GCN4 Binds with High Affinity to DNA Sequences Containing a Single Consensus Half-Site†

Jessica J. Hollenbeck and Martha G. Oakley*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received November 24, 1999; Revised Manuscript Received March 14, 2000

ABSTRACT: bZip proteins contain a bipartite DNA-binding motif consisting of a "leucine zipper" dimerization domain and a highly charged "basic region" that directly contacts DNA. These transcription factors form dimeric complexes with each monomer recognizing half of a symmetric or nearly symmetric DNA site. We have found that the bZip protein GCN4 can also bind with high affinity to DNA sites containing only a single GCN4 consensus half-site. Because several recent lines of evidence have suggested a role for monomeric DNA binding by bZip proteins, we investigated the structure of the GCN4·half-site complex. Quantitative DNA binding and affinity cleaving studies support a model in which GCN4 binds as a dimer, with one monomer making specific contacts to the consensus half-site and the other monomer forming nonspecific contacts that are nonetheless important for binding affinity. We also examined the folding transition induced in the basic regions of this complex upon binding DNA. Circular dichroism (CD) studies demonstrate that the basic regions of both monomers are helical, suggesting that a protein folding transition may be required for both specific and nonspecific DNA binding by GCN4.

The basic region leucine zipper (bZip)[1] family of DNA-binding proteins was first identified by McKnight and co-workers in 1988, on the basis of sequence conservation within a group of transcriptional activators and protooncogenes (*1*). This motif is now recognized as a bipartite DNA-binding domain, consisting of a coiled-coil leucine zipper dimerization domain and a highly charged basic region that directly contacts DNA (for reviews, see refs *2*−*4*). In the presence of DNA, the bZip domain forms a dimer of uninterrupted α-helices (*5*−*7*). The leucine zipper region is both necessary and sufficient for dimerization (*8*−*10*), forming a stable, two-stranded, parallel coiled coil (*11*−*13*). The basic region, on the other hand, is largely unstructured in the absence of DNA, but adopts a helical structure upon DNA binding (*14*−*22*).

Although the leucine zipper region does not contact the DNA directly, DNA binding has been thought to be absolutely dependent on dimerization. This assumption is based on evidence that most mutations that prevent dimerization also prevent DNA binding (*9*, *23*−*25*). Indeed, although the isolated basic region of the yeast transcriptional activator GCN4 does not bind to its DNA recognition site, replacement of the dimerization region with a flexible disulfide linkage restores DNA binding (*14*).

As would be expected for a dimeric DNA-binding protein, GCN4 recognizes a DNA sequence with approximate 2-fold symmetry. The pseudosymmetric sequence 5′-TGA(C/G)-TCA-3′ has been identified from a comparison of enhancer sites in GCN4-dependent promoters (*26*) and from in vitro selection experiments (*27*, *28*). However, unlike other dimeric DNA-binding proteins, identical contacts are not made with both half-sites of the consensus sequence. In the crystal structure of the GCN4·AP-1 complex, Arg-243 of one GCN4 monomer specifically contacts the central guanine nucleotide whereas Arg-243 of the other monomer forms nonspecific hydrogen bonds with the DNA backbone (*5*). Similarly, mutations in one DNA half-site have stronger effects on protein binding than symmetrically related mutations both in vitro and in vivo (*26*, *27*). Taken together, these results suggest that DNA binding by GCN4 is inherently asymmetric with respect to the optimal binding site and that specific recognition of a single half-site by one GCN4 monomer may be more important than recognition by the second monomer.

Due to the observation that many bZip proteins dimerize in the absence of DNA (*12*, *29*), it has been assumed that preformed bZip protein dimers are responsible for sequence-specific DNA recognition. Recently, however, several groups have observed that bZip peptides can bind to DNA sequentially as monomers that subsequently dimerize on the DNA (*30*−*32*). Similarly, footprinting assays have shown that monomeric Jun basic regions protect the DNA binding site of the Jun-Fos heterodimer from cleavage by DNase I (*33*), suggesting that a 1:1 bZip·DNA complex can exist. Finally, Schepartz and co-workers have demonstrated that when the

† Supported by Grant GM57571 from the National Institutes of Health and Grant 32029-G4 from the American Chemical Society, Petroleum Research Fund.

* To whom correspondence should be addressed.

[1] Abbreviations: bZip, basic region leucine zipper; CRE, cAMP-responsive element; CREB, CRE-binding protein; EDTA, ethylenediaminetetraacetic acid; CD, circular dichroism; bp, base pair; poly[dI·dC], polydeoxyinosinic-deoxycytidylic acid (poly[dI-dC]·poly[dI-dC]); CT DNA, calf thymus DNA; PCR, polymerase chain reaction; IPTG, isopropyl 1-thio-β-D-galactopyranoside; PMSF, phenylmethanesulfonyl fluoride; HPLC, high-performance liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; PBS, phosphate-buffered saline; DTT, DL-dithiothreitol; BSA, bovine serum albumin; θ_{app}, fraction DNA bound; $[L]_{tot}$, total protein monomer concentration; K_a, apparent monomeric equilibrium association constant; K_d, apparent monomeric equilibrium dissociation constant; $[\theta]_{222}$, molar ellipticity at 222 nm; NMR, nuclear magnetic resonance.

10.1021/bi992705n CCC: $19.00
Published on Web 05/02/2000

GCN4 basic region residues that contact DNA are incorporated into a stable, helical hairpin motif, high-affinity monomeric binding occurs (*34*). Taken together, these data suggest a role for monomer·DNA complex formation in vivo.

We have observed high-affinity, specific binding by the GCN4 bZip domain to DNA sites containing only a single consensus half-site, 5′-TGAC-3′. High-affinity binding to DNA half-sites has been observed with other DNA-binding proteins, including the *Escherichia coli* LexA repressor protein (*35*) and the rat glucocorticoid receptor (*36*). In fact, the observed half-site recognition by GCN4 may be biologically significant. A number of promoters that are responsive to bZip transcription factors contain only a single half-site of their consensus sequence. For example, the *his*3 gene, the *jun* proto-oncogene, the HIV-LTR, and the genes for human DNA polymerase α and transforming growth factor β1 (TGF-β1) contain isolated 5′-TGAC-3′ sequences that have been reported to confer responsiveness to GCN4 or AP-1 (*37*−*42*). In addition, upon phosphorylation, the bZip protein CREB binds with high affinity in vivo to a half-site of the cAMP-responsive element (CRE) (*43*, *44*).

We have explored the structure of the GCN4·half-site complex to determine whether GCN4 binds as a monomer or as a dimer. We have monitored protein·DNA complex formation at DNA sites containing zero, one, or two half-sites by equilibrium electrophoretic mobility shift assays. In addition, we have used the sequence-specific DNA-cleaving molecule Fe·EDTA-GCN4(226−281) (*45*) to map the positions of both amino termini in the dimeric GCN4 bZip domain relative to this DNA binding site. These studies indicate that GCN4 binds as a dimer to DNA sites containing only a single consensus half-site.

We have also investigated the structure of the two basic regions in the GCN4·half-site complex. Previous experiments have shown that a coil → helix transition is induced in the basic region of bZip peptides upon binding specific DNA (*14*−*17*, *19*−*22*). We therefore used circular dichroism (CD) spectroscopy to compare the degree of α-helicity in the GCN4 bZip domain in the absence of DNA and in the presence of AP-1, half-site, and nonspecific DNA duplexes. Our results suggest that a fully folded GCN4 dimer binds to an isolated DNA half-site, making specific contacts in one major groove and nonspecific contacts to neighboring DNA bases.

EXPERIMENTAL PROCEDURES

Materials

Deionized water was further purified with a Corning MegaPure Filtration System. Deproteinized calf thymus DNA was purchased from Sigma and dissolved in TE buffer [10 mM Tris·HCl and 1 mM EDTA (pH 8.4)] to a concentration of 2 mM in base pairs. Polydeoxyinosinic-deoxycytidylic acid (poly[dI-dC]·poly[dI-dC]) was purchased from ICN Pharmaceuticals, Inc., and dissolved in water to a concentration of 2 mM in base pairs. Enzymes were obtained from New England Biolabs and used with the supplied buffers. Deoxyadenosine 5′-[α-^{33}P]triphosphate was purchased from Amersham, and adenosine 5′-[γ-^{32}P]triphosphate was purchased from ICN Pharmaceuticals, Inc. Oligonucleotides were supplied by Genosys Biotechnologies, Inc. Cerenkov

a. **Basic Region**

MKDPAALKRARNTEAARRSRARKLQRM

Leucine Zipper Region

KQLEDKVEELLSKNYHLENEVARLKKLVGER

b.

5′-AGTGGA**GATGACTCATC**TCGTGC-3′
3′-TCACCT**CTACTGAGTAG**AGCACG-5′ **AP-1**

5′-AGTGGA**GATGACGTCATC**TCGTGC-3′
3′-TCACCT**CTACTGCAGTAG**AGCACG-5′ **CRE**

5′-AGTGGA**GATGAC**AGCTACTCGTGC-3′
3′-TCACCT**CTACTG**TCGATGAGCACG-5′ **CREHS**

5′-ATCCTTATCCTTAAGGATAAGGAT-3′
3′-TAGGAATAGGAATTCCTATTCCTA-5′ **NS24**

FIGURE 1: (a) Sequence from the N- to the C-terminus of GCN4-56. (b) Sequences of the oligonucleotides used in the mobility shift assays and CD studies. The consensus half-sites are in bold type, and the arrows indicate the relative orientations of each half-site within the full binding site. The center of the pseudosymmetric AP-1 site is indicated with a black circle. The sequences of the AP-1, CRE, and CREHS oligonucleotides are identical to those used previously (*30*, *55*).

radioactivity was measured with a Packard Tri-Carb 1600TR Liquid Scintillation Analyzer. A Molecular Dynamics 400E PhosphorImager and ImageQuant version 3.3 were used to obtain data from storage phosphor screens purchased from Molecular Dynamics.

Methods

Plasmid Construction. DNA manipulations were carried out by standard methods (*46*). All mutations were made by Kunkel mutagenesis (*47*) or PCR mutagenesis (QuikChange Site-Directed Mutagenesis Kit, Stratagene). The primary sequence of each construct was confirmed by DNA sequencing (*48*). Plasmid pGCN4-56, encoding the basic region and leucine zipper of GCN4 (residues 226−281), was constructed from vector pGG63 (kindly provided by R. V. Talanian and P. S. Kim, unpublished observations). pGG63, derived from the pAED4 cloning vector (*49*), contains the coding sequence for residues 222−281 of GCN4, preceded by a Met residue and followed by C-terminal Gly-Gly-Cys residues. Successive rounds of mutagenesis were used to replace residues 222−225 with a lysine codon for better expression (*50*) and to delete the coding sequence for the three C-terminal residues. The resulting plasmid, pGCN4-56, contains the coding sequence for a 58-amino acid peptide containing the 56 C-terminal residues of GCN4, preceded by codons for Met and Lys (Figure 1a).

Two plasmids containing the DNA binding sites CRE and CREHS were constructed for Fe·EDTA-GCN4(226−281) affinity cleaving experiments. The oligonucleotides 5′-TATGG**ATGACGTCAT**CG-3′ and 5′-GATCCG**ATGACGTCAT**CCA-3′ containing the CRE sequence (in bold) were treated with T4 polynucleotide kinase and subcloned into the multiple cloning site of pGG63. Plasmid pCREHS was subsequently constructed by site-directed mutagenesis. A 186 bp fragment containing each binding site was amplified by PCR with 5′-end-labeled primers and purified by preparative nondenaturing (5% polyacrylamide, 19:1 acrylamide:bis-

acrylamide) gel electrophoresis. Chemical sequencing A and G reactions were carried out as described previously (*51*, *52*).

Protein Expression and Purification. GCN4-56 was overexpressed in *E. coli* strain BL21(DE3) pLysS using the T7 expression system (*53*). Cells derived from a single colony were grown in Luria broth containing 20 μg/mL ampicillin and 25 μg/mL chloramphenicol at 37 °C. Protein expression was induced when A_{590} reached 0.4 by addition of isopropyl 1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 0.4 mM. Cells were harvested after 2 h by centrifugation and lysed by freezing and sonication in 10 mM phosphate, 1 mM EDTA, and 500 mM NaCl (pH 6.9) in the presence of phenylmethanesulfonyl fluoride (PMSF, ca. 50 μM) and DNase I (ca. 2.5 μg/mL). Cell lysates were treated with polyethyleneimine (0.6% v/v), and the precipitated DNA was removed by centrifugation. The protein-containing supernatant was purified by size-exclusion chromatography (Sephadex G-50), followed by reverse-phase HPLC, using a Vydac C_{18} column and a linear water/acetonitrile gradient containing 0.1% trifluoroacetic acid. The protein concentration was determined by UV absorbance in 6 M guanidinium hydrochloride and 0.02 M phosphate (pH 6.5) assuming an extinction coefficient of 1470 M^{-1} cm^{-1} for Tyr at 275 nm (*54*). The molecular mass of the purified protein, as determined by MALDI-TOF mass spectrometry, is well within experimental error of the calculated mass of GCN4-56 (expected $[M + H]^+$, 6875.0 Da; observed, 6875.9 Da).

Quantitative Electrophoretic Mobility Shift Assays. These assays were performed essentially as described by Schepartz and co-workers (*55*), with minor modifications. The sequences of the strands containing DNA binding sites AP-1, CRE, and CREHS are identical to those described previously (*30*, *55*). An additional self-complementary oligonucleotide (NS24) was synthesized containing the noncognate sequence 5′-ATCCTTATCCTTAAGGATAAGGAT-3′. Each oligonucleotide was purified by preparative denaturing (19% polyacrylamide, 19:1 acrylamide:bisacrylamide) gel electrophoresis and 5′-end-labeled by standard procedures (*46*). Radiolabeled oligonucleotides were annealed to their unlabeled complement by heating an equimolar mixture of the two fragments in 100 mM KCl and 10 mM phosphate (pH 7.4) to 90 °C for 2 min and then cooling slowly to room temperature. Double-stranded NS24 was prepared in a similar fashion.

Complex formation was initiated by adding labeled DNA (20 000 cpm) to known concentrations of peptide in PBS binding buffer [137 mM NaCl, 2.7 mM KCl, 4.3 mM sodium phosphate, 1.4 mM potassium phosphate, 1 mM EDTA, 1 mM DTT, 0.1% Igepal CA-630, and 0.4 mg/mL BSA (pH 7.4)] (*55*). The solutions were allowed to equilibrate for 30 min at 4 °C followed by addition of 6× glycerol loading buffer (30% glycerol, 0.25% bromophenol blue, and 0.25% xylene cyanole FF). Each sample was loaded on a nondenaturing 8% polyacrylamide (79:1 acrylamide:bisacrylamide) gel pre-equilibrated in 45 mM Tris-borate and 1 mM EDTA at 300 V and 4 °C. After electrophoresis (1 h), gels were dried and analyzed using storage phosphor technology. Storage phosphor screens were pressed against dried gels and exposed overnight. A Molecular Dynamics 400E PhosphorImager was used to obtain data from the storage screens, and the amounts of free and bound DNA were quantified using ImageQuant version 3.3. The fraction of DNA bound (θ_{app}) was calculated as the volume contained within the band corresponding to the bound DNA divided by the sum of the volume present in the bands corresponding to the bound and free DNAs within each lane. Data were fit to the modified Hill equation (eq 1) describing formation of a 2:1 GCN4·DNA complex:

$$\theta_{app} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{K_a^2[L]_{tot}^2}{1 + K_a^2[L]_{tot}^2} \qquad (1)$$

where $[L]_{tot}$ corresponds to the total protein monomer concentration, K_a corresponds to the apparent monomeric equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied and saturated, respectively. Data were fit using the nonlinear least-squares fitting procedure of KaleidaGraph software (version 3.0.5, Abelbeck Software) running on a Macintosh PowerPC computer with K_a, θ_{max}, and θ_{min} as the adjustable parameters. Data sets that fit eq 1 with an R value of <0.97 were discarded; all individual data points from data sets with an R value of >0.97 were retained. Each dissociation constant was determined from the average of three independent data sets.

The AP1, CRE, and CREHS duplexes contain a single site for specific complex formation. In contrast, we estimate that the 24 bp NS24 duplex contains 13−14 overlapping nonspecific sites (*56*), assuming a binding site size for the GCN4 dimer of 11−12 bp (*5*−*7*). Thus, the observed macroscopic apparent monomeric dissociation constant must be corrected by a statistical factor to derive a microscopic binding constant for nonspecific DNA. Because we report apparent monomeric binding constants, the square of the microscopic monomeric association constant is related to the square of the observed monomeric association constant by this statistical factor. Thus, our apparent monomeric dissociation constant overestimates the affinity of GCN4 for nonspecific DNA by a factor of $13^{1/2}$−$14^{1/2}$, or approximately 3.7-fold.

Affinity Cleavage Reactions. Reaction conditions included 30 mM Tris·HCl, 3 mM sodium acetate, 20 mM NaCl, 5 mM DTT, 200 μM bp calf thymus DNA, 10 μM Fe·EDTA-GCN4(226−281), and 50 000 cpm of labeled DNA (pH 7.9). Fe·EDTA-GCN4(226−281) was generated immediately prior to use by equilibration of 600 μM EDTA-GCN4(226−281) (kindly provided by M. G. Oakley and P. B. Dervan) with an equal volume of 600 μM ferrous ammonium sulfate for 10 min at room temperature. Fe·EDTA-GCN4(226−281) was then diluted to a final concentration of 10 μM and allowed to equilibrate with the DNA for 30 min at room temperature prior to the addition of DTT. Reactions were quenched after 30 min by ethanol precipitation. Cleavage products were analyzed by electrophoresis on 6% polyacrylamide denaturing gels (19:1 acrylamide:bisacrylamide, 3.8 M urea). After electrophoresis, gels were dried and analyzed using storage phosphor technology as described above.

CD Spectroscopy. CD spectra were acquired with a Jasco J-715 spectropolarimeter. Samples were prepared in 50 mM KCl and 12.5 mM potassium phosphate (pH 7.0). To minimize the error in these experiments, a single sample of

GCN4-56 (30 μM) was prepared, and aliquots of this sample were mixed with an equal volume of 15 μM double-stranded AP-1, CREHS, or NS24 DNAs. For each sample, the wavelength dependence of [θ] was monitored at 25 °C in 1 nm increments with a sampling time of 4 s. Spectra were averaged five times and smoothed using KaleidaGraph software (version 3.0.5, Abelbeck Software). The CD spectrum shown for each protein·DNA complex was obtained by subtracting the CD spectrum of the corresponding DNA solution from the CD spectrum of the protein·DNA complex. The self-complementary nonspecifc DNA sample (NS24) was determined to be >90% double-stranded by nondenaturing polyacrylamide gel electrophoresis before CD spectra were recorded. The helix content was calculated by the method of Chen et al. (*57*). We find from repeated measurements that the error in $[\theta]_{222}$ for protein samples prepared from the same concentrated protein stock solution is approximately 1%. We estimate that the error in the concentration determination of the DNA stock solutions is approximately 5−10%.

RESULTS

GCN4-56 contains the 56 C-terminal residues (226−281) of GCN4, including both the leucine zipper dimerization domain and the basic region (Figure 1a) (*45*, *58*). This peptide also contains a Met-Lys dipeptide at the N-terminus for efficient translation initiation of the recombinant protein fragment in *E. coli* (*50*). The resulting 58-residue peptide is the same GCN4 construct that was used for crystallization of the peptide complex with AP-1 DNA (*5*).

We studied the DNA-binding affinity of GCN4-56 for four different DNA sites using a quantitative mobility shift assay. The pseudosymmetric AP-1 sequence, 5′-TGA(C/G)TCA-3′, is the optimal recognition site for GCN4 (*26*−*28*). In addition, GCN4 binds with high affinity to the perfectly symmetric CRE site, 5′-ATGACGTCAT-3′ (*59*). Equivalent protein−DNA contacts are observed in both complexes despite the additional G·C base pair at the center of the CRE site (*5*−*7*). In addition to the AP-1 and CRE sites, we examined the binding affinity of GCN4-56 for an isolated half-site (CREHS) 5′-ATGAC-3′ (*30*), and for a fourth duplex containing no specific half-sites (NS24, Figure 1b). Oligonucleotides containing the AP-1, CRE, and CREHS binding sites are identical to those described previously (*30*, *55*), and the self-complementary nonspecific DNA was carefully designed to avoid overlap with the AP-1 and CRE sites.

Quantitative Electrophoretic Mobility Shift Assays. We measured the equilibrium dissociation constants of the four GCN4-56·DNA complexes using an electrophoretic mobility shift assay (*60*, *61*). For our initial experiments, we monitored protein·DNA complex formation in the presence of approximately physiological concentrations of monovalent salt (137 mM NaCl, 2.7 mM KCl, 4.3 mM sodium phosphate, and 1.4 mM potassium phosphate), but in the absence of nonspecific competitor DNA. The binding isotherms we obtained are shown in Figure 2a. For each DNA duplex, eq 1, describing formation of a 2:1 peptide·DNA complex, provides a good fit to the fraction of DNA bound plotted as a function of total peptide monomer concentration. Calculated values for the apparent monomeric equilibrium dissociation constants are shown in Table 1. GCN4-56 binds AP-1 DNA with a dissociation constant of (2.2 ± 0.9) × 10^{-8} M, in agreement with previously published values (*16*, *55*, *62*, *63*). There appears to be little difference in the affinity of GCN4-

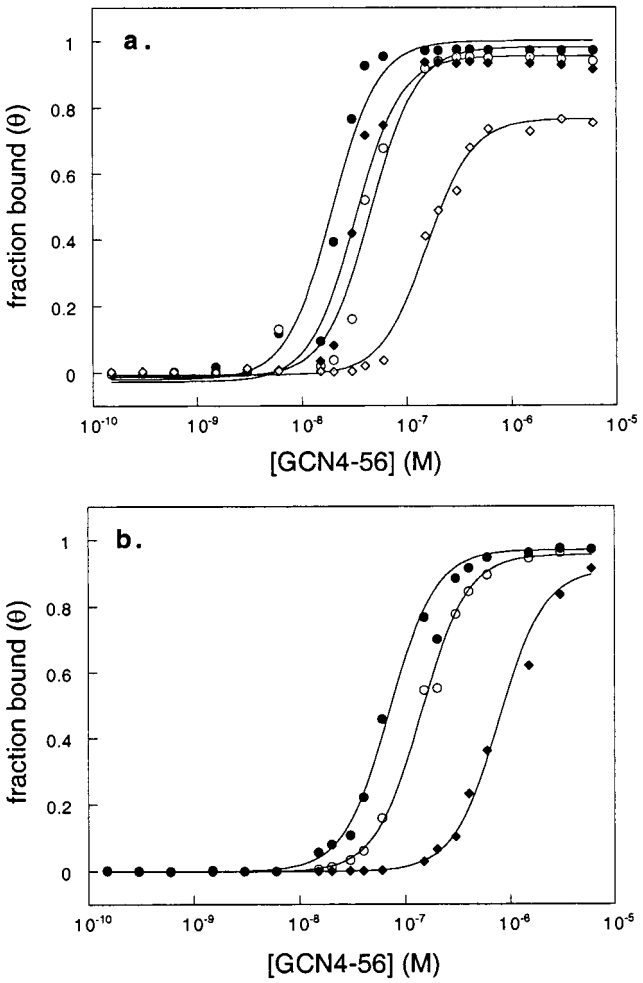

FIGURE 2: (a) Equilibrium binding isotherms for GCN4-56 binding to AP-1 (●), CRE (○), CREHS (◆), and nonspecific DNA (◇) in PBS buffer. (b) Equilibrium binding isotherms for GCN4-56 binding to AP-1 (●), CRE (○), and CREHS (◆) in PBS buffer supplemented with calf thymus DNA (100 μM in base pairs). Protein concentrations for both sets of experiments are 6 μM, 3 μM, 1.5 μM, 600 nM, 400 nM, 300 nM, 200 nM, 150 nM, 60 nM, 40 nM, 30 nM, 20 nM, 15 nM, 6 nM, 3 nM, 1.5 nM, 600 pM, 300 pM, and 150 pM. Each data point represents the average of the fraction bound (θ) values at each concentration of GCN4-56 from three data sets, and the binding isotherms were obtained using eq 1 and the mean K_a values for each data set.

Table 1: Binding Constants for GCN4-56 Binding to Four Different DNA Sites Measured by an Electrophoretic Mobility Shift Assay

	K_d (M)[a]	
DNA-binding site	without excess nonspecific DNA	with excess nonspecific DNA
AP-1	(2.2 ± 0.9) × 10^{-8}	(9 ± 6) × 10^{-8}
CRE	(5 ± 3) × 10^{-8}	(1.4 ± 0.3) × 10^{-7}
CREHS	(3 ± 1) × 10^{-8}	(8 ± 2) × 10^{-7}
NS24	(1.6 ± 0.5) × 10^{-7} [b]	

[a] Values reported are the mean values ± the standard deviation in the values measured from three individual data sets. [b] The macroscopic binding constant reported for GCN4-56 binding to NS24 DNA overestimates the microscopic binding constant for a single nonspecific site by a statistical factor of ca. 3.7 (see the text).

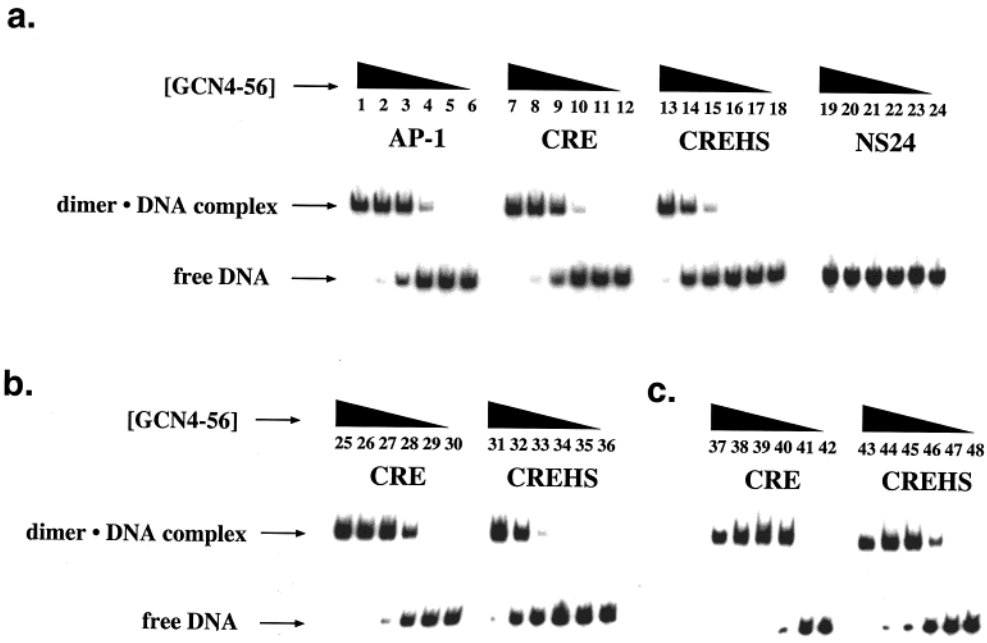

FIGURE 3: (a) Equilibrium binding of GCN4-56 to four different DNA sequences in PBS buffer in the presence of poly[dI·dC] (100 μM in base pairs): AP-1, lanes 1−6; CRE, lanes 7−12; CREHS, lanes 13−18; and nonspecific DNA (NS24), lanes 19−24. (b) Equilibrium binding of GCN4-56 to the CRE and CREHS sequences in PBS buffer in the presence of calf thymus DNA (100 μM in base pairs): CRE, lanes 25−30; and CREHS, lanes 31−36. (c) Equilibrium binding of GCN4-56 to the CRE and CREHS sequences in PBS buffer in which the salt concentration has been doubled (274 mM NaCl, 5.4 mM KCl, 8.6 mM sodium phosphate, and 2.8 mM potassium phosphate): CRE, lanes 37−42; and CREHS, lanes 43−48. Protein concentrations are 6 μM (lanes 1, 7, 13, 19, 25, 31, 37, and 43), 1.2 μM (lanes 2, 8, 14, 20, 26, 32, 38, and 44), 240 nM (lanes 3, 9, 15, 21, 27, 33, 39, and 45), 48 nM (lanes 4, 10, 16, 22, 28, 34, 40, and 46), and 9.6 nM (lanes 5, 11, 17, 23, 29, 35, 41, and 47). Lanes 6, 12, 18, 24, 30, 36, 42, and 48 contained ^{32}P-end-labeled DNA only.

56 for the AP-1, CRE, and CREHS sites. This result is surprising, as the CREHS site does not contain the full 5′ to 3′ GCN4 recognition sequence. Similarly, we observe binding to the nonspecific duplex at submicromolar protein concentrations. Unlike the other three DNA duplexes, in which a single specific binding site for GCN4-56 is expected, NS24 contains 13−14 nonspecific binding sites, assuming a binding site size of 11−12 bp for the GCN4 dimer (*5*−*7*, *56*). The observed macroscopic binding constant for NS24 is therefore an overestimate of the microscopic binding constant for a single nonspecific site. After a correction for this statistical factor (see Experimental Procedures), we estimate a K_d(nonspecific) of approximately 6×10^{-7} M. Thus, we observe only a 30-fold lower affinity for completely nonspecific DNA relative to the optimal AP-1 site under these conditions.

Protein·DNA Complex Formation in the Presence of Excess Nonspecific DNA. One explanation for this apparent lack of specificity is that the conditions of our assay are not sufficiently stringent to allow GCN4-56 to distinguish among specific and nonspecific sites. We therefore examined DNA binding by GCN4-56 in the presence of excess nonspecific DNA (poly[dI·dC] or calf thymus DNA, 100 μM in base pairs). As expected, only specific protein·DNA complexes were observed under these conditions (Figure 3a). No complex formation was observed with the NS24 duplex even at the highest protein concentration (6 μM) (Figure 3a). Furthermore, GCN4-56 can now discriminate between the consensus sequences and CREHS, which contains only a single consensus half-site (Figure 3a,b). The binding isotherms we obtained for these three sites in the presence of calf thymus DNA are shown in Figure 2b. Calculated values for the apparent monomeric equilibrium dissociation constants are shown in Table 1. The affinity of GCN4-56 for the consensus AP-1 and CRE sites decreases slightly under these conditions (3−4-fold). In contrast, the affinity of GCN4-56 for the CREHS site decreases by a factor of ∼30, leading to an approximately 10-fold difference in the affinities of GCN4-56 for the AP-1 and CREHS sites. Thus, while high-affinity binding by GCN4 to the CREHS site is still observed in the presence of excess nonspecific DNA, the full recognition sequence is preferred.

Both Coulombic and hydrophobic interactions play an important role in the formation of specific protein·DNA complexes (*64*, *65*), while the burial of nonpolar surface area appears to be less important for the formation of nonspecific complexes (*66*). Therefore, we would expect higher salt concentrations to destabilize nonspecific protein·DNA complex formation preferentially. When the ionic strength of the buffer is doubled (274 mM NaCl, 5.4 mM KCl, 8.6 mM sodium phosphate, and 2.8 mM potassium phosphate), the affinity of GCN4-56 for the CREHS site appears to be somewhat reduced relative to the full CRE site (Figure 3c). However, GCN4-56·CREHS complex formation is observed at lower protein concentrations under these conditions than in the presence of excess nonspecific DNA (compare lanes 15 and 16 of Figure 3a, lanes 33 and 34 of Figure 3b with lanes 45 and 46 of Figure 3c).

Affinity Cleaving Assays with Fe·EDTA-GCN4. Data from the electrophoretic mobility shift assays described above are consistent with dimeric binding by GCN4-56 not only to the AP-1 and CRE sites but also to the CREHS site. The electrophoretic mobility of the GCN4-56·CREHS complex is similar to that of the GCN4-56·AP-1 and GCN4-56·CRE complexes (Figure 3), to which GCN4 clearly binds as a dimer (*5*−*7*). Similarly, the shape of the binding isotherms

for the CREHS complex does not differ from the shapes of those for the AP-1 and CRE complexes (Figure 2), again suggesting that the protein binds all three sites as a dimer. However, recent evidence indicates that specific DNA binding by GCN4 monomers occurs, at least as an intermediate, on the pathway to dimer·DNA complex formation (*30*−*33*). These results suggest that it would be possible for a GCN4 monomer to bind the CREHS site in a sequence-specific manner, resulting in the observed affinity. To distinguish more clearly between possible monomeric and dimeric binding modes for the GCN4·CREHS complex, we have used an affinity cleaving assay.

A sequence-specific DNA-binding molecule can be converted into a sequence-specific DNA-cleaving molecule by the attachment of the nonspecific DNA-cleaving moiety, Fe·EDTA. In the presence of a reducing agent such as dithiothreitol (DTT), Fe·EDTA localized at a specific DNA-binding site cleaves both DNA strands via a diffusible, non-sequence-specific radical species (*67*−*71*). The cleavage pattern generated by a protein appended with Fe·EDTA provides information about the orientation of the bound protein and the position, relative to the DNA binding site, of the amino acid residue to which Fe·EDTA is attached (*45*, *69*−*73*). Thus, affinity cleaving is a sensitive probe of changes in the structure of a protein−DNA complex.

Fe·EDTA has been attached to the N-terminus of the GCN4 DNA-binding domain to produce the DNA-cleaving protein, Fe·EDTA-GCN4(226−281) (*45*). Reaction of the Fe·EDTA-GCN4(226−281) dimer with a DNA fragment containing AP-1-like binding sites generates a distinct tripartite cleavage pattern, resulting from the placement of *two* Fe·EDTA moieties in successive major grooves of the DNA binding site (*45*). In contrast, a *bipartite* affinity cleaving pattern, resulting from a single Fe·EDTA species, would be expected if GCN4 binds to DNA as a monomer.

Such a cleaving pattern has been observed previously for Jun-Fos heterodimers in which Fe·EDTA has been appended to only one monomer (*73*). When the orientation of the Jun-Fos heterodimer on the DNA is fixed, the Fe·EDTA-Jun-Fos and the Jun-Fe·EDTA-Fos heterodimers generate bipartite cleavage patterns, resulting from a single Fe·EDTA moiety bound in the major groove of the DNA (*73*). The cleavage sites are localized asymmetrically with respect to the binding site, reflecting the orientation of the Jun-Fos heterodimer in the protein·DNA complex (*73*). Thus, if Fe·EDTA-GCN4(226−281) binds to CREHS as a dimer, a tripartite cleavage pattern centered around the consensus half-site and the following nonconsensus base pairs would be expected. If GCN4-56 binds as a monomer, a bipartite pattern localized distinctly around the consensus half-site would be observed.

Affinity cleaving assays were performed on DNA fragments containing the CRE and CREHS sites. As expected, Fe·EDTA-GCN4(226−281) cleavage of the perfectly symmetric CRE site results in a tripartite pattern very similar to the one observed with AP-1-like DNA (Figure 4a,b,e) (*45*). When Fe·EDTA-GCN4(226−281) is allowed to react with CREHS DNA (Figure 4c,d), the extent of cleavage is substantially lower than that observed at the full CRE site (Figure 4a,b), reflecting the reduced binding affinity of GCN4-56 for the CREHS site in the presence of high concentrations of calf thymus DNA (200 μM bp). Nonetheless, a very similar tripartite pattern is observed for both DNA sites, suggesting that GCN4 binds to CREHS as a dimer under these conditions (Figure 4e). Importantly, the spacing of the three regions of cleavage on each strand is similar for both sites, indicating that the number of bases in contact with the protein dimer is approximately the same in each case. In addition, the ratios of cleavage along each strand are similar for both sites. Along the top strand of the CRE site, as viewed in Figure 4e, the ratio of cleavage intensities for each region in the tripartite pattern is 1.5:2:1. For the CREHS site, the ratio is 1.4:1.6:1. Importantly, in each case, the two outer loci resulting from cleavage due to a single Fe·EDTA moiety on each GCN4 monomer are of roughly equal intensity. The monomer making nonspecific contacts appears to bind exclusively to the right of the CREHS site, as would be expected for a dimer in which the orientation is fixed by specific contacts with the consensus half-site.

CD Spectroscopy. Equilibrium mobility shift analysis and affinity cleaving experiments suggest that GCN4-56 binds CREHS as a dimer that makes specific contacts to the sequence 5′-ATGAC-3′ and nonspecific contacts to adjacent DNA base pairs. While we expected the basic region making specific contacts to be folded in the presence of DNA, it was not clear if the basic region making nonspecific contacts would be α-helical or unstructured. We therefore used circular dichroism (CD) spectroscopy to measure the degree of α-helicity in GCN4-56 in the absence of DNA and in the presence of AP-1, CREHS, and nonspecific DNAs. In the absence of DNA, GCN4−56 is ca. 80% α-helical at 25 °C based on a mean residue ellipticity of −30100 deg cm^2 dmol^{-1} at 222 nm (Figure 5) (*57*). This is consistent with a helical leucine zipper domain and a largely unstructured basic region, as others have observed (*14*−*19*). In the presence of AP-1 DNA, the basic region folds into an α-helical conformation, and the apparent helix content of GCN4 increases to ca. 97% (Figure 5), in accord with previously published results (*15*−*17*). A change in the CD spectrum of GCN4-56 also occurs in the presence of CREHS DNA; the CD signal at 222 nm indicates that the protein is ca. 94% helical in this complex (Figure 5). Finally, an increase in the magnitude of the signal at 222 nm is also observed upon addition of NS24 to the extent that GCN4-56 appears to be fully helical in this nonspecific complex (Figure 5).

The CD difference spectra shown in Figure 5 were obtained by subtracting the CD spectrum of the corresponding DNA solution from the CD spectrum of the protein·DNA complex. The DNA spectra for all three DNA duplexes are consistent with B-form DNA, with a minimum at ca. 247 nm and a maximum at ca. 221 nm. Because there is no change upon complex formation in the CD spectrum at 247 nm, where the protein contributes little, if any, to the observed signal, it is reasonable to assume that the DNA conformation is B-form in all three complexes and that the increased magnitude of the signal at 222 nm is due to a helix → coil transition in the basic region of GCN4-56.

The change in $[\theta]_{222}$ upon addition of DNA is significant. A single sample of GCN4-56 was prepared for these experiments, and aliquots of this sample were mixed with each of the three double-stranded DNAs (see Experimental Procedures). The concentration of protein in each sample is therefore the same, and any change in the signal at 222 nm

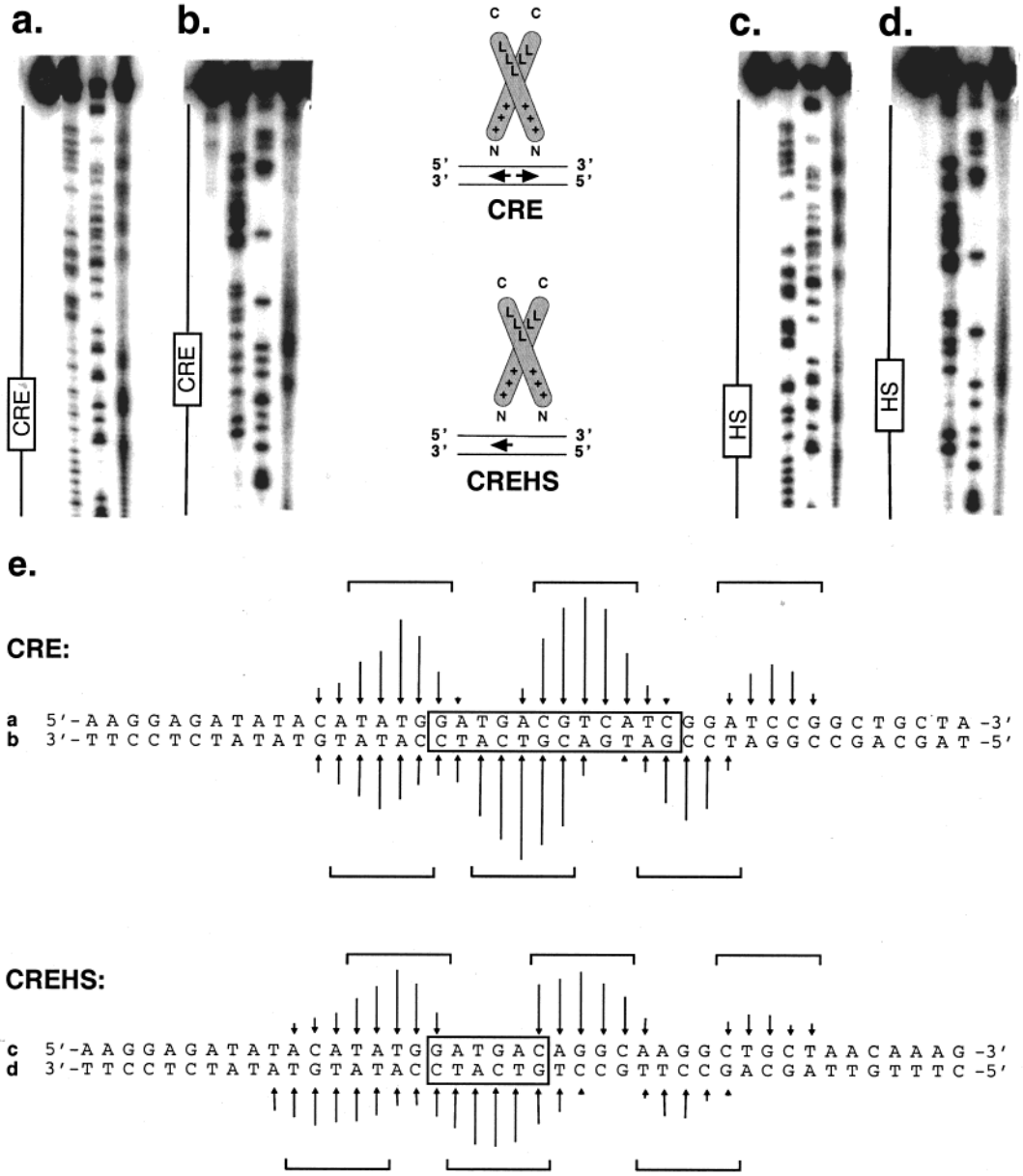

FIGURE 4: Fe•EDTA-GCN4(226−281) affinity cleaving reactions in the presence of CRE (a and b) and CREHS (c and d) DNA. (a−d) Phosphorimages of denaturing polyacrylamide gels: (left) intact DNA control lane, (center) A- and G-specific chemical sequencing lanes, and (right) Fe•EDTA-GCN4(226−281) cleavage at a concentration of 10 μM. The top strands, as shown in panel e, are labeled in panels a and c; the bottom strands are labeled in panels b and d. (e) DNA sequences surrounding each of the binding sites in a−d. The arrows represent the extent of cleavage at the indicated base position. All DNA strands are 5′-end-labeled; thus, the DNA sequence from 5′ to 3′ can be read from the bottom to the top in panels a−d. The cleavage intensities at the bracketed base positions were summed to give the ratios of cleavage intensities along each strand (see the Discussion).

is a result of a conformational change in the complex. CD and NMR data in the presence of AP-1 DNA suggest that this conformational change corresponds to a helix → coil transition in the basic region of GCN4 upon DNA binding (*14*−*19*, *74*).

The extent of the change in $[\theta]_{222}$ upon DNA binding appears to differ slightly for the three DNA duplexes (Figure 5), although it is not clear if these differences are outside the expected error associated with these experiments (see Experimental Procedures). The apparent magnitude of the change in helical content observed for GCN4-56 upon binding to the CREHS site is somewhat smaller than that observed for binding to the AP-1 site. On the basis of these data, we cannot exclude the possibility that this difference reflects a lower degree of helix formation in the GCN4-56•CREHS complex, presumably due to the inclusion of a nonconsensus half-site. However, we observe the largest apparent change in $[\theta]_{222}$ when GCN4-56 is added to nonspecific DNA (NS24). The simplest interpretation of these results is that a helix → coil transition occurs in both basic regions of the GCN4-56 dimer whether specific or nonspecific DNA sequences are bound.

DISCUSSION

We have examined the interaction of the GCN4 DNA-binding domain with DNAs containing zero, one, or two consensus half-sites. In the absence of excess nonspecific DNA, there is very little difference in the affinity of GCN4-56 for the four DNAs that were studied, with only a ca. 30-fold difference in the equilibrium binding constants for the

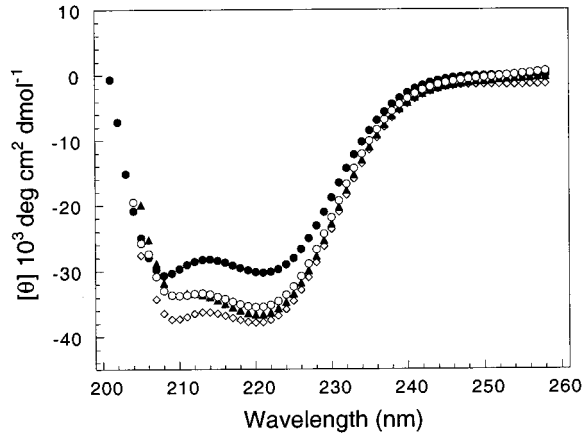

FIGURE 5: CD spectra of 15 μM GCN4-56 in the absence (●) and presence of 7.5 μM AP-1 (▲), CREHS (○), and nonspecific DNA (NS24) (◇) in 50 mM KCl and 12.5 mM potassium phosphate at pH 7.0 and 25 °C. The DNA signal has been subtracted from each spectrum to compare the protein structures in the presence and absence of DNA. Values of $[\theta]_{222}$ (deg cm^2 dmol^{-1}) are −30100 for GCN4-56 alone, −36500 for GCN4-56 bound to AP-1 DNA, −35500 for GCN4-56 bound to CREHS DNA, and −37700 for GCN4-56 bound to nonspecific DNA. These values correspond to folded peptides that are 80, 97, 94, and 100% helical, respectively (*57*). In no case did adding more DNA increase the ellipticity of the peptide.

optimal AP-1 site and completely nonspecific DNA (Figure 2a and Table 1). Under these conditions, GCN4 fails to discriminate among the AP-1, CRE, and CREHS sites, even though specific contacts can be made by both monomers only in the AP-1 and CRE complexes. This apparent lack of selectivity is unexpected. For transcription factors such as GCN4, the difference between the equilibrium dissociation constant for specific binding and the average dissociation constant for nonspecific binding must be large (a factor of ≫30) to ensure specific DNA binding in vivo (*75*).

We therefore decided to investigate DNA binding by GCN4-56 under more stringent conditions. In the presence of excess, nonspecific DNA, only specific complexes are observed (Figure 3a,b). Under these conditions, GCN4-56 binds with higher affinity to the AP-1 and CRE sites than to the CREHS site. Nonetheless, high-affinity, specific binding is observed for the CREHS site (Figure 2b and Table 1). As a transcription factor must bind to its cognate site in the presence of a large number of possible nonspecific binding sites, the inclusion of excess nonspecific DNA more closely approximates in vivo binding conditions.

High-affinity, specific binding by GCN4-56 to the CREHS site is surprising, as specific DNA contacts can be made by only one basic region in the GCN4-56·CREHS complex. One possible explanation for these observations is that monomeric GCN4-56 is able to bind with high affinity to a consensus half-site. Recently, footprinting assays have shown that monomeric Jun basic regions can protect the DNA binding site of the Jun-Fos heterodimer from cleavage by DNase I (*33*), and it has been demonstrated that some bZip peptides can bind their DNA sites sequentially as monomers that subsequently dimerize on the DNA (*30*, *32*). These data suggest that it would be possible for a GCN4 monomer to bind the 5′-ATGAC-3′ site in a sequence-specific manner, resulting in the observed affinity for the CREHS site. In fact, Bosshard and co-workers have reported that unfolded GCN4 monomers are able to bind DNA specifically as intermediates on the pathway to dimeric bZip·DNA complex formation (*31*).

However, our data suggest that GCN4-56 binds to the CREHS site as a dimer and that the resulting protein·DNA complex is very similar in overall structure and orientation to the well-characterized GCN4·AP-1 and GCN4·CRE complexes. Equilibrium mobility shift analyses are consistent with dimeric binding by GCN4-56 to all four of the duplexes we tested (Figures 2 and 3). Similarly, affinity cleaving assays suggest that GCN4-56 binds as a dimer to CREHS (Figure 4). The tripartite cleavage pattern observed when Fe·EDTA-GCN4(226−281) reacts with CREHS in the presence of DTT is indicative of two basic regions bound in successive DNA major grooves (*45*). As only one basic region can make specific contacts with the consensus half-site, the other basic region must make nonspecific contacts to neighboring DNA bases. Finally, CD experiments demonstrate that both GCN4 monomers are largely helical when bound to CREHS DNA (Figure 5), suggesting that bZip basic regions undergo a random coil → α-helix folding transition to bind both specific and nonspecific DNA.

Indeed, GCN4-56 is largely helical when bound to nonspecific (NS24) DNA. Several groups have reported that the addition of nonspecific DNA induces an increase in the helicity of the GCN4 basic region (*16*−*18*, *21*, *22*). However, the extent of this folding transition is quite variable depending on the specific DNA sequence and structure, the experimental conditions, and the GCN4 derivative.[2] Because the nonspecific DNA site (NS24) is fully saturated at micromolar concentrations of GCN4-56 in the absence of calf thymus DNA (Figure 2), we expect full complex formation under the conditions used in our CD experiment. The resulting data indicate that helix formation correlates well with protein·DNA complex formation, even when nonspecific DNA sequences are bound.

A folding transition in the presence of nonspecific DNA has also been observed with model peptides. Both double-stranded DNA oligonucleotides and mononucleosomal calf thymus DNA promote random coil → α-helix transitions in short peptides that can form positively charged, amphipathic helices (*76*, *77*). von Hippel and co-workers have suggested that the induction of α-helix formation by nonspecific DNA is thermodynamically important for bZip and related transcription factors for at least two reasons (*76*). First, the specificity of DNA binding would be reduced if the free energy diverted to helix formation were only required for specific binding. In addition, the use of a flexible domain for complexation of nonspecific DNA would be expected to allow the bZip motif to avoid unfavorable contacts arising at nonspecific sites, again reducing the specificity of DNA binding. Record and co-workers have provided support for the first argument. They have observed that extended, 17-residue tetracationic peptides bind nonspecifically to DNA with lower affinity than do compact, 5-residue tetracationic peptides. This result suggests an energetic penalty for helix formation and other conformational changes required for the 17-residue peptides to bind to DNA (*77*). Finally, NMR

[2] Perhaps the presence of the helix-capping residues Asp and Pro at the N-terminus of GCN4-56 contributes to the degree of helicity we observe in the presence of random sequence DNA.

studies have demonstrated that the basic region of GCN4 has partial, but highly dynamic, helical character in the absence of DNA (*18*, *74*). Thus, it is not surprising that nonspecific DNA appears to complete the transition to an α-helical conformation in this region.

The optimal DNA binding site for the intact GCN4 protein (*27*) as well as the isolated GCN4 DNA-binding domain (*28*, *78*) is the 7 bp symmetric sequence 5′-TGA(C/G)TCA-3′. Interestingly, data from random-sequence selection experiments suggest that one-half of the asymmetrical binding site is more tolerant to mutation than the other half of the site. The majority of deviations from the 7 bp core are observed to the right of the central base at positions 1 and 2 (*27*, *28*, *78*). These results are consistent with our observation that, although two symmetrical half-sites are required for optimal DNA binding by GCN4, high-affinity binding may nonetheless be achieved with a DNA site containing only the sequence 5′-TGAC-3′.

Thermodynamic studies of DNA binding by the Lac repressor and by *Eco*RI endonuclease demonstrate that the first mutation in a DNA half-site can lead to a large decrease in the stability of a protein·DNA complex (*79*−*81*), while a second mutation in each half-site is less detrimental than the first (*66*). In fact, mutation of two base pairs in each half-site can lead to an operator site that binds Lac repressor with essentially the same affinity as nonspecific DNA (*66*). Thus, complete conversion of a half-site to a noncognate sequence can be no more deleterious than a 2 bp substitution. Given that one of the two-half-sites in the consensus AP-1 site appears to be more important for GCN4 binding, multiple substitutions in the second half-site may have only a modest effect on complex stability.

Dimeric DNA-binding proteins, in general, bind most tightly to specific palindromic target sequences. However, high-affinity binding to a single consensus half-site is not unprecedented. For example, the *E. coli* LexA repressor protein protects an isolated operator half-site from DNase I cleavage (*35*). Unlike GCN4, LexA binds to a single half-site as a monomer, and protection of nonoperator sequences is not observed (*35*). Significantly, LexA is primarily a monomer in solution at physiologically relevant concentrations (*82*, *83*), and it appears that nonconsensus base pairs are not sufficient to promote dimerization in this case.

In contrast, a crystal structure has been determined of the DNA-binding domain of the glucocorticoid receptor in which one monomer makes specific contacts with a consensus half-site of the glucocorticoid response element (GRE), while the other monomer interacts nonspecifically with a noncognate element (*36*). Remarkably, the DNA site used for the structure determination contains two consensus half-sites, but the spacing between them is not optimal. In this case, then, although the glucocorticoid receptor is primarily monomeric in solution (*84*, *85*), maintenance of the appropriate protein−protein contacts appears to be more important for complex stability than specific interactions by both monomers with their DNA half-sites. GCN4, which can bind both the AP-1 and CRE sites, is more flexible with regard to spacing between half-sites than is the glucocorticoid receptor. Nonetheless, when GCN4-56 binds to a DNA sequence containing only a single consensus half-site, it appears to form a complex similar to that observed in the glucocorticoid receptor structure, in which one monomer makes specific contacts, while the other contacts nonconsensus base pairs in a position determined by the dimer interface.

Half-site recognition by bZip proteins may be biologically significant. Several GCN4- and AP-1-responsive promoters have binding sites that contain only one-half of the consensus core sequence (*37*−*42*). In addition, binding by the bZip transcription factor CREB to an isolated half-site occurs in a phosphorylation-dependent manner. The cAMP-responsive element (CRE)-binding protein (CREB) recognizes the symmetric CRE sequence with high affinity in vivo. However, weak binding to sites containing only one-half of the CRE element, 5′-CGTCA-3′, has been observed (*43*, *44*, *86*). Interestingly, protein binding to this site in the tyrosine aminotransferase (TAT) gene is phosphorylation-dependent. The level of CREB binding to TATCRE increases upon cAMP stimulation (*43*, *44*). However, phosphorylation has no effect on the binding affinity of CREB for the symmetric CRE site in the phosphoenolpyruvate carboxykinase (PEPCK) gene (*44*). These results suggest that half-site binding may play a role in the regulation of gene activation in vivo.

ACKNOWLEDGMENT

We thank P. Dervan for the gift of EDTA-GCN4(226-281), R. Talanian and P. Kim for plasmid pGG63, S. Sharrow and S. Kan for assistance with mass spectrometry, and J. Richardson, T. Widlanski, and L. Kiessling for careful reading of the manuscript.

REFERENCES

1. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science 240*, 1759.
2. Kerppola, T. K., and Curran, T. (1991) *Curr. Opin. Struct. Biol. 1*, 71.
3. Hu, J. C., and Sauer, R. T. (1992) *Nucleic Acids Mol. Biol. 6*, 82.
4. Hurst, H. C. (1995) *Protein Profile 2*, 105.
5. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223.
6. König, P., and Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139.
7. Keller, W., König, P., and Richmond, T. J. (1995) *J. Mol. Biol. 254*, 657.
8. Gentz, R., Rauscher, F. J., Abate, C., and Curran, T. (1989) *Science 243*, 1695.
9. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1989) *Science 243*, 1681.
10. Ransone, L. J., Visvader, J., Wamsley, P., and Verma, I. M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3806.
11. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538.
12. O'Shea, E. K., Rutkowski, R., Stafford, W. F., III, and Kim, P. S. (1989) *Science 245*, 646.
13. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science 254*, 539.
14. Talanian, R. V., McKnight, C. J., and Kim, P. S. (1990) *Science 249*, 769.
15. Weiss, M. A. (1990) *Biochemistry 29*, 8020.
16. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) *Nature 347*, 575.
17. O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990) *Science 249*, 774.
18. Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991) *Biochemistry 30*, 1310.
19. Cuenoud, B., and Schepartz, A. (1993) *Science 259*, 510.
20. Patel, L., Abate, C., and Curran, T. (1990) *Nature 347*, 572.
21. O'Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991) *Biochemistry 30*, 9030.
22. Talanian, R. V., McKnight, C. J., Rutkowski, R., and Kim, P. S. (1992) *Biochemistry 31*, 6871.

23. Ransone, L. J., Visvader, J., Sassone-Corsi, P., and Verma, I. M. (1989) *Genes Dev. 3*, 770.
24. Neuberg, M., Adamkiewicz, J., Hunter, J. B., and Müller, R. (1989) *Nature 341*, 243.
25. Turner, R., and Tjian, R. (1989) *Science 243*, 1689.
26. Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986) *Science 234*, 451.
27. Oliphant, A. R., Brandl, C. J., and Struhl, K. (1989) *Mol. Cell. Biol. 9*, 2944.
28. Mavrothalassitis, G., Beal, G., and Papas, T. S. (1990) *DNA Cell Biol. 9*, 783.
29. Hope, I. A., and Struhl, K. (1987) *EMBO J. 6*, 2781.
30. Metallo, S. J., and Schepartz, A. (1997) *Nat. Struct. Biol. 4*, 115.
31. Berger, C., Piubelli, L., Haditsch, U., and Bosshard, H. R. (1998) *FEBS Lett. 425*, 14.
32. Kohler, J. J., Metallo, S. J., Schneider, T. L., and Schepartz, A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11735.
33. Park, C., Campbell, J. L., and Goddard, W. A., III (1996) *J. Am. Chem. Soc. 118*, 4235.
34. Zondlo, N. J., and Schepartz, A. (1999) *J. Am. Chem. Soc. 121*, 6938.
35. Kim, B., and Little, J. W. (1992) *Science 255*, 203.
36. Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991) *Nature 352*, 497.
37. Arndt, K., and Fink, G. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8516.
38. Angel, P., Hattori, K., Smeal, T., and Karin, M. (1988) *Cell 55*, 875.
39. Franza, B. R., Jr., Rauscher, F. J., III, Josephs, S. F., and Curran, T. (1988) *Science 239*, 1150.
40. Spandidos, D. A., Yiagnisis, M., and Pintzas, A. (1989) *Anticancer Res. 9*, 383.
41. Pearson, B. E., Nasheuer, H.-P., and Wang, T. S.-F. (1991) *Mol. Cell. Biol. 11*, 2081.
42. Kim, S.-J., Denhez, F., Kim, K. Y., Holt, J. T., Sporn, M. B., and Roberts, A. B. (1989) *J. Biol. Chem. 264*, 19373.
43. Weih, F., Stewart, A. F., Boshart, M., Nitsch, D., and Schütz, G. (1990) *Genes Dev. 4*, 1437.
44. Nichols, M., Weih, F., Schmid, W., DeVack, C., Kowenz-Leutz, E., Luckow, B., Boshart, M., and Schütz, G. (1992) *EMBO J. 11*, 3337.
45. Oakley, M. G., and Dervan, P. B. (1990) *Science 248*, 847.
46. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
47. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367.
48. Sanger, F. S. N., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
49. Doering, D. S. (1992) Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
50. Looman, A. C., Bodlaender, J., Comstock, L. J., Eaton, D., Jhurani, P., de Boer, H. A., and van Knippenberg, P. H. (1987) *EMBO J. 6*, 2489.
51. Iverson, B., and Dervan, P. (1987) *Nucleic Acids Res. 15*, 7823.
52. Maxam, A., and Gilbert, W. (1980) *Methods Enzymol. 65*, 499.
53. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60.
54. Edelhoch, H. (1967) *Biochemistry 6*, 1948.
55. Metallo, S. J., and Schepartz, A. (1994) *Chem. Biol. 1*, 143.
56. McGhee, J. D., and von Hippel, P. H. (1974) *J. Mol. Biol. 86*, 469.
57. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) *Biochemistry 13*, 3350.
58. Hope, I. A., and Struhl, K. (1986) *Cell 46*, 885.
59. Sellers, J. W., Vincent, A. C., and Struhl, K. (1990) *Mol. Cell. Biol. 10*, 5077.
60. Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505.
61. Garner, M. M., and Revzin, A. (1981) *Nucleic Acids Res. 9*, 3047.
62. Hope, I. A., and Struhl, K. (1985) *Cell 43*, 177.
63. Berger, C., Jelesarov, I., and Bosshard, H. R. (1996) *Biochemistry 35*, 14984.
64. Record, M. T., Jr., Ha, J.-H., and Fisher, M. A. (1991) *Methods Enzymol. 208*, 291.
65. Spolar, R. S., and Record, M. T., Jr. (1994) *Science 263*, 777.
66. Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M., and Record, M. T., Jr. (1997) *J. Mol. Biol. 267*, 1186.
67. Hertzberg, R. P., and Dervan, P. B. (1984) *Biochemistry 23*, 3934.
68. Taylor, J., Schultz, P., and Dervan, P. (1984) *Tetrahedron 40*, 457.
69. Sluka, J. P., Griffin, J. H., Mack, D. P., and Dervan, P. B. (1990) *J. Am. Chem. Soc. 112*, 6369.
70. Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., and Dervan, P. B. (1990) *Biochemistry 29*, 6551.
71. Griffin, L. C., and Dervan, P. B. (1989) *Science 245*, 967.
72. Mack, D. P., Sluka, J. P., Shin, J. A., Griffin, J. H., Simon, M. I., and Dervan, P. B. (1990) *Biochemistry 29*, 6561.
73. Chen, L., Oakley, M. G., Glover, J. N. M., Jain, J., Dervan, P. B., Hogan, P. G., Rao, A., and Verdine, G. L. (1995) *Curr. Biol. 5*, 882.
74. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) *J. Mol. Biol. 285*, 2133.
75. Kodadek, T. (1995) *Chem. Biol. 2*, 267.
76. Johnson, N. P., Lindstrom, J., Baase, W. A., and von Hippel, P. H. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4840.
77. Padmanabhan, S., Zhang, W., Capp, M. W., Anderson, C. F., and Record, M. T., Jr. (1997) *Biochemistry 36*, 5193.
78. Lee, Y.-s., and Oakley, M. G. Unpublished results.
79. Mossing, M. C., and Record, M. T., Jr. (1985) *J. Mol. Biol. 186*, 295.
80. Lesser, D. R., Kurpiewski, M. R., and Jen-Jacobsen, L. (1990) *Science 250*, 776.
81. Jen-Jacobsen, L. (1995) *Methods Enzymol. 259*, 305.
82. Schnarr, M., Pouyet, J., Granger-Schnarr, M., and Duane, M. (1985) *Biochemistry 24*, 2812.
83. Schnarr, M., Granger-Schnarr, M., Hurstel, S., and Pouyet, J. (1988) *FEBS Lett. 234*, 56.
84. Freedman, L., Yamamoto, K., Luisi, B. F., and Sigler, P. (1988) *Cell 54*, 444.
85. Härd, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J-Å., and Kaptein, R. (1990) *Science 249*, 157.
86. Boshart, M., Weih, F., Schmidt, A., Fournier, R. E. K., and Schütz, G. (1990) *Cell 61*, 905.

BI992705N