Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 2920–2925, April 1996
Biochemistry

In vivo selection of basic region–leucine zipper proteins with altered DNA-binding specificities

TAKASHI SERA AND PETER G. SCHULTZ*

Howard Hughes Medical Institute, Department of Chemistry, University of California, Berkeley, CA 94720

Contributed by Peter G. Schultz, December 4, 1995

ABSTRACT **A transcription interference assay was used to generate mutant basic region–leucine zipper proteins with altered DNA-binding specificities. A library of mutants of a CCAAT/enhancer binding protein was constructed by randomizing five DNA-contacting amino acids in the basic region Asn^{-18}, Ala^{-15}, Val^{-14}, Ser^{-11}, and Arg^{-10}. These mutants were then selected for their ability to bind mutant recognition sequences containing substitutions at the 2 and 3 positions of the wild-type sequence 5′-$A^5T^4T^3G^2C^1G^{1'}C^{2'}A^{3'}A^{4'}T^{5'}$-3′. Mutants containing the sequence $Leu^{-18}Tyr^{-15}Xaa^{-14}$-$Tyr^{-11}Arg^{-10}$, in which four of the five contact residues are altered, were identified that recognize the palindromic sequence 5′-ATCYCGY′GAT-3′ (Xaa = asparagine when Y = G; Xaa = methionine when Y = A). Moreover, in a selection against the sequence 5′-ATTACGTAAT-3′, mutants were obtained containing substitutions not only in the basic region but also in the hinge region between the basic and leucine zipper regions. The mutant proteins showed high specificity in a functional transcription interference assay. A model for the interaction of these mutants with the target DNA sequences is discussed.**

The basic region–leucine zipper (bZip) DNA-binding proteins, which consist of a basic DNA-binding region and leucine-zipper dimerization domain, constitute a major class of eukaryotic transcription factors (1, 2). These proteins bind selectively as dimers to 7- to 10-bp palindromic operator sequences with nanomolar affinities (2, 3). Recently the x-ray crystal structures of DNA complexes of two bZip proteins, GCN4 (4, 5) and the cFos/cJun heterodimer (6), have been solved. In contrast to many other DNA-binding proteins, such as zinc finger (7), helix-turn-helix (8), or β-ribbon proteins (9), the recognition motif is composed of an isolated α-helix that contacts the bases in the major groove of B-DNA (4–6). Analysis of the cocrystal structures (4–6), as well as a sequence comparison of a large number of bZip proteins (2), suggests that DNA recognition is mediated, to a large degree, by hydrogen-bonding and hydrophobic interactions between the amino acid side chains at positions −18, −15, −14, −11, and −10 (the threonine at the first α-helix position of the leucine-zipper domain is +1) and the bases in the major groove. Consequently, this motif is among the simplest in which to explore the effects of mutations on DNA-binding affinity and specificity, as well as to generate new DNA-binding proteins with new specificities for potential applications as molecular biological or therapeutic reagents.

Several approaches have been reported for generating DNA-binding proteins with altered specificities. These approaches include site-directed and random mutagenesis experiments (10–12), the construction of chimeric DNA-binding proteins from different DNA-binding domains (13, 14), and the generation and screening of large libraries of mutant DNA-binding proteins (15–18). Phage-display systems (15–17) and *in vivo* selection experiments (18) have been used to screen libraries of zinc finger and helix-turn-helix proteins, respectively, for new specificities. In the former case, it was shown that a number of new specificities can be generated by altering amino acids in the α-helical region of the finger.

To determine to what degree the basic region of leucine-zipper proteins can serve as a general framework for DNA recognition we have generated a library of 3.2×10^6 mutants of a truncated CCAAT/enhancer binding protein (C/EBP) bZip protein in which five DNA contacting residues were randomized. Mutants were then selected for their ability to bind DNA-recognition sites using a transcription interference assay. A number of mutant proteins with new specificities have been generated and analyzed in terms of the specificity determinants for selective protein–DNA recognition.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. *Escherichia coli* JM109 was used for the selection experiments. *E. coli* NovaBlue(DE3) (Novagen) was used for overexpression of proteins. The gene encoding a truncated C/EBP (80 amino acids from Gly^{-40} to Ser^{40}) was synthesized by ligating synthetic oligonucleotides. The resulting fragment was ligated with the large *Eco*RI–*Hin*dIII fragment of pUC18 to produce plasmid pTS-19. The reporter plasmids used here, in which the 5′ side of each recognition site was placed at the +2 position, relative to the start of transcription, were constructed according to a described protocol (18).

Construction of C/EBP Library. pTS-47, the precursor plasmid for library construction, was constructed by ligating a 37-bp DNA linker, containing *Bsi*WI and *Xho* I ends and an internal *Swa* I site, with the large *Bsi*WI–*Xho* I fragment of pTS-19. A phosphorylated 85-nt oligonucleotide (14 pmol) with the sequence, 5′-GTA CGT CGT GAA CGT NNK AAT ATT NNK NNK CGT AAA NNK NNK GAT AAA GCT AAA CAG CGT AAC GTT GAA ACT CAG CAG AAA GTT C-3′ [where N is A, G, C, or T (equimolar) and K is G or C (1:1)], encoding the C/EBP library was annealed with two phosphorylated primers, 5′-ACGTTCACGAC-3′ (15 pmol) and 5′-TCGAGAACTTTCTGCT-3′ (15 pmol), and ligated to 10 μg (5.6 pmol) of the large *Bsi*WI–*Xho* I fragment of pTS-47 using T4 DNA ligase (2000 units, incubation overnight at 16°C). Sequenase version 2.0 (United States Biochemical, 13 units) was then added to the ligation mixture, which was then incubated at 37°C for 1.5 hr, followed by *Swa* I digestion. Half of the mixture was then electroporated into freshly prepared electrocompetent *E. coli* MC1061. After 1 hr of nonselective growth, a library of 2×10^9 independent transformants was obtained. The cells were incubated overnight in 1 liter of Luria broth containing 50 mg of ampicillin (Amp). Plasmids were purified by ultracentrifugation. The library DNA was se-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bZip, basic region–leucine zipper; WT, wild type; C/EBP, CCAAT/enhancer binding protein; AMP, ampicillin; Sp, spectinomycin; Cm, chloramphenicol.

*To whom reprint requests should be addressed.

quenced as a pool and revealed equal band intensities in each of the sequencing lanes at the randomized sites.

***In Vivo* Selection Against Wild-Type (WT) and Mutant Operator Sequences.** The library DNA (5 μg) was transformed into 200 μl of freshly prepared electrocompetent JM109 containing a reporter plasmid. After 1 hr of nonselective growth in 5 ml of SOC medium followed by 2 hr of selective growth in the presence of Amp (50 μg/ml) and chloramphenicol (Cm) (10 μg/ml) at 37°C, IPTG (final concentration 1mM) was added for induction. After incubation at 37°C for 2 hr, the resulting culture was plated onto 40 plates (150 × 15 mm) of Luria broth agar containing Amp (80 μg/ml), spectinomycin (Sp) (60 μg/ml), and IPTG (1 mM), and incubated at 37°C for 40 hr. Colonies from the selection plates were pooled, and the plasmids were isolated and purified on 0.8% agarose gel to remove reporter plasmid. The selection protocol was repeated three to five times until an $\approx 10^4$-fold enrichment was obtained. Individual clones were then isolated and sequenced.

Expression and Purification of Mutant Proteins. The coding regions of selected clones were transferred into expression vector pET-14b (Novagen) with the PCR. Resulting plasmids were then introduced into *E. coli* NovaBlue(DE3) for protein overexpression. A 600-ml culture was grown to $OD_{600} = 0.6$ at 37°C, induced with 1 mM IPTG for 3 hr, and lysed using a ultrasonicator in cold lysis buffer [100 mM Tris·HCl, pH 8.0/1 M NaCl/2 mM EDTA/1 mM dithiothreitol containing 1 mM phenylmethylsulfonyl fluoride, pepstatin, and leupeptin (the latter each at 1 μg/ml)]. After treatment with polyethyleneimine (pH 7, 0.6%) and precipitation with 40% $(NH_4)_2SO_4$, the resulting pellet was redissolved in 50 mM Tris·HCl, pH 8.0/1 mM EDTA/1.4 mM 2-mercaptoethanol buffer and purified by chromatography on a Bio-Rex 70 column, eluting with 600 mM NaCl buffer. All purified proteins were >95% homogeneous as judged by SDS/PAGE.

DNA-Binding Assays. The *in vivo* DNA-binding activities of selected clones were measured as described (18). Thirty base-pair synthetic oligonucleotides (see legend, Table 4), labeled at the 5′-end with [γ-^{32}P]ATP, were used in the gel-retardation assays. Proteins were preincubated at room temperature for 30 min in 10 mM imidazole phosphate, pH 7.0/100 mM NaCl/2 mM $MgCl_2$/10% glycerol buffer containing 1 μg of poly(dI-dC)$_2$ per 15 μl of buffer. The end-labeled probe (2.5 fmol) was then added, and incubation was continued for 30 min before loading onto a 4% nondenaturing polyacrylamide gel (10 mM imidazole phosphate, pH 7.0) and electrophoresing at 115 V for 1.5 hr at room temperature.

RESULTS

Library Construction and Selection Against WT Operator. GCN4 is the most extensively studied of the bZip DNA-binding proteins (19). X-ray crystal structures of GCN4–DNA complexes (4, 5) reveal an almost continuous α-helix making up both the dimerization and DNA-binding domains. These structures together with a sequence analysis of a large family of bZip proteins (2) suggest that five residues, Asn^{-18}, Ala^{-15}, Ala^{-14}, Ser^{-11}, and Arg^{-10}, are important in determining sequence specificity by interacting with bases in the major groove of B-DNA.

To select for substitutions at these residues that might alter DNA-binding specificities, an *in vivo* selection system based on a previously developed transcription interference assay was used (20, 21). In this assay, the strong ConII bacterial polymerase promoter interferes with the expression of the Sp resistance gene, aadA (encoding aminoglycoside 3′-adenyltransferase), which is transcribed in the opposite direction. Functional bZip protein expressed in a cell containing the reporter plasmid disrupts transcription from the ConII promoter when the bZip DNA recognition sequence is inserted downstream of the ConII promoter. The result is a Sp^r phenotype in the host. By selecting for Sp^r on the transformation plates, clones encoding a bZip protein that binds a particular DNA recognition sequence can be selectively enriched. The dimeric nature of both the bZip protein and the recognition site should strongly favor binding to the palindromic recognition site over other sites in the reporter plasmid. Moreover, the use of an *in vivo* selection system eliminates difficulties associated with generating libraries of dimeric proteins fused to a surface protein of filamentous phage.

Initially, a reporter plasmid containing the palindromic GCN4 binding site, 5′-ATGACTCAT-3′, was constructed. Unfortunately, in the absence of GCN4 expression, the ratio of colonies on Sp plates versus Cm plates with *E. coli* JM109 transformed only with the reporter plasmid, was ≈1. This result illustrates a potential limitation of this selection technique in that some sequences interfere with transcription from the ConII promoter. Although changes in the size and sequence of the intervening DNA sequence may overcome this problem, we instead chose to examine another member of the bZip family, C/EBP. In this case, transformation of *E. coli* JM109 with the corresponding reporter plasmid containing the palindromic recognition site, 5′-ATTGCGCAAT-3′, led to a Sp^r/Cm^r ratio of 5×10^{-5} indicating that this was a viable system for selection using a library of mutant C/EBP proteins.

A library of mutant proteins was generated in which the key specificity-determining residues Asn^{-18}, Ala^{-15}, Val^{-14}, Ser^{-11}, and Arg^{-10} were completely randomized in the bZip element of C/EBP (residues Gly^{-40} to Ser^{40}) from rat (22). The codons for these amino acids were substituted with NNK (N = an equimolar mixture of A, G, C, or T and K = G or C) to afford a library of 3.2×10^7 gene sequences (3.2×10^6 protein sequences). The library was constructed from the precursor plasmid pTS-47, which harbors the gene encoding C/EBP with a *Bsi*WI–*Xho* I insert (containing an 8-bp *Swa* I restriction site) replacing the region coding for the DNA-binding domain. A synthetic oligonucleotide encoding the library and containing *Bsi*WI–*Xho* I termini was then ligated to the large fragment of *Bsi*WI/*Xho* I-digested pTS-47, followed by treatment with DNA polymerase and *Swa* I. The resulting library DNA was then transformed into competent *E. coli* MC1061, and 2×10^9 independent recombinants were obtained, covering the entire library of mutant C/EBP sequences.

The library was then subjected to selection using the reporter plasmid pTS-4, which contains the WT recognition sequence, 5′-A^{5}T^{4}T^{3}G^{2}C^{1}G$^{1'}$C$^{2'}$A$^{3'}$A$^{4'}$T$^{5'}$-3′. Three rounds of selection were done by isolating plasmid from Amp^{r}Sp^r colonies and retransforming this plasmid into JM109 bearing the reporter plasmid. The ratio of Amp^{r}Sp^r/Amp^r colonies increased from 5×10^{-4} after round one to 4×10^{-2} after round three of selection. Twenty colonies were then isolated, and the gene encoding C/EBP was sequenced, revealing the presence of WT C/EBP and a mutant protein containing a Ser$^{-11} \rightarrow$ Cys substitution. Retransformation of this mutant protein into JM109 cells harboring the reporter plasmid afforded an Amp^{r}Sp^r/Amp^r of 0.1, similar to that of WT protein. That no other proteins were isolated from the selection experiment confirms that the α-helical sequence Asn^{-18}, Ala^{-15}, Val^{-14}, (Ser/Cys)$^{-11}$, Arg^{-10} is optimal for binding to the WT recognition sequence.

Selection Against Mutant Operators. To select mutant proteins with new specificities, a series of reporter plasmids were constructed containing substitutions in the recognition sequence. The A-T bp at position 5 does not appear to be important in C/EBP recognition (unpublished work), whereas virtually every member of the bZip family recognizes a C-G bp at position 1 (2). Consequently, we initially constructed a series of nine mutant reporter plasmids containing substitutions at positions 2–4 of the WT sequence. Among them, four mutants

Table 1. Selection of C/EBP library against mutant operator sequences

	Amp^rSp^r/Amp^r ratio			
Round	pTS-62 5′-ATTACGTAAT	pTS-64 5′-ACTACGTAGT	pTS-65 5′-ATCACGTGAT	pTS-68 5′-ATCGCGCAT
1st	1×10^{-4}	7×10^{-4}	3×10^{-4}	1×10^{-4}
2nd	2×10^{-3}	5×10^{-5} to 1×10^{-3}	1×10^{-3}	3×10^{-4}
3rd	0.12	1.3×10^{-4}	7×10^{-3}	3.5×10^{-3}
4th		4.4×10^{-4}	0.02–0.13	0.37
5th			0.61	

Underlining represents mutation.

showed very low backgrounds when assayed for Sp^r/Cm^r in the transcriptional interference assay. Sequences with higher G-C/A-T ratios (5′-ACTGC . . . -3′, 5′-ACGAC . . . -3, 5′-ATGGC . . . -3′, 5′-ACGGC- . . . 3′, and 5′-ACCGC . . . -3′) tended to have Sp^r/Cm^r ratios > 0.1, again reflecting a current limitation of the transcriptional interference assay. *In vivo* selections were then done on the four sequences with the lowest backgrounds, and the results are summarized in Table 1. Although no enrichment could be obtained in the selection using pTS-64 (5′-ACT ACGTAGT-3, where underlining represents substitutions in WT sequences), enrichments were seen with the three other sequences after several selection rounds. Twenty colonies were isolated from each selection plate, and the DNA was sequenced (Table 2). In the library selected by using reporter plasmid pTS-62 (5′-ATTACGTAAT-3′), the residues Asn^{-18}, Ala^{-15}, Val^{-14}, and Arg^{-10} were conserved, but Ser^{-11} was substituted with the aromatic residues, tryptophan and histidine to afford mutants MT1 and MT2, respectively. Interestingly, fortuitous mutations were also found at residues other than those targeted in the initial library. These mutations included the relatively nonconservative changes, $Asp^{-9} \rightarrow$ Ala in MT1 and $Ala^{-7} \rightarrow$ Val, $Asn^{-3} \rightarrow$ Tyr in MT2. These mutations are located between the DNA-contacting region and the leucine-zipper moiety.

Clones isolated in selection experiments using reporter plasmid pTS-68 (5′-ATCGCGCGAT-3′) had the consensus sequence $(Leu/Cys)^{-18}$, Tyr^{-15}, Asn^{-14}, Tyr^{-11}, Arg^{-10}. The clones selected using the reporter plasmid pTS-65 (5′-ATCACGTGAT-3′), in which an additional substitution is made at position 2, had a remarkably similar sequence, Leu^{-18}, Tyr^{-15}, Met^{-14}, Tyr^{-11}, Arg^{-10}. In both cases, there are significant changes in the polarity and size for four of the five randomized amino acids when compared to WT C/EBP—only Arg^{-10} is conserved with WT C/EBP protein.

To characterize the DNA-binding properties of these mutant proteins, *in vivo* DNA-binding experiments were done. Each C/EBP mutant protein was transformed into *E. coli* JM109 carrying reporter plasmids containing the WT and three mutant recognition sequences. The Amp^rSp^r/Amp^r ratios were then measured to determine to what degree the proteins bind to the target sequence and block transcription (Table 3). Transformation of each selected clone into *E. coli* JM109 containing the reporter plasmid against which that mutant was originally selected afforded Amp^rSp^r/Amp^r ratios of >0.50. On the other hand, transformation of each selected clone into cells containing either the WT operator or the other reporter plasmids (against which that particular mutant was not selected) afforded Amp^rSp^r/Amp^r ratios of $<10^{-4}$. These results indicate that all mutants discriminate their target sites with very high specificities *in vivo*. Interestingly, $Ser^{-11} \rightarrow$ Trp and $Ser^{-11} \rightarrow$ His mutant C/EBP proteins, generated independently, showed lower affinities for the mutant sequence 5′-ATTACGTAAT-3′ (Amp^rSp^r/Amp^r ratio, 0.014 and 4×10^{-4}, respectively) than did the mutants MT1 and MT2, which in addition to the Ser^{-11} substitution, contained the additional mutations in the region between the DNA-binding and leucine-zipper regions.

The five mutant C/EBP proteins, MT1–MT5, were then subcloned into a T7-based expression vector and overexpressed. The resulting proteins were purified to homogeneity, and their binding affinities were determined by a gel-

Table 2. Sequences of selected clones

	Residue							
	−18	−15	−14	−11	−10	−9	−7	−3
				pTS-62 (5′-ATTACGTAAT-3′)				
	Asn	Ala	Val	Trp	Arg	Ala		
MT1	AAC	GCG	GTG	TGG	CGG	GCT		
	Asn	Ala	Val	His	Arg		Val	Tyr
MT2	AAC	GCC	GTG	CAC	CGC		GTT	TAC
				pTS-68 (5′-ATCGCGCGAT-3′)				
	Leu	Tyr	Asn	Tyr	Arg			
MT3	TTG	TAC	AAC	TAC	CGC			
	CTG	TAC	AAC	TAC	CGC			
	Cys	Tyr	Asn	Tyr	Arg			
MT4	TGC	TAC	AAC	TAC	AGG			
				pTS-65 (5′-ATCACGTGAT-3′)				
	Leu	Tyr	Met	Tyr	Arg			
MT5	CTG	TAC	ATG	TAC	AGG			
	TTG	TAC	ATG	TAC	CGG			
	CTC	TAC	ATG	TAC	AGG			
	TTG	TAC	ATG	TAC	AGG			
WT	Asn	Ala	Val	Ser	Arg	Asp	Ala	Asn

Underlining represents mutation.

Table 3. *In vivo* DNA-binding assay of C/EBP mutants

	Amp^rSp^r/Amp^r ratio			
	pTS-4 (WT) 5′-ATTGCGCAAT	pTS-62 5′-ATTACGTAAT	pTS-68 5′-ATCGCGCGAT	pTS-65 5′-ATCACGTGAT
pTS-62				
MT1	7.0×10^{-5}	0.61	ND	7.8×10^{-5}
MT2	7.0×10^{-5}	0.65	ND	3.0×10^{-5}
pTS-68				
MT3	4.8×10^{-5}	ND	0.50	1.9×10^{-4}
MT4	7.0×10^{-5}	ND	0.44	1.5×10^{-4}
pTS-65				
MT5	5.0×10^{-5}	ND	1.7×10^{-4}	0.50

Underlining represents mutation. ND, not determined.

retardation assay (Table 4). Each mutant protein binds to the corresponding mutant target site with affinity similar to that of WT C/EBP for the recognition site. WT C/EBP protein showed high specificity, and mutants MT3 and MT4 showed significant specificity for position 3 of the recognition site. Mutant MT5 showed specificity for position 2 of the binding site. Mutants MT1 and MT2 demonstrated relatively low specificity. This contrasts to the specificity seen in the *in vivo* functional assay and may reflect either the differences in assay conditions *in vivo* and *in vitro* or differential sensitivities of the two assay methods to mutations.

DISCUSSION

Recent mutagenesis studies have shown in the case of *Drosophila melanogaster* C/EBP that residues Arg^{-21}, Glu^{-20}, Asn^{-18}, Asn^{-17}, Ser^{-11}, and Arg^{-10}, but not Ala^{-15}, contribute to DNA binding (23). Another study identified Tyr^{-25}, Asn^{-17}, and Val^{-14} as important determinants of DNA specificity (24). The results from the library selection experiments described here in which Asn^{-18}, Ala^{-15}, Val^{-14}, Ser^{-11}, and Arg^{-10} were randomized and selected against the WT recognition sequence show that all five residues are important in selective binding. The only viable mutation that resulted from the selection was that of $Ser^{-11} \rightarrow$ Cys. In the x-ray structure of the GCN4–DNA complex, the side-chain methylene group of the corresponding Ser^{-11} makes a hydrophobic contact with the 5-methyl group of thymine 4. If the same is true of Ser^{-11} in C/EBP, then the methylene side chain of Cys^{-11} in C/EBP might be expected to make a similar contact.

The mutant proteins isolated in selection experiments in which substitutions were made at positions 2, 3, and 4 of the WT recognition sequence can be categorized into two groups: those with amino acid sequence Asn^{-18}, Ala^{-15}, Val^{-14}, Xaa^{-11}, Arg^{-10} that recognize the sequence ATTPuCG-PyAAT) and those with the amino acid sequences Leu/Cys^{-18}, Tyr^{-15}, Xaa^{-14}, Tyr^{-11}, Arg^{-10} (where Xaa-Met or Asn) that recognize the sequence ATCPuCGPyGAT. For both sequences the arginine side chain at position −10 of C/EBP is essential for recognition. Purine-base substitution at position 2 of the recognition site can be accommodated by a one-amino acid substitution at either position −11 or −14 of C/EBP. On the other hand, pyrimidine-base substitution at position 3 results in nonconservative changes in four of the five amino acids of the recognition helix. In addition, in the case of mutants MT1 and MT2, additional mutations were found in the region between the DNA-binding and leucine-zipper domains.

A model for the interaction of these proteins with their recognition site on DNA can be constructed (Fig. 1) from the GCN4–DNA cocrystal structure. In this model the side chain of conserved Arg^{-10} forms specific hydrogen bonds to guanine at position 1′. The side chains of Leu^{-18} and Tyr^{-11} can form a hydrophobic pocket to accommodate the 5-methyl group of thymine at position 4 (in the WT protein, Asn^{-18} appears able to hydrogen-bond to the C4 carbonyl oxygen of T4). The side chain of Tyr^{-15} can be positioned to hydrogen-bond to the sugar phosphate backbone. The base at position 2′ of the recognition site appears to interact with amino acid side chains at position −14 or −11 of the bZip protein. Asn^{-14} in mutant MT3 can hydrogen-bound to the C4 amino group of cytidine 2′, and Met^{-14} in mutant MT5 can make a van der Waal contact with the 5-methyl group of thymine 2′ (in WT C/EBP, Val^{-14} appears to be positioned in close proximity to C5 of cytidine 2′) (24). Trp^{-11} (MT1) and His^{-11} (MT2) appear to be able to interact with thymine 2′ via either hydrophobic or hydrogen-bonding interactions, respectively. Although this model awaits verification by x-ray crystallography or NMR, it suggests that mutant bZip proteins with very little homology to the WT protein can form surfaces that are complementary to the DNA-recognition sites containing one- or two-bp substitutions at positions 2, 3, or 4.

In addition to substitutions in the recognition helix, C/EBP mutants were isolated in the selection against the sequence 5′-ATTACGTAAT-3′, which contained substitutions between the recognition helix and the leucine-zipper domain. These mutations significantly affect the binding interaction with DNA because the corresponding $Ser^{-11} \rightarrow$ Trp or $Ser^{-11} \rightarrow$ His single-site C/EBP mutants (lacking the additional $Asp^{-9} \rightarrow$ Ala mutation in MT1, and $Ala^{-7} \rightarrow$ Val

Table 4. *In vitro* DNA-binding properties

	Concentration of half-maximal binding, M			
	5′-ATTGCGCAAT (WT, pTS-4)	5′-ATTACGTAAT (pTS-62)	5′-ATCGCGCGAT (pTS-68)	5′-ATCACGTGAT (pTS-65)
MT1	4.0×10^{-8}	2.7×10^{-8}	ND	ND
MT2	1.5×10^{-8}	6.4×10^{-9}	ND	ND
MT3	3.8×10^{-7}	ND	4.6×10^{-8}	7.6×10^{-8}
MT4	1.9×10^{-6}	ND	1.0×10^{-7}	1.6×10^{-7}
MT5	6.1×10^{-8}	ND	2.7×10^{-7}	2.1×10^{-8}
WT(Gly^{-40}–Ser^{40})	7.1×10^{-8}	9.5×10^{-8}	2.4×10^{-5}	1.2×10^{-5}

Underlining indicates mutation. Synthetic duplexes consisted of the sequence 5′-TCGACTACAGN_{10}CTGGCAAATT-3′. ND, not determined.

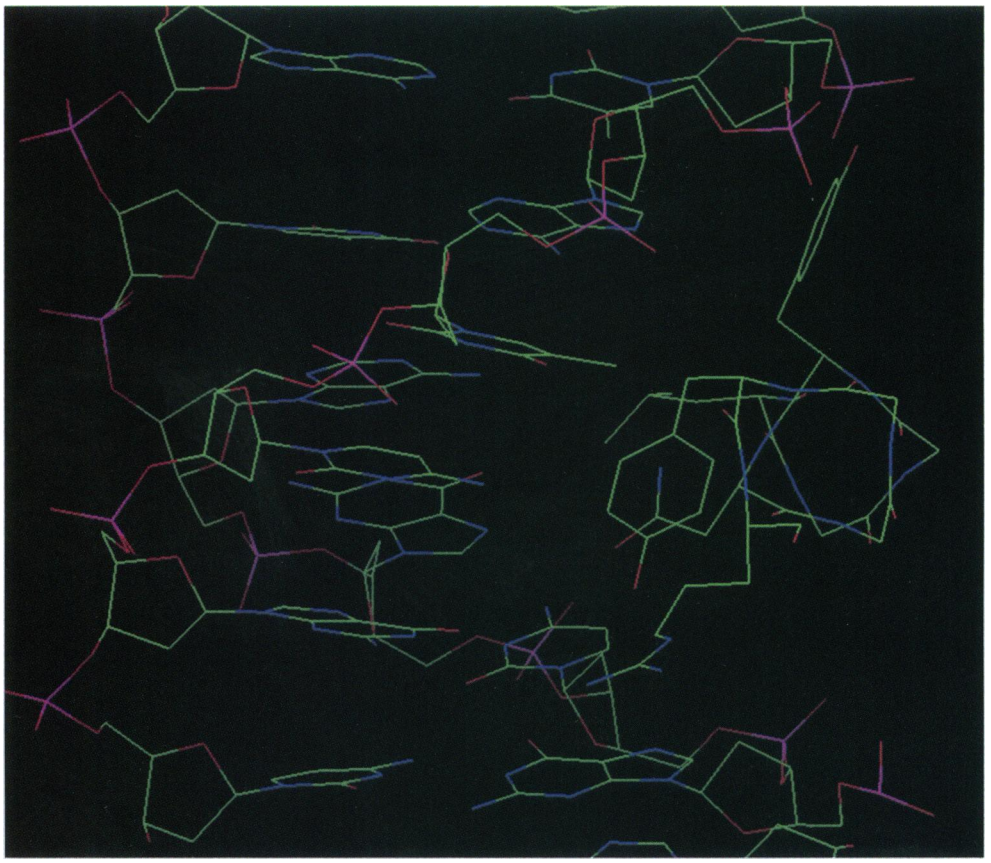

FIG. 1. Model for the interaction of the mutant C/EBP protein, Leu^{-18}, Tyr^{-15}, Asn^{-14}, Tyr^{-11}, Arg^{-10} (MT3), with the recognition site 5′-ATCGCGCGAT-3′.

and $Asn^{-3} \rightarrow Tyr$ mutations in MT2) showed lowered affinities in the *in vivo* transcription interference assay. These substitutions may act by altering the structure of the α-helix, via either side-chain or backbone interactions, to better fit DNA-binding site. Alternatively, an analysis based on the GCN4–DNA complex (4) suggests that tyrosine at position -3 in C/EBP may stack with the corresponding Tyr^{-3} in the paired α-helix. This interaction may change the angle between the recognition helixes and leucine-zipper domain, resulting in improved complementarity to the DNA target site. Previous studies have also emphasized the role of the hinge region in controlling bending and half-site spacing of the target DNA (25, 26).

Although the mutant proteins show high selectivity in the functional *in vivo* transcription interference assay, they show various degrees of specificity in the *in vitro* gel-retardation assays. This behavior has also been reported in other instances (24). *In vitro* competition experiments using plasmid DNA containing the recognition sequences showed no apparent effect of DNA supercoiling on binding or specificity. In addition, no changes in specificity due to changes in salt (27), pH, temperature, bovine serum albumin, or polyethylene glycol (macromolecular crowding) (28) were observed. Other factors not yet apparent from our *in vitro* assays are clearly affecting the *in vivo* activity of these mutant proteins. Alternatively, small changes in binding affinity may lead to large functional differences in the mutant C/EBP proteins *in vivo*.

CONCLUSION

The results of these experiments demonstrate that bZip proteins with altered specificities can be selected from large random libraries. Even in the case of a relatively simple helical protein this may require a series of interdependent substitutions in the protein, indicating that the protein–DNA recognition can involve highly cooperative interactions. The isolation of hinge mutants that also affect affinity reinforces this notion and points out the importance of expanding the libraries to include randomization of additional residues, both at the DNA–protein interface and elsewhere in the protein. The fact that recognition helices containing very different side-chain residues selectively bind operators differing by a single-base substitution also points to the generality of the α-helix framework in DNA recognition. Libraries of other natural or synthetic α-helical peptides may also prove useful sources of selective DNA-binding ligand and provide increased insight into the requirements for selective DNA recognition.

This work was supported by the National Institutes of Health (Grant no. R01GM41679) and the Director, Office of Energy Research, Office of Biological and Environmental Research, General Life Sciences Division, of the U.S. Department of Energy under contract DE-AC03-76SF00098. P.G.S. is a Howard Hughes Medical Institute Investigator.

1. Landschultz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240,** 1759–1764.
2. Hu, J. C. & Sauer, R. T. (1992) *Nucleic Acids Mol. Biol.* **6,** 82–101.
3. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249,** 769–771.
4. Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) *Cell* **71,** 1223–1237.
5. Konig, P. & Richmond, J. T. (1993) *J. Mol. Biol.* **233,** 139–154.
6. Glover, J. N. & Harrison, S. C. (1995) *Nature (London)* **373,** 257–261.
7. Rhodes, D. S. & Klug, A. (1993) *Sci. Am.* **268,** 56–65.
8. Harrison, S. C. & Aggarwal, A. K. (1990) *Annu. Rev. Biochem.* **59,** 933–969.
9. Raumann, B. E., Brown, B. M. & Sauer, R. T. (1994) *Curr. Opin. Struct. Biol.* **4,** 36–43.

10. Wharton, R. P. & Ptashne, M. (1987) *Nature (London)* **316,** 601–605.
11. Suckow, M., Madan, A., Kisters-Woike, B. Wilcken-Bergmann, B. V. & Muller-Hill B. (1994) *Nucleic Acids Res.* **22,** 2198–2208.
12. Nelson, H. C. & Sauer, R. T. (1985) *Cell* **42,** 549–558.
13. Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. (1995) *Science* **267,** 93–96.
14. Park, C., Campbell, J. L. & Goddard, W. A., III (1995) *J. Am. Chem. Soc.* **117,** 6287–6291.
15. Rebar, E. J. & Pabo, C. O. (1994) *Science* **263,** 671–673.
16. Jamieson, A. C., Him, S.-H. & Wells, J. A. (1994) *Biochemistry* **33,** 5689–5695.
17. Choo, Y. & Klug, A. (1991) *Proc. Natl. Acad. Sci. USA* **91,** 11163–11167.
18. Huang, L., Sera, T. & Schultz, P. G. (1994) *Proc. Natl. Acad. Sci. USA* **90,** 6340–6344.
19. Struhl, K. (1992) *Transcription Regulation* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 833–859.
20. Elledge, S. & Davis, R. (1989) *Genes Dev.* **3,** 185–197.
21. Elledge, S. & Davis, R. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3689–3693.
22. Landschultz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. (1988) *Genes Dev.* **2,** 786–800.
23. Rorth, P. (1994) *Science* **266,** 1878–1881.
24. Johnson, P. F. (1993) *Mol. Cell. Biol.* **13,** 6919–6930.
25. Paolella, D. N., Palmer, R. C. & Schepartz, A. (1994) *Science* **264,** 1130–1133.
26. Kim, J., Tzamaris, D., Ellenberger, T., Harrison, S. C. & Stuhl, K. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6340–6344.
27. Mossing, M. C. & Record, M. T., Jr. (1985) *J. Mol. Biol.* **186,** 295.
28. Jarvis, T. C., Ring, D. M., Daube, S. S. & von Hippel, P. H. (1990) *J. Biol. Chem.* **265,** 15160–15167.