The EMBO Journal vol.12 no.3 pp.1193-1200, 1993

Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding

Manfred Suckow, Brigitte von Wilcken-Bergmann and Benno Müller-Hill

Institut für Genetik der Universität zu Köln, Weyertal 121, D-5000 Köln 41, Germany

Communicated by B.Müller-Hill

The bZIP regions of the eukaryotic transcription factors GCN4 and C/EBP have similar protein sequences but they recognize different DNA sequences. In order to understand their specificity, a vector was constructed which permits overexpression in *Escherichia coli* of those domains of GCN4 that are necessary and sufficient for specific DNA binding i.e. the basic region and the leucine zipper. Specific DNA binding was monitored with gel shift experiments. The residues of the basic region of GCN4 were systematically replaced by those of C/EBP to transform GCN4 into C/EBP with respect to DNA binding. Residues −17, −16 and −14 were found to be responsible for switching GCN4 to C/EBP binding specificity (we define as residue +1 the first leucine of the first leucine heptad repeat of GCN4). We broadened the specificity of GCN4 to TAF-1 by replacing residues −15 and −17 and we changed the specificity of C/EBP to TAF-1 by swapping residue −17 of a particular hybrid. Thus residues positioned from −14 to −17 of the basic region play a key role in recognizing specific DNA sequences.

Key words: basic region/bZIP proteins GCN4, C/EBP and TAF-1/change of DNA binding specificity/protein−DNA recognition

Introduction

The eukaryotic transcription factor GCN4 (Hope and Struhl, 1986; Oliphant *et al.*, 1989) is a member of the large family of eukaryotic bZIP proteins which have similar protein sequences and bind preferentially to two palindromic DNA sequences, termed TRE (5′-ATGACTCAT-3′) and CRE (5′-ATGACGTCAT-3′). TRE and CRE can be seen to differ in the spacing of the TGA half-site elements (Pathak and Sigler, 1992; O'Neil *et al.*, 1990). Two bZIP proteins have been described which recognize different DNA sequences: the C/EBP protein (Landschulz *et al.*, 1988; Agre *et al.*, 1989) and the TAF-1 protein (Oeda *et al.*, 1991). C/EBP has been shown to bind specifically to the sequence: 5′-ATTGCGCAAT-3′ (Agre *et al.*, 1989) while the TAF-1 protein has been reported to bind *in vitro* with high affinity to the G-box motif, 5′-GCCACGTGGC-3′ (Oeda *et al.*, 1991).

Two domains are necessary for the proper binding of bZIP proteins to DNA: the basic region and the leucine zipper (Hope and Struhl, 1986). The basic region is directly involved in DNA recognition and binding whereas the leucine zipper is a device used to aggregate two monomers. For the bZIP proteins fos and jun it has been demonstrated that the target DNA is bent by the specifically binding bZIP-proteins (Kerppola and Curran, 1991a,b). Two models of bZIP protein−DNA complexes have been discussed so far: (i) the scissors grip model (Vinson *et al.*, 1989) and (ii) the induced helical fork model (O'Neil *et al.*, 1990). In both models the basic region is viewed as forming α-helices which recognize fine structural details in the major groove of DNA.

There are three ways to elucidate the structural details of protein−DNA interaction in such complexes: (i) Physical X-ray or NMR studies, (ii) biochemical studies and (iii) genetic complementation analysis. Genetic complementation analysis provides a means of analysing protein−DNA interactions in detail. It requires the demonstration that a mutant protein which has an amino acid change in the presumptive recognition domain does not interact with the wild-type target sequence and does specifically recognize a mutant form of this target—a mutant form which is not recognized by the wild-type protein.

The mutation in the DNA binding protein has thus been complemented by a change in the DNA target sequence. In reality one expects and finds not all or none effects but quantitative differences in binding strength of the various protein−DNA complexes (Figure 1). This type of analysis cannot differentiate between direct effects through contacts to bases and indirect effects through contacts to the phosphate backbone.

The analysis of the *lac* repressor−operator interaction may serve as an example of how such complementation provides insights into details of the protein−DNA interaction (Lehming *et al.*, 1990, 1991). Here the genetic analysis confirmed and extended a NMR analysis of the protein−DNA complex (Boelens *et al.*, 1988). As demonstrated in the *lac* case, three strategies of such a genetic analysis are possible: (i) testing of physical models by synthesizing specific pairs of mutant proteins and target DNAs (Kisters-Woike *et al.*, 1991), (ii) selection of complementing mutants from libraries expressing various protein variants (Sartorius *et al.*, 1989, 1991) and (iii) reduction of differences in sequence of homologous proteins and DNA targets to the essential differences by synthesizing and testing systematically various partially homologous proteins and DNA targets (Lehming *et al.*, 1987). The first two strategies have already been used in the GCN4 case (Pu and Struhl, 1991; Tsamarias *et al.*, 1992). We have used the third strategy to identify three residues of the basic region which determine differences between the specificities of GCN4, C/EBP and TAF-1. After our experiments had been completed, the X-ray structure of a GCN4 protein−DNA complex was solved (Ellenberger *et al.*, 1992). This structure was kindly made available to us prior to publication by T.E.Ellenberger. Its knowledge allowed us to differentiate between putative direct base contacts and indirect backbone contacts.

Results

The test system

We synthesized and cloned DNA coding for residues 222–277 of GCN4 which form the basic region and the leucine zipper, and are sufficient for specific DNA binding (Hope and Struhl, 1986, 1987). In order to protect the peptide against N- or C-terminal proteolytic degradation in *E.coli*, we embedded it between the two N-terminal and the four C-terminal residues of the Lac repressor (Figure 2). The synthetic gene was linked at its 3′ end to the *Sau*3A fragment of phage fd which carries an efficient transcription termination signal (Gentz *et al.*, 1981). This construct was inserted between the *Eco*RI and *Bam*HI sites of plasmid pPLc28 (Remaut *et al.*, 1981) such that a peptide of 62 amino acids was expressed in *E.coli* K12ΔH1 from λPL after heat induction (Figure 2). Mutant derivatives of the GCN4 bZIP peptide were created by replacing short segments between unique restriction sites with synthetic DNA fragments. Aliquots of crude extracts were analysed on Tricine–SDS protein gels (Schägger and von Jagow, 1987) for their contents of the GCN4 peptide or the mutant derivatives (Figure 3). All peptides are clearly visible and except the C/EBP peptide (PC) and the TAF-1 peptide (PT), which are expressed at higher amounts, the respective yields of the hybrid peptides varied by not more than a factor of five. The peptides have molecular weights of ~6.8 kDa. Note that they do not run according to their size. Their bands appear next to the 8.16 kDa band of the molecular weight marker.

We also synthesized and cloned the symmetric GCN4 consensus target with the central base pair (Hill *et al.*, 1986; Oliphant *et al.*, 1989) and a fully symmetrical target where the axis of symmetry runs between two base pairs (Figure 4) which is also known as CRE binding site and has been shown to function as GCN4 binding site *in vitro* (Sellers *et al.*, 1990). We compared the ability of crude extracts containing GCN4 peptide (PG) to bind to these sequences in electrophoretic mobility shift essays and found that both could serve almost equally well as targets for GCN4. The GCN4 peptide (PG) binds well to the sequences 5′-ATGACTCAT-3′ and 5′-ATGACGTCAT-3′, but not to 5′-ATGATCAT-3′, suggesting that the spacing between the canonical TGA blocks can be one or two base pairs. Since we wanted to compare the DNA binding behaviour of C/EBP and GCN4, we first replaced the complete basic region of GCN4 with the corresponding sequence of C/EBP (PC, Figures 4 and 5) and tested the binding of the C/EBP/GCN4 hybrid peptide to various binding sites (Figure 4). The C/EBP peptide PC is not as tolerant as GCN4 with respect to spacing: it can bind only to 5′-ATTGCGCAAT-3′, and not to 5′-ATTGCCAAT-3′. A C/EBP/GCN4 hybrid peptide (P1, Figure 4), where only residues −12 to −25 of GCN4 are replaced by those of C/EBP, shows the same C/EBP specificity, but with higher affinity (we define as residue +1 the first leucine of the first leucine heptad repeat of GCN4, see Figure 4). Thus we decided to use for further studies the fully symmetric sequence which was designated GCN4* (Figure 4). We then synthesized all variants of the GCN4* target with single symmetric base pair exchanges in positions 0–4 (Figure 5) and tested their binding to the GCN4 peptide by gel retardation. It can be seen that the binding of GCN4 is abolished by all base pair exchanges in positions 0–3 except by the G to T exchange in position 2 which is moderately well tolerated (Figures 5 and 7, lanes 5–16).

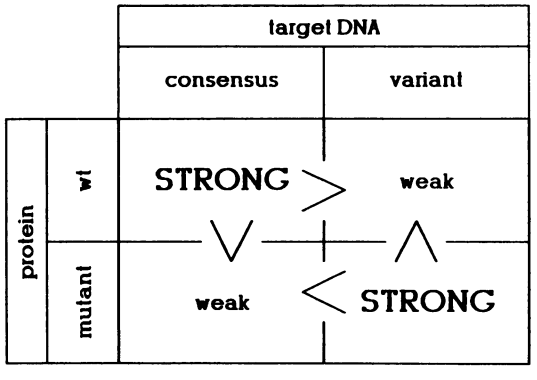

Fig. 1. The requirements for a change of specificity. (1) The mutant protein must bind to the variant target (i) better than to the consensus target and (ii) better than the wild-type (wt) protein. (2). The wild-type protein must bind to the consensus target (i) better than to the variant target and (ii) better than the mutant protein.

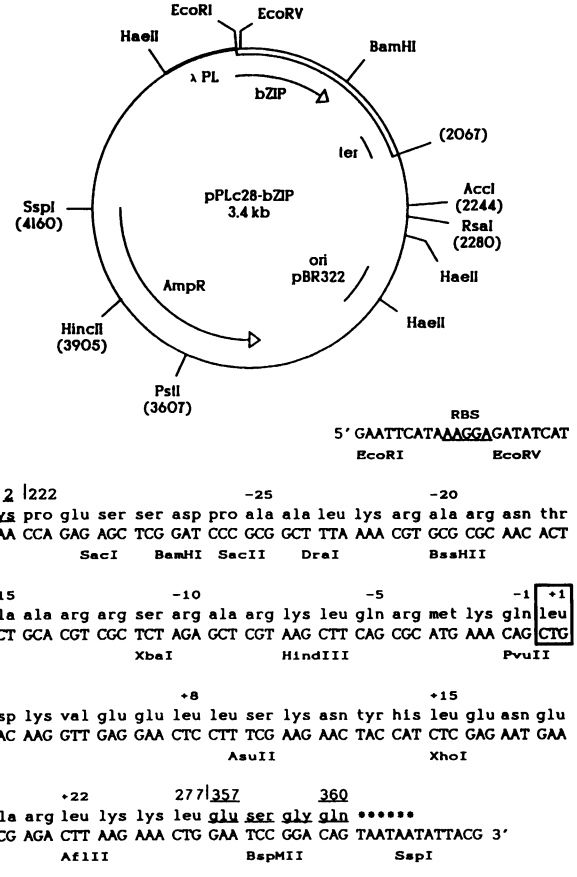

Fig. 2. Plasmids used to express bZIP peptides. (**A**) Physical map of pPLc28-bZIP. The fragment carrying a synthetic bZIP gene and transcription termination signal (ter) is indicated as an open bar, the bacteriophage λPL promoter as a filled bar. The bZIP and the ampicillin resistance genes are shown as arrows. The numbers below the restriction sites refer to pBR322 coordinates. (**B**) DNA sequence of the synthetic bZIP gene with the basic region of GCN4. N- and C-terminal residues (1,2 and 357–360) taken from the lac repressor are underlined. The numbers at the borders of these N- or C-terminal residues refer to the numbering of the complete GCN4 protein. The positive and negative numbers indicate the numbering of residues used here. Residue +1, the first leucine of the zipper sequence, is boxed. The ribosomal binding site (RBS) is underlined, and important restriction sites are indicated below the DNA sequence.

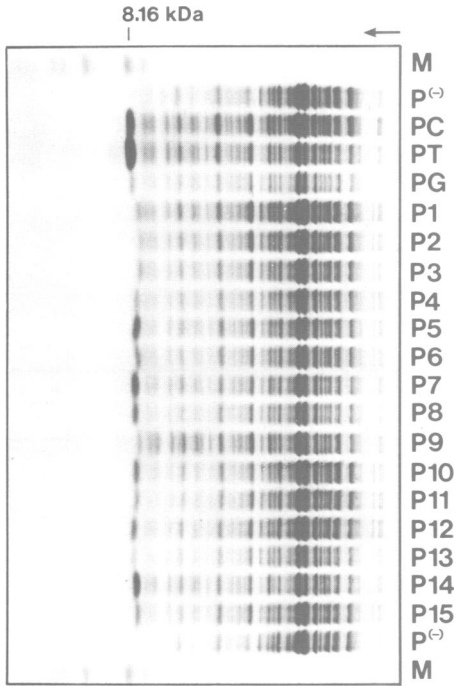

Fig. 3. Tricine−SDS protein gel of the GCN4 peptide and its variants. 10 μl of each crude extract containing 20 mg protein/ml were analysed on a protein gel as described by Schägger and von Jagow (1987). Abbreviations on the right are as follows: M: molecular weight marker MW-SDS-17S, Sigma; the 8.16 kDa band is indicated; $P^{(-)}$: crude extract of strain K12 ΔHI without plasmid after heat induction; PC, PT, PG and P1−P15: crude extracts of strain K12 ΔHI containing the peptides PC to P15 encoded by the various pPLc28-bZIP derivatives. The amino acid sequences of peptides PC−P15 are shown in Figures 5 and 6 (lower left panels). The arrow on the top indicates the direction of gel electrophoresis.

In agreement with previous reports (Oliphant *et al.*, 1989; Pu and Struhl, 1991; Tzamarias *et al.*, 1992) little specificity is observed for position 4. After prolonged exposition of the film, faint binding to target variants with G or T in position 0 becomes visible (data not shown). The C/EBP target is only very weakly retarded, the TAF-1 and the TAF-1* targets are not retarded at all by the GCN4 peptide (PG) (Figures 4, 5 and 7). The C/EBP peptide (PC) binds only to the C/EBP target 5′-ATTGCGCAAT-3′ (Figure 5). When the basic region of GCN4 is replaced with that of TAF-1, the resulting peptide is unable to bind any target (data not shown). We suspected that residues close to the transition from the GCN4 zipper to the TAF-1 basic region could disturb each other in this chimeric peptide. We therefore replaced the first heptad repeat of the GCN4 leucine zipper (LEDKVEE) with that of TAF-1 (LAIRVQS) in addition to the basic region (PT, Figures 5 and 8). Peptide PT does not recognize only the TAF-1 and the TAF-1* targets, but binds even more strongly to the GCN4* sequence and some of its variants with single base pair substitutions (Figures 5 and 8). Both peptides PC and PT served as positive controls for our further studies with hybrid basic regions.

Substitution of residues of the basic region of GCN4 with the corresponding residues of C/EBP

We replaced residues −12 to −25 of the basic region of GCN4 with the corresponding residues of the basic region of C/EBP. The resulting peptide (P1) specifically binds to the C/EBP target which differs from the GCN4 target in positions 1 and 2 (Figure 5 and 7). Having thus confirmed that all amino acids necessary for C/EBP target recognition are located within this region, we exchanged amino acid after amino acid in a stepwise fashion (peptides P2−P6), and analysed the DNA binding properties of the resulting peptides with the hybrid basic regions by gel retardation experiments

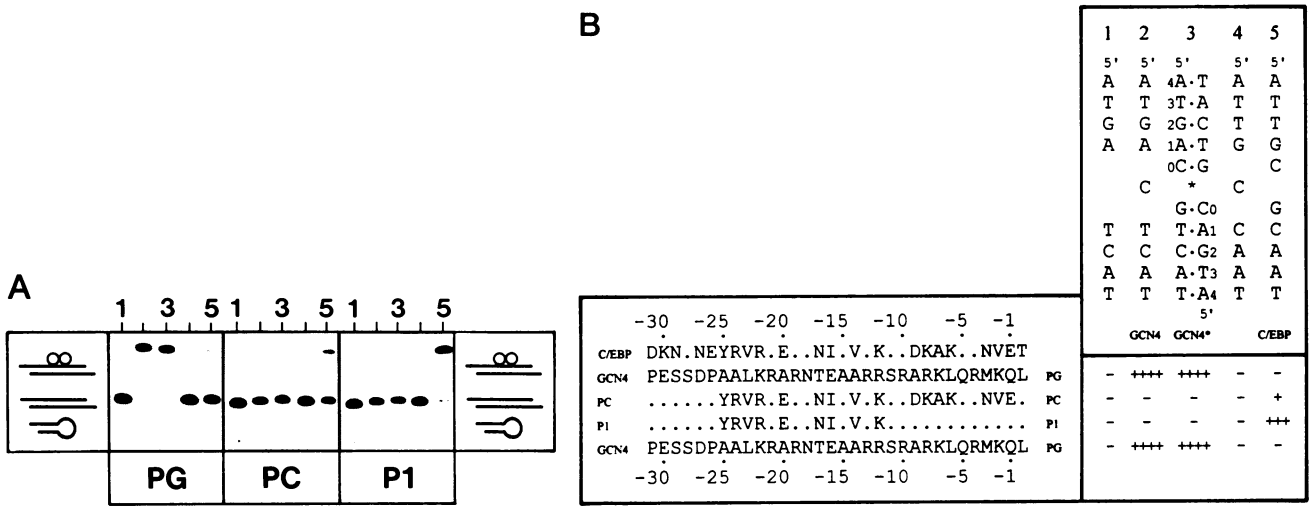

Fig. 4. Alignment of the GCN4 and the C/EBP targets. (**A**) Autoradiograph of electrophoretic mobility shift assays. The targets and petpides used are as summarized in panel B. The diagrams on the left and right of the autoradiographs symbolize the DNA structures (hairpin or double-stranded free DNA) and the protein−DNA complexes that cause the respective bands. (**B**) Upper right panel: five targets were analysed. In columns 1−3, variants of the GCN4 target (column 2) are shown either without spacing (column 1) or with an additional G (GCN4*, column 3) between the canonical TGA blocks. Column 4 shows a derivative of the C/EBP target with only one spacing base pair between the TTG blocks, column 5 the C/EBP target (where the TTG blocks are separated by two base pairs). Lower left panel: amino acid sequences of the basic regions including the first L or T of the first heptad repeat of the leucine zipper of C/EBP, GCN4 and the hybrid peptides in the one letter code. Only those amino acids that differ from the GCN4 amino acid sequence are printed. Numbering is from right to left and starts immediately N-terminal of the first leucine of the leucine zipper. Lower right panel: the affinities of the individual peptides for the respective targets were estimated from the relative intensities of the bands of the complexes and of the free double-stranded DNA: ++++, very high; +++, high; ++, moderate; +, low; −, very low to not measurable.

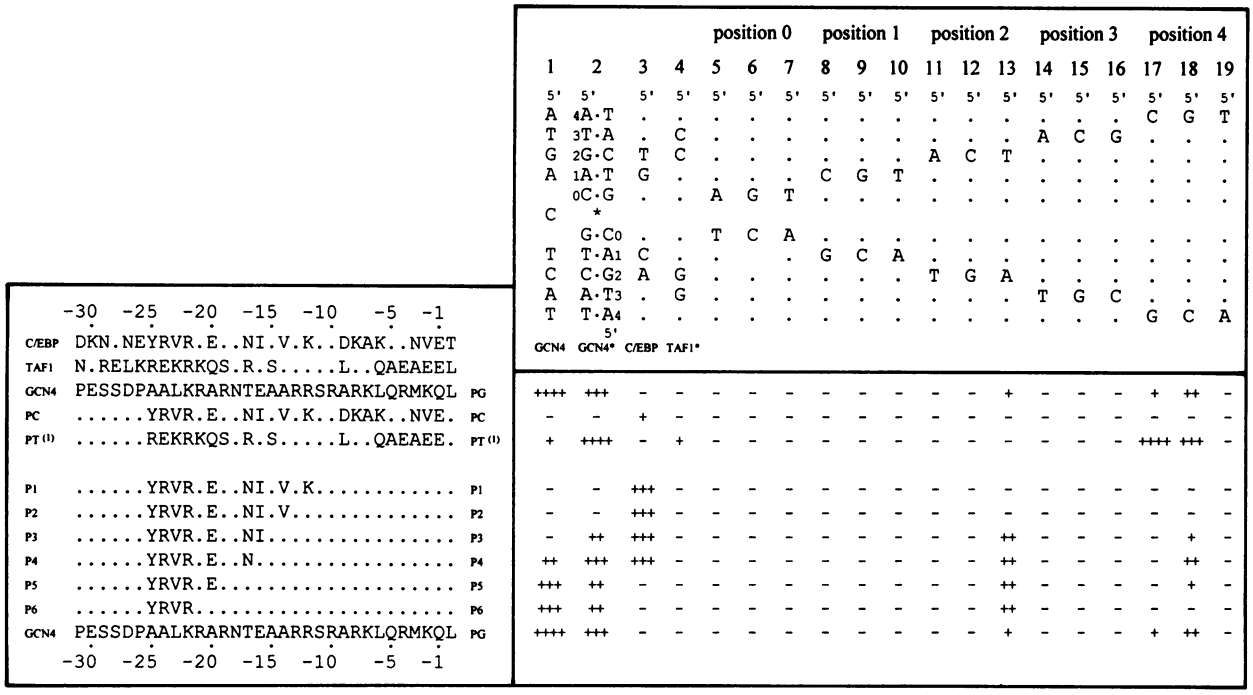

Fig. 5. Residues that discriminate between GNC4 and C/EBP targets reside in positions −17, −16 and −14 of the basic region. Upper right panel: 19 targets were analysed. Their numbering is the same as in Figure 7 where the gel shifts are shown. In column 2 the GCN4* target, the symmetrical variant of the natural GCN4 target is shown as double-stranded sequence. It carries two base pairs instead of one in the centre. The base pairs are numbered from the centre towards the borders. The centre of symmetry is indicated by an asterisk. In columns 5−19 the target variants derived from GNC4* are indicated by the bases that have been replaced. TAF-1* (column 4) differs from the best known TAF-1 target by an A to G exchange in base pair 4 (see Figure 8). Lower left panel: amino acid sequences of the basic regions including the first L or T of the first heptad repeat of the leucine zipper of C/EBP, TAF-1, GCN4 and the hybrid peptides in the one letter code. Scheme and symbols are the same as in Figure 4B. The first heptad repeat of the GCN4 leucine zipper is replaced with the corresponding sequence of TAF-1 in peptide PT(1) in addition to the amino acid exchanges in the basic region. The gel shifts are shown in Figure 7.

(Figure 7). The results are summarized in Figure 5. A comparison of the gel retardation experiments with peptides P1 and P2 shows that arginine −12 of GCN4 and lysine −12 of C/EBP are not involved in the discrimination between the GCN4* and C/EBP targets. The binding properties of P1 and P2 are virtually identical. When valine −14 of C/EBP is replaced with alanine the resulting peptide P3 recognizes the GCN4* and C/EBP targets almost equally well (Figures 5 and 7). The same behaviour is observed with peptide P4 where residue −16 has been replaced (isoleucine to glutamic acid). The binding properties of the latter two peptides are again very similar. Both bind fairly well and better than the GCN4 peptide (PG) to the GCN4* target variant with a T in position 2. The ability to bind to the C/EBP target is lost when asparagine −17 is replaced with threonine. Peptides P5 and P6 (glutamic acid −20 of P5 replaced by alanine in P6) exhibit the same binding specificity as the parent peptide PG with the GCN4 basic region (Figure 5 and 7). From this we conclude that residues −14 to −17 contain information which is required to discriminate between the GCN4* and the C/EBP targets.

Single amino acid exchanges at positions −15 and −17 of the basic region of GCN4

In order to learn more about the function of residue −17 in specific target recognition, we replaced threonine −17 of the basic region of GCN4 by asparagine (P7), glutamine (P8), arginine (P9) or methionine (P10). The gel shift experiments shown in Figure 7 and summarized in Figure 6 indicate that the single threonine to asparagine exchange at position −17 broadens DNA recognition with respect to the targets with G or T in position 0 and allows additional tight binding to the C/EBP target which differs from the GCN4* sequence at positions 1 and 2. Peptide P8 with glutamine in position −17 is not able to bind to the C/EBP target, although glutamine carries the same functional group as asparagine. Peptide P9 exhibits the same specificity as PG except for weak binding to the target variant with a T in position 0 and a slight preference for the GCN4* target versus the GCN4 consensus (Figures 6 and 7). When glutamine −17 is combined with residues −20 and −22 to −25 of the C/EBP basic region, the resulting peptide P15 becomes highly specific for the GCN4* target (Figures 6 and 7). This hints that this region is also somehow involved in DNA binding. Methionine in position −17 (P10) also increases the specificity of the GCN4 basic region for base pairs 0−3, but P10 recognizes the GCN4 and GCN4* targets equally well (Figures 6 and 7).

Because of the overall similarity of the basic regions, we reasoned that the same region of amino acids which determines the change of specificity between GCN4 and C/EBP (i.e. the region between residues −17 and −14) might also determine the specificity of TAF-1. In this region TAF-1 differs from GCN4 only by two residues: −17 and −15 (Figures 5 and 6). We thus replaced threonine −17 alone or alanine −15 alone with the corresponding residues

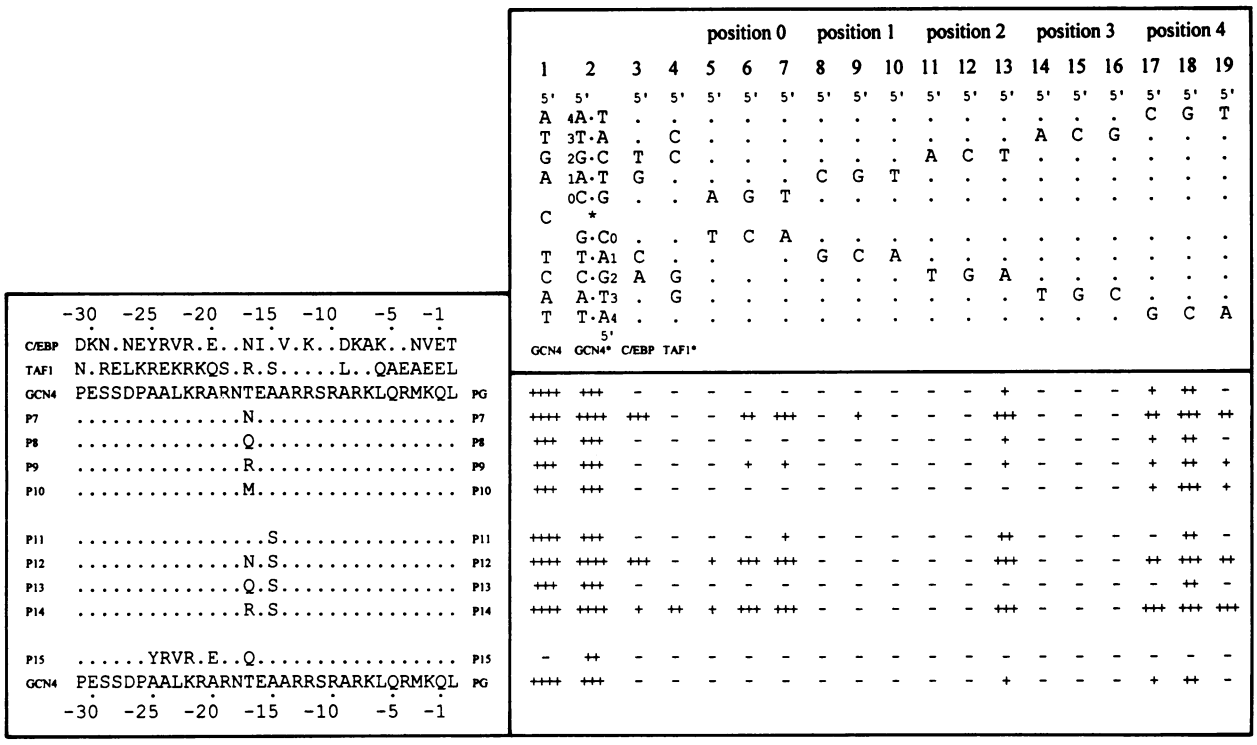

Fig. 6. The importance of residues −15 and −17. Single and double exchanges in positions −17 and −15 of GCN4 are indicated. Scheme and symbols are the same as in Figures 4B and 5. For PT[(1)] see legend of Figure 5. The gel shifts are shown in Figure 7.

of the TAF-1 basic regions. We found the specificity of the respective peptides P9 and P11 slightly broadened (Figures 6 and 7), i.e. they are able to bind with low affinity to the TAF-1 target (Figure 8; the affinity of P11 to the TAF-1 target becomes visible after longer exposure of the film).

Changing the specificity from C/EBP to TAF-1

We then constructed the double mutant P14 where both alanine −15 and threonine −17 are replaced with serine −15 and arginine −17 of the TAF-1 basic region and tested its binding properties with the various targets (Figures 6 and 7). P14 binds very well to the GCN4*, weakly to the C/EBP and moderately well to the TAF-1* target. It does not discriminate between the base pairs in position 0 and 4 of the GCN4* target variants, but it binds significantly better to the G-box motif (the TAF-1 target with a G in position 4) than to the TAF-1* target with an A in position 4 (Figure 8). It behaves like the basic construct PT, which includes the whole basic region of TAF-1. If arginine −17 in P14 is replaced with the corresponding asparagine from C/EBP (P12), none of the TAF-1 targets is recognized, but the C/EBP target is bound very well. The ability to bind to the C/EBP target is lost in peptide P13 with glutamine −17 combined with serine −15 (see Figures 6 and 7). Thus a single amino acid substitution is sufficient to switch the specificity from C/EBP to TAF-1. A comparison of the GCN4 peptide (PG) and the double mutant P14 (arginine −17 and serine −15) shows that the double exchange broadens the specificity from GCN4 to TAF-1 in the same manner as does the insertion of the whole basic region of TAF-1 including the first leucine heptad repeat (PT, Figures 5 and 8).

Discussion

The test system

The use of crude extracts without further purification for band shift experiments needs to be discussed. As shown in Figure 3, there are differences in the expression rates between the various peptides. Peptides PG and P1−P15 differ maximally by a factor of 5. The peptides which are compared directly are often expressed at similar amounts (for example P2/P3/P4, P9/P11 or P12/P14, see Figures 6, 7 and 8). In some cases the peptide expressed at a higher amount shows lower activity than one expressed at a lower level (for example PC/PG, PC/P1 or PT/P14; see Figures 4, 5 and 8). Finally the slight differences between individual gel shift experiments should be mentioned. For example, the combination of peptide PG and the target 5′-ATGACGTCAT-3′ can result in a complete DNA shift (++++, see Figure 4) or in a 90% shift (+++, see Figure 5). Individual experiments differed by maximally ±10%.

The role of residues −14 to −17 of the basic regions of GCN4, C/EBP and TAF-1

Substitution of residues −14 to −17 of the basic region of GCN4 with the corresponding residues of C/EBP and TAF-1 allows the mutant peptides to recognize the C/EBP and TAF-1 targets, respectively (Figures 5, 6 and 8). We thus conclude that some of these residues participate directly in target recognition. In the experiments summarized in Figure 5, peptides P1 and P2 recognize only the C/EBP target, whereas peptides P5 and P6, like the GCN4 peptide (PG), bind only to the GCN4 targets. Peptides P3 and P4 bind to

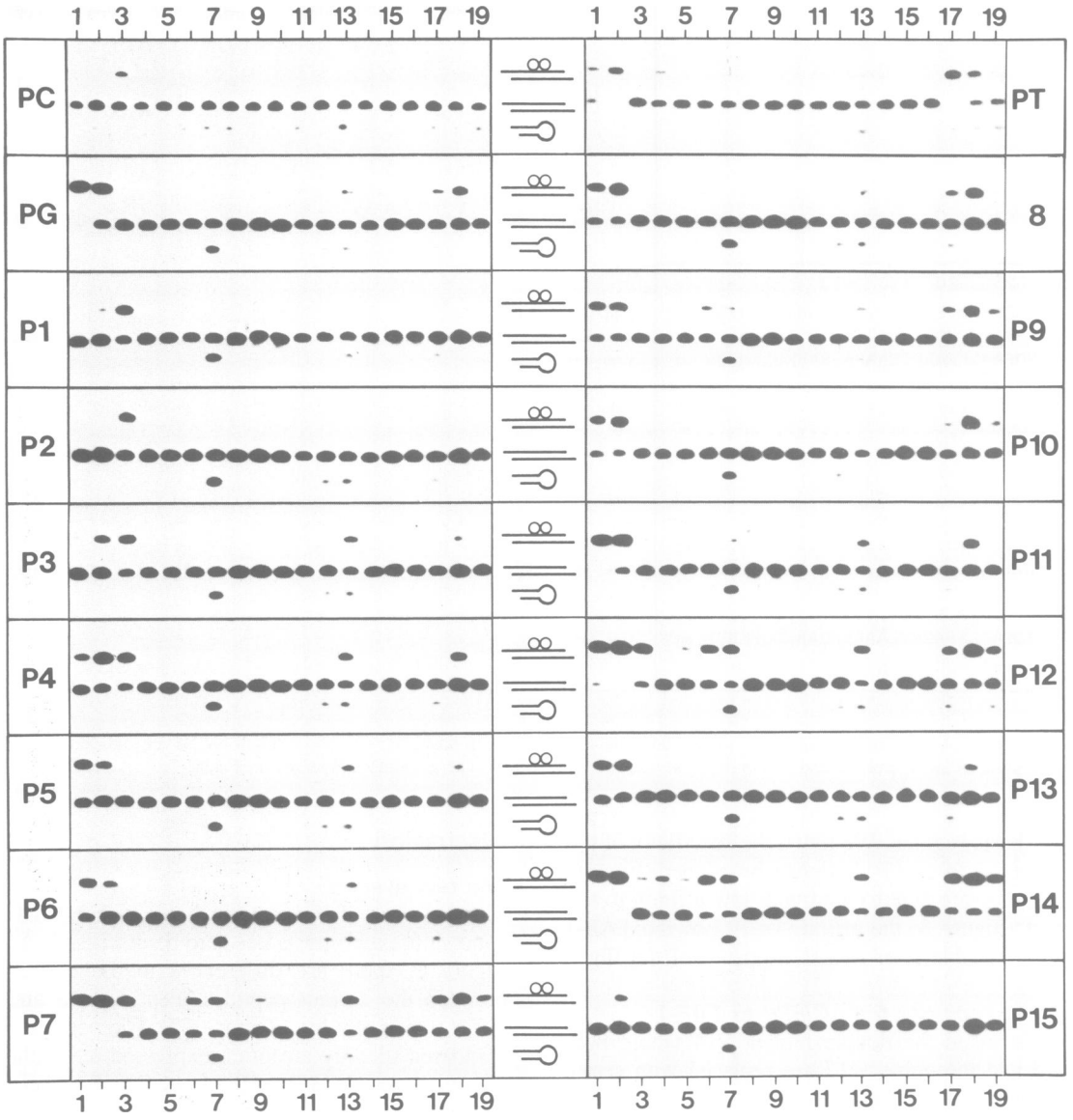

Fig. 7. Autoradiographs of mobility shift assays with crude extracts containing bZIP peptides. The numbers above and below the lanes refer to the targets used. They are the same as in Figures 5 and 6. Thus lane 1 refers to GCN4 consensus, lane 2 to the symmetric GCN4* target with one additional base pair in the centre, lane 3 to the C/EBP target, lane 4 to a variant of the best possible TAF-1 target which carries an A to G exchange in base pair 4 (see Figures 5, 6 and 8). Lanes 5–19 correspond to symmetric variants of the fully symmetric GCN4* target (Figures 5 and 6). The bZIP peptides are indicated on the left and right side respectively. PG corresponds to the parental GCN4 peptide. The protein sequences of PC, PT and P1–P6 are listed in Figure 5 and those of P7–P15 in Figure 6. The symbols between the autoradiographs are the same as in Figure 4A.

both targets. One common feature of peptides P1, P2, P3 and P4 is asparagine −17, and one common feature of peptides P3, P4, P5 and P6 is alanine −14. Our results thus suggest that both valine −14 and asparagine −17 are involved in switching the specificity from the GCN4 to C/EBP target.

Recently the structure of a GCN4 protein–DNA cocrystal has been solved (Ellenberger *et al.*, 1992). The authors used as target the asymmetric GCN4 binding site 5′-ATGACTCAT-3′. Their X-ray data show that alanine −14 of GCN4 contacts the methyl group of thymine in position 1 of the target, while threonine −17 anchors the helix to the phosphate backbone near the centre of the binding site. In general there are two possible ways of influencing the specificity of DNA binding: the direct way is to alter residues that contact base pairs directly, but changes in the positioning of a recognition helix may also lead indirectly to an altered specific binding. Our genetic approach cannot distinguish between these two possibilities, but X-ray analysis of a protein–DNA complex can. Since both asparagine −17 and valine −14 (Figure 5, P2–P5) are necessary for switching the specificity of the bZIP domain from GCN4* to C/EBP, we have here an example for a specificity change where a new amino acid–base pair contact has to be accompanied by an adjustment of the position of the recognition helix. Peptides P4, P5 and P7 show that an asparagine in position −17 alone is sufficient to broaden the specificity of GCN4-derived peptides such that the C/EBP

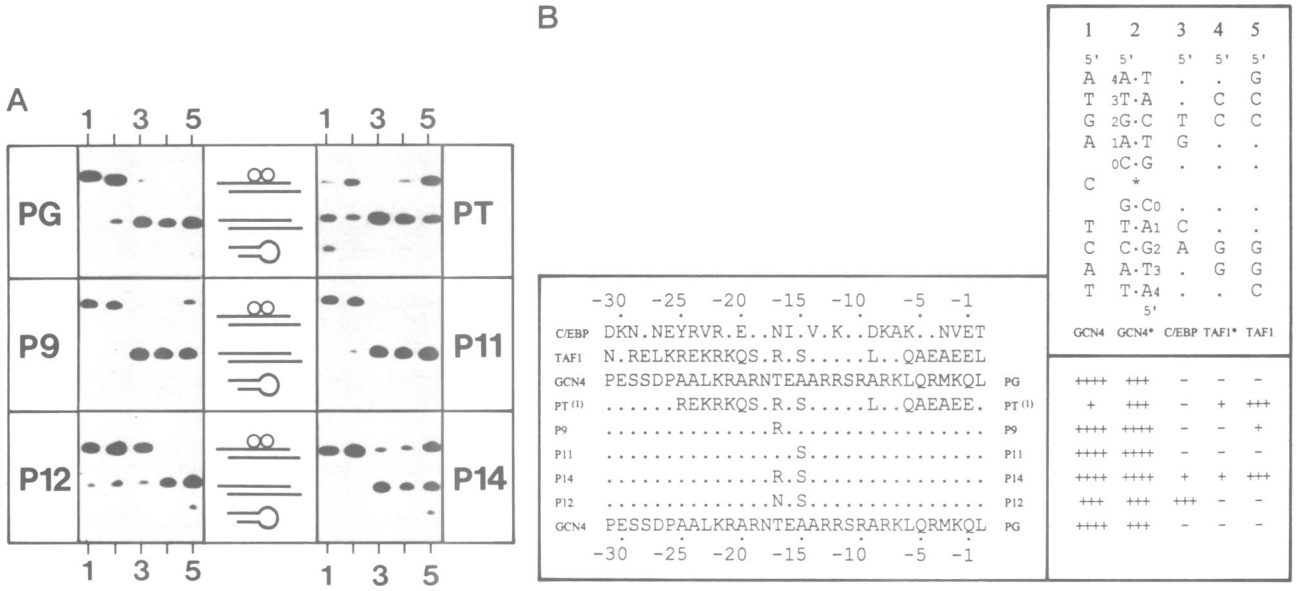

Fig. 8. The role of residues −15 and −17 with respect to the TAF-1 and the C/EBP binding sites. (**A**) Autoradiograph of electrophoretic mobility shift assays. Lanes 1–4 are the same as in Figure 7. In lane 5 the TAF-1 consensus target (Oeda *et al.*, 1991) is used. The symbols between the autoradiographs are the same as in Figure 4A. (**B**) Summary of the results shown in part A. Scheme and symbols are the same as in Figure 4B. For $PT^{(1)}$ see legend of Figure 5.

binding site is recognized. Only in combination with valine −14 can the binding to GCN4 targets be avoided (Figure 5, P2).

Similar arguments can be made in the case of TAF-1. While C/EBP and GCN4 bind different DNA targets, peptide PT binds all targets bound by GCN4 except the variant of the GCN4* target with a T in position 2 in addition to the TAF-1 targets. The sequence 5′-GCCACGTGGC-3′, which has been described as a high affinity binding site for the TAF-1 protein (Oeda *et al.*, 1991), is indeed recognized better than the TAF-1* target by peptide PT (Figure 8). A comparison of PT and P14 shows that the residues responsible for the differences in the DNA binding specificities of GCN4 and TAF-1 are located in positions −17 and −15 (P14, Figures 6 and 8). Thus the residues responsible for the altered DNA binding specificity are found again in positions from −14 to −17 of the basic region. The single substitution of threonine −17 of GCN4 with an arginine leads to weak but detectable binding to the TAF-1 target (P9, Figure 8), substitution of alanine −15 of GCN4 by a serine has a similar effect, but P11 binds even more weakly to the TAF-1 target than P9. The complex becomes visible only after longer exposition of the film (P11, Figure 8). The crystal structure of the GCN4–DNA complex (Ellenberger *et al.*, 1992) shows that alanine −15 of GCN4 contacts the methyl group of thymine in position 3 of the TRE site. Blatter *et al.* (1992) recently also detected this contact by using a site-specific bromouracil-mediated photocrosslinking method. The GCN4* and the TAF-1 targets differ in positions 2, 3 and 4. Changing one direct base contact (serine −15) and adjusting the position of the helix in the major groove of the DNA (arginine −17) may be necessary to alter the GCN4 binding specificity to that of the TAF-1. Residue −17 again plays a key role in changing the specificity.

Thus it seems that the backbone contact near the centre of the binding site is important for positioning residues that contact base pairs directly. This is further emphasized by peptides P12 and P14 (Figure 8) which differ only in their residue in position −17 of their basic regions: P12, which has asparagine at −17, binds well to the C/EBP target, but not to the TAF-1 and the TAF-1* targets, while P14, which has arginine at −17, shows the opposite binding behaviour. Since the X-ray analysis of the GCN4–DNA complex has been done with the asymmetric GCN4 target, it cannot be excluded that some details in the network of direct and indirect contacts may be different when GCN4 is complexed with the fully symmetric GCN4* binding site. One could also imagine that replacing the relative short side chain of threonine −17 in GCN4 by the longer ones of either arginine or asparagine could lead to a change in the mode of recognition of the basic region. This would explain why the changes of specificity reported here are not additive: only the complete C/EBP or TAF-1 sites are recognized, not the single base pair substitutions in the GCN4* target leading to them. Residues −19 to −25 probably contribute to the stability of the complexes and in some manner also to the specificity. The accumulation of four (TAF-1, C/EBP) or three (GCN4) basic amino acids in this short region suggests that these parts of the basic regions may also serve to anchor the protein to the DNA by means of salt bridges with phosphates of the DNA backbone. Ellenberger *et al.* (1992) found in their analysis of the GCN4–DNA complex a salt bridge between arginine −19 of GCN4 and a phosphate group beyond the base pair in the target position 4. This could account for the requirement of these residues for the complete specificity change in the C/EBP case.

The changes of specificity between GCN4 and C/EBP, C/EBP and TAF-1, and the broadening of specificity from GCN4 to TAF-1 are mediated by residues −14, −15 and −17 of the respective basic regions. In all three cases, residue −17 of the respective protein plays a key role in

altering the target recognition, most probably by altering the position of either residue −14 (GCN4/C/EBP) or residue −15 (GCN4/TAF-1). In the case of C/EBP/TAF-1, residue −17 alone is sufficient for the change of binding specificity.

Materials and methods

Bacterial strains and plasmids

E.coli K12ΔH1 Sm^R*lacZ*am*Δbio-uvrB ΔtrpEA2*(λ*N*am7-*N*am53*c*I857ΔH1) and pPLc28 (Remaut *et al.*, 1981) 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, polynucleotide kinase and *Taq* polymerase were from Boehringer (Mannheim, Germany) and 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); all other chemicals including the SDS molecular weight marker no. MW-SDS-17S from Sigma (München, Germany) or Merck (Darmstadt, Germany).

Oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer and purified on denaturing polyacrylamide gels prior to use.

Methods

Standard techniques were used for *in vitro* manipulation of DNA, as described by Sambrook *et al.* (1989). The synthetic gene encoding the GCN4 basic region and leucine zipper sequences was composed of eight overlapping, complementary oligonucleotides. The strategy for their simultaneous cloning has been described (Wilcken-Bergmann *et al.*, 1986). For the preparation of crude protein extracts small cultures of freshly transformed colonies of *E.coli* K12ΔH1 were grown to saturation over night at 30°C, diluted 1:25 in dYT with 0.5 mg ampicillin/ml and incubated at 30°C until the OD_{600} reached 0.5. The cultures were then placed into a 70°C waterbath until they reached a temperature of 42°C and shaken at 42°C for 20 min. After further 90 min incubation at 37°C the cultures were quickly cooled on ice. The OD_{600} was determined and the cells were harvested by centrifugation. The cell pellet was resuspended at 10^{11} cells per ml in extraction buffer (200 mM Tris−HCl, pH 8; 10 mM $MgCl_2$; 1 mM EDTA; 7 mM β-mercaptoethanol) and sonicated (W220F, microtip; Heat Systems-Ultrasonics, Inc.) for 1 min at 4.5 scale units. After 30 min centrifugation 30% glycerol (v/v) was added to the clear supernatant, and aliquots were stored at −70°C. Electrophoretic mobility shift assays were performed as described by Hope and Struhl (1985): 1 μl of a 1:20 dilution of the crude extracts in extraction buffer was added to 7 μl binding buffer (20 mM Tris−HCl, pH 7; 50 mM KCl; 3 mM EDTA; 2 mM DTT; 100 μg gelatin/ml; 1 mg BSA/ml) which contained 1 fmol of the respective radiolabelled target DNA fragments and a 1000-fold excess of poly (dI:dC). This mixture was incubated on ice for 15 min. The samples were loaded onto a running 10% polyacrylamide gel after adding of 3 μl of 15% Ficoll, 0.1% bromophenol blue and 0.1% xylene cyanol in binding buffer.

The fragments used for gel shift experiments differ in position 0−4 of their binding sites. The flanking regions were optimized up to position 11 according to the results of Oliphant *et al.* (1989) and *Xba*I cohesive ends were added: 5′-CTAGACGGGCGATGACTCATCGCCCGT-3′ (the GCN4 target is underlined). Synthetic oligonucleotides with the fully symmetric targets were purified by denaturing polyacrylamide gel electrophoresis and labelled without prior cloning. Oligonucleotides carrying palindromes with central base pairs were cloned into the *Xba*I site of the polylinker of pWB9 (Kolkhof *et al.*, 1992) and were excised with *Xba*I to generate fragments of identical length. A typical labelling reaction contained 200 fmol target fragments, 10 μCi [^{32}P]dCTP (Amersham Buchler) and 0.2 U of Klenow. The mixture was passed through a 10% native polyacrylamide gel for final purification. Protein gel electrophoresis was performed as described by Schägger and von Jagow (1987).

Acknowledgements

We thank W.Fiers for providing plasmid pPLc28 and *E.coli* K12ΔH1, K.Otto for expert technical assistance, U.Ringeisen for excellent photography, B.Kisters-Woike and R.Fickert for stimulating discussions, and R.S.Jack for critically reading the manuscript. In particular we are indebted to T.E.Ellenberger for communicating results prior to publication. This work was supported by DFG through SFB 274.

References

Agre,P., Johnson,P.F. and McKnight,S.L. (1989) *Science*, **246**, 922−926.
Blatter,E.E., Ebright,Y.W. and Ebright,R.H. (1992) *Nature*, **359**, 650−652.
Boelens,R., Lamerichs,R.M.J.N., Rullmann,J.A.C., van Boom,J.H. and Kaptein,R. (1988) *Protein Sequence Data Anal.*, **1**, 487−498.
Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) *Cell*, **171**, 1223−1237.
Gentz,R., Langner,A., Chang,A.C.Y., Cohen,S.T. and Bujard,H. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4936−4940.
Hill,D.E., Hope,I.A., Macke,J.P. and Struhl,K. (1986) *Science*, **234**, 451−457.
Hope,I.A. and Struhl,K. (1985) *Cell*, **43**, 177−188.
Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885−894.
Hope,I.A. and Struhl,K. (1987) *EMBO J.*, **6**, 2781−2784.
Kerppola,T.K. and Curran,T. (1991a) *Cell*, **66**, 317−326.
Kerppola,T.K. and Curran,.T. (1991b) *Science*, **254**, 1210−1214.
Kisters-Woike,B., Lehming,N., Sartorius,J., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1991) *Eur. J. Biochem.*, **198**, 411−419.
Kolkhof,P., Teichmann,D., Kisters-Woike,B., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1992) *EMBO J.*, **11**, 3031−3038.
Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759−1764.
Lehming,N., Sartorius,J., Niemöller,M., Genenger,G., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1987) *EMBO J.*, **6**, 3145−3153.
Lehming,N., Sartorious,J., Kisters-Woike,B., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1990) *EMBO J.*, **9**, 615−621.
Lehming,N., Sartorius,J., Kisters-Woike,B., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1991) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Moleculr Biology*. Springer Verlag, Berlin, Vol. 5, pp. 114−125.
Oeda,K., Salinas,J. and Chua,N.-H. (1991) *EMBO J.*, **10**, 1793−1892.
Oliphant,A.R., Brandl,C.J. and Struhl,K. (1989) *Mol. Cell. Biol.*, **9**, 2944−2949.
O'Neil,K.T., Hoess,R.H. and DeGrado,W.F. (1990) *Science*, **249**, 774−778.
Pathak,D. and Sigler,P. (1992) *Curr. Biol.*, **2**, 116−123.
Pu,W.T. and Struhl,K. (1991) *Mol. Cell. Biol.*, **11**, 4918−4926.
Remaut,E., Stanssens,P. and Fiers,W. (1981) *Gene*, **15**, 81−93.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sartorius,J., Lehming,N., Kisters-Woike,B., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1989) *EMBO J.*, **8**, 1265−1270.
Sartorius,J., Lehming,N., Kisters-Woike,B., Wilcken-Bergmann,B.v. and Müller-Hill,B. (1991) *J. Mol. Biol.*, **218**, 313−321.
Schägger,H. and von Jagow,G. (1987) *Anal. Biochem.*, **166**, 368−379.
Sellers,J.W., Vincent,A.C. and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 5077−5086.
Tsamarias,D., Pu,W.T. and Struhl,K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2007−2011.
Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) *Science*, **246**, 911−916.
Wilcken-Bergmann,B.v., Tils,D., Sartorius,J., Auerswald,E.A., Schröder,W. and Müller-Hill,B. (1986) *EMBO J.*, **5**, 3219−3225.

Received on July 23, 1992; revised on November 16, 1992