A GCN4 Variant with a *C*-Terminal Basic Region Binds to DNA with Wild-Type Affinity†

Jessica J. Hollenbeck, Daniel G. Gurnon, Gia C. Fazio, Jennifer J. Carlson, and Martha G. Oakley*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received May 29, 2001; Revised Manuscript Received September 7, 2001

ABSTRACT: Basic-region leucine zipper (bZip) proteins contain a bipartite DNA-binding motif consisting of a leucine zipper dimerization domain and a basic region that directly contacts DNA. In all naturally occurring bZip proteins, the basic region is positioned *N*-terminal to the leucine zipper. We have designed a series of model bZip peptides in which the basic region of the yeast transcriptional activator GCN4 is placed *C*-terminal to its leucine zipper. DNA-binding studies demonstrate that the optimal *reverse* GCN4 (rGCN4) peptide is able to bind specifically and with wild-type affinity to DNA despite this unnatural arrangement of the two subdomains. These results suggest that a thermodynamic basis for the observed *N*-terminal positioning of the basic region relative to the dimerization domain is unlikely.

Naturally occurring basic-region leucine zipper (bZip)[1] proteins contain a bipartite DNA-binding motif consisting of a coiled-coil dimerization domain and a highly charged basic region that directly contacts DNA (for reviews, see *1, 2*). The basic region is always *N*-terminal to the leucine zipper, and the residues that contact DNA lie near the ends of a pair of continuous α-helices. A similarly uniform arrangement of the basic region and dimerization domain is found in basic-region helix−loop−helix (bHLH) proteins. It is not clear if there is a thermodynamic basis for this evolutionary trend.

Many classes of DNA-binding proteins, including helix−turn−helix and zinc finger proteins, use residues at or near the *N*-terminus of an α-helix to contact DNA (*3*). It has been suggested that this arrangement allows favorable electrostatic interactions between the α-helix macrodipole (*4*) and the negatively charged phosphodiester backbone of DNA (*5*). The α-helix macrodipole is a result of the parallel alignment of individual peptide dipoles along the helix axis (*4*) and the combination of four unpaired hydrogen-bond donors with partial positive charges at the *N*-terminus and four unpaired hydrogen-bond acceptors with partial negative charges at the *C*-terminus (*6, 7*). Local electrostatic interactions with the helix dipole are clearly important for protein stability, enzyme catalysis, and substrate binding (*8*−*27*). However, the importance of the helix dipole in long-range interactions is less clear. Even for the well-studied class of four-helix bundle proteins, there is no clear consensus as to whether helix dipole interactions stabilize the all-antiparallel arrangement of helices. Indeed, theoretical and experimental estimates for the magnitude of dipole interaction energies between a pair of antiparallel helices range over an order of magnitude, from 0.2 to 2.5 kcal/mol (*28*−*32*).

In a number of protein·DNA complexes, including prokaryotic helix−turn−helix protein·DNA complexes (*3*), residues at the *N*-terminal end of an α-helix make critical base contacts, suggesting a potential interaction between the α-helix macrodipole and the negatively charged DNA backbone. However, in the case of bZip proteins, the α-helix extends for roughly two turns beyond the residues that contact DNA (*33, 34*). Thus, a long-range interaction would be required for the α-helix macrodipole to play a role in stabilizing the bZip·DNA complex. Nonetheless, macrodipole−phosphate interactions have been proposed explicitly to explain the exclusive *N*-terminal positioning of the basic region relative to the leucine zipper in bZip proteins (*35*).

Here we describe the design of a series of GCN4 variants in which the GCN4 basic region is placed *C*-terminal to the leucine zipper. These "reverse" GCN4 (rGCN4) proteins have been tested for their ability to bind to DNA sites in which the individual consenus half-sites are inverted relative to the GCN4 recognition site (Figure 1). This design allows conservation of all local interactions between the protein and its DNA target, despite the unnatural positioning of the basic region relative to the leucine zipper. The DNA-binding domain, however, is moved from a position close to the *N*-terminal ends of a pair of long, continuous α-helices to a position equally close to the *C*-terminal ends of the dimer (Figure 1). This unique design allows us to test experimentally whether there is a thermodynamic basis, due to helix dipole interactions or to any other factor, for the rigid evolutionary bias for placing the basic region *N*-terminal to the dimerization domain in bZip and related proteins.

† 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. Telephone: (812) 855-4843; fax: (812) 855-8300; e-mail: oakley@indiana.edu.

[1] Abbreviations: bZip, basic-region leucine zipper; bHLH, basic-region helix−loop−helix; rGCN4, reverse GCN4; PCR, polymerase chain reaction; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; PBS, phosphate-buffered saline; EDTA, ethylenediaminetetraacetic acid; DTT, DL-dithiothreitol; BSA, bovine serum albumin; θ_{app}, fraction of DNA bound; $[L]_{tot}$, total protein monomer concentration; K_a, apparent monomeric equilibrium association constant; K_d, apparent monomeric equilibrium dissociation constant; CD, circular dicroism; $[\theta]_{222}$, molar ellipticity at 222 nm; DMS, dimethyl sulfate; CRE, cAMP-responsive element.

10.1021/bi011088b CCC: $20.00
Published on Web 10/23/2001

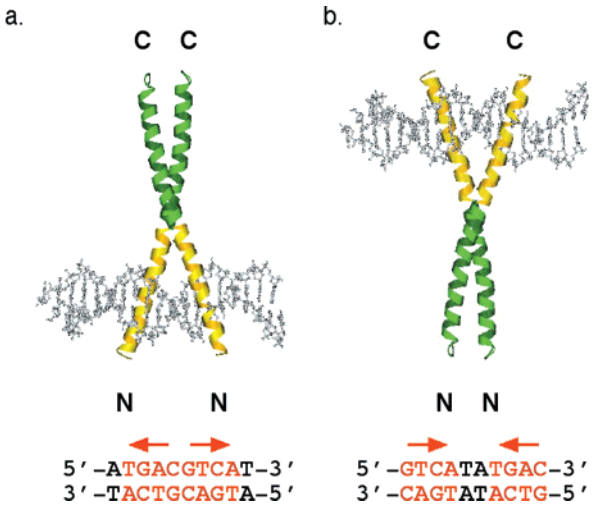

FIGURE 1: Schematic representation of (a) GCN4−56 binding to CRE DNA and (b) rGCN4(+3) binding to INV2 DNA. The leucine zipper is shown in green and the basic region in yellow. The arrows above the DNA sequences indicate the relative orientations of each half-site within the full binding site. As a general rule, in the GCN4·CRE complex, residues near the *N*-terminal end of the basic region contact base pairs on the outside of the DNA site, while those near the *C*-terminal end of the basic region, closer to the coiled-coil domain, contact residues at the center of the DNA site. In the rGCN4 peptides, residues at the *C*-terminal end of the basic region are next to the coiled-coil domain, while those at the *N*-terminal end are located on the ends of the bZip structure. Thus, to maintain specific DNA contacts, each DNA half-site must be inverted.

EXPERIMENTAL PROCEDURES

Plasmid Construction. DNA manipulations were carried out by standard methods (*36*). All mutations were made by Kunkel mutagenesis (*37*) or QuikChange Site-Directed Mutagenesis (Stratagene). The primary sequence of each construct was confirmed by DNA sequencing (*38*). Plasmid pBRBR, encoding two consecutive basic regions, was constructed as a precursor to the plasmid pRGCN4 for expression of *reverse* GCN4 (rGCN4). The oligonucleotide primers 5′-GGTGAACTGCAGCGTGTTAAACGTGCTCG-TAAC-3′ and 5′-CGCGGATCCTCAACCCTTCAGAGCG-GCTTTACGAGCACGAGA-3′, designed to incorporate a "linker" region and helix-stabilizing residues at the *C*-terminus of rGCN4, were used for amplification of the GCN4 basic region from the plasmid pGG63 (kindly provided by R. V. Talanian and P. S. Kim; see *39* and references cited therein). The PCR product was subcloned into the *Pst* I-*Bam H* I restriction fragment of pGG63, affording the intermediate pBRBR. PCR amplification of the leucine zipper coding region of pGG63 followed by subcloning into pBRBR resulted in the desired plasmid, pRGCN4. Six additional recombinant plasmids, with insertions or deletions in the coding sequence for the "linker" region, were constructed by site-directed mutagenesis.

Protein Expression and Purification. Each of the rGCN4 proteins was overexpressed in *E. coli* strain BL21(DE3) pLysS using the T7 expression system (*40*) and purified from cell lysates as described previously for the GCN4 bZip domain (*39*). Protein concentrations were determined by UV absorbance in 6 M guanidinium chloride and 0.02 M phosphate (pH 6.5) assuming an extinction coefficient of 1470 M^{-1} cm^{-1} for Tyr at 275 nm (*41*). The molecular mass of each purified protein was confirmed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry.

Electrophoretic Mobility Shift Assays. Oligonucleotides (Genosys) containing the DNA-binding sites INV0, INV1, INV2, INV3, and INV4 (Figure 2d) were purified by preparative denaturing (19% polyacrylamide, 19:1 acrylamide:bisacrylamide) gel electrophoresis and 5′-end-labeled by standard procedures (*36*). Radiolabeled oligonucleotides were annealed to their complement by heating an equimolar mixture of the two single-stranded DNAs to 90 °C for 2 min and then cooling slowly to room temperature.

Protein·DNA complex formation was initiated by adding labeled DNA to known concentrations of protein 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)] (*39*, *42*) supplemented with calf thymus DNA (100 μM in base pairs). These conditions are sufficiently stringent to distinguish between specific and nonspecific protein·DNA interactions (*39*). 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, 0.25% xylene cyanole FF). Each sample was loaded on a nondenaturing 8% polyacrylamide (79:1 acrylamide:bisacrylamide) gel preequilibrated 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 using ImageQuant version 3.3 software.

An apparent equilibrium dissociation constant was determined for the rGCN4(+3)·INV2 DNA complex. The fraction of INV2 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 rGCN4(+3)·INV2 DNA complex (*39*):

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

$[L]_{tot}$ 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 or saturated, respectively. Data were fit using the nonlinear least-squares fitting procedure of KaleidaGraph software (version 3.0.5, Abelbeck software) with K_a, θ_{max}, and θ_{min} as the adjustable parameters. The dissociation constant was determined from the average of four independent data sets.

CD Spectroscopy. CD spectra were acquired with a Jasco J-715 spectropolarimeter. Samples were prepared in 200 mM KCl and 50 mM potassium phosphate (pH 7.0). The wavelength dependence of $[\theta]$ was monitored at 25 °C by 5 scans acquired in 1 nm increments with a sampling time of 4 s. The helix content was determined by the method of Chen et al. (*43*).

Dimethyl Sulfate (DMS) Footprinting Assays. Two plasmids containing the DNA-binding sites CRE and INV2 were

constructed for dimethyl sulfate (DMS) footprinting experiments. The oligonucleotides 5′-TATGG**ATGACGTCAT**CG-3′ and 5′-GATCCG**ATGACGTCAT**CCA-3′ containing the CRE sequence (in boldface type) were treated with T4 polynucleotide kinase and subcloned into the multiple cloning site of pGG63. Plasmid pINV2 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:bisacrylamide) gel electrophoresis. Chemical sequencing A reactions were carried out as described previously (*44*).

DMS reaction conditions included 50 mM sodium cacodylate, 1 mM EDTA, 200 μM bp calf thymus DNA, pH 8.0 (*45*). Protein samples were allowed to equilibrate with labeled DNA for 30 min prior to addition of dimethyl sulfate to a final concentration of 1%. DMS digestion was quenched after a 10 min incubation period (4 °C) by the addition of excess β-mercaptoethanol followed by precipitation from ethanol. Strand scission was induced by treatment with aqueous piperidine (1 M, 30 min, 90 °C). Reaction products were analyzed by electrophoresis on 8% 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.

RESULTS AND DISCUSSION

Design of rGCN4. The *C*-terminal 56 residues of GCN4 contain both the leucine zipper and the basic region (Figure 2a). We defined the leucine zipper as residues 251−281 and the basic region as Lys 231 through Lys 246. A short "linker" sequence separates these two domains. The *N*-terminal residues, DPAAL, though not part of the conserved basic region, are important for DNA binding. Deletion of these residues leads to a reduction in DNA-binding affinity relative to the *C*-terminal 56-mer (*46*). Asp and Pro are often found at the *N*-termini of α-helices, and they are followed by the excellent helix-forming residues Ala and Leu (*47*).

In the rGCN4(0) sequence (Figure 2b), the leucine zipper is at the *N*-terminus, followed by the "linker" and basic regions defined for GCN4−56. The *C*-terminal end of the basic region, which no longer abuts the leucine zipper, is stabilized by helix-forming residues, followed by a lysine and a glycine, both of which occur preferentially at the *C*-termini of α-helices (*47*). Although the *order* of the two subdomains is reversed, the sequence of amino acid residues *within* each subdomain is unchanged.

The spacing between the leucine zipper and basic region is rigidly conserved in all naturally occurring bZip proteins (*1, 2, 48*), suggesting that DNA binding requires a fixed three-dimensional relationship between the dimerization interface and the residues that contact DNA. DNA binding by the bZip protein C/EBP is abolished upon deletion or insertion of five amino acid residues (*49*), and GCN4 proteins containing two-, four-, or six-residue insertions are not functional in vivo (*48*). However, when seven amino acids are inserted, corresponding to an integral number of heptad repeats within the coiled-coil domain, the correct spatial relationship between the leucine zipper and basic region is restored, and DNA binding is observed (*48*). Importantly, a wide variety of seven-amino acid insertions are tolerated (*48*),

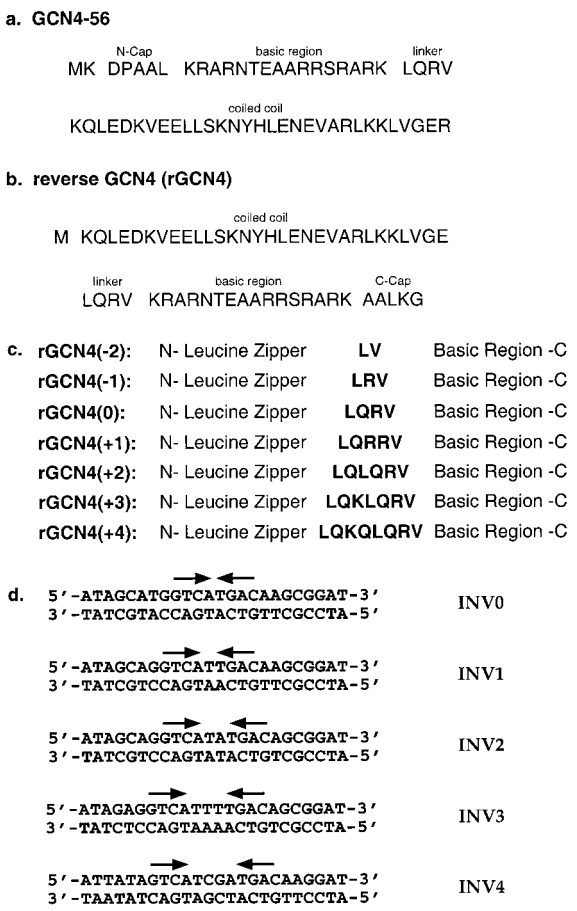

FIGURE 2: (a) Sequence from the *N*- to the *C*-terminus of GCN4−56. (b) Sequence from the *N*- to the *C*-terminus of the parent "reverse" GCN4 (rGCN4) protein, rGCN4(0). (c) Sequences from the *N*- to the *C*-termini of the seven rGCN4 proteins designed to determine the optimal spacing between the leucine zipper and the basic region for a bZip protein with a *C*-terminal DNA-binding domain. (d) Sequences of the oligonucleotides used for electrophoretic mobility shift assays. Each DNA binding site contains an inverted repeat of the half-site 5′-GTCA-3′ with a different number of base pairs between the two half-sites. The INV4-binding site is the same DNA sequence used previously by Park et al. (*66*).

suggesting that conserving the spatial orientation of the basic region with respect to the DNA is more important than inserting specific amino acid residues into the protein sequence. Similarly, the introduction of extra amino acid residues between the homeodomain and the leucine zipper in homeodomain-leucine zipper proteins significantly reduces DNA-binding affinity, indicating that the leucine zipper is also responsible for the correct positioning of the DNA-binding residues in this class of proteins (*50*).

As the appropriate spacing between an *N*-terminal leucine zipper and a *C*-terminal basic region was unknown, we designed an additional six rGCN4 peptides, each with a different number of amino acids in the "linker" region, to span the full 360° of rotational space (Figure 2c). We expected that only one of the seven rGCN4 peptides would position the basic region residues appropriately for specific and high-affinity DNA binding.

In all of the rGCN4 proteins, the amino acid residues immediately adjacent to both the leucine zipper and the basic region are conserved. The *C*-terminal residue in the "linker" region of GCN4 is the final residue in a coiled-coil conformation in the crystal structure of the GCN4·AP-1

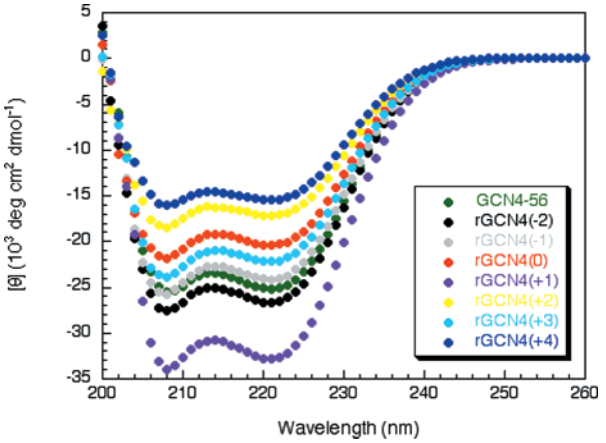

FIGURE 3: CD spectra of each rGCN4 peptide in 200 mM KCl, 50 mM potassium phosphate, pH 7.0, 25 °C. The CD spectrum of GCN4−56 under the same conditions is shown in green. Values of $[\theta]_{222}$ (deg cm^2 dmol^{-1}) are −25 058 for GCN4−56, −26 564 for rGCN4(−2), −23 982 for rGCN4(−1), −20 316 for rGCN4-(0), −32 664 for rGCN4(+1), −17 078 for rGCN4(+2), −22 074 for rGCN4(+3), and −15 407 for rGCN4(+4). These values correspond to folded peptides that are 66%, 71%, 64%, 54%, 87%, 45%, 58%, and 41% helical, respectively (*43*).

complex (*33*). The Leu residue adjacent to the leucine zipper in rGCN4 was maintained as it falls at a **d** position and may play a role in stabilizing the coiled-coil. For the four rGCN4 proteins containing amino acid insertions, residues with a high helical propensity, such as Arg, Lys, and Leu, were chosen (*51*). As all the side chains in the GCN4 "linker" region are bulky, amino acid residues with short side chains were avoided. The basic Arg in the GCN4 "linker" region engages in phosphate contacts with DNA in one of the two published crystal structures (*33, 34*). The position of this residue relative to the basic region was therefore maintained by making insertions in the *N*-terminal end of the "linker" in each protein except for rGCN4(+1). In this case, an additional Arg residue was inserted near the basic region to probe the importance of additional basic residues in the "linker" region.

Secondary Structure of the rGCN4 Peptides. Each of the rGCN4 peptides was overexpressed in *E. coli* and purified as described previously (*39*). The CD spectrum of each peptide in the absence of DNA (200 mM potassium chloride, 50 mM potassium phosphate, pH 7.0 and 25 °C) is shown in Figure 3. Under these conditions, GCN4−56 is ca. 66% helical based on a mean residue ellipticity of −25 058 at 222 nm (*43*). This observation is consistent with a helical leucine zipper domain and a largely unstructured basic region, as others have observed (*39, 52−60*). Interestingly, the helicity of the rGCN4 peptides varies significantly. Because the isolated GCN4 leucine zipper is fully helical (*61*), a bZip domain with a helical leucine zipper and with completely unstructured "linker" and basic regions should be approximately 50−55% helical, depending on the linker length. However, NMR studies have demonstrated that the GCN4 basic region adopts a fluctuating helical structure in the absence of DNA (*58, 62*) leading to the observed higher degree of helicity. The CD data for four of the seven peptides: rGCN4(−2), rGCN4(−1), rGCN4(0), and rGCN4-(+3), are consistent with this model, and minor differences in the degree of helicity in these rGCN4 variants may result from small changes in the lifetime of partially folded states. In contrast, two of the variants, rGCN4(+2) and rGCN4-(+4), are significantly less helical, suggesting that the leucine zipper domain is not fully folded in these peptides. As all of the rGCN4 variants contain the same leucine zipper sequence, the reason for this observation is not clear.

The remaining variant, rGCN4(+1), is approximately 87% helical, suggesting that its basic and "linker" regions are largely folded. It has been proposed that a helical conformation of bZip basic regions in the absence of DNA would lead to repulsive interhelical interactions between positively charged residues in the two basic region subdomains (*55, 57, 63*). The observation that the rGCN4(+1) bZip domain is largely folded therefore suggests that the positively charged residues important for DNA binding are rotated to the outside of the helices in this structure, diminishing the Coulombic repulsions between the two basic regions. Such an arrangement of these positively charged residues would be expected to place them out of phase for sequence-specific DNA binding.

Determination of the Appropriate Spacing between the Subdomains of rGCN4. GCN4 binds with high affinity to the pseudosymmetric AP-1 site, 5′-**TGACTCA**-3′ (*64*), as well as to the perfectly symmetric CRE site, 5′-A**TGACGT-CA**T-3′ (*65*). In the rGCN4·DNA complex, each basic region lies in the opposite orientation with respect to the DNA. The protein·DNA contacts observed in the GCN4 complexes can therefore be conserved if rGCN4 binds to an inverted repeat with both half-sites in the opposite orientation of those in AP-1/CRE DNA (Figure 1). Two v-Jun basic regions covalently attached at their *N*-termini through a flexible, disulfide-containing linker recognize such a DNA sequence (*66*).

Because the optimal half-site spacing for rGCN4·DNA complex formation was unknown, we used DNA duplexes in which the two half-sites, 5′-GTCA-3′ and 5′-TGAC-3′, are separated by zero, one, two, three, or four base pairs (Figure 2d). We monitored the affinity of each rGCN4 peptide for the DNA duplexes using a mobility shift assay. To minimize nonspecific binding, approximately physiological concentrations of monovalent salt were used in the presence of nonspecific competitor DNA. These conditions have been shown previously to differentiate between specific and nonspecific interactions (*39*). Qualitative results are shown in Figure 4. Of the 35 protein·DNA combinations tested, high-affinity binding was observed only for rGCN4-(+3) to the INV2 site. Thus, the placement of seven residues in the "linker" region of rGCN4 leads to the appropriate positioning of the basic regions for DNA binding.

These results highlight the importance of the spacing between the leucine zipper and basic region for specific DNA binding by bZip proteins. Only one of seven rGCN4 peptides is able to properly position both basic regions in the DNA major groove. As expected, changes in the "linker" region significantly diminish DNA-binding affinity, presumably by rotating the basic regions out of register to maintain sequence-specific contacts. Modest DNA-binding affinity is observed only for rGCN4(−1), which is related to rGCN4-(+3) by approximately one turn of helix. All other proteins show little or no affinity for any of the designed binding sites. Interestingly, the two worst binders, rGCN4(+1) and rGCN4(+4), are also related by approximately one turn of helix.

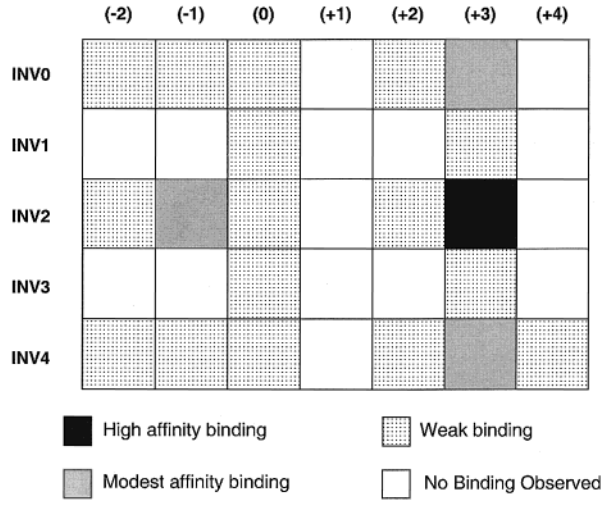

FIGURE 4: Qualitative results of electrophoretic mobility shift assays with each of the seven rGCN4 proteins and the five INV DNA-binding sites. Each protein·DNA combination was tested at least twice. Black boxes, binding observed at 60 nM [protein]; gray boxes, binding observed at 600 nM [protein]; stippled boxes, weak binding observed at 6 μM [protein]; white boxes, no binding observed at any protein concentration.

Characterization of the rGCN4(+3)·INV2 Complex. The structure of the rGCN4(+3)·INV2 complex was probed qualitatively by dimethyl sulfate (DMS) footprinting and circular dichroism (CD) spectroscopy. To investigate whether the same protein·DNA contacts are made in the GCN4·CRE and rGCN4(+3)·INV2 complexes, we treated radiolabeled DNA fragments containing the CRE- or INV2-binding sites with DMS in the presence and absence of protein. Because DMS modifies guanine residues in the major groove of DNA, bases contacted by these proteins should be protected from DMS modification (*67*). As expected, in the presence of GCN4−56, the two guanine residues in the CRE-binding site are protected from DMS modification, and no cleavage is observed at those positions (Figure 5a). Similarly, no cleavage is observed at the two guanines in the INV2 site (Figure 5b), suggesting that the protein·DNA interactions are similar in both complexes. Finally, CD measurements indicate that rGCN4(+3) is ca. 58% helical in the absence of DNA (Figure 3), and that the helicity increases to approximately 80% in the presence of INV2 (data not shown). This increase in helicity is indicative of the coil → helix transition observed in bZip peptides upon binding DNA (*39*, *52*−*60*).

DNA-Binding Affinity and Specificity. The affinity of rGCN4(+3) for the INV2-binding site was determined using a quantitative mobility shift assay (Figure 6). The fraction of DNA bound was plotted as a function of total peptide monomer concentration, and the data were fit to an equation describing formation of a 2:1 peptide·DNA complex (*39*). rGCN4(+3) binds to the INV2 site with an apparent monomeric dissociation constant (K_d) of $(2.9 \pm 0.8) \times 10^{-8}$ M. This dissociation constant compares favorably to that of GCN4 for its optimal AP-1 site, $(9 \pm 6) \times 10^{-8}$ M, under the same experimental conditions (*39*).

DNA binding by rGCN4(+3) is also specific. No secondary sites are apparent in the 186 bp fragment of DNA used for footprinting analysis (Figure 5b). Indeed, rGCN4(+3) appears to bind to DNA with increased specificity compared

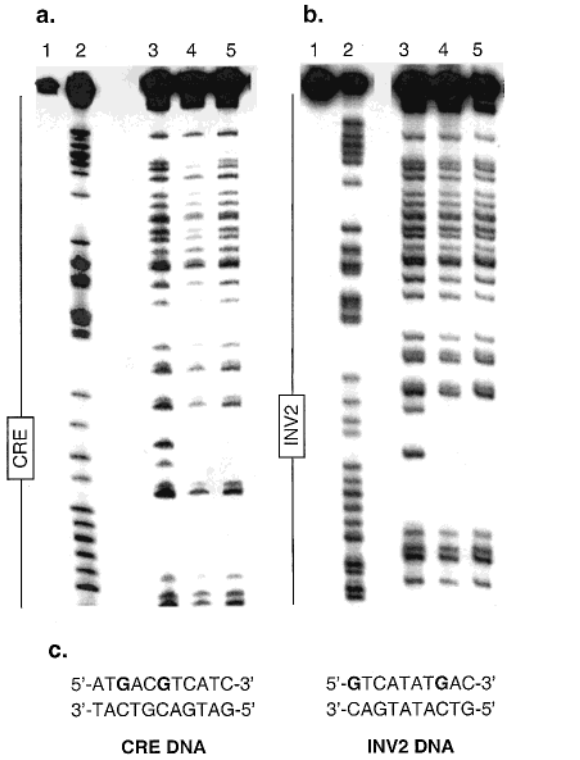

FIGURE 5: DMS footprinting reactions with (a) GCN4−56 on CRE DNA and (b) rGCN4(+3) on INV2 DNA. Lane 1, intact DNA; lane 2, A-specific chemical sequencing reaction; lanes 3−5, DMS cleavage in the presence of 0, 6, and 3 μM protein, respectively. (c) DNA sequence of the CRE- and INV2-binding sites. Only the top strand (as shown) is 5′-end-labeled; thus, the sequence from 5′ to 3′ can be read from the bottom to the top in panels a and b. DMS modification is observed at the G positions indicated in boldface type.

to GCN4−56. GCN4 binds with near-equal affinity to two DNA sites, AP-1 (*64*) and CRE (*65*). The perfectly symmetric CRE site contains an additional C·G base pair compared to the pseudosymmetric AP-1 site. rGCN4(+3), on the other hand, has only one high-affinity ligand, INV2 (Figure 4). Crystal structures of the GCN4·AP-1 and the GCN4·CRE complexes demonstrate that, at least in the solid phase, DNA flexibility allows equivalent protein·DNA contacts in both complexes with minimal changes to the protein structure (*33*, *34*). The clear preference of rGCN4-(+3) for a single INV site may therefore result from properties of the DNA target, rather than from the protein.

GCN4−56 also binds with moderate affinity to sites containing only a single consensus half-site, showing approximately a 10-fold reduction in affinity with respect to the AP-1 site (*39*, *68*). Binding to a single half-site is likely to involve specific contacts by one monomer to the consensus half-site and nonspecific contacts by the second monomer (*39*). Because rGCN4(+3) can make specific contacts at one half-site even when the spacing between the consensus half-sites does not allow specific contacts at both half-sites, we expected to observe moderate affinity binding to all of the INV sites by rGCN4(+3). Such moderate affinity binding is seen for the INV0 and INV4 sites (Figure 4). However, rGCN4(+3) does not bind with the expected affinity to the INV3 and INV1 sites. This result suggests that, in contrast to GCN4, recognition by one monomer of rGCN4(+3) of a single consensus half-site along with nonspecific binding by

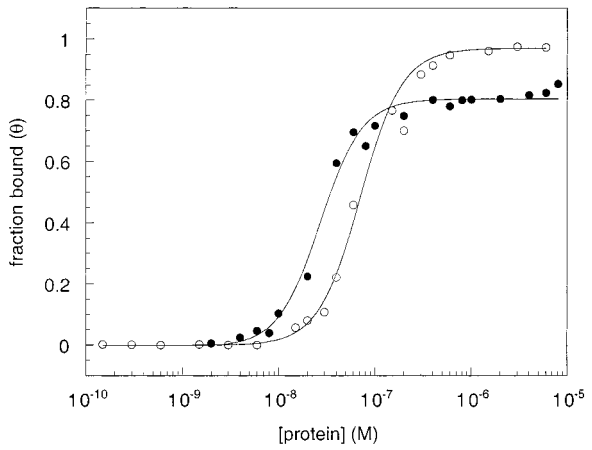

FIGURE 6: Equilibrium binding isotherms for GCN4−56 binding to AP-1 DNA (○) (*39*) and rGCN4(+3) binding to INV2 DNA (●) in 137 mM NaCl, 2.7 mM KCl, 4.3 mM sodium phosphate, 1.4 mM potassium phosphate, 1 mM EDTA, 1 mM DTT, 0.1% Igepal CA-630, 0.4 mg mL^{-1} BSA, 100 μM bp calf thymus DNA, pH 7.4. GCN4−56 concentrations 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. rGCN4(+3) concentrations are 8 μM, 6 μM, 4 μM, 2 μM, 1 μM, 800 nM, 600 nM, 400 nM, 200 nM, 100 nM, 80 nM, 60 nM, 40 nM, 20 nM, 10 nM, 8 nM, 6 nM, 4 nM, and 2 nM. Each data point represents the average of the fraction bound (θ) values at each concentration from three or four data sets, and the binding isotherms were obtained using eq 1 and the mean K_a values for each set. The apparent monomeric equilibrium dissociation constant for GCN4−56 binding to AP-1 DNA is $(9 \pm 6) \times 10^{-8}$ M (*39*) and for rGCN4(+3) binding to INV2 DNA $(2.9 \pm 0.8) \times 10^{-8}$ M.

the second monomer is not sufficient for modest DNA-binding activity.

Our results suggest it is unlikely that there is a thermodynamic basis for the observed *N*-terminal position of the basic region relative to the leucine zipper. We cannot definitively rule out the possibility that there is a modest thermodynamic cost to placing the basic region at the *C*-terminus, and that an unanticipated interaction occurs in the rGCN4(+3)·DNA complex that restores binding affinity. However, the "linker" region of rGCN4(+3) was designed to avoid unanticipated interactions. The four residues closest to the basic region are identical to the linker used in GCN4−56, and the closest additional basic residue is almost two full turns of helix away, making additional phosphate contacts unlikely. In addition, as rGCN4(+1), which contains an extra basic residue in close proximity to the linker region, fails to bind DNA, it is clear that incorporation of an additional positive charge is insufficient to allow DNA binding. Three-dimensional structural analysis will be required to address this issue.

Our overall design strategy has been remarkably successful, as the optimal rGCN4 variant binds DNA with at least wild-type specificity and affinity. Importantly, no further rounds of optimization were required to achieve wild-type activity after identifying the correct "register" for DNA binding, even though, unlike GCN4, rGCN4(+3) is not a product of natural selection. In addition, we have been able to target a DNA site not contacted by any naturally occurring bZip domain. As v-Jun basic regions covalently attached in a head-to-tail fashion through a flexible, disulfide-containing linker can also recognize a direct repeat of the AP-1 half-site (*69*), it should be possible to engineer a bZip protein with an antiparallel coiled-coil to recognize such a DNA sequence. These experiments can also be extended to the bHLH class of DNA-binding proteins.

Our results clearly demonstrate that the exclusive *N*-terminal positioning of the basic region relative to the coiled-coil is not necessary for specific, high-affinity DNA binding. Because bHLH proteins contain a similar bipartite DNA-binding domain (*3*), it is also unlikely that the basic region must be positioned *N*-terminal to the dimerization domain in this class of proteins. It is unclear whether the α-helix macrodipole contributes to the overall stability of complexes between zinc finger or helix−turn−helix proteins and their DNA targets. In these cases, DNA-contacting residues are located at or near the *N*-terminus of an α-helix (*3*), possibly allowing local interactions with the DNA. However, the recognition helix in both classes of proteins is tightly packed against additional secondary structural elements, precluding an experimental test of the importance of the helix dipole without disrupting the structure of these domains. It is interesting to note that homeodomain proteins, which have a structure similar to that of the helix−turn−helix proteins, make protein·DNA contacts with residues located in the middle of the recognition helix (*3*), suggesting that helix−dipole interactions are unnecessary for DNA binding by this class of proteins as well.

CONCLUSION

We have examined the DNA-binding properties of GCN4 variants in which the basic region is placed *C*-terminal to the leucine zipper. The spacing between the leucine zipper and the basic region is strictly conserved in naturally occurring bZip proteins. We find that only one of seven rGCN4 variants binds to DNA with high affinity, suggesting that there are also strict spatial requirements for positioning the basic region *C*-terminal to the leucine zipper. The affinity of the optimal rGCN4 peptide, rGCN4(+3), for DNA is at least as high as that of GCN4 for its optimal site, and rGCN4-(+3) appears to bind more specifically to its target site than does GCN4. These observations suggest that there is no thermodynamic basis for the exclusive *N*-terminal placement of the basic region relative to the leucine zipper in bZip proteins.

ACKNOWLEDGMENT

We thank R. Talanian and P. Kim for plasmid pGG63, A. Trudel for assistance with plasmid construction, S. Kan and A. Hansen for assistance with mass spectrometry, and T. Widlanski for careful reading of the manuscript.

REFERENCES

1. Hu, J. C., and Sauer, R. T. (1992) *Nucleic Acids Mol. Biol. 6*, 82−101.
2. Hurst, H. C. (1995) *Protein Profile 2*, 105−168.
3. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
4. Wada, A. (1976) *Adv. Biophys. 9*, 1−63.
5. Hol, W. G. J., van Duijnen, P. T., and Berendsen, H. J. C. (1978) *Nature 273*, 443−446.
6. Åqvist, J., Luecke, H., Quiocho, F. A., and Warshel, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2026−2030.
7. Tidor, B. (1994) *Proteins: Struct., Funct., Genet. 19*, 310−323.
8. Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol. 45*, 149−195.

9. Pflugrath, J. W., and Quiocho, F. A. (1985) *Nature 314*, 257−260.
10. Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., and Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349−2353.
11. Wierenga, R. K., DeMaeyer, M. C. H., and Hol, W. G. J. (1985) *Biochemistry 24*, 1346−1357.
12. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., and Baldwin, R. L. (1987) *Nature 326*, 563−567.
13. Nicholson, H., Becktel, W. J., and Matthews, B. W. (1988) *Nature 336*, 651−656.
14. Sali, D., Bycroft, M., and Fersht, A. R. (1988) *Nature 335*, 740−743.
15. Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M., and Baldwin, R. L. (1989) *Proteins: Struct., Funct., Genet. 5*, 1−7.
16. Takahashi, S., Kim, E.-H., Hibino, T., and Ooi, T. (1989) *Biopolymers 28*, 995−1009.
17. Nicholson, H., Anderson, D. E., Dao-pin, S., and Matthews, B. W. (1991) *Biochemistry 30*, 9816−9828.
18. Eijsink, V. G. H., Vriend, F., van den Burg, V., van der Zee, J. R., and Venema, G. (1992) *Protein Eng. 5*, 165−170.
19. Sancho, J., Serrano, L., and Fersht, A. R. (1992) *Biochemistry 31*, 2253−2258.
20. Armstrong, K. M., and Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11337−11340.
21. Huyghues-Despointes, B. M. P., Scholtz, J. M., and Baldwin, R. L. (1993) *Protein Sci. 2*, 1604−1611.
22. Lockhart, D. J., and Kim, P. S. (1993) *Science 260*, 198−202.
23. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1997) *J. Pept. Sci. 3*, 209−223.
24. Malakauskas, S. M., and Mayo, S. L. (1998) *Nat. Struct. Biol. 5*, 470−475.
25. Trickey, P., Wagner, M. A., Jorns, M. S., and Mathews, F. S. (1999) *Structure 7*, 331−345.
26. Arndt, K. M., Pelletier, J. N., Müller, K. M., Alber, T., Michnick, S. W., and Plückthun, A. (2000) *J. Mol. Biol. 295*, 627−639.
27. Strop, P., Marinescu, A. M., and Mayo, S. L. (2000) *Protein Sci. 9*, 1391−1394.
28. Sheridan, R. P., Levy, R. M., and Salemme, F. R. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4545−4549.
29. Chou, K.-C., Némethy, G., and Scheraga, H. A. (1983) *J. Phys. Chem. 87*, 2869−2881.
30. Chou, K.-C., Maggiora, G. M., Némethy, G., and Scheraga, H. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4295−4299.
31. Gilson, M. K., and Honig, B. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1524−1528.
32. Robinson, C. R., and Sligar, S. G. (1993) *Protein Sci. 2*, 826−837.
33. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223−1237.
34. König, P., and Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139−154.
35. Pathak, D., and Sigler, P. B. (1992) *Curr. Opin. Struct. Biol. 2*, 116−123.
36. Sambrook, J., Fritsch, E. F., and Maniatis, R. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
37. Kunkel, R. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
38. Sanger, F. S. N., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
39. Hollenbeck, J. J., and Oakley, M. G. (2000) *Biochemistry 39*, 6380−6389.
40. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
41. Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
42. Metallo, S. J., and Schepartz, A. (1994) *Chem. Biol. 1*, 143−151.
43. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) *Biochemistry 13*, 3350−3359.
44. Iverson, B., and Dervan P. (1987) *Nucleic Acids Res. 15*, 7823−7830.
45. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.
46. Oakley, M. G., and Dervan, P. B. (1990) *Science 248*, 847−850.
47. Richardson, J. S., and Richardson, D. C. (1988) *Science 240*, 1648−1652.
48. Pu, W. T., and Struhl, K. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6901−6905.
49. Agre, P., Johnson, P. F., and McKnight, S. L. (1989) *Science 246*, 922−926.
50. Sessa, G., Morelli, G., and Ruberti, I. (1993) *EMBO J. 12*, 3507−3517.
51. O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990) *Science 250*, 646−651.
52. Talanian, R. V., McKnight, C. J., and Kim, P. S. (1990) *Science 249*, 769−771.
53. Weiss, M. A. (1990) *Biochemistry 29*, 8020−8024.
54. Patel, L., Abate, C., and Curran, T. (1990) *Nature 347*, 572−575.
55. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) *Nature 347*, 575−578.
56. O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990) *Science 249*, 774−778.
57. O'Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991) *Biochemistry 30*, 9030−9034.
58. Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991) *Biochemistry 30*, 1310−1317.
59. Talanian, R. V., McKnight, C. J., Rutkowski, R., and Kim, P. S. (1992) *Biochemistry 31*, 6871−6875.
60. Cuenoud, B., and Schepartz, A. (1993) *Science 259*, 510−513.
61. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538−542.
62. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) *J. Mol. Biol. 285*, 2133-2146.
63. Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989) *Science 246*, 911−916.
64. Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986) *Science 234*, 451−457.
65. Sellers, J. W., Vincent, A. C., and Struhl, K. (1990) *Mol. Cell. Biol. 10*, 5077−5086.
66. Park, C., Campbell, J. L., and Goddard, W. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9094−9096.
67. Gilbert, W., Maxam, A., and Mirzabekov, A. (1976) Contacts between the LAC Repressor and DNA Revealed by Methylation. in *Control of Ribosome Synthesis* (Kjelgaard, N. O., and Maaloe, O., Eds.) pp 139−148, Munksgaard, Copenhagen, Denmark.
68. Metallo, S. J., and Schepartz, A. (1997) *Nat. Struct. Biol. 4*, 115−117.
69. Park, C., Campbell, J. L., and Goddard, W. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4892−4896.

BI011088B