Yeast general transcription factor GFI: sequence requirements for binding to DNA and evolutionary conservation

J.C.Dorsman, W.C.van Heeswijk and L.A.Grivell*
Section for Molecular Biology, Department of Molecular Cell Biology, University of Amsterdam, Kruislaan 318, 1098 SM, Amsterdam, The Netherlands

Received November 17, 1989; Revised and Accepted March 19, 1990

ABSTRACT

GFI is an abundant DNA binding protein in the yeast *S.cerevisiae*. The protein binds to specific sequences in both ARS elements and the upstream regions of a large number of genes and is likely to play an important role in yeast cell growth. To get insight into the relative strength of the various GFI-DNA binding sites within the yeast genome, we have determined dissociation rates for several GFI-DNA complexes and found them to vary over a 70-fold range. Strong binding sites for GFI are present in the upstream activating sequences of the gene encoding the 40 kDa subunit II of the QH2:cytochrome *c* reductase, the gene encoding ribosomal protein S33 and in the intron of the actin gene. The binding site in the ARS1-TRP1 region is of intermediate strength. All strong binding sites conform to the sequence 5' RTCRYYYNNNACG-3'. Modification interference experiments and studies with mutant binding sites indicate that critical bases for GFI recognition are within the two elements of the consensus DNA recognition sequence. Proteins with the DNA binding specificities of GFI and GFII can also be detected in the yeast *K. lactis*, suggesting evolutionary conservation of at least the respective DNA-binding domains in both yeasts.

INTRODUCTION

Several abundant sequence specific DNA -binding proteins have recently been identified in the yeast *S.cerevisiae*. These proteins bind to upstream regions of a large number of genes and to genetic elements as diverse as centromeres, telomeres and DNA replication origins. The involvement of these proteins in transcriptional activation or repression, DNA replication and chromosome segregation suggests that they play a central role in cell growth, the different effects they exert probably being dependent on the context of the various binding sites and interaction with other proteins. The further study of these factors should shed light on the mechanisms which govern gene expression and chromosome maintenance.

Based on their DNA sequence requirements, proteins belonging to this class can be divided in several groups. The first of these consists of RAP1, alias GRFI or TUF (1−3). Binding sites are present in upstream activating sequences (UASs), silencers, telomeres and transcribed regions. The DNA sequence reguirements for binding have been extensively studied and high affinity sites conform to the sequence 5'-RMACCCANN-CAYY-3', where R, Y and M represent a purine, a pyrimidine and an A or C respectively (2,4).

We have described the characterization of two other abundant DNA-binding proteins, designated as GFI and GFII. These proteins bind to the 5' flanks of many nuclear genes coding for proteins of the mitochondrial respiratory chain, to a number of other genes involved in processes important for cell growth, and to elements important for cell division, ARS and CEN sequences (5,6). Proteins whose binding sites can be accommodated within the recognition sequences for GFI and GFII have been identified by others and have been variously named SBF-B, ABFI, TAF or SUF in the case of the GFI related factors (6−9) and CP1, or CBP-I, in the case of the GFII related factors (10−12). We and others have previously presented evidence that GFI, SUF, TAF, ABFI and SBF-B are likely to be the same protein (6,8,13).

GFII binding sites contain a motif identical to the CDEI element of yeast centromeres, namely 5'-RTCACRTG-3' (14). In contrast, the DNA sequence requirements of GFI appear more complex and are only partly understood. All the GFI binding sites detected thus far conform to the consensus 5'-RTCRYNNNNNACG-3' (6) which is characterized by the presence of two conserved elements, separated by a region which is variable in sequence, but constant in length. However, information about the relative strengths of the many possible GFI binding sites is so far lacking. This information is important for an understanding of the role played by GFI in the yeast cell.

We have now determined the relative strength of several GFI binding sites and have studied the DNA sequence requirements for binding in more detail. In addition, we have investigated whether DNA-binding proteins with sequence specificities similar to those of GFI and GFII are present in the distantly related yeast *K. lactis*. The implications of these findings are discussed.

* To whom correspondence should be addressed

Table I: Compilation of experimentally examined GFI-DNA binding sites in *S.cerevisiae*

	Sequence	Gene
Binders	ATCATTCCCAACG	CIII-subII
	GTCGCTTTGGACG	CIII-subVI
	GTCATCGTCCACG	CIII-subVI
	GTCACGTGGAACG	CIII-subVIII
	GTCGTCTCACACG	CYC1 (A)
	ATCATATTCGACG	CYC1 (B)
	AGCATTTTTGACG	ARS1-TRP1
	ATCGTTAATGACG	PHO5
	GTCACTCTAGACG	S33
	ATCGTTTTGTACG	L3
	ATCACTTATCACG	ACT
Non-binders	ATCAAGAAAAACG	CYC1 (C)
	ATCGGTACATACG	Mn-SOD

Data taken from refs 5 and 6. Invariant bases are indicated in bold type.

Table II: Frequency table of nucleotides in GFI binding sites

	Position													
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**		
5	1		4		1	1	1	3	5			11	G	
														R
6			7		1	1	2	2	2	11			A	
	10			7	7	7	6	3	1				T	
														Y
		11		4	2	2	2	3	3		11		C	
Consensus														
R	**T**	**C**	**R**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**A**	**C**	**G**		

MATERIALS AND METHODS

Recombinant DNA methods

Restriction enzymes were obtained from Boehringer Mannheim and New England Biolabs, T4DNA ligase was from New England Biolabs and DNA polymerase I Klenow fragment from Boehringer Mannheim and dam methylase from Biolabs. Recombinant DNA techniques were carried out according to standard procedures (15). Plasmids were propagated in the *Escherichia coli* strain HB101 or K12 3902(15). The K12 3902 strain is deficient in the dam methylase.

Plasmids and probes

Plasmids pUCS33-Taq-Taq, pEMBL9, pUC18, YRp7 and YEp351-RIP1, have been described (6, 16–19). Plasmids pEMBL9–24, pEMBL9-D24, pEMBL9–274, p31Y and p9-CYC1-II have also been described (5). Plasmid pBR-ACT was constructed by the insertion of a 3.2 kbp BamHI-EcoRI fragment containing the actin gene into the BamHI-EcoRI backbone of pBR322 (20). Plasmid p-cyc1-III was constructed by insertion of the 606 bp ClaI-EcoRI fragment containing the CYC1 gene coding region into the AccI-EcoRI backbone of pUC18 (21). The oligonucleotide used for testing of GFII binding contains the CDEI region plus flanking sequences of CENIV (12). The oligonucleotides used for GFI binding are designated as 40-1/40-2 and 40-3/40-4 and are shown in Table V.

DNA fragments used in retardation, footprint and dissociation experiments were labelled at the 3′ end by filling in the overhanging ends with the Klenow fragment of DNA polymerase I in the presence of [α^{32}P] dATP. Oligonucleotides were labelled at their 5′-termini by polynucleotide kinase and [γ^{32}P] ATP or at their 3′-termini by the Klenow fragment of DNA polymerase and [α^{32}P] dATP. DNA fragments were isolated according to standard procedures. DNA fragments were as follows: a 320 bp EcoRI-HindII fragment of pEMBL9-D24 (labelled at the EcoRI site), a 305 bp BamHI-SmaI fragment of p9-CYC1-II (labelled at the BamHI site), a 254 bp BamHI-HinfI fragment of p9-CYC1-II (labelled at the BamHI site), a 298 bp EcoRI-HindIII fragment of pEMBL9–274 (labelled at the EcoRI and the HindIII site), a 212 bp HindIII-PstI fragment of YRP7 (labelled at the HindIII site), a 204 bp XhoI-ClaI fragment of pBR-ACT (labelled at the XhoI site), a 387 bp EcoRI-PstI fragment of 31Y (labelled at the EcoRI site), a 302 bp EcoRI-PVUII fragment of pUC S33-Taq-Taq (labelled at the EcoRI site).

Table III: Stability of different GFI-DNA complexes

Gene	Fragment size (bp)	Dissociation half time (min)
ACT	204	70
S33	302	34.3
CYC1-A	305	21.2
ARS1-TRP1	212	14.1
CIII-subVIII	298	10.1
CIII-subII	320	6.5
CYC1 -B	305	3.5
PHO5	387	<1

Measurements were performed with cell extracts prepared as described by Arcangioli and Lescure (24). The standard deviation for duplicate experiments was 10% or less in all cases. For the CYC1 gene, the 305 and 254 bp fragments contain either both A and B GFI-binding regions, or only the A region. See text for discussion. Note also that the result for the CIII-subII gene 5′ flank is influenced by the methylation of the A present at position 1 of the GFI binding site. Sites are listed in descending order of binding strength.

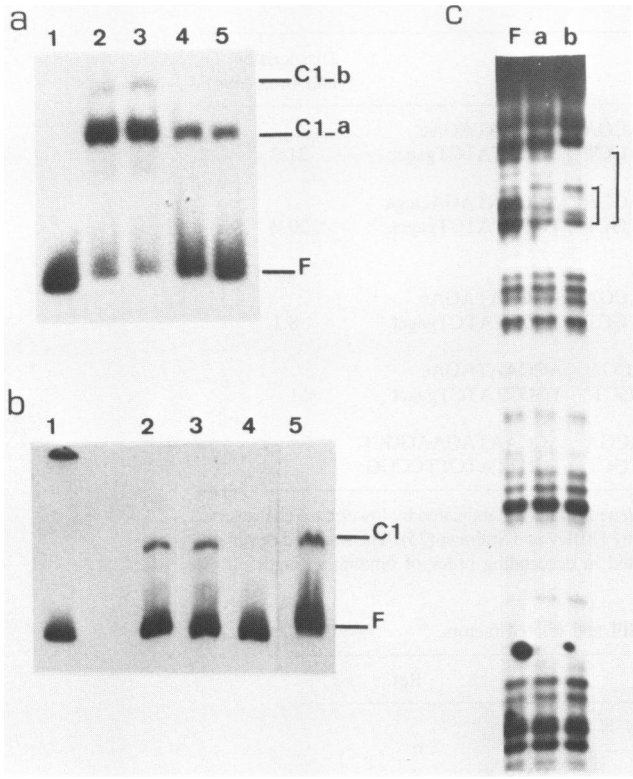

Figure 1: Mobility shift analysis of GFI-CYC1 gene DNA complexes. a) A ^{32}P-labelled fragment encompassing the 5′ flanking region of the CYC1 gene (−680 to −382 relative to the ATG) was incubated with a crude yeast cell lysate and the resulting complexes were analysed on a non-denaturing polyacrylamide gel. F indicates the position of the free DNA fragments in the gel. C1-a and C1-b indicate the positions of the two prominent DNA-protein complexes. Lanes 2 +3 show the result of an analysis in the absence of yeast competitor DNA. Lanes 4 + 5 show the results of an identical analysis in the presence of unlabelled GFI binding oligonucleotide 40-1/40-2 (see Table V) in a 20 molar excess. b) A ^{32}P-labelled fragment encompassing the 5′ flanking region of the gene coding for subunit II of complex III (−408 to −102 relative to the ATG) was used. Lanes 2 + 3 show the result of an analysis in the absence of yeast competitor DNA. Lanes 4 and 5 show the results of an identical analysis in the presence of the unlabelled 5′-flank, or coding region of the CYC1 gene in a 20-fold molar excess. c) DNaseI footprint analysis of the two GFI-DNA complexes, C1-a and C1-b on the 5′ flanking region of the CYC1 gene. The mixture of free and protein bound DNA fragments was digested with limited amounts of DNaseI prior to electrophoresis on a native acrylamide gel. Bands corresponding to the two complexes were excised, DNA was eluted and run on a denaturing polyacrylamide gel. Lanes F, a and b show DNA eluted from the corresponding bands. Brackets indicate the protected regions.

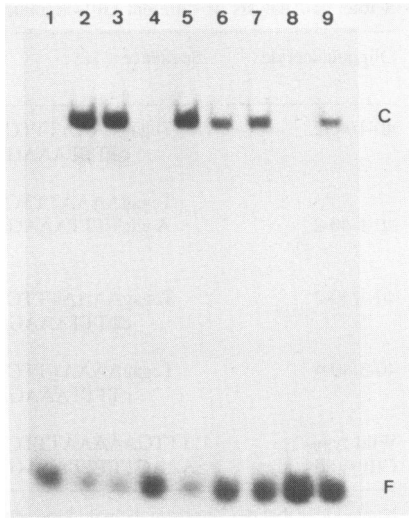

Figure 2: Retardation-competition experiments using either crude or partially purified preparations of GFI. The 40-1/40-2 oligonucleotide, which contains the GFI binding site of the CIII-subII gene was used as a probe for GFI binding. Lanes 2−5: whole cell extract. Lanes 6−9: Phenyl-Sepharose CL-4B fraction. Unlabelled oligonucleotides were present in a 50-molar excess as follows: lanes 4, 8: 40-1/40-2 oligonucleotide; lanes 5, 9: 40-3/40-4 oligonucleotide. This last oligonucleotide differs from 40-1/40-2 by a single base change (Table V), which reduces binding by GFI.

Table IV: GFI binding sites and their relative affinities

Site	Sequence	Cell Extr.	Phenyl-Seph. Fr.
S33	GTCACTCTAGACG	100	100
40-1/40-2	ATCATTCCCAACG	56	60
ARS-TRP1	AGCATTTTTGACG	14	12
CIII-subVIII	GTCACGTGGAACG	6	8

Measurements were performed with either whole cell extracts, or a partially purified GFI fraction, obtained by following the procedure described by Diffley and Stillman (25) as far as the Phenyl-Sepharose step. Relative binding affinities were calculated from the dissociation half-times of the complexes (see METHODS).

Yeast strains and media

The protease-deficient *S.cerevisiae* yeast strain BJ1991 (22) and the *K. lactis* yeast strain JA6 (23) was used. Cells were grown in YPD medium (1% yeast extract, 2% bacto-peptone, 2% glucose) to an A_{600nm} of 2.

Extract preparation and DNA-binding assays

Crude cell extracts prepared as described by Arcangioli and Lescure (24), were used for the dissociation experiments described in Table III, while for those described in Tables IV and V, extracts prepared as described by Diffley and Stillman (25) were used . The partially purified GFI preparation used in some experiments was a fraction obtained after the Phenyl-Sepharose CL-4B step described by these authors. A cell extract was prepared from *K. lactis* according to the procedure of Olesen et al. (26). Extracts were routinely dialyzed against standard binding buffer (5). Procedures for retardation-competition and footprint analysis have been described previously (5).

Dissociation rate constant analysis

Dissociation half times were determined by a modified version of the technique of Fried and Crothers (27). Standard binding reactions were incubated until equilibrium (15 min at 30°C). A 200-fold molar excess of GFI binding oligonucleotide was then added and samples representing six to eight time points were loaded onto standard 4% acrylamide gels while the gels were running. The gels were electrophoresed at 4°C, dried and autoradiographed. Binding activity was quantitated by counting excised bands. Data were plotted according to the first-order rate equation of Riggs et al. (28).

RESULTS

Comparison of different naturally occurring GFI binding sites

With the exception of sites present in the intron of the actin gene, and one of the regions present in the 5′ flank of the CYC1 gene, the GFI binding DNA regions used in this study have been

Table V: Analysis of different GFI-oligonucleotide complexes

Oligonucleotide	Sequence			Dissociation half time (min)
40-1/40-2	TcgagAAAATTTCTG cTTTTAAAGAC	ATCATTCCCAACGA TAGTAAGGGTTGCT	ACCAATAGAc TGGTTATCTgagct	21.2
40-1/40-2	TcgagAAAATTTCTG AgctcTTTTAAAGAC	ATCATTCCCAACGA TAGTAAGGGTTGCT	ACCAATAGActcga TGGTTATCTgagct	20.8
40-1/40-2	TcgagAAAATTTCTG cTTTTAAAGAC	CH_3 ATCATTCCCAACGA TAGTAAGGGTTGCT	ACCAATAGAc TGGTTATCTgagct	8.1
40-3/40-4	TcgagAAAATTTCTG cTTTTAAAGAC	ATCAgTCCCAACGA TAGTcAGGGTTGCT	ACCAATAGAc TGGTTATCTgagct	<1
Wild type CIII-subII	TTTGAAAAATTTCTG AAACTTTTTAAAGAC	ATCATTCCCAACGA TAGTAAGGGTTGCT	ACCAATAGAAGGCC TGGTTATCTTCCGG	

Nucleotides that deviate from the wild type sequence of the CIII-subII gene 5′ flank are indicated by lower case characters. Measurements were performed with cell extracts prepared as described by Diffley and Stillman (25). The standard deviation for duplicate experiments was 10% or less in all cases. Sites are listed in descending order of binding strength.

Table VI: Compilation of mutant binding sites for GFI and related factors

	Consensus	Gene	Ref.
	R T **C** R Y Y Y N N N **A C G** N N N N N		
M1	– A – – – – – – – – – – – – – – – –	L3	9
M2	– – A – – – – – – – – – – – – – – –	HMR-E	8
M3	– – – – G – – – – – – – – – – – – –	CIII-subII	this study
M4	– – – – – – – – – – – – C – – – – –	S33	6
M5	– – – – – – – – – – – – – – – G – –	L3	9
M6	– – – – – – – – – – – – – – – – – –	HMR-E	8

Bold residues indicate invariant bases. In mutant M6, each half site has been retained, but the geometry relating them has been altered by the addition of an extra A residue (8).

described previously (5,6). A compilation of all the GFI-binding sites experimentally examined thus far in *S.cerevisiae* is given in Table I. Based on the analysis of these binding sites, the consensus sequence 5′-RTCRYYYNNNACG-3′ can be derived (Table II).

To determine the dissociation half time of different GFI-DNA complexes, we used a modified version of the gel retardation method, as described by Fried and Crothers (27). For all sites, we first checked whether the factor giving rise to the retarded complex or complexes (CYC1 305 bp fragment) displayed the sequence specificity characteristic of GFI (results not shown). The results of the stability analysis are presented in Table III and show that the dissociation half times vary over a 70-fold range, the site present at the 3′ end of the coding region of the PHO5 gene being the weakest, with a dissociation half-time of less than 1 min. The ability to observe even these weak interactions by gel retardation suggests that the gel matrix stabilizes these complexes, a phenomenon also noted by others (29,30).

In the case of the CYC1 gene, there are 3 DNA regions, A, B and C which conform to the sequence RTCR $(N)_6$ ACG, but differ from each other at positions 5–7 of this motif. The A and B regions are located upstream of the UAS in the 5′ flank of the CYC1 gene, while the C region is present in the coding region. These three regions were further compared. Figure 1a shows the result of an experiment in which a fragment containing element A and B was used as a probe. At higher protein concentrations, this probe gives rise to the formation of two complexes, C1-a and C1-b. Specific competition was given by

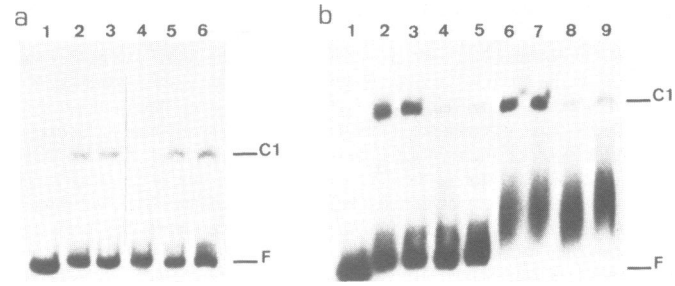

Figure 3: Retardation-competition studies with single-strand and closed circular DNA. a) A ^{32}P-labelled fragment encompassing the 5′ flanking region of the gene coding for subunit II of complex III (−408 to −102 relative to the ATG) was used. Lanes 2 + 3 show the result of an analysis in the absence of yeast competitor DNA. Lanes 4, 5 + 6 show the results of an identical analysis with unlabelled ds-GFI-binding oligonucleotide 40-1/40-2 in a 20 molar excess, unlabelled ss-oligonucleotide 40-1 in a 20 molar excess, unlabelled ss-oligonucleotide 40-2 in a 20 molar excess. b) The same labelled fragment was used as indicated under a). As competitor DNAs were used: pEMBL9 cleaved with HinfI (lanes 2, 3) pEMBL9-D24 cleaved with HinfI (lanes 4,5), pEMBL9 (lanes 6,7), pEMBL9-D24 (lanes 8,9).

a 20-molar excess of unlabelled GFI binding oligonucleotide 40-1/40-2 (lanes 4–5). The C region cannot compete for GFI binding in a retardation competition experiment, using the CIII-subII gene 5′ flank as a probe (Figure 1b, lane 5) and a fragment containing only the C region does not give rise to significant

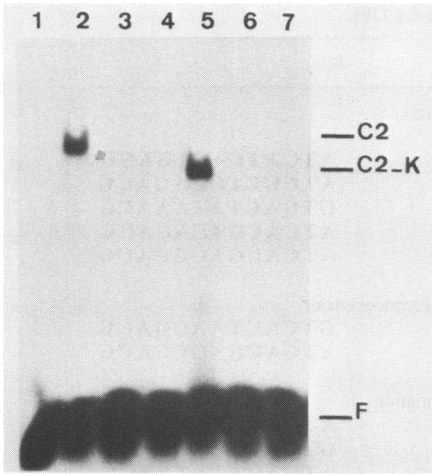

Figure 4: Identification of a protein with the binding specificity of *S.cerevisiae* GFII in *K. lactis*. The oligonucleotide containing the CDEI region of CENIV was used as a probe. Lanes 2, 3 and 4 show the results of mobility shift assays obtained with a cell extract prepared from *S.cerevisiae*. Lanes 5, 6 and 7 show the results obtained with a cell extract prepared from *K. lactis*. The following yeast competitor DNAs were used: CENVI sequences (lanes 3, 6) and CENIV sequences (lanes 4 and 7).

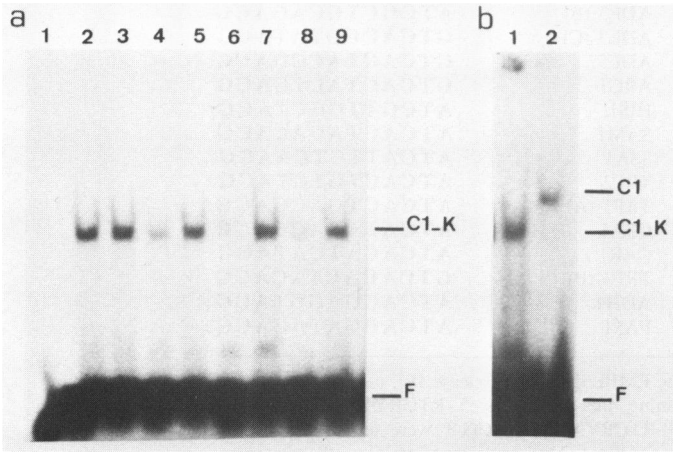

Figure 5: Identification of a protein with the binding specificity of *S.cerevisiae* GFI in *K. lactis*. a) The 40-1/40-2 oligonucleotide was used as a probe. Lanes 2 + 3 show the results of a mobility shift assay in the absence of yeast competitor DNA. For the remaining lanes, the following competitor DNAs were used: 40-1/40-2 oligonucleotide (lane 4), 40-3/40-4 oligonucleotide (lane 5), unlabelled S33 gene sequences, either intact or cleaved in the GFI binding site by XbaI (lanes 6, 7), unlabelled sequences of the ARS-TRP1 region, either intact or cleaved in the GFI binding site by DdeI (lanes 8, 9). b) Mobility shift assays were performed using the same oligonucleotides as in (a), with either a cell extract form *K. lactis* (lane 1), or a partially purified preparation of *S. cerevisiae* GFI (lane 2)

complex formation under the same conditions (results not shown). C1-a and C1-b were further analyzed by an indirect footprint experiment. A probe was incubated with protein and then treated lightly with DNaseI. Both free and complexed DNA were then separated on a native acrylamide gel, excised, extracted and re-run on a sequencing gel. As Figure 1c shows, the DNaseI digestion pattern of the A region is affected in C1-a, while the pattern of the B region is only slightly affected. C1-a must therefore arise from the occupation of site A by GFI. At higher protein concentrations C1-b arises by the occupation of both the A and B regions. The results obtained in dissociation experiments parallel these findings. When a fragment containing only the A region is used (CYC1 254 bp), a dissociation half time of 28 min. is found (results not shown). A comparable value (21 min.) is found for the dissociation of GFI from the A region when a fragment containing both A and B regions is used (CYC1 305 bp). However, GFI dissociates much faster from the B region (3.5 min.).

The A region contains the strongest GFI binding site of the CYC1 gene and conforms exactly to the consensus sequence. The same holds true for all the other strong binding sites analyzed thus far. These are the binding sites present in the UAS regions of the S33 gene and the site present in the intron of the actin gene. For the site present in the 5′-flank of the CIII-subII gene, the result presented in Table III is influenced by dam methylation (see below). This binding site is, however, also relatively strong. The weaker site in the B region of the CYC1 gene deviates from the consensus sequences at position 6. A deviation at the same position occurs in the CIII-subVIII gene-GFI binding DNA region and this is also a relatively weak binding site.

These results suggest that the presence of a pyrimidine, with a possible preference for a T, at position 6 is important for the binding process. The presence of a pyrimidine at position 5 is also implicated as important for binding from the results of an analysis of a point mutant of the CIII-subII gene 5′ flank binding site (see below). Taking these findings into account, it is understandable why both the C region of the CYC1 gene and the region of the SOD gene 5′ flank, as indicated in Table I do not bind GFI under standard conditions. These two regions deviate from the consensus at positions 5−7 and positions 5 and 7 respectively.

The validity of these observations was confirmed by additional controls. First, a number of the DNA regions mentioned above were also tested in retardation competition experiments (not shown) and for these, analogous conclusions can be derived from the two types of experiments: strong binding sites as measured with the dissociation assay function as good competitor DNAs in retardation competition experiments, while weak sites display short dissociation half-times and act as poor competitors. Second, we also tested some of the regions in experiments using preparations of GFI purified approximately 200-fold according to the procedure of Diffley and Stillman (25) in order to exclude a possible influence of proteins other than GFI on binding strength, or complex stability. As Figure 2 shows, a contribution of such additional factors to complex formation is unlikely. Regardless of whether whole cell extracts, or partially purified preparations are used for mobility shifts, the 40-1/40-2 oligonucleotide yields only a single major complex and this is retarded to exactly the same extent with both preparations. Competition experiements (Figure 2, lanes 4,5,8,9; see also below) also show that in each case, the binding observed is specific for GFI. The dissociation experiments using the partially purified preparation yielded the same relative affinities as those measured with cell extracts i.e. S33 > CIII-subII > ARS-TRP1 > CIII-subVIII (see Table IV).

DNA sequence requirements of GFI

The fragments used for the dissociation experiments presented in the first section were prepared from DNA isolated from an *Escherichia coli* strain which contains both dam and dcm site specific methylases (31−34). 5′-CCAGG-3′ and 5′-CCTGG-3′,

Table VII: Genes whose upstream regions contain (potential) binding sites for GFI

Gene	Site	Gene	Site
Mitochondrial functions		Translation (contd)	
CYC1 (A)	GTCGTCTCACACG		
CIII-subII	ATCATTCCCAACG	ATS	ATCATCAGTGACG
COX6	ATCGCTCCATACG	eIF-4E	ATCGCTGCAGACG
HTS1	ATCGCCTTACACG	TIF1	GTCACTATAAACG
MRS3 (A)	ATCATTTAATACG	RPL2	ATCACGTCACACG
MRS3 (B)	ATCGTTTGTAACG	MES1	ATCACGAGGCACG
MAS1	ATCACTAAATACG		
OMPMI-1 (A)	ATCACTGTATACG	Glycolysis/gluconeogenesis	
OMPMI-1 (B)	ATCATTATCTACG	ENO2	GTCACTAACGACG
PUT1 (A)	GTCGCTACCGACG	PGK	ATCACGAGCGACG
PUT1 (B)	GTCGCTACACACG		
MSW	GTCATATTTCACG	Cell differentiation	
CYC1 (B)	ATCATATTCGACG	STE7	ATCACTTATAACG
PUT1 (C)	ATCGCGCACGACG	YP2 (A)	GTCACTGTACACG
CIII-subVIII	GTCACGTGGAACG	ARFA	ATCACATATAACG
PET54	ATCACGATTAACG	CDC25	ATCGCGGGAGACG
		CMD1	ATCACAGTACACG
Cell growth/division		MAT4 (HMR)	ATCATAAAATACG
CDC16	ATCATCTTTAACG	YP2 (B)	GTCACGATATACG
BTUB (A)	GTCACTGTACACG		
GST1	ATCATATTATACG	Other Metabolic pathways	
BTUB (B)	GTCACGATATACG	DED1	ATCATTCTATACG
		DED1	GTCATTCTGAACG
Nucleic acids/transcription		DFR1	GTCGCTCTTGACG
RBP2	ATCATTTTGAACG	PIS	GTCGCTCCGCACG
SSB1	GTCACCTGGGACG	ADE3 (A)	GTCGCTTGACACG
PHO2	ATCACTTCGGACG	ADE3 (B)	ATCGCTGCAGACG
PPA (A)	GTCATCGCAGACG	ADE3 (C)	GTCACCGTATACG
PPA (B)	GTCATTAGGTACG	ADE57	GTCACTACCGACG
RPC40	GTCACTATAAACG	ARG1	GTCACTACTGACG
RPO31	ATCACTATATACG	HIS1	ATCGTTGGCTACG
PABPG	ATCATGTTTAACG	SAM1	ATCACTACACACG
SPT2	ATCATGTTAAACG	SSA1	ATCATTGTCAACG
RAD3	ATCATGGCCGACG	TRP2	ATCACTGTGTACG
SPT3	ATCACAGCAGACG	TRP3 (A)	ATCACTGACGACG
		CPA1	ATCATACACGACG
Translation		CAR1	ATCACATCATACG
RPS33	GTCACTCTAGACG	TRP3 (B)	GTCACACTACACG
HTS1	ATCGCCTTACACG	ADH4	ATCACGAGGTACG
RPL3	ATCGTTTTGTACG	FAS1	ATCACGATACACG

The list is based on a survey of yeast sequences contained in the EMBL data bank (release 19), carried out with the sequence 5'-RTCR$(N)_6$ACG-3' as string. Regions containing the sequences 5'-RTCRYYYNNNACG-3', 5'-RTCRYYNNNNACG-3', 5'-RTCRYNYNNNACG-3' or 5'-RTCRYNNNNNACG-3' were selected for further analysis. Of these, only the sites lying in the 5'-flanks of genes are listed. Nucleotides conforming to the consensus sequence 5'-RTCRYYYNNNACG-3' are given in bold type.

the recognition sequences of the dcm methylase are not present in the binding sites tested, whereas 5'-GATC-3', the recognition sequence of the dam methylase is present in the 5' flank of the CIII-subII gene. To test whether methylation of this site influences binding, we compared the dissociation of GFI from the oligonucleotide 40-1/40-2 before and after methylation by dam methylase in vitro. The results of this analysis are presented in Table V and they show that dam methylation results in a reduced dissociation half time. At the same time, these results also suggest that a purine at position 1 (an A for the CIII-subII gene 5' flank) is important for GFI binding.

The analysis of the different binding sites, as presented above, suggested that the presence of a pyrimidine at position 5, may be important for the binding of GFI. To test this supposition we used the 40-1/40-2 oligonucleotide already mentioned and another oligonucleotide 40-3/40-4 which differs from 40-1/40-2 by a single base change: the pyrimidine T at position 5 of the consensus is replaced by the purine G. Table V shows that this alteration indeed decreased the dissociation half time and the effect of this mutation was larger than the effect observed for methylation of the A at position 1. Table VI gives an overview of all the mutant binding sites for GFI and related factors. All mutations, apart from that in M5, involve conserved nucleotides and all reduce binding, irrespective of whether whole cell extracts, or partially purified fractions are used (see also Figure 2).

Interaction of GFI with single-stranded and covalently-closed circular DNA

Results of competition studies with single-strand DNA and covalently closed circular DNA containing the CIII-subII gene 5' flank binding sites as competitor DNAs are shown in Figure 3. These indicate that GFI has either no, or only low affinity for single-strand DNA (Figure 2a, lanes 5, 6). This result was verified by the observation that the 40-1 and 40-2 oligonucleotides, when used as single-strand probes, did not give significant retardation (results not shown). In addition, Figure 3b shows that the competing abilities of linear fragments containing the CIII-subII gene 5' flank and uncleaved plasmid

DNA containing the same flank are not significantly different.

Factors with the target specificities of GFI and GFII are also present in the yeast *K. lactis*

To get more insight in the evolutionary conservation of GFI and GFII, we tested whether proteins with the target specificity of *S. cerevisiae* GFI and GFII are also present in the distantly related yeast *K. lactis*. For GFII, a protein with the same specificity was easily detected. Incubation of an oligonucleotide containing the CENIV CDEI region plus flanking sequences as a probe with a whole cell extract of *K. lactis* result in the formation of one prominent complex. Competition studies show that the protein involved can also bind CENVI sequences (Figure 4, lane 6). As expected the 5′ flanks of the FeS gene, the CIII-subVI and CIII-subVIII gene also bind this factor (results not shown, see also ref 5).

In the first instance, we were unable to detect a protein with the binding specificity of GFI. The use of higher protein concentrations in the incubation mixture, however, resulted in the formation of a complex with the 40-1/40-2 oligonucleotide, although at lower levels than observed with extracts from*S. cerevisiae*, suggesting that the abundance of the *K. lactis* factor, and/or its affinity for this target is lower than in *S. cerevisiae*. To ascertain that the protein involved displays the target specificity of GFI, we performed an extensive array of competition experiments, the results of which are presented in Figure 5. This shows that the protein also binds to the 5′-flank of the S33 gene and to the ARS-TRP1 region, while binding less efficiently to the 40-3/40-4 oligonucleotide, which contains a mutant GFI site. Binding is specific, as indicated by the fact that in each case competition is eliminated by restriction cleavage within the various sites. The factor from *K. lactis* thus indeed binds to the same sequences as *S. cerevisiae* GFI. Both *K. lactis* factors retard the oligonucleotide probes to a lesser extent than their *S. cerevisiae* counterparts, suggesting that although the DNA-binding domains may have been conserved, other differences probably exist between the factors in the two yeasts.

DISCUSSION

Since our original description of the DNA-binding factors GFI and GFII (5) it has become increasingly clear that both are likely to be identical with proteins variously named SBF-B, ABFI, TAF or SUF in the case of the GFI related factors (6−9) and CP1, or CBP-I, in the case of the GFII related factors (10−12). We and others have previously presented evidence that GFI, SUF, TAF, ABFI and SBF-B are likely to be the same protein (6,8,13). While this paper was being prepared for publication, two reports appeared, dealing with the cloning and sequencing of what is likely to be the same (essential) gene coding for the factors ABFI (35) and BAF1 (36). Accordingly, we think it probable that all independently identified activities in fact refer to a single multifunctional protein.

The characterization in this paper of several strong binding sites for GFI has provided additional information about the sequence specificities of this factor and has permitted the formulation of a consensus sequence for high affinity sites, namely 5′- RTCRYYYNNNACG -3′. Weaker binding sites are characterized by changes at positions 5−7. Nucleotides at position 1, 3 and 5 of the first element and at positions 12 and 13 of the second element have been implicated as involved in DNA-protein interactions by mutational analysis and/or modification-interference analysis procedures (see also refs 5,6). There are several separate lines of evidence, which suggest that nucleotides outside the consensus DNA recognition domain of GFI do not play a significant role in the binding process. First, methylation interference footprinting of the GFI-S33 gene 5′ flank interaction indicates that modification of G nucleotides outside the consensus region does not impair binding (6). Second, dam methylation of an A nucleotide in the vicinity of the consensus region in the CIII-sub VIII gene 5′ flank does not influence the dissociation half time (results not shown). Additional support for this idea can be derived from recent work on a factor designated OBF1 (37). This factor binds to the ARS1-TRP1 region and two other ARS regions, ARS120 and ARS121. The three binding sites conform to the consensus DNA recognition sequence of GFI. Linker-substitution mutagenesis of the ARS121 binding sites suggested that nucleotides outside the two conserved elements are not important for binding.

The GFI binding site present in the 5′ flank of the CIII-subII gene contains the recognition sequence for dam methylase (GATC) and we have found that dam methylation results in a reduced dissociation half time of the GFI-DNA complex formed at this site in vitro. The consequences of dam and/or dcm methylation for protein-DNA interactions are often overlooked. For example, the interaction of HAP1 with the 5′ flank of the CYC1 gene has only recently been found to be decreased by dam methylation (38). In another case, involving a systematic study of the sequence requirements for binding of the transcription factor NFI (39), dcm sites were created by mutational events, but the effects of methylation on binding were not taken into account.

It is interesting to note that strong binding sites for GFI and its namesakes are located in DNA regions important for transcriptional activation and replication. In a number of cases, mutational studies indicate existence of a correlation between strength of factor binding and extent of transcriptional activation. This holds both for the strong binding sites in the UAS regions of the L3 and S33 genes (9, 40) as well as for more recently characterized sites in the intergenic region of the YPT1-TUB2 genes and a Ty2 element (41,42). For DNA replication, a strong binding site present in an ARS region is that recently identified in the ARS121 region. This site is considerably stronger than the one in the ARS1-TRP1 region (37), which in our analysis is of intermediate strength. Like the ARS-TRP1 site, the ARS121 binding site has been shown to be important for normal ARS function, as an origin of replication.

One of the aims of the present study was to provide criteria by which potentially high affinity sites could be recognized. Application of these criteria to yeast DNA sequences in the EMBL sequence databank results in the list given in Table VII and it may be significant that of the 18 sites displaying a perfect match to the consensus 5′-RTCRYYYNNNACG-3′, all but one lie in a narrow region of approximately −150 to −450 relative to the translational start of the relevant gene. Further weight is lent to the predictive value of the list by our observation that the site present upstream of the CIII-subVIII gene is necessary for basal level transcriptional activity of the gene during growth of cells on glucose (de Winde and Grivell, unpubl. obs). Moreover, a similar situation seems to hold for the predicted strong site present in the UAS of the COX6 gene (43). Finally, one of the two sites predicted in the 5′ flank of the DED1 gene has already been shown to bind the GFI related factor ABFI (44). In vivo, this site acts synergistically with a T-rich element to

activate transcription. We predict that GFI, like the RAP1 factor, will turn out to interact with a variety of genes whose expression is essential for growth (45). In the case of genes coding for mitochondrial proteins, we have suggested that the factor may play a role in the coupling of the rate of synthesis of these proteins to the rate of cell growth (13).

Although the role of GFI in the 5′ flanks and ARS regions has become more clear from the studies discussed above, the role of GFI-binding sites in coding regions of genes is unclear. A very weak binding site, such as that found PHO5 may be fortuitous. However, a strong binding site as present in the intron of the actin gene may not. For the factor RAP1, both strong and weak sites are also found in coding regions (2). Their significance is not yet understood.

GFII binding sites conform to a motif identical to the CDEI element of yeast centromeres, namely 5′-RTCACRTG-3′(14). As discussed above for GFI, all evidence so far suggests that nucleotides outside the consensus region do not significantly contribute to the binding process (11). CDEI sites are required for optimal centromere function (11,46). Other GFII binding sites occur in the 5′ flanking regions of genes coding for proteins of the mitochondrial respiratory chain, in the 5′ flank of the GAL2 gene, and in the UAS region of the MET25 gene (5,10,47), thus suggesting that GFII also plays a role in several different cellular processes.

We have tested whether proteins with the sequence specificities of GFI and GFII are present in the distantly related yeast *K. lactis*. Factors with the target specificity of GFII have been reported to be evolutionarily conserved from fungi to man (10,48). We therefore expected to be able to detect a protein with the sequence specificity of GFII in *K. lactis*. This turns outs to be the case, and like its *S.cerevisiae* counterpart, this is an abundant DNA-binding protein. In contrast, detection of in *K. lactis* of a protein with the target specificity of GFI turned out to be more difficult, suggesting that that either the abundance of this protein, or its affinity for the DNA sequences used is lower than in *S. cerevisiae*. It will thus be of interest to examine the distribution of the binding sites for this protein in the nuclear genome of this yeast and to determine its exact DNA sequence requirements for binding.

ACKNOWLEDGEMENTS

We thank Dr K. Breunig (Universität Düsseldorf) for the gift of *K. lactis* strain JA6 and for stimulating discussions. We are grateful to Dr M. Meijer and Mr J.H. de Winde for their critical reading of the manuscript. We also acknowledge the help of Mr C. Bisschops (Isogen Bioscience bv) for the synthesis of oligonucleotides used. This work was supported in part by grants from the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Research (NWO).

REFERENCES

1. Shore, D. and Nasmyth, K. (1987) Cell, 51, 721–732.
2. Buchman, A.R., Lue, N.F. and Kornberg, R.D. (1988) Mol. Cell. Biol., 8, 5086–5099.
3. Vignais, M.-L., Woudt, L.P., Wassenaar, G.M., Mager, W.H., Sentenac, A. and Planta, R.J. (1987) EMBO J., 6, 1451–1457.
4. Nieuwint, R.T.M., Mager, W.H., Maurer, K.C.T. and Planta, R.J. (1989) Curr. Genet., 15, 247–251.
5. Dorsman, J.C., Van Heeswijk, W.C. and Grivell, L.A. (1988) Nucl. Acids Res., 16, 7287–7301.
6. Dorsman, J.C., Doorenbosch, M.M., Maurer, C.T.C., De Winde, J.H., Mager, W.H., Planta, R.J. and Grivell, L.A. (1989) Nucl. Acids Res., 17, 4917–4923.
7. Shore, D., Stillman, D.J., Brand, A.H. and Nasmyth, K.A. (1987) EMBO J., 6, 461–467.
8. Buchman, A.R., Kimmerly, W.J., Rine, J. and Kornberg, R.D. (1988) Mol. Cell. Biol., 8, 210–225.
9. Hamil, K.G., Nam, H.G. and Fried, H.M. (1988) Mol. Cell. Biol., 8, 4328–4341.
10. Bram, R.J. and Kornberg, R.D. (1987) Mol. Cell. Biol., 7, 403–409.
11. Baker, R.E., Fitzgerald-Hayes, M. and O'Brien, T.C. (1989) J. Biol. Chem., 264, 10843–10850.
12. Cai, M. and Davis, R.W. (1989) Mol. Cell. Biol., 9, 2544–2550.
13. Grivell, L.A., Dorsman, J.C., Gozdzicka-Jozefiak, A., van Heeswijk, W.C. and de Winde, J.H. (1990) in Structure, Function and Biogenesis of Energy Transfer Systems (E. Quagliariello, S. Papa, F. Palmieri and C, Saccone, eds) Elsevier, in press.
14. Hieter, P., Pridmore, D., Hegemann, J.H., Thomas, M., Davis, R.W. and Philippsen, P. (1985) Cell, 42, 913–921.
15. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
16. Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl. Acids Res., 11, 1645–1655.
17. Yanish-Perron, C., Vieira, J. and Messing, J. (1985) Gene, 33, 103–119.
18. Struhl, K., Stinchcomb, D.T., Scherer, S. and Davis, R.W. (1979) Proc. Natl. Acad. Sci. USA, 76, 1035–1039.
19. Beckmann, J.D., Ljungdahl, P.O., Lopez, J.L. and Trumpower, B.L. (1987) J. Biol. Chem., 262, 8901–8909.
20. Ng, R. and Abelson, J. (1980) Proc. Natl. Acad. Sci. USA, 77, 3912–3916.
21. Lowry, C.V., Weiss, J.L., Walthall, D.A. and Zitomer, R.S. (1983) Proc. Natl. Acad. Sci. USA, 80, 151–155.
22. Jones, E.W. (1977) Genetics, 85, 23–33.
23. Breunig, K.D. and Kuger, P. (1987) Mol. Cell. Biol., 7, 4400–4406.
24. Arcangioli, B. and Lescure B. (1985) EMBO J., 4, 2627–2633.
25. Diffley, J.F.X. and Stillman, B. (1988) Proc. Natl. Acad. Sci. USA, 85, 2120–2124.
26. Olesen, J., Hahn, S. and Guarente, L. (1987) Cell, 51, 953–961.
27. Fried, M. and Crothers, D.M. (1981) Nucl. Acids Res., 9, 6505–6525.
28. Riggs, A.D., Bourgeois, S. and Cohn, M. (1970) J. Mol. Biol., 53, 401–417.
29. Peterson, C.L. and Calame,. K.L. (1987) Mol. Cell. Biol., 7, 4194–4203.
30. Huet, J. and Sentenac, A. (1987) Proc. Natl. Acad. Sci. USA, 84, 3648–3652.
31. Marinus, M.G. and Morris, N.R. (1973) J. Bacteriol., 114, 1143–1150.
32. Geier, G.E. and Modrich, P. (1979) J. Biol. Chem., 254, 1408–1413.
33. May, M.S. and Hattman, S. (1975) J. Bacteriol., 123, 768–770.
34. Backman, K. (1980) Gene, 11, 167–171.
35. Diffley, J.F.X. and Stillman, B. (1989) Science, 246, 1034–1038.
36. Halfter, H., Kavety, B., Vanderkerckhove, J., Kiefer, F. and Gallwitz, D. (1989) EMBO J., 8, 4265–4272.
37. Walker, S.S., Francesconi, S.C., Tye, B-K. and Eisenberg, E. (1989) Mol. Cell. Biol., 9, 2914–2921.
38. Sousa, R. and Arcangioli, B. (1989) EMBO J., 8, 1801–1808.
39. Gronostajski, R.M. (1987) Nucl. Acids Res., 15, 5545–5559.
40. Herruer, M.H., Mager, W.H., Doorenbosch, T.M., Wessels, P.L.M., Wassenaar, T.M. and Planta, R.J. (1989) Nucl. Acids Res., 17, 7427–7439.
41. Halfter, H., Müller, U., Winnacker, E-L. and Gallwitz, D. (1989) EMBO J., 8, 3029–3037.
42. Goel, A. and Pearlman, R.E. (1988) Mol. Cell. Biol., 8, 2572–2580.
43. Trawick, J.D., Rogness, C. and Poyton, R.O. (1989) Mol. Cell. Biol., 9, 5350–5358.
44. Lue, N.F., Buchman, A.R. and Kornberg, R.D. (1989) Proc. Natl. Acad. Sci., USA, 86, 486–490.
45. Capieaux, E., Vignais, M-L., Sentenac, A. and Goffeau, A. (1989) J. Biol. Chem., 264, 7437–7446.
46. Hegemann, J.H., Shero, J.H., Cottarel, G., Philippsen, P. and Hieter, P. (1988) Mol. Cell. Biol., 8, 2523–2535.
47. Thomas, D., Cherest, H. and Surdin-Kerjan, Y. (1989) Mol. Cell. Biol., 9, 3292–3298.
48. Chodosh, L.A., Buratowski, S. and Sharp, P.A. (1989) Mol. Cell. Biol., 9, 820–822.