Nucleic Acids Research, 1993, Vol. 21, No. 9 2081–2086

The DNA binding specificity of the basic region of the yeast transcriptional activator GCN4 can be changed by substitution of a single amino acid

Manfred Suckow, Brigitte von Wilcken-Bergmann and Benno Müller-Hill*
Institut für Genetik der Universität zu Köln, Weyertal 121, D-5000 Köln 41, Germany

Received February 22, 1993; Revised and Accepted April 6, 1993

ABSTRACT

The X-ray structure of a GCN4 DNA complex (1) shows, that specific DNA binding of the GCN4 basic region is mediated by a complicated network of base pair and DNA backbone contacts. According to the X-ray structure, alanine −14 of the basic region of GCN4 (we define the first leucine of the leucine zipper as +1) makes a hydrophobic contact to the methyl group of the thymine next to the center of the GCN4 binding site 5′ ATGACTCAT 3′. We tested the DNA binding properties of the nineteen derivatives of GCN4, which carry all possible amino acids in position −14 of the basic region. Substitution of alanine −14 of GCN4 by either asparagine or cysteine changes the DNA binding specificity. Serine in this position broadens the specificity for position 1 of the target, whereas other amino acids either retain or decrease GCN4 specificity.

INTRODUCTION

The yeast transcription activator GCN4 (2, 3) is a member of the family of bZIP proteins. It activates many genes of yeast coding for enzymes of amino acid biosynthetic pathways during amino acid starvation (4). GCN4 dependent genes carry one or more GCN4 binding sites in their upstream regions (5). Their consensus is the asymmetric palindrome 5′ ATGACTCAT 3′, also known as AP1 site, which has been shown to be the optimal GCN4 target (6, 7). GCN4 binds *in vitro* nearly equally well to the fully symmetric sequence 5′ ATGACGTCAT 3′ (8, 9), also termed CRE (cAMP response element). GCN4 binds as a dimer to its DNA targets, such that the two protein monomers interact symmetrically with each half site.

Dimerisation and specific DNA binding are mediated by the bZIP motif, which consists of a basic region (10) followed by a leucine zipper (11). The basic region is directly involved in target recognition and binding (10), whereas the leucine zipper permits parallel aggregation of the two monomers (11, 12). Both form a functional unit, which is highly conserved throughout the whole bZIP family (10, 13). Short peptides comprising only the bZIP motif of GCN4 or of other proteins of the bZIP family are able to dimerize and to bind to DNA with the same specificity as the full length proteins (3, 9, 14, 15).

The X-ray structure of the GCN4-DNA complex has recently been solved at high resolution (1). The results of Ellenberger and coworkers show that the basic region and the leucine zipper form a dimer of two continous extended α-helices, which diverge with a small angle at the first leucine of the leucine zipper. This structure is placed on the center of the DNA target, such that the basic regions are positioned almost parallel to the plane of the bases. Since the DNA binding site is a palindrome, both basic regions form identical contacts in the major grooves of both half sites. Residues from positions −4 to −19 of the basic regions form a complicated network of specific contacts to bases and to the phosphates of the DNA backbone (we define the first leucine of the leucine zipper as +1). All contacts to bases are mediated by residues in positions −10 to −18.

We chose position −14 for specificity changes, because according to the X-ray structure (1) the contact between alanine −14 and the thymine next to the center of the target is not complicated by additional interactions with other bases or amino acids.

MATERIALS

Bacterial strains and plasmids

E.coli (K12ΔH1 SmRlacZamΔbio-uvrB ΔtrpEA2(λNam7-Nam53cI857ΔH1)) and pPLc28 (16) were kindly supplied by W. Fiers.

Chemicals and enzymes

Restriction enzymes were purchased from Boehringer-Mannheim (Germany), New England Biolabs (Bad Schwalbach, Germany) and Bethesda Research Laboratories (Eggenstein, Germany). DNA polymerase large fragment was obtained from Boehringer-Mannheim (Germany). All enzymes were used according to the suppliers' instructions. (^{32}P)deoxyribonucleotides were obtained from Amersham Buchler (Braunschweig, Germany), the chemicals for automated DNA synthesis from Applied Biosystems (Pfungstadt, Germany), all other chemicals including the SDS weight marker no. MW-SDS-17S from Sigma (München, Germany) or Merck (Darmstadt, Germany).

Oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer and purified on denaturating polyacrylamide gels.

* To whom correspondence should be addressed

Mutagenesis of the GCN4 basic region in position −14

The parental plasmid pPLc28-bZIP carries a synthetic gene coding for residues 222−277 of GCN4 flanked by two N-terminal (M, K) and four C-terminal codons (E, S, G, Q) of Lac repressor (9). In order to construct a library of bZIP genes with all amino acids in position −14 of the basic region, we replaced the region between the unique BssHII and XbaI sites of the synthetic bZIP gene by a synthetic DNA fragment, which was composed as follows:

coding strand: 5′ CGCGGAACACTGAAGCTNNNCGTCGCT 3′
noncoding strand: 3′ CTTGTGACTTCGAIIIGCAGCGAGATC 5′

The coding strand carried in the three positions of codon −14 an equal mixture of A, C, G, and T (N), the noncoding strand inosine (I). Thus all codons except that for residue −14 code for the wildtype GCN4 amino acids. We enriched positive clones by a BssHII restriction of the ligation reaction mixture prior to transformation and found fourteen out of nineteen possible amino acids in position −14 of the basic region after sequencing 68 individual BssHII resistant plasmids. Plasmids coding for GCN4 derivatives with the remaining five amino acids (aspartic acid, cysteine, histidine, phenylalanine, and tyrosine) in position −14 were obtained by cloning the respective synthetic DNA fragments.

Heat induction of the bZIP peptides

2ml cultures of freshly transformed colonies of *E.coli* K12ΔH1 were grown to saturation over night in YT with 0.5mg ampicilline/ml at 30°C and diluted 1:25 in dYT with 0.5mg ampicilline/ml to a final volume of 26ml. The cultures were transfered to a 42°C waterbath for 20′ after 2h growth at 30°C, at an OD_{600} of approximately 0.5. They were quickly cooled on ice after further 90′ of incubation at 37°C. The cells were harvested by centrifugation and resuspended in 2.5ml extraction buffer (30% glycerol; 200mMTris−HCl, pH8; 10mM $MgCl_2$; 1mM EDTA; 7mM β-mercaptoethanol) and sonicated (W220F, microtip, Heat Systems-Ultrasonics, Inc.) for 1′ at 4.5 scale units. Aliquots of the supernatant were stored frozen at −70°C after centrifugation (30′, 50000×g).

Protein gels

15μl of each crude extract were analysed for their contents of bZIP peptides on Tricine-SDS proteine gels (17).

Gel mobility shift assays

Electrophoretic mobility shift assays were performed according to (5) with minor modifications (9): 1μl undiluted crude extract was added to 5μl binding buffer (20mM Tris−HCl, pH 7; 50mM KCl; 3mM EDTA; 2mM DTT; 100mg gelatine/ml; 1mg BSA/ml) which contained 6.5fmole of the respective radiolabeled target DNA fragments and a 1000 fold excess of poly(dI:dC). The mixtures were incubated on ice for 5′ and loaded on a running 10% polyacrylamide gel after 3μl of 15% Ficoll, 0.1% bromophenol blue and 0.1% xylene cyanol in binding buffer had been added. The fragments used for mobility shift assays are the same as described (9) with added XbaI cohesive ends. They are 28 base pairs long and differ only in positions 0, 1, or 2 of the binding site:

5′CTAGACGGGCG A_4 T_3 G_2 A_1 C_0 $G_{0'}T_{1'}C_{2'}A_{3'}T_{4'}$ CGCCCGT 3′
3′ TGCCCGC $T_{4'}A_{3'}C_{2'}T_{1'}G_{0'}$ C_0 A_1 G_2 T_3 A_4 GCGGGCAGATC 5′

(the fully symmetric GCN4* binding site is numbered). Synthetic oligonucleotides carrying the GCN4* binding site or one of its variants were purified by denaturating polyacrylamide gelelectrophoresis, and labeled in a fill in reaction with DNA polymerase large fragment and α(^{32}P)dCTP. Double stranded fragments were purified on native polyacrylamide gels.

RESULTS

Analysis of the DNA binding properties of the GCN4 bZIP derivatives with amino acid exchanges in position −14 of the basic region

In order to study the DNA binding properties of the mutant GCN4 bZIP peptides, we prepared crude extracts from *E.coli* K12ΔH1 harbouring the respective expression plasmids. Aliquots of these extracts were analysed on Tricine-SDS protein gels (17) with respect to their contents of bZIP peptides (Fig. 1). All peptides were clearly visible and their concentrations did not differ by more than a factor of 3. We then tested their binding to the fully symmetric GCN4 binding site 5′ ATGACGTCAT 3′ (termed GCN4*) and all its variants with single symmetric base pair substitutions in positions 0, 1, or 2 (see Materials) by

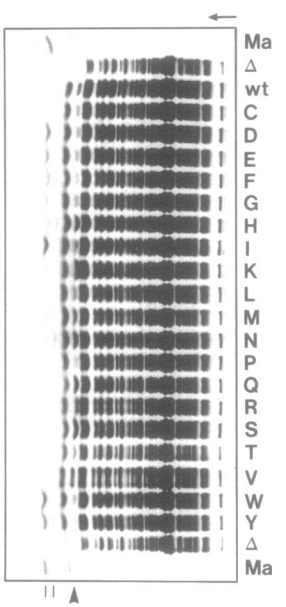

Figure 1. Tricine-SDS protein gel of the GCN4 bZIP peptide and its variants in position −14 of the basic region. 15μl of each crude extract containing 50mg protein/ml were analysed on a protein gel as described in (17). Note that the respective peptides do not run according to their calculated molecular weight of approximately 6.8kDa. Their position corresponds to the 8.16 kDa band of the molecular weight marker. Abbreviations on the right are as follows: Ma: molecular weight marker MW-SDS-17S, Sigma (the 8.16 kDa band is indicated by an arrow); all other lanes show crude extracts of strain K12ΔH1 after heat induction containing either a GCN4 peptide with a stop codon in position −14 of the basic region (Δ), the wildtyp GCN4 bZIP peptide (wt), or one of its variants with the indicated amino acid in position −14 of the basic region (single letter code). The arrow on the top indicates the direction of electrophoresis; the positions of the degradation products of the bZIP peptides are indicated by bars.

electrophoretic mobility shift assays. The autoradiographs of the retardation gels are shown in Figure 2. The relative amounts of retarded DNA fragments were estimated from at least three independent experiments and classified as follows: '++++' = more than 90% retarded; '+++' = 50−90% retarded; '++' = 10−50% retarded; '+' = up to 10% retarded; '−' = no retarded band visible. The results are summarized in figure 3.

The wildtype GCN4 bZIP peptide (with an alanine in position −14 of the basic region) binds best to the GCN4* binding site (++++), but it also recognizes with lower relative affinities the variants with guanine (+) or thymine (++) in position 0 and with thymine in position 2 (+++) of the target. Complexes with other targets do not become visible even after prolonged exposition of the X-ray film (up to six days). Eleven of the peptides, namely those with the amino acids aspartate, glutamate, phenylalanine, histidine, isoleucine, leucine, methionine, proline, arginine, tryptophan, and tyrosine in position −14, do not bind to any of the targets tested (Fig. 2, 3). The remaining peptides which bind specifically to DNA fall into four groups according to their binding properties:

1. Three peptides which bind weakly to the GCN4 target (G, K, T).* The GCN4 bZIP derivatives with glycine, lysine, or threonine in position −14 only recognize the GCN4* target but to a much lesser extent (+). In the cases of glycine and threonine complexes sometimes become visible only after prolonged exposition of the film (Fig. 2, 3).

2. Two peptides which bind in a similar manner like GCN4 to its targets (Q, V). The peptides with glutamine or valine in position −14 have similar binding properties as the GCN4 bZIP peptide. The V-14 peptide differs from GCN4 in so far that it does not recognize the target variant with a thymine in position 0. The Q-14 peptide differs from GCN4 in its affinities for the target variants with different base pairs in position 0: It binds better to a target with guanine (+++) in position 0 than to a target with thymine (+) in that position, whereas the GCN4 bZIP peptide prefers thymine over guanine in position 0 (Fig. 2, 3).

3. One peptide with broadened specificity (S). This group contains only the peptide with serine in position −14 of the basic region. This peptide binds to all targets recognized by the GCN4 bZIP peptide with similar relative affinities. But in addition it retards substantial amounts of the target variants with guanine (+++) or thymine (++) in position 1, which are not bound by the GCN4 bZIP peptide (Fig. 2, 3).

4. Two peptides with changed specificities (C, N). We found two peptides with changed DNA binding specificities: They carry cysteine or asparagine in position −14 of the basic region. The C-14 peptide binds only to the target variants with guanine (++) or thymine (++) in position 1 of the binding site, targets which are also recognized by the S-14 peptide. The N-14 peptide binds

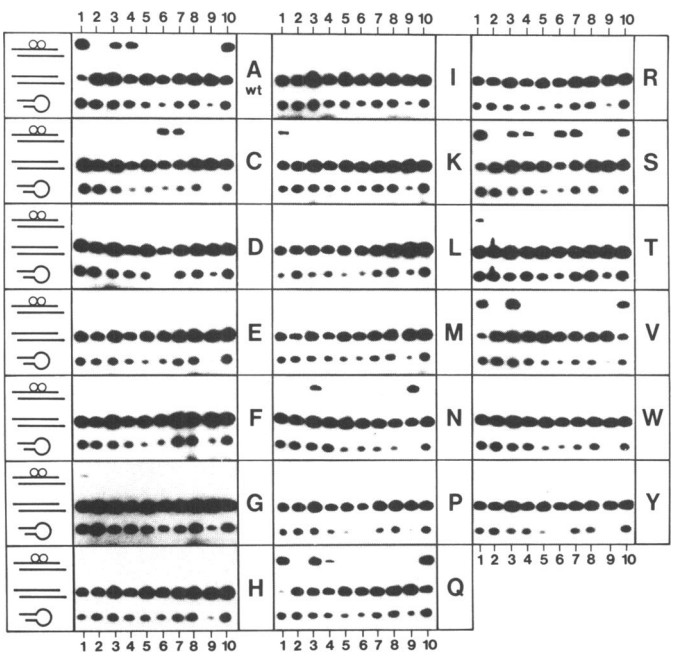

Figure 2. Typical autoradiographs of mobility shift assays with crude extracts containing the GCN4 bZIP peptide or its variants with the indicated amino acids in position −14 of the basic region. The numbers above and below the lanes refer to the targets used. They are the same as in fig. 3. The mutant bZIP peptides are indicated on the right sides of the respective autoradiographs. The diagrams on the left symbolize hairpin or double-stranded free DNA and the bZIP-DNA complexes.

Peptides \ Targets		GCN4*	Exchanges in position 0			position 1			position 2		
	4	5' A	5'	5'	5'	5'	5'	5'	5'	5'	5'
	3	T	.	.	.	.	.	.	.	.	.
	2	G	.	.	.	.	.	.	A	C	T
	1	A	.	.	.	C	G	T	.	.	.
	0	C	A	G	T	.	.	.	.	.	.
	0'	* G	T	C	A	.	.	.	.	.	.
	1'	T	.		.	G	C	A	.	.	.
	2'	C	.	.	.	.	.	.	T	G	A
	3'	A	.	.	.	.	.	.	.	.	.
	4'	T	.	.	.	.	.	.	.	.	.
-30 -25 -20 -15 -10 -5 -1		1	2	3	4	5	6	7	8	9	10
PESSDPAALKRARNTEAARRSRARKLQRMKQL		++++	−	+	++	−	−	−	−	−	+++
.....C.....		−	−	−	−	++	++	−	−	−	−
.....D.....		−	−	−	−	−	−	−	−	−	−
.....E.....		−	−	−	−	−	−	−	−	−	−
.....F.....		−	−	−	−	−	−	−	−	−	−
.....G.....		+	−	−	−	−	−	−	−	−	−
.....H.....		−	−	−	−	−	−	−	−	−	−
.....I.....		−	−	−	−	−	−	−	−	−	−
.....K.....		+	−	−	−	−	−	−	−	−	−
.....L.....		−	−	−	−	−	−	−	−	−	−
.....M.....		−	−	−	−	−	−	−	−	−	−
.....N.....		−	−	+	−	−	−	−	−	++	−
.....P.....		−	−	−	−	−	−	−	−	−	−
.....Q.....		++++	−	+++	+	−	−	−	−	−	+++
.....R.....		−	−	−	−	−	−	−	−	−	−
.....S.....		++++	−	+	+	−	+++	++	−	−	++
.....T.....		+	−	−	−	−	−	−	−	−	−
.....V.....		++++	−	+++	−	−	−	−	−	−	+++
.....W.....		−	−	−	−	−	−	−	−	−	−
.....Y.....		−	−	−	−	−	−	−	−	−	−
PESSDPAALKRARNTEAARRSRARKLQRMKQL		++++	−	+	++	−	−	−	−	−	+++
-30 -25 -20 -15 -10 -5 -1		1	2	3	4	5	6	7	8	9	10
position			lane								

Figure 3. Summary of the results shown in fig. 2. Left panel: the amino acid sequence of the GCN4 bZIP peptide is printed in single letter code from position −31 of the basic region to +1 (first leucine of the first leucine heptad repeat). For the mutant peptides only the amino acids in position −14 of the basic region are indicated. Upper right panel: the sequences of the GCN4* target and its variants. The complete GCN4* binding site is numbered is as in (1) from the center towards the borders. The target variants are indicated by the base pairs which differ from the GCN4* binding site. Lower right panel: Quantification of the individual peptide/target complexes. The relative affinities of the individual peptides for the respective targets were estimated from the relative intensities of the bands of the complexed and free double-stranded DNA in at least three independent experiments: '++++' = more than 90% retarded; '+++' = 50−90% retarded; '++' = 10−50% retarded; '+' = up to 10% retarded; '−' = no retardation. The numbers on the top and the bottom refer to the numbering of the lanes of the corresponding autoradiographs shown in fig. 2.

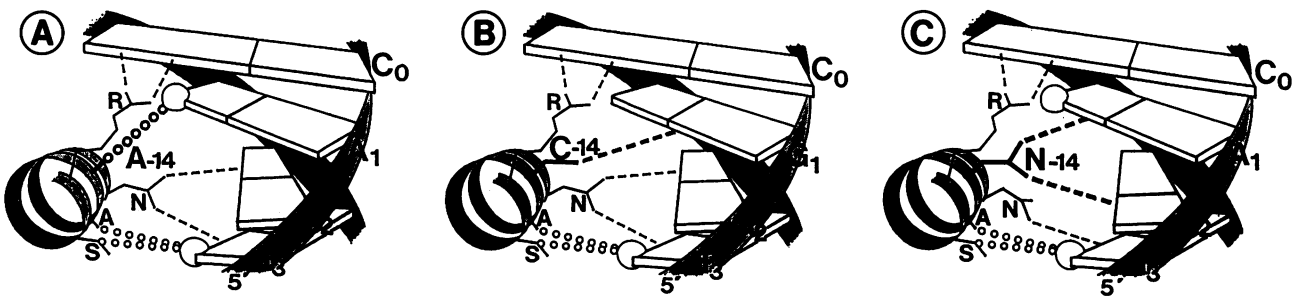

Figure 4. Schematic representation of specific amino acid-base pair interactions adapted from Ellenberger et al. (1). The amino acid in position −14 of each basic region are in bold print. Only those side chains are shown, which make specific contacts to base pairs. These are arginine −10, asparagine −18, alanine −15, and serine −11 (our numbering). Circles indicate hydrophobic interactions, dashed lines symbolize hydrogen bonds. (A) The GCN4-DNA complex according to the X-ray structure (1). (B) Model of a putative complex between the C-14 peptide and the target variant with guanine in position 1. (C) Model of a putative complex between the N-14 peptide and the target variant with cytosine in position 2. Note that asparagine −18 establishes only one hydrogen bond in our view of this complex.

with moderate affinity to the target with cytosine in position 2 (++). In addition it recognizes the target variant with guanine in position 0 (+) to an extent comparable to the GCN4 bZIP peptide (Fig. 2, 3).

DISCUSSION

The test system

Figure 1 shows that the steady state levels of the various peptides in the crude extracts differ by not more than a factor of 3. Many of the crude extracts contain various amounts of two degradation products in addition to the full length peptides (Fig. 1). It is likely that some of the truncated peptides contain an intact leucine zipper without a functional basic region. Formation of heterodimers of full length and truncated peptides would reduce the DNA binding activity of such an extract. In particular the peptides with aspartate, glutamate, isoleucine, tryptophan, or tyrosine in position −14 of the basic region are extensively degraded (Fig. 1) and they have no detectable DNA binding activity either (Fig. 2, 3). In these five cases we can not exclude the possibility that sufficient amounts of intact, full length homodimers would display a low affinity for one or the other target DNA.

All other crude extracts contain significantly lower amounts of degraded peptides (Fig. 1). For example the crude extracts with the V-14 or the R-14 peptide contain similar amounts of full length peptides, but only in the former degradation products are visible (Fig. 1). The V-14 peptide binds very well to some of the DNA targets, whereas the R-14 peptide does not bind to any of the binding sites tested (Fig. 2, 3). From this we conclude that moderate amounts of degradation products are not critical for DNA binding.

We used five times more crude extract than in our earlier experiments (9) to enhance weak complexes. Thus complexes formed with the GCN4 bZIP peptide and the variants of the GCN4* binding site with a guanine or a thymine in position 0 are now better visible than in our previous work (9).

For practical reasons we analysed in our experiments complex formation between the various peptides and the fully symmetric GCN4* binding site (5′ ATGACGTCAT 3′) and its variants with symmetric substitutions of single base pairs. In contrast the recently published X-ray structure of a GCN4-DNA complex (1) was derived from the asymmetric GCN4 binding site (5′ ATGACTCAT 3′). Ellenberger and coworkers found that the contacts

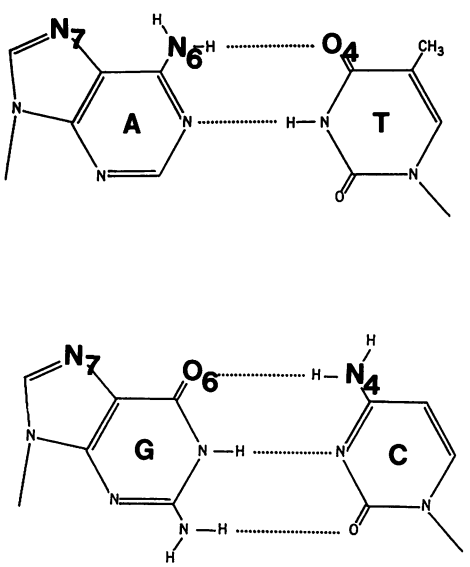

Figure 5. Schematic representation of base pairs. Atoms available for hydrogen bonding in the major groove are in bold print. Hydrogen bond acceptors are N7 and O6 of guanine, N7 of adenine, and O4 of thymine. Hydrogen bond donors are the N4 amino group of cytosine and the N6 amino group of adenine.

of amino acids to the phosphate backbone and to the base pairs were fully symmetric except for the hydrogen bonds between arginine −10 (our numbering) and the central guanine 0. Since there is only one guanine at the center, only one of the monomers can establish this contact. The other arginine contacts the DNA backbone (1). Since the GCN4* and the GCN4 targets are recognized almost equally well by GCN4 *in vitro* (8, 9), the complex of the GCN4 dimer and the fully symmetric GCN4* target is most probably fully symmetric such that now both arginines −10 are able to form hydrogen bonds to the central guanines.

The binding properties of the wildtype GCN4 bZIP peptide

As shown in figures 2 and 3 the wildtype GCN4 bZIP peptide recognizes four of the ten targets tested. This reflects the role of the GCN4 protein in yeast: It regulates the expression of 30−40 enzymes involved in amino acid biosyntheses (4). The *in vivo* binding sites are suboptimal variations of the optimized

targets, most of them contain only one perfect half site (3, 4). GCN4 binds with various affinities to its different control elements, such that it is able to modulate the expression rates of its genes according to the actual metabolic needs of yeast.

With the knowledge of the X-ray data of the GCN4-DNA complex (1) it is tempting to explain on the molecular level, why the wildtype basic region of GCN4 recognizes some of the variants of the GCN4* binding site. As shown in figure 4a, asparagine −18 forms hydrogen bonds with 1.) the O4 keto group of thymine 3 as donor and 2.) the N4 amino group of cytosine 2′ as acceptor (Fig. 5). Our experiments show that the GCN4 bZIP peptide binds well to the variant of the GCN4* binding site with a thymine in position 2 (+++, see figures 2 and 3). We suggest that in this case the N6 amino group of adenine 2′ can substitute for the N4 amino group of cytosine 2′ and accept the proton (Fig. 5). The fact that both amino groups are not identically positioned within the major groove might explain the weaker complex between the GCN4 bZIP peptide and the target variant with thymine in position 2 (+++) compared with the GCN4* complex (++++, see figures 2 and 3).

The moderate binding of the GCN4 bZIP peptide to the variant of the GCN4* binding site with a thymine in position 0 (++, figures 2 and 3) may also be explained: According to the X-ray structure of the GCN4-DNA complex (1) arginine −10 of the GCN4 basic region forms two hydrogen bonds with the guanine 0′ in the center of the target. In both cases the guanidinium group is the proton donor, whereas N7 and O6 of guanine serve as acceptors (see figure 4a and 5). If the cytosine-guanine pair of the GCNN4* target is exchanged for a thymine-adenine pair, N7 of adenine 0′ is in the same position as N7 of guanine 0′ and may accept one of the protons of the guanidinium group of arginine −10. The second hydrogen bond can not be formed because the N6 amino group of adenine 0′ is a proton donor. The missing hydrogen bond might explain the weaker binding of the GCN4 bZIP peptide to this binding site. The weak binding of the GCN4 peptide to the target variant with a guanine in position 0 (+, figures 2 and 3) may then be explained by one hydrogen bond to the O6 keto group of guanine 0 (now on the opposite strand, but still within reach), while the second hydrogen bond again is lost.

Peptides with abolished or strongly reduced activity

11 of the 20 possible amino acids in position −14 of the basic region of the GCN4 bZIP peptide destroy its ability to bind to any of the tested binding sites (Fig. 2 and 3). To this group belong the acidic amino acids (aspartate and glutamate), the neutral hydrophobic amino acids with extended or helix destabilizing side chains (leucine, isoleucine, tryptophan, phenylalanine, methionine, and proline), the two basic amino acids with the most bulky side chains (arginine and histidine), but only one amino acid with a neutral polar side chain (tyrosine). Three amino acids in position −14 of the GCN4 basic region allow weak interactions with the GCN4* binding site: Glycine, threonine, and lysine (Fig. 2 and 3).Our data do not indicate, how the function of an individual bZIP peptide is abolished or reduced in each particular case.

Peptides with DNA binding properties similar to the GCN4 bZIP peptide

The bZIP peptides with a valine or a glutamine in position −14 of the basic region bind in a similar manner as the GCN4 peptide to the GCN4* binding site and its variants. The V-14 peptide recognizes the target variant with a guanine in position 0 significantly better (+++) than the GCN4 peptide (+), but fails to bind to the variant with a thymine in this position. All other complexes are formed with similar relative affinities as observed for the GCN4 peptide. We suggest that the short but bulky side chain of valine may establish a similar hydrophobic contact to the methyl group of thymine 1′ as does alanine −14 in the wildtype situation. The second methyl group might interact with the side chain of arginine −10, and thereby stabilize its contacts to guanine 0′ but impede a hydrogen bond to an adenine in this position.

It is difficult to explain, how the longer, polar side chain of glutamine −14 can functionally replace the methyl group of alanine. The Q-14 peptide binds to all targets, even to those with exchanges in position 0, with similar relative affinities as the GCN4 peptide does. The only difference is the preference of guanine over thymine in position 0 of the binding site (Fig. 2 and 3). We speculate that the β-methylen group establishes a hydrophobic contact to the methyl group of thymine 1′. It is difficult to explain why the side chain of glutamine can fit into the network of the other interactions without disturbing them while lysine in position −14 cannot. We could imagine that in the case of the K-14 bZIP peptide a destabilisation of the helix might be caused by the juxtaposition of K-14 and R-10.

Peptides with broadened or changed specificities

The S-14 peptide. The S-14 peptide forms complexes with all targets which are bound by the GCN4 bZIP peptide with very similar relative affinities. Possibly the β-methylene group of serine can functionally replace the methyl group of alanine without disturbing the DNA binding surface. But the S-14 peptide recognizes in addition the two variants of the GCN4* binding site with a guanine (+++) or a thymine (++) in position 1 which are not bound by the GCN4 bZIP peptide (Fig. 2 and 3). We suggest that the additional complexes are stabilized by a hydrogen bond between the hydroxyl group of serine −14 and the O6 keto group of guanine 1 or the O4 keto group of thymine 1, respectively.

The C-14 peptide. Replacement of alanine −14 by cysteine leads to a change in DNA binding specificity of the basic region. The C-14 peptide does not bind to any of the binding sites which are recognized by the GCN4 peptide, but it forms moderate complexes (++) with the target variants with a guanine or a thymine in position 1 (Fig. 2 and 3). Both targets are also recognized by the S-14 peptide, suggesting that the sulfur hydrogen forms similar hydrogen bonds to the O6 keto group of guanine 1 or the O4 keto group of thymine 1, respectively. The absence of all GCN4 like complexes is presumably caused by the difference between oxygen and sulfur: the larger sulfur atom might sterically hinder a hydrophobic contact between the β-methylen group of cysteine and the methyl group of thymine 1′. Our view of the possible interactions between the basic region of the C-14 peptide and the variant of the GCN4* target with guanine in position 1 is depicted in figure 4b. The results obtained with the peptides carrying serine or cystein in position −14 of the basic region are in agreement with the X-ray structure of the GCN4-DNA complex (1) with respect to the hydrophobic interaction between the methyl group of alanine −14 and the methyl group of thymine 1′.

The N-14 peptide. A different change of DNA binding specificity is observed after the replacement of alanine −14 by asparagine. The N-14 peptide does not recognize the targets which are bound

by the GCN4 peptide except the variant with guanine in position 0 (+), but binds moderately to the variant with cytosine in position 2 (++, fig. 2 and 3). We explain the complex formation with this target variant by proposing that the side chain of asparagine −14 establishes hydrogen bonds to both the O4 keto group of thymine 1′ and the N4 amino group of cytosine 2 in a manner very similar to the contacts made by asparagine −18 and base pairs 2 and 3 in the wildtype GCN4 DNA complex (compare fig. 4a and 4c). In contrast to the wildtype situation (Fig. 4a) asparagine −18 can not establish its hydrogen bond to base pair 2, because both N7 and O6 of guanine 2′ are proton acceptors (Fig. 4c and 5). Thus the moderate relative affinity of the complex between the N-14 peptide and this binding site would then be due to this missing hydrogen bond (Fig. 4c).

Similarities with other systems

Most of the known bZIP proteins resemble GCN4 in so far that they carry an alanine in position −14 of their respective basic regions. Only a few of the amino acids which we have found to function in position −14 of the basic region of GCN4 also occur in other natural bZIP proteins: serine −14 (for example Cys3, 15), valine −14 (for example C/EBP, 18), and glutamine −14 (for example yAP1, 19). In all three cases the flanking residues from position −10 to −17 of the respective basic regions differ substantially from the GCN4 sequence. The precise Cys3 target is not yet known, C/EBP has been shown to bind specifically to the fully symmetric sequence 5′ ATTGCGCAAT 3′ (9, 18). YAP1 seems to bind best to the asymmetric GCN4 binding site 5′ ATGACTCAT 3′. This agrees with our observations concerning the Q-14 peptide which binds also to the asymmetric GCN4 target with high relative affinity (data not shown). The other residues which we detected to be functional in position −14 of the basic region of GCN4, *i.e.* cysteine or asparagine, have not been reported to occur in natural bZIP proteines in this position.

ACKNOWLEDGEMENTS

We thank W.Fiers for providing plasmid pPLc28 and *E.coli* K12ΔH1, K.Otto for expert technical assistance, U.Ringeisen for excellent photography, R.Ehring, B.Kemper and M.Cramer for critically reading the manuscript. We also thank twenty undergraduate students who performed part of the screening of the library as their course work. This work was supported by DFG through SFB 274.

REFERENCES

1. Ellenberger, T.E., Brandle, C.J., Struhl, K. & Harrison, S.C. (1992) *Cell* **71**, 1223−1239.
2. Hinnebusch, A.G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6442−6446.
3. Hope, I.A. & Struhl, K. (1986) *Cell* **46**, 885−894.
4. Hinnebusch, A.G. (1990) *TIBS* **15**, 148−152.
5. Hope, I.A. & Struhl, K. (1985) *Cell* **43**, 177−188.
6. Hill, D.E., Hope, I.A., Macke, J.P. & Struhl, K. (1986) *Science* **234**, 451−457.
7. Oliphant, A.R., Brandle, C.J. & Struhl, K. (1989) *Mol. Cell. Biol.* **9**, 2944−2949.
8. Sellers, J.W., Vincent, A.C. & Struhl, K. (1990) *Mol. Cell. Biol.* **10**, 5077−5086.
9. Suckow, M., Wilcken-Bergmann, B.v. & Müller-Hill, B. (1993), *EMBO J.*, in press.
10. Busch, S.J. & Sassone-Corsi, P. (1990) *TIG* **6**, 36−40.
11. Landschulz, W.H., Johnson, P.F. & McKnight, S.L. (1988) *Science* **240**, 1759−1764.
12. Oas, T.G., McIntosh, L.P., O'Shea, E.K., Dahlquist, F.W. & Kim, P.S. (1990) *Biochemistry* **29**, 2891−2894.
13. Vinson, C.R., Sigler, P.B. & McKnight, S.L. (1989) *Science* **246**, 911−916.
14. Kouzarides, T. & Ziff, E. (1989) *Nature* **340**, 568−571.
15. Kanaan, M.N. & Marzluf, G.A. (1991) *Mol. Cell. Biol.* **11**, 4356−4362.
16. Remault, E., Stanssens, P. & Fiers, W. (1981) *Gene* **15**, 81−93.
17. Schägger, H. & Jagow, G.v. (1987) *Anal. Biochem.* **166**, 368−379.
18. Agre, P., Johnson, P.F. & McKnight, S.L. (1989) *Science* **246**, 922−926.
19. Moye-Rowley, W.S., Harshman, K.D. & Parker, C.S. (1989) *Genes & Development* **3**, 283−292.