Volume 16 Number 15 1988 Nucleic Acids Research

Identification of two factors which bind to the upstream sequences of a number of nuclear genes coding for mitochondrial proteins and to genetic elements important for cell division in yeast

J.C.Dorsman, W.C.van Heeswijk and L.A.Grivell

Section for Molecular Biology, Department of Molecular Cell Biology, Kruislaan 318, 1098 SM Amsterdam, The Netherlands

Received June 3, 1988; Accepted July 14, 1988

ABSTRACT

Two abundant factors, GFI and GFII which interact with the 5' flanking regions of nuclear genes coding for proteins of the mitochondrial respiratory chain have been identified. In one case (subunit VIII of QH_2: cytochrome c oxidoreductase) the binding sites for both factors overlap completely and their binding is mutually exclusive. For the other 5' regions tested the GFI and GFII binding sites do not coincide. Interestingly, binding sites for GFI and GFII are also present in or at the 3' ends of the coding regions of two genes of the PHO gene family and in DNA elements important for optimal ARS and CEN function respectively. The sites recognized by GFI conform to the consensus RTCRNNNNNNACGNR, while those recognized by GFII contain the element RTCACGTG. We speculate that GFI and GFII may play a role in different cellular processes, dependent on the context of their binding sites and that one of these processes may be the coordination of the expression of genes involved in mitochondrial biogenesis with the progress of the cell cycle.

INTRODUCTION

In the yeast Saccharomyces cerevisiae, biosynthesis of a functional mitochondrion is regulated in response to the need for mitochondrial function, with control being exerted by oxygen, haem and catabolite repression. Synthesis of mitochondrial proteins, including components of the respiratory chain is regulated mainly at the level of transcription and recent work has led to the demarcation of upstream sequence elements responsible for either activation or repression of the transcription of a number of genes. In the case of CYC-1, the gene encoding iso-1 cytochrome c, proteins capable of interacting specifically with adjacent upstream sequences responsible for activating transcription in response to haem and non-fermentable carbon sources (UAS1, UAS2) have been characterized (1-3). The picture emerging is a complex one, each UAS being responsive to specific stimuli and capable of interacting with at least two different proteins (UAS1:HAP1-product, factor RC2; UAS2:HAP2 and HAP3-products). Sequences resembling both UAS1 and UAS2 are present upstream of genes for other mitochondrial proteins and it seems reasonable to suppose that the proteins involved are capable of exercising control on these genes too.

Genes coding for four of the 8-9 imported subunits of the QH_2: cytochrome c oxido-reductase (complex III) in yeast have previously been isolated and characterized (4-6). The expression of these genes is regulated in parallel both with each other and with genes for other

mitochondrial proteins. Upstream sequences important for induction of expression on a non-fermentable carbon source have been identified for subunits II (40 kDa) and VIII (11 kDa) (Dorsman et al., manuscript in preparation; 7).

We have used direct biochemical approaches to identify DNA binding proteins and now report on the identification of two such factors, designated GFI and GFII. These interact both with the 5' flanking sequences of genes coding for subunits II, V, VI and VIII of complex III, and with other genetic elements. The function of these factors is as yet unknown, but the analysis of their binding sites suggests that they play roles in both transcription and DNA replication.

MATERIALS AND METHODS

Recombinant DNA methods

Restriction enzymes were obtained from Boehringer Mannheim and New England Biolabs, T4 DNA ligase was from New England Biolabs and DNA polymerase I Klenow fragment from Boehringer Mannheim. Recombinant DNA techniques were carried out according to standard procedures (8).

Plasmids and probes

Plasmids pEMBL9, pUC18, YRp7, pLG669-Z and pYEP351-RIP-1 have been described (9-13). Plasmids pEMBL9-D24 and pEMBL9-24 were derived from pEMBL9-SH (14). They contain HindIII-Bal31 deletion fragments of the 5' flanking region of the gene coding for subunit II of complex III (-530 to -102 relative to ATG, -530 -2 relative to ATG, respectively) in a HindIII-BamHI blunt backbone of pEMBL9. Plasmid pEMBL9-274 was constructed by the insertion of a HindIII-XhoII fragment containing sequences of the 5' region and of the coding region of the gene coding for subunit VIII of complex III (-274 to +7 relative to the ATG; 15) into the HindIII-BamHI backbone of pEMBL9. The detailed construction of this plasmid is described elsewhere (7). Plasmid p9-cyc1-II was constructed by the insertion of the 433 bp XhoI fragment of pLG669-Z (12) containing sequences of the 5' region of the CYC1 gene into the SalI site of pEMBL9. M13mp7-17-Sau3A was constructed by the insertion of the 718 bp Sau3A fragment of p209 (17k) (6) containing sequences of the 5' flanking region and of the coding region of the gene coding for subunit VI of complex III into the BamHI site of M13mp7. Plasmid p31Y is a yeast expression vector which contains the 5' region of the PHO5 acid phosphatase gene, the 3' region of the PHO5 gene (-540 to -3 relative to ATG, +1113 to +1490 relative to ATG; 16) separated by the polylinker sequence GGAATTCCCGG in the HindIII-BamHI backbone of pAT153.

DNA fragments used in retardation and footprint 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 and isolated according to standard procedures. DNA fragments were as follows: a 321 bp EcoRI-HindII fragment of pEMBL9-D24 (labelled at the EcoRI site), a 298 bp EcoRI-HindIII fragment of pEMBL9-274 (labelled at the HindIII site), a 305 bp BamHI-SmaI fragment of p9-cyc1-II (labelled at the BamHI site), a 528 bp HindIII-PstI fragment of pYEP351-RIP-1 (labelled at the HindIII site), a 371 bp EcoRI-AvaII fragment of M13mp7-17-Sau3A (labelled at the EcoRI site), a 368 bp EcoRI-AvaII fragment of M13mp7-17-Sau3A (labelled at the EcoRI site), a 212 bp EcoRI-FokI fragment of M13mp7-17-Sau3A (labelled at the EcoRI site), a 387 bp EcoRI-PstI fragment of p31Y (labelled at the RI site), a 211 bp HindIII-PstI fragment of pYRp7 (labelled at the HindIII site).

Yeast strains and media

The protease-deficient yeast strain BJ1991 (17) was used to prepare yeast extracts. 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

Cell extracts were prepared as described by Arcangioli and Lescure (2).

DNA-binding assays were performed under a variety of conditions. Typically 10 µg of a yeast cell lysate was incubated with 0.2-1 ng of probe and 0.5 µg pEMBL9 or other plasmid DNA and/or poly[d(I-C)] (Sigma) for 20 min. at 30°C in 20 mM Hepes, pH 7.8, 100 mM NaCl, 2 mM EDTA, 7 mM mercaptoethanol, 1 mM PMSF and 10% glycerol and the sample was immediately loaded on a 4% vertical polyacrylamide gel in TBE buffer (TBE: 90 mM Tris, 90 mM boric acid, 2.5 mM EDTA) at 4°C. The sample was electrophoresed at 185V for 5 min. and then at 135V for 2 to 4 hrs. Gels were treated for 10 min. with 10% methanol, 10% acetic acid,

dried and autoradiographed. For competition assays a 10-100 molar excess of unlabelled DNA over the probe was added prior to the addition of the extract. The total amount of DNA present in the reaction mixture was kept constant at 0.5 µg. When plasmid DNA containing yeast inserts was used as specific competitor DNA, it was routinely cut with suitable restriction enzymes which cleave outside the yeast fragments, unless otherwise indicated in the text.

For footprinting, standard binding reactions were scaled up 10 fold, adjusted to 5 mM $CaCl_2$ and digested with empirically determined amounts of DNaseI for 1-2 min. The reaction was stopped by adding EDTA to a final concentration of 25 mM and the incubation mixtures were immediately loaded on a 4% polyacrylamide gel. Following electrophoresis, the gel was autoradiographed overnight, free and bound probe was excised from the gel and the DNA was eluted by incubation overnight at 37°C in 10 mM Tris, pH 7.8, 1 mM EDTA, 0.2% SDS and 0.3 M NaCl. The eluate was filtered through glasswool, phenol extracted, ethanol precipitated and analysed by electrophoresis in a 6% polyacrylamide gel containing 6 M urea.

RESULTS

GFI binds to the 5' region of the gene coding for subunit II of QH_2-cytochrome c oxidoreductase

In an attempt to detect proteins that interact specifically with the 5' region of the gene coding for subunit II of complex III, the DNA fragment retardation technique of Garner and

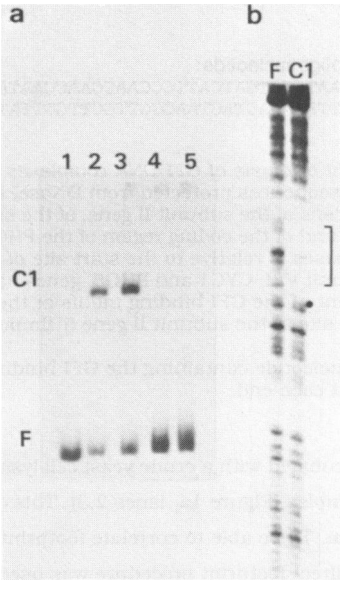

Fig.1: *Analysis of GFI in a crude yeast cell lysate by gel retardation*

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 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. Lanes 2-3 show the result of an analysis in the absence of yeast competitor DNA. One prominent retarded band is observed (C1). Lanes 4-5 show the results of an identical analysis in the presence of unlabelled GFI binding oligonucleotide in a 20 molar excess.

b) DNaseI footprint analysis of complexes in band C1. The mixture of free and protein bound DNA fragments was digested with DNase I prior to electrophoresis on a native polyacrylamide gel. Bands corresponding to F and C1 were excised, DNA was eluted and run on denaturing polyacrylamide gel. Lanes F and C1 show DNA eluted from the corresponding bands. A bracket indicates the protected region.

a

```
                  -203                    -179
GFI          ATTTCTGATCATTCCCAACGAACCAATAG
CIII-subII   TAAAGACTAGTAAGGGTTGCTTGGTTATC

                  -231          -248
GFI(GFII)    TGGACTAGTCACGTGGAACGGTAGATCGA
CIII-subVIII ACCTGATCAGTGCACCTTGCCATCTAGCT

                  -446          -463
GFI          CGACGATGTCGTCTCACACGGAAATATAA
CYC1         GCTGCTACAGCAGAGTGTGCCTTTATATT

                  +1201         +1184
GFI          CAACAGCATCGTTAATGACGTATCTGACG
PHO5         GTTGTCGTAGCAATTACTGCATAGACTGC

Consensus GFI       RTCRNNNNNNACGNR
Consensus GFII      RTCACGTG
```

b

GFI binding oligonucleotide:

```
5' TCGAGAAAATTTCTGATCATTCCCAACGAACCAATAGAC 3'
3'     CTTTTAAAGACTAGTAAGGGTTGCTTGGTTATCTGAGCT 5'
```

Fig.2: *Summary of the footprint analysis of GFI DNA complexes*
a) Brackets cover DNA sequences protected from DNaseI by binding of protein. Sequences of the upstream regions of the subunit II gene, of the subunit VIII gene and of the CYC1 iso-1-cytochrome c gene and of the coding region of the PHO5 acid phosphatase gene are presented. Numbering of the bases is relative to the start site of translation. The orientation of the GFI binding sites at subunit VIII, CYC1 and PHO5 genes has been inverted relative to the TATA box to allow the alignment of the GFI binding motifs at the four different sites. The BclI site present in the GFI binding site in the subunit II gene 5' flanking region is indicated by the black bar below the sequence.
b) The synthetic oligonucleotide containing the GFI binding site is also shown. The duplex has an XhoI overhang at each end.

Revzin (18) was used. When incubated with a crude yeast cell lysate, this fragment is capable of forming a single prominent complex (Figure 1a, lanes 2,3). This complex was analysed in more detail by DNaseI digestion. To be able to correlate footprinting results directly with specific retarded bands, an indirect footprint procedure was used: radiolabelled DNA was incubated with a crude cell lysate and then treated lightly with DNaseI, free DNA and DNA-protein complexes were separated on a native acrylamide gel, excised , extracted and re-run on a sequencing gel. The result of such an experiment is presented in Figure 1b. The size of the protected region is ± 24 bp and the sequence is shown in Figure 2a.

We have designated the factor which binds to this region as GFI. In order to better define the sequences with which it interacts, an oligonucleotide containing the binding region of the subunit II gene promoter along with flanking sequences, was synthesized (Figure 2b). This oligonucleotide is indeed able to bind GFI (data not shown), suggesting that no long

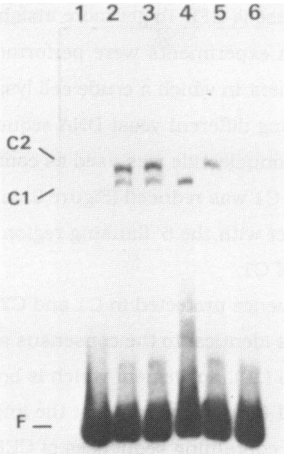

Fig.3: *Analysis of GFI and GFII in a crude yeast cell lysate by gel retardation*
A ^{32}P-labelled fragment encompassing the 5' flanking region of the gene coding for subunit VIII of complex III (-274 to +7 relative to the ATG) was used. F indicates the position of the free DNA fragments in the gel. Lanes 2-3 show the results of an analysis in the absence of yeast competitor DNA. Two prominent retarded complexes are observed (C1 and C2). Lanes 4,5 and 6 show the results of an identical analysis with unlabelled CENVI sequences, unlabelled GF1 binding oligonucleotide (20 molar excess) and a mixture of both (each in a 10 molar excess) respectively.

regions of sequences flanking the protected region are necessary for recognition and binding, a feature which is shared by other DNA binding proteins. This oligonucleotide was also used in the retardation competition experiment, shown in Figure 1a. A crude cell lysate was pre-incubated with an excess of the GFI binding oligonucleotide (20-fold on a molar basis), before addition of the labelled subunit II gene promoter fragment. As expected, preincubation effectively prevented the formation of the retarded complex, thus establishing the value of the oligonucleotide as a specific competitor DNA in assays of GFI binding.

Two factors GFI and GFII bind to the same 5' flanking region of the gene coding for subunit VIII

Preliminary retardation competition experiments using the 5' flank of the gene coding for subunit II as a probe, (data not shown) had already suggested that GFI can also interact with the 5' flanking sequence of the gene coding for subunit VIII of the same complex. To verify this directly, retardation experiments were performed, this time using a fragment derived from the 5' flank of this gene as a probe. When incubated with a crude yeast cell lysate, this fragment is capable of forming two prominent complexes, designated C1 and C2 (Figure 3, lanes 2, 3). The indirect footprint procedure described above shows that the factors in C1 and C2 protect the same DNA region (7). The results of this analysis are summarized in Figure 2a. A reasonable explanation is that two different proteins are involved in the formation of the

two complexes and that one of these is GFI. To get more insight into the identity of the factors involved, retardation competition experiments were performed.

Figure 3 shows an experiment in which a crude cell lysate was preincubated with an excess of DNA fragments containing different yeast DNA sequences (20-fold excess on a molar basis). When the GFI binding oligonucleotide was used as competitor DNA, the formation of only the faster migrating complex C1 was reduced (Figure 3, lane 5). This experiment thus confirms that GFI can also interact with the 5' flanking region of this gene and it shows that it is responsible for the formation of C1.

Inspection of the DNA sequence protected in C1 and C2 shows that this region contains the sequence GTCACGTG which is identical to the consensus sequence of yeast centromere DNA element I (CDEI), RTCACGTG (19), an element which is bound by a protein designated CP1, centromere binding protein 1 (20). In order to test the importance of this sequence motif for complex formation, fragments containing sequences of CENVI were used as competitor. As shown in Figure 3, lane 4, competition is observed with the amount of the slower migration complex C2 being reduced and this suggests that this complex is formed by a protein capable of recognizing both elements. We have designated the factor which gives rise to C2 as GFII. We suggest that it is related or identical to CP1 identified by Bram and Kornberg (20). As might be expected, a mixture of CENVI sequences and the GFI binding oligonucleotide diminished the formation of both complexes (Figure 3, lane 6).

These experiments suggest that two different factors interact with the same DNA region of the 5'-flank of the subunit VIII gene in a mutually exclusive manner (Figures 2 and 3).

GFI and GFII also bind to the 5' flanking sequences of nuclear genes coding for other proteins of the mitochondrial respiratory chain

The concentration of GFI is evidently higher than the concentration of regulatory factors such as GCN4 and GAL4 which cannot be detected with retardation assays using crude yeast cell lysates (21,22). This abundancy suggests that the binding sites of GFI are not confined to the 5' flanking regions of this specific set of genes. To test this supposition and to get more insight in the sequence requirements for GFI binding, several other DNA fragments were tested.

A fragment was scored positive for GFI binding if: incubation of the radiolabelled fragment with a crude yeast cell lysate gave rise to complex formation, if both the GFI binding oligonucleotide and subunit II gene promoter fragments competed for this binding, and if this competition was prevented by cleavage of the subunit II promoter fragments in the GFI binding site (achieved with BclI, see Figure 2a). We also tested the same fragments for GFII binding. These were scored positive if: CENVI sequences diminished complex formation, and if the consensus sequence of CDEI RTCACGTG was present. The results of these experiments are summarized in Table I. The sequence organization of the gene regions tested is shown in Figure 4.

Table I: DNA fragments used to assay different DNA sequences for GFI and GFII binding

Gene	5'/3' flank	Region of gene contained		Binding of GFI	GFII	Reference
CIII-subII	5'	-408	-102	+	-	14
CIII-subVIII	5'	-274	+7	+	+	15
CYC1	5'	-680	-382	+	-	M.Smith, unpublished data
	5'	-381	-244	-	-	
	5'	-243	+4	-	-	
CIII-FeS	5'	-321	+204	-	+	13
CIII-subVI	5'	-489	-131[1)]	+	+	24
	5'	-127	+229	-	-	
PHO4	3'	-238	-40	+	-	24
PHO5	3'	-293	+81	+	-	16
TRP1	3'	-162	+46	+	-	34
URA3	5'	-226	-22	-	-	35
PHO5	5'	-540	-3[2)]	-	-	16
Mn-SOD	5'	-222	+19	-	-	36
	5'	-558	+19[2)]	-	-	

Regions of the genes contained are indicated relative to the translation start codon (5'-flank) and stop codon (3'-flank) respectively.
1) This fragment also contains a part of the 3' end of the coding region of the PHO4 gene (see for the sequence organization of the PHO4 -CIII-subVI gene region, Figure 4).
2) These regions were tested indirectly, using these fragments as specific competitor DNA and the CIII-subII gene, or the CIII-FeS gene 5' flanks as probes.

The first region tested was the intergenic region of the ANB1 and the CYC1 genes. These are transcribed in divergent directions and are oppositely regulated by oxygen (23). Fragments containing the TATA boxes and transcription initiation signals, or the UAS sequences of the CYC1 gene did not show either GFI or GFII binding (data not shown). However, a fragment containing sequences derived from the region upstream of the UAS elements binds GFI

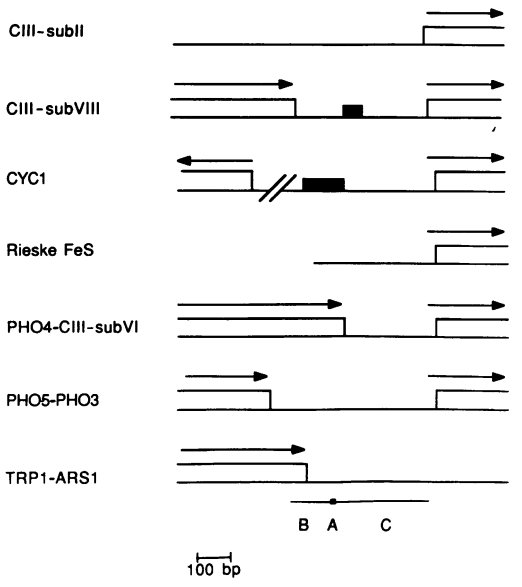

Fig.4: *The sequence organization of DNA regions tested for GFI and GFII binding*
Open boxes indicate protein-coding regions. Arrows indicate the direction of transcription. Filled boxes indicate UAS elements important for the induction of expression on a non-fermentable carbon source. For simplicity, the two UAS elements of the CYC1 gene (Guarente et al.,1984), are represented as a single bar. See Table I for references and text for further information.

according to the criteria given above (Figure 5). The 5' flank of the gene coding for the Rieske FeS protein of complex III was also analysed. The complex formed in this case was not influenced by the addition of the GFI binding oligonucleotide, but the addition of CENVI sequences resulted in competition (Figure 6a). An indirect footprint analysis shows that the protein involved binds to a region containing the sequence GTCACGTG, suggesting that the factor binding to the 5' flank of this gene is GFII (Figure 6b, Table III).

Fragments containing overlapping parts of the intergenic region of the PHO4-gene and the gene coding for subunit VI (17 kDa) of complex III were also examined (6,24). These two genes are separated by only 252 bp and transcribed in the same direction (Figure 4). A fragment containing sequences of the 5' and coding region of the gene coding for subunit VI did not show GFI or GFII binding. A fragment containing sequences located more upstream relative to the translational startcodon of the gene coding for subunit VI shows specific binding of GFI and GFII however (Figure 6c). This fragment contains only one region with sequences which resemble the CDEI consensus and it seems reasonable to assume that this region is the GFII binding site. The consensus sequence of CDEI is located just downstream of

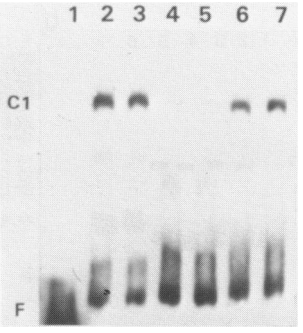

Fig.5: *Analysis of GFI complexes by gel retardation*
A ^{32}P-labelled fragment encompassing the 5' flanking region of the CYC1 gene (-680 to -382 relative to the ATG) was used as probe. As competitor DNAs were used: pEMBL9 (Lanes 2-3), unlabelled GFI binding oligonucleotide in a 20 molar excess (Lanes 4-5), and unlabelled subunit II gene 5' flanking sequences cleaved in the GFI binding site in a 20 molar excess (Lanes 6-7).

the translational stopcodon of PHO4. For a further analysis of the coding region of PHO4, see below.

The binding sites of GFI and GFII on the 5' flanking sequences of the CYC1 gene and the gene coding for subunit VIII are located upstream of their respective UAS sequences (1,7). Accordingly, these factors are clearly distinct from the yeast HAP1 activator and from RC2 which both bind to the UAS sequences of the CYC1 gene (2,25).

GFI can also bind to the 3' end of coding regions of genes, at least one of which is associated with ARS function

A fragment containing the 3' end of the PHO4-coding region and sequences downstream binds both GFI and GFII (Table I, CIII-subVI gene 5' flank, -489 -131 relative to ATG). It is thus possible that GFI has binding sites in the 3' end of the coding region of the PHO4 gene. To test this, a fragment containing only this region was used, with positive result (Figure 7a). The fact that in this case additional complexes are found is in agreement with the observation that several GFI binding motifs are present in this region (Table II).

This result prompted us to test other 3' ends of coding regions for GFI binding sites. Two genes of the PHO gene family, PHO5 and PHO3 are arranged in a tandem array and are transcribed in the same direction (16). The 3' end of the coding region of the PHO5 gene was tested for GFI binding, with positive result (data not shown). An indirect footprint analysis shows that the GFI binding site is located in the coding region of the gene (Figure 8b). The occurrence of binding sites in the coding regions of genes is not unprecedented in yeast. A factor designated SBF-B initially identified by its specific binding to the silencer region HMR-E, can also bind in the 3' end of the coding region of the TRP1 gene (26) and the binding site of this factor maps to a region, region B which is necessary for optimal ARS function (27). We

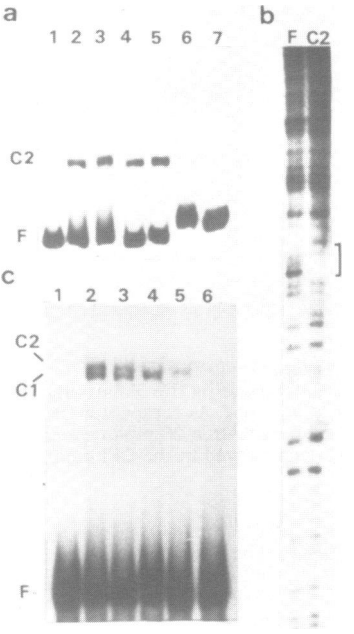

Fig.6: *Analysis of GFI and GFII complexes by gel retardation*

a) A ^{32}P-labelled fragment encompassing the 5' flanking region of the gene coding for the Rieske FeS protein of complex III (-321 to +204 relative to the ATG) was used. As competitor DNAs were used: pEMBL9 (Lanes 2-3), unlabelled GFI binding oligonucleotide in a 20 molar excess (Lanes 4-5), unlabelled CENVI sequences in a 20 molar excess (Lanes 6-7).

b) DNaseI footprint analysis of complexes in band C2. Lanes F and C2 show DNA eluted from the corresponding bands. Brackets indicate the protected region.

c) A ^{32}P-labelled fragment encompassing the 5' flanking region of the gene coding for subunit VI of complex III (-489 to -131 relative to the ATG) was used. As competitor DNAs were used: pEMBL9 (lanes 2-3), unlabelled CEN VI sequences in a 20 molar excess (Lane 4), unlabelled GFI binding oligonucleotide in a 20 molar excess (Lane 5), or a mixture of both, each in a 20 molar excess (Lane 6).

tested a fragment containing this region for GFI binding, again with positive result (Figure 7b). Furthermore, when the subunit II 5' flanking region was used as a probe, the resulting complex could be competed away by ARS1-TRP1 sequences, but not by the same sequences which were cut in the SBF-B binding site (data not shown). Thus, both factors GFI or GFII, identified by their interaction with the 5' flanking regions of nuclear genes coding for proteins of the mitochondrial respiratory chain, also bind to elements which play a role in the processes of cell division.

The location of these binding sites suggests that these elements might also play a role in termination of transcription extending from an upstream gene. The URA3 promoter region contains a sequence element that is capable of functioning as an efficient transcription terminator (28). The test of this fragment was, however, negative. In fact, we did not detect any

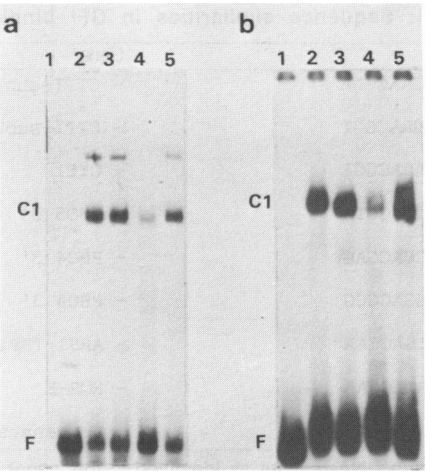

Fig.7: *Analysis of GFI complexes by gel retardation*

a) A ^{32}P-labelled fragment encompassing the 3' end of the coding region of the PHO4 gene (-238 to -40 relative to TGA) was used. As competitor DNAs were used: pEMBL9 (Lanes 2-3), unlabelled GFI binding oligonucleotide in a 20 molar excess (Lane 4), and unlabelled subunit II gene 5- flanking sequences cleaved in the GFI binding site in a 20 molar excess (Lane 5).

b) A ^{32}P-labelled fragment encompassing the 3' end of the coding region of the TRP1 gene (-162 to +46 relative to TAG) was used. The competitor DNAs used, were as indicated under a.

prominent binding, using the standard conditions (data not shown), suggesting that GFI and GFII are not involved in the transcription termination in this case. Two other 5' flanking regions, of the MnSOD and PHO5 genes, also lacked GFI and GFII binding sites (data not shown).

DNA sequence requirements for GFI and GFII binding

The analysis of the different binding sites of GFI permitted the identification of a common sequence motif. All sites contain two conserved elements, separated by a region which is variable in sequence but constant in length. This motif is also present in other regions shown to display GFI binding (Table II). Notably, the binding site of SBF-B which binds to region B of ARS1 also contains this motif (26).

This motif is necessary, but may not be sufficient to constitute a GFI binding site. For example, it is present in the 5' flanking sequence of the Mn-SOD gene, but we were unable to detect GFI binding. Further experiments will be necessary to clarify the matter.

The GFII binding regions (Table III) were also compared. The sites associated with CEN regions are flanked by A-T rich DNA regions, whereas those in the 5' flanking regions of genes are not. To what extent the A-T rich regions contribute to GFII binding is not known at present.

Table II: Sequence similarities in GFI binding sites

Sequence	Gene
ATCATTCCCAACGAA	- CIII-subII
GTCACGTGGAACGGT	- CIII-subVIII
GTCGTCTCACACGGA	- CYC1
ATCGTTAATGACGTA	- PHO5 3'
GTCATCGTCCACGAG	- PHO4 3'
GTCGCTTTGGACGCG	- PHO4 3'
AGCATTTTTGACGAA	- ARS1-TRP1
ATCATAAAATACGAA	- HMR-E
RTCRNNNNNNACGNR	- consensus sequence

Similarities in the GFI binding DNA regions as identified in this paper are compared with the sequences of the SBF-B binding sites in ARS-TRP1 and HMR-E, as identified by Shore et al. (26). The HMR-E sequences are negative regulatory elements, with properties opposite to those of enhancer sequences.

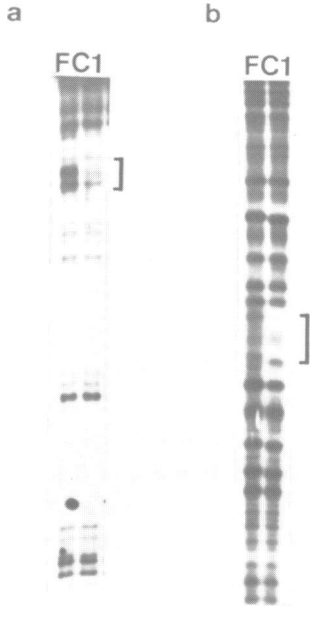

Fig.8: *Analysis of GFI complexes by indirect footprint analysis procedures*

a) A ^{32}P-labelled fragment encompassing the 5' flanking region of the CYC1 gene (-680 to -382 relative to the ATG) was used and a DNaseI footprint analysis of the complex in band C1 is shown. Lanes F and C1 show DNA eluted from the corresponding bands. A bracket indicates the protected region.

b) A ^{32}P-labelled fragment encompassing the 3' end of the coding region of the PHO5 gene (-293 to + 81 relative to the translational stop codon, TAG) was used and a DNaseI footprint analysis of the complex in band C1 is shown. Lanes F and C1 show DNA eluted from the corresponding bands. A bracket indicates the protected region.

Table III: Sequence similarities in GFII binding sites

Sequence	Gene
GTCACGTG	- CIII-subVIII
GTCACGTG	- CIII-subVI
GTCACGTG	- CIII-FeS
GTCACGTG	- GAL2
RTCACGTG	- consensus sequence

Similarities in the GFII binding DNA regions as identified in this paper, are compared with the CP1 binding site as identified by Bram and Kornberg (20). The orientation of the GFII binding sites in the subunit VIII and subunit VI gene 5' flanking regions has been inverted relative to the TATA box to allow optimal alignment.

DISCUSSION

We initially identified two factors, now designated as GFI and GFII as sequence-specific DNA binding proteins that interact with the 5' flanking sequences of genes coding for subunit II and subunit VIII of complex III. The analysis of additional binding sites shows that both factors can also bind to other genes and elements important for ARS and CEN activity (27,29).

Comparison of the sequences of the first two GFI binding sites identified did not provide enough information to formulate a consensus sequence. The discovery and analysis of additional sites has, however, permitted the identification of a fairly complex sequence motif, which is not only present in all the GFI binding DNA regions, but also in two sequences previously identified as binding sites for the factor SBF-B (26). This latter factor may thus be either identical with, or closely related to GFI. Based on the analysis of the two sites Brand et al. (30) proposed that SBF-B recognizes the sequence ACATCATAAAATA, which differs from the motif presented by us for GFI. This difference is likely to be entirely attributable to the limited number of SBF-B sites analyzed.

In the case of GFII, an important clue that this factor can also interact with other genetic elements came from the observation that the binding site present in the 5' flank of the gene coding for subunit VIII contains a motif identical to the CDEI element of yeast centromeres, namely RTCACGTG (19). A retardation competition experiment subsequently confirmed that GFII can interact with both elements and further screening resulted in the identification of additional sites in the 5' flanking regions of two other genes. CDEI is recognized by the protein CP1 (20), which may thus be identical with, or related to GFII.

Potential binding sites for both GFI and GFII are widespread in the yeast genome. A computer-assisted search of the EMBL data bank using the consensus sequences proposed in Tables II and III respectively, revealed approximately 30 matches for GFI and 6 for GFII in the vicinity of 190 genes transcribed by RNA polymerase II. These include 33 genes encoding proteins imported by mitochondria and of these, 11 (34%) contain binding sites for GFI and GFII. This contrasts with the situation for the remaining 157 genes, of which only 25 (16%)

contain binding sites for GFI and GFII. Although there is as yet no proof that all these sites are functional in vivo, this distribution suggests that the sites may be preferentially associated with genes coding for imported mitochondrial proteins, but not with any other immediately recognizable gene family. Obviously, further analysis is required to evaluate the significance of this finding, but it is interesting to speculate that the factors may be play a role in the coordination of the expression of genes involved in mitochondrial biosynthesis with progress of the cell cycle.

GFI and GFII clearly differ from regulatory DNA binding proteins initially identified by genetic approaches on which attention has mainly been focussed in yeast, e.g. GCN4, GAL4, and HAP1. The latter are non-essential for cell viability, are expressed at low levels and influence the expression of only a small set of genes (22,25,31). GFI and GFII are abundant, recognize a large set of target sequences and may play several roles in transcription and DNA replication, dependent on the context of their binding sites. Such context dependency is clearly seen in the case of the GFI-related factor SBF-B. When bound to its binding site, element B, SBF-B activates transcription. However, when bound to DNA in combination with a factor designated RAP1, which recognizes an adjacent element E repression of transcription is observed (30). A single basic mechanism may underly such different effects. For example, both GFI and GFII may facilitate interaction of DNA with additional proteins, thus allowing formation of the the higher order structures necessary for transcription, replication or segregation, respectively.

During preparation of this manuscript, publications describing the DNA sequence requirements of factors similar to GFI appeared (32,33). ABF1, the factor characterized by Buchman et al. (32), was first detected by its ability to recognize and bind to the negative regulatory element of the silent mating type locus HMR. Subsequent analysis showed that this factor is capable of binding to sites associated with a number of ARS sequences and with the HIS3/DED1 gene cluster. These sites appear to be variations on the consensus proposed in Table II. Interestingly, these authors were able to find only a single non-ARS binding site (in HIS3/DED1) amongst a selection of 10 yeast genes screened, whereas our analysis, directed primarily at genes coding for imported mitochondrial proteins has uncovered a further four sites. This observation lends support to the idea that GFI sites are preferentially associated with genes required for mitochondrial biogenesis.

The second publication describes a telomeric ARS-binding factor, which interacts with a DNA sequence resembling the consensus proposed for GFI (33). The fact that this factor shows no discernible affinity for either ARS1 or HMR-E could mean that it is either a modified form of ABF1 or GFI, or that it differs from both.

Acknowledgements

We are grateful to Drs M.Meijer and S.de Vries for critical reading of the manuscript and to Dr.M.Smith (University of British Columbia) for communicating to us the sequence of the 5' flank of the CYC1 gene prior to publication.

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 Guarente, L., Lalonde, B., Gifford, P. and Alani, E. (1984) Cell 36, 503-511.
2 Arcangioli, B. and Lescure, B. (1985) EMBO J. 4, 2627-2633.
3 Olesen, J., Hahn, S. and Guarente, L. (1987) Cell 51, 953-961.
4 Van Loon, A.P.G.M., De Groot, R.J., Van Eijk, E., Van der Horst, G.T.J. and Grivell, L.A. (1982) Gene 20, 323-337.
5 Van Loon, A.P.G.M., Maarse, A.C., Riezman, H. and Grivell, L.A. (1983) Gene 26, 261-272.
6 Van Loon, A.P.G.M., De Groot, R.J., De Haan, M., Dekker, A. and Grivell, L.A. (1984) EMBO J. 3, 1039-1043.
7 Maarse, A.C., De Haan, M., Bout, A. and Grivell, L.A. (1988) Nucl. Acids Res., in press.
8 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
9 Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl.Acids Res. 11, 1645-1655.
10 Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
11 Struhl, K., Stinchcomb, D.T., Scherer, S. and Davis, R.W. (1979)Proc.Natl.Acad Sci. USA 76, 1035-1039.
12 Guarente, L. and Ptashne, M. (1981) Proc.Natl.Acad.Sci.USA 78, 2199-2203.
13 Beckmann, J.D., Ljungdahl, P.O., Lopez, J.L. and Trumpower, B.L. (1987) J.Biol.Chem. 262, 8901-8909.
14 Oudshoorn, P., Van Steeg, H., Swinkels, B.W., Schoppink, P.J. and Grivell, L.A. (1987) Eur.J.Biochem. 163, 97-103.
15 Maarse, A.C. and Grivell, L.A. (1987) Eur.J.Biochem. 165, 419-425.
16 Bajwa, W., Meyhack, B., Rudolph, H., Schweingruber, A.-M. and Hinnen, A. (1984) Nucl.Acids Res. 12, 7721-7739.
17 Jones, E.W. (1977) Genetics 85, 23-33.
18 Garner, M.M. and Revzin, A. (1981) Nucl.Acids Res.9, 3047-3060.
19 Hieter, P., Pridmore, D., Hegemann, J.H., Thomas, M., Davis, R.W. and Philippsen, P. (1985) Cell 42, 913-921.
20 Bram, R.J. and Kornberg, R.D. (1987) Mol.Cell.Biol. 7, 403-409.
21 Arndt, K. and Fink, G.R. (1986) Proc.Natl.Acad.Sci.USA 83, 8516-8520.
22 Bram, R.J. and Kornberg, R.D. (1985) Proc.Natl.Acad.Sci.USA 82, 43-47.
23 Lowry, C.V., Weiss, J.L., Walthall, D.A. and Zitomer, R.S. (1983) Proc.Natl. Acad.Sci. USA 80, 151-155.
24 Legrain, M., De Wilde, M. and Hilger, F. (1986) Nucl.Acids Res. 14, 3059-3073.
25 Pfeifer, K., Arcangioli, B. and Guarente, L. (1987) Cell 49, 9-18.
26 Shore, D., Stillman, D.J., Brand, A.H. and Nasmyth, K.A. (1987) EMBO J. 6, 461-467.
27 Celniker, S.E., Sweder, K., Srienc, F., Bailey, J.E., Campbell, J.L. (1984) Mol.Cell. Biol. 4, 2455-2466.
28 Yarger, J.G., Armilei, G. and Gorman, M.C. (1986) Mol.Cell.Biol. 6,1095-1101.
29 Panzeri, L., Landonio, L., Stotz, A. and Philippsen, P. (1985) EMBO J. 4,1867-1874.
30 Brand, A.H., Micklem, G. and Nasmyth, K. (1987) Cell 51, 709-719.
31 Hope, I.A. and Struhl, K. (1985) Cell 43, 177-188.
32 Buchman, A.R., Kimmerly, W.J., Rine, J. and Kornberg, R.D. (1988) Mol.Cell.Biol. 8, 210-225.
33 Eisenberg, S., Civalier, C. and Tye, B.-K. (1988) Proc.Natl.Acad. Sci.USA 85, 743-746.
34 Tschumper, G. and Carbon, J. (1980) Gene 10, 157-166.
35 Rose, M., Grisafi, P. and Botstein, D. (1984) Gene 29, 113-124.
36 Marres, C.A.M., Van Loon, A.P.G.M., Oudshoorn, P., Van Steeg, H., Grivell, L.A. and Slater, E.C. (1985) Eur.J.Biochem. 147, 153-161.