4348–4355 Nucleic Acids Research, 1993, Vol. 21, No. 18

Dimerization of leucine zippers analyzed by random selection

William T.Pu and Kevin Struhl*
Department Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Received April 8, 1993; Revised and Accepted July 16, 1993

ABSTRACT

The leucine zipper is a coiled coil that mediates specific dimerization of bZIP DNA-binding domains. A hydrophobic spine involving the conserved leucines runs down the coiled-coil and is thought to stabilize the dimer. We used the method of random selection to further define the primary sequence requirements for homodimer formation and heterodimer formation with Fos. When positions on either side of the hydrophobic spine of GCN4 are diversified to include the corresponding residues of Jun, a large percentage of the resulting sequences form homodimers, and a large percentage form heterodimers with Fos. Basic residues were preferred, but not essential, at position e of zippers which heterodimerize with Fos. When random sequences containing 5 heptad repeat of leucines are subject to a selection for homodimer formation, a diverse set of sequences is isolated. Certain residues are preferred at each position in the heptad repeat, although no essential primary sequence determinants could be identified. No pair of residues not involving the conserved leucines could be identified which strongly promotes homodimerization. These results suggest that factors determining leucine zipper dimerization are complex, with numerous interactions contributing to the association.

INTRODUCTION

A number of eukaryotic transcriptional regulatory proteins belong to the bZIP class of DNA-binding proteins. The bZIP motif consists of two independent subdomains. A region rich in basic amino acids directly contacts DNA, while an adjacent heptad repeat of leucines, the leucine zipper, mediates dimer formation (1, 2). bZIP proteins bind to dyad-symmetric binding sites as dimers, and often have distinct dimerization properties. For example, GCN4 forms homodimers, Jun dimerizes with itself or with Fos, and Fos only dimerizes with Jun (3–9). Domain-swap experiments indicate that the leucine zipper is necessary and sufficient to determine dimerization specificity (10, 11). These dimerization properties create a large repertoire of distinct complexes with different DNA-binding specificities and transcriptional regulatory properties. Several instances in which the composition of a regulatory dimer has physiological consequences have been described (12–15).

The leucine zipper mediates dimer formation by forming a coiled-coil (16). In standard coiled-coil notation, the positions in the heptad repeat are labelled **a**–**g**; the leucines occur at the fourth position of a heptad, position **d** (Fig. 1). A high resolution crystal structure of a GCN4 leucine zipper peptide homodimer shows that the stability of the leucine zipper is mainly due to a hydrophobic spine created by groups at positions **a** and **d**. Systematic mutagenesis of leucines at position **d** and hydrophobic residues at position **a** demonstrate that alteration of two or more of these residues has a significant functional effect (17–19). Methylene groups of predominantly charged residues at positions **e** and **g** also pack against this hydrophobic spine and contribute to the stability of the coiled-coil. The terminal charged groups of residues at positions **e** and **g′** (prime indicates the partner monomer) of the preceding heptad can make interhelical contacts. Positions **b**, **c**, and **f** are on the solvent-exposed face of the coiled-coil and are not in position to make interhelical interactions (20).

Despite this structural information, the features of a leucine zipper in an intact bZIP domain which permit or prevent it from associating with the zipper of another bZIP domain are incompletely understood. Transplanting the 8 residues from Fos and from Jun at positions **e** and **g** into the corresponding positions of a GCN4 leucine zipper peptide results in preferential heterodimer formation (21). Co-immunoprecipitation experiments involving full-length Fos and Jun containing systematically mutagenized leucine zippers generally support the importance of residues **e**, and **g** for preferential heterodimer formation (18). However, these experiments were limited in a number of ways. First, because only specific sequences were tested for the ability to dimerize, the results may not be generalizable to other zippers. Second, the experiments were performed *in vitro* at concentrations much higher than present under physiological conditions. Third, only dimerization, not DNA-binding, was examined; the requirements for each may be distinct. In the crystal structure of a GCN4 leucine zipper peptide, the subunits of the leucine zipper dimer pack closely together at their N-terminal ends (20);

* To whom correspondence should be addressed

in comparison, the cocrystal structure of the GCN4 bZIP domain bound to DNA shows the N-terminal ends of the zipper smoothly diverging so that the basic regions can make appropriate DNA contacts (22).

To better define the sequence determinants of leucine zipper dimer stability and specificity, we used two approaches. First, we examined the primary sequence differences between Jun and GCN4 which allow Jun but not GCN4 to dimerize with Fos. Second, we sought to identify primary sequence patterns which contribute to homodimer formation. For both approaches, we used the method of random selection (23). Degenerate oligonucleotides were used to generate a large collection of sequences, from which functional zipper sequences were isolated using genetic selections for DNA-binding or dimerization *in vivo*.

MATERIALS

DNAs

Oligonucleotides used in this study are shown in Fig. 2. GJ12, GJ13, and ZIP-LIB are degenerate, double-stranded DNAs synthesized by mutually primed synthesis (24). GJ13 was constructed by annealing and extending the complementary 3′ ends of GJ1 and GJ3. GJ12 and ZIP-LIB were constructed similarly by combining GJ1 and GJ2, and ZIP-TOP and ZIP-BOT, respectively. YCp88-Sc4400 is a centromeric yeast plasmid which expresses GCN4 (25). YCp88-Sc4324 is a similar plasmid which expresses the C-terminal 131 amino acids of GCN4 (ΔAct-GCN4) and therefore lacks an activation domain (26). SO12, a centromeric plasmid which expresses c-Fos in yeast cells, has been previously described (27). pRW88 express 434 repressor in *E.coli* using a lacUV5 promoter (28).

CS15S was derived from YCp88-Sc4400 by using oligonucleotides S1 and S2 to eliminate the unique *Sac*I site and introduce a new *Sac*I site after the first leucine of the leucine zipper. CS18 was constructed by cloning the *Kpn*I−*Eco*RI fragment of CS15S into YCp88-Sc4324 and then inserting oligonucleotide FLU-1 encoding the influenza-derived epitope HA-1 (29) into the *Kpn*I site.

Libraries expressing GCN4 derivatives containing randomly chimeric GCN4/Jun leucine zippers were constructed by cloning the *Sac*I−*Eco*RI fragment of GJ13 into similarly digested CS15S or CS18, resulting in libraries named Act-GJ13 and ΔAct-GJ13, respectively. The corresponding library constructed using the *Sac*I−*Hin*dIII fragment of GJ12 is named ΔAct-GJ12. The libraries Act-GJ13 and ΔAct-GJ13 each contain 5×10^5 independent clones, while the library ΔAct-GJ12 contains only 2000. The actual degeneracy of the 9 diversified positions in each library was determined by sequencing clones chosen at random. The nucleotide distribution at these degenerate positions is skewed in a position dependent manner, and there is a dearth of adenines in each of the synthesized strands. Since libraries Act-GJ13 and ΔAct-GJ13 were constructed using the same pair of oligonucleotides, the skew is the same in these two libraries.

To fuse DNA-binding domain of 434 repressor to GCN4, oligonucleotides LacUV5-3 and 434-BOT were used to PCR amplify pRW88. The *Pst*I−*Bam*HI digested PCR product was cloned a derivative of pTZ18R (Pharmacia) in which the *Eco*RI site had been destroyed. The *Eco*RI site in the cloned fragment was then eliminated by oligonucleotide-directed mutagenesis using 434-MUT. The resulting plasmid has a modified 434 gene in which the region encoding the N-terminal 110 residues is fused to an oligonucleotide containing *Xho*I and *Eco*RI sites; the C-terminal dimerization domain is deleted. The random leucine zipper library ZIP-LIB, or the GCN4 leucine zipper, was inserted into these sites to create 434-ZIP-LIB or 434-GCN4, respectively. 434-ZIP-LIB contains 5×10^7 clones, and sequences of several randomly chosen clones indicate that the nucleotide distribution is not position-dependent or excessively skewed.

Phenotypic analysis

Yeast strain KY372 is the same as KY803 (*trp1*-Δ1 *ura3*-52 *leu2*-P1 *gcn4*-Δ1) (26), except there are two GCN4 binding sites upstream of the *his3* promoter (27). For phage analysis, *E.coli* strain XA90 ($lacI^{Q1}$), and phage λimm^{P22}(clear) and λimm^{434}(clear) (28) were used.

Function of GCN4 derivatives in which a test zipper replaced the GCN4 zipper was monitored in yeast by measuring the ability of the hybrid protein to activate transcription of *his3*. To monitor function as a homodimer in yeast, the standard complementation assay (26) was used. In this assay, KY803 expressing a derivative of GCN4 is grown in the presence of aminotriazole (AT), a competitive inhibitor of the *his3* gene product. To monitor function as a heterodimer in yeast, a modified complementation assay was used (27). Strain KY372 was transformed with plasmids expressing c-Fos (SO12) and a derivative of ΔAct-GCN4. The AT resistance of this doubly transformed strain was then monitored as described previously (26). Plasmids of interest were isolated from yeast, and the regions encoding the leucine zipper were sequenced.

A phage immunity assay was used to select and test 434 repressor-leucine zipper fusion proteins. To test a zipper fusion, *E.coli* strain XA90 transformed with the appropriate fusion protein expression plasmid was cross-streaked against 10^6-10^{11} pfu/ml of λimm^{434}(clear) on LB-agar plates containing ampicillin and 10 μM IPTG. For selection of functional proteins from a library of fusions containing a random zipper, XA90 transformed with 434-ZIP-LIB was plated on LB-agar containing ampicillin, 1.5mM IPTG, and 10^{10} pfu/ml λimm^{434}(clear). DNA recovered from colonies growing on these plates were reintroduced into XA90 and replated. DNA was prepared from individual colonies growing on these plates, reintroduced into XA90, and tested for immunity to infection by λimm^{434}(clear) as described above.

Statistical analysis

The distribution of amino acids present in functional zippers was analyzed by helical position. To determine if position-dependent variations in the frequency of an amino acid were significant, we calculated the predicted frequency at which each residue should have occurred in the unselected pool based on the measured nucleotide frequencies in the library. The observed frequency in functional zippers at each helical position was then compared to the predicted frequency using a one-sample test for a binomial proportion, assuming the sampling error conformed to a normal distribution: λ = abs {(po−pc)/(sqrt (pc * (1−pc)/n)}, and p-value = 2 * (1− $\Phi(\lambda)$), where po is the observed frequency, pc is the calculated frequency, n is the number of samples, $\Phi()$ is the standard normal distribution, and the p-value is the likelihood the difference is due to chance sampling error.

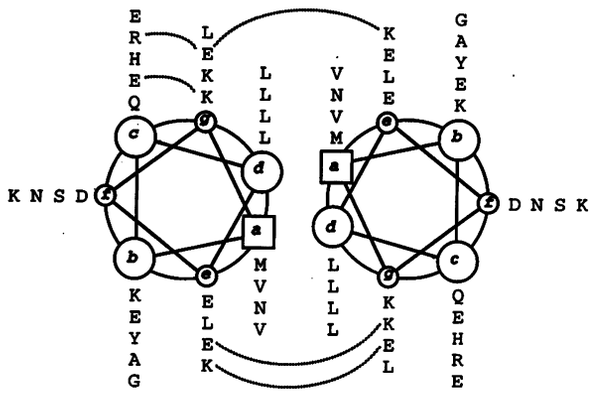

Figure 1. Helical wheel representation of a GCN4 homodimer. View is from the N-terminus. Residues from in the first two helical turns are boxed or circled. Helical positions are labeled **a**–**g**. Residues that form ion pairs in the crystal structure are connected with dashed lines. The dimer interface ('inside residues') is made up of hydrophobic residues from positions **a** and **d**, as well as residues at positions **e** and **g**. Residues at positions **b**, **c**, and **f** ('outside residues') are not in position to make interhelical contacts, but can make intrahelical contacts. Adapted from (20).

GJ1	GAAGAGCTCNNAAGCNAAACTACCACCTGGNGACNNGGNGGCACGTCTG
GJ2	ACTGAATTCTCATTAACGCTCACCCACCNGCTTCNTCAGACGTGC
GJ3	GGCGAATTCATTAGTTCATCACCTTTTGCTTCAGTTGAGCCACCNGCTTCNTCAGACGTGC
ZIP-TOP	GGCTCGAG(NNB)$_5$CTG(NNB)$_6$CTG(NNB)$_4$GAGGAACTG
ZIP-BOT	CCGAATTCATTAV(NNV)$_5$NNCAGV(NNV)$_5$NNCAGTTCCTC
S1	GGGAAGCTTTTTCAGACGTGCCACTTCGTTTTCTAAGTGGTAGTTTTT
S2	TAGCTCGAGGATAAAGTGGAAGAGCTCTTAAGCAAAAACTACCAC
FLU-1	TAGGTACCATATCCCTATGACGTGCCCGACTATGCGTCCCTCAGATCTCTGGTACCAG
434-MUT	GTCCAAGCTGAATCCTTTTGCTTTTTACC
434-BOT	CCGGATCCGAATTCGGCCCTCGAGGTCACTGTCATACCA
LacUV5-3	CCCTGCAGCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGT

Figure 2. Oligonucleotides. Degenerate oligonucleotides used to construct libraries amd for PCR mutagenesis of 434 repressor (N, 25% each G, A, T, C; V, 33% each G A C; B, 33% each G, T, C). Greek letters: labels for diversified nucleotides.

Heptad	1	2	3	4	5	
Position	abcdefg	abcdefg	abcdefg	abcdefg	abcdefg	
GCN4	MKQLEDK	VEELLSK	NYHLENE	VARLKKL	VGER	
Jun	IARLEEK	VKTLKAQ	NSELAST	ANMLREQ	VAQLKQK	VMN
GJ13	MKQLEDK	VEEL1S2	NYHL3N4	5ARL6E7	VAQLKQK	VMN
GJ12	MKQLEDK	VEEL1S2	NYHL3N4	5ARL6E7	VGER	

1 = GEAVRKTIMQPLS#
2 = EKQ#
3 = GEAV
4 = GEAVRKTIMQPLS#
5 = GEAV
6 = RKTM
7 = RQPL

Figure 3. Design of GCN4/Jun randomly chimeric zippers. Amino acid sequence of GCN4, Jun, and GCN4/Jun randomly chimeric zipper libraries GJ12 and GJ13. Underlined, numbered residues in GJ13 indicate 'inside' positions which are divergent between GCN4 and Jun. These positions were diversified to include the indicated residues.

RESULTS

GCN4/Jun random chimeric zippers

In coiled-coiled structures, residues at positions **a**, **d**, **e**, and **g** ('inside' residues) are in position to make interhelical contacts (20, 22). Out of 17 shared inside positions, Jun and GCN4 differ at only seven, and six of these divergent residues occur at positions **e** and **g**. To further study the role of the seven divergent inside residues, we constructed 'random chimeras' in which the codon for each divergent residue of the GCN4 zipper is diversified so that it can encode a number of other residues, including the corresponding residue of Jun (Fig. 3). Jun and GCN4 also differ in the length of their leucine zippers: GCN4 has only four heptad repeats of leucines, compared to five in Jun and Fos. It was not clear if the last heptad repeat is crucial for the interaction between Jun and Fos. Therefore, in addition to the four heptad repeat random chimera made by diversifying the GCN4 zipper (GJ12), we made a five heptad repeat random chimera by fusing the fifth heptad repeat of Jun onto the diversified GCN4 zipper (GJ13).

To identify random chimeric zippers which form homodimers, we constructed Act-GJ13, a library of GCN4 derivatives in which the GCN4 zipper is replaced by the random chimeric zipper library GJ13. Derivatives in which the random chimeric zipper supports DNA-binding by the GCN4 bZIP domain are selected by their ability activate transcription of *his3*, thereby making the yeast cells resistant to aminotriazole, a competitive inhibitor of the *his3* gene product. To isolate random chimeric zippers which interact with Fos, we coexpressed Fos and GCN4 derivatives from the ΔAct-GJ13 library in yeast. In addition to containing the random chimeric zipper library GJ13 in place of the native zipper, these GCN4 derivatives also lack an activation domain and therefore do not activate *his3* expression. Fos contains an activation domain, but because Fos cannot bind DNA efficiently as a homodimer, it cannot activate *his3* expression independently. However, DNA-binding heterodimers between Fos and a ΔAct-GJ13 derivative should be able to activate his3 expression, since a dimer containing a single activation domain is able to activate transcription (27). Thus, we found that yeast cells which coexpress Fos and ΔAct-GCN4 are not aminotriazole resistance, but cells coexpressing Fos and a derivative of ΔAct-GCN4 which contains the Jun zipper in place of the GCN4 zipper are aminotriazole resistant.

Approximately one third of yeast cells transformed with the library Act-GJ13 expressed zippers which allowed the proteins to form DNA-binding homodimers. Of the remaining two thirds, approximately half contained stop codons. Similarly, when yeast cells expressing Fos were transformed with the library ΔAct-GJ13, one third of transformants expressed zippers which were capable of forming DNA-binding heterodimers with Fos, one third were nonfunctional because of stop codons, and one third were nonfunctional for other reasons.

	2	3	4
	abcdefg	abcdefg	abcdefg
Homodimers:			
Act-GJ13-10	VEELLSE	NYHLVNS	AARLKER
-12	....L.Q	....V.L	G...K.P
-14	....L.Q	....V.S	A...R.Q
-16	....L.E	....V.S	V...M.R
-26	....L.K	....V.L	V...K.Q
-27	....S.E	....V.L	V...R.Q
-28	....L.Q	....V.L	V...K.Q
-29	....S.E	....V.L	V...K.Q
-33	....L.E	....V.L	G...R.Q
-34	....S E	...V.L	V...K.Q
-35	....L.Q	....V.V	A...R.R
Heterodimers			
ΔAct-GJ13-14	....L.Q	....G.W	A...K.L
-15	....S.Q	....V.V	A...R.Q
-16	....R.Q	....A.L	V...R.Q
-17	....S.E	....V.L	V...K.Q
-18	....R.Q	....A.V	G...R.R
-19	....R.E	....G.L	V...K.Q
-20	....R.Q	....A.V	G...R.R
-22	....R.E	....A.L	V...R.Q
-23	....R.E	....G.V	V...K.Q
-24	....R.Q	....G.W	V...K.Q
-25	....L.Q	....V.L	A...R.Q
-26	....L.Q	....V.L	V...K.Q
-27	....R.E	....A.L	V...K.Q
-28	....R.E	....V.L	A...R.L
-29	....L.Q	....A.W	A...K.Q
-30	....R.Q	....V.L	A...R.R
-31	....L.Q	....G.L	V...R.Q
-32	....L.Q	....G.L	V...R.Q
-33	....G.E	....V.L	A...K.Q
-34	....S.Q	....V.L	A...R.Q
-35	....S.Q	....G.L	A...R.Q
-36	....R.Q	....V.L	V...R.Q
-37	....L.Q	....A.L	A...R.Q
-38	....R.Q	....A.W	A...K.Q
Not Homodimers			
Act-GJ13-5	....R.Q	....V.V	G...K.Q
-7	....G.Q	....A.P	A...M.R
-25	....E.Q	....V.L	E...K.Q
Not Heterodimers			
ΔACT-GJ13-40	....L.Q	....V.L	A...K.Q
-41	....L.E	....V.A	A...K.Q
-42	....L.Q	....V.W	G...R.R

Figure 4. Amino acid sequences of heptads 2–4 of zippers isolated from GCN4/Jun randomly chimeric zipper libraries. Zippers selected for homodimerizing activity and heterodimerizing activity with Fos are listed under the headings 'homodimers' and 'heterodimers', respectively. Zippers which were not functional as homodimers or heterodimers are listed under 'Not homodimers' and 'Not heterodimers', respectively. Nonfunctional zippers containing stop codons are omitted from the table. Sequences of GCN4 and Jun native zippers are shown for comparison. In the zippers selected from the libraries, only diversified positions (marked by asterisks) are listed. Other residues, indicated by '.', are identical to those shown for Act-GJ13-10. Helical positions are labelled a–g. '?' indicates an ambiguous amino acid residue due to poor sequence quality.

Zipper	a	b	c	d	e	f	g	Fn
GCN4	M	K	Q	L	E	D	K	11
	V	E	L	L	S	K	Q	
	N	Y	H	L	E	N	E	
	V	A	R	L	K	K	L	
	V	G	E	R				
FOS	T	D	T	L	Q	A	E	**
	T	D	Q	L	E	D	K	
	K	S	A	L	Q	T	E	
	I	A	N	L	L	K	E	
	K	E	K	L	E	F		
ZIP-4	M	K	Q	L	E	R	A	10
	L	H	G	L	R	E	V	
	H	A	S	L	R	S	I	
	T	E	E	L	W	F	I	
	L	R	S	L	D	A	H	
	N	G	A					
ZIP-6	M	K	Q	L	E	V	M	11
	T	R	R	L	D	D	V	
	R	S	L	L	Q	A	I	
	G	E	E	L	S	L	L	
	Q	A	L	L	G	S	H	
	R	F	T					
ZIP-8	M	K	Q	L	E	G	G	9
	L	S	S	L	A	T	A	
	T	A	S	L	H	D	R	
	V	E	E	L	Y	S	R	
	L	L	L	L	L	R	D	
	R	G	V					
ZIP-10	M	K	Q	L	E	V	H	7
	N	T	V	L	T	C	W	
	T	S	T	L	V	P	G	
	N	E	E	L	L	A	L	
	G	I	F	L	M	Y	V	
	T	L	H					
ZIP-12	M	K	Q	L	E	E	S	11
	L	R	T	L	G	T	H	
	L	E	R	L	G	A	L	
	V	E	E	L	E	S	L	
	T	S	V	T	L	C	A	D
	T	L	T					
ZIP-14	M	K	Q	L	E	Q	T	9
	L	S	L	L	G	S	R	
	A	H	C	L	G	A	R	
	V	E	E	L	K	E	L	
	V	S	R	L	A	A	A	
	L	A	H					
ZIP-16	M	K	Q	L	E	S	T	11
	R	D	R	L	E	R	V	
	I	R	F	L	T	W	A	
	V	E	E	L	A	S	F	
	S	A	S	L	D	T	R	
	S	R	P					
ZIP-18	M	K	Q	L	E	V	N	9
	A	G	T	L	S	L	E	
	T	R	S	L	E	T	L	
	A	E	E	L	L	I	L	
	I	R	K	L	N	R	Y	
	P	I	P					

Zipper	a	b	c	d	e	f	g	Fn
JUN	I	A	R	L	E	E	K	NT
	V	K	T	L	K	A	Q	
	N	S	E	L	A	S	T	
	A	N	M	L	R	E	Q	
	V	A	Q	L	K	Q		
ZIP-3	M	K	Q	L	E	R	A	9
	R	A	T	L	P	A	A	
	L	R	D	L	R	G	G	
	L	E	E	L	S	V	A	
	I	G	N	L	A	S	G	
	C	R	D					
ZIP-5	M	K	Q	L	E	Y	K	10
	H	V	A	L	Y	V	L	
	C	D	D	L	G	Q	R	
	E	E	E	L	I	H	L	
	L	Y	E	L	P	#	P	
	G	H	T					
ZIP-7	M	K	Q	L	E	D	A	8
	V	D	R	L	N	A	A	
	V	A	R	L	G	G	N	
	A	E	E	L	L	R	C	
	W	L	H	L	T	D	K	
	L	C	#					
ZIP-9	M	K	Q	L	E	G	R	9
	#	T	F	L	E	A	S	
	A	W	G	L	G	T	T	
	V	E	E	L	H	S	S	
	G	D	G	L	Q	G	F	
	L	P	E					
ZIP-11	M	K	Q	L	E	W	A	9
	I	K	R	L	C	A	I	
	N	A	F	L	A	R	L	
	E	E	E	L	S	D	L	
	P	S	T	L	Q	D	L	
	L	P	R					
ZIP-13	M	K	Q	L	E	W	A	9
	I	K	R	L	C	A	I	
	N	A	F	L	A	R	L	
	E	E	E	L	S	D	L	
	P	S	T	L	Q	D	L	
	L	P	R					
ZIP-15	M	K	Q	L	E	E	L	9
	F	S	G	L	Q	G	A	
	Y	A	H	L	A	M	V	
	V	E	E	L	G	K	V	
	I	Q	A	L	G	A	R	
	L	P	P					
ZIP-17	M	K	Q	L	E	K	G	11
	I	W	I	L	V	T	W	
	G	K	#	L	G	E	L	
	A	E	E	L	T	G	S	
	N	D	L	L	N	I	K	
	P	S	H					
ZIP-21	M	K	Q	L	E	Q	R	11
	M	T	N	L	S	G	V	
	L	L	T	L	R	K	M	
	R	E	E	L	D	C	V	
	T	H	V	L	E	S	Y	
	F	N	R					

Figure 5. Amino acid sequences of zippers isolated from the random zipper library. Fusions of the DNA-binding domain of 434 repressor to a random zipper library were subjected to selection for homodimer function as described in the text. Zippers which supported homodimer formation were sequenced, and the inferred amino acid sequences are displayed as heptad repeats, with helical positions labelled A–G. The function of each zipper was quantitated by cross-streaking against 10^6–10^{11} pfu/ml phage, and recorded under 'Fn' as the highest titer to which cells expressing the 434-zipper fusion were immune in the presence of 10 μM IPTG. **, nonfunctional fusion protein. #, amber stop codon. Prolines and stop codons are shown in boldface.

0/72 zippers from the ΔAct-GJ12 library functioned as heterodimers with Fos. The chimeric zippers in this library contain four heptad repeats, while those in ΔAct-GJ13 contain five. This result suggest that heterodimer formation with Fos, which contains five heptad repeats, requires a zipper that also contains five repeats. The final repeat may provide essential stabilizing contacts, or alternately, mismatched zipper lengths may be destabilizing.

Sequences of zippers that were functional and nonfunctional as homodimers or heterodimers are shown in Fig. 4. There are no stop codons in the functional zippers, despite being present at high frequency in the unselected library. Helix-breaking prolines also do not occur in the functional zippers, except at the position prior to the final heptad repeat for zippers which form homodimers. Interestingly, they do not occur at this position for zippers which form heterodimers, in agreement with our finding that the terminal heptad repeat is important for heterodimer formation. Comparison of the sequences shows that a large number of residues at each position is compatible with both homodimer and heterodimer formation, and no single residue or group of residues uniquely determines whether the zipper will function as a homodimer or as a heterodimer with Fos.

Certain nonfunctional zippers are only slightly different from functional ones, so that it is not possible to predict whether a zipper will be functional or not. For instance, functional zipper ΔAct-GJ13-25 differs from the nonfunctional zipper ΔAct-GJ13-40 by a single residue, and this divergent residue is present in several functional zippers. Several zippers are nonfunctional despite having the hydrophobic spine of GCN4 or Jun (e.g. Act-GJ13-25 and ΔAct-GJ13-40), suggesting that either the residues outside of this hydrophobic spine destabilize the dimer, or fail to provide essential, stabilizing interactions.

Table 1. Amino acid distribution of GCN4/Jun chimeric zippers[a]

	e_2	g_2	e_3	g_3	a_4	e_4	g_4
G	4,0,4	–	8,0,29	0	46,18,8	–	4,0,0
E	–	17,55,29	0	0	4,0,0	–	–
A	–	–	12,0,33	0	8,27,50	–	–
V	12,0,0	–	75,100,38	29,9,17	21,55,42	–	–
R	4,0,50	–	–	4,0,0	–	21,36,58	29,27,12
K	0	0,9,0	–	0	–	58,55,42	4,0,0
T	0	–	–	0	–	8,0,0	–
Q	0	42,36,71	–	0	–	–	33,64,79
P	4,0,0	–	–	12,0,0	–	–	4,9,0
L	21,73,29	–	–	21,64,67	4,0,0	–	17,0,8
S	12,27,17	–	–	8,27,4	4,0,0	–	–
M	0	–	–	0	–	4,9,0	–
W	–	–	–	17,0,0	–	0	–
Stop	21,0,0	38,0,0	4,0,0	0	–	–	–

[a]Amino acid usage at each diversified position of unselected, homodimerizing, and heterodimerizing zippers is shown as percentages. Columns are labelled by the diversified position as indicated in Fig. 3. Rows are labelled by amino acid residue. Residues which were not programmed at a position are indicated by '–'. Residues which were not included at a position due to skewed nucleotide distribution in the oligonucleotides are indicated by 0.

Table 2. Amino acid distribution of homodimerizing random zippers[a]

	A	B	C	D	E	F	G	Predicted
A	7	15 #	4	0	10	15 #	14 *	8
F	2	1	7 @	0	0	1	2	2
I	8 @	3	1	0	1	2	6 *	3
L	20 @	7	7	100	6	2 *	20 @	7
V	11 #	1	6	0	3	6	9	5
W	1	3	0	0	1	4	2	2
Y	1	1	0	0	3	2	2	3
G	6	6	6	0	16 @	9	6	8
C	2	1	1	0	4	2	1	3
S	2 #	15	9	0	9	12	5	10
N	7	1	3	0	4	0 *	2	4
Q	1	1	0	0	7 #	4	0	3
T	11	4	15 @	0	6	7	4	7
H	2	6	7	0	3	1 #	5	4
M	1	0	0	0	1	1	2	2
D	0 *	7	4	0	6	9 #	2	4
E	4	1	3	0	7 @	6 *	1	3
K	0	4	1	0	1	4	4	2
R	7	13	16	0	6	9	11	11
P	5	6	4	0	3 #	1 @	1 @	9
Stop	1	0	1	0	0	1	0	2

[a]Amino acid usage, displayed as percentages, at each helical position of zippers selected for homodimer formation from the random zipper library. Columns are labelled by helical position. The column labelled 'predicted' indicates the calculated frequency, based upon the measured nucleotide distribution and the genetic code. Rows are labelled by amino acid residue. * indicates significant with $p \leq 0.10$; #, $p \leq 0.05$; @, $p \leq 0.01$.

We compared the amino acid distribution at each degenerate position in unselected zippers, zippers selected for homodimer formation, and zippers selected for heterodimer formation with Fos (Table 1). We found that at each position different residues are preferred for formation of heterodimers with Fos or for formation of homodimers. In most cases the preferred residues for homodimers are similar to the native residue found in GCN4, and the preferred residues for heterodimers are similar to the native residue in Jun. At position a_4 (subscripts denote a specific heptad repeat of the coiled-coil, e.g. a_4 refers to position **a** in the fourth heptad from the N-terminus of the zipper), the only component of the hydrophobic spine (residues **a** and **d**) which is not conserved between GCN4 and Jun, the native GCN4 residue valine is preferred over the native Jun residue alanine for homodimer formation. However, for heterodimerization with Fos, alanine is preferred over valine. The β-branched aliphatic chain of valine may pack more efficiently than alanine in the homodimer but sterically clash with the isoleucine residue from position a_4 of Fos in the heterodimer (20). At positions e_2 and e_4, basic residues are favored in zippers selected for heterodimer formation with Fos, while no such preference is seen in zippers selected for homodimer formation. We could not test for such a preference at position e_3, since basic residues were not represented at that position due to the skewed nucleotide distribution. Charge neutralization between basic residues at position **e** of a Jun zipper and acidic residues at position **g** of

a Fos zipper has been shown to be important for preferential heterodimer formation by isolated zipper peptides (21). At position $\mathbf{g_2}$, the charged residues glutamate and lysine are preferred over the polar residue glutamine for homodimer formation, whereas this is not the case for heterodimer formation with Fos. Position $\mathbf{g_2}$ of GCN4 contains a lysine, and position $\mathbf{g_2}$ of Jun contains glutamine.

While there are preferred residues at each position, there is no identifiable set of residues which is crucial for homodimerization or heterodimerization with Fos. For instance, while basic residues at position **e** favor heterodimer formation with Fos, some of the isolated heterodimerizing zippers, like GCN4, had only one basic residue at this position (excluding the undiversified fifth heptad). Similarly, at position **a**, alanine is favored for heterodimer formation but valine is tolerated in numerous zippers.

Selection of functional 434 repressor-zipper fusions

In a different approach to better understand primary sequence requirements for homodimer formation, we constructed a random leucine zipper library with 5 complete heptad repeats of leucines. Every position other than **d** was made random, except for two 'outside' residues which were held constant to facilitate construction of the random zipper (Fig. 1).

Sequences which encoded homodimerizing zippers were selected using a variant of a previously described *in vivo* dimerization assay (17). *E.coli* expressing the repressor protein of phage 434 are immune to infection by lambda phages containing the 434 immunity region. The ability of the 434 repressor to confer immunity is dependent upon a C-terminal dimerization domain that can be functionally replaced by a leucine zipper dimerization element. Thus, *E.coli* expressing 434-Δ or 434-Fos, which lack homodimerization elements, can be infected by λimm^{434}(clear), while cells expressing 434-GCN4 or 434-Jun are immune to infection. The immunity is specific since 434-GCN4 does not confer resistance to λimm^{P22}(clear), a phage containing a different immunity region. Sequences which support dimerization in this context are not necessarily functional in supporting dimerization and DNA-binding in the context of the intact bZIP domain.

The random leucine zipper library containing five heptad repeats of leucines was fused to the N-terminal 110 residues of 434 repressor to create 434-ZIP-LIB. We predict that 90% of the molecules in the library do not encode a functional protein for trivial reasons, such as nonsense codons and absence of a zipper insert. *E.coli* strain XA90 was transformed with the library, and unselected colonies were tested for immunity to λimm^{434}(clear); 0/100 colonies were immune. To isolate sequences that promote dimerization, transformants were grown on plates impregnated with λimm^{434}(clear). DNA isolated from cells which grew on these plates were reintroduced into XA90 and subject to a second round of selection. After the first round, 10% of individually tested colonies were immune to infection. After the second round, the percentage increased to 50%. By comparing the number of colonies on plates containing and lacking phage, we estimate that 1 in 10^5 to 10^6 zipper sequences is functional.

DNA was prepared from individual colonies after the second round of selection, and reintroduced into XA90. The activity of the fusion protein expressed by each clone was assayed by cross-streaking against different titers of phage. The sequence of each functional zipper was also determined (Fig. 5).

Although functional zippers are rare, they contain a surprisingly diverse collection of sequences. Some anomalous sequences were isolated. One class of unusual sequences contained prolines, which would be predicted to disfavor coiled-coil formation by disrupting the α-helix. Prolines were uncommon residues, and most of these prolines occur either at the end or at the beginning of the zipper sequence, perhaps leaving enough sequence to form a stable coiled-coil. One exception, 434-ZIP-10, contains a centrally located proline, and this sequence is the least functional. Another class contained nonsense codons. 434-ZIP-5 is easy to reconcile, since the stop codon occurs at the end of the putative zipper; however, 434-ZIP-9 and 434-ZIP-17 are highly functional yet have stop codons early in the putative zipper. Two explanations are plausible: efficient read-through of the amber codon, or a small energetic contribution from the zipper region being sufficient to confer phage immunity. The former is known to occur in a context-dependent manner. We think the later unlikely, because of the rarity of functional sequences, and the fact that 434-Δ is sensitive to phage infection even at much higher levels of protein expression (10mM IPTG, compared to 10μM IPTG used for these experiments).

The total charge on functional zippers ranged from −2 to +3, with an average charge of ±1.8. This result is consistent with the the suggestion that interhelical electrostatic repulsion is a major driving force in destabilizing homodimers of Fos, which has a net negative charge of −6.

The distribution of amino acids present in functional zippers was analyzed by helical position (Table 2). At most of the positions, three or four residues are more frequent, so that a loose consensus sequence can be defined. In most cases these do not coincide with the preferred residues at degenerate positions in the GCN4/Jun chimeric zippers, most likely because the degeneracy at each position was limited and biased by the starting sequence, and because these residues had to function in the context of a GCN4 zipper. Along the hydrophobic spine created by residues at positions **a** and **d**, position **d** was not randomized. Position **a** was significantly enriched in the branched-chain, neutral amino acids leucine, isoleucine, and valine. Another branched-chain amino acid, threonine was also common, but at the borderline of statistical significance. Branched-chain residues optimally pack with the constant leucines and with hydrophobic residues at positions **e** and **g** (20). Serine, which has a short, neutral side chain, is significantly discriminated against at position **a**.

Positions **e** and **g**, which in the GCN4 crystal structure pack against the hydrophobic spine as well as participate in intrahelical and interhelical salt bridges via terminal charged groups on basic and acidic residues, also contained a high frequency of neutral and hydrophobic residues, and the distribution of residues at each position was different. At position **e**, the most common residue was, surprisingly, glycine, and the preference for glycine at this position was highly significant statistically. The basis for this preference is unclear, but clearly shows that residues that do not efficiently pack against the hydrophobic spine are well tolerated at position **e**. Glutamate and glutamine were also significantly enriched at this position. At position **g**, the only significantly enriched residues were hydrophobic: alanine, leucine, and isoleucine. The basic residue arginine also occurred frequently, although they were not enriched to a significant degree. Differences between amino acid distributions at positions **e** and **g** may reflect differences in spatial positioning and neighboring residues at the two positions.

Based on the crystal structure of the GCN4 zipper (20), the outside positions **b**, **c**, and **f** do not make interhelical contacts and therefore cannot directly influence dimerization specificity. However, these residues can influence dimer stability by affecting the solvent-exposed surface and by forming intrahelical salt bridges. Residues which were enriched to a significant degree by selection for function are glutamate, aspartate, alanine, phenylalanine, and threonine. Residues significantly discriminated against include leucine, asparagine, and histidine. The reason that some nonpolar residues are preferred and some polar residues are disfavored at these solvent exposed positions is not clear.

Covariant analysis

Neighboring residues in the coiled-coil interact to stabilize the dimer. Residues at positions **a** and **d** fit like 'knobs into holes' (20), forming an efficiently packed hydrophobic spine. For example, a residue at position $\mathbf{a}_2$ packs into a hole surrounded by residues $\mathbf{d}_1'$, $\mathbf{g}_1'$, $\mathbf{a}_2'$, and $\mathbf{d}_2'$, while a residue at $\mathbf{d}_2$ is surrounded by residues $\mathbf{a}_2'$, $\mathbf{d}_2'$, $\mathbf{a}_3'$, and $\mathbf{e}_2'$. Interhelical salt bridges occur between terminal charged groups at positions $\mathbf{e}_{n+1'}$ and $\mathbf{g}_n$ in the structure determined for the GCN4 zipper and have been suggested to play a minor role in stabilizing the Fos-Jun heterodimer. Intrahelical salt bridges between residues $\mathbf{g}_n$ and $\mathbf{c}_{n+1}$ also are present in the GCN4 zipper crystal structure.

We examined the importance of these proposed interactions by looking for residues which frequently occur in pairs at positions which putatively interact. We found no significant co-occurrence of pairs of residues or classes of residues at pairs of positions described above, except for the association of basic residues at $\mathbf{e}_n$ and glutamate of Fos at $\mathbf{g}_{n+1}'$. Although glutamate is enriched at position **e** in zippers selected from the random zipper library for homodimer function, there was no significant pairing of these residues with positively charged residues at position **g**. Pairs involving position **d** were excluded because the invariant leucines confounded the analysis.

Thus we found no pairwise interaction which is so beneficial for dimer formation that it is enriched by selection for function, with the exception of positions $\mathbf{e}_n$ and $\mathbf{g}_{n+1}$ in heterodimers with Fos. This likely reflects the overwhelming importance of interactions in the hydrophobic spine, positions which could not be included in this analysis. Alternately, interactions involving three or more residues may stabilize the dimer, and these more complex interactions would be missed by this pairwise analysis.

DISCUSSION

Specificity is a common feature of interactions between macromolecules. In some cases, specificity results from a small number of critical intermolecular contacts. For instance, a small number of pairwise interactions largely accounts for the specific binding proteins to their DNA target sites; alteration of any of these contacts has a dramatic effect on binding affinity. In other cases, factors determining binding specificity cannot be so clearly delineated. Multiple residues may combine to form a single contact, and each contact may contribute only a small fraction of the total binding energy. In this situation, loss of any one contact may not dramatically alter binding affinity.

We used the method of random selection to examine the primary sequence requirements for dimer formation by leucine zippers. We found that an extremely diverse set of zippers are capable of forming homodimers or of heterodimerizing with Fos. Analysis of the zipper sequences revealed that certain residues are preferred at certain positions. These preferences were in general consistent with prior predictions. At position **a** of homodimerizing zippers selected from both the random zipper and random GCN4/Jun chimeric libraries, neutral, branched-chain amino acids were preferred. This class of residues optimally packs the hydrophobic dimer interface (20). At position **e**, zippers heterodimerizing with Fos were enriched for basic residues, which have previously been shown to make electrostatic interactions with acidic residues of Fos that contribute to dimer formation. Among homodimerizing zippers, positions **e** and **g** are not enriched for acidic or basic residues, and pairs of such residues do not occur at increased frequency, suggesting that electrostatic interactions are not crucial for homodimer stability. Instead, residues containing aliphatic side chains at these positions were preferred, consistent with the predicted role of these residues in contributing to the hydrophobic dimer interface. Positions **e** and **g** are spatially distinct, since different residues are preferred at each position. The high frequency of glycine at position **e** indicates that residues which fail to efficiently pack against the hydrophobic spine at this position are well tolerated.

Despite these sequence preferences, none of the preferred residues are essential for function. For example, although there is a bias towards basic residues at position **e** of zippers which heterodimerize with Fos, basic residues at this position are not essential to heterodimer formation since some heterodimerizing zippers contained the same complement of basic residues at this position as GCN4, which does not heterodimerize. Zippers selected for homodimerization from the random zipper library are significantly enriched for residues with branched chain aliphatic side chains at position **a**, yet some highly functional zippers lack these residues at this position. Although alanine is preferred over valine at position $\mathbf{a}_4$ in zippers which form heterodimers with Fos, zippers containing valine are still capable of forming Fos heterodimers. Previous studies have shown that although leucines are highly preferred at position **d**, mutation of individual leucines has a minimal effect on dimer formation (17, 19). The absence of essential residues or pairs of residues argue that numerous additive interactions stabilize leucine zipper dimers.

A previous study of dimer formation by leucine zippers suggested that the basis of preferential formation of Jun-Fos, but not GCN4-Fos, dimers could largely be understood in terms of pairwise electrostatic interactions between residues at positions **e** and **g** (21). However, our results indicate that the rules determining the ability of zippers to heterodimerize with Fos are more complex. Some zippers that form Fos heterodimers have the same number of basic residues at position **e** as GCN4. We have identified at least two other factors which are important for heterodimer formation with Fos: the presence of five heptad repeats, and alanine at position $\mathbf{a}_4$. The diversity of heterodimerizing sequences suggests that many other interactions not yet identified also participate.

In summary, a diverse set of sequences is capable of forming homodimers or heterodimers with Fos. Although certain residues are preferred at certain helical positions, no residue is essential for homodimer or heterodimer formation. Analysis of pairwise occurrences of residues also did not identify any pairwise interaction which is so beneficial for dimer formation that it is enriched by selection for function. We conclude that numerous, additive interactions contribute to the formation of a leucine zipper dimer.

ACKNOWLEDGEMENTS

This work was supported by a Johnson and Johnson fellowship to W.T.P. and by a grant to K.S. from the National Institutes of Health (GM 46555).

REFERENCES

1. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246,** 911–916.
2. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249,** 774–778.
3. Hope, I. A. & Struhl, K. (1987) *EMBO J.* **6,** 2781–2784.
4. Struhl, K. (1987) *Cell* **50,** 841–846.
5. Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. & Tjian, R. (1987) *Science* **238,** 1386–1392.
6. Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., Hunter, T. & Karin, M. (1988) *Nature* **332,** 166–171.
7. Nakabeppu, Y., Ryder, K. & Nathans, D. (1988) *Cell* **55,** 907–915.
8. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. (1988) *Cell* **55,** 917–924.
9. Sassone-Corsi, P., Ransone, L. J., Lamph, W. W. & Verma, I. M. (1988) *Nature* **336,** 692–695.
10. Kouzarides, T. & Ziff, E. (1989) *Nature (Lond).* **340,** 568–571.
11. Sellers, J. W. & Struhl, K. (1989) *Nature* **341,** 74–76.
12. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & Weintraub, H. (1990) *Cell* **61,** 49–59.
13. Ron, D. & Habener, J. F. (1992) *Genes & Dev.* **6,** 439–453.
14. Abate, C., Luk, D. & Curran, T. (1991) *Mol. Cell. Biol.* **11,** 3624–3632.
15. Diamond, M., Miller, J. N., Yoshinaga, S. K. & Yamamoto, K. R. (1990) *Science* **249,** 1266–1272.
16. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* **243,** 538–542.
17. Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990) *Science* **250,** 1400–1403.
18. Schuermann, M., Hunter, J. B., Henig, G. & Muller, R. (1991) *Nucl. Acids Res.* **19,** 739–746.
19. van Heeckeren, W. J., Sellers, J. W. & Struhl, K. (1992) *Nucl. Acids Res.* **20,** 3721–3724.
20. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991) *Science* **254,** 539–544.
21. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1992) *Cell* **68,** 699–708.
22. Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) *Cell* **71,** 1223–1237.
23. Oliphant, A. R. & Struhl, K. (1987) *Methods in Enzymology* **155,** 568–582.
24. Oliphant, A. R., Nussbaum, A. L. & Struhl, K. (1986) *Gene* **44,** 177–183.
25. Pu, W. T. & Struhl, K. (1991) *Mol. Cell. Biol.* **11,** 4918–4926.
26. Hope, I. A. & Struhl, K. (1986) *Cell* **46,** 885–894.
27. Oliviero, S. & Struhl, K. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88,** 224–228.
28. Wharton, R. P. & Ptashne, M. (1985) *Nature* **316,** 601–605.
29. Field, J., Nikawa, J.-I.. Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., Lerner, R. A. & Wigler, M. (1988) *Mol. Cell. Biol.* **8,** 2159–2165.