The EMBO Journal vol.8 no.10 pp.3029–3037, 1989

# Isolation and DNA-binding characteristics of a protein involved in transcription activation of two divergently transcribed, essential yeast genes

**Hartmut Halfter, Ulrike Müller[1], Ernst-L. Winnacker[1] and Dieter Gallwitz**

Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Genetics, PO Box 2841, D-3400 Göttingen and [1]University of Munich, Department of Biochemistry, Karlstrasse 23, D-8000 München, FRG

Communicated by D.Gallwitz

**We have identified a protein, BAF1, which has two oppositely oriented, partially overlapping binding sites within a symmetrical sequence located midway between and upstream of the divergently transcribed *YPT1* and *TUB2* genes of the yeast *Saccharomyces cerevisiae*. The 120 kd BAF1 protein was purified to near homogeneity and used to delineate the two binding sites and to identify apparent protein contact sites by the missing contact technique, methylation interference and by site-directed mutagenesis. The BAF1-recognition sequence contains a conserved $TCN_7ACG$ element recently identified at autonomously replicating sequences (*ARS*) and in the 5′ and 3′ flanking region of other yeast genes. The symmetrical sequence of the *YPT1/TUB2* intergene region seems not to be involved in DNA replication but activates transcription in an orientation-independent fashion.**

*Key words:* BAF1 protein/bidirectional transcription activation/upstream activating sequence/yeast/*YPT1*–*TUB2* intergene region

## Introduction

In the yeast *Saccharomyces cerevisiae*, the genes encoding the *ras*-related, guanine nucleotide-binding *YPT1* protein (Gallwitz *et al.*, 1983) and β-tubulin (Neff *et al.*, 1983) are essential for cell viability. The *YPT1* and *TUB2* genes, located on chromosome VI, are divergently transcribed and only some 195 bp separate their transcription start sites. We had previously noted that a 30 bp region of dyad symmetry is located midway between the two genes and suggested a possible role of this sequence in transcription regulation (Gallwitz *et al.*, 1983). The deletion of a 38 bp fragment from the *YPT1/TUB2* intergene region containing this symmetrical sequence did indeed negatively affect the transcription of both genes (Donath, 1985). This is reminiscent of the well-studied *GAL* upstream activating sequence ($UAS_G$), a GC-rich sequence that is located in the *GAL1/GAL10* intergene region and is responsible for the galactose-induced *GAL4* protein binding resulting in the activation of the two divergently transcribed genes (Johnston and Davis, 1984; West *et al.*, 1984; Bram and Kornberg, 1985; Giniger *et al.*, 1985).

In this report we describe the purification and DNA-binding characteristics of a 120 kd protein, called BAF1 (for *b*idirectionally *a*ctivating *f*actor), that binds to the symmetrical sequence within the *YPT1/TUB2* intergene region as well as to a sequence element of the autonomously replicating sequence *ARS1* that has recently been shown to bind a protein factor, ABF1 (Buchman *et al.*, 1988; Diffley and Stillman, 1988). We also show that the BAF1 protein-binding region, regardless of its orientation, serves to activate the transcription of a downstream gene.

## Results

### *Protein binding to part of the region of dyad symmetry*

The sequence of the *YPT1/TUB2* intergene region is presented in Figure 1A. The centre of this region harbours an imperfect palindrome with 6 bp long perfectly matching arms enclosing 14 bp of alternating purine and pyrimidine nucleotides.

To test the potential role of this sequence in protein binding and transcription regulation, a 5′ end-labelled 142 bp *Nco*I–*Bgl*II fragment (Figure 1A) was probed for specific complex formation with heparin–agarose-binding proteins of the proteinase-deficient *Saccharomyces cerevisiae* strain ABYS1 (Achstetter *et al.*, 1984). Such a complex could be easily identified in a gel-retardation assay (Fried and Crothers, 1981) and binding of the $^{32}P$-labelled DNA fragment could be specifically competed with by pre-incubation of the proteins with an excess of unlabelled fragment (data not shown). The sequence recognized by the protein was determined by DNase I protection mapping (Galas and Schmitz, 1978) using the same DNA fragment labelled at its *Bgl*II restriction site. As can be seen in Figure 1B, a sequence of 22 bp (nucleotides −172 to −193 upstream of the *YPT1* initiation codon, see Figure 1A), encompassing one arm of the inverted repeat, was protected from DNase I digestion.

For further analysis of protein–DNA interaction and for purification of the DNA-binding protein, a 53 bp double-stranded DNA fragment was synthesized. It contained nucleotides −150 to −200 of the intergene region (Figure 1A) and was extended by a GC base pair on either side (WT, in Figure 2A) to allow the recovery of single fragments from cloned tandem repeats by *Sma*I or *Xma*I restriction cutting. Gel-retardation assays proved that this end-labelled fragment gave rise to specific complexes with the same protein fractions eluted from a heparin–agarose column that formed specific complexes with the *Nco*I–*Bgl*II fragment (see above). The labelled complex vanished after pre-incubation of the binding protein(s) with a 100-fold molar excess of unlabelled 'wild-type' fragment (WT) or of mutated fragments (Mir L and Mir R) with either one of the inverted repeat sequences destroyed (Figure 2A,B). Destruction of the repeat sequence on the side of the *TUB2* gene (fragment Mir L), however, made the fragment a less potent competitor for labelled complex formation than the mutation of the repeat sequence facing the *YPT1* gene (fragment Mir R, see Figure 2B). Specific complex formation, although less

©IRL Press

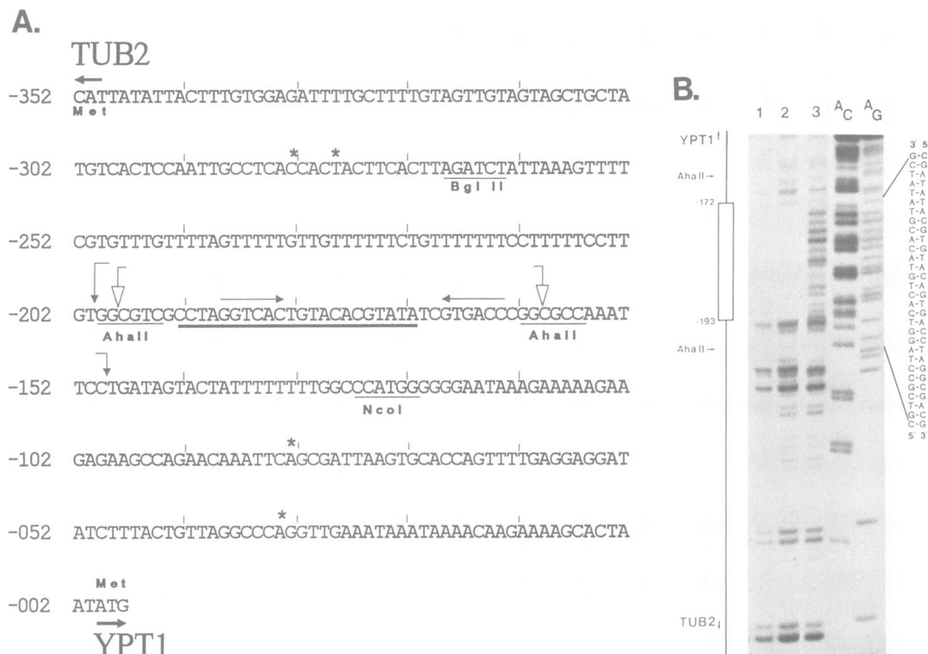



**Fig. 1.** *YPT1/TUB2* intergene region on chromosome VI (Gallwitz *et al.*, 1983) and footprint analysis of DNA−BAF1 protein complex. (**A**) Transcription start sites are marked by asterisks. The perfect matching arms of the inverted repeat are shown by horizontal arrows. The sequence protected from DNase I digestion of the DNA−protein complex is underlined. Open vertical arrows define the segment whose deletion leads to an impairment of transcription of both genes (Donath, 1985). Closed vertical arrows show the borders of oligonucleotides used in gel-retardation assays. (**B**) Footprint analysis showing the region of the 142 bp *Bgl*II−*Nco*I fragment protected from DNase I digestion in the DNA−BAF1 protein complex. Between 2 and 5 ng of fragment, 5′ end-labelled at its *Bgl*II restriction site, was complexed with heparin−agarose-purified protein and protein-bound DNA (**lanes 1,2**) and free DNA (**lane 3**) were digested with DNase I (Worthington). A sequence ladder generated by chemical degradation of the fragment (A/C and A/G reactions) and the sequence protected are shown to the right. The protected region and its border nucleotides are schematically represented to the left.

efficient than with the wild-type sequence, was indeed observed with the mutated fragment Mir L but not with Mir 1, the latter of which carried several base substitutions to abolish the alternation of purine and pyrimidine nucleotides (Figures 2B and 8).

### *Purification of the DNA-binding protein*

The isolation of a 120 kd protein binding to the central part of the *YPT1/TUB2* intergene region was achieved by several successive purification steps. Fractions containing the DNA-binding activity were identified by complex formation with the 53 bp fragment (Figure 2B) in gel-retardation assays. Total cellular protein was first passed over a heparin−agarose column from which the binding activity was eluted with 0.2−0.3 M $(NH_4)_2SO_4$. After $(NH_4)_2SO_4$ precipitation, the proteins were either passed over DEAE−cellulose [elution of the DNA-binding protein(s) between 0.25 and 0.35 M NaCl], phosphocellulose (elution of active protein at 0.4−0.6 M NaCl) or a calf thymus DNA−Sepharose column (elution of active protein at 0.15−0.25 M NaCl). Extensive purification was then achieved by affinity purification using tandem repeats (20−30 copies) of the 53 bp fragment (WT in Figure 2A) covalently linked to Sepharose. The active fractions after DNA affinity purification enriched for a protein of ~120 kd were finally subjected to FPLC on a Mono-Q column. As shown in Figure 3, a protein of ~120 kd was purified to near homogeneity (fractions 34−42). Similar to other purified transcription factors (Sorger and Pelham, 1987; Wu *et al.*, 1987), the 120 kd protein appeared as doublet in a 7.5% polyacrylamide gel. Band-shift activity was observed with fractions containing the 120 kd protein (fractions 34−42) (Figure 3).

To prove further that the 120 kd protein is the factor binding to the *YPT1/TUB2* intergene region, the $^{32}P$-labelled 53 bp fragment containing bromodeoxyuridine (BUdR) was used to form complexes with DNA affinity-purified protein and subsequently crosslinked to the bound protein with UV light (Ogata and Gilbert, 1977; Treisman, 1987). Separation of the crosslinked material by polyacrylamide gel electrophoresis (Figure 4) led again to the identification of a protein of ~120 kd as the prominent DNA-binding factor. Competition of complex formation with an excess of the WT- and the Mir R-fragment, and to a lesser extent with the Mir L-fragment (Figure 2B), underlined the sequence specificity of this protein−DNA interaction. In addition, the upstream activating sequence of the *S.cerevisiae TEF2* gene, which is known to bind the RAP1 protein also having an apparent molecular mass of 120 kd (Huet and Sentenac, 1987; Shore and Nasmyth, 1987), proved unable to compete for BAF1 binding (Figure 2B).

### *Identification of apparent protein contact sites*

A more detailed analysis of the DNA−protein interaction was performed with chemical modification experiments, methylation interference (Siebenlist and Gilbert, 1980) and missing contact probing using partially depurinated and depyrimidated DNA (Brunelle and Schleif, 1987). As shown in Figure 5A, pre-methylated *Bgl*II−*Nco*I fragment complexed with affinity-purified binding protein and subsequently cleaved with piperidine led to the identification of G residues −176 and −188 of the coding strand and G

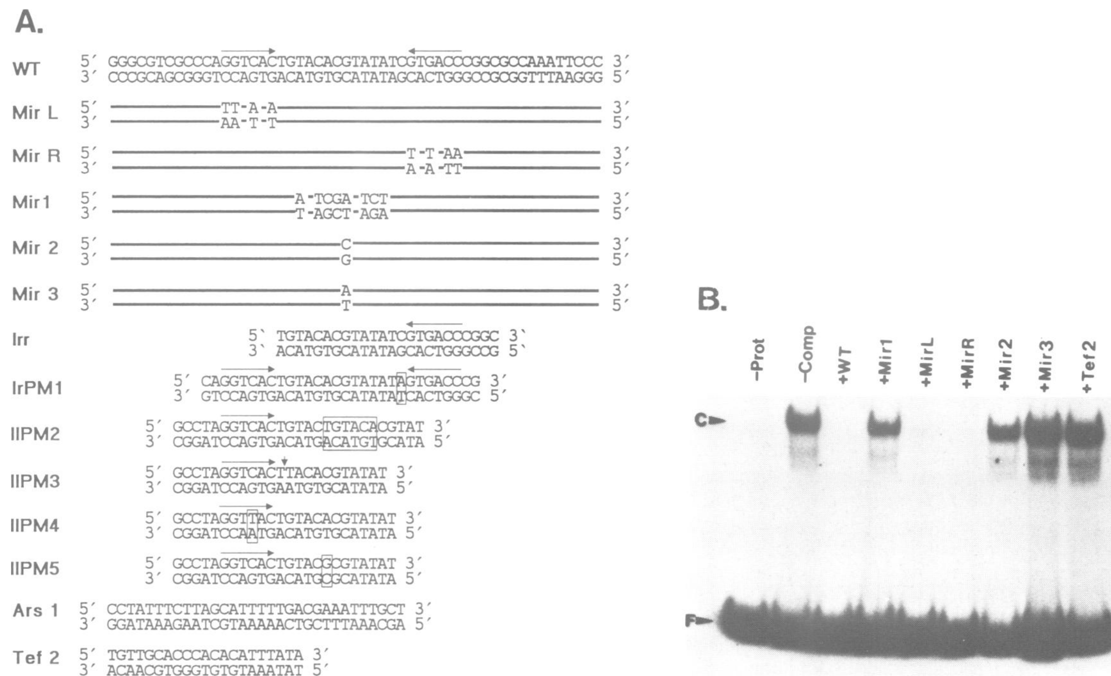



**Fig. 2.** Chemically synthesized oligonucleotides used for band-shift analyses. (**A**) The arms of the symmetrical sequence within the *YPT1/TUB2* intergene region are indicated by horizontal arrows. The base changes only are shown for the 53 bp mutant oligonucleotides. The sequences of oligonucleotides WT and Mir deviate from the chromosomal sequence in position −192 (T → C transition; see also Figure 1); this transition did not interfere in any way with protein binding. For shorter oligonucleotides, the entire sequence is shown and mutations introduced are boxed. To facilitate comparison of oligonucleotides, identical sequences are aligned. A vertical arrow (fragment I1PM3) indicates the position where a GC base pair was deleted. Oligonucleotides representing segments of the upstream region of the *TEF2* gene (Huet and Sentenac, 1987) and of the *ARS1* region and shown to be able to form specific DNA−protein complexes were designed according to Shore and Nasmyth (1987) and Diffley and Stillman (1988). (**B**) Formation of specific DNA−protein complexes (C) with a segment of the *YPT1/TUB2* intergene region (oligonucleotide WT) and heparin−agarose-purified protein. Competition experiments were performed with a 100-fold molar excess of different unlabelled oligonucleotides. The labelled oligonucleotide WT (F) not bound to protein (−Prot) or bound to protein without addition of competitor oligonucleotide (−Comp) were run in the two left lanes of the polyacrylamide gel.

residues −177, −184 and −186 of the non-coding strand (with respect to the *YPT1* gene) as apparent contact sites. Complex formation with DNA missing either G and A or C and A bases was in addition used to identify protein contact sites. This method is based on the assumption that the removal of bases involved in protein binding leads to an enrichment of fragments corresponding to these bases in chemically degraded free versus protein-bound DNA (Brunelle and Schleif, 1987). By comparing the intensities of labelled fragments of free and protein-bound DNA (Figure 5B), significant differences were observed for several residues, all of them residing in the region protected from DNase I digestion. In our hands, this method proved helpful to delineate the core binding region but it was problematic in the identification of potential protein contact sites. For instance, missing contact probing but not repeated methylation interference experiments pointed to G residue −182 being a protein contact site (Figure 5). Another discrepancy became apparent in a missing contact analysis using the 142 bp *Bgl*II−*Nco*I fragment, end-labelled at its *Bgl*II restriction site: according to partial depurination (guanines and adenines), the A residue in position −180 would be predicted to contact the protein (Figure 5B, lanes 1 and 2), but in contrast, after limited depurination and depyrimidation (adenines and cytosines), this A residue was not recognized as a contact site (Figure 5B, lanes 3 and 4).

Supporting evidence for the G residue in position −176 being involved in protein binding was obtained from a mutational analysis of the binding region. Replacement of

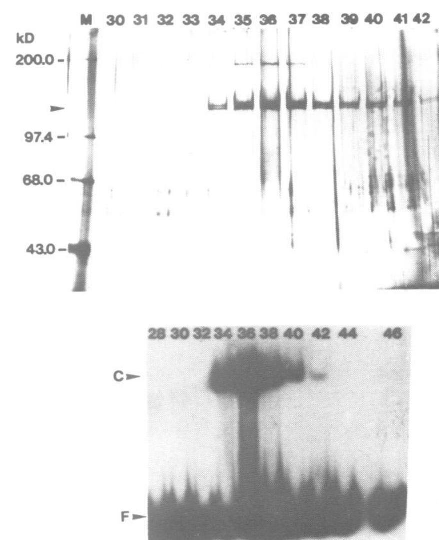



**Fig. 3.** FPLC purification of the BAF1 protein and the ability of the protein to form specific DNA−protein complexes. DNA affinity-purified protein was eluted from an ion-exchange column by FPLC using a 50−500 mM NaCl gradient. Aliquots (10−50 μl) of the 0.5 ml fractions were separated by SDS−PAGE on a 7.5% gel and proteins were identified by silver staining (upper part of figure). Band-shift activity using the protein-binding oligonucleotide WT was tested with 1 μl aliquots of the column fractions indicated (lower part of figure). Molecular mass standards (M) are shown to the left. The position of the BAF1 protein is indicated by an arrow.

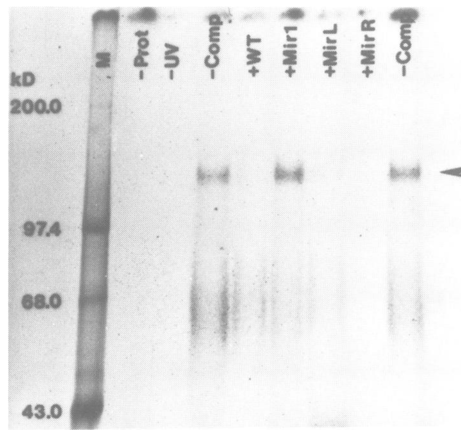



**Fig. 4.** UV-crosslinking of protein to a sequence segment (oligonucleotide WT) of the *YPT1/TUB2* intergene region. BUdR-substituted, labelled oligonucleotide WT not complexed with protein (−Prot) or bound to affinity-purified protein in the absence (−Comp) or in the presence of a 100-fold molar excess of different competitors (see Figure 2A) were UV-irradiated and, after digestion of the DNA, the proteins were separated by SDS−PAGE. The arrow points to the labelled protein of ~120 kd identified in the specific complexes. The protein is not visible in identically treated but unirradiated samples (−UV). M, molecular mass standards.

the GC base pair in position −176 with either a CG base pair (oligonucleotide Mir 2, Figure 2A) or an AT base pair (oligonucleotide Mir 3, Figure 2A) rendered the mutant DNA fragments incapable of effectively competing with the wild-type DNA fragment for protein binding (Figure 2B). The GC to AT transition (position −176) was one of several other mutations generated in fragment Mir 1 (Figure 2A), an oligonucleotide originally designed to interfere with the alternation of purine and pyrimidine nucleotides. As expected, this DNA fragment was highly ineffective in forming specific complexes with the affinity-purified protein (Figure 6B), and it did not compete with the wild-type fragment in complex formation (Figure 2B).

The elucidation by methylation interference of apparent protein contact sites within the DNase I footprint allowed the core sequence involved in BAF1 protein binding to be delineated. To characterize further the binding region, we introduced several mutations into the core sequence and tested the synthetic fragments for their ability to bind the affinity-purified BAF1 protein (Figures 2A and 6A). The duplication of five base pairs (−179 to −183, fragment I1PM2) as well as the deletion of one base pair (position −182, fragment I1PM3) between the critical protein contact-forming G residues rendered the DNA fragments unable to bind the protein. Whereas an AT to GC transition at position −178 (fragment I1PM5) resulted in the loss of binding activity, the CG to TA transition at position −186 (fragment I1PM4), which removed one of the apparent contact sites, resulted in a significant drop of the relative binding affinity (Figure 6A).

### *BAF1 has ARS1-binding activity*

By comparing the binding region of BAF1 with sequences recognized by other yeast DNA-binding proteins, we noted a remarkable similarity to the consensus recognition sequence of factor ABF1 (SBF-B) which has been shown to bind to a specific domain of autonomously replicating sequences (*ARS*) of the silent mating type loci *HMRE, HMR1* and

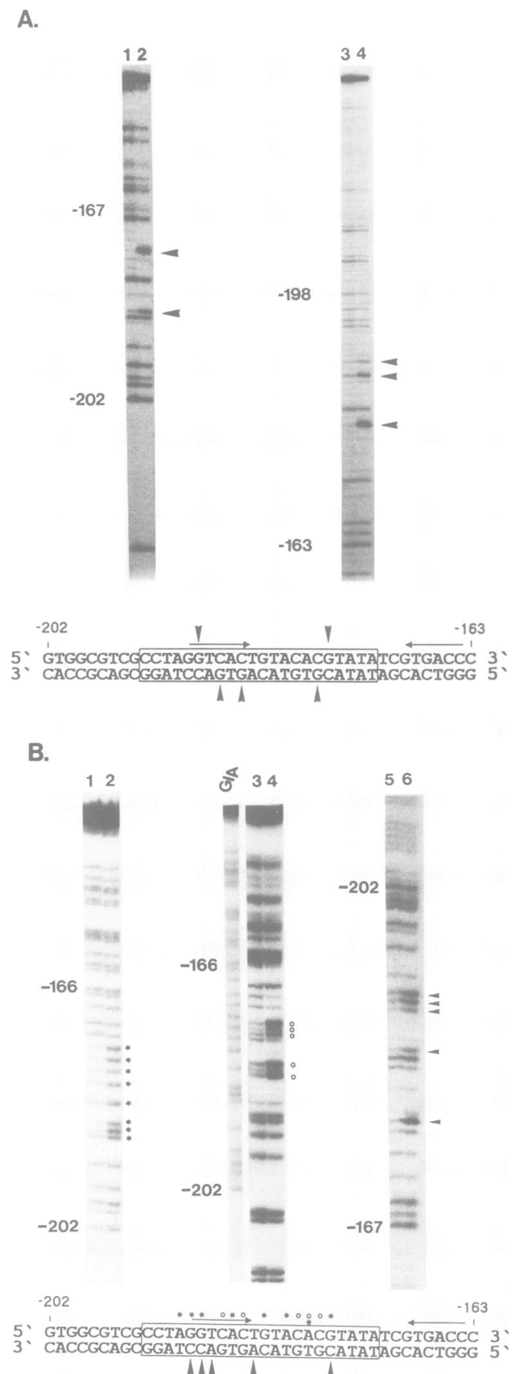



**Fig. 5.** Methylation interference (**A**) and missing contact probing (**B**) of DNA−BAF1 complexes. (**A**) The 142 bp *Bgl*II−*Nco*I fragment, 5′ end-labelled at either its *Bgl*II (**lanes 1** and **2**) or its *Nco*I restriction site (**lanes 3** and **4**) and partially methylated with DMS, was complexed with DNA affinity-purified protein and subjected to agarose gel electrophoresis to separate protein-bound and free DNA. After piperidine treatment, the fragments generated from free DNA (**lanes 2** and **4**) and from protein-bound DNA (**lanes 1** and **3**) were separated on a 6% sequencing gel. Apparent protein contact sites (G bases) are indicated by arrows on the autoradiogram and in the DNA sequence below. The protein-binding region identified by DNase I-footprinting analysis is framed. (**B**) After partial removal of bases (G + A, **lane 2**; C + A, **lanes 4** and **6**) from the 142 bp *Bgl*II−*Nco*I fragment, 5′ end-labelled at either the *Bgl*II (**lanes 1−4**) or *Nco*I restriction site (**lanes 5** and **6**), complexes were formed with BAF1 protein (**lanes 1**, **3** and **5**). Protein-associated and free DNA (**lanes 2**, **4** and **6**) were treated and analysed as in (**A**). Missing bases apparently interfering with protein binding are marked on the autoradiogram and in the sequence below.

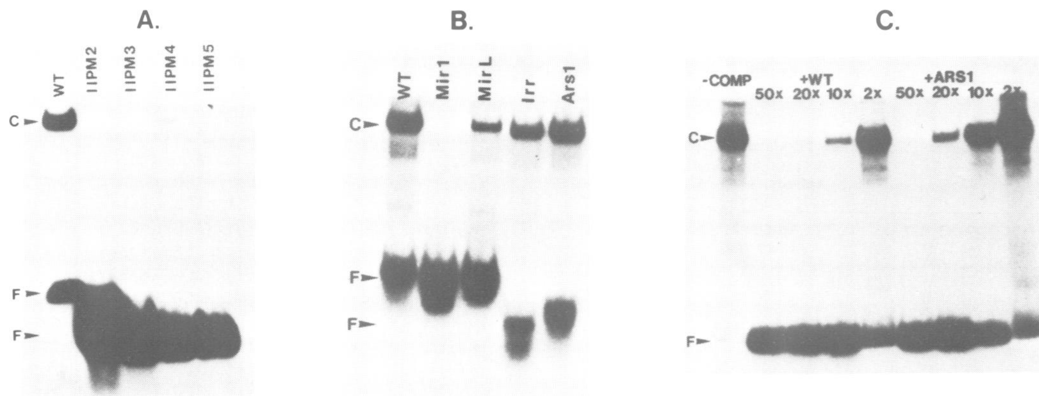



**Fig. 6.** (**A**,**B**) Gel-retardation analysis of complexes formed with DNA affinity-purified BAF1 protein and synthetic oligonucleotides representing either the protein-binding segment of the *YPT1/TUB2* intergene region (oligonucleotide WT) and several mutated derivatives (Mir 1, Mir L, Irr and I1PM2–5) or the *ARS1* sequence known to bind factor ABF1 (Buchman *et al.*, 1988; Diffley and Stillman, 1988) (for sequences, see Figure 2A). Note that mutations (Mir L and I1PM4) or the complete lack of the left-sided arm of the symmetrical region residing between the *YPT1* and*TUB2* gene (Irr) did not completely abolish complex formation. (**C**) Competition of BAF1 protein/oligonucleotide WT complex formation with 2-, 20-, 50- and 200-fold molar excess of unlabelled fragment (WT) or of *ARS1* sequences (see Figure 2A). C, complexed DNA; F, free DNA.

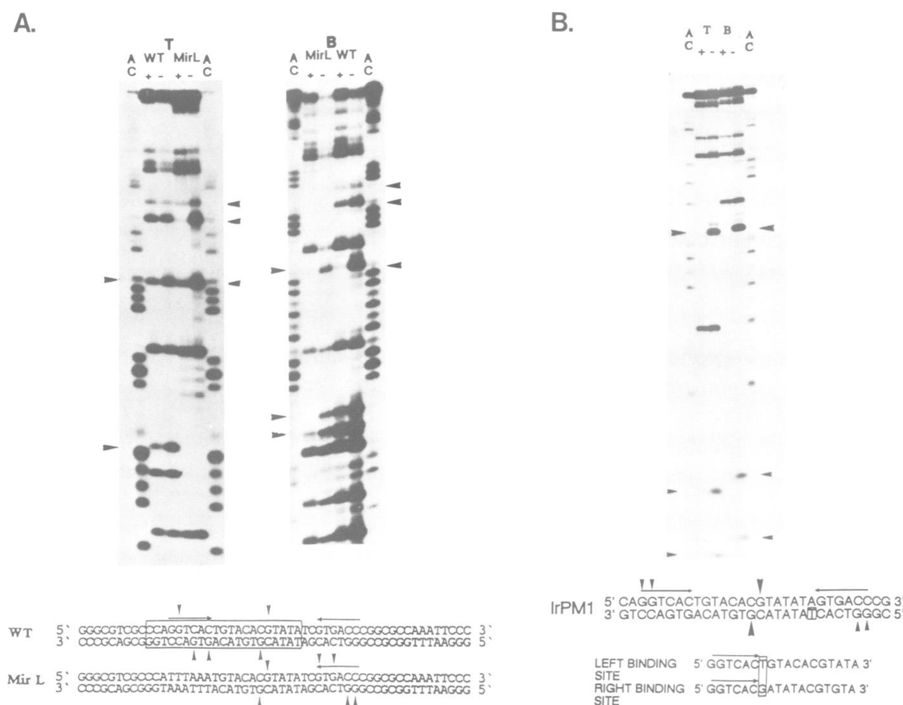
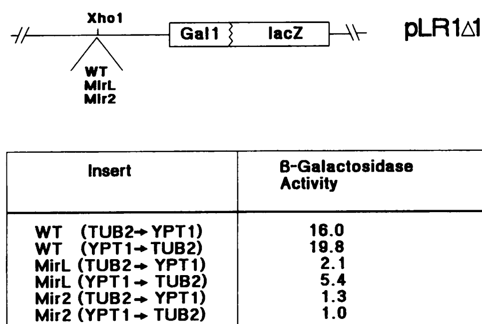



**Fig. 7.** Identification of methylation interference of two overlapping BAF1-binding sites. (**A**) Apparent protein contact sites at the top (T) and the bottom (B) strands of the wild-type sequence (WT) and the mutant sequence (Mir L), with several base changes in the high-affinity binding site, are marked with short arrows. The DNase I-protected region of the wild-type sequence is boxed. (**B**) The GC to TA transversion of the right BAF1-binding site (boxed in the sequence of fragment IrPM1) makes this site similarly efficient for protein binding as the strong binding site. Apparent contact sites indicated by different intensities of fragments generated by chemical degradation of premethylated protein-bound (+) and free DNA (−) are marked with short, thick arrows. A + C sequencing reactions of the same fragments were used for comparison. The residues, which seem to affect critically BAF1 protein-binding efficiency, are boxed in the conserved sequence motifs and shown below the autoradiogram.

*HML1* and of *ARS1* located near the *TRP1* gene (Shore *et al.*, 1987; Buchman *et al.*, 1988; Diffley and Stillman, 1988). Of this consensus sequence, 5′-TATCATTN$_4$A-CGA-3′ (Buchman *et al.*, 1988), which is also present and recognized by ABF1 near several polymerase II-transcribed yeast genes, the sequence 5′-TCN$_7$ACG-3′ is highly conserved in all binding regions identified, and it is likewise contained in the BAF1-binding sequence midway between the *YPT1* and the *TUB2* genes. Indeed, this sequence motif is present on both strands of the palindromic region with the 3′ located and conserved C and G residues overlapping at the axis of symmetry (Figure 1A).

To test whether the BAF1 protein binds to the ABF1-binding *ARS1* region (Diffley and Stillman, 1988), a 32 bp double-stranded DNA fragment of the protein-binding *ARS1* domain B (Figure 2A) was synthesized and used as a competitor in a band shift assay with BAF1 protein and the labelled binding fragment of the *YPT1/TUB2* intergene region. As shown in Figure 6C, the *ARS1* fragment successfully competed in this binding assay although some-




| Insert | ß-Galactosidase Activity |
|---|---|
| WT (TUB2→YPT1) | 16.0 |
| WT (YPT1→TUB2) | 19.8 |
| MirL (TUB2→ YPT1) | 2.1 |
| MirL (YPT1→ TUB2) | 5.4 |
| Mir2 (TUB2→ YPT1) | 1.3 |
| Mir2 (YPT1→ TUB2) | 1.0 |

**Fig. 8.** Transcription activation mediated by the BAF1-binding sequence element. Plasmid pLR1Δ1 (West *et al.*, 1984) harbouring a *GAL1–lacZ* fusion gene with a deletion of the upstream activating sequence ($UAS_G$) was used to monitor the ability of the BAF1-binding element (WT) and two mutant derivatives (Mir L and Mir 1) to activate transcription. The synthetic DNA fragments (for sequences, see Figure 2A) were inserted in both orientations into the *Xho*I restriction site located at the point of $UAS_G$ deletion. The orientation of the inserted fragments, given in parentheses in the lower part of the figure, is represented by arrows indicating the 5′ → 3′ direction of the strong $TCN_7ACG$ recognition sequence with respect to the *TUB2* and *YPT1* gene. Units of β-galactosidase activity in yeast transformants grown in glucose-containing minimal medium were determined according to Guarente and Ptashne (1981); activity measured in transformants carrying plasmid pLR1Δ1 without BAF1-binding elements (<0.5 units) was subtracted from values obtained with other constructions.

what less efficiently than unlabelled WT oligonucleotide. As expected, the *ARS1* DNA fragment could in addition be shown to form a specific complex with the purified BAF1 protein (Figure 6B).

### *Two overlapping BAF1-binding regions in the YPT1/TUB2 intergene region exhibit different apparent affinities for protein binding in vitro*

As pointed out in the preceding paragraph, the region of dyad symmetry in the *YPT1/TUB2* intergene region contains two sequence motifs conforming to the ABF1-binding consensus sequence (Buchman *et al.*, 1988). These sequences have opposite orientation and overlap with residues −176 and −177 in the centre of the symmetrical sequence (Figure 1A). The DNAse I-protection experiment had clearly shown that the conserved sequence motif of the upper strand only was protected by protein binding (Figure 1B). Mutation of several nucleotides of the left-sided inverted repeat sequence (nucleotides −189 to −184) within the DNase I-footprint area, however, did not abolish the formation of specific complexes with the BAF1 protein (Figure 6B). This finding led us to investigate whether the oppositely oriented conserved sequence now became accessible for protein binding. A 24 bp double-stranded fragment (Irr, Figure 2A), lacking a significant part of the symmetrical sequence protected from DNase I digestion by the bound BAF1 protein (Figure 1B), was still able to form a specific complex with this protein (Figure 6B). In a methylation interference experiment with fragment Mir L it became evident that the BAF1 protein now bound to the conserved sequence element of the symmetrical region facing the *YPT1* gene, apparently contacting G residues −167, −169 and −176 of the upper strand (Figure 7A).

This finding strongly suggests that the region of dyad symmetry contains two BAF1-binding sites that exhibit significantly different affinities for this protein. A comparison

| Localisation | Protein Binding Sequence (DNaseI-Footprint) | Designation of Protein | Molecular Mass (kD) | Source |
|---|---|---|---|---|
| ARS1 | TTTCTTAGCATTTTTTGACGAAATT | ABF1,SFB-B | 135 | 1,2,3,4 |
| HML I | ATATCATTGCAAACGTTCGGGCT | ABF1 | | 2 |
| HMR I | TATCGCCATATACGAAAATG | ABF1 | | 2 |
| HMR E | CAATACATCATAAAATACGAACGA | ABF1,SFB-B | | 1,2 |
| 2μ ARS | TAGCATATCTTTGTTAACGAAGCATC | ABF1 | | 1 |
| ARS120 | ACCCACATCATTATGCACGGCACTTG | OBF1 | | 5 |
| DED1′ (5′) | TAATGCATCATTCTATACGTGTCATT | ABF1 | | 2 |
| DED1″ (5′) | ATACGTGTCATTCTGAACGAGGCGCT | ABF1 | | 2 |
| CIII-SubII (5′) | CTGATCATTCCCAACGAACCAATAG | GF1 | | 6 |
| CIII-SubVII (5′) | CTAGTCACGTGGAACGGT | GF1 | | 6 |
| CYC1 (5′) | GATGTCGTCTCACACGGA | GF1 | | 6 |
| PHO5 (3′) | GCATCGTTAATGACGTAT | GF1 | | 6 |
| Ty2 (5′) [1)] | GTCATCATAGACG | TyBF | | 7 |
| TCM1 (5′) | GTAATCGTTTTGTACGTTTTTCAAG | TAF | 147 | 8 |
| TUB2/YPT1 (5′) | CCTAGGTCACTGTACACGTATA | BAF1 | 120 | 9 |
| YPT1/TUB2 (5′) | GGTCACGATATACG | BAF1 | 120 | 9 |
| Consensus | RTCRYYNNNNACG | | | |

**Fig. 9.** Compilation of *S.cerevisiae* protein-binding elements containing the conserved $TCN_7ACG$ sequence. With the exception of the Ty2 and the low-affinity *YPT1/TUB2* recognition sequence, which were analysed by the methylation interference technique only, all other sequences represent the protein-binding region as revealed by DNase I protection. The strictly conserved TC and ACG residues are shown in large bold letters. The location of the recognition elements, the designation and, where determined, the approximate molecular mass of the binding proteins are also given. R, purine; Y, pyrimidine. References: (1) Shore *et al.* (1987); (2) Buchman *et al.* (1988); (3) Diffley and Stillman (1988); (4) Sweder *et al.* (1988); (5) Eisenberg *et al.*, 1988; (6) Dorsman *et al.* (1988); (7) Goel and Pearlman (1988); (8) Hamil *et al.* (1988); (9) this paper.

of the two binding regions shows that the position three residues 3′ of the conserved TC dinucleotide is occupied by by a pyrimidine in the strong but by a purine in the weak binding site (Figure 7B). As almost all protein binding sites described that conform to the $TCN_7ACG$ consensus sequence contain a pyrimidine residue in the same position (Figure 9), we substituted the G of the weak binding site for a T residue in a 30 bp oligonucleotide (IrPM1). This DNA fragment was used for protein binding and methylation interference experiments. As can be seen in Figure 7B, methylation of guanine −176 of the top strand and of guanine −177 of the bottom strand clearly interfered with the formation of specific DNA–protein complexes. This was expected since the C and G residues of the conserved trinucleotides ACG of the two oppositely oriented binding regions overlap. Most importantly, complex formation appeared similarly affected by methylation of guanines −164/−165 and −188/−189 and, less pronounced, of guanines −167/−169 and −184/−186. This suggests that the G to T transversion (position −170, bottom strand; see Figure 7B) increased the affinity for BAF1 protein of the right binding region such that both binding sites became equally efficient. Why the methylation of guanine −189 of fragment IrPM1 (Figure 7B) but not of fragment WT (Figure 7A) interfered with complex formation is not known, but it might be due to the shorter IrPM1 fragment which does not include the entire sequence protected in a DNase I-footprinting assay (Figure 1B).

### *The BAF1-binding sequence element activates transcription in both directions*

Earlier experiments in our laboratory had already suggested that the deletion of the region of dyad symmetry inhibited

the transcription of both the β-tubulin and the *YPT1* gene (Donath, 1985). To prove the transcription-activating property of the sequence element encompassing both BAF1-binding regions, the 53 bp 'wild-type' fragment WT or the mutated fragments Mir L and Mir 1 (see Figure 2A) were inserted in place of the $UAS_G$ 5′ of the *GAL1–lacZ* fusion gene of vector pLR1Δ1 (West *et al.*, 1984). Independent of its orientation, the WT fragment led to the activation of transcription of the fusion gene in transformed yeast cells as judged by β-galactosidase activity (Figure 8). Disruption of the left-sided repeat structure (fragment Mir L) as well as a substitution of A for G in position −176 (Mir 3), shown above to be a contact site for the BAF1 protein, led to a drop of transcriptional activity. Not surprisingly, the interference with transcription activation was significantly more pronounced by the point mutation in position −176 as this affected both binding regions.

## Discussion

The suspicious location and structure of a 26 bp sequence of nearly perfect dyad symmetry midway between the divergently transcribed *YPT1* and *TUB2* genes encoding a *ras*-like, GTP-binding protein and β-tubulin respectively (Gallwitz *et al.*, 1983), and the previous finding that the deletion of part of the *YPT1/TUB2* intergene region encompassing this symmetrical sequence inhibits the transcription of the two adjacent genes (Donath, 1985), led us to investigate more thoroughly its possible role as an upstream activating sequence.

Our results show that a protein with a molecular mass of ~120 kd binds *in vitro* to part of this symmetrical sequence. The region protected in a DNase I-footprint experiment contains a conserved sequence element 5′-$TCN_7ACG$-3′, first noted by Buchman *et al.* (1988) to be included in DNA segments recognized by a protein called ABF1 or SFB-B (Shore *et al.*, 1987) at *ARS* sequences of the yeast mating type loci *HMR1*, *HMRE* and *HML1*, within *ARS1* and the 2 μm plasmid *ARS* as well as 5′ upstream of the *DED1* gene. The identical sequence motif is present in protein-binding sites recently identified in the 5′ flanking region of several *S.cerevisiae* nuclear genes (Dorsman *et al.*, 1988; Hamil *et al.*, 1988), within the *ARS120* region (Eisenberg *et al.*, 1988) and the 5′ region of the Ty2 element (Goel and Pearlman, 1988). All of these binding sites, including those within the *YPT1/TUB2* intergene region described here, were independently discovered and different names have been proposed for the binding protein(s). A compilation of DNA sequences containing the conserved sequence motif and shown by DNase I footprint and methylation interference analysis to interact with protein(s) is presented in Figure 9. Taking into account the larger collection of available binding sites, the consensus 5′-$PuTCPuPyPyN_4ACG$-3′ can be deduced, which is a modified version of the recognition sequence proposed by Buchman *et al.* (1988) and Dorsman *et al.* (1988).

Although different possible functions have been assigned to these binding proteins, it might be that it is the same protein which recognizes all of these different sequences. It is also possible, of course, that members of a family of functionally distinct proteins bind to the $TCN_7ACG$ recognition sequences *in vitro*. A hint to the existence of different proteins binding to the same sequence motif, but with different affinities, comes from the observation reported by Eisenberg *et al.* (1988), that a protein called OBF1 binds to an *ARS120* element (conforming to the ABF1 consensus sequence) even in excess of a 40-fold molar excess of ABF1-(GF1-,BAF1-)binding *ARS1* or *HMRE ARS*. An example of different yeast transcriptional activators binding to the same DNA sequence element are the members of the jun family, GCN4 (Hope and Struhl, 1985) and yAP-1 (Moye-Rowley *et al.*, 1989).

The location of $TCN_7ACG$ binding sites at *ARS* elements, including those of the silent mating-type loci which also mediate *SIR*-dependent transcriptional repression, have led to the speculation that protein binding to this conserved sequence might be of importance for DNA replication and silencer function (Shore *et al.*, 1987; Buchman *et al.*, 1988; Diffley and Stillman, 1988). Evidence for these functions has recently been obtained in an analysis with a mutated ABF1-binding site at the *HMRE* silencer (Kimmerly *et al.*, 1988). The occurrence of $TCN_7ACG$ elements in regions 5′ or 3′ of protein-coding genes devoid of *ARS* activity was taken to suggest an additional role of this sequence in transcription activation and termination (Buchman *et al.*, 1988; Dorsman *et al.*, 1988). Besides our own results presented here, two recent reports clearly show that sequence elements conforming to the ABF1-binding site associate with a *trans*-acting protein *in vitro* and function as upstream activating sequences *in vivo*: transcription activation of genes adjacent to a Ty2 element (Goel and Pearlman, 1988) as well as activation of the ribosomal protein gene *TCM1* (Hamil *et al.*, 1988) was demonstrated to be dependent on a $TCN_7ACG$ binding site in Ty2 and in the 5′ flanking region of the *TMC1* gene respectively. In both of these analyses, the exchange of one of the strictly conserved nucleotides, $TCN_7ACG$ → $TCN_7ATG$ (Goel and Pearlman, 1988) and $TCN_7ACG$ → $ACN_7ACG$ (Hamil *et al.*, 1988), negatively affected factor binding *in vitro* and transcription activation *in vivo*.

Of special interest is the sequence arrangement within the *YPT1/TUB2* intergene region where two oppositely oriented $TCN_7ACG$ sequence elements with significantly different affinities for a protein, which we have called BAF1, overlap partially. Both *YPT1* and *TUB2* are constitutively expressed and essential genes. Overexpression of β-tubulin interferes with cell viability (Schatz *et al.*, 1986) and high expression of the YPT1 protein from plasmid-borne *YPT1* genes has been observed to be regulated against (Schmitt *et al.*, 1986). It thus appears that the transcription of the two genes is regulated in some way to furnish the cell with the appropriate amount of gene products. This could be achieved via the BAF1 protein-binding region in the context of neighbouring sequences. As *in vitro* the $TCN_7ACG$ sequence facing the *TUB2* gene only was found to bind the *trans*-acting protein, it remains to be seen whether *in vivo* both binding regions can associate with the BAF1 protein. It is clear, however, that after mutating the strong binding site, the low-affinity binding region facing the *YPT1* gene binds BAF1 protein *in vitro* and activates transcription of an adjacent *GAL1–lacZ* fusion gene *in vivo*. Of the different activities that have been ascribed to this conserved sequence element, the only function for the BAF1-binding region described here appears to be transcription activation. A fragment encompassing the entire *YPT1/TUB2* intergene region and short segments of the protein-coding part of both genes was found to be devoid

of *ARS* activity (unpublished results of the Göttingen laboratory). As the deletion of the BAF1-binding DNA segment from the *YPT1/TUB2* intergene region inhibits the transcription of the two genes (Donath, 1985) and as the same segment activates the transcription in an orientation-independent fashion, it seems justified to attribute to the BAF1 protein a bidirectionally activating function. It is this function that led us to the designation of the *trans*-acting factor as BAF1 protein.

As revealed by the methylation interference approach, the three guanine residues of the stringently conserved CG or GC base pairs of the $TCN_7ACG$ sequence seem to be of importance for the interaction with the *trans*-acting factors, TyBF (Goel and Pearlman, 1988) and BAF1 (this report) alike. This assumption is strengthened by the results obtained with mutant binding sites which were changed at three positions and exhibited a significantly lower relative protein-binding affinity. We also observed that the spacing between the highly conserved TC and ACG residues of the binding sequence is critical for the interaction with BAF1 protein. Not only the extension but also the shortening of the distance between TC and ACG by one or several nucleotides interfered with protein binding. A similar observation was made by Kimmerly *et al.* (1988) who showed that *ARS* activity is lost following the insertion of one base pair between the TC and ACG of the ABF1-binding region at the *HMRE* silencer.

The apparent molecular mass of 120 kd which we determined for the purified BAF1 protein is slightly different from the 135 kd estimated for the *ARS1*-binding ABF1 protein (Diffley and Stillman, 1988; Sweder *et al.*, 1988) and the 147 kd calculated for the $UAS_T$-binding TAF protein (Hamil *et al.*, 1988). Whether indeed all these proteins are functionally and structurally distinct species binding to DNA segments containing the conserved $TCN_7ACG$ motif can only be answered after the proteins and the genes encoding them have been characterized.

## Materials and methods

### DNA probes

Oligonucleotides were synthesized on a Biosearch 8700 DNA synthesizer and purified either by HPLC or PAGE. For footprint analysis and mapping of protein contact sites, a 142 bp *Bgl*II/*Nco*I fragment was 5′ end-labelled with T4 polynucleotide kinase at either the *Bgl*II or the *Nco*I restriction site. For UV crosslinking, the 53 bp fragment (nucleotides −150 to −200 of the *YPT1/TUB2*V intergene region) was recovered by *Sma*I cutting from plasmid pIRWT which contained multiple copies of the respective fragment.

### Isolation of the BAF1 protein

Two hundred and fifty grams (wet weight) of *S.cerevisiae* strain ABYS1 (a *pra1 prb1 prc1 cps1 ade*), kindly provided by D.H.Wolf, Freiburg, were suspended in 200 ml of a buffer consisting of 0.2 M Tris−Cl, pH 8.0, 10 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 0.3 M $(NH_4)_2SO_4$, 20% glycerol (v/v) and 1 mM PMSF. After adding an equal volume of glass beads (0.45 mm diameter), the cells were disrupted in glass vessels chilled with $CO_2$ using a cell grinder (Braun, Melsungen, FRG). The glass beads were removed by filtration and the homogenate was centrifuged at 100 000 *g* for 1 h. Proteins of the supernatant were precipitated with $(NH)_2SO_4$ (0.35 g/ml of protein solution) and, after thorough dialysis against buffer A50 [20 mM Tris−Cl, pH 8.0, 1 mM EDTA, 1 mM DTT, 50 mM $(NH_4)_2SO_4$, 10% glycerol, 1 mM PMSF], the precipitated proteins were applied to a 150 ml column of heparin−agarose (Sigma) equilibrated with buffer A50 and fractionated with a linear 50−600 mM $(NH_4)_2SO_4$ gradient in buffer A50. Aliquots of fractions obtained were analysed for binding activity using the band-shift assay. Active fractions were pooled and, after precipitating the proteins with $(NH_4)_2SO_4$, they were desalted on a PD10 column (Pharmacia) equilibrated with buffer B50 (20 mM Tris−Cl, pH 8.0, 1 mM EDTA, 1 mM DTT, 10% glycerol, 50 mM NaCl). Proteins were then chromatographed either on a calf thymus DNA column (Pharmacia); or they were separated by HPLC on a DEAE column (Spherogel TSK DEAE PW, 7.5 × 75 mm) using a linear gradient of 50−800 mM NaCl in buffer B (B50 lacking NaCl). Fractions containing proteins active in the band-shift assay and eluting at salt concentrations >0.3 M were pooled and diluted to 0.2 M NaCl. Pooled fractions were applied to a specific DNA column consisting of the synthetic double-stranded oligonucleotide 5′-CGTCGCCCAGGTCACTGTACACGTATAT-3′ and its complementary strand, multimerized to DNA molecules of 0.2−1.5 kd in size and coupled to activated Sepharose 4B (Kadanoga and Tjian, 1986). After protein binding, the column was extensively washed with buffer B200 (buffer B containing 200 mM NaCl), and bound proteins were eluted with a linear 0.2−1.5 M NaCl gradient. Fractions containing the DNA-binding factor (recovered between 0.4 and 1 M NaCl) were diluted to 0.1 M NaCl with buffer B, loaded on a Mono-Q HR5 column (Pharmacia) and proteins were fractionated by FPLC with a salt gradient resulting in a highly purified, active BAF1 protein, which was eluted with ~0.3 M NaCl.

### Band-shift assay

Tests for protein binding to DNA were carried out in buffer B100 (buffer B containing 100 mM NaCl) in a final volume of 20 μl. Proteins (1−4 μl of column fractions) were first incubated with 0.7−1 μg of poly d(IC) (Boehringer Mannheim) for 5 min at room temperature. Following the addition of the 5′ end-labelled fragments (10 000−15 000 c.p.m.), the incubation was continued for a further 10 min, after which the assay mixture was immediately loaded onto a 4% native polyacrylamide gel in 50 mM Tris−borate, pH 8.3, 1 mM EDTA, for electrophoretic separation of free and protein-bound DNA. Gels were dried and subjected to autoradiography for 2−12 h. For competition assays, the protein was pre-incubated with an excess of competitor DNA and 1 μg of poly d(IC). After 5 min, the assay mixture was treated as above.

### DNase I-protection experiments

Following 10 min in incubation of heparin−agarose-purified binding protein with the 142 bp *Bgl*II−*Nco*I fragment (5′ end-labelled on its *Bgl*II restriction site) under conditions described for the band-shift assay, $MgCl_2$ and $CaCl_2$ were added to final concentrations of 10 and 0.5 mM respectively. Two nanograms of DNase I (Worthington) were added to the incubation mixture and, after 10−30 s at room temperature, the action of the enzyme was stopped by adding 100 μl of 12.5 mM EDTA and 100 μl of 10 mM Tris−Cl, pH 7.5, 1 mM EDTA (TE buffer). The DNA fragments were subsequently extracted with phenol and chloroform, precipitated with ethanol and separated on an 8% polyacrylamide gel under denaturing conditions.

### Methylation interference

The 142 bp *Bgl*II−*Nco*I fragment, 5′ end-labelled at one of the two restriction sites, was treated with dimethylsulphate (DMS) as described by Maxam and Gilbert (1980). The methylated DNA fragment was incubated with affinity-purified binding protein as described, and the complexed DNA was separated from free DNA by electrophoresis in a native polyacrylamide gel. The gel was exposed to X-ray film (Kodak X-Omat), the bands corresponding to bound and unbound DNA were cut out of the gel and the DNA was electrophoretically eluted in a Biotrap elution chamber (Schleicher and Schuell), extracted with phenol, treated with piperidine and analysed on a 7.5% polyacrylamide gel under denaturing conditions.

### Missing contact probing

Approximately 200 ng of the 142 bp *Bgl*II−*Nco*I fragment, 5′ end-labelled at either the *Bgl*II restriction site (GA pre-modification reaction) or at its *Nco*I cutting site (C/A and G/A pre-modification reaction) were partially depurinated or depyrimidated according to Maxam and Gilbert (1980) using a kit (Merck, Darmstadt) and following the instructions of the suppliers. In a total volume of 35 μl containing 25 mM Hepes, pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM DTT and 2.5 μg poly d(IC), 20 ng of pre-modified DNA fragment and 3 μg of crude extract protein, active in band-shift assay and DNase I-footprint analysis, were allowed to form complexes for 10 min at room temperature. Bound and unbound DNA were separated on a 6% native polyacrylamide gel, electroeluted, treated with piperidine and subjected to a 6% sequencing gel.

### UV crosslinking

The DNA probe was generated from plasmid pIRWT by primed synthesis in the presence of 50 μCi [$\alpha$-$^{32}$P]dCTP (3000 Ci/mmol) and 50 μM each of dATP, dGTP and 5-bromodeoxyuridine triphosphate according to the method of Chen and Seeburg (1985). The labelled fragment was obtained by *Sma*I digestion of the plasmid and gel purification, and 20 000 c.p.m.

of fragment were incubated with DNA affinity-purified protein under binding assay conditions described above. After 10 min of incubation, the mixtures were spotted onto Saranwrap and irradiated for 30 min with UV light (256 nm) at room temperature. Following irradiation, the reaction mixtures were transferred back to Eppendorf tubes and treated with 2 ng of DNAse I for 10 min at 37°C in the presence of 1 mM $MgCl_2$ and 5 mM $CaCl_2$. After stopping the reaction with EDTA at a final concentration of 10 mM, the samples were precipitated with 10% trichloroacetic acid and the proteins were analysed by SDS–PAGE according to Laemmli (1970).

#### *Analysis of transcription activation*

To analyse the potential for transcription activation of the BAF1-binding sequence, plasmid pLR1Δ1 (West *et al.*, 1984), kindly provided by R.Brent (Boston), was cut with *Xho*I, the overhanging ends were filled in with Klenow enzyme and different non-phosphorylated, synthetic, double-stranded fragments were inserted by blunt-end ligation. The orientation of inserted fragments was checked by DNA sequencing. *Saccharomyces cerevisiae* strain YNN27 (*trp1-289 ura3-52 gal2*) was transformed by the LiCl method (Ito *et al.*, 1983). Positive transformants were grown in minimal medium containing glucose to a density of ~1.0 at $OD_{600}$ and analysed for β-galactosidase activity following the protocol of Guarente and Ptashne (1981) and Miller (1972).

## Acknowledgements

We are indebted to Roger Brent (Boston) for providing plasmid pLR1Δ1. We thank Hans-Peter Geithe for synthesizing oligonucleotides, Peter Mienkus for technical assistance and Ingrid Balshüsemann for secretarial help. This work was supported by grants from Fonds der Chemischen Industrie (to D.G.) and Deutsche Forschungsgemeinschaft (to E.W.).

## References

Achstetter,T., Emter,O., Ehmann,C. and Wolf,D.H. (1984) *J. Biol. Chem.*, **259**, 13334–13343.
Bram,R.J. and Kornberg,R.D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 43–47.
Brunelle,A. and Schleif,R.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6673–6676.
Buchman,A.R., Kimmerly,W., Rine,J. and Kornberg,R.D. (1988) *Mol. Cell. Biol.*, **1**, 210–225.
Chen,E.J. and Seeburg,P.H. (1985) *DNA*, **4**, 165–170.
Diffley,J.F.X. and Stillman,B. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2120–2124.
Donath,C. (1985) *Ph.D thesis, University of Marburg, FRG.*
Dorsman,J.C., van Heeswijk,W.C. and Grivell,L.A. (1988) *Nucleic Acids Res.*, **16**, 7287–7301.
Eisenberg,S., Civalier,C. and Tye,B.-T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 743–746.
Fried,M. and Crothers,D.M. (1981) *Nucleic Acids Res.*, **9**, 6505–6524.
Galas,D. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
Gallwitz,D., Donath,C. and Sander,C. (1983) *Nature*, **306**, 704–707.
Giniger,E., Varnum,S.M. and Ptashne,M. (1985) *Cell*, **40**, 767–774.
Goel,A. and Pearlman,R.E. (1988) *Mol. Cell. Biol.*, **8**, 2572–2580.
Guarente,L. and Ptashne,M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2199–2203.
Hamil,K.G., Nam,H.G. and Fried,H.M. (1988) *Mol. Cell. Biol.*, **8**, 4328–4341.
Hope,I. and Struhl,K. (1985) *Cell*, **43**, 177–188.
Huet,J. and Sentenac,A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3648–3652.
Ito,H., Fukuda,Y., Murat,K. and Kimura,A. (1983) *J. Bacteriol.*, **153**, 163–168.
Johnston,M. and Davis,R.W. (1984) *Mol. Cell. Biol.*, **4**, 1440–1448.
Kadonaga,J.T. and Tjian,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5993.
Kimmerly,W., Buchman,A., Kornberg,R. and Rine,J. (1988) *EMBO J.*, **7**, 2241–2253.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Miller,J.H. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Moye-Rowley,W.S., Harshman,K.D. and Parker,C.S. (1989) *Genes Dev.*, **3**, 283–292.
Neff,N.F., Thomas,J.H., Grisafi,P. and Botstein,D. (1983) *Cell*, **33**, 211–219.
Ogata,R. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4973–4976.
Schatz,P.J., Solomon,F. and Botstein,D. (1986) *Mol. Cell. Biol.*, **6**, 3722–3733.
Schmitt,H.D., Wagner,P., Passf,E. and Gallwitz,D. (1986) *Cell*, **47**, 401–412.
Shore,D. and Nasmyth,K. (1987) *Cell*, **51**, 721–732.
Shore,D., Stillman,D.J., Brand,A. and Nasmyth,K.A. (1987) *EMBO J.*, **6**, 461–467.
Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 122–126.
Sorger,P.K. and Pelham,H.R. (1987) *EMBO J.*, **6**, 3035–3041.
Sweder,K.S., Rhode,P.R. and Campbell,J.L. (1988) *J. Biol. Chem.*, **263**, 17270–17277.
Treisman,R. (1987) *EMBO J.*, **6**, 2711–2717.
West,R.W., Yocum,R.R. and Ptashne,M. (1984) *Mol. Cell. Biol.*, **4**, 2467–2478.
Wu,C., Wilson,S., Walker,B., Dawid,I., Paisley,T., Zimarino,V. and Ueda,H. (1987) *Science*, **238**, 1247–1252.

*Received on May 9, 1989; revised on June 28, 1989*