2198–2208 Nucleic Acids Research, 1994, Vol. 22, No. 12

Creating new DNA binding specificities in the yeast transcriptional activator GCN4 by combining selected amino acid substitutions

Manfred Suckow, Anup Madan+, Brigitte Kisters-Woike, Brigitte von Wilcken-Bergmann and Benno Müller-Hill*
Institut für Genetik der Universität zu Köln, Weyertal 121, 50931 Köln, Germany

Received May 4, 1994; Accepted May 18, 1994

ABSTRACT

The specificity of the GCN4/DNA complex is mediated by a complicated network of interactions between the basic regions of both GCN4 monomers and their target halfsites. According to X-ray analyses (1, 2) one particular thymine of the target sequence is recognized by serine −11 and alanine −15 (we define the leucine in the first d-position of the heptad repeats as +1). We replaced serine −11 or alanine −15 with all other amino acids and analysed the DNA binding properties of the resulting stable GCN4 derivatives by electrophoretic mobility shift assays. Among these, mutants with tryptophan in position −11, or glutamic acid and glutamine in position −15, differ significantly from GCN4 in their DNA binding specificities. We then constructed selected double mutants, which differ from GCN4 in positions −11, −15 or −14 (3) of the basic region. The double mutants with tryptophan in position −11 and asparagine or serine in position −14 show drastically altered DNA binding specificities, presumably due to additive effects.

INTRODUCTION

The transcription activator GCN4 of yeast is one of the best understood eukaryotic regulatory proteins (4). The 281 residue protein, whose expression is controlled at the translational level, plays a key role in the regulation of amino acid anabolism (5). GCN4-dependent genes generally carry several GCN4 binding sites in their upstream sequences (6, 7). GCN4 binds to its target sites as a dimer and stimulates transcription via a short acidic activation domain located between residues 107 and 125 (8, 9). The natural binding sites are more or less degenerate variants of the optimal GCN4 target 5'ATGACTACT3' (10, 11), also known as the AP1 site (12). *In vitro*, GCN4 has been shown to bind with similar high affinity to the fully palindromic sequence

5' A4 T3 G2 A1 C0 * G0'T1'C2'A3'T4' 3'
3' T4'A3'C2'T1'G0' * C0 A1 G2 T3 A4 5'

the so-called ATF/CREB site (13, 14, 15; this numbering of the bases will be used subsequently). The C-terminal 60 residues of GCN4 were found to be responsible for specific DNA binding as well as for dimerisation (8, 9). They form a structural element, the bZip motif, which also occurs in several other eukaryotic transcription factors (16).

The bZip element is one of the most recently detected DNA binding motifs (16). It has been found in more than 50 eukaryotic transcription factors which make up the bZip protein family. The bZip motif is characterized by an area of approximately 30 residues with a predominant positive net charge, the so-called basic region (8), followed by a leucine zipper, which usually consists of 3–5 heptad repeats (17). The location of the bZip motifs within the primary sequences of the different members of the bZip protein family varies considerably (16). In order to facilitate their comparison, we use a common numbering system where position +1 corresponds to the first d-position of the leucine zipper, which in most cases is occupied by leucine, but threonine or isoleucine may also occur. The residues of the basic regions are counted towards the N-terminus and hence are assigned negative numbers, whereas the residues of the leucine zippers are counted towards the C-terminus (18). BZip leucine zippers have been shown to assemble in parallel orientation into a coiled coil structure (19, 20), thereby adjusting the basic regions of both monomers such that they can enter the major groove right and left of the center of symmetry of their target DNA (21). The basic region contains all residues necessary for specific DNA binding. Chimeric bZip peptides with basic regions and leucine zippers from different proteins show the DNA binding specificities of the donors of the basic regions (18, 22). Furthermore the bZip domains of GCN4 and other bZip proteins function independently of the rest of the protein. Short bZip peptides show the same DNA binding specificities as the full length proteins (8, 9, 18, 23, 24). In contrast to the leucine zipper, the basic region has been shown to adopt a stable α-helical conformation only upon interaction with the DNA target (25, 26).

Sequence comparisons of bZip motifs (for example: 16, 27) reveal characteristic features of the bZip protein family: two

*To whom correspondence should be addressed

+Present address: Chemical Physics Group of the Tata Institute of Fundamental Research, Homi Bhabha, Colaba, Bombay-4000 005, India

residues are absolutely conserved, namely asparagine −18 and arginine −10. In addition there are less strictly conserved residues: the basic regions start with an accumulation of 3 to 6 basic amino acids between residues −26 to −19. In positions −15 and −14 alanine predominates, whereas positions −13 and −12 are usually occupied by basic amino acids. In position −11 serine or cysteine are often found. A second cluster of 2 to 4 basic side chains occurs between positions −9 to −4. BZip proteins can be divided into several subgroups, whose members are characterized by a particularly high degree of homology between positions −4 to −26. In several cases residues −18 to −10 are identical. The similarities between the basic regions of members of the same subgroup are reflected in their DNA binding specificities: they often recognize identical target sites (28).

The crystal structures of the complexes between GCN4 peptides and the AP1 site (1) or the ATF/CREB site (2) have been solved. As predicted in the 'induced helical fork' model (25), both monomers of the bZip dimer form uninterrupted α-helices. In order to accommodate the target DNA, the basic regions deviate slightly from the coiled coil, which is oriented perpendicular to the target DNA. Recognition of the AP1 target carrying a central base pair is asymmetric. Arginine −10 of one monomer contacts the central base pair, while arginine −10 of the other monomer binds to the phosphate backbone (1). In the complex with the fully symmetric ATF/CREB site, which in contrast to the AP1 site is bent by about 20° upon contact with GCN4, both arginines establish hydrogen bonds with the central guanines and contact the DNA backbone. Thus, they adopt a conformation which is intermediate to those observed with the two GCN4 monomers in the AP1 site complex. Residues from positions −4 to −21 of each basic region are involved in DNA binding, with five of them contacting base pairs: asparagine −18, alanines −15 and −14, serine −11, and arginine −10. The other residues which contact DNA are engaged in backbone contacts (1, 2). Except for the arginines −10, the structural details of the protein−DNA interactions are nearly identical in both complexes.

We previously replaced alanine −14 with all other amino acids and tested the binding of the resulting mutants to variants of the fully symmetric ATF/CREB site (3). We found that serine in position −14 broadens the DNA binding specificity of GCN4, whereas cysteine and asparagine change the specificity. These results agree well with both X-ray studies (1, 2).

We now report a similar analysis focusing on positions −15 and −11. We have replaced alanine −15 and serine −11 with all other amino acids and tested the DNA binding properties of the resulting mutants. Furthermore, we have constructed double mutants which combine exchanges in positions −14 (3), −15, or −11, and screened them for additive effects.

MATERIALS AND METHODS

Bacterial strains and plasmids

E.coli (K12ΔH1: SmR lacZam Δbio-uvrB ΔtrpEA2 λNam7-Nam53 cI857 ΔH1) and pPLc28 (32) 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 supplier's instructions. (^{32}P)deoxyribonucleotides were obtained from Amersham Buchler (Braunschweig, Germany), the chemicals for automated DNA synthesis from Applied Biosystems (Pfungstadt, Germany), and all other chemicals from Sigma (München, Germany) or Merck (Darmstadt, Germany). Oligonucleotides were synthesized on an Applied Biosystems 380A synthesizer and purified on denaturating polyacrylamide gels.

Mutagenesis of the GCN4 basic region

The parental plasmid pPLc28-bZip carries a synthetic gene coding for residues 222−277 of GCN4 flanked by codons for the two N-terminal (M, K) and the four C-terminal residues (E, S, G, Q) of Lac repressor (18). In order to construct two libraries of GCN4-bZip genes with codons for all possible amino acids in positions −15 or −11 of the basic region, we replaced the regions between the unique BssHII site and the unique XbaI or HindIII sites of the synthetic bZip gene by synthetic DNA fragments, which were composed as follows:

Pos. −15:
coding strand: 5′ CGCGGAACACTGAANNNGCACGTCGCT 3′
non-coding strand: 3′ CTTGTGACTTIIICGTGCAGCGAGATC 5′

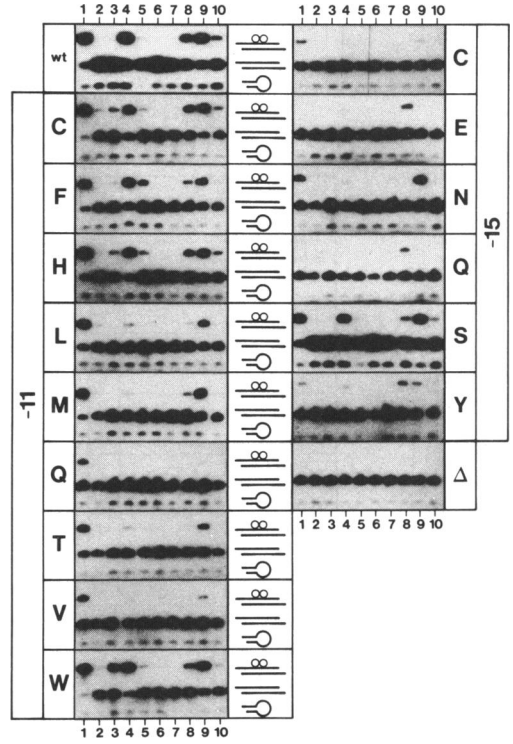


Figure 1. Mobility shift assays with cell extracts containing the wildtype GCN4 peptide and all single mutants in positions −11 or −15 of the basic region, which bind specifically to DNA. The respective amino acid replacements are indicated on the left or right side of the autoradiographs, respectively. A truncated GCN4 peptide terminating at position −14 of the basic region (Δ) was used as negative control. The numbers above and below the lanes refer to the targets used. They are the same as in figure 2 (between the upper and the lower right panels). The diagrams between the autoradiographs indicate hairpin or double-stranded free DNA and the bZip/DNA complexes.

Pos. −11:
coding strand: 5′ CGCGTAACACTGAAGCTGCACGTCGCNNNCGCGCTCGTA 3′
non-coding
strand: 3′ ATTGTGACTTCGACGTGCAGCGIIIGCGCGAGCATTCGA 5′

Both coding strands carried an equal mixture of A, C, G, and T (N) in the three positions of codon −15 or −11, respectively, while the non-coding strands contained inosine (I) residues in the complementary positions. Thus all codons except those in the variable positions −11 or −15 code for the wildtype GCN4 amino acids. We enriched positive clones by a BssHII restriction of the ligation mixtures prior to transformation. We found bZip genes coding for fourteen of the nineteen possible amino acids in position −15 or −11 of the basic region after sequencing seventy individual BssHII resistant plasmids for each position. Synthetic genes coding for GCN4 derivatives with the missing codons in positions −15 or −11 as well as those coding for the double mutants were obtained by cloning defined synthetic DNA fragments. The preparation of cell extracts has been described in detail previously (3). The presence of the respective bZip peptides in the extracts was determined by analysis on Tricine−SDS protein gels (29), and extracts were used for gel mobility shift assays without further purification. For a particular binding mixture 1μl of the respective crude extract was added to 7μl binding buffer (20mM Tris−HCl, pH7; 50mM KCl; 3mM $MgCl_2$; 1mM EDTA; 2mM DTT; 100mg gelatine/ml; 1mg BSA/ml) containing 6.5fmole of the respective radiolabeled target DNA fragment and a 1000 fold excess of poly (dI:dC). (We deeply regret that the binding buffer has been described incorrectly in two previous publications (3, 18). Please note that it is not 3mM EDTA as stated there.) After 10′ incubation on

		Symmetric single base pair exchanges in								
ATF/CREB		position 2			position 3			position 4		
5′	5′	5′	5′	5′	5′	5′	5′	5′	5′
A4	.	.	.	.	.	.	C	G	T
T3	.	.	.	A	C	G	.	.	.
G2	A	C	T	.	.	.	.	.	.
A1	.	.	.	.	.	.	.	.	.
C0	.	.	.	.	.	.	.	.	.
G0′	.	.	.	.	.	.	.	.	.
T1′	.	.	.	.	.	.	.	.	.
C2′	T	G	A	.	.	.	.	.	.
A3′	.	.	.	T	G	C		.	.
T4′	.	.	.	.	.	.	G	C	A
3′	3′	3′	3′	3′	3′	3′	3′	3′	3′

-18			-15	-14			-11	-10
N	T	E	A	A	R	R	S	R
.	.	.	.	.	.	.	C	.
.	.	.	.	.	.	.	F	.
.	.	.	.	.	.	.	H	.
.	.	.	.	.	.	.	L	.
.	.	.	.	.	.	.	M	.
.	.	.	.	.	.	.	Q	.
.	.	.	.	.	.	.	T	.
.	.	.	.	.	.	.	V	.
.	.	.	.	.	.	.	W	.
.	.	.	C	.	.	.	.	.
.	.	.	E	.	.	.	.	.
.	.	.	N	.	.	.	.	.
.	.	.	Q	.	.	.	.	.
.	.	.	S	.	.	.	.	.
.	.	.	Y	.	.	.	.	.

1	2	3	4	5	6	7	8	9	10
++++	-	-	+++	-	-	-	++	+++	+
++++	-	+	+++	+	-	-	++	+++	+
++++	-	-	+++	++	-	-	+	++	-
++++	-	+	+++	+	-	-	++	+++	+
+++	-	-	-	-	-	-	-	++	-
++++	-	-	-	-	-	-	+	+++	-
++	-	-	-	-	-	-	-	-	-
+++	-	-	-	-	-	-	-	++	-
++	-	-	-	-	-	-	-	+	-
++++	-	++	++++	+	-	-	++	+++	+
+	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	+	-	-
++	-	-	-	-	-	-	-	++	-
-	-	-	-	-	-	-	+	-	-
+++	-	-	++	-	-	-	+	++	+
+	-	-	-	-	-	-	+	+	-

Figure 2. Summary of the results shown in figure 1. Left panel: the amino acid sequence of the GCN4 bZip peptide is printed in single letter code from position −18 of the basic region to −10 (counted backwards from the first leucine of the leucine zipper, they correspond to residues 235−243 of the complete GCN4 protein). Only the altered amino acids are indicated in the mutant peptides. Upper right panel: the sequences of the ATF/CREB site and its symmetric variants with substitutions in positions 2, 3 and 4. The ATF/CREB site is numbered according to Ellenberger *et al.* (1), from the center towards the borders. For target variants, only those bases are printed which differ from the ATF/CREB site. The numbering of the binding sites between the upper and the lower right panels corresponds to the numbering of the lanes of the autoradiographs shown in figure 1. Lower right panel: quantification of the individual peptide-target complexes. The relative affinities of the individual peptides for the respective targets were estimated by comparing the relative intensities of the bands of complexed and free double-stranded DNA in at least three independent experiments: '++++' = more than 90% shifted; '+++' = 50−90% shifted; '++' = 10−50% shifted; '+' = up to 10% shifted; '−' = visible only upon unusually prolonged exposition of the film or no retardation detectable.

ice 3μl of 15% Ficoll, 0.1% bromophenol blue and 0.1% xylene cyanol in binding buffer were added before separating the mixtures on pre-run 10% polyacrylamide gels. The 28 base pair DNA fragments containing the various target sequences are the same as described (18).

Computer modelling

Modelling was done on an ESV30 workstation with Insight II and Discover (Biosym Inc., San Diego, CA.) starting with the coordinates of the GCN4/AP1 site cocrystal (1). In order to examine the possible interactions of the mutant peptides, we replaced the respective amino acids and base pairs in both half-complexes. Steric clashes were removed manually and the resulting mutant complexes were relaxed. The amber force field was used for the minimization of the mutant complexes.

RESULTS

The test system

The various peptides were expressed in *E.coli* K12 Δ H1 under the control of λ PL (3). Aliquots of crude cell extracts were analysed for their contents of the respective bZip peptides on Tricine–SDS gels (29) as shown previously [(3), figure 1; (18), figure 3]. Again, all peptides were clearly visible and their amounts differed by less than a factor of five (data not shown). The DNA binding properties of the bZip peptides were monitored by gel shift experiments as described (3).

Single mutants in positions −15 and −11 of the basic region of GCN4

Alanine −15 and serine −11 of the wildtype GCN4 peptide are both involved in hydrophobic contacts with the methyl group of thymine 3 (1, 2). We therefore determined the DNA binding specificities of the mutant bZip peptides with a set of ten targets including the ATF/CREB site and nine derivatives with all possible single symmetric base pair substitutions in positions 2, 3, and 4. Typical gel shift experiments are shown in figure 1, the results are summarized in figure 2.

The double mutants

Mutations in positions −15 or −11, which alter the DNA binding specificity of the GCN4 bZip peptide and promise to exhibit synergistic effects upon combination, are tryptophan −11, asparagine −15 and glutamine −15. These as well as cysteine, asparagine and serine in position −14 (3) were used to construct double mutants. They were tested with a set of binding sites consisting of: 1.) the ATF/CREB site; 2.) those symmetric single exchange targets which were bound by the respective single mutants; and 3.) the respective double exchange targets. In order to demonstrate possible additive effects in their DNA binding specificities not only the double mutant bZip peptides were tested in combination with the new collection of binding sites, but also their parental single mutants. The results are shown in figures 3 and 4.

DISCUSSION

The DNA binding properties of the wildtype GCN4 bZip peptide

Electrophoretic mobility shift assays are convenient for the analysis of specific protein–DNA interactions. Figures 1 and 2 show that the wildtype GCN4 peptide binds best to the symmetric ATF/CREB site, but also with reduced relative affinity to target variants with substitutions in position 2 or 4. Possible molecular interactions between side chains of the wildtype GCN4 peptide and the variant of the ATF/CREB site with thymine in position 2 have been discussed (3). Figure 6A shows a schematic drawing of a half-complex between GCN4 and the ATF/CREB site.

In position 3 of the target a thymine / adenine pair is required. The X-ray structures (1, 2) show that the methyl group of thymine 3 contacts the methyl group of alanine −15 and the methylene carbon of serine −11. In addition, the H-bond between O4 of thymine 3 and Nδ of asparagine −18 is disrupted when thymine 3 is replaced by any other base. Thus the simultaneous loss of three contacts prevents complex formation between the wildtype GCN4 peptide and targets with a substitution in position 3.

Asparagine −18 might also be involved in the recognition of base pair 4. It has been suggested that a water molecule might stabilize the complex by bridging N6/N7 of adenine 4 and Nδ of asparagine −18 (1, 2). The observed binding of the wildtype

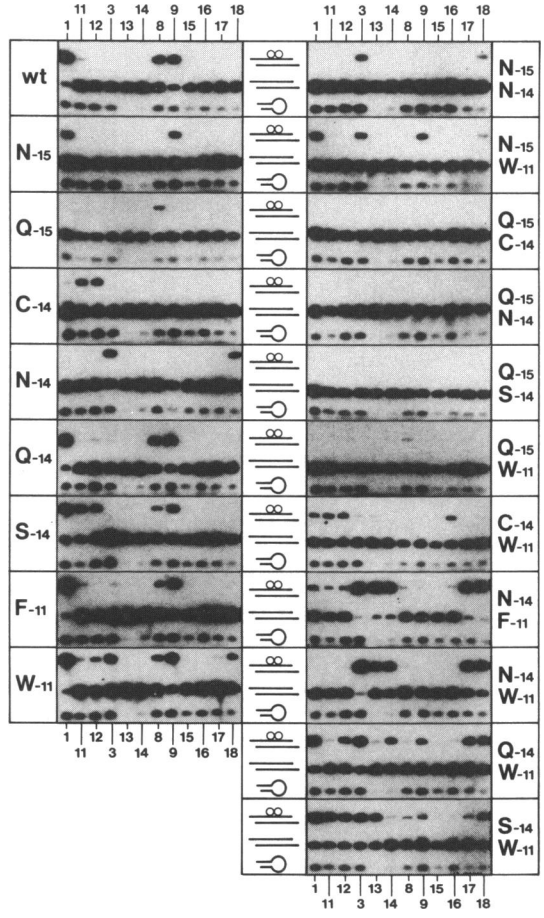


Figure 3. Mobility shift assays of cell extracts containing the wildtype GCN4 bZip peptide or the indicated single and double mutants in combination with a new set of binding sites. The numbers above and below the binding sites refer to the targets used. Targets, which are also used in figure 1, have identical numbers in both figures. Additional targets, which do not occur in figure 1, are numbered from 11 to 18. As in figure 1, the respective amino acid replacements are indicated on the left (single mutants) or on the right side (double mutants) of the respective autoradiographs.

GCN4 peptide to all target variants with substitutions in position 4 agrees with the proposed indirect DNA–protein interaction, with purines preferred over pyrimidines. Similar results were obtained with a different test system (30). The authors analysed the relative affinities of full lenght GCN4 to a set of symmetric single exchange target variants of the pseudo-palindromic AP1 site. In contrast to our results they did not observe binding of GCN4 to their target variant with thymine in position 2.

Single mutants in positions −15 or −11 of the basic region

In the wildtype GCN4/DNA complexes, both alanine −15 and serine −11 make hydrophobic contacts with the methyl group of thymine 3. Serine −11 might in addition be involved in a H-bond with an unesterified phosphate oxygen between adenine 4 and thymine 3 (1, 2). Residue −15 of the basic region seems to be critical for stable complex formation: although all peptides were expressed in sufficient amounts (data not shown), only six of the −15 single mutants show detectable DNA binding activity. In contrast nine of the −11 single mutants are able to bind to DNA (figures 1 and 2).

Peptides with GCN4-like DNA binding specificities. The peptides with leucine, methionine, glutamine, threonine or valine in position −11, or asparagine, cysteine, or serine in position −15 (figures 1 and 2) resemble the wildtype GCN4 peptide in their function. They bind best to the ATF/CREB site, although mostly with more or less reduced relative affinity in comparison to the GCN4 peptide, and they recognize some or all of the target variants which are bound by the GCN4 peptide with lowered relative affinities. Except for the serine −15 peptide they do not complex the target variant with thymine in position 2 (figures 2, 3). In the case of the larger aliphatic side chains in position −11 (leucine, methionine, threonine and valine) steric hindrance by the methyl groups in positions 2 and 3 might prevent recognition of the target variant with thymine in position 2.

The failure of the asparagine −15 peptide to recognize the target variant with thymine in position 2 might possibly be explained by a rearrangement of the side chain of asparagine −18, which could direct its amide function towards N7 of adenine 2′ in order to allow an interaction between Oδ of asparagine −18 and Nδ of asparagine −15. This might result in the repulsion

	Symmetric base pair exchanges in position										
ATF/CREB	1		2	1 and 2		4		1 and 4		2 and 4	
5′	5′	5′	5′	5′	5′	5′	5′	5′	5′	5′	5′
A4	.	.	.	.	.	C	G	C	C	C	G
T3	.	.	.	.	.	.	.	.	.	.	.
G2	.	.	C	C	C	.	.	.	.	C	C
A1	G	T	.	G	T	.	.	G	T	.	.
C0	.	.	.	.	.	.	.	.	.	.	.
G0′	.	.	.	.	.	.	.	.	.	.	.
T1′	C	A	.	C	A	.	.	C	A	.	.
C2′	.	.	G	G	G	.	.	.	.	G	G
A3′	.	.	.	.	.	.	.	.	.	.	.
T4′	.	.	.	.	.	G	C	G	G	G	C
3′	3′	3′	3′	3′	3′	3′	3′	3′	3′	3′	3′

-18			-15	-14				-11	-10
N	T	E	A	A	R	R	S	R	
.	.	.	N	.	.	.	.	.	
.	.	.	Q	.	.	.	.	.	
.	.	.	.	C	.	.	.	.	
.	.	.	.	N	.	.	.	.	
.	.	.	.	Q	.	.	.	.	
.	.	.	.	S	.	.	.	.	
.	.	.	.	.	.	.	F	.	
.	.	.	.	.	.	.	W	.	
.	.	.	N	N	.	.	.	.	
.	.	.	N	.	.	.	W	.	
.	.	.	Q	C	.	.	.	.	
.	.	.	Q	N	.	.	.	.	
.	.	.	Q	S	.	.	.	.	
.	.	.	Q	.	.	.	W	.	
.	.	.	.	C	.	.	W	.	
.	.	.	.	N	.	.	F	.	
.	.	.	.	N	.	.	W	.	
.	.	.	.	Q	.	.	W	.	
.	.	.	.	S	.	.	W	.	

1	11	12	3	13	14	8	9	15	16	17	18
++++	-	-	-	-	-	++	+++	-	-	-	-
++	-	-	-	-	-	-	++	-	-	-	-
-	-	-	-	-	-	+	-	-	-	-	-
-	++	++	-	-	-	-	-	-	-	-	-
-	-	-	++	-	-	-	-	-	-	-	++
++++	-	-	-	-	-	++	+++	-	-	-	-
+++	++	++	-	-	-	+	++	-	-	-	-
++++	+	-	-	-	-	++	+++	-	-	-	-
++++	+	+	++	-	-	++	+++	-	-	-	++
-	-	-	++	-	-	-	-	-	-	-	+
++	-	-	++	-	-	-	++	-	-	-	+
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	+	-	-	-	-	-
++	++	++	-	-	-	-	-	-	++	-	-
++	+	++	++++	++++	++++	-	-	-	-	++++	++++
-	-	-	++++	++	+++	-	-	-	-	+++	+++
+++	-	++	+++	-	++	-	++	-	-	++	++
+++	+++	+++	+++	+++	-	+	+	-	-	+	++

Figure 4. Summary of the results shown in figure 3. For a detailed explanation, see the legend to figure 2. The numbering of the binding sites between the upper and lower right panels corresponds to the numbering of the lanes in figure 3.

of Oδ of asparagine −15 by O4 of thymine 3 and thus prevent complex formation.

The serine −15 peptide binds to all targets which are bound by the GCN4 peptide with very similar relative affinities (figures 2, 3). This suggests that the methylene carbon of serine can substitute well for the methyl group of alanine. A dipole−dipole interaction with asparagine −18 seems possible and could stabilize the complex. As shown previously, both serine in position −15 and arginine in position −17 are necessary to allow recognition of a TAF1 binding site, which differs from the ATF/CREB site in positions 2, 3 and 4 (18).

In contrast to the other peptides of this group, the tyrosine −15 peptide has equal relative affinities (+) for the target variants with guanine or cytosine in position 4. The ATF/CREB site is recognized with a slightly reduced relative affinity when compared with the variants in position 4 (figures 1 and 2). The loss of binding to the target variant with thymine in position 2 might be explained in a similar manner to the aliphatic side chains in position −11 (see above). The altered preference of cytosine or guanine over adenine in position 4 as compared to the wildtype GCN4 peptide and the other peptides of this group (figures 1 and 2) might be caused by an interference of tyrosine −15 and asparagine −18. In a recently published work (30) the full lenght GCN4 variant with tyrosine −15 was tested for binding to symmetric single exchange targets derived from the pseudo-palindromic AP1 site. The authors observed equally strong binding to the AP1 site and to its variant with guanine in position 4, and reduced complex formation with the cytosine 4 variant, in agreement with our results. However, they found additional complexes with AP1 site variants with thymine in position 4, or cytosine or guanine in position 3 (30), while we could not detect any of the corresponding complexes with ATF/CREB site variants. This discrepancy might be caused by the different target geometries within the complexes of GCN4 with either the AP1 site (1) or the ATF/CREB site (2).

Peptides with broadened DNA binding specificities. The GCN4 mutants with cysteine, phenylalanine, histidine, or tryptophan in position −11 of the basic region bind with similar relative affinities to the binding sites which are bound by the GCN4 peptide. However, they also recognize with low or moderate relative affinities the target variants with cytosine in position 2 or adenine in position 3 (figures 1 and 2; the complex between the phenylalanine −11 peptide and the cytosine 2 target becomes visible only after prolonged exposure of the film). The tryptophan −11 peptide exhibits the highest relative affinity (++) for the cytosine 2 binding site and thus has one of the most significantly broadened DNA binding specificity among the single mutant bZip peptides in position −11 or −15 of the basic region (figures 2, 3). Since the common feature of tryptophan, histidine, and phenylalanine is the extended hydrophobic surface of the aromatic ring systems, we expected a similar pattern of hydrophobic interactions in all three complexes.

In order to analyse the interaction on a molecular level, we modelled the complex of the tryptophan −11 peptide with the cytosine 2 target on the basis of the GCN4/AP1 complex (1). Figures 7A and B show how tryptophan −11 is pushed towards base pair 2 mainly by the methyl group of thymine 3, thereby getting close to the more hydrophobic parts of cytosine 2. No steric clashes are observed with tryptophan −11 after minimization. The dihedral angle (N−Cα−Cβ−Cγ) is 162°, which is allowed even for large, bulky side chains. In accordance with our hypothesis, interactions between the extended hydrophobic areas of both tryptophan −11 and cytosine 2 rather than interactions between functional groups seem to play the crucial role for the new DNA binding specificty: the nonpolar region between C2 and C4 of the indol ring system as well as the β-methylene group of tryptophan −11 come close to the nonpolar area between C2′ of the deoxyribose and C5, C6 of cytosine 2 [figures 7A and B; (31)]. The distance between N1 of tryptophan and N4 of cytosine 2 suggests weak dipole−dipole interactions, while N4 of cytosine 2 may still form its H-bond to Oδ of asparagine −18 (figures 7A and B). Such a mechanism would also offer an explanation for the weak relative affinities of the peptides with histidine, phenylalanine or cysteine in position

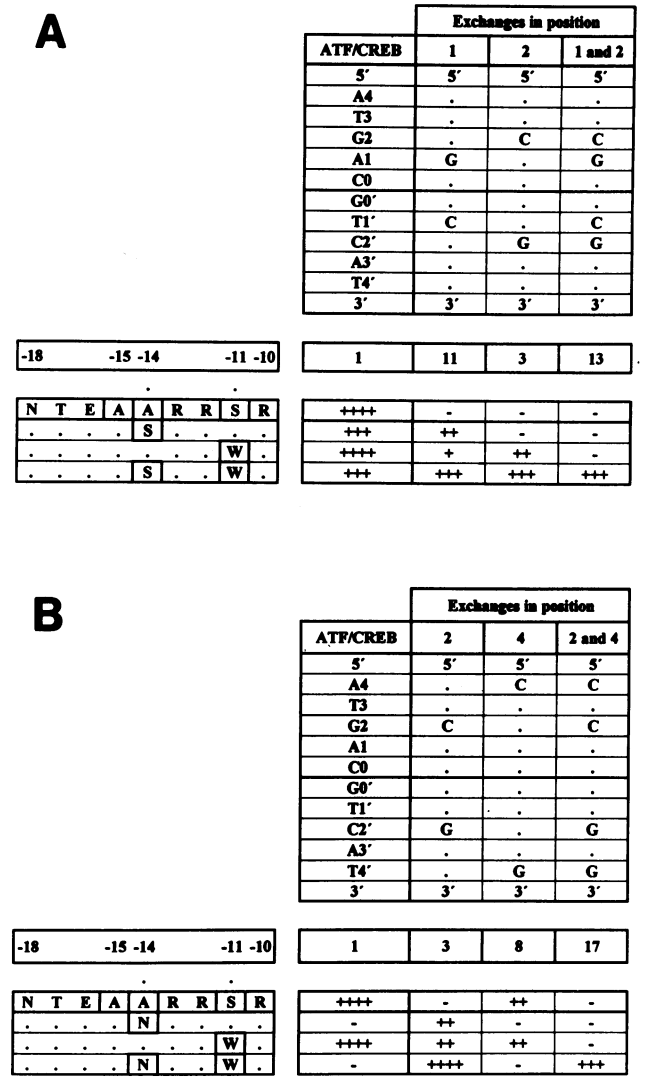


Figure 5. Direct comparison of single and double mutants shows additivity of individual amino acids. The data are extracted from figure 4 and are based on the autoradiographs shown in figure 3. The numbers of the binding sites between the upper and the lower right panels correspond to the respective lanes in figure 3 and to the numbering in figure 4. For a detailed explanation, see the legend to figure 2. (**A**) Serine −14 and tryptophan −11 act additively and recognize the double exchange target with guanine in position 1 and cytosine in position 2. (**B**) Asparagine −14 and tryptophan −11 act additively and recognize the target variants with cytosine in position 2.

−11 of the basic region for the cytosine 2 binding site: these residues either have comparable nonpolar areas, or, as cysteine, a slightly polar function. However, as indicated by the weaker complexes, the fitting may be suboptimal in comparison to tryptophan −11.

The moderate binding (++) of the phenylalanine −11 peptide to the adenine 3 target may be explained by a strong hydrophobic interaction between phenylalanine −11 and the methyl groups of thymine 3′ and thymine 4′. Since alanine −15 and asparagine −18 should not be able to contact adenine in position 3, the new hydrophobic interaction can obviously compensate for the loss of two base pair contacts. A schematic drawing of the complex between the phenylalanine −11 peptide and the adenine 3 target is shown in figure 6B.

Peptides which bind exclusively to the cytosine 4 target variant. The single mutant GCN4 peptides with either glutamic acid or glutamine in position −15 of the basic region do not bind to the ATF/CREB site or to any other binding site which is recognized by the GCN4 peptide, except the variant with cytosine in position 4 (+; figures 1 and 2). Since the wildtype GCN4 peptide also complexes this binding site (++; figures 1 and 2), the DNA binding properties of these mutants can not be regarded as altered specificity. It seems that glutamic acid or glutamine in position −11 disrupt all GCN4 complexes except for those with the cytosine 4 target variant. The low relative affinity for this target could hint at interactions between N4 of cytosine 4 and the functional groups of glutamic acid or glutamine −15.

The double mutants

In order to test if altered DNA binding properties of particular single mutant bZip peptides can be combined, we chose as a basis for double mutants mainly those single mutants in positions −14 (3), −15 and −11 (figures 1 and 2) of the basic region which differ in their DNA binding pattern from the GCN4 peptide most significantly: asparagine, glutamic acid or glutamine in position −15; serine, cysteine, glutamine or asparagine in position −14; and tryptophan or phenylalanine in position −11. A set of twelve binding sites consisting of the ATF/CREB site, symmetric single exchange variants, which are bound by the respective single mutant peptides [figures 1 and 2; (3)], and double exchange targets, which we expected to be bound by double mutant peptides (figure 4, upper panel), were chosen as targets. In addition, the wildtype GCN4 peptide and the parental single mutant peptides were also tested for their relative affinities to the new set of binding sites (figures 3 and 4).

As depicted in figures 3 and 4, neither the wildtype GCN4 peptide, nor the single mutants form complexes with the double exchange targets, with the exceptions of the asparagine −14 or the tryptophan −11 peptides, which recognize with moderate relative affinities (++) the guanine 4 / cytosine 2 variant. In contrast, several double mutant bZip peptides bind to double exchange targets.

Double mutants with asparagine −15. The main difference between the asparagine −15 single mutant and the wildtype GCN4 peptide is the inability of the asparagine −15 peptide to bind the thymine 2 single exchange target (figures 1 and 2). If asparagine −15 is combined with asparagine −14, the resulting double mutant exhibits the same DNA binding specificity as the asparagine −14 single mutant with only slightly reduced relative affinities: it binds moderately (++) to the cytosine 2 target and weakly (+) to the cytosine 2 / guanine 4 target, but not to the ATF/CREB site nor to the guanine 4 target variant, which are both bound by the asparagine −15 single mutant (figures 3 and 4). It is remarkable that the accumulation of three asparagines (−18, −15 and −14) within a five residue segment of the recognition helix does not result in the loss of specific DNA binding. A possible explanation for the complex formation with the cytosine 2 target could be a network of hydrogen bonds involving O4 of thymine 3 and the Nδ's of asparagines −18 and −15 beside Nδ and Oδ of asparagine −14, and O4 of thymine 1′ and N4 of cytosine 2, respectively (figure 6C). Such an arrangement should not interfere with the indirect contact between

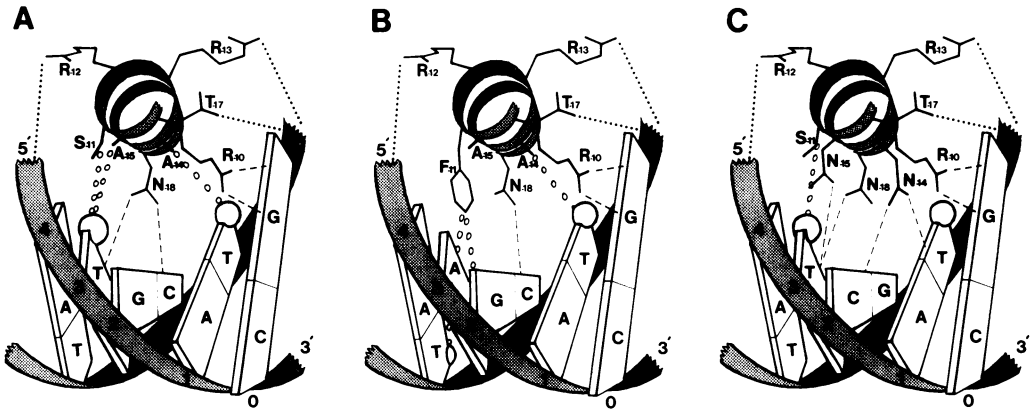


Figure 6. Schematic drawings of proposed amino acid-base pair interactions in mutant complexes compared with the wildtype GCN4 half-complex [**A**; adapted from Ellenberger *et al.* (1)]. All side chains of the −18 to −10 segment of the basic region are shown, which are believed to be involved in specific contacts with base pairs or unspecific contacts with the phosphate backbone. Circles indicate hydrophobic interactions, hydrogen bonds are symbolized either as dashed lines, when bases are involved, or as dotted lines to the phosphate backbone. The complexes were checked for their steric probabilities with Insight II on the basis of the coordinates of the GCN4/AP1 site cocrystal (1), without minimization. The numbering of bases and amino acids corresponds to the numbering in figures 2, 4, and 5. (**B**) The complex between the phenylalanine −11 mutant and the adenine 3 target variant. (**C**) the complex between the asparagine −14 / asparagine −15 double mutant and the cytosine 2 target variant.

asparagine −18 and base pair 4, as indicated by the weak binding of this double mutant to the cytosine 2 / guanine 4 target.

The peculiarity of the double mutant with asparagine −15 and tryptophan −11 is the equally moderate binding (++) to the ATF/CREB site, the cytosine 2 variant and the guanine 4 variant. In addition, it weakly recognizes the double exchange target with cytosine 2 and guanine 4 (figures 3 and 4). This double mutant seems to combine features of the respective single mutants: moderate relative affinities for the ATF/CREB site and for the guanine 4 target and no detectable interaction with the cytosine 4 target may be caused by asparagine −15, whereas the recognition of the cytosine 2 and the cytosine 2 / guanine 4 targets could be attributed to tryptophan −11.

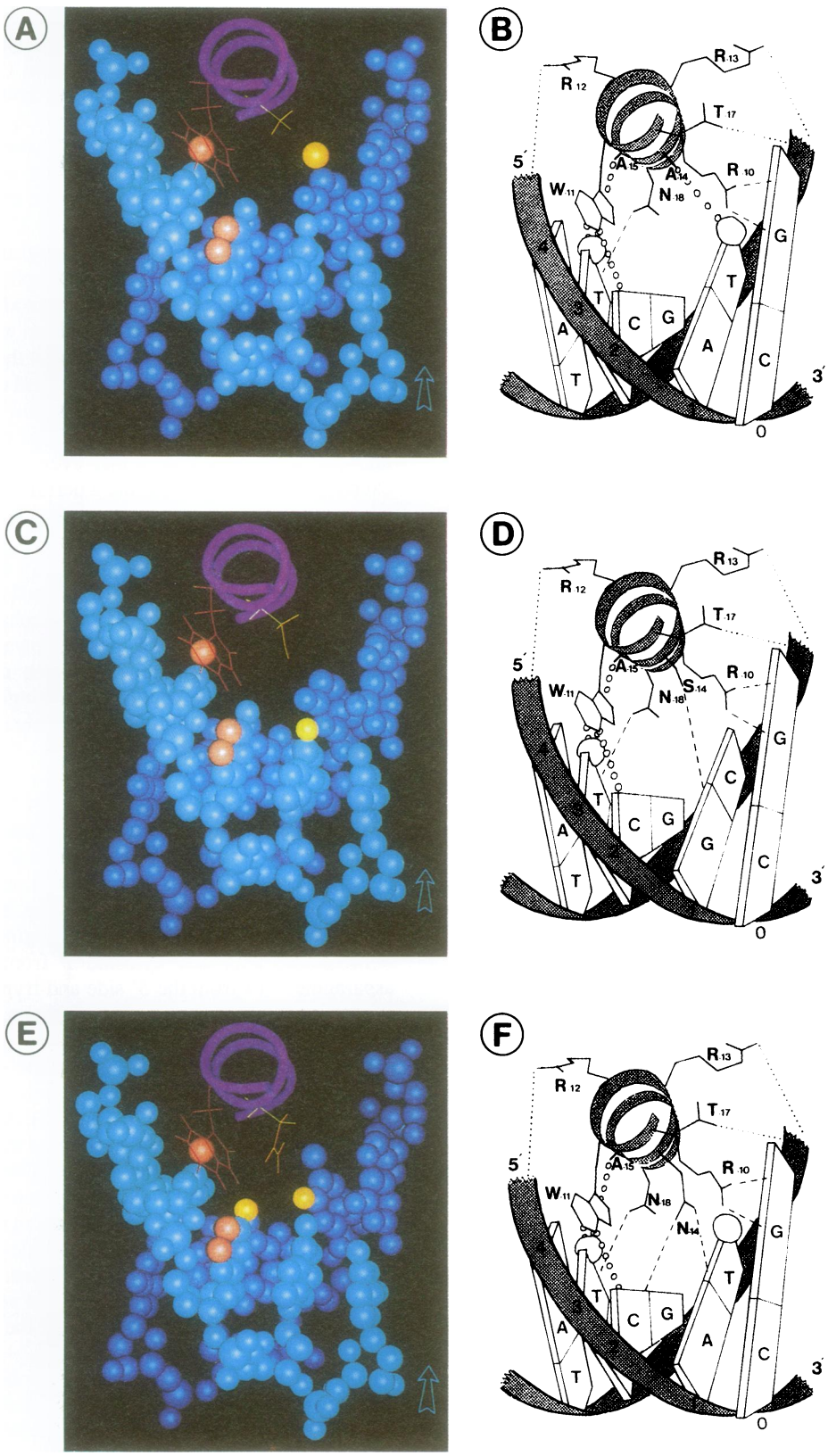
Double mutants with glutamine −15. The glutamine −15 single mutant binds weakly but exclusively to the cytosine 4 target (figures 1, 2, 3, and 4). Therefore, a direct interaction between glutamine −15 and cytosine 4 seems possible. The side chains of two residues, which are adjacent in the primary sequence, are well separated on the surface of an α-helix. Thus they should be able to establish independent base pair contacts. We therefore expected the double mutants with glutamine −15 and serine or cysteine −14 to recognize double exchange targets with cytosine 4 and guanine or thymine 1. For the same reasons we expected complex formation between the glutamine −15 / asparagine −14 double mutant and the cytosine 4 / cytosine 2 double exchange target. However, none of these double mutants shows detectable binding to any of the binding sites tested, nor does the double mutant with glutamine −15 and tryptophan −11 bind to the cytosine 4 / cytosine 2 target variant, as was predicted from the DNA binding properties of the parental single mutants. On the other hand, this double mutant binds weakly to the cytosine 4 single exchange target as does the glutamine −15 single mutant (figures 3 and 4). We also tested the corresponding four double mutants with glutamic acid −15 instead of glutamine, but none of them binds to any of the targets (data not shown). The loss of the DNA binding function of these double mutants is possibly the consequence of the per se low relative affinities (+) of the parental single mutants with glutamic acid or glutamine in position −15 for the cytosine 4 target variant (figures 1, 2, 3, and 4).

Serine / cysteine −14 and tryptophan −11. Since the serine or the cysteine −14 single mutants bind to the target variants with thymine or guanine in position 1 [figures 3 and 4; (3)], and since the tryptophan −11 single mutant binds to the cytosine 2 target (figures 1, 2, 3, and 4), we expected the resulting double mutant to bind to the double exchange targets with cytosine 2 and guanine or thymine 1. The serine −14 / tryptophan −11 double mutant indeed recognizes the cytosine 2 / guanine 1 target with a high relative affinity (+++) and thus clearly shows the expected additive effect (figure 5A), whereas the complex with the cytosine 2 / thymine 1 target is much weaker and becomes visible only after prolonged exposure of the film (figures 3 and 4). As summarized in figure 5A, the serine −14 and the tryptophan −11 single mutants bind with high relative affinities to the ATF/CREB side. In addition, the serine −14 peptide binds moderately (++) to the guanine 1 target, but not to the cytosine 2 or the cytosine 2 / guanine 1 targets, whereas the tryptophan −11 single mutant recognizes moderately (++) the cytosine 2 target, but not the cytosine 2 / guanine 1 target and only weakly the variant with guanine 1. The wildtype GCN4 peptide binds to none of these binding sites (figure 5A). Thus the serine −14 / tryptophan −11 double mutant combines the features of the respective single mutants: it binds well (+++) to the ATF/CREB site (similar to the GCN4 wildtype peptide and both single mutants), to the guanine 1 target (similar to the serine −14 single mutant), and to the cytosine 2 target (similar to the tryptophan −11 single mutant). In addition it recognizes well (+++) the cytosine 2 / guanine 1 double exchange target (figure 5A), most probably by independent interactions between Oγ of serine −14 and O6 of guanine 1 on the one hand, and between the hydrophobic areas of tryptophan −11 and cytosine 2 on the other hand. A computer-generated model as well as a schematic drawing of the proposed complex is shown in figures 7C and D.

Surprisingly the corresponding synergistic effect is not observed in the case of the cysteine −14 / tryptophan −11 double mutant. As shown in figures 3 and 4, it forms moderate (++) complexes with the ATF/CREB site and its variants with guanine or thymine 1. Unexpectedly, it also interacts with the cytosine 4 / thymine 1 double exchange variant, which we had designed for the double mutants with glutamine −15 and serine or cysteine −14. The lack of binding of the double exchange targets with cytosine 2 and guanine or thymine 1 and even of the cytosine 2 single exchange target might indicate a perturbation of the interactions between tryptophan −11 and cytosine 2. Such an interpretation would be supported by the moderate (++) complex formation of the cysteine −14 / tryptophan −11 double mutant and the cytosine 4 / thymine 1 target. In contrast to the tryptophan −11 single mutant or the serine −14 / tryptophan −11 double mutant, in which tryptophan −11 is pushed towards base pair 2 by the methyl group of thymine 3, cysteine might inhibit such an arrangement so that a new DNA binding specificity results: Sγ of cysteine −14 could possibly contact thymine 1, while tryptophan −11 contacts cytosine 4.

Double mutants with asparagine −14 / tryptophan −11 and similar combinations. Both the asparagine −14 and the tryptophan −11 single mutants are able to bind (++) to the cytosine 2 target (figures 3 and 4), but, as discussed previously (3) and above, we propose different mechanisms which involve different regions of cytosine 2. Asparagine −14 and tryptophan −11 should approach cytosine 2 from opposite directions: asparagine −14 from the 3′ side and tryptophan −11 from the 5′ side. Both amino acids might recognize the same base without disturbing each other and thus we expected enhanced binding to the cytosine 2 target for the asparagine −14 / tryptophan −11 double mutant.

As shown in figures 3, 4, and 5B, the asparagine −14 / tryptophan −11 double mutant shows a striking change in DNA binding specificity compared to the wildtype GCN4 peptide. It fails to recognize the ATF/CREB site, but indeed binds with very high relative affinity (++++) to the cytosine 2 target. It also binds well (+++) to most of the double exchange targets with cytosine in position 2 which we have tested. None of these binding sites is recognized by the wildtype GCN4 peptide or the respective single mutants, with the exception of the cytosine 2 / guanine 4 target, which is moderately bound (++) by both parental mutants (figures 3 and 4). The slight preference of the guanine 4 / cytosine 2 target over the cytosine 4 / cytosine 2 target (figure 3) is qualitatively paralleled by the preferences of the wildtype GCN4 peptide with respect to base pair 4 of the ATF/CREB site (figures 1, 2, 3, and 4) and may indicate that asparagine −18 makes similar contacts to base pair 4 in both complexes (1, 2). The cytosine 2 / guanine 1 target is also well (+++) bound


A
B
R13
R12
T17
5′
A15
A14
R10
W11
N18
G
T
T
C
G
A
A
C
T
3′
0
C
D
R13
R12
T17
5′
A15
R10
W11
S14
N18
G
C
T
C
G
A
G
C
T
3′
0
E
F
R13
R12
T17
5′
A15
R10
W11
N18
N14
G
T
T
C
G
A
A
C
T
3′
0

by the serine −14 / tryptophan −11 double mutant, but in this case most probably by different interactions (see preceeding section and figures 7C and D).

The non-binding of the asparagine −14 / tryptophan −11 double mutant to the ATF/CREB site may be due to asparagine in position −14, whereas the very high relative affinity (+++++) of the double mutant for the cytosine 2 target seems to indicate independent interactions between cytosine 2 and asparagine −14 on the one side and tryptophan −11 on the other side (figure 5B). Furthermore, the double mutant recognizes, as does the wildtype GCN4 peptide, cytosine (or guanine) in position 4, but in contrast to the GCN4 peptide, which requires guanine in position 2 for the corresponding complexes, only in combination with cytosine 2 (figure 5B).

In order to take a closer look at the interactions of asparagine −14 and tryptophan −11 with cytosine 2, we modelled the corresponding complex on the basis of the GCN4/AP1 site structure (1). The results are shown in figures 7E and F. Tryptophan −11 may be held close to cytosine 2 by hydrophobic interactions with the methyl group of thymine 3, the hydrophobic parts of asparagine −14 and the hydrophobic area of cytosine 2. At the same time asparagine −14 is able to form H-bonds with O4 of thymine 1′ and N4 of cytosine 2. The wildtype hydrogen bond between arginine −10 and guanine 0', the hydrophobic interactions between alanine -15 and thymine 3, and the putative indirect interaction between asparagine −18 and base pair 4 are still possible.

If tryptophan −11 is replaced by phenylalanine, the resulting peptide behaves very similarly with respect to the binding sites which carry cytosine 2, but it can also form weak or moderate complexes with the ATF/CREB site and its variants with guanine or cytosine in position 1 (figures 3 and 4). This confirms that the major determinants for the recognition of cytosine 2 by tryptophan −11 are indeed hydrophobic interactions, as discussed above. Since phenylalanine does not have any functional groups which could be involved in hydrogen bonds, a possible dipole−dipole interaction between N1 of asparagine −11 and N4 of cytosine 2 does not seem to contribute significantly to the complex stability.

If asparagine −14 of the asparagine −14 / tryptophan −11 double mutant is replaced by glutamine, the resulting peptide binds equally well (+++) to the ATF/CREB site and the cytosine 2 target. Furthermore, it recognizes (++) the double exchange targets with cytosine 2 and cytosine or guanine 4. Moderate complexes are also observed with the thymine 1, the cytosine 2 / thymine 1 and the guanine 4 binding sites (figures 3 and 4). These results underline the importance of asparagine in position −14 of the asparagine −14 / tryptophan −11 double mutant for the observed specificity switch. Furthermore, they indicate the important role of surrounding residues for the function of a particular amino acid. Although the DNA binding specificities of the wildtype GCN4 peptide and the glutamine −14 single mutant appear to be very similar [figures 3 and 4; (3)], those of the tryptophan −11 single mutant and the glutamine −14 / tryptophan −11 double mutant differ significantly (figures 3 and 4). Thus the potential of glutamine in position −14 to contribute to specific recognition remains hidden in combination with serine −11, but is uncovered with the assistance of tryptophan −11.

ACKNOWLEDGEMENTS

We thank W.Fiers for the gift of plasmid pPLc28 and *E.coli* K12DH1, K.Otto and G.Zimmer for syntheses of oligonucleotides, P.Niland and F.Zeiss for part of the screening of the −11 single mutants, A.Barker for critical reading of the manuscript, and U.Ringeisen for photography. We especially thank T.E.Ellenberger for stimulating discussions and for providing the GCN4 crystal structure data. This work was supported by DFG through SFB 274 including a stipend to A.Madan for three months, through SFB 243, and by BMFT through 'Genzentrum Köln'.

REFERENCES

1. Ellenberger, T.E., Brandle, C.J., Struhl, K. and Harrison, S.C. (1992) *Cell* 171, 1223−1237.
2. König, P. and Richmond, T.J. (1993) *J. Mol. Biol.* 233, 139−154.
3. Suckow, M., v.Wilcken-Bergmann, B. and Müller-Hill, B. (1993) *Nucleic Acids Res.* 21, 2081−2086.
4. Hinnebusch, A.G. (1984) *Proc. Nat. Acad. Sci. U.S.A.* 81, 6442−6446.
5. Hinnebusch, A.G. (1990) *TIBS* 15, 148−152.
6. Hope, I.A. and Struhl, K. (1985) *Cell* 43, 177−188.
7. Hinnebusch, A.G. (1988) *Microbiol. Rev.* 52, 248−273.
8. Hope, I.A. and Struhl, K. (1986) *Cell* 46, 885−894.
9. Hope, I.A. and Struhl, K. (1987) *EMBO J.* 6, 2781−2784.
10. Hill, D.E., Hope, I.A., Macke, J.P. and Struhl, K. (1986) *Science* 234, 451−457.
11. Oliphant, A.R., Brandle, C.J. and Struhl, K. (1989) *Mol. Cell. Biol.* 9, 2944−2949.
12. Angel, P., Allegretto, E.A., Okino, S.T., Hattori, K., Boyle, W.J., Hunter, T. and Karin, M., (1988) *Nature* 332, 166−177.
13. Sellers, J.W., Vincent, A.C. and Struhl, K. (1990) *Mol. Cell. Biol.* 10, 5077−5086.
14. Hai, T., Liu, F.,Coukos, W.J. and Green, M.R. (1989) *Genes Develop.* 3, 2083−2090.
15. Dwarki, V.L., Montminy, M. and Verma, I.M. (1990) *EMBO J.* 9, 225.232.

Figure 7. Minimized computer models of mutant GCN4 half-complexes. The complexes were modelled on the basis of the coordinates of the GCN4/AP1 site complex (1) with Insight II and minimized with the amber force field of Discover. The illustrations on the left (A, C, and E) only show the side chains of tryptophan −11 (red) and of the respective amino acids in position −14 (yellow). Atoms on the DNA surface, which may be contacted, are shown in the same colour as the interacting amino acids. The backbone of the protein α-helix is coloured magenta, whereas the DNA is coloured blue. The corresponding schematic drawings on the right (B, D, and F) are based on a cartoon from Ellenberger *et al.* (1), and illustrate the complete network of interactions. Bases and amino acids are numbered as in figures 2, 4 and 5. The symbols for the different kinds of interactions are described in the legend to figure 6. (**A**) The complex between the tryptophan −11 single mutant and the cytosine 2 target variant. Tryptophan −11 (red) is pushed towards the hydrophobic parts of cytosine 2 (C5, C6; lighter red) by the methyl group of thymine 3 (lighter red), while alanine −14 (yellow) contacts the methyl group of thymine 1′ (yellow) as in the wildtype complex. (**B**) The corresponding schematic drawing. (**C**) The complex between the tryptophan −11 / serine −14 double mutant and the cytosine 2 / guanine 1 double exchange target. The function of tryptophan −11 (red) is the same as depicted in (A). Oγ of serine −14 (yellow) contacts O6 of guanine 2 (yellow). (**D**) The corresponding schematic drawing. (**E**) The complex between the tryptophan −11 / asparagine −14 double mutant and the cytosine 2 target variant. Tryptophan −11 (red) acts as depicted in (A), while Oδ and Nδ of asparagine −14 (yellow) contact N4 of cytosine 2 (yellow) and O4 of thymine 1′ (yellow), respectively. (**F**) The corresponding schematic drawing.

16. Vinson, C.R., Sigler, P.B. and McKnight, S.L. (1989) *Science* 246, 911–916.
17. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) *Science* 240, 759–1764.
18. Suckow, M., v.Wilcken-Bergmann, B. and Müller-Hill, B. (1993) *EMBO J.* 12, 1193–1200.
19. O'Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T. (1991) *Science* 254, 539–544.
20. Vinson, C.R., Hai, T. and Boyd, S.M. (1993) *Genes Develop.* 7, 1047–1058.
21. Talanian, R.V., McKnight, C.J. and Kim, P.S. (1990) *Science* 249, 769–771.
22. Agre, P., Johnson, P.F. and McKnight, S.L. (1989) *Science* 246, 922–926.
23. Kouzarides, T. and Ziff, E. (1989) *Nature* 340, 568–571.
24. Kanaan, M.N. and Marzluf, G.A. (1991) *Mol. Cell. Biol.* 11, 4356–4362.
25. O'Neil, K.T., Hoess, R.H. and DeGrado, W.F. (1990) *Science* 249, 774–778.
26. O'Neil, K.T., Shuman, J.D., Ampe, C. and DeGrado, W.F. (1991) *Biochemistry* 30, 9030–9034.
27. Busch, S.J. and Sassone-Corsi, P. (1990) *TIG* 6, 36–40.
28. Johnson, P.F. (1993) *Mol. Cell. Biol.* 13, 6919–6930.
29. Schägger, H. and v.Jagow, G. (1987) *Anal. Biochem.* 166, 368–379.
30. Kim, J., Tzamarias, D., Ellenberger, T.E., Harrison, S.C. and Struhl, K. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 4513–4517.
31. Orozco, M. and Lugue, F.J. (1993) *Biopolymers* 33, 1851–1869.
32. Remault, E., Stanssens, P. and Fiers, W. (1981) *Gene* 15, 81–93.