The EMBO Journal vol.15 no.3 pp.598–606, 1996

Mutant bZip–DNA complexes with four quasi-identical protein–DNA interfaces

Manfred Suckow[1], Michael Lopata, Anke Seydel, Brigitte Kisters-Woike, Brigitte von Wilcken-Bergmann and Benno Müller-Hill[2]

Institut für Genetik der Universität zu Köln, Weyertal 121, 50931 Köln, Germany

[1]Present address: Institut für Mikrobiologie der Heinrich-Heine-Universität zu Düsseldorf, Universitätsstrasse 1, Geb. 26.12, 40225 Düsseldorf, Germany

[2]Corresponding author

The complex between the yeast transcriptional activator GCN4 and the palindromic ATF/CREB site 5′-A4T3G2A1C0*G0′T1′C2′A3′T4′-3′ shows dyad symmetry. The basic region of GCN4 contains a segment of 18 amino acids with a partially palindromic sequence: N-LKRARNTEA*ARRSRARKL-C. Symmetric residues are underlined. Apart from the ATF/CREB site, GCN4 also binds well to the symmetric variants with guanine in position 4 (5′-G4T3G2A1C0*G0′T1′C2′A3′C4′-3′) or thymine in position 0 (5′-A4T3G2A1T0*A0′T1′C2′A3′T4′-3′). The half-sites of these sequences can be regarded as short pseudo-palindromes with central guanine 2/cytosine 2′ base pairs. We investigated whether the geometry of the peptide of the basic region of GCN4 could be functionally related to the pseudo-palindromic character of some target half-sites. Since inspection of the X-ray structures of GCN4–DNA complexes reveals that several amino acid–DNA interactions are symmetric within the wild-type half-complexes, we introduced mutations into a GCN4 bZip peptide that improve the symmetry of the peptide. We found that most of the constructs retain specific DNA recognition. For one mutant, we conclude that it is not only capable of forming DNA complexes showing the well-known overall dyad symmetry, but that the protein–DNA interface of each half-complex can be divided further into two quasi-identical, quasi-symmetric substructures.

Keywords: amino acid–base pair interactions/bZip–DNA complexes/GCN4/palindromic peptide sequence/quasi-symmetric peptide

Introduction

Dyad symmetry is a common feature of protein–DNA complexes, in which a palindromic or a pseudo-palindromic DNA sequence is bound by a protein homodimer. Here, both monomers interact in an identical manner with each half-site of the DNA target. Thus, the overall structure of the complex is rotationally symmetric. This is rendered possible by the structural features of DNA: since the two strands are assembled complementarily in an antiparallel orientation, palindromic sequences necessarily exhibit dyad symmetry (Watson and Crick, 1953). Examples of proteins forming such symmetric DNA complexes are the prokaryotic helix–turn–helix (HTH) proteins (Aggarwal *et al.*, 1988; Mondragon *et al.*, 1989; Schuhmacher *et al.*, 1994), and the eukaryotic basic helix-loop-helix (bHLH) or bHLH leucine zipper (bHLHZip; Ferre-D'Amare *et al.*, 1993), and basic leucine zipper (bZip) proteins (Ellenberger *et al.*, 1992; König and Richmond, 1993) as well as zinc finger proteins of class 2 (Luisi *et al.*, 1991) or 3 (Pan and Coleman, 1990). As a common DNA-recognizing structure, the DNA binding domains of these transcription factors include an α-helix (for a review, see for example Harrison, 1991). In each half-complex, this recognition helix is specifically anchored inside the major groove of the respective target half-site, oriented more or less in parallel to the staples of base pairs. Consequently, a typical half-site is made up of no more than five subsequent base pairs, corresponding to a half turn of DNA. The two half-sites can be abutted, as in the targets for bZip, bHLH and bHLHZip proteins, or separated by about one half turn, as in the targets for HTH proteins and the zinc finger proteins of class 2 or 3. In all cases, the requirements for a certain type of half-site spacing are determined by the position of the two DNA binding domains within the protein dimer. The development of dimeric DNA binding proteins may be regarded as a simple, but efficient evolutionary strategy for enlarging the protein–DNA interface and thus for increasing the DNA binding specificity and affinity of the respective monomeric systems.

To date, >60 transcription factors have been described (for a review, see Hurst, 1994) that include as a common, conserved structural element the so-called bZip domain (Vinson *et al.*, 1989) which enables them to dimerize and to bind specifically to their DNA targets (Hope and Struhl, 1986). bZip domains fulfill their functions autonomously: short peptides comprising only the bZip domain dimerize and bind to DNA with the same specifity as the respective full-length proteins (Hope and Struhl, 1986; Agre *et al.*, 1989; Kouzarides and Ziff, 1989). When complexed with DNA, bZip domains form continous α-helices: their leucine zipper segments assemble to form a parallel coiled-coil structure, while the basic regions diverge and enter the major groove on each target half-site, thereby anchoring the aggregate sequence specifically over the center of symmetry of the binding site (Ellenberger *et al.*, 1992; König and Richmond, 1993; Glover and Harrison, 1995). Thus, the optimal binding sites for bZip proteins are 10 bp palindromes or 9 bp pseudo-palindromes with overlapping half-sites.

The basic region of the transcriptional activator GCN4

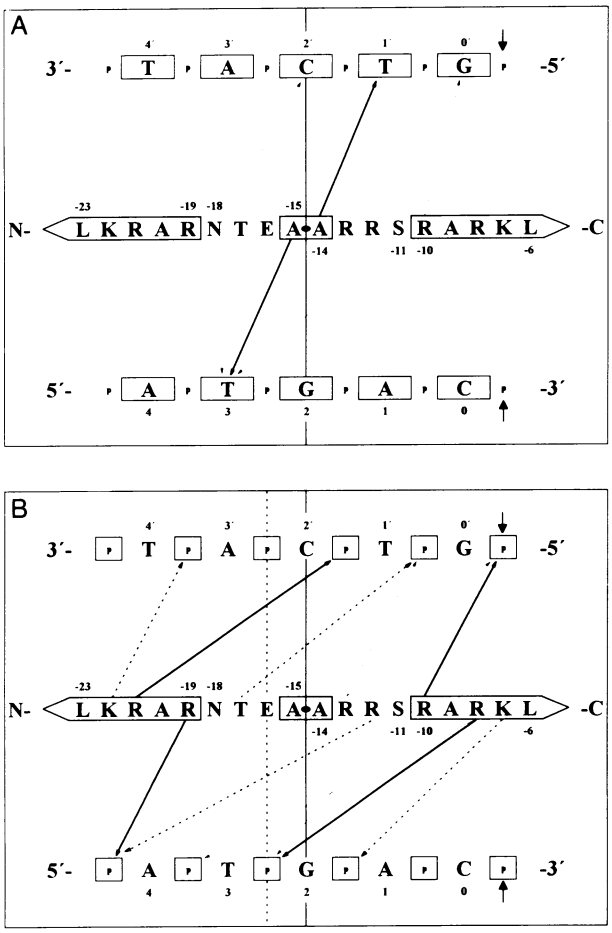

Fig. 1. 2-D projection of the protein–DNA interactions within a half-complex between GCN4 and the ATF/CREB site according to König and Richmond (1993). Residues –23 to –6 of the GCN4 basic region (counted in negative numbers from the first leucine of the leucine zipper towards the N-terminus; see Figure 2, top) are placed between the (+) strand (bottom) and the (–) strand (top) of the target half-site, with the numbering of bases used in this study. Residues which fit into the degenerate palindrome of the sequence (with its center of symmetry between Ala –15 and Ala –14) are boxed. The center of symmetry of the complete ATF/CREB site is indicated by small arrows in the upper and lower right corner of each drawing. A vertical line through the centers of symmetry of the amino acid sequence and of the target half-site (base pair 2/2′) symbolizes the putative axis of symmetry within the protein–DNA interface of a half-complex. Interactions which are qualitatively symmetric with regard to this axis are indicated by thick arrows, asymmetric interactions by thin arrows. (**A**) Amino acid–base interactions. (**B**) Amino acid–phosphate backbone interactions. In (B), some interactions are indicated with thick, dotted arrows. These are qualitatively symmetric with regard to an axis through the 5′ phosphates of guanine 2 and adenine 3′ (symbolized by a dotted, vertical line).

from *Saccharomyces cerevisiae* (Hinnebusch and Fink, 1983) contains a segment of 18 residues with the approximate palindromic peptide sequence: N-LKRARNTEA*ARRSRARKL-C (the center is marked; residues which fit into the palindrome are underlined). The degenerate palindrome of the peptide sequence extends from positions –23 to –6 of the GCN4 basic region (we define the start of the leucine zipper as +1; see Figure 1) and thus correlates with the area of the recognition helix. In contrast to palindromic DNA sequences which possess perfect dyad symmetry, a palindromic amino acid sequence which forms an α-helix is only approximately symmetric. In the single-helical molecule, the asymmetry of the peptide bonds, as well as the Cα–Cβ bonds which point towards the N-terminus, prevents the formation of a fully symmetric surface (Pauling *et al.*, 1951).

Two DNA sequences have been shown to be optimal for GCN4 binding. *In vivo*, GCN4 binds best to the 9 bp pseudo-palindromic AP1 site 5′-A4T3G2A1C0T1′-C2′A3′T4′-3′ (Hill *et al.*, 1986); *in vitro* it binds nearly equally well to the 10 bp palindromic ATF/CREB site 5′-A4T3G2A1C0*G0′T1′C2′A3′T4′-3′ (Sellers *et al.*, 1990), thereby forming a complex of perfect dyad symmetry (König and Richmond, 1993). However, GCN4 also binds to the symmetric single exchange variants of the ATF/CREB sequence with guanine in position 4 (5′-G4T3G2A1C0*G0′T1′C2′A3′C4′-3′; Suckow *et al.*, 1994a) or thymine in position 0 (5′-A4T3G2A1-T0*A0′T1′C2′A3′T4′-3′; Suckow *et al.*, 1993b) whose half-sites can be regarded as short pseudo-palindromes with central guanine 2/cytosine 2′ base pairs.

Considering the degenerate palindrome of the amino acid sequence of GCN4 on the one hand and the pseudo-palindromic character of these half-sites on the other hand, we wondered about a possible functional relationship. Might it be possible that, in the case of GCN4, not only the overall protein–DNA complex shows dyad symmetry, but also the protein–DNA interface of each half-complex can be further divided into two approximately symmetric substructures? If this were the case, then a single GCN4 recognition helix would superficially mimic a DNA binding dimer, and the organization of the complete protein–DNA interface would have remote similarity to the organization of a fractal. Here we present evidence for the proposition that such a structure is indeed functional.

Results

Symmetric protein–DNA interactions within a half-complex of the GCN4–ATF/CREB site; X-ray structure

Two X-ray studies of GCN4–DNA complexes are available, in which short peptides comprising the bZip domain of GCN4 are bound either to the pseudo-palindromic AP1 site 5′-A4T3G2A1C0T1′C2′A3′T4′-3′ (Ellenberger *et al.*, 1992) or to the palindromic ATF/CREB site 5′-A4T3G2-A1C0*G0′T1′C2′A3′T4′-3′ (König and Richmond, 1993). We scrutinized the X-ray structure of the complex between GCN4 and the ATF/CREB site (König and Richmond, 1993) with regard to possible symmetric protein–DNA interactions within one half-complex, for two reasons: (i) for the experiments shown below, we used a set of 10 bp palindromic targets which are derived from the ATF/CREB site; (ii) the GCN4–ATF/CREB site complex shows perfect dyad symmetry (König and Richmond, 1993), in contrast to the complex between GCN4 and the pseudo-palindromic AP1 (Ellenberger *et al.*, 1992). The complete network of protein–DNA contacts between the basic region of GCN4 and a half-site of the ATF/CREB sequence, according to the crystal structure of König and Richmond (1993), is shown in Figure 1A (amino acid–base contacts) and B (amino acid–DNA backbone contacts). In these schematic two-dimensional (2-D) projections, we placed the center of symmetry of

the GCN4 amino acid sequence between guanine 2 and cytosine 2′, which marks the center of the target half-site. Figure 1 shows that the palindromic peptide sequence (extending from position –23 to –6 of the basic region) correlates perfectly with the recognition helix (position –22 to –7).

Five amino acids of the GCN4 basic region contact bases within each half-complex (König and Richmond, 1993). Among these, two are not only symmetrically positioned within the degenerate palindrome of the peptide sequence but interact symmetrically with regard to the center of the target half-site (Figure 1A). Ala –15 binds to the methyl group of thymine 3 and Ala –14 to the methyl group of thymine 1′. In contrast, the H bonds between Nδ and Oδ of Asn –18 and O4 of thymine 3 and N4 of cytosine 2′, respectively, the hydrophobic interaction between Cβ of Ser –11 and the methyl group of thymine 3, and the H bond between the guanidinium group of Arg –10 and N7 of guanine 0′ (König and Richmond, 1993), do not have a symmetric counterpart in the respective other half of the half-complex (Figure 1A). Among the 10 amino acids of the GCN4 basic region which are engaged in contacts with the sugar phosphate backbone (König and Richmond, 1993), four occupy symmetric positions within the peptide sequence and at the same time establish symmetric backbone contacts with regard to base pair 2/2′ (Figure 1B). Arg –19 contacts the 5′ phosphate of adenine 4, Arg –10 contacts the 5′ phosphate of guanine 0′; Arg –21 contacts the 5′ phosphate of cytosine 2′, and Arg –8 the corresponding phosphate of guanine 2 (Figure 1B). Two further corresponding pairs of amino acids act symmetrically with regard to an axis through the 5′ phosphates of adenine 3′ and guanine 2 of the target half-site which, compared with the axis through guanine 2/cytosine 2′, is displaced towards the left half of the target half-site. Thr –17 contacts the sugar of guanine 0′ and the 5′ phosphate of thymine 1′, Arg –12 interacts with the 5′ phosphate of adenine 4; Lys –22 contacts the phosphate of thymine 4′, while Lys –7 contacts the phosphate of adenine 1 (Figure 1B). As in the case of the amino acid–base contacts, some interactions do not have a symmetrically related counterpart in the respective other half of the half-complex, such as the contact between Ser –11 and the phosphate of thymine 3, and the interaction between Arg –13 and the phosphates of cytosine 0 and guanine 0′ (Figure 1B).

The inspection of the crystal structure of the GCN4–ATF/CREB site complex (König and Richmond, 1993) thus indeed reveals features which may indicate a functional relationship between the degenerate palindrome of the peptide sequence of the GCN4 basic region and the geometry of the target half-site: (i) the recognition helix (residues –22 to –7; König and Richmond, 1993) correlates with the segment containing the degenerate palindrome of the peptide sequence (residues –23 to –6; Figure 1); (ii) two of the five base-contacting and eight of the 10 backbone-contacting amino acids (König and Richmond, 1993) are symmetrically positioned within the degenerate palindrome of the peptide sequence, and establish contacts which are at least similar in their respective halves of the half-complex (Figure 1A and B). However, it should be noted that the distances between the interacting functional groups (König and Richmond, 1993) are not identical for corresponding pairs of interactions, which are therefore partially, but not fully, symmetric, as shown in the 2-D projections in Figure 1.

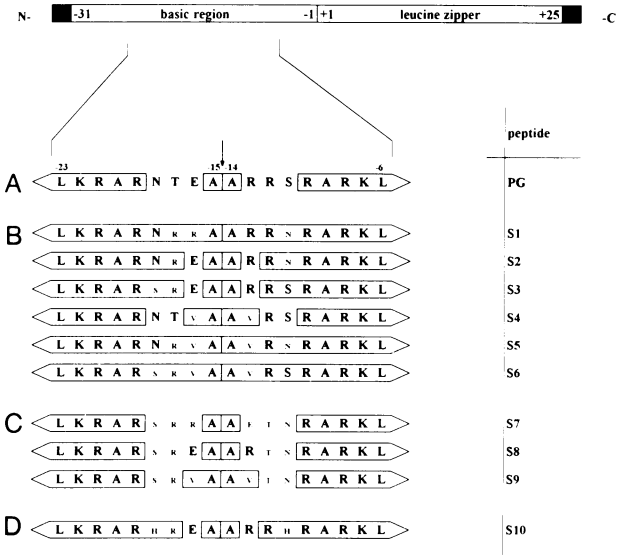

Fig. 2. The protein constructs. All peptides are derived from the GCN4 bZip peptide PG (Suckow *et al.*, 1993a; schematically shown on the top) and differ only in the region from residues –18 to –11. The black boxes at the N- and the C-terminus symbolize the two N-terminal and the four C-terminal residues from the Lac repressor which flank the various bZip domains in order to protect them against degradation by *Escherichia coli* proteases (Suckow *et al.*, 1993a). Residues –23 to –6 of the wild-type GCN4 peptide PG (Suckow *et al.*, 1993a) and of the various mutant peptides are shown in one-letter code. Large capital letters indicate that the respective residue in this position is equivalent to the wild-type amino acid, small capital letters symbolize amino acid exchanges, as compared with wild-type. The center of symmetry of the various palindromes of the peptide sequences is between Ala –15 and Ala –14, as indicated by an arrow on the top of the sequence of peptide PG and by a vertical line within each sequence; residues which are identical in symmetric positions of the two halves of the respective palindromic peptide sequences are boxed. (**A**) The wild-type GCN4 peptide PG (see also Suckow *et al.*, 1993a). (**B–D**) Mutant peptides (S1–S10) with altered palindromic peptide sequences are ordered in groups as discussed in the text.

The palindrome of the sequence of the GCN4 peptide can be extended without loss of function

The next step was to analyze what happens if the palindrome of the sequence of the GCN4 peptide (Figure 2A) is extended artificially by amino acid substitutions. We introduced mutations into the GCN4 basic region and characterized the DNA binding specificities of the various peptides (see Figure 2) by measuring their relative affinities for the ATF/CREB site and its 15 single exchange variants in positions 0–4 (see Figure 3A). Since crude bacterial cell lysates were used for the gel retardation assays (see also Suckow *et al.*, 1994a), we analyzed all extracts on Tricine–SDS protein gels (Schägger and von Jagow, 1987) to determine the amounts of the respective bZip peptide they contained. All peptide bands were clearly visible and their intensities differed only moderately (data not shown). As shown and discussed previously (Suckow *et al.*, 1993b, 1994a), the wild-type GCN4 peptide PG (Figure 2A) not only binds very well to the ATF/CREB site (++++), but also binds with reduced relative affinities to the symmetric variants with guanine 0 (++) or thymine 0

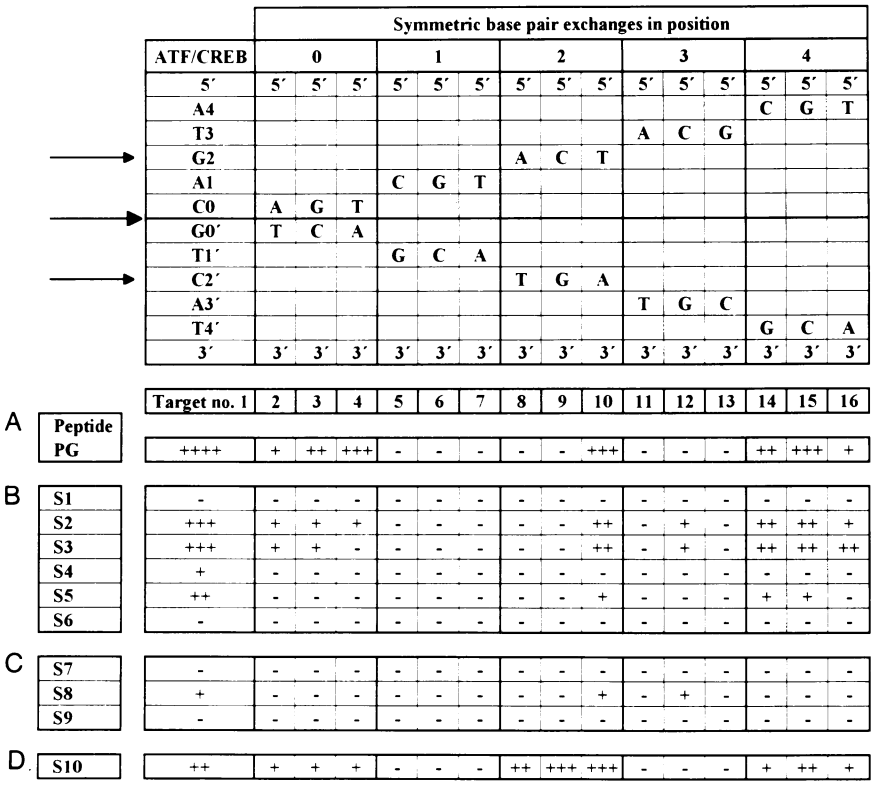

Fig. 3. Summary of the results of the gel retardation assays shown in Figure 4. Upper right panel: the ATF/CREB sequence and its 15 symmetric single exchange variants in positions 0–4. For the variants, only those bases are printed which differ from the ATF/CREB site. The large arrow on the left points to the center of symmetry of the ATF/CREB site, the two small arrows to the center of the respective target half-site. The numbers below the binding sites correspond to the numbers shown in Figure 4. Lower left panel: the protein constructs are the same as shown in Figure 2. Lower right panel: the relative affinities of the various peptides for the respective binding sites were estimated from the intensities of bands corresponding to complexed and free double-stranded DNA (see Figure 4). ++++: very high (>90% shifted); +++: high (50–90% shifted); ++: moderate (15–50% shifted); +: low (up to 15% shifted); –: very low (<1%) or no complex formation detectable.

(+++), thymine 2 (+++), and cytosine 4 (++), guanine 4 (+++), or thymine 4 (+; Figures 3A and 4). The weak complex with the adenine 0 variant, which is also visible in Figure 3A, was not detected in previous studies where the retardation gels were analyzed from autoradiographs (e.g. Suckow *et al.*, 1994b).

In a first attempt to generate a mutant GCN4 peptide with a perfect palindrome in the amino acid sequence, we replaced Thr –17, Glu –16 and Ser –11 with Arg –17, Arg –16 and Asn –11, respectively, resulting in peptide S1 (Figure 2B). However, peptide S1 does not bind to any of the targets tested, as shown in Figure 3B. According to the crystal structures of GCN4–DNA complexes (Ellenberger *et al.*, 1992; König and Richmond, 1993), Arg –13 of the GCN4 basic region not only contacts the DNA backbone (see Figure 1B), but also forms an intramolecular salt bridge with Glu –16 which was proposed to be essential for stable monomer conformation (Ellenberger *et al.*, 1992). In the case of peptide S1, which contains Arg –16 instead of Glu, this intrahelical salt bridge cannot be formed.

Therefore, we replaced Arg –16 of peptide S1 with Glu, resulting in peptide S2 which, except for Glu –16/Arg –13, contains a palindrome (Figure 2B). Compared with the wild-type GCN4 peptide PG (Figure 2A), the basic region of peptide S2 contains two further pairs of identical, corresponding amino acids: Asn –18 of the left half has its symmetric counterpart in Asn –11 on the right half, and Arg –12 of the right half has its counterpart in Arg –17 of the left half (Figure 2B). As shown in Figure 3B, peptide S2 indeed binds specifically to DNA: it recognizes all targets which are bound by the wild-type GCN4 peptide PG with similar relative affinities and, in addition, it binds weakly to the symmetric variant of the ATF/CREB sequence with cytosine 3 (+; Figures 3B and 4) which is not bound by peptide PG (Figures 3A and 4).

The next mutant which we tested, peptide S3, like peptide S2 contains a palindrome in its peptide sequence which, except for Glu –16/Arg –13, is perfectly symmetric. In peptide S3, Ser –11 of the right half of the palindrome is projected to position –18 of the left half, in contrast to peptide S2 in which Asn –18 of the left half is projected to position –11 of the right half (Figure 2B). Peptide S3 also binds specifically to DNA: it behaves very similarly to peptide S2, but fails to recognize the symmetric variant of the ATF/CREB sequence with thymine 0 (Figures 3B and 4).

The finding that the mutant peptides S2 and S3 bind specifically to DNA is surprising, since previous studies revealed that single mutants of the GCN4 bZip peptide PG with Asn –11 or Ser –18 are inactive (Suckow *et al.*, 1994a,b). However, these single mutants differ from peptides S2 and S3, respectively, in position –17 of their basic regions: they contain Thr –17 instead of Arg. Obviously, the presence of Arg –17, which naturally occurs in the basic regions of several bZip proteins, helps

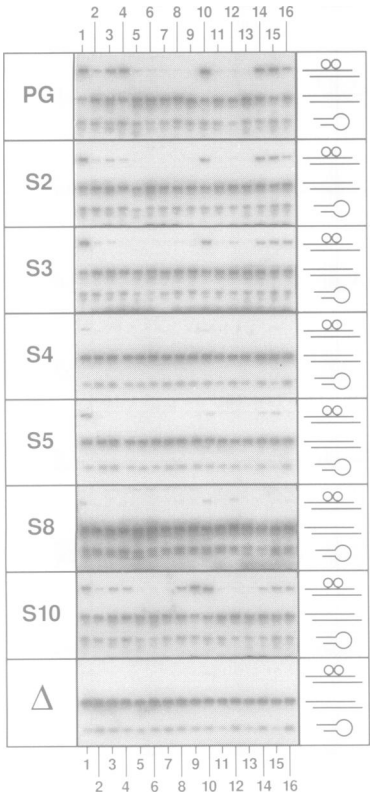

Fig. 4. PhosporImager prints of mobility shift assays with cell extracts containing the wild-type GCN4 peptide PG or mutant peptides, as indicated on the outside of each block. Only those peptides are shown which bind to at least one of the targets tested. The numbers above and below the slots indicate the respective binding site; they correspond to the numbers shown in Figure 3 between the upper and the lower right panel. The symbols between the autoradiographs indicate hairpin, bZip-complexed or free double-stranded DNA. The results are summarized in Figure 3. The Δ extract contains a truncated peptide terminating in position −14 of the GCN4 basic region. With this extract, a faint band becomes visible with target No. 10, which is caused by an *E.coli* protein. This complex moves slightly faster than the bZip complexes (data not shown).

peptides S2 and S3 by compensating for the inactivating effects of Asn −11 or Ser −18, possibly by a different positioning of the recognition helix in the major groove. Arg −17 of the basic region of jun establishes a backbone contact with the 5′ phosphate of guanine 0′, according to the crystal structure of a jun–fos heterodimer bound to the pseudo-palindromic AP1 site (Glover and Harrison, 1995); in contrast, Thr −17 of the GCN4 basic region interacts with the sugar of guanine 0′ and the 5′ phosphate of thymine 1′, according to the crystal structures of GCN4–DNA complexes (Ellenberger *et al.*, 1992; König and Richmond, 1993; see Figure 1B).

The finding that peptide S1 does not bind to DNA while peptide S2 does (Figure 3B; see above) underlines the importance of the intramolecular salt bridge between Glu −16 and Arg −13 for the structural stability of the bZip monomer (Ellenberger *et al.*, 1992). We wondered whether this salt bridge could be replaced with a hydrophobic interaction, and thus introduced the medium sized hydrophobic amino acid valine into positions −16 and −13 of the basic region of peptide PG. The resulting peptide S4 not only has an improved palindrome in its peptide sequence in comparison with the wild-type GCN4 peptide PG (Figure 2A and B), it also binds specifically to DNA: from the 16 targets tested, it still binds weakly to the ATF/CREB site (+; Figures 3B and 4). Compared with the complexes formed by the wild-type GCN4 peptide PG, the spectrum of DNA targets recognized by peptide S4 is smaller, and the relative affinity for the ATF/CREB site is reduced (Figure 3A and B). However, these findings indicate that: (i) an intramolecular hydrophobic interaction between two valines in positions −16 and −13 of the GCN4 basic region can indeed functionally replace the electrostatic interaction between Glu −16 and Arg −13 and (ii) the backbone contact performed by Arg −13 (Ellenberger *et al.*, 1992; König and Richmond, 1993; see Figure 1B) is not essential for DNA binding. We now introduced valine into positions −16 and −13 of the basic regions of peptides S2 and S3. The resulting peptides S5 and S6 have perfect amino acid palindromes (Figure 2B). While peptide S6 does not bind to any of the targets tested, peptide S5 binds moderately to the ATF/CREB site (++) and weakly to the variants with thymine 2, and cytosine or guanine 4 (+/+/+; Figures 3B and 4). Thus, it behaves similarly to the corresponding peptide S2 with Glu −16/Arg −13 (Figures 3B and 4).

As shown in the preceding section, the protein–DNA interface between wild-type GCN4 and the ATF/CREB site (König and Richmond, 1993) includes several interactions which are approximately symmetric within a half-complex with regard to an axis of symmetry through base pair 2/2′ and to the center of the degenerate palindromic peptide sequence (Figure 1A and B). Peptides S2 and S3, as shown above, contain palindromic sequences which are perfect except for Glu −16 and Arg −13 (Figure 2B) and still form complexes with the symmetric variants of the ATF/CREB site with guanine in position 4 or thymine in position 0 (Figures 3B and 4) whose half-sites can be regarded as short pseudo-palindromes with central guanine 2/cytosine 2′ base pairs. We supposed that in these cases the amount of corresponding symmetric interactions may be increased further, in comparison with the GCN4–ATF/CREB site half-complexes, and performed computer modeling studies of these complexes, based on the coordinates of the crystal structure of the symmetric GCN4–DNA complex (König and Richmond, 1993). To our surprise, it appeared to be impossible to position the basic region of peptide S2 in the major groove of a target half-site such that Asn −11 and Asn −18 perform symmetric base contacts with regard to base pair 2/2′ without disrupting the coiled-coil structure of the bZip dimer. Asparagine in position −11 seemed much more inclined to establish a contact with the 5′ phosphate of thymine 3, which clearly would not be a symmetric counterpart for the interactions of Asn −18 with thymine 3 and cytosine 2′ (data not shown). Similarly, we failed to arrange the basic region of peptide S3 in the major groove of the target half-site such that the two serines in positions −18 and −11 can act symmetrically with regard to base pair 2/2′ (data not shown). However, Arg −17 of both peptide S2 and S3 is easily put in a position adapted from Arg −17 of jun in the jun/fos–DNA complex (Glover and Harrison, 1995) such that it reaches the 5′ phosphate of the respective base in position 0′ of the half-site (data not shown) and thus may act symmetrically to Arg −12 with regard to base pair 2/2′ (see Figure 1B). In principle, these findings

are consistent with the observed pattern of target sites which are recognized by peptides S2 and S3: if the protein–DNA interfaces within the half-complexes between these peptides and the respective binding sites could indeed be subdivided further into two quasi-identical symmetric halves, one would expect concerted binding or non-binding to pairs of variants of the ATF/CREB site with complementary exchanges in positions 3/1 or position 2. This is not the case (Figures 3B and 4): both peptides S2 and S3 bind weakly to the cytosine 3 variant, but not to the guanine 1 variant, and both peptides bind to the ATF/CREB site (with guanine 2) and to the variant with thymine 2, but not to the variants with cytosine or adenine 2 (Figures 3B and 4).

Peptides with inverted palindromic sequences

In a second approach to probe the degree of symmetry within the interfaces between the GCN4 basic region and the respective target half-sites, we constructed a series of mutant peptides with inverted degenerate palindromic sequences (Figure 2C). If there are sufficient symmetric interactions in the two halves of a half-complex, they should be interchangeable. In the first attempt, we inverted the natural GCN4 sequence from position –23 to position –6 completely; the resulting peptide S7 (Figure 2C) does not bind to any of the DNA targets tested (Figures 3C and 4). However, weak DNA binding activity is observed if Glu –16 and Arg –13 are excluded from the inversion of the palindromic peptide sequence: peptide S8 (Figure 2C) forms weak but detectable specific complexes with the ATF/CREB site and the variants with thymine 2 or cytosine 3 (+/+/+; Figures 3C and 4). If Glu –16 and Arg –13 of peptide S8 are replaced with valine (peptide S9; Figure 2C), no binding to any of the targets tested is detectable (Figure 3C).

Peptides S7 and S8 merely differ in the orientation of residues –16 and –13; the finding that peptide S7 is inactive while peptide S8 still binds weakly to DNA indicates that the intrahelical salt bridge between Glu –16 and Arg –13 cannot be inverted without loss of DNA binding function. Furthermore, the DNA binding activity of peptide S8 may indicate some symmetry within the interface between a GCN4 basic region and a target half-site. However, the weakness of the complexes formed by this peptide may underline our conclusion from the preceding section that, with some exceptions, a considerable number of the interactions within the GCN4–DNA half-complex (see Figure 1) are qualitatively, but not quantitatively, symmetric in the 3-D structure.

Construction of a mutant bZip half-complex with quasi-symmetric protein–DNA interactions

We then wondered whether it is possible to generate bZip–DNA complexes with more symmetric interactions within a half-complex by choosing a suitable composition of amino acids. In previous work (Suckow *et al.*, 1994a,b), we described two single mutants of a GCN4 bZip peptide with altered DNA binding specificities. (i) If Asn –18 is replaced with His, the resulting peptide does not bind to the ATF/CREB site but to the symmetric variant with cytosine 2; we explained this change of specificity by proposing the formation of H bonds between the nitrogens of His –18 and N6 of adenine 3′ and O6 of guanine 2′ of this particular binding site (Suckow *et al.*, 1994b). (ii) If Ser –11 of the GCN4 peptide is replaced with His, the DNA binding specificity is not changed but broadened: the resulting peptide binds to all targets bound by GCN4 and, in addition, weakly to the cytosine 2 variant. In this case, we explained the broadened specificity, on the basis of the GCN4–AP1 site complex (Ellenberger *et al.*, 1992), with hydrophobic interactions between His –11 and C5/C6 of cytosine 2 and the methyl group of thymine 3 (Suckow *et al.* 1994a). Thus, His –11 and His –18 of the respective mutants are both involved in interactions with base pair 2/2′ which, according to our considerations, may be regarded as the center of a half-site. Since, in addition, residues –18 and –11 of the basic region occupy corresponding positions within the amino acid palindrome (see Figure 1), His appeared to be best suited for creating a GCN4 mutant capable of performing quasi-symmetric protein–DNA interactions within a half-complex.

Peptide S10 (with His –18, Arg –17 and His –11; Figure 2D) not only contains a palindromic sequence which, except for Glu –16 and Arg –13, shows perfect symmetry, but, in addition, it has a remarkable pattern of recognized target variants: it binds weakly as well to the ATF/CREB site as to all symmetric variants in positions 0, 2 and 4, and it fails to bind to any of the variants in position 1 or 3 (Figures 3D and 4). Since positions 0 and 1 correspond to positions 4 and 3, respectively, in contrast to all other peptides shown so far (Figures 3 and 4), peptide S10 produces a symmetric pattern of bound and unbound binding sites, with regard to base pair 2/2′, which marks the center of the half-site. In the complexes between the His –11 and His –18 single mutants and the variant of the ATF/CREB site with cytosine 2, the respective His interacts with base pair 2/2′ (Suckow *et al.*, 1994a,b). Therefore, the most striking feature of peptide S10 is its completely relaxed DNA binding specificity for position 2 of the ATF/CREB site, which might be explained by simultaneous contacts between both His –18 and His –11 and base pair 2/2′. This, however, would be equivalent to an increase of symmetric interactions within the protein–DNA interface of a half-complex.

In order to obtain some insight into some of these protein–DNA complexes at the atomic level, we modeled the complexes between peptide S10 and either the ATF/CREB site (with guanine 2) or the variants with adenine, cytosine or thymine in position 2 (Figures 3D and 4) on the basis of the coordinates of the GCN4–ATF/CREB site complex (König and Richmond, 1993). The results are shown schematically in Figure 4: in the cases of the ATF/CREB site (with guanine 2) and the thymine 2 variant, Nδ of His –18 may establish a H bond with N4 of cytosine or N6 of adenine 2′, respectively, while Nε of His –11 might interact with N7 of guanine or O4 of thymine 2, respectively (Figure 5A and B). On the other hand, in the complexes between peptide S10 and the variants with adenine or cytosine 2, the rings of the two His may be turned around: here Nε of His –18 might interact with O4 of thymine or O6 of guanine 2′, and Nδ of His –11 with N6 of adenine or N4 of cytosine 2 (Figure 5C and D).

These models not only explain why peptide S10 binds as well to the ATF/CREB site as to all position 2 variants (Figures 3D and 4) but, in addition, suggest indeed a significant increase of symmetric interactions within a

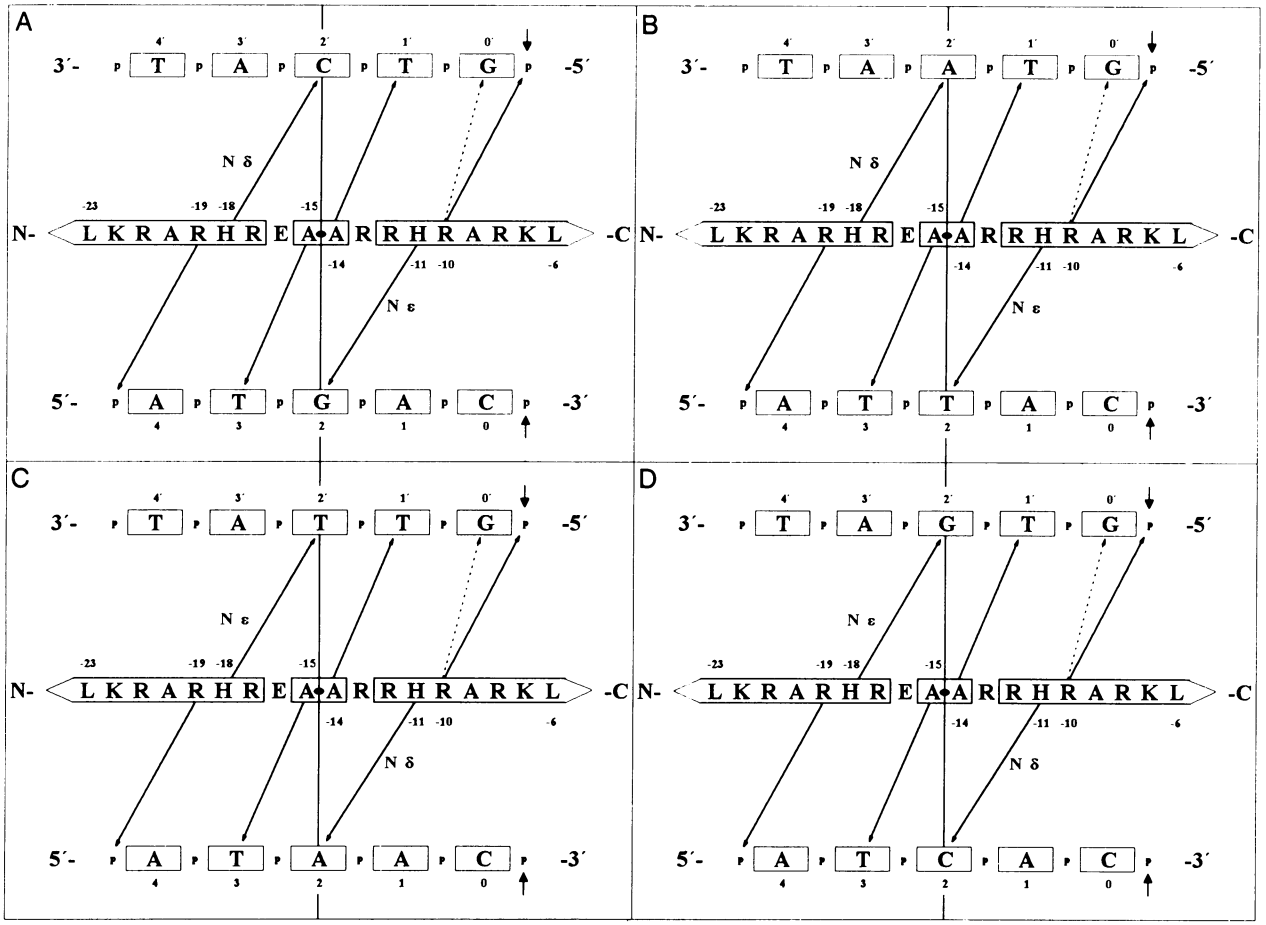

Fig. 5. 2-D projections of the modeled half-complexes between peptide S10 (see Figure 2) and the ATF/CREB site (**A**) and its three variants of position 2 (**B–D**), according to our modeling studies. Thick arrows symbolize interactions which are qualitatively symmetric in the two halves of a half-complex, with regard to an axis through the center of the palindromic amino acid sequence (between Ala −15 and Ala −14) and through base pair 2/2′ (vertical line). Dotted arrows indicate interactions without symmetric counterparts in the respective other half of the half-complex. Except for the interactions between Arg −19 and the 5′ phosphate of adenine 4 and between Arg −10 and the 5′ phosphate of guanine 0′ on the opposite strand, only the putative amino acid–base interactions are indicated. In all cases, Arg −19, Ala −15, Ala −14 and Arg −10 interact as in the GCN4–ATF/CREB site complex (König and Richmond, 1993; see Figure 1). (**A**) and (**B**) Nδ of His −18 interacts with N4 of cytosine 2′ (A) or N6 of adenine 2′ (B), while Nε of His −11 interacts with N7 of guanine 2 (A) or O4 of thymine 2 (B), respectively. (**C**) and (**D**) The rings of His are turned around in comparison with (A) and (B): Nε of His −18 interacts with O4 of thymine 2′ (C) or O6 of guanine 2′ (D), while Nδ of His −11 interacts with N6 of adenine 2 (C) or N4 of cytosine 2 (D), respectively.

half-complex, in comparison with the GCN4–ATF/CREB site complex (König and Richmond, 1993). With regard to base pair 2/2′ and to the center of the palindromic peptide sequence, the interaction between Ala −15 and thymine 3 would be approximately symmetric to the interaction between Ala −14 and thymine 1′, as shown already for the wild-type complex (see Figure 1A). Furthermore, the interactions between His −18 and the respective base in position 2′ would be approximately symmetric to the interactions of His −11 and the respective base in position 2, as shown in Figure 5. Thus, except for the interactions performed by Arg −19 and Arg −10 which are merely partially symmetric in the two halves of a target half-site (see Figures 1A and B, and 5), all amino acid–base interactions appear to have increased in symmetry in all four half-complexes (see Figure 5A–D).

Discussion

The width of the B-DNA major groove is 1.2 nm and is thus complementary to the diameter of a typical α-helix, including side chains. The perfect correlation between the sizes of these structures may have been the most important constraint for the evolution of various DNA binding domains which use α-helices as DNA-recognizing elements. Several classes of DNA binding proteins have been described whose optimal binding sites are palindromic DNA sequences. Generally, these proteins form rotationally symmetric homodimers when complexed with DNA, and thus the two DNA binding domains are also arranged in a rotationally symmetric manner. Consequently, the overall protein–DNA complex is rotationally symmetric too, with regard to an axis pointing through the centers of symmetry of the binding site and of the protein dimer. Formation of such complexes necessitates two symmetric components: a palindromic DNA sequence is *per se* symmetric, while a protein monomer is *per se* asymmetric. Thus, the symmetric DNA binding protein structure is generated by the assembly of two protein monomers into a rotationally symmetric dimer. The subject of this study is the question of whether a single α-helix

with a palindromic sequence, which at the atomic level is strictly speaking asymmetric, can interact with DNA in a quasi-symmetric manner.

As shown in the Results section, there are hints that the complex between GCN4 and the ATF/CREB sequence (König and Richmond, 1993) may represent an evolutionary attempt to use such a single α-helix with an almost symmetric peptide sequence for the specific recognition of a short DNA pseudo-palindrome. (i) The area of the recognition helix is almost identical to the area of the degenerate palindrome in the peptide sequence (see Figure 1). (ii) The half-sites of some variants of the ATF/CREB site bound by GCN4 (those with thymine 0 or guanine 4; see Figure 3, upper panel) can be regarded as short pseudo-palindromes. (iii) The interface between a GCN4 basic region and a target half-site (König and Richmond, 1993) contains a considerable amount of qualitatively symmetric interactions, with regard to the center of the half-site and to the center of the palindrome of the peptide sequence (see Figure 1). (iv) By reflection of wild-type amino acids from the left half of the degenerate palindromic amino acid sequence to the right half, or vice versa, the sequence from position –23 to –6 of the GCN4 basic region can be converted into a nearly perfect (peptides S3 and S4 with Glu –16/Arg –13; Figure 2B) or perfect palindrome (peptide S5 with Val –16/Val –13; Figure 2B). However, although these peptides still bind specifically to DNA (Figures 3B and 4), our modeling studies did not indicate an increase in qualitatively symmetric interactions within the respective half-complexes as compared with the GCN4–ATF/CREB site complex (König and Richmond, 1993). (v) Except for Glu –16 and Arg –13, the degenerate palindromic amino acid sequence can be inverted without complete loss of function (peptide S8; Figures 3C and 4). (vi) The number of residues which fit into the palindromic peptide sequence, and at the same time the number of symmetric interactions within the protein–DNA interface of each half-complex, can be increased by insertion of His into positions –18 and –11 of the basic region, as indicated by the DNA binding properties of peptide S10 (Figures 3D and 4). According to our modeling studies, peptide S10 may well use its –18 and –11 His in a quasi-symmetric manner to bind equally well to any base pair in position 2/2′ of the ATF/CREB site (Figure 5).

The asymmetry of an α-helix with a palindromic peptide sequence is mainly caused by: (i) its dipole character; and (ii) the direction of the Cα–Cβ bonds, which are angled towards the N-terminus. The symmetric arrangement of amino acid–DNA interactions which we describe in the Results section for the interfaces between the GCN4 or some mutant recognition helices and their target half-sites is only part of the structure of the respective complete bZip–DNA complexes. In the bZip dimer, the coiled-coil structure formed by the leucine zippers (O'Shea *et al.*, 1989; Oas *et al.*, 1990) dictates a certain divergence angle between the two basic regions suitable for enabling them to enter the major groove of two adjacent target half-sites (Vinson *et al.*, 1989; O'Neill *et al.*, 1990). This fixed angle may impose one of the limitations which prevent an ideally symmetric positioning of the basic region over its target half-site: the C-terminal part of the recognition helix (around residue –10) is forced into a closer proximity to the major groove than the N-terminal part (around residue –18; Ellenberger *et al.*, 1992; König and Richmond, 1993; Glover and Harrison, 1995). In addition, the conserved helical repeat phase between leucine zipper and recognition helix (Vinson *et al.*, 1989) hinders an ideally symmetric positioning of the basic region over the center of a target half-site. In fact, according to the crystal structures of GCN4–DNA complexes (Ellenberger *et al.*, 1992; König and Richmond, 1993) and the jun/fos–DNA complex (Glover and Harrison, 1995), the recognition helices appear to be slightly turned away from this ideally symmetric positioning, clockwise, when viewed along their long axes from the C- to the N-terminus. Other important contributors to the asymmetry observed are (i) the interaction of Asn –18 with the center of the half-site, which with 5 bp is short and has an odd number of base pairs; and (ii) the symmetrically positioned pair Glu –16 / Arg –13. Both are important for stabilizing the bZip monomer, but only Arg –13 can establish backbone contacts, whereas Glu –16 must be repelled (Ellenberger *et al.*, 1993; König and Richmond, 1994). Thus, the GCN4 basic region seems to be positioned just as symmetrically in the major groove of a target half-site as is allowed by the governing architecture of the bZip–DNA complex.

An interesting aspect of DNA binding α-helices carrying a palindromic sequence is the question of their genesis. On the one hand, we can hardly imagine that the GCN4 degenerate peptide palindrome is a product of chance. (i) Starting from a sequence of 18 amino acids, the probability that six of nine residues in one half become symmetric with respect to the other half is very small. (ii) Alignment of the basic regions of 65 known bZip proteins with GCN4 shows three members with five, seven with four, seven with three, six with two, 32 with one and four without palindromically arranged amino acids (data not shown; for a complete list of sequences, see Hurst, 1994). This distribution may indicate a tendency towards the formation of partially palindromic sequences within the class of bZip proteins. (iii) In the complex between GCN4 and the ATF/CREB site (König and Richmond, 1993), many protein–DNA interactions are qualitatively symmetric within a half-complex, which means that the palindromic character of the GCN4 basic region is indeed functionally connected with the pseudo-palindromic character of the half-site (see Figure 1). On the other hand, several other bZip proteins, for example CREB (Gonzales *et al.*, 1989), ATF2 (Maekawa *et al.*, 1989), ATF3, ATF4 (Hai and Curran, 1991) and ATF6 (Hai *et al.*, 1989), have been described which bind well to the very same ATF/CREB site, but whose amino acid sequences are either not palindromic, or are so to a lesser extent than the sequence of the basic region of GCN4. Whatever may have been the evolutionary pressure which favored the GCN4 degenerate palindromic peptide sequence is not obvious. One, speculative, possibility might have been the following: in the basic region of GCN4, Lys –22 is the most N-terminally positioned basic residue; in the basic regions of many other bZip proteins, basic residues occur up to position –28, or even further (for a complete list of sequences, see Hurst, 1994). We could imagine that the complexes between these proteins and their DNA targets, unlike the GCN4–ATF/CREB site complex (König and Richmond, 1993), might be stabilized further by contacts between the additional basic residues beyond position

–22 and the phosphate backbone of the respective binding sites. These hypothetical, additional backbone contacts would be absent in the GCN4–DNA complexes; instead, the contacts between residues from position –22 to –7 of the basic region and the DNA target (see Figure 1) may have been evolutionarily optimized, with the degenerate 'peptide palindrome' as a result.

Materials and methods

Bacterial strains and plasmids

Escherichia coli K12ΔH1 (*SmR lacZam Δbio-uvrB ΔtrpEA2 λNam7–Nam53 cI857 ΔH1*) was kindly supplied by W.Fiers. Plasmid pPLc28-bZip has been described (Suckow *et al.*, 1993a).

Chemicals and enzymes

Restriction enzymes were obtained from Boehringer Mannheim (Mannheim, Germany) and New England Biolabs (Bad Schwalbach, Germany). DNA polymerase large fragment was from Boehringer-Mannheim; [α-^{32}P]deoxyribonucleotides from Amersham Buchler (Braunschweig, Germany); chemicals for automated DNA synthesis from Applied Biosystems (Pfungstadt, Germany); DNA sequencing kits from Pharmacia (Freiburg, Germany); plasmid preparation kits from Promega (Heidelberg, Germany) or Qiagen (Hilden, Germany) and all other chemicals were from Sigma (München, Germany) or Merck (Darmstadt, Germany).

Methods

Oligonucleotides were synthesized on an Applied Biosystems 380A synthesizer and purified on denaturating polyacrylamide gels prior to use. Expression vectors coding for mutant GCN4 bZip peptides were generated by replacement of the *Bss*HII–*Hin*dIII fragment within the GCN4 bZip gene of plasmid pPLc28-bZip (Suckow *et al.*, 1993a) with the respective synthetically produced DNA double strands. Cell extracts containing the various bZip peptides were prepared as described (Suckow *et al.*, 1993b) and used without further purification for gel retardation assays (Suckow *et al.*, 1994b) after 1:3 dilution in extraction buffer (200 mM Tris–HCl, pH 8; 10 mM $MgCl_2$; 1 mM EDTA; 7 mM β-mercaptoethanol; 30% glycerol). The DNA fragments containing the various binding sites, with total lengths of 28 bp, were the same as shown in Suckow *et al.* (1994b). The relative affinities of the peptides for the various targets tested were derived from PhosphorImager analyses of the respective retardation gels. All cell extracts were analyzed for the expression of the respective bZip peptide on Tricine–SDS protein gels, as described by Schägger and von Jagow (1987).

Computer modeling

Modeling was performed on an ESV30 workstation with Insight II and Discover (Biosym Inc., San Diego, CA), on the basis of the coordinates of the GCN4–ATF/CREB site complex (König and Richmond, 1993).

Acknowledgements

We thank R.Hühne for help with the PhosphorImager, K.Otto and G.Zimmer for the synthesis of the oligonucleotides used, U.Ringeisen for excellent photography and A.Barker for critical reading of the manuscript. Furthermore, we thank M.Suzuki for stimulating discussions. This work was supported by the DFG through SFB 274, SFB 243 and Schwerpunkt 'Protein Design'.

References

Aggarwal,A.K., Rodgers,D.W., Drottar,M., Ptashne,M. and Harrison,S. (1988) *Science*, **242**, 899–907.
Agre,P., Johnson,P.F. and McKnight,S.L. (1989) *Science*, **246**, 922–926.
Ellenberger,T.E., Brandle,C.J., Struhl,K. and Harrison,S.C. (1992) *Cell*, **71**, 1223–1237.
Ferre-D'Amare,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K (1993) *Nature*, **363**, 38–45.
Glover,J.N.M. and Harrison,S.C. (1995) *Nature*, **373**, 257–261.
Gonzales,G.A., Yamamoto,K.K., Fischer,W.H., Karr,D., Menzel,P., Biggs,W., Vale,V.V. and Montminy,M.R. (1989) *Nature*, **337**, 749–752.
Hai,T. and Curran,T. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3720–3724.
Hai,T., Liu,F., Coukos,W.J. and Green,M.R. (1989) *Genes Dev.*, **3**, 2083–2090.
Harrison,S.C. (1991) *Nature*, **353**, 715–719.
Hill,D.E., Hope,I.A., Macke,J.P. and Struhl,K. (1986) *Science*, **234**, 451–457.
HinnebuschA.G. and Fink,G.R. (1983) *J. Biol. Chem.*, **258**, 5238–5247.
Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.
Hurst,H.C. (1994) *Protein profile*, **1**, 123–168.
König,P. and Richmond, T.J. (1993) *J. Mol. Biol.*, **233**, 139–154.
Kouzarides,T. and Ziff,E. (1989) *Nature*, **340**, 568–571.
Luisi,B.F., Xu,W.X., Otwinowski,Z., Freedman,L.P., Yamamoto,K.R. and Sigler,P.B. (1991) *Nature*, **352**, 497–505.
Maekawa,T., Sakura,H., Kanei Ishi,C., Sudo,T., Yoshimura,T., Fujisawa,J., Yoshida,M. and Ishii,S. (1989) *EMBO J.*, **8**, 2023–2028.
Mondragon,A., Wolberger,C. and Harrison,S. (1989) *J. Mol. Biol.*, **205**, 179–188.
Oas,T.G., McIntosh,L.P., O'Shea,E.K., Dahlquist,F.W. and Kim,P.S. (1990) *Biochemistry*, **29**, 2891–2894.
O'Neill,K.T., Hoess,R.H. and DeGrado,W.F. (1990) *Science*, **249**, 774–778.
O'Shea,E.K., Rutkowski,R. and Kim,P.S. (1989) *Science*, **243**, 538–542.
Pan,T. and Coleman,J.E. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 2077–2081.
Pauling,L., Corey,R.B. and Branson,H.R. (1951) *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
Schägger,H. and von Jagow,G. (1987) *Anal. Biochem.*, **166**, 368–379.
Schuhmacher,M.A., Choi,K.Y., Zalkin,H. and Brennan,R.G. (1994) *Science*, **266**, 763–770.
Sellers,J.W., Vincent,A.C. and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 5077–5086.
Suckow,M., von Wilcken-Bergmann,B. and Müller-Hill,B. (1993a) *EMBO J.*, **12**, 1193–1200.
Suckow,M., von Wilcken-Bergmann,B. and Müller-Hill,B. (1993b) *Nucleic Acids. Res.*, **21**, 2081–2086.
Suckow,M., Madan,A., Kisters-Woike,B., von Wilcken-Bergmann,B. and Müller-Hill,B. (1994a) *Nucleic Acids. Res.*, **22**, 2198–2208.
Suckow,M., Schwamborn,K., Kisters-Woike,B., von Wilcken-Bergmann,B. and Müller-Hill,B. (1994b) *Nucleic. Acids. Res.*, **22**, 4395–4404.
Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) *Science*, **246**, 911–916.
Watson,J.D. and Crick,F.H.C. (1953) *Nature*, **171**, 737–738.

Received on June 19, 1995; revised on October 3, 1995