

The centromere and promoter factor 1 of yeast contains a dimerisation domain located carboxy-terminal to the bHLH domain

Simon J.Dowell⁺, Jimmy S.H.Tsang and Jane Mellor*

Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK

Received May 20, 1992; Revised and Accepted July 13, 1992

ABSTRACT

CPF1 is a basic helix–loop–helix (bHLH) protein required for optimal centromere function and for maintaining methionine independent growth in yeast. In this work, we show that the region carboxy-terminal to the bHLH domain of CPF1 is essential for CPF1 function in the cell and for dimerisation of CPF1 in solution. The C-terminus of CPF1 contains a potential long amphipathic helix with a hydrophobic face which could provide a suitable protein:protein interface. Point mutations in residues forming this hydrophobic face are sufficient to weaken the interaction between the protein and DNA. By fusing the DNA binding domain or the transcriptional activation domain of GAL4 to the C-terminal 87 amino acids of CPF1, we show that this region is sufficient for mediating protein:protein interactions *in vivo*. The C-terminal domain of CPF1 can be replaced by the leucine repeat region of the bHLH-ZIP protein USF and the hybrid CPF1-USF protein functions *in vivo* to provide normal centromere function and methionine independent growth. However, the CPF1-USF hybrid protein is unable to interact with CPF1 suggesting that a dimer of CPF1 is sufficient for maintaining methionine independent growth and normal centromere function.

INTRODUCTION

The centromere and promoter factor (CPF1) binds *in vivo* and *in vitro* to the CDEI octanucleotide (RTCACRTG) found in centromeres and several promoters of genes in the budding yeast *Saccharomyces cerevisiae* (1–8). CPF1 contains a basic helix–loop–helix (bHLH) domain which is a DNA binding and dimerisation motif recently identified in a rapidly expanding family of proteins found in yeast, *Drosophila*, plants and mammals (9–12). More than 60 genes potentially encoding HLH proteins have been reported and almost every member of this class has been implicated in transcriptional regulation, oncogenesis and/or cell type determination and differentiation (for reviews see 13, 14). One feature of some HLH proteins is their

capacity to form either homodimeric or heterodimeric complexes through the HLH domain with other members of this family thereby modulating their DNA binding capacity (10, 15). It has been suggested that the heteromers that form between the ubiquitously expressed HLH proteins (such as E12 and *daughterless*) and the tissue specific HLH proteins (such as myoD and achaete-scute T3) bind with higher affinity than the homodimer to the tissue-specific target sequence (CANNTG) common to the bHLH proteins thus providing the potential for a complex regulatory network for directing gene expression in eukaryotic cells (10).

An analogous pattern of specialized protein–protein oligomerisation is also a feature of the leucine zipper transcriptional regulators (such as jun, fos and GCN4) which contain a dimerisation domain comprising an alpha-helical coiled-coil-like leucine repeat structure adjacent to a basic DNA-binding motif (16–19, for reviews see 20–22). A feature of the leucine zipper is the repeat of hydrophobic non-leucine residues at the 3–4 position of the heptad leucine repeat and laterally aligned charged amino acids that are proposed to effect the specificity of interactions and prevent promiscuous dimer formation (18, 23–25). A subgroup of the bHLH proteins, the bHLH-ZIP proteins such as the myc family (N- C- L- myc and max), USF, TFE3 and AP4, also contain leucine repeat structures which function as dimerisation domains (26–30).

Gene disruptions of the *CPF1* locus on chromosome X result in two clear phenotypes; a partial loss of centromere function revealed by the increased rate of loss of a centromere based plasmid in strains lacking CPF1 and an absolute requirement for methionine for growth (4, 7, 8). As duplicate CDEI motifs are found in the promoters of at least four co-regulated genes (*MET25*, *MET2*, *MET3* and *SAM2*) whose products are involved in methionine biosynthesis, it was suggested that CPF1 might be functioning as a transcriptional regulator of this pathway (4, 7, 8, 31). Strains of yeast containing a deletion of CPF1 extending from amino acids 10 to 209 that removes two acidic rich regions, which are a feature of many transcriptional activator proteins (32, 33, 34), have normal centromere function and are methionine prototrophs (8). Although CPF1 may be required, together with

* To whom correspondence should be addressed

⁺ Present address: ICRF Clare Hall Laboratories, Blanche Lane, South Mimms, Potter's Bar, Herts EN6 3LD, UK

MET4, for activation of transcription through elements normally found in the promoters of some methionine biosynthetic genes, DNA bound CPF1 is not a transcriptional activator (35). Furthermore, strains expressing a non-DNA binding form of CPF1 made by introducing a point mutation in the basic domain, are methionine prototrophs but have sub-optimal centromere function (36). This result indicates that DNA binding of CPF1 is not essential for maintaining methionine independent growth but is essential for optimal centromere function. Therefore, there may be multiple functions associated with CPF1 and each function may require CPF1 in a different conformation. In this paper we begin a functional analysis of CPF1 and show that the protein contains an essential dimerisation domain located C-terminal to the bHLH.

MATERIALS AND METHODS

Strains, media and growth conditions

S. cerevisiae strains. DBY745 (alpha, *leu2-3*, *leu2-112*, *ura3-52*, *ade1-100*). Strain JT1 is a haploid obtained by sporulation of a diploid strain crossing YAG93 with GGY1::171 (37). The phenotype of JT1 is *cpf1-*, *gal4-*, *leu2-*, *his3-*, *GAL1-LacZ::URA3*. YAG90 is isogenic with DBY745 but carries a *URA3* insertion in the *MscI* site of *CPF1*. Versions of the *CPF1* gene carrying point mutations or deletions were integrated at the *CPF1* locus of YAG90 using 5' fluoro-orotic acid selection (38, 39). 30 μ g of linear fragments containing the *CPF1* gene or mutant derivatives were transformed using lithium acetate into YAG90 cells at a density of no more than 5×10^6 /ml. Cells were replica plated from YEPD onto defined minimal plates containing 5'FOA at 750 μ g/ml and uracil at 40 μ g/ml after 12 to 16 hours growth at 30°C. After 2 to 3 days on 5'FOA plates, the cells were replica plated back to YEPD. Strains were transformed following the procedure of Hinnen *et al.* to leucine prototrophy with the *CEN* based plasmid pRS315 (40).

DNA analysis

Genomic DNA was isolated from uracil auxotrophs using the method of Holm and the *CPF1* locus analysed by PCR using 0.1 μ g of oligonucleotides BB4357 5'-CACCATGTCTCACCCAC-3' and BB4356 5'-CTCAAGCCTCATGTGG-3', 1 to 2 μ g of genomic DNA, 2.5 U of Taq Polymerase (Promega), 0.2 mM dNTPs (Pharmacia) in a total reaction volume of 100 μ l in reaction buffer provided for 38 cycles at 94°C for 90 seconds, 55°C for 60 seconds and 72°C for 30 seconds. The PCR product was isolated from agarose, subcloned into pSP73 and sequenced to confirm the presence of point mutations. Large deletions in the *CPF1* locus were detected by Southern blotting *Bgl*II cut genomic DNA.

DNA manipulations and plasmid constructions

Restriction enzymes, T4 DNA ligase and Klenow were obtained from BRL, Inc., T7-GEN *in vitro* mutagenesis kit and sequenase V2.0 from USB, Erase-a base kit from Promega. A *Sph*I/*Rsa*I and an *Pst*I/*Eco*RI fragment from *CPF1* on pSP73-22 were cloned into M13mp18 as templates for mutagenesis. Radioactively labelled isotopes were obtained from Amersham International. Deletions were introduced into a *CPF1* fragment cloned into pGEM5Zf(+) after cutting with *Nsi*I and *Bgl*II or *Apa*I and *Nco*I and ligating a *Bgl*II linker (CAAAAGATCTTTTG). Different half molecules of *CPF1* were ligated together at the *Bgl*II linker and *Eco*RI/*Hind*III fragments used to replace the *Eco*RI/*Hind*III

fragment in pSP73-22. Molecule 93 was made by filling-in the *Hind*III site at amino acid 10 and ligating it to the *Hpa*I site located in the 3' non-coding region of *CPF1*. A *Bam*HI fragment (the 3 Kb fragment but containing the deletion of the open reading frame) was used to transform strain YAG90 to uracil auxotrophy. Deletion 317 was made by linking the *Pst*I/*Mae*I sites in *CPF1* in frame using oligonucleotides BB4943 5'-TAGATCTTCC-TGCA-3' and BB4942 5'-GGAAAGATC-3'. Molecule 323 was created by inserting BB4711 5'-GAGATCTCTGCA-3' at the *Pst*I site of *CPF1*. Molecule 341 was made by ligating a *Pst*I fragment from USF containing amino acids 272 to 310 at the *Pst*I site of *CPF1*. The newly introduced amino acids at deletion end points and junctions are shown below for each molecule (number underlined). The amino acids underlined are those encoded by the linkers inserted at end points or as a result of truncations. The ten amino acids preceding and after the junctions are shown together with the position of the first and last amino acids shown.

300:	1-MNSLANNNKLRSEATTDEWKKQR-223
309:	207-GRKPTTLATTAKDRSFDINTAINVLS-247
324:	233-RRRENINTAITKDLQIFWARAAEYIQKL-270
317:	293-SQLASANEKLRQERSRSDNPHEA-351
323:	293-SQLASANEKLRSLQEELGNAYKE-312
311:	334-KKQENERKSTTK
312:	333-HKKQENERKSTTK
313:	305-ELGNAYKEIEAK
314:	284-QKLLSEQNASPK
315:	233-RRRENINTAITK

All versions of *CPF1* were subcloned into pSP73 (Promega), linearised with *Bgl*II or *Eco*RI prior to *in vitro* transcription with T7 RNA polymerase. Fragments for yeast transformation were isolated from versions of the *CPF1* gene in pSP73 or pGEM5Zf(+). Oligonucleotides used to introduce point mutations into *CPF1* are:

Leu-Met 306	BB6357	5'-GGAAGAAATGGGAAATGC-3'
Ile-Met 313	BB5765	5'-CAAAGAAATGGAGTAC-3'
Leu-Met 320	BB5766	5'-GCGTTATGAGGAAGG-3'

Plasmid pGAD1F was supplied by S.Fields (41). Plasmid p13H was constructed by replacing a *Bgl*II fragment, which contains a *TRP1* marker on p13T, by a *Bam*HI fragment which contains a *HIS3* marker from pMA700. Plasmid p13T is a derivative of pKW13 carrying DNA sequences coding for the GAL4₁₋₁₄₇, which was made by replacing an *Eco*RI/*Hind*III fragment with the *Eco*RI to *Hind*III polylinker region of pSP73. Plasmids p13-DD and pG-DD were constructed by cloning a *Bam*HI/*Bgl*II fragment, which contains DNA sequences coding for CPF1₂₆₆₋₃₅₁ from pSP73-DD, into the *Bam*HI site of p13H and pGAD1F, respectively. In these two constructs CPF1₂₆₆₋₃₅₁ was inserted in frame with the GAL4₁₋₁₄₇ and the GAL4₇₆₈₋₈₈₁, respectively. Plasmid pSP73-DD was made by cloning a 1 Kb *Xmn*I fragment containing DNA sequences coding for CPF1₂₆₆₋₃₅₁ into the *Sma*I site of pSP73.

Analysis of mutations in CPF1

In vitro transcription and translation, nuclear protein preparations, band shift analysis, growth rate analysis and the mitotic stability assay were carried out as described earlier (36).

Cross-linking with dimethylsuberimidate (DMSI)

This method was adapted from that published by Hu *et al.* (23). 2 μ l of translation mix (containing approx. 100 μ g total protein) was diluted into 200 μ l of 200 mM phosphate pH 8.5 containing DMSI (Pierce) at 12 mM. Samples were incubated for 30 minutes

at room temperature, concentrated by precipitation with an equal volume of 20% trichloroacetic acid, washed with acetone and subjected to electrophoresis on either 10% or 12.5% SDS polyacrylamide gels.

RESULTS

The bHLH domain and the carboxyterminus are essential for CPF1 function in yeast

Working with a simple eukaryote such as yeast, it is possible to investigate how mutations affect the function of a protein in the cell by replacing a wild-type version of the gene with a defective one. A number of different deletions of CPF1 were created and analysed in yeast but only four, 300, 309, 324 and 317 are discussed here (42; Figure 1). To assess the phenotypes associated with mutations in *CPF1* *in vivo*, the technique of gene replacement together with 5'-fluoro-orotic acid (FOA) selection was used to target the mutated versions of the *CPF1* gene into the *CPF1* locus marked with the *URA3* gene in strain YAG90 (38, 39). The presence of mutations in the *CPF1* locus was confirmed by restriction digestion of PCR products (data not shown). We analysed the effect of a deletion of amino acids 302 to 343 at the C-terminus of CPF1 in yeast strain YAG317. This strain was compared with strains YAG300, YAG309 and YAG324 containing deletions of the amino terminal amino acids 10 to 209, the basic domain (amino acids 217 to 237) and the HLH domain (amino acids 243 to 260) of CPF1 respectively (Figure 1).

We had shown previously that the first 209 amino acids are dispensable for CPF1 function (8). Strain YAG300 contains a deletion of amino acids 10 to 209 of CPF1 and has essentially wild type characteristics, similar to the parental strain DBY745. However, strains YAG309, YAG324 and YAG317 are methionine auxotrophs and show suboptimal centromere function, similar to strain YAG90 containing an interruption of the CPF1 gene (8; Figure 1). These results indicate that the first 200 amino acids of CPF1 are dispensable for function *in vivo* but that the basic region, the HLH and the carboxyterminus are all essential for CPF1 function in the cell.

The C-terminus is required for dimerisation of CPF1

Studies with previous bHLH proteins indicate that the HLH domain is required for dimerisation in solution (12, 15). We had previously shown that CPF1, purified from yeast, binds to DNA as a dimer (8). In order to localise the dimerisation domain for CPF1 and to investigate the essential function associated with the C-terminal region of the protein we produced ³⁵S-labelled protein by transcription and translation of mutant versions of *CPF1* and analysed these proteins for DNA binding and dimerisation *in vitro*. In order to demonstrate that dimerisation of *CPF1* purified from yeast does not require auxiliary factors, the original experiment (8) was repeated using the protein produced by *in vitro* transcription and translation of the *CPF1* gene and analysed by gel retardations using a ³²P labelled CDEI oligonucleotide. The result of co-translating increasing amounts of CPF1 deletion 300 (Figure 1) with decreasing amounts of full length CPF1 RNA is shown in Figure 2. A retarded band of intermediate mobility is seen, indicating the presence of heterologomers, most likely heterodimers made up of the short and long versions of the protein (lanes 2, 3 and 4). No intermediate bands are seen if short and full length CPF1 proteins are mixed after individual *in vitro* translation suggesting that specific protein-protein complexes form in solution before binding to the CDEI sequence (data not shown). These results suggest that the DNA binding characteristics of *in vitro* produced CPF1 is similar to that of the protein purified from the cell and that CPF1 dimers in solution are stable.

In order to map the dimerisation domain(s) a number of internal deletions and carboxy-terminal truncations of *CPF1* were generated by exonuclease III digestion (Figure 1). The ability of the mutant forms of the CPF1 protein to bind to DNA was assessed by gel retardation using a CDEI oligonucleotide (Figure 3A, B). Deletion of the last 9 amino acids of CPF1 (molecule 312; Figure 1) does not affect DNA binding (Figure 3A, lane 4) whereas a carboxy terminal truncation removing 37 amino acids of CPF1 (molecule 313 containing a truncation at amino acid 313; Figure 1) is sufficient to prevent the protein produced by *in vitro* translation from binding to the CDEI oligonucleotide

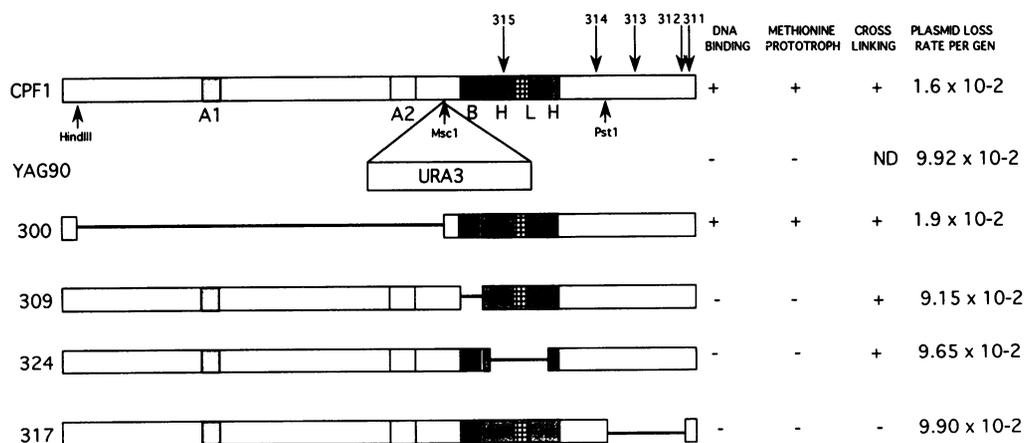


Figure 1. Internal deletions define three domains on CPF1. Schematic representation of the intact CPF1 protein and positions of internal deletions or carboxy-terminal truncations. The two acidic patches (A), the basic region (B) and the helix-loop-helix (HLH) are indicated together with key restriction sites. + indicates a wild-type phenotype; - indicates loss of phenotype; ND, not determined. Cross-linking indicates the presence or absence of a dimer after treatment with DMSI. Plasmid loss rate per generation is the average from measurements of 8 individual transformants (containing plasmid pRS315; 40) over ten generations in non-selective medium. YAG90 contains an insertion of the *URA3* gene at the *MscI* site of CPF1 (8).

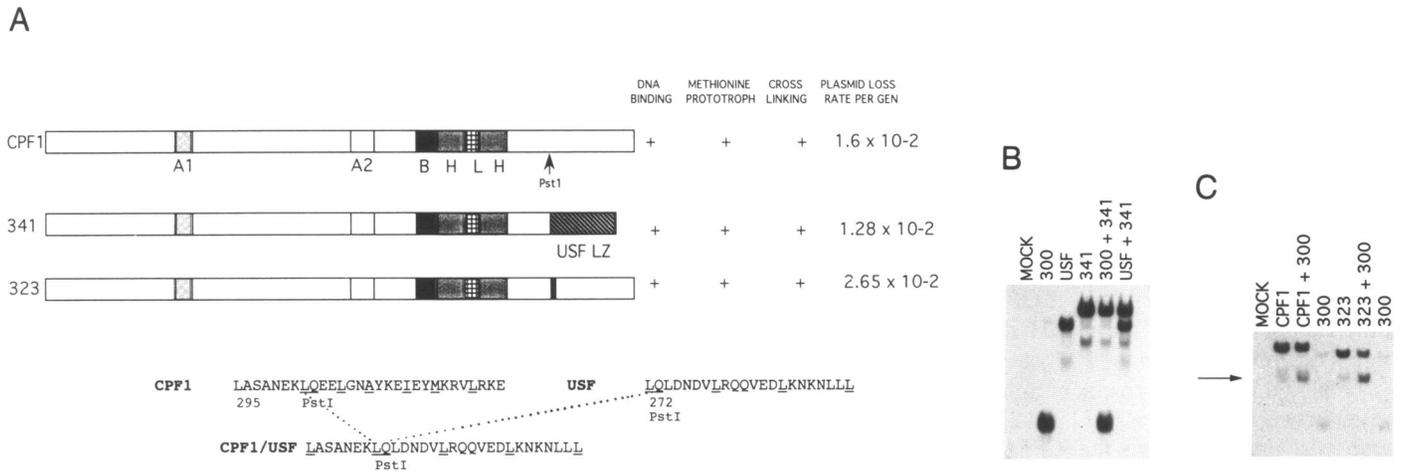


Figure 5. The dimerisation domain of USF functions in CPF1. **A.** Schematic representation of the intact CPF1 protein and derivatives manipulated at the *PstI* site. Molecule 341 contains the dimerisation domain of USF inserted at the *PstI* site and molecule 323 contains a four amino acid insertion at the *PstI* site. The two acidic patches (A), the basic region (B) and the helix–loop–helix (HLH) are indicated together with key restriction sites. + indicates a wild-type phenotype; – indicates loss of phenotype. Plasmid loss rate per generation is the average from measurements of 8 individual transformants (containing plasmid pRS315; 40) over ten generations in non-selective medium. The sequence around the junction of the fusion between CPF1 and USF used to make the hybrid protein 341 is shown. **B.** Gel retardation analysis using ³²P labelled CDEI oligonucleotide and ³⁵S labelled proteins. RNAs were produced from templates linearised with *Bam*HI. RNAs were co-translated in tracks containing two proteins. 300 is the truncated derivative of CPF1, USF is the full length USF template on T7βNco (27) and 341 is the CPF1-USF hybrid protein. **C.** Gel retardation analysis of heterodimers formed between protein 323 and 300. 2 μl of 323, 300 and CPF1 RNA were translated singly or together and the ³⁵S labelled products incubated with ³²P labelled CDEI oligonucleotide. The position of the heterodimer is indicated with an arrow.

of protein 317 lacking the C-terminal amino acids 302 to 343 of CPF1 even on a long exposure indicating that, as expected from the C-terminal truncations, this region is essential for dimerisation of CPF1 in solution (Figure 3D). We interpret the presence of this new protein species D to indicate that CPF1 exists as a dimer in solution, that dimers are required for DNA binding and that the ability to dimerise is located at the carboxyterminus of CPF1. These results indicate that the essential function associated with the C-terminal region of CPF1 in the cell is protein:protein dimerisation.

³⁵S labelled proteins from CPF1 templates containing internal deletions in the bHLH domain were subject to the same cross-linking assay (Figure 3D). Protein 309, lacking amino acids 217 to 237 including the basic domain, forms a D species similar to that seen in the wild-type protein after treatment with DMSI. This indicates that the basic region is dispensable for dimerisation in solution. We expected protein 324, lacking amino acids 243 to 260 including the HLH, not to dimerise after treatment with DMSI. However, a dimer species was evident on the gels (Figure 3D). This result indicates that CPF1 can dimerise without a complete HLH domain. However, the quality and quantity of the cross-linked product was different from that seen with the wild type protein. In contrast to the wild type protein and protein 309, not all the M form of protein 324 is converted to the slower migrating D form indicating that the CPF1 protein carrying this deletion does not dimerise as efficiently as wild-type protein or protein 309. This reduction in the amount of the D species and in its quality was observed repeatedly. It is possible that the deletion of the HLH removes essential residues for DMSI cross-linking or that the HLH deletion alters the conformation of the protein, resulting in a reduced efficiency of dimerisation. Alternatively, it is possible that the HLH plays an auxiliary but important role in dimerisation.

These results indicate that the carboxy-terminal region of CPF1 is essential for dimerisation in solution and that the HLH may

have a secondary role. We decided to investigate in more detail the role of these two regions of the CPF1 protein in DNA binding, protein dimerisation, centromere function and methionine independent growth.

The carboxy-terminal region comprising amino acids 266 to 351 of CPF1 is sufficient for mediating protein:protein interactions in vivo

As deletion of the C-terminal region of CPF1 on molecule 317 always results in a non-functional protein (Figure 1) it is difficult to assess if this region alone is sufficient to dimerise CPF1 *in vivo* or if dimerisation also requires the helix–loop–helix domain. To investigate this we used an assay system developed by Fields and Song (41) where a protein that has the ability to dimerise can be used to associate the DNA binding and transcriptional activation domains of the transcription factor G-AL4. The interaction between the GAL4 DNA binding and activation domains can then activate transcription of a reporter gene. We fused the carboxy-terminal 85 amino acids of CPF1 (amino acids 266 to 351) to the GAL4 DNA binding domain (amino acids 1–147) on plasmid p13-DD and the same 85 amino acids to the GAL4 activation domain (768–881) on plasmid pG-DD (Figure 4). A *cpf1-gal4* yeast strain carrying a β-galactosidase reporter gene, under the control of the *GAL1* promoter and integrated at the *URA3* locus, was transformed with both plasmids and expression of the hybrid proteins induced by growth on ethanol. The levels of β-galactosidase activity were induced over fifty fold compared to the same strain grown on glucose (Table 1). Expression of the individual GAL4-CPF1 fusions from plasmids p13-DD or pG-DD in strain JT1 showed only basal levels of β-galactosidase activity after growth on ethanol indicating that the induction is dependent on the presence of both fusion proteins in the same cell. Any basal level of β-galactosidase activity seen in these strains is also seen in strains containing plasmids expressing only the DNA binding or transcriptional

activation domains of GAL4 (p13H or pGAD1F). This result indicates that the carboxy-terminal 85 amino acids of CPF1 are sufficient for dimerisation of GAL4 protein domains *in vivo*.

The leucine zipper repeat of USF functions in place of the C-terminal dimerisation domain of CPF1

The results presented so far indicate that the C-terminus of CPF1 contains a domain that is essential for CPF1 function and that this region alone can mediate dimer formation *in vivo*. In order to show that this region contains an essential dimerisation domain, we substituted the C-terminus of CPF1 with the leucine zipper dimerisation domain of the bHLH-ZIP protein USF (27). The region downstream from the *Pst*I site of CPF1 at amino acid 302 was replaced with a *Pst*I fragment from USF encoding the leucine repeat between amino acids 272 and 310 (Figure 5A). The hybrid gene was integrated into the *CPF1* locus using homologous recombination and 5'FOA selection to create yeast strain YAG341. Expression of the CPF1:USF hybrid protein (molecule 341) at the *CPF1* locus results in a phenotypically wild-type strain indicating that the carboxy-terminal region of USF containing a leucine repeat can function as a dimerisation domain for CPF1 (Figure 5A). A second strain 323 was also constructed to test the effect of increasing the spacing between the C-terminal domain and the bHLH and to act as a control for the CPF1-USF hybrid protein. This molecule contains an insertion of four amino acids at the *Pst*I site at amino acid 302. Strain 323 was a methionine prototroph but showed a small but reproducible increase in the rate at which a centromere based plasmid is lost (Figure 5A).

Proteins 341 and 323 were tested for their ability to bind to the CDEI oligonucleotide as homo- or heterodimers. Protein 341 was tested for heterodimer formation with protein 300, the truncated version of CPF1, and with full length USF in a gel retardation assay (Figure 5B). No products of intermediate mobility were seen when protein 341 was co-translated with either protein 300 or USF and incubated with CDEI oligonucleotide indicating that DNA binding heterodimers are not formed (Figure 5B). These data indicate that the 341 protein may be functioning as a homodimer rather than a heterodimer in the cell since it functions as a wild-type protein in yeast, but is unable to form a heterodimer with CPF1.

In contrast, protein 323 was able to bind to DNA and form a heterodimer of intermediate mobility with the amino terminal truncated version of CPF1, protein 300 (Figure 5C). This indicates that the dimerisation domain in protein 323 is likely to be functional and that the distance from the HLH to the dimerisation domain is not critical in heterodimer formation *in vitro*.

Individual residues within the carboxy-terminal 85 amino acids of CPF1 are important for the dimerisation function

As the leucine repeat of USF is able to complement the essential function associated with the C-terminus of CPF1 we compared the sequence around the *Pst*I site of the CPF1 protein with the USF protein (Figure 5A). Structural predictions indicated a long potential alpha-helix extending from helix 2 to close to the end of the protein, covering the region defined by deletion 317 as a dimerisation domain (43, 44). When this region was plotted as a 3–4 helix it revealed a potential amphipathic arrangement with a hydrophobic interaction face and a charged face (45; Figure 6A). The hydrophobic face is clearly a good candidate for an interaction domain and shows some resemblance to the

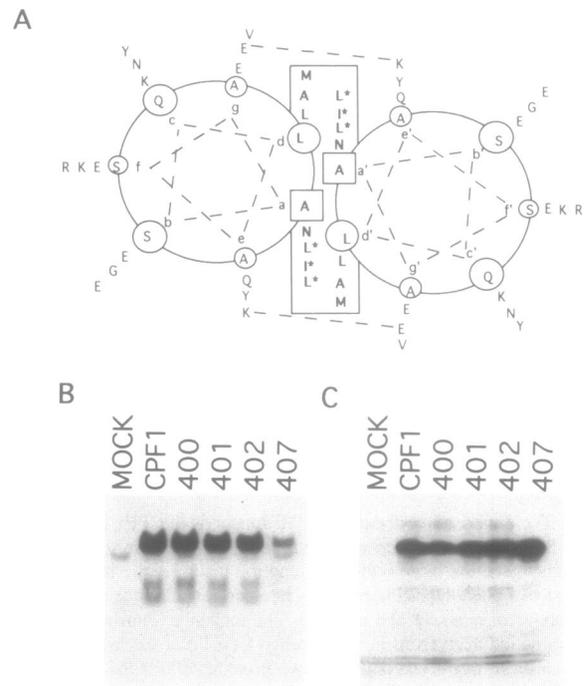


Figure 6. The dimerisation domain of CPF1 contains a hydrophobic face. **A.** A pair of 3–4 helical plots of amino acids 292 to 320 of CPF1. A potential hydrophobic interaction face is outlined and a potential salt bridge by a broken line. The residues, leucine 306, isoleucine 313 and leucine 320, mutated to methionine are marked with an asterisk. **B.** Gel retardation analysis with ^{32}P labelled CDEI oligonucleotide and ^{35}S labelled proteins containing point mutations. **C.** 10% SDS-PAGE of ^{35}S labelled proteins used for gel retardation in B. Molecules contained the following single or double point mutations. 400, I313M; 401, L320M; 402, I313M + L320M; 407, I313M + L306M.

leucine repeat (17). For some bZIP proteins such as GCN4, C/EBP or the Fos/Jun heterodimer a hydrophobic face is not sufficient to provide the dimerisation function as point mutations changing two or more leucine residues in the heptad repeat to hydrophobic residues result in loss of function (46–50). However, other proteins such as CPC1, do not require an aligned leucine heptad repeat for dimerisation; a hydrophobic face appears to be sufficient to dimerise a bZIP protein (51). We decided to investigate whether leucine residues 306 and 320 and isoleucine residue 313 are required for dimerisation and DNA binding of CPF1 or whether a hydrophobic face is sufficient (Figure 6A). Previous studies with the GCN4, Jun and c-fos leucine rich repeats had indicated that generally, one conservative change mutating leucine to another hydrophobic residue could be tolerated but more than one could not (23, 47, 49, 50). Leucine 306, isoleucine 313 and leucine 320 were mutated to methionine either singly or in pairs. No effect on DNA binding to CDEI or dimerisation as monitored by gel retardation were seen for individual mutations or a double mutation at positions 313 and 320 (Figures 6B). However, a double mutation at positions 306 and 313 resulted in a version of CPF1 that bound to DNA less avidly than wild type protein (Figures 6B). Since methionine is a conservative change, we conclude that residues 306 (leucine) and 313 (isoleucine), although not part of an obvious leucine heptad repeat, must make a significant contribution to the homodimerisation function of this domain.

DISCUSSION

In this paper we show that the region C-terminal to the bHLH domain of CPF1 contains an essential dimerisation domain. There are four pieces of evidence to support our claim. First, deletion analysis shows that the C-terminus is required for CPF1 function in the cell (Figure 1) and for dimerisation in solution and DNA binding to CDEI (Figure 3). Second, point mutations changing leucine 306 and isoleucine 313 to methionine are sufficient to markedly reduce the efficiency of DNA binding (Figure 6). Third, the C-terminus alone, when tagged to the GAL4 DNA binding and to the transcriptional activation domains is able to mediate a protein-protein interaction resulting in the activation of a reporter gene in yeast (Table 1). Fourth, the leucine repeat dimerisation domain of USF can function in place of the C-terminal domain (Figure 5). However, we have been unable to demonstrate unequivocally that this C-terminal domain is sufficient for dimerisation of CPF1 monomers. The HLH domain may be required to ensure optimal dimerisation of CPF1 but it is not essential for this function and the basic domain of CPF1 is dispensable for dimerisation. A number of point mutations in the bHLH domain of CPF1 have been made that adversely affect the ability of the protein to maintain methionine independent growth or normal centromere function but do not affect the ability of the protein to cross-link after treatment with DMSI (52). This indicates the bHLH domain of CPF1 does not play a critical role in dimerisation.

The nature of the C-terminal dimerisation domain of CPF1 is not clear. A number of proteins that belong to the HLH family contain one or more leucine heptad repeats or zippers at the end of the second helix. In most cases this domain is required for protein dimerisation and DNA binding (e.g AP4, 28; USF, 27; TFE3, 29; c-myc/max, 30). Structural predictions indicate an alpha helix running from the loop of CPF1 at amino acid 250 to close to the end of the protein at amino acid 340 (43, 44). Parts of this helix, between amino acids 256 and 270 (helix 2 of the HLH) and between amino acids 295 and 320 have an amphipathic nature with a hydrophobic face and a polar face (helix 2) or a hydrophobic face and a charged face (amino acids 295 and 320). These hydrophobic faces may provide protein-protein interaction domains in an analogous way to the leucine repeat in the bZIP and bHLH-ZIP proteins. A limited analysis by point mutation shows that introducing conservative changes at leucine 306 and isoleucine 313 is sufficient to markedly reduce DNA binding and dimerisation of CPF1. We chose to replace these residues by methionine so that we could distinguish between a contribution by specific residues and the requirement for a hydrophobic face for CPF1 interactions. However, substitution of leucine 306 and isoleucine 313 by a number of different hydrophobic residues is required to rule out any effect of substituting methionine at positions 306 and 313 on the protein structure. It is possible that helix 2 and the C-terminus of CPF1 form a continuous hydrophobic face for dimerisation. This might explain why deletion of the HLH region reduces the efficiency of dimerisation but point mutations have no effect on dimerisation.

From this work it appears that CPF1 may resemble the bHLH-ZIP proteins although there is no good homology to the leucine repeat in the C-terminal domain. However, the dimerisation domain of CPF1 appears to confer characteristics which are different from those seen in some bHLH-ZIP proteins. In contrast to AP4, where the leucine repeats are dispensable for DNA binding and instead serve strictly as a dimer interface, the

dimerisation domain of CPF1 is absolutely required for all functions of the protein, including DNA binding. In the full length USF protein, DNA binding requires the leucine repeat but if a large amino terminal deletion of the first 180 residues is made, the leucine repeat is now no longer required for DNA binding. It is likely that this truncated version of USF binds to DNA as a monomer or that the HLH, as with AP-4, is sufficient to mediate protein:protein interactions but only in the presence of DNA. Truncating CPF1 deletion 300 (lacking amino acids 10 to 209) at the *Pst*I site (amino acid 302) results in a protein that is unable to bind to CDEI indicating that a dimer may be critical for recognition of the binding site and that the HLH in the absence of the amino and carboxy-terminal regions is not sufficient to mediate dimer formation even in the presence of DNA (42). The leucine repeat in TFE3, like that of CPF1, is absolutely required for DNA binding *in vitro* and the HLH when separated from the leucine repeat is incapable of mediating dimer formation in the presence of DNA (29). However, in direct contrast to CPF1, the leucine repeat alone in TFE3 is incapable of forming stable protein:protein interactions *in vivo* or *in vitro*. TFE3 is unable to form heterodimers promiscuously. We have found no evidence that CPF1 can form heterodimers with other HLH proteins in solution or in the presence of DNA. CPF1 is unable to dimerise in the presence or absence of DNA with proteins such as PHO4, c-myc and USF, all of which bind to a common DNA sequence (CACGTG; 53-58).

The essential role of the C-terminal dimerisation domain for CPF1 function in the cell can be replaced by a region of the bHLH-ZIP protein USF containing the leucine heptad repeat, indicating that a leucine zipper can function as a dimerisation domain for CPF1. We used the CPF1-USF hybrid protein 341 to ask questions about protein-protein interactions *in vitro*. If CPF1 exists as a heterodimer with other proteins in yeast, in the way in which the bHLH proteins c-myc and max form a heterodimer, then we would expect protein 341 to form similar interactions to the wild type protein because yeast expressing the 341 protein show wild-type phenotypes. We tested for interactions between the CPF1-USF hybrid protein and the truncated version of CPF1 on protein 300 but were unable to find any evidence for a heterodimer. This may be because of incompatibility in the dimerisation domains (a leucine repeat in USF and an unknown domain in CPF1) or because the distance from helix 2 to the dimerisation domain is different in CPF1 and the CPF1-USF hybrid (29). CPF1 protein 323 which contains a four amino acid insertion, including a leucine, at the *Pst*I site at the beginning of the leucine rich hydrophobic face is able to form DNA binding heterodimers with CPF1 *in vitro*. As the insertion in protein 323 theoretically causes a rotation in the putative helix, it suggests that optimal alignment of the hydrophobic faces is not essential for dimerisation in solution. This preliminary analysis of the domains required for CPF1 function indicate that while CPF1 has the structure of a bHLH protein, it may reveal more complex interactions than expected for a protein of this class.

ACKNOWLEDGMENTS

We would like to thank Polly Gregor and Robert Roeder for the USF clone; Stanley Fields for pGAD1F and strain GGY1::171; Mark Ptashne for pKW13; Colin Goding for the PHO4 clone, Trevor Littlewood for a c-myc clone; Louise Bird and Nick Kent for reading the manuscript and for many useful discussions and Ken Johnson for photography. J.M. is a Wellcome Trust Senior

Research Fellow in Basic Biomedical Science. S.D. acknowledges the support of a SERC studentship. This work was supported by the Wellcome Trust and the SERC (grant GR/F48067).

REFERENCES

1. Panzeri, L., Landonio, L., Stotz, A. and Philippsen, P. (1985) *EMBO J.* **1**, 1605–1611.
2. Bram, R.J. and Kornberg, R.D. (1987) *Mol. Cell. Biol.* **7**, 403–409.
3. Cai, M. and Davis, R.W. (1989) *Mol. Cell. Biol.* **9**, 2544–2550.
4. Cai, M. and Davis, R.W. (1990) *Cell* **61**, 437–446.
5. Baker, R.E., Fitzgerald-Hayes, M. and O'Brien, T.C. (1989) *J. Biol. Chem.* **264**, 10843–10850.
6. Jiang, W. and Philippsen, P. (1989) *Mol. Cell. Biol.* **9**, 5585–5593.
7. Baker, R.E. and Masison, D.C. (1990) *Mol. Cell. Biol.* **10**, 2458–2467.
8. Mellor, J., Jiang, W., Funk, M., Rathjen, J., Barnes, C.A., Hinz, T., Hegemann, J.H. and Philippsen, P. (1990) *EMBO J.* **9**, 4017–4026.
9. Murre, C., McCaw, P.S. and Baltimore, D. (1989) *Cell* **56**, 777–783.
10. Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lasser, A.B., Weintraub, H. and Baltimore, D. (1989) *Cell* **58**, 537–544.
11. Ludwig, S.R., Habera, L.F., Dellaporta, S.L. and Wessler, S.R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7092–7096.
12. Davis, R.L., Cheng, P.-F., Lasar, A.B. and Weintraub, H. (1990) *Cell* **60**, 733–746.
13. Jones, N. (1990) *Cell* **61**, 9–11.
14. Garrell, J. and Campuzano, S. (1991) *BioEssays* **13**, 493–498.
15. Voronova, A. and Baltimore, D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4722–4726.
16. Hope, I.A. and Struhl, K. (1987) *EMBO J.* **6**, 2781–2784.
17. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) *Science* **240**, 1759–1764.
18. O'Shea, E.T., Rutkowski, R. and Kim, P. (1989) *Science* **243**, 538–542.
19. Rasmussen, R., Benvengnu, D., O'Shea, E.K., Kim, P.S. and Alber, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 561–564.
20. Busch, S.J. and Sassone-Corsi, P. (1990) *Trends Genet.* **6**, 36–40.
21. Johnson, P.F. and McKnight, S.L. (1989) *Annu. Rev. Biochem.* **58**, 799–839.
22. Lamb, P. and McKnight, S.L. (1991) *Trends Biochem. Sci.* **16**, 417–422.
23. Hu, J.C., O'Shea, E.K., Kim, P.S. and Sauer, R.T. (1990a) *Science* **250**, 1400–1403.
24. O'Shea, E.T., Klemm, J.D., Kim, P.S. and Alber, T. (1991) *Science* **254**, 539–544.
25. Schuermann, M., Hunter, J., Hennig, G. and Müller, R. (1991) *Nucleic Acids Res.* **19**, 739–746.
26. Dang, C.V., McGuire, M., Buckmire, M. and Lee, W.M.F. (1989) *Nature* **337**, 664–666.
27. Gregor, P.D., Sawadogo, M. and Roeder, R.G. (1990) *Genes Dev.* **4**, 1730–1740.
28. Hu, Y.-F., Luescher, B., Admon, A., Mermod, N. and Tjian, R. (1990b) *Genes Dev.* **4**, 1741–1752.
29. Beckman, H. and Kadesch, T. (1991) *Genes Dev.* **5**, 1057–1066.
30. Blackwood, E.M. and Eisenman, R.N. (1991) *Science* **251**, 1211–1217.
31. Thomas, D., Cherest, H. and Surdin-Kerjan, Y. (1989) *Mol. Cell. Biol.* **9**, 3292–3298.
32. Ma, J. and Ptashne, M. (1987) *Cell* **48**, 847–853.
33. Giniger, E. and Ptashne, M. (1987) *Nature* **330**, 670–672.
34. Hope, I.A., Mahadevan, S. and Struhl, K. (1988) *Nature* **333**, 635–637.
35. Thomas, D., Jacquemin, I. and Surdin-Kerjan, Y. (1992) *Mol. Cell. Biol.* **12**, 1719–1727.
36. Mellor, J., Rathjen, J., Jiang, W., Barnes, C. and Dowell, S.D. (1991) *Nucleic Acids Res.* **19**, 2961–2969.
37. Gill, G. and Ptashne, M. (1987) *Cell* **51**, 121–126.
38. Rothstein, R.J. (1983) *Methods Enzymol.* **101**, 202–211.
39. Boeke, J.D., Lacroute, F. and Fink, G.R. (1984) *Mol. Gen. Genet.* **197**, 345–346.
40. Sikorski, R.S. and Heiter, P. (1989) *Genetics* **122**, 19–27.
41. Fields, S. and Song, O.-K. (1989) *Nature* **340**, 245–246.
42. Dowell, S.J. (1992) D.Phil Thesis, University of Oxford.
43. Chou, P.Y. and Fasman, G.D. (1974) *Biochemistry* **13**, 211–222.
44. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
45. Cohen, C. and Parry, D.A.D. (1990) *Proteins* **7**, 1–15.
46. Kouzarides, T. and Ziff, E. (1988) *Nature* **336**, 646–651.
47. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1989) *Science* **243**, 1681–1688.
48. Neuberger, M., Schuermann, M., Hunter, J.B. and Muller, R. (1989) *Nature* **338**, 589–590.
49. Ransone, L.J., Visvader, J., Sassone-Corsi, P. and Verma, I.M. (1989) *Genes Dev.* **3**, 770–781.
50. Smeal, T., Angel, P., Meek, J. and Karin, M. (1989) *Genes Dev.* **3**, 2091–2100.
51. Paluh, J. and Yanofsky, C. (1991) *Mol. Cell. Biol.* **11**, 935–944.
52. McKenzie, E.A., Moreno, F. and Mellor, J. Unpublished observations.
53. Sawadogo, M. and Roeder, R.G. (1985) *Cell* **43**, 165–175.
54. Carthew, R.W., Chodosh, L.A. and Sharp, P.A. (1985) *Cell* **43**, 439–448.
55. Chodosh, L.A., Buratowski, S. and Sharp, P.A. (1989) *Mol. Cell. Biol.* **9**, 820–822.
56. Ogawa, N. and Oshima, Y. (1990) *Mol. Cell. Biol.* **10**, 2224–2236.
57. Prendergast, G.C. and Ziff, E.B. (1991) *Science* **251**, 186–189.
58. Mellor, J. Unpublished data.