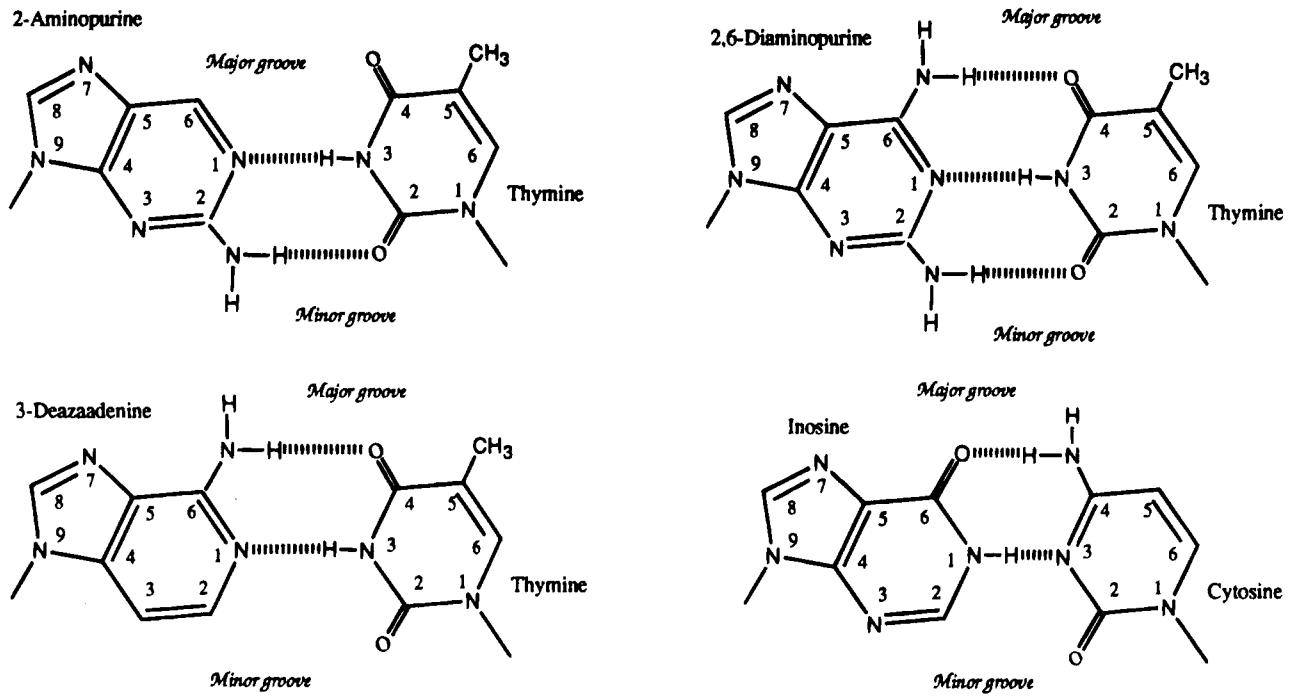

FIGURE 2: Base pairing schemes containing the four analogues used in this study. Base pairs are drawn with the major groove on top and the minor groove at the bottom.

sonicated calf thymus DNA). The nonspecific DNA concentration in all binding reactions was 70 μM (base pairs) except in those experiments measuring the nonspecific binding constant where it was varied.

The binding-assay products were separated on 7% polyacrylamide gels [acrylamide:bis(acrylamide) ratio of 29:1]. The gels were loaded with reaction samples lacking loading dyes and run in 1× TBE (89 mM Tris−borate, 2 mM EDTA) at 10 V/cm for 2.5 h at room temperature and then dried onto Whatmann 3MM paper. The dried gels were subjected to autoradiography at −70 °C with an intensifying screen or to a Phosphoimager cassette at room temperature.

Quantification of Equilibrium Binding Constants. The equilibrium binding constants of IHF for its target sites were determined by two methods. The first method was a modification of that of Baker et al. (1986), and was used for the H1-, H2-, and H′-binding sites as well as for nonspecific sites. Since the stoichiometry of IHF binding is one dimer per site (Yang & Nash, 1989), IHF binds DNA both specifically and nonspecifically with the equilibrium binding constants, K_s, for specific and, K_n, for nonspecific binding being

$$I + D_s \overset{K_s}{\leftrightarrow} D_sI \quad K_s = \frac{[D_sI]}{[D_s][I]}$$

$$I + D_n \overset{K_n}{\leftrightarrow} D_nI \quad K_n = \frac{[D_nI]}{[D_n][I]}$$

where I is free IHF; D_s and D_n are specific and nonspecific DNA sites, respectively; D_sI and D_nI are the corresponding DNA−IHF complexes. In the reaction mixture, the total amount of IHF, $[I^0]$, can be expressed as $[I^0] = [I] + [D_sI] + [D_nI]$. After rearrangement:

$$[D_sI] = \frac{[I^0]K_{app}[D_s]}{1 + K_{app}[D_s]}$$

and

$$[D_sI]/[D_s] = -K_{app}[D_sI] + K_{app}[I^0] \qquad \text{(i)}$$

where

$$K_{app} = \frac{K_s}{1 + [D_n]K_n}$$

K_{app} is the apparent equilibrium binding constant. Because the concentration of nonspecific binding sites (~70 μM) far exceeds that of IHF (~55 nM), we can approximate $[D_n] \approx [D_n^0]$. Consequently, the expression for K_{app} can be rewritten as

$$K_{app} = \frac{K_s}{1 + [D_n^0]K_n} \qquad \text{(ii)}$$

This equation can be rearranged as

$$\frac{1}{K_{app}} = \frac{1}{K_s} + \frac{K_n}{K_s}[D_n^0] \qquad \text{(iii)}$$

After separating the free DNA and DNA−IHF complex by electrophoresis and visualizing them by radioautography, the densities of corresponding bands were determined and the areas under the peaks established using an LKB Ultrascan XL scanning laser densitometer connected to an IBM PC. According to eq i, the slope of a plot of $[D_sI]/[D_s]$ as a function of $[D_sI]$ equals $-K_{app}$, a value which is governed by the specific binding constant K_s, the nonspecific binding constant K_n, and the concentration of nonspecific DNA sites in the assay. The relationship between K_{app} and nonspecific DNA is revealed by eq iii. If K_{app} values were measured at

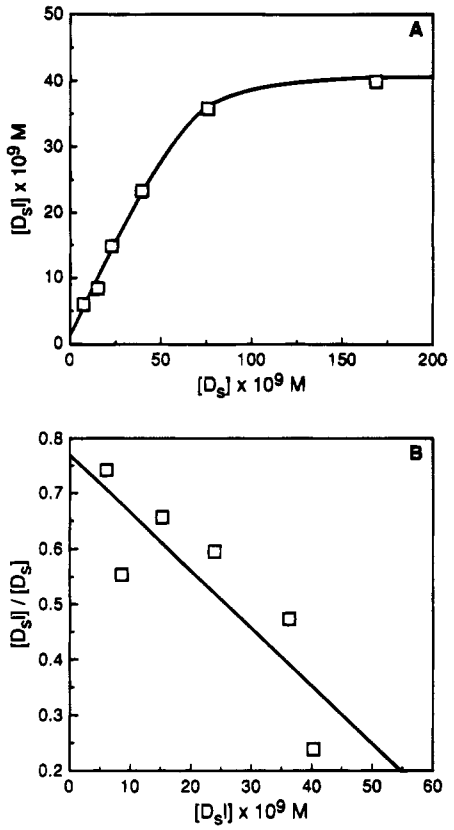

FIGURE 3: Determination of K_{app}. (A) Plot of the specific DNA concentration [D_s] and IHF–DNA complex concentration [D_sI]. (B) Plot of [D_sI] and its fractional specific DNA site occupancy [D_sI]/[D_s]. The slope of this curve gives the K_{app} according to eq i (see Materials and Methods).

different nonspecific DNA concentrations, the y intercept of a plot of $1/K_{app}$ versus $[D_n^0]$ is $1/K_s$. In turn, K_n can be calculated by dividing the slope, K_n/K_s, by the y intercept, $1/K_s$.

The second method is described by Carey (1991), where the $K_{d(app)}$ equals the concentration of IHF at which 50% of the DNA is in complexed form. The dried acrylamide gels were directly exposed to a Molecular Dynamics cassette and analyzed using Phosphorimager SP (Molecular Dynamics). The pixel values were within the linear range of the screen. The titration curves were then plotted, and $K_{d(app)}$ values were determined from the plot. This method was used with the analogue-substituted oligonucleotides. The molar amount of IHF was at least 10–100 times more than that of DNA to ensure that the amount of complexed form was negligible with respect to the total IHF (Carey, 1991). Within experimental error, the two methods gave identical results.

RESULTS

Determination of Binding Constants of IHF. We measured the equilibrium binding constants of IHF for its three λ *attP* sites (H1, H2, and H′) as well as for nonspecific DNA to determine the relative specificity of binding. Preliminary experiments with ^{32}P-labeled oligonucleotides and IHF showed that 55 nM IHF binds approximately 50% of 1 nM H′ DNA (data not shown). A series of experiments varying the amount of DNA were then performed at a fixed concentration of IHF (55 nM). An example of an experiment is shown in Figure 3. The amount of DNA–IHF complex increases linearly as the H′ DNA concentration raises from

Table 1: Equilibrium Association Constants of the Three IHF-Binding Sites in the *attP* Region

K_{app}[a] ($\times 10^7$ M^{-1})		
H′ site	H2 site	H1 site
1.6	0.69	2.78
1.9	0.64	3.13
1.4	0.83	5.48
1.5	0.94	4.20
mean ± SD 1.6 ± 0.22	0.78 ± 0.14	3.90 ± 1.22

[a] K_{app} is defined under Materials and Methods.

Table 2: Relationship between Nonspecific DNA Concentration, [D_n], and K_{app}[a]

[D_n] ($\times 10^{-6}$ M)	K_{app} ($\times 10^7$ M^{-1}) ± SD	$1/K_{app}$ ($\times 10^{-7}$ M)
0.0	123 ± 16.10	0.01
18.9	6.96 ± 2.24	0.14
37.9	2.23 ± 0.55	0.45
70.0	1.60 ± 0.22	0.63

[a] D_n and K_{app} are defined under Materials and Methods. The K_{app} values are averages of four or five independent experiments.

0 to 75 nM, approaching a saturation value of ~40 nM complex (Figure 3A). The individual equilibrium binding constants were determined as the slope of [D_sI]/[D_s] versus [D_sI] (Figure 3B). The results for all three sites are summarized in Table 1. IHF binds to its three *attP* sites with affinities (K_{app}) on the order of 10^7 M^{-1}. The H1 site had slightly higher affinity. It is noteworthy that in a previous study using footprinting analysis (Craig & Nash, 1984) IHF bound the H1 more weakly than H′ and H2. Surprisingly, computer analysis of 27 putative IHF-binding sites predicted that of the three IHF *attP* sites H1 was most similar to the consensus sequence (Goodrich et al., 1990). Our data support the prediction suggested by computer analysis. Since the DNA fragment (~400 base pairs) used in the footprinting study was longer than the oligonucleotides (56 base pairs) used here, we assume that this difference must be due to the flanking sequences bounding the consensus target sequence.

The apparent equilibrium binding constant we measured is a function of both specific and nonspecific binding and is thus affected by the amount of nonspecific DNA present in the assay (eq ii under Materials and Methods). In order to perform direct measurements of specific and nonspecific binding of IHF, we determined the K_{app} of the H′ site at different concentrations of a nonspecific DNA (calf thymus). The results are shown in Table 2. The K_{app} increases from (1.6 ± 0.22) × 10^7 M^{-1} to (123 ± 16.10) × 10^7 M^{-1} as nonspecific [DNA] decreases from 70 μM to zero (Figure 4). As defined in eq iii under Materials and Methods, the slope and the y intercept were measured. The nonspecific equilibrium binding constant (K_{ans}) of IHF is 5.8 × 10^5 M^{-1}. The specific binding of IHF to the H′ site is 6.8 × 10^8 M^{-1}. The value of the specific K_d agrees well with those reported previously (Schneider et al., 1992; Yang & Nash, 1994). If we define K_s/K_n as the relative specificity of IHF binding, IHF prefers the H′ site over an average random site by about 1000-fold.

Base-Analogue Studies. Three adenine analogues were used to modify the minor groove of the DNA since their edges are likely to be in direct contact with IHF (Figure 2). Analogues were placed singly into 41 different positions in

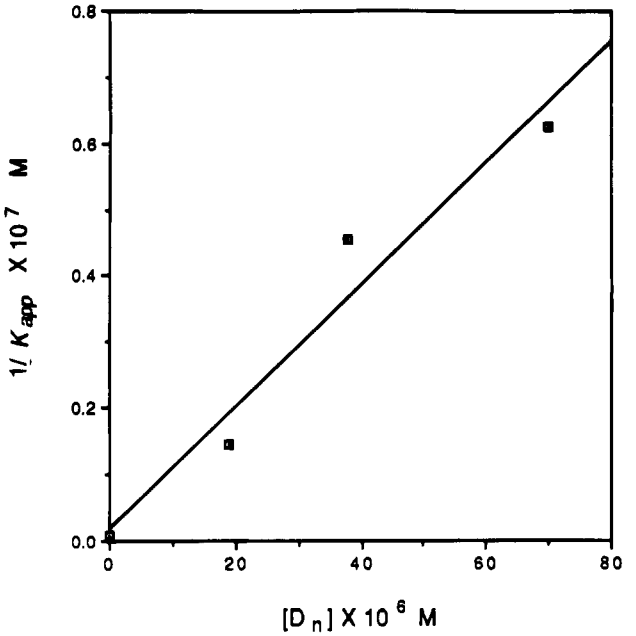

FIGURE 4: Titration curve for $1/K_{app}$ and $[D_n]$. The nonspecific binding constant was calculated by dividing the slope by the y-intercept according to eq iii under Materials and Methods.

```
DAP        0        0           +  +0
3DA        0        0           0  +0
2AP        0       00     0     0  00        0
 TAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCAC
      |        |        |        |        |        |
     +10      +20      +30      +40      +50      +60
      |        |        |        |        |        |
 ATTCAACCGTAATATTTTTTCGTAACGAATAGTTAAACAACGTTGCTTGTCCAGTG
2AP              0        0     +  +     +  +  ++           0
3DA                       0  +     0  0  00              0
DAP                       0  +     0  0  ++              0
```

FIGURE 5: Summary of the analogue substitution effects on IHF binding to the H′ site. The (0) indicates that a substitution at this position failed to result in a significant affinity change, whereas a (+) means that a reduction in affinity was detected.

the H′ site by synthesis and the resultant oligonucleotides tested for their abilities to bind IHF *in vitro*.

(1) 2-Aminopurine (2-AP). After incorporation into DNA, 2-aminopurine differs from adenine in that the 6-amino group is absent from the major groove and a 2-amino group is present in the minor groove (Figure 2). Thus, the surfaces of both grooves are modified. The particular sites of substitution were carefully chosen based on previous biochemical (Craig & Nash, 1984; Yang & Nash, 1989) and genetic (Lee et al., 1991) studies. We replaced 9 of 10 adenines on both strands in the consensus region of the H′ site, including positions 33, 34, 35, 37, 38, 39, 41, 43, and 44 (Figure 1). Six adenines at positions 14, 18, 27, 28, 47, and 57 were substituted as controls since neither footprinting nor genetic studies suggested that they are important for IHF binding. Previous studies showed that the dA+dT-rich sequence in the H′ site is important for IHF binding *in vitro* and *in vivo* (Craig & Nash, 1984; Hales et al., 1994a; Lee et al., 1991; Yang & Nash, 1989). We therefore also replaced adenines at positions 21 and 22 to test the possibility of specific-sequence recognition by IHF in this part of the H′ site. Among 17 singly-substituted H′ sites, only 6 (positions 33, 35, 39, 41, 43, and 44) showed reduced affinity for IHF (Figure 5). These sites are all within the proposed consensus sequence. The affinities of the sites not implicated in binding (positions 14, 18, 27, 28, 47, and 57) were unaltered. Substitutions in the consensus sequence at positions 34, 37, and 38 failed to reduce affinity as did the substitutions in the dA+dT-rich region at positions 21 and 22 (Figure 5).

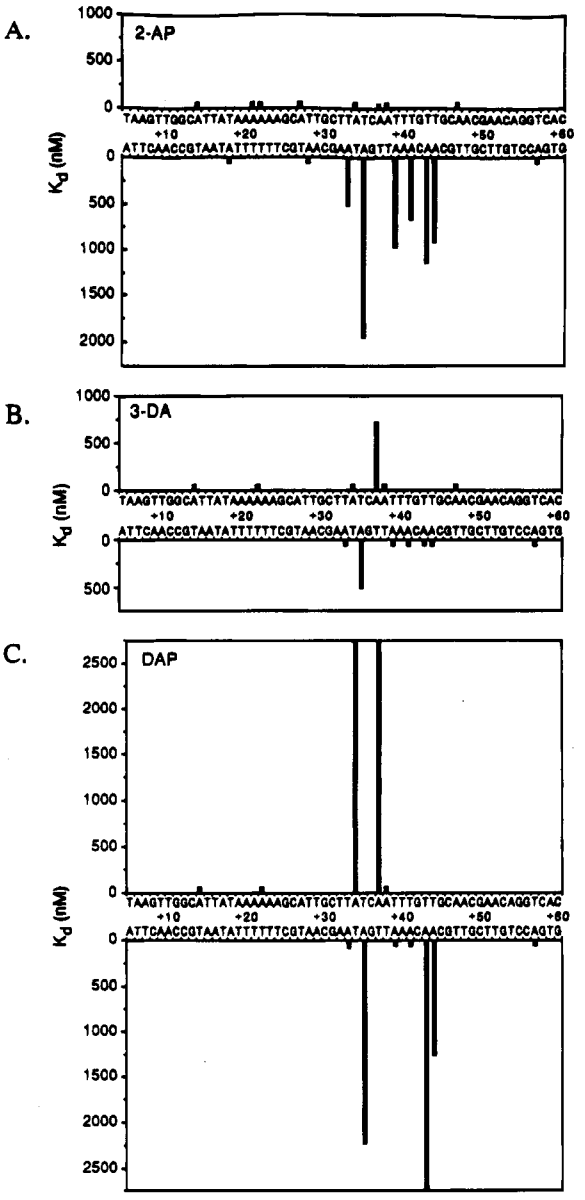

FIGURE 6: Effects of structural alterations of adenine on IHF binding to the H′ site. Panels A, B, and C summarize results for 2-aminopurine (2-AP), 3-deazaadenine (3-DA), and 2,6-diaminopurine (DAP), respectively. Both strands are indicated, and in each panel, the ordinate shows the binding affinity expressed as the dissociation constant (K_d) and the abscissa displays the H′ sequence with the consensus portion underlined. In panel C, the data shown for substitutions at positions 34, 37, and 43 indicate the minimum possible values. See Results for an explanation.

The quantitation of IHF binding to those sites with reduced affinities revealed that the affinities decreased from 10-fold to 35-fold (Figure 6A). When adenine was replaced with 2-aminopurine, H′35 had the strongest effect. The K_d value increased more than 35-fold—a change which corresponds to a free energy difference ($\Delta\Delta G$) of 2.1 kcal/mol. Substitution at positions 39, 43, and 44 decreased the affinity approximately 20-fold whereas substitutions at positions 33 and 41 had approximate 10-fold effects.

(2) 3-Deazaadenine (3-DA). This adenine analogue replaces the N-3 nitrogen of the heterocyclic ring system with a carbon and thus removes a potential hydrogen bond

acceptor in the minor groove (Figure 2). Single substitutions of 12 adenines in the H′ site revealed only 2 positions (35 and 37) where binding was disrupted (Figure 5). Recall that substitution of 2-AP at position 35 had a major effect; we thus conclude that the integrity of the minor groove at this position is vital for IHF binding. Although substitutions with 2-AP at position 37 failed to disrupt IHF binding, the results with 3-DA and other analogues (see below) and base-pair substitutions with standard bases confirm that the minor groove at this position is important for binding. The lack of effects at other sites substituted with 3-DA indicates that minor groove interactions between IHF and the N-3 position of adenine in its λ H′ site are likely localized to the WATCAR segment of the consensus sequence. The two 3-DA substitutions at the H′ site reduced affinities for IHF about 10-fold (Figure 6B), an energetic difference of 1.3−1.5 kcal/mol. These values are consistent with the disruption of one hydrogen bond.

(3) 2,6-Diaminopurine (DAP). The third adenine analogue, 2,6-diaminopurine, conserves the major groove surface of adenine but adds an amino group in the minor groove. Substitution of this analogue disrupted binding to 5 of the 12 sites tested [positions 34, 35, 37, 43, and 44 (Figure 5)]. Sites with substitutions at positions 34, 37, and 43 had such weak affinities (Figure 6C) that a binding constant could not be determined. The affinity of IHF for these sites was reduced by at least 50-fold. Position 35 showed an approximate 40-fold reduction in affinity, and position 44 had a 23-fold reduction. These substitutions are all located within the consensus region. DAP substitution did not affect IHF binding at positions 39 and 41—sites that lie in the region (4 base pairs) between the WATCAR and TTR consensus elements.

(4) Other Analogue and Natural Base Substitutions. The systematic replacements of adenines with 2-AP, 3-DA, and DAP in the H′ site confirm the involvement of the minor groove in IHF specific binding to DNA. Results of analogue substitutions at position 37 suggested that the major groove might also be involved because removal of the 6-amino group from the major groove, i.e., substitution of adenine by 2-AP, improved IHF binding (Figures 5 and 6A−C). To further monitor the major groove involvement, we constructed two substrates with base-pair substitutions at position 37. In one experiment, a G·C base pair replaced the A·T base pair at position 37. Although a G·C base pair has a minor groove structure like that of a 2-AP·T base pair (Figure 2), the substitution almost completely abolished IHF binding (Figure 7). This result supports the involvement of the major groove at this site.

The transcription factor TFIID and the Hin recombinase both continue to interact with the minor groove of their DNA-binding sites when an I (inosine)·C base pair replaces an A·T base pair (Hughes et al., 1992; Star & Hawley, 1991), because the I·C base pair has a minor groove structure like that of an A·T base pair. In contrast, the major groove structure of an I·C base pair is like that of a G·C base pair. Thus, in a second experiment we replaced the A·T with an I·C base pair at position 37 and found a moderate, 3-fold reduction in binding affinity (Figure 7). A similar result was obtained by the independent study of Yang and Nash (1995). These results, in which the base pair (I·C) substitution maintains the wild-type (A·T) minor groove but alters the major groove with respect to A·T, again suggest that the major groove is involved in IHF binding at position 37.

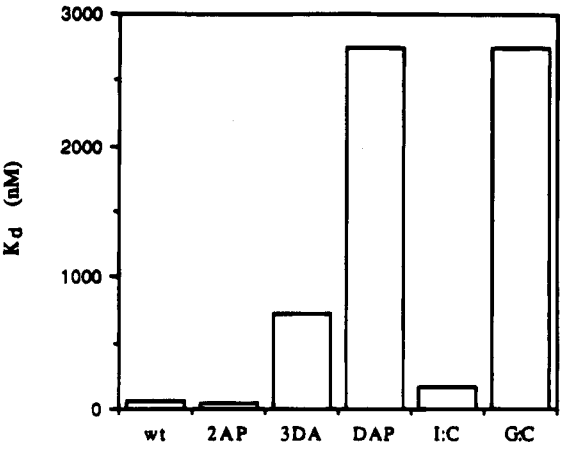

FIGURE 7: Results from base substitutions at position 37. The affinities for each analogue-substituted site are expressed as the dissociation constant (K_d). 2-AP, 3-DA, and DAP were base changes, whereas I·C and G·C were base pair changes. The dissociation constants shown for DAP and G·C are minimum possible values. See Results for an explanation.

DISCUSSION

In this study, we determined the equilibrium binding constants for IHF with its λ *attP* H1, H2, and H′ sites and for random sequences (sonicated calf thymus DNA). The results indicate that IHF binds with dissociation constants of 1.5 nM to the H′ site and of 1.7 μM for the random sequences. The value of the ratio of these dissociation constants (≈1100) agrees well with the 1000-fold difference between specific and nonspecific binding sites obtained by other laboratories (Mengeritsky et al., 1993; Yang & Nash, 1994). The amount of IHF within the cell is relatively high with up to 17 000 IHF molecules per cell having been estimated (Ditto et al., 1994). The affinity of IHF for random DNA and the large number of potential IHF-binding sites in the bacterial chromosome (Freundlich et al., 1992) suggest that most IHF molecules are bound to DNA in the cell.[2] The 1000-fold preference of IHF for its target with respect to random sequences lies between the preference of the Lac (Linn & Riggs, 1975) and Trp repressors (Carey, 1988), which are 10^8-fold and 10^2-fold, respectively.

We used a series of adenine analogue substitutions to reveal the atomic determinants important for specific binding. To test the hypothesis that IHF interacts with the minor groove of the DNA, the base analogues were chosen so that the minor groove was altered. The 3-deazaadenine substitution removed the N3, a potential hydrogen bond acceptor, from the minor groove, and 2,6-diaminopurine and 2-aminopurine added an amino group to the minor groove when they substituted for adenine. The 2-aminopurine analogue caused changes in both major and minor grooves by switching the amino group from the 6 position (major groove) to the 2 position (minor groove). Single substitution of adenines with these analogues at various positions in the H′ site resulted in a spectrum of results varying from no

[2] Assuming the *E. coli* cells are about 1 μm^3 in volume. The IHF concentration in the cell would then be 30 μM. This is about 20 times the K_{dns} = 1.7 μM at which half of the IHFs are in the bound form. Assuming free sites exist on the DNA, most of the IHF would be bound, a case analogous to that of the lac repressor (Linn & Riggs, 1975).

change to more than 50-fold reductions in binding affinity. The substituted sites that disrupted IHF binding were all within the proposed consensus sequence and substitutions outside this region had little effect on IHF binding. We discuss below the analogue results when substituted at four arbitrarily defined subsites of the consensus sequence: the WATCAR, TTR, NNNN, and dA+dT-rich regions (Figure 1).

The WATCAR Region. The genetic studies of Lee et al. (1992) indicated that the WATCAR region is contacted by the arm of α subunit. The minor groove of this half-turn of helix presumably faces IHF and the major groove faces away from the protein. This region is composed of six base pairs, and the four base pairs in the center of the sequence are usually ATCA. Computer alignment of the known IHF-binding sites (Goodrich et al., 1990) indicates that the bases next to ATCA are also distinguished during binding, i.e., W (A or T) at the 5′ end and R (A or G) at the 3′ end.

In our experiments, 2 of 13 3-DA substituted λ *attP* H′ sites showed a reduction in affinity for IHF. These are substitutions at position 35 in the bottom strand and position 37 in the top strand (Figure 1). The removal of N3 from the minor groove at these two positions reduced the IHF-binding affinity about 10-fold. For each position, the calculated free energy change for the disruption is approximately 1 kcal/mol, a value consistent with the loss of a hydrogen bond. The nucleotides containing DAP, which adds an amino group into the minor groove at position 35, also disrupted IHF binding. The IHF affinity of this DAP-containing oligonucleotide was reduced approximately 40-fold. The substitution of A35 by 2-AP caused a level of disruption similar to that of DAP, indicating that alteration of the minor groove both by removal of a hydrogen bond acceptor (N3 of adenine) and by addition of an amino group is highly disruptive. The similar effects caused by these three base analogues clearly demonstrated the importance of the minor groove in IHF binding at position 35.

Unlike position 35, the substitution of A37 with 2-AP unexpectedly failed to disrupt IHF binding. Adding a 2-amino group to the minor groove was detrimental when DAP was substituted at this position. Although 2-AP also introduces a 2-amino group in the minor groove and likely disrupts direct interactions between the protein and the adjacent N3, it had no effect on IHF binding. Thus, in this case IHF binding to this 2-AP-substituted site could occur by compensation for the disruptive effect of adding a 2-amino group in the minor groove by removing the 6-amino group from the major groove. Possibly, the absence of a 6-amino group in the major groove resulting from a 2-AP substitution facilitated the binding of IHF so as to lead to no change in the overall affinity. In the model proposed by Yang and Nash, the bound DNA was sharply bent toward the minor groove around this region (Yang & Nash, 1989). As a result, the major groove would be widened and the minor groove compressed. We speculate that the removal of the 6-amino group in the major groove by a 2-AP substitution could favor this widening and thus might complement the deleterious effect of the 2-amino group in the minor groove in an indirect way that does not involve the amino group on the base interacting with amino acids on the protein. The 2-AP•T base pair is somewhat less stable than an A•T base pair, and a kinetic study of base pair opening rates in a decamer, CTGA(2-AP)TTCAG, demonstrated that the 2-AP•T base pair has a higher opening rate than the corresponding A•T base pair (Lycksell et al., 1987). This apparent "loose" pairing by 2-AP•T might confer a flexible site for bending in the DNA and thus facilitate IHF binding to and bending this site.

The substitution of A33 (the W in the consensus WATCAR) with 2-AP had only a small deleterious effect on IHF binding, and no change was seen when 3-DA or DAP was introduced at this site. These results suggest that the modest effect was likely due to a change in the major groove. The major groove of the base pair at position 33 should be close to IHF and is about a half-turn helix away from the one at position 37, a site which we propose to have a strong interaction with IHF in the minor groove. The sequence between positions 33 and 37 would probably also have its minor groove close to IHF. Consistent with this interpretation, substitutions at positions 34 and 35 within this half-turn helix also had strong effects when the minor groove structure was altered. Thus, our data suggest that the center of the WATCAR element interacts with IHF closely in the minor groove.

The finding of Yang and Nash (1995) that replacing the G•C at position 36 with I•C (deleting the 2-amino group of guanine in the minor groove) resulted in a free energy difference ($\Delta\Delta G$) of 2.3 kcal/mol (an affinity change of about 50-fold) is further evidence that IHF interacts with the minor groove in this region, since the substitution of I•C with G•C changes substituents in the minor groove.

The Region between WATCAR and TTR. This is a less conserved region in the IHF consensus sequence. Although initially listed as NNNN (Gardner & Nash, 1986; Gamas et al., 1987; Friedman, 1988), some bases in this region were later shown to be preferred (Kur et al., 1989; Goodrich et al., 1990). Genetic studies indicated that mutations in this region can have from weak to strong effects on IHF binding (Lee et al., 1991). Footprinting studies indicated that this region is susceptible to hydroxyl radical attack, but is protected from both DMS methylation and DNase I attack (Craig & Nash, 1984; Yang & Nash, 1989). These results show that although this region is close to IHF it remains accessible to small molecules like hydroxyl radicals. In our study, substitutions with DAP and 3-DA did not interfere with IHF binding, indicating that direct read-out in the minor groove in this part of the sequence is unlikely. However, the replacement of adenine by 2-AP at positions 39 and 41 reduced the binding affinity moderately, i.e., 10-fold to 20-fold, indicating the composition of the major groove in this region matters. Since other substitutions support the model (Yang & Nash, 1989) suggesting that the minor groove of the WATCAR region faces toward IHF, the region between the WATCAR and TTR sequences would have its major groove turned toward IHF.

The TTR Region. In the model proposed by Yang and Nash (1989), the minor groove of this region (positions 43, 44, and 45) is juxtaposed to IHF. Their model is further supported by the observation that the insertion of a 2-amino group into the minor groove at positions 43 and 44 reduced the affinity for IHF. The extent of disruption at the two sites was unequal. The substitution of either 2-AP or DAP at the former position had a stronger effect than the substitution at the latter, suggesting a stronger interaction at position 43. The results also agree with previous methylation protection and interference studies in that the adenine at

position 43 was consistently protected upon IHF binding and, if methylated, interfered with IHF binding, whereas at position 44 only methylation interference was observed (Craig & Nash, 1984; Yang & Nash, 1989).

The dA+dT-Rich Region and Surrounding Sequences. A dA+dT-rich element (positions 19−24 in the H′ site) is found in several IHF-binding sites in the region 5′ to the WATCAR consensus element. It has been suggested that dA+dT-rich sequences are favored where the minor groove faces inward due to the anisotropic flexibility of the DNA double helix (Drew & Travers, 1985; Gartenberg & Crothers, 1988; Satchwell et al., 1986). Although natural base changes in this region caused defects in the binding of IHF to mutant H′ sites *in vivo* (Lee et al., 1991), none of our single-analogue substitutions reduced the binding affinities significantly. We believe, therefore, that there are few, if any, specific base contacts with IHF in this region for either the major or the minor groove. The lack of specific contacts in the major groove is further supported by the observation that the replacement of all six A·T base pairs in the H′ site with I·C base pairs caused little change in IHF binding (Yang & Nash, 1995).

As pointed out in an analysis of restriction enzyme−DNA interactions using DNA analogues (Aiken & Gumport, 1991), some of the affinity changes caused by the analogue substitutions may be due to indirect effects on the DNA structure. The argument that the replacement of a base by a base analogue could influence the curvature of the DNA double helix has been addressed in several structural studies (Chazin et al., 1991; Diekmann et al., 1987; Nordlund et al., 1989; Schneider et al., 1992; Seela & Green, 1992; Sowers et al., 1986). A comparative 2D NMR analysis has shown that substitution by DAP in oligonucleotides appeared not to disturb the global or local conformation of the B-DNA duplex (Chazin et al., 1991). When DAP replaced the adenine in a Z-DNA hexamer, it also retained the conformation of Z-DNA (Schneider et al., 1992). Studies using 2D NMR and the intrinsic fluorescence of 2-AP also demonstrated that 2-AP-substituted DNA is B-form in solution (Nordlund et al., 1989; Sowers et al., 1986). Studies using gel electrophoretic analysis showed that the replacement of adenine by 2-AP and DAP (Diekmann et al., 1987) or its isosteric analogue 3-DA (Seela & Green, 1992) may alter the intrinsic bending of poly(dA)·poly(dT) sequences. It is hypothesized that the introduction of a 2-amino group on purines may prevent the DNA from bending into the minor groove. The substitution of 2-AP for A in an oligonucleotide duplex destabilizes the helix (Eritja et al., 1986), but not enough to significantly affect the native to denatured ratio of our oligonucleotides, which are 56 base pairs long. Substitution of adenine by 3-DA may disrupt a "water-spine" in the minor groove and thus affect the bending of the DNA duplex. In our study we failed to detect significant affinity changes when we replaced the adenine at H′22 in the center of the A_6-tract of the H′ site. The 3-DA-substituted oligonucleotide, both free and in its IHF complex, appears to have the same mobilities as those of the free and IHF-bound unsubstituted sites (data not shown). The DAP and 2-AP substitutions at the center position in the A_6-tract of the H′ site also had no effect on IHF binding. Since it has been proposed that the role of this dA+dT-rich region is to assist IHF binding by facilitating bending (Hales et al., 1994a), we conclude that the substitutions of 2-AP, DAP, and 3-DA in this region do not affect IHF binding or the DNA conformation to an extent that allows detection by electrophoretic analysis.

In summary, we measured the specific and nonspecific binding constants of IHF. The three IHF-binding sites in *attP* have similar affinities, 62.5 nM, 128 nM, and 25.6 nM for H′, H2, and H1, respectively. The results suggest that IHF is a sequence-specific DNA-binding protein with a binding specificity intermediate between the Lac (high specificity) and Trp (low specificity) repressor proteins. Its relative specificity is closer to that of Trp than of Lac. Systematic substitutions of adenine with analogues revealed important contacts between IHF and its binding site in both the minor groove and the major groove. Therefore, although IHF appears to be primarily a minor groove recognizing protein, our current data also suggest an involvement of the major groove. The data also indicated that the energetic contribution of direct read-out is not equally distributed along the binding surface between IHF and its target DNA site. The strongest interactions lie in the WATCAR and TTR regions. Our data are consistent with the model originally proposed by Yang and Nash (1989) and further refined by our laboratory and others (Granston & Nash, 1993; Lee et al., 1992; Mengeritsky et al., 1993). The results reported here should prove useful in interpreting data arising from structural studies on IHF and its complex with DNA.

ACKNOWLEDGMENT

We are grateful to Dr. Howard Nash for his gifts of purified IHF and comments on the manuscript. We thank Dr. Janet Carey for suggestions about gel-retardation assays. We also thank Saw Kyin for excellent technical assistance in the synthesis of some oligonucleotides.

REFERENCES

Aiken, C. R., & Gumport, R. I. (1991) in *Methods in Enzymology* (Sauer, R. T., Ed.) pp 433−457, Academic Press, New York.
Aiken, C. R., McLaughlin, L. W., & Gumport, R. I. (1991) *J. Biol. Chem. 266*, 19070−19078.
Baker, R. E., Gabrielsen, O., & Hall, B. D. (1986) *J. Biol. Chem. 261*, 5275−5282.
Carey, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 975−979.
Carey, J. (1991) in *Methods in Enzymology* (Sauer, R. T., Ed.) pp 103−117, Academic Press, New York.
Chazin, W. J., Rance, M., Chollet, A., & Leupin, W. (1991) *Nucleic Acids Res. 19*, 5507−5513.
Cosstick, R., Li, X., Tuli, D. K., Williams, D. M., Connolly, B. A., & Newman, P. C. (1990) *Nucleic Acids Res. 18*, 4771−4778.
Craig, N., & Nash H. A. (1984) *Cell 39*, 707−716.
Diekmann, S., von Kitzing, E., McLaughlin, L., Ott, J., & Eckstein, F. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8257−8261.
Ditto, M. D., Roberts, D., & Weisberg, R. A. (1994) *J. Bacteriol. 176*, 3738−3748.
Drew, H. R., & Travers, A. A. (1985) *J. Mol. Biol. 186*, 773−790.
Drlica, K., & Rouvier-Yaniv, J. (1987) *Microbiol. Rev. 51*, 301−319.
Eritja, R., Kaplan, B. E., Mhaskar, D., Sowers, L. C., Pertruska, J., & Goodman, M. F. (1986) *Nucleic Acids Res. 14*, 5869−5884.
Freundlich, M., Ramani, N., Mathew, E., Sirko, A., & Tsui, P. (1992) *Mol. Microbiol. 6*, 2557−2563.
Friedman, D. I. (1988) *Cell 55*, 545−554.
Gait, M. J., & Sproat, B. S. (1984) in *Oligonucleotide synthesis* (Gait, M. J., Ed.) pp 83−115, IRL Press, Oxford and Washington, D.C.
Gartenberg, M. R., & Crothers, D. M. (1988) *Nature 333*, 824−829.
Goodman, S. D., & Nash, H. A. (1989) *Nature 341*, 251−254.

Goodrich, J. A., Schwartz, M. L., & McClure, W. R. (1990) *Nucleic Acids Res. 18*, 4993−5000.
Granston, A. E., & Nash H. A. (1993) *J. Mol. Biol. 234*, 45−59.
Hales, L. M., Gumport, R. I., & Gardner, J. F. (1994a) *J. Bacteriol. 176*, 2999−3006.
Hales, L. M., Gumport, R. I., & Gardner, J. F. (1994b) *Biochimie 76*, 1103−1126.
Hughes, K. T., Gaines, P. C. W., Karlinsey, J. E., Vinayak, R., & Simon, M. I. (1992) *EMBO J. 11*, 2695−2705.
Lee, E. C., MacWilliams, M. P., Gumport, R. I., & Gardner, J. F. (1991) *J. Bacteriol. 173*, 609−617.
Lee, E. C., Hales, L. M., Gumport, R. I., & Gardner, J. F. (1992) *EMBO J. 11*, 305−313.
Linn, S. A., & Riggs, A. D. (1975) *Cell 4*, 107−111.
Lycksell, P. O., Graslund, A., Claesens, F., McLaughlin, L. W., Larson, U., & Rigler, R. (1987) *Nucleic Acids Res. 15*, 9011−9015.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, pp C.4−C.10, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mengeritsky, G., Goldenberg, D., Mendelson, I., Giladi, H., & Oppenheim, A. B. (1993) *J. Mol. Biol. 231*, 646−657.
Nordlund, T. M., Anderson, S., Nilsson, L., Rigler, R., Graslund, A., & McLaughlin, L. M. (1989) *Biochemistry 28*, 9095−9103.
Satchwell, S. C., Drew, H. R., & Travers, A. A. (1986) *J. Mol. Biol. 191*, 659−675.
Schneider, B., Ginell, S., Jones, R., Gaffney, B., & Berman, H. M. (1992) *Biochemistry 31*, 9622−9628.
Seela, F., & Green, T. (1992) *Nucleic Acids Res. 20*, 2297−2306.
Sowers, L. C., Fazakerley, G. V., Eritja, R., Kaplan, B. E., & Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5434−5438.
Star, D. B., & Hawley, D. K. (1991) *Cell 67*, 1231−1240.
Tanaka, I., Appelt, K., Dijk, J., White, S. W., & Wilson, K. S. (1984) *Nature 310*, 376−381.
Yang, C. C., & Nash, H. A. (1989) *Cell 57*, 869−880.
Yang, S., & Nash, H. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12183−12187.
Yang, S., & Nash, H. A. (1995) *EMBO J.* (in press).

BI950962W