The EMBO Journal vol.9 no.12 pp.4017-4026, 1990

CPF1, a yeast protein which functions in centromeres and promoters

J.Mellor, W.Jiang[1], M.Funk[1], J.Rathjen, C.A.Barnes, T.Hinz[1], J.H.Hegemann[1] and P.Philippsen[1]

Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK and [1]Institute of Microbiology and Molecular Biology, Frankfurter-Strasse 107, D-6300 Giessen, FRG

Communicated by P.Philippsen

Centromeres and several promoters of *Saccharomyces cerevisiae* contain a highly conserved octanucleotide, RTCACRTG, called CDEI. Using biochemical, genetic and structural analyses, we show that the same protein binds *in vivo* to CDEI sites in centromeres and in promoters. This protein, called CPF1 for centromere promoter factor, binds DNA as a dimer. Inactivation of the gene is not lethal but leads to a partial loss of the centromere function and to a Met$^-$ phenotype. Changes of the chromatin structure due to inactivation of CPF1 are seen at centromeres and at several CDEI-carrying promoters (e.g. *MET25, TRP1, GAL2*). However promoter activities are affected in diverse ways making it presently difficult to describe a function for CPF1 in gene expression. The sequence of the cloned gene reveals in the carboxy-terminal part two potential amphipathic helices preceded by a positively charged stretch of amino acids very similar to the helix-loop-helix domains recently identified in factors controlling tissue specific transcription in higher eukaryotes. Carboxy-terminal truncations of CPF1 lacking this domain no longer bind to CDEI. The amino-terminal half of CPF1 carries two clusters of negatively charged amino acid residues. Surprisingly, deletions of these clusters still render cells Met$^+$ and lead only to a marginal decrease in centromere activity.

Key words: centromere promoter factor/DNA binding proteins/*Saccharomyces cerevisiae*

Introduction

The purification of centromere binding proteins from the yeast *Saccharomyces cerevisiae* has been reported by three groups. A 16 kd protein (Cai and Davis, 1989), a 58 kd protein (Baker *et al.*, 1989) and 64 kd and 37 kd proteins (Jiang and Philippsen, 1989) were all identified by their sequence specific interaction with a highly conserved octanucleotide which marks the left boundary of the 120 bp centromere DNA (Fitzgerald-Hayes *et al.*, 1982; Hieter *et al.*, 1985; Cottarel *et al.*, 1989). This octanucleotide RTCACRTG, which is called centromere DNA element I (CDEI), is important for assembly of optimally functional *S.cerevisiae* centromeres since small but significant increases in mitotic loss rates were measured for centromere plasmids or chromosomes carrying deletions or point mutations in CDEI (Panzeri *et al.*, 1985; Cumberledge and Carbon, 1987; Hegemann *et al.*, 1988; Baker *et al.*, 1989).

Soon after CDEI was recognized as an important *cis*-acting DNA element in centromeres, this octanucleotide was found to be present in promoter regions of several *S.cerevisiae* genes and the possibility was discussed that the same protein binds to all CDEI sequences (Bram and Kornberg, 1987). In some cases structural or functional roles could be attributed to these promoter elements. For example, one of the three DNase I footprints seen with the *GAL2* promoter DNA mapped at a CDEI sequence (Bram *et al.*, 1986; Bram and Kornberg, 1987). Analysis of nuclear genes coding for subunits of the mitochondrial cytochrome *c* oxidoreductase revealed DNase I footprints at CDEI sites in two promoters (Dorsman *et al.*, 1988). Methylation of the CDEI sequence close to a site of RNA initiation in the *TRP1* promoter prevents a protein in yeast nuclear extracts binding to the site (Mellor *et al.*, 1988). Four genes in the pathway for S-adenosyl-methionine metabolism carry CDEI sequences in their promoters, and in the case of *MET25*, this sequence has been shown to be associated with an upstream activation site (Thomas *et al.*, 1989).

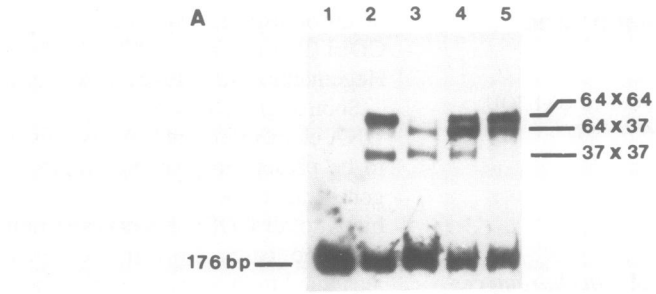
DNA motifs very similar to CDEI are functional elements in mammalian promoters, e.g. GTCATGTG in the immunoglobulin heavy chain enhancer (Ephrussi *et al.*, 1985), and specific protein binding to this sequence was observed in nuclear extracts of *S.cerevisiae* (Beckman and Kadesch, 1989). Furthermore, a mammalian DNA binding protein, called USF or MLTF, that functions to stimulate transcription from CDEI-like motifs in mammalian promoters such as the adenovirus major late promoter, also binds to CDEI sequences in *S.cerevisiae* (Carthew *et al.*, 1985; Sawadogo and Roeder, 1985; Bram and Kornberg, 1987; J.Mellor and S.J.Dowell, unpublished observation).

In *S.cerevisiae* it is not clear how the function of the protein binding to CDEI in centromeres relates to the proteins binding to CDEI-like sequences in promoters. Proteins of different sizes were identified which bind *in vitro* to CDEI or CDEI-like sequences (Bram and Kornberg, 1987; Baker *et al.*, 1989; Beckman and Kadesch, 1989; Cai and Davis, 1989; Jiang and Philippsen, 1989). It is possible that *in vivo* different proteins bind to centromeres and to promoters. However, it is also possible that the smaller proteins are proteolytic degradation products and that one and the same protein acts on functionally unrelated chromosomal loci. In order to clarify this and to learn more about centromeres and possibly transcription factors in *S.cerevisiae*, we compared the 37 kd and the 64 kd CDEI binding proteins previously isolated by us, cloned the gene encoding the 64 kd protein, constructed mutant strains with gene inactivating insertions or in-frame deletions, and analysed the phenotypic changes.

Results

Relationship between 37 kd and 64 kd CDEI binding proteins

Proteins binding to CDEI DNA sequences were isolated from *S.cerevisiae* as previously described (Jiang and Philippsen, 1989). The affinity purified material contained two major species of 64 kd and 37 kd. When both proteins were mixed, denatured in SDS and renatured, a hetero-oligomer formed as recognized by band-shift assays (Figure 1A). A new band appeared migrating between the bands representing the complexes with the homo-oligomers. This new band was not seen when the two proteins were mixed but not denatured (Figure 1A, lane 2). The most likely interpretation of this result is that the 64 kd and 37 kd proteins exist as stable dimers in solution, that they bind as dimers to DNA, and that they have similar or identical dimerization and DNA



B

```
1    GGATCCTTCGGGACCACCATGTCTACCCACCAAGGTCAAAAAAAAATAAAATCGGACGAG
61   AAAAGTACTGGGCAAACAACAATTATCCACGTGCTTATCACGTTTTTTTATTTAATGTAG
121  AATGTAAACCATTATACCTTAAAGATAAACACTTTGTTAATTATTCGTTTTAGTTGGCGG
181  TCATTTTCATTTCTGCTACAAGTACCAACATCAAGTGCTTAAAATATAATACGGTTTTCC
                          M  N  S  L  A  N  N  N  K  L  S  T  E  D  E   15
241  ACACTTTTATTAACGATGACCTCTCTGGCAAATAATAATAAGCTTTCTACTGAGGATGAG
      E  I  H  S  A  R  K  R  G  Y  N  E  E  Q  N  Y  S  E  A  R   35
301  GAAATCCATTCGGCAAGAAAAAGAGGCTATAACGAGGAGCAGAACTACAGCGAGGCCAGG
      K  K  Q  R  D  Q  G  L  L  S  Q  E  S  N  D  G  N  I  D  S   55
361  AAAAAGCAAAGAGATCAAGGCTTGTTGTCACAAGAGAGCAATGACGGAAATATTGACTCT
      A  L  L  S  E  G  A  T  L  K  G  T  Q  S  Q  Y  E  S  G  L   75
421  GCTCTGTTGAGCGAAGGGGCTACGCTAAAAGGGACTCAAAGTCAGTATGAAAGCGGACTG
      T  S  N  K  D  E  K  G  S  D  D  E  D  A  S  V  A  E  A  A   95
481  ACGTCTAACAAAGATGAGAAAGGAAGTGATGACGAAGATGCGTCAGTGGCTGAGGCTGCT
      V  A  A  T  V  N  *Y  T  D  L  I  Q  G  Q  E  D  S  S  D  A  115
541  GTTGCCGCTACCGTCAATTATACAGACTTAATCCAGGGCCAGGAAGATAGTTCTGATGCC
      H  T  S  N  Q  T  N  A  N  G  E  H  K  D  S  L  N  G  E  R  135
601  CACACATCTAATCAAACGAACGCGAATGGTGAGCACAAGGATTCCCTTAATGGGGAAAGA
      A  I  T  P  S  N  E  G  F  K  P  N  T  S  L  E  G  M  T  S  155
661  GCCATAACCCCATCAAATGAAGGATTCAAGCCAAACACCTCCCTGGAGGGAATGACGTCT
      S  P  M  E  S  T  Q  Q  S  K  N  D  M  L  I  P  L  A  E  H  175
721  TCACCAATGGAATCTACACAGCAATCTAAAAACGACATGCTCATACCATTGGCAGAACAC
      D  R  G  P  E  H  Q  Q  D  D  E  D  N  D  D  A  D  I  D  L  195
781  GATCGTGGGCCAGAACACCAACAGGATGATGAAGACAATGATGACGCTGACATTGATCTC
      K  K  D  I  S  M  Q  P  G  R  R  G  R  K  P  T  T  L  A  T  215
841  AAAAAGGATATAAGCATGCAGCCGGGTCGTCGTGGAAGAAAACCTACTACTTTGGCCACA
      T  D  E  W  K  K  Q  R  K  D  S  H  K  E  V  E  R  R  R  R  235
901  ACAGACGAGTGGAAAAAGCAAAGAAAAGACTCCCATAAAGAAGTCGAAAGGCGTCGCCGC
      E  N  I  N  T  A  I  N  V  L  S  D  L  L  P  V  R  E  S  S  255
961  GAAAACATCAACACTGCAATCAACGTTTTAAGCGACCTCCTGCCCGTGAGAGAATCAAGT
      K  A  A  I  L  A  R  A  A  E  Y  I  Q  K  L  K  E  T  D  E  275
1021 AAGGCAGCAATATTGGCGCGTGCCGCTGAATACATTCAAAAATTGAAAGAAACCGATGAA
      A  N  I  E  K  W  T  L  Q  K  L  L  S  E  Q  N  A  S  Q  L  295
1081 GCAAACATCGAAAAGTGGACGTTACAAAAATTGCTTAGCGAGCAAAACGCATCGCAATTA
      A  S  A  N  E  K  L  Q  E  E  L  G  N  A  Y  K  E  I  E  Y  315
1141 GCCAGTGCAAATGAGAAACTGCAGGAAGAACTGGGAAATGCTTACAAAGAAATCGAGTAC
      M  K  R  V  L  R  K  E  G  I  E  Y  E  D  M  H  T  H  K  K  335
1201 ATGAAACGCGTTTTAAGGAAGGAGGGAATAGAATACGAGGATATGCACACTCACAAGAAA
      Q  E  N  E  R  K  S  T  R  S  D  N  P  H  E  A                351
1261 CAAGAGAACGAAAGAAAAAGCACTAGGAGCGATAATCCACATGAGGCTTGAGAGTTAAAA
1321 ATAGATAGC
```

Fig. 1. Relationship between the CDEI binding proteins of 37 kd and 64 kd. **A.** Band-shift assay with 5 fmol of a 176 bp CEN6 DNA fragment (Cottarel *et al.*, 1989) and mixtures of the 37 kd protein (p37) and the 64 kd protein (p64). Lane 1, no protein added; lane 2, 10 ng p37 + 20 ng p64; lane 3, 10 ng p37 + 7 ng p64, denatured and renatured; lane 4, 10 ng p37 + 20 ng p64, denatured and renatured; lane 5, 3 ng p37 + 20 ng p64, denatured and renatured. The protein amounts represent estimates. **B.** Alignment of peptide sequences with the open reading frame of the *CPF1* gene. The nucleotide sequence starts at a *Bam*HI site, 255 bp upstream of the start codon and ends 18 bp downstream of the stop codon. Underlined nucleotides mark restriction sites relevant to experiments discussed in this paper. The portion of peptides which were sequenced are also underlined. Peptides isolated from p64 start at amino acids 24, 40, 66, 209 and 302, respectively. Peptides isolated from p37 start at amino acids 102 (amino-terminal peptide), 166, 209 and 236. Potential amphipathic helices are shaded.

binding domains in common. Both proteins were subject to tryptic digestion and partial sequence was obtained for five peptides from the 64 kd protein and four peptides from the 37 kd protein. One peptide sequence was found in both the 37 kd and the 64 kd protein indicating that the 37 kd protein is probably a degradation product of the 64 kd protein.

This was confirmed when the peptide sequences were compared with the amino acid sequence deduced from the cloned and sequenced gene of the 64 kd protein (Figure 1B). The tryptic peptide which starts at amino acid 102 is not preceded by an arginine or lysine. It therefore represents the amino terminus of the 37 kd protein. For reasons stated below the gene will, from now on, be referred to as the *CPF1* gene (centromere and promoter factor) and the product as CPF1 protein.

Characterization of the CPF1 gene

A restriction map surrounding the *CPF1* open reading frame is shown in Figure 2A. The length of the mRNA as determined by Northern analysis is ~1.7 kb (Figure 2B, right lanes). The transcription initiation sites were determined by primer extension. Multiple sites mapping from approximately −400 bp to +64 bp relative to the start codon were detected in three separate experiments with a major initiation site at −71 bp (data not shown).

The open reading frame of *CPF1* codes for a protein of 351 amino acids with a calculated molecular size of 39.5 kd. This is much smaller than the observed size of 64 kd for the CDEI binding protein isolated from *S.cerevisiae*. The difference may be due to post-translational modification or to an unusual migration behaviour in SDS gels or to both. The predicted amino acid sequence has some unusual properties which may influence the migration in SDS gels. The protein is substantially hydrophilic and has disproportionately high amounts of acidic (20%), basic (14%) and hydroxyl (15%) residues. The net overall negative charge is −21 and within the first 200 amino acids it is −25. When a fragment containing the open reading frame and 54 bp upstream of the ATG (see Figure 2A) is subjected to *in vitro* transcription and translation the majority of proteins synthesized migrate in SDS−PAGE at 60 kd. When a slightly shorter fragment, missing the 54 bp upstream of the ATG and the first 24 bp of the open reading frame, is analysed only background protein synthesis is observed (data not shown). This indicates that the ATG predicted as the start of the 39.5 kd open reading frame is sufficient to produce a protein with apparent molecular weight of 60 kd, very close to the mol. wt observed for the protein isolated from *S.cerevisiae*. The same result was obtained when the gene was expressed in *Escherichia coli* (Baker and Masison, 1990; J.Mellor, unpublished data). Therefore, no substantial post-translational modifications occur *in vivo*.

Inactivation of the CPF1 gene is not lethal

The DNA binding domain resides in the carboxy-terminal part of CPF1 since proteins synthesized by *in vitro* transcription and translation using *Msc*I cleaved fragment 22 (Figure 2A) no longer binds to CDEI (J.Mellor, unpublished data). A 1.1 kb fragment containing the yeast *URA3* gene was inserted upstream of the DNA binding domain at the *Msc*I site of the *CPF1* gene (Figure 2A). The resulting 4.1 kb *Bam*HI gene fragment was isolated and used to transform a homozygous *ura3* diploid strain (699/700)

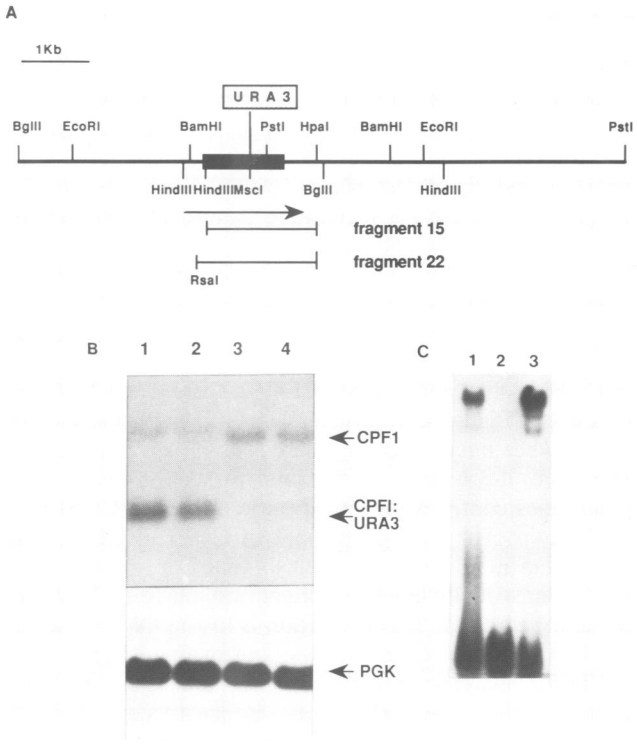


Fig. 2. Map of the *CPF1* locus and insertional inactivation of CPF1. **A** Restriction sites around the *CPF1* gene. The open reading frame is indicated as a black bar and the transcribed region as an arrow. The thin bars represent the extent of fragment 15 (*Hin*dIII−*Hpa*I) and fragment 22 (*Rsa*I−*Hpa*I), which were used as templates in *in vitro* translation assays as described in Materials and methods. For inactivation of the gene a 1.1 kb *URA3* fragment was inserted at the *Msc*I site. **B** Northern blot of total RNA from YAG90 (*cpf1::URA3*, lanes 1 and 2) and DBY745 (*CPF1*, lanes 3 and 4) grown in the presence of 1 mM L-methionine (lanes 1 and 3) or 0.01 mM L-methionine (lanes 2 and 4) probed with fragment 15. The *CPF1* transcript is ~1.7 kb. The same filter was reprobed with ^{32}P-labelled PGK DNA in order to visualize the 1.45 kb PGK mRNA which served as internal loading control. **C** DNA−protein complexes formed with nuclear extracts prepared from DBY745 (*CPF1*, lane 1); YAG90 (*cpf1::URA3*, lane 3); and YAG90 transformed with pMA3a (Dobson *et al.*, 1982) containing the *CPF1* gene (lane 3). Nuclear extracts were incubated with ^{32}P-labelled 23 bp CDEI oligonucleotide.

and a *ura3* haploid strain (DBY745) to uracil prototrophy and therefore to inactivate the *CPF1* gene by homologous recombination (Rothstein, 1983). Viable transformants were readily obtained after transformation of the haploid strain. Correct integration of the disrupted gene at the *CPF1* locus was verified by Southern blot analysis (data not shown). Northern blot analysis revealed two transcripts of ~1.8 kb and 1.0 kb (Figure 2B, left lanes). These probably represent transcripts promoted in *CPF1* and terminating at the 5′ end of the *URA3* gene (1.0 kb) or transcripts reading through and terminating at the 3′ end of the *URA3* gene (1.8 kb). In nuclear extracts prepared from the *cpf1* disruption strain YAG90 a CPF1−CDEI complex was no longer observed in a band-shift assay (Figure 2C, lane 2). No other complexes between DNA containing the CDEI sequence and proteins in the nuclear extract from the disruption strain were evident even after a long exposure.

Structural changes in centromeres and promoters of cpf1 strains

Genetic inactivation of the CPF1 protein could lead to measurable changes in the chromatin structure of CDEI sites. We looked for such structural changes in four centromeres and three promoters using as tools DNase I, micrococcal nuclease and dimethylsulphate. Examples are presented in Figure 3. The DNase I cleavage pattern of *S.cerevisiae* centromeres shows 150−160 bp of DNA protected from cleavage flanked by 60 bp regions which are hypersensitive to cleavage (Funk *et al.*, 1989). In a *cpf1* strain (YAG90), the protected region is ~10 bp shorter and the sequences which are additionally accessible to DNase I include part of CDEI as shown for CEN6, CEN10, CEN11, CEN14 (see example in Figure 4A). Structural changes are not so clearly visible when micrococcal nuclease is used (data not shown). However, the *in vivo* methylation protection analysis of guanosines confirms the site specific change at CDEI in a *cpf1* strain. The four guanosines in CDEI of CEN6 and one guanosine directly adjacent to CDEI are protected *in vivo* against methylation by dimethylsulphate (P.Wenink, unpublished data). This protection is lost in the *cpf1* strain as shown for the lower DNA strand in Figure 3B.

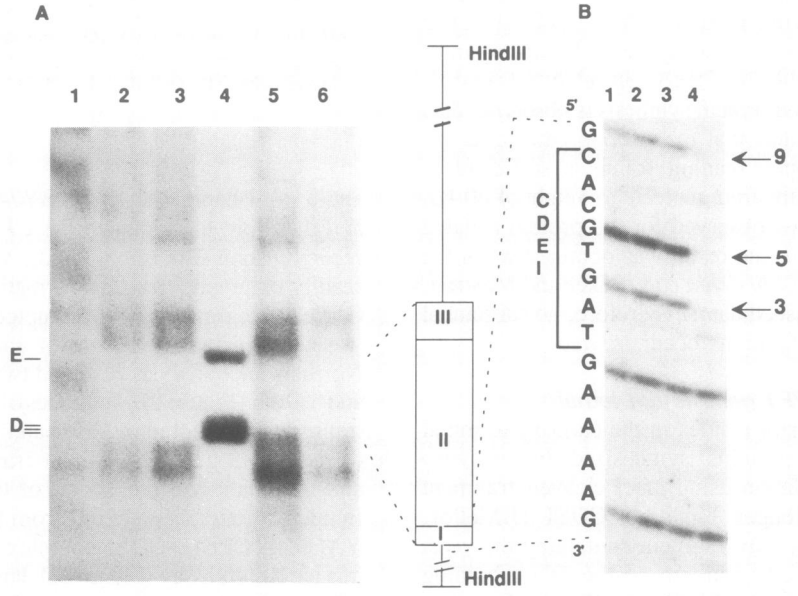
CDEI-containing promoters also undergo structural changes when the *CPF1* gene is inactivated. This is seen, for example, by probing the *TRP1* promoter with micrococcal nuclease (Figure 3C). One hypersensitive site at CDEI and one hypersensitive site 190 bp downstream marked with asterisks are lost in a *cpf1* strain. Similarly, hypersensitive sites close to the two CDEI sites of the *MET25* promoter disappear in the absence of CPF1 (data not shown). The micrococcal nuclease analysis of the *GAL2* promoter showed very few differences between *CPF1* and *cpf1* strains. However, *in vivo* methylation protection experiments clearly revealed structural changes at CDEI of *GAL2*. The protection of the guanosines in CDEI against methylation is lost in *cpf1* strains as shown in Figure 3D for the lower DNA strand of the *GAL2* promoter.

Functional changes in cpf1 strains

In order to quantify the influence of a *CPF1* inactivation on centromere activity, a mitotic CEN plasmid stability test was performed. To exclude strain background problems sometimes associated with plasmid assays we analysed the segregation behaviour in two pairs of isogenic strains (each *CPF1* and *cpf1*) with two different plasmids. The results presented in Table I show an ~5-fold increase in the rate at which the plasmid YCpL2 is lost and a 4-fold increase for the plasmid YEp42 in the *cpf1* strain. This result is similar to the rate at which plasmids carrying point mutations in, or deletion of, CDEI are lost in a *CPF1* strain (Panzeri *et al.*, 1985; Cumberledge and Carbon, 1987; Hegemann *et al.*, 1988; Baker *et al.*, 1989) and shows that the CPF1 protein is required for optimal centromere activity.

Colonies of the *cpf1* strain YAG90 were variable in size and the growth rate in rich (YPD) or synthetic complete media was slightly slower compared with the isogenic parental strain DBY745. When tested on minimal plates lacking one of each of the 14 amino acids present in synthetic complete media YAG90 could grow in the absence of tryptophane. This indicates that binding of CPF1 to the *TRP1* promoter is not essential for expression of the *TRP1* gene. However, these experiments revealed a methionine auxotrophy for *cpf1* strain YAG90. It is known that the promoter regions of *MET2, MET3, MET25* and *SAM2* each contain two CDEI or CDEI-like motifs (Thomas *et al.*, 1989). Therefore, CPF1 plays a role in the overall regulation of methionine and S-adenosyl methionine biosynthesis. We checked whether starvation for methionine affected *CPF1* gene expression. Northern blot analysis (Figure 2B) showed that the levels of the 1.7 kb *CPF1* transcript are indistinguishable in cells grown in the presence or absence of L-methionine.

No significant differences were observed between the isogenic *CPF1* and *cpf1* strains DBY745 and YAG90 during growth either on galactose or glycerol. This shows that the binding of CPF1 to the promoter of the galactose permease



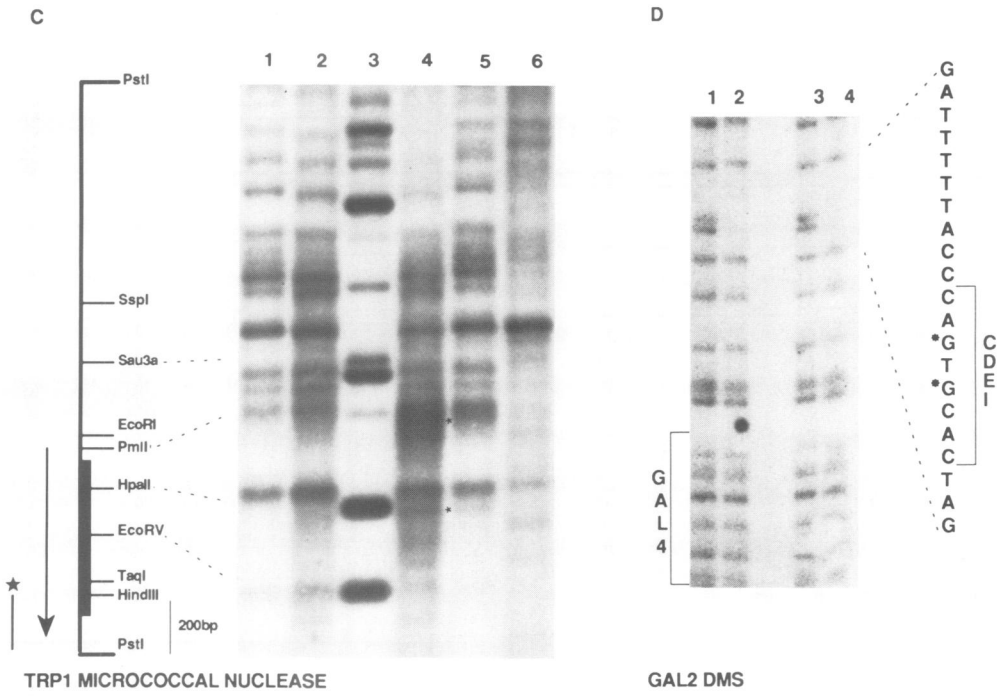


Fig. 3. Structural changes at CDEI sites in centromeres and promoters in *cpf1* strains. **A.** Mapping DNase I sensitive sites in the CEN14 region. Nuclei from DBY745 (*CPF1*, lanes 2 and 3) and from YAG90 (*cpf1*, lanes 5 and 6) were incubated for 10 min with 10 U/ml (lanes 2 and 6) or 20 U/ml (lanes 3 and 5) of DNase I. Aliquots of the isolated DNA were cleaved to completion with *Hin*dIII. As control purified DBY745 DNA was partially digested with DNase I followed by complete digestion with *Hin*dIII (lane 1). Size markers (lane 4) are CEN14 *Hin*dIII−*Dra*I fragments (D) of 422, 432 and 441 bp and a CEN14 *Hin*dIII−*Eco*RV fragment (E) of 560 bp. The box in the map represents the 120 bp long CEN DNA with its three conserved elements. *Dra*I sites lie directly adjacent to CDEI in the CDEII region of CEN14. The *Eco*RV site lies 27 bp outside of the CDEIII region. The DNA fragments were separated on a 1.8% agarose gel and probed with a 70 bp long *Hin*dIII−*Hin*cII fragment which maps adjacent to CDEI of CEN14. **B.** *In vivo* methylation protection of guanosines at CDEI in CEN6. Isolated DNA from strain YPH266 (lane 1) and cells from strain YAG90 (*cpf1*, lane 2), YSS90 (*cpf1*, lane 3) and YPH266 (*CPF1*, lane 4) were incubated with dimethylsulphate. Total DNA was afterwards cleaved at *Taq*I sites and in addition at methylated guanosines. The fragments were separated on an 8% sequencing gel, blotted and probed with ^{32}P-labelled RNA complementary to the lower strand of CEN6. **C.** Micrococcal nuclease sensitive sites around CDEI in the *TRP1* promoter. Nuclei from YAG90 (*cpf1*, lanes 1 and 2) and from DBY745 (*CPF1*, lanes 4 and 5) were incubated for 10 min with 6 U/ml (lanes 1 and 5) or 20 U/ml (lanes 2 and 4) micrococcal nuclease. Aliquots of the isolated DNA were cleaved to completion with *Pst*I. As a control purified DBY745 DNA was partially digested with micrococcal nuclease followed by complete digestion with *Pst*I (lane 6). Size markers (lane 3) were generated from *Pst*I cleaved chromosomal DNA by secondary partial cleavage with *Sau*3A (880 bp), *Pml*I which cuts in CDEI (759 bp), *Hpa*II (571 bp) and *Eco*RV (436 bp). The DNA fragments were separated in a 1.2% agarose gel and probed with a ^{32}P-labelled 208 bp *Pst*I−*Hin*dIII fragment from the 3′ end of the *TRP1* gene. The restriction map (Kim *et al.*, 1986) adjacent to the autoradiogram shows the position of the TRP1 open reading frame and the direction of transcription. **D.** *In vivo* methylation protection of guanosines around CDEI in the *GAL2* promoter. Cells from strain YAG90 (*cpf1*, lanes 1 and 3) and strain DBY745 (*CPF1*, lanes 2 and 4) were incubated with dimethylsulphate. Purified DNA was cleaved at *Hpa*I sites and at methylated guanosines. The fragments were separated on a 6% sequencing gel, blotted and probed with ^{32}P-labelled RNA complementary to the lower strand of a 145 bp *Hpa*I−*Eco*RV fragment which maps close to the start of the *GAL2* gene (Bram *et al.*, 1986). The sequence around CDEI and the position of the most upstream binding site for the positive regulator *GAL4* were taken from the same paper.

Table I. Mitotic plasmid loss rate in CPF1 and cpf1 strains

Strain background	Genotype	Plasmid	% LEU$^+$ cells after		No. of generations	Average plasmid loss rate per cell and generation
			0 gen.	*x* gen.		
DBY745	CPF1	YCpL2	94 ± 2	87 ± 1	10.0	7.7×10^{-3}
DBY745	cpf1	YCpL2	82 ± 3	56 ± 2	10.0	3.8×10^{-2}
YPH266	CPF1	YEp42	91 ± 2	83 ± 3	9.3	9.9×10^{-3}
YPH266	cpf1	YEp42	76 ± 4	53 ± 4	9.6	3.7×10^{-2}

Details of the mitotic plasmid stability assays are given in Materials and methods. The calculation of the average plasmid loss rate per cell and generation was done as follows: $Ln/L0 = (1 - R)^n$, Ln = number of colonies after n generations; L0 = number of colonies after 0 generations; R = loss rate; n = number of generations.

gene (*GAL2*) is not essential for the induction of this gene. Indeed, the *GAL2* mRNA levels are high in *CPF1* and *cpf1* strains grown on galactose. When grown on glucose, repression of transcription seems to be leaky in the *cpf1* strain (W.Jiang, unpublished data). The growth of the *cpf1* strain on a gluconeogenic carbon source, like glycerol, indicates that CPF1 is not essential for the expression of genes such as *LPD* (Ross, 1988) and two of the genes encoding subunits of the cytochrome *c* oxidoreductase which all carry CDEI sequences in their promoters (Dorsman *et al.*, 1988).

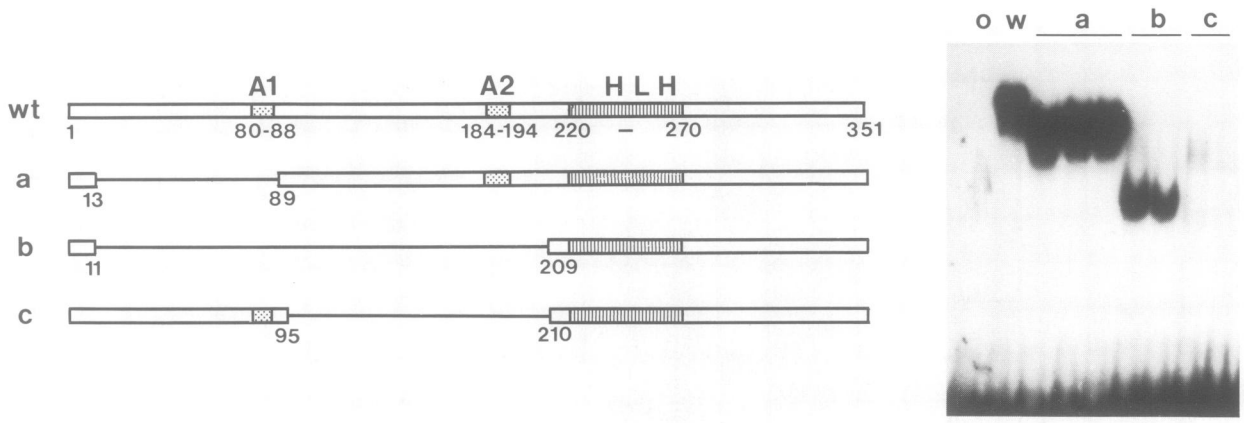


Fig. 4. Deletions in the amino-terminal half of *CPF1*. The drawings show the position of the two clusters of acidic amino acids (6 in A1 and 8 in A2) and the helix-loop-helix domain (HLH) which is preceded by a cluster of basic amino acids (9 out of 16). The extent of each deletion is indicated by the single lines. Numbers refer to amino acids in the wild-type protein. The autoradiogram is an example of a band-shift assay identifying binding activities to CDEI DNA. Crude extracts from several YSS91 transformants carrying plasmids with wild-type CPF1 (w) or one of the deletions (a, b, c) were incubated with a ^{32}P-labelled 75 bp CDEI fragment (Jiang and Philippsen, 1989) and separated on a 6% polyacrylamide gel. Lane 0 contains only the DNA fragment.

Table II. Phenotypes of CPF1 in-frame deletions expressed in strain YSS91

CEN plasmid	Plasmid-carried CPF1 derivative	Binding to CDEI[a]	Methionine phenotype	Loss of CEN plasmid in non-selective medium		
				No. of gen.	% cells with plasmid[b]	Loss rate[c]
pWJ306	CPF1 (wt)	+++	Met$^+$	11.8	85 ± 2.1 (8)	9.4×10^{-3}
pWJ306a	CPF1 (Δ14−88)	+++	Met$^+$	11.5	83 ± 3.6 (8)	11.7×10^{-3}
pWJ306b	CPF1 (Δ12−208)	+++	Met$^+$	10.9	80 ± 3.0 (6)	15.6×10^{-3}
pWJ306c	CPF1 (Δ96−209)	+	Met$^+$	11.6	84 ± 3.3 (8)	10.6×10^{-3}
pRS316	no CPF1	−	Met$^-$	11.4	32 ± 3.3 (4)	66×10^{-3}[d]

[a]Determined by band-shift assays as described in Figure 4.
[b]Several independent transformants were assayed, the numbers of which are given in parentheses.
[c]Calculated according to the formula given in Table I. At 0 generations ~95% of the cells carried a CEN plasmid with CPF1 or CPF1 derivative
[d]This number is two times higher than the loss rates in *cpf1* strains given in Table I. This difference is accounted for by the smaller size of pRS316 (4.8 kb) compared with Yep42 (11 kb) and YCpL2 (8 kb).

Characterization of in-frame deletions of the CPF1 gene

The experiments described so far leave no doubt that the CPF1 protein is a constituent of centromeres and several promoters in *S.cerevisiae*. In a first attempt to identify regions in CPF1 which are important for its function in addition to the carboxy-terminal DNA binding domain, we deleted the cluster of six and of eight negative charges at amino acids 80−88 and 184−194, respectively. The endpoints of the in-frame deletions are shown in Figure 4. The choice of the endpoints was governed by the available restriction sites and by the known protease sensitive regions around amino acid 102 (37 kd fragment, Figure 1) and amino acid 214 (16 kd fragment, Cai and Davis, 1990). We assumed that these regions may represent hinges between structural domains of the protein and that the introduced changes would not interfere with the folding of the DNA binding domain which starts after amino acid 220 (Cai and Davis, 1990).

The wild-type and the three mutated CPF1 genes were cloned into the CEN plasmids pRS315 (*LEU2* marker) and pRS316 (*URA3* marker) and subsequently expressed in strain YAG90 (*leu2, cpf1*) and YSS91 (*ura3, cpf1*), respectively. The chromosomal copy of *CPF1* was inactivated in YAG90 by *URA3* insertion and in YSS91 by replacement of 85% of the open reading frame with a *LEU2* carrying plasmid sequence.

CPF1 proteins lacking either one or both of the acidic clusters still bind to CDEI sequences as shown by band-shift assays. Examples are given in Figure 4 for transformants of strain YSS91. Only weak signals were obtained with the mutant lacking the second acidic cluster. This may be due to instability of the protein, low expression levels or weak CDEI binding affinity. Nevertheless, transformants with all mutant genes were Met$^+$ and cells retained the Met$^-$ phenotype after loss of the plasmids. Changes in centromere activity due to the in-frame deletions were determined by measuring CEN plasmid loss rate. Data are summarized in Table II. Surprisingly, no measurable differences were found between CPF1 wild-type and those mutants lacking one acidic cluster. Even deletion of amino acids 12−208 leads only to a marginal increase in CEN plasmid loss rate. This loss rate is still 2−4 times smaller than loss rates in strains lacking CPF1.

Discussion

Proteins of different sizes have previously been shown to bind *in vitro* specifically to fragments of DNA containing a short octanucleotide, RTCACRTG, found at the CDEI region of *S.cerevisiae* centromeres and upstream of a number of transcriptional units (Bram *et al.*, 1986; Bram and Kornberg, 1987; Baker *et al.*, 1989; Beckmann and

Kadesch, 1989; Cai and Davis, 1989; Jiang and Philippsen, 1989). In our hands there appear to be at least two forms of the CDEI binding protein that can be isolated from *S. cerevisiae*, a 64 kd and a 37 kd protein. This is in agreement with the observation that protein−DNA complexes formed with *S.cerevisiae* extracts and DNA fragments containing the CDEI octanucleotide often resolve in band-shift analyses into a doublet. Here we demonstrate that the 37 kd protein is an amino-terminal truncation of 101 amino acids of the 64 kd protein. This was concluded from a comparison of peptide sequence data with the sequence of the cloned gene. It is not clear whether there is any functional significance associated with a truncated form of the protein or whether the 37 kd protein represents a degradation product. No phenotypic changes are associated with an in-frame deletion mutant which expresses a protein very similar to the 37 kd protein. Nevertheless, we only refer to the 64 kd protein as centromere and promoter factor CPF1. This is the same protein called CP1 (centromere protein 1) by Bram and Kornberg (1987) and by Baker *et al.* (1989) who recently also characterized the gene (Baker and Masison, 1990). The 16 kd protein isolated by Cai and Davis (1989) which was named CBF1 (centromere binding factor 1) represents a degradation product as concluded from a comparison of its amino-terminal sequence with the gene sequence (Cai and Davis, 1990) and consists of the carboxy-terminal part of the 64 kd protein. The *CPF1* gene isolated by these two laboratories and by us encodes the same protein of 351 amino acids. It is open as to whether the 33−41 kd *S.cerevisiae* protein (YEB-3) which binds *in vitro* to a CDEI-like sequence (Beckmann and Kadesch, 1989) is a degradation product of CPF1 or a different protein.

Upon insertional inactivation of the unique *CPF1* gene no binding to CDEI DNA can be detected in crude extracts under our assay conditions. Despite this result it is not clear if *CPF1*, or a derivative thereof, represents the only RTCACRTG binding protein in *S.cerevisiae*. Recent data on the PHO4 protein, a transcriptional activator of the acid phosphatase gene (*PHO5*) in *S.cerevisiae*, suggests that it binds to two CDEI-like sequences, AGCACGTT and CACACGTG, on the *PHO5* promoter (Vogel *et al.*, 1989; Ogawa and Oshima, 1990). It is interesting that CPF1 in *S.cerevisiae* nuclear extracts binds tightly and specifically to the PHO4 binding sites on the *PHO5* promoter as seen in band-shift assays, yet experiments in *cpf1* and *pho4* strains show that CPF1 has little or no influence on the regulation or expression of *PHO5* (J.Mellor, unpublished data). This may indicate that factors in addition to primary DNA sequence influence CPF1−DNA interactions *in vivo*.

Although there may exist less abundant RTCACRTG binding proteins in *S.cerevisiae*, it is clear from the experiments presented in this paper that CPF1 binds to the CDEI sequence in centromeres. The supposition that the mitotic function of CDEI is mediated through CPF1 is supported by the data which show that the rate of loss of centromere-containing plasmids is increased 4- to 5-fold in *cpf1* strains. This rate of loss is within the 2- to 10-fold range found by Panzeri *et al.* (1985) and Hegemann *et al.* (1988) for point mutations and deletions in CDEI. It is also clear that at least in the *GAL2* promoter, probably also in the *TRP1* and *MET25* promoter, CPF1 is bound to the CDEI site *in vivo*. Thus, CPF1 does indeed encode a protein acting at *S.cerevisiae* centromeres and promoters as suggested by Bram and Kornberg (1987). However, the extent and significance of CPF1 interactions *in vivo* with CDEI-like sites identified close to *S.cerevisiae* genes remains to be seen.

The CPF1 protein carries in its amino-terminal part two clusters of negative charged amino acids which may well provide sites for interactions with other proteins. Such negatively charged domains, required for activation of transcription presumably through interactions with other proteins, have been well documented in yeast proteins such as *GAL4* (Gill and Ptashne, 1987; Ma and Ptashne, 1987). However, in contrast to *GAL4*, secondary structure predictions reveal no tendency of the acidic domains in the CPF1 protein to form negative amphipathic helices. The phenotype of *cpf1* strains suggests at first sight that the CPF1 protein may function directly or indirectly in the activation of transcription in a family of genes concerned with methionine metabolism. If so, such function does not require the two clusters of acidic amino acids since a mutant just expressing the carboxy-terminal 142 amino acids with the CDEI binding domain is Met^+. CDEI octanucleotides are found in the promoters of four co-regulated genes whose expression is induced in the absence of exogenously supplied methionine (Thomas *et al.*, 1989). Deletion of a region of the *MET25* promoter containing one of two CDEI octanucleotides is sufficient to cause a 90% decrease in enzyme activity and a similar reduction in RNA levels (Thomas *et al.*, 1989) strongly suggesting that UAS activity is associated with these motifs. However, previous reports and work in our laboratories have failed to find any evidence for transcriptional activation associated with DNA elements to which CPF1 binds. In fact, transcription is considerably repressed if the motif is positioned between the UAS and TATA box of a *GAL1−HIS3* gene fusion (Bram and Kornberg, 1987), or between the UAS and TATA box of the *S.cerevisiae PGK* gene. In addition, no activation of transcription is seen if the CDEI motif from *TRP1*, *SAM2* or *CEN3* is inserted into a UAS-less PGK promoter (J.Rathjen and J.Mellor, manuscript in preparation). Furthermore, it is possible to demonstrate induction of *MET25* transcription in strains such as YAG90 (*cpf1*) grown in very low concentrations of methionine which indicates that CPF1 is not essential for induction. The patterns of gene expression from other promoters known to contain CDEI motifs that bind to CPF1 *in vivo* seem to be unaffected by the presence or absence of the CPF1 protein. For instance, the presence or absence of the CPF1 protein has no influence on *TRP1* gene expression either on a plasmid or in the chromosome (J.Mellor and J.Rathjen, unpublished data). The expression of *GAL2* and *LPD1* also seem to be unaffected as both wild-type and *cpf1* strains grow on galactose and glycerol. Thus the question of the role of the protein in transcriptional activation awaits further experimentation, but we would favour an indirect role for CPF1, perhaps by modulating the chromatin structure of promoters allowing or preventing other proteins required for transcription to access the DNA.

Such a change in chromatin structure in the absence of the CPF1 protein is seen clearly on the *TRP1* and *MET25* promoters. The octanucleotide is found coincident with the initiation site for transcription from the downstream *TRP1* promoter (Kim *et al.*, 1986, 1988; Mellor *et al.*, 1988) and at this downstream promoter, the changes seen upon inactivation of *CPF1* mimic those seen close to a CDEI

```
                    basic      amphipathic                                          amphipathic
                   domain        helix I                                             helix II
                 ------------ ------------                                         ------------
                    *   *   ** *   *    *  *  *                                  ** **  **  **  *
I.    E12        KERRVANNAR ERLRVRDINE AFKELGRMCQ LHLNSEKP----------QT KLLILHQAVS VILNLEQQVR
I.    da         KERRQANNAR ERIRIRDINE ALKELGRMCM THLKSDKP----------QT KLGILNMAVE VIMTLEQQVR
I.    E47        RERRMANNAR ERVRVRDINE AFRELGRMCQ MHLKSDKA----------QT KLLILQQAVQ VILGLEQQVR
II.   MyoD       ADRRKAATMR ERRRLSKVNE AFETLKRCTS SNP---------------NQRLP KVEILRNAIR YIEGLQALLR
II.   myf-5      MDRRKAATMR ERRRLKKVNQ AFETLKRCTT TNP---------------NQRLP KVEILRNAIR YIESLQELLR
II.   AS-CT3     --ARR--NAR ERNRVKQVNN GFVNLRQHLP QTVVNSLSNG----GRGSSKKLS KVDTLRIYVE YIRGLQDMLD
II.   AS-CT4     -VQRR--NAR ERNRVKQVNN SFARLRQHIP QSIITDLTKG--G-GRGPHKKIS KVDTLRIAVE YIRSLQDLVD
II.   AS-CT5     -VIRR--NAR ERNRVKQVNN GFSQLRQHIP AAVIADLSNGRRGIGPGANKKLS KVSTLKMAVE YIRRLQKVLH
III.  lyl-1      -ARRVFTNSR ERWRQQNVNG AFAELRKLLP THP-------------PDRKLS KNEVLRLAMK YIGFLVRLLR
III.  N-myc      -ERRRNHNIL ERQRRNDLRS SFLTIRDHVP -----ELVK---------NEKAA KVVILKKATE YVHSLQAEEH
III.  L-myc      -TKRKNHNFL ERKRRNDLRS RFLAIRDQVP -----TLAS---------CSKAP KVVILSKALE YLQALVGAEK
III.  C-myc      -VKRRTHNVL ERQRRNELKR SFFALRDQIP -----ELEN---------NEKAP KVVILKKATA YILSVQAEEQ
      PHO4       DDKRESHKHA EQARRNRLAV ALHELASLIP ----AEWKQQNVSAA------PS KATTVEAACR YIRHLQQNGS
      CPF1       KQRKDSHKEV ERRRRENINT AINVLSDLLP -------VRESS---------- KAAILARAAE YIQKLKETDE
```

Fig. 5. Amino acid similarities between CPF1 and other helix-loop-helix proteins. Amino acids shared by yeast CPF1 (residues 221–275) and various other proteins are indicated. Identical amino acids found in most or all of the proteins are marked with asterisks. Those residues that form the proposed basic region and amphipathic helices are overlined with bars. The proteins are divided into classes as proposed by Murre *et al.* (1989). Additional data are taken from Chen *et al.* (1989). AS-C refers to the proteins of the *Drosophila achaete–scute* complex; da is the product of the *Drosophila* daughterless gene.

octanucleotide in the *MET25* promoter. A micrococcal nuclease sensitive chromatin structure is lost in the absence of the CPF1 protein. The DNase I accessibility at CDEI sites on centromere DNA actually increases in the absence of the CPF1, the opposite of the result seen with micrococcal nuclease in promoters. This apparent contradiction may be explained by the different proteins binding next to CDEI sites in centromeres and in promoters, by the different nucleases used, or by both. It is not yet clear how or whether the changes in nuclease sensitivity in the *TRP1* and *MET25* promoters influence gene expression. We favour the view that CPF1 may be functioning in a similar way to a factor known as REBI (Fedor *et al.*, 1988). REBI protects only 25 bp of DNA *in vitro*, yet appears to exclude nucleosomes from a 160 bp region *in vivo*. CPF1 and REBI may be members of a family of general transcription factors in yeast which function at a wide variety of sites. In mammals a transcription factor (USF, MLTF) has been characterized which stimulates transcription by binding to CCACGTGA, a CDEI-like sequence (Carthew *et al.*, 1985; Sawadogo and Roeder, 1985). The interaction of the mammalian USF protein and the *S.cerevisiae* CPF1 protein with this CDEI sequence are indistinguishable as judged by methylation interference binding assays (Chodosh *et al.*, 1989; Jiang and Philippsen, 1989). Although this does not necessarily indicate that CPF1 is the yeast homologue of the mammalian transcription factor, recent data concerning the molecular function of USF are intriguing. *In vitro*, binding of this factor can successfully compete with the assembly of promoter sequences into nucleosomes (Workman *et al.*, 1990).

The region of the CPF1 protein extending from amino acids 220 to 275 shows marked homology to a DNA binding and dimerization motif known as the helix-loop-helix (HLH) identified in a group of transcription regulators from mammalian cells and *Drosophila* involved in growth control and differentiation (Murre *et al.*, 1989 and Figure 5). These proteins can be divided into groups by their ability to bind *in vitro* to CDEI-like DNA as homodimers or heterodimers. The ubiquitous class I factors can form dimers with the tissue specific class II factors. The class II factors do not form dimers with other class II members. Proteins in class III, for instance the myc proteins, do not form heterodimers capable of binding to CDEI-like DNA with either class I or class II proteins. CPF1 is one of the first *S.cerevisiae* proteins identified with substantial homology to this type of DNA binding protein (Baker and Masison, 1990; Cai and Davis, 1990; this work). We show here that CPF1 is most likely to bind to DNA as a homodimer. Recent correction of the *S.cerevisiae PHO4* sequence also reveals a carboxy-terminal HLH domain preceded by a basic region in this transcription factor (Ogawa and Oshiwa, 1990; Figure 5). It may be possible that CPF1 and PHO4 form heterodimers under certain conditions. Both proteins recognize a similar octanucleotide present in yeast promoters and besides the HLH domain the PHO4 protein shows significant homology to the basic domain of CPF1 (Figure 5). SFL1, encoding a DNA binding protein involved in the suppression of flocculation, has been reported to have significant homology to the myc family of proteins but only two of the 13 residues in the HLH region are conserved in SFL1 (Fujita *et al.*, 1989). It is possible that more HLH proteins will be found in *S.cerevisiae*, which may form a family of homo- and heterodimeric DNA binding proteins equivalent to those reported in mammals and *Drosophila*.

Materials and methods

S.cerevisiae strains

DBY745 (α *leu2-3 leu2-112 ura3-52 ade2-1*) and 666/700 (a/α *ade2-1/ade2-1 trp1-1/trp1-1 can1-100/can1-100 leu2-3, 112/leu2-3, 112 his3-11, 15/his3-11, 15 ura3-52/ura3-52*) were obtained from S.Kearsey. YAG90 is isogenic with DBY745 and carries a URA3 insertion in the *Msc*I site of *CPF1*. YPH266 (a *ura3-52 lys2-801* amber *ade2-101* ochre *leu2-1*) was obtained from Philip Hieter. YSS90 is isogenic with YPH266 and carries a *URA3* insertion in the *Msc*I site of *CPF1*. YSS91 is isogenic with YPH266 and has 880 bp of the *CPF1* gene (*Hin*dIII–*Pst*I) replaced by the plasmid pRS305, which carries a *LEU2* marker (Sikorski and Hieter, 1989). Strains were grown in YPD or in selective medium as described by Sherman *et al.* (1986).

Peptide analysis of CPF1 protein

Proteins, purified from yeast as described by Jiang and Philippsen (1989), were digested with trypsin and peptides further purified by HPLC essentially as described by Aebersold *et al.* (1987) except that PVDF membranes were substituted for nitrocellulose (Matsudaira, 1987). Sequences were determined with an Applied Biosystems gas-phase sequenator.

CPF1 gene cloning and sequence analysis

CPF1 gene isolation was started initially by screening a λgt11 *S.cerevisiae* cDNA expression library using ^{32}P-labelled double stranded DNA with

three to five CPEI sites (Singh *et al.*, 1988; Vinson *et al.*, 1988). A number of presumptive positive clones were identified. During the analysis of these clones partial protein sequence information of CPF1 became available (this work and M.Cai and R.W.Davis, personal communication). This allowed the synthesis of redundant DNA oligonucleotides, which were used to rescreen the clones and in addition to screen a *S.cerevisiae* genomic DNA library. This library contained partially cleaved *Sau*3A fragments of 10–14 kb cloned into the *Bam*HI site of pMA3a (Dobson *et al.*, 1982). Because of the relative ease of handling plasmid DNA compared with phage λ DNA all further work was performed with the three positive clones obtained from the plasmid library. These clones contained overlapping restriction fragments, which were subcloned in M13 vectors and sequenced according to Sanger *et al.* (1977).

Primer extension analysis

RNA start site mapping was performed using primer extensions as modified by Rathjen and Mellor (1990). 7 ng of the end labelled oligonucleotide 5′-CGCTGTAGTTCTGCTCCTCG (positions +97 to +78 of the CPF1 open reading frame) was annealed to 5 μg of total RNA. Products were separated on 7.5% polyacrylamide urea gels with sequencing reactions as markers.

In vitro transcription and translation

Plasmids for generation of transcripts *in vitro* from the *CPF1* gene were constructed using the vector pSP73 (Promega). The *Pvu*II site was used to insert a filled in *Hin*dIII–*Hpa*I fragment (fragment 15; pSP73-15) or two fragments extending from *Rsa*I to *Hin*dIII and from *Hin*dIII to *Hpa*I (fragment 22; pSP73-22). Plasmids were linearized with *Bam*HI, *Pst*I (3′ overhang removed with mung bean nuclease) or *Msc*I prior to generation of *in vitro* transcribed mRNA using T7 RNA polymerase in the presence of m^7GpppG as described by the suppliers (TransProbe T kit; Pharmacia). The RNA transcribed from 1.5 μg of template DNA in a total reaction volume of 25 μl was redissolved in 12.5 μl of water. *In vitro* translation was performed using a micrococcal nuclease-treated rabbit reticulocyte lysate (Promega). A reaction mix was prepared using 17.5 μl of lysate, 0.5 μl of 0.1 M dithiothreitol, 1 μl of 1 mM amino acid mix minus methionine, 0.5 μl of RNasin (Promega), 2.5 μl of [^{35}S]methionine (Amersham, SJ204) and 2.5 μl of RNA, 1.5 μl water and incubated for 60 min at 37°C. 1 μl of each mix was analysed for complex formation and 3 μl resolved on a 7.5% SDS–polyacrylamide gel. Gels were fixed in 20% methanol, 7% acetic acid, treated with Amplify (Amersham), dried and autoradiographed at −70°C.

Band-shift analysis

Band-shift analyses were performed essentially as described by Jiang and Philippsen (1989) except that different ^{32}P-labelled CDEI fragments were used (see figure legends). The mixture of 37 kd and 64 kd proteins were denatured in 1% SDS, ethanol precipitated in the presence of bovine serum albumin, renatured from 6 M guanidine hydrochloride and passed through 250 μl Sephadex G-50 columns as described by Bram and Kornberg (1987).

Chromatin analysis

Cultures of strain DBY745 and YAG90 were grown to a density of 2×10^7 cells in YPD. Nuclei isolation, incubation and the DNA hybridizations were carried out as described by Funk *et al.* (1989). The hybridization probe for the *TRP1* gene was prepared by subcloning the 208 bp *Hin*dIII–*Pst*I fragment of the coding region into pBIISK+ (Stratagene), linearizing the DNA with *Pvu*II, and primer extending the denatured DNA with the sequencing and reverse sequencing primers in the presence of [α-^{32}P]dCTP (3000 Ci/mmol).

In vivo footprinting with DMS

The method used was adapted from the original protocol of Church and Gilbert (1984) and modifications of Saluz and Jost (1986). 50 ml cultures were grown to a cell density of 5×10^7 cells/ml. The cells were washed, concentrated to 5×10^9 cells/ml and treated with 0.5% DMS for 2 min. The reaction was stopped by dilution with 20 ml TEN (40 mM NaCl in TE). DNA was isolated and cut with *Taq*I for analysis of CEN6 DNA and with *Hpa*I for analysis of *GAL2* DNA. For control 5 μg of DNA isolated from strain S288C were methylated *in vitro* by treatment with 0.5% DMS in 200 μl of DMS buffer (50 mM Na-cacodylate, 1 mM EDTA) and then cut with the appropriate restriction nuclease. After piperidine cleavage, 5 μg of each DNA sample was separated on a 8% denaturing polyacrylamide gel, transferred to a Genescreen membrane and UV crosslinked. Probes were prepared in the following way. For CEN6, a *Taq*I–*Mbo*II fragment close to CDEI (Panzeri *et al.*, 1985) was subcloned into pGEMII to give pPW-95 (P.Wenink, unpublished data). For *GAL2*, an *Eco*RV–*Hpa*I (positions 104–248 on the sequence of Bram *et al.*, 1986) was subcloned into the *Sma*I site of pGEM1 (Promega). SP6 RNA polymerase and [α^{32}P]UTP were used to prepare labelled riboprobes from each linearized vector.

Construction of in-frame deletion mutants of CPF1

The 1.16 kb *Bam*HI–*Pst*I fragment of *CPF1* (see Figure 2A) was cloned into *Bam*HI and *Pst*I cleaved pUC19 which had its *Hin*dIII site inactivated. This plasmid (pWJ102) was start point for three in-frame deletions. For removal of amino acids 14–88 (deletion a) the gene was cleaved with *Hin*dIII (nucleotide 280) and at *Alw*NI (nucleotide 525) and the gap filled with the synthetic oligo duplex 5′-AGCTTTCTACTGAGGCGTCAGTGG/ 3′-AAGATGACTCCGCAGTC which has one *Hin*dIII and one *Alw*NI sticky end. For removal of amino acids 12–208 (deletion b) the gene was cleaved at *Hin*dIII (nucleotide 280) and *Bal*I (nucleotide 893) and the gap filled with the synthetic oligo duplex 5′-AGCTTTCTAAACCTACTACTTTGG/ 3′-AAGATTTGGATGATGAAACC which as one sticky *Hin*dIII end and one blunt *Bal*I end. For removal of amino acids 96–209 (deletion c) the gene was cleaved with *Alw*NI (nucleotide 525) and *Bal*I (nucleotide 893) and the gap filled with the synthetic oligo duplex 5′-CTGAGGCTGCTC-CTACTACTTTGG/3′-ACCGACTCCGACGAGGATGATGAAACC which has one sticky *Alw*NI end and one blunt *Bal*I end. All constructions were verified by restriction nuclease cleavage patterns. The *Bam*HI–*Pst*I fragments of deletions a, b and c were used to replace the *Bam*HI–*Pst*I fragment of the wild-type gene in pWJ303. This plasmid carries the 5 kb *CPF1 Bgl*II fragment cloned into the *Bam*HI site of pRS315 (*CEN, ARS, LEU2*). In a second series the *Bam*HI–*Esp*I fragments of deletions a, b and c were used to replace the *Bam*HI–*Esp*I fragment of the wild-type gene in pWJ306. This plasmid contains the 5 kb *CPF1 Bgl*II fragment cloned into the *Bam*HI site of pRS316 (*CEN, ARS, URA3*). The vectors pRS315 and pRS36 were described by Sikorski and Hieter (1989). The pWJ303 series was analysed in strain YAG90 and the pWJ306 series in strain YSS91. The integrity of plasmids after transformation was verified by Southern hybridizations to restriction fragments.

Mitotic stability assay

The stability assay was performed as described earlier (Hegemann *et al.*, 1988). The plasmids YCpL2 and YEp42 carry the *LEU2* gene, *CEN3* and *ARS1* and are 8 kb and 11 kb respectively (M.v.Pein and C.Hollenberg, unpublished). Plasmids pWJ306wt, a, b and c carry the *URA3* gene, *CEN6* and *ARSH4* and are 9.2 and 9.8 kb, depending on the size of the deletion. Plasmids were transformed into yeast and transformants were colony purified on minimal plates. After selective growth, eight independent transformants per yeast strain were taken for assay. The percentages of *LEU2* or *URA3* cells present in cultures were determined after ~10 generations of growth in non-selective liquid medium by plating on full medium and replica plating on selective medium. The number of cell doublings was calculated by counting the total cell number at the beginning and end of non-selective growth.

Acknowledgements

We would like to thank Simon Dowell for many useful discussions and access to his unpublished data; M.F.Tuite for yeast genomic library; M.v.Pein and C.Hollenberg for plasmids YEp42 and YCpL2; Colin Goding for useful discussions; Neil Wedlock for help with DNA sequencing; H.-G.Walker, M.Dreisbach and D.Linder for help with separation and sequencing of peptides; C.Reitz and K.Johnson for photographic work. We also acknowledge the exchange of peptide and DNA sequence data with M.Cai and R.Davis. W.J. thanks EMBO for support to attend the EMBO Course on protein microsequencing organized by J.Vandekerckhove in Gent. The experiments were done in partial fulfilment of the PhD theses of W.J. and M.F. J.M. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science. This work was supported by the Science and Engineering Research Council (GR/E85867), the Wellcome Trust and Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 272).

EMBL accession number

The accession number for the sequence reported in this paper is X52137.

References

Aebersold,R., Leavitt,J., Saavedra,R.A., Hood,L.E. and Kent,S.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6970–6974.

Baker,R.E. and Masison,D.C. (1990) *Mol. Cell. Biol.*, **10**, 2458–2467.

Baker,R.E., Fitzgerald-Hayes,M. and O'Brien,T.C. (1989) *J. Biol. Chem.*, **264**, 10843–10850.

Beckmann,H. and Kadesch,T. (1989) *Mol. Cell. Biol.*, **9**, 4535–4550.
Bram,R.J. and Kornberg,R.D. (1987) *Mol. Cell. Biol.*, **7**, 403–409.
Bram,R.J., Lue,N.F. and Kornberg,R.D. (1986) *Proc. Natl. Acad. Sci. USA*, **82**, 43–47.
Buchman,A.R. and Kornberg,R.D. (1990) *Mol. Cell. Biol.*, **9**, 887–897.
Cai,M. and Davies,R.W. (1989) *Mol. Cell. Biol.*, **9**, 2544–2550.
Cai,M. and Davis,R.W. (1990) *Cell*, **61**, 437–446.
Carthew,R.W., Chodosh,L.A. and Sharp,P.A. (1985) *Cell*, **43**, 439–448.
Chen,Q., Cheng,J.-T., Tsai,L.-H., Schneider,N., Buchanan,G., Carroll,A., Crist,W., Ozanne,B., Siciliano,M.J. and Baer,R. (1990) *EMBO J.*, **9**, 415–424.
Chodosh,L.A., Buratowski,S. and Sharp,P.A. (1989) *Mol. Cell. Biol.*, **9**, 820–822.
Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
Cottarel,G., Shero,J.H., Hieter,P. and Hegemann,J.H. (1989) *Mol. Cell. Biol.*, **9**, 3342–3349.
Cumberledge,S. and Carbon,J. (1987) *Genetics*, **117**, 203–212.
Dobson,M.J., Tuite,M.F., Roberts,N.A., Kingsman,A.J., Kingsman,S.M., Perkins,R.E., Conroy,S.C., Dunbar,B. and Fothergill,L.A. (1982) *Nucleic Acids Res.*, **10**, 2625–2637.
Dorsman,J.C., van Heeswijk,W.C. and Grivell,L.A. (1988) *Nucleic Acids Res.*, **16**, 7287–7301.
Ephrussi,A., Church,G.M., Tonegawa,S. and Gilbert,W. (1985) *Science*, **227**, 134–140.
Fedor,M.J., Luc,N.F. and Kornberg,R.D. (1988) *J. Mol. Biol.*, **204**, 109–127.
Fitzgerald-Hayes,M., Clarke,L. and Carbon,J. (1982) *Cell*, **29**, 235–244.
Fujita,A., Kikuchi,Y., Kuhara,S., Misumi,Y., Matsumoto,S. and Kobayashi,H. (1989) *Gene*, **85**, 321–328.
Funk,M., Hegemann,J.H. and Philippsen,P. (1989) *Mol. Gen. Genet.*, **219**, 153–160.
Gill,G. and Ptashne,M. (1987) *Cell*, **57**, 121–126.
Hegemann,J.H., Shero,J.H., Cottarel,G., Philippsen,P. and Hieter,P. (1988) *Mol. Cell. Biol.*, **8**, 2523–2535.
Hieter,P., Pridmore,D., Hegemann,J.H., Thomas,H., Davis,R.W. and Philippsen,P. (1985) *Cell*, **42**, 913–921.
Jiang,W. and Philippsen,P. (1989) *Mol. Cell. Biol.*, **9**, 5585–5593.
Kim,S., Mellor,J., Kingsman,A.J. and Kingsman,S.M. (1986) *Mol. Cell. Biol.*, **6**, 4251–4258.
Kim,S., Mellor,J., Kingsman,A.J. and Kingsman,S.M. (1988) *Mol. Gen. Genet.*, **211**, 472–476.
Ma,J. and Ptashne,M. (1987) *Cell*, **48**, 847–853.
Matsudaira,P. (1987) *J. Biol. Chem.*, **206**, 10035–10038.
Mellor,J., Dowell,S.J. and Rathjen,J. (1988) *Yeast*, **4**, S411.
Murre,C., McCaw,P.S, Vaessin,H., Caudy,M., Jan,L.Y., Jan,Y.N., Cabrerar,C.V., Buskin,J.N., Hauschka,S.D., Lasser,A.B., Weintraub,H. and Baltimore,D. (1989) *Cell*, **58**, 537–544.
Ogawa,N. and Oshima,Y. (1990) *Mol. Cell. Biol.*, **10**, 2224–2236.
Panzeri,L., Landonio,L., Stotz,A. and Philippsen,P. (1985) *EMBO J.*, **4**, 1867–1874.
Rathjen,J. and Mellor,J. (1990) *Nucleic Acids Res.*, **18**, 3219–3225.
Ross,J. (1988) *Regulation of the gene encoding lipoamide dehydrogenase in Saccharomyces cerevisiae*. Ph.D Thesis, University of Edinburgh.
Rothstein,R.J. (1983) *Methods Enzymol.*, **101**, 202–211.
Salluz,H. and Jost,J.-P. (1986) *Gene*, **42**, 151–157.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165–175.
Sherman,F., Fink,G.R. and Hicks,J.B. (1986) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sikorski,R. and Hieter,P. (1989) *Genetics*, **122**, 19–27.
Singh,H., LeBowitz,J.H., Baldwin,A.S. and Sharp,P.A. (1988) *Cell*, **52**, 415–423.
Thomas,D., Cherest,H. and Surdin-Kerjan,Y. (1989) *Mol. Cell. Biol.*, **9**, 3292–3298.
Vinson,C.R., LaMarco,K.L., Johnson,P.P., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.
Vogel,K., Hörz,W. and Hinnen,A. (1989) *Mol. Cell. Biol.*, **9**, 2050–2057.
Workman,J.L., Roeder,R.G. and Kingston,R.G. (1990) *EMBO J.*, **9**, 1299–1308.

Received on July 9, 1990; revised on September 12, 1990